UNIVERSIDAD DE LOS LAGOS

VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
DIRECCIÓN DE POSTGRADO
DEPARTAMENTO DE CIENCIAS EXACTAS
DOCTORADO EN EDUCACIÓN MATEMÁTICA

NIVELES DE RAZONAMIENTO INFERENCIAL SOBRE LOS ESTADÍSTICOS CHI-CUADRADA Y T-STUDENT

Tesis doctoral

JESÚS GUADALUPE LUGO ARMENTA

Director: Dr. Luis R. Pino-Fan

Codirector: Dra. Blanca R. Ruiz H.

Osorno, Chile, 2021

UNIVERSIDAD DE LOS LAGOS

VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
DIRECCIÓN DE POSTGRADO
DEPARTAMENTO DE CIENCIAS EXACTAS
DOCTORADO EN EDUCACIÓN MATEMÁTICA

NIVELES DE RAZONAMIENTO INFERENCIAL SOBRE LOS ESTADÍSTICOS CHI-CUADRADA Y T-STUDENT

Tesis Doctoral presentada por Jesús Guadalupe Lugo Armenta dentro del programa de Doctorado en Educación Matemática para aspirar al grado de Doctor en Educación Matemática por la Universidad de Los Lagos, dirigida por el Dr. Luis Roberto Pino Fan, académico de la Universidad de Los Lagos; y codirigida por la Dra. Blanca Rosa Ruiz Hernández, académica del Tecnológico de Monterrey.

Jesús Gpe. Lugo A.

Jesús Guadalupe Lugo Armenta

Dr. Luis R. Pino-Fan

Dra. Blanca R. Ruiz Hernández

UNIVERSIDAD DE LOS LAGOS

VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
DIRECCIÓN DE POSTGRADO
DEPARTAMENTO DE CIENCIAS EXACTAS
DOCTORADO EN EDUCACIÓN MATEMÁTICA

NIVELES DE RAZONAMIENTO INFERENCIAL SOBRE LOS ESTADÍSTICOS CHI-CUADRADA Y T-STUDENT

Esta Tesis de Doctorado ha sido desarrollada al seno del Grupo de Investigación sobre Didáctica de la Matemática (GIDMAT-ULAGOS) del Departamento de Ciencias Exactas de la Universidad de Los Lagos, en el marco del Proyecto de Investigación Fondecyt 1200005 titulado “Desarrollo de competencias profesionales clave para la práctica pedagógica de profesores de matemáticas de enseñanza media”, el cual es dirigido por el Dr. Luis R. Pino-Fan; y se inscribe en las líneas de investigación ‘Didáctica de los Diversos Marcos Matemáticos' e 'Historia, epistemología y aspectos socioculturales de las matemáticas’.

A mis padres,
Ma. Luisa Armenta López y
Rubén A. Lugo Velázquez.
A mi hijo,
Emmanuel Cruz Lugo.

AGRADECIMIENTOS

A Dios, por brindarme la oportunidad de vivir la maravillosa experiencia que ha sido estudiar el doctorado y realizar esta investigación.

A mis padres, Rubén y Ma. Luisa, quienes son mi fuente de inspiración y que siempre me han apoyado y alentado a perseverar por mis sueños. A mis hermanos, Omar y Javier, por su apoyo incondicional en las diversas etapas de mis estudios. A mi hijo, Emmanuel, por acompañarme, ser mi motor de vida y por su inmenso amor que ilumina mis días.

A mi director, el Dr. Luis R. Pino-Fan, por haber aceptado dirigir esta tesis y ser mi mentor, por sus orientaciones y apoyo, por compartir su experiencia y conocimiento en las sesiones de discusión académica y hacer posible esta investigación. Del mismo modo quiero agradecer a la Dra. Blanca Ruiz, mi codirectora de tesis, por su orientación, disponibilidad y valiosos consejos. Gracias por todas las experiencias que aportaron para desarrollar esta tesis.

Al Dr. Juan Díaz Godino, por permitir la realización de mi estancia de investigación en el Departamento de Didáctica de la Matemática de la Universidad de Granada, España, por su hospitalidad durante la misma, pero sobre todo por compartir su experiencia y conocimientos lo cual contribuyó significativamente al desarrollo de esta investigación. Asimismo, a la Dra. Carmen Batanero, por las conversaciones y valiosas orientaciones que aportaron considerablemente en mi formación.

A todos los profesores del Programa de Doctorado en Educación Matemática de la Universidad de Los Lagos, quienes fueron parte de mi proceso formativo; especialmente a la Dra. Ismenia Guzman, por su confianza, amabilidad y por compartir conmigo sus experiencias y consejos académicos y de vida.

A mis amigas, Edna Granillo y Guadalupe Morales, por su sincera amistad, conversaciones interminables y de alguna forma estar a mi lado. Edna, gracias por tus palabras de aliento y por levantar mi ánimo cuando más lo necesité; Guadalupe (Lupita), gracias por tantas conversaciones enriquecedoras, grandes consejos e ideas.

A mis compañeros del doctorado, con quienes compartí cursos, experiencias y momentos que me ayudaron a crecer en lo personal y profesional.

A la Universidad de Los Lagos, institución que me otorgó una beca de exclusividad e hizo posible la realización de mis estudios de doctorado en Chile.

Finalmente, agradezco a todas las personas que, de alguna forma, estuvieron conmigo, apoyándome en este proceso.

RESUMEN

El estudio sobre cómo promover el razonamiento inferencial formal (RIF) a partir de un razonamiento inferencial informal (RII) ha ido tomando un creciente interés en los últimos años. Sin embargo, aun es necesario contar con propuestas que permitan explorar y desarrollar progresivamente (del RII al RIF) el razonamiento inferencial de estudiantes y profesores. En esta investigación caracterizamos niveles progresivos de razonamiento inferencial (de lo informal a lo formal) sobre los estadísticos Chi-cuadrada y t-Student, a partir de la riqueza matemática subyacente a los significados parciales que se le han conferido a tales estadísticos en su evolución histórico-epistemológica y de las sugerencias de la investigación desarrollada sobre razonamiento inferencial.

Para lograr el objetivo general de la investigación, el estudio se llevó a cabo en cuatro fases: (1) mediante un estudio histórico-epistemológico se determinan los significados parciales y el significado holístico de los estadísticos Chi-cuadrada y t-Student; (2) vinculación de los significados parciales de los estadísticos Chi-cuadrada y t-Student con la revisión de literatura científica sobre el razonamiento inferencial informal y formal y estos estadísticos; (3) propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, tomando como base los resultados del estudio de los significados de estos estadísticos y los aportes de la literatura científica; (4) análisis de las prácticas matemáticas realizadas por profesores en formación y profesores en ejercicio de matemáticas de educación media al resolver actividades que involucran a los dichos estadísticos.

Como resultados de nuestra investigación, se logra una *propuesta de niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student*. A partir de estas propuestas teóricas de niveles de razonamiento y de la caracterización de las practicas desarrolladas por los profesores en ejercicio y en formación, fue posible realizar la *propuesta general de niveles de razonamiento inferencial*. Los niveles de razonamiento inferencial podrían constituir directrices para la gestión del aprendizaje de los estudiantes, el diseño de actividades y el análisis del currículo.

ABSTRACT

The study on how to promote formal inferential reasoning (FIR) from an informal inferential reasoning (IIR) has been gaining a growing interest in recent years. However, it is still necessary to have proposals that allow to explore and progressively develop (from IIR to FIR) the inferential reasoning of students and teachers. In this research we characterize progressive levels of inferential reasoning (from informal to formal) about the Chi-square and t-Student statistics, based on the mathematical richness underlying the partial meanings that have been conferred on such statistics in their historical-epistemological evolution and on the suggestions of the investigation developed on inferential reasoning.

To achieve the general objective of the research, the study was carried out in four phases: (1) a historical-epistemological study determines the partial meanings and the holistic meaning of Chi-square and t-Student statistics; (2) linking the partial meanings of the Chi-square and t-Student statistics with the review of scientific literature on informal and formal inferential reasoning and these statistics; (3) proposal of progressive levels, from informal to formal, of inferential reasoning on Chi-square and t-Student statistics, based on the results of the study of the meanings of these statistics and the contributions of the scientific literature; (4) Analysis of the mathematical practices performed by teachers in prospective and practicing mathematics teachers of high school when solving activities that involve the Chi-square and t-Student statistics.

As results of our research, a proposal of inferential reasoning levels of the Chi-square and t-Student statistics is achieved. Based on these theoretical proposals for reasoning levels and the characterization of the practices developed by practicing and prospective teachers, it was possible to carry out the general proposal of inferential reasoning levels. Inferential reasoning levels could constitute guidelines for student learning management, activities design, and curriculum analysis.

TABLA DE CONTENIDO

CAPÍTULO 3

EL ESTADÍSTICO CHI-CUADRADA (χ^2).
RECONSTRUYENDO EL SIGNIFICADO HOLÍSTICO DE REFERENCIA

CAPÍTULO 4

EL ESTADÍSTICO T-STUDENT.
RECONSTRUYENDO EL SIGNIFICADO HOLÍSTICO DE REFERENCIA

CAPÍTULO 5

PROPUESTA DE NIVELES PARA LOS ESTADÍSTICOS χ^2 Y -STUDENT

CAPÍTULO 6

NIVELES DE RAZONAMIENTO INFERENCIAL SOBRE LOS ESTADÍSTICOS CHI-CUADRADA Y T-STUDENT DE PROFESORES DE ENSEÑANZA MEDIA

CAPÍTULO 7

CONCLUSIONES

REFERENCIAS 403

ANEXOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

En las últimas dos décadas, ha cobrado especial atención la formación de ciudadanos que posean la capacidad de interpretar el mundo en el que se desenvuelven, es por ello que la estadística se ha ido incorporando en los currículos de matemática de diversos países. En particular, en los currículos de enseñanza media recientemente existen tópicos de inferencia, tópicos que anteriormente se habían privilegiado para los cursos universitarios. Sin embargo, diversas investigaciones han reportado que las nociones involucradas en inferencia estadística suelen ser complejas de comprender para los estudiantes y profesores, estas investigaciones destacan que al trabajar inferencia se presentan errores y dificultades frecuentemente en la comprensión del nivel de significancia, lógica de las pruebas de hipótesis, planteamiento de las hipótesis nula y alternativa, distribuciones muestrales, valor-p, estadístico-parámetro, criterio para la toma de decisión y los errores tipo I y tipo II (e.g., Watson, 2004; Bakker y Gravemeijer, 2004; Carver, 2006; Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Garfield y Ben-Zvi, 2008; Harradine, Batanero y Rossman, 2011; Batanero, Vera y Díaz, 2012).

Para afrontar estas dificultades, las investigaciones han propuesto, por una parte, aproximarnos desde una perspectiva informal, denominada razonamiento inferencial informal (RII), desde edades tempranas (e.g., Zieffler, delMas, Garfield & Reading, 2008; Makar y Rubin, 2009); y por otra parte, promover un razonamiento inferencial formal sobre las bases del RII (e.g., Pfannkuch, Arnold & Wild, 2015; Makar & Rubin, 2018). En otras palabras, introducir la inferencia por etapas.

Como resultado de dichas investigaciones se han realizado propuestas de marcos de trabajo para desarrollar y caracterizar el razonamiento inferencial informal de estudiantes, así como propuestas de actividades; también, han indicado la necesidad de promover progresivamente un razonamiento inferencial formal y propuesto tipos de actividades y nociones que podrían ayudar a promover el razonamiento inferencial formal desde el informal. No obstante, a pesar de los avances sobre cómo promover el razonamiento inferencial informal desde edades tempranas y sobre cómo promover progresivamente el razonamiento inferencial formal, aún está abierta la discusión sobre cómo podemos promover progresivamente el razonamiento inferencial formal sobre las bases del razonamiento inferencial informal.

En este sentido, el interés de nuestra investigación se centra en realizar una propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal. Consideramos que en la búsqueda de la progresión del RII al RIF, los estadísticos Chi-cuadrada y t-Student nos permitirán brindar una trayectoria desde lo intuitivo a lo formal. Estos estadísticos ocupan un lugar importante dentro de las pruebas no paramétricas y paramétricas, respectivamente; esto se debe a que, resultaron clave en el desarrollo histórico de la inferencia y en las aplicaciones que tienen hoy en día.

Así, en el presente trabajo de investigación nos cuestionamos ¿cómo los significados parciales de los estadísticos Chi-cuadrada y t-Student pueden contribuir en la construcción de una propuesta de niveles progresivos de razonamiento inferencial sobre dichos estadísticos?, entonces, el objetivo de esta investigación consiste en establecer vínculos o conexiones entre los significados de los estadísticos Chi-cuadrada y t-Student y la literatura científica para proponer niveles de razonamiento inferencial, de lo informal a lo formal, sobre estos estadísticos.

La investigación que a continuación se presenta se encuentra estructurada en siete capítulos, a través de los cuales se consigue gradualmente el objetivo de esta investigación. En el Capítulo 1, antecedentes, realizamos un recorrido por las investigaciones sobre razonamiento inferencial informal, razonamiento inferencial formal, transición de lo informal a lo formal y sobre las investigaciones vinculadas con los errores y dificultades que presentan estudiantes y profesores al trabajar inferencia estadística y en especial los estadísticos Chi-cuadrada y t-Student. También, en el primer capítulo nos aproximamos a la problemática de investigación planteando preguntas de investigación iniciales.

En el Capítulo 2, abordamos las nociones y herramientas teóricas del Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos –EOS– (Godino y Batanero, 1994; Godino y Batanero, 1998; Godino, Batanero y Font, 2007; 2019) que utilizamos a lo largo de nuestra investigación y que le dan soporte. Una vez presentadas las nociones teóricas planteamos la pregunta y los objetivos de investigación que orientan el estudio y, finalmente, como parte de la metodología describimos las fases de la investigación, los sujetos de estudio y el instrumento de indagación.

Los Capítulos 3 y 4, tratan sobre los estudios históricos-epistemológicos de los estadísticos Chi-cuadrada y t-Student, respectivamente, a través de los cuales identificamos las grandes problemáticas donde surgieron y evolucionaron estos estadísticos; realizamos una descripción de las prácticas matemáticas que se desarrollaron en la historia para resolver estas grandes problemáticas, las cuales dan paso a los significados. A partir de dichas prácticas matemáticas se identificaron doce significados parciales para el estadístico Chi-cuadrada y doce significados parciales para el estadístico t-Student, los cuales constituyen el significado holístico para cada uno de estos estadísticos. Para realizar los estudios históricos-epistemológicos recurrimos a las nociones de práctica y configuración ontosemiótica epistémica del EOS.

En el Capítulo 5, presentamos una propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. Las propuestas de niveles de razonamiento inferencial están conformadas por cuatro niveles que van de lo informal a lo formal y en cada nivel se determinaron diversos indicadores vinculando criterios epistémicos identificados con los estudios de tipo histórico-epistemológico sobre estos estadísticos y los aportes de la literatura científica desarrollada sobre razonamiento inferencial.

En el Capítulo 6, mostramos con detalle el instrumento de indagación, para el diseño de las actividades que conforman el instrumento se consideraron los resultados de los Capítulos 3, 4 y 5, esto debido a que buscamos que las actividades tuvieran representatividad de los niveles de razonamiento inferencial, y representatividad de los significados de los estadísticos Chi-cuadrada y t-Student. Asimismo, presentamos la caracterización del razonamiento inferencial que evidenciaron profesores cuando realizan prácticas para resolver problemas sobre los estadísticos Chi-cuadrada y t-Student. Esta caracterización tiene doble finalidad, por una parte, estudiar los elementos primarios que utilizan en sus prácticas los profesores y si dichos elementos permiten asociar las prácticas a los distintos niveles de razonamiento inferencial; y por otra parte, explorar empíricamente si los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, propuestos a nivel teórico, son predictores del nivel de razonamiento inferencial asociado a las prácticas que desarrollan los profesores al resolver problemas sobre estos estadísticos. Además, a partir de los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, propuestos a nivel teórico, y de la caracterización de las prácticas de los profesores, formulamos una propuesta general de niveles de razonamiento inferencial.

Finalmente, en el Capítulo 7, presentamos las conclusiones de esta investigación; para ello presentamos los objetivos específicos y cómo es que se ha dado cumplimiento a cada uno de ellos. También destacamos las principales contribuciones de esta investigación y posibles líneas de continuidad para futuros estudios.

CAPÍTULO 1

Antecedentes y área problemática

"La literatura es siempre una expedición a la verdad". (Franz Kafka)

INTRODUCCIÓN

En este capítulo presentamos un panorama general sobre los estudios realizados acerca del razonamiento inferencial, tanto informal, formal como de la transición de lo informal a lo formal, en el campo de la Educación Matemática. En estos estudios, a los cuales hacemos referencia en este Capítulo, se encuentran desarrollos que nos ayudan a situar nuestro problema de investigación, el estado actual de la literatura de investigación y, a orientar y brindar directrices para el desarrollo de la investigación.

Los antecedentes se han dividido por secciones, en la primera se encuentran aquellos estudios que abordan nociones que se consideran fundamentales para el desarrollo de razonamiento inferencial, así como los que generan propuestas de niveles sobre este razonamiento. En la segunda sección, aquellos estudios vinculados con los errores y dificultades que presentan estudiantes y profesores al trabajar inferencia estadística. Mientras que, en la tercera sección, podemos encontrar la aproximación al problema de investigación.

1.1 La importancia del razonamiento inferencial para el estudio de la estadística

Recientemente, en Educación Estadística ha cobrado especial atención la formación de ciudadanos que posean la capacidad de interpretar el mundo que los rodea y hemos observado cómo se produce una gran cantidad de datos provenientes de diferentes actividades que, además, es común verlos en los medios de comunicación. Llama especialmente la atención, aquellos informes estadísticos que provienen de encuestas que forman opinión o que ayudan en la toma de decisiones. Sin embargo, comprender este tipo de reportes estadísticos no es una tarea sencilla para el ciudadano, debido a que se requiere disponer de un razonamiento que le permita valorar la generación, el análisis y la interpretación de la información, así como examinar objetiva y críticamente las inferencias, producto de estos procesos estadísticos. Es por ello, que en los currículos de enseñanza básica y media de diversos países han incorporado tópicos de Probabilidad y Estadística, destacamos especialmente la incorporación de tópicos de inferencia estadística en los currículos de enseñanza media, tal es el caso del currículo chileno (Mineduc, 2019), tópicos que anteriormente se habían privilegiado para los cursos universitarios. Sin embargo, nos preguntamos si los estudiantes, y sobre todo los profesores, se encuentran preparados para afrontar este reto.

Sobre esta incorporación de la inferencia en los currículos, Callingham y Watson (2017) y Pfannkuch (2018), realizaron estudios donde han evidenciado que existen rasgos de la inferencia estadística en años previos a la educación universitaria. En términos generales, Moore (2000) describe que la inferencia estadística "*va más allá de los datos disponibles y obtiene conclusiones sobre un universo más amplio, teniendo en cuenta la omnipresencia de la variabilidad y la incertidumbre de las conclusiones*" (p. XXXIV).

En diversas investigaciones se han identificado dificultades que presentan tanto estudiantes como profesores al trabajar inferencia estadística, por ejemplo, sobre la comprensión del nivel de significancia, los errores tipo I y tipo II, el planteamiento de las hipótesis estadísticas y las distribuciones muestrales (e.g., Watson, 2004; Bakker y Gravemeijer, 2004; Carver, 2006; Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Garfield y Ben-Zvi, 2008; Harradine, Batanero y Rossman, 2011; Batanero, Vera y Díaz, 2012). Además, se suele comprender erróneamente la lógica de las pruebas de hipótesis, la relación entre el estadístico y el parámetro, y realizar interpretaciones incorrectas de los resultados; y, de acuerdo con Batanero (2013), la mayoría de las interpretaciones incorrectas y de errores en inferencia estadística se relacionan con el contraste de hipótesis. Aunado a lo anterior, Pfannkuch (2005) señala que la falta de proximidad de los estudiantes con eventos de estocástica también dificulta el desarrollo del razonamiento inferencial. En la sección 1.2.1 describiremos con detalle las dificultades.

Estas dificultades pueden llevar al estudiante a enfocarse más en los algoritmos que en comprender realmente las nociones involucradas para llevar a cabo la inferencia. De acuerdo con Rossman (2008), además de la conclusión, la inferencia debe incluir la evidencia y el razonamiento sobre los cuales se realiza; es decir, los desarrollos algorítmicos no son suficientes para realizar inferencias.

Para afrontar estas dificultades, han surgido propuestas sobre cómo promover el razonamiento inferencial (RI); por una parte, se encuentran aquellas propuestas que se enfocan en cómo aproximarnos a la inferencia desde edades tempranas bajo una perspectiva informal, denominada razonamiento inferencial informal -RII- (e.g., Zieffler, delMas, Garfield y Reading, 2008; Makar y Rubin, 2009; Doerr, delMas y Makar, 2017), y, por otra parte, para promover un razonamiento inferencial formal -RIF- sobre las bases del RII (e.g., Jacob y Doerr, 2014; Pfannkuch, Arnold y Wild, 2015; Makar y Rubin, 2018).

El desarrollo de investigaciones sobre el razonamiento estadístico, RII y RIF es debido a la importancia que actualmente se les otorga, junto con el proceso estadístico, en la literatura de educación estadística. Burrill y Camden (2005, p. 4) sugieren que "*los estudiantes parecen estar dominando procedimientos y vocabulario estadístico pero no son capaces de utilizar el razonamiento estadístico en una forma significativa*" y que hay "*un énfasis excesivo en los programas de estudios escolares en responder a las preguntas en lugar de formularlas y tomando decisiones basadas únicamente en lo datos desplegados lo cual produce un enfoque basado en el totalitarismo de los datos que limita el desarrollo del pensamiento estadístico*", este énfasis puede obstaculizar ver tanto el antes como el después de los datos y su tratamiento.

Entendemos que aquí tiene injerencia el razonamiento estadístico pues "*se espera que los estudiantes vayan más allá de los procedimientos estadísticos para evaluar o reflexionar sobre estos procedimientos*" (Newton, Dietiker y Horvath, 2011, p. 11) y más aun teniendo presente que "*el razonamiento estadístico permite tomar decisiones adecuadas o efectuar predicciones a partir de datos y en presencia de la incertidumbre*" (Batanero, Díaz, Contreras y Roa, 2013, p. 8).

A continuación, presentamos aportaciones que se han realizado en diversas investigaciones sobre el RII y el RIF, abordamos las nociones fundamentales para estos razonamientos y las propuestas de niveles que han surgido y que permiten caracterizar el RII en los estudiantes. Además, presentamos diversos posicionamientos y usos sobre el razonamiento, la alfabetización y el pensamiento estadístico que encontramos en investigaciones sobre educación estadística.

1.1.1 Alfabetización, Razonamiento y Pensamiento Estadístico

Los términos de alfabetización estadística, razonamiento estadístico y pensamiento estadístico han ido cobrando relevancia en las investigaciones de educación estadística, sin embargo, no existe un consenso en cuanto a las definiciones o usos de alfabetización, razonamiento y pensamiento. A continuación, describimos diversas posturas.

1.1.1.1 Alfabetización Estadística

La presencia que actualmente tiene la estadística en los currículos de enseñanza básica y media de diversos países, se debe a la necesidad proclamada por instituciones como la UNESCO, de proporcionar una alfabetización estadística que permita a los ciudadanos participar de forma informada en la sociedad actual. De acuerdo con Ido Gal, la alfabetización estadística se caracteriza por dos rasgos distintivos que son el contexto social y el uso individual; así pues, la alfabetización estadística incluye:

> a) capacidad para interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos, y b) capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante (Gal, 2002, pp. 2-3).

Por su parte, Watson (2006) propusó una jerarquía de niveles de alfabetización estadística:

- El desarrollo del conocimiento básico de los conceptos estadísticos y probabilísticos.
- La comprensión de los razonamientos y argumentos estadísticos cuando se presentan dentro de un contexto más amplio de algún informe en los medios de comunicación o en el trabajo.
- Una actitud crítica que se asume al cuestionar argumentos que estén basados en evidencia estadística.

Schield (2010, p. 140), define a la alfabetización estadística como "*la habilidad de leer e interpretar resúmenes estadísticos en los medios cotidianos: en gráficos, tablas, afirmaciones y ensayos. La alfabetización estadística es la necesaria para los consumidores de datos*". La noción de competencia estadística es necesaria para los consumidores de datos, este autor también define la competencia estadística como "*la habilidad de producir, analizar y resumir estadísticas detalladas en estadísticas y estudios*".

Por otra parte, Batanero (2004) utilizó el término de cultura estadística y señaló que "*quiere resaltar el hecho de que la estadística se considera hoy en día como parte de la herencia cultural necesaria para el ciudadano educado*" (Batanero, 2004, p. 28).

Como se puede observar existen diversas definiciones de alfabetización estadística, Ben-Zvi y Garfield (2004) realizaron una revisión de la literatura y a partir de ella proponen la siguiente definición:

> La alfabetización estadística incluye habilidades básicas e importantes que pueden usarse para comprender la información estadística o los resultados de la investigación. Estas habilidades incluyen ser capaz de organizar datos, construir y mostrar tablas y trabajar con diferentes representaciones de datos. La alfabetización estadística también incluye una comprensión de conceptos, vocabulario y símbolos, e incluye una comprensión de la probabilidad como medida de incertidumbre. (Ben-Zvi y Garfield, 2004, p. 7)

1.1.1.2 Razonamiento Estadístico

En los planes y programas de estudio de Estadística o del eje temático 'Datos y azar', podemos observar diversos tipos de razonamiento que se consideran importante que los estudiantes desarrollen. Por ejemplo, el razonamiento sobre muestras, el razonamiento sobre datos (Garfield, 2002), el razonamiento sobre asociación (Garfield, 2003).

De acuerdo con Chance (2002), a diferencia del razonamiento matemático, la investigación en estadística depende de los datos y, usualmente, se basa en un determinado contexto. Es importante resaltar que el razonamiento y el pensamiento estadístico suelen ser utilizados indistintamente para describir al mismo tipo de habilidades, a continuación, describiremos algunas posturas que se tienen en la literatura de educación estadística sobre el razonamiento estadístico.

Según delMas (2004), se puede decir que el pensamiento estadístico y el razonamiento están involucrados cuando se trabaja en la misma tarea, es por ello que, los dos tipos de actividad cognitiva no necesariamente pueden distinguirse por el contenido de un problema; sin embargo, se les puede distinguir por la naturaleza de la tarea. De acuerdo con la postura de delMas se puede explicar por qué es apropiado seleccionar un modelo o representación en particular, o se ha producido un resultado, explicaciones que requiere comprensión de los procesos que producen tales datos.

Por otra parte, Lovett (2001, p. 351) definió el razonamiento estadístico como "*el uso de herramientas y conceptos (e.g., prueba de hipótesis, variación, correlación) para resumir, hacer predicciones y sacar conclusiones de los datos*".

Jones, Langrall, Mooney y Thornton (2004) propusieron un modelo cognitivo de desarrollo en el razonamiento estadístico que refiere a una teoría de diferentes niveles de crecimiento en el razonamiento estadístico. Estos niveles son resultado de efectos de maduración o interacción en entornos de aprendizaje estructurados y no estructurados.

La postura que presentó Ben-Zvi (2004), refiere al desarrollo del razonamiento estadístico mediante entornos de aprendizaje donde los estudiantes primero adquieren habilidades de una visión global de datos y sus representaciones y posteriormente llegan a comprender el análisis exploratorio de datos (EDA).

Ben-Zvi y Garfield (2004, p. 7) defininieron el razonamiento estadístico realizando las distinciones con respecto al pensamiento estadístico.

> El razonamiento estadístico puede definirse como la forma en que las personas razonan con ideas estadísticas y le dan sentido a la información estadística. Esto implica realizar interpretaciones basadas en conjuntos de datos, representaciones de datos o resúmenes estadísticos de datos. El razonamiento estadístico puede involucrar la conexión de un concepto con otro (por ejemplo, centro y extensión), o puede combinar ideas sobre datos y azar. Razonar significa comprender y ser capaz de explicar los procesos estadísticos y ser capaz de interpretar plenamente los resultados estadísticos.

1.1.1.3 Pensamiento Estadístico

De acuerdo con Batanero y Borovcnik (2016), el pensamiento es algo que hacemos en todo momento y se ve influenciado por nuestras experiencias y conocimientos que hemos adquirido a lo largo de la vida. Por lo tanto, el pensamiento estadístico es un tipo pensamiento que se encuentra influenciado por las experiencias y conocimientos estadísticos que tenemos.

Watson (1997) abordó la necesidad de evaluar el pensamiento estadístico y sugirió una jerarquía para el pensamiento estadístico, esta jerarquía se compone por: (1) comprensión básica de la terminología probabilística y estadística; (2) integración del lenguaje probabilístico y estadístico y los conceptos en un contexto más amplio; y (3) cuestionamiento de reclamaciones, en este nivel los estudiantes poseen la confianza para desafiar lo que leen en los medios.

Mientras que Mooney (2002) caracterizó el pensamiento estadístico de los estudiantes de educación media de acuerdo con los siguientes cuatro procesos: describir datos, organizar y reducir datos, representar datos y analizar e interpretar datos.

Por otra parte, Wild y Pfannkuch (1999) propusieron un marco de trabajo sobre el pensamiento estadístico que ha impactado en diversas investigaciones de educación estadística. En este marco ellos discuten el proceso de pensamiento estadístico que se encuentra implicado al resolver problemas, desde la formulación del problema hasta la conclusión. Wild y Pfannkuch (1999), conciben al pensamiento estadístico como la suma de cuatro dimensiones: (1) el ciclo de investigación, este ciclo consiste en una serie de pasos que van desde que plantear un problema estadístico hasta resolverlo o modificarlo; (2) los modos fundamentales de razonamiento estadístico, en los que destaca reconocer la necesidad de los datos, transnumeración, percepción de la variación, razonamiento con modelos estadísticos e integración de la estadística y el contexto; (3) el ciclo de interrogación, este ciclo se aplica en cada paso y consiste en la búsqueda y comprobación sucesiva de preguntas o hipótesis, los datos, los análisis realizados o los resultados; y (4) una serie de actitudes, tales como el escepticismo, la imaginación, la perseverancia, el espíritu crítico o la curiosidad. La figura 1.1 resume este marco de pensamiento estadístico.

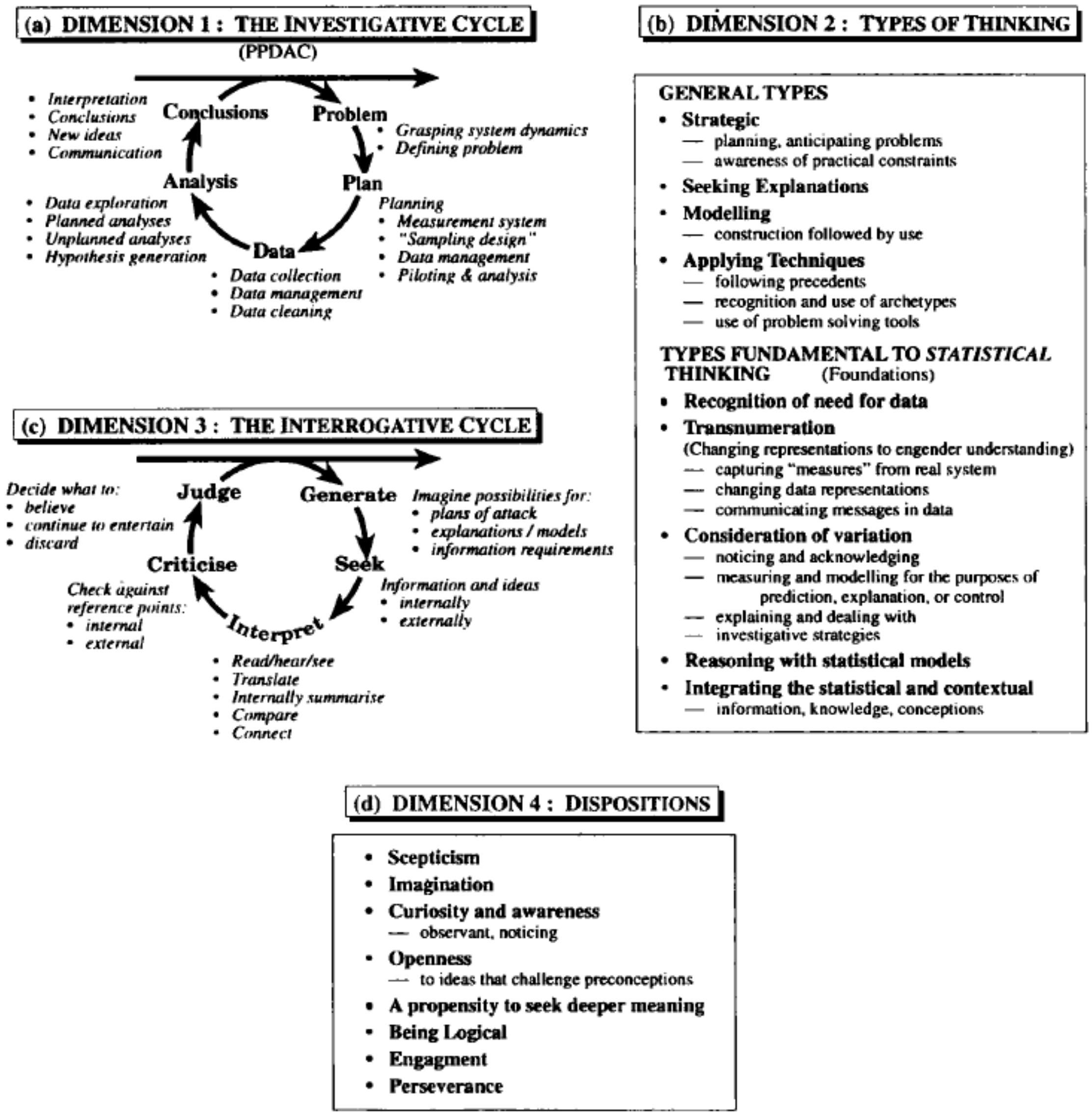


Figura 1. Un marco de cuatro dimensiones para el pensamiento estadístico (fuente: Wild y Pfannkuch, 1999, p. 226)

Batanero (2013), utilizó el término de sentido estadístico como unión de los componentes de cultura estadística y razonamiento y pensamiento estadístico. Además, señaló que este debe construirse en forma progresiva desde la educación básica, media y universitaria. De acuerdo Batanero, Contreras y Arteaga (2011), con las nuevas propuestas curriculares se puede introducir gradualmente ideas estadísticas desde la educación básica, aumentando el nivel de formalización progresivamente.

Finalmente, queremos referirnos a la postura de Ben-Zvi y Garfield (2004) sobre el pensamiento estadístico, quienes describieron que este tipo de pensamiento

> implica la comprensión de por qué y cómo se llevan a cabo las investigaciones estadísticas y las "grandes ideas" que subyacen a las investigaciones estadísticas. Estas ideas incluyen la naturaleza omnipresente de la variación y cuándo y cómo utilizar métodos apropiados de análisis de datos, como resúmenes numéricos y presentaciones visuales de datos. El pensamiento estadístico implica una comprensión de la naturaleza del muestreo, cómo hacemos inferencias de muestras a poblaciones y por qué se necesitan experimentos diseñados para establecer la causalidad. Incluye una comprensión de cómo se usan los modelos para simular fenómenos aleatorios, cómo se producen los datos para estimar probabilidades y cómo, cuándo y por qué las herramientas inferenciales existentes pueden usarse para ayudar en un proceso de investigación. El pensamiento estadístico también incluye ser capaz de comprender y utilizar el contexto de un problema para formar investigaciones y sacar conclusiones, y reconocer y comprender todo el proceso (desde la formulación de preguntas hasta la recopilación de datos, la elección de análisis y la prueba de supuestos, etc.). Finalmente, los pensadores estadísticos pueden criticar y evaluar los resultados de un problema resuelto o de un estudio estadístico. (p. 7)

Otras posturas sobre la alfabetización, razonamiento y pensamiento estadístico pueden verse en delMas (2002) y Rumsey (2002).

1.1.2 Razonamiento Inferencial Informal (RII)

Para promover el razonamiento de inferencial, recientemente surge una propuesta denominada razonamiento inferencial informal, las investigaciones han sugerido que las ideas de inferencia podrían ser introducidas desde el inicio del curso de estadística o desde edades tempranas de manera informal, buscando que los estudiantes se familiaricen con dichas ideas y que cuando aborden tópicos de inferencia estadística puedan comprender mejor las ideas formales, los procedimientos y el lenguaje. En la literatura de investigación para referirse al razonamiento inferencial informal también se suele encontrar como inferencia estadística informal (IEI).

El razonamiento inferencial informal, de un estudiante, concierne al esfuerzo que éste despliega entre la inferencia y la inferencia estadística. Para establecer la diferencia entre ellas, retomamos la postura de Ramsey y Schafer (2012), quienes señalan que "*Una inferencia es una conclusión de aquellos patrones en los datos que se encuentran presentes en algunos* *contextos más amplios. Una inferencia estadística es una inferencia justificada por un modelo de probabilidad vinculando los datos con el contexto más amplio*" (p. 8).

De acuerdo con Zieffler, et al. (2008), el RII se define como "*la forma en la cual los estudiantes usan su conocimiento estadístico informal para plantear argumentos que apoyen inferencias sobre poblaciones desconocidas basándose en muestras observadas*" (p.44), en donde el conocimiento estadístico informal es aquel conocimiento informal que poseen los estudiantes previamente a su contacto con la Estadística Inferencial.

Por su parte, Pfannkuch (2006) indica que el RII puede interconectar ideas estadísticas, por ejemplo, la distribución, muestreo y medidas de tendencia central, dentro de un ciclo de razonamiento 'empírico'. Mientras que, Rubin, Hammerman y Konold (2006) señalan que el RII tiene propiedades de "*agregación en lugar de propiedades de los casos individuales en sí, señales y ruido, varias formas de variabilidad, tamaño de la muestra, control de sesgos y tendencia*" (p. 2). Esta propiedad de agregación también es retomada en la investigación de Makar y Rubín (2009), junto con la incertidumbre y la evidencia utilizada. Como podemos observar, el RII tiene un carácter integrador y multifacético. Desde la perspectiva de Rossman (2008), se describe el RII como ir más allá de los datos disponibles y buscar eliminar o cuantificar el azar como una explicación de los datos observados mediante un argumento sin un método formal.

Dentro de la postura de RII, han surgido algunas propuestas de cómo investigar y trabajar el RII en el aula de clases, tanto con el objetivo de integrar y dar sentido a las nociones estadísticas, como de generar un acercamiento temprano a la Estadística Inferencial por parte de los estudiantes, a continuación, se comentan los niveles y las nociones que se han propuesto para el desarrollo del RII.

1.1.2.1 Propuestas de marcos de trabajo y niveles del RII

Algunas de las propuestas para investigar y promover el RII versan sobre marcos de trabajo (e.g., Zieffler, et al., 2008; Makar y Rubin, 2009; Bakker y Derry, 2011; Makar, Bakker y Ben-Zvi, 2011). Por una parte, Zieffler et al. (2008), proponen un marco de trabajo con un entramado de tareas y componentes con los cuales se puede caracterizar el RII. En dicho marco de trabajo, se contemplan tres tipos de tareas claves (Ibíd., p. 47):

1. Estimar y graficar una población basados en una muestra;

2. Comparar dos o más muestras de datos para inferir si existe una verdadera diferencia entre las poblaciones de las que se obtuvieron las muestras, y
3. Juzgar cuál de dos modelos en competencia o afirmaciones es más probable sea el verdadero.

Estas tareas son pertinentes tanto para indagar sobre el RII como para desarrollar formas de pensamiento involucradas con él, en un proceso que puede incluir (Zieffler et al., 2008, p. 45):

- Razonamiento acerca de las posibles características de una población (por ejemplo, forma, centro) basados en una muestra de datos.
- Razonamiento acerca de las posibles diferencias entre dos poblaciones basados en las diferencias observadas entre dos muestras de datos (es decir, ¿las diferencias son debidas a un efecto en lugar de ser debidas solo al azar?); y
- Razonamiento sobre si una muestra particular de datos (y su resumen estadístico) es o no probable (o sorprendente) dada una expectativa o pretensión en particular.

Estos autores indicaron los siguientes tres componentes como parte esencial del RII, cada uno de los tres tipos de tareas incorpora estos tres componentes, formando así el entramado; es decir, en las respuestas de las actividades, que obedezcan a cada uno de los tres tipos de tareas, pueden tener lugar los siguientes componentes (Zieffler et al., 2008, p. 45):

1. Hacer juicios, afirmaciones, o predicciones acerca de las poblaciones basados en muestras, pero sin utilizar procedimientos y métodos estadísticos formales (por ejemplo, p-valor, prueba t);
2. Aprovechar, utilizar e integrar el conocimiento previo (por ejemplo, conocimiento formal acerca de conceptos fundamentales, tal como distribución o media; conocimiento informal sobre inferencia tal como el reconocimiento de que una muestra puede ser sorprendente dada una pretensión en particular; uso del lenguaje estadístico), a medida que este conocimiento esté disponible; y
3. Articular evidencia basada en argumentos para elaborar juicios, pretensiones o predicciones sobre poblaciones basados en muestras.

Con este entramado de tareas y componentes se puede caracterizar el RII.

Por otra parte, en el marco de trabajo de Makar y Rubin (2009), se establecieron los siguientes conceptos, los cuales fueron considerados como críticos para acceder, por medio de datos, al razonamiento inferencial (Makar y Rubin, 2009, p. 85):

- Noción de incertidumbre y variabilidad articulada a través del lenguaje que rompió con la convención matemática de afirmaciones de certeza;
- Confianza en el concepto de agregado (en oposición a puntos individuales) mediante el uso de generalizaciones sobre el grupo;
- Reconocimiento de un mecanismo o tendencia que se extendió más allá de los datos disponibles; y
- Evidencia de razonamiento basado en el uso intencionado de datos.

A partir de estos conceptos, que inicialmente consideraron críticos, proponen tres principios que son esenciales en una inferencia informal: (1) la generalización, ir más allá de describir los datos; (2) usar los datos como evidencia en las generalizaciones; y (3) utilizar un lenguaje probabilístico al describir la generalización.

Asimismo, Makar, Bakker y Ben-Zvi (2011) realizaron una investigación con estudiantes de sexto grado, en la cual pusieron en escena una propuesta propia del marco de trabajo para dar soporte al estudio del RII, considerando como elementos y estrategias: "*cuestionamientos, conocimiento estadístico, conocimiento del contexto, normas, hábitos, la incertidumbre, las creencias y la explicación*" (p. 159). Además, también señalan la distinción entre IEI y RII, siendo para ellos la primera una declaración o conclusión y la segunda el razonamiento en el que se sustenta dicha declaración.

En el marco del RII también han surgido propuestas de niveles con la taxonomía SOLO (Structure of Observed Learning Outcome) adaptándola para las tareas y concepciones de distintos marcos de trabajo de inferencia informal (e.g., Watson, 2008; Reading, 2009; Juarez e Inzunza, 2014). Reading (2009) tomó como base las investigaciones presentadas en la STRLT5 (Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy), para afirmar que "*muchos acercamientos de educadores estadísticos definiendo el RII empiezan por señalar que es un razonamiento inferencial que no involucra pruebas estadísticas sofisticadas*" (p. 2). También señaló la necesidad de contar con un marco de trabajo para el estudio del RII, para lo cual propuso como punto de partida medir el progreso de los estudiantes en su trayectoria del desarrollo cognitivo implicado en el RII. Específicamente, planteó utilizar dos ciclos del modelo taxonómico SOLO teniendo como precedente el estudio realizado por Watson (2008), en el cual se utilizó dicha taxonomía para introducir descriptores de los niveles de la cognición inicial en el marco de trabajo para el RII, usando los niveles: Uniestructural, Multiestructural y Relacional. En forma análoga, Reading (2009) utilizó los mismos niveles de la taxonomía SOLO en ambos ciclos. En el primer ciclo se enfoca en los conceptos fundamentales mientras que en el segundo ciclo se basa en el azar e implica un razonamiento contextual. Finalmente, Reading (2009) propuso que su marco de trabajo del desarrollo cognitivo del RII de doble ciclo basado en la taxonomía SOLO, puede ser utilizado como punto de partida por investigadores para formular sus propios marcos de trabajo para valorar el grado de RII de los estudiantes.

Juárez e Inzunza (2014), realizaron una investigación sobre la comprensión y el razonamiento acerca de los conceptos estadísticos básicos de los profesores de matemáticas de bachillerato. Para ello, utilizan dos adaptaciones del modelo taxonómico SOLO, la primera para caracterizar el razonamiento estadístico y la segunda para caracterizar las inferencias informales. Ellos concluyen:

> Los resultados del estudio muestran algunas similitudes con resultados de otras investigaciones donde se han observado dificultades por parte de profesores —o futuros profesores— para razonar adecuadamente con conceptos estadísticos, lo que pone en evidencia que la comprensión de conceptos estadísticos y su razonamiento no son procesos triviales, y que para su desarrollo se requiere poner en juego competencias de mayor nivel cognitivo que agrupar datos, construir gráficas y utilizar fórmulas y procedimientos para resumir datos y obtener resultados, como era característico en el enfoque tradicional de enseñanza de la estadística. (Juárez e Inzunza, 2014, p. 27)

En resumen, como podemos observar en los diversos marcos de trabajo y niveles de RII anteriormente comentados, el RII se define y/o utiliza en el sentido de integrador de nociones estadísticas para realizar inferencias informales sin utilizar nociones y procedimientos de Estadística Inferencial (e.g., Prueba de hipótesis, valor-p, valor crítico, nivel de significancia).

1.1.2.2 Nociones fundamentales para el desarrollo del RII

Dada la relevancia que tiene el RII en la inferencia estadística, los investigadores también han desarrollado estudios con propuestas de actividades para comprender y apoyar el surgimiento de la inferencia estadística informal en los primeros años de escolarización, promoviendo ideas y nociones estadísticas, por ejemplo, la idea de predicción, el desarrollo de modelos, la relación entre muestra y población, la media, distribuciones muestrales y la variabilidad del muestreo.

En este sentido, para Pfannkuch (2006) el RII está interconectado con el razonamiento de distribuciones, medidas de tendencia central y muestreo dentro de un ciclo de investigación empírica; aspectos del razonamiento que se basan en la consideración de un elemento fundamental de pensamiento estadístico: la variación. De forma similar, Reading (2009) indicó como nociones básicas de estadística que deben comprenderse para promover el RII a la variación, distribución, media, dispersión y gráficos.

Pfannkuch (2005, 2006, 2007), desarrolló estudios sobre la inferencia informal con diagramas de caja y bigote donde participaron estudiantes y profesores, estos estudios no solo le permitieron determinar nociones fundamentales en las inferencias informales, sino percatarse de que el currículo no brinda una trayectoria para desarrollar el razonamiento inferencial formal, es decir, el currículo parece no contemplar una transición entre el pensamiento inferencial informal y formal. Otras investigaciones (e.g., Zieffler et al., 2008; Watson, 2012) han utilizado diversos tipos de gráficos (puntos, dispersión, caja y bigote, barras, histograma, pastel) para el análisis de datos, en su búsqueda del desarrollo de un RII en los estudiantes.

Tal como Zieffler et al. (2008) y Rossman (2008) hacen referencia a la importancia del argumento en el RII, Ben-Zvi (2006) incluye como componente clave a la argumentación. Él comenta que la argumentación es un "*discurso para la persuasión, la prueba lógica y la creencia basada en la evidencia y, de manera más general, una discusión en la que se presentan los desacuerdos y el razonamiento*" (Ben-Zvi, 2006, p. 2). Para este autor, la integración de la argumentación a las inferencias informales es esencial para construir el razonamiento de los estudiantes y, también propone que este tipo de razonamiento se ve enriquecido cuando utilizamos contextos que a los estudiantes les resulten interesantes y motivadores. En las actividades que se trabajan en el marco de trabajo de Zieffler y en la propuesta de Pfannkuch (2015) las representaciones son fundamentales, debido a que, a partir de los diferentes gráficos se realiza la inferencia informal.

Por su parte, Rossman (2008) identificó algunos conceptos claves en la IEI y cuestiones relacionadas. Si bien concibe que "*la inferencia requiere ir más allá de los datos en la mano*" (p. 5), al realizar una inferencia estadística plantea a la variabilidad aleatoria como fundamental; esto es debido a que la valoración de la variabilidad es el rasgo distintivo y formalmente esto se lleva a cabo mediante un modelo de probabilidad adecuado. Otras investigaciones (e.g., Doerr, delMas y Makar, 2017; Aridor y Ben-Zvi, 2017; Braham, H. y Ben-Zvi, 2017; Kazak y Pratt, 2017; English y Watson, 2018) también han trabajado con modelos de probabilidad tanto para favorecer el RII como para justificar las inferencias informales. Además, Rossman (2008) propuso actividades escolares para el nivel universitario, en las cuales se razonen elementos como azar, recolección de datos, modelos de probabilidad, simulación de muestras e inferencia, para responder preguntas con la posibilidad de utilizar en ello alguna estimación subjetiva, empírica o prefigurada del p-valor involucrado, lo que dé el matiz de informal.

Harradine, Batanero y Rossman (2011), señalan la importancia de la inferencia estadística y retoman algunos caminos alternativos propuestos en diversas investigaciones para introducir a los estudiantes en tales tópicos, refiriéndose a la inferencia informal, mencionan que "*una característica constante de la inferencia informal es que se sugieren actividades para envolver a los estudiantes en el proceso de razonamiento de la inferencia estadística sin depender de distribuciones de probabilidad y fórmulas*" (p. 243). Para llevar a los estudiantes a este tipo de razonamiento se suelen auxiliar de simulaciones donde los estudiantes pueden visualizar el fenómeno y extraer conjeturas; así como en dicha investigación proponen el uso de recursos tecnológicos para promover el RII, otras investigaciones proponen el uso de estos recursos, específicamente con simulaciones y applets (e.g., Rossman, 2008; Watson, 2008; Batanero y Díaz, 2015; Noll, Gebresenbet y Glover, 2016; Ben-Zvi, y Aridor-Berger, 2016).

Por lo que refiere a Batanero y Díaz (2015), consideran el razonamiento inductivo implicado en una inferencia estadística, señalando que la inferencia en sí, es "*generalizar lo observado en casos particulares (la muestra) a algo más general (la población)*" (p. 136), esto se ve complementado por una validación de esta generalización sustentada en el uso de la probabilidad. Entonces, de acuerdo con estas autoras puede ser considerada una inferencia estadística sólo si la inferencia incluye una validación probabilística. Para ejemplificar su planteamiento, Batanero y Díaz (2015) muestran cómo sería una aproximación informal al contraste de hipótesis, una propuesta que "*trata de introducir algunas ideas principales y el razonamiento del contraste y, a la vez, liberar al alumno de los cálculos asociados, recurriendo a la simulación*" (p. 141); ideas como muestra y población, media muestral y poblacional, distribución de la media muestral, hipótesis nula y valor-p. De acuerdo a lo anterior argumentan que,

> la simulación (en vez del cálculo formal) puede utilizarse al comenzar la enseñanza del contraste de hipótesis para poder concentrar al alumno en el aprendizaje de la lógica del proceso y en los conceptos que, todavía necesita: muestra y población, media muestral y poblacional, distribución de la media muestral, hipótesis nula, valor p. (Batanero y Díaz, 2015, p. 142)

Otras investigaciones han realizado aportes sobre una aproximación informal a las pruebas estadísticas, no sólo al contraste de hipótesis sino sobre intervalos de confianza, correlación y regresión lineal y análisis de varianza (e.g., Weinberg, Wiesner y Pfaff, 2010; Stohl Lee, Angotti y Tarr, 2010; Dierdorp, Bakker, Eijkelhof, y van Maanen, 2011; Trumpower, 2013; Trumpower, 2015; Dolor y Noll, 2015).

De acuerdo con Ben-Zvi y Gil (2010) el contexto es un factor que influye en el desarrollo del RII, dado que el contexto desempeña un rol crucial para resolver conflictos entre las expectativas y los datos a analizar, esto se debe a que pueden ayudar en la comprensión de los gráficos. Al respecto de la importancia del contexto también se han pronunciado otras investigaciones como Ben-Zvi (2006), Gil y Ben-Zvi (2011), Pfannkuch (2011), Madden (2011) y Ben-Zvi y Aridor-Berger (2016).

Dentro de los trabajos que han planteado actividades buscando promover el RII, se encuentra el de Makar (2016), donde se realiza un estudio exploratorio para comprender y apoyar el surgimiento de la inferencia estadística informal en los primeros años de escolarización; por medio de actividades logró que los niños desarrollaran la idea de predicción y algunas características clave de la inferencia estadística informal.

Asmismo, Doerr, delMas y Makar (2017), implementan una secuencia instructiva de tareas centradas en el desarrollo de modelos generalizados de estudiantes para hacer inferencias informales al comparar dos conjuntos de datos. Argumentan que las actividades que promueven el razonamiento basado en modelos abordan dos desafíos adicionales: 1) proporcionar una secuencia coherente de temas; y 2) promover la aplicación del conocimiento a situaciones novedosas. En este sentido, Doerr, delMas y Makar (2017), argumentan que las actividades que promueven este tipo de razonamiento abordan desafíos adicionales, como proporcionar una secuencia coherente de temas.

Con nuestra revisión de la literatura sobre el RII y estudios que se han llevado a cabo sobre este tema, podemos observar las siguientes nociones que han resultado fundamentales para desarrollar el RII:

- Variación (e.g., Pfannkuch, 2006; Reading, 2009; Pfannkuch, Arnold yWild, 2015).
- Incertidumbre (e.g., Makar y Rubin, 2009; Ben-Zvi, Aridor, Makar y Bakker, 2012; Henriques y Oliveira, 2016).
- Variabilidad muestral (e.g., Rubin, Hammerman y Konold, 2006; Wild, Pfannkuch, Regan y Horton, 2011; Zieffler et al., 2008; Rossman, 2008; Pfannkuch, 2011; Arnold, Pfannkuch, Wild, Regan y Budgett, 2011; Garfield, Le, Zieffler y Ben-Zvi, 2015).
- Distribución (e.g., Bakker y Gravemeijer, 2004; Pfannkuch, 2006; Reading, 2009; Noll y Hancock, 2015; Kazak, Fujita y Wegerif, 2016).
- Media (e.g., Pfannkuch, 2006; Reading, 2009; Makar, 2014; Watson y English, 2017).
- Dispersión (e.g., Reading, 2009).
- Generalización –de la muestral a la población– (e.g., Zieffler et al., 2008; Makar y Rubin, 2009: Batanero y Díaz, 2015; Makar, 2016).
- Comparación de dos poblaciones por medio de dos muestras (e.g., Watson y Moritz, 1999; Makar y Rubin, 2009; Pfannkuch, 2005, 2006, 2007; Wild et al., 2011; Zieffler et al., 2008).
- Probabilidad (e.g., Rossman, 2008; Makar y Rubin, 2009; Konold et al., 2011; English, 2012; Prodromou, 2013; Batanero y Díaz, 2015; Noll, Gebresenbet y Glover, 2016; Doerr, delMas y Makar, 2017; Aridor y Ben-Zvi, 2017; Braham, H. y Ben-Zvi, 2017; Kazak y Pratt, 2017; Büscher y Schnell, 2017; English y Watson, 2018).
- Agregación (e.g., Makar y Rubin, 2009; Rubin, Hammerman y Konold, 2006; Aridor y Ben-Zvi, 2017).
- Pruebas –prueba de hipótesis, prueba t, intervalos de confianza, correlación y regresión lineal, análisis de varianza y contraste de hipótesis– (e.g., Bakker, Kent, Derry, Noss y Hoyles, 2008; Weinberg, Wiesner y Pfaff, 2010; Stohl Lee, Angotti y Tarr, 2010; Dierdorp, et al., 2011; Trumpower, 2013; Batanero y Díaz, 2015; Trumpower, 2015; Dolor y Noll, 2015).
- Gráficos (e.g., Pfannkuch, 2005, 2006, 2007; Zieffler et al., 2008; Reading, 2009; Watson, 2012).
- Argumentación (e.g., Ben-Zvi, 2006; Zieffler et al., 2008; Rossman, 2008).

- Contextos (e.g., Ben-Zvi, 2006; Gil y Ben-Zvi, 2011; Pfannkuch, 2011; Madden, 2011; Ben-Zvi y Aridor-Berger, 2016).

A partir de estudios como los señalados anteriormente, se considera que la enseñanza de la estadística puede abordar el desafío de introducir las nociones de una manera coherente desde una perspectiva del RII. Sin embargo, Batanero (2013) hace un llamado a reflexionar sobre la dosis exacta de formalización que se requiere para enseñar las nociones estadísticas.

1.1.3 Razonamiento Inferencial Formal (RIF)

Si bien, en los últimos años se ha incrementado la investigación acerca del RII, también se han realizado investigaciones sobre la comprensión de los estudiantes sobre los métodos formales de inferencia estadística con la intención de incidir en las dificultades relacionadas con nociones y técnicas de la Estadística Inferencial que presentan estudiantes y profesores, abordando, por ejemplo, diferentes momentos de la toma de decisiones, la lógica de las pruebas de hipótesis, análisis de datos paramétricos y no paramétricos (e.g., Rochowicz, 2010; Riemer y Seebach, 2014; Tarlow, 2016). Otras investigaciones se han enfocado en tópicos específicos, tales como el valor-p, el tamaño de la muestra e intervalos de confianza (e.g., Lane-Getaz, 2017; Taylor y Doehler, 2015).

Makar y Rubin (2009, p. 85), explicaron que, por inferencia estadística formal, se refieren "*a enunciados de inferencia que se utilizan para realizar estimaciones puntuales o de intervalo de parámetros de población o pruebas de hipótesis formales (generalizaciones), utilizando un método que es aceptado por los estadísticos y la comunidad de investigación*". Por su parte, Park (2012) describe la inferencia estadística formal como un dominio de contenido que involucra los conceptos e ideas relacionados con la estadística inferencial; lo cual incluye fundamentos de la inferencia estadística como la distribución de muestreo, y la inferencia estadística formal, por ejemplo, las pruebas de hipótesis.

Dentro de las investigaciones que revisamos no se encuentran propuestas de marcos de trabajo, pero sí énfasis en algunas nociones que se corresponden con los Lineamientos para la Evaluación y Enseñanza en Educación Estadística, el reporte de la Guidelines for Assessment and Instruction in Statistics (GAISE, 2005, 2016), sobre la importancia de que los estudiantes comprendan las siguientes nociones: las distribuciones muestrales y su importancia al hacer inferencias; estimaciones (intervalos de confianza), su interpretación y el margen de error asociado a la estimación; y significancia estadística y el rol que tiene el valor-p al determinar la significancia estadística.

A continuación, se comentan algunos de los estudios que han surgido sobre cómo trabajar la inferencia estadística en el aula destacando las nociones que se han considerado fundamentales para promover en los estudiantes un RIF.

1.1.3.1 Nociones fundamentales para el desarrollo del RIF

En libros de texto (e.g., Devore, 2008) las muestras, muestreo, distribuciones muestrales, distribuciones poblacionales, distribución de frecuencias y el teorema del límite central se explican como fundamentos de la estadística inferencial; estas nociones son necesarias para que los estudiantes puedan desarrollar un razonamiento inferencial formal, es decir, comprender estas nociones es necesario para que puedan comprender y desarrollar pruebas estadísticas de forma formal, tales como las pruebas de hipótesis.

Se han desarrollado algunas investigaciones que versan sobre la validación de las hipótesis, por ejemplo, Riemer y Seebach (2014) proponen una forma de trabajar la validación de hipótesis por medio de datos reales, donde primero se investiga por simulación, luego por cálculo de probabilidad, después compara las expectativas con los datos reales y, finalmente, los resultados de las 'pruebas de hipótesis'. Estos investigadores señalan que, si las desviaciones son demasiado grandes, se debe dudar de la validez de la hipótesis y en consecuencia rechazarla.

Al igual que en el estudio anterior, Rochowicz (2010) también utilizó software en su propuesta de actividades, y proporcionó algunas técnicas de aproximación numérica que pueden utilizarse para analizar datos y hacer inferencias sobre poblaciones a partir de muestras. La técnica de estimación que se aborda en dicha investigación son los intervalos de confianza, con la ayuda del software; además, realiza una discusión entre el análisis de datos paramétrico y no paramétrico. Asimismo, siguiendo con las propuestas que incluyen recursos tecnológicos, Dinov, Palanimalai, Khare y Christou (2018) realizaron una aplicación web de aleatorización y remuestreo, la cual se puede utilizar para el análisis de datos, así como para el aprendizaje y la enseñanza formal de diversas nociones, por ejemplo, el muestreo, variación aleatoria, y en general la inferencia estadística asistida por computadora y realizar análisis basado en datos.

Por otra parte, diversos estudios han desarrollado propuestas de actividades o problemas para trabajar con los estudiantes y profesores las pruebas de hipótesis (e.g., Leigh y Dowling, 2010; Gibbs y Goossens, 2013; Seir, 2014; DePaolo, Robinson y Jacobs, 2016; Tarlow, 2016; Woodard, Lee y Woodard, 2020; Fellers y Kuiper, 2020). Por ejemplo, Leihg y Dowling (2010) presentaron datos que recolectaron de dos pruebas sobre el sabor del agua en una universidad y cómo éstos pueden ser utilizados para realizar analisis inferenciales con las pruebas del estadístico χ^2, de bondad de ajuste y de independencia. En el mismo sentido, Gibbs y Goossens (2013) proponen analizar un conjunto de datos sobre la eficacia de HPV vaccines con la prueba del estadístico χ^2 de homogeneidad. Asimismo, en la actividad sobre la muestra de 200 transacciones en DePaolo, Robinson y Jacobs (2016) es abordada con la prueba de hipótesis con el estadístico χ^2. Mientras que Tarlow (2016), realiza una propuesta de actividades para ilustrar cómo funciona la prueba de hipótesis con un analisis de varianza (ANOVA), para el desarrollo de estas actividades en el aula de clases Tarlow propone poner énfasis en cómo pensar críticamente sobre el razonamiento inferencial.

También, se han desarrolado investigaciones sobre la lógica de las pruebas de hipótesis (e.g., Vallecillos, 2002; Thompson, Liu y Saldanha, 2007; Liu y Thompson, 2009); por ejemplo, la investigación de Liu y Thompson (2009) se desarrolló con profesores, y mediante entrevistas llegaron a la conclusión que los profesores tenían concepciones incompletas sobre probabilidad, esto generaba que los profesores tuvieran dificultades para comprender las inferecias mediante las pruebas de hipótesis. Además, Thompson, Liu y Saldanha (2007), señalaron que para que los estudiantes puedan comprender la lógica de la inferencia, es necesario que conciban al muestro como un proceso estocástico imaginando la recolección repetida de muestras con el "*resultado de que los valores de un estadístico se distribuyen de alguna manera con un rango de posibilidades*" (Íbid, p. 209). Por su parte, en la investigación de Vallecillos (2002) que se llevó a cabo con estudiantes universitarios, se encontró que los estudiantes no consideran la prueba de hipótesis como un proceso de toma de decisiones para aceptar o rechazar una hipótesis, y se identificó cuatro concepciones que tenían los estudiantes sobre las pruebas de hipótesis.

También se han realizado investigaciones en torno a nociones como el valor-p, el nivel de significancia, el error tipo I y tipo II (e.g., Wilkerson y Olson, 1997; Lane-Getaz,2007; Kaplan, 2009; Lane-Getaz, 2013; Lane-Getaz, 2017; Lyu, Peng y Hu, 2018). Lane-Getaz (2007), diseñó un test sobre el valor-p y el nivel de confianza y posteriormente incluyó otras nociones a su test, por ejemplo, las pruebas de hipótesis y la probabilidad de cometer el error tipo I y el tipo II; entonces, con este nuevo diseño Lane-Getaz (2013) investigó sobre la definición de un valor-p, uso de pruebas de significancia, interpretación de los resultados a través de pruebas de hipótesis, establecer conclusiones sobre la significancia estadística, la vinculación del valor-p con las hipótesis y sobre realizar conexiones inferenciales. Kaplan (2009), realizó entrevistas a estudintes universitarios y les presento actividades que contenían la descripción del estudio experimental, las conclusiones estadísticas, el valor-p y las interpretaciones de los resultados de la prueba. Lyu, Peng y Hu (2018), realizaron actividades, en las cuales primeramente describían un escenario de investigación ficticio, donde el valor-p era igual a 0.01; después, presentaron seis declaraciones sobre el valor-p y les pidieron a los participantes que marcaran cada una de las declaraciones como verdadera o falsa. En este caso, Falso significaba que la declaración no seguía lógicamente las premisas.

A partir de la revisión de literatura sobre el RIF observamos las siguientes nociones que han resultado fundamentales para su desarrollo:

- Pruebas de hipótesis (e.g., Leigh y Dowling, 2010; Kamin, 2010; Gibbs y Goossens, 2013; Lane-Getaz, 2013; Seir, 2014; DePaolo, Robinson y Jacobs, 2016; Tarlow, 2016; Woodard, Lee y Woodard, 2020; Fellers y Kuiper, 2020).
- Valor-p (e.g., Cumming, 2006; Lane-Getaz, 2007; Kaplan, 2009; Lane-Getaz, 2013; Lyu, Peng y Hu, 2018).
- Nivel de significancia (e.g., Wilkerson y Olson, 1997; Lane-Getaz, 2007; Lane-Getaz, 2013; Lyu, Peng y Hu, 2018).
- Error tipo I y tipo II (e.g., Wilkerson y Olson, 1997; Lane-Getaz, 2013).
- Lógica de las pruebas de hipótesis (e.g., Vallecillos, 2002; Thompson, Liu y Saldanha, 2007; Liu y Thompson, 2009).
- Validación de la inferencia (e.g., Wilkerson y Olson, 1997; Riemer y Seebach, 2014).
- Intervalos de confianza (e.g., Cumming, Williams y Fidler, 2004; Rossman y Chance, 2004; Cumming, 2006; Cumming, 2015; Rochowicz, 2010).
- Muestras y muestreo (e.g., Saldanha y Thompson, 2002; Devore, 2008; Dinov et al., 2018).
- Distribuciones muestrales, poblacionales y de frecuencias (e.g., Lipson, 2003; Chance, delMas y Garfield, 2004; Devore, 2008).
- Teorema del límite central (e.g., Sotos et al., 2007; Devore, 2008).

1.2 Aspectos cognitivos subyacentes al desarrollo del razonamiento inferencial en los estudiantes

En esta sección abordaremos los errores y dificultades que presentan los estudiantes y profesores al realizar inferencias estadísticas que se han reportado en la literatura de investigación de educación estadística. Finalmente, enlistaremos las principales dificultades que observamos.

1.2.1 Errores y dificultades en el estudio de la inferencia estadística

La importancia de la inferencia estadística, dentro de la Estadística, se debe a su relevancia para realizar análisis, emitir juicios y tomar decisiones en estudios estadísticos. Sin embargo, diversas investigaciones (e.g., Watson, 2004; Bakker y Gravemeijer, 2004; Carver, 2006; Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Garfield y Ben-Zvi, 2008; Harradine, Batanero y Rossman, 2011; Batanero, Vera y Díaz, 2012; Makar y Rubin, 2018) han reportado las dificultades que presentan los estudiantes y profesores al trabajar con inferencia estadística. A continuación, describimos las dificultades identificadas.

Vallecillos (1994), clasificó los siguientes errores relacionados con la inferencia estadística, especialmente con las nociones de hipótesis y nivel de significación y, sobre la lógica global de la prueba de hipótesis:

- Parámetro, estadístico y distribuciones muestrales
- Confusión entre las hipótesis
- Interpretación de las probabilidades de error y sus relaciones
- Nivel de significación
- Criterio de decisión (confución entre las regiones de rechazo y no rechazo)
- Interpretación de resultados

En resumen, Vallecillos (2002) encontró que los estudiantes confunden el nivel de significación con la probabilidad de obtener un resultado correcto. Además, en relación con las dificultades que presentaron los estudiantes sobre el nivel de identificación, identificó que, por una parte, los estudiantes presentaron confusión con la hipótesis nula y la hipótesis alternativa, ellos definen como hipótesis nula aquella que se desea probar; por otra parte, los estudiantes tuvieron dificultades con la idea de estadístico como variable aleatoria y de la relación con la distribución muestral, su variabilidad, y su relación con el parámetro poblacional. También identificó concepciones erróneas sobre el contraste de hipótesis como una prueba probabilística de la hipótesis, y la determinista como una prueba matemática de la hipótesis que es aceptada.

Garfield y Ben-Zvi (2008) y Saldanha y Thompson (2002), sugieren que los estudiantes poseen una comprensión incompleta de la distribución, la variación, el muestreo y las distribuciones del muestreo, conceptos que consideran fundamentales para el razonamiento inferencial. Aunado a lo anterior, Pfannkuch (2005) señala que la falta de proximidad de los estudiantes con eventos de estocástica también dificulta el desarrollo de este tipo de razonamiento. Además, algunos estudios han demostrado que los estudiantes no distinguen entre la distribución de una muestra y la distribución muestral de un estadístico (e.g., Lipson, 2003).

El valor-p es una de las principales dificultades en inferencia que se ha reportado en diversas investigaciones y comúnmente se relaciona con errores y dificultades en la comprensión del nivel de significancia (e.g., Williams, 1999; Cohen, 1994; Vera y Díaz, 2013; Inzunza y Jiménez, 2013; Biehler, Frischemeier y Podworny, 2015; López-Martín, Batanero y Gea, 2018). A continuación, describimos los hallazgos de algunas de ellas. Williams (1999) investigó sobre el conocimiento conceptual de las pruebas de hipótesis de los estudiantes estadística novatos, centró su investigación en cuatro conceptos mayores: hipótesis, nivel de significancia, valor-p y significancia. Para realizar la investigación le resultó útil los mapas conceptuales y encontró que la comprensión de los estudiantes sobre la definición de nivel de significancia varió entre verlo como un nivel para la toma de decisiones, una medida de significancia, un nivel de confianza o como el error. Mientras que para estudiar el conocimiento procedimental utilizó tareas de prueba de hipótesis y observó que los estudiantes demostraron una confusión entre el valor-p y el nivel de significancia.

Por otra parte, Cohen (1994) revisó los problemas con las prácticas del valor-p e identificó interpretaciones erróneas del valor-p como la probabilidad de que la hipótesis nula sea falsa y que el complemento, sea, por lo tanto, la probabilidad de que la siguiente replicación sea exitosa y que, además, se realizan suposiciones erróneas en cuanto a que, si uno rechaza la hipótesis nula, entonces afirma la teoría que llevó a prueba. De acuerdo con Haller y Krauss, (2002), los conceptos más complicados asociados a los resultados de las pruebas de hipótesis son el valor-p y el nivel de significancia.

Liu y Thompson (2009), identificaron desafíos conceptuales bajo la lógica de la prueba de hipótesis que tienen los profesores, como concepciones incompletas sobre los resultados de probabilidad, y destacaron que esta concepción incompleta se vuelve un desafío cuando se trata de comprender las inferencias en las pruebas de hipótesis. Además, observaron que los profesores presentaban dificultades para concebir el rol de las pruebas de hipótesis, como una herramienta útil para la toma de decisiones. Asimismo, Vallecillos (2002) investigó la comprensión de estudiantes universitarios sobre la prueba de hipótesis y encontró que los estudiantes no consideran la prueba de hipótesis como un proceso de toma de decisiones para aceptar o rechazar una hipótesis. Vallecillos (2002), también identificó cuatro concepciones distintas en cuanto a las pruebas de hipótesis: (1) la prueba como regla de toma de decisiones; (2) la prueba como un procedimiento para obtener soporte empírico de la hipótesis que se investiga; (3) la prueba como prueba probabilística de las hipótesis; y, (3) la prueba como prueba matemática de la verdad de la hipótesis.

Por su parte, Vera y Díaz (2013) investigaron sobre la asignación de hipótesis estadísticas, diferenciación entre hipótesis nula y alternativa, diferenciación entre contraste unilateral y bilateral, asignación del valor-p y nivel de significación, determinación de la zona de rechazo y aceptación y regla de decisión. En dicha investigación encontraron que los estudiantes de psicología presentaron dificultades con la diferenciación entre el contraste unilateral derecho e izquierdo, pues los estudiantes confundieron el signo del valor crítico correspondiente para un nivel de significación específico. Además, observaron que los estudiantes confundieron las hipótesis nula y alternativa, errores tipo I y tipo II y la regla de decisión en el contraste de hipótesis.

El planteamiento de las hipótesis nula y alternativa es un aspecto fundamental en las pruebas de hipótesis y ha sido reportado como una de las principales dificultades que presentan los estudiantes y profesores en inferencia estadística (e.g., Vallecillos, 1994, 1997, 2002; Vallecillos, y Batanero, 1997; Sotos et al., 2007; Batanero, 2013; Vera y Díaz, 2013; López-Martín, Batanero y Gea, 2019). En dichas investigaciones se observó que las dificultades con el planteamiento de las hipótesis estadísticas pueden afectar la comprensión del proceso que implica realizar la prueba de hipótesis e interpretar los resultados. Por ejemplo, Vallecillos y Batanero (1997), encontraron que los participantes presentaron dificultades para plantear las hipótesis nula y alternativa para un problema contextualizado (intercambiaron las hipótesis), aunque comprendían teóricamente que la hipótesis nula se plantea con la intención de encontrar evidencia en su contra. Además, identificaron concepciones erróneas sobre las hipótesis, por ejemplo, considerar que la hipótesis nula y la región de aceptación son lo mismo, y que una hipótesis puede referirse tanto a una población como a una muestra.

López-Martín, Batanero y Gea (2019), realizaron una investigación con fututos profesores e identificaron que confundieron los errores tipo I y tipo II e indicaron que una vez que los profesores toman la decisión sobre la hipótesis nula, existe la posibilidad de cometer errores dependiendo de la decisión que hayan tomado. Además, puntualizan que la importancia de diferenciar los errores tipo I y tipo II se debe a que la probabilidad asociada al primer error queda determinada antes de la prueba, mientras que la probabilidad asociada al segundo error es variable y dependiente de diversos factores. Por su parte, diversas investigaciones (e.g., Falk, 1986; Shaughnessy y Dick, 1991; Sotos et al., 2007) señalan que los estudiantes confunden las probabilidades condicionales que intervienen en el nivel de significación, el valor-p y los errores tipo I y tipo II con probabilidades de un solo evento, esto revela una dificultad mucho mayor que es inherente a la probabilidad.

De forma análoga, Hoekstra (2015) realiza una investigación que permite observar que las dificultades en la inferencia estadística no son exclusivas de los estudiantes, sino también se presentan con investigadores. Este autor utiliza el concepto de riesgo como factor explicativo de las interpretaciones inferenciales, pues considera que las técnicas estadísticas protegen a sus usuarios contra las malas interpretaciones respecto de la incertidumbre. Sin embargo, los investigadores a menudo pretenden estar demasiado seguros de la presencia o ausencia de un efecto, y los datos se analizan de manera selectiva, lo que afecta la validez de las conclusiones o inferencias que pueden esbozar a partir de las técnicas que se utilizan.

En resumen, en las investigaciones que reportan errores y dificultades sobre inferencia estadística, describieron errores en la interpretación del valor-p y del nivel de significancia (eg., Vallecillos, 1994), y con los errores tipo I y tipo II (e.g., Vera y Díaz, 2013; Sotos et al., 2007; López Martín, Batanero y Gea, 2019), estos errores se encuentran relacionados especialmente con la probabilidad. También se observaron problemas en identificar la población bajo estudio, en discriminar contrastes unilaterales y bilaterales, y con el criterio de decisión asociado al tipo de contraste (Vallecillos, 2002; Sotos et al. 2007; Vera, Díaz y Batanero, 2011); estas dificultades se encuentran relacionadas con otra dificultad, que refiere a la comprensión de la lógica que subyace al contraste de hipótesis (Liu y Thompson, 2009; Harradine, Batanero y Rossman, 2011).También se reportan errores al confundir el estadístico y el parámetro, esto al plantear las hipótesis en términos de un estadístico y no del parámetro (e.g., Vallecillos, 1994; Sotos et al., 2007; López Martín, Batanero y Gea, 2019). Asimismo, se identificó que existen confusiones entre el planteamiento de la hipótesis nula y alternativa. A continuación, enlistamos las principales dificultades que hemos identificado, a partir de nuestra revisión de la literatura de investigación:

- Lógica de las pruebas de hipótesis (e.g., Liu y Thompson, 2009; Vera y Díaz, 2013; Harradine, Batanero y Rossman, 2011)
- Planteamiento de las hipotesis nula y alternativa (e.g., Vallecillos, 1994, 1997, 2002; Vallecillos, y Batanero, 1997; Sotos et al., 2007; Batanero, 2013; Vera y Díaz, 2013; López-Martín, Batanero y Gea, 2019)
- Valor-p (e.g., Williams, 1999; Cohen, 1994; Vera y Díaz, 2013; Inzunza y Jiménez, 2013; Biehler, Frischemeier y Podworny, 2015; López-Martín, Batanero y Gea, 2019)
- Nivel de significancia (e.g., Williams, 1999; Vera y Díaz, 2013; Inzunza y Jiménez, 2013; Biehler, Frischemeier y Podworny, 2015; López-Martín, Batanero y Gea, 2019)
- Criterio de desición (e.g., Vallecillos, 2002; Sotos et al. 2007; Vera, Díaz y Batanero, 2011).
- Errores tipo I y tipo II (e.g., Falk, 1986; Shaughnessy y Dick, 1991; Sotos et al., 2007; Vera y Díaz, 2013; López-Martín, Batanero y Gea, 2019)
- Estadístico-parámetro (e.g., Vallecillos, 1994; Sotos et al., 2007; López Martín, Batanero y Gea, 2019).
- Fundamentos –e.g., distribución, la variación, el muestreo y distribuciones muestrales– (e.g., Garfield y Ben-Zvi, 2008; Saldanha y Thompson, 2002).

1.2.2 Problemáticas subyacentes al estudio de los estadísticos Chi-cuadrada y t-Student

En este estudio nos vamos a enfocar en los estadísticos Chi-cuadrada y t-Student; estos estadísticos se han ganado un lugar importante dentro de las pruebas no paramétricas y paramétricas, respectivamente; esto se debe a que, por una parte, el estadístico Chi-cuadrada fue valioso en la construcción de una metodología para las pruebas de hipótesis y actualmente, las pruebas de hipótesis con este estadístico forman parte de la estadística aplicada; contribuyendo de manera importante en la medicina, psicología, genética, acuicultura, biología, análisis financiero, econometría, industria e investigaciones de mercadeo. Por otra parte, una de las bondades principales de las pruebas con el estadístico t-Student es que pueden ser utilizadas con muestras pequeñas, muestras de este tipo son muy frecuentes tanto en estudios del área industrial, farmacéutica y agronómica como en los de carácter social. A pesar de ello, realizar inferencias basados en pruebas tanto con el estadístico t-Student como con el estadístico Chi-cuadrada requiere una comprensión profunda de las nociones que se encuentran relacionadas con estos estadísticos, ya que aunque los profesores y futuros profesores pueden realizar los procedimientos aparentemente de forma adecuada cuando resuelven un problema guiado del libro texto, pueden tener dificultades para comprender o conectar las nociones estadísticas involucradas, lo cual podría llevar a no saber identificar el tipo de datos, elegir el estadístico y la distribución o la prueba necesaria para resolver un problema real (Batanero, 2013).

Otras investigaciones han documentado errores y dificultades en la comprensión de los estadísticos Chi-cuadrada y t-Student (e.g., Batanero, Estepa, Godino y Green, 1996; Garfield, 2003; Moritz, 2004; Cañadas, Batanero, Díaz y Gea, 2012). A continuación, describimos los hallazgos de dichas investigaciones.

Cañadas, Batanero, Díaz y Gea (2012), realizaron una investigación sobre la comprensión de las pruebas de hipótesis con el estadístico Chi-cuadrada, para el caso de la prueba de homogeneidad, con estudiantes universitarios de psicología. En este estudio identificaron las siguientes dificultades: Por un lado, los estudiantes presentaron dificultades en el planteamiento de las hipótesis al hacer referencia a la media poblacional y en vez de usar la proporción (dado que los datos son cualitativos), también se identificaron errores en el planteamiento de la hipótesis alternativa y al igual que en Vera, Díaz y Batanero (2011), en este estudio se logró observar que las dos hipótesis no cubren el espacio paramétrico. Además, al igual que en las dificultades descritas en el apartado anterior (e.g., Vallecillos, 1994; Vera y Díaz, 2013), los estudiantes confunden las hipótesis nula y alternativa. Asimismo, identificaron que los estudiantes confundieron el parámetro y el estadístico, puesto que plantearon hipótesis en términos del estadístico.

Por otro lado, los estudiantes confundieron el contraste de homogeneidad con el de independencia e indicaron que la asociación se da entre filas o entre columnas. Los estudiantes confunden los grados de libertad, este error con los grados de libertad se ha identificado en diversas investigaciones con la distribución t-Student t, Chi-cuadrada y F (e.g., Alvarado, 2007; Olivo, 2008). También, identificaron que los estudiantes confunden el valor crítico con el valor-p, y que no pueden reconocer el valor de la probabilidad asociada al valor obtenido del estadístico.

En cuanto a la interpretación de resultados, Cañadas, Batanero, Díaz y Gea (2012), encontraron que los estudiantes tuvieron dificultades para interpretar en el contexto del problema como se les pide, o no lograron realizar la interpretación en el contexto y toman una decisión correcta haciendo uso de una interpretación genérica de la prueba de hipótesis.

Batanero, Estepa, Godino y Green (1996), realizaron una investigación sobre las concepciones de estudiantes pre-universitarios acerca de la asociación en tablas de contingencia. Ellos indicaron que aunque muchos estudiantes demostraron juicios correctos o parcialmente correctos y estrategias de solución correctas, algunos estudiantes utilizaron conceptos erróneos y estrategias incorrectas. De acuerdo con Batanero, Estepa, Godino y Green (1996), a los estudiantes se les dificultó la comparación de distribuciones de frecuencias condicionales absolutas proporcionales, esto generó que se les dificulte la comparación de probabilidades. Esto cuando los estudiantes encontraron una asociación empírica e intentaron justificar su respuesta. También comentaron que una típica respuesta involucró la comparación de dos probabilidades sobre los casos favorable y no favorable. Además, observaron que los estudiantes presentan una *concepción unidireccional de asociación* en la que los estudiantes sólo admiten asociación directa o independencia (asociación inversa); en otras palabras, los estudiantes concibieron erróneamente la dependencia de dos variables dicotómicas como una coincidencia de la presencia o ausencia de estas dos variables. Los estudiantes también presentaron una *concepción determinista de asociación*, los investigadores indicaron que los estudiantes esperan que exista una correspondencia que asigne sólo un valor en la variable dependiente para cada valor de la variable independiente, cuando esto no sucede, consideran que no existe dependencia entre las variables.

Por otra parte Batanero, Estepa, Godino y Green (1996), indicaron que los eventos negativos tienen un menor impacto en los estudiantes, de lo cual extraen que no consideran un evento negativo de la misma manera que uno positivo. Asimismo, que los estudiantes presentan una *concepción localista de asociación*, puesto que a menudo forman su juicio utilizando sólo parte de los datos proporcionados en la tabla de contingencia. Si esta información parcial (una distribución condicional o una sola celda) sirve para confirmar un tipo de asociación dado, adoptan este tipo de asociación en su respuesta.

Por su parte, Garfield (2003) observó concepciones erroneas que involucran promedios. Sobre esta concepción erronea identificó algunos aspectos, por ejemplo, (1) identificó que los promedios son el número más común; (2) no tiene en cuenta los valores atípicos al calcular la media; (3) compara grupos enfocándose exclusivamente en la diferencia en sus promedios; y (4) suelen confundir la media con la mediana.

Dada la complejidad que involucra el estudio de las pruebas con la Chi-cuadrada y la t-Student, y que hemos comentado, en este estudio hemos considerado focalizarnos en estos dos estadísticos; considerando también que la investigación sobre los procesos de enseñanza y de aprendizaje de éstos, no ha recibido suficiente atención por parte de la comunidad de investigación. Además, como hemos visto, el estudio de los estadísticos es un elemento clave para promover el desarrollo del razonamiento inferencial tanto formal como informal.

1.3 Una aproximación al problema de investigación

Como hemos evidenciado a lo largo de este capítulo, la importancia que tiene el RII para los estudiantes radica en que pueda ayudarle a tener una transición continua hacia el RIF. Como argumentan Zieffler, et al., (2008): "*es importante que los estudiantes construyan sobre su conocimiento informal para reinventar conceptos formales y representaciones y al mismo tiempo ampliar su comprensión del sentido común de los fenómenos del mundo real*" (p. 42).

Sin embargo, poco se conoce de la transición del RII al RIF. Sobre esta problemática, de transición continua hacia el RIF, algunas investigaciones (e.g., Weinberg, Wiesner y Pfaff, 2010; Jacob y Doerr, 2014; Pfannkuch, Arnold y Wild, 2015; Makar y Rubin, 2018), indican la necesidad de introducir la inferencia por etapas, en otras palabras, promover progresivamente el Razonamiento Inferencial Formal (RIF) en los estudiantes. Weinberg, Wiesner y Pfaff (2010), presentaron el análisis de una actividad práctica que está diseñada para ayudar a los estudiantes, en un curso introductorio de Estadística, a realizar inferencias informales sobre una bolsa de fichas de bingo y conectar estas ideas informales con lo formal del 'T-test' e 'intervalos de confianza'.

Por su parte, Jacob y Doerr (2014) efectúan un estudio acerca del razonamiento estadístico informal en torno a las 'distribuciones muestrales' para realizar el análisis inferencial formal. El estudio establece que "*el conocimiento de las características de una distribución muestral* *y las experiencias con la generación de distribuciones muestrales no proporcionan bases suficientes para este razonamiento*” (p. 1).

En relación con esta problemática, Pfannkuch y Wild (2015) señalan que la inferencia estadística debe ser parte de las experiencias de aprendizaje de los estudiantes y los fundamentos conceptuales necesarios para una introducción por etapas de la inferencia; reflexión derivada de la implementación de las etapas con estudiantes de 13 a 20 años. Dicha implementación en las aulas requirió del desarrollo de nuevas visualizaciones dinámicas, verbalizaciones, formas de razonamiento, trayectorias de aprendizaje y recursos materiales. Estos autores indicaron algunos conceptos fundamentales, para la inferencia informal (‘Making a call’, ideas de muestra-población, variabilidad del muestreo) y para la inferencia formal (método de remuestreo y método de aleatorización), los cuales se pueden trabajar en diversos momentos del currículo escolar. También presentan algunas actividades de aprendizaje sobre la comparación de diagramas de caja y el uso de software que ayudan a explicar cada uno de los conceptos fundamentales. Sin embargo, aún sigue abierta la discusión sobre cómo podemos construir un RIF sobre las bases que se han desarrollado del RII.

Los resultados y reflexiones de las investigaciones apuntan al análisis del razonamiento inferencial informal de los estudiantes y la necesidad de promover progresivamente un razonamiento inferencial formal. En este análisis el estadístico puede brindar una trayectoría desde lo intuitivo a lo formal, ya que se puede observar tanto en estudios de RII como de RIF (e.g., Pfannkuch, 2006; Rossman, 2008; Weinberg, Wiesner y Pfaff, 2010; Rochowicz, 2010; Vera y Díaz, 2013; Riemer y Seebach, 2014; Cañadas, Batanero, Díaz y Gea, 2012; Dolor y Noll, 2015; Taylor y Doehler, 2015). No obstante, en Educación Estadística aun no hay un consenso y sigue abierta la discusión sobre cómo promover el razonamiento inferencial de forma progresiva, es por ello que es necesario contar con propuestas que permitan explorar y desarrollar progresivamente (del IIR al FIR) el razonamiento inferencial de estudiantes y profesores, considerando los aspectos clave tanto para el desarrollo tanto del RII como del RIF (los cuales hemos identificado en las secciones 1.1.2.2 y 1.1.3.1).

Por la relevancia de los estadísticos Chi-cuadrada y t-Student, mencionada en el apartado anterior, consideramos que en esta búsqueda de la integración del RII al RIF, estos nos permitirán brindar una trayectoria o una transición continua desde el RII al RIF; de momento nos permitimos cuestionamos ¿cómo podemos ayudar a los estudiantes a transitar del RII al RIF?, ¿cuáles son los significados de las nociones de los estadísticos Chi-cuadrada y t-Student que los profesores y estudiantes utilizan en sus prácticas matemáticas/estadísticas?, ¿cuáles son los vínculos que establecen los estudiantes entre los diversos significados de estos estadísticos?, ¿tales significados de los estadísticos Chi-cuadrada y t-Student, se pueden relacionar en niveles progresivos de razonamiento inferencial? Estas preguntas orientarán nuestra investigación, y las retomaremos y reformularemos, para proponer nuestro problema y objetivos de investigación, una vez hayamos introducido las nociones teóricas que utilizaremos para nuestro estudio.

Con esta investigación se pretende realizar una propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, sobre los estadísticos Chi-cuadrada y t-Student, tomando como base tanto la riqueza matemática que puede recuperar mediante los estudios histórico-epistemológicos sobre estos estadísticos, como los aportes de la literatura de educación estadística sobre el razonamiento inferencial. Pretendemos que los niveles de RI presenten 'indicadores o criterios' que vayan de lo informal a lo formal. También se pretende explorar empíricamente los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student.

1.4 Reflexiones finales

En este capítulo, presentamos un panorama general de la literatura de investigación sobre el razonamiento inferencial informal, formal y sobre la transición del RII al RIF. Además, describimos diversas posturas y usos sobre la alfabetización, el razonamiento y el pensamiento estadístico.

Cuando hablamos de alfabetización estadística a menudo se relaciona con aspectos cuantitativos, sin embargo, hemos observado que también se refiere a comprender la información estadística o los resultados de la investigación; mientras que pensamiento y razonamiento suelen usarse para abordar los mismos aspectos. Ante esta problemática, Ben-Zvi y Garfield (2004), presentan las definiciones de dichas nociones, rescatando las diferencias que existen entre ellas. En cuanto a la alfabetización estadística, estos autores señalan que incluye las habilidades básicas para comprender información estadística o resultados de una investigación, mientras que el razonamiento estadístico lo definen como "*la forma en que las personas razonan ideas estadísticas y dan sentido a la información estadística*" (p. 7); y, el pensamiento estadístico implica comprender por qué y cómo se llevan a cabo las investigaciones estadísticas, la naturaleza del muestreo, cómo utilizar modelos para simular fenómenos aleatorios, cómo, cuándo y por qué pueden usarse las herramientas inferenciales, utilizar el contexto de un problema para formar una investigación y esbozar conclusiones, comprender el proceso de una investigación estadística; así mismo, ser capaz de criticar y evaluar tanto los resultados como el estudio estadístico.

Como pudimos observar en las secciones 1.1, los términos inferencia estadística informal y razonamiento inferencial informal suelen utilizarse en el mismo sentido, aunque en algunas investigaciones plantean que la siguiente diferencia: la inferencia estadística informal es una declaración o conclusión y el razonamiento inferencial informal es el razonamiento en el que se sustenta la conclusión. No obstante, en esta investigación para referirnos al RII tomamos como referente tanto la conclusión como el razonamiento que la sustenta, esta postura coincide con la declaración que realiza Rossman (2008) sobre qué es una inferencia.

En la sección 1.1 también identificamos las nociones claves, tanto para el RII como para el RIF, entre estas nociones, para el RIF, se destacan: el nivel de significancia, valor-p, los errores tipo I y tipo II, planteamiento de las hipótesis estadísticas, distribuciones muestrales, estadístico y parámetro y, la lógica de las pruebas de hipótesis. Mientras que para el RII destacan nociones tales como, variación, incertidumbre, variabilidad muestral, distribución, dispersión, media, probabilidad, las propiedades de generalización y agregado, comparación de dos poblaciones por medio de dos muestras y aspectos como los gráficos, argumentos y contextos. Consideramos que el razonamiento sobre las nociones que se destacaron en el RII, es fundamental que se promueva antes de desarrollar un RIF.

Los principales errores y dificultades que se observaron cuando estudiantes y profesores trabajan con la inferencia estadística y que se abordaron en la sección 1.2, son: la lógica de las pruebas de hipótesis, planteamiento de las hipótesis nula y alternativa, valor-p, nivel de significancia, criterio de decisión, los errores tipo I y tipo II, estadístico-parámetro, y los fundamentos (e.g., distribución, la variación, el muestreo y distribuciones muestrales).

Los posicionamientos, propuestas de marcos de trabajo y de actividades sobre el razonamiento inferencial informal y formal, así como las nociones que se abordan en dichas investigaciones y que hemos denominado como nociones fundamentales para el desarrollo del RII y del RIF, y los errores y dificultades anteriormente mencionados, nos ayudan a tener una visión global del razonamiento inferencial, tanto informal como formal, y consideramos que serán relevantes en la constitución de los criterios para una transición continua hacia el RIF.

Finalmente, sobre dicha transición continua hacia el RIF, observamos que en los resultados y reflexiones de las investigaciones de enseñanza y aprendizaje de inferencia estadística, sugieren la necesidad de promover progresivamente un razonamiento inferencial formal a partir del razonamiento inferencial informal. Pensamos que las características matemáticas de los estadísticos Chi-cuadrada y t-Student nos podrían ayudar a brindar una trayectoria para desarrollar el razonamiento inferencial, desde lo intuitivo a lo formal.

CAPÍTULO 2

Marco teórico, Pregunta y objetivos de investigación, y Metodología

"Conozca todas las teorías. Domine todas las técnicas, pero al tocar un alma humana sea apenas otra alma humana".
(Carl G. Jung)

INTRODUCCIÓN

En este capítulo se encuentran tres grandes secciones, las cuales están en concordancia con los antecedentes y problemática presentada en el Capítulo 1. En la primera de ellas, abordamos las nociones y herramientas teóricas del "Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos", las cuales dan soporte a nuestra investigación. En la segunda sección planteamos la pregunta y los objetivos de investigación que orientan el estudio. Finalmente, la tercera sección corresponde a la metodología del estudio, en la cual describimos las fases de la investigación, los sujetos de estudio y el instrumento de indagación.

2.1 Marco Teórico

2.1.1 Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos (EOS)

Para alcanzar el objetivo general, hemos adoptado la postura teórica del Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos (EOS), la cual surge a inicio de los 90's en el grupo de investigación de Educación Matemática de la Universidad de Granada y se ha desarrollado en diversos trabajos (Godino y Batanero, 1994; Godino y Batanero, 1998; Godino, Batanero y Font, 2007; 2019). Este enfoque es un sistema teórico inclusivo que trata de articular diversas aproximaciones y modelos teóricos que se utilizan en la investigación en Educación Matemática a partir de supuestos antropológicos, socioculturales y semióticos sobre las matemáticas y su enseñanza (Presmeg, 2014). En el EOS se asumen dos postulados, el antropológico y el semiótico, el primero señala que las matemáticas son una actividad humana, y el segundo, que las entidades involucradas en la actividad emergen de las acciones o discursos a través de los cuales se expresan y comunican.

En esta investigación utilizamos el EOS porque provee herramientas teórico-metodológicas que permiten realizar análisis pormenorizados de las prácticas matemáticas que desarrolla un sujeto (estudiante o profesor) y de aquellas prácticas que dan cuenta de la evolución de un objeto matemático. En otras palabras, nos permiten estudiar los significados del objeto matemático, en nuestro caso, de los estadísticos Chi-cuadrada y t-Student. A continuación, describimos brevemente algunas nociones del EOS que consideramos de utilidad para el desarrollo de esta investigación.

2.1.1.1 Sistema de prácticas

En el EOS se reconoce una doble naturaleza para las matemáticas: como sistema de objetos y como sistema de prácticas. La noción de sistema de prácticas tiene un papel muy importante dentro de este enfoque, ya que permite hacer operativo el postulado antropológico, entendiéndolas como refieren a la práctica matemática como "*toda actuación o manifestación (lingüística o no) realizada por alguien para resolver problemas matemáticos, comunicar a otros la solución, validar la solución y generalizarla a otros contextos y problemas*" (Godino y Batanero, 1994, p. 334).

Las prácticas matemáticas pueden ser idiosincrásicas de una persona (practicas personales) o compartidas por un grupo en el seno de una institución (prácticas institucionales), en el primer caso nos referimos a ellas como sistema de prácticas personales y de acuerdo con Godino y Batenero (1994, p. 339) "*Los sistemas de prácticas personales, asociadas a un campo de problemas, está constituido por las prácticas prototípicas que una persona realiza en su intento de resolver un campo de problemas C. Representamos este sistema por la notación* $P_p(C)$". Mientras que en el segundo caso, cuando hablamos de prácticas institucionales, nos referimos a ellas como sistema de prácticas institucionales, y conforme a los mismos autores "*El sistema de prácticas institucionales, asociadas a un campo de problemas, está constituido por las prácticas consideradas como significativas para resolver un campo de problemas* C *y compartidas en el seno de la institución* I" (Ibíd., p. 337).

Las prácticas operativas y discursivas de las personas son las que, al resolver determinados tipos de problemas, dan lugar a los "saberes matemáticos". Con esto se asume en el EOS el postulado antropológico de la relatividad socioepistémica de los sistemas de prácticas, de los objetos emergentes de ellas y sus significados.

2.1.1.2 Objetos matemáticos

En las prácticas matemáticas que se realizan para resolver un campo de problemas, intervienen y emergen los objetos matemáticos que desde el EOS son vistos como entidades. Estos objetos matemáticos, que intervienen en la practica matemática, pueden ser ostensivos (símbolos, gráficos, etc.) y no ostensivos (conceptos, proposiciones, etc., que evocamos al hacer matemáticas) y que son representados en forma textual, oral, gráfica o incluso gestual (Godino y Batanero, 1994). De acuerdo con Godino y Batanero (1994, p. 338) "*El objeto institucional* O_I *es un emergente del sistema de prácticas sociales asociadas a un campo de problemas, esto es, un emergente de* $P_I(C)$*. Los elementos de este sistema son los indicadores empíricos de* O_I "; por consiguiente, si los objetos matemáticos emergen en un sistema de prácticas institucionales entonces se consideran objetos institucionales, y si emergen de un sistema de prácticas personales se consideran objetos personales, en palabras de los mismos autores "*El objeto personal* O_P *es un emergente del sistema de prácticas personales significativas asociadas a un campo de problemas, esto es, un emergente de* $P_P(C)$" (Ibíd., p. 339).

Entonces, de los sistemas de prácticas matemáticas operativas y discursivas emergen al menos seis nuevos objetos (o entidades primarias) que provienen de las mismas y dan cuenta de su organización y estructura: (Godino, Batanero y Font, 2007; 2019):

- Elementos lingüísticos: términos, expresiones, anotaciones, gráficas, etc. en sus diferentes registros (escritos, orales, etc.).
- Situaciones/problemas: aplicaciones matemáticas, ejercicios, etc.
- Conceptos/definiciones: introducidos por medio de definiciones o descripciones (punto, número, recta, media, función, etc.).

- Proposiciones/Propiedades: declaraciones sobre conceptos (postulados, teoremas, etc.).
- Procedimientos: operaciones, algoritmos, técnicas de cálculo, etc.
- Argumentos: discursos utilizados para validar y explicar proposiciones y procedimientos (deductivo, inductivo, etc).

Estos seis objetos matemáticos (tipología de objetos o entidades) que en el EOS se les denomina como *objetos matemáticos primarios* están relacionados entre sí formando redes denominadas configuraciones, es decir, los objetos matemáticos primarios interactúan para configurar la actividad matemática (ver Figura 2.2).

2.1.1.3 Significado de los objetos matemáticos

Cuando nos preguntamos ¿qué es el objeto Chi-cuadrada? ¿qué significa la media? ¿qué es la t-Student? Y en general, ¿Cuál es el significado de un objeto matemático concreto? Desde el EOS se propone como respuesta: Es el sistema de prácticas (operativas o discursivas) que realiza una persona (significado personal), o compartidas en el seno de una institución (significado institucional) para resolver un tipo de situaciones-problemas. Entonces, el significado de un objeto matemático se puede ver desde dos perspectivas, personal e institucional (ver Figura 2.1), Godino y Batanero (1994, p.340), definen al significado de un objeto institucional O_I como "*el sistema de prácticas institucionales asociadas al campo de problemas de las que emerge O_I en un momento dado*", asimismo, definen al significado de un objeto personal O_P como "*el sistema de prácticas personales de una persona p para resolver el campo de problemas del que emerge el objeto O_P en un momento dado*" (Ibíd., p. 341).

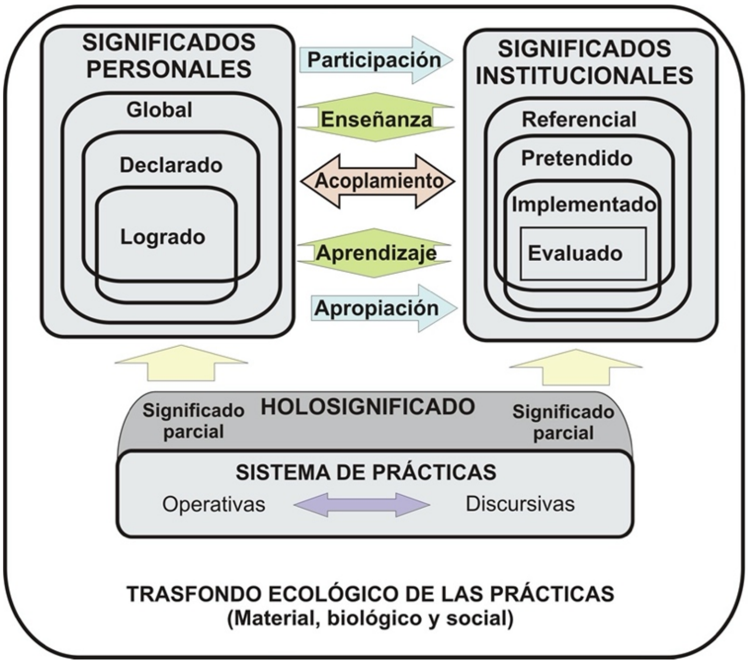


Figura 2.1. Tipología de significados (fuente: Godino, 2014, p. 13)

La noción de significado institucional permite el estudio de las prácticas donde emergen los objetos matemáticos, su evolución histórica y además atiende a los contextos. A este tipo de estudios se les denomina histórico-epistemológicos y por medio de ellos se determina el significado holístico o también denominado holosignificado. El significado holístico comprende varios significados parciales de un objeto matemático, mientras que los significados parciales están asociados a una configuración ontosemiótica (Pino-Fan, Godino y Font, 2011).

De este modo, es evidente que un determinado objeto matemático no constituye un único saber a ser enseñado, sino que cada objeto matemático, cada contenido, tiene múltiples significados, graduables en procesos de generalidad y formalización, cuya emergencia es progresiva como resultado de la actividad (práctica) de las personas en diferentes momentos para dar respuesta a situaciones-problemas diversos (Font, Godino y Gallardo, 2013). Cada concepto matemático es entendido de manera antropológica, pragmatista y sistémica, implicando diversos significados parciales o sentidos (e.g., Pino-Fan, et al., 2017), los cuales se relacionan de manera compleja y atendiendo a los contextos en los que son usados y marcos institucionales.

2.1.1.4 Configuración ontosemiótica

Para llevar a cabo el análisis de la actividad matemática es necesaria la elaboración de una tipología de objetos y procesos, y es la noción de configuración ontosemiótica quien "*responde a la necesidad de identificar los tipos de objetos y procesos que intervienen y emergen en las* *prácticas matemáticas que se realizan para la resolución de las situaciones-problemas*" (Godino, Batanero y Font, 2019, p. 39).

La noción de configuración ontosemiótica se ha utilizado desde dos perspectivas, la primera es la personal que refiere a una configuración ontosemiótica cognitiva, y la segunda es la institucional, la cual da paso a la configuración ontosemiótica epistémica. Podemos entender la configuración como una red de objetos matemáticos primarios que están relacionados entre sí (Figura 2.2), y será una configuración ontosemiótica epistémica si las redes de objetos matemáticos configuran una práctica institucional, mientras que si tales redes de objetos matemáticos configuran una práctica personal se trata de una configuración ontosemiótica cognitiva.

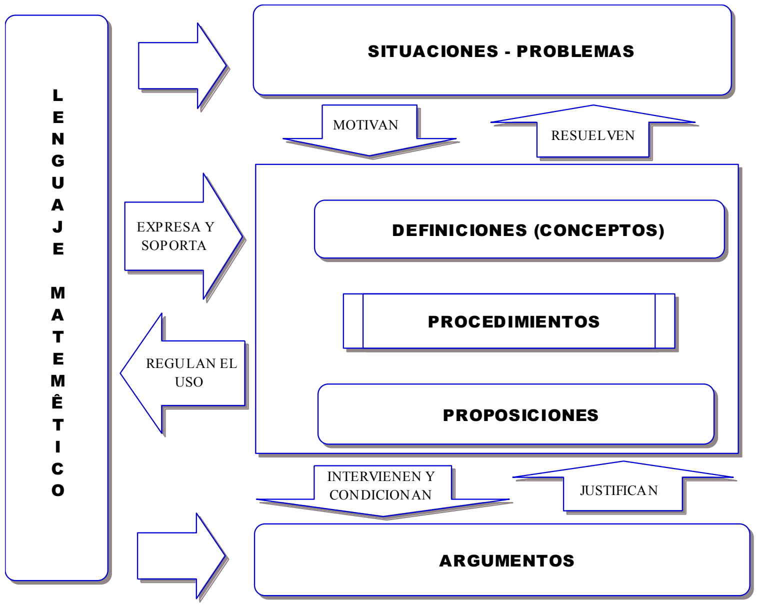


Figura 2.2. Configuración de objetos matemáticos primarios (fuente: Godino, 2008, p. 7)

Las situaciones-problemas son el origen o razón de ser de la actividad, el lenguaje permite representar las restantes entidades y sirve de instrumento para la acción; los argumentos justifican los procedimientos y proposiciones que relacionan los conceptos entre sí (Pino-Fan, 2013). Godino, Font, Wilhelmi y Lurduy (2011), señalan que estos objetos matemáticos primarios pueden ser analizados desde una perspectiva proceso-producto, lo que conlleva a considerar los siguientes procesos: comunicación (con el uso de los elementos lingüísticos), problematización (tipos de problemas), definición (de conceptos), enunciación (de proposiciones/propiedades), algoritmización (que permite la obtención del procedimiento) y argumentación. Otros procesos considerados en el EOS y que permiten entender la naturaleza compleja y progresiva de los objetos matemáticos son: generalización, particularización (ejemplificación), idealización (esquematización), materialización, representación, significación, reificación, descomposición y modelización (Font y Rubio, 2017; Pino-Fan, Godino y Font, 2015).

La Figura 2.3 da cuenta de los objetos matemáticos primarios, de las dualidades desde los cuales pueden ser vistos estos objetos, así como de los procesos que llevan asociados. En la misma figura, también podemos observar el papel central que tienen los problemas, las prácticas (operativas y discursivas) que se desarrollan entorno al problema y el contexto institucional donde se desenvuelven dichas prácticas.

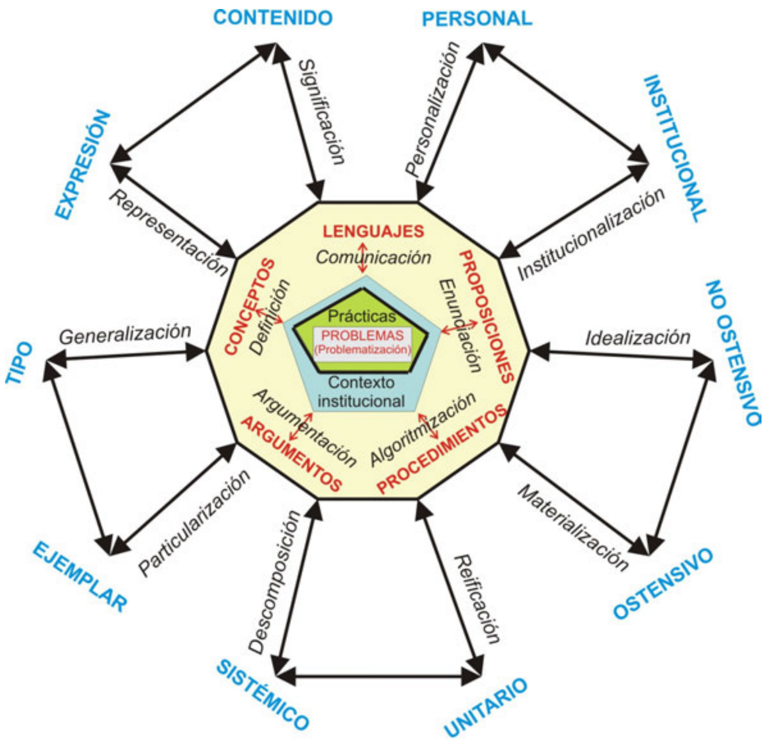


Figura 2.3. Objetos y procesos matemáticos (fuente: Godino, 2014, p. 23)

De acuerdo con Godino, Font, Wilhelmi y Lurduy (2011), la emergencia de los objetos matemáticos es un fenómeno complejo y que para ser explicado es necesario como mínimo dos niveles de objetos que emergen de la actividad. En el primer nivel tenemos, a los objetos matemáticos primarios que podemos observar en un texto matemático. Mientras que en un segundo nivel tenemos una tipología de objetos que emerge de las distintas formas de ver a los objetos matemáticos primarios (identificados en el primer nivel), con esto nos referimos a las dualidades o facetas duales (ver Figura 2.3) desde las cuales pueden ser vistos los objetos primarios que intervienen en las prácticas matemáticas:

Personal – Institucional

Como indicamos anteriormente, de acuerdo con Godino y Batanero (1994), si los sistemas de prácticas son compartidos dentro de una institución, entonces los objetos matemáticos que emergen se consideran *objetos institucionales*; en cambio, si se trata de sistemas de prácticas de un individuo, se entienden como *objetos personales*. Estas facetas duales tienen asociado un proceso, institucionalización y personalización respectivamente.

Ostensivo – No ostensivo

Cuando nos referimos a los objetos ostensivos anteriormente mencionamos que se pueden ver como símbolos, gráficos, etc., en otras palabras, son aquellos objetos que son públicos y que se pueden mostrar a otros. Y cuando nos referimos a los no ostensivos señalamos a los conceptos, proposiciones, etc., entonces, podríamos decir que son aquellos que no son visibles o perceptibles por sí mismos y son usados en la práctica por medio de sus ostensivos asociados (e.g., símbolos y gráficos).

Unitario – Sistémico

En determinado contexto los objetos matemáticos pueden ser tratados como una entidad unitaria, esto es debido a que son objetos ya conocidos. Mientras que, en otro contexto pueden ser introducidos como sistemas que se deben descomponer para su estudio. Usualmente este último caso sucede en los primeros cursos, cuando el estudiante se encuentra por primera vez con el objeto matemático.

Ejemplar – Tipo

Esta dualidad centra la atención entre lo particular (ejemplar) y lo general (tipo). Cuando en la práctica podemos observar la resolución de un problema con un ejemplo particular (para un caso) y por otra parte tenemos la resolución del mismo problema, pero en un sentido general, es decir, se cumple para todos los casos de este tipo de problema. "*La generalización es esencial porque este es el proceso que distingue la creatividad matemática de la conducta mecanizable o algorítmica*" (Otte, 2003, p. 187).

Expresión – Contenido

Los objetos matemáticos no se deben concebir como entidades aisladas sino como entidades que se relacionan. Dicha relación se establece por medio de funciones semióticas, entendidas como una relación entre un *antecedente* (expresión, significante) y un *consecuente* (contenido, significado) establecida por un sujeto (persona o institución) de acuerdo con un cierto criterio o código de correspondencia.

Para este estudio, utilizaremos, principalmente, la perspectiva epistémica para realizar el análisis desde el punto de vista histórico-epistemológico. La noción de configuración ontosemiótica epistémica ha sido empleada en investigaciones con la finalidad de caracterizar el significado holístico de diversas nociones matemáticas (e.g., Pino-Fan, Godino y Font, 2011; Crisóstomo, 2012; Rondero y Font, 2015; Gordillo y Pino-Fan, 2016; Pino-Fan, Font, Gordillo, Larios y Breda, 2017).

De acuerdo con Pino-Fan (2017), esta herramienta consiste en identificar y caracterizar los objetos matemáticos primarios y sus significados, a partir de las prácticas matemáticas que se han desarrollado en la historia, las cuales han permitido el surgimiento, evolución, formalización y generalización de las nociones de estudio.

Se debe localizar las problemáticas que resultaron claves para el desarrollo del objeto matemático y cómo es que se solucionaron; en otras palabras, las prácticas matemáticas desplegadas. Las mencionadas problemáticas y su práctica matemática, llevan asociada una configuración epistémica y ésta nos ayuda a determinar un significado parcial del objeto matemático. El conjunto de significados parciales identificados conforma el significado holístico de la noción bajo estudio.

2.1.2 El razonamiento en el enfoque ontosemiótico

En el Capítulo uno abordamos distintas concepciones, por una parte, sobre la alfabetización, razonamiento y pensamiento estadístico; y, por otra parte, sobre el razonamiento inferencial informal y formal. Aunque como pudimos observar, aún no existe un consenso en la literatura sobre el razonamiento estadístico ni sobre el inferencial. Entonces, teniendo las diversas concepciones como antecedente, nosotros partiremos de la postura que se tiene en el EOS sobre razonamiento para intentar dar una definición de razonamiento inferencial.

Aké (2013) plantea una concepción sobre el razonamiento algebraico basada, por una parte, en la literatura de investigación sobre la inclusión del algebra en la escuela primaria y la algebrización del currículo. Y, por otra parte, desde la perspectiva del EOS considera las herramientas que le permiten distinguir los objetos que tienen una naturaleza algebraica y en qué medida se pueden introducir en la escuela primaria.

Aké (2013, p. 99), considera las expresiones "*razonamiento algebraico, sentido algebraico y pensamiento algebraico como perspectivas equivalentes del mismo objeto álgebra escolar desde un enfoque transdisciplinar*". También, asocia al razonamiento algebraico una perspectiva epistemológica, al sentido algebraico una perspectiva semiótica y al pensamiento algebraico una perspectiva cognitiva; sin embrago, sugiere que todas ellas se relacionan en la visión del algebra escolar. Así pues, la caracterización del razonamiento algebraico elemental, que propone, se realiza en términos de los tipos de tareas, objetos y procesos algebraicos implicados en la práctica.

En el estudio de las normas que influyen en los procesos de argumentación de Molina (2019), establece algunas acepciones que han tenido la prueba y la argumentación como tipos de razonamiento. Por una parte, para argumentación, indica que para Duval es un tipo de razonamiento no deductivo que tiene reglas implícitas provenientes tanto de la estructura del lenguaje como del conocimiento de quien realiza el argumento. Mientras que para Krummheuer y Yackel tiene una dimensión social y es una explicación intencional del razonamiento de una solución. Por otra parte, para la prueba, en el escenario donde se le vincula con un proceso argumentativo que se utiliza para justificar la verdad de los procesos matemáticos, menciona que en las ultimas décadas la concepción de prueba con carácter formal ha ido variando, esto debido a que tanto matemáticos como educadores matemáticos han otorgado un lugar más significativo al razonamiento por deducción y las respectivas pruebas formales. También plantea que, reconociendo la realidad de la práctica matemática, las pruebas pueden tener varios grados de validez formal y con el mismo nivel de aceptación. Finalmente, expone que si entendemos los términos prueba y probar como un proceso, nos referíamos a un proceso psicológico de razonamiento o un proceso social, mientras que, si se entiende probar como un objeto, su acepción dentro del marco de un proceso psicológico de razonamiento, se refiere entonces a un razonamiento estrictamente deductivo. Entonces podríamos decir que accedemos al razonamiento por medio de los objetos y procesos matemáticos involucrados o que se encuentran presentes en el argumento que realiza un sujeto al resolver un problema.

Con base en lo anterior, podríamos decir que, en el EOS, el razonamiento se asume como un "macro proceso social y epistémico", que involucra poner en juego tanto los objetos matemáticos primarios como los procesos matemáticos (generalización, particularización –ejemplificación–, idealización –esquematización–, materialización, representación, significación, reificación, descomposición, modelización), para la solución de una situación-problema (e.g., Molina, 2019; Aké, 2013). En otras palabras, que para decir que un sujeto "comprende" la χ^2 debemos observar que en su razonamiento asociado a sus prácticas (para resolver distintos tipos de situaciones/problemas), emerjan de manera gradual, sistemática y progresiva, objetos matemáticos primarios y procesos vinculados a los significados de esta noción (Lugo-Armenta y Pino-Fan, en prensa).

Cabe señalar que la posición sobre 'razonamiento' que se considera en nuestra propuesta, recurre a una visión pragmatista sobre la construcción del conocimiento matemático (así como el conocimiento matemático escolar), que contempla los postulados semióticos, antropológicos y pragmáticos del EOS (Godino, Batanero y Font, 2017; 2019). Es decir, en concordancia con los señalamientos de Aké (2013) para el álgebra, y Molina (2019) para la geometría, para hablar de estadística escolar se debe recurrir a una visión integrada y transdisciplinar, que involucra pensamiento estadístico (Statistical thinking), razonamiento estadístico (Statistical reasoning) y alfabetización estadística (Statistical literacy), pues cada uno de estos enfoques de la estadística escolar se desarrolla desde las perspectivas psicológica, epistémica y semiótica, respectivamente. Así, al considerar el razonamiento en términos de prácticas, objetos (matemáticos primarios) y procesos matemáticos utilizados en ellas, se transita en cierto sentido por las definiciones dadas por Ben-Zvi y Garfield (2004) para statistical literacy, statistical reasoning y statistical thinking.

2.2 Pregunta y objetivos de investigación

Una vez que hemos introducido las nociones y herramientas teóricas y hemos generado el contexto dentro del cual se enmarca esta investigación, y retomando la problemática esbozada en el Capítulo 1, estamos en condiciones de replantear la pregunta de investigación y formular los objetivos del estudio.

Pregunta de investigación:

> *¿Cómo los significados parciales de los estadísticos Chi-cuadrada y t-Student, y sus articulaciones, pueden contribuir en la propuesta de niveles progresivos de razonamiento inferencial, de lo informal (RII) a lo formal (RIF), sobre dichos estadísticos?*

Objetivo General de investigación:

> *Caracterizar niveles progresivos de razonamiento inferencial (de lo informal a lo formal) sobre los estadísticos Chi-cuadrada y t-Student, a partir de la riqueza matemática subyacente a los significados parciales que se le han conferido a tales estadísticos en su evolución histórico-epistemológica.*

Para logar el objetivo general de la investigación y así dar respuesta a la pregunta de investigación, es necesario plantearnos una serie de objetivos específicos (OE):

OE1: Reconstruir el significado holístico de los estadísticos Chi-cuadrada y t-Student, con base en un estudio histórico-epistemológico, determinando los pares <Prácticas matemáticas/estadísticas, Configuraciones de objetos matemáticos primarios>.

OE2: Establecer las posibles relaciones entre los significados parciales, esto es entre los pares <Prácticas matemáticas/estadísticas, Configuraciones de objetos matemáticos primarios>, de los estadísticos Chi-cuadrada y t-Student, para determinar los elementos que podrían constituir una propuesta de niveles de razonamiento inferencial

OE3: Determinar niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student, con base en la postura adoptada en este estudio sobre el razonamiento inferencial (informal y formal), y las relaciones entre los significados parciales de ambos estadísticos y los aportes sobre el tema de la literatura científica.

OE4: Caracterizar las prácticas matemáticas de profesores y futuros profesores de educación media, determinando los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student que movilizan con tales prácticas.

2.3 Metodología

La perspectiva metodológica dentro de la cual se desarrolla esta investigación es de carácter cualitativa (Cohen, Manion y Morrison, 2011; Hernández, Fernández y Baptista, 2014) ya que se realizó un estudio de tipo histórico-epistemológico, con el uso de las nociones del EOS descritas anteriormente, a partir del cual se identificaron una serie de características o elementos epistémicos que permitieron, junto con las investigaciones de educación estadística, construir niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. A partir de los niveles para estos estadísticos, se realizó una propuesta de niveles de razonamiento inferencial, de lo informal a lo formal, en términos más generales. Además, se busca describir y caracterizar el razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student de profesores en formación y en ejercicio de educación media, mediante el diseño y la aplicación de un instrumento de indagación que permita observar las prácticas matemáticas/estadísticas que realizan los profesores. Dichas prácticas se caracterizarán mediante los niveles progresivos de razonamiento inferencial sobre tales estadísticos.

La investigación cualitativa de acuerdo con Hernández, Fernández y Baptista (2014) "*nos proporciona profundidad a los datos, dispersión, riqueza interpretativa, contextualización del ambiente o entorno, detalles y experiencias únicas*" (p. 16), por medio de la comprención del fenomeno de estudio desde el entorno donde se desarrolla naturalmente y desde la perspectiva de los sujetos involucrados.

De acuerdo con las características de nuestro estudio y como hemos señalado, la metodología a seguir es de tipo cualitativa y a continuación se detallan las fases de esta investigación.

2.3.1 Fases de la investigación

Para alcanzar los objetivos específicos planteados anteriormente, nos hemos propuesto las siguientes fases y actividades de investigación:

Fase 1: Estudio histórico-epistemológico para determinar los significados parciales y el significado holístico de los estadísticos Chi-cuadrada y t-Student.

- Revisar y analizar las fuentes primarias principalmente, tales como artículos originales dónde surgen las propuestas sobre estos estadísticos como los de Galton, Pearson, Fisher, Gosset y Yates, y fuentes secundarias (artículos de interpretación sobre estos estadísticos o libros de historia), donde se identifican las grandes problemáticas que resultaron claves para el desarrollo del objeto matemático y cómo estas se solucionaron.
- Describir las mencionadas problemáticas y las prácticas matemáticas que se realizaron para solucionarlas. Esto con la noción de práctica matemática del EOS.
- Identificar y caracterizar los objetos matemáticos primarios mediante la noción de configuración ontosemiótica (epistémica), cada configuración nos ayuda a determinar los significados parciales que conforman el significado global para los estadísticos Chi-

cuadrada y t-Student.

Fase 2: Vinculación de los resultados de la Fase 1 con la revisión de literatura científica sobre los estadísticos Chi-cuadrada y t-Student, el RII y RIF.

- Analizar e identificar las relaciones entre los significados de los estadísticos Chi-cuadrada y t-Student en las configuraciones ontosemióticas de estos estadísticos.
- Revisar y analizar la literatura científica sobre RII, RIF y sobre los estadísticos Chi-cuadrada y t-Student para identificar diferentes posturas teóricas sobre el RII y RIF, así como los errores y dificultades, y propuestas de actividades o de enseñanza que se han reportado sobre inferencia estadística y también sobre dichos estadísticos.
- Brindar una postura sobre *Razonamiento* en términos del EOS, así como qué se entiende, en esta investigación, por un razonamiento inferencial informal, preformal y formal.

Fase 3: Propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, con base en los resultados del estudio de los significados de estos estadísticos y lo que reporta la literatura (i.e., resultados de Fase 1 y Fase 2).

- Identificar, a partir de las configuraciones epistémicas de los estadísticos Chi-cuadrada y t-Student, los elementos o características para una transición continua del RII al RIF, teniendo como referente la postura sobre razonamiento basada en el EOS.
- Construir los niveles de razonamiento inferencial sobre los estadísticos por medio de criterios, los cuales vinculan elementos de los significados y de la literatura de investigación, que den cuenta de una progresión de lo informal a lo formal, pasando por lo preformal, de este tipo de razonamiento.

Fase 4: Análisis de las prácticas matemáticas realizadas por futuros profesores y profesores de educación media al resolver actividades que involucran a los estadísticos Chi-cuadrada y t-Student.

- Diseñar, validar y aplicar actividades que se pueden resolver con cualquier nivel de razonamiento inferencial, tanto para el estadístico Chi-cuadrada como para el estadístico t-Student.
- Describir la práctica matemática que realizan los profesores para resolver las

actividades.

- Identificar los objetos matemáticos primarios que emergen en la práctica matemática de los profesores.
- Determinar el nivel de razonamiento inferencial que se encuentra asociado a la práctica matemática que despliegan los profesores para resolver cada actividad.
- Determinar la correspondencia entre los niveles progresivos de razonamiento inferencial (resultado de Fase3) y las prácticas matemáticas (y sus características) observadas.
- Realizar una propuesta general de niveles de razonamiento inferencial a partir de la identificación de conexiones entre los niveles de los estadísticos Chi-cuadrada y t-Student, y de los resultados del estudio experimental con profesores y futuros profesores.

2.3.2 Sujetos de estudio

Los sujetos que participaron en la fase 4 de este estudio se conforman por cuatro grupos de profesores. El primer grupo (G1), corresponde a 28 profesores en formación de una universidad mexicana que se encontraban tomando su primer curso de probabilidad y estadística, aunque aún no habían iniciado con tópicos sobre inferencia. Como parte del curso, los profesores en formación tomaron un taller de razonamiento estadístico con una duración de dos semanas, el cual se llevó a cabo de forma virtual y con modalidad sincrónica (utilizando la plataforma Zoom). Durante el taller los profesores en formación resolvieron en equipos actividades sobre el estadístico Chi-cuadrada y discutieron sus soluciones con sus pares y el formador del taller. La conformación de los equipos fue de forma aleatoria. Las sesiones sincrónicas fueron grabadas y tuvieron una duración de 12 horas.

El segundo grupo que participó (G2), fue de 22 profesores en formación de diversas universidades de Costa Rica y previamente habían llevado los cursos de probabilidad y estadística de sus universidades. Estos profesores en formación se inscribieron en un taller de razonamiento estadístico, con duración de dos semanas de forma virtual y con modalidad asincrónica. En este taller los participantes resolvieron actividades sobre los estadísticos Chi-cuadrada y t-Student de forma individual, por la modalidad del taller, y tuvieron oportunidad de interactuar con sus pares y con el formador por medio de un foro y videos. La información se recolectó por medio de una plataforma tipo Moodle provista por la universidad que organizó el taller (y que invitó a dictar el taller a la autora de esta tesis), y concretamente se consideraron los foros, los chats, y los documentos escritos (Word, Excel y PDF) que enviaron los estudiantes como evidencia del desarrollo de las actividades.

El tercer grupo (G3) está conformado por 59 profesores en ejercicio que se inscribieron a un taller de razonamiento estadístico organizado por una universidad chilena y dirigido a profesores de enseñanza media de matemáticas de Chile. Los profesores en ejercicio eran de distintos colegios (municipales, subvencionados y particulares) con distintas características, en su mayoría, de una región del sur de Chile, y algunos de otras regiones.

Por su parte, el cuarto grupo (G4) estaba conformado por 41 profesores en ejercicio de distintas nacionalidades de Latinoamérica, concretamente de Argentina, Chile, Colombia, Guatemala, México y Perú. Ellos se inscribieron a un taller sobre razonamiento estadístico organizado por una universidad chilena y dirigido a profesores de matemáticas de enseñanza media.

Los talleres para los profesores en ejercicio (G3 y G4) tuvieron una duración de una semana y se llevaron a cabo de forma virtual con modalidad sincrónica y asincrónica. Los profesores en ejercicio, del grupo 3 y 4, tuvieron la oportunidad de resolver actividades sobre la t-Student y Chi-cuadrada, respectivamente, y discutir sus soluciones con sus pares y los formadores que impartieron el taller (la autora de esta tesis y el académico que dirige esta tesis). Las actividades propuestas en los talleres se resolvieron en equipos y la conformación de los equipos fue de forma aleatoria. Para el desarrollo de las sesiones sincrónicas se utilizó la plataforma Zoom, y dichas sesiones fueron grabadas para su análisis posterior; además, mediante la plataforma tipo Moodle, utilizada para la modalidad asincrónica, los profesores enviaron documentos escritos (Word, Excel y PDF) con el desarrollo de las actividades.

En la Tabla 2.1 se resume la información sobre el número de participantes con cada estadístico.

Tabla 2.1. *Distribución de los participantes en el estudio*

	Chi-cuadrada	**t-Student**
Profesores en formación (PF)	28 PF mexicanos 22 PF costarricenses	 22 PF costarricenses
Profesores en ejercicio (PE)	41 PE latinoamericanos	59 PE chilenos

En el capítulo 6, se brindarán mayores detalles sobre la participación de los sujetos de estudio a medida que demos los pormenores del instrumento de indagación y analicemos las prácticas matemáticas desarrolladas para resolver las actividades.

2.3.3 Instrumentos de indagación

Para caracterizar las prácticas matemáticas que realizan los profesores participantes sobre los estadísticos Chi-cuadrada y t-Student (fase 4), hemos considerado el diseño de actividades (situaciones-problemas), sobre los estos estadísticos, que admitan distintas prácticas matemáticas para resolver la actividad.

2.3.3.1 Criterios para el diseño

Un primer criterio para el diseño de las actividades es la representatividad de los niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student; esto implica, que las actividades se pueden resolver con prácticas que contienen elementos intuitivos, preformales o formales.

El segundo criterio para el diseño es la representatividad de los significados de los estadísticos Chi-cuadrada y t-Student. Para ello, se han diseñado actividades que atiendan a cada significado de estos estadísticos. Finalmente, se consideraron aspectos más generales como el uso de contextos accesibles e interesantes y uso de diversas representaciones (lenguaje natural, tablas, figuras).

En el Capítulo 6, se muestran mayores detalles sobre los criterios que se consideraron para el diseño de las actividades que conforman el instrumento de indagación, esto se debe a que es necesario primero desarrollar el estudio histórico-epistemológico (Capítulos 3 y 4) y las propuestas de niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student (Capítulo 5).

2.3.3.2 Validación

Aunque no se trata de un estudio cuantitativo, podemos hablar de validez y confiabilidad realizando una analogía para el paradigma cualitativo; algunos autores como Cohen, Manion y Morrison (2011) nos han brindado algunas pautas. Es importante destacar que lo que nos interesa es la capacidad del instrumento para explicar el fenómeno de estudio en profundidad y no la generalización.

Para la validación del instrumento consideramos la triangulación metodológica y la validación por contenido; la triangulación metodológica usa el mismo método en diferentes ocasiones, o diferentes métodos sobre el mismo objeto de estudio (Cohen, Manion y Morrison, 2011), mientras que podemos hablar de validación por contenido cuando mostramos que el instrumento cubre de manera justa y completa el dominio o los elementos que se propone cubrir (Ibíd, 2011).

Para hablar de confiabilidad en investigación cualitativa se han utilizado diversos términos como credibilidad, neutralidad, confirmabilidad, consistencia, aplicabilidad, fiabilidad, integridad, entre otros, dependiendo de la postura del autor que lo aborda; nosotros utilizaremos el término confiabilidad.

Mientras que los métodos cuantitativos requieren un grado de control y manipulación del fenómeno, en los métodos cualitativos cobra relevancia lo único y la idiosincrasia de las situaciones, por esta característica el estudio no podría ser replicado en un método cualitativo (lo cual se considera una debilidad de los estudios cualitativos). Sin embargo, Denzin y Lincoln (1994) sugieren que la confiabilidad como replicabilidad en la metodología cualitativa podría abordarse desde diversos caminos, como la estabilidad de las observaciones, formas paralelas, confiabilidad inter-evaluador.

Por una parte, la estabilidad de las observaciones se refiere a si el investigador pudiera realizar las mismas observaciones e interpretaciones, si realiza observaciones en diferentes momentos o en diferentes lugares. Por otra parte, la confiabilidad inter-evaluador se da cuando otro investigador con el mismo marco teórico y que observa el mismo fenómeno puede realizar las mismas interpretaciones.

En el capítulo 6, cuando presentemos las actividades del instrumento de indagación y mostremos el tipo de prácticas que se podrían realizar para dar solución a las actividades, podremos valorar la validez del instrumento. Mientras que, cuando analicemos las prácticas matemáticas desarrolladas por los profesores para resolver las actividades sobre los estadísticos Chi-cuadrada y t-Student, podremos referirnos con mayor detalle a la confiabilidad en el sentido de la estabilidad de las observaciones y también de la confiabilidad inter-evaluador.

2.4 Reflexiones finales

En este capítulo, presentamos las nociones teóricas del Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos que consideramos para el desarrollo de nuestra investigación. Estas nociones teóricas nos ayudaron a reformular las preguntas que habíamos planteado en la aproximación al problema de investigación (Capítulo 1) en la pregunta y objetivos de investigación.

En la metodología destacamos como aspectos relevantes: (1) las fases de la investigación, donde podemos prever la operativización de las nociones teóricas y cómo cada una de las acciones de las fases contribuirán a la consecución de los objetivos específicos de investigación y a dar respuesta a la pregunta de investigación; (2) los sujetos de estudio que participarán en nuestra investigación (en la fase 4); (3) el instrumento de investigación, y criterios relevantes considerados para su diseño. En el capítulo 6 brindaremos más detalles sobre los criterios del diseño y la validez del instrumento.

CAPÍTULO 3

El Estadístico Chi-cuadrada (χ^2).

Reconstruyendo el Significado Holístico de Referencia

"Un estadístico es un valor calculado a partir de una observación muestral con intención a caracterizar la población de la que se extrae". (Fisher, 1934, p. 41)

INTRODUCCIÓN

¿Por qué el estudio de los significados?

Al abordar la investigación de los estadísticos Chi-cuadrada y t de Student, surge de manera natural preguntarse ¿qué es el estadístico Chi-cuadrada?, y ¿qué es el estadístico t de Student? Para responder estas preguntas se realiza un estudio de los significados de referencia de dichos estadísticos.

El estudio de significados de referencia puede ser abordado desde dos perspectivas, la primera refiere a la indagación de los currículos oficiales del Ministerio de Educación, entendido como la dupla de planes y programas y libros de texto. La segunda perspectiva trata de un estudio de tipo histórico-epistemológico, donde se recurre principalmente a las fuentes primarias.

Para estudiar los significados hemos recurrido al de tipo histórico-epistemológico que da cuenta del origen y evolución de los estadísticos Chi-cuadrada y t de Student, lo cual nos permite realizar la reconstrucción del significado holístico de referencia.

Conocer el significado holístico de referencia de los objetos matemáticos es de suma importancia, debido a que a partir de ello la institución y/o el profesor determina cuáles significados se pretenden, implementan y evalúan en el aula de clases.

Estudio histórico-epistemológico sobre el estadístico χ^2

En esta sección se encuentran (1) la descripción de las prácticas matemáticas que se desarrollaron en la historia para resolver las grandes problemáticas sobre el estadístico Chi-cuadrada, las cuales dan paso a los significados. (2) A partir de dichas prácticas matemáticas se identificaron doce significados parciales para el estadístico Chi-cuadrada y finalmente, (3) una síntesis de los significados del estadístico Chi-cuadrada.

3.1 Análisis de las prácticas históricas sobre el estadístico χ^2

El análisis de las prácticas matemáticas corresponde, de acuerdo con el EOS, al primer nivel de análisis de la actividad matemática. Y por medio de este análisis se han identificado tres campos de problemas o grandes problemáticas en el surgimiento y evolución del estadístico χ^2, que corresponden a las pruebas de: bondad de ajuste, independencia y homogeneidad. Adicionalmente, se considera la importancia que en estas pruebas tiene la distribución χ^2 como la distribución asintótica del estadístico χ^2.

3.1.1 Los trabajos previos al estadístico Chi-cuadrada

Las principales contribuciones para la construcción de una metodología empírica y conceptual que proveyera un sustito para el control experimental fueron de Francis Galton, Ysidro Edgeworth y Karl Pearson; sus trabajos iniciaron en la década de los 80's del siglo XIX y marcaron un notable cambio.

A Galton se le considera "*el hombre de la idea, un pensador altamente imaginativo con considerable energía y una curiosidad direccionadora*" (Stigler, 1986, p. 266). Para el desarrollo matemático de las ideas de Galton, Edgeworth, que era un teorista, tuvo una contribución importante debido a que tomó algunas de las ideas de Galton y las traslado a una forma matemática generalizable. Sin embargo, Edgeworth fue incapaz de llegar a una audiencia general, además no construyó una base empírica para probar o mostrar el valor de su método. Posteriormente, Pearson reconoció la importancia de las formulaciones que había realizado

Edgeworth sobre las ideas de Galton, a partir de las cuales realizó desarrollos y generó una metodología. Se profundizará en algunas de las contribuciones de Pearson en el apartado siguiente.

Francis Galton estudió medicina en Cambridge. Cuando dejó la profesión de médico emprendió una expedición en solitario por el sur de África (1850 - 1852), realizando producciones cartográficas, después se interesó por las estaciones climatológicas de Europa y creó intrincados mapas del clima que incluían simbología sobre la dirección del viento, temperatura, presión barométrica de cada estación, además a partir de una serie de observaciones usó métodos gráficos para el análisis de datos multivariados. Posteriormente se interesó en la psicología, antropología, sociología, educación y las huellas dactilares, pero lo que dominó su trabajo fue el estudio de la herencia que inició en 1865, donde destaca la naturaleza estadística que tenía. De este trabajo resaltan más que sus cálculos y uso de tablas, los análisis y las aproximaciones conceptuales que realizaba.

Galton fue discípulo de Quetelet, pero con su trabajo 'Hereditary Genius' de 1869 se separa de él en grandes aspectos. En este trabajo Galton intentó demostrar que el talento se hereda, debido a que hombres eminentes con frecuencia tienen parientes eminentes. Con respecto a esta obra, su importancia no únicamente recae en la adaptación que realizó Galton de la metodología de Quetelet, sino en los primeros avistamientos de la idea de regresión, la cual desarrolló durante los veinte años posteriores (Stigler, 1986).

Galton estaba fascinado por la ley teórica de la desviación del promedio[1] en múltiples casos y siguiendo a Quetelet[2], quien fue el primero en aplicar esta ley a mediciones humanas, propuso un test para una clasificación adecuada de los datos en un solo grupo, en donde la no aparición de la curva normal era indicativo de que los datos no debían tratarse juntos. En los trabajos de Quetelet sobre el hombre tipo o medio se consideraba que en un pueblo existe un hombre tipo que representa a dicho pueblo por una característica como su estatura y los demás hombres de ese pueblo se consideran en relación con él como desviaciones, traduciendo al promedio en un ente ideal en las mediciones de los humanos por medio de trasladar la idea del promedio, de la

[1] Segunda ley del error de Laplace o ley de Gauss también conocida como ley de frecuencia del error en la que se expresa la relación que existe entre la magnitud de un error y la frecuencia con la que se cometerá ese error cuando se realizan un gran número de mediciones cuidadosas.
[2] Quetelet realizó tablas para facilitar los cálculos cuando se recurra a la ley de frecuencias del error. Estas tablas son parecidas a las actuales de la distribución normal estándar.

curva de los errores, como un concepto que indicaba magnitud física. Posteriormente, Galton reconoció que aún un grupo de datos que provienen de la misma especie, por ejemplo, los habitantes de las islas británicas, podían agruparse sólo al inicio del análisis para posteriormente investigar las diferencias entre los subgrupos y no verlos a todos como un conjunto homogéneo[3]. Y que donde "*Quetelet había utilizado la aparición de una curva normal para demostrar la homogeneidad, Galton basó una disección de la población en ella. Usando una escala inferida, pudo distinguir entre las habilidades de los hombres en una escala numérica en vez de decir que eran indistinguibles*" (Stigler, 1986, p. 271). Galton pudo reconocer la curva normal en los datos con los que trabajaba, sin embargo, no logró conectar la curva con la transmisión de habilidades de generación en generación. No obstante, este problema le llevaría a los inicios de los trabajos de regresión y correlación.

En 1889, en su libro 'Natural Inheritance', los desarrollos de Galton sobre el concepto de regresión y sus bases probabilísticas alcanzan su máximo progreso. Stigler (1986) distingue cuatro etapas. La primera es su investigación inicial en 1874 sobre las condiciones que producirían la ley del error. Galton se interesó en la naturaleza de esta ley, partiendo de las condiciones clásicas de Laplace que resultaban muy restrictivas para el uso que les estaba dando Galton, por lo que tuvo que demostrar que las condiciones de Laplace se encontraban lejos de ser necesarias para la curva normal. Y en 1875 presentó un método de ajuste (scaling o intercomparación) que podía producir la distribución observada, para lo que utilizó la curva normal y proporcionó las condiciones que podrían implicar sí realmente es apropiada la curva normal. Galton explicó su método de intercomparación de la siguiente manera:

> Considerando la importancia de los resultados que admiten ser derivados siempre que se pueda demostrar que la ley de frecuencia del error se aplica, yo daré algunas razones por las que su aplicabilidad es más general de lo que se podría haber esperado de las hipótesis altamente artificiales sobre las cuales se basó la ley. Se recordará que estos son el efecto de los errores individuales de la observación o las diferencias individuales en objetos pertenecientes al mismo grupo genérico, se deben totalmente a la acción agregada de influencias variables en diferentes combinaciones, y estas influencias deben ser (1) todas independientes en sus efectos, (2) todas iguales, (3) todas

[3] Las diferencias individuales de deben a influencias y éstas admiten ser tratadas como alternativas sobre o debajo del promedio.

> admitiendo ser tratadas como simples alternativas 'sobre el promedio' o 'debajo del promedio'; y (4) las tablas usuales se calculan sobre la suposición adicional de que las influencias variables son infinitamente numerosas. (Galton, 1875, p. 38)

La idea fue mostrar los datos de una manera diferente, Galton los ordenó de forma ascendente y graficó los valores de los datos versus los rangos (cuartiles); la Figura 3.1 muestra la curva ideal, si los datos fueran homogéneos, a la cual Galton llamó ojiva, pero actualmente la conocemos como la función de distribución acumulativa de la normal inversa.

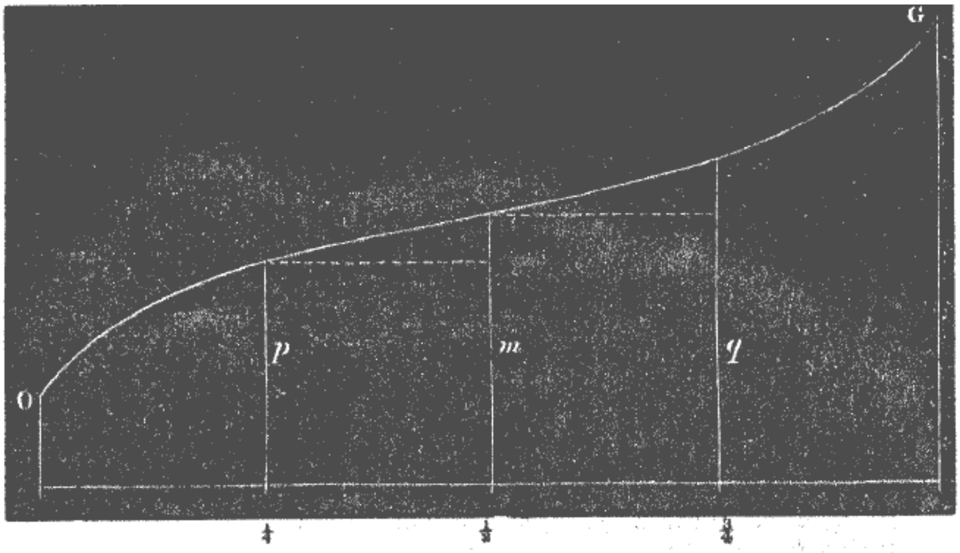


Figura 3.1. Primera concepción de Galton sobre la ojiva, muestra la mediana y cuartiles (fuente: Galton, 1875 p. 36)

En la Figura 3.1, muestra el proceso de estadística por intercomparación, donde si "$n =$ *longitud de la base de la ojiva, cuya ordenada* y *representa la magintud del objeto que se encuentra a una distancia* x *desde el final de la base donde las ordenadas son más pequeñas*" (Galton, 1875, p. 36). Entonces m en $\frac{1}{2}$ representa el valor medio de la serie (la mediana), p en $\frac{1}{4}$ y q en $\frac{3}{4}$ brindan información para estimar la divergencia al relacionarlos con m, por lo que $q - m$ es la divergencia o el error probable (de la parte de la serie que excede a la mediana), mientras que $m - p$ corresponde a la divergencia de la otra parte de la serie. De acuerdo con Galton (1875) cuando la serie es simétrica $q - m = m - p$.

En 1875, Galton trabajó en sentido opuesto de lo establecido en la ley de frecuencia del error, expresando que se puede decir, "*dado que tales y tales magnitudes ocurren con tales y tales grados de frecuencia, por lo tanto, las diferencias entre ellas y el valor medio son tal y tal como las expresadas en unidades del error probable*" (1875, p. 38). Entonces, de acuerdo con lo que estableció, la primera división de la escala de divergencia corresponde a los valores que se encuentran a una unidad de error probable por encima o debajo del valor medio, estos serían p en $\frac{1}{4}$ y q en $\frac{3}{4}$. Mientras que la segunda división serían aquellos valores a dos unidades de error probable del valor medio y de forma análoga en la tercera división.

De acuerdo con Galton (1875), para distinguir entre los procesos que realizó, en el primero se denominó binomial y realizó los cálculos bajo un número considerable de elementos que denominó (r) y cuya divergencia es expresada por los coeficientes de la expansión de la binomial $(a+b)^r$, mientras que el segundo proceso lo llamó exponencial y se llevó a cabo bajo el supuesto de que el número es infinito y fue expresado por la exponencial $e^{-\frac{x^2}{c^2}}$. En la Figura 3.2 se puede apreciar los resultados prolongados de ambos procesos, donde la unidad de medida vertical en el caso de una serie de grado binomial es de un solo grado y la unidad de la curva exponencial será $q-m$, o el error probable, y Galton señaló que está unidad es igualmente aplicable a la curva de la binomial.

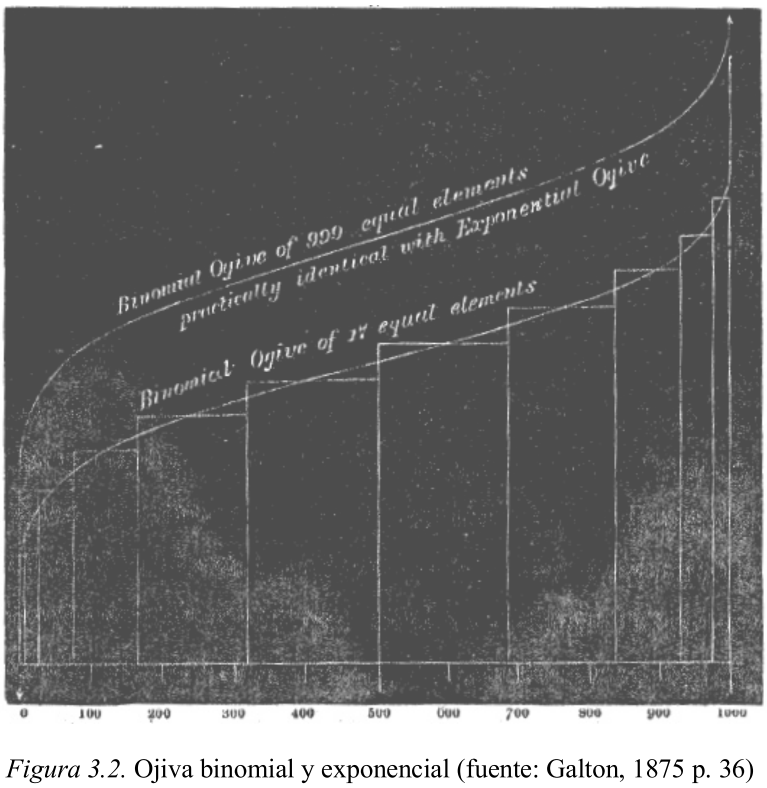


Figura 3.2. Ojiva binomial y exponencial (fuente: Galton, 1875 p. 36)

Galton (1875), respecto a la figura anterior, señaló que se puede observar cuánto se asemeja la ojiva de la exponencial a la de la binomial.

Para ilustrar el método de intercomparación, Galton recurrió a lo que ocurre en la naturaleza y consideró las causas que determinan el tamaño de una fruta, señaló que existen factores de gran importancia y otros de menor y que cuanto mayor sea su influencia se convierte en un elemento primario o de primer orden en los cálculos, como el 'pedigree' del árbol o la calidad del suelo donde se encuentra el árbol.

Galton retoma la Figura 3.2, señalando que en ella se expresa una serie de 1000 objetos ordenados de acuerdo a sus magnitudes y supone que se deben a varias combinaciones de 17 alternativas (que se encuentran en la ojiva de la binomial), entonces los términos que se graficaron se obtuvieron mediante varios términos de la expansión de $(1+1)^{17}$, todo multiplicado a $\frac{1000}{2^{17}}$, formando la siguiente serie $=$ $1000{:}-0,0,1,5,18,47,95,148,186,186,148,95,47,18,5,1,0,0$. En la ojiva se puede encontrar p y q, el valor de $q-m$ o $m-q$ y también $m+2(q-m)$ y $m+3(q-m)$. Galton (1875), consideró que es lo más lejos que la escala de la figura admite y que las tablas de la ley de frecuencia del error le dirían dónde poner los otros puntos para finalizar con la curva. También señaló que encontró la posición de varios puntos en las dos ojivas, mediciones apropiadas de la base, que se expresan en la siguiente tabla:

Tabla 3.1. *Posición de las mediciones en ambas ojivas* (fuente: Galton, 1875, p. 42)

	En la ojiva binomial de 17 elementos	**En la ojiva exponencial o en la ojiva binomial de 1000 elementos**
Valor medio	500	500
Valor medio $\pm$ 1 unidad	250	250
Valor medio $\pm$ 2 unidades	71	82
Valor medio $\pm$ 3 unidades	16	17

Al obtener estos valores para ambas ojivas, Galton (1875, p. 42) consideró que "*la cercanía de la semejanza es sorprendente*". Posteriormente procedió a mostrar cómo un conjunto o mezcla de pequeñas causas podrían verse como un número moderado de influencias pequeñas e iguales. Propuso también algunas aproximaciones para trabajar con dichas influencias, que pueden ser tratadas como simples alternativas mayores o menores al valor medio, para lo que recurrió a los lanzamientos de una moneda, señalando que "*la ojiva exponencial es, a primera vista, falaz en una gran cantidad de casos, y que podemos aprender cuál es el mayor número* *posible de elementos en la binomial cuya ojiva más cercana representa la serie genérica*" (Ibíd., p. 43). Asimismo, añadió que el valor de $\frac{m}{q-m}$ es directamente dependiente del número de elementos, observó las diferencias entre series simples y series compuestas de 1024 eventos, teniendo entre ellas una correspondencia muy cercana.

Por consiguiente, concluyó que "*mientras cada serie estadística puede ser juzgada de acuerdo a sus peculiaridades, la ley de frecuencia del error basado en una ojiva binomial es mucho más probable que sea aproximadamente cierta que cualquier otra que pueda especificarse a priori*" (Galton, 1875, p. 46) y que si se adopta un sistema como el que sugirió,

> las magnitudes de cualidades para la medición de partes iguales de las cuales no existe una escala, dicho sistema puede razonablemente basarse en una aplicación inversa de la ley del error de frecuencia, de la manera que he descrito, a series estadísticas obtenidas por el proceso de intercomparación (Ibíd., p. 46).

Cuando los datos graficados bajo el procedimiento de intercomparación se asemejan a ojiva significa que los datos se distribuyen normalmente, dejando claro que existe variación incluso en poblaciones homogéneas (Stigler, 1986). Este método buscaba indagar si un conjunto de datos observados se ajustaba a la distribución normal.

En la segunda etapa del trabajo de Galton, se encuentran los experimentos que realizó con chicharos (ver Figura 3.3), los cuales le condujeron a la formulación de una ley empírica de la reversión (regresión) en 1877. En estos experimentos solicitó la ayuda de sus amigos para cultivar los chícharos, les proporcionó paquetes de semillas que había separado por peso en siete grupos, de las más pesadas a las más ligeras, para investigar la herencia (sí el tamaño del padre influye en el hijo). Cada grupo de los hijos se distribuía normalmente, además se encontraban centrados en diferentes pesos donde había la misma dispersión entre los grupos.

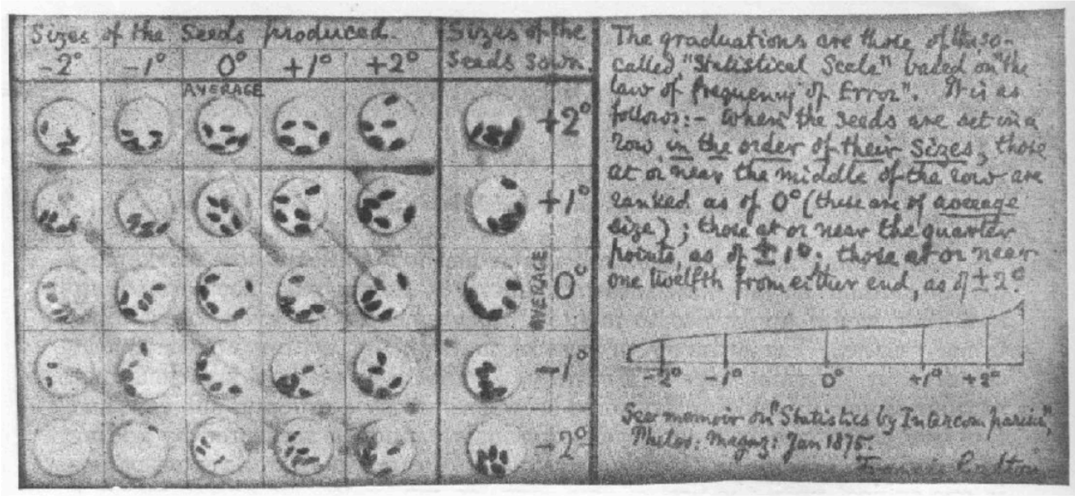


Figura 3.3. Experimentos con chicharos (fuente: galton.org)

Sus guisantes producían semillas con una variación (normal) de tamaños que retrocedía de la distribución de sus padres.

En la tercera etapa resalta su marco matemático que abarca la regresión con datos de población humana en 1885; y en la cuarta etapa el enlace de sus trabajos sobre regresión en su libro 'Natural Inheritance'. En la misma publicación, 'Natural Inheritance', también aborda la comparación entre lo observado con la distribución normal retomando aspectos del método de intercomparación de 1875, como el valor medio m y que en 1889 trabajó como M, el error probable $q - m$ que en ese momento lo denominó Q y la consideró la desviación probable, estableció que la medición de grado 25 era $M - Q$, mientras que la de grado 75 era $M + Q$ y M se encontraba en el grado 50. Finalmente, añadió que se podían "*aplicar varias propiedades hermosas de la ley de frecuencia del error para los valores observados de Q*" (Galton, 1889, p. 60). Sobre este trabajo presentó las tablas de las Figuras 3.4 y 3.5.

200 NATURAL INHERITANCE.

TABLE 2.

DATA FOR SCHEMES OF DISTRIBUTION of various qualities and faculties among the persons measured at the Anthropometric Laboratory in the International Exhibition of 1884.

Subject of measurement.	Age.	Unit of measurement.	Sex.	No. of persons in the group.	Values at the undermentioned Grades, from 0° to 160°. 5°	10°	20°	30°	40°	50°	60°	70°	80°	90°	95°
Height, standing, without shoes	23–51	Inches	M.	811	63·2	64·5	65·8	66·5	67·3	67·9	68·5	69·2	70·0	71·3	72·4
			F.	770	58·9	59·9	61·3	62·1	62·7	63·3	63·9	64·6	65·3	66·4	67·3
Height, sitting, from seat of chair	23–51	Inches	M.	1013	33·6	34·2	34·9	35·3	35·4	36·0	36·3	36·7	37 1	37·7	38·2
			F.	775	31·8	32·3	32·9	33·3	33·6	33·9	34·2	34·6	34·9	35·6	36·0
Span of arms . . .	23–51	Inches	M.	811	65·0	66·1	67·2	68·2	69·0	69·9	70·6	71·4	72·3	73·6	74·8
			F.	770	58·6	59·5	60·7	61·7	62·4	63·0	63·7	64·5	65·4	66·7	68·0
Weight in ordinary indoor clothes	23–26	Pounds	M.	520	121	125	131	135	139	143	147	150	156	165	172
			F.	276	102	105	110	114	118	122	129	132	136	142	149
Breathing capacity.	23–26	Cubic Inches	M.	212	161	177	187	199	211	219	226	236	248	277	290
			F.	277	92	102	115	124	131	138	144	151	164	177	186
Strength of pull as archer with bow.	23–26	Pounds	M.	519	56	60	64	68	71	74	77	80	82	89	96
			F.	276	30	32	34	36	38	40	42	44	47	51	54
Strength of squeeze with strongest hand	23–26	Pounds	M.	519	67	71	76	79	82	85	88	91	95	100	104
			F.	276	36	39	43	47	49	52	55	58	62	67	72
Swiftness of blow .	23–26	Ft. per second	M.	516	13·2	14·1	15·2	16·2	17·3	18·1	19·1	20·0	20·9	22·3	23·6
			F.	271	9·2	10·1	11·3	12·1	12·8	13·4	14·0	14·5	15·1	16·3	16·9
Sight, keenness of — by distance of reading diamond test-type .	23–26	Inches	M.	398	13	17	20	22	23	25	26	28	30	32	34
			F.	433	10	12	16	19	22	24	26	27	29	31	32

Figura 3.4. Datos por esquemas de la distribución de varias cualidades y facultades (fuente: Galton, 1889, p. 200)

TABLE 3.

DEVIATIONS from M in each of the series in Table 2, after reduction to a Scale in which Q′ = 1, where Q′ is the *Mean* of the observed Deviations at the Grades 20°, 30°, 70°, and 80°.

Subject of measurement.	Values of Q′	Unit of measurement in Table 2.	Sex.	No. of persons	Deviations reckoned in units of Q′. 5°	10°	20°	30°	40°	50°	60°	70°	80°	90°	95°
Height, standing, without shoes . .	1·72	Inches	M.	811	2·73	1·98	1·22	0·81	0·35	0	0·35	·76	1·22	1·98	2·61
	1·62		F.	770	2·71	2·10	1·23	·74	·37	0	·37	·80	1·23	1·91	2·46
Height, sitting, from seat of chair	0·95	Inches	M.	1013	2·52	1·89	1·15	·73	·63	0	·31	·73	1·15	1·79	2·31
	0·82		F.	775	2·55	1·95	1·22	·73	·36	0	·36	·85	1·22	2·07	2·55
Span of arms . . .	2·07	Inches	M.	811	2·36	1·83	1·30	·82	·43	0	·33	·72	1·16	1·79	2 36
	1·87		F.	770	2·35	1·87	1·23	·69	·32	0	·37	·80	1·28	1·98	2·67
Weight in ordinary indoor clothes . .	10·00	Pounds	M.	520	2·20	1·80	1·20	·80	·40	0	·40	·70	1·30	2·20	2·90
	11·00		F.	276	1·80	1·60	1·10	·70	·40	0	·60	·90	1·30	1·80	2·40
Breathing capacity .	24·50	Cubic Inches	M.	212	2·32	1·68	1·28	·80	·32	0	·28	·68	1·16	2·32	2·84
	19·00		F.	277	2·39	1·87	1·20	·73	·36	0	·31	·67	1·35	2·03	2·49
Strength of pull as archer with bow .	7·50	Pounds	M.	519	2·39	1·86	1·33	·80	·40	0	·40	·80	1·06	1·99	2·92
	5·22		F.	276	1·92	1·06	·80	·53	·27	0	·27	53	·93	1·46	1·86
Strength of squeeze with strongest hand	7·75	Pounds	M.	519	2·32	1·81	1·16	·77	·39	0	·39	·77	1·29	1·93	2·45
	7·50		F.	276	2 12	1·73	1·20	·66	·40	0	·40	·80	1·33	1·99	2·66
Swiftness of blow .	2·37	Ft. per second	M.	516	2·06	1·68	1·22	·80	·34	0	·42	·80	1·18	1·77	2·31
	1·55		F.	271	2·71	2·13	1·35	·84	·38	0	·38	·71	1·10	1·87	2·26
Sight, keenness of — by distance of reading diamond test-type .	4·00	Inches	M.	398	3·00	2·00	1·25	·75	·50	0	·25	·75	1·25	1·75	2·25
	5·22		F.	433	2·66	2·28	1·52	·95	·38	0	·38	·57	·95	1·33	1·52
SUMS . .					43·11	33·12	21·96	13·65	7·00	0	6·57	13·34	21·46	33·96	43·82
MEANS . .					2·40	1·84	1·22	0·76	0·39	0	0·37	0·74	1·19	1·89	2·43
MEANS multiplied by 1·015, to change unit to Q = 1					2·44	1·87	1·24	0·77	0·40	0	0·38	0·75	1·21	1·92	2·47
Normal Values, when Q = 1					2·44	1·90	1·25	0·78	0·38	0	0·38	0·78	1·25	1·90	2·44

TABLES. 201

Figura 3.5. Desviaciones de M de las series de la Figura 3.4 (fuente: Galton, 1889, p. 200)

Al respecto, Galton (1889) indicó que se encontraba "*sorprendido por la extraordinaria coincidencia entre las dos líneas inferiores de la Tabla 3* [Figura 3.5]*, considerando la gran* *variedad de facultades contenidas en los 18 esquemas… no me atreví a esperar que sus irregularidades se habrían equilibrado entre sí tan bellamente como lo han hecho* ” (p. 56). Aunque en la Figura 3.5 se puede apreciar en las últimas filas la diferencia entre las desviaciones observadas y las esperadas de acuerdo con la normal del conjunto de las 18 series, si se centra la atención en los valores de Q, de acuerdo con Galton, la mitad de las personas se estarían desviando de M en no más de diez u once libras (para el caso de las series donde se está trabajando con el peso).

En 1885 también publicó sobre su método gráfico, en parte utilizando el método de intercomparación, lo que marcaría el inicio de lo explicado anteriormente sobre las mediciones observadas y las desviaciones que presentan con respecto a M en diversos percentiles y cómo son estas desviaciones cuando las comparamos con las desviaciones de los mismos percentiles de la normal. Realizó una descripción del método gráfico auxiliándose del ejemplo que presentamos a continuación, sobre la altura (sentada) en pulgadas de 775 mujeres de entre 23 y 50 años, los datos observados se encuentran en la fila de A de la Tabla 3.2.

Ejemplo 1: El método gráfico sobre la altura de mujeres entre 23 y 50 años, tomado de Galton, (1885, pp. 262-265)

Tabla 3.2. *Altura, sentada, en pulgadas de mujeres adultas entre 23 y 50 años* (fuente: Galton, 1885, p. 264)

	29	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	
A	2	8	52	116	226	227	108	31	5	Total 775
B	2	10	62	178	404	631	739	770	775	Abscisa 0 a 775
C	0.2	1.3	8.0	23.0	52.2	81.4	95.4	99.4	100.0	Abscisa 0 a 100
D	30	31	32	33	34	35	36	37	38	Ordenada correspondiente

De acuerdo con el método gráfico se deben ordenar, por sus magnitudes, en la fila A los valores observados. Galton (1885) explicó que el significado de A es que del total de las 775 observaciones 2 casos corresponden a medidas de 29 pulgadas y menores a 30 pulgadas (frecuencia). Mientras que la fila B contiene la suma de A (frecuencia acumulada). En la fila C se encuentran los porcentajes de acuerdo a las sumas de la fila B (frecuencia relativa acumulada). En la fila D se acepta 0.2 como la abscisa correspondiente de 30 pulgadas y así sucesivamente, esto es lo que Galton graficaba en su curva de distribución y de donde él medía las ordenadas a $10°, 20°, 25°$ ... como en la Figura 3.6.

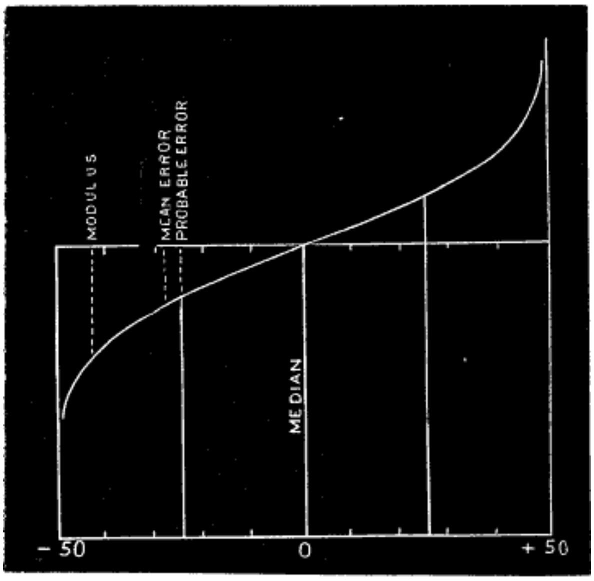


Figura 3.6. Curva de distribución (ojiva) (fuente: Galton, 1885, p. 262)

Respecto a su método graficó, Galton (1885) señaló "*nos permite probar muy fácilmente la presunta conformidad en cualquier caso dado, y, si existe, la curva graficada no solo nos da el error probable de estas series, sino también el error medio y el módulo*" (p. 264). En relación con esta curva de distribución también señalaba, al igual que en su método de intercomparación de 1875, que la ordenada central nos conduce al valor mediano que es prácticamente igual al valor medio, así como la identificación de los cuartiles. Indicó que el error probable era la medida más conveniente de variabilidad de la serie y se podía encontrar como la mitad de la diferencia entre los cuartiles. Cabe destacar que la curva de distribución de la Figura 3.6 presentada en 1885 es muy similar a la presentada en 1875 con el método de intercomparación y que se puede apreciar en la Figura 3.1; sin embargo, en la Figura 3.6 resalta la introducción de los percentiles y que se considera al cuartil dos o la mediana como 0, en el sentido de los trabajos realizados por Quetelet donde los otros elementos de la serie son considerados deviaciones del representante hasta ± 50.

Galton (1885) señaló que para realizar esta prueba había que cambiar la línea de referencia sobre la cual se originaron las ordenadas, partiendo la curva de la Figura 3.6 en dos partes, "e invirtiendo el más bajo", como se muestra en la Figura 3.7.

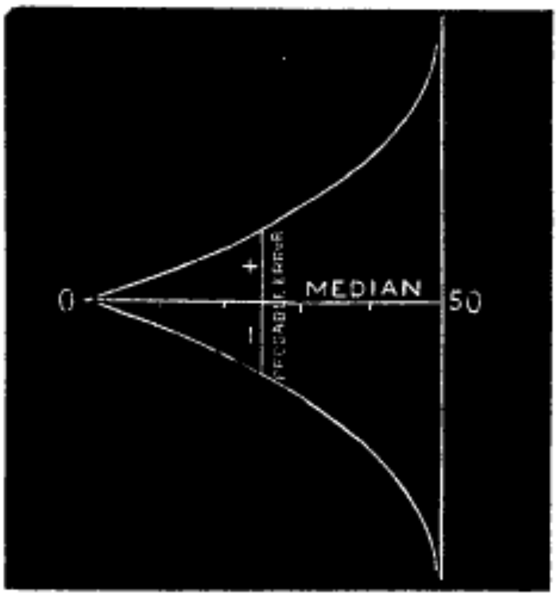


Figura 3.7. Segunda curva de distribución (fuente: Galton, 1885, p. 262)

Adicionalmente brindó una tabla de la curva normal de distribución del error (ver Tabla 3.3) y comentó que

> para llevar los valores observados a una forma adecuada para la comparación con esta tabla [*Tabla 3.3*], debemos comenzar midiendo las desviaciones observadas en $\pm 10°, 20°, 25°, 30°, 40°$, y $45°$ [...], si la serie es "normal", los valores así obtenidos serán idénticos a aquellos de la columna B, y si es aproximadamente normal, ellos corresponderán aproximadamente. (Galton 1885, p. 265)

Tabla 3.3. *Curva Normal de Distribución del Error* (fuente: Galton, 1885, p. 265)

Abscisa, calculada de 0° a $\pm$ 50° (valor de la integral)	**Ordenadas correspondientes**	
	Valor de la desviación en la que el $Modulo = 1$	Valor reducido proporcionalmente al $Error\ Probable = 1$
	A	B
……………………10	0.179	0.38
……………………20	0.371	0.78
Error Probable….25	0.477	1.00
Error Medio……..28.7	0.564	1.18
……………………30	0.595	1.25

……………………40	0.906	1.90
Módulo………..42.1	1.000	2.10
……………………45	1.163	2.44
……………………47	1.330	2.79
……………………49	1.651	3.46
……………………50	Infinito	Infinito

Galton añadió que el valor de las desviaciones en 28.7 será exactamente el valor del error medio, o en último caso, un valor muy cercano; así como, que la desviación en 42.1 dará el módulo.

En 1892, Galton realizó trabajos sobre las huellas dactilares en su libro Finger Prints en el que desarrolló una fórmula para calcular las frecuencias esperadas en cada celda con los totales marginales bajo el supuesto de independencia (Stigler, 1995; Batanero y Borovcnik, 2016), esta forma de calcular las frecuencias esperadas fue retomada por Pearson en 1904 cuando utilizó el estadístico χ^2 para medir la contingencia. También formuló el coeficiente de correlación estadística por otra ruta indirecta, graficando minuciosamente y volviendo a graficar sus datos sobre distribuciones normales bivariadas hasta que se dio cuenta de que las fórmulas para las curvas elípticas (un tema popular en las matemáticas del siglo XIX) podrían proporcionarle un método para resumir con un número la relación gráfica que observó (Stigler, 1986). Este número podría usarse para razonar sobre la relación y formar una base para las comparaciones.

De acuerdo con Magnello (2005), antes de que Pearson en 1900 realizará su propuesta de bondad de ajuste con el estadístico χ^2, se valoraba si las frecuencias observadas se ajustaban a la normal, "*el procedimiento habitual consistía en comparar errores de observación, con una tabla de distribuciones basadas en la curva normal, o gráficamente por medio de un diagrama de frecuencias*" (p. 727).

3.1.2 Los inicios de la prueba de bondad de ajuste con el estadístico χ^2de Karl Pearson

Entre 1830 y 1900, la aplicación de los métodos estadísticos se extendió gradualmente a las ciencias sociales, la biología y la genética. Quetelet y Galton utilizaban la distribución normal para describir la variación entre los individuos o de las características de los individuos, lo que les permitía diseccionar el conjunto. Sin embargo, con las nuevas aplicaciones de las técnicas estadísticas observaron que los datos con los que habían estado trabajando seguían distribuciones sesgadas y por lo tanto los subgrupos que se formaban basados en la distribución normal tenían elementos que no pertenecían realmente a ese subgrupo, lo que llevó a los estadísticos de la época a la búsqueda de distribuciones no normales. Se partió de las propiedades de la distribución normal para buscar nuevas distribuciones que se ajustaran al comportamiento de distintos conjuntos de datos. En esta búsqueda hicieron aportaciones Thiele, Gram, Fechner, Galton, Edgeworth, Kapteyn, Hagen y Karl Pearson (Hald, 2007).

En 1884, cuando Karl Pearson fue profesor de la University College of London, recibió una gran influencia del trabajo de Francis Galton y su amigo, el zoólogo W. R. F. Weldon sobre sus estudios de la evolución y genética.

En 1890, Weldon intentó ajustar los datos sobre las longitudes del caparazón del cangrejo de Plymouth a la curva normal siguiendo el método gráfico presentado por Galton en 1885 y en ese momento el método fue apropiado para los datos, pero en un trabajo posterior de 1892, cuando analizaba los caparazones de curvas dobles en cangrejos femeninos de Naples, se percató que el método de Galton no era lo suficientemente sensible con la no normalidad de su distribución de datos.

En 1892, Pearson ayudó a Weldon con un análisis de datos del zoológico con el propósito de dilucidar la teoría de la evolución de Darwin. En ese momento se dio un cambio en los intereses científicos de Pearson y empezó a instruirse en Estadística Matemática (Hald, 2007). Al igual que otros contemporáneos, Pearson sintió la necesidad de una serie de distribuciones continuas para poder describir los fenómenos biológicos que estaba estudiando.

Generalizando la aproximación de Hagen, Pearson en 1895 publicó un sistema o una familia de distribuciones de probabilidad continuas (siete tipos de curvas de frecuencia), para lo cual resolvió una ecuación diferencial de la misma forma que la ecuación de diferencia que satisfizo a la distribución hipergeométrica (Hald, 2007). En este sistema de distribuciones utilizó las mismas medidas de asimetría y de curtosis que Thiele, con la diferencia que Pearson las expresó en momentos.

Para estimar los parámetros de las nuevas distribuciones, se desarrollaron varios métodos que eran más sencillos que el de mínimos cuadrados. Galton utilizó dos percentiles para ajustar la distribución normal a sus datos, de manera similar, Kapteyn utilizó cuatro percentiles para estimar el parámetro en su modelo. Thiele usó cumulantes empíricos como estimador de los cumulantes[4] teóricos y Pearson utilizó los momentos empíricos como estimador de los momentos teóricos (método de los momentos) y obtuvo tantas ecuaciones no lineales como parámetros para estimar (Hald, 2007).

Pearson en 1893 mostró que, si se tiene un conjunto de líneas a una determinada unidad de distancia entre sí, se encuentra el primer momento con la suma de sus longitudes multiplicada por sus respectivas distancias de una línea recta paralela; en otras palabras, se calcula la media. El segundo momento es la suma de sus longitudes multiplicada por los cuadrados de sus distancias. A este último, Pearson lo llamó el cuadrado de la desviación estándar y en 1918 Fisher utilizó el término de varianza. El tercer momento es la suma de sus longitudes multiplicada por el cubo de sus distancias, esto determina el sesgo de la distribución. Mientras que el cuarto momento mide la curtosis y se encuentra multiplicando las longitudes por la cuarta potencia, esta medida de forma puede categorizarse en leptocúrtica (curtosis >0), platicúrtica (curtosis < 0) y mesocúrtica (curtosis =0) (Magnello, 2005).

Magnello (2005), señala que Pearson en 1893 también introdujo un sexto momento, de su método de los momentos para el ajuste de curvas simétricas y asimétricas, lo que le proporcionaría una base empírica para la prueba de bondad de ajuste con el estadístico χ^2. En 1894, en una consideración previa de la prueba de bondad de ajuste con el estadístico χ^2 "*demostró cómo encontrar $\sum s/y$, donde s equivalía a la diferencia entre el polígono observado y la curva teóricamente esperada y y su ordenada correspondiente*" (Magnello, 2005, p. 727), pero que Pearson prefirió medir la relación del área completa entre la curva y el polígono, donde todos los valores positivos equivalían a W y A sería el área bajo la curva. Teniendo la siguiente expresión $W/A = (\sum errores\ de\ ajuste)/(\sum ordenadas) = \sum s/\sum y$ como una medida razonable (hasta ese momento) de bondad de ajuste; donde s son las diferencias entre las frecuencias observadas y las frecuencias teóricas e y su correspondiente ordenada.

Tabla 3.4. *Distribuciones de Pearson* (fuente: Hald, 2007, p. 122)

Tipo	Ecuación y=	Origen para x	Límites para x	Criterio

[4] Los cumulantes (k_n) de una distribución de probabilidad representan una alternativa a los momentos de una distribución y se pueden definir como el logaritmo natural de la función característica como $H(t) = \log E[e^{itX}] = \sum_{n=1}^{\infty} k_n \frac{(it)^n}{n!} = \mu it - \sigma^2 \frac{t^2}{2} + \cdots$

I	$y_0\left(1+\frac{x}{a_1}\right)^m\left(1-\frac{x}{a_2}\right)^m$	Moda	$-a_1 \leq x \leq a_2$	$k < 0$
II	$y_0\left(1-\frac{x^2}{a^2}\right)^m$	Media (= Moda)	$-a \leq x \leq a$	$k = 0$
III	$y_0 e^{\gamma x}\left(1+\frac{x}{a}\right)^{\gamma a}$	Moda	$-a \leq x < \infty$	$k = \infty$
IV	$y_0 e^{\nu tan^{-1}\frac{x}{a}}\left(1+\frac{x^2}{a^2}\right)^{-m}$	$Media + \frac{\nu a}{r}$, $r = 2m - 2$	$-\infty < x < -\infty$	$0 < k < 1$
V	$y_0 e^{\frac{-\gamma}{x}} x^{-p}$	Al iniciar la curva	$0 \leq x < \infty$	$k = 1$
VI	$y_{0(x-a)^{q_2}x^{q_1}}$	Antes o al iniciar la curva	$a \leq x < \infty$	$k > 1$
VII	$y_0\left(1+\frac{x^2}{a^2}\right)^{-m}$	Media (= Moda)	$0 < x < \infty$	$k = 0$

Pearson probó su sistema de distribuciones de probabilidad continuas (ver Tabla 3.4) con una variedad de datos de meteorología, antropometría, zoología, botánica, economía, demografía y mortalidad estadística, demostrando así la habilidad de sus distribuciones de ajustarse a los datos (Hald, 2007); en las Figuras 3.8 y 3.9 se presentan algunos ejemplos. Estaba buscando distribuciones con soporte tanto infinito como finito, así como sesgo a la derecha o a la izquierda. A partir de esta propuesta, surgieron otras preguntas y llamó especialmente la atención de Pearson, el de una medición objetiva o razonable de la bondad de ajuste, problema que intentaría solucionar en 1900 con su propuesta del test de bondad de ajuste por medio del estadístico χ^2.

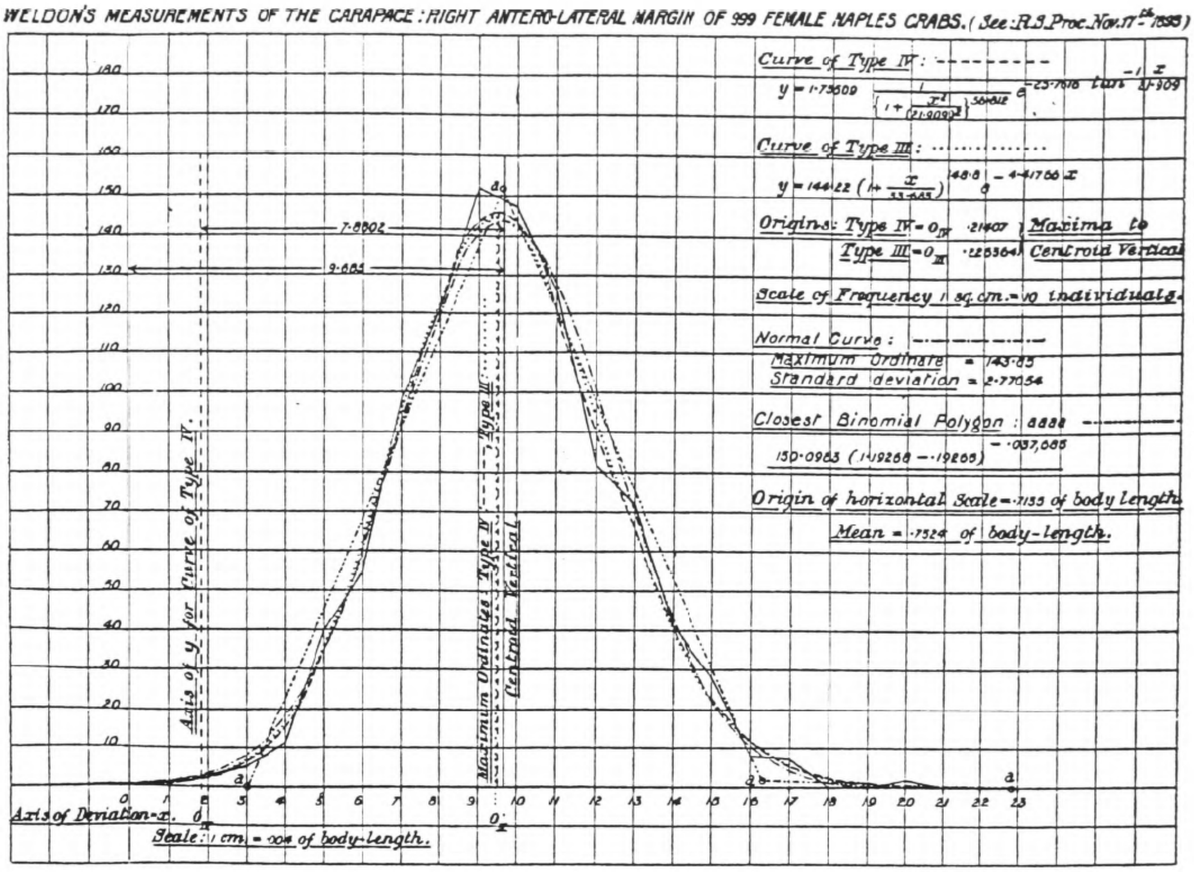


Figura 3.8. Distribución sesgada del margen anterior lateral derecho de 999 cangrejos femeninos de Weldon de 1895 (fuente: Magnello, 1993, p. 227)

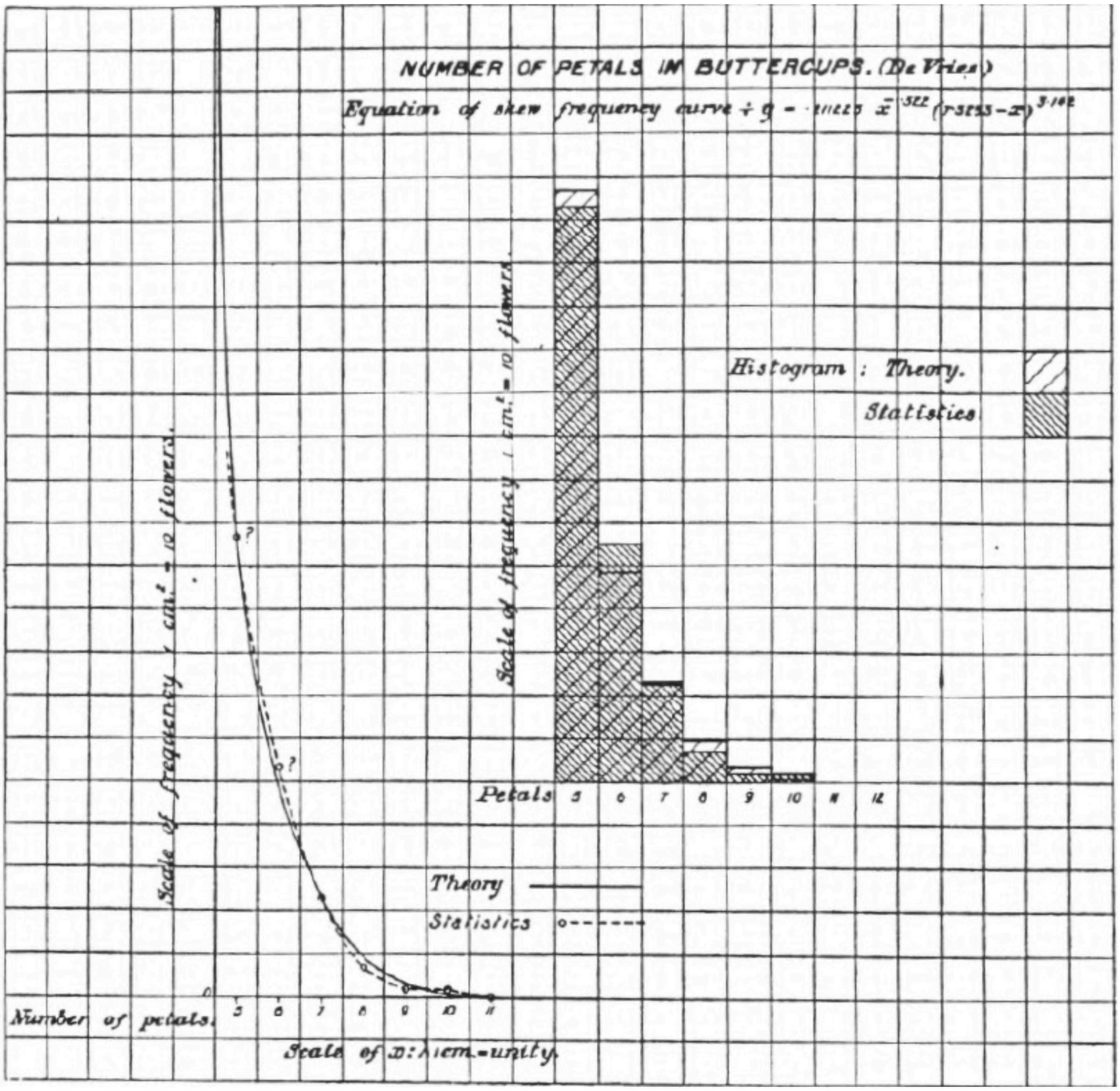


Figura 3.9. Curva continua e histograma de pétalos de buttercups de 1895 (fuente: Magnello, 1993, p. 241)

En 1900, Pearson publicó la prueba de bondad de ajuste con el estadístico χ^2 que permite, a manera de una prueba de hipótesis, valorar en qué medida se ajusta un grupo de datos observados a cierta distribución teórica preestablecida, por medio del contraste de las frecuencias observadas y las frecuencias esperadas. El estadístico χ^2 sigue una distribución de probabilidad del mismo nombre y es dada por la curva de Pearson de Tipo III. En la primera parte del artículo de Pearson (1900) se señala que el objetivo es "*investigar un criterio de probabilidad de cualquier teoría de un sistema de errores observados y aplicarlo a la determinación de bondad de ajuste en el caso de las curvas de frecuencia*" (p. 157).

Posteriormente, Pearson parte de la expresión de la función de densidad normal para un vector n, de variables x, con media $\mu = 0$ y una matriz de dispersión V. De acuerdo con Barnard

(1992), en una notación actual es: $K\ exp^{-\frac{1}{2}}(x-\mu)'V^{-1}(x-\mu)$, donde $'$ denota transposición y $K\ es\ P(x)=1$. Pearson señala que:

$$\chi^2 = x'V^{-1}x$$

Estableciendo que χ^2 es igual a una constante y representa a una ecuación de un elipsoide generalizada en el espacio muestral y que los valores que toma χ^2 deben de darse para el rango de 0 a ∞, es decir se calcula $P(x>x_0)$. El elipsoide, por medio de una transformación lineal podría convertirse en una esfera y las $X's$ referirían ahora a coordenadas, entonces las posibilidades de un sistema de errores con frecuencias muy grandes estarían denotadas por χ y dado por:

$$P = \frac{\left[\int\int\int\int \dots e^{\frac{\chi^2}{2}}\, dX_1 dX_2 \dots dX_n\right]_{\chi}^{\infty}}{\left[\int\int\int\int \dots e^{\frac{\chi^2}{2}}\, dX_1 dX_2 \dots dX_n\right]_{0}^{\infty}}$$

De acuerdo con Pearson, una vez que conocemos las desviaciones observadas, los errores probables (o $\sigma's$) y las correlaciones de los errores podríamos encontrar χ de la formula anterior.

Pearson realizó una transformación a coordenadas polares generalizadas, donde χ puede ser tratada como una de las líneas que diverge de un centro común, teniendo factores comunes en el denominador y numerador que representan la generalización de ángulos sólidos y con límites idénticos. Al reducir el resultado obtuvo:

$$P = \frac{\int_{\chi}^{\infty} e^{\frac{\chi^2}{2}} \chi^{n-1} d\chi}{\int_{0}^{\infty} e^{\frac{\chi^2}{2}} \chi^{n-1} d\chi}$$

Pearson (1900) indicó que:

> Esta es la medida de probabilidad de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema de observado […] y una evaluación de P nos da lo que parece ser un criterio bastante razonable de la probabilidad de que tal error ocurra en una selección realizada aleatoriamente. (p. 158)

Después aplicó los resultados al problema de ajustar distribuciones de frecuencias observadas a distribuciones de frecuencias teóricas. Si los datos se agrupan en $n+1$ grupos, entonces la frecuencias observadas en cada uno serían: $m'_1, m'_2, m'_3, \dots, m'_n, m'_{n+1}$; mientras que las frecuencias teóricas previamente conocidas son: $m_1, m_2, m_3, \dots, m_n, m_{n+1}$. Como $e = m' - m$ proporciona el error, entonces $e_1, e_2, e_3, \dots, e_n, e_{n+1} = 0$ si la frecuencia observada corresponde a la teórica.

Pearson estableció, sin probar, la desviación estándar y la correlación del error aleatorio y al insértalos en la expresión inicial, después de hacer uso de algebra obtiene:

$$\chi^2 = S\left(\frac{e^2}{m}\right)$$

La suma es ahora extendida a todos los errores $(n+1)$, y no únicamente al primer n. Actualmente, es común expresar al estadístico χ^2 de la siguiente forma:

$$\chi^2 = \sum \frac{(Frecuencia\ Obervada - Frecuencia\ Esperada)^2}{Frecuencia\ Esperada}$$

La χ^2 no sólo contiene el promedio de los valores de X y su desviación estándar, también tiene las siguientes propiedades, como las que tiene la varianza: (1) se pueden adherir dos independientes $\chi^2's$ para formar una nueva χ^2, (2) de forma inversa, una χ^2 puede algunas veces ser dividida significativamente en dos o más componentes. Si utilizáramos solo χ perdería estas ventajas (Barnard, 1992). Esta cantidad es lo que Pearson (1900) denominó "*un resultado muy simple y muy sencillo de aplicar*" (p. 163):

$$\chi = \sqrt{S\left(\frac{e^2}{m}\right)}$$

"*Es una medida de bondad de ajuste y las etapas de nuestra investigación son muy claros"* (Pearson, 1900, pp. 163-164). Las etapas que señala son las siguientes:

1. Encontrar χ de la ecuación:

$$\chi^2 = S\left(\frac{e^2}{m}\right)$$

2. Sí el número de errores, $n' = n + 1$, es impar, encuentre la improbabilidad del sistema observado de:

$$P = e^{\frac{\chi^2}{2}}\left(1 + \frac{\chi^2}{2} + \frac{\chi^4}{2 \cdot 4} + \frac{\chi^6}{2 \cdot 4 \cdot 6} + \cdots + \frac{\chi^{n'-3}}{2 \cdot 4 \cdot 6 \ldots n' - 3}\right)$$

Sí el número de errores, $n' = n + 1$, es par, encuentre la probabilidad del sistema observado de:

$$P = \sqrt{\frac{2}{\pi}}\int_{\chi}^{\infty} e^{\frac{\chi^2}{2}} dX + \sqrt{\frac{2}{\pi}} e^{\frac{\chi^2}{2}}\left(\frac{\chi}{1} + \frac{\chi^3}{1 \cdot 3} + \frac{\chi^5}{1 \cdot 3 \cdot 5} + \cdots + \frac{\chi^{n'-3}}{1 \cdot 3 \cdot 5 \ldots n' - 3}\right)$$

3. Si n (variables correlacionadas) es menor que 13, entonces, Pearson propuso auxiliarse de la tabla que brinda al finalizar el artículo, para determinar la probabilidad o improbabilidad general del sistema observado sin la necesidad de utilizar las expresiones del punto dos para el cálculo de P.

Pearson consideraba que era despreciable el efecto que provoca m en el estadístico χ^2, cuando la frecuencia teórica era desconocida y se debía derivar de estimaciones de parametros de la población (m_*). Suponía que estas deviaciones que causaba en el estadístico eran muy pocas o pequeñas y se podían ignorar en la práctica, si las muestras eran suficientemente grandes (Pearson, 1900).

En la segunda parte de la publicación presentó algunos ejemplos donde aplicó la prueba de bondad de ajuste. También incluyó una tabla para una apreciación general de la probabilidad de ocurrencia de un sistema de desviaciones, tan o más grande que el atípico en cuestión, definida por la distribución χ^2, para valores del estadístico χ^2 de 1 a 70 y n' de 3 a 20. Dentro de los ejemplos que proporcionó se encuentra el siguiente con frecuencia de la población general no conocida previamente:

Ejemplo 2: Prueba de bondad de ajuste con el estadístico χ^2 (Pearson, 1900, pp. 172-173)

> El profesor Merriman en su tratado sobre los mínimos cuadrados[5] inicia de la forma correcta, no con la teoría, sino con la experiencia actual, y después desde sus datos se

[5] 'A Textbook on the Method of Least Squares,' 1891, p. 14. (nota del texto original)

deducen tres axiomas. De esos axiomas él obtuvo, por medio del análisis de la curva normal, como resultado teórico: pero si estos axiomas son verdaderos, sus datos pueden sólo diferir de la ley normal de frecuencias por un sistema de desviaciones como las que razonablemente surgirían si una selección aleatoria se hiciera a partir de material que en realidad obedece a la ley normal. Ahora el profesor Merriman pone en el lugar de honor 1000 disparos en la línea de fuego sobre un objetivo en la práctica para el gobierno de los Estados Unidos. Las desviaciones fueron agrupadas de acuerdo con los cinturones golpeados, en el objetivo fueron dibujados los cinturones de igual amplitud y paralelo a la línea. La siguiente tabla [*Tabla 3.5*] muestra la distribución de aciertos y la distribución teórica de frecuencias calculada a partir de tablas del área de la curva normal[6].

Tabla 3.5. *Distribución de las frecuencias* (fuente: Pearson, 1900, p. 173)

Cinturón	**Frecuencia Observada**	**Frecuencia Teórica**	$\boldsymbol{e}$	$\boldsymbol{\frac{e^2}{m}}$
1	1	1	0	0
2	4	6	-2	0.667
3	10	27	-17	10.704
4	89	67	+22	7.224
5	190	162	+28	4.839
6	212	242	-30	3.719
7	204	240	-36	5.400
8	193	157	+36	8.225
9	79	70	+9	1.157
10	16	26	-10	3.846
11	2	2	0	0
Σ	1000	1000	$\chi^2 = 45.811$	

Y señala que por lo tanto se deduce: $P = .000{,}00155$

En otras palabras, si los disparos son distribuidos sobre el objetivo de acuerdo con la ley normal, entonces una distribución como la señalada por el señor Merriman sólo se podría esperar que ocurra, en promedio, algunas 15 o 16 veces en 10,000,000 de ensayos. Ahora, seguramente es muy desafortunado citar, a manera de ilustración, que

[6] Se tomó esta ilustración con la gentileza de Mr. W. R. Macdonell, M. A., LL. D. (nota del texto original)

> el fundamento de aquellos axiomas ¡debe venir de la curva normal! Porque sí la curva de la normal fluye desde los axiomas, entonces los datos deberían ser un sistema probable de desviaciones de la curva normal. Pero estos ciertamente no lo son. Ahora me parece que, si los primeros escritores de probabilidad no hubiesen procedido enteramente desde el punto de vista matemático, sino que hubieran procedido empíricamente, intentando primero clasificar las desviaciones del promedio, y después medir la bondad de ajuste a la actual curva normal, dicha curva no sería tan utilizada en la teoría de los errores. Incluso hoy hay quienes la consideran una especie de obsesiva devoción; y aunque admiten a los errores como un medio para describir generalmente la distribución de la variación de una cantidad x desde su media, afirman que debe existir una cantidad desconocida z de la cual x es una función desconocida y que ¡z realmente obedece a la ley normal! Esto podría ser razonable si hubiera pocas excepciones a esta ley universal del error; pero la dificultad es encontrar incluso las pocas variables que la obedecen, ¡y esas pocas no son aquellas usualmente citadas como ilustraciones por los escritores sobre el tema!

En la Tabla 3.5, por una parte, el cálculo de la columna de los errores (e) se procede a encontrar las diferencias entre la frecuencia observada y la frecuencia esperada ($e = m' - m$). En la última columna se calculan los cocientes de las diferencias al cuadrado entre la frecuencia esperada $\frac{(m'-m)^2}{m}$. Al realizar la sumatoria de la última columna se obtiene el valor del estadístico χ^2.

Sí las frecuencias observadas son iguales a las esperadas, el estadístico χ^2 tomaría el valor de cero, es decir, $\chi^2 = S\left(\frac{e^2}{m}\right) = 0$, mientras que sí $\chi^2 > 0$ no coinciden exactamente. A mayor valor del estadístico χ^2 las desviaciones entre las frecuencias observadas y esperadas son mayores. La pregunta obligada es qué tan grande tiene que ser el valor del estadístico χ^2 para considerar que la distribución observada no corresponde a la distribución teórica. Para responder esta pregunta se hizo uso de la probabilidad de ocurrencia de un sistema de desviaciones, tan grande o más grande que el atípico en cuestión, definida por la distribución χ^2.

En la primera parte de la publicación Pearson estableció una fórmula para el cálculo de la probabilidad para el $n' = n + 1$ cuando es par y otra para cuando es impar. En este caso son 12 cinturones por lo que debemos encontrar P de la siguiente formula:

$$P = \sqrt{\frac{2}{\pi}}\int_{\chi}^{\infty} e^{\frac{\chi^2}{2}} dX + \sqrt{\frac{2}{\pi}} e^{\frac{\chi^2}{2}} \left(\frac{\chi}{1} + \frac{\chi^3}{1\cdot 3} + \frac{\chi^5}{1\cdot 3\cdot 5} + \cdots + \frac{\chi^{n'-3}}{1\cdot 3\cdot 5 \ldots n'-3}\right)$$

Podemos obtener el valor de χ del estadístico χ^2, entonces $\chi = 6.768$. Si evaluamos χ en la expresión anterior obtenemos:

$$\begin{aligned} P = {} & \sqrt{\frac{2}{\pi}}\int_{6.768}^{\infty} e^{\frac{\chi^2}{2}} dX \\ & + \sqrt{\frac{2}{\pi}} e^{\frac{6.768^2}{2}} \left(\frac{6.768}{1} + \frac{6.768^3}{1\cdot 3} + \frac{6.768^5}{1\cdot 3\cdot 5} + \frac{6.768^7}{1\cdot 3\cdot 5\cdot 7} + \frac{6.768^9}{1\cdot 3\cdot 5\cdot 7\cdot 9}\right) \\ & = 0.000{,}00155 \end{aligned}$$

Esta probabilidad nos indica que tendríamos oportunidad de ver un conjunto de datos con un sistema de errores de tan amplias desviaciones de la distribución normal, únicamente en 15 o 16 ocasiones en 10,000,000 de ensayos.

Finalmente, Pearson realizó algunas críticas enfocadas hacia el uso indiscriminado de la distribución normal para ser utilizada como modelo de los fenómenos que se estaban trabajando, haciendo énfasis en el uso que se hacía de ella sin preguntarse si realmente X se distribuía como una normal (Z), señalando que existían pocas variables que obedecen la ley universal del error.

De acuerdo con Pearson (1900, p. 174), "*la curva de la normal no posee un ajuste especial para describir errores o desviaciones como los que surgen ya sea observando la práctica o en la naturaleza*". Pearson nuevamente critica el uso desmedido que se le daba a distribución normal y sugería la prueba de bondad de ajuste con el estadístico χ^2 para valorar y describir las desviaciones de lo observado en la práctica con diversas distribuciones teóricas.

3.1.3 Los inicios de la prueba de independencia con el estadístico χ^2

George Udny Yule se encontraba buscando una forma de medir estadísticamente la asociación de atributos, por ejemplo, sí existía alguna asociación (en el sentido de correlación, pero para variables discretas) entre enfermedades como sordera, ceguera y retraso mental o en los casos de mortalidad por enfermedades y la aplicación de nuevas antitoxinas. Abordar estudios sobre las nuevas antitoxinas y las variaciones de las dosis que se aplicaban era de gran importancia para el avance de la ciencia médica, sin embargo, eran poco abordados por los estadistas de la época.

Para abordar esta problemática, Yule (1900) parte de la teoría de correlación estadística, sea normal[7] o no, señaló que siempre se supone que la variación está involucrada en los datos que trabajamos, ya sea una variación continua o una variación por un número considerable de pasos discontinuos. Indicó ejemplos de correlación de los dos tipos, siendo la primera la que corresponde a variables continuas (e.g., medición de las proporciones de las partes del cuerpo) y la segunda a variables discretas (e.g., número de niños en una familia).

Inicialmente, Yule retomó ideas del Método de Jevons; aunque este método se había trabajado por logistas, no había aplicaciones prácticas y Yule lo consideraba prácticamente intuitivo. En este método se denota con $A, B, C, \&c.$ a los objetos o individuos que tienen las cualidades $A, B, C, \&c.$ y con los términos $(A), (B), (C), \&c.$ denota la frecuencia de los indivuduos que poseen la cualidad $A, B\ o\ C$, y como (U) la frecuencia total. Yule retomó esta simbología y adoptó algunas convenciones como tomar las letras inglesas para las cualidades positivas y las letras griegas para las negativas (e.g., $A, \propto$), es decir, para la propiedad de la igualdad de contrarios (simetría en los grupos entre las cualidades positivas y cualidades negativas).

Para fundamentar y desarrollar su trabajo de asociación, Yule (1900) partió de teoremas desde teoría de conjuntos de probabilidad, donde señala que "dos cualidades o atributos, A y B, son definidos como independientes si la probabilidad de encontrarlos juntos es el producto de las probabilidades de encontrarlos también por separado, por ejemplo, si $\frac{(AB)}{U} = \frac{(A)}{(U)} \cdot \frac{(B)}{(U)}$, o $(AB)(U) = (A)(B)$" (p. 270).

Yule señaló que pensaba que éste era el único legítimo test de independencia o asociación, entendiendo asociación como dependencia. Y para demostrarlo recurrió al siguiente teorema:

$$(AB)(U) = (A)(B)$$

entonces,

[7] Medida de asociación entre dos variables aleatorias normalmente distribuidas

$$(A\beta)(U) = (A)(\beta)$$

$$(\alpha B)(U) = (\alpha)(B)$$

$$(\alpha\beta)(U) = (\alpha)(\beta)$$

Además, Yule (1900) argumenta que

> si la oportunidad de encontrar dos de las cualidades juntas es el producto de las posibilidades de encontrarlas también por separado, la oportunidad de encontrar una sin la otra es la posibilidad de encontrar una multiplicada por la posibilidad de no encontrar la otra. Cualquiera de las relaciones implica todas las demás. (p. 270)

De lo anterior se deduce que, si los productos cruzados son iguales, las variables son independientes, lo que se entiende de la siguiente manera $(AB)(\alpha\beta) = (A\beta)(\alpha B)$.

Yule (1900), señaló también la necesidad de una especie de coeficiente asociación para analizar este tipo de datos, que tome el lugar del coeficiente de correlación que se utiliza con variables continuas; y que este coeficiente de asociación "*sea una medida que aproxime la asociación, por una parte, hacia una completa independencia y, por otra parte, hacia una completa asociación*" (p. 271).

Y que el coeficiente de asociación debería:

1. Ser cero cuando las variables o atributos A, B, son independientes, y sólo cuando son independientes
2. Deberían ser $+1$ cuando, y solo cuando, A y B son completamente asociados, i.e., cuando cualquiera

$$\left.\begin{matrix} todos\ los\ A's\ son\ B \\ todos\ los\ \ \beta's\ son\ \alpha \end{matrix}\right\}$$

o

$$\left.\begin{matrix} todos\ los\ B's\ son\ A \\ todos\ los\ \ \alpha's\ son\ \beta \end{matrix}\right\}$$

O cuando ambas afirmaciones son verdaderas, lo que solo puede ser cuando $(A) = (B),\ (\alpha) = (B)$

Los siguientes tres diagramas ejemplifican los tres casos que corresponden a

$$(A\beta) = 0, \qquad (\alpha B) = 0, \qquad (A\beta) = (\alpha B) = 0$$

	(B)	(β)
(A)	(AB)	0
(α)	(αB)	$(\alpha\beta)$

	(B)	(β)
(A)	(AB)	$(A\beta)$
(α)	0	$(\alpha\beta)$

	(B)	(β)
(A)	(AB)	0
(α)	0	$(\alpha\beta)$

3. Deberían ser -1 cuando, y sólo cuando, A y β o B y α son completamente asociados, i.e., cuando cualquiera

$$\left.\begin{matrix} todos\ los\ A's\ son\ \beta \\ todos\ los\ \ B's\ son\ \alpha \end{matrix}\right\}$$

$$\left.\begin{matrix} todos\ los\ \beta's\ son\ A \\ todos\ los\ \ \alpha's\ son\ B \end{matrix}\right\}$$

O cuando ambas de estas afirmaciones son verdaderas, que de nuevo solo puede ser si

$$(A) = (\alpha), \qquad (B) = (\beta)$$

Los siguientes tres diagramas ejemplifican estos casos de asociación negativa que corresponden a

$$(AB) = 0, \qquad (\alpha\beta) = 0, \qquad (AB) = (\alpha\beta) = 0$$

	(B)	(β)
(A)	0	$(A\beta)$
(α)	(αB)	$(\alpha\beta)$

	(B)	(β)
(A)	(AB)	$(A\beta)$
(α)	(αB)	0

	(B)	(β)
(A)	0	$(A\beta)$
(α)	(αB)	0

Finalmente, de los teoremas anteriores surge Q como el coeficiente de asociación, tal que $Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$

Q sirve como un coeficiente de asociación cuando:

1. A y B son independientes el numerador es cero y por lo tanto Q es cero; y a la inversa cuando Q es cero las variables son independientes.

2. $(A\beta) = 0$ o $(\alpha B) = 0$, o ambas, $Q = +1$; y a la inversa cuando $Q = +1$ $(A\beta) = 0$, $(\alpha B) = 0$, o ambas.
3. $(AB) = 0$ o $(\alpha\beta) = 0$, o ambas $Q = -1$; y a la inversa cuando $Q = -1$ $(AB) = 0$ o $(\alpha\beta) = 0$, o ambas. (Yule, 1900, p. 272)

Posteriormente, Yule señaló que era perfectamente posible que se idearan otras formas de medir la asociación y que estas debían cumplir con las mismas propiedades. Además, intenta establecer las relaciones entre el coeficiente de correlación y el de asociación que propone, para esto también hace referencia al teorema de Mr. W. F. Sheppard de 1898 y forman una conexión con el coeficiente de correlación de Bravais y su coeficiente de asociación para el caso de correlación normal. Añadiendo que para una completa conexión entre los coeficientes es necesario mayor investigación.

Para mostrar la aplicación del coeficiente de asociación Q que propone, Yule (1900) presentó algunos ejemplos, a continuación, se aborda uno de ellos y trata sobre el índice de ataque de viruela y vacunación.

Ejemplo 3: Prueba de independencia con el coeficiente de asociación Q (Yule 1900, pp. 288 - 290)

> Al inicio de este documento sugerimos los índices de muertes, con y sin la administración de una antitoxina, como ejemplos de una adecuada accesibilidad de "asociación". Los índices de muertes por viruela entre los vacunados y los no vacunados podrían formar una instancia, pero en las figuras yo encontré más adecuado para mi propósito los índices de ataques – no los índices de muertes. La siguiente tabla [*Tabla 3.6*] muestra el índice (porcentaje) de ataque de viruela, *en casas actualmente invadidas por viruela*, de personas debajo y sobre 10 años de edad, en cinco pueblos en los cuales han ocurrido recientes epidemias de viruela.

Tabla 3.6. *Índices de ataque de viruela en casas infectadas* (fuente: Yule, 1900, p. 289)

		Índice de ataque a menores de 10		**Índice de ataque a mayores de10**	
Pueblo	Fecha	Vacunado	No vacunado	Vacunado	No Vacunado
Sheffield	1887-88	7.9	67.6	28.3	53.6
Warrington	1892-93	4.4	54.5	29.9	57.6
Dewsbury	1891-92	10.2	50.8	27.7	53.4

Leicester	1892-93	2.5	35.3	22.2	47.0
Gloucester	1895-96	8.8	46.3	32.2	50.0

A partir de estos datos nosotros podemos trabajar la asociación entre la "ausencia de la vacuna" y "ataque", para niños y personas mayores de 10 años de edad. Si nosotros llamamos A "ataque", B "no vacunados", los datos dados son $100\,(A\beta)/(\beta)$ y $100\,(AB)/(B)$; restando cada porcentaje de 100, nosotros obtenemos $100\,(\alpha\beta)/(\beta)$ y $100\,(\alpha B)/(B)$. Entonces para el coeficiente de asociación en Sheffield para niños tenemos:

$$Q = \frac{67.6 \times 92.1 - 32.4 \times 7.9}{67.6 \times 92.1 + 32.4 \times 7.9} = 0.92$$

Donde hemos dividido a través del numerador y denominador de la expresión ordinaria de Q por $(B)(\beta)$, dejando su valor inalterado. Esto parece un caso bastante interesante, ya que la forma en la cual los datos son presentados no arrojan la relación de excedentes[8] para los no vacunados, por ejemplo, la relación de no vacunados a vacunados, pero nos arrojan el coeficiente de asociación Q. La serie completa de valores son dados a continuación y forma una sorprendente adición a la tabla anterior. La asociación entre no vacunados y los ataques es sorprendentemente muy alta para niños jóvenes – 0.8 a 0.9 – pero decae bruscamente a 0.5 (presumiblemente debido a la protección decreciente de la vacunación realizada en la infancia) en el grupo de mayor edad [*Tabla 3.7*].

Tabla 3.7. *Asociación entre los no vacunados y ataques en hogares infectados* (fuente: Yule, 1900, p. 289)

Pueblo	**Niños menores de 10**	**Personas mayores de 10**
Sheffield	0.92	0.49
Warrington	0.93	0.52
Dewsbury	0.80	0.50
Leicester	0.91	0.51
Gloucester	0.80	0.36

[8] La relación de excedentes era necesaria para el cálculo del coeficiente de correlación

> La constancia de la asociación en los pueblos con índices de ataque con amplias diferencias es un punto digno de notar. Sheffield, Warrington, y Leicester exhiben asociaciones prácticamente idénticas, a pesar de que los índices de ataques varían de 7.9 a 2.5 y de 67.6 a 35.3.

Yule había consultado a Pearson para su publicación de 1900 sobre el coeficiente de asociación Q, y en 1904 Karl Pearson realizó una publicación sobre la contingencia y su relación con la asociación y la correlación normal, donde retoma su trabajo con el estadístico χ^2 que realizó en 1900 y el trabajo de asociación que llevó a cabo Yule en el mismo año.

Acerca del problema de la relación entre dos atributos, Pearson (1904) señaló que es posible clasificarlos en diferentes grupos en una tabla, de tal manera que esté formada por s columnas y t filas con el universo (frecuencia de la población) distribuida en los subgrupos que corresponden a los compartimentos $s \times t$. También Pearson reconoció el problema al que se enfrentaba cuando no se tiene un caso de asociación simple y no se puede aplicar una división fourfold[9] del universo, sino que se tienen grandes números de atributos, como la clasificación del color del ojo humano en ocho tipos y la correlación de éstos con seis clases de color de cabello. Para abordar este tipo de problemas introduce el concepto de contingencia y para realizar las conexiones con las nociones de asociación y correlación normal partió de la independencia de probabilidad, señalando que "*si p es la probabilidad de cualquier evento, y q la probabilidad de un segundo evento, entonces se dice que los dos eventos son independientes, si la probabilidad del evento combinado es $p \times q$*". (Pearson 1904, p. 5)

De acuerdo con Pearson (1904), si consideramos $A\ y\ B$ como atributos que pueden ser clasificados respectivamente en los grupos $A_1, A_2, \dots, A_s$ y $B_1, B_2, \dots, B_t$ de un número total de individuos N, el número de individuos que tiene cada uno de estos atributos son $n_1, n_2, \dots, n_s$ y $m_1, m_2, \dots, m_t$ respectivamente. Entonces la probabilidad de que un individuo se encuentre en alguno de estos grupos será $\frac{n_1}{N}, \frac{n_2}{N}, \dots, \frac{n_s}{N}$ y $\frac{m_1}{N}, \frac{m_2}{N}, \dots, \frac{m_t}{N}$ respectivamente. "*Y de acuerdo con el número de combinaciones de B_v con A_u que son esperados en la teoría de*

[9] Método para obtener el coeficiente de correlación dividiendo cualquier tabla de contingencia en tablas de 2 × 2. (Brownlee, 1910, pp. 476 - 480)

independencia en probabilidad si N pares de atributos son examinados es $N \times \frac{n_u}{N} \times \frac{m_v}{N} = \frac{n_u \cdot m_v}{N} = \nu_{uv}$" (p. 5)

Donde el número de los observados es n_{uv}. Atendiendo a los errores aleatorios en el muestreo se tiene la desviación de la probabilidad independiente en la ocurrencia de los grupos A_u, B_v. Y esta desviación esta en función de $n_{uv} - \frac{n_u m_v}{N} = n_{uv} - \nu_{uv}$.

Pearson (1904), denominó a cualquier medida de la desviación total de la clasificación de independencia probabilística una medida de su contingencia. A mayor medida de contingencia corresponde una mayor asociación o correlación entre dos atributos. Para realizar el cálculo de la medida de contingencia, Pearson retomó su prueba de bondad de ajuste con el estadístico χ^2 de 1900, vinculando lo anterior con "*si* $m'_1, m'_2, \ldots, m'_n$ *es un sistema de frecuencias observadas y donde* $m_1, m_2, \ldots, m_n$ *es un sistema de frecuencias teóricas conocidas a priori*" (p. 6), entonces $\chi^2 = Sum\left\{\frac{(m'_q - m_q)^2}{m_q}\right\}$ con $q = 0\ a\ n$, mientras que la probabilidad (P de χ^2) de que cualquier sistema de frecuencias observadas $m''_1, m''_2, \ldots, m''_n$ sea factible que ocurra con mayores desviaciones de $m_1, m_2, \ldots, m_n$, como el sistema realmente observado lo hace.

Aunado a esto, Pearson alude que al medir contingencia realmente estamos midiendo las desviaciones de los resultados observados de la probabilidad independiente, y por lo tanto si $m_1, m_2, \ldots, m_n$ corresponde al sistema teórico ν_{uv} y $m'_1, m'_2, \ldots, m'_n$ pertenece al sistema observado n_{uv}, entonces $\chi^2 = S\left\{\frac{(n_{uv} - \nu_{uv})^2}{\nu_{uv}}\right\}$. En este caso *P* es una medida de cuan lejos el sistema observado es o no compatible con las bases de la independencia probabilística. Estableciendo, por una parte, que cuando P es grande las posibilidades están a favor de que el sistema provenga de una probabilidad independiente y, por otra parte, cuando P es pequeña hay ciertamente una asociación entre los atributos. Por lo tanto, Pearson declaró como una medida apropiada de la contingencia a $1 - P$, lo que denominó grado de contingencia. En consecuencia, declaró que "*Claramente, a mayor contingencia, mayor debe ser la asociación o correlación entre los dos atributos, ya que dicha asociación o correlación es únicamente una medida desde otro punto de vista del grado de desviación de ocurrencia de independencia*". (Pearson, 1904, p. 6)

Pearson (1904) también estableció una función muy relacionada con el estadístico χ^2 y lo denominó contingencia cuadrada media $\emptyset^2 = \frac{\chi^2}{N}$ que, asumiendo normalidad para las frecuencias, es igual a $\frac{r^2}{(1-r^2)}$ y por lo tanto $r = \pm\sqrt{\frac{\emptyset^2}{1+\emptyset^2}}$. Precisando que esta relación se da entre la contingencia cuadrada media y la correlación normal.

Debido al método que Pearson (1904) adoptó para abordar el problema, no tuvo que considerar el signo de la contingencia, además la contingencia cuadrada media se encuentra basada en una suma de cuadrados extendida a todos los compartimentos de la tabla de contingencia $s \times t$. Sin embargo, cuando se realiza la sumatoria de todas las contingencias $n_{uv} - \nu_{uv}$ ésta debe ser cero, por lo que estableció como la contingencia media a la suma de todas las contingencias positivas $\psi = \sum \frac{n_{uv}-\nu_{uv}}{N}$. Pearson indicó que "*cualquiera de las dos funciones, $\emptyset^2$ o ψ podría servir como medida de contingencia*" (Ibíd., p. 7), y denominó primer coeficiente de contingencia a la expresión $\sqrt{\frac{\emptyset^2}{1+\emptyset^2}}$.

Sobre la relación entre la contingencia media y la correlación, Pearson (1904) brindó una tabla donde se integra ψ para los valores de r. Reconoció que a partir de esta tabla la correlación puede ser sencillamente encontrada por los valores de la contingencia media; sin embargo, también señala que este método le parece claramente inferior al de contingencia cuadrada media.

Las expresiones que estableció Pearson (1904) son formas de medida de la desviación de un sistema de probabilidad independiente, dicho de otro modo, de la cantidad de asociación entre los atributos involucrados e indicó como funciones para dichas mediciones:

a. El grado de contingencia, que es $1 - P$, donde P se encuentra a partir del estadístico χ^2 y con las tablas probabilidad para bondad de ajuste.
b. El coeficiente de contingencia cuadrada media $C_1 = \sqrt{\frac{\emptyset^2}{1+\emptyset^2}}$
c. El coeficiente de contingencia media C_2, que se encuentra a partir de la tabla que brinda, donde se integra ψ para los valores de r. Y en el caso de tener una agrupación lo suficientemente pequeña y con correlación normal señaló que $C_1 = C_2 = coeficiente\ de\ correlación.$

El caso de coeficientes iguales es sumamente limitante y no un caso general. No obstante, la igualdad de C_1 y C_2 puede significar una buena medida de la normalidad de la distribución. Para el cálculo de la probabilidad estableció una $n' = (x-1)(\lambda-1)/N - \emptyset^2$ donde x son el número de renglones y λ el número de columnas.

Pearson añadió el caso de poder dividir los datos sólo en cuatro grupos y establece la siguiente nomenclatura (Tabla 3.8):

Tabla 3.8. *Nomenclatura de tabla de contingencia de* $\mathbf{2 \times 2}$ (fuente: Pearson, 1904, p. 21)

a	c	$a+c$
d	b	$d+b$
$a+d$	$c+b$	N

Expresó $\emptyset^2$ y ψ en los terminos de asociación para dos atributos establecidos por Yule (1900), obteniendo:

$$\psi = \frac{2(ab-cd)}{N^2}$$

$$\emptyset^2 = \frac{(ab-cd)^2}{(a+d)(c+b)(a+c)(d+b)}$$

Pearson mostró de esta forma cómo $\emptyset^2$ y ψ se encontraban intimamente conectadas con la concepción de asociación de Yule (1900). También señaló que los coeficientes de contingencia C_1 y C_2, aunque sirven para medir asociación, no era útiles en este caso tan simple.

Para ejemplificar su propuesta, hizo uso de datos de una epidemia de viruela que había sucedido en 1890, y mostró la siguiente tabla (Tabla 3.9):

Ejemplo 4: La prueba de independencia con el estadístico χ^2 y los coeficientes de contingencia (Pearson, 1904, pp. 21-22)

Tabla 3.9. *Presencia de cicatriz por vacuna y recuperados de viruela* (fuente: Pearson, 1904, p. 21)

Cicatriz	**Recuperados**	**Muertos**	**Total**
Presente	1562	42	1604
Ausente	383	94	477

Total	1945	136	2081

Esto nos da $\emptyset^2 = 0.0845, \chi^2 = 175.76, \psi = 0.0604$. A partir de esto encontramos

$C_1 = 0.279, \qquad C_2 = 0.190$

Coeficiente de asociación de Yule = 0.803

Coeficiente de correlación por división fourfold = 0.595

Grado de contingencia = $1 - P$, [10] Donde $P = \frac{718}{10^{40}}$

Ahora en lo que respecta a los valores numéricos son totalmente diferentes. C_1, C_2, y el coeficiente de asociación dependen en gran medida de donde se toma la división fourfold[11]. Por lo tanto, es extremadamente difícil usarlos con fines comparativos. Por otra parte, sin embargo, el coeficiente de correlación con el supuesto de normalidad se encuentra libre de dicha restricción; esto nos alinea con otras cosas para fines comparativos. El grado de contingencia también es independiente en el sentido de la división, i.e., tiene definido un significado físico. Lo que esto nos dice es que la desviación de la independencia en relación entre el resultado, un caso de viruela, y la presencia o ausencia de cicatriz son tales que la tabla anterior solo podría ocurrir 718 veces en 1040 casos si los dos eventos fueran absolutamente independientes.

Si, en lugar de una tabla como la anterior, tomamos una serie de posibilidades alternativas para cada atributo, el coeficiente de asociación pierde su significado único; C_1 y C_2 aún mantendrán su importancia, y a medida que el número de alternativas se incrementa se unen al coeficiente de correlación. El grado de contingencia, por otra parte, conserva completamente el mismo significado. (Pearson, 1904, p. 22)

3.1.4 Los aportes posteriores a las pruebas con χ^2 de bondad de ajuste e independencia

[10] Cuando el número de grupos = 4, tenemos ('Phil. Mag.,' vol. 50, p. 157 et seq.):-

$$P = \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-\frac{1}{2}x^2} dx + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2} x = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2} x \left\{1 - \frac{1}{x^2} + \frac{1}{x^4} - \frac{3}{x^6} + \frac{15}{x^8}\right\}$$

Cuando P es sencillamente encontrado si χ^2 es grande. (nota del texto original)

[11] YULE, *loc. cit.*, p. 276. (nota del texto original)

Major Greenwood y Undy Yule trabajaron con asociación en tablas de contingencia con el estadístico χ^2 y mostraron que para una tabla de contingencia de 2×2, la distribución χ^2 tiene únicamente un grado de libertad (Barnard, 1992). Mientras que Fisher (1922a), señaló que, de forma general, la prueba con el estadístico χ^2 introducida por Pearson es adecuada, siempre y cuando el número de observaciones sea grande. Fisher retomó las tablas de bondad de ajuste que realizó W. Palin Elderton, publicadas en una colección de tablas para estadísticos y biométricos que editó Pearson en 1914, e introdujo una variación en la forma de calcular n', indicó que para las tablas de contingencia con c columnas y r filas se debe tener: $n' = (c-1)(r-1)+1$ en lugar de $n' = cr$. Esta variación provoca una gran diferencia en la probabilidad que se calcula a partir del valor del estadístico χ^2. Esta n' es una unidad mayor que la actual definición de grados de libertad.

Previamente Greenwood y Yule (1915), probaron el efecto de la inoculación contra la fiebre tifoidea y el cólera. Trabajaron con dos tablas de contingencia. En la primera tabla, con quienes fueron vacunados incluso durante el periodo de exposición a la infección; y en la segunda tabla, con los que fueron vacunados al momento en que llegaban del extranjero. Se calcularon los valores del χ^2 y la probabilidad, con las tablas de Elderton para $n' = 4$, para ambas tablas de contingencia y obtuvieron divergencias de las distribuciones esperadas en caso de que la vacuna no tuviera efecto en la posibilidad de contraer la enfermedad y que eran poco probables que se debieran a un error en el muestreo. Además, señalan que si se calcula el coeficiente normal de Pearson (cuando se calcula el coeficiente de correlación bajo el supuesto de una correlación normal entre dos variables) se llega a la misma conclusión.

Greenwood y Yule (1915), también notaron que, si se calcula la proporción atacada entre los vacunados y los no vacunados, la diferencia de las proporciones $(p' - p)$, dividida entre su error probable, podría ser una prueba de independencia. Sin embargo, puntualizaron una discrepancia cuando se comparan los resultados obtenidos por medio de la prueba con el estadístico χ^2 y la prueba mediante la diferencia de las proporciones y su error probable, y añadieron que dicha discrepancia "*es sorprendente y no se debe a un descuido de la correlación en los errores entre las frecuencias de los subgrupos*" (p. 118).

Por consiguiente, Fisher (1922a), parte de la dificultad que habían destacado Greenwood y Yule en 1915 e indica que "*cuando reconocemos que debemos tomar* $\mathrm{n}' = 2$ *, la dificultad desaparece*" (p. 90). Y utiliza la siguiente nomenclatura para una tabla de contingencia (Tabla 3.10) de 2×2:

Tabla 3.10. *Nomenclatura de tabla de contingencia de* **2** × **2** (fuente: Fisher, 1922a, p. 90)

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d$

Para Greenwood y Yule (1915), $p = \frac{a}{a+b}$ y $p' = \frac{c}{c+d}$. Fisher (1922a) propone que para el error estándar se tiene que $p = \sqrt{\frac{(a+c)(b+d)}{(a+b+c+d)^2(a+b)}}$ y $p' = \sqrt{\frac{(a+c)(b+d)}{(a+b+c+d)^2(c+d)}}$, entonces si $x = p' - p$ obtenemos:

$$\frac{x^2}{\sigma_x^2} = \frac{(bc-ad)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)} = \chi^2$$

Siendo χ, para $n' = 2$, distribuida normalmente en el lado positivo de la curva unicamente y con una desviacion estandar unitaria.

Así que el estadístico χ^2 se aplica, por una parte, en la prueba de bondad de ajuste para reconocer si una distribución de frecuencias observadas sigue cierta distribución teórica y, por otra parte, para establecer si existe asociación alguna entre dos variables discretas con tablas de contingencia, Fisher (1922a, p. 91) menciona que "las dos pruebas son idénticas cuando la prueba es aplicada correctamente". Posteriormente, Fisher presenta los cálculos para una tabla de contingencia con dos filas y s columnas, puntualizando que hasta el año anterior Pearson no había reconocido que "*las restricciones lineales*[12] *impuestas a las frecuencias de la población muestreada, por nuestros métodos de reconstrucción de esa población, tienen exactamente el mismo efecto sobre las distribuciones de la* χ^2 *que las restricciones impuestas sobre los contenidos de cada celda de la muestra*" (Ibíd., p. 92). Al trabajar con tablas de contingencia de forma análoga obtiene:

$\chi^2 = S\left\{\frac{NN'}{f+f'}\left(\frac{f}{N} - \frac{f'}{N'}\right)^2\right\}$, donde el número de grados de libertad es $s-1$.

[12] Número de parámetros que se pueden calculan bajo el supuesto de independencia directamente de los datos observados y a partir del cual se obtienen las otras frecuencias esperadas.

Por lo que Fisher concluye enfatizando que la prueba de bondad de ajuste con el estadístico χ^2 se puede aplicar a tablas de contingencia, pero se debe tomar a n' como el número de grados de libertad más uno para el cálculo de la probabilidad. Además, señala que esta prueba incluye tres casos especiales de los cuales nos interesan los siguientes (Fisher, 1992a, p. 93):

1. La comparación de las asociaciones en una tabla de 2×2.
2. El método de comparación de las distribuciones de muestras de Pearson.

En 1922, en la misma publicación que Fisher, Yule escribió un artículo donde llega a la misma conclusión para el cálculo de n'; sin embargo, en ese momento no logró probar ese resultado matemáticamente. Aunque le parecía muy razonable, señala que lo más cerca que estuvo de esa prueba fue reducir la expresión del estadístico χ^2 en términos de correlaciones y desviaciones estándar para una tabla de 3×3 (Yule, 1922). Yule también indicó que la expresión $P = \sqrt{\frac{2}{\pi}}\int_{\chi}^{\infty} e^{\frac{\chi^2}{2}} d\chi$ corresponde a $n' = 2$ y no a $n' = 4$, dependiendo de la prueba a aplicar. Para probar este supuesto en la práctica, realizó experimentos con tablas de 2×2 y obtuvo la probabilidad para los dos casos. El primero, es cuando el valor de χ^2 es calculado a partir de conocer a priori la distribución de frecuencias teóricas ($n' = 4$). En el segundo caso, encuentra el valor del estadístico para divergencia de los valores independientes, es decir, con una distribución teórica desconocida a priori ($n' = 2$). Yule (1922) señaló que para el segundo caso se obtienen valores más pequeños del χ^2, exhibiendo "*lo incorrecto de aplicar la misma ley en ambos casos*" (p. 100); y sugirió como resultado general para tablas de r renglones y c columnas que $n' = rc$ para el cálculo de la probabilidad en caso de conocer a priori las frecuencias teóricas (prueba de bondad de ajuste) y para el caso del cálculo de probabilidad de las divergencias a partir de "*calcular valores independientes del total de renglones y columnas el número de valores independientes de δ*[13] *es solo $(c-1)(r-1)$ y n' estaría dada por* $n' = (c-1)(r-1)+1$" (Ibíd., p. 97) para la prueba de independencia.

Las tablas de probabilidad de la distribución propuesta por Pearson para las tablas de contingencia de 2×2 se habían calculado con $n' = 4$, pero después de estos estudios se

[13] De acuerdo con (Yule, 1922) δ denota la diferencia entre la frecuencia (AB) y su valor independiente $\frac{(AB)}{N}$.

establece que $n' = 2$. Fisher propuso utilizar las calculadas por Elderton, siempre y cuando $n' = 2$. Aunque Yule, Greenwood, Pearson y Elderton realizaron trabajos y aproximaciones sobre n' para la distribución χ^2, se le atribuye a Fisher el introducir el número de grados de libertad en este contexto y los define, para una distribución de frecuencias con $n + 1$ clases, como el número de filas menos el número de restricciones lineares independientes (r) en las frecuencias, es decir, $gl = n - r$.

Fisher en 1925 publicó 'Statistical Methods for Research Workers', cuyo objetivo era brindar los medios para aplicar pruebas estadísticas con precisión, en el definía los contrastes de significación, donde se tenía una sola hipótesis, la hipótesis nula, que era la del investigador donde se establecía el valor numérico del parámetro. Las ideas centrales de los experimentos de Fisher, que se daban en el sector de la agricultura, era la aleatorización y la significación. La aleatorización no fue bien recibida por las dificultades que implicaba en el campo y fue rechazado por considerarse poco práctico, en cambio, la significación fue bien recibida pues parecía ser mucho más seguro cuando algunas diferencias pueden declararse significativas y las otras no (Pearce, 1992); el método de Fisher aparentemente brindaba certeza. De acuerdo con Pearce (1992) la prueba de significancia que proponía Fisher era "*la reducción al absurdo en un sentido probabilístico*" (p. 59), Fisher tenía su hipótesis nula en la que suponía que no había cambios, por ejemplo, el efecto de algún nuevo fertilizante era nulo, entonces si se quería demostrar que un nuevo fertilizante era bueno se suponía que no lo era y se intentaba falsear la hipótesis nula. En los años subsecuentes, Fisher siguió trabajando sobre las pruebas de significancia y en 1935 publicó 'The Design of Experiments' donde había desarrollado una teoría de inferencia basado en el análisis y diseño de experimentos. Fisher utilizó significancia "*como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria, no en el que se estableció su existencia*" (Pearce, 1992, p. 62), y señaló que él no inventó la idea de significancia, sino que se encontraba implícita en la prueba con el estadístico χ^2 que había introducido Karl Pearson años atrás.

Acerca de la importancia y uso de los grados de libertad y la significancia al aplicar la prueba de bondad de ajuste, Fisher proporcionó algunos ejemplos de los cuales se retoma el siguiente:

Ejemplo 5: Prueba de bondad de ajuste con el estadístico χ^2 y los grados de libertad (Fisher, 1934, pp. 84-85)

Comparación con la expectativa de las frecuencias de clase Mendelianas. - En un cruce que involucra a dos factores mendelianos nosotros esperaríamos obtener la generación de un híbrido (F1) de cuatro clases en la proporción $9 : 3 : 3 : 1$; la hipótesis en este caso es que los dos factores segregan de forma independiente, y que las cuatro clases de descendencia son igualmente viables. ¿Se encuentran las siguientes observaciones [*Tabla 3.11*] sobre las primaveras de jardín (de Winton y Bateson) en concordancia con la hipótesis?

Tabla 3.11. *Distribución de las frecuencias* (fuente: Fisher, 1934, p. 84)

	Hojas planas		**Hojas curvas**		**Total**
	Normal Eye	Primrose Queen Eye	Lee's Eye	Primrose Queen Eye	
Observado ($m + x$)	328	122	77	33	560
Esperado (m)	315	105	105	35	560
χ^2/m	0.537	2.752	7.467	0.114	10.870

Los valores esperados se calculan a partir del total observado, de modo que las cuatro clases deben estar de acuerdo en su suma, y si tres clases se completan arbitrariamente el cuarto, por lo tanto, es determinado; por lo tanto $n = 3$; $\chi^2 = 10.87$ la posibilidad de exceder tal valor es entre 0.01 y 0.02; si nosotros tomamos $P = 0.05$ como límite de desviación significativa, diremos que en este caso las desviaciones de las expectativas son claramente significativas.

Fisher (1934) consideraba que la prueba de independencia es útil cuando se tiene un mismo grupo que se clasifica en dos o más formas y se requiere saber si las dos clasificaciones son independientes. Además, señaló que el caso más simple es cuando la clasificación son dos clases, es decir, cuando trabajamos con una tabla de 2×2. Y retoma las clasificaciones (vacunados, no vacunados, atacados y no atacados) y los datos que utilizaron Greenwood y Yule sobre la tifoidea, brindando el siguiente ejemplo:

Ejemplo 6: Prueba de independencia con el estadístico χ^2 y los grados de libertad (Fisher, 1934, pp. 88-89)

Tabla 3.12. *Distribución de las frecuencias observadas* (fuente: Fisher, 1934, p. 88)

	Atacado	**No atacado**	**Total**

Inoculado	56	6759	6815
No inoculado	272	11396	11668
Total	328	18155	18483

Tabla 3.13. *Distribución de las frecuencias esperadas* (fuente: Fisher, 1934, p. 88)

	Atacado	**No atacado**	**Total**
Inoculado	120.93	6694.07	6815
No inoculado	207.07	11460.93	11668
Total	328	18155	18483

> Al probar independencia, debemos comparar los valores observados con los calculados para que las cuatro frecuencias sean proporcionales; ya que deseamos probar únicamente independencia, y no alguna hipótesis en cuanto a la números totales atacados, o inoculados, los valores 'esperados' se calculan a partir de los totales marginales observados, para que los números esperados estén de acuerdo a los números observado en los totales; solo se necesita calcular un valor, por ejemplo, $\frac{328\times 6815}{18483} = 120.93$; los otros se escriben a la vez por sustracción desde los totales. Por lo tanto, es obvio que los valores observados pueden diferir de los esperados en solo un grado de libertad, de modo que al probar la independencia en una tabla de 2×2, $n = 1$. Debido a que $\chi^2 = 56.234$ las observaciones son claramente opuestas a la hipótesis de independencia. Sin calcular los valores esperados χ^2, para tablas de 2×2, puede ser directamente calculada por medio de la formula: $\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$, donde a, b, c y d son los cuatro números observados. (Fisher, 1934, pp. 88-89)

Fisher (1934), señala que la prueba de independencia con el estadístico χ^2 "*no intenta medir el grado de asociación, pero como prueba de significación es independiente de todas las hipótesis adicionales en cuanto a la naturaleza de la asociación*" (p. 94).

Como la prueba con el estadístico χ^2 es aproximada y al aplicarla, como prueba de independencia en tablas de contingencia (e.g., 2×2), se requería que cada uno de los valores de las celdas fueran grandes y se distribuyeran como una normal. Para aplicar la prueba de independencia cuando se trabaja con número pequeños, en 1934 Yates propone un factor de corrección de continuidad para el estadístico χ^2. Analiza las discrepancias con el estadístico χ^2 y su probabilidad asociada posterior a la corrección de continuidad en tablas de contingencia que involucran números pequeños con uno o más grados de libertad.

Yates (1934, p. 217), indica que "*la precisión de la aproximación de la prueba depende de los números en las celdas, y en la práctica se ha considerado χ^2 como suficientemente exacta si ninguna celda tiene una expectativa menor a 5*". Así, enfoca su trabajo en la aplicabilidad de la prueba para tablas de contingencia con expectativas menores, debido a su importancia en la práctica de muchas ramas de la experimentación biológica, por la limitación que se tiene en los materiales experimentales.

Para Yates era importante la identificación de límites en las colas de la distribución del estadístico χ^2, así que asignó para n' un 2.5% (u otro) nivel de significancia para cada cola sujeta a discontinuidad, ya que se pueden clasificar todos los conjuntos posibles de observaciones que tienen $N - n$ y n de acuerdo al valor de n'. Todas las observaciones que cayeran fuera de estos límites las consideró significativas; añadiendo que "*la prueba de significación propuesta es una prueba válida, en el sentido de que, si no hay asociación, una proporción fija de los conjuntos de observaciones se considerará significativa, cualquiera que sea la p*" (Yates, 1934, p. 220). El hecho de tener un grado de libertad brinda una distribución definida y sus colas proporcionan las regiones apropiadas de significancia. Yates también elaboró una tabla donde presentó los valores críticos de la cola para diferentes m y p donde $m = \frac{nn'}{N}$ con $n' \geq n$ y $p = \frac{m}{n}$ y añadió que esta tabla se podría considerar para la construcción práctica de las pruebas de significancia, debido a que se podía comparar el valor calculado de χ' con los valores de la tabla al 2.5% o al 0.5%.

De acuerdo con Yates (1934), las discrepancias se deben a que χ es de una distribución continua, mientras que la distribución que se intenta aproximar es discreta. Si se calculan los valores de χ^2 para las desviaciones menos la mitad de una unidad que las desviaciones verdaderas, se le denomina corrección de continuidad y el valor resultante para χ es χ' y con una $P = (\chi')$. En otras palabras, esta corrección de continuidad consiste en restar 0.5 a las desviaciones positivas y sumar 0.5 a las deviaciones negativas. Actualmente se podría expresar el estadístico χ^2 con corrección por continuidad de la siguiente forma:

$$\chi_c^2 = \sum \frac{(|Frecuencia\ Obervada - Frecuencia\ Esperada| - 0.5)^2}{Frecuencia\ Esperada}$$

Yates (1934) define las tablas de contingencia de 2×2 bajo la siguiente notación, donde $n \leq n' \leq \frac{1}{2}N$ (Tabla 3.14):

Tabla 3.14. *Notación tabla contingencia* (fuente: Yates, 1934, p. 218)

	A	**No A**	**Total**
B	a	b	$N-n$
No B	c	d	n
Total	$N-n'$	n'	N

Yates (1934) también propone una forma de calcular la probabilidad de la distribución exacta con los totales marginales, donde la probabilidad correspondiente a cualquier termino $a, b, c, d,$ es: $\frac{n!\,n'!\,(N-n)!\,(N-n')!}{N!\,a!\,b!\,c!\,d!}$; es decir, "*el producto de los factoriales de los cuatro totales marginales dividido por el producto de los factoriales del gran total y los cuatro números de las celdas*" (p. 219). Para calcular los factoriales, sugiere el uso de una tabla de factoriales para cualquier término de la distribución.

Mientras que parte de $\chi^2 = \sum \frac{(a \times d - b \times c)^2\, N}{n \times n'(N-n)(N-n')}$ para establecer que el estadístico con corrección es $\chi_c^2 = \sum \frac{\left(\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right)^2 N}{n \times n'(N-n)(N-n')}$.

Yates (1934) proporciona un ejemplo acerca de la maloclusión, es decir, de la alteración de la posición de los dientes, de los bebés, tomando los datos obtenidos por M. Hellman, quien concluye que la alimentación con biberón es uno de los factores que causan la maloclusión.

Ejemplo 7: Prueba de independencia con factor de corrección de Yates

Tabla 3.15. *Maloclusión de dientes en infantes* (fuente: Yates, 1934, p. 230)

	Dientes normales	**Maloclusión**	**Total**
Amamantar	4	16	20
Biberón	1	21	22
Amamantar y biberón	3	47	50
Total	8	84	92

Primeramente, sólo consideró los datos de los bebés a los que únicamente se les había amamantado o a los que habían alimentado solo por biberón, como se muestra en la Tabla 3.16:

Tabla 3.16. *Maloclusión de dientes en infantes amamantados Vs alimentados por biberón* (fuente: Yates, 1934, p. 230)

	Dientes normales	Maloclusión	Total
Amamantados	4	16	20
Biberón	1	21	22
Total	5	37	42

Y realizó los siguientes cálculos (Yates, 1934, p. 231):

$$\chi^2 = \frac{(4 \times 21 - 1 \times 16)^2 \times 42}{5 \times 37 \times 20 \times 22} = 2.386,$$

$$\chi = 1.545, \qquad P(\chi) = 0.0612, \qquad P(\chi^2) = 2P(\chi) = 0.1224$$

$$\chi'^2 = \frac{(3.5 \times 20.5 - 1.5 \times 16.5)^2 \times 42}{5 \times 37 \times 20 \times 22} = 1.140$$

$$\chi' = 1.068, \qquad P(\chi') = 0.1427, \qquad P(\chi'^2) = 2P(\chi') = 0.2854$$

Tomando en cuenta que $P(\chi)$ y $P(\chi')$ se obtienen de χ y χ' por referencia a una tabla de la integral de probabilidad normal, la distribución exacta se calcula de la siguiente manera. La probabilidad de no obtener un hijo normal sí es amamantado es:

$$\frac{5!\ 37!\ 20!\ 22!}{42!\ 0!\ 20!\ 5!\ 17!} = 0.309568$$

Realizando los sucesivos multiplicadores mostrados se obtiene la tabla [Tabla 3.17] completa:

Tabla 3.17. *Probabilidad de No. niños amamantados* (fuente: Yates, 1934, p. 231)

No. de niños normales amantados	Multiplicador	Probabilidad
0	$-$	0.309568
1	$5.20/1.18$	0.171982
2	$4.19/2.19$	0.343964
3	$3.18/3.20$	0.309568

4	$2.17/4.21$	0.125301	
5	$1.16/5.22$	0.018226	0.143527
		0.999998	

> Por lo tanto, la probabilidad real de obtener cuatro o más niños normales alimentados con leche materna (amamantados) es 0.1435. $P(\chi')$ da 0.1427, una excelente aproximación, mientras que $P(\chi)$ da 0.0612, que, aunque en sí misma no es significativa, es menor que la mitad del verdadero valor; esto sería sumamente engañoso si se combinaran varias de estas probabilidades de diferentes clases de experimentos. (Yates, 1934, p. 231).

"χ' está tan lejos del 2.5% para cualquier m y p […] El valor menor del 2.5% de χ' en la Tabla III es 1.68" (Yates, 1934, p. 231). La Tabla III presenta los valores críticos de la cola para diferentes m y p.

Aunque el procedimiento para realizar el cálculo del χ^2 de Fisher es diferente, obtiene el mismo valor que Yates antes de realizar la corrección de continuidad, como se ejemplifica a continuación:

$$\chi^2 = \frac{(bc - ad)^2(a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}$$

$$\chi^2 = \frac{\big((16 \times 1) - (4 \times 21)\big)^2(4 + 16 + 1 + 21)}{(4 + 16)(1 + 21)(4 + 1)(16 + 21)} = 2.386$$

Siendo $n' = (c - 1)(r - 1) + 1 = (2 - 1)(2 - 1) + 1 = 2$, por lo tanto $P(\chi) = 0.0612$.

Por lo que no existe independencia entre la maloclusión y la forma en que se alimenta al niño (amamantándolo o por biberón).

Para ilustrar la aplicación de la corrección de continuidad de Yates, en la prueba de bondad de ajuste, y el impacto que tiene sobre el estadístico χ^2, cuando se presentan frecuencias esperadas menores a cinco, se muestra el ejemplo 8 que se ha retomado de la publicación de 1900 de Pearson (pp. 169-170):

> En mis memorias sobre sesgo-variación (Phil. Trans, vol. clxxxvi. p. 401) he ajustado las estadísticas para la frecuencia de pétalos en 222 ranunculus con la curva sesgada $y = 0.211225x^{-0.322}(7.3253 - x)^{3.142}$.
>
> El rango posible es de 5 a 11 pétalos, y las frecuencias son:

Tabla 3.18. *Distribución de frecuencias del No. de pétalos de ranunculus* (fuente: Pearson, 1900, p. 169)

No Pétalos	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Observados	133	55	23	7	2	2	0
Teóricos	136.9	48.5	22.6	9.6	3.4	0.8	0.2

> Estos conducen a $\chi^2 = 4.885{,}528$; de donde encontramos la probabilidad de un sistema de desviaciones tan o más remoto de lo más probable $P = 0.5586$.
>
> En una selección aleatoria de 56 casos de un centenar de tales ensayos deberíamos obtener resultados más improbables que los que hemos obtenido. Por lo tanto, podemos considerar el ajuste notablemente bueno. (Pearson, 1900, pp. 169-170)

Con la distribución de frecuencias presentadas en la Tabla 3.18 se obtiene el valor del estadístico a partir de $\chi_c^2 = \sum \frac{(|Frecuencia\ Obervada - Frecuencia\ Esperada| - 0.5)^2}{Frecuencia\ Esperada}$; siendo $\chi_c^2 = 2.587262$ menor a el valor de χ^2 obtenido por Pearson en 1900. Además, si se considera los grados de libertad como $k = n - r = 7 - 1 = 6$ y se busca la probabilidad en la tabla de la distribución χ^2, se obtiene que $P = .858576$. Entonces, de acuerdo a la probabilidad obtenida se puede decir que la curva sesgada propuesta por Pearson al inicio del problema presenta un buen ajuste para la distribución de los pétalos de las ranunculus.

3.1.5 La prueba de homogeneidad con el estadístico χ^2

Pearson (1911), retoma como base el problema que detonó el tratamiento de la bondad de ajuste en 1900, sin embargo, señaló que es un poco diferente esta vez porque se "*tienen dos muestras, y a priori pueden ser de la misma población o de diferentes poblaciones; deseamos averiguar cuál es la probabilidad de que sean muestras aleatorias de la misma población. Sin embargo, de esta población no se tiene experiencia a priori*" (p. 250). Pearson indicó algunos ejemplos del tipo de problemas en que sería deseable tener dicha información sobre las muestras, como cuando se tenían registros sobre el número de habitaciones en casas donde alguno de los habitantes (1) había tenido cáncer (2) había tenido tuberculosis; añadiendo que la cantidad de casos de cada la enfermedad puede ser bastante diferente y que posiblemente no conoce la distribución de frecuencia del número de habitaciones en el distrito dado. En concordancia con este problema, planteó la siguiente pregunta ¿Cuál es la posibilidad de que haya una diferencia significativa en las casas donde se presentaron casos de tuberculosis y las de cáncer?

Pearson (1911) parte suponiendo que la población, de la cual son las dos muestras, puede estar dada por las frecuencias de clase $\mu_1, \mu_2, \mu_3, \mu_4 \ldots \mu_p, \mu_q \ldots \mu_g$, siendo el total de la población M. Las muestras estarían dadas entonces por frecuencias en las mismas clases:

									Total
Muestra 1	f_1	f_2	f_3	...	f_p	f_q	...	f_g	N
Muestra 2	$f_1{}'$	$f_2{}'$	$f_3{}'$	...	$f_p{}'$	$f_q{}'$	...	$f_g{}'$	N'

Pearson también supuso que las dos muestras eran absolutamente independientes y por lo tanto no habría una correlación entre cualquier desviación en alguna frecuencia de la primera fila y de la segunda fila. Además, sobre las dos distribuciones de frecuencia se tiene que:

$$\sum\nolimits_p 2 = N\frac{\mu_p}{M}\left(1-\frac{\mu_p}{M}\right), \sum\nolimits_p {}'2 = N'\frac{\mu_p}{M}\left(1-\frac{\mu_p}{M}\right)$$

$$\Sigma_p \Sigma_q \mathrm{R}_{pq} = -N\frac{\mu_p}{M}\frac{\mu_q}{M}, \qquad \Sigma_p{}'\Sigma_q{}'\mathrm{R}_{pq}{}' = -N'\frac{\mu_p}{M}\frac{\mu_q}{M}$$

Siendo $\Sigma_p, \Sigma_q, \Sigma_p', \Sigma_q'$ las desviaciones estandar de las frecuencias de la pth y qth frecuencias de las dos muestras y R_{pq}, R_{pq}' las correlaciones de las mismas frecuencias. A partir de reducir las frecuencias de las dos muestras a un total estándar común n, consideró un nuevo sistema de variables x_p, x_q y sobre la hipótesis de que ambas frecuencias son de muestras aleatorias de la misma población obtuvo:

$$\frac{\bar{f}_p}{N} = \frac{f_p'}{N'} = \frac{\mu_p}{M}, \qquad \frac{\bar{f}_q}{N} = \frac{f_q'}{N'} = \frac{\mu_q}{M}$$

Y $\bar{x}_q$ y $\bar{x}_p$ serán cero. Entonces, σ_p y σ_q son las desviaciones estandar de x_p y x_q, mientras que r_{pq} es la correlación de las mismas. Después, Pearson tomó en cuenta la independencia de las muestras y obtuvo $\sigma_p{}^2 = n^2\left(\frac{\Sigma_p{}^2}{N^2} + \frac{\Sigma_p{}^{2\prime}}{N'^2}\right)$ y $\sigma_p\sigma_q r_{pq} = n^2\left(\frac{\Sigma_p\Sigma_q\mathrm{R}_{pq}}{N^2} + \frac{\Sigma_p'\Sigma_q'\mathrm{R}_{pq}'}{N'^2}\right)$. Además, si

${\sigma_p}^2 = n^2\left(\frac{1}{N}+\frac{1}{N'}\right)\frac{\mu_p}{M}\left(1-\frac{\mu_p}{M}\right)$ y $\sigma_p\sigma_q r_{pq} = -n^2\left(\frac{1}{N}+\frac{1}{N'}\right)\frac{\mu_p\mu_q}{M^2}$ entonces ${\mu_p}' = n^2\left(\frac{1}{N}+\frac{1}{N'}\right)\frac{\mu_p}{M}$, ${\mu_q}' = n^2\left(\frac{1}{N}+\frac{1}{N'}\right)\frac{\mu_q}{M}$ y $M' = n^2\left(\frac{1}{N}+\frac{1}{N'}\right)$; por lo tanto, $\frac{{\mu_p}'}{M'} = \frac{\mu_p}{M}$, y como ${S_1}^g(\mu_p) = M$ por consiguinete ${S_1}^g({\mu_p}') = M'$ se tiene que M' es la población total de una distribución de frecuencias ${\mu_1}', {\mu_2}', {\mu_3}', {\mu_4}' \dots {\mu_p}', {\mu_q}' \dots {\mu_g}'$.

Se tiene también que ${\sigma_p}^2 = {\mu_p}'\left(1-\frac{{\mu_p}'}{M'}\right)$ y $\sigma_p\sigma_q r_{pq} = -\frac{{\mu_p}'{\mu_q}'}{M'}$, como medida de un sistema teórico ${\mu_1}', {\mu_2}', {\mu_3}', {\mu_4}' \dots {\mu_p}', {\mu_q}' \dots {\mu_g}'$, tienen idéntico tipo de desviación estándar y de correlación que la distribución de frecuencias del sistema $x_1, x_2, x_3, x_4 \dots x_p, x_q \dots x_g$.

Pearson (1911), señala que se tiene la siguiente forma del estadístico $\chi^2 = {S_1}^g\left(\frac{{x_p}^2}{{\mu_p}'}\right)$ debido a que las ecuaciones $\sigma_p\sigma_q r_{pq}$ y ${\mu_p}'$ estan acorde con lo que había expresado en Pearson (1900).

Entonces $\chi^2 = {S_1}^g\left\{\frac{\left(\frac{f_p}{N}-\frac{{f_p}'}{N'}\right)^2}{\left(\frac{1}{N}+\frac{1}{N'}\right)\frac{\mu_p}{M}}\right\}$

En cuanto a la mejor hipótesis sobre la constitución de la población, Pearson (1911, p. 252) señaló que “*la indicada será la* pth *frecuencia de clase para las dos muestras combinadas*”; expresado de otra forma: $f_p + {f_p}'$ es proporcional a μ_p, o $(f_p + {f_p}')/N + N' = \mu_p/M$.

$$\chi^2 = {S_1}^g\left\{\frac{NN'\left(\frac{f_p}{N}-\frac{{f_p}'}{N'}\right)^2}{f_p + {f_p}'}\right\}$$

Pearson (1911) puntualizó que con esta expresión el estadístico χ^2 no presenta dificultades en casos reales. También brinda algunos ejemplos para mostrar cómo aplicar la prueba, tomando los datos de todos ellos de estudios del Dr. Macdonald, pretendiendo determinar si había una enfermedad selectiva. A continuación, retomamos uno de ellos.

Ejemplo 9 (Pearson, 1911, pp. 252-253): Prueba de homogeneidad con el estadístico χ^2.

> Preguntemos si el color del cabello ejerce una selección diferencial con respecto a la incidencia de fiebre escarlatina y sarampión.

Los siguientes datos son proporcionados por el Dr. Macdonald, Biometrika, vol. VIII págs. 28, para todos los casos de fiebre escarlatina y sarampión:

Tabla 3.19. *Distribución de frecuencias de dos muestras* (fuente: Pearson, 1911, p. 252)

		Color de cabello						
		Negro	Oscuro	Medio	Claro	Rojo		Totales
Fiebre escarlatina	(i)	12	289	1109	360	94	f	1864
Sarampión	(ii)	0	85	367	184	25	f'	661
$(\mathbf{i}) + (\mathbf{ii})$	(iii)	12	374	1476	544	119	$f + f'$	2525
$(\mathbf{i})/\mathbf{1864}$	(iv)	0.0064	0.1551	0.5950	0.1931	0.0504	f/N	1.0000
$(\mathbf{ii})/\mathbf{661}$	(v)	0.0000	0.1286	0.5552	0.2784	0.0398	f'/N'	1.0000
$(\mathbf{iv}) - (\mathbf{v})$	(vi)	+0.0064	+0.0265	+0.0398	-0.0853	+0.0126	$f/N - f'/N'$	-
$(\mathbf{vi})^2$	(vii)	0.000,041	0.000,702	0.001,584	0.007,276	0.000,159	$(f/N - f'/N')^2$	-
$(\mathbf{vii}) \div (\mathbf{iii})$	(viii)	0.000,0034	0.000,0019	0.000,0011	0.000,0134	0.000,0013	$\frac{(f/N - f'/N')^2}{f + f'}$	0.000,0211

Por lo tanto, $\chi^2 = NN' \times 0.000{,}0211 = 1864 \times 661 \times 0.000{,}0211 = 26.00$

P de las tablas (Biometrika, Vol. I. p. 155) es aproximadamente 0.000,03. En otras palabras, las probabilidades son más de 33,000 a 1 contra la aparición de dos muestras de color de cabello tan divergentes, si fueran muestras aleatorias de la misma población. Considero que podemos concluir que son muestras realmente diferenciadas, o que la fiebre escarlatina y el sarampión no atacan indiferentemente a todos los individuos cualquiera que sea la pigmentación de su cabello; o, que la fiebre escarlatina y el sarampión son diferenciales en su selección.

Pearson finalizó agregando que existen muchos otros problemas en los cuales este método puede ser aplicado.

Por su parte, George W. Snedecor (1930) presentó una técnica estadística que denominó experimental, la prueba estadística de homogeneidad. Indicó que la prueba era aplicable al tipo de datos conocido como estadísticas homogéneas o las estadísticas de atributos y que trabajaba con este tipo de datos cuando los elementos o individuos de una muestra se pueden clasificar en categorías.

Inicialmente, Snedecor ilustró como se ha utilizado la prueba cuando se tiene un gran número de observaciones con un ejemplo sobre la distribución de las camadas de cerdos de acuerdo al número de machos en la camada (de cero a ocho) y se calculó la frecuencia esperada bajo el supuesto de una distribución binomial. El problema que expresó fue "*probar si este conjunto de desviaciones de lo esperado es un conjunto tal como se podría esperar que ocurra en el muestreo aleatorio, o si las desviaciones son demasiado grandes para justificar tal suposición*" (Ibíd., p. 280). Para abordar el problema usó el estadístico χ^2 y concluyó que la variación de la distribución observada con respecto a la esperada no es mayor a lo usual en muestras de ese tamaño, además agregó que "*esta es simplemente una muestra que puede ser normalmente extraída de una población homogénea, en la que hay una probabilidad uniforme de masculinidad. Esta es la prueba de homogeneidad*" (Ibíd., p. 280). Sin embargo, presentó también otra distribución de las camadas de cerdos perteneciente una muestra más grande, donde de acuerdo con el estadístico χ^2 obtenido, la muestra no pudo haber sido extraída al azar de una población homogénea. Con respecto a esta segunda distribución de las camadas de cerdos, señaló que con los resultados obtenidos se podía afirmar que la variación de la distribución observada con respecto a la esperada era mayor de lo que puede ser explicado por las exigencias del muestreo aleatorio, o que también se podría señalar que la población no era homogénea para la masculinidad de los cerdos, e incluso que el muestreo no se llevó a cabo correctamente. También consideraba que el método anteriormente descrito era la forma clásica de aplicar la prueba en muestras grandes, pero para el caso de números pequeños de submuestras retomó el método de Fisher.

En 1933, G. W. Snedecor y M. R. Irwin, en la publicación 'On the Chi-square test for homogeneity', presentaron una propuesta para la prueba de homogeneidad con el estadístico χ^2 y retoman el método de Fisher (1932) para los factores que expliquen la heterogeneidad de la población.

Snedecor e Irwin (1933), destacaron los resultados de experimentos que se llevan a cabo en laboratorio donde las frecuencias de las observaciones que provienen de distintos subconjuntos son desiguales, ya que con este tipo de datos había estado trabajando Snedecor, al respecto señalaron algunos ejemplos, como la mortalidad en epidemias inducidas a animales de laboratorio e infestaciones controladas a campos de cultivo. Aunado a lo anterior, indicaron que, si los datos pueden dividirse en subconjuntos, se podría determinar la probabilidad de que la serie de submuestras se haya extraído de una población homogénea, en el sentido de que la probabilidad del evento es uniforme a lo largo del material experimental. Esta forma de expresar la probabilidad es diferente a interpretar la probabilidad a favor o en contra de un evento y como la probabilidad media de todo el conjunto, como se había estado utilizando. Los autores añadieron que cuando se trabaja con material experimental en el laboratorio es difícil tener acceso a subconjuntos del mismo tamaño en una serie de experimentos, este problema lo había resuelto Pearson (1911) y Arnee Fisher en 1922. Snedecor (1930), había propuesto un método especialmente adaptado para datos que se obtienen en pruebas de inmunología y este método lo retomó Fisher en 1932 en la cuarta edición de 'Statistical methods for research workers'.

Para mostrar su propuesta de prueba de homogeneidad, brindaron los siguientes datos (Snedecor e Irwin, 1933, pp. 76-77):

Ejemplo 10: Prueba de homogeneidad con el estadístico χ^2 aplicado a datos de experimentos de inyecciones de bacilo Danysz a roedores.

> Los datos […] provienen de experimentos realizados por el autor menor (Irwin, 1929), en el que se inyectaron dosis controladas de un organismo específico en grupos de animales de laboratorio. Los tamaños de estos grupos (submuestras) fueron dictados por el número de animales experimentales disponibles en cada fecha de inyección.

Tabla 3.20. *Resultados obtenidos de inyectar quince submuestras de una población de ratas con una dosis estándar de bacilo Danysz* (fuente: Snedecor e Irwin, 1933, p. 76)

No. submuestra	**No. inyectados n**	**No. muertos s**	**Porcentaje de mortalidad p**
1	40	31	77.50000
2	12	10	83.33333
3	22	19	86.36364
4	26	22	84.61538
5	43	36	83.72093
6	25	21	84.00000
7	17	14	82.35294

8	20	17	85.00000
9	11	10	90.90909
10	37	35	94.59459
11	39	35	89.74360
12	47	47	100.00000
13	29	22	75.86207
14	43	31	72.09302
15	20	15	75.00000
	$\sum n = 431$	$\sum s = 365$	$\bar{p} = 84.68678\%$
$\bar{\boldsymbol{p}} = \mathbf{15.31322}\%$		$\chi^2 = 22.62$	$P = 0.07$

En cualquier submuestra la probabilidad de muerte está dada por $p = \frac{100s}{n}$ %. La probabilidad promedio (media ponderada) para la muestra completa es $\bar{p} = \frac{100 \sum s}{\sum n}$ %, mientras que la probabilidad media de supervivencia es $\bar{q} = 100 - \bar{p}$ %. El problema es probar la hipótesis de homogeneidad en la reacción de estos grupos de ratas a la inyección. La fórmula para Chi-cuadrada es $\chi^2 = \frac{100\,(\sum sp - \bar{p} \sum s)}{\bar{p}\bar{q}}$ en el que $\sum sp$, es la suma de los productos de la pareja de números de la tercer y cuarta columna de la tabla [*Tabla 3.20*], se calcula fácilmente en una calculadora como sigue, $\sum sp = (31)(77.50000) + (10)(83.33333) + \cdots = 31{,}203.95$.

Sustituyendo este y los otros valores especificados tomados de la tabla en la fórmula para Chi-cuadrada, tenemos

$$\chi^2 = \frac{100[31{,}203.95 - (84.68678)(365)]}{(84.68678)(15.31322)} = 22.62$$

La prueba se completa utilizando una tabla la distribución χ^2 con grados de libertad, $15 - 1 = 14$, y observando la probabilidad, $P = 0.07$, que muestras similares de chi-cuadrada, extraídas al azar de la misma población homogénea, sería más grande. Aunque esta es una probabilidad bastante pequeña, todavía varios grupos inoculados pueden considerarse tentativamente como submuestras extraídas de una población en la que existe una probabilidad uniforme de muerte. Dentro de los límites del error experimental, la población animal puede por el momento asumirse como homogénea y la técnica adecuada.

También mostraron otro ejemplo donde los datos pertenecían a una población heterogénea y señalan que es una ventaja probar homogeneidad antes de un análisis estadístico adicional (Ibíd., pp. 79- 80):

Ejemplo 11: Prueba de homogeneidad con el estadístico χ^2 aplicado para probar la eficiencia de dos métodos de aplicación de insecticidas

Estos datos se deben a Hartzell (1929), se toman de un experimento diseñado para probar la eficacia de dos métodos de aplicación de insecticidas. Las recogieron manzanas de ocho árboles en cada parcela.

Tabla 3.21. *Lesión de la polilla de manzana en dos parcelas con tratamiento diferent*e (fuente: Snedecor e Irwin, 1933, p. 79)

Parcela roseada Azufre de cal y arseniato de plomo			**Parcela espolvoreada 90 - 10 polvo de arseniato de azufre de plomo**		
No. manzanas examinadas	No. manzanas lesionadas	Porcentaje de lesionadas	No. manzanas examinadas	No. manzanas lesionadas	Porcentaje de lesionadas
1804	102	5.65	1083	118	10.90
1811	88	4.86	1011	48	4.75
860	2	0.23	946	128	13.53
1671	7	0.42	840	37	4.41
1078	11	1.02	2347	41	1.75
1204	9	0.75	2404	69	2.87
1199	17	1.42	2548	38	1.49
2149	13	0.60	2376	38	1.60
11776	249	2.114	13555	517	3.814
$\boldsymbol{\chi^2 = 257}$		P = 0	$\chi^2 = 497$		P = 0

Los valores de χ^2 indican que no existe una probabilidad constante de lesión bajo cualquier método. La diferencia entre las probabilidades promedio (ponderadas), $3.814 - 2.114 = 1.700\%$ puede deberse a la diferencia en el tratamiento, pero, por otro lado, puede deberse a causas desconocidas que producen la heterogeneidad dentro de las dos muestras. Bajo estas circunstancias, una prueba de significancia de esta diferencia es de dudoso valor. Sin embargo, si uno desea hacerlo, está limitado en su elección de un método. No puede usar el valor de $\bar{p}$ en la fórmula de Pearson (1907) porque se acaba de demostrar que no tiene valor uniforme en las submuestras. Si debe usar la prueba χ^2 para independencia en una tabla de 2×2, ignoraría la heterogeneidad de los subconjuntos demostrada. Como se trata de pequeñas muestras el método apropiado es el de Fisher (1932). Para esto la suma del cuadrado de las desviaciones de la media es requerida para cada parcela. Estos son fácilmente obtenidos de la fórmula para la varianza ponderada [...] $\sigma^2 = \frac{100(\sum sp - \bar{p}\sum s)}{\sum n}$.

Como el numerador es el mismo que el de la fracción de χ^2, el cálculo adicional es trivial. Las variaciones en las dos parcelas de manzanos son $\sigma_1^2 = 4.5104$ y $\sigma_2^2 = 13.4542$.

De estos, las sumas de cuadrados se derivan simplemente por multiplicar la varianza ponderada por ocho, el número de submuestras en cada parcela. Usando el método de cálculo de Fisher se obtiene la siguiente tabla [*Tabla 3.22*].

Tabla 3.22. *Cálculos para la estimación de la varianza agrupada (*fuente: Snedecor e Irwin, 1933, p. 80)

	No. de árboles, N	**Grados de libertad**	**Varianza, σ^2**	**Suma de cuadrados, $N\sigma^2$**
Parcela roseada	8	7	4.5104	36.083
Parcela espolvoreada	8	7	13.4542	107.634
Totales		14		143.717

Estimación de la varianza agrupada $= \frac{143.717}{14} = 10.265$

Varianza de la diferencia de medias $= 10.265\left({}^{1}/_{8} + {}^{1}/_{8}\right) = 2.566$

Desviación estándar de la diferencia de medias $= \sqrt{2.566} = 1.602\%$

$t = \frac{1.700}{1.602} = 1.061 \qquad P = 0.3$

> La conclusión es que la diferencia entre los porcentajes de manzanas lesionadas en las dos parcelas es demasiado pequeña para ser atribuida a otra cosa que no sea variaciones al azar en submuestras de la misma población. La falta de la homogeneidad puede deberse a la técnica de aplicación del insecticida o a algunas características desconocidas del ataque severo de los insectos. No es posible una prueba crítica del tratamiento en presencia de tan pronunciada heterogeneidad en el porcentaje de frutos lesionados.

De acuerdo con Fisher (1934), las pruebas de homogeneidad y de independencia con el estadístico χ^2 son matemáticamente idénticas. También señaló que "*el índice de dispersión χ^2 podría ser equivalente a la χ^2 obtendida de la tabla de contingencia*" (p. 95). La prueba de homogeneidad se puede aplicar a muestras de igual o diferente tamaño, y presenta algunos ejemplos sobre su aplicación en casos donde las muestras no son del mismo tamaño. El ejemplo 12 es sobre homogeneidad de diferentes familias con respecto a la proporción ojos color negro y rojo.

Los siguientes datos (Tabla 3.23) muestran los números de ojos rojos y negros en 33 familias de Gammarus (datos de Huxley).

Tabla 3.23. *No. de ojos rojos y negros en familias de Gammarus* (fuente: Fisher, 1934, p. 95)

Negro	79	120	24	117	62	79	66	45	61	64	208	154	31	158	21	105	28
Rojo	14	31	6	29	17	20	12	11	14	13	52	45	4	45	4	28	7
Total	93	151	30	146	79	99	78	56	75	77	260	199	35	203	25	133	35
Negro	58	81	25	95	47	67	30	70	139	179	129	44	24	19	45	91	2565
Rojo	19	27	8	29	16	21	11	28	57	62	44	17	9	8	23	41	772
Total	77	108	33	124	63	88	41	98	196	241	173	61	33	27	68	132	3337

> Los totales negro 2565 y rojo 772 claramente no son en la proporción $3 : 1$; la discrepancia se atribuye al vínculo. La pregunta que tenemos ante nosotros es si todas las familias indican la misma proporción entre negro y rojo, o si la discrepancia se debe

únicamente a algunas familias. Para toda la tabla $\chi^2 = 35.620, n = 32$. Esto es más allá del rango de la tabla, entonces aplicamos el método[14] explicado en la p. 62:

$$\sqrt{2\chi^2} = 8.44;$$

$$\sqrt{2n-1} = 7.94;$$

$$Diferencia = +0.50 \pm 1$$

Por lo tanto, la serie no es significativamente heterogénea; efectivamente todas las familias están de acuerdo y confirman entre sí la proporción negro-rojo observada en el total.

Exactamente el mismo procedimiento se adoptaría si los números negros y rojos representaran dos muestras distribuidas de acuerdo con algún carácter o caracteres en las 33 clases. La pregunta "¿Son estas muestras de la misma población?" es en efecto idéntico con la pregunta "¿Es la proporción de negro a rojo igual en cada familia?" Reconocer esta identidad es importante, ya que ha sido ampliamente ignorada. (Fisher, 1934, pp. 95-96)

Con la finalidad de mostrar el efecto que tiene la corrección de continuidad de Yates sobre el estadístico χ^2 en la prueba de homogeneidad, se retoma el ejemplo 9 que se presentó en la página 106 que Pearson en 1911 utilizó con la finalidad de ilustrar cómo funcionaba la prueba de homogeneidad cuando se intentaba conocer si dos muestras pertenecen a la misma población. Para efectos del análisis de las configuraciones a este ejemplo, que ahora implica la corrección de Yates, se le denomina ejemplo 13. Con la distribución de las frecuencias observadas de la Tabla 3.23 se han calculado las frecuencias esperadas, específicamente los valores fueron calculados a partir de los marginales totales observados y pueden observarse en la Tabla 3.24.

Ejemplo 13: Prueba de bondad de ajuste con la corrección para el estadístico χ^2.

[14] Para grandes valores de n se hace uso del hecho de que la distribución de χ se vuelve casi la distribución normal. Una buena aproximación se da asumiendo que la expresión $\left(\sqrt{2\chi^2} - \sqrt{2n-1}\right)$ esta normalmente distribuida alrededor de cero con desviación estándar unitaria. Si el valor de la expresión es mayor a dos, el valor de χ^2 no está de acuerdo a las expectativas (Fisher, 1934, p. 62).

Tabla 3.24. *Distribución de frecuencias de color del cabello en enfermos de fiebre escarlatina y sarampión*

	Color de cabello					
	Negro	Oscuro	Medio	Claro	Rojo	Total
Fiebre escarlatina	8,85861	276,09347	1089,60950	401,59050	87,84792	1864
Sarampión	3,14139	97,90653	386,39050	142,40950	31,15208	661
Total	12	374	1476	544	119	2525

Con las frecuencias observadas y las esperadas se obtiene el valor del estadístico $\chi^2 = 23.8388$. Existe diferencia del valor obtenido por Pearson (1911), que era $\chi^2 = 26$. Se busca en la tabla de la distribución χ^2 con 4 grados de libertad, debido a que para tablas de contingencia de r renglones y c columnas $n = (r-1)(c-1)$, para $n = 4$ se tiene que $P = 0.000{,}086039$. Por lo tanto, las probabilidades son alrededor de 11,628 a 1 contra la aparición de dos muestras de color de cabello tan divergentes, si fueran muestras aleatorias de la misma población.

La prueba de homogeneidad con es el estadístico χ^2 como la trabajaron Snedecor e Irwin estaba enfocada a probar la homogeneidad de la población evaluando la varianza en las muestras, mientras que las pruebas de homogeneidad que realizaron Pearson y Fisher con el mismo estadístico buscaban si dos muestras de datos pertenecen a misma población, o bien, si la distribución de la variable para dos grupos datos es igual a través de las proporciones.

3.2 Configuraciones ontosemióticas del estadístico χ^2

A continuación, se describen las configuraciones ontosemióticas epistémicas identificadas con el recorrido histórico del estadístico χ^2. Se parte de las tres principales problemáticas que permitieron el surgimiento, desarrollo y generalización de dicho objeto matemático: prueba de bondad de ajuste, prueba de independencia y prueba de homogeneidad. Estas problemáticas se encuentran estrechamente vinculadas a la problemática: distribución χ^2. Cada una de estas problemáticas da cuenta de un significado para el estadístico χ^2.

El significado holístico del estadístico χ^2 esta conformado por doce significados parciales, para determinar cada significado parcial se ha realizado una configuración ontosemiótica a partir de la práctica matemática desplegada para solucionar las mencionadas problemáticas.

La herramienta de configuración ontosemiótica corresponde al segundo nivel de análisis de la actividad matemática y con esta herramienta se han analizado las redes epistémicas o configuraciones que forman los objetos primarios (elementos lingüísticos, situaciones/problemas, conceptos/definiciones, proposiciones/propiedades, procedimientos y argumentos) cuando se relacionan entre sí.

3.2.1 Significado 1: Chi-cuadrada en la prueba de bondad de ajuste

Una prueba de bondad de ajuste con el estadístico χ^2 refiere a una medición objetiva o razonable sobre si las frecuencias observadas de una muestra corresponden a determinada distribución teórica.

En el desarrollo histórico de la prueba de bondad de ajuste con el estadístico χ^2, se identificaron cuatro significados parciales. El primero de ellos, corresponde a una prueba de bondad de ajuste intuitiva, pues en las prácticas matemáticas aún no se utilizaba explícitamente el estadístico χ^2.

3.2.1.1 Significado parcial 1: El método gráfico de Galton (1885)

En este significado se retoma el ejemplo 1 que se presentó en la página 69. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar el método gráfico Galton utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* se muestran las frecuencias observadas de la altura (sentada) de 775 mujeres de entre 23 y 50 años. Se hace uso de la *representación gráfica* (ojiva) bajo el método de intercomparación y una gráfica que muestra la distribución de las desviaciones con respecto al percentil 50°.

Situación/problema:

Galton buscaba una forma de establecer si era apropiado trabajar un conjunto de datos observados bajo las condiciones de la distribución normal. Para lo que en un inicio trabajó el método de intercomparación (1875) y en 1885 presentó el método gráfico.

Para explicar el método gráfico, Galton en 1885 trabajó con una serie de datos observados en 775 mujeres, los datos versaban sobre la altura que tenían mujeres de entre 23 y 50 años cuando se encontraban sentadas. Para probar que los datos seguían una distribución de acuerdo a la normal se aplicó el método gráfico.

Conceptos/definiciones:

CD1: *Error probable* es una medida de variabilidad de la serie observada. Definido como la mediana de los errores para datos distribuidos normalmente.

CD2: *Ojiva* es la curva de distribución de los valores (en porcentaje) observados ordenados.

CD3: *Cuartiles* determinan los valores correspondientes al 25%, 50% (mediana) y 75% de los datos, la base de la ojiva se encuentra dividida en cuatro partes iguales.

CD4: *Percentiles* determinan los valores correspondientes del 1% al 100% de los datos, se divide la base de la ojiva en 100 partes.

CD5: *Módulo* es el término en que son calculados los valores de las desviaciones.

CD6: *Error medio* es un estimador que mide el promedio de los errores, las desviaciones de la estimación.

CD7: *Desviaciones* son las nuevas ordenadas del segundo gráfico (del método gráfico) y representan las diferencias entre cada valor observado y el valor de m de todos los valores observados. También se les ha denominado como divergencias o errores.

Propiedades/proposiciones:

PP1: *Ley del error* (Sevilla, 1993)

Primer postulado:

En una serie de un gran número de observaciones efectuadas en idénticas circunstancias y con el mismo esmero, los distintos errores posibles se presentan con tanta mayor frecuencia cuanto menor sea su valor absoluto.

Por lo que cada fenómeno se presenta un número de veces proporcional a su probabilidad, mientras que la probabilidad se encuentra en función de la magnitud del error: $P_i = \omega(\varepsilon_i)$.

Segundo postulado:

Postulado de la obra Theoria Motus Corporum Caelestium: Obtenidos por observación n valores distintos de una magnitud, siempre que todos hayan sido obtenidos en idénticas condiciones y con el mismo esmero, el valor más probable (medida) de la magnitud es la media aritmética de los valores observados.

Éste postulado después fue sustituido por el mismo Gauss, por el siguiente:

En una serie de un gran número de observaciones efectuadas en idénticas circunstancias y con el mismo esmero, los errores del mismo valor absoluto, pero de distinto signo, se presentan con la misma frecuencia y son, por lo tanto, igualmente probables.

La distribución de los errores es simétrica respecto al cero y siempre que se trate de un gran número de observaciones se puede decir que $\sum_{i=1}^{n} \varepsilon_i = 0$

PP2: *Distribución normal* es una distribución de probabilidad de variable continua cuya representación gráfica es una curva con forma de campana, es una distribución de los errores de medición, es simétrica y se distribuye alrededor de la media; la altura y ancho de la distribución depende de la varianza de las mediciones. La distribución normal estándar tiene $\mu = 0$ y $\sigma = 1$.

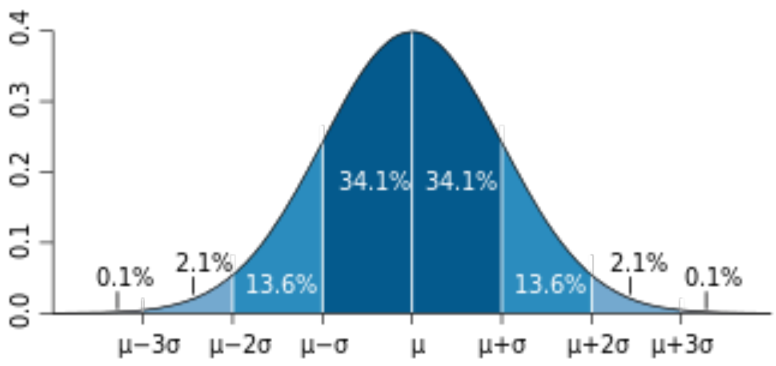


Figura 3.10. Distribución de probabilidad N(μ,σ)

PP3: m en $\frac{1}{2}$ o $0°$.

PP4: p en $\frac{1}{4}$ o $-25°$.

PP5: q en $\frac{3}{4}$ o $25°$.

PP6: *Error probable* es la mitad de la diferencia entre los cuartiles, también estableció a $q - m$ como la divergencia o el error probable (de la parte de la serie que excede a la mediana) y a $m - p$ como la divergencia de la otra parte de la serie.

PP7: Si la *serie es simétrica* entonces $q - m = m - p$.

Procedimientos:

Graficar la ojiva de frecuencias acumuladas y posteriormente el gráfico de la distribución de las desviaciones entre cada valor observado y de m. El método gráfico se lleva a cabo por medio de la secuencia de las siguientes acciones:

1. Calcular la suma de los observados hasta cada medida en pulgadas (frecuencias acumuladas) a partir de las frecuencias observadas en las mediciones. Esto corresponde a la abscisa de 0 a 775 mujeres.
2. Encontrar los porcentajes de acuerdo a las sumas del cálculo del punto 1 (frecuencia relativa acumulada). Corresponde a la abscisa de 0 a 100.
3. Establecer la ordenada correspondiente a cada frecuencia acumulada relativa.
4. Con los valores de los puntos 3 y 4 realizar la gráfica de la ojiva o curva de frecuencias correspondiente.

   a. En la ojiva se pueden identificar p, m y q, y con estos establecer si la serie es simétrica mediante $q - m = m - p$.
   b. Si la curva sigue la forma de la ojiva establecida por Galton es un indicativo de que el conjunto de datos observados sigue un a distribución normal.
5. Realizar un cambio de la línea de referencia sobre la cual se originaron las ordenadas, partiendo la curva de la ojiva en dos partes e invirtiendo el más bajo. Este gráfico muestra las desviaciones entre cada valor observado y de m.
6. Encontrar los valores del error probable, error medio y el modulo con ayuda de la tabla de Galton sobre la curva normal de distribución del error.

Argumentos:

Se argumenta utilizando el gráfico de la curva de la distribución y el gráfico de la distribución de las desviaciones con respecto a m.

3.2.1.2 Significado parcial 2: El estadístico χ^2 de Pearson (1900)

En este significado se retoma el ejemplo 2 que se presentó en la página 80. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de bondad de ajuste se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* se muestran las distribuciones de frecuencias (esperadas y observadas) así como el cálculo del error y del error al cuadrado entre la frecuencia esperada. Se hace uso de la *representación simbólica* en la tabla y en caso de realizar los cálculos para obtener la probabilidad mediante la fórmula.

Situación/problema:

A partir de las preguntas que surgen sobre la familia de distribuciones sesgadas de Pearson en 1900 intenta resolver el problema de una medición objetiva o razonable de la bondad de ajuste, con su propuesta del test de bondad de ajuste por medio del estadístico χ^2.

Para ilustrar la utilidad y cómo aplicar la prueba de bondad de ajuste Pearson (1900) presenta algunos problemas prototipo como el que se describe a continuación:

Merriman quien era profesor y economista estadounidense participó en la guerra civil española y adquirió conocimientos militares en el Cuerpo de Entrenamiento de Oficiales en la Reserva, de ahí su interés por el análisis de las desviaciones de 1000 disparos durante entrenamientos militares, utilizando el método de mínimos cuadrados y señalando que los datos provienen de un sistema probable de desviaciones de la curva normal.

Sin embargo, Pearson (1900) retoma este problema para probar si los datos de los 1000 disparos siguen una distribución normal, como lo sugería el profesor Merriman, por medio de la prueba de bondad de ajuste χ^2

Conceptos/definiciones:

CD1: *Experimento* como cualquier acción o proceso cuyo resultado está sujeto a incertidumbre.

CD2: *Ensayo* es un experimento más pequeño, es decir, el experimento consta de una secuencia de n experimentos más pequeños llamados ensayos.

CD3: *Frecuencia observada* es el número de veces que ocurre un valor en el conjunto de datos.

CD4: *Frecuencia teórica* es el número de veces que de acuerdo a una distribución se espera ocurra un valor en un conjunto de datos.

CD5: *Probabilidad* es una medida de la certidumbre asociada a un evento, en este tipo de problema se entiende como una medida de ocurrencia de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema de observado.

CD6: *Estadístico* χ^2 es una medida de la divergencia entre la distribución de frecuencias esperadas y las teóricas.

Propiedades/proposiciones:

PP1: *Ley del error*, en el mismo sentido que PP1 del significado parcial 1.

PP2: *Distribución normal*, en el mismo sentido que PP2 del significado parcial 1.

PP3: *Error* $e = m' - m$. Son las diferencias entre la frecuencia observada (m') y la frecuencia esperada (m).

PP4: *Distribución de probabilidad* χ^2 es una distribución de probabilidad continua cuya ecuación es: $y = y_0 e^{\gamma x}\left(1+\frac{x}{a}\right)^{\gamma a}$, tiene origen para x en la Moda y esta dada por la Distribución de Pearson Tipo III.

PP5: *Probabilidad*

Para n' cuando es par

$$P = \sqrt{\frac{2}{\pi}}\int_{x}^{\infty} e^{\frac{\chi^2}{2}} dX + \sqrt{\frac{2}{\pi}} e^{\frac{\chi^2}{2}}\left(\frac{\chi}{1}+\frac{\chi^3}{1\cdot 3}+\frac{\chi^5}{1\cdot 3\cdot 5}+\cdots+\frac{\chi^{n'-3}}{1\cdot 3\cdot 5\ldots n'-3}\right)$$

Para n' cuando es impar

$$P = e^{\frac{\chi^2}{2}}\left(1+\frac{\chi^2}{2}+\frac{\chi^4}{2\cdot 4}+\frac{\chi^6}{2\cdot 4\cdot 6}+\cdots+\frac{\chi^{n'-3}}{2\cdot 4\cdot 6\ldots n'-3}\right)$$

PP6: *Número de errores* $n' = n+1$.

PP7: *Estadístico* χ^2

$$\chi^2 = \sum \frac{(Frecuencia\ Obervada - Frecuencia\ Esperada)^2}{Frecuencia\ Esperada}$$

$$\chi^2 = S\left(\frac{e^2}{m}\right) = S\left(\frac{(m'-m)^2}{m}\right)$$

Procedimientos:

Encontrar la probabilidad de ocurrencia del sistema de desviaciones del conjunto de frecuencias observadas con respecto a las frecuencias esperadas o teóricas de cierta distribución (en este caso de la distribución normal), para lo cual es necesario calcular el valor del estadístico χ^2. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 (concatenación de fórmulas):
    a. Para los errores (e) se procede a encontrar las diferencias entre la frecuencia observada y la frecuencia esperada ($e = m' - m$).
    b. Las diferencias al cuadrado para cada cinturón ($e^2 = (m'-m)^2$).

c. Los cocientes de las diferencias al cuadrado entre la frecuencia esperada $\frac{(m'-m)^2}{m}$.
d. La sumatoria de los cocientes al cuadrado entre la frecuencia observada $\left(\chi^2 = S\left(\frac{(m'-m)^2}{m}\right)\right)$.
e. Obtener la raíz cuadrada positiva del estadístico χ^2, en otras palaras χ.

2. Obtener n', que en este caso es igual al número de cinturones más 1.
3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 (con χ y n') de Pearson u obtenerlo mediante el desarrollo de la formula propuesta de probabilidad.

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia de tener un conjunto de datos (frecuencias observadas) con un sistema de errores de determinadas desviaciones de cierta distribución teórica.

Por ejemplo:

La probabilidad que obtuvimos nos indica que tendríamos oportunidad de ver un conjunto de datos con un sistema de errores de tan amplias desviaciones de la distribución normal, únicamente en 15 o 16 ocasiones en 10,000,000 de ensayos.

3.2.1.3 Significado parcial 3: Los grados de libertad de Fisher

En este significado se retoma el ejemplo 5 que se presentó en la página 97. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de bondad de ajuste se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* se muestran las distribuciones de frecuencias (esperadas y observadas), así como el cálculo del valor del estadístico χ^2. Se hace uso de la *representación simbólica* en la tabla, en el planteamiento del problema y en la conclusión.

Situación/problema:

La problemática que aborda Fisher y otros estadísticos de la época es referente a los grados de libertad para el cálculo de la probabilidad asociada al valor del estadístico χ^2 en la distribución χ^2 en las pruebas de bondad de ajuste y de independencia.

Fisher (1925a), publicó algunos problemas sobre la aplicación de la prueba de bondad de ajuste con el estadístico χ^2 y con $n' = k + 1$, donde k indica el número de grados de libertad.

Se retoma el ejemplo 5 donde Fisher realizó una comparación con la expectativa de las frecuencias de clase Mendelianas en la generación de un híbrido (F1) de cuatro clases en la proporción $9 : 3 : 3 : 1$. Introduce la hipótesis nula o del investigador como: los factores segregan de forma independiente, y que las cuatro clases de descendencia son igualmente viables.

Conceptos/definiciones:

CD1: *Experimento*, en el mismo sentido que CD1 del significado parcial 2.

CD2: *Ensayos*, en el mismo sentido que CD2 del significado parcial 2.

CD3: *Frecuencia observada*, en el mismo sentido que CD3 del significado parcial 2.

CD4: *Frecuencia teórica*, en el mismo sentido que CD4 del significado parcial 2.

CD5: *Probabilidad*, en el mismo sentido que CD5 del significado parcial 2.

CD6: *Estadístico* χ^2, en el mismo sentido que CD6 del significado parcial 2.

CD7: *Grados de libertad* es el número de filas menos el número de restricciones lineares independientes en las frecuencias.

CD8: *Significancia* como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria.

Propiedades/proposiciones:

PP1: *Ley del error*, de acuerdo con lo establecido en el significado parcial 1 en PP1.

PP2: *Distribución de probabilidad* χ^2, en el mismo sentido que lo establecido en PP4 del significado parcial 2.

PP3: *Grados de libertad* $k = n - r$.

PP4: n' ahora es el número de grados de libertad más uno $n' = k + 1$, donde k indica el número de grados de libertad.

PP5: *Estadístico* χ^2

$$\chi^2 = \sum \frac{\left((m+x) - m\right)^2}{m}$$

O bien, de acuerdo a lo establecido en PP7 del significado parcial 2.

PP6: *Regla de decisión* compara la probabilidad obtenida a partir del valor del estadístico calculado, con el límite preestablecido como desviación significativa ($P = 0.05$) y si la posibilidad de exceder tal valor es menor al límite preestablecido diremos que las desviaciones de las expectativas son claramente significativas.

Procedimientos:

Encontrar la probabilidad de ocurrencia del sistema de desviaciones del conjunto de frecuencias observadas con respecto a las frecuencias esperadas, para lo cual es necesario calcular el valor del estadístico χ^2 y obtener la probabilidad de las tablas de Elderton. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 (concatenación de formulas):
    a. Los cocientes de las diferencias de la frecuencia observada y esperada al cuadrado entre la frecuencia esperada $\frac{\left((m+x)-m\right)^2}{m}$.
    b. La sumatoria de los cocientes al cuadrado entre la frecuencia observada $\left(\chi^2 = \sum \frac{\left((m+x)-m\right)^2}{m}\right)$.
2. Obtener n', que en este caso es igual al número de grados de libertad más 1, $n' = k + 1$ y $k = n - r$.
3. Buscar el valor de la probabilidad en las tablas de la distribución χ^2 (con χ^2 y n') de Elderton.

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia de tener un conjunto de datos con un sistema de errores de tan amplias desviaciones de cierta distribución teórica y comparándolo con el límite preestablecido como desviación significativa ($P = 0.05$). Por ejemplo, con n = 3 y $\chi^2 = 10.87$:

> La posibilidad de exceder tal valor es entre 0.01 y 0.02; si nosotros tomamos $P = 0.05$ como límite de desviación significativa, diremos que en este caso las desviaciones de las expectativas son claramente significativas.
>
> Además, argumenta también los grados de libertad utilizados para este problema de la siguiente manera: los valores esperados se calculan a partir del total observado, de modo que las cuatro clases deben estar de acuerdo en su suma, y si tres clases se completan arbitrariamente el cuarto es por lo tanto determinado; por lo tanto $n = 3$.

3.2.1.4 Significado parcial 4: La prueba de bondad de ajuste con la corrección de Yates (1934)

En este significado se retoma el ejemplo 8 que se presentó en la página 103. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de bondad de ajuste se utilizó *lenguaje natural* en el planteamiento e interpretación de los resultados, por medio de la *representación tabular* se muestran las distribuciones de frecuencias (esperadas y observadas) y se hace uso de la *representación simbólica* al señalar el valor del estadístico χ_c^2, los grados de libertad y la probabilidad.

Situación/problema:

El problema que abordó Yates es referente a la corrección de continuidad que requiere el estadístico χ^2 cuando se trabaja con números pequeños, señaló que la discrepancia que se hace notoria cuando se trabaja con números pequeños y se debe a que χ es de una distribución continua, mientras que la distribución que se intenta aproximar es discreta.

Se retoma el ejemplo 8, del artículo de Pearson en 1900, para ilustrar el método de Yates en una prueba de bondad de ajuste al trabajar con frecuencias menores a cinco. Este ejemplo es sobre el número de pétalos de una flor denominada ranunculus, las distribuciones esperadas se calcularon bajo la curva sesgada $y = 0.211225x^{-0.322}(7.3253 - x)^{3.142}$.

Conceptos/definiciones:

CD1: *Experimento*, en el mismo sentido que CD1 del significado parcial 2.

CD2: *Ensayos*, en el mismo sentido que CD2 del significado parcial 2.

CD3: *Frecuencia observada*, en el mismo sentido que CD3 del significado parcial 2.

CD4: *Frecuencia teórica*, en el mismo sentido que CD4 del significado parcial 2.

CD5: *Probabilidad*, en el mismo sentido que CD5 del significado parcial 2.

CD6: *Estadístico* χ_c^2, en el mismo sentido que CD6 del significado parcial 2 y además considerando el factor de corrección de continuidad.

CD7: *Grados de libertad* en el mismo sentido que CD7 del significado parcial 3.

CD8: *Factor de corrección de continuidad* es un ajuste que se realiza al estadístico χ^2 cuando se trabaja con números pequeños, debido a que se intenta aproximar una distribución discreta mediante una distribución continua.

CD9: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 3.

Propiedades/proposiciones:

PP1: *Ley del error*, de acuerdo con lo establecido en el significado parcial 1 en PP1.

PP2: *Error*, en el mismo sentido que PP3 del significado parcial 2.

PP3: *Distribución de probabilidad* χ^2, en el mismo sentido que PP4 del significado parcial 2.

PP4: *Grados de Libertad*, de acuerdo con lo establecido en PP3 del significado parcial 3.

PP5: n', de acuerdo con lo establecido en PP4 del significado parcial 3.

PP6: *Estadístico* χ_c^2 con corrección de continuidad.

$$\chi_c^2 = \sum \frac{(|Frecuencia\ Obervada - Frecuencia\ Esperada| - 0.5)^2}{Frecuencia\ Esperada}$$

$$\left(\chi_c^2 = \sum \frac{(|m' - m| - 0.5)^2}{m}\right)$$

PP7: *Regla de decisión* compara el valor del estadístico y el valor crítico de la distribución, sí el valor del estadístico calculado es más grande que el valor crítico entonces las desviaciones son significativas.

Procedimientos:

Encontrar la probabilidad de ocurrencia del sistema de desviaciones del conjunto de frecuencias observadas con respecto a las frecuencias esperadas o teóricas de cierta distribución (en este caso de la distribución normal), para lo cual es necesario calcular el valor del estadístico χ_c^2 . Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ_c^2 (concatenación de fórmulas):
    a. Para los errores (e) se procede a encontrar el valor absoluto de las diferencias entre la frecuencia observada y la frecuencia esperada ($e = |m' - m|$).
    b. Restar al error 0.5 como factor de corrección ($e_c = |m' - m| - 0.5$).
    c. El valor absoluto de las diferencias (con corrección) al cuadrado para cada cinturón ($e_c^2 = (|m' - m| - 0.5)^2$).
    d. Los cocientes del error con corrección al cuadrado entre la frecuencia esperada $\frac{(|m'-m|-0.5)^2}{m}$.
    e. La sumatoria de los cocientes al cuadrado entre la frecuencia observada $\left(\chi_c^2 = \sum \frac{(|m'-m|-0.5)^2}{m}\right)$.
2. Obtener n', que en este caso es igual al número de grados de libertad más 1.
3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 (con χ_c^2 y n').

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia del tener un conjunto de datos con un sistema de errores de tan amplias desviaciones de cierta distribución teórica. Por ejemplo: Siendo $\chi_c^2 = 2.587262$ y $P = .858576$ se puede decir que la curva sesgada $y =$

$0.211225x^{-0.322}(7.3253-x)^{3.142}$ propuesta por Pearson es un buen ajuste para la distribución de los pétalos de las ranunculus.

3.2.2 Significado 2: El estadístico Chi-cuadrada en la prueba de independencia

La prueba de independencia con el estadístico χ^2 determina si dos variables, de n individuos que conforman una muestra, se encuentran asociadas.

Se identificaron cuatro significados parciales en la evolución de la prueba de independencia con el estadístico χ^2. Nos referiremos a ellos continuando con la numeración de los significados anteriores. Así, el significado parcial 5 se puede considerar una prueba de independencia intuitiva, ya que en la configuración ontosemiótica se destacó la noción de asociación, cuya posterior generalización dio lugar a la prueba de independencia con el estadístico χ^2.

3.2.2.1 Significado parcial 5: Los inicios de la prueba de independencia mediante el coeficiente de asociación de Yule

En este significado se retoma el ejemplo 3 que se presentó en la página 87. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

En el planteamiento del problema se utilizó *lenguaje natural* y una *representación tabular* para proporcionar los índices de ataque de viruela, de personas menores y mayores de diez años de edad, en las casas que se encuentran afectadas en los pueblos donde han ocurrido recientes epidemias de viruela. Se utilizó una *representación simbólica* al realizar los cálculos del coeficiente de asociación y *representación tabular* para señalar los coeficientes de asociación. También hace uso del *lenguaje natural* al explicar cómo es que realiza la prueba de asociación y al momento de concluir si existe o no asociación.

Situación/problema:

Yule (1900), buscaba una forma de probar asociación entre dos variables discretas tal como se medía la asociación para las variables continuas. Utilizó el ejemplo 3 para probar su coeficiente de asociación Q.

Yule retoma datos de la publicación ‘Vaccination and Small-Pox Statistics’ de Mr. Noel A. Humphrey de 1897. Se muestra una tabla con los índices de ataque de viruela, en casas que en el momento de la recolección de datos estaban invadidas por viruela, de personas menores y mayores de diez años de edad de Sheffield, Warrington, Dewsbury, Leicester y Gloucester, ciudades de Inglaterra en las que recientemente habían ocurrido epidemias de viruela. Lo anterior con la finalidad de mostrar si existe una asociación entre los no vacunados y los ataques en los hogares infectados de personas menores y mayores de dicha edad.

Conceptos/definiciones:

CD1: *Variable* es cualquier característica cuyo valor puede cambiar de un objeto o individuo a otro en la población. Esta variable puede ser cuantitativa o bien referir a cualidades.

CD2: *Cualidades o atributos* son los valores no numéricos, positivos o negativos, que adquirieren las variables en un objeto o individuo. Se denotan por A, B, C.

CD3: *Frecuencia* es el número de objetos o individuos que poseen una o varias cualidades y se denotan por $(A), (B), (C)$. Yule (1900) trabajó esta frecuencia como porcentaje.

CD4: *Los conjuntos en probabilidad* son agrupaciones de elementos con características similares.

CD5: *Coeficiente de asociación* es una medida que aproxime la asociación, por una parte, hacia una completa independencia y, por otra parte, hacia una completa asociación. Es una función simétrica de los atributos, con rangos entre ± 1 y es cero cuando los atributos no estan asociados.

CD6: *El coeficiente de correlación* es una función simétrica de las variables, con rangos entre ± 1 y es cero cuando las variables son independientes.

CD7: *Asociación* se entiende en términos de correlación para variables continuas.

Propiedades/proposiciones:

PP1: *Dos cualidades o atributos, A y B, son definidos como independientes* si la probabilidad de encontrarlos juntos es el producto de las probabilidades de encontrarlos también por separado, por ejemplo, si $\frac{(AB)}{U} = \frac{(A)}{(U)} \cdot \frac{(B)}{(U)}$.

Lo anterior nos lleva al siguiente teorema:

$$(AB)(U) = (A)(B)$$

entonces,

$$(A\beta)(U) = (A)(\beta)$$

$$(\alpha B)(U) = (\alpha)(B)$$

$$(\alpha\beta)(U) = (\alpha)(\beta)$$

PP2: *Si los productos cruzados son iguales, las variables son independientes*, lo que se entiende de la siguiente manera $(AB)(\alpha\beta) = (A\beta)(\alpha B)$.

PP3: *El coeficiente de asociación debería*:

1. Ser cero cuando las variables o atributos A, B, son independientes, y sólo cuando son independientes
2. Deberían ser +1 cuando, y sólo cuando, A y B son completamente asociados,
$$(A\beta) = 0, \qquad (\alpha B) = 0, \qquad (A\beta) = (\alpha B) = 0$$
3. Deberían ser -1 cuando, y solo cuando, A y β o B y α son completamente asociados,
$$(AB) = 0, \qquad (\alpha\beta) = 0, \qquad (AB) = (\alpha\beta) = 0$$

PP4: *Coeficiente de asociación* $Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$.

PP5: *Sirviendo Q como un coeficiente de asociación para*:

1. Cuando A y B son independientes el numerador es cero y por lo tanto Q es cero; y a la inversa cuando Q es cero las variables son independientes.
2. Cuando $(A\beta) = 0$ o $(\alpha B) = 0$, o ambas, $Q = +1$; y a la inversa cuando $Q = +1$ $(A\beta) = 0$, $(\alpha B) = 0$, o ambas.
3. Cuando $(AB) = 0$ o $(\alpha\beta) = 0$, o ambas $Q = -1$; y a la inversa cuando $Q = -1$ $(AB) = 0$ o $(\alpha\beta) = 0$, o ambas.

Procedimientos:

Encontrar la medida de asociación entre los no vacunados y los ataques en los hogares infectados de personas menores y mayores de diez años de edad por medio del coeficiente de asociación Q, mediande la concatenación de las siguientes acciones:

1. Formar la tabla de contingencia de los atributos con los que interesa establecer si existe una asociación.

	(B)	(β)
(A)	(AB)	$(A\beta)$
(α)	(αB)	$(\alpha\beta)$

   Para este caso A son las personas atacadas y B a las personas no vacunadas (Se forman distintas tablas de contingencia para cada pueblo tanto para los menores de diez años como para los mayores de esta edad).

   En la tabla de contingencia donde se muestran los datos al presentar el ejemplo se tiene $(AB) = 67.6$ y $(A\beta) = 7.9$, mientras que para obtener (αB) es necesario a 100 restarle (AB), de manera analoga se realiza este procedimiento para (αB).
2. Calcular el coeficiente de asociación $Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$, para cada una de las ciudades del estudio y para los rangos de edad.
3. Mostrar en forma tabular la asociación entre los no vacunados y ataques en hogares infectados.

Argumentos:

Se argumenta con qué tan elevado o no es el coeficiente de asociación por grupo de edad, además se realiza una observación en cuanto a los índices presentados inicialmente en el planteamiento del problema y el valor obtenido de los coeficientes de asociación de las diferentes ciudades. En este caso, se concluye el problema con los siguientes argumentos:

La asociación entre no vacunados y los ataques es sorprendentemente alta para niños jóvenes – 0.8 a 0.9 – pero decae bruscamente a 0.5 en el grupo de mayor edad presumiblemente debido a la protección decreciente de la vacunación realizada en la infancia.

La constancia de la asociación en los pueblos con índices de ataque con amplias diferencias es un punto digno de notar. Sheffield, Warrington, y Leicester exhiben prácticamente asociaciones idénticas, a pesar de que los índices de ataques varían de 7.9 a 2.5 y de 67.6 a 35.3.

3.2.2.2 Significado parcial 6: La prueba de independencia con el estadístico χ^2 y los coeficientes de contingencia de Pearson (1904)

En este significado se retoma el ejemplo 4 que se presentó en la página 92. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para abordar este problema se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular*, con una tabla de contingencia de dos por dos, se muestran las frecuencias observadas sobre la presencia o ausencia de cicatriz con respecto a los sobrevivientes o muertos a causa de la epidemia de viruela. Se hace uso de la *representación simbólica* al mostrar los resultados de los cálculos de la contingencia cuadrada media, contingencia media, estadístico χ^2 y los coeficientes de asociación.

Situación/problema:

Pearson en 1904 buscó alternativas para obtener una medida de asociación más certera que las técnicas que hasta ese momento se tenían.

El problema que plantea retoma datos de la epidemia de viruela de 1890, con el fin de mostrar con un ejemplo numérico su propuesta y la viabilidad sobre otras, como el coeficiente de asociación de Yule.

Conceptos/definiciones:

CD1: *Una variable* es cualquier característica cuyo valor puede cambiar de un objeto o individuo a otro en la población.

CD2: *Los atributos* no adquirieren mediciones numéricas y se pueden clasificar en cierto número de grupos.

CD3: *Frecuencia observada* es el número de veces que ocurre un atributo en un grupo.

CD4: *Frecuencia esperada* es el número de veces que, de acuerdo a la teoría de independencia, se espera que ocurra un atributo en un grupo.

CD5: *Tabla de contingencia* es una tabla formada por s columas y t filas con una frecuencia total N distribuida en subgrupos correspondientes a estos $s \times t$ compartimentos (este término fue introducido por Pearson en 1904).

CD6: *La contingencia* generaliza la noción de asociación de dos atributos que había desarrollado Yule. Se puede clasificar individuos no sólo en dos grupos sino en tantos grupos con exclusividad de atributos como queramos y donde el orden de los subgrupos no tiene importancia.

CD7: *La contingencia media* nos permite medir hasta qué punto estos dos sistemas son contingentes o no contingentes.

CD8: *La asociación* es otro punto de vista de una medida del grado de desviación de ocurrencia de la independencia.

CD9: *La correlación* es otro punto de vista de una medida del grado de desviación de ocurrencia de la independencia.

CD10: *Probabilidad* es una medida de cuán lejos el sistema observado es o no compatible con las bases de independencia probabilística.

Propiedades/proposiciones:

PP1: $n' = (x-1)(\lambda-1)/N - \emptyset^2$ donde x son el número de renglones y λ el número de columnas.

PP2: *Probabilidad*. Si P es grande, las posibilidades están a favor de que el sistema proviene de probabilidades independientes; si P es pequeña hay una certera asociación entre los atributos.

PP3: *Independencia* (probabilística). Si p es la probabilidad de cualquier evento, y q la probabilidad de un segundo evento, decimos que los dos eventos son independientes, si la probabilidad de la combinación de los eventos es $p \times q$.

PP4: *Grado de contingencia* $1 - P$, a mayor contingencia, mayor es el monto de asociación o de correlación entre los dos atributos.

PP5: *Estadístico Chi-cuadrada* $\chi^2 = S\left\{\frac{(n_{uv}-v_{uv})^2}{v_{uv}}\right\}$, donde n_{uv} son los números observados permitiendo los errores de azar de muestreo, y $v_{uv} = \frac{n_u \times m}{N} = N \times \frac{n_u}{N} \times \frac{m_v}{N}$ son las observaciones esperadas de acuerdo a la teoría de independencia probabilística si N pares de atributos son examinados. En caso de trabajar con una tabla de contingencia de 2×2 en correspondencia con los terminos de Yule de asociación de dos atributos se tiene $\chi^2 = \frac{(ab-cd)^2(a+b+c+d)}{(a+d)(c+b)(a+c)(d+b)}$.

PP6: $n_{uv} - v_{uv}$ esta magnitud será absolutamente independiente del orden de clasificación.

PP7: *Contingencia Cuadrada Media* $\emptyset^2 = \frac{\chi^2}{N} = S\left\{\frac{(n_{uv}-v_{uv})^2}{Nv_{uv}}\right\}$ y para el caso específico de tablas de contingencia de 2×2 en correspondencia con los términos de Yule de asociación de dos atributos se tiene $\emptyset^2 = \frac{(ab-cd)^2}{(a+d)(c+b)(a+c)(d+b)}$.

PP8: *Contingencia Media* $\psi = \sum \frac{n_{uv}-v_{uv}}{N}$, donde Σ refiere a la sumatoria de todas las contingencias positivas. Y para el caso específico de tablas de contingencia de 2×2 en correspondencia con los términos de Yule de asociación de dos atributos se tiene $\psi = \frac{2(ab-cd)}{N^2}$.

PP9: *Coeficiente de Contingencia Cuadrada Media* $C_1 = \sqrt{\frac{\emptyset^2}{1+\emptyset^2}}$.

PP10: *Coeficiente de Contingencia Media* C_2 se busca en tabla de integrales I_r, Q_r y la contingencia ψ para valores de r.

P11: $C_2 = C_1 = coeficiente\ de\ correlación$ si la agrupación es lo suficientemente pequeña y con correlación normal.

PP12: *Coeficiente de asociación de Yule* $Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$.

Procedimientos:

Encontrar las medidas de asociación, correlación y contingencia, y compararlas para reconocer cuál de ellas es apropiada para medir la desviación de los resultados observados de las frecuencias esperadas bajo el supuesto de independencia (probabilística). A partir de esta medida de desviación, establecer si existe contingencia entre los atributos. Lo anterior se realiza bajo las siguientes acciones:

1. Por medio de los valores de la tabla de contingencia encontrar los valores de la contingencia cuadrada media ($\emptyset^2 = \frac{(ab-cd)^2}{(a+d)(c+b)(a+c)(d+b)}$), estadístico $\chi^2 = \frac{(ab-cd)^2(a+b+c+d)}{(a+d)(c+b)(a+c)(d+b)}$ y contingencia media $\psi = \frac{2(ab-cd)}{N^2}$.
2. A partir de los valores de $\emptyset^2$, χ^2 y ψ encontrar:
    a. Coeficiente de Contingencia Cuadrada Media $C_1 = \sqrt{\frac{\emptyset^2}{1+\emptyset^2}}$.
    b. Coeficiente de Contingencia Media C_2 se busca en tabla de integrales I_r, Q_r y la contingencia ψ para valores de r.
    c. Coeficiente de asociación de Yule $Q = \frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$.
    d. Coeficiente de correlación por división cuádruple.
    e. Grado de contingencia $1 - P$.
3. Comparar los valores de los coeficientes.

Argumentos:

Se argumenta con respecto a la efectividad de los coeficientes en el contexto que se está utilizando. Se hacen explicitas las bondades de utilizar el grado de contingencia, ya que se encuentra libre de la restricción que tienen $C_1\ y\ C_2$; señalando con el grado de contingencia la desviación de la independencia (probabilidad independiente) en relación con el resultado. Por ejemplo, el caso de viruela y la presencia o ausencia de cicatriz son tales que la tabla anterior solo podría ocurrir 718 veces en 1040 casos si los dos eventos fueran absolutamente independientes.

3.2.2.3 Significado parcial 7: La prueba de independencia y los grados de libertad de Fisher

En este significado se retoma el ejemplo 6 que se presentó en la página 98. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de independencia se utilizó *lenguaje natural* en la introducción a los datos presentados, descripción de algunos procedimientos e interpretación de los resultados, por medio de la *representación tabular* (tabla de doble entrada o de contingencia) se muestran las distribuciones de frecuencia de dos variables. También se hace uso de la *representación simbólica* al realizar los cálculos para obtener la probabilidad mediante la fórmula y el valor del estadístico.

Situación/problema:

La problemática que aborda Fisher y otros estadísticos de la época es referente a los grados de libertad para el cálculo de la probabilidad asociada al valor del estadístico χ^2 en la distribución χ^2 en las pruebas de bondad de ajuste y de independencia. De acuerdo con Fisher (1922a), para el caso de la prueba de independencia con tablas de contingencia 2×2 debemos tomar $n' = 2$ en vez de $n' = 4$ y para el cálculo de la probabilidad se pueden usar las tablas calculadas por Elderton.

Fisher retoma los datos que utilizaron Greenwood y Yule en 1915 sobre la tifoidea, también para medir asociación entre las personas atacadas y la vacuna, pero Fisher realizó la prueba con la noción de grados de libertad que introdujo en 1922.

Conceptos y/o definiciones:

CD1: *Variante*, en el mismo sentido que el CD1 del significado parcial 6.

CD2: *Frecuencia observada*, en el mismo sentido que CD3 del significado parcial 2.

CD3: *Frecuencia teórica* es el número de veces que de acuerdo a una distribución bajo el supuesto de independencia se espera ocurra un valor en un conjunto de datos.

CD4: *Distribución de frecuencias* especifica con qué frecuencia la variante toma cada uno de sus valores posibles.

CD5: *Distribución Marginal* da información unidimensional sobre cada clasificación y no dicen nada sobre la asociación entre las dos variables.

CD6: *Probabilidad* en el mismo sentido que el CD10 del significado parcial 6.

CD7: *Asociación* como divergencia de la independencia.

CD8: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 3.

Propiedades/proposiciones:

PP1: *Distribución de probabilidad* χ^2 de acuerdo con lo establecido en el significado parcial 2 en PP4.

PP2: *Grados de libertad* como el número de parámetros libres menos el número de parámetros a estimar, $k = cr - 1 - (c-1) - (r-1) = (c-1)(r-1)$, donde c indica el número de columnas y r el número de renglones.

PP3: $n' = (c-1)(r-1) + 1$.

PP4: *Estadístico* χ^2 :

para tabla de contingencia de 2×2

$$\chi^2 = \frac{(bc-ad)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

para tabla de contingencia de $s \times t$

$$\chi^2 = \sum \frac{\left((m+x)-m\right)^2}{m}$$

PP5: *Regla de decisión* de acuerdo con lo establecido en el PP6 del significado parcial 3.

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar independencia entre dos variables de una misma población. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 utilizando la formula propuesta por Fisher (1922a) para una tabla de contingencia de 2×2. También se puede calcular por medio de $\chi^2 = \sum \frac{\left((m+x)-m\right)^2}{m}$, para lo cual es necesario obtener las frecuencias esperadas bajo el supuesto de independencia a partir de las frecuencias observadas mediante de los totales marginales observados.
2. Obtener n', que en este caso es igual al número de grados de libertad más uno $n' = (c-1)(r-1) + 1$.

3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 (con χ^2 y n') de Elderton para bondad de ajuste.

Argumentos:

Se puede argumentar utilizando la probabilidad asociada χ^2 y comparandolo con el límite preestablecido como desviación significativa ($P = 0.05$). En este ejemplo debido a que $\chi^2 = 56.234$ Fisher señaló que las observaciones son claramente opuestas a la hipótesis de independencia.

Realiza otras argumentaciones referentes a los grados de libertad y el cálculo de la frecuencia esperada:

Para que los números esperados estén de acuerdo a los números observados en los totales; solo se necesita calcular un valor, por ejemplo, $\frac{328\times6815}{18483} = 120.93$; los otros se escriben a la vez por sustracción desde los totales. Por lo tanto, es obvio que los valores observados pueden diferir de los esperados en solo un grado de libertad, de modo que al probar la independencia en una tabla de $2 \times 2, n = 1$.

3.2.2.4 Significado parcial 8: La prueba de independencia y la corrección de continuidad de Yates (1934)

En este significado se retoma el ejemplo 7 que se presentó en la página 101. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de independencia se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* (tabla de doble entrada o de contingencia) se muestran las distribuciones de frecuencia de dos variables y la probabilidad de número de niños amamantados. También se hace uso de la *representación simbólica* al realizar los cálculos para obtener el valor del estadístico, el estadístico con corrección por continuidad y la probabilidad de la distribución de número de niños amamantados.

Situación/problema:

El Dr. Milo Hellman quien fue un ortodontista estadounidense que realizó contribuciones al campo de la ortodoncia sobre la relación entre los dientes, la mandíbula y cara, recopiló datos durante sus investigaciones referentes a la maloclusión, que son alteraciones de la posición de los dientes, de los bebés y la forma en que el bebé era alimentado. M. Hellman concluye que la alimentación con biberón es uno de los factores que causan la maloclusión.

Yates (1934) retoma este ejemplo para probar cómo afecta la corrección que propuso cuando se trabaja con frecuencias menores a cinco.

Conceptos/definiciones:

CD1: *Una Variable* en el mismo sentido que el CD1 del significado parcial 6.

CD2: *Frecuencia observada* en el mismo sentido que el CD2 del significado parcial 7.

CD3: *Frecuencia teórica* en el mismo sentido que el CD3 del significado parcial 7.

CD4: *Distribución de frecuencias* en el mismo sentido que el CD4 del significado parcial 7.

CD5: *Distribución Marginal* en el mismo sentido que el CD5 del significado parcial 7.

CD6: *Probabilidad* en el mismo sentido que el CD6 del significado parcial 7.

CD7: *Factor de corrección de continuidad* en el mismo sentido que el CD8 del significado parcial 4.

CD8: *Asociación* es una medida de divergencia de la independencia.

CD8: *Independencia* entre dos variables indica que no existe relación (asociación) entre las variables que se estudian. Dichas variables tienen una distribución bajo independencia probabilística.

CD9: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 3.

Propiedades/proposiciones:

PP1: *La probabilidad* correspondiente a cualquier termino $a, b, c, d,$ es: $\frac{n!\,n'!\,(N-n)!\,(N-n')!}{N!\,a!\,b!\,c!\,d!}$, es decir, el producto de los factoriales de los cuatro totales marginales dividido por el producto de los factoriales del gran total y de los números de las cuatro celdas.

PP2: *Grados de libertad*, de acuerdo con lo establecido en el significado parcial 7 en PP2.

PP3: n', en en el mismo sentido que PP3 del significado parcial 7.

PP4: *Estadístico* χ^2 sin factor de correción de continuidad

$$\chi^2 = \frac{(a \times d - b \times c)^2\, N}{n \times n'(N-n)(N-n')}$$

PP5: *Estadístico* χ^2 *con factor de correción de continuidad*:

Para tablas de 2×2

$$\chi_c^2 = \frac{\left(\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right)^2 N}{n \times n'(N-n)(N-n')}$$

Para tablas de $r \times c$

$$\chi_c^2 = \sum \frac{(|Frecuencia\ Obervada - Frecuencia\ Esperada| - 0.5)^2}{Frecuencia\ Esperada}$$

PP6: *Regla de decisión* en el mismo sentido que el PP7 del significado parcial 4.

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar independecia entre dos variables de una misma población. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 con correción de continuidad, restando o sumando, según corresponda, la mitad de la unidad a cada una de las frecuencias observadas y utilizando los marginales.
2. Establecer los grados de libertad.
3. Calcular la probabilidad exacta mediante el desarrollo de la fórmula propuesta por Yates, también se puede utilizar una aproximación con la probabilidad expresada en las tablas de la distribución χ^2 de Elderton.

Para este ejemplo y probar la utilidad de la corrección a la prueba, Yates realizó estos procedimientos adicionales:

1. Calcular el valor del estadístico χ^2 sin corrección de continuidad.
2. Encontrar en las tablas de la distribución χ^2 la probabilidad asociada a la raiz cuadrada positiva del valor del estadístico y con un grado de libertad.
3. Multiplicar por dos la probabilidad obtenida en el punto dos para la probabilidad correspondiente a χ^2.
4. Obtener la probabilidad de la raíz cuadrada positiva del valor del estadístico χ^2 con corrección de continuidad y con un grado de libertad con las tablas de la distribución χ^2.
5. Multiplicar por dos la probabilidad obtenida en el punto cuatro para la probabilidad correspondiente a χ^2 con corrección de continuidad.
6. Realizar la comparación entre la probabilidad obtenida con la corrección de continuidad y sin ella.

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia de un evento, por ejemplo, la probabilidad real de obtener cuatro o más niños normales alimentados con leche materna (amamantados) es 0.1435.

También se puede argumentar con respecto al valor del estadístico y el valor crítico de la distribución establecidos en la Tabla III que proporciona Yates.

3.2.3 Significado 3: El estadístico Chi-cuadrada en la prueba de homogeneidad

En este apartado, describiremos las características de los cuatro significados parciales que se identificaron en la evolución histórica de la prueba de homogeneidad con el estadístico χ^2. La prueba de homogeneidad con el estadístico χ^2 determina si es posible que varias muestras provengan de la misma población.

3.2.3.1 Significado parcial 9: Los inicios de la prueba de homogeneidad con Pearson (1911)

En este significado se retoma el ejemplo 9 que se presentó en la página 106. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de homogeneidad se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* (tabla de contingencia) se muestran las frecuencias de color de cabello de dos muestras. También se hace uso de la *representación simbólica* en la tabla de contingencia y al realizar los cálculos para obtener el valor del estadístico y al referirse a la probabilidad.

Situación/problema:

Pearson (1911) deseaba averiguar la probabilidad de que dos muestras conocidas a priori puedan pertenecer a la misma población, pero sin tener conocimiento a priori de la población. Para lo cual retomó su estadístico χ^2 que había propuesto para bondad de ajuste en 1900 y con el que también estaban trabajando pruebas de independencia.

Para mostrar cómo funciona la prueba de homogeneidad con el estadístico χ^2, Pearson retoma datos de los estudios del Dr. Macdonald. Abordamos uno sobre color de cabello de dos muestras. La primera muestra corresponde a personas enfermas de fiebre escarlatina y la segunda muestra de enfermos de sarampión.

Conceptos/definiciones:

CD1: *Muestra* es el subconjunto de una población, el subconjunto está constituido por ciertas observaciones que representan al total.

CD2: *Las muestras independientes* son muestras que se seleccionan de manera aleatoria para que sus datos no dependan de otras observaciones, este tipo de muestras se realizan a dos conjuntos de elementos distintos.

CD3: *Población* es el conjunto de elementos que tienen características comunes.

CD4: *Una Variable* en el mismo sentido que el CD1 del significado parcial 6.

CD5: *Los atributos o clases* en el mismo sentido que el CD2 del significado parcial 6.

CD6: *Frecuencia observada* en el mismo sentido que el CD4 del significado parcial 6.

CD7: *Probabilidad* es una medida de la certidumbre asociada a un evento, en este tipo de problema se entiende como una medida de ocurrencia de que ambas sean muestras aleatorias de la misma población.

CD8: *Sobre la probabilidad de que dos distribuciones de frecuencias sean realmente muestras aleatorias de la misma población* es una medida para reconocer si es posible que las muestras provengan de la misma población.

Propiedades/proposiciones:

PP1: *Si las frecuencias pertenecen a muestras aleatorias de la misma población* se tiene que

$$\frac{\bar{f}_p}{N}=\frac{f_p'}{N'}=\frac{\mu_p}{M}, \qquad \frac{\bar{f}_q}{N}=\frac{f_q'}{N'}=\frac{\mu_q}{M}$$

PP2: *Si las muestras son independientes* entonces

$${\sigma_p}^2=n^2\left(\frac{{\Sigma_p}^2}{N^2}+\frac{{\Sigma_p}^{2\prime}}{N'^2}\right), \qquad \sigma_p\sigma_q r_{pq}=n^2\left(\frac{\Sigma_p\Sigma_q \mathrm{R}_{pq}}{N^2}+\frac{\Sigma_p'\Sigma_q'\mathrm{R}_{pq}'}{N'^2}\right)$$

Siendo $\Sigma_p, \Sigma_q, \Sigma_p', \Sigma_q'$ las desviaciones estandar de las frecuencias de la pth y qth frecuencias de las dos muestras y R_{pq}, R_{pq}' las correlaciones de las mismas frecuencias.

PP3: Sí M es la población total de una distribución de frecuencias $\mu_1, \mu_2, \mu_3, \mu_4 \dots \mu_p, \mu_q \dots \mu_g$ y como ${S_1}^g(\mu_p)=M$ y además $\frac{{\mu_p}'}{M'}=\frac{\mu_p}{M}$, entonces ${S_1}^g({\mu_p}')=M'$ y por lo tanto M' es la población total de una distribución de frecuencias ${\mu_1}', {\mu_2}', {\mu_3}', {\mu_4}' \dots {\mu_p}', {\mu_q}' \dots {\mu_g}'$.

PP4: *Distribución de probabilidad* χ^2 de acuerdo con lo establecido en el significado parcial 2 en PP4.

PP5: n', en el mismo sentido que en el significado parcial 2 en PP6.

PP6: *Probabilidad*. Si *P* es grande, las posibilidades están a favor de que las muestras provengan de la misma población.

PP7: *Estadístico* χ^2

$$\chi^2 = S_1{}^g \left\{ \frac{NN'\left(\frac{f_p}{N} - \frac{f_p{}'}{N'}\right)^2}{f_p + f_p{}'} \right\}$$

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar si dos muestras provienen de la misma población. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 por medio de $\chi^2 = S_1{}^g \left\{ \frac{NN'\left(\frac{f_p}{N} - \frac{f_p{}'}{N'}\right)^2}{f_p + f_p{}'} \right\}$. Esto para detallar el cálculo del estadístico Pearson evaluó las siguientes expresiones para cada frecuencia de clase f de las dos muestras
   a. $f + f'$
   b. f/N
   c. f'/N'
   d. $f/N - f'/N'$
   e. $(f/N - f'/N')^2$
   f. $\frac{(f/N - f'/N')^2}{f + f'}$
2. Encontrar P de la distribución χ^2 con la ayuda de las tablas de bondad de ajuste de Elderton (χ^2 y n').

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia de que ambas muestras correspondan a la misma población. Por ejemplo:

Las probabilidades son más de 33,000 a 1 contra la aparición de dos muestras de color de cabello tan divergentes, si fueran muestras aleatorias de la misma población. Considero que podemos concluir que son muestras realmente diferenciadas.

3.2.3.2 Significado parcial 10: La prueba de homogeneidad de Snedecor (1930 y 1933)

En este significado se retoma el ejemplo 11 que se presentó en la página 111. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de homogeneidad se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* (tabla de contingencia) se proporciona el número de manzanas de cada submuestra, número de manzanas lesionadas y porcentaje de la submuestra al que corresponden las manzanas lesionadas para ambos métodos de fumigación. También se hace uso de la *representación simbólica* al presentar los valores del estadístico y la probabilidad en la tabla de contingencia y en las conclusiones.

Situación/problema:

Snedecor e Irwin (1933), presentaron una propuesta para la prueba de homogeneidad que fuera aplicable a los resultados de experimentos que se llevan a cabo en laboratorio donde las frecuencias de las observaciones que provienen de distintos subconjuntos son desiguales. Mostraron cómo funciona la prueba de homogeneidad con el estadístico χ^2 y brindaron algunos ejemplos de la aplicación de la prueba para permitir al investigador evitar conclusiones erróneas si la heterogeneidad aparece.

Conceptos/definiciones:

CD1: *Muestra,* en el mismo sentido que el CD1 del significado parcial 9. Sin embargo, para este caso la muestra se encuentra compuesta por la combinación de diversas submuestras tomadas de diferentes lugares.

CD2: *Población,* en el mismo sentido que el CD2 del significado parcial 9.

CD3: *Variable,* en el mismo sentido que el CD1 del significado parcial 6.

CD4: *Frecuencia observada,* en el mismo sentido que el CD2 del significado parcial 7.

CD5: *Probabilidad* es una medida de la certidumbre asociada a un evento, en este tipo de problema se entiende como una medida de ocurrencia de que la serie de submuestras se haya extraído de una población homogénea, en el sentido de que la probabilidad del evento es uniforme a lo largo del material experimental.

CD6: *Homogeneidad* es una medida para reconocer si es posible que las muestras provengan de una población homogénea.

CD7: *Varianza* es una medida de dispersión de un grupo de datos con respecto a su media, también representa el cuadrado de la desviación estándar.

Propiedades/proposiciones:

PP1: *Distribución de probabilidad* χ^2 de acuerdo con lo establecido en el significado parcial 2 en PP4.

PP2: *Grados de libertad* como $n = (c-1)(r-1)$.

PP3: En cada submuestra la *probabilidad del atributo* está dada por $p = \frac{100s}{n}$ %.

PP4: *La probabilidad promedio* (media ponderada) para la muestra completa es $\bar{p} = \frac{100\sum s}{\sum n}$ %.

PP5: L*a probabilidad promedio del complemento del atributo* es $\bar{q} = 100 - \bar{p}$ %.

PP6: *Estadístico* χ^2

$$\chi^2 = \frac{100\left(\sum sp - \bar{p}\sum s\right)}{\bar{p}\bar{q}}$$

En el que $\sum sp$, es la suma de los productos de la pareja de la frecuencia del atributo por su probabilidad en cada submuestra.

Adicionalmente, para intentar explicar el factor al que se debe la heterogeneidad de la población utilizó:

PP5: *Varianza ponderada* $\sigma^2 = \frac{100(\sum sp - \bar{p}\sum s)}{\sum n}$.

PP6: *Suma de cuadrados* $N\sigma^2$.

PP7: *Estimación de la varianza agrupada* $\sigma_p^2 = \frac{\sum N\sigma^2}{\sum n}$.

PP8: *Varianza de la diferencia de medias* $\sigma_{\bar{p}}^2 = \sigma_p^2\left({}^{1}/_{N_1} + {}^{1}/_{N_2}\right)$.

PP9: *Desviación estándar de la diferencia de medias* $\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2}$.

PP10: $t = \frac{\bar{p}_2 - \bar{p}_1}{\sigma_{\bar{p}}}$.

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar si la población de la que provienen las dos muestras es homogénea. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 para cada muestra por medio de $\chi^2 = \frac{100\,(\sum sp - \bar{p}\sum s)}{\bar{p}\bar{q}}$, para lo cual es necesario obtener la probabilidad del atributo en cada submuestra $p = \frac{100s}{n}$ %, la probabilidad promedio para cada muestra $\bar{p} = \frac{100\,\sum s}{\sum n}$ % y $\bar{q} = 100 - \bar{p}$ %.
2. Obtener n, que es igual al número de grados de libertad $n = (c-1)(r-1)$
3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 en las tablas de distribución de probabilidad χ^2.

Para este ejemplo e intentar explicar el factor de la heterogeneidad en la población Snedecor e Irwin retoman el método de Fisher y realizaron estos procedimientos adicionales:

1. La varianza ponderada $\sigma^2 = \frac{100(\sum sp - \bar{p}\sum s)}{\sum n}$ para las dos muestras.
2. Las sumas de cuadrados se derivan de multiplicar la varianza ponderada por el número de submuestras en cada parcela, $N\sigma^2$.
3. Estimar la varianza agrupada, $\sigma_p^2 = \frac{\sum N\sigma^2}{\sum n}$.
4. Calcular la varianza de la diferencia de medias $\sigma_{\bar{p}}^2 = \sigma_p^2 \left({}^1/_{N_1} + {}^1/_{N_2}\right)$.
5. Obtener la desviación estándar de la diferencia de medias $\sigma_{\bar{p}} = \sqrt{\sigma_{\bar{p}}^2}$.
6. Finalmente encontrar $t = \frac{\bar{p}_2 - \bar{p}_1}{\sigma_{\bar{p}}}$ y su probabilidad asociada.

Argumentos:

Se argumenta utilizando la probabilidad asociada al valor del estadístico χ^2. Por ejemplo:

Los valores de χ^2 indican que no existe una probabilidad constante de lesión bajo cualquier método. La diferencia entre las probabilidades promedio (ponderadas), $3.814 - 2.114 = 1.700\%$ puede deberse a la diferencia en el tratamiento, pero, por otro lado, puede deberse a causas desconocidas que producen la heterogeneidad dentro de las dos muestras.

3.2.3.3 Significado parcial 11: La prueba de homogeneidad con los grados de libertad y tabla de la distribución χ^2 de Fisher (1934)

En este significado se retoma el ejemplo 12 que se presentó en la página 113. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de homogeneidad se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* (tabla de contingencia) se muestran las frecuencias de color de ojos por familia. También se hace uso de la *representación simbólica* al realizar los cálculos para obtener el valor del estadístico y la probabilidad.

Situación/problema:

Fisher (1934), para mostrar cómo funciona la prueba de homogeneidad con el estadístico χ^2 retoma los datos sobre crustaceos (Gammarus) que previamente había analizado a petición de Huxley.

Conceptos/definiciones:

CD1: *Muestra,* en el mismo sentido que el CD1 del significado parcial 9.

CD2: *Población,* en el mismo sentido que el CD1 del significado parcial 9.

CD3: *Una Variable,* en el mismo sentido que el CD1 del significado parcial 6.

CD4: *Frecuencia observada,* en el mismo sentido que el CD2 del significado parcial 7.

CD5: *Frecuencia esperada,* en el mismo sentido que el CD3 del significado parcial 7.

CD6: *Distribución de frecuencias,* en el mismo sentido que el CD4 del significado parcial 7.

CD7: *Distribución Marginal,* en el mismo sentido que el CD5 del significado parcial 7.

CD8: *Probabilidad,* en el mismo sentido que el CD7 del significado parcial 9.

CD9: *Homogeneidad,* en el mismo sentido que el CD8 del significado parcial 9.

CD10: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 3.

Propiedades/proposiciones:

PP1: *Distribución de probabilidad* χ^2 de acuerdo con lo establecido en el significado parcial 2 en PP4.

PP2: *Frecuencia esperada* $e_{ij} = \frac{n_i \times n_j}{n}$.

PP3: *Grados de libertad* $n = (c-1)(r-1)$.

PP4: *Estadístico* χ^2, en el mismo sentido que PP4 del significado parcial 7.

PP5: Cuando $n > 30$ se tiene que

Para grandes valores de n se hace uso del hecho de que la distribución de χ^2 se vuelve casi la distribución normal. Una buena aproximación se da asumiendo que la expresión $\left(\sqrt{2\chi^2} - \sqrt{2n-1}\right)$ esta normalmente distribuida alrededor de cero con desviación estándar unitaria. Si el valor de la expresión es mayor a dos, el valor de χ^2 no está de acuerdo a las expectativas.

PP5: *Regla de decisión,* de acuerdo con lo establecido en el PP6 del significado parcial 3.

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar si dos muestras provienen de la misma población. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 por medio de $\chi^2 = \sum \frac{\left((m+x)-m\right)^2}{m}$, para lo cual es necesario obtener las frecuencias esperadas bajo el supuesto de independencia a partir

de las frecuencias observadas mediante de los totales marginales observados $e_{ij} = \frac{n_i \times n_j}{n}$.

2. Obtener n, que es igual al número de grados de libertad $n = (c-1)(r-1)$.
3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 en las tablas de distribución de probabilidad χ^2de Fisher. En este caso el valor de n no se encuentra dentro de los valores tabulados así que se procede a utilizar la siguiente expresión $\left(\sqrt{2\chi^2} - \sqrt{2n-1}\right)$.

Argumentos:

Se argumenta utilizando la $Diferencia = +0.50 \pm 1$ por lo que, la serie no es significativamente heterogénea; efectivamente todas las familias están de acuerdo y confirman entre sí la proporción negro-rojo observada en el total.

3.2.3.4 Significado parcial 12: La prueba de homogeneidad con corrección de continuidad de Yates

En este significado se retoma el ejemplo 13 que se presentó en la página 114. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de homogeneidad se utilizó *lenguaje natural* en el planteamiento y la interpretación de los resultados, por medio de la *representación tabular* (tabla de contingencia) se muestran las frecuencias de color de cabello de dos muestras. También se hace uso de la *representación simbólica* al referirse a los valores que toma el estadístico y la probabilidad.

Situación/problema:

Pearson (1911) deseaba averiguar la probabilidad de que dos muestras conocidas a priori puedan pertenecer a la misma población, pero sin tener conocimiento a priori de la población. Para ello retomó su estadístico χ^2 que había propuesto para bondad de ajuste en 1900 y con el que también estaban trabajando pruebas de independencia.

Para mostrar cómo funciona la prueba de homogeneidad con la corrección de continuidad de Yates al estadístico χ^2, se ha retomado el problema 8 que presentó Pearson en 1911, dónde se utilizaron datos de los estudios del Dr. Macdonald. Abordamos uno sobre color de cabello de dos muestras. La primera muestra corresponde a personas enfermas de fiebre escarlatina y la segunda muestra de enfermos de sarampión.

Conceptos/definiciones:

CD1: *Muestra,* en el mismo sentido que el CD1 del significado parcial 9.

CD2: *Una Variable,* en el mismo sentido que el CD1 del significado parcial 6.

CD3: *Frecuencia observada,* en el mismo sentido que el CD2 del significado parcial 7.

CD4: *Frecuencia esperada,* en el mismo sentido que el CD3 del significado parcial 7.

CD5: *Distribución de frecuencias,* en el mismo sentido que el CD4 del significado parcial 7.

CD6: *Distribución Marginal,* en el mismo sentido que el CD5 del significado parcial 7.

CD7: *Probabilidad,* en el mismo sentido que el CD7 del significado parcial 9.

CD8: *Homogeneidad,* en el mismo sentido que el CD8 del significado parcial 9.

CD9: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 3.

Propiedades/proposiciones:

PP1: *Distribución de probabilidad* χ^2 de acuerdo con lo establecido en el significado parcial 2 en PP4.

PP2: *Frecuencia esperada*, en el mismo sentido que el PP2 del significado parcial 11.

PP3: *Grados de libertad*, en el mismo sentido que el PP3 del significado parcial 11.

PP4: *Estadístico* χ^2, en el mismo sentido que PP5 del significado parcial 8.

PP5: *Regla de decisión,* en el mismo sentido que el PP7 del significado parcial 4.

Procedimientos:

Encontrar la probabilidad de ocurrencia del valor del estadístico χ^2, cuando éste es utilizado en tablas de contingencia para probar si dos muestras provienen de la misma población. Lo anterior se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico χ^2 por medio de $\chi_c^2 = \sum \frac{(|Frecuencia\ Obervada - Frecuencia\ Esperada| - 0.5)^2}{Frecuencia\ Esperada}$, para lo cual es necesario obtener las frecuencias esperadas bajo el supuesto de independencia a partir de las frecuencias observadas mediante de los totales marginales observados $e_{ij} = \frac{n_i \times n_j}{n}$.
2. Obtener n, que es igual al número de grados de libertad $n = (c-1)(r-1)$.
3. Buscar el valor de la probabilidad en la tabla de la distribución χ^2 en las tablas de distribución de probabilidad χ^2de Elderton.

Argumentos:

Se argumenta utilizando la probabilidad de ocurrencia de que ambas muestras correspondan a la misma población.

Por ejemplo: para $\chi^2 = 26$ y $n = 4$ la tabla de la distribución χ^2 muestra que $P = 0.000{,}086039$. Por lo tanto, las probabilidades son alrededor de 11,628 a 1 contra la aparición de dos muestras de color de cabello tan divergentes, si fueran muestras aleatorias de la misma población.

También se puede argumentar con respecto al valor del estadístico y el valor crítico de la distribución establecidos en la Tabla III que proporciona Yates.

3.2.4 Significado 4: La distribución Chi-cuadrada

De acuerdo con Heyde y Seneta (1977), Karl Pearson fue el primero en obtener la distribución χ^2 como la distribución asintótica del estadístico χ^2. Pearson realizó dicha conexión en su publicación de 1900 cuando trabajaba el problema de bondad de ajuste para una curva de frecuencia.

En este capítulo se ha descrito previamente cómo fue evolucionando tanto el estadístico χ^2 para resolver diversos tipos de problemas, así como la aplicación de la distribución χ^2 para resolver dichos problemas. Dentro de las variaciones a la aplicación de la distribución se abordó lo que hoy conocemos como grados de libertad y su importancia para determinar tanto la probabilidad como las regiones críticas en la distribución χ^2.

Además, se reconoce la importancia que tuvieron las tablas de la distribución χ^2 que facilitaron el uso de las pruebas de contraste del estadístico χ^2, y fue Pearson (1900) quien brindó una primera versión de ellas, en las que se podía encontrar la probabilidad con (χ, n'); además, estableció las fórmulas para calcular la probabilidad en la distribución para cualquier n'. Posteriormente, en 1914, Elderton trabajó en una extensión de dichas tablas también para la probabilidad con (χ^2, n') que se publicaron en un libro dedicado a tablas para estadísticos de Pearson (1914). Fisher (1934) también publicó tablas para la distribución con (n, P), donde se obtiene χ^2 para establecer las regiones de aceptación y de rechazo y estableció una expresión para calcular del valor crítico de χ^2 en la distribución para los casos de $n > 30$.

Aunque en 1900 Pearson nombra a la distribución χ^2, que está dada por la curva de Pearson de Tipo III, de la familia de curvas de frecuencia de Pearson (1895), en ese momento no hace ninguna referencia a Helmert, quien ya había estado trabajando con dicha distribución. Es hasta 1931 que Pearson reconoce el precedente de la distribución χ^2. Acontinuación, abordaremos algunos de los precedentes de la distribución.

De acuerdo con Lancaster (1966), la distribución χ^2 de Pearson se puede considerar la culminación del trabajo de mínimos cuadrados de Laplace. Laplace trabajó con la distribución gamma en el contexto de la teoría de los errores, siendo la distribución χ^2 un caso especial de la distribución gamma. En 1836, Laplace realizó una discusión del error con parámetros ajustados donde señaló que si tres ángulos de un triángulo se pueden medir con igual precisión, la probabilidad de los errores α, β y γ se podía escribir como $\exp(-h\alpha^2 - h\beta^2 - h\gamma^2)$ y la probabilidad de que el error del ángulo $A^{(n)}$ se encuentrara dentro de los límites $\pm\frac{2}{3}\theta r'$ era $2\pi^{-\frac{1}{2}} \int \exp -r'^2 dr'$, la integral será tomada de $r' = 0$. De acuerdo con Lancaster (1966), Laplace derivó un tipo de distribución gamma para h, es una distribución para las observaciones bajo una hipótesis Bayesiana.

Laplace proporcionó con su trabajó las técnicas necesarias para el trabajo posterior de Bienaymé de 1838, donde desarrolló la distribución de una forma lineal en los éxitos y fracasos del binomio como preliminar a una extensión de la teoría a la multinomial. Partió de que p_i, donde $i = 1,2,3$ son probabilidades desconocidas de una observación sobre pertenecer a una de las n clases mutuamente excluyentes y, además, que $p_1 + p_2 + p_3 = 1$ y $N_1 + N_2 + N_3 = N$; suponiendo que se realizaron N_i observaciones independientes que caen en la clase i^{th}. Bienaymé trabajó bajo la probabilidad Bayesiana y señaló que $P(a \leq V \leq a')$ donde $V = \gamma p_1 + \gamma p_2 + \gamma p_3$ y $\gamma_1 \leq \gamma_2 \leq \gamma_3$. También asumió como conocida la probabilidad multinomial para $P(\{N_i\}|N, \{p_i\})$ y entonces la probabilidad de la hipótesis, $V = v$, $P(V = v) = p_1^{N_1} p_2^{N_2} p_3^{N_3} / \sum p_1^{N_1} p_2^{N_2} p_3^{N_3}$, a partir de la sumatoria de todos los valores posibles obtuvo que $\gamma p_1 + \gamma p_2 + \gamma p_3 = v$. Siguiendo a Laplace, encontró que la integral expandida a los exponenciales es $N_1^{N_1} N_2^{N_2} N_3^{N_3} N^{-N} \times \int_{b'}^{b} \int_0^1 \exp[-N^2(z_1^2 N_1^{-1} + z_2^2 N_2^{-1} + z_3^2 N_3^{-1} + S)]\, dz_3 du$, y proporciona $v = N^{-1} \sum_1^m N_i \gamma_i \pm c N^{-1} [\sum_1^m N_i (\gamma_i - \bar{\gamma})^2]^{\frac{1}{2}}$ como una estimación de V, donde $\bar{\gamma} = \sum N_i \gamma_i / N$. De acuerdo con Lancaster (1966), Bienaymé obtuvo la distribución χ^2 como resultado asintótico sin el supuesto de normalidad.

En 1852, Bienaymé utilizó el modelo de mínimos cuadrados donde ε_j son los errores en las observaciones y los errores en las estimaciones son $r = K\varepsilon$. Propusó obtener la distribución conjunta de los errores r_i mediante la transformación de Laplace y obtuvo la expresión para las discrepancias de r_i de su verdadero valor como $P = \left(\frac{2}{(\sqrt{\pi})}\right)^m \int_0^{\gamma} u \exp(-u^2) du \ldots \int_{m-1} \ldots \int \frac{dt_1 \ldots dt_{m-1}}{\sqrt{(u^2 - t_1^2 - t_2^2 \ldots - t_{m-1}^2)}}$. En concordancia con Lancaster (1966), Bienaymé obtuvo la distribución de una suma de cuadrados en la teoría de los mínimos cuadrados.

En cuanto al trabajo de Ellis, en 1844 Lancaster (1966) señaló que por medio de un problema sugerido en un estudio de Isaac Newton sobre la duración de los reinados, proveyó un método que le permitía determinar la distribución de la suma de n varibales aleatorias distribuidas independientemente por el uso de transformadas de Laplace. Para el problema Ellis supuso que la distribución era exponencial $f(x) = e^{-x}$, $0 \leq x \leq \infty$, y mostró que la suma de las n variables tenía como función de frecuencia a la distribución gamma. En el artículo 'On the method of least squares' de Ellis de 1844, hizo una crítica revisión de la teoría de mínimos cuadrados, donde mostró que si los $(2m)^{th}$ momentos de la media de un gran número de observaciones, entonces $a_{i_1}^2 E X_{i_1}^2 a_{i_2}^2 E X_{i_2}^2 \ldots a_{i_m}^2 E X_{i_m}^2$, donde $\{i_1, i_2, \ldots, i_m\}$ es una combinación m de los números $1, 2, \ldots, N$, que necesitan ser considerados. Ellis señaló también que estos momentos de la suma se comportan como si estuvieran compuestos de variables normales.

En 1875, Fredrich Robert Helmert publicó un artículo titulado sobre el cálculo del error probable a partir de un número finito de nuestros errores de observación, en el que de acuerdo con Lancaster (1966) y Hald (2007), brinda una distribución de frecuencia de la suma de cuadrados $\sum \varepsilon_i^2 = n\sigma_2$, donde ε_i son variables normales mutuamente independientes, con un valor esperado de cero, donde la expresión es $f\left(\frac{v}{2}\right) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{v}{2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{v}{2}\right)$, entonces v tiene una distribución χ^2 con n grados de libertad, en 1876 Helmert reemplaza n por $n-1$. En especial este trabajo de 1876 resulta interesante debido a que también trabajó con números pequeños de n. Helmert, consideró distribuciones conjuntas de z_i a las que aplicó máxima verosimilitud y obtuvo $\hat{\sigma}^2 \sum z_1^2/n$ y obtuvo las distribuciones conjuntas de las $n-1$ diferencias $u_j = x_j - \bar{x}$. También realizó una transformación a $\sum (x-\bar{x})^2/\sigma^2$ para las $x's$ y las $v's$, que es conocida como la transformación de Helmert, con la cual demostró que $\sum_1^n (x_i - \bar{x})^2/\sigma^2 = \sum_1^{n-1} v_i^2/\sigma^2$ se distribuye como χ^2_{n-1}, donde $x_i = \mu + z_i$ (Lancaster, 1966).

En Lancaster (1966), se menciona que también se le conoce como distribución de Helmert a la distribución conjunta de la media muestral y desviación estándar para muestras de observaciones aleatorias independientes, que provienen de una población normal.

En 1934, en la ampliación de su libro 'Statistical Methods for Research Workers', en donde publicó sus tablas para la distribución χ^2, Fisher puntualizó que la distribución χ^2 es una distribución continua,

> sin embargo, la distribución de las frecuencias debe ser siempre discontinuas. En consecuencia, el uso de la distribución χ^2, en la comparación de las frecuencias observadas con las esperadas solo puede ser de precisión aproximada, la distribución continua sería de hecho el límite hacia el cual la verdadera distribución discontinua tiende a medida que la muestra es cada vez más grande (Fisher, 1934, p. 96).

3.3 Síntesis de los significados del estadístico χ^2

Mediante el estudio histórico-epistemológico sobre el estadístico Chi-cuadra fue posible identificar cuatro grandes problemáticas, cuyo abordaje dan paso a cuatro grandes "significados de referencia" para este estadístico: 1) Prueba de bondad de ajuste, 2) Prueba de independencia, 3) Prueba de homogeneidad y, 4) Distribución Chi-cuadrada. Además, cada uno de estos grandes significados de referencia se componen de significados parciales, que dan cuenta de la evolución progresiva, de lo informal a lo formal, de este estadístico (ver Figura 3.11).

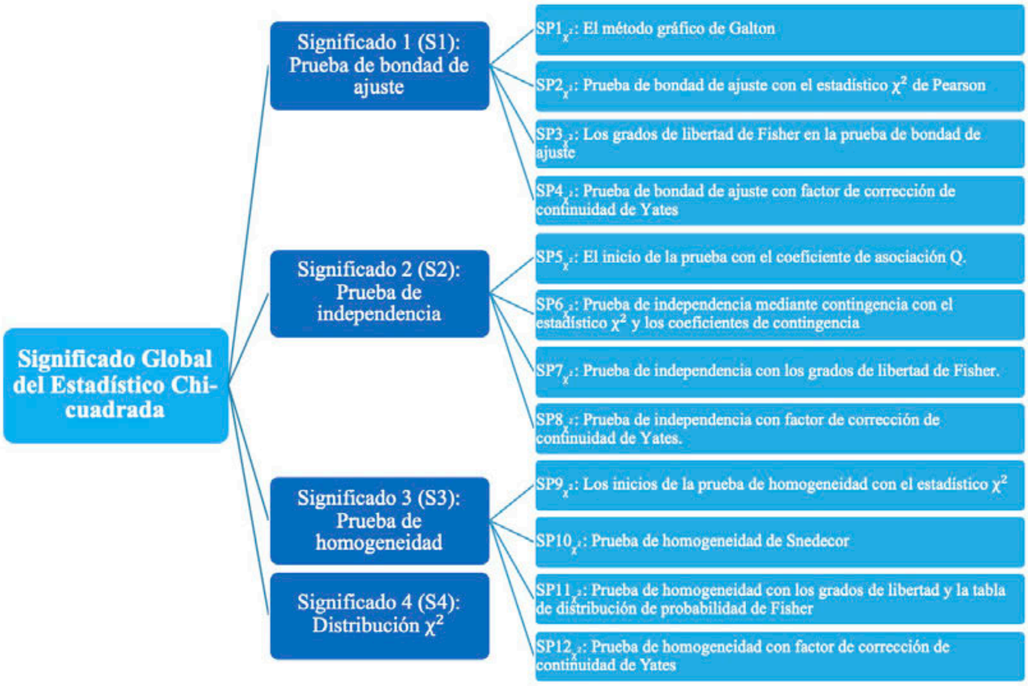


Figura 3.11. Significado global del Estadístico Chi-cuadrada

Una prueba de bondad de ajuste con el estadístico Chi-cuadrada, a manera de prueba de hipótesis, busca valorar en qué medida se ajusta un grupo de datos observados a cierta distribución teórica previamente establecida, por medio del contraste de frecuencias observadas y esperadas. El SP1 corresponde a una prueba de bondad de ajuste intuitiva, donde se analizan las desviaciones de cada dato de la muestra con respecto al valor medio de la misma muestra.

El segundo gran significado, la prueba de independencia, nos permite determinar si dos variables se encuentran asociadas. Cabe destacar que en el SP6 se generalizó la noción de asociación (del SP5), dando lugar a la prueba de independencia con el estadístico χ^2. Aunque la prueba de la de independencia y la de homogeneidad son matemáticamente idénticas, esta última determina si es posible que varias muestras provengan de la misma población. A medida que fue evolucionando el estadístico χ^2, para resolver diversos problemas, también fue evolucionando la distribución Chi-cuadrada, esto a partir de que, en 1900, Pearson la obtuviera como la distribución asintótica del estadístico χ^2 cuando trabajaba el problema de bondad de ajuste para una curva de frecuencia.

Esta caracterización de los significados del estadístico Chi-cuadrada nos permite acceder a la riqueza matemática y visualizar la variedad de caminos para la enseñanza y aprendizaje de esta noción. Además, nos puede ayudar a identificar elementos para una transición continua del RII al RIF sobre el estadístico χ^2.

3.4 Reflexiones finales del estudio histórico-epistemológico sobre el estadístico Chi-cuadrada

De acuerdo con E. L. Lehmann (1992), en el siglo XIX las pruebas de hipótesis tendían a ser informales y ocurrían primordialmente en el contexto específico de las aplicaciones y fue hasta los trabajos de K. Pearson, R. A. Fisher y W. S. Gosset que la atención se centró en las pruebas de hipótesis como una metodología. En la búsqueda para fortalecer la lógica de las pruebas de hipótesis, de los trabajos de Fisher, J. Neyman y E. S. Pearson surgieron algunas ideas sobre la metodología de las pruebas de hipótesis.

Gosset le había sugerido a Egon S. Pearson que la única razón válida para rechazar una hipótesis estadística era tener una hipótesis alternativa que pudiera explicar los eventos con un mayor grado de probabilidad (Lehmann, 1992), lo cual daría paso a la hipótesis alternativa.

Jerzy Neyman y E. S. Pearson en 1928 introdujeron una teoría de pruebas de hipótesis desde el punto de vista frecuentista (Hald, 2007), donde ya trababan aspectos como el error tipo I y tipo II, potencia de prueba, hipótesis simple y compuesta; sin embargo, en ese momento se encontraba centrada la prueba en la razón de verosimilitud que le había propuesto Gosset a Pearson, como "*un criterio que le permitía rechazar la hipótesis nula cuando la máxima verosimilitud de la muestra observada es lo suficientemente grande comparada con su valor bajo hipótesis*" (Lehmann, 1992, p. 68).

Y aunque E.S. Pearson consideraba que la razón de verosimilitud les proporcionaba el método que habían estado buscando, Neyman no estaba aún satisfecho, lo que los llevó a seguir trabajando sobre ese criterio y en 1930, Neyman planteó la solución, que se convirtió en el lema fundamental de Neyman y Pearson, utilizado para prueba de hipótesis con hipótesis simples, estableciendo el criterio para rechazar la hipótesis con las funciones de verosimilitud y pre asignando el nivel de la prueba, es decir, la probabilidad de rechazo de la hipótesis nula. Este nuevo enfoque maximiza la potencia de prueba.

En 1933, Neyman y E. S. Pearson publicaron algunos ejemplos de la aplicación de su metodología lo que les permitió sentar las bases de su teoría de prueba de hipótesis, esta teoría es una de las más usadas actualmente en inferencia. De acuerdo con Lehmann, la formulación de Neyman y E. S. Pearson marcó el inicio de un nuevo paradigma, "*una teoría de muestras pequeñas en la cual los procedimientos estadísticos son derivados acordes con algunos criterios de optimización y, por lo tanto, como solución a problemas matemáticos claramente establecidos*" (Lehmann, 1992, p. 69).

Actualmente, la prueba de bondad de ajuste con el estadístico χ^2 requiere que las frecuencias esperadas sean iguales o mayores a cinco, esta regla la impuso Yates en 1934 y se aborda en el apartado sobre contribuciones posteriores; y podría resolverse por medio de un paquete estadístico. En caso de que las frecuencias sean inferiores a cinco es necesario utilizar el factor de corrección de continuidad, el cual se abordó en los significados parciales 4, 8 y 12. Para realizar la prueba es necesario establecer las hipótesis (H_0 y H_1), por ejemplo:

H_0: $Las\ frecuencias\ observadas\ se\ ajustan\ a\ la\ distribución\ teórica$

H_1: $Las\ frecuencias\ observadas\ no\ se\ ajustan\ a\ la\ distribución\ teórica$

Se considera que en la época de K. Pearson la H_0 estaba implícita en la prueba. A diferencia de los procedimientos que se pueden observar en el ejemplo 2 que se brinda de la prueba de bondad de ajuste que realizó Karl Pearson, para el cálculo de la probabilidad actualmente se utilizan los grados de libertad (que fueron introducidos por Fisher y se comentaron previamente) y se requiere contrastar el valor-p con el nivel de significancia de la prueba de bondad de ajuste que se habría definido previamente (se supone un $\alpha = .05$, este valor de alfa está definido implícitamente). Si el valor-p es mayor al α no se rechaza H_0 (esta regla de decisión se introdujo posteriormente por Neyman y Pearson). En el caso del ejemplo 2, es menor, el estadístico χ^2 calculado se puede localizar en la región de rechazo de la curva, lo cual refuerza el comentario que realizó Pearson sobre que en 15 o 16 ocasiones en 10,000,000 ensayos este sistema de errores se podría encontrar en una distribución normal. También se puede recurrir a obtener el valor crítico (19.675 para este ejemplo) para el alfa de 0.05, y si el estadístico es más grande que el valor crítico, cae en el área de rechazo y por la tanto se rechaza la hipótesis nula.

Mientras que una práctica común en el contraste de independencia y homogeneidad actual es calcular las frecuencias esperadas suponiendo independencia y se requiere que estas frecuencias sean iguales o mayores a cinco. Para realizar la prueba es necesario establecer las hipótesis $(H_0 \ y \ H_1)$, se considera que la H_0 estaba implícita en la prueba, desde que Yule (1900) trabajaba el coeficiente de asociación hasta el momento en que Yates establece su factor de corrección de continuidad en 1935. Por ejemplo, siguiendo con el problema 7 que se trabajó previamente con la prueba de independencia con el factor de corrección de Yates:

$$H_0: Las\ distribuciones\ de\ malocusión\ y\ tipo\ de\ alimentación\ son\ independientes$$

$$H_1: Las\ distribuciones\ de\ malocusión\ y\ tipo\ de\ alimentación\ se\ encuentran\ asociadas$$

Al utilizar un software para realizar estas pruebas se debe tener precaución pues algunos no tienen la opción de utilizar la corrección de continuidad, que correspondería al SP4, SP8 y SP12. Cuando se trabaja con frecuencias inferiores a cinco las pruebas con el estadístico χ^2 requieren esta corrección, pero los softwares (e.g., Minitab) suelen dar una advertencia como una nota al pie en los resultados. Entonces, es importante la comprensión de lo que se encuentra de trasfondo de los procesos o procedimientos que realizan algunos softwares en el análisis de los datos.

CAPÍTULO 4

El Estadístico t $- Student$.

Reconstruyendo el Significado Holístico de Referencia

Un estadístico es una variable aleatoria, en su significado como función, cuyo valor depende de la observación muestral.

INTRODUCCIÓN

Al igual que en el Capítulo 3, en este Capítulo también se encuentran: (1) la descripción de las prácticas matemáticas que se desarrollaron en la historia para resolver las grandes problemáticas sobre el estadístico t-Student; (2) A partir de dichas prácticas matemáticas desplegadas se identificaron las redes de objetos primarios o configuraciones epistémicas para dicho estadístico y finalmente; y (3) una síntesis de los significados del estadístico t-Student.

4.1 Análisis de las prácticas históricas sobre el estadístico $t - Student$

Como se señaló anteriormente, el análisis de las prácticas matemáticas corresponde al primer nivel de análisis de la actividad matemática. En este nivel se describen las acciones o prácticas matemáticas que se realizaron para dar solución a las grandes problemáticas que permitieron el surgimiento y evolución del estadístico t -Student. Se identificaron tres grandes problemáticas, que corresponden a las pruebas t-Student: para una muestra, para dos muestras y para dos muestras dependientes. Adicionalmente, se considera la importancia que tiene en estas pruebas la distribución t-Student.

4.1.1 Los trabajos previos al estadístico $t - Student$

Para abordar los trabajos que sentaron un precedente para las pruebas con el estadístico t-Student, se retoma de la sección 3.2.1.1 el método de intercomparación que desarrolló Galton en 1875.

Recordemos que Galton buscaba una forma de establecer si era apropiado trabajar un conjunto de datos observados bajo las condiciones de la distribución normal y en 1875 trabajó el método de intercomparación. Galton ordenó los datos que tenía de forma ascendente y procedió a graficar los datos versus los cuartiles, en la Figura 3.1 se muestra la curva ideal. En el proceso de estadística por intercomparación se tiene que n es la longitud de la base de la ojiva, en $\frac{1}{4}$ se localiza a p, m se encuentra en $\frac{1}{2}$ representando al valor medio de la serie, y q en $\frac{3}{4}$. Cuando se relaciona a p y q con m, los dos primeros brindan información que permite estimar la divergencia. Entonces $q - m$ es la divergencia o el error probable de la parte de la serie que excede al valor medio, y $m - p$ es la divergencia de la otra parte de la serie. Se debe cumplir que $q - m = m - p$ para que la serie se considere simétrica.

Galton expresó la variación o divergencia de las frecuencias con respecto al valor medio en términos del error probable. Por lo que en p en $\frac{1}{4}$ y en q en $\frac{3}{4}$ se encontrarían los valores de la serie que tienen una divergencia de una unidad de error probable ya sea por encima o debajo del valor medio.

En 1880, Galton publicó algunos ejemplos del uso de su método, estadística por intercomparación. Como se ha mencionado, las investigaciones de Galton versaban sobre la herencia, enfermedades y la raza; con los ejemplos de 1880 sobre intercomparación buscaba proporcionar datos confiables en cuanto a la frecuencia relativa con la que se heredan diferentes facultades en diferentes grados. Específicamente proporcionó ejemplos donde deseaba definir los diferentes grados de viveza con los que diferentes personas tienen la facultad de recordar escenas familiares bajo la forma de imágenes mentales, y las peculiaridades de las visiones mentales de diferentes personas (iluminación, definición y color).

Para ejemplificar cómo funcionaría esta intercomparación que estaría además sujeta a la interpretación y valoración de la persona que analizaría los datos, utilizó un ejemplo más sencillo, este referiría a la estatura de 1000 hombres.

Ejemplo 14: Método de intercomparación sobre las alturas de ingleses y franceses (Galton, 1880, pp. 6-10)

> Si deseamos comparar las alturas de ingleses y franceses […] En mi proceso, no hay necesidad de un estándar externo. Claramente se trata de lo mismo si tomo once hombres y, midiéndolos uno contra otro, los clasifico en orden, comenzando con el más alto y terminando con el más bajo, o si los mido por separado con una medida de pie y los clasifico en el orden de la magnitud de las medidas registradas en mi cuaderno. En cada caso, el hombre más alto se parará primero, el próximo segundo más alto, y así sucesivamente hasta el último. En cada caso, el mismo hombre ocupará el sexto o el lugar más intermedio y, por lo tanto, representará la altura media de todos ellos. No quiero dar a entender que "el valor medio o media" es idéntico a "media" o "promedio", ya que no es necesariamente así. Pero sí digo que valor medio o media puede estar estrictamente definida y, por lo tanto, si deseamos comparar las alturas de los ingleses con los franceses, procederemos de manera tan científica si comparamos sus alturas medias como si comparamos sus alturas promedio. Ahora se observará que tenemos las alturas medias sin una regla de pie o cualquier estándar externo; Lo hemos hecho completamente por el método de intercomparación. […] Supongamos que, como antes, están todos dispuestos en orden de estatura, a distancias iguales en una larga línea $A\ B$, con la espalda vuelta hacia nosotros. Si hay mil hombres, debemos suponer que $A\ B$ se dividirá en 1000 partes iguales, y un hombre se colocará en el medio de cada parte. El hombre más alto tendrá A cerca de su izquierda, y el hombre más bajo tendrá B cerca de su derecha. Formarán una serie como se muestra en la Fig. I. [*Figura 4.1*], donde las subdivisiones de $A\ B$ están indicadas por las líneas verticales, y las posiciones donde los hombres están parados se muestran mediante puntos a medio camino entre esas líneas.

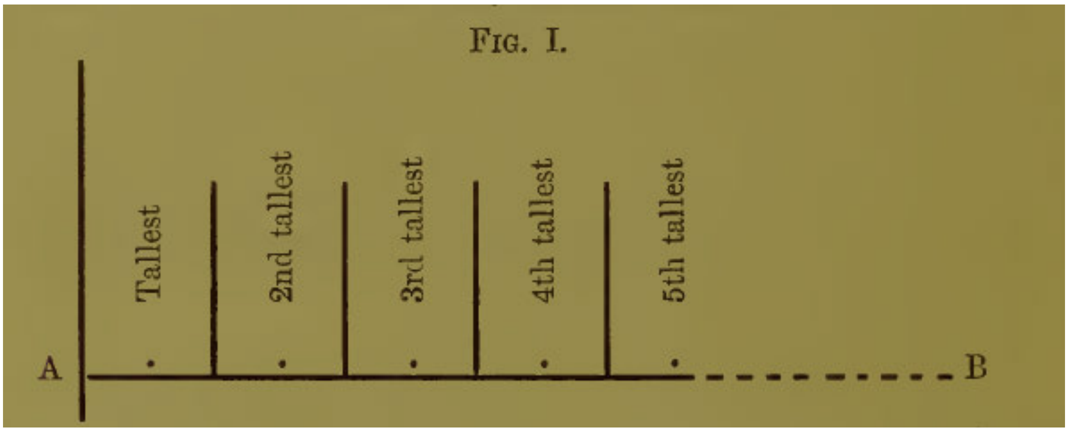


Figura 4.1. Hombres ordenados por estatura, del más alto al más bajo (fuente: Galton, 1880, p. 8)

Debido a la continuidad de cada serie estadística, la línea imaginaria dibujada a lo largo de la parte superior de las cabezas de los hombres formará una curva regular, y si podemos registrar esta curva, se nos proporcionarán datos para determinar la altura de cada hombre en toda la serie Dibujando una curva de este tipo para los ingleses y otra para los franceses, y superponiendo los dos, deberíamos poder comparar la estatura de las dos naciones en los detalles más mínimos. Se registra una curva midiendo sus ordenadas. Si dividimos *A B* por un número suficiente de subdivisiones equidistantes y medimos las ordenadas en cada una de ellas como se hizo en la figura II [*Figura 4.2*].

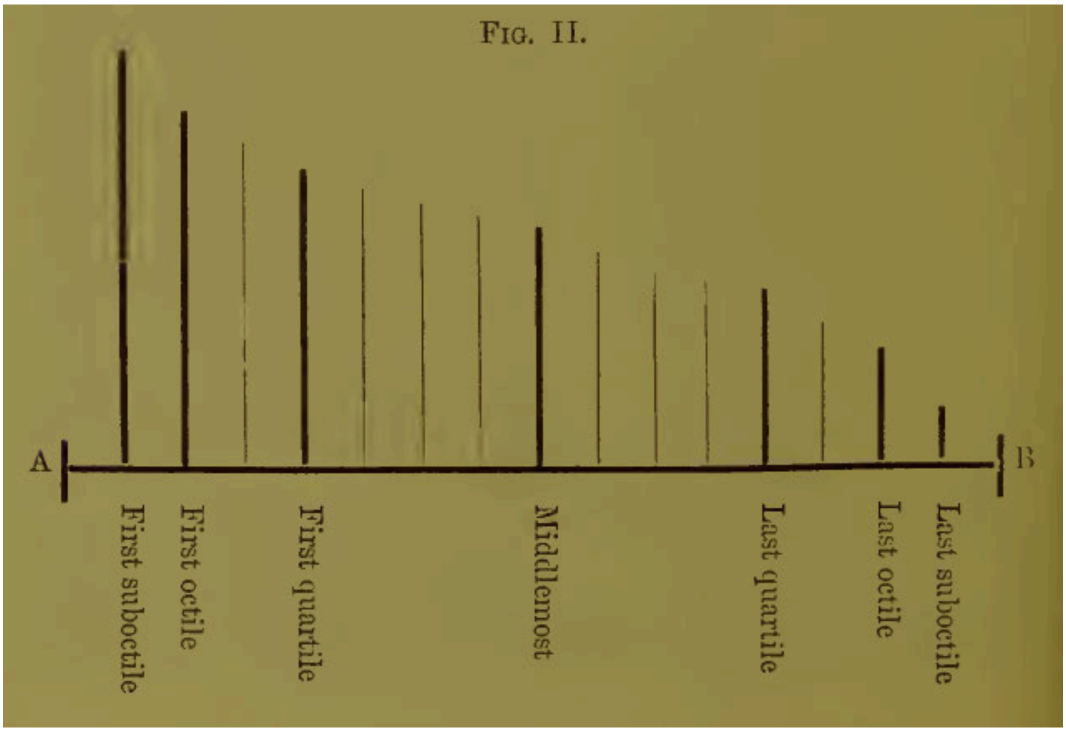


Figura 4.2. Ordenadas en A B sobre la altura de los hombres (fuente: Galton, 1880, p. 8)

(donde las ordenadas solo se muestran, y no la curva), en cualquier momento podemos trazarlas a escala, y al trazar una línea libre que toque sus cimas, podemos reproducir con mayor o menor precisión la curva. Sucede, sin embargo, por el carácter peculiar de todas las curvas estadísticas, que las ordenadas a distancias iguales no son las más adecuadas. Los casos mediocres son siempre tan numerosos que la curva fluye en una línea constante y casi recta alrededor de su centro, y se convierte en un desperdicio de esfuerzo tomar muchas medidas a su alrededor. Por otro lado, su forma varía rápidamente en cada extremo, y allí las observaciones deberían ser numerosas. Las estaciones más adecuadas son aquellas que corresponden a ordenadas que difieren en altura en grados iguales, y estos lugares admiten ser descubiertos por consideraciones a priori en ciertos supuestos generales.

Sin embargo, haremos bien en ignorar esas minucias en las que hice mucho hincapié en la Memoria, y adoptar el plan más simple de subdivisiones sucesivas de AB, y de medir las ordenadas que se muestran con líneas oscuras en la figura II. 'valor medio', primer y último 'cuartil', primer y último 'octil', y primer y último 'suboctil'. Esto es lo suficientemente lejos para nuestros deseos actuales, aunque el sistema admite una extensión indefinida. Al medir la "ordenada", me refiero a medir el "hombre" cuyo lugar en la serie está más cerca de la verdadera posición de esa ordenada. No se necesita una coincidencia absoluta en un trabajo tan grosero como este; así, en una serie de 100 hombres, ya sea el 50 o el 51 cumplirán el deber para el más medio. Los lugares que he tomado en la serie de 100 hombres para las varias estaciones, son, 6 y 94 para el primer y último suboctiles, 12 y 88 para los octiles, 25 y 75 para los cuartiles, y 50 para el valor medio.

Siete hombres se convierten así en representantes eficientes de una clase muy numerosa. Como regla general, se encontrará que estos siete representantes seleccionados diferirán entre sí en intervalos aproximadamente iguales, la diferencia entre el suboctil y el octil suele ser aproximadamente la misma que entre el octil y el cuartil, y entre el cuartil y el valor medio.

Como cuestión de interés, y para la posibilidad de encontrar casos muy excepcionales, también registro el más alto y el más bajo de la serie, pero debe entenderse claramente que estos no tienen un valor sólido para fines de comparación. En primer lugar, su posición como ordenadas es incierta a menos que se indique el número del grupo de

> casos, ya que cuando el número es grande, la posición de los más altos y más bajos estará más cerca de A y B, respectivamente, que cuando es pequeño. En segundo lugar, siendo los casos externos más altos y más bajos, es más probable que tengan un carácter excepcional que cualquiera de los que se interponen entre vecinos, uno a cada lado.
>
> La comparación de cualquiera de los dos grupos se realiza al cotejar sus siete representantes cada uno con cada uno, el primer suboctil de uno con el primer suboctil del otro, el primer octil con el primer octil, el primer cuartil con el primer cuartil, etc. También comparto el más alto de cada uno, y nuevamente el más bajo de cada uno, como un mero asunto de interés, pero no como una operación estadística precisa, por las razones ya expuestas. (Galton, 1880, pp. 6-10)

Stigler (2017), destaca el método de intercomparación como uno de los pilares de la sabiduría estadística. En especial rescata la idea de que "*las comparaciones estadísticas se pueden realizar estrictamente en términos de la variación interna de los datos sin referencia a o de criterios externos*" (p. 77). Y es precisamente esta idea del método de intercomparación la que retoman en diferentes décadas Edgeworth, William S. Gosset y Fisher para proponer nuevos métodos y/o actualizaciones.

En 1885, Edgeworth desarrolló un método de análisis de tablas estadísticas basado en la variación interna de los datos (intercomparación), con el cual pretendía eliminar el azar de la teoría de los errores. En este método para determinar las variaciones se apoyó en el constructo que llamó fluctuación, por ejemplo, para medir la variación de la tasa de mortalidad de una ciudad en el tiempo (de un año a otro o en una serie de años) o la variación de la misma tasa pero para dos ciudades.

Edgeworth señalaba que este método tenía un carácter lógico pero también deseaba probar que era indispensable. Para lo cual, parte de la lógica deductiva y da una serie de ejemplos donde si bien es cierto podemos deducir empíricamente, basados en el conocimiento cotidiano, para precisar la inferencia se hace evidente reducir las aparentes irregularidades de los fenómenos y para ello es necesario recurrir a la probabilidad, error probable, error medio y fluctuación. También observó una constancia de la fluctuación, señalando el ejemplo del hexámetro Virgiliano, a lo que llamó un fenómeno insospechado y sorprendente. Y añadió que "*esta constancia de fluctuación, además de ser un dato necesario para la eliminación del azar,* *parece ser en sí misma una circunstancia muy curiosa, no predecible a partir de los principios generales del cálculo de las probabilidades*" (Edgeworth, 1885, p. 633).

En 1885, Edgeworth dirigió su investigación con problemas, principalmente, sobre tasas de defunción, de matrimonio u otras tasas que se podían trabajar por lugares y tiempo, de estas estadísticas buscaba encontrar un cierto coeficiente o constante denominado fluctuación (módulo al cuadrado). Edgeworth explicó su método a través de dos problemas principales, hemos retomado algunos extractos indispensables para visualizar cómo evoluciono y utilizó el método de intercomparación de Galton y comparó dos poblaciones dentro de su método de análisis de fluctuación.

Ejemplo 15: Método de análisis de Edgeworh aplicado a la mortalidad (Edgeworth, 1885, pp. 639-643)

> […] Se requiere calcular a partir de los retornos del registrador general la fluctuación de las tasas de mortalidad en los siguientes casos: (1) se toman al azar varios grupos del mismo tamaño (digamos 100 o 1,000) de la población residente en el mismo lugar en el mismo año, y las tasas de mortalidad calculadas para cada grupo; (2) las tasas de mortalidad se relacionan con grupos que se han formado por selección aleatoria de individuos de diferentes años y lugares.
>
> La siguiente metáfora puede ayudarnos a formar las concepciones apropiadas. El sitio de una ciudad consta de varias terrazas, producidas por las más gentiles acciones geológicas. Las terrazas se encuentran paralelas entre sí, este y oeste. Están intersectadas perpendicularmente por crestas que han sido producidas por desplazamiento ígneo. Podríamos suponer que la agencia volcánica viaja a una velocidad uniforme de oeste a este, produciendo cada año una cresta de la misma amplitud. No se sabe antes de la observación si el desplazamiento de un año se asemeja (en cualquiera, o en todas las terrazas) a los desplazamientos en los años próximos. Tampoco se sabe si el desplazamiento en una terraza puede ser idéntico al de las terrazas vecinas. Sobre el terreno, así intersectado y escarpado, se construyen varias casas. La altura de cada techo de la casa sobre el nivel del mar se puede determinar barométricamente o de otra manera. Se registra la altura media sobre el mar de las azoteas de cada acre [*ver Tabla 4.1*]. A partir de estos datos, se requiere obtener la fluctuación de la altura de las casas sobre el suelo.

Tomemos primero el caso de mortalidad a todas las edades; y luego el de períodos de edades especiales.

Tabla 4.1. *Análisis de la mortalidad de Edgeworth* (fuente: Edgeworth, 1885, p. 640)

	1876	**1877**	**1878**	**1879**	**1880**	**1881**	**1882**	**1883**	**Suma**	**Media**	**Fluctuación**
Berks	175	172	187	186	181	153	169	166	1389	**178.5**	*224*
Herts	174	165	185	184	176	186	163	188	1401	**175**	*176*
Bucks	182	171	186	195	179	162	177	183	1435	**179.5**	*172*
Oxford	179	182	194	183	180	169	167	166	1420	**177.5**	*162*
Bedford	196	174	203	195	198	171	181	184	1502	**188.5**	*246*
Cambridge	173	177	190	191	187	165	171	181	1435	**179.5**	*158*
Suma	1079	1041	1145	1134	1101	986	1028	1068	8582	**1073**	*1138*
Media	180	173.5	191	189	183.5	164	171	178	1630	**179**	**190** *146*
Fluctuación	*124*	*55*	*77*	*50*	*107*	*68*	*73*	*152*	__	*46* **88**	__

Aquí hay una serie de datos extraídos de las estadísticas del registrador general acerca de las tasas de mortalidad en diferentes condados (Report, 1883, p. 53), sobre las cuales he operado (a.) de la manera explicada en p. 634. Sea C^2 la fluctuación que se requiere determinar; coloca $C^2 + C_t^2$ para la fluctuación de una fila. El valor medio de $C^2 + C_t^2$ es 189. La fluctuación de la fila marginal (en figuras negras) es $\frac{C^2}{6} + \alpha C_t^2$; donde α es un coeficiente que depende de los grados de identidad prevaleciendo entre las fluctuaciones de las diferentes filas. Si trazamos las curvas formadas por las filas, aparecerá en ellas una semejanza considerable. Sin embargo, α debe quedarse corta de la unidad. El valor de la fluctuación marginal es 146. Eliminando C_t^2 de las dos ecuaciones que se han establecido, tenemos $C^2 = \frac{\alpha 190 - 146}{\alpha - \frac{1}{6}}$. Está claro que C^2 incrementa con α. Por lo tanto, como α no puede exceder 1, tenemos un límite superior de C^2, $\frac{44}{\frac{5}{6}}$ o 53.

Procediendo de manera similar con las columnas y los marginales verticales, tenemos $C^2 + C_p^2 = 88$.

$$\frac{C^2}{8} + C_p^2 = 46$$

De donde $C^2 = \frac{\beta 88-46}{\beta-\frac{1}{8}}$. Y como β no puede ser mayor que la unidad, tenemos un límite superior de C^2, $\frac{42}{\frac{7}{8}}$ o 48.

[…] Estos resultados se confirman mediante un método que puede denominarse intercalación o intercomparación. [Retomando la metáfora] Suponga, en la formación del sitio de nuestra ciudad imaginaria, que los desplazamientos ígneos habían sido exactamente los mismos a lo largo de cada cresta; es decir, que cada acre a lo largo de la misma cresta se haya levantado exactamente en la misma medida, podríamos entonces llevar las dos hectáreas que pertenecen a la misma terraza y diferentes crestas al mismo nivel, si supiéramos la diferencia entre la altura de las crestas. Luego, una comparación de la altura media sobre el mar (corregida) de las casas en los dos acres proporcionaría un dato para calcular la fluctuación de la altura de la casa. El cuadrado de la diferencia entre las dos alturas medias corregidas debe equipararse a la fluctuación requerida. Por supuesto, este cálculo es muy inexacto. No se puede suponer que las crestas sean crestas de montañas afiladas ininterrumpidas, ni, incluso si así fuera, deberíamos tener las medias para determinar la diferencia entre las alturas de las dos grandes crestas. La diferencia entre la altura media dada (de la parte superior de la casa sobre el mar) se vería afectada por la fluctuación de las casas. En consecuencia, la operación indicada daría solo un límite superior para la fluctuación buscada.

La siguiente tabla [*Tabla 4.2*] ilustra este proceso […] la primera columna muestra las tasas de mortalidad durante ocho años en Berkshire; y en la segunda columna, la de Cambridgeshire, la tercera, da la diferencia por la cual la segunda excede a la primera; la cuarta columna, corrige las diferencias por la resta de 6 (179.5-173.5); la quinta columna contiene los cuadrados de las diferencias, cuya media es un valor para el (límite superior del) módulo buscado. […]

Tabla 4.2. *Análisis de mortalidad de dos poblaciones de Edgeworth* (fuente: Edgeworth, 1885, p. 643)

	Berkshire	**Cambridgeshire**	**-**	**+**	**-**	**+**	e^2
1876	175	173	2		8		64
‘77	172	177		5	1		1

‘78	187	190	3	3			9
‘79	186	191	5	1			1
‘80	181	187	6		0		0
‘81	153	165	12			6	36
‘82	169	171	2	4			16
‘83	166	181	15			9	81
Suma	1,389	1,435		17		15	208
Media	173.5	179.5					26

En 1885, Edgeworth trabajó comparando las medias de dos poblaciones, lo que se podría considera un precursor de la prueba t-Student para dos muestras, método que decadas más adelante idearía y trabajaría ampliamente Fisher.

4.1.2 Los inicios del estadístico $t - Student$

En el siglo XIX, las pruebas de hipótesis sobre la media de una población basados en una muestra de la misma se trabajaban bajo el enfoque de una muestra grande, este enfoque se remontaba a Laplace (Lehmann, 1992). En los inicios del siglo XX, la forma de llevar a cabo dicha prueba de hipótesis cambió con las aportaciones de Gosset y Fisher, principalmente.

En 1899, Gosset empezó a trabajar como químico cervecero en la cervecería Guinness de Dublín, la empresa buscaba mejorar las materias primas para obtener una mayor calidad de la cerveza, para lo cual se requería determinar qué variedades, qué cultivo y manipulación de la cosecha, qué condiciones de secado y almacenado se los proporcionaría. Todos los trabajos que publicó Gosset lo hizo bajo el seudónimo de ‘Student’, debido a su trabajo en la cervecera.

De acuerdo con Gosset (Student, 1908a), una muestra tiene valor en la medida que permite realizar juicios sobre las constantes estadísticas de la población de la cual se extrajo, por ejemplo, el valor de la media. Se consideraba que cuando se tenían muestras grandes la información sobre el valor de la media era preciso; sin embargo, cuando se trabajaba con muestras pequeñas, Gosset señaló que existían dos fuentes de incertidumbre, la primera se debe al ‘error de muestreo aleatorio’, y es precisamente esta fuente de incertidumbre sobre la cual trabajó en 1908. La segunda refiere a que la muestra no es lo suficientemente grande para determinar cuál es la ley de distribución de los individuos. Gosset asumió la distribución normal y añadió que sus conclusiones no eran estrictamente aplicables a poblaciones que no se distribuyeran como una normal. Aunado a esto, indicó que el método que se utilizaba usualmente para:

> determinar la probabilidad de que la media de la población se encontrara dentro de una distancia dada de la media de la muestra, es asumir una distribución normal sobre la media de la muestra con una desviación estándar igual a $\mathrm{s}/\sqrt{\mathrm{n}}$, donde s es la desviación estándar de la muestra, y usar las tablas de probabilidad. (Student, 1908a, p. 1)

Para lidiar con la dificultad que surge debido a que a medida que el tamaño de la muestra va disminuyendo también va incrementando el error sobre el valor de la desviación estándar de la muestra, en 1908 Gosset propuso dos alternativas. La primera consistía en repetir el experimento muchas veces, entonces la desviación estándar se determina con suficiente exactitud y puede ser utilizada para otros experimentos similares con menores repeticiones. La segunda, radicaba en realizar experimentos por duplicado donde el cuadrado medio de la diferencia entre los pares correspondientes determina la desviación estándar de la población multiplicada por $\sqrt{2}$. Además, Gosset puntualizó aquellos experimentos que no son sencillos de repetir, como es el caso de los experimentos químicos, biológicos y de la agricultura, donde es necesario juzgar la certeza de los resultados con muestras pequeñas.

Gosset buscaba determinar "*el punto en el que podemos utilizar las tablas de probabilidad para juzgar la significancia de la media de una serie de experimentos, y proveer tablas alternativas para cuando el número de experimentos es pequeño*" (Student, 1908a, p. 2). Para lograr lo anterior, buscó una ecuación que representara la frecuencia de las desviaciones estándar de las muestras, teniendo en cuenta que las muestras (con n individuos) provenían de una población normal. Gosset (Student, 1908a) tomó s, que es la desviación estándar de la muestra, y sumó las desviaciones de todas las muestras y las dividió entre el número de ellas obteniendo el valor medio de s^2 expresado de la siguiente forma: $\bar{s}^2 = \frac{n\mu_2}{n} - \frac{n\mu_2}{n^2} = \frac{\mu_2(n-1)}{n}$, donde μ_2 es el coeficiente del segundo momento en la distribución normal original de x.

De acuerdo con Lehmann (1992), Gosset percibía que el numerador de t era normal y a pesar de que, en 1876, Friedrich R. Helmert había publicado la distribución de s^2, Gosset lo desconocía y para obtener la distribución del denominador calculó algebraicamente los primeros cuatro momentos de la distribución de s^2, primero sobre cero, donde $s^2 = 0$ y obtuvo:

$$M_1' = \mu^2 \frac{(n-1)}{n}, M_2' = \mu_2^2 \frac{(n-1)(n+1)}{n^2}, M_3' = \mu_2^3 \frac{(n-1)(n+1)(n+3)}{n}, M_4' = \mu_2^4 \frac{(n-1)(n+1)(n+3)(n+5)}{n}$$

Sin embargo, Gosset no pudo encontrar una forma general. Entonces, procedió a establecer los momentos de s^2 sobre su media y obtuvo:

$$M_2 = 2\mu_2^2 \frac{(n-1)}{n^2}, M_3 = 8\mu_2^3 \frac{(n-1)}{n^3}, M_4 = 12\mu_2^4 \frac{(n-1)(n+3)}{n^4}$$

Al observar una forma simple para conectar los momentos sucesivos descubrió la curva que se ajustaba a sus momentos. Y expresó que la curva de Pearson tipo III (que es la distribución χ^2), $y = Cx^p e^{-\gamma^x}$, podría contemplarse como un ajuste para la distribución de s^2 como $y = Cx^{(n-3)/2} e^{-nx/2\mu_2}$, entonces el área bajo la curva es $C \int_0^\infty x^{(n-3)/2}\, e^{-nx/2\mu_2} = I$. A partir de esta ecuación de la curva encontró que los momentos más grandes de la distribución coincidieron con los de s^2 (Lehmann, 1992). Acerca de esto, Gosset (Student, 1908a, p. 5) expresó que "*es probable que la curva encontrada represente la distribución teórica de* s^2" a pesar de no tener pruebas reales lo asumió para encontrar la distribución de s, considerando que la frecuencia de s es igual a la de s^2. Y obtuvo la expresión $y_2 = 2Cs^{(n-2)} e^{-ns^2/2\mu_2}$, por lo tanto $y = Ax^{(n-2)} e^{-nx^2/2\sigma^2}$ dará la distribución de frecuencias de las desviaciones estándar de muestras de *n* individuos, que provienen de poblaciones que siguen una distribución normal con desviación estándar σ. Calculó la constante A tanto para cuando n es par como impar, obteniendo las siguientes dos expresiones sobre dicha distribución de frecuencias:

$$y = \frac{N}{(n-3)(n-5)\ldots 3.1} \sqrt{\frac{2}{\pi}} \left(\frac{n}{\sigma^2}\right)^{(n-1)/2} x^{n-2} e^{-nx^2/2\sigma^2} \text{ (cuando } n \text{ es par)}$$

$$y = \frac{N}{(n-3)(n-5)\ldots 4.2} \left(\frac{n}{\sigma^2}\right)^{(n-1)/2} x^{n-2} e^{-nx^2/2\sigma^2} \text{ (cuando } n \text{ es impar),}$$

donde *N* representa la frecuencia total.

Después de obtener las ecuaciones que determinan la distribución de frecuencia de la desviación estándar de muestras de una población normal, Gosset procedió a mostrar que no existe correlación entre la media (la distancia entre la media de la muestra y la media de la población) y la desviación estándar de dicha muestra. Si se tiene que $R_{\mu^2 s^2}$ es la correlación entre μ^2 y s^2 entonces $R_{\mu^2 s^2}\sigma_{\mu^2}\sigma_{s^2} + \mu_2^2 \frac{(n-1)}{n^2} = \mu_2^2 \frac{(n-1)}{n^3}\{3 + n - 3\} = \mu_2^2 \frac{(n-1)}{n^2}$, por lo tanto $R_{\mu^2 s^2}\sigma_{\mu^2}\sigma_{s^2} = 0$. Gosset creyó que al probar que no existía correlación entre el numerador y el denominador significaba que eran independientes, y sobre esta base posteriormente calculó la distribución de t (Lehmann, 1992).

Para encontrar la ecuación de la distribución de frecuencias de las medias de muestras que provienen de una población normal, Gosset (Student, 1908a), expresó la media en términos de la desviación estándar de la muestra. Partió de $y = \frac{C}{\sigma^{n-1}} s^{n-2} e^{-ns^2/2\sigma^2}$ como la ecuación de la distribución de s y de $y = \frac{\sqrt{n}N}{\sqrt{2\pi}\sigma} e^{-nx^2/2\sigma^2}$ como la ecuación de la distribución de las medias de las muestras e indicó que "*no existe correlación entre* x*, la distancia de la media de la muestra, y* s*, la desviación estándar de la muestra*" (Student, 1908a, p. 7).

Una vez que estableció la distribución de frecuencias tanto del numerador como del denominador (por separado), se propuso encontrar la distribución de su estadístico $z = x/s$, en otras palabras, la distribución conjunta, suponiendo que x se puede medir en términos de s. Siendo x, la distancia entre la media de la muestra y la media de la población y s, la desviación estándar de la muestra. Entonces, de acuerdo a lo que establece Gosset como estadístico z, éste se puede entender bajo la expresión $z = (\bar{x} - \mu) / \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$. Gosset encontró las siguientes ecuaciones para la distribución de z:

$y = \frac{1}{2} \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \frac{5}{4} \cdot \frac{3}{2} (1 + z^2)^{-n/2}$ (cuando n es impar)

$y = \frac{1}{\pi} \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \frac{4}{3} \cdot \frac{2}{1} (1 + z^2)^{-n/2}$ (cuando n es par)

Gosset (Student, 1908a), puntualizó que estas ecuaciones son independientes de σ y que la distribución a la que refieren esta expresada en términos de la desviación estándar de la muestra para cualquier población normal.

De acuerdo con Zabell (2008), la distribución t la habían trabajado previamente en 1876 Jakob Lüroth, para el caso general de regresión lineal, y en 1883 Francis Y. Edgeworth, para el caso de una muestra, como es el caso que trabajó Gosset.

Gosset (Student, 1908a) también facilitó una tabla que proporcionaba el valor de la probabilidad de que la media se encuentre entre $-\infty$ y z. Tablas de $\frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \begin{pmatrix} \frac{3}{2} \cdot \frac{1}{2} n \ impar \\ \frac{2}{1} \cdot \frac{1}{\pi} n \quad par \end{pmatrix} \int_{-(\pi/2)}^{tan^{-1} z} cos^{n-2}\theta \ d\theta$, para valores de n desde cuatro a diez.

Gosset (Student, 1908a), llevó a cabo un estudio Monte Carlo usando datos reales para mostrar el ajuste de su distribución por medio de la ecuación teórica $y = \frac{2}{\pi}\cos^4\theta\,, z = \tan\theta$ a los datos que reales que tenía en las tarjetas. Dicho ajuste lo realizó con la prueba de bondad de ajuste con el estadístico χ^2, y obtuvo que era muy bueno con valores de $\chi^2 = 7.39$ y $P = 0.92$. De acuerdo con Zabell (2008), la simulación que realizó Gosset fue una de las primeras que existieron y de las pocas que se llevaron a cabo antes de 1927, año en el cual las denominadas "computadoras humanas" realizaron la tabla de números aleatorios.

En su documento Gosset presentó algunos ejemplos para ilustrar la adecuada aplicación de su prueba; hemos seleccionado el siguiente:

Ejemplo 16: Prueba de la media de una muestra (Student, 1908a, pp. 20-21)

> Como ejemplo del tipo de uso que se puede realizar de las tablas, tomo las siguientes figuras de una tabla [*Tabla 4.3*] de A. R. Cushny y A. R. Peebles del Journal of Physiology de 1904, que muestra los diferentes efectos de los isómeros ópticos del hidrobromuro de hiosciamina en la producción del sueño. Se midió el sueño de diez pacientes antes y después del tratamiento (1) con D. hidrobromuro de hiosciamina, (2) con L. hidrobromuro de hiosciamina. El promedio del número de horas de sueño ganado por el uso de los fármacos es tabulado a continuación.
>
> Se llegó a la conclusión de que, en la dosis habitual, 2 era de valor como somnífero, pero 1 no.

Tabla 4.3. *Horas adicionales de sueño ganadas por el uso de hidrobromuro de hiosciamina* (fuente: Student, 1908a, p. 20)

Paciente	**1 (Dextro-)**	**2 (Laevo-)**	**Diferencia $(2-1)$**
1	+0.7	+1.9	+1.2
2	-1.6	+0.8	+2.4
3	-0.2	+1.1	+1.3
4	-1.2	+0.1	+1.3
5	-0.1	-0.1	0
6	+3.4	+4.4	+1.0
7	+3.7	+5.5	+1.8

8	+0.8	+1.6	+.8
9	0	+4.6	+4.6
10	+2.0	+3.4	+1.4
MEDIA	+0.75	+2.33	+1.58
S. D.	1.70	1.90	1.17

Primero, veamos cuál es la probabilidad de que 1, en promedio, incrementara el tiempo de sueño; es decir, cuál es la probabilidad de que la media de la población, de la cual estos experimentos son una muestra, sea positiva. $\frac{+0.75}{1.70} = 0.44$ y buscando la tabla z$=$ 0.44 para diez experimentos encontramos por interpolación entre 0.8873 a 0.113 que la media es positiva.

Eso es aproximadamente 8 a 1 y correspondería en la curva normal a aproximadamente 1.8 veces el error probable. Es muy probable que 1 produzca un aumento del sueño, pero no sorprendería si los resultados fueran revertidos por otros experimentos.

Si ahora consideramos la probabilidad de que 2 sea realmente un somnífero, tenemos el aumento medio del sueño $= \frac{2.33}{1.90}$ o 1.23 veces la desviación estándar. Desde la tabla de probabilidad correspondiente a esto es $.9974$, es decir, las probabilidades son cerca de 400 a 1 de que tal sea el caso. Esto corresponde a cerca de 4.15 veces el error probable en la curva normal. Pero supongo que el punto real de los autores es que 2 es mejor que 1. Esto debemos probarlo haciendo una nueva serie, restando 1 de 2. La media de esta serie es $+1.58$ mientras que la S. D. es 1.17, la media sería $+1.35$ veces la S. D. De la tabla de probabilidad se obtiene 0.99859 o que las probabilidades son alrededor de 666 a 1 de que 2 es mejor somnífero. El bajo valor de la S. D. probablemente se deba a que los diferentes fármacos reaccionan de manera similar en el mismo paciente, de modo que existe una correlación entre los resultados.

Por supuesto, las probabilidades de este tipo hacen que sea casi seguro que 2 sea el mejor somnífero, y en la vida práctica una probabilidad tan alta se considera en la mayoría de los casos como una certeza. (Student, 1908a, pp. 20-21)

De acuerdo con Zabell (2008), en 1934 el Dr. Isidor Greenwald de la New York University y del Bellevue Hospital Medical College le escribió una carta a Fisher, ya que en 1925 este último había retomado este ejemplo de Gosset, señalando que los dos fármacos correspondientes a los datos del estudio de Cushny y Peebles eran de hidrobromato de L-hiosciamina e hidrobromato de L-hioscina y no isómeros ópticos. Además del error señalado anteriormente, Stigler (2017) identificó otros como la cita incorrecta de la fuente, que los datos identificados como pacientes en realidad eran medias de muestras de diferentes tamaños y por ende de diferentes varianzas.

Sin embargo, lo que es interesante rescatar de la prueba (t-Student) de Gosset y que se puso de manifiesto en este ejemplo, es que la comparación de la media muestral se realizó con la desviación estándar de la muestra, sin atender a algún agente externo. En otras palabras, no hubo intervención de la desviación estándar de la población, ya que como señala Stigler (2017), hacer referencia a la desviación estándar de la población era lo generalmente aceptado en las investigaciones de este tipo.

Gosset señala que, si se han hecho dos observaciones y no tenemos otra información, es una posibilidad equitativa de que la media de la población (normal) se encuentre entre ellas (Student, 1908a, p. 13) esto puede considerarse un antecedente de los intervalos de confianza, para Fisher fue un ejemplo temprano de la probabilidad fiducial.

En 1912, Fisher envía una carta a Gosset sobre su método del error probabilístico de la media, que posteriormente se le conocería como la prueba t-Student, pues consideraba que no había una rigurosa prueba matemática que sustentara el trabajo de Gosset y le sugería plantear el problema desde la geometría multidimensional arrojaría una prueba sencilla y rigurosa (Stigler, 2017). En 1913, Mr. H. E. Soper publicó un artículo sobre el problema de la distribución de frecuencias del coeficiente de correlación, y llamó la atención de Fisher que se podría atacar este problema con ideas geométricas, lo cual iba en correspondencia con lo que un año antes le había planteado a Gosset.

En 1915, Fisher publicó sobre la distribución del coeficiente de correlación r, retomando el trabajo de 1908 de Gosset, que ya hemos dilucidado. Al respecto del trabajo de 1908, comentó que, desde su perspectiva, Gosset deseaba principalmente "*obtener una estimación justa de la precisión que se atribuye a la media de una muestra pequeña*" (Fisher, 1915, p. 507) y que si $x_1, x_2, \dots, x_n$ son los miembros de una muestra, entonces,

$$n\bar{x} = x_1 + x_2 + \cdots + x_n,$$

$$\text{Y } n\mu^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2$$

Por lo que la frecuencia con la que el error cuadrático medio se encuentra en el rango $d\mu$, es proporcional a

$$\mu^{n-2}e^{-\frac{n\mu^2}{2\sigma^2}}d\mu$$

Sobre este resultado, Fisher señaló que "*aunque se llegó por métodos empíricos, se estableció casi fuera de toda duda razonable en el primero de los documentos de 'Student'. Sin embargo, es interesante notar que la forma se establece instantáneamente, cuando la distribución de la muestra se ve geométricamente*" (1915, p. 507).

Además, también hace referencia al segundo artículo de Gosset, también del año 1908, 'Probable Error of a Correlation Coefficient', donde señala que el problema de mayor dificultad que abordó refiere a la distribución de frecuencias del coeficiente de correlación. Gosset trabajó con muestras de dos, donde la distribución de frecuencias sólo podía tomar los valores -1 y $+1$, ya que estos puntos siempre deben estar en la línea de regresión que los une. Gosset partió de trabajo empírico y argumentó que es posible calcular a priori la distribución de frecuencias cuando el tamaño de la muestra es dos y señaló que:

> He tratado con muestras de dos con cierta longitud, porque es posible que este valor limitante de la distribución, con su media de $\frac{2}{\pi}\sin^{-1}\rho$ y su coeficiente del segundo momento de 1-$\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2$, puede proporcionar una pista de la distribución cuando n es más grande que 2. (Student, 1908b, p. 304)

Donde ρ es la correlación de la población. Además de esta serie Gosset trabajo con otra más corta donde trabajó con muestras de 30, 4 y 8, proporcionó una tabla con la distribución de r para dichas muestras y las comparó con la ecuación $y = y_0\left(1-\frac{x^2}{.9802}\right)^{2.021}$ que calculó de los momentos.

Respecto a este trabajo de Gosset, Fisher mencionó que el problema de la distribución de frecuencias del coeficiente de correlación r, que proviene de una muestra de n pares y son tomados de una población infinita, se puede resolver con concepciones de la geometría multidimensional, siempre y cuando la población puede representarse por una superficie normal. El objetivo del documento de Fisher de 1915 fue demostrar la forma general y para algunos casos importantes se obtendrán los momentos, para hacerlo parte de la forma que especifica a la distribución de frecuencias de la población

$df = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{1-\rho^2}\left\{\frac{(x-m_1)^2}{2\sigma_1^2} - \frac{2(x-m_1)(y-m_2)}{2\sigma_1\sigma_2} + \frac{(y-m_2)^2}{2\sigma_2^2}\right\}} dx\,dy$, donde df es la probabilidad de que cualquier observación se encuentre dentro del rango $dx\,dy$, entonces la probabilidad de que n pares se encuentren dentro de sus elementos especificados es:

$$\frac{1}{\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^n} e^{-\frac{1}{1-\rho^2}\sum_1^n\left\{\frac{(x-m_1)^2}{2\sigma_1^2} - \frac{2(x-m_1)(y-m_2)}{2\sigma_1\sigma_2} + \frac{(y-m_2)^2}{2\sigma_2^2}\right\}} dx_1\,dy_1 \ldots dx_n\,dy_n$$

Fisher interpretó esta última como una distribución de densidad simple en $2n$ dimensiones. Posteriormente presentó las ecuaciones $n\bar{x} = \sum_1^n(x)$, $n\bar{y} = \sum_1^n(y)$, $n\mu_1^2 = \sum_1^n(x-\bar{x})^2$, $n\mu_2^2 = \sum_1^n(y-\bar{y})^2$y $nr\mu_1\mu_2 = \sum_1^n(x-\bar{x})(y-\bar{y})$ para x y y.

Fisher (1915), consideró el espacio de n dimensiones donde se representan las variaciones de x, la media y el error cuadrático medio de n observaciones están determinadas por las relaciones de P. La longitud de la perpendicular PM es $\mu_1\sqrt{n}$ y la expresó con la Figura 4.3.

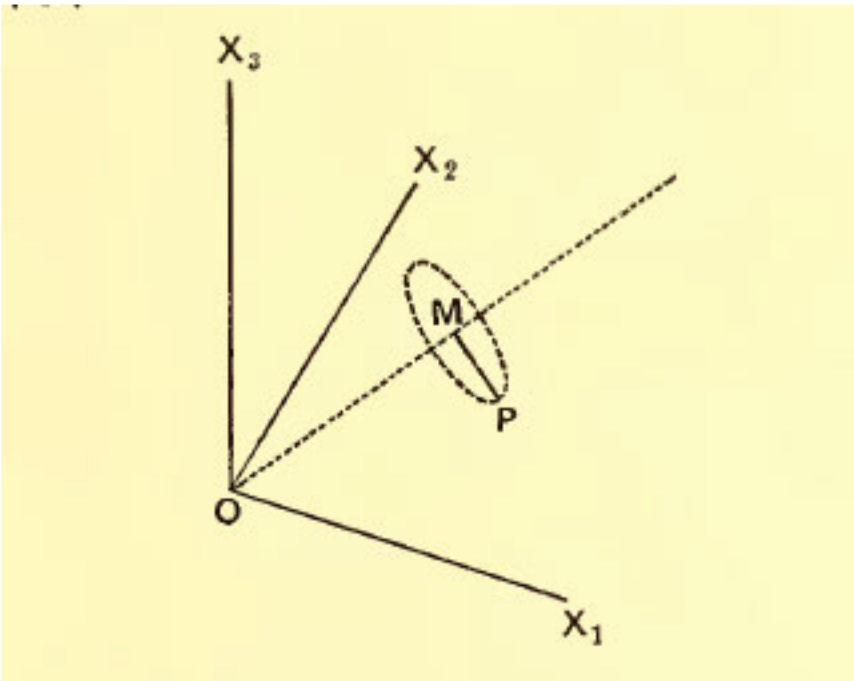


Figura 4.3. Longitud PM en n dimensiones (fuente: Fisher, 1915, p. 509)

Fisher también especificó a la expresión $C\mu_1^{n-2}d\mu_1 d\bar{x}$ como el elemento de volumen en este espacio de n dimensiones; posteriormente procedió a simplificar la fórmula de la probabilidad expresada anteriormente y descartó en cada etapa cualquier factor que no implicara a r y obtuvo la siguiente expresión:

$$\frac{1}{\sigma_1\sigma_2(1-\rho^2)}\int_0^\infty\int_0^\infty e^{-\frac{1}{(1-\rho^2)}\left(\frac{\mu_1^2}{\sigma_1^2}-\frac{2\rho r\mu_1\mu_2}{\sigma_1\sigma_2}+\frac{\mu_2^2}{\sigma_2^2}\right)}d\mu_1 d\mu_2=\frac{\theta}{\sin\theta}$$

Fisher añadió, que esta expresión se podía aplicar a los pequeños valores pares de n y que era adecuada para el calculo de los momentos. Además, encontró la hiperbólica cuando $n = 2$ en las cercanías de, concordando con lo que había encontrado Gosset en 1908, también encontró la curva cuando $n = 4$ y calculó el número de valores que se encuentran dentro de cualquier rango y los tabuló con la integral de $\frac{1}{\sin\theta}(1-\theta\cot\theta)$. Después, señaló que era necesario encontrar expresiones generales para las integrales en términos de n y p, para lo cual introdujo la cantidad $\emptyset$ y encontró la forma general $\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}dr = n-3\pi tan^{n-1}\alpha$.

Sobre el uso del coeficiente de correlación r como variable independiente de las curvas de frecuencia, que tenía hasta ese momento, Fisher encontró que en algunos puntos se distorsionaba incluso de forma extrema; por esto buscó si las transformaciones le proporcionarían variables más naturales para el tratar sus fórmulas y las expresó como:

$$t=\tan\beta=\frac{r}{\sqrt{1-r^2}}$$

y

$$\tau=\tan\alpha=\frac{\rho}{\sqrt{1-\rho^2}}$$

Con $\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$ como la expresión de la frecuencia de la curva; entonces con esta expresión el rango de la curva se extendió desde $-\infty$ a ∞. Fisher (1915), destacó que cuando $r = 0$ la frecuencia de la curva es $\frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$. Estas expresiones coinciden con las que Gosset en 1908 cálculo su tabla de probabilidad para z. Posteriormente, obtuvo los cuatro momentos para las curvas de frecuencia, los cuales se presentan a continuación:

Primer momento $\bar{t}=\frac{n-2}{n-3}\frac{\rho}{\sqrt{1-\rho^2}}=\frac{n-2}{n-3}\tau$.

Segundo momento $\bar{t}^2 + \sigma^2 = \frac{1}{n-4}\{1 + (n-1)\tau^2\}$ y $\sigma^2 = \frac{1}{n-4}\left\{1 + \tau^2 + \frac{(n-2)}{(n-3)^2}\tau^2\right\}$.

Tercer momento $\sqrt{\beta_1}\sigma^3 = \frac{(n-2)\tau}{(n-3)(n-4)(n-5)}\left\{3(1+\tau^2) + \frac{2\tau^2(n-1)}{(n-3)^2}\right\}$.

Cuarto momento $\beta_2\sigma^4 = \frac{3}{(n-4)(n-6)}\left\{(1+\tau^2)^2 + \frac{6(n-2)\tau^2}{(n-3)(n-5)}(1+\tau^2) + \frac{6(n-2)(3n^2-11n+12)\tau^4}{(n-3)^4(n-5)}\right\}$.

Sobre esto Fisher señaló que,

> para altos valores de n, todos los términos tienden a desaparecer menos el primero; β_1 tiende a variar como ρ^2, y β_2 tiende a ser independiente de ρ. En efecto para valores altos de t, donde ρ^2 es muy cercano a la unidad la forma de la curva es casi constante, pero la asimetría medida por β_1 disminuye a cero en el origen, y cambia su sentido, cuando τ y ρ cambian de signo. (Fisher, 1915, p. 519)

Fisher incluyó tablas donde se podía observar los cambios en la forma de las curvas. Finalmente realizó una aproximación de r y la relación que existe entre t y τ, siendo $r = \rho\left(1 + \frac{1+r^2}{2n}\right)$ y $t = \tau\left(1 + \frac{1}{2n}\right)$.

El método geométrico de Fisher (1915), obtuvo la distribución conjunta del numerador y el denominador de z (de Gosset), pero en términos t. Fisher trabajó con la variable t por las bondades para el cálculo de los momentos y el sesgo de la curva de frecuencias. Además, utilizó el mismo enfoque para derivar la distribución del coeficiente de correlación (como el error estándar de la media). Este trabajó fue un complemento al de Gosset de 1908. El editor del artículo de 1915 de Fisher, hizo notar que la t que había encontrado Fisher era la raíz cuadrada de la contingencia cuadrada media $\emptyset$ de Pearson.

4.1.3 La generalización del estadístico $t - Student$

En 1918, Fisher volvió a escribir a Gosset para analizar la aplicación del método de correlación de diferencias de Gosset con datos de natalidad y matrimonio, con esta carta se inició una larga serie de correspondencia entre Gosset y Fisher donde expusieron sus problemas ‘prácticos’ el uno al otro y contestando cómo podrían resolverlos. Esta colaboración llevó a la formalización y expansión de las aplicaciones del estadístico t-Student y su distribución (Fisher, 1981).

A raíz de esta correspondencia, Fisher (1922b) propuso un test de los coeficientes de regresión y su distribución basados en z y cuyas bases están sentadas en Student (1908a), para dar respuesta a una de las cartas de Gosset, donde le expresaba estar interesado en conocer la distribución de frecuencias de $r(\sigma_x/\sigma_y)$ para pequeñas muestras; y en otra de sus cartas añade que desea conocer el error probable de los coeficientes de regresión y correlación para pequeñas muestras (Eisenhart, 1979).

Según Fisher (1922b), a pesar de que no se tenían métodos para probar la bondad de ajuste de líneas de regresión para datos extendidos, si era capaz de proporcionar una solución exacta para la distribución de los coeficientes de regresión y que en Student (1908a), se sentaban las bases exactas para su formulación. Fisher partió de una formula simple de regresión lineal $Y = a + b(x - \bar{x})$, asumiendo que a cada valor de x corresponde uno de y, y señaló que los coeficientes pueden ser calculados por las ecuaciones $a = \bar{y}$, $b = \frac{S(y(x-\bar{x}))}{S(x-\bar{x})^2}$, notando que a y b son funciones ortogonales. Fisher retomó cómo se podía calcular el error probable de a, del trabajo de Gosset (Student, 1908a), y entonces, indicó que si a se distribuye normalmente se tiene una varianza ${\sigma_a}^2 = \frac{\sigma^2}{n}$. Puntualizó que "*Si α es el valor poblacional de a, y $\tau = \frac{a-\alpha}{\sigma}\sqrt{n}$, entonces τ se distribuye normalmente alrededor de cero con una desviación estándar unitaria*" (Fisher, 1922b, 608). El continuó expresando que en caso de desconocer la varianza la mejor estimación podía provenir de la muestra con $s^2 = \frac{1}{n-2}S(y - Y)^2$. Entonces indicó que, si $\chi^2 = (n-2)\frac{s^2}{\sigma^2}$, la distribución de s^2 es

$$df = \frac{1}{\frac{n-4}{2}}\left(\frac{\chi^2}{2}\right)^{\frac{n-4}{2}} e^{-\frac{1}{2}\chi^2}\, d\left(\frac{\chi^2}{2}\right)$$

Por lo que s^2 se distribuye como una χ^2 con $n-2$ grados de libertad. La noción de grados de libertad fue introducida por Fisher (1922a), inicialmente para la distribuciónχ^2 y ha sido descrita y analizada en la sección 3.2. Fisher (1922b), concluyó que s y a son completamente independientes y que siguiendo a Gosset "*nosotros podemos encontrar la distribución de una cantidad completamente calculable de la muestra*" (Fisher, 1922b, 608):

$$z = \frac{\tau}{\chi} = \frac{(a-\alpha)\sqrt{n}}{\sqrt{S(y-Y)^2}}$$

Para el df integrado con respecto a χ^2de 0 a ∞, obtuvo:

$$\frac{1}{\sqrt{\pi}} \frac{\frac{n-3}{2}!}{\frac{n-4}{2}!} \frac{dz}{(1+z^2)^{\frac{n-1}{2}}}$$

Fisher indicó que esto concuerda con la curva tipo VII encontrada por Gosset, con una n reducida a la unidad, debido al ajuste para una regresión lineal de primer grado. De forma similar lo hizo para b, considerando que se encuentra normalmente distribuida sobre β y con varianza ${\sigma_b}^2 = \frac{\sigma^2}{S(x-\bar{x})^2}$, y obtuvo:

$$z = \frac{(b-\beta)\sqrt{S(x-\bar{x})^2}}{\sqrt{S(y-Y)^2}}$$

De acuerdo con Eisenhart (1979), Fisher llegó a la misma distribución z de Gosset y la cual había señalado previamente para a, pero ahora se ha reemplazado n por $n-1$.

Fisher (1922b), generalizó su enfoque para líneas de regresión de cualquier tipo y con cualquier número de coeficientes, con la ecuación de regresión:

$$Y = a + bX_1 + cX_2 + \cdots + kX_p$$

donde $X_1, X_2, \ldots X_p$ son funciones ortogonales de x, para los valores observados, por lo que:

$$z = \frac{(k-\kappa)\sqrt{S(X)^2}}{\sqrt{S(y-Y)^2}}$$

Fisher señaló que z se distribuye de cuerdo con la curva tipo VII y se puede encontrar la probabilidad auxiliándose de las tablas de probabilidad (Student, 1917) para valores de $n-p$ de 0 a 30; y que "*estas tablas son adecuadas para probar la significancia de un coeficiente de regresión observado*" (Fisher, 1922b, 610).

Esta aplicación dio paso a la generalización de la prueba; de acuerdo con Fisher (1922b), así como se estimaba a partir de la media de una muestra se podía incrementar o ampliar, en términos de que esta misma distribución era útil para las diferencias de las medias. E indicó que "*si $\bar{x}$ y $\bar{x}'$ son las medias de las muestras de* n *y* n′, *y deseamos probar si las medias están* *suficientemente de acuerdo para garantizar la creencia de que las muestras provienen de la misma población, podemos calcular*" (Fisher, 1922b, p. 610).

$$z = \frac{\bar{x} - \bar{x}'}{\sqrt{S(x-\bar{x})^2 + S'(x'-\bar{x}')^2}} \sqrt{\frac{nn'}{n+n'}}$$

Y z se distribuye de acuerdo con

$$df = \frac{1}{\sqrt{\pi}} \frac{\frac{n+n'-3}{2}!}{\frac{n+n'-4}{2}!} \frac{dz}{(1+z^2)^{\frac{n+n'-1}{2}}}$$

Finalizó indicando que "*este método de comparación puede aplicar directamente a los coeficientes de regresión, cuando se observa la misma serie de valores de x en cada caso*" (Fisher, 1922b, p. 611).

En 1923 y 1924, Fisher propagó estas pruebas y en 1925 dio una versión unificada de las pruebas en 'Statistical Methods for Research Workers', parece ser que en este último trabajo fue la primera vez que aparece el estadístico t en vez de z, como $t = z\sqrt{n'-1}$ o $t = z\sqrt{n}$, donde n indica los grados de libertad, uno menos que el tamaño de la muestra. Utilizar n' para indicar el tamaño de la muestra generó después cierta confusión ya que era usual utilizar n para indicar el tamaño de la muestra. La ecuación $t = z\sqrt{n}$, es la que Gosset adoptaría posteriormente para el cálculo de las nuevas tablas de la distribución t (Fisher, 1925a, pp. 105-106).

En Fisher (1925b), presenta también algunas aplicaciones de la distribución t-Student, por ejemplo: (1) significancia de la diferencia entre medias con $t = \frac{(\bar{x}_1-\bar{x}_2)\sqrt{n_1+n_2-2}}{\sqrt{S_1+S_2}} \sqrt{\frac{n_1 n_2}{n_1+n_2}}$, donde t se distribuye de acuerdo con la distribución t-Student para $n = n_1 + n_2 - 2$; (2) Significancia de los coeficientes de regresión con $t = \frac{(b-\beta)\sqrt{n'-2}\,\sqrt{S(x-\bar{x})^2}}{\sqrt{S(y-Y)^2}}$ y se distribuye de acuerdo con la distribución t-Student para $n = n' - 2$; (3) Coeficientes de regresión no lineal con $t = \frac{(a_i-\alpha_i)\sqrt{n'-p-1}\,\sqrt{S(X^2{}_i)}}{\sqrt{S(y-Y)^2}}$ con distribución t-Student para $n = n' - p - 1$; y (4) Coeficientes de regresión múltiple con $t_i = \frac{(b_i-\beta_i)\sqrt{n'-p-1}\,\sqrt{\Delta}}{\sqrt{S(y-Y)^2}\,\sqrt{\Delta_{ii}}}$, donde β_i es el valor de la población correspondiente al valor observado b_i, con t_i distribuyéndose como una distribución t-Student para $n = n' - p - 1$.

Hemos retomado algunos problemas que presentó Fisher (1934, 1925b) para ejemplificar las pruebas.

Ejemplo 17. Significancia de la media de una muestra pequeña (Fisher, 1934, pp. 118-120)

> Las siguientes cifras (Cushny and Peebles' data), que cito del trabajo de "Student", muestra el resultado de un experimento con diez pacientes sobre el efecto de los isómeros ópticos del hidrobromuro de hiosciamina en la producción de sueño [*Tabla 4.4*].

Tabla 4.4. *Horas adicionales de sueño ganadas por el uso de hidrobromuro de hoscina* (fuente: Fisher, 1934, p. 119)

Paciente	**1 (Dextro -)**	**2 (Laevo -)**	**Diferencia (2 – 1)**
1	$+\,0.7$	$+\,1.9$	$+\,1.2$
2	$-\,1.6$	$+\,1.8$	$+\,2.4$
3	$-\,0.2$	$+\,1.1$	$+\,1.3$
4	$-\,1.2$	$+\,0.1$	$+\,1.3$
5	$-\,0.1$	$-\,0.1$	0.0
6	$+\,3.4$	$+\,4.4$	$+\,1.0$
7	$+\,3.7$	$+\,5.5$	$+\,1.8$
8	$+\,0.8$	$+\,1.6$	$+\,0.8$
9	0.0	$+\,4.6$	$+\,4.6$
10	$+\,2.0$	$+\,3.4$	$+\,1.4$
Media ($\bar{x}$)	$+\,0.75$	$+\,2.33$	$+\,1.58$

> La última columna ofrece una comparación controlada de la eficacia de los dos fármacos como somníferos, los mismos pacientes fueron utilizados para evaluar cada uno de los fármacos; de la serie de diferencias encontramos

$$\bar{x} = +1.58,$$

$$\frac{S^2}{10} = 0.1513,$$

$$S/\sqrt{10} = 0.3890,$$

$$t = 4.06$$

Para $n = 9$, se tiene la probabilidad de que solo un valor en cien excederá 3.250, por lo que la diferencia entre los resultados son claramente significativos. Por los métodos de los capítulos anteriores deberíamos, en este caso, haber sido guiados a la misma conclusión con casi igual certeza; porque si los dos medicamentos hubieran sido igualmente efectivos, aparecerían signos positivos y negativos en la última columna con igual frecuencia. Sin embargo, de los 9 valores distintos de cero, todos son positivos y parecen de la distribución binomial, $\left(\frac{1}{2} + \frac{1}{2}\right)^9$, todos serán de la misma forma, con una probabilidad de, sólo dos veces en 512 experimentos.

Para Fisher (1934), si $x_1, x_2, \ldots, x_{n'}$ es una muestra de n' valores de la variable x, podemos aplicar la prueba para valorar si la media de x difiere significativamente de cero por medio del estadistico t. Si:

$$\bar{x} = \frac{1}{n'} S(x)$$

$$\frac{s^2}{n'} = \frac{1}{n'(n'-1)} S(x-\bar{x})^2$$

$$t = \bar{x} \Big/ \sqrt{\frac{s^2}{n'}}$$

$$n = n' - 1$$

También proporcionó las tablas de probabilidad para la distribución t-Student para cada valor de n, en la que se pueden encontrar los valores del estadístico t para la probabilidad P de encontrarse fuera del rango $\pm t$.

Fisher (1925b, 1934), reconoce que en el trabajo experimental cada vez eran más necesarias las pruebas sobre si existe una diferencia significativa entre las medias de muestras diferentes o si es posible considerar que estas provienen de la misma población, y señaló que este tipo de problemas puede ser abordado con la comparación de los valores de las dos medias. Fisher (1925b) proporcionó una aplicación del estadístico t-Student y su distribución sobre la significancia de medias de dos muestras pequeñas de diferente tamaño, el cual presentamos a continuación.

Ejemplo 18. Significancia de la diferencia de medias de dos muestras de diferente tamaño (Fisher 1925b, p. 95)

> Como un ejemplo de la aplicación de este método en el trabajo experimental, nosotros podemos tomar una parte de los datos de un experimento de electro-cultivo realizado en 1922 en Rothamsted [*Tabla 4.5*]. Ocho macetas con tres plantas de cebada cada una, fueron expuestas a una descarga eléctrica de alta tensión, mientras que nueve macetas similares se encerraron en una jaula de alambre con conexión a tierra. El número de macollos en cada maceta fue el siguiente:

Tabla 4.5. *Número de macollos por maceta* (fuente: Fisher, 1925b, p. 95)

Electrificada	16,	16,	20,	16,	20,	17,	15,	21		Media 17.625= $\overline{x}_1$
Enjaulada	17,	27,	18,	25,	27,	29,	27,	23,	17	Media 23.333= $\overline{x}_2$

> Por lo tanto, la diferencia entre las medias es 5.708; también
>
> $$s_1 = 37.875, \qquad s_2 = 184, \qquad s_1 + s_2 = 221.875$$
>
> Multiplicando $(s_1 + s_2)$ por 17 y dividiéndolo por 15, 8 y 9 sucesivamente, tenemos como la varianza estimada de la diferencia entre las medias, 3.4925; y para la desviación estándar estimada, 1.8688. El valor de t, que es la razón de la diferencia de su desviación estándar estimada es, 3.054. Para $n = 15$, la Tabla 1 muestra que este valor excede la posibilidad de 41 veces en $10{,}000$. En otras palabras, una diferencia, positiva o negativa, mayor que la observada tiene la probabilidad de ocurrir únicamente alrededor de 8 veces en $1{,}000$ experimentos. Por lo tanto, la diferencia debe ser juzgada entonces como significativa. Las dos series son definitivamente diferentes en cuanto a macollaje; sin embargo, la posibilidad de que la diferencia en la variabilidad, así como la diferencia en la media, hayan contribuido al resultado no está excluida por esta prueba.

Además, Fisher (1925b) explicó que la posibilidad de obtener valores significativos de t, cuando las muestras tienen una media similar y diferencia en la variabilidad, es apenas importante en relación con el uso práctico de la prueba. De acuerdo con Fisher, para los casos en que se requiere aplicar la prueba para muestras con el mismo tamaño y en que cada a individuo de una muestra le corresponde un individuo particular de la otra muestra se puede aplicar el método anteriormente descrito o el método de la prueba para una muestra como lo vimos en el ejemplo 16 y 17, y lo retomaremos en el ejemplo 20. Aunque señala que cuando se quiere observar la diferencia de las medias, a veces, un método o el otro puede ser más sensible. Sobre aplicar método de la prueba para una muestra en la diferencia señaló que

> como comúnmente pasa en el trabajo experimental los valores correspondientes de las dos muestras se encuentran positivamente correlacionados, la desviación estándar de las diferencias se reducirá por esta circunstancia; contra esta ventaja, debemos resaltar el hecho de que, al tratar los resultados como una sola muestra, el valor de n es solo la mitad de grande que si las dos muestras se hubieran tratado por separado. (Fisher, 1925b, p. 47)

Fisher (1934), estableció que si $x_1, x_2, \ldots, x_{n_{1+1}}$ y $x'_1, x'_2, \ldots, x'_{n_{2+1}}$son dos muestras, se puede calcular si la diferencia entre las medias es significativa por medio de:

$$\bar{x} = \frac{1}{n_1+1} S(x), \qquad \bar{x}' = \frac{1}{n_2+1} S(x')$$

$$s^2 = \frac{1}{n_1+n_2}\{S(x-\bar{x})^2 + S(x'-\bar{x}')^2\}$$

$$t = \frac{\bar{x}-\bar{x}'}{s}\sqrt{\frac{(n_1+1)(n_2+1)}{n_1+n_2+2}}$$

$$n = n_1 + n_2$$

A diferencia del estadístico t-Student para muestras con diferente tamaño que presentó Fisher (1925b), donde para estimar la desviación estándar se utiliza $s = \sqrt{\frac{\frac{(S_1+S_2)(n_1+n_2)}{n_1+n_2-2}/n_1}{n_2}}$, en Fisher (1934) para el estadístico t-Student cuando se trabaja con dos muestras del mismo tamaño se utiliza una estimación por medio de la desviación estándar agrupada (pooled S.E.) $s = \sqrt{\frac{1}{n_1+n_2}\{S(x-\bar{x})^2 + S(x'-\bar{x}')^2\}}$. Sobre la suposición de que las dos varianzas deben ser iguales para la aplicación de este método Fisher (1934, p. 122) indicó que:

> Esta es una declaración incorrecta; la igualdad de las varianzas es una parte necesaria de la hipótesis a probar, esto es, que las dos muestras provienen de la misma población normal. La validez de la prueba t, como prueba de esta hipótesis, es, por lo tanto, absoluta, y no requiere suposición alguna. Sería, por supuesto, legítimo hacer una prueba de significancia, diferente que sea apropiada para la pregunta, ¿podrían estas muestras haber sido extraídas de diferentes poblaciones normales que tienen la misma media? Este problema, de hecho, se ha resuelto, pero en relación con situaciones reales que surgen en la investigación, la respuesta a esta pregunta parece ser completamente académica.

Respecto a la aplicación de la prueba t-Student cuando se tienen varianzas poblacionales diferentes, Bernard Lewis Welch desarrolló trabajos en la década de los 30's y 40's, los cuales analizaremos más adelante.

Ejemplo 19: Significancia de diferencia de medias de pequeñas muestras (Fisher, 1934, pp. 122-123):

> Supongamos que las figuras de la tabla 27 [*Tabla 4.4*] se obtuvo con diferentes pacientes para los dos fármacos; el experimento hubiera sido menos controlado, y deberíamos esperar obtener resultados menos seguros del mismo número de observaciones, porque es a priori probable, y las cifras anteriores sugieren, que las variaciones personales en respuesta a los fármacos estarán correlacionadas en cierta medida.
>
> Tomando, entonces, las cifras para representar dos grupos diferentes de pacientes, tenemos

$$\bar{x} - \bar{x}' = +1.58$$

$$s^2\left(\frac{1}{10} + \frac{1}{10}\right) = 0.7210$$

$$t = +1.861$$

$$n = 18$$

El valor de P es, por lo tanto, entre 0.1 y 0.05, y no puede considerarse como significativo Este ejemplo muestra claramente el valor del diseño en experimentos a pequeña escala, y que la eficacia de tal diseño es capaz de medición estadística.

Ejemplo 20: Significancia de diferencia de medias de pequeñas muestras sobre el cambio del número de bacterias (Fisher, 1934, pp. 124-125):

La siguiente tabla [*Tabla 4.6*] muestra la media de colonias bacterianas por placa obtenida por cuatro métodos ligeramente diferentes de muestras de suelo tomado a las 4 p.m. y 8 p.m. respectivamente (datos de H. G. Thornton):

Tabla 4.6. *Cambio en el número de bacterias* (fuente: Fisher, 1934, p. 125)

Método	**4 p.m.**	**8 p.m.**	**Diferencia**
A	29.75	39.20	+9.45
B	27.50	40.60	+13.10
C	30.25	36.20	+5.95
D	27.80	42.20	+14.60
Media	28.825	39.60	+10.775

Para las series de diferencias tenemos $\bar{x} = +10.775, \frac{1}{4}s^2 = 3.756, t = 5.560, n = 3,$ por lo que la tabla muestra que P esta entre 0.01 y 0.02. Si, por el contrario, nosotros usamos el método del ejemplo [18], y tratamos a las dos series por separado, encontramos $\bar{x} - \bar{x}' = +10.775, \frac{1}{2}s^2 = 2.188, t = 7.285, n = 6$; esto no es solo un valor más grande de n sino un valor más grande de t, que ahora está mucho más allá del rango de la tabla, que muestra que P es extremadamente pequeño. En este caso de los efectos diferenciales de los diferentes métodos son insignificantes o han actuado de manera bastante diferente en las dos series, por lo que se perdió precisión al comparar cada valor con su contraparte en la otra serie.

En casos como este, a veces ocurre que un método no muestra diferencia significativa, mientras que el otro lo trae afuera; si cualquiera de los métodos indica una diferencia definitivamente significativa, su testimonio no puede ser ignorado, incluso si el otro método no muestra el efecto.

A lo que Fisher denominó significancia de diferencia de medias, por una parte, para el caso específico de la serie 'Diferencia' (ver ejemplo en la cuarta columna de la Tabla 4.4), donde obtuvo que " $\bar{x} = +10.775, \frac{1}{4}s^2 = 3.756, t = 5.560, n = 3,$ *por lo que la tabla muestra que P esta entre 0.01 y 0.02*" (1934, p. 125), es lo que hoy en día se conoce como la prueba t-Student para muestras dependientes. Por otra parte, al tratar las dos series por separado Fisher aplicó lo que hoy se conoce como el método de la prueba t-Student para dos muestras independientes, sin embargo, las muestras en el ejemplo anterior son dependientes. Para ese momento no se tenía esta diferencia al aplicar la prueba de significancia de medias de dos muestras pequeñas.

En el mismo sentido de lo descrito previamente sobre la prueba de los coeficientes de regresión y su distribución que propuso Fisher(1922b), en Fisher (1934) indicó que el método de la prueba de significancia de medias de dos muestras no era únicamente aplicable a los valores de las medias sino también a los coeficientes de regresión. También brindó un ejemplo cualitativo para familiarizarnos con el concepto de regresión.

De acuerdo con Fisher (1934), es común que la estatura de un niño dependa de su edad, sin embargo, si conocemos su edad no podemos precisar su estatura debido a que, para cada edad podemos tener una distribución de frecuencias de las estaturas características. Mediante una función de regresión (representada gráficamente por la línea de regresión) de la estatura sobre la edad podemos obtener la media de la estatura para cualquier edad. Para la línea de regresión debemos considerar que la edad es la variable independiente y la estatura es la variable dependiente. Si ocurren errores en las estaturas estas no influenciaran a la línea de regresión, mientras que si éstos ocurren en las edades se alteraría la forma de la línea de regresión. También indicó que

> cuando la función de regresión es lineal, la especificación de la regresión se simplifica mucho, ya que además de las medias generales sólo tenemos que indicar la razón con la que incrementa la media de la variable dependiente que lleva al correspondiente incremento de la variable independiente. Tales proporciones se denominan coeficientes de regresión. (Fisher, 1934, p. 128)

Para la función de regresión $Y = a + b(x - \bar{x})$, b es el coeficiente de regresión de y sobre x.

Ejemplo 21. Significancia de los coeficientes de regresión por medio del efecto de los fertilizantes nitrogenados en la conservación del rendimiento (Fisher, 1934, pp. 133-137).

> Los rendimientos (en fanegas) de granos preparados por acre que se muestran en la [*Tabla 4.7*] se obtuvieron de dos parcelas de trigo en Broadbalk durante treinta años; la única diferencia en el tratamiento de estiércol fue que '9 a' recibió nitrato de sodio, mientras que '7 b' recibió una cantidad equivalente de nitrógeno como sulfato de amonio. En el curso del experimento, la parcela '9 a' parece estar ganando rendimiento sobre la parcela '7 b'. ¿Es este aparente incremento significativo?
>
> Una gran parte de la variación en el rendimiento de un año a otro es evidentemente similar en las dos parcelas; en consecuencia, la serie de diferencias dará un resultado más claro. En cierto sentido estos datos son especialmente simples, los treinta valores de la variable independiente forman una serie con intervalos iguales entre los valores sucesivos, con solo un valor de la variable dependiente correspondiente a cada uno. En tales casos, el trabajo se simplifica utilizando la fórmula $S(x-\bar{x})^2 = \frac{1}{12}n'\left(n'^2 - 1\right)$.

Tabla 4.7. *Rendimiento de granos preparados por acre (en fanegas)* (fuente: Fisher, 1934, p. 134)

Año de cosecha	**9 a**	**7 b**	**9 a – 7 b**	
1855	29.62	33.00	-3.38	
1856	32.38	36.91	-4.53	
1857	43.75	44.84	-1.09	
1858	37.56	38.94	-1.38	
1859	30.00	34.66	-4.66	
1860	32.62	27.72	+4.90	$S(x-\bar{x})^2 = \frac{n'\left(n'^2-1\right)}{12} = 2247.5$
1861	33.75	34.94	-1-19	$b = 0.2668$
1862	43.44	35.88	+7.56	$S(y-\bar{y})^2 = 1020.56$
1863	55.56	53.66	+1.90	$b^2S(x-\bar{x})^2 = 159.99$
1864	51.06	45.78	+5.28	$S(y-Y)^2 = 860.57$
1865	44.06	40.22	+3.84	$s^2 = 30.73$
1866	32.50	29.91	+2.59	$s^2/S(x-\bar{x})^2 = 0.013675$
1867	29.13	22.16	+6.97	$= (0.1169)^2$

1868	47.81	39.19	+8.62	$t = 2.282$
1869	39.00	28.25	+10.75	$n = 28$
1870	45.50	41.37	+4.13	
1871	34.44	22.31	+12.13	
1872	40.69	29.06	+11.63	
1873	35.81	22.75	+13.06	
1874	38.19	39.56	-1.37	
1875	30.50	26.63	+3.87	
1876	33.31	25.50	+7.81	
1877	40.12	19.12	+21.00	
1878	37.19	32.19	+5.00	
1879	21.94	17.25	+4.69	
1880	34.06	34.31	-0.25	
1881	35.44	26.13	+9.31	
1882	31.81	34.75	-2.94	
1883	43.38	36.31	+7.07	
1884	40.44	37.75	+2.69	
Media	37.50	33.03	+4.47	

Donde n' es el número de términos, o 30 en este caso. Para evaluar b es necesario calcular la suma de productos $S\{y(x-\bar{x})^2\}$; el cual tiene la misma relación con la covarianza de dos variables que la suma de los cuadrados con la varianza de una sola variable; […] al probar la hipótesis de que $\beta = 0$, es decir, que la parcela '9 a' no ha estado ganando sobre la parcela '7 b', dividimos b por esta cantidad y encontramos $t = 2.282$. Dado que s se encontró a partir de 28 grados de libertad $n = 28$, y el resultado de t muestra que P está entre 0.02 y 0.05.

El resultado debe ser considerado significativo, aunque apenas; a la vista de los datos, no podemos ignorar la posibilidad de que, en este campo, y en conjunto con los otros abonos utilizados, el nitrato de sodio haya conservado la fertilidad mejor que el sulfato

> de amonio; sin embargo, los datos no demuestran este punto más allá de toda posibilidad de duda.
>
> El error estándar de $\bar{y}$, calculado a partir de estos datos, es 1.012, por lo que no cabe duda de que la diferencia en los rendimientos medios es significativa […]

De acuerdo con Fisher (1934) así como la prueba de significancia de la diferencia de medias se puede aplicar cuando las muestras son de diferentes tamaños, "*si obtenemos una estimación del error combinando las sumas de cuadrados derivadas de las dos muestras diferentes*" (p. 137), entonces, también podemos aplicar el método para comparar los coeficientes de regresión cuando las series de la variable independiente no son idénticas, y en caso de que lo fueran se puede omitir el comparar los coeficientes de regresión. Fisher mostró el siguiente problema para ejemplificar cómo funciona esta prueba en caso de que las series no sean idénticas.

Ejemplo 22: Prueba de significancia de la diferencia de los coeficientes de regresión por medio de la comparación de la tasa de crecimiento relativo de dos cultivos de un alga (Fisher, 1934, pp. 137-139).

> La [*Tabla 4.8*] muestra el logaritmo (base 10) de los volúmenes ocupados por células de algas en días sucesivos, en cultivos paralelos, cada uno de ellos tomado en un período durante el cual la tasa de crecimiento relativo fue aproximadamente constante. En el cultivo A hay nueve valores disponibles, y ocho en el cultivo B (datos del Dr. M. Bristol-Roach).

Tabla 4.8. *Logaritmo de los volúmenes ocupados por células de algas* (fuente: Fisher, 1934, p. 134)

Valores Log		**Valores de la sumatoria**			
A	B	A	B		
3.592	3.538	43.426	38.358	$S(y-Y)^2$, A	0.05089
3.823	3.828	39.834	34.820	" B	0.07563
4.174	4.349	36.011	30.992	ns^2	0.12652
4.534	4.833	31.837	26.643	s^2	0.009732
4.955	4.911	37.303	21.810	$s^2/60$	0.0001622
5.163	5.297	22.347	16.899	$s^2/42$	0.0002317

	5.495	5.566	17.184	11.602		
	5.602	6.036	11.689	6.036		
	6.087	…	6.087	…		
Total	43.426	38.358	235.718	187.160	Error estándar	0.01985
Media	4.8251	4.7947	217.130	172.111	$b' - b$	0.0366
		$Sy(x - \bar{x})$	18.588	14.549	t	1.844
		b	0.3098	0.3464	n	13

El método para encontrar $Sy(x - \bar{x})$ mediante la sumatoria se muestra en el segundo par de columnas: los valores originales se suman desde la parte inferior, dando totales sucesivos de 6.087 a 43.426; el valor final debe, por supuesto, coincidir con el total debajo de los valores originales. De la suma de la columna de totales se resta la suma de los valores originales multiplicada por 5 para A y por 4.5 para B. Las diferencias son $Sy(x - \bar{x})$; estos deben dividirse por los valores respectivos de $S(x - \bar{x})^2$, a saber, 60 y 42, para obtener los valores de b, que miden las tasas de crecimiento relativo de los dos cultivos. Para probar si la diferencia es significativa, calculamos en los dos casos $S(y^2)$, y restamos sucesivamente el producto de la media con el total, y el producto de b con $Sy(x - \bar{x})$; este proceso deja los dos valores de $S(y - Y)^2$, que se agregan como se muestra en la tabla anterior, y la suma se divide por n, para dar s^2. El valor de n se encuentra sumando los 7 grados de libertad de la serie A y los 6 grados de la serie B, por lo tanto es 13. Las varianzas estimadas de los dos coeficientes de regresión se obtienen dividiendo s^2 por 60 y 42, y la varianza de su diferencia es la suma de estos. Tomando la raíz cuadrada, encontramos que el error estándar es 0.01985, y $t = 1.844$. La diferencia entre los coeficientes de regresión, aunque es relativamente grande, no puede considerarse significativa. No hay evidencia suficiente para afirmar que el cultivo B estaba creciendo más rápidamente que el cultivo A.

Welch (1938, 1947), propuso una adaptación para la prueba t-Student cuando las varianzas de las poblaciones, de las cuales se extrajeron las muestras, son desiguales; aunque también se puede utilizar cuando las muestras tienen diferente tamaño. A esta prueba se le conoce como ‘Prueba t de varianzas desiguales de Welch’ o ‘Prueba t de varianzas desiguales’.

Welch (1938), inició indicando que si tenemos dos muestras de tamaño n_1 y n_2, las cuales provienen de las poblaciones normales π_1 y π_2, respectivamente. Mientras que π_1 tiene por media y desviación estándar a α_1 y σ_1, la media y desviación estándar para π_2 son α_2 y σ_2. Si se requiere probar que $\alpha_1 = \alpha_2$ podemos tener dos casos. En el primero, σ_1 y σ_2 son iguales, el segundo caso es que las desviaciones estándar sean desiguales. Para el primer caso definió el criterio $u = \frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\left(\frac{\Sigma_1+\Sigma_2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, para una distribución t-Student con $f = (n_1 + n_2 - 2)$. Para el de varianzas desiguales, cuando se desconoce la relación entre ellas, presentó el criterio $v = \frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\left(\frac{\Sigma_1}{n_1(n_1-1)}\right)\left(\frac{\Sigma_2}{n_2(n_2-1)}\right)}}$, que a menudo era utilizado en estos casos. Para este último criterio se utilizaban las tablas de la distribución normal cuando la muestra es lo suficientemente grande, sin embargo, no existía una aproximación para trabajar con muestras pequeñas.

El trabajo de Welch (1938), partió de algunas consideraciones que había realizado Fisher y primero se planteó el problema ¿hasta qué punto es válido el criterio u incluso cuando las varianzas son desiguales?, también consideró la validez de la prueba de hipótesis utilizando v para la distribución t-Student con $f = (n_1 + n_2 - 2)$. Después procedió a encontrar una aproximación de las distribuciones de u y v; y cuando $\sigma_1 = \sigma_2$ se tiene que $(\bar{x}_1 - \bar{x}_2) = \chi'\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$; $\Sigma_1 = \chi_1^2\sigma_1^2$; $\Sigma_2 = \chi_2^2\sigma_2^2$, donde χ'^2, χ_1^2 y χ_2^2 son independientemente distribuidas como una χ^2 con un grado de libertad, $(n_1 - 1)$ y $(n_2 - 1)$ respectivamente. Por lo que fue posible escribir u y v de la forma $y = \frac{\chi'}{\sqrt{a\chi_1^2+b\chi_2^2}} = \frac{\chi'}{\sqrt{w}}$, donde a y b son contantes que dependen de las $n's$ y $\sigma's$, mientras que w se distribuye independientemente de χ', pero cuando las constantes son iguales o iguales a cero, w se distribuye de acuerdo con la distribución χ^2 multiplicada por una constante. Entonces, para estos casos, y se distribuirá de acuerdo a una t-Student multiplicada por una constante. Cuando no se tienen los casos señalados anteriormente, según Welch (1938) una aproximación de la distribución puede ser muy útil y señaló que la curva de Pearson tipo III puede ser la primera aproximación de w como $p(w) = \frac{1}{(2g)^{\frac{1}{2}f}\Gamma\left(\frac{1}{2}f\right)} w^{\frac{1}{2}f-1}e^{-\frac{w}{2g}}$, donde f y g provienen de los primeros dos momentos de la curva de acuerdo con los verdaderos momentos de w. Con dichos valores de f y g obtuvo que w/g se distribuye aproximadamente como una distribución χ^2 con f grados de libertad, e indicó que para el propósito de la comparación de este trabajo esta aproximación es suficiente y señaló que u es de la forma

$$a = \frac{\sigma_1^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{(n_1+n_2-2)\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$$

$$b = \frac{\sigma_2^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{(n_1+n_2-2)\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$$

Mientras que $u = ct_f$, donde $f = \frac{\{(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2\}^2}{\{(n_1-1)\sigma_1^4+(n_2-1)\sigma_2^4\}}$; $c = \sqrt{\frac{(n_1+n_2-2)\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\{(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2\}}}$ y para v se tiene que

$$a = \frac{\frac{\sigma_1^2}{n_1(n_1-1)}}{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$$

$$b = \frac{\frac{\sigma_2^2}{n_2(n_2-1)}}{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$$

Entonces, podemos decir que $v = ct_f$, donde $c = 1$ y $f = \frac{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2}(n_1-1)+\frac{\sigma_2^4}{n_2^2}(n_2-1)}$.

Posteriormente, Welch (1938) procedió a la validación primero para u y después para v. En la validación de u obtuvo que, cuando $\sigma_1 \neq \sigma_2$ por medio de la conexión entre la distribución t-Student y la función Beta se tiene que $P(|u| > u_0) = I_{z_0}\left(\frac{1}{2}f, \frac{1}{2}\right)$, donde $z_0 = \frac{f}{\left(f+\frac{u_0^2}{c^2}\right)}$; es importante señalar que c y f dependen de la relación $\theta = \sigma_1^2/\sigma_2^2$.

De acuerdo con Welch (1938), para la validación del procedimiento de la prueba de hipótesis con el criterio v y la distribución t-Student con (n_1+n_2-2) grados de libertad, se procede de la misma forma que para validar u (aunque los tamaños muestrales sean diferentes); mientras que si los tamaños muestrales son iguales no hay necesidad de discutir estos criterios por separado, dado que, entonces u y v son iguales. Supongamos $n_1 = 5$ y $n_2 = 15$, entonces tenemos que $f = \frac{28(3\theta+1)^2}{(63\theta^2+2)}$ y $c = 1$. Por lo que de forma similar a $P(|u| > u_0) = I_{z_0}\left(\frac{1}{2}f, \frac{1}{2}\right)$ obtenemos $P(|v| > 2.101)$ que se ilustra en la Figura 4.4, al respecto indicó que es notorio que esta prueba es imparcial cuando $\theta = 2/21$, sin embargo, no existe la posibilidad de sesgos tan grandes utilizando el criterio v como cuando se utiliza u para algunos valores de θ. Por lo que añadió "*si existe la posibilidad de que σ_1 y σ_2 difieran, entonces puede dar muy resultados engañosos y será más seguro usar v*" (Welch, 1938, p. 355).

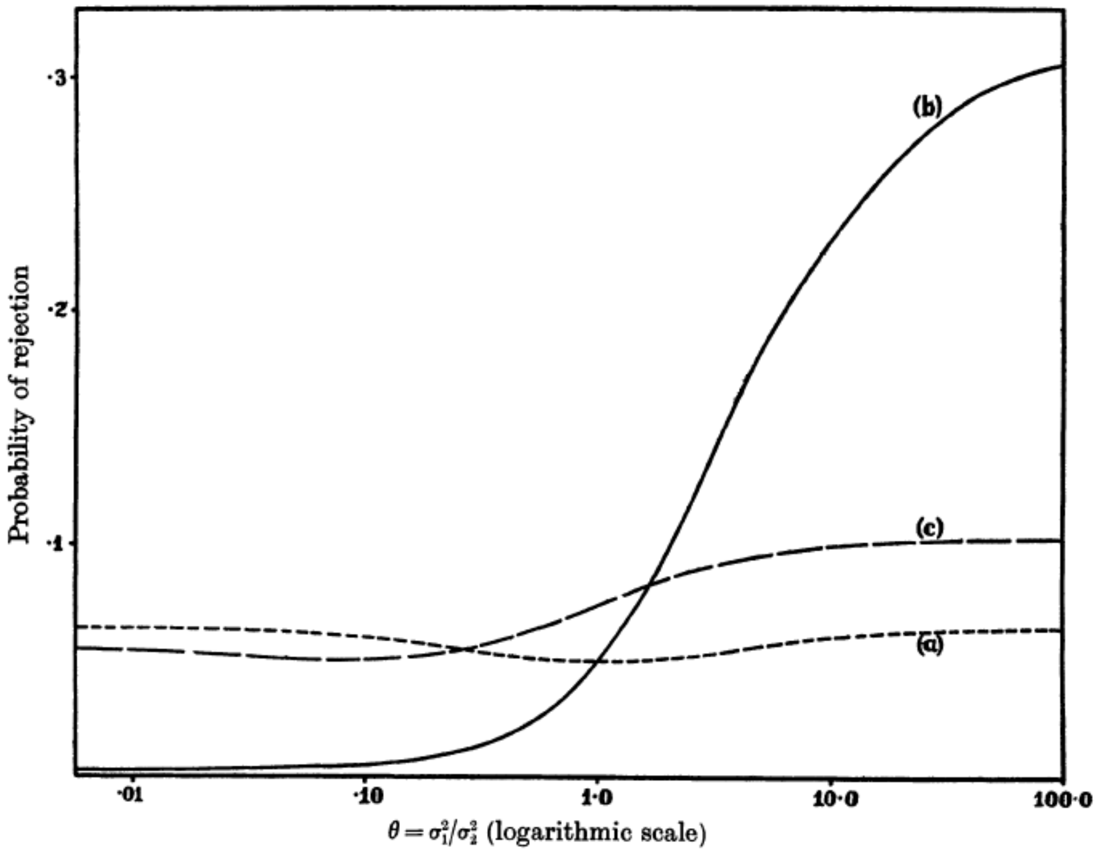


Fig. 1. Probability of rejection of hypothesis $\alpha_1 = \alpha_2$ when true plotted against θ. (*a*) $n_1 = n_2 = 10$. $P(|u| > 2{\cdot}101)$; (*b*) $n_1 = 5$, $n_2 = 15$, $P(|u| > 2{\cdot}101)$; (*c*) $n_1 = 5$, $n_2 = 15$, $P(|v| > 2{\cdot}101)$.

Figura 4.4. Probabilidad de rechazo de la hipótesis $\alpha_1 = \alpha_2$ cuando es trazado contra θ

Welch también se cuestionó que si utilizamos v, ¿cuál es el mejor valor de los grados de libertad? Al respecto, planteó que una vez que obtuvo v buscar P en tabla de la distribución t-Student considerando $(f_1 + f_2)$ grados de libertad sólo para cuando θ tiene un valor particular $(\theta = V_2 f_1 / V_1 f_2)$, mientras que cuando consideramos que se encuentra próximo a cierto valor es preferible $f = \frac{(V_1\theta + V_2)^2}{\frac{V_1^4\theta^2}{f_1} + \frac{V_2^4}{f_2}}$.

Welch (1947) retomó este problema y realizó la generalización para cundo se tienen k poblaciones y partió indicando que la distribución muestral de s_i^2 es $p(s_i^2) = ds_i^2 = \frac{1}{\Gamma\left(\frac{1}{2}f_i\right)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i - 1} exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right] d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)$ y que un caso muy sencillo de esta configuración ocurre cuando tenemos dos muestras, n_1 y n_2, las cuales provienen de poblaciones normales con medias verdaderas α_1 y α_2, y desviaciones estándar σ_1 y σ_2. Además, si consideramos que η es la diferencia verdadera entre las medias $(\alpha_1 - \alpha_2)$, entonces la diferencia estimada es $y = (\bar{x}_1 - \bar{x}_2)$; mientras que la varianza estimada es $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, donde $\lambda_1 = 1/n_1$ y $\lambda_2 = 1/n_2$. Para estimar los valores de las varianzas σ_1^2 y σ_2^2 se recurre a $s_1^2 = \Sigma_1/f_1$ y $s_2^2 = \Sigma_2/f_2$, donde Σ_1 y Σ_2 son las sumas del cuadrado de las desviaciones de las observaciones individuales de la media de la muestra respectiva y $f_1 = n_1 - 1$ y $f_2 = n_2 - 1$. Para generalizar los declarado en Welch (1938), ahora era necesario determinar una cantidad h que se pudiera calcular a partir de las observaciones, "*con la propiedad de que la probabilidad de la diferencia* $(y - \eta)$ *sea inferior a* h, *es una probabilidad dada por P*" (Welch, 1947, p. 28). De acuerdo con Welch, este problema lo satisface la ecuación $Pr.[(y - \eta) < h(s_1^2, s_2^2, \dots, s_k^2, P)] = P$, la derivación matemática de la solución exacta a este problema esta dada, para términos de orden $1/f_1^2$ por

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_1^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_1^2)^2} - \frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_1^4}{f_i^2}\right)}{(\Sigma\lambda_i s_1^2)^2} + \frac{(3+5\xi^2+\xi^4)}{3}\frac{\left(\Sigma\frac{\lambda_i^3 s_1^6}{f_i^2}\right)}{(\Sigma\lambda_i s_1^2)^3} - \frac{(15+32\xi^2+9\xi^4)}{32}\frac{\left(\Sigma\frac{\lambda_i^2 s_1^4}{f_i^2}\right)}{(\Sigma\lambda_i s_1^2)^4}\right]$$

Welch (1947), retomó el criterio v, que hemos descrito previamente, para el problema especifico de $k = 2$ y considerando que si $z = (a\chi_1^2 + b\chi_2^2)$, la distribución aproximada es $p(z)dz = \frac{1}{\Gamma\left(\frac{1}{2}f\right)} e^{-\frac{1}{2}(z/g)}\left(\frac{z}{2g}\right)^{\frac{1}{2}f-1} d\left(\frac{z}{2g}\right)$, por lo que obtuvo que

$$v = \frac{(y - \eta)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}}$$

y sigue aproximadamente una distribución t-Student con $f = \frac{\left(\lambda_1\sigma_1^2+\lambda_2\sigma_2^2\right)^2}{\frac{\lambda_1^2\sigma_1^4}{f_1}+\frac{\lambda_2^2\sigma_2^4}{f_2}}$ grados de libertad.

Mientras que para el caso en que k no se restringe a 2, obtuvo el criterio

$$v = \frac{(y-\eta)}{\sqrt{(\lambda_i s_i^2)}}$$

Este criterio también sigue aproximadamente una distribución t-Student con $f = \frac{\left(\lambda_i\sigma_i^2\right)^2}{\Sigma\frac{\lambda_i^2\sigma_i^4}{f_i}}$ grados de libertad. Si las $\sigma_i's$ se desconocen podemos utilizar $f = \frac{\left(\Sigma\lambda_i s_i^2\right)^2 - 2\left(\Sigma\frac{\lambda_i s_i^4}{f_i+2}\right)}{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i+2}\right)}$.

Considerando el criterio v, Welch (1947) reformuló su aproximación de la cantidad h como

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_1^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2} - \frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2}\right.$$
$$\left. + \frac{(51+64\xi^2+5\xi^4)}{96}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4} + \cdots\right]$$

Mediante el cálculo de $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, que depende de las varianzas observadas, y que si tenemos una P de que $(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)$, de acuerdo con Welch (1947) esto nos provee de un método para probar consistencia entre la y observada y η. Pero si queremos estimar η y proporcionar una medida de la incertidumbre de la estimación, podemos cambiar la interrogante, por ejemplo, si tenemos una sola muestra la P de que η sea mayor que $\{y - h(s_1^2, s_2^2, \ldots, s_k^2, P)\}$. Por lo que una estimación del intervalo con dos niveles de P "sería $(P_1 - P_2)$ de que η se encuentre entre $\{y - h(s_1^2, s_2^2, \ldots, s_k^2, P_1)\}$ y $\{y - h(s_1^2, s_2^2, \ldots, s_k^2, P_2)\}$.

Así como en las anteriores pruebas t-Student, la prueba de Welch sigue la asunción de que las dos poblaciones siguen una distribución normal; sin embargo, cambia en cuanto a que ahora se asumen varianzas desiguales. Además, esta prueba de varianzas desiguales, también tiene la bondad de que se puede utilizar si los tamaños muestrales son iguales o diferentes.

Con la finalidad de ilustrar cómo funciona la prueba de varianzas desiguales de Welch con el estadístico t-Student (criterio v de Welch) hemos retomado el ejemplo 18 que se presentó en la página 186 que Fisher (1925b) utilizó con la finalidad de aplicar la prueba t-Student de significancia de la diferencia de medias de dos muestras con diferente tamaño. Para efectos del análisis de las configuraciones a este ejemplo, que ahora implica lo que se puede denominar la corrección de Welch, se le denomina ejemplo 23. Con el número de macollos presentados en la Tabla 4.9 hemos realizado los procedimientos para realizar esta prueba, los cuales se detallan a continuación.

Ejemplo 23. Significancia de la diferencia de medias de dos muestras con diferente tamaño y con desigualdad de varianzas.

Tabla 4.9. *Número de macollos por maceta* (fuente: Fisher, 1925b, p. 95)

Experimento	frecuencia									$\bar{x}$	Σ	s^2
Electrificada	16,	16,	20,	16,	20,	17,	15,	21		17.625	37.875	5.410
Enjaulada	17,	27,	18,	25,	27,	29,	27,	23,	17	23.333	184	23

El valor del estadístico t-Student o criterio $v = 3.175$ y el de los grados de libertad $f = 11.846$, nos indican que se tiene una probabilidad $P = 0.00441764$ de que no exista una diferencia entre las medias, lo cual quiere decir que la posibilidad de que las medias sean iguales es de 44 veces en 10,000. Por lo tanto, la diferencia entre las medias es significativa y podemos decir que el número de macollos de las plantas de las macetas electrificadas y de las enjauladas es diferente.

En estudios recientes (e.g. Ruxton, 2006) se ha puesto de manifiesto que esta prueba t de varianzas desiguales es una alternativa para la prueba t-Student y para la Prueba U de Mann-Whitney (es una prueba no paramétrica que se aplica a dos muestras independientes, es decir, como la versión de la prueba t-Student para datos no paramétricos).

4.2 Configuraciones ontosemióticas del estadístico $t - Student$

Asimismo, se describen las configuraciones ontosemióticas identificadas con el recorrido histórico del estadístico t-Student por medio de los objetos primarios: elementos lingüísticos, situaciones/problemas, conceptos/definiciones, proposiciones/propiedades, procedimientos y argumentos.

Se parte de las tres grandes problemáticas que permitieron el surgimiento y evolución de dicho estadístico: prueba t-Student para una muestra, prueba t-Student para dos muestras y prueba t-Student para dos muestras apareadas o dependientes. Estas problemáticas se encuentran estrechamente vinculadas a la problemática: distribución t-Student. Cada una de estas problemáticas da cuenta de un significado para el estadístico t-Student y a su vez en cada significado se encuentran diversos significados parciales. El significado holístico del estadístico t-Student está conformado por 12 significados parciales.

4.2.1 Significado 1: Prueba t-Student de la media de una muestra

La prueba t-Student para una muestra determina si la media de una población, que se distribuye de acuerdo a la normal, es igual a un valor μ_0.

4.2.1.1 Significado parcial 1 (SP1): El método de intercomparación de Galton (1875)

En este significado se retoma el ejemplo 14 que se presentó en la página 163. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar el método de intercomparación, Galton utilizó *lenguaje natural* en el planteamiento, la explicación de cada uno de los pasos del método y la interpretación de los resultados. También utilizó la *representación gráfica* (ojiva) para mostrar cómo se ordenarían y graficarían los datos sobre las estaturas de los hombres.

Situación/problema:

Galton buscaba una forma de establecer si era apropiado trabajar un conjunto de datos observados bajo las condiciones de la distribución normal. Para lo que en un inicio trabajó el método de intercomparación (1875), además este método le permitía realizar un análisis de los datos sin algún agente externo a la serie observada. Este método le permitía comparar con mucho detalle los datos de la misma serie o con otra serie.

Para explicar el método intercomparación, en 1880, Galton lo detalló tratando el ejemplo de las estaturas de los hombres ingleses (ver sección 4.1.1).

Conceptos/definiciones:

CD1: *Error probable,* es una medida de variabilidad de la serie observada. Definido como la mediana de los errores para datos distribuidos normalmente. [En el mismo sentido del CD1 del SP1 del estadístico χ^2].

CD2: *Ojiva,* es la curva de distribución de los valores (en porcentaje) observados ordenados. [En el mismo sentido del CD2 del SP1 del estadístico χ^2].

CD3: *Cuartiles,* determinan los valores correspondientes al 25%, 50% (mediana) y 75% de los datos, la base de la ojiva se encuentra dividida en cuatro partes iguales.

CD4: *Valor medio,* corresponde al dato o individuo que se encuentra justo en el centro de la serie o grupo de datos, en otras palabras m.

CD5: *Percentiles,* determinan los valores correspondientes del 1% al 100% de los datos, se divide la base de la ojiva en 100 partes. [En el mismo sentido del CD4 del SP1 del estadístico χ^2].

Propiedades/proposiciones:

PP1: *Ley del error* (Sevilla, 1993) [En el mismo sentido del PP1 del SP1 del estadístico χ^2].

Primer postulado:

En una serie de un gran número de observaciones efectuadas en idénticas circunstancias y con el mismo esmero, los distintos errores posibles se presentan con tanta mayor frecuencia cuanto menor sea su valor absoluto.

Por lo que cada fenómeno se presenta un número de veces proporcional a su probabilidad, mientras que la probabilidad se encuentra en función de la magnitud del error: $P_i = \omega(\varepsilon_i)$.

Segundo postulado:

Postulado de la obra Theoria Motus Corporum Celestium: Obtenidos por observación n valores distintos de una magnitud, siempre que todos hayan sido obtenidos en idénticas condiciones y con el mismo esmero, el valor más probable (medida) de la magnitud es la media aritmética de los valores observados'.

Éste postulado después fue sustituido por el mismo Gauss, por el siguiente:

En una serie de un gran número de observaciones efectuadas en idénticas circunstancias y con el mismo esmero, los errores del mismo valor absoluto, pero de distinto signo, se presentan con la misma frecuencia y son, por lo tanto, igualmente probables.

La distribución de los errores es simétrica respecto al cero y siempre que se trate de un gran número de observaciones se puede decir que $\sum_{i=1}^{n} \varepsilon_i = 0$.

PP2: *Distribución normal* [En el mismo sentido del PP2 del SP1 del estadístico χ^2].

Es una distribución de probabilidad de variable continua cuya representación gráfica es una curva con forma de campana, es una distribución de los errores de medición, es simétrica y se distribuye alrededor de la media; la altura y ancho de la distribución depende de la varianza de las mediciones. La distribución normal estándar tiene $\mu = 0$ y $\sigma = 1$.

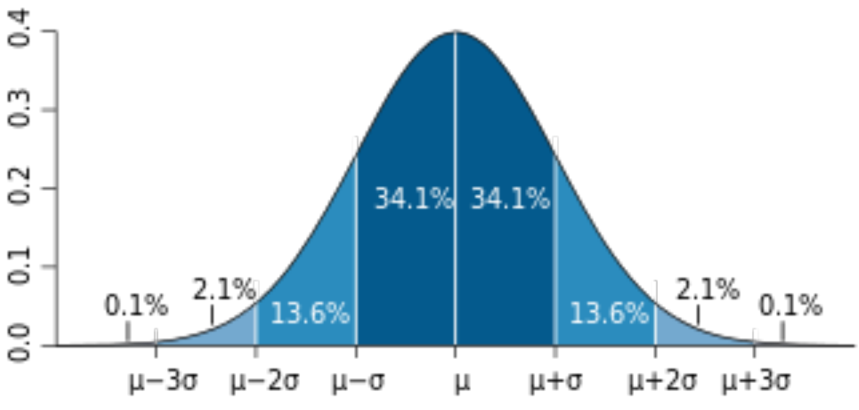


Figura 4.5. Distribución de probabilidad $N(\mu, \sigma)$

PP3: m en $\frac{1}{2}$ o $0°$ [En el mismo sentido del PP3 del SP1 del estadístico χ^2].

PP4: p en $\frac{1}{4}$ o $-25°$ [En el mismo sentido del PP4 del SP1 del estadístico χ^2].

PP5: q en $\frac{3}{4}$ o $25°$ [En el mismo sentido del PP5 del SP1 del estadístico χ^2].

PP6: *Error probable,* es la mitad de la diferencia entre los cuartiles, también estableció a $q - m$ como la divergencia o el error probable (de la parte de la serie que excede a la mediana) y a $m - p$ como la divergencia de la otra parte de la serie. [En el mismo sentido del PP6 del SP1 del estadístico χ^2].

PP7: Si la *serie es simétrica* entonces $q - m = m - p$. [En el mismo sentido del PP7 del SP1 del estadístico χ^2].

Procedimientos:

Graficar la ojiva de frecuencias acumuladas mediante el método intercomparación bajo la secuencia de las siguientes acciones:

1. Calcular la suma de los observados hasta cada medida en pulgadas (frecuencias acumuladas) a partir de las frecuencias observadas en las mediciones. Esto corresponde a la abscisa de A a B.
2. Encontrar los porcentajes de acuerdo a las sumas del cálculo del punto 1 (frecuencia relativa acumulada). Corresponde a la abscisa de 0 a 100.
3. Establecer la ordenada correspondiente a cada frecuencia acumulada relativa.
4. Con los valores de los puntos 3 y 4 realizar la gráfica de la ojiva o curva de frecuencias correspondiente:
    a. En la ojiva se pueden identificar p, m y q, y con estos establecer si la serie es simétrica mediante $q - m = m - p$.
    b. Si la curva sigue la forma de la ojiva establecida por Galton es un indicativo de que el conjunto de datos observados sigue un a distribución normal.

Argumentos:

Se argumenta utilizando el gráfico de la curva de la distribución y los valores representantes como los cuartiles.

Por ejemplo:

> Siete hombres se convierten así en representantes eficientes de una clase muy numerosa. Como regla general, se encontrará que estos siete representantes seleccionados diferirán entre sí en intervalos aproximadamente iguales, la diferencia

> entre el suboctil y el octil suele ser aproximadamente la misma que entre el octil y el cuartil, y entre el cuartil y el valor medio. (Galton, 1880, p. 9)

Si se va a comparar con otra serie de datos Galton señaló que:

La comparación de cualquiera de los dos grupos se realiza al cotejar sus siete representantes cada uno con cada uno, el primer suboctil de uno con el primer suboctil del otro, el primer octil con el primer octil, el primer cuartil con el primer cuartil, etc. También comparto el más alto de cada uno, y nuevamente el más bajo de cada uno, como un mero asunto de interés, pero no como una operación estadística precisa, por las razones ya expuestas. (Galton, 1880, p. 10)

4.2.1.2 Significado parcial 2 (SP2): Significancia de la media con el estadístico $z = \frac{x}{s}$

En este significado parcial se identificaron los siguientes elementos primarios a partir de la práctica matemática desarrollada en la resolución del ejemplo 16 descrito en la página 174.

Elementos lingüísticos:

Para llevar a cabo la prueba bajo su método, Gosset recurrió al *lenguaje natural* para el planteamiento del problema, para ir describiendo cada paso de su método, así como para las conclusiones de la prueba. Por medio de la *representación tabular* mostró los datos con los que se trabajarían y, además, en la misma representación mostró el valor de la media y la desviación estándar para cada columna (Student, 1908a). También hizo uso de la representación simbólica cuando se refirió al valor del estadístico.

Situación/problema:

Alrededor de 1900, las pruebas de hipótesis sobre la media aún se trabajaban únicamente bajo en enfoque de grandes muestras, pero por el trabajo que desarrollaba Gosset en la industria cervecera, requería trabajar con muestras pequeñas. La desventaja de trabajar con medias pequeñas consistía en que a medida que el tamaño de la muestra va disminuyendo también va incrementando el error sobre el valor de la desviación estándar de la muestra. Para abordarlo, en 1908 Gosset propuso trabajar con $z = x/s$ y proveyó tablas de probabilidad para z para cuando n es pequeña, lo que permitiría juzgar la significancia de la media de la muestra.

Para ilustrar su método retomó datos sobre los efectos que surtían dos fármacos en diez personas, los datos fueron publicados en el Journal of Physiology de 1904 por Cushny y

Peebles. Los datos versaban sobre el número de horas de sueño ganadas por el uso de cada uno de los fármacos.

Conceptos/definiciones:

CD1: *Experimento,* como cualquier acción o proceso cuyo resultado está sujeto a incertidumbre. [En el mismo sentido del CD1 del SP2 del estadístico χ^2].

CD2: *Muestra,* es el subconjunto de una población, el subconjunto está constituido por ciertas observaciones que representan al total. [En el mismo sentido del CD1 del SP9 del estadístico χ^2].

CD3: *Población,* es el conjunto de elementos que tienen características comunes. [En el mismo sentido del CD3 del SP9 del estadístico χ^2].

CD4: *Media,* es una medida de posición y es el promedio de un conjunto de observaciones. Se consideraba que la media era el valor más probable de la magnitud, por ejemplo, en las mediciones astronómicas.

CD5: *Desviación estándar,* es una medida de dispersión de los datos con respecto a la media que en 1893 sustituyó al error probable (ver CD1 del SP1).

CD6: *Estadístico* z, es una medida que indica a cuantas desviaciones estándar se encuentra la media de cero.

CD7: *Probabilidad,* es una medida de la certidumbre asociada a un evento, en este contexto se entiende como una medida de que la media se encuentre entre $-\infty$ y z. Para juzgar la significancia de la media de una serie de experimentos se utiliza la probabilidad de que la media de una muestra de tamaño n, con una medición en términos de la desviación estándar de la muestra, exceda el valor del estadístico z.

CD8: *Significancia,* como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria [En el mismo sentido del CD8 del SP3 del estadístico χ^2].

Propiedades/proposiciones:

PP1: *Distribución normal,* en el mismo sentido que la PP2 del SP1.

PP2: *Distribución del estadístico* $z = \frac{x}{s}$, es la distribución conjunta, suponiendo que x se puede medir en términos de s. Cuando n es grande esta distribución tiende a la normalidad.

- Teniendo las siguientes ecuaciones para la distribución de z:

Cuando n es impar

$$y = \frac{1}{2}\frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \ldots \frac{5}{4} \cdot \frac{3}{2}(1+z^2)^{-n/2}$$

Cuando n es par

$$y = \frac{1}{\pi}\frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \ldots \frac{4}{3} \cdot \frac{2}{1}(1+z^2)^{-n/2}$$

- Fórmula para encontrar el valor de la probabilidad de que la media se encuentre entre $-\infty$ y z:

Cuando n es impar

$$P = \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \ldots \left(\frac{3}{2} \cdot \frac{1}{2}\right) \int_{-(\pi/2)}^{tan^{-1}z} cos^{n-2}\theta \ d\theta$$

Cuando n es par

$$P = \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \ldots \left(\frac{2}{1} \cdot \frac{1}{\pi}\right) \int_{-(\pi/2)}^{tan^{-1}z} cos^{n-2}\theta \ d\theta$$

PP3: *Media* $\bar{x} = \frac{1}{n}\sum x_i$.

PP4: *Desviación estándar* $s = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$.

PP5: *Estadístico t-Student*

$$z = \frac{x}{s} = \frac{(\bar{x} - \mu)}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}}$$

Procedimientos:

Encontrar la probabilidad de ocurrencia de que la media se encuentre entre $-\infty$ y z, para lo cual es necesario calcular el valor del estadístico z para encontrar la probabilidad asociada a él. En este ejemplo, al tener dos fármacos Gosset aplicó su prueba para cada fármaco y además para la diferencia entre los ellos, bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico $z = \frac{x}{s}$, para lo cual es necesario:
    a. Calcular la media bajo la siguiente formula $\bar{x} = \frac{1}{n}\sum x_i$.
    b. Encontrar el valor de la desviación estándar de la muestra por medio de $s = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$.
2. Encontrar el valor de la probabilidad en la tabla de probabilidad para valores de z y n de Gosset o realizar los cálculos con $P = \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \cdots \begin{pmatrix} \frac{3}{2}\cdot\frac{1}{2} n \ impar \\ \frac{2}{1}\cdot\frac{1}{\pi} n \quad par \end{pmatrix} \int_{-(\pi/2)}^{tan^{-1}z} cos^{n-2}\theta \ d\theta.$

Argumentos:

Se argumenta señalando que la distancia media se encuentra a z (el valor del estadístico) deviaciones estándar de cero y con la probabilidad asociada al estadístico que se puede encontrar en la tabla, lo que permite juzgar la significancia de la media de esta serie de experimentos.

Por ejemplo:

> La media de esta serie es $+1.58$ mientras que la S. D. es 1.17, la media sería $+1.35$ veces la S. D. De la tabla de probabilidad se obtiene 0.99859 o que las probabilidades son alrededor de 666 a 1 de que 2 es mejor somnífero. (Student, 1908, p. 21)

Reflexión:

Si vemos este problema que utilizó Gosset en Student (1908a) con una mirada actual, podríamos decir que en realidad este problema trata sobre dos muestras dependientes. La prueba es similar a lo que Gosset realizó con la diferencia (ver los resultados para la cuarta columna de la Tabla 4.3). Este problema fue retomado por Fisher (1934), quien lo resolvió por este mismo método y por el de diferencia de medias de dos muestras (ver ejemplo 17 y 18 en las páginas 184 y 186, respectivamente).

4.2.1.3 Significado parcial 3 (SP3): Significancia de la media de una muestra con el estadístico t-Student

En este significado se retoma el ejemplo 17 que se presentó en la página 184. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar esta prueba de significancia Fisher, recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos del experimento, y, además en la misma representación mostró el valor de la media para cada columna. También hizo uso de la *representación simbólica* al realizar los cálculos para obtener el valor del estadístico t-Student.

Situación/problema:

Para ilustrar cómo funciona la prueba de significancia de la media de una muestra pequeña con el estadístico t-Student, Fisher (1934) retoma el ejemplo de Gosset (Student, 1908a) sobre los efectos que surtían dos fármacos en diez personas, como se describió en el SP2. Con la diferencia que Fisher (1934) utiliza el estadístico t-Student y su propia tabla de la distribución t-Student.

Conceptos/definiciones:

CD1: *Experimento,* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra,* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población,* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media,* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar,* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student, de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad,* es una medida de la certidumbre asociada a un evento, en este contexto se entiende como una medida de que la media se encuentre fuera del rango $\pm t$, en la tabla de Fisher (1934) tomó valores de 0.9, …, .01; para juzgar la significancia de la media de una serie de experimentos.

CD8: *Distribución t-Student,* es para muestras aleatorias de una población normal que se distribuye alrededor de cero, como su media. Fisher (1934), realizó una tabla de esta distribución donde podemos encontrar los valores del estadístico t-Student para determinada probabilidad P y grados de libertad n.

CD9: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Varianza,* es una medida de dispersión de un grupo de datos con respecto a su media [En el mismo sentido del CD7 del SP10 del estadístico χ^2].

CD11: *Error estándar,* es una desviación estándar de la distribución de las medias.

Propiedades/proposiciones:

PP1: *Distribución normal,* en el mismo sentido que en la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student, en el mismo sentido que en la PP2 del significado parcial 2, precisando que se sustituye el estadístico z por el estadístico $t = z/\sqrt{n}$, donde n es el número de grados de libertad, y que cuando n' es grande la distribución t-Student tiende a la normalidad.

PP3: *Media,* $\bar{x} = \frac{1}{n'} S(x)$.

PP4: La raíz de $\frac{s^2}{n'} = \frac{1}{n'(n'-1)} S(x-\bar{x})^2$ sustituye a la desviación estándar s del estadístico z.

PP5: *Grados de libertad* como $n = n' - 1$.

PP6: *Estadístico t-Student*

$$t = \bar{x} / \sqrt{\frac{s^2}{n'}}$$

PP7: *Regla de decisión,* compara valor del estadístico calculado con el valor del estadístico t que se ha localizado en la tabla de probabilidad de la distribución t-Student con respecto a cierta P y n, y si el valor del estadístico calculado lo excede diremos que la diferencia es claramente significativa.

Procedimientos:

Determinar si la media de la muestra (de las diferencias entre los efectos de los medicamentos) es significativa, para lo cual es necesario calcular el valor del estadístico t-Student y buscar en las tablas de la probabilidad sí el valor del estadístico calculado se encuentra fuera del rango para cierta P. Esto se realiza bajo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico $t = \bar{x}/\sqrt{\frac{s^2}{n'}}$
    a. Calcular la media bajo la siguiente formula $\bar{x} = \frac{1}{n'}S(x)$, donde n' es el tamaño de la muestra.
    b. Encontrar el valor de $\sqrt{\frac{s^2}{n'}} = \sqrt{\frac{1}{n'(n'-1)}S(x-\bar{x})^2}$ para el cálculo de las desviaciones.
2. Calcular los grados de libertad mediante $n = n' - 1$.
3. Buscar en las tablas de la probabilidad de Fisher (1934), sí el valor del estadístico t-Student calculado se encuentra fuera del rango para cierta P con el numero de grados de libertad n determinado previamente.

Argumentos:

Se argumenta de acuerdo al contraste entre el valor teórico del estadístico t-Student, que se ha localizado en la tabla de probabilidad de la distribución t-Student con respecto a cierta P y n, y el valor del estadístico t-Student calculado.

Por ejemplo, si se considera una $P = 0.01$, "*para* $n = 9$*, se tiene la probabilidad de que solo un valor en cien excederá 3.250, por lo que la diferencia entre los resultados es claramente significativos*" (Fisher, 1934, p. 119).

4.2.2 Significado 2: Distribución t-Student

Situación/problema:

Como se comentó en la sección 3.3.2.1, Gosset buscaba juzgar la significancia de la media de una muestra en el contexto de muestras pequeñas y proporcionar tablas de probabilidad (Student, 1908a). Para lograrlo, era necesario determinar la ecuación que representaba la frecuencia de las desviaciones estándar de las muestras, y una vez que la determinó como $z = x/s$, fue necesario *conocer cómo se distribuía*. Este problema, sobre cómo se distribuía el estadístico z, constituyó el problema principal que detonó este significado.

Inicialmente fue Gosset quien abordó este problema, así como el del error probable del coeficiente de correlación, de forma empírica, principalmente; y posteriormente Fisher los retomó logrando su generalización. Para el primero de ellos (Student, 1908a), proporcionó una solución matemática a través de la representación geométrica, mientras que cuando trabajó en el problema del segundo trabajo de Gosset (Student, 1908b) formuló la generalización de la distribución del coeficiente de correlación.

Las prácticas de Gosset y de Fisher sobre la distribución de este estadístico se han descrito previamente en la sección 3.3.1. A continuación, analizamos los elementos de las configuraciones subyacentes a ambas prácticas.

4.2.2.1 Significado parcial 4 (SP4): La distribución empírica del estadístico z

Bajo el supuesto de que las muestras provenían de una distribución normal, Gosset consideró que x, el numerador del estadístico z, seguía una distribución normal. Y para encontrar la distribución de s recurrió a los siguientes *conceptos/definiciones*:

CD1: *Media,* en el mismo sentido que en el CD4 del significado parcial 2.

CD2: *Desviación estándar,* de acuerdo con lo expresado en el CD5 del significado parcial 2.

CD3: *Sesgo,* nos permite medir la asimetría de una distribución, en otras palabras, si se encuentra el mismo número de elementos a la derecha y a la izquierda de la distribución.

CD4: *Curtosis,* permite indicar la concentración de los elementos de la distribución.

Y a las siguientes *propiedades/proposiciones* de los momentos (ver sección 3.2.1.1):

- PP1: *Media,* es el primer momento y se encuentra con la suma de sus longitudes multiplicada por sus respectivas distancias de una línea recta paralela.
- PP2: *El cuadrado de la desviación estándar,* es la suma de sus longitudes multiplicada por los cuadrados de sus distancias.
- PP3: *Sesgo de la distribución,* es el tercer momento y corresponde a la suma de sus longitudes multiplicada por el cubo de sus distancias.
- PP4: *Curtosis,* es el cuarto momento y se encuentra multiplicando las longitudes por la cuarta potencia, esta medida de forma puede categorizarse en leptocúrtica (curtosis >0), platicúrtica (curtosis < 0) y mesocúrtica (curtosis =0).

Gosset aplicó *procedimientos* algebraicos orientados al cálculo de los primeros cuatro momentos de la distribución de s^2, primero sobre cero, donde $s^2 = 0$ y obtuvo:

$$M_1' = \mu^2 \frac{(n-1)}{n}, M_2' = \mu_2^2 \frac{(n-1)(n+1)}{n^2}, M_3' = \mu_2^3 \frac{(n-1)(n+1)(n+3)}{n}, M_4' = \mu_2^4 \frac{(n-1)(n+1)(n+3)(n+5)}{n}$$

Y posteriormente obtuvo los momentos de s^2 sobre su media:

$$M_2 = 2\mu_2^2 \frac{(n-1)}{n^2}, M_3 = 8\mu_2^3 \frac{(n-1)}{n^3}, M_4 = 12\mu_2^4 \frac{(n-1)(n+3)}{n^4}$$

Estas expresiones de los momentos constituyen, por su naturaleza, *propiedades/proposiciones* (PP5 y PP6, respectivamente) que Gosset utilizó para descubrir la curva que se ajusta a los momentos, expresando que la

PP7: *Curva de Pearson Tipo III* $y = Cx^p e^{-\gamma^x}$.

Y podría contemplarse como un ajuste para la distribución de s^2. Para obtener la ecuación de la curva de la distribución recurrió a la PP6 y realizó los siguientes *procedimientos*:

si $y = Cx^p e^{-\gamma^x}$, donde $\gamma = 2\frac{M_2}{M_3} = \frac{4\mu_2^2(n-1)n^3}{8n^2\mu_2^3(n-1)} = \frac{n}{2\mu_2}$, y $p = \frac{4}{\beta_1} - 1 = \frac{n-1}{2} = \frac{n-3}{2}$, entonces

$y = Cx^{(n-3)/2} e^{-nx/2\mu_2}$, cuya el área bajo la curva es $C \int_0^\infty x^{(n-3)/2}\, e^{-nx/2\mu_2} = I$. Nuevamente recurrió a las *propiedades/proposiciones* de los momentos (PP1, PP2, PP3 y PP4) y a *procedimientos* algebraicos para calcular los cuatro momentos de la ecuación de la curva y encontró que los momentos más grandes de la distribución coincidieron con los de s^2, y señaló que "*es probable que la curva encontrada represente la distribución teórica de* s^2" (Student, 1908a, p. 5). Además, consideró que la frecuencia de s es igual a la de s^2, por lo que obtuvo la expresión $y_2 = 2Cs^{(n-2)}e^{-ns^2/2\mu_2}$, por lo tanto $y = Ax^{(n-2)}e^{-nx^2/2\sigma^2}$, esta ecuación nos dará la distribución de frecuencias de las desviaciones estándar de las muestras.

Posteriormente, calculó la constante A para cuando n es tanto par como impar, por medio de la identificación del área de la curva (PP8) $Area = A\int_0^\infty x^{n-2}e^{-nx^2/2\sigma^2}\,dx$. Obtuvo las siguientes dos expresiones sobre dicha distribución de frecuencias, las cuales también constituyen las *propiedades/proposiciones* 9 y 10, respectivamente:

- PP9: $y = \frac{N}{(n-3)(n-5)\ldots 3.1}\sqrt{\frac{2}{\pi}}\left(\frac{n}{\sigma^2}\right)^{(n-1)/2} x^{n-2}e^{-nx^2/2\sigma^2}$ (cuando n es par)
- PP10: $y = \frac{N}{(n-3)(n-5)\ldots 4.2}\left(\frac{n}{\sigma^2}\right)^{(n-1)/2} x^{n-2}e^{-nx^2/2\sigma^2}$ (cuando n es impar),

donde N representa la frecuencia total.

Para mostrar que no existe correlación entre la media (la distancia entre la media de la muestra y la media de la población) y la desviación estándar de dicha muestra Gosset hace evidente el uso del siguiente *concepto/definición*:

CD5: *Correlación* refiere a la existencia de alguna relación entre dos variables, se mide por medio del coeficiente de correlación (ver CD6 del SP5 del estadístico χ^2)

Y de las siguientes *propiedades/proposiciones*:

- PP11: Cuadrado de la distancia $\mu^2 = \left(\frac{S(x_1)}{n}\right)^2$
- PP12: Cuadrado de la desviación estándar $s^2 = \frac{S(x_1^2)}{n} - \left(\frac{S(x_1)}{n}\right)^2$
- PP13: El valor medio de μ^2 como $m_1' = \frac{\mu_2}{n}$

Y retomó a $M_1' = \mu^2\frac{(n-1)}{n}$, de la PP5. También recurrió a *procedimientos* algebraicos para calcular $\mu^2 s^2$, operaciones para sumar todos los valores y dividirlo por el número de casos, obtuvo que $R_{\mu^2 s^2}\sigma_{\mu^2}\sigma_{s^2} + m_1 M_1 = \frac{\mu_4}{n^2} + \mu_2^2\frac{(n-1)}{n^2} - \frac{\mu_4}{n^3} - 3\mu_2^2\frac{(n-1)}{n^3}$, y consideró que si se tiene que $R_{\mu^2 s^2}$ es la correlación entre μ^2 y s^2 entonces $R_{\mu^2 s^2}\sigma_{\mu^2}\sigma_{s^2} + \mu_2^2\frac{(n-1)}{n^2} = \mu_2^2\frac{(n-1)}{n^3}\{3 +$

$n-3\} = \mu_2^2 \frac{(n-1)}{n^2}$, por lo tanto $R_{\mu^2 s^2}\sigma_{\mu^2}\sigma_{s^2} = 0$; aspecto que constituye el *argumento* central para establecer que no existe correlación entre μ^2 y s^2, probando así que no hay correlación entre la distancia de la media de la muestra y la desviación estándar de la muestra (el numerador y el denominador del estadístico z), que constituye la PP14 y es retomada a continuación.

Ahora Gosset estaba interesado en encontrar la ecuación que representara la distribución de frecuencia de las medias de muestras de tamaño n provenientes de una población normal (la distribución de $z = x/s$) y utiliza las siguientes *propiedades/proposiciones* para lograrlo:

- PP15: Se tiene que $y = \frac{C}{\sigma^{n-1}} s^{n-2} e^{-ns^2/2\sigma^2}$ como la ecuación de la distribución de s.
- PP16: Las medias de las muestras son distribuidas como $y = \frac{\sqrt{n}N}{\sqrt{2\pi}\sigma} e^{-nx^2/2\sigma^2}$.
- PP14: No existe correlación entre la distancia de la media de la muestra (x) y la desviación estándar de la muestra (s).
- PP17: Suponemos que x se mide en términos de s.
- PP18: Ecuación representante de la frecuencia de x $y_1 = \emptyset(x)$.
- PP19: Ecuación representante de la frecuencia de z $y_2 = \psi(z)$.

Así como el siguiente *procedimiento* que también utiliza como *argumento* para la ecuación de la distribución del estadístico z:

$$y_1 dx = y_2\, dz = y_2 \frac{dx}{s} \therefore y_2 = s y_1$$

por lo que $y = \frac{N\sqrt{ns}}{\sqrt{2\pi}\sigma} e^{-ns^2/2\sigma^2}$ es la ecuación representante de la distribución de z.

Para obtener la distribución de z, inició por indicar que la probabilidad de que s se encuentre entre s y $s+ds$ es $\frac{\int_s^{s+ds} \frac{C}{\sigma^{n-1}} s^{n-2} e^{-ns^2/2\sigma^2} ds}{\int_0^{\infty} \frac{C}{\sigma^{n-1}} s^{n-2} e^{-ns^2/2\sigma^2} ds}$, la cual representa a N de la ecuación de la distribución de z y conforma la PP20 necesaria para determinar la distribución de z con respecto a los valores de s que se encuentran entre s y $s+ds$, que es $y = \frac{\int_s^{s+ds} \frac{C}{\sigma^n} \sqrt{\frac{n}{2\pi}} s^{n-1} e^{-ns^2(1+z^2)/2\sigma^2} ds}{\int_0^{\infty} \frac{C}{\sigma^{n-1}} s^{n-2} e^{-ns^2/2\sigma^2} ds} = \frac{\sqrt{\frac{n}{2\pi}} \int_s^{s+ds} s^{n-1} e^{-ns^2(1+z^2)/2\sigma^2} ds}{\sigma \int_0^{\infty} s^{n-2} e^{-ns^2/2\sigma^2} ds}$. Indicó que sumando todos los valores de s se tiene que $y = \frac{\sqrt{\frac{n}{2\pi}}}{\sigma} \frac{\int_0^\infty s^{n-1} e^{-ns^2(1+z^2)/2\sigma^2} ds}{\int_0^\infty s^{n-2} e^{-ns^2/2\sigma^2} ds}$ es la ecuación de la distribución del estadístico z y la reduce a las ecuaciones:

$y = \frac{1}{2} \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \frac{5}{4} \cdot \frac{3}{2} (1+z^2)^{-n/2}$ (cuando n es impar).

$y = \frac{1}{\pi} \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \frac{4}{3} \cdot \frac{2}{1} (1+z^2)^{-n/2}$ (cuando n es par).

Las ecuaciones anteriores constituyen la propiedad/proposición PP21 y se pueden localizar también en la PP2 del significado parcial 2.

Retomó la PP21 para establecer el área bajo la curva y estableció que si $z = \tan\theta$ entonces $y = \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \text{etc.} \times \cos^2\theta$ (PP22), además $dz = \frac{d\theta}{\cos^2\theta}$ (PP23), por lo que para encontrar el área bajo la curva entre cualquier límite realizó los siguientes *procedimientos*:

$\frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \text{etc.} \times \cos^2\theta$

$= \frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \text{etc.} \left\{ \frac{n-3}{n-2} \int \cos^{n-4}\theta \, d\theta + \left[\frac{\cos^{n-3}\theta \sin\theta}{n-2} \right] \right\}$

$= \frac{n-4}{n-5} \cdot \frac{n-6}{n-7} \dots \text{etc.} \int \cos^{n-4}\theta \, d\theta + \frac{1}{n-3} \cdot \frac{n-4}{n-5} \dots \text{etc.} \left[\cos^{n-3}\theta \sin\theta \right]$

- PP23: Estableció que toda el área bajo la curva es igual a $\frac{n-2}{n-3} \cdot \frac{n-4}{n-5} \dots \text{etc.} \times \int_{-\pi/2}^{\pi/2} \cos^{n-2}\theta \, d\theta$.

Gosset procedió a evaluar la curva y proporcionó la tabla de probabilidad de la distribución de z, donde $\theta = -\frac{\pi}{2}$ y $\theta = \tan^{-1} z$.

4.2.2.2 Significado parcial 5 (SP5): Fundamentación matemática y generalización de la distribución

Fisher tomó como basé los documentos de Gosset (Student, 1908a, 1908b), y buscaba brindar una rigurosa prueba matemática que sustentara el trabajo de Gosset. Fisher partió interpretando que Gosset deseaba "*obtener una estimación justa de la precisión que se atribuye a la media de una muestra pequeña*" (Fisher, 1915, p. 507), señaló los procedimientos que realizó Gosset (Student, 1908a) para establecer que:

si $x_1, x_2, \dots, x_n$ son los miembros de una muestra, entonces, $n\bar{x} = x_1 + x_2 + \cdots + x_n$, y $n\mu^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2$.

Por lo que la frecuencia con la que el error cuadrático medio se encuentra en el rango $d\mu$, es proporcional a

$$\mu^{n-2} e^{-\frac{n\mu^2}{2\sigma^2}} d\mu$$

Sobre este resultado, Fisher señaló que "*aunque se llegó por métodos empíricos, se estableció casi fuera de toda duda razonable en el primero de los documentos de 'Student'. Sin embargo, es interesante notar que la forma se establece instantáneamente, cuando la distribución de la muestra se ve geométricamente*" (Fisher, 1915, p. 507).

Mientras que con respecto al trabajo de Gosset, sobre el error probable de los coeficientes de correlación (Student, 1908b), Fisher señala que el problema de mayor dificultad que abordó refiere a la distribución de frecuencias del coeficiente de correlación.

Fisher (1915), estableció que el problema de la distribución de frecuencias del coeficiente de correlación r, se puede resolver con concepciones de la geometría multidimensional, siempre y cuando la población puede representarse por una superficie normal; y que su objetivo era demostrar la forma general, para lo cual recurrió a los siguientes *conceptos/definiciones*:

CD1: *Media,* en el mismo sentido que en el CD4 del significado parcial 2.

CD2: *Desviación estándar,* de acuerdo con lo expresado en el CD5 del significado parcial 2.

CD3: *Sesgo,* según lo establecido en el CD3 del significado parcial 4.

CD4: *Curtosis,* el mismo sentido que en el CD4 del significado parcial 4.

CD5: *Correlación,* en el sentido del CD5 del significado parcial 4.

CD6: *Coeficiente de correlación,* es una función simétrica de las variables, con rangos entre ± 1 y es cero cuando las variables son independientes [En el mismo sentido del CD6 del SP5 del estadístico χ^2].

CD7: *Espacio de* n *dimensiones,* es un tipo de espacio geométrico donde se satisfacen los axiomas de Euclides de la geometría.

Y partió indicando las siguientes *propiedades/proposiciones*:

- PP1: La distribución de frecuencia de la población es especificada por $df = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{1-\rho^2}\left\{\frac{(x-m_1)^2}{2\sigma_1^2} - \frac{2(x-m_1)(y-m_2)}{2\sigma_1\sigma_2} + \frac{(y-m_2)^2}{2\sigma_2^2}\right\}} dx\, dy$, donde df es la probabilidad de que cualquier observación se encuentre dentro del rango $dx\, dy$.

- PP2: La probabilidad de que n pares se encuentren dentro de sus elementos especificados es

$$\frac{1}{\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^n} e^{-\frac{1}{1-\rho^2} \sum_1^n \left\{\frac{(x-m_1)^2}{2\sigma_1^2} - \frac{2(x-m_1)(y-m_2)}{2\sigma_1\sigma_2} + \frac{(y-m_2)^2}{2\sigma_2^2}\right\}} dx_1\, dy_1 \dots dx_n\, dy_n.$$

Fisher interpretó la PP2 como una distribución de densidad simple en $2n$ dimensiones, y para las variables x y y sustituyó las formulaciones estadísticas definidas por las siguientes *propiedades/proposiciones*:

- PP3: $n\bar{x} = \sum_1^n(x), n\bar{y} = \sum_1^n(y)$.
- PP4: $n\mu_1^2 = \sum_1^n(x-\bar{x})^2, n\mu_2^2 = \sum_1^n(y-\bar{y})^2$.
- PP5: $nr\mu_1\mu_2 = \sum_1^n(x-\bar{x})(y-\bar{y})$.

A continuación, examinó la interpretación de las PP3, PP4 y PP5 en el espacio generalizado, bajo los siguientes *procedimientos*:

Considerando que "el espacio de n dimensiones en el que se representan las variaciones de x, la media y el error cuadrático medio de n observaciones están determinadas por las relaciones de P, el punto que representa las n observaciones, con la línea $x_1 = x_2 = x_3 = \cdots = x_n$" (Fisher, 1915, p. 509).

Para ubicar la región donde se encuentra la perpendicular PM dibujada a partir de P utilizó la PP3 y obtuvo que $x_1 + x_2 + x_3 + \cdots + x_n = n\bar{x}$, al encontrarse esta línea con el punto M, donde se tiene que $x_1 = \bar{x}, x_2 = \bar{x}, \dots, x_n = \bar{x}$.

Además, utilizando la PP6: $PM^2 = (x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2$, estableció que la longitud de la perpendicular PM es $\mu_1\sqrt{n}$. Así como utilizó *lenguaje natural* y *simbólico* también lo expresó con una *representación gráfica*, como se puede ver en la Figura 4.6.

A partir de lo anterior, especificó a la expresión $C\mu_1^{n-2}d\mu_1 d\bar{x}$ como el elemento de volumen en este espacio de n dimensiones.

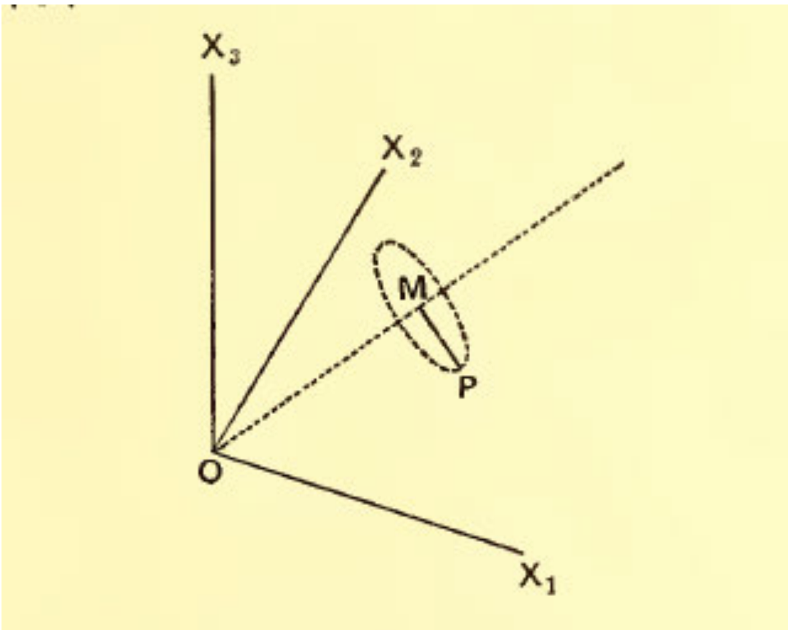


Figura 4.6. Longitud PM en n dimensiones (fuente: Fisher, 1915, p. 509)

PP7: un punto en el espacio de $2n$ dimensiones que está representado por n pares de observaciones debe ser tal que su proyección en el espacio de n dimensiones, en el que se representa x, se encuentre sobre una determinada esfera de radio $\mu_1\sqrt{n}$, y sobre el espacio en el que y se representa, en otra esfera de radio $\mu_2\sqrt{n}$.

Bajo la PP7, establece la siguiente expresión, donde a cada punto en la primera esfera debería corresponder un cierto punto en la segunda esfera, $\frac{x_1-\bar{x}}{y_1-\bar{y}} = \frac{x_2-\bar{x}}{y_2-\bar{y}} = \cdots = \frac{x_n-\bar{x}}{y_n-\bar{y}}$, pero dicha relación no se cumple para los n pares. Por lo anterior, Fisher opta por girar una de las esferas para ocupar el mismo espacio que la otra esfera, y que coincidan las líneas sobre las cuales son medidos x_1 y y_1, y los otros pares de coordenadas. Entonces, el coeficiente de correlación r mide el coseno del ángulo entre la radio y los dos puntos especificados por las observaciones.

PP8: Si tomamos una de las proyecciones como fija en cualquier punto de la esfera que tiene radio de μ_2, la región donde r se encuentra en el rango dr, es una zona de la otra esfera en $n - 1$ dimensiones, con radio $\mu_1\sqrt{n}\sqrt{1-r^2}$ y amplitud de $\frac{\mu_1\sqrt{n}dr}{\sqrt{1-r^2}}$.

Utilizó la PP8 para establecer el volumen de la región donde r se encuentra en el rango dr es ${\mu_1}^{n-2}(1-r^2)^{\frac{n-4}{2}}dr$.

Posteriormente Fisher, procedió a la simplificación de la PP2 y eliminó cualquier factor que no implicara a r, para llevarlo a cabo realizó *procedimientos algebraicos* y obtuvo como distribución de frecuencias a la expresión $(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-2}}{\partial r^{n-2}}\left(\frac{\theta}{\sin\theta}\right)dr$, expresión que constituye la PP9.

Fisher añadió, que esta expresión se podía aplicar a los pequeños valores pares de n y que era adecuada para el calculo de los momentos. Y utilizó la PP9 para indicar que cuando $n = 2$ la ordenada de la curva, con abscisa r, es $\frac{\theta}{(1-r^2)\sin\theta}$, que es se vuelve hiperbólica en las cercanías de -1 y $+1$. A partir de esto, dilucidó las siguientes expresiones que se vuelven *propiedades/proposiciones*:

- PP10: La relación de las áreas infinitas incluidas con las asíntotas de la curva anterior es $\frac{\cos^{-1}\rho}{\cos^{-1}(-\rho)}$, donde de ρ depende la proporción con que ocurren los valores de r.

- PP11: El valor medio de varias observaciones es $\frac{\sin^{-1}\rho}{\frac{\pi}{2}}$.

A partir de las propiedades anteriores establece que cuando $n = 4$, la curva toma la forma $\frac{1}{\sin^3\theta}(\theta - 3\cot\theta + 3\theta\cot^2\theta)$ y para calcular el número de valores que se encuentran dentro de cualquier rango, la integral, $\frac{1}{\sin^2\theta}(1-\theta\cot\theta)$.

Posteriormente Fisher (1915), indicó que el procedimiento directo de integración por partes aplicado a expresiones como la siguiente que constituye la PP12:

$\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}dr$, y $\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}r\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}dr$, cuando n es par, simplemente se introduce las sumas y diferencias de los términos $\frac{\partial^p}{\partial r^p}\frac{\theta^2}{2}$ en los extremos, donde r es -1 o $+1$. Y realiza elsiguiente *procedimiento* de integración por partes para cuando $n = 6$:

$$\int_{-1}^{+1}(1-r^2)\frac{\partial^5}{\partial r^5}\frac{\theta^2}{2}dr = \left[(1-r^2)\frac{\partial^4}{\partial r^4}\frac{\theta^2}{2}\right]_{-1}^{+1} + \left[2r\frac{\partial^3}{\partial r^3}\frac{\theta^2}{2}\right]_{-1}^{+1} - \left[2\frac{\partial^2}{\partial r^2}\frac{\theta^2}{2}\right]_{-1}^{+1}$$

$$= 2 \times suma\ de\ los\ valores\ extremos\ de\ \frac{\rho^3}{\sin^3\theta}(\theta - 3\cot\theta + 3\theta\cot^2\theta)$$
$$- 2 \times la\ diferencia\ de\ los\ valores\ extremos\ de\ \frac{\rho^2}{\sin^2\theta}(1 - \theta\cot\theta)$$

Y señaló que si $\rho = \sin\alpha$, de modo que los valores extremos de θ son $\frac{\pi}{2} - \alpha$ y $\frac{\pi}{2} + \alpha$, las sumas y las diferencias pueden expresarse fácilmente en términos de α, las primeras series de sumas y diferencias las presentó en una tabla e indicó que la más simple puede expresarse de la forma $\frac{\pi}{2}\sin^p\alpha\left(\frac{\partial}{\cos\alpha\partial\alpha}\right)^p\alpha$.

Sin embargo, Fisher deseaba obtener expresiones generales para dichas integrales en términos de n y ρ, y evaluarlos para cuando n es impar. Para lo cual introdujo la cantidad $\emptyset$, como vemos en la PP13.

PP13: $\cos\emptyset = \cos\theta - k$.

PP14: Teorema de Taylor. Sea $k \geq 1$ un entero y la función $f : R \to R$ diferenciable k veces en el punto $a \in R$. Entonces existe una función $h_k : R \to R$ tal que $f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^k(a)}{k!}(x-a)^k + h_k(x)(x-a)^k$ y $\lim_{x\to a} h_k(x) = 0$.

Retomó la PP13 e indicó que cuando k es o suficientemente pequeña, podemos utilizar la PP14 para expandir a $\emptyset^2$, de modo que realizó los siguientes *procedimientos*:

$$\frac{\emptyset^2}{2} = \frac{\theta^2}{2} + k\frac{\partial}{\sin\theta\partial\theta}\frac{\theta^2}{2} + \frac{k^2}{\lfloor 2}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2} + \cdots$$

Ahora, si $k = \rho h\sqrt{1-r^2}$,

entonces $\frac{\emptyset^2}{2} = \frac{\theta^2}{2} + \rho h\sqrt{1-r^2}\frac{\partial}{\sin\theta\partial\theta}\frac{\theta^2}{2} + \frac{\rho^2 h^2(1-r^2)}{\lfloor 2}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2} + \cdots$,

y diferenciando dos veces con respecto a h $\rho^2(1-r^2)\left(\frac{\partial}{\sin\emptyset\partial\emptyset}\right)^2\frac{\emptyset^2}{2} = \rho^2(1-r^2)\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2} + h\rho^3(1-r^2)^{\frac{3}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^3\frac{\theta^2}{2} + \cdots$, de donde, dividiendo por $(1-r^2)^{\frac{3}{2}}$, obtenemos $\frac{\rho^2}{\sqrt{1-r^2}}\left(\frac{\partial}{\sin\emptyset\partial\emptyset}\right)^2\frac{\emptyset^2}{2} = \frac{\rho^2}{(1-r^2)^{\frac{1}{2}}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^2\frac{\theta^2}{2} + h\rho^3\left(\frac{\partial}{\sin\theta\partial\theta}\right)^3\frac{\theta^2}{2} + \frac{\rho^4 h^2}{\lfloor 2}(1-r^2)^{\frac{1}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^4\frac{\theta^2}{2} + \cdots$,

así que $\int_{-1}^{+1} r^p\,(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}dr$

se puede obtener multiplicando por $n-3$ el coeficiente de h^{n-3} en $\rho^2\int_{-1}^{1}\frac{r^p dr}{\sqrt{1-r^2}}\frac{1-\emptyset\cot\emptyset}{\sin^2\emptyset}$, expresión que constituye la PP15, también estableció la PP16 mediante la ecuación $\cos\emptyset = \cos\theta - \rho h\sqrt{1-r^2} = -\rho\left(r + h\sqrt{1-r^2}\right)$.

Añadió que el objetivo también se podría lograr por medio de la evaluación de la integral $\rho^2\int_{-1}^{1}\frac{r^p dr}{1-r^2}\frac{\emptyset}{\sin\emptyset} - \frac{\theta}{\sin\theta}$ (PP17). Mientras que la cantidad $\emptyset$ es determinada mediante la ecuación de la PP16.

PP18: Si $r = \sin\beta$ y $h = \tan\epsilon$, entonces $\cos\theta = -\rho\sin\beta$.

Con lo establecido ahora en la PP18 retomó la PP16 y se reformula como la PP19 donde $\cos\emptyset = -\rho\sqrt{1-h^2}\sin(\beta+\epsilon) = -\rho\sqrt{1-h^2}\sin\beta'$, donde r pasa de -1 a $+1$, β pasa de $-\frac{\pi}{2}$ a $+\frac{\pi}{2}$, θ desde $\frac{\pi}{2}-\alpha$ a $\frac{\pi}{2}+\alpha$, β' desde $-\frac{\pi}{2}+\epsilon$ a $\frac{\pi}{2}$ y de ahí a $\frac{\pi}{2}+\epsilon$, y $\emptyset$ desde $\frac{\pi}{2}-\alpha$ a $\frac{\pi}{2}+\alpha'$ y de vuelta a $\frac{\pi}{2}+\alpha$. Donde $\alpha' = \rho\sqrt{1-h^2}$, $\emptyset$ oscila de la misma forma que θ con una amplitud un poco mayor, y ligeramente por adelantado con respecto a la fase.

Después procedió a reducir la expresión $\rho^2\int_{-1}^{1}\frac{1-\emptyset\cot\emptyset}{\sin^2\emptyset}\frac{dr}{\sqrt{1-r^2}}$ por medio de *procedimientos algebraicos* y obtuvo que $\rho^2\int_{-1}^{1}\frac{1-\emptyset\cot\emptyset}{\sin^2\emptyset}\frac{dr}{\sqrt{1-r^2}} = \frac{\pi\tan^2\alpha}{1-h\tan\alpha}$.

A partir de la expresión anterior, se deduce la forma general como $\int_{-1}^{+1}(1-r^2)^{\frac{n-4}{2}}\frac{\partial^{n-1}}{\partial r^{n-1}}\frac{\theta^2}{2}dr = \lfloor n-3\,\pi\tan^{n-1}\alpha$, que constituye la PP20. Y a partir de la PP20 estableció que la expresión para determinar la frecuencia absoluta df con la r se encuentra en el rango dr es $\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi\lfloor n-3}(1-r^2)^{\frac{n-4}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-2}\frac{\theta}{\sin\theta}dr$.

Sin embargo, no encontró cómo integrar otras expresiones del tipo de la PP15, por lo que procedió a determinar la expresión general para el segundo momento mediante una fórmula de reducción. Para ello, llevó a cabo *procedimientos de integración por partes* y obtuvo, cuando n es par, la expresión $I_{n.2} = I_{n.0} - \pi\lfloor n - 2 \int_0^{\alpha} \tan^{n-2} x dx$, e indicó que esta expresión bien puede sostenerse cuando n es impar.

No obstante, Fisher señalaba que el uso del coeficiente de correlación como variable independiente de estas curvas de frecuencia es en algunos aspectos altamente insatisfactorio, presentando dificultades cuando se tienen grandes valores de r y de n, por lo que buscó si las transformaciones le proporcionarían variables más naturales para tratar sus fórmulas y las expresó como $t = \tan\beta = \frac{r}{\sqrt{1-r^2}}$ y $\tau = \tan\alpha = \frac{\rho}{\sqrt{1-\rho^2}}$ (PP21).

Ahora la expresión de la curva de frecuencias $(1-r^2)^{\frac{n-4}{2}} \left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1} \frac{\theta^2}{2} dr$, se puede expresar como $\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1} \frac{\theta^2}{2} \frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$(PP22), y el rango de la curva se extiende de $-\infty$ a $+\infty$. También indicó que en caso de que $r = 0$ la frecuencia se expresa como $\frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$ y que las curvas con idénticas que las encontradas por Gosset para z (Student, 1908a) con las cuales realizó las tablas de probabilidad de la distribución.

Posteriormente Fisher, obtuvo los cuatro momentos para las curvas de frecuencia, por medio de las expresiones $\int_{-\infty}^{\infty} \left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1} \frac{\theta^2}{2} \frac{dt}{(1+t^2)^{\frac{n-1}{2}}}$ y $\int_{-\infty}^{\infty} \left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1} \frac{\theta^2}{2} \frac{tdt}{(1+t^2)^{\frac{n-1}{2}}}$, que surgen de la PP22. Los momentos que obtuvo se presentan a continuación:

Primer momento $\bar{t} = \frac{n-2}{n-3} \frac{\rho}{\sqrt{1-\rho^2}} = \frac{n-2}{n-3}\tau$.

Segundo momento $\bar{t}^2 + \sigma^2 = \frac{1}{n-4}\{1 + (n-1)\tau^2\}$ y $\sigma^2 = \frac{1}{n-4}\left\{1 + \tau^2 + \frac{(n-2)}{(n-3)^2}\tau^2\right\}$.

Tercer momento $\sqrt{\beta_1}\sigma^3 = \frac{(n-2)\tau}{(n-3)(n-4)(n-5)}\left\{3(1+\tau^2) + \frac{2\tau^2(n-1)}{(n-3)^2}\right\}$.

Cuarto momento $\beta_2\sigma^4 = \frac{3}{(n-4)(n-6)}\left\{(1+\tau^2)^2 + \frac{6(n-2)\tau^2}{(n-3)(n-5)}(1+\tau^2) + \frac{6(n-2)(3n^2-11n+12)\tau^4}{(n-3)^4(n-5)}\right\}$.

Fisher (1915), incluyó tablas donde se podía observar los cambios en la forma de las curvas. Finalmente realizó una aproximación de r y la relación que existe entre t y τ. Para establecerlas partió de la expresión $(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1}\frac{\theta^2}{2}dr$ y realizó los siguientes *procedimientos*:

Buscó el valor de ρ para el cual esta cantidad es un máximo, y así obtener la ecuación $\frac{\partial}{\partial\rho}\left\{(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1}\frac{\theta^2}{2}\right\}=0$

Donde $\int_0^\infty \frac{dx}{(\cos hx+\cos\theta)^{n-1}}=\frac{1}{\lfloor n-1}\left(\frac{\partial}{\sin\theta\partial\theta}\right)^{n-1}\frac{\theta^2}{2}$

Tenemos que $\int_0^\infty \frac{\partial}{\partial\rho}\left\{(1-\rho^2)^{\frac{n-1}{2}}\frac{dx}{(\cos hx+\cos\theta)^{n-1}}\right\}=0$, y por medio de una simplificación obtuvo que $\int_0^\infty \frac{dx}{(\cos hx+\rho r)^n}(r-\rho\cos h\,x)=0$.

Declaró que el integrando es insignificante, salvo cuando x es muy pequeño, y estableció $1+\frac{x^2}{2}$ para $\cos h\,x$; y $(1-\rho r)^n e^{\frac{nx^2}{2(1-\rho r)}}$ para $(\cos h\,x+\rho r)^n$.

Entonces $r\int_0^\infty e^{-\frac{nx^2}{2(1-\rho r)}}dx=\rho\int_0^\infty\left(1+\frac{x^2}{2}\right)e^{-\frac{nx^2}{2(1-\rho r)}}dx$, y en consecuencia obtuvo la aproximación de $r=\rho\left(1+\frac{1+r^2}{2n}\right)$ y la correspondiente relación entre t y τ que es determinada por $t=\tau\left(1+\frac{1}{2n}\right)$.

Concluyó argumentando que, "*ahora es evidente que el valor más probable de la correlación será en general menor que el observado, pero la diferencia será solo la mitad de la sugerida por la media,* $\bar{t}$*"(*Fisher, 1915, p. 521).

4.2.3 Significado 3: Prueba t-Student para dos muestras

La prueba t-Student para dos muestras determina si las medias de dos poblaciones distribuidas de forma normal son iguales, en otras palabras, si no existe diferencia significativa entre las medias.

4.2.3.1 Significado parcial 6 (SP6): El método de Edgeworth (1885)

En este significado se retoma el ejemplo 15 que se presentó en la página 167. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Al realizar su método de análisis, Edgeworth utilizó *lenguaje natural* en el planteamiento, la explicación de cada uno de los pasos del método y la interpretación de los resultados. Utilizó *representación tabular* para mostrar las tasas de mortalidad de diferentes lugares y años. También hizo uso de la *representación simbólica* al resolver el problema que planteó, por ejemplo, para realizar el cálculo de la fluctuación.

Situación/problema:

En 1885, Edgeworth buscaba establecer un método que le permitiera eliminar el azar de la teoría de los errores, para lo que desarrolló un método basado únicamente en la variación interna de los datos. Para mostrar cómo funciona su método tomó datos sobre la tasa de mortalidad de la oficina de registros, estos datos eran sobre seis ciudades inglesas por un periodo de ocho años. Calculó para cada columna y cada fila la media la fluctuación.

Conceptos/definiciones:

CD1: *Error probable,* en el mismo sentido que CD1 del significado parcial 1.

CD2: *Población,* en el mismo sentido que CD3 del significado parcial 2.

CD3: *Media,* se entiende en concordancia con lo explicitado en el CD4 del significado parcial 2.

CD4: *Fluctuación,* es una medida de la variación de una serie de datos con respecto a otra serie.

Propiedades/proposiciones:

PP1: *Ley del error,* en el mismo sentido que PP1 del significado parcial 1.

PP2: *Distribución normal,* en el mismo sentido que PP1 del significado parcial 1.

PP3: *Error probable normal,* en el mismo sentido que PP6 del significado parcial 1.

PP4: *Media* $\bar{x} = \frac{1}{n}\sum x_i$, de acuerdo con en el PP4 del significado parcial 2.

PP5: Se^2 es la suma de las diferencias al cuadrado.

PP6: Fluctuación $2\frac{Se^2}{n}$, donde n es la sumatoria de las diferencias.

Procedimientos:

Encontrar la variación de las tasas de defunción de seis ciudades inglesas en varios años, mediante el método de análisis de Edgeworth. Para rescatar la serie de acciones que es necesario realizar nos vamos a centrar en lo que trabajó Edgeworth en la tabla 26, sobre el análisis de mortalidad de dos poblaciones.

1. En la primer y segunda columna de la tabla se colocan las tasas de mortalidad con las que se va a trabajar.
2. Calcular la media para la columna 1($\bar{x}$) y para la columna 2 ($\bar{y}$).
3. Calcular las diferencias de la segunda columna con respecto a la primera ($y_i - x_i$).
4. Calcular las diferencias corregidas (e), para lo que se debe restar de los valores obtenidos en la acción 2 ($y_i - x_i$) la diferencia entre la media de la primera y la segunda columna ($\bar{y} - \bar{x}$). $e = (y_i - x_i) - (\bar{y} - \bar{x})$.
5. Obtener el cuadrado de cada una de las diferencias corregidas e^2.
6. Calcular el límite superior, mediante la media de las diferencias corregidas e^2.
7. Calcular la fluctuación mediante la formula $2\frac{Se^2}{n}$, donde Se^2 indica la sumatoria de las diferencias corregidas al cuadrado y n es la suma de las diferencias obtenidas en la acción 3 (sin importar el signo).
8. Comparar el valor del límite con el de la fluctuación.

Para este ejemplo e intentar explicar el método de análisis de Edgeworth, también se realizaron estos procedimientos:

1. Obtener la fluctuación por fila $C^2 + C_t^2$.
2. Calcular el valor medio de $C^2 + C_t^2$.
3. Calcular la fluctuación de la fila marginal por medio de $\frac{C^2}{6} + \alpha C_t^2$.

4. Obtener la fluctuación general de la tabla por fila, para el ejemplo abordado se tiene $C^2 = \frac{\alpha 190-146}{\alpha - \frac{1}{6}}$.

5. Realizar las acciones anteriores de manera similar con las columnas y los marginales verticales. Ahora tenemos $C^2 + C_p^2$.

6. Calcular los límites superiores para la fluctuación por filas y por columnas.

Argumentos:

Se argumentó con respecto a la fluctuación que se obtuvo, tanto para el caso de la tasa de mortalidad de dos poblaciones a través del tiempo o para el caso más amplio de la tasa de mortalidad de seis ciudades con respecto a 8 años. Además, se señaló cuál es el límite superior para la fluctuación buscada.

4.2.3.2 Significado parcial 7 (SP7): Significancia de la diferencia de medias de dos muestras con diferente tamaño con el estadístico t-Student

En este significado se retoma el ejemplo 18 que se presentó en la página 186 sobre la prueba t-Student de medias de dos muestras de diferente tamaño. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de significancia, Fisher recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos del número de bacterias y el valor de la media para cada columna. También hizo uso de la *representación simbólica* para presentar los resultados.

Situación/problema:

Fisher reconoció que en el trabajo experimental cada vez eran más necesarias las pruebas sobre si existe una diferencia significativa entre las medias de muestras de diferente tamaño. Para realizar esa prueba se auxilió del estadístico t-Student y su distribución.

Para ilustrar la aplicación de esta prueba, Fisher (1925b) trabajó con datos de un experimento de electro-cultivo realizado en 1922 en Rothamsted. En estos experimentos se medían el número de macollos de la planta de cebada. Por una parte, estaban aquellas plantas que recibían una descarga eléctrica de alta tensión y, por otra parte, las que se encerraron en una jaula de alambre con conexión a tierra.

Conceptos/definiciones:

CD1: *Experimento,* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra,* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población,* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media,* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar,* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student, de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad,* de acuerdo con lo establecido en el CD7 del significado parcial 3, precisando que en este caso se trata de una diferencia de medias, entonces se podría expresar como una medida que la diferencia de las medias se encuentre fuera del rango $\pm t$.

CD8: *Distribución t-Student,* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD9: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Tamaño muestral,* es el número de objetos o sujetos que componen a la muestra, es decir, el número de observaciones del subconjunto de la población.

CD11: *Error estándar,* en un sentido similar al expresado en el CD11 del significado parcial 3 pero en vez de tratarse con respecto a la media, hora es con respecto a la diferencia de medias, es decir, se entiende como la desviación estándar de la distribución de la diferencia entre las medias.

Propiedades/proposiciones:

PP1: *Distribución normal,* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student, de acuerdo con la PP2 del significado parcial 3.

PP3: *Media para cada muestra,* en un sentido similar que la PP3 del significado parcial 2, pero ahora tendríamos $\bar{x}_1$ y $\bar{x}_2$.

PP4: *Las desviaciones al cuadrado* S_1 y S_2 se calculan por medio de $\mathrm{S}_i = S(x - \bar{x})^2$.

PP5: *Desviación estándar* $s = \sqrt{\frac{\frac{(S_1+S_2)(n_1+n_2)}{n_1+n_2-2}/n_1}{n_2}}$.

PP6: *Grados de libertad,* como $n = n_1 + n_2 - 2$, donde n es el tamaño de la muestra.

PP7: *Estadístico t-Student*

$$t = \frac{(\bar{x}_1 - \bar{x}_2)\sqrt{n_1 + n_2 - 2}}{\sqrt{S_1 + S_2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

Procedimientos:

Determinar si la diferencia entre las medias de las muestras pequeñas es significativa por medio del estadístico t-Student. Para realizarlo se llevó acabo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico $t = \frac{(\bar{x}_1-\bar{x}_2)\sqrt{n_1+n_2-2}}{\sqrt{S_1+S_2}} \sqrt{\frac{n_1 n_2}{n_1+n_2}} = \frac{\bar{x}_1-\bar{x}_2}{s}$
    a. Calcular la media por medio de $\bar{x} = \frac{1}{n}\sum x_i$, tanto para la muestra 1 como para la muestra 2.
    b. Calcular la diferencia de las medias de las muestras $(\bar{x}_1 - \bar{x}_2)$.
    c. Calcular el valor de la desviación estándar para cada muestra por medio de $s = \sqrt{\frac{\frac{(S_1+S_2)(n_1+n_2)}{n_1+n_2-2}/n_1}{n_2}}$.
    d. Obtener la razón de la diferencia de medias entre la desviación estándar $\frac{(\bar{x}_1-\bar{x}_2)}{s}$.
2. Determinar los grados de libertad mediante $n = n_1 + n_2 - 2$.

3. Buscar en las tablas de la probabilidad de Gosset (Student, 1925) la probabilidad de ocurrencia para el estadístico t-Student calculado y con los grados de libertad de la prueba. O en Fisher (1934), sí el valor del estadístico t-Student calculado se encuentra fuera del rango para cierta P con el número de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a la probabilidad de que exista una diferencia mayor que la expresada por el estadístico t-Student calculado. En este problema Fisher (1934) proporcionó el siguiente argumento:

Por ejemplo:

> Este valor [$t = 3.054$] excede la posibilidad de 41 veces en $10{,}000$. En otras palabras, una diferencia, positiva o negativa, mayor que la observada tiene la probabilidad de ocurrir únicamente alrededor de 8 veces en $1{,}000$ experimentos. Por lo tanto, la diferencia debe ser juzgada entonces como significativa. Las dos series son definitivamente diferentes en cuanto a macollaje; sin embargo, la posibilidad de que la diferencia en la variabilidad, así como la diferencia en la media, hayan contribuido al resultado no está excluida por esta prueba. (Fisher 1925b, p. 95)

Reflexión:

De acuerdo con Fisher (1934), la prueba de significancia de la diferencia de medias era igualmente aplicable si se trataba de muestras del mismo tamaño o de diferente tamaño. Sólo había que hacer una pequeña modificación al estimar el error, combinando las sumas al cuadrado de las dos muestras. Siguiendo la misma lógica, señaló que incluso se podía aplicar a los coeficientes de regresión cuando las series de valores de la variable independiente no son idénticas.

4.2.3.3 Significado parcial 8 (SP8): Significancia de la diferencia de medias de dos muestras con igual tamaño con el estadístico t-Student

En este significado se retoma el ejemplo 19 que se presentó en la página 188. Fisher proporcionó dos métodos para solucionar este problema, en este significado parcial analizamos una de ellos y el segundo método lo analizaremos en el SP9. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de significancia, Fisher recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos del número de bacterias y el valor de la media para cada columna. También hizo uso de la *representación simbólica* al realizar los cálculos para obtener el valor del estadístico t-Student.

Situación/problema:

Debido al trabajo de campo experimental que Fisher desarrollaba en Rothamsted, reconoció que en el trabajo experimental cada vez eran más necesarias las pruebas sobre si existe una diferencia significativa entre las medias de muestras diferentes. Para realizar esa prueba se auxilió del estadístico t-Student y su distribución.

Para ilustrar la aplicación de esta prueba. Fisher (1934) trabajó con datos de H. G. Thornton, sobre el cambio en el número de bacterias en un cultivo por medio de muestras que fueron tomadas con cuatro métodos y en dos momentos distintos.

Conceptos/definiciones:

CD1: *Experimento,* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra,* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población,* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media,* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar,* en el sentido del CD5 del significado parcial 2.

CD6: *Desviación estándar agrupada,* es un método para calcular la desviación estándar agrupando las sumas de cuadrados de las dos muestras y dividiendo por el número total de los grados de libertad.

CD7: *Estadístico* t-Student, de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD8: *Probabilidad,* de acuerdo con lo establecido en el CD7 del significado parcial 7.

CD9: *Distribución t-Student,* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD10: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD11: *Tamaño muestral,* en el mismo sentido que el CD10 del significado parcial 7.

CD12: *Varianza,* de acuerdo a lo señalado en el CD10 del significado parcial 3.

CD13: *Error estándar,* en el sentido que el CD11 del significado parcial 7.

Propiedades/proposiciones:

PP1: *Distribución normal,* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student, de acuerdo con la PP2 del significado parcial 3.

PP3: *Medias* $\bar{x} = \frac{1}{n_1+1} S(x),\ \bar{x}' = \frac{1}{n_2+1} S(x')$.

PP4: *Varianza* $s^2 = \frac{1}{n_1+n_2}\{S(x-\bar{x})^2 + S(x'-\bar{x}')^2\}$.

PP5: *Grados de libertad* $n = n_1 + n_2$, donde $n_i = n_i' - 1$. Expresado de otra forma, $n = n_1' + n_2' - 2$, lo cual concuerda con lo expresado en la PP5 del significado parcial 7.

PP6: *Estadístico t-Student*

$$t = \frac{\bar{x}-\bar{x}'}{s}\sqrt{\frac{(n_1+1)(n_2+1)}{n_1+n_2+2}} = \frac{\bar{x}-\bar{x}'}{\sqrt{s^2\left(\frac{2}{n}\right)}}$$

PP7: *Regla de decisión,* compara la probabilidad obtenida a partir del valor del estadístico calculado, con el límite preestablecido como desviación significativa ($P = 0.05$) y si la posibilidad de exceder tal valor es menor al límite preestablecido diremos que las desviaciones de las expectativas son claramente significativas. [En el mismo sentido del PP6 del SP3 del estadístico χ^2].

Procedimientos:

Para determinar si la diferencia entre las medias de las muestras pequeñas con igual tamaño es significativa por medio del estadístico t-Student se realizaron las siguientes acciones:

1. Encontrar el valor del estadístico $t = \frac{\bar{x}-\bar{x}\prime}{s}\sqrt{\frac{(n_1+1)(n_2+1)}{n_1+n_2+2}}$
   a. Calcular las medias de las muestras bajo las siguientes formulas $\bar{x} = \frac{1}{n_1+1}S(x)$ y $\bar{x}' = \frac{1}{n_2+1}S(x')$, donde $n_i = n_i' - 1$ y n_i' es el tamaño de la muestra i.
   b. Calcular el valor de la desviación estándar por medio de $s = \sqrt{\frac{1}{n_1+n_2}\{S(x-\bar{x})^2 + S(x'-\bar{x}')^2\}}$.
2. Determinar los grados de libertad mediante $n = n_1 + n_2$.
3. Buscar en las tablas de la probabilidad de Fisher (1934), sí el valor del estadístico t-Student calculado se encuentra fuera del rango para cierta P con el numero de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a si la diferencia de medias es significativa, para lo cual fue necesario calcular el valor del estadístico t-Student y buscar en las tablas de la probabilidad sí el valor del estadístico calculado se encuentra fuera del rango para cierta P dado los grados de libertad. En este problema Fisher (1934) proporcionó el siguiente argumento:

> Si […] tratamos a las dos series por separado, encontramos $\bar{x} - \bar{x}' = +10.775$, $\frac{1}{2}s^2 = 2.188$, $t = 7.285$, $n = 6$; esto no es solo un valor más grande de n sino un valor más grande de t, que ahora está mucho más allá del rango de la tabla, que muestra que P es extremadamente pequeño. (Fisher, 1934, p. 125)

Reflexión:

Aunque hoy en día este problema se trataría de acuerdo a lo que exponemos en el SP9, Fisher (1934) también lo resolvió por medio de lo que hoy se conoce como el método de la prueba t-Student para dos muestras independientes, ya que para ese momento no se tenía esta diferencia al aplicar la prueba de significancia de medias de dos muestras pequeñas. Además, Fisher comparó los resultados obtenidos por ambos métodos y en ambos casos las diferencias de las medias de las dos muestras no eran significativas.

4.2.3.4 Significado parcial 9 (SP9): Significancia de diferencia de medias con el estadístico t-Student

En este significado se retoma el ejemplo 20 que se presentó en la página 189. Como se mencionó en el SP8, Fisher (1934) resolvió este problema por dos métodos, los elementos primarios que a continuación se describen son los involucrados en el desarrollo de esta actividad bajo el segundo método.

Elementos lingüísticos:

Para realizar la prueba de significancia Fisher, recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos del número de bacterias y el valor de la media para cada columna. También hizo uso de la *representación simbólica* al realizar los cálculos para obtener el valor del estadístico t-Student.

Situación/problema:

Debido al trabajo de campo experimental que Fisher desarrollaba en Rothamsted, reconoció que en el trabajo experimental cada vez eran más necesarias las pruebas sobre si existe una diferencia significativa entre las medias de muestras diferentes. Para realizar esa prueba se auxilió del estadístico t-Student y su distribución.

Para ilustrar la aplicación de esta prueba Fisher (1934), trabajó con datos de H. G. Thornton, sobre el cambio en el número de bacterias en un cultivo por medio de muestras que fueron tomadas con cuatro métodos y en dos momentos distintos.

Conceptos/definiciones:

CD1: *Experimento,* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra,* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población,* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media,* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar,* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student, de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad,* en un sentido similar a lo establecido en el CD7 del significado parcial 7, específicamente se interpreta como una medida de ocurrencia de que la media de las diferencias se encuentre fuera del rango $\pm t$.

CD8: *Distribución t-Student* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD9: *Significancia,* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Tamaño muestral,* en el mismo sentido que el CD10 del significado parcial 7.

CD11: *Varianza,* de acuerdo a lo señalado en el CD10 del significado parcial 3.

CD12: *Error estándar,* en el sentido que el CD11 del significado parcial 7.

Propiedades/proposiciones:

PP1: *Distribución normal,* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student, de acuerdo con la PP2 del significado parcial 3.

PP3: *Media de la diferencia* $\bar{x} = \frac{1}{n+1} S(x)$ que es igual a lo expresado en la PP3 del significado parcial 3.

PP4: *Las desviaciones* $\frac{s^2}{n'}$ en el mismo sentido de la PP4 del significado parcial 3.

PP5: *Grados de libertad* n, de acuerdo con lo expresado en la PP5 del significado parcial 3.

PP6: *Estadístico t-Student* se puede entender en el mismo sentido que la PP6 del significado parcial 3.

$$t = \bar{x} \Big/ \sqrt{\frac{s^2}{n'}}$$

Nota: Es importante recordar, que tanto la media como la desviación estándar no corresponde a las muestras sino a la diferencia de las muestras. Entonces podríamos expresar el estadístico como: $t = \frac{\bar{x}_D - \mu_0}{s_D / \sqrt{n}}$.

PP7: *Regla de decisión,* según lo expresado en el PP7 del SP8.

Procedimientos:

Determinar si la diferencia entre las medias de las muestras pequeñas es significativa por medio del estadístico t-Student. Por medio de la siguiente concatenación de acciones:

4. Encontrar el valor del estadístico $t = \bar{x}/\sqrt{\frac{s^2}{n'}}$
    a. Calcular la media de las diferencias de las muestras por medio de $\bar{x} = \frac{1}{n+1}S(x)$, donde $n = n' - 1$
    b. Calcular las desviaciones bajo la expresión $\frac{s^2}{n'} = \frac{1}{n'(n'-1)}S(x - \bar{x})^2$
5. Determinar los grados de libertad mediante $n = n' - 1$.
6. Buscar en las tablas de la probabilidad de Fisher (1934), sí el valor del estadístico t-Student calculado se encuentra fuera del rango para cierta *P* con el numero de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a si la diferencia de medias es significativa; para lo cual era necesario calcular el valor del estadístico t-Student y buscar en las tablas de la probabilidad sí el valor del estadístico calculado se encuentra fuera del rango para cierta *P* dado los grados de libertad, o cuál es la probabilidad asociada al valor del estadístico calculado. En este problema Fisher (1934) proporcionó el siguiente argumento:

"*Para las series de diferencias tenemos* $\bar{x} = +10.775$, $\frac{1}{4}s^2 = 3.756$, $t = 5.560$, $n = 3$, *por lo que la tabla muestra que P esta entre 0.01 y 0.02*" (Fisher, 1934, p. 125).

Reflexión:

La prueba t-Student para muestras apareadas o dependientes demuestra que la diferencia entre dos respuestas que son medidas en las mismas unidades estadísticas es cero.

4.2.3.5 Significado parcial 10 (SP10): Significancia de los coeficientes de regresión (significancia de la diferencia entre b*, y cualquier valor hipotético,* β*) con el estadístico* t*-Student*

En este significado se retoma el ejemplo 21 que se presentó en la página 191 sobre la prueba t -Student de significancia de los coeficientes de regresión. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de significancia, Fisher recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos sobre el rendimiento de los cultivos en las dos parcelas, el valor de la media para cada columna y algunos cálculos para obtener el estadístico t-Student. También hizo uso de la *representación simbólica* en la tabla y para presentar los resultados.

Situación/problema:

Fisher indicó que se podía seguir la misma lógica de la prueba de significancia de medias de dos muestras para aplicar la prueba a los coeficientes de regresión cuando las series de valores de la variable independiente no son idénticas. Para realizar esa prueba se auxilió del estadístico t-Student y su distribución.

Para ilustrar la aplicación de esta prueba, Fisher (1934) trabajó con datos que versaban sobre los rendimientos (en fanegas) de granos preparados por acre, de dos parcelas de trigo en Broadbalk. Estos datos se recolectaron durante treinta años; la parcela '9 a' recibió nitrato de sodio, mientras que '7 b' recibió una cantidad equivalente de sulfato de amonio. En el cuál se deseaba conocer si el rendimiento 'aparentemente' superior de la parcela '9 a' era significativo. Por lo que generó la cuarta columna de la Tabla 4.7 sobre las diferencias entre las parcelas (ver páginas 191-192), es decir, 9 a - 7 b, y sobre la cual se calculará el coeficiente de regresión b.

Conceptos/definiciones:

CD1: *Experimento* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad* de acuerdo con lo establecido en el CD7 del significado parcial 7.

CD8: *Distribución t-Student* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD9: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Tamaño muestral* en el mismo sentido que el CD10 del significado parcial 7.

CD11: *Varianza* de acuerdo a lo señalado en el CD10 del significado parcial 3.

CD12: *Error estándar* en el sentido que el CD11 del significado parcial 7.

CD13: *Variable independiente* es aquella que el experimentador puede controlar para observar los cambios que genera en la variable dependiente y no se ve influida por la variable dependiente.

CD14: *Variable dependiente* es sobre la cual tiene efecto la variable independiente, mientras que la variable dependiente no causa cambios en la variable independiente. En el ejemplo cualitativo de Fisher (1934), sobre la edad y la estatura, la variable dependiente es precisamente la estatura.

CD15: *Regresión* se utiliza para estimar la relación entre dos o más variables, nos ayuda a obtener información sobre una de ellas mediante los datos de las otras variables.

Propiedades/proposiciones:

PP1: *Distribución normal* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student de acuerdo con la PP2 del significado parcial 3.

PP3: *Media para cada muestra* en el mismo sentido que la PP3 del significado parcial 7.

PP4: *Función de regresión* $Y = a + b(x - \bar{x})$.

PP5: *Varianza* $s^2 = \frac{1}{n'-2} S(y - Y)^2$.

PP6: Desviación estándar $s = \sqrt{\frac{1}{n'-2} S(y - Y)^2}$.

PP7: *Grados de libertad* $n = n' - 2$, donde n' es el tamaño de la muestra.

PP8: *Suma de las deviaciones de la media al cuadrado* $S(x-\bar{x})^2 = S(x^2) - n\bar{x}^2$ *o* $S(x-\bar{x})^2 = \frac{n'\left(n'^2-1\right)}{12}$.

PP9: *Estimación del coeficiente de regresión* b de y sobre x $\quad b = \frac{S\{y(x-\bar{x})^2\}}{S(x-\bar{x})^2}$.

PP10: *Error estándar de b* $\sqrt{s^2/S(x-\bar{x})^2}$.

PP11: *Estadístico t-Student*

$$t = \frac{(b-\beta)\sqrt{S(x-\bar{x})^2}}{\mathrm{s}}$$

PP12: *Regla de decisión* según lo expresado en el PP7 del SP8.

Procedimientos:

Determinar si es significativa la diferencia de los coeficientes de regresión por medio del estadístico t-Student. Para realizarlo se llevó acabo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico $t = \frac{(b-\beta)\sqrt{S(x-\bar{x})^2}}{\mathrm{s}}$. Para este caso se puede calcular el estadístico t-Student como el coeficiente de regresión entre el error estándar $t = \frac{b}{\sqrt{s^2/S(x-\bar{x})^2}}$.
   a. Establecer las diferencias entre los rendimientos de las parcelas, 9 a - 7 b.
   b. Calcular la sumatorias y las medias por medio de $\bar{x} = \frac{1}{n}\sum x_i$ para los datos de las dos parcelas y de las diferencias.

   Nota: como se intenta conocer si el rendimiento 'aparentemente' superior de la parcela '9 a' era significativo se realizará el cálculo del coeficiente de regresión b para los datos de la columna '9 a - 7 b'.

   c. Calcular la suma de las deviaciones de la media al cuadrado para x con $S(x-\bar{x})^2 = \frac{n'\left(n'^2-1\right)}{12}$.

d. Calcular las diferencias $Sy(x-\bar{x})$ para lo cual los valores originales (de la cuarta columna) se suman desde abajo, dando totales sucesivos de 40.44 a 134.01 (sumatoria de dicha columna); el valor final debe coincidir con valor final de los valores originales. A la suma de los valores de esta nueva columna se le resta la suma de los valores originales multiplicada por $\frac{n'+1}{2}$.
e. Calcular el coeficiente b mediante la razón de las covarianzas de las dos variables según $b=\frac{S\{y(x-\bar{x})\}}{S(x-\bar{x})^2}$.
f. Calcular la suma de las deviaciones de la media al cuadrado para y $S(y-\bar{y})^2 = S(y^2)-n\bar{y}^2$.
g. Obtener los residuales, es decir la sumatoria de los cuadrados de las diferencias de la fórmula de regresión de acuerdo con $S(y-Y)^2 = S(y-\bar{y})^2 - b^2S(x-\bar{x})^2 = S(y^2)-n'\bar{y}^2 - b^2S(x-\bar{x})^2$.
h. Calcular el valor de la varianza por medio de $\mathrm{s}^2=\frac{1}{n'-2}S(y-Y)^2$.
i. Estimar el error estándar de b mediante $\sqrt{s^2/S(x-\bar{x})^2}$.

2. Determinar los grados de libertad mediante $n=n'-2$.
3. Buscar en las tablas de la probabilidad de Fisher (1934), cuál es la *P* para el valor del estadístico t-Student calculado con el número de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a la probabilidad de que no exista una diferencia entre las mediciones de las parcelas, es decir, que la parcela '9 a' no ha obtenido un incremento sobre la parcela '7 b'.

En este problema Fisher (1934, p. 136) proporcionó el siguiente argumento:

> Al probar la hipótesis de que $\beta=0$, es decir, que la parcela '9 a' no ha estado ganando sobre la parcela '7 b', dividimos b por esta cantidad y encontramos $t=2.282$. Dado que s se encontró a partir de 28 grados de libertad $n=28$, y el resultado de t muestra que *P* está entre 0.02 y 0.05. El resultado debe ser considerado significativo, aunque apenas.

4.2.3.6 Significado parcial 11 (SP11): Significancia de la diferencia de los coeficientes de regresión con el estadístico t-Student

En este significado se retoma el ejemplo 22 que se presentó en la página 193 sobre la prueba t-Student donde se comparan dos coeficientes de regresión. Los elementos primarios involucrados en el desarrollo de esta actividad son:

Elementos lingüísticos:

Para realizar la prueba de significancia, Fisher recurrió al *lenguaje natural* para plantear la situación problema e interpretar los resultados; a una *representación tabular* para presentar los datos sobre los volúmenes ocupados por células de algas en dos cultivos, el valor de la media para cada columna y algunos cálculos para obtener el estadístico t-Student. También hizo uso de la *representación simbólica* en la tabla, al presentar los resultados y al explicar los procedimientos necesarios para realizar la prueba.

Situación/problema:

Así como Fisher indicó que se podía seguir la misma lógica de la prueba de significancia de medias de dos muestras para aplicar la prueba a los coeficientes de regresión cuando las series de valores de la variable independiente no son idénticas; también señaló que tal como la prueba de significancia de la diferencia de medias se puede aplicar cuando las muestras son de diferentes tamaños, obteniendo una estimación del error combinado, podemos aplicar el método para comparar los coeficientes de regresión cuando las series de la variable independiente no son idénticas.

Para mostrar cómo funciona esta prueba, Fisher (1934) retomó datos de Dr. M. Bristol-Roach que trataban sobre los volúmenes que se encuentran ocupados por células de algas en dos cultivos en paralelo (A y B). En los periodos en que se tomaron los datos la tasa de crecimiento relativo fue constante. Se deseaba conocer si el cultivo B estaba creciendo más rápidamente que el cultivo A. Esto se llevó a cabo por medio de la prueba de significancia de diferencia de los coeficientes de regresión, para probar si existe una diferencia significativa que indique tal crecimiento superior del cultivo B sobre el A.

Conceptos/definiciones:

CD1: *Experimento* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad* de acuerdo con lo establecido en el CD7 del significado parcial 7.

CD8: *Distribución t-Student* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD9: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Tamaño muestral* en el mismo sentido que el CD10 del significado parcial 7.

CD11: *Varianza* de acuerdo a lo señalado en el CD10 del significado parcial 3.

CD12: *Error estándar* en el sentido que el CD11 del significado parcial 7.

CD13: *Variable independiente* según lo establecido en el CD13 del significado parcial 10.

CD14: *Variable dependiente* en concordancia con lo señalado en el CD14 del significado parcial 10.

CD15: *Regresión* en el mismo sentido que en el CD15 del significado parcial 10.

Propiedades/proposiciones:

PP1: *Distribución normal* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student de acuerdo con la PP2 del significado parcial 3.

PP3: *Media para cada muestra* en el mismo sentido que la PP3 del significado parcial 7.

PP4: *Función de regresión* Y según lo establecido en la PP4 del significado parcial 10.

PP5: *Varianza* $\mathrm{s}^2 = \frac{1}{n} S_A(y - Y)^2 + S_B(y - Y)^2$, donde n son los grados de libertad.

PP6: *Varianza de la diferencia de los coeficientes de regresión* $(s^2/S_A(x-\bar{x})^2) + (s^2/S_B\,(x-\bar{x})^2)$.

PP7: Error estándar $\sqrt{(s^2/S_A(x-\bar{x})^2) + (s^2/S_B\,(x-\bar{x})^2)}$.

PP8: *Grados de libertad* $n = n_A + n_B$, es decir, la sumatoria de los grados de libertad de cada serie, mientras que cada uno de ellos se obtiene restando una al tamaño de la serie n', de modo que $n_i = n' - 1$, esto de acuerdo a lo expresado en la PP6 del significado parcial 7.

PP9: *Suma de las deviaciones de la media al cuadrado* $S(x-\bar{x})^2$ en el mismo sentido que la PP8 del significado parcial 7.

PP10: *Estimación del coeficiente de regresión b* de acuerdo a lo establecido en la PP9 del significado parcial 10.

PP11: *Estadístico t-Student* como la diferencia de los coeficientes de regresión entre el error estándar

$$t = \frac{b' - b}{\sqrt{(s^2/S_A(x-\bar{x})^2) + (s^2/S_B\,(x-\bar{x})^2)}}$$

PP12: *Regla de decisión* según lo expresado en el PP7 del SP8.

Procedimientos:

Determinar si es significativa la diferencia de los coeficientes de regresión por medio del estadístico y la distribución t-Student. Para realizarlo se llevó acabo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico $t = \frac{b'-b}{\sqrt{(s^2/S_A(x-\bar{x})^2)+(s^2/S_B(x-\bar{x})^2)}}$, es decir, la diferencia de los coeficientes de regresión entre el error estándar.
   a. Calcular la sumatorias y las medias por medio de $\bar{x} = \frac{1}{n}\sum x_i$ para las dos series
   b. Calcular la suma de las deviaciones de la media al cuadrado para x con $S(x-\bar{x})^2 = \frac{n'\left(n'^2-1\right)}{12}$.
   c. Calcular las diferencias $Sy(x-\bar{x})$ para lo cual los valores originales se suman desde abajo, dando totales sucesivos de 6.087 a 43.426, el valor final debe

coincidir con valor final de los valores originales. A la suma de los valores de esta nueva columna se le resta la suma de los valores originales multiplicada por $\frac{n'+1}{2}$.

d. Calcular el coeficiente b mediante la razón de las covarianzas de las dos variables según $b = \frac{S\{y(x-\bar{x})\}}{S(x-\bar{x})^2}$.
e. Obtener los residuales para cada serie, es decir la sumatoria de los cuadrados de las diferencias de la fórmula de regresión de acuerdo con $S(y-Y)^2 = S(y-\bar{y})^2 - b^2S(x-\bar{x})^2 = S(y^2) - n'\bar{y}^2 - b^2S(x-\bar{x})^2$.
f. Calcular el valor de la varianza por medio de $\mathrm{s}^2 = \frac{1}{n}S_A(y-Y)^2 + S_B(y-Y)^2$
g. Estimar el error estándar de b mediante $\sqrt{(\mathrm{s}^2/S_A(x-\bar{x})^2) + (\mathrm{s}^2/S_B\,(x-\bar{x})^2)}$.

2. Determinar los grados de libertad mediante $n = n_A + n_B$ o $n = n'_A + n'_B - 2$.
3. Buscar en las tablas de la probabilidad de Fisher (1934), cuál es la P para el valor del estadístico t-Student calculado con el número de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a la probabilidad de que el cultivo B no tenga un rendimiento mayor al del cultivo A por medio de la diferencia entre los volúmenes ocupados por células de algas entre los cultivos A y B.

Por ejemplo:

> Tomando la raíz cuadrada, encontramos que el error estándar es 0.01985, y $t = 1.844$. La diferencia entre los coeficientes de regresión, aunque es relativamente grande, no puede considerarse significativa. No hay evidencia suficiente para afirmar que el cultivo B estaba creciendo más rápidamente que el cultivo A. (Fisher, 1934, p. 139)

4.2.3.7 Significado parcial 12 (SP12): Significancia de la diferencia de medias cuando las varianzas poblacionales son diferentes con el estadístico t-Student o t-Welch

En este significado se retoma el ejemplo 23 que se presentó en la página 200 sobre la prueba de Welch con el estadístico y distribución t-Student donde se desea conocer si las medias de las poblaciones son iguales cuando existe una desigualdad de varianzas. A continuación, detallamos los objetos matemáticos primarios involucrados en el desarrollo de esta actividad.

Elementos lingüísticos:

Para realizar la prueba t-Student de desigualdad de varianzas se utilizó *lenguaje natural* para plantear la situación problema e interpretar los resultados; por medio de la *representación tabular* se presentan los datos del número de bacterias y el valor de la media para cada columna. También se hace uso de la *representación simbólica* en la tabla y para presentar los resultados.

Situación/problema:

La problemática que abordó Welch (1938 y 1947) refiere a la forma de abordar la prueba de significancia de diferencia de medias cuando existe una desigualdad de varianzas poblacionales. La prueba también se aplica si los tamaños muestrales son iguales o diferentes. Para realizar esta prueba propuso criterios (u y v) y probó que se distribuyen aproximadamente como una distribución t-Student con f grados de libertad.

Para ilustrar la aplicación de esta prueba hemos retomado un ejemplo de Fisher (1925b), que también hemos utilizado en el SP7, en este ejemplo se trabajó con datos de un experimento de electro-cultivo realizado en 1922 en Rothamsted, donde se medía el número de macollos de la planta de cebada. Por una parte, estaban aquellas plantas que recibían una descarga eléctrica de alta tensión, y, por otra parte, las que se encerraron en una jaula de alambre con conexión a tierra.

Conceptos/definiciones:

CD1: *Experimento* de acuerdo con lo establecido en el CD1 del significado parcial 2.

CD2: *Muestra* en el mismo sentido del CD2 del significado parcial 2.

CD3: *Población* de acuerdo con lo establecido en el CD3 del significado parcial 2.

CD4: *Media* en concordancia con el CD4 del significado parcial 2.

CD5: *Desviación estándar* en el sentido del CD5 del significado parcial 2.

CD6: *Estadístico* t-Student de acuerdo con lo establecido en el CD6 del significado parcial 2.

CD7: *Probabilidad* de acuerdo con lo establecido en el CD7 del significado parcial 7.

CD8: *Distribución t-Student* de acuerdo con lo establecido en el CD8 del significado parcial 3.

CD9: *Significancia* de acuerdo con lo establecido en el CD8 del significado parcial 2.

CD10: *Tamaño muestral* en el mismo sentido que el CD10 del significado parcial 7.

CD11: *Varianza* de acuerdo a lo señalado en el CD10 del significado parcial 3.

CD12: *Error estándar* en el sentido que el CD11 del significado parcial 7.

Propiedades/proposiciones:

PP1: *Distribución normal* en el mismo sentido que la PP2 del significado parcial 1.

PP2: *Distribución del estadístico* t-Student de acuerdo con la PP2 del significado parcial 3.

PP3: *Media para cada muestra* en el mismo sentido que la PP3 del significado parcial 7.

PP4: *Verdadera diferencia de medias* $\eta = \alpha_1 - \alpha_2$.

PP5: *Diferencia de medias estimada* $y = (\bar{x}_1 - \bar{x}_2)$.

PP6: *Varianza estimada* $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, donde $\lambda_1 = 1/n_1$ y $\lambda_2 = 1/n_2$.

PP7: *Varianza de la muestra* $s_i^2 = {}^{\Sigma_i}/_{f_i}$, donde $\Sigma_i = S(x_i - \bar{x})^2$ y $f_i = n_i - 1$. La varianza muestral se utiliza para estimar a σ^2. En otra notación, como la de Gosset y que es como la podemos encontrar actualmente se tiene que $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$, donde n es el tamaño de la muestra.

PP8: *Grados de libertad*:

- Si $k = 2$

$$f = \frac{(\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)^2}{\frac{\lambda_1^2\sigma_1^4}{f_1} + \frac{\lambda_2^2\sigma_2^4}{f_2}}$$

Con s^2 como estimación de σ^2, f es igual a como comúnmente conocemos v hoy en día para esta prueba:

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1)^2 n_1 - 1} + \frac{s_2^4}{(n_2)^2 n_2 - 1}}$$

- Si k no se restringe a 2

$$f = \frac{(\lambda_i \sigma_i^2)^2}{\Sigma \frac{\lambda_i^2 \sigma_i^4}{f_i}}$$

- Si las $\sigma_i{}'s$ se desconocen

$$f = \frac{(\Sigma \lambda_i s_i^2)^2 - 2\left(\Sigma \frac{\lambda_i s_i^4}{f_i + 2}\right)}{\left(\Sigma \frac{\lambda_i^2 s_i^4}{f_i + 2}\right)}$$

PP9: *Estadístico t-Student o también llamado criterios por Welch*:

- Si $k = 2$

$$v = \frac{(y - \eta)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}}$$

El criterio v, asumiendo que $\eta = \alpha_1 - \alpha_2 = 0$, es igual a como comúnmente conocemos al estadístico t actualmente para esta prueba:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{1}{n_1} s_1^2 + \frac{1}{n_2} s_2^2\right)}}$$

- Si k no se restringe a 2

$$v = \frac{(y - \eta)}{\sqrt{(\lambda_i s_i^2)}}$$

Procedimientos:

Determinar si la diferencia entre las medias de las muestras pequeñas es significativa por medio del estadístico t-Student. Para realizarlo se llevó acabo la siguiente concatenación de acciones:

1. Encontrar el valor del estadístico t mediante el criterio $v = \frac{(y-\eta)}{\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}}$.
    a. Dado que la hipótesis que se quiere comprobar es que no existe una diferencia significativa entre las medias poblacionales se asume que $\eta = \alpha_1 - \alpha_2 = 0$.
    b. Calcular la media por medio de $\bar{x} = \frac{1}{n}\sum x_i$, tanto para la muestra 1 como para la muestra 2.
    c. Calcular la diferencia de las medias de las muestras $y = (\bar{x}_1 - \bar{x}_2)$.
    d. Obtener la varianza muestral $s_i^2 = {}^{\Sigma_i}/_{f_i}$, donde $\Sigma_i = S(x_i - \bar{x})^2$ y $f_i = n_i - 1$.
    e. Calcular la varianza estimada $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, donde $\lambda_1 = 1/n_1$ y $\lambda_2 = 1/n_2$ y la varianza muestral se utiliza para estimar a σ^2.
2. Determinar los grados de libertad mediante $f = \frac{(\lambda_1\sigma_1^2+\lambda_2\sigma_2^2)^2}{\frac{\lambda_1^2\sigma_1^4}{f_1}+\frac{\lambda_2^2\sigma_2^4}{f_2}}$.
3. Buscar en las tablas de la probabilidad de Gosset (Student, 1925) la probabilidad de ocurrencia para el estadístico t-Student calculado y con los grados de libertad de la prueba. O en Fisher (1934), sí el valor del estadístico t-Student calculado se encuentra fuera del rango para cierta P con el numero de grados de libertad n determinado previamente.

Argumentos:

Se argumenta con respecto a la probabilidad, asociada al estadístico t-Student calculado, de que no exista una diferencia significativa entre las medias.

Por ejemplo:

De acuerdo al valor del estadístico t-Student calculado, $t = 3.175$, se tiene una probabilidad $P = 0.00441764$ de que no exista una diferencia entre las medias, lo cual quiere decir que la posibilidad de que las medias sean iguales es de 44 veces en 10,000. Por lo tanto, la diferencia entre las medias es significativa y podemos decir que el número de macollos de las plantas de las macetas electrificadas y de las enjauladas es diferente.

4.3 Síntesis de los significados del estadístico t − *Student*

Con el estudio histórico-epistemológico que se realizó para el estadístico t-Student, se identificaron tres grandes problemáticas que han resultado clave en el surgimiento, desarrollo y generalización de este estadístico; estas grandes problemáticas dan paso a tres grandes “significados de referencia”: 1) Prueba de la media de una muestra, 2) Distribución t-Student y, 3) Prueba de la diferencia de medias para dos muestras. En las prácticas matemáticas que se realizaron para resolver estas problemáticas, fue posible identificar diversos significados parciales donde se movilizaron distintas situaciones-problema, elementos lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos. De esta forma, los doce significados parciales conforman el significado global del estadístico t-Student, lo cual se esquematiza en la Figura 4.7.

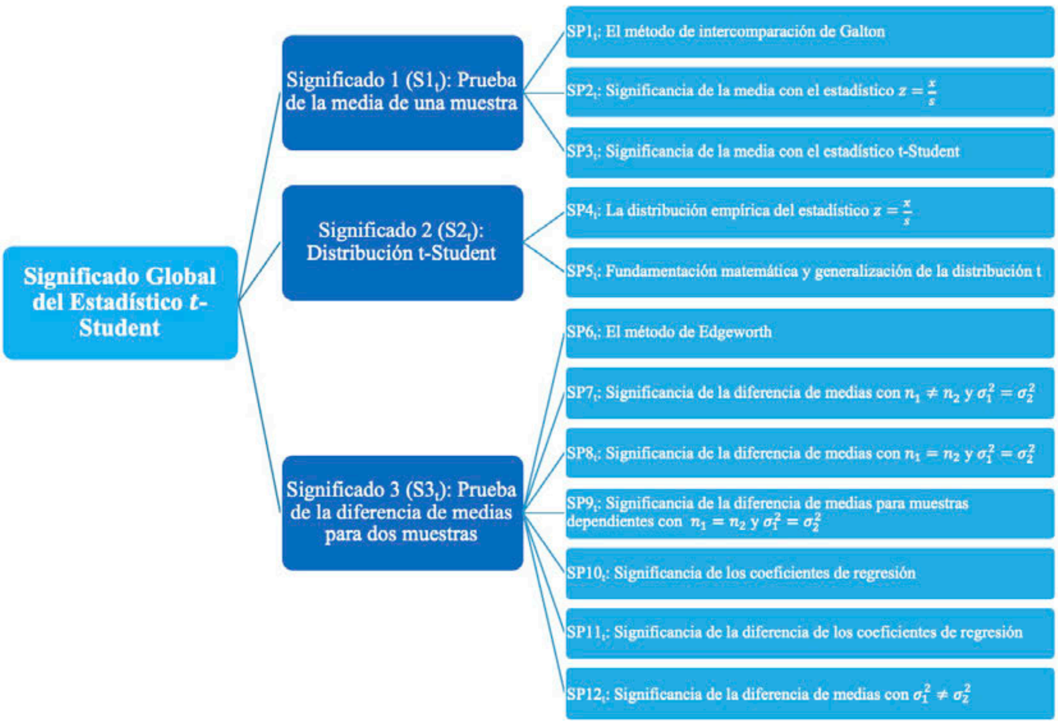


Figura 4.7. Significado global del Estadístico t-Student

La prueba de la media de una muestra (S1) nos permite determinar si la media de una población, que se distribuye de acuerdo a la normal, es igual a un valor μ_0. Mientras que la interrogante ¿cómo se distribuye el estadístico t-Student?, constituyó el problema principal que detonó el S2. Además, en la prueba de la diferencia de medias para dos muestras (S3), se encuentra la generalización de algunos elementos del S1, con esta prueba podemos determinar si las medias de dos poblaciones distribuidas de forma normal son iguales, en otras palabras, si no existe diferencia significativa entre las medias.

El significado parcial uno y seis corresponden a una versión intuitiva de estas pruebas, por ejemplo, en el SP1 se puede analizar la variación interna de un conjunto de datos, mientras que en el SP6 se analiza la fluctuación como una medida de la variación de una serie de datos con respecto a otra serie.

Cada uno de estos significados parciales del estadístico Chi-cuadrada, dan cuenta de la evolución progresiva, de lo informal a lo formal, de este estadístico. Esta caracterización de los significados del estadístico t-Student nos permite acceder a elementos matemáticos para una transición continua del RII al RIF y también mediante estos elementos podríamos visualizar la variedad de caminos para la enseñanza y aprendizaje de esta noción.

4.4 Reflexiones finales del estudio histórico-epistemológico sobre el estadístico $t-Student$

La idea de intercomparación de Galton (1875), es decir, de que las comparaciones se puedan realizar únicamente en términos de la variación interna del conjunto de datos y no tomando como referencia un criterio externo al conjunto, es lo que da surgimiento a los desarrollos de la t-Student. Además, la extensión de dicha idea, en los trabajos de Francis Edgeworth, William Gosset y Roland Fisher, hacen a la intercomparación un sostén de la estadística moderna (Stigler, 2017).

Galton (1875), presentó la intercomparación como un método de comparación donde hacía uso de percentiles, cuartiles y la media (principalmente), los cuales se determinaban al ordenar y graficar los datos, sin la necesidad de cálculos aritméticos complejos. Mientras que el uso más elaborado del método de intercomparación o más matematizado lo realizó Gosset (Student, 1908a), en su artículo "El error probabilístico de la media".

De acuerdo con Stigler (2017), por más de un siglo se había utilizado la media aritmética en astronomía y se describía su precisión en términos de "error probable" (definido en el SP1). Para 1893, Pearson había propuesto una escala de la desviación estándar que era proporcional al error probable, esta propuesta ($ep \approx 0.6745\sigma$) pronto se volvió la regla y cuando se trabajaba con muestras grandes era frecuente remplazar σ por $\sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}$ o por $\sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}$. Sin embargo, Gosset cuestionó que tan factible era utilizar dicha aproximación cuando se esta trabajando con muestras pequeñas y si se reemplazaba σ por $\sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}$, cómo sería la distribución de $z = \frac{\bar{X}}{\sqrt{\frac{1}{n}\sum(X_i-\bar{X})^2}}$. Posteriormente, Fisher planteó y resolvió el problema en términos de geometría multidimensional, probando con rigurosidad el test de Gosset y posteriormente lo reescribió como $t = \sqrt{n-1}z$. Fisher logró liberar a la distribución t de la dependencia respecto a σ e inventó también la prueba t de dos muestras.

40 años antes de la extensión a dos muestras de las pruebas t de Student que realizó Fisher, Edgeworth había desarrollado un método de análisis para tablas estadísticas que forman un precedente para estas pruebas, utilizando lo que llamó "fluctuación". Este análisis se basaba solamente en la variación interna de los conjuntos de datos, siguiendo la idea de Galton.

Para Fisher, continuar y extender los trabajos de Gosset y Edgeworth fue el inicio para el "análisis de varianza" con el estadístico y distribución F de Fisher o también conocido como F de Fisher-Snedecor.

Las pruebas t-Student actualmente se aplican cuando estamos trabajando con muestras pequeñas y éstas provienen de una población normal, y desconocemos el valor de la varianza poblacional. Tal como lo hemos planteado en los significados 1 y 3, se utilizan con una o dos muestras, y para este último caso existen algunas variaciones, (1) tamaños muestrales iguales, (2) tamaños muestrales desiguales y (3) varianzas desiguales; todos ellos abordados en los significados parciales de la sección 4.2.3.

Al igual que con el estadístico χ^2, en la época en que emergen y se desarrollan las pruebas t-Student la hipótesis nula se encontraba implícita en la prueba; sin embargo, con el desarrollo de las pruebas de hipótesis como una metodología, la hipótesis nula y alternativa se hacen explícitas. Por ejemplo, hoy en día podemos encontrar las hipótesis para el ejemplo 16 (ver página 174) de la siguiente forma:

Hipótesis nula $Ho: \mu = \mu_0$

Hipótesis alternativa $Ha: \mu > \mu_0$

Actualmente, el estadístico t-Student también se utilizan para formar intervalos de confianza y un esbozo de esto se pudo apreciar en la práctica de la generalización del estadístico t-Student cuando Welch sugiere utilizar el estadístico si queremos estimar η (la media poblacional) y proporcionar una medida de la incertidumbre de la estimación.

CAPÍTULO 5

Propuesta de Niveles para los Estadísticos χ^2 y $\mathrm{t}-Student$

"El conocimiento es, pues, un sistema de transformaciones que se vuelven progresivamente adecuadas". (Jean Piaget)

INTRODUCCIÓN

En este Capítulo presentamos una propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. La propuesta se formuló a partir de criterios epistémicos identificados con el estudio de tipo histórico-epistemológico sobre estos estadísticos, que fueron detallados en los capítulos 3 y 4, y de la investigación desarrollada sobre razonamiento inferencial (informal, formal y de progresión). En los indicadores de los distintos niveles de razonamiento aquí expuestos vinculamos atributos matemáticos de los diversos significados de los estadísticos Chi-cuadrada y t-Student.

5.1 Niveles de razonamiento inferencial para el estadístico Chi-cuadrada (χ^2)

En esta sección presentamos cuatro niveles de razonamiento inferencial para el estadístico χ^2, los cuales vinculan tanto los aportes de la literatura científica en didáctica de la estadística, como los objetos matemáticos primarios y procesos identificados en cada uno de los significados parciales resultantes del estudio histórico-epistemológico.

5.1.1 Nivel 1

En este nivel se brindan 'indicadores o elementos' que corresponden a un razonamiento inferencial informal, donde se trabaja inicialmente la visualización de gráficos para establecer conjeturas y, posteriormente, se analizan datos. Intervienen objetos matemáticos de Estadística Descriptiva y Probabilidad para hacer conjeturas fundamentadas. En este nivel las conjeturas no se apoyan en un modelo de probabilidad.

5.1.1.1 Subnivel 1.1: Visualización

Se espera que el estudiante pueda conjeturar y argumentar si los datos de una muestra siguen una distribución normal a través de los elementos presentes en las gráficas (e.g., forma, dispersión, amplitud de los cuartiles, mediana, sesgo); de una forma similar a como se aborda la primera parte de la tarea 1 para promover el RII de Zieffler et al. (2008), donde se les pide a los estudiantes que especulen sobre las características gráficas de la población desconocida basándose únicamente en la gráfica de una muestra. La Figura 5.1 muestra el tipo de gráficos que se les pueden presentar a los estudiantes. Estos gráficos se retoman de los elementos lingüísticos identificados en el SP1χ^2[15] de la prueba de bondad de ajuste.

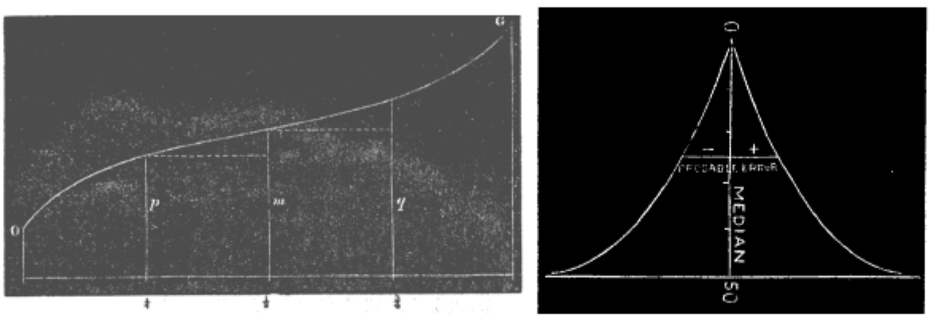


Figura 5.1. Gráficos pertenecientes al SP1χ^2 de la prueba de bondad de ajuste (Galton, 1775; Galton, 1885)

5.1.1.2 Subnivel 1.2: Trabajar con datos

El estudiante puede analizar los datos de una muestra bajo el método gráfico de Galton para conjeturar si el grupo de datos sigue una distribución normal; esto se puede ver en dos partes:

[15] Para referirnos a los significados parciales del estadístico Chi-cuadrada utilizaremos la notación SP1χ^2, SP2χ^2, SP3χ^2, …, SP12χ^2, mientras que, para referirnos a los significados parciales del estadístico t-Student, utilizaremos la notación SP1$_t$, SP2$_t$, SP3$_t$, …, SP12$_t$.

La primera, corresponde al denominado método de intercomparación, donde intervienen objetos matemáticos primarios del SP1χ^2 (e.g., *conceptos/definiciones* como ojiva –CD2–, cuartiles –CD3–, percentiles –CD4– y desviaciones –CD7–; y *propiedades/proposiciones* como distribución normal –PP2–, m en $\frac{1}{2}$ o 0°–PP3–, p en $\frac{1}{4}$ o $-25°$–PP4–, q en $\frac{3}{4}$ o 25°–PP5–, y $q - m = m - p$ como condición para una serie simétrica –PP7–). Realizando los procedimientos correspondientes el estudiante puede obtener la gráfica de la izquierda de la Figura 5.1.

En la segunda parte del método gráfico, el estudiante es capaz de utilizar los resultados del método de intercomparación para calcular y graficar las desviaciones, positivas y negativas, con respecto a la mediana (como se describió en el *procedimiento* del SP1χ^2), para lo cual hace uso del concepto error medio –CD6– y la propiedad error probable –PP6– del SP1χ^2. El tipo de representación gráfica que obtendría el estudiante corresponde al gráfico de la derecha de la Figura 5.1.

Se puede hacer la conexión de las desviaciones con respecto a la mediana con el error ($e = FO - FE$) del estadístico χ^2 de Pearson (SP1χ^2 y SP2χ^2). Esto debido a que, en ambos casos, se están midiendo desviaciones de las observaciones con respecto a cierto valor (esperado), en este caso el valor contra el que se contrastan los observados es la m del conjunto de datos que se están trabajando, mientras que el valor esperado cuando trabajamos en una prueba con el estadístico χ^2 depende de una distribución teórica.

5.1.1.3 Subnivel 1.3: Asociación intuitiva

El estudiante puede establecer si existe asociación entre dos variables, con un atributo de cada variable, por medio del coeficiente de asociación Q. Para lograrlo, se movilizan objetos matemáticos primarios correspondientes al SP5χ^2 del S2χ^2[16]. Concretamente:

- Comprende *conceptos/definiciones* básicos de probabilidad y sus *propiedades* –e.g., variable (CD1), cualidades o atributos (CD2), frecuencia (CD3), conjuntos de

[16] Para referirnos a los grandes significados del estadístico Chi-cuadrada utilizaremos la notación S1χ^2, S2χ^2, S3χ^2 y S4χ^2, mientras que, para referirnos a los grandes significados del estadístico t-Student, utilizaremos la notación $S1_t$, $S2_t$, $S3_t$. Recordemos que estos grandes significados se constituyen con significados parciales, por ejemplo, el S1χ^2 esta constituido por el SP1χ^2, SP2χ^2, SP3χ^2 SP4χ^2.

probabilidad (CD4), asociación (CD7), la PP1 donde se indica que $(AB)(U) = (A)(B)$, y $(AB)(\alpha\beta) = (A\beta)(\alpha B)$–,

- Puede trabajar con tablas de contingencia de 2×2, como se muestra en la Figura 5.2.

	(B)	(β)
(A)	(AB)	$(A\beta)$
(α)	(αB)	$(\alpha\beta)$

Figura 5.2. Tabla de 2×2 con nomenclatura de Yule

En el ejemplo 3, presentado en el capítulo 3, podemos observar que se trabaja con dos variables, donde se buscaba encontrar si existía una asociación entre la "ausencia de la vacuna" y "ataque", donde (A) indica si fue atacado por la viruela y (B) que no estaba vacunado, si (α) y (β) indican el complemento de (A) y (B) respectivamente, se pueden describir en términos del ejemplo como (α) aquellos que no fueron atacados de viruela y (β) quienes si estaban vacunados.

- Es capaz de relacionar la *propiedad* del coeficiente de asociación Q (PP4) con conceptos del coeficiente de correlación de Pearson (CD6), debido a que el sentido de la asociación con este coeficiente es en términos de la correlación para variables continuas.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

Cabe recalcar que los objetos matemáticos primarios que se encuentran en este nivel corresponden al Significado parcial 5 de la Chi cuadrada (SP5χ^2) lo cual refiere al significado del estadístico Chi-cuadrada como prueba de independencia (S2χ^2).

Es importante recordar la importancia de trabajar con tipos de situaciones-problema en contextos cercanos a los estudiantes, debido a que en diversas investigaciones (e.g., Makar, Bakker & Ben-Zvi, 2011; Bakker & Derry, 2011; Gil & Ben-Zvi, 2011; Makar &Ben-Zvi, 2011; Ben-Zvi & Aridor-Berger, 2016; Bakker, Ben-Zvi & Makar, 2017) se ha puesto de manifiesto la importancia del contexto al promover un RII. Este tipo de razonamiento no se basa únicamente en el conocimiento estadístico de los estudiantes, también considera el papel que tiene el razonamiento informal y el conocimiento informal, tal como lo declaran Zieffler et al. (2008).

5.1.2 Nivel 2

Algunos indicadores de este nivel tienen rasgos del RII; por ejemplo, la forma en que se aproxima la hipótesis nula, pues ésta se encuentra implícita en el problema, y el sentido que se le da a la probabilidad. Es por ello, que este nivel se puede considerar *preformal*.

Estas formas de aproximarse a las pruebas las podemos ver en la historia, por ejemplo, en el inicio de la prueba de bondad de ajuste; Pearson (1900) hacía una pregunta sobre el problema que estaba planteando, a manera de cierre, esta pregunta podemos entenderla como la hipótesis nula.

5.1.2.1 Subnivel 2.1: Identificar la prueba no paramétrica necesaria para analizar los datos

Para que un estudiante pueda identificar la prueba χ^2 adecuada para analizar los datos, son necesarios los siguientes indicadores:

- Reconocer el tipo de datos que está trabajando (e.g., se trata de una muestra o de una población, son cualitativos o cuantitativos, se clasifican de acuerdo a una o dos variables, tipo de muestreo y número de muestras).
- Comprender el problema a resolver, en otras palabras, reconoce cuál es la incógnita a responder.
  Polya (1989) define ésta como la primera de las cuatro fases para resolver problemas, y que al comprender el problema el alumno será capaz de separar las principales partes del problema, la incógnita, los datos y la condición.
- Comprender la lógica de las pruebas de bondad de ajuste (S1χ^2), independencia (S2χ^2) y homogeneidad (S3χ^2). Por ejemplo, la prueba de bondad de ajuste busca determinar si los datos siguen cierta distribución teórica; la prueba de independencia se utiliza para determinar si dos variables tienen asociación o son independientes; y la prueba de homogeneidad permite estudiar si diferentes poblaciones son homogéneas con respecto a alguna variable.
- Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla. Para seleccionar la prueba adecuada, el estudiante es capaz de establecer una relación entre los elementos del problema, como la incógnita y los datos, con la lógica de las pruebas Chi-cuadrada.

5.1.2.2 Subnivel 2.2: Una aproximación de las pruebas con el estadístico χ^2

Una vez que se ha identificado la prueba, para dar respuesta al problema planteado el estudiante debe:

- Identificar la hipótesis nula que se encuentra implícita en el problema.

  Como se indicó previamente, la identificación de la hipótesis en forma de pregunta es un rasgo clave en este nivel. Además, esta forma presenta una primera aproximación a las hipótesis estadísticas formales que se trabajan en el nivel 4. Algunas investigaciones han reportado las dificultades que presentan los estudiantes al construir las hipótesis estadísticas e incluso en la lógica que estas tienen (e.g. Vallecillos, 1997; Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Batanero, 2013; López-Martín, Batanero y Gea, 2019); por lo cual, introducirlas en forma de pregunta y después ir evolucionando, como lo veremos en el nivel tres, puede ayudar a los estudiantes a comprenderlas.

  En este sentido, los ciclos de investigación estadística tal como el PPDAC (the statistical enquiry cycle) tienen como primer componente la generación de una pregunta de investigación, la cual debe estar dada sobre un contexto particular que es sobre el cual se desea indagar. Algunas investigaciones (e.g., Pfannkuch & Wild, 2004; Pfannkuch, Budgett, Fewster, Fitch, Pattenwise, Wild & Ziedins, 2016; Stohl, Angotti & Tarr, 2010), han retomado este primer componente, reconociendo que la mayoría de estas preguntas tienen forma de conjetura o hipótesis.
- Conocer la distribución χ^2 (e.g., es asimétrica positiva, tiene como único parámetro a los grados de libertad, a medida que incrementan los grados de libertad se aproxima a la curva normal y no puede tomar valores negativos) de acuerdo con lo que se establece en la *propiedad/proposición* PP4 del SP2χ^2.

  Para que el estudiante comprenda la noción de distribución, algunas investigaciones se han apoyado en la tecnología, generando, por ejemplo, simulaciones (eg., Bakker & Gravemeijer, 2004; Reading & Reid, 2006; Rossman, 2008; Dinov, Palanimalai, Khare & Christou, 2018). En la Figura 5.3 podemos observar cómo varía la forma de la distribución al aumentar los grados de libertad.

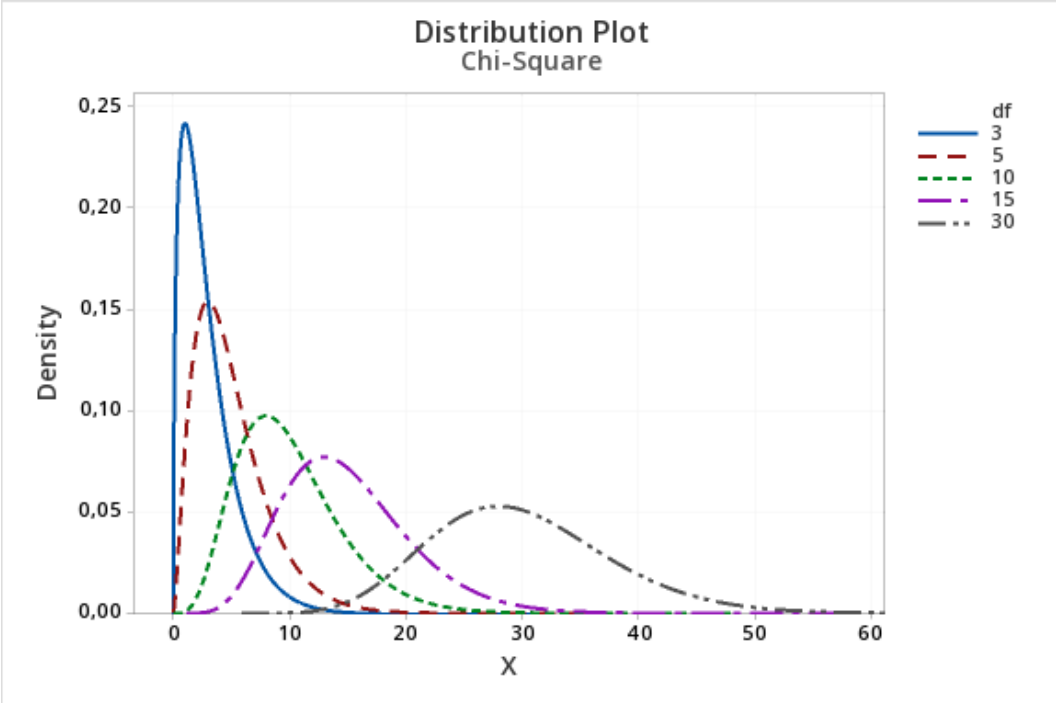


Figura 5.3. Distribución Chi-cuadrada

Los indicadores anteriores son para las tres pruebas de la χ^2 (S1, S2 y S3). Pero, además, otros indicadores a considerar por prueba, son los contemplados en las secciones 5.1.2.2.1, 5.1.2.2.2, 5.1.2.2.3 y 5.1.2.2.4, las cuales de enuncian a continuación.

5.1.2.2.1 Prueba de bondad de ajuste (S1χ^2)

El estudiante puede valorar en qué medida se ajusta un grupo de datos observados a cierta distribución teórica preestablecida, por medio del contraste de las frecuencias observadas y las frecuencias esperadas. Por ejemplo, si deseamos conocer si el conjunto de datos que tenemos sigue una distribución normal. Para hacer esta valoración, el estudiante:

- Calcula y comprende qué indica el estadístico χ^2 por medio de *conceptos/definiciones* y *propiedades/proposiciones* del SP2χ^2 (e.g., frecuencia observada –CD3–, frecuencia teórica o esperada –CD4–, estadístico Chi-cuadrada –PP7– visto como la propiedad $\chi^2 = \sum \frac{\mathrm{e}^2}{m} = \sum \frac{(FO-FE)^2}{FE}$ y el error –PP3– como $e = \mathrm{FO} - FE$).
- Calcula y comprende los grados de libertad de acuerdo con su *definición* –CD7–, indica que refiere al número de filas menos el número de restricciones lineares independientes en las frecuencias y la propiedad –PP3–, $k = n - r$ (SP3χ^2).

- Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad del suceso y puede interpretarla como una medida de ocurrencia de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema observado (para lo que se utiliza el *concepto/definición* de probabilidad –CD5– y la *propiedad* de distribución de probabilidad χ^2–PP4–, correspondientes al SP2χ^2). De acuerdo con Stohl, Angotti y Tarr (2010), los estudiantes estarán tomando decisiones de forma natural, tanto para mantener su hipótesis actual o para alterarla basándose en la probabilidad obtenida.

  Esta forma de utilizar la probabilidad es otro de los indicadores clave de este nivel. Como vimos en los estudios históricos-epistemológicos presentados en los capítulos previos, esta forma de interpretar y utilizar la probabilidad se empleó en este tipo de pruebas a inicios del siglo XX; esto permitió interpretar por medio de la probabilidad los valores obtenidos con los diferentes estadísticos (e.g., Chi-cuadrada y t-Student).

Proceso de generalización internivel

El nivel 1.2, donde se trabaja con el SP1χ^2, sólo se puede aplicar cuando se desea saber si el conjunto de datos sigue una distribución normal; en cambio, cuando se trabaja con los elementos del nivel 2.2 (SP2χ^2) se pueden contrastar los datos con cualquier distribución teórica.

5.1.2.2.2 Prueba de independencia (S2χ^2) y homogeneidad (S3χ^2)

Debido a que las pruebas de independencia y homogeneidad son matemáticamente idénticas, comparten los siguientes indicadores:

- Comprende y utiliza frecuencia observada con tablas de $r \times$ c (CD3 del SP6χ^2). Para el caso específico de las tablas de contingencia de 2×2, reconoce la siguiente representación tabular y simbología:

$$\begin{array}{ccc} a & b & a+b \\ c & d & c+d \\ a+c & b+d & a+b+c+d \end{array}$$

- Comprende y utiliza la frecuencia esperada bajo independencia probabilística (CD4 del SP6 y PP2 del SP11).
  - Para tablas de contingencia de 2×2:

$$\begin{array}{ccc} \frac{(a+b)(a+c)}{a+b+c+d} & \frac{(a+b)(b+d)}{a+b+c+d} & a+b \\ \frac{(c+d)(a+c)}{a+b+c+d} & \frac{(c+d)(b+d)}{a+b+c+d} & c+d \\ a+c & b+d & a+b+c+d \end{array}$$

  - Para tablas de contingencia de $r \times$ c:

$$e_{ij} = \frac{n_i \times n_j}{n}$$

- Puede calcular el estadístico χ^2 (*propiedad* –PP5– del SP6χ^2):
  - Para tablas de contingencia de 2×2:

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+c)(b+d)(a+b)(c+d)}$$

  - Para tablas de contingencia de $r \times$ c:

$$\chi^2 = \sum \left\{ \frac{(n_{uv} - v_{uv})^2}{v_{uv}} \right\} = \sum \frac{(FO - FE)^2}{FE}$$

- Calcula y comprende los grados de libertad $k = (c-1)(r-1)$ (*propiedades* –PP2 y PP3– del SP7).

Proceso de generalización intranivel

En el nivel 2.2, al realizar el cálculo de las frecuencias esperadas hay un proceso de generalización, así como al calcular el estadístico χ^2, de las tablas de contingencia de 2×2 a las de $r \times c$.

5.1.2.2.3 Prueba de independencia

El estudiante es capaz de determinar si existe asociación entre dos variables (con n atributos cada variable), para lo cual valora en qué medida las frecuencias observadas difieren de la independencia probabilística. Para determinar dicha asociación, intervienen objetos matemáticos primarios del SP6χ^2 (e.g., *conceptos* de variable –CD1–, tabla de contingencia –CD5–, asociación –CD8–). Una vez que el estudiante ha calculado el valor del estadístico y los grados de libertad:

- Comprende qué indica el Estadístico χ^2 para la prueba de independencia.
- Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad y puede interpretarla como una medida de cuan lejos el sistema observado es o no compatible con las bases de independencia probabilística, según lo establecido en el CD10 y la PP2 del SP6χ^2.

Proceso de generalización internivel

El nivel 1.3, donde se trabaja con el SP5χ^2, sólo se puede aplicar cuando se desea saber si existe asociación entre dos variables, con un atributo cada variable; en cambio, cuando se trabaja con los elementos del nivel 2.2, SP6χ^2, se puede analizar la asociación entre dos variables con n atributos cada variable.

5.1.2.2.4 Prueba de homogeneidad

El estudiante es capaz de determinar si dos muestras conocidas pueden pertenecer a la misma población, sin tener conocimiento a priori de la población. Algunos objetos matemáticos primarios del SP9 χ^2 que intervienen son los *conceptos* de muestra –CD1–, muestras independientes –CD2– y población –CD3–. Además:

- Comprende qué indica el Estadístico χ^2 para homogeneidad.
- Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad y puede interpretarla como una medida de que las muestras efectivamente sean muestras aleatorias de la misma población, de acuerdo con lo que se estableció en el CD7, CD8 y la PP6 del SP9χ^2.

5.1.3 Nivel 3

Los indicadores que se encuentran en este nivel se pueden considerar *preformales*, pero con un mayor grado de formalidad que en el nivel 2. Los aspectos que marcan cierto grado preformal son, por ejemplo, la forma de trabajar y comprender la significancia y el lenguaje utilizado en las hipótesis.

En este nivel se espera que el estudiante sea capaz de reconocer la necesidad de utilizar un factor de corrección de continuidad bajo ciertas condiciones de los datos con los que está trabajando. También se espera que argumente su inferencia basándose en el contraste con del valor obtenido, de la probabilidad o del estadístico, y el límite significativo establecido.

5.1.3.1 Subnivel 3.1: Restricciones de las pruebas χ^2

El estudiante puede reconocer las restricciones de las pruebas de hipótesis con el estadístico χ^2 y aplicar el factor de corrección de continuidad cuando es necesario.

- Comprende que la distribución χ^2 brinda una aproximación. Esto es debido a que la distribución χ^2 es una distribución continua, mientras que la distribución que se intenta aproximar es discreta.
- Comprende las limitaciones que causa en la prueba esta aproximación.
  Como nos encontramos trabajando con una variable discreta e intentamos aproximarla a una distribución continua, esta aproximación resulta de utilidad cuando se trabaja con conjuntos de datos que son grandes, pues las desviaciones que causa esta aproximación son muy escasas y se consideran despreciables. Sin embargo, se identifican discrepancias cuando se trabaja con números pequeños, como es muy usual en la práctica. En caso de no realizar la corrección de continuidad cuando se trabaja con números pequeños, la desviación de la aproximación a la realidad es importante.
  Es un caso similar a cuando se intenta aproximar una variable discreta que se ajusta con una distribución binomial e intentamos aproximarla a la distribución normal.
- Reconoce cuándo aplicar el factor de corrección de continuidad (*concepto* factor de corrección de continuidad –CD8– del SP4χ^2).
  Usualmente se realiza la corrección de continuidad para las pruebas χ^2 cuando se tienen frecuencias esperadas inferiores a cinco, aplicar esta corrección mejora la correspondencia entre los resultados del cálculo y la distribución Chi-cuadrada. La corrección de continuidad de Yates es simétrica y consiste en restar 0.5 al valor absoluto de las diferencias de las frecuencias observadas y esperadas.
- Es capaz de hacer explícita la hipótesis nula y alternativa en lenguaje natural.

Como vimos en el nivel dos, se trabajó con la identificación de la hipótesis nula, en este nivel podemos ver un segundo momento del trabajo con las hipótesis, pues en este nivel se pretende que el estudiante sea capaz de expresar en una representación con lenguaje natural tanto la hipótesis nula como la hipótesis alternativa; para realizarlo es importante que el estudiante comprenda la lógica de las hipótesis. Para comprender a la hipótesis nula, podemos partir de lo que comúnmente conocemos como *hipótesis de investigación*, y esto es, simplemente, lo que el investigador establece como probable respuesta a una pregunta de investigación; mientras que la prueba se realiza para valorar la veracidad de la hipótesis de investigación. En el enfoque de Neyman y Pearson, podemos ver a la hipótesis alternativa en contraposición con la hipótesis nula (ver sección 3.4).

Además de los indicadores señalados anteriormente, se proponen para este subnivel algunos indicadores por prueba, los cuales se presentan en las secciones 5.1.3.1.1 y 5.1.3.1.2.

5.1.3.1.1 Prueba de bondad de ajuste

- Comprende y aplica el factor de corrección de continuidad (*conceptos* y *propiedades* del SP4χ^2):

$$\chi_c^2 = \sum \frac{(|\mathrm{e}| - 0.5)^2}{m} = \sum \frac{(|FO - FE| - 0.5)^2}{FE}$$

5.1.3.1.2 Prueba de independencia y homogeneidad

- Comprende y aplica el factor de corrección de continuidad (CD6 del SP4χ^2 y PP5 del SP8χ^2)
  - Para tablas de contingencia de 2 × 2:

$$\sum \chi_c^2 = \sum \frac{\left(\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right)^2 N}{(a+c)(b+d)(a+b)(c+d)}$$

  - Para tablas de contingencia de r × c:

$$\chi_c^2 = \sum \frac{(|FO - FE| - 0.5)^2}{FE}$$

Proceso de generalización intranivel

En el nivel 3.1, al realizar el cálculo del valor del estadístico χ^2 hay un proceso de generalización, de la forma de calcularlo cuando se trabaja con tablas de contingencia de 2×2 a las de $r \times c$.

5.1.3.2 Subnivel 3.2: Conexiones y argumentos

En este subnivel se espera que el estudiante trabaje con la significancia como un límite significativo y el estadístico teórico, y sea capaz de rechazar o no la hipótesis nula.

- Comprende la significancia como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria (definición de significancia –CD8– del SP3χ^2). Esta consideración de un límite significativo puede ser visto como una versión intuitiva del nivel de significancia, el cual lo abordamos en el Nivel 4. También es importante resaltar que la significancia es un concepto clave en la Inferencia Estadística, en este nivel se introduce únicamente como un límite que le permite al estudiante tener un punto crítico para la toma de decisiones.
- Puede encontrar el valor del estadístico teórico en la tabla de probabilidad de la distribución χ^2, con respecto a cierta P y n, y lo compara contra el valor del estadístico χ^2 calculado.
  Por ejemplo, si establece $P = 0.05$ como límite significativo y con $n = 2$, obtendremos un valor del estadístico teórico de $\chi^2_{0.05,1} = 3.841$, este valor es el que se contrasta con el valor del estadístico que se ha calculado con los datos observados. Supongamos que este valor es $\chi^2 = 4.861$, entonces tenemos que $3.841 < 4.861$.
- Es capaz de rechazar o no rechazar la hipótesis nula (y comprende lo que significa) bajo un contraste con un límite preestablecido como desviación significativa, de acuerdo con las reglas de decisión con respecto a la probabilidad y al valor del estadístico (propiedades/proposiciones del SP3χ^2 y del SP4χ^2 respectivamente).
  Según la PP6 del SP3χ^2, la regla de decisión indica que si la probabilidad que obtenemos a partir del estadístico χ^2, que hemos calculado, es menor que el límite que previamente establecimos como significativo, podemos considerar a las desviaciones

como significativas y, por lo tanto, rechazar la hipótesis nula. En caso de suceder lo contrario, no tenemos argumentos suficientes para rechazar la hipótesis nula.

Por su parte, la PP7 del SP4χ^2, la regla de decisión, indica que si el valor del estadístico calculado es más grande que el valor crítico entonces las desviaciones son significativas. Si retomamos el ejemplo anterior, donde tenemos que $3.841 < 4.861$, claramente podemos rechazar la hipótesis nula de acuerdo con esta regla de decisión.

- Logra argumentar, con base en la significancia, por qué rechaza o no rechaza la hipótesis nula.

  El nivel previo para tomar y argumentar la decisión se recurría a la probabilidad, en el sentido de número de casos posibles, en este nivel se espera que el estudiante apoye su decisión en el límite significativo.
- Es capaz de conectar los resultados de la prueba con el contexto del problema.

  Es importante que los estudiantes puedan interpretar en el contexto del problema las conclusiones de sus análisis estadísticos. La interpretación junto con los argumentos, evidencian la comprensión y las conexiones que el estudiante realiza de las nociones estadísticas (Makar, Bakker y Ben-Zvi, 2011; Lane-Getaz, 2013).

En los indicadores de los niveles dos y tres podemos encontrar los tres principios claves (generalización, uso de los datos como evidencia y el empleo de lenguaje probabilístico) que Makar y Rubin (2009) indican como esenciales para la inferencia estadística informal. Estos tres principios claves se encuentran en diversa profundidad en los dos niveles, ya que como señalan en este marco para la inferencia informal, se pueden aplicar estos principios en diversos momentos del currículo escolar con diferente profundidad.

5.1.4 Nivel 4

En este nivel se presentan indicadores que corresponden a un razonamiento inferencial formal. En correspondencia con lo trabajado en niveles previos y tal como se observó en la historia, primero se aborda la prueba de hipótesis con el *valor-p* y posteriormente con la región crítica. Los últimos indicadores corresponden a trabajar el error tipo I y II, y la potencia de la prueba.

5.1.4.1 Subnivel 4.1: Criterio para la toma de decisión

Se espera que el estudiante pueda tomar decisiones basado en las técnicas estadísticas de la metodología de las pruebas de hipótesis.

5.1.4.1.1 Valor-p:

- Puede plantear la hipótesis nula y alternativa con lenguaje simbólico y natural.
  Las hipótesis son uno de los elementos claves en los niveles, esta forma de plantear las hipótesis –con lenguaje simbólico– es el tercer momento, el primero lo abordamos en el nivel 2 con la identificación de la hipótesis nula en el problema y hacerla explícita en lenguaje natural; el segundo en el nivel 3 con el planteamiento de las hipótesis en lenguaje natural.
- Conoce y comprende los valores comunes del nivel de significancia, por ejemplo, 0.05 es muy común y se popularizó en las pruebas que realizó Fisher; también se utilizan 0.10 y 0.01 como nivel de significancia, cuya elección depende de la magnitud del error que se desea asumir.
  En el nivel tres, el estudiante trabaja con la significancia como un límite significativo, mientras que en el nivel 4 el nivel de significancia además de proporcionarle un punto crítico ('límite') para la toma de decisión, es importante que el estudiante comprenda el valor del nivel de significancia como la probabilidad de concluir que existe una desviación o diferencia cuando en realidad no existe, en otras palabras, es la probabilidad de fallar en la estimación.
- Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$.
  Siendo la probabilidad de certeza 1 y el nivel de significancia (α), el nivel de confianza $(1-\alpha)$ como la probabilidad de que el verdadero valor del parámetro (de la población) se encuentre dentro de nuestra estimación.
- Conoce, comprende y es capaz de aplicar la regla de decisión. Si el $valor - p < \alpha,$ se rechaza H_0, (*propiedad/proposición* del SP3χ^2).
  El valor-p se puede interpretar como una medida (probabilidad) de que el valor del estadístico calculado sea posible dada la hipótesis nula, por lo que solemos ver que cuando tenemos un valor-p alto no se rechaza la hipótesis nula, mientras que cuando es bajo se rechaza; ¿qué tan bajo?, depende del nivel de significancia (Cohen, 1994).
  Posibles errores en la interpretación del valor-p:

  > No brinda una probabilidad de que la hipótesis nula sea cierta, así como tampoco de que la hipótesis alternativa sea cierta. Sin embargo, sí nos permite definir el rechazo o no rechazo de la hipótesis nula al contrastarlo con el valor del nivel de significancia (Podworny, 2015; López-Martín, Batanero y Gea, 2018).

Si tomamos como ejemplo valor-p y el nivel de significancia de la Figura 5.4, que es de 0.02746 y 0.05 respectivamente, tenemos que el valor-p es menor que el del nivel de significancia, entonces, de acuerdo con la regla de decisión, se rechaza la hipótesis nula.

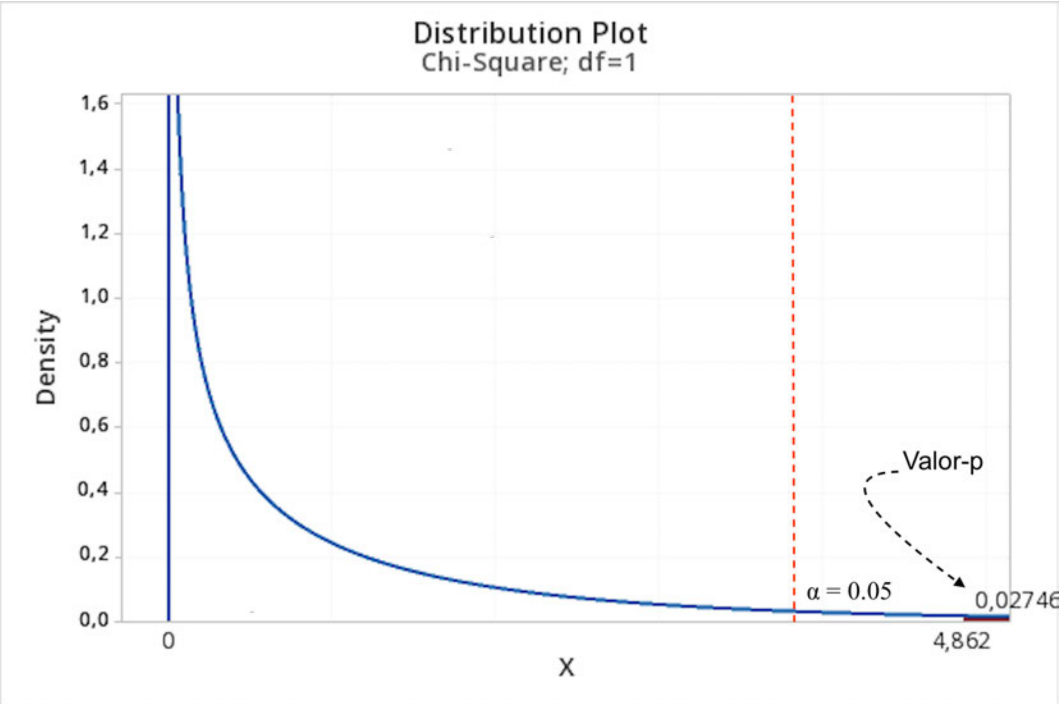


Figura 5.4. Ejemplo de valor-p inferior al nivel de significancia

5.1.4.1.2 Valor crítico:

- Es capaz de identificar el valor teórico del estadístico χ^2, de acuerdo con el nivel de significancia y los grados de libertad.
  Es importante que el estudiante comprenda que este valor teórico del estadístico no depende de la muestra con la que está trabajando, sino de la distribución teórica del estadístico cuando la hipótesis nula es verdadera (Vallecillos, 2002; Vera, Díaz y Batanero, 2011).
- Puede representar gráficamente las regiones de rechazo y de no rechazo, y comprende sus relaciones con el nivel de confianza y la hipótesis nula.
  Según Vallecillos (1994) es común que en las pruebas de hipótesis los estudiantes confundan las regiones de rechazo y no rechazo; además, este error se encuentra relacionado con la falta de comprensión de la lógica que subyace al contraste de hipótesis (Liu y Thompson, 2009; Harradine, Batanero y Rossman, 2011).

A la región de rechazo, que podemos ver en rojo en la Figura 5.5, también se le denomina región crítica, esta región no depende del estadístico calculado con los datos de la muestra, sino del estadístico teórico.

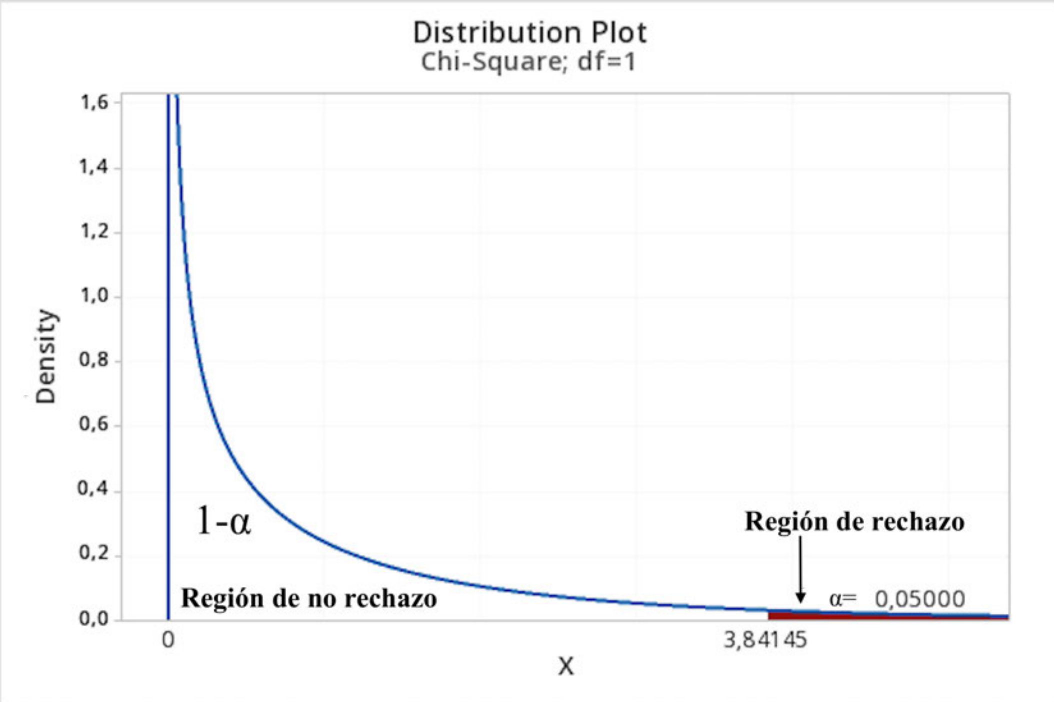


Figura 5.5. Representación de las regiones de rechazo y no rechazo para $\boldsymbol{\chi^2_{0.05,1}}$

- Conoce, comprende y es capaz de aplicar la regla de decisión. Sí $\chi^2 \geq \chi^2_{\alpha,gl}$, se rechaza H_0 (PP7 del SP4χ^2).

  Si tomamos como ejemplo los valores de la Figura 5.6, tenemos que el valor del estadístico calculado excede el valor crítico, 4.862 > 3.84145, por lo que se rechaza la hipótesis nula.

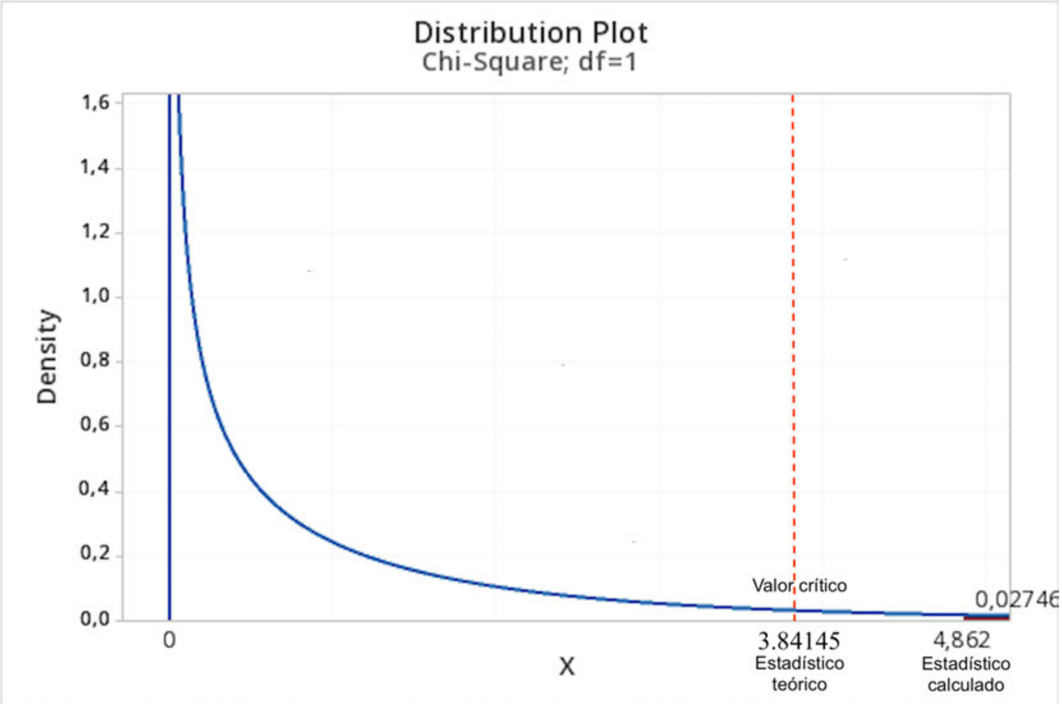


Figura 5.6. Representación gráfica sobre la aplicación de la regla de decisión con valor crítico

- Es capaz de brindar una respuesta al problema utilizando los resultados de la prueba y desarrolla argumentos con fundamentación estadística.

5.1.4.2 Subnivel 4.2: Error tipo I y II, y Potencia de la prueba

El estudiante es capaz de evaluar la validez de los procedimientos y las inferencias realizadas con base en dichos procedimientos, para lo cual:

- Comprende cuándo se comete el Error tipo I y la probabilidad de cometerlo.

$$P\begin{bmatrix} Rechazar\ H_0 | H_0 \\ es\ verdadera \end{bmatrix} = \alpha$$

- Comprende cuándo se comete el Error tipo II y la probabilidad de cometerlo.

$$P\begin{bmatrix} No\ rechazar\ H_0 | H_0 \\ es\ falsa \end{bmatrix} = \beta$$

- Comprende las relaciones entre el Error tipo I y el Error tipo II. Por ejemplo, puede auxiliarse de una representación gráfica y reconocer en ella que cuanto mayor es α, menor es β.

  Diversas investigaciones han planteado la importancia de una comprensión profunda de la probabilidad en las pruebas de hipótesis, especialmente destacan que los estudiantes no confundan las probabilidades condicionales que intervienen en los errores tipo I y tipo II, con probabilidades de un solo evento (Birnbaum, 1982; Falk, 1986; Shaughnessy y Dick, 1991; Sotos et al., 2007).

- Es capaz de calcular la probabilidad de tomar la decisión correcta cuando H_0 es verdadera y cuando H_0 es falsa.
    - $P\begin{bmatrix} No\ rechazar\ H_0 | H_0 \\ es\ verdadera \end{bmatrix} = 1 - \alpha$
    - $P\begin{bmatrix} Rechazar\ H_0 | H_0 \\ es\ falsa \end{bmatrix} = 1 - \beta$

  Solemos buscar un test donde los valores de α y β sean pequeños de forma simultánea, aunque no siempre es posible.
- Comprende la potencia de la prueba y es capaz de calcularla. Por ejemplo, reconoce que la potencia está relacionada con el tamaño de la muestra y el nivel de significancia.

$$P\begin{bmatrix} Decidir\ H_1 | H_1 \\ es\ verdadera \end{bmatrix} = 1 - \beta$$

  La potencia nos permite medir la sensibilidad de la prueba para detectar las desviaciones o diferencias significativas. Se entiende como la probabilidad de rechazar la hipótesis nula cuando es realmente falsa y por ende como la probabilidad de aceptar la hipótesis alternativa cuando es verdadera. Algunas consideraciones a tener en cuenta con la potencia son:
    - Cuando aumentamos el valor del nivel de significación (α) también aumenta la potencia de la prueba, pero debemos tener cuidado pues al aumentar el nivel de significación estamos asumiendo un mayor riesgo de equivocarnos si rechazamos la hipótesis nula, lo que comúnmente se conoce como un 'falso positivo'.
    - También cuando aumentamos el tamaño de la muestra aumenta la potencia de la prueba, al tener una muestra grande es más probable la detección de diferencias significativas.
    - Mientras que cuando aumentamos el tamaño del efecto también aumenta la potencia de la prueba. Cuando tenemos un tamaño del efecto pequeño con una muestra pequeña la potencia es reducida también, pero si tenemos un tamaño del efecto grande con una muestra pequeña puede ser probable obtener resultados significativos. Asimismo, si tenemos un tamaño del efecto pequeño a medida que se incrementa la muestra también aumenta la probabilidad de identificar resultados significativos.
- Puede argumentar sobre la validez de la inferencia realizada.

> Una vez que el estudiante ha realizado la inferencia y ha dado respuesta al problema en el contexto propio del problema, también puede argumentar sobre la prevalencia de la estimación en la población, mediante la potencia de la prueba. Hoekstra (2015) y Riemer y Seebach (2014) han trabajado actividades con datos reales, motivando a los estudiantes a ir más allá de la inferencia por medio de las pruebas de hipótesis, realizando la validación de sus inferencias.

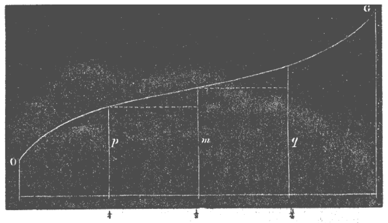
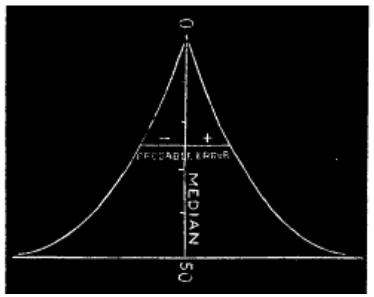
De acuerdo con Wild, Utts y Horton (2018), además de los aspectos técnicos es muy importante el razonamiento que hay detrás de las pruebas de hipótesis, ya que incluso se puede utilizar para tomar decisiones sin datos cuantitativos. Así mismo, comprender la importancia de los valores esperados nos puede ayudar a tomar mejores decisiones incluso de nuestra vida diaria.

5.1.5 Sumario de los niveles de razonamiento inferencial para el estadístico Chi-cuadrada (χ^2)

En la Tabla 5.1 recogemos de forma sucinta la propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, para el estadístico Chi-cuadrada que describimos detalladamente a lo largo del apartado 5.1.

Tabla 5.1. *Propuesta de niveles de razonamiento inferencial para el estadístico Chi-cuadrada*

RII → **RIF**

Nivel 1	Nivel 2	Nivel 3	Nivel 4
1.1 Visualización (S1χ^2) Elementos del SP1χ^2	2.1 Identificar la prueba no paramétrica necesaria para analizar los datos	3.1 Restricciones de las pruebas χ^2	4.1 Criterios para la toma de decisión
• A través de la forma de las gráficas puede conjeturar si los datos siguen una distribución normal.  	• Reconoce el tipo de datos que está trabajando • Comprende el problema a resolver • Comprende la lógica de las pruebas de bondad de ajuste (S1χ^2), independencia (S2χ^2) y homogeneidad (S3χ^2). • Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla.	(S1χ^2, S2 χ^2y S3χ^2) • Comprende que la distribución χ^2 brinda una aproximación. • Comprende las limitaciones que causa en la prueba esta aproximación. • Reconoce cuándo aplicar el factor de corrección de continuidad. (CD8 del SP4χ^2) • Es capaz de hacer explícita la hipótesis nula y alternativa en lenguaje natural.	Valor-p • Puede plantear las hipótesis nula y alternativa con lenguaje simbólico y natural. • Conoce y comprende los valores comunes del nivel de significancia. • Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$. • Conoce, comprende y es capaz de aplicar la regla de decisión. Si el $valor-p<\alpha$ se rechaza H_0. (PP6 del SP3χ^2).
	2.2 Una aproximación de las pruebas con el estadístico χ^2	(S1χ^2) Elementos del SP4χ^2	Valor crítico
	(S1χ^2, S2χ^2 y S3χ^2) • Es capaz de identificar la hipótesis nula que se encuentra implícita en el problema. • Conoce la Distribución χ^2. (PP4 del SP2χ^2)		• Es capaz de identificar y comprende el valor teórico del estadístico χ^2, de acuerdo con α y los gl.

 1.2 Trabajar con datos ($S1\chi^2$) **Elementos del $SP1\chi^2$** Para conjeturar si el grupo de datos sigue una distribución normal utiliza: *El método de intercomparación (CD2, CD3, CD4, CD7, PP2, PP3, PP4, PP5 y PP7 del $SP1\chi^2$) *El método gráfico (CD6 y PP6 del $SP1\chi^2$). Notas: 1. Solo aplica para la distribución normal. 2. En el método gráfico se puede hacer la conexión con el error ($e = FO - FE$) del estadístico χ^2 de Pearson. ($SP1\chi^2$ y $SP2\chi^2$) **Generalización** → $SP1\chi^2$-N1.2 Distribución teórica normal $SP2\chi^2$-N2.2 Cualquier distribución teórica	($S1\chi^2$) Elementos del $SP2\chi^2$ y $SP3\chi^2$ Puede valorar en qué medida se ajusta un grupo de datos observados a cierta distribución teórica preestablecida, por medio del contraste de las frecuencias observadas y las frecuencias esperadas. $\chi^2 = \sum \frac{e^2}{m} = \sum \frac{(FO-FE)^2}{FE}$ • Calcula y comprende qué indica el Estadístico χ^2. (CD3, CD4, CD6, PP3 y PP7 del $SP2\chi^2$) • Calcula y comprende los grados de libertad $k = n - r$. (CD7 y PP3 del $SP3\chi^2$) • Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema de observado. (CD5 y PP4 del $SP2\chi^2$) Nota: Aplica para contrastar los datos con cualquier distribución teórica.	• Comprende y aplica el factor de corrección de continuidad. (CD6 y PP6 del $SP4\chi^2$) $\chi_c^2 = \sum \frac{(\lvert e\rvert - 0.5)^2}{m} = \sum \frac{(\lvert FO-FE\rvert - 0.5)^2}{FE}$ ($S2\chi^2$ y $S3\chi^2$) Elementos del $SP4\chi^2$ y $SP8\chi^2$ • Comprende y aplica el factor de corrección de continuidad. (CD6 del SP4 y PP5 del $SP8\chi^2$) *Para tablas de contingencia de 2×2: $\sum \chi_c^2 = \sum \frac{\left(\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right)^2 N}{(a+c)(b+d)(a+b)(c+d)}$ *Para tablas de contingencia de $r \times$ c: $\chi_c^2 = \sum \frac{(\lvert FO-FE\rvert - 0.5)^2}{FE}$ **3.2 Conexiones y argumentos** • Comprende la significancia. (CD8 del $SP3\chi^2$) • Puede encontrar el valor del estadístico teórico, en la tabla de probabilidad de la distribución χ^2, con respecto a	• Puede representar gráficamente las regiones de rechazo y de no rechazo y comprende sus relaciones con el nivel de confianza y la hipótesis nula. • Conoce, comprende y es capaz de aplicar la regla de decisión. Sí $\chi^2 \geq \chi^2_{\alpha,gl}$ se rechaza H_0. (PP7 del $SP4\chi^2$) • Es capaz de brindar una respuesta al problema utilizando los resultados de la prueba y realiza argumentos con fundamentación estadística. **4.2 Error tipo I y II, y Potencia de la prueba** • Comprende cuándo se comete el Error tipo I y la probabilidad de cometerlo. $P\left[\begin{matrix} Rechazar\ H_0 \mid H_0 \\ es\ verdadera \end{matrix}\right] = \alpha$ • Comprende cuándo se comete el Error tipo II y la probabilidad de cometerlo. $P\left[\begin{matrix} No\ rechazar\ H_0 \mid H_0 \\ es\ falsa \end{matrix}\right] = \beta$ • Es capaz de calcular la probabilidad de tomar la decisión correcta cuando: * H_0 es verdadera * H_0 es falsa
1.3 Asociación intuitiva ($S2\chi^2$)	($S2\chi^2$ y $S3\chi^2$)		

Elementos del SP5χ^2

Puede establecer si existe asociación entre dos variables, con un atributo de cada variable, por medio del coeficiente de asociación Q. (CD1, CD2, CD3, CD4, CD7, PP1, PP2 y PP4)

- Comprende conceptos básicos de probabilidad.
- Es capaz de relacionar Q con el coeficiente de correlación de Pearson. (CD5 y CD6)
- Trabaja con tablas de contingencia de 2 × 2.

	(B)	(β)
(A)	(AB)	$(A\beta)$
(α)	(αB)	$(\alpha\beta)$

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

- Calcula el coeficiente de asociación e interpreta su valor

Generalización

SP5-N1.3 Asociación entre dos variables, con un atributo de cada variable

SP6-N2.2 Asociación entre dos variables, con n atributos cada variable

- Comprende y utiliza frecuencia observada con tablas de $r \times$ c. (CD3 del SP6χ^2)

 *Para tablas de contingencia de 2 × 2:

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d$

- Comprende y utiliza la frecuencia esperada bajo independencia probabilística. (CD4 del SP6 y PP2 del SP11χ^2)

 *Para tablas de contingencia de 2 × 2:

$\frac{(a+b)(a+c)}{a+b+c+d}$	$\frac{(a+b)(b+d)}{a+b+c+d}$	$a+b$
$\frac{(c+d)(a+c)}{a+b+c+d}$	$\frac{(c+d)(b+d)}{a+b+c+d}$	$c+d$
$a+c$	$b+d$	$a+b+c+d$

2

 *Para tablas de contingencia de $r \times$ c:

$$e_{ij} = \frac{n_i \times n_j}{n}$$

- Calcula el estadístico χ^2 (PP5 del SP6χ^2):

 *Para tablas de contingencia de 2 × 2

cierta P y gl. Y lo compara contra el valor χ^2 calculado.

- Es capaz de rechazar o no rechazar la hipótesis nula (y comprende lo que significa) bajo el contraste con un límite preestablecido como desviación significativa. (PP6 del SP3 y PP7 del SP4χ^2)
- Logra argumentar, con base en la significancia por qué rechaza o no rechaza la hipótesis nula.
- Es capaz de conectar los resultados de la prueba con el contexto del problema.

- Comprende las relaciones entre los dos tipos de error
- Comprende la potencia de la prueba

$$P\begin{bmatrix} Decidir\ H_1|H_1 \\ es\ verdadera \end{bmatrix} = 1 - \beta$$

 y es capaz de calcularla.
- Puede argumentar sobre la validez de la inferencia realizada.

	$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+c)(b+d)(a+b)(c+d)}$ *Para tablas de contingencia de $r \times$ c $\chi^2 = \sum \left\{ \frac{(n_{uv}-v_{uv})^2}{v_{uv}} \right\} = \sum \frac{(FO-FE)^2}{FE}$ • Calcula y comprende los grados de libertad $k = (c-1)(r-1)$. (PP2 y PP3 del SP7χ^2)		
	(S2χ^2) Elementos del SP6χ^2 y SP7χ^2		
	Es capaz de determinar si existe asociación entre dos variables (con n atributos cada variable), para lo cual valora en qué medida las frecuencias observadas difieren de la independencia probabilística. (CD1, CD5, CD8 del SP6χ^2) • Comprende qué indica el Estadístico χ^2 para la prueba de independencia. (PP5 del SP6χ^2) • Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad y puede interpretarla como una medida de cuan lejos el sistema observado es o no compatible con las bases de		

	independencia probabilística. (CD10 y PP2 del SP6χ^2)		
	(S3χ^2) Elementos del SP6, SP7, SP9 y SP11		
	Es capaz de determinar si dos muestras conocidas pueden pertenecer a la misma población, sin tener conocimiento a priori de la población. (CD1, CD2, CD3 del SP9χ^2) • Comprende qué indica el Estadístico χ^2 para homogeneidad. • Utiliza la tabla de probabilidad de la distribución χ^2 para determinar la probabilidad y puede interpretarla como una medida de que las muestras efectivamente sean muestras aleatorias de la misma población. (CD7, CD8 y PP6 del SP9χ^2)		

PARTICULAR **GENERAL**

5.2 Niveles de razonamiento inferencial para el estadístico t-Student

A partir de los aportes de la literatura científica en didáctica de la estadística y de los objetos matemáticos primarios y procesos identificados en cada uno de los significados parciales resultantes del estudio histórico-epistemológico, fue posible caracterizar cuatro niveles de razonamiento inferencial para el estadístico t-Student.

5.2.1 Nivel 1

Los indicadores de este nivel corresponden a un razonamiento inferencial informal, éstos van más allá de simples interpretaciones de los gráficos o cálculos de medidas como la media; están enfocados en razonar inferencialmente y realizar conjeturas a partir de los datos. Los argumentos que apoyan las conjeturas están basados en la variación interna de los datos, inicialmente dicha variación es observable en las gráficas y posteriormente en el análisis que se realiza de una o dos muestras. Los objetos matemáticos que intervienen para analizar los datos, así como para establecer y argumentar las conjeturas pertenecen a Estadística Descriptiva y Probabilidad.

5.2.1.1 Subnivel 1.1: Visualización

Se espera que el estudiante sea capaz de identificar la variación interna de un conjunto de datos a través de observar e interpretar la forma y los cuartiles de la gráfica (e.g., Figura 5.7), y razone sobre cómo podría comportarse la población de donde se extrajeron estos datos.

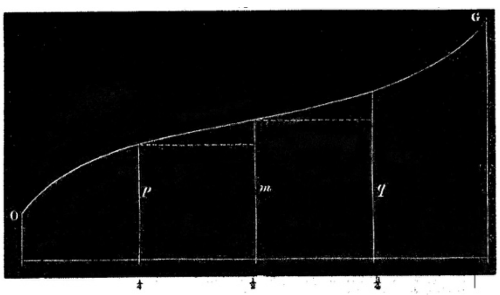


Figura 5.7. Ojiva resultante del método de intercomparación (fuente: Galton, 1875, p.36)

Cabe destacar que los objetos matemáticos primarios que se encuentran en este nivel corresponden al Significado parcial 1 ($SP1_t$) lo cual refiere al significado de prueba de media de una muestra del estadístico t-Student ($S1_t$); por ejemplo: cuartiles –CD3–, valor medio –CD4– y la simetría de la serie –PP7–.

En este nivel, además de identificar las características de la muestra que puede observar mediante la gráfica, es importante trabajar con la relación muestra-población, tal como lo expresan diversas investigaciones (e.g., Zieffler et al., 2008; de Vetten, Schoonenboom, Keijzer, y van Oers, 2018). También podría trabajar con gráficos, como gráficos de caja y bigote, de dos conjuntos de datos donde el estudiante además de identificar las características de los gráficos los comparará para tomar una decisión (e.g., Pfannkuch, 2007; Zieffler et al., 2008; Arnold, Pfannkuch, Wild, Regan, y Budgett, 2011).

5.2.1.2 Subnivel 1.2: Trabajar con datos de una muestra

El estudiante puede utilizar el método de intercomparación, donde identifica valores representantes (e.g., cuartiles, octiles, deciles y percentiles); realiza la gráfica correspondiente y compara e interpreta la dispersión de los intervalos, los intervalos son formados por valores representantes, como los cuartiles (e.g., Figura 5.8).

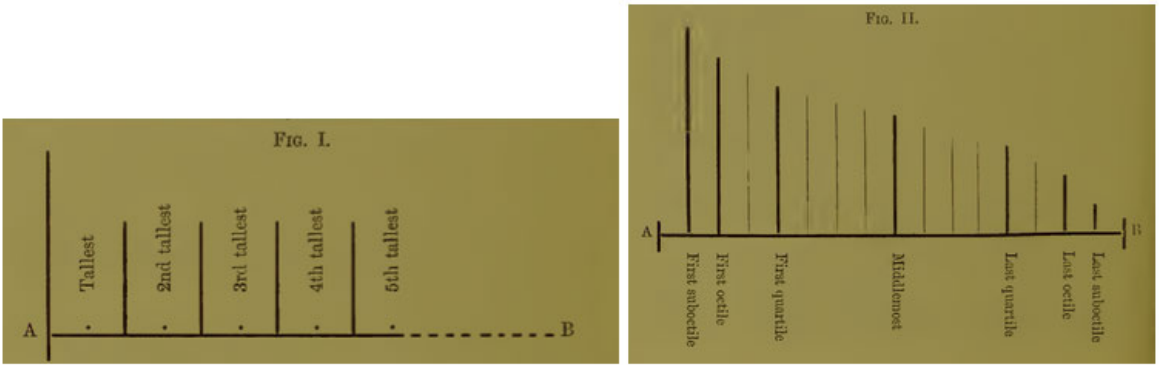


Figura 5.8. Gráficos del método de intercomparación con diversos representantes (fuente: Galton, 1880, p. 8)

Los *conceptos/definiciones* que se movilizan son error probable como una medida de variabilidad de la serie observada –CD1–, ojiva –CD2–, cuartiles –CD3–, valor medio –CD4– y percentiles –CD5–; mientras que ejemplo de *propiedades/proposiciones* son m en $\frac{1}{2}$ o $0°$ –PP3–, p en $\frac{1}{4}$ o $-25°$ –PP4–, q en $\frac{3}{4}$ o $25°$ –PP5–, error probable como $q - m$ y $m - p$ –PP6–, y la propiedad –PP7– sobre la simetría de la serie como $q - m = m - p$; todos ellos correspondientes al $SP1_t$.

Es importante que, posteriormente, el estudiante establezca una conjetura de la población fundamentada en la interpretación de los gráficos y en la dispersión de los intervalos que ha calculado.

5.2.1.3 Subnivel 1.3: Trabajar con datos de dos muestras

El estudiante puede establecer una conjetura argumentada en la variación interna de las dos muestras. Para lograrlo:

- Debe ser capaz de analizar bajo el método de intercomparación cada una de las muestras y comparar los valores representantes (e.g., cuartiles) de cada una de las muestras, puede realizar gráficos (e.g., diagrama de caja y bigote) y conjetura las semejanzas y/o diferencias entre las muestras.

Esto se realiza a manera de conexión con el nivel 1.2 y posteriormente dar paso al siguiente indicador.

- Puede analizar la variación interna de los datos a través de las gráficas que elabora (e.g., diagramas de caja y bigote, y ojiva) y de la fluctuación $\left(2\frac{\sum e^2}{n}\right)$, donde n es la suma de las diferencias $(y_i - x_i)$ y $e = (y_i - x_i) - (\bar{y} - \bar{x})$.

Intervienen elementos primarios del SP6$_t$ como los *conceptos/definiciones* de población –CD2–, media –CD3– y fluctuación entendida como una medida de la variación de una serie de datos con respecto a otra serie –CD4–; mientras que algunas *propiedades/proposiciones* que intervienen son $\bar{x} = \frac{1}{n}\sum x_i$ –PP4– y fluctuación como $2\frac{Se^2}{n}$, donde n es la sumatoria de las diferencias –PP6–.

En el nivel 1 podemos encontrar características del marco de trabajo de Zieffler et al. (2008), como es el caso de los tres componentes del RII (hacer juicios o predicciones, usar o integrar el conocimiento previo y articular evidencia basada en argumentos), y también la similitud con las tareas 1 y 2 (estimar y esbozar un gráfico de la población y comparar dos muestras de datos). En diversas investigaciones se ha proclamado la importancia de trabajar con situaciones en contextos cercanos a los estudiantes y en especial al promover un RII, ya que este tipo de razonamiento no contempla únicamente el conocimiento estadístico de los estudiantes sino también su conocimiento informal (e.g., Makar y Ben-Zvi, 2011; Ben-Zvi y Aridor-Berger, 2016; Bakker, Ben-Zvi y Makar, 2017).

Al abordar los indicadores de este nivel es importante que el estudiante también razone sobre ideas como variabilidad y distribución, error, muestreo e inferencia y predicción; estas ideas se han promovido dentro de la inferencia informal en diversas investigaciones (e.g., Pfannkuch, Arnold, y Wild, 2015; Noll y Hancock, 2015; Garfield, Le, Zieffler y Ben-Zvi, 2015; de Vetten, Schoonenboom, Keijzer, y van Oers, 2018).

5.2.2 Nivel 2

En este nivel se encuentran indicadores que corresponden a un nivel de tipo preformal, ya que podemos identificar rasgos del RII tales como: la forma en que se utiliza la probabilidad, como una medida de certidumbre asociada a un evento sin compararla con un nivel de significancia o un límite preestablecido; y cómo se observa la hipótesis nula enmarcada dentro del problema (se puede considerar una primera aproximación a la hipótesis ya que esta se encuentra implícita). Esta forma de utilizar la probabilidad se encuentra en los inicios de las pruebas t-Student (ver sección 4.1.2).

5.2.2.1 Subnivel 2.1: Identificar la prueba paramétrica necesaria para analizar los datos

Para que el estudiante identifique la prueba t-Student adecuada para analizar los datos que tiene y pueda dar solución al problema, es necesario que los siguientes indicadores estén presentes:

- Reconoce el tipo de datos que está trabajando y que provienen de poblaciones normales (e.g., número de muestras, tamaño de la muestra, son datos cualitativos o cuantitativos, estadístico –muestral–, parámetro –poblacional– y tipo de muestreo).
- Comprende el problema a resolver.
  Según Schoenfeld (1982), si el estudiante no tiene acceso a un esquema de solución no tiene un camino claro de cómo iniciar. Entonces, es posible que no comprenda completamente el problema y que la alternativa que tome sea explorarlo hasta que se sienta cómodo con el problema y que posiblemente el estudiante intente emparejarlo con problemas familiares. Para Schoenfeld, las heurísticas de resolución de problemas son personales y dependen del conocimiento previo y de la comprensión del problema por parte del estudiante (Liljedahl, Santos-Trigo, Malaspina y Bruder, 2016). Así pues, este criterio es de suma importancia para que el estudiante pueda establecer un camino para resolver el problema.
- Comprende la lógica de las pruebas t-Student para una muestra ($S1_t$) y para dos muestras ($S3_t$). Por ejemplo, la prueba t-Student para una muestra se utiliza para determinar si la media de una población, que se distribuye de acuerdo a la normal, es igual a un valor μ_0; mientras que la prueba t-Student para dos muestras busca

determinar si las medias de dos poblaciones distribuidas de forma normal son iguales, en otras palabras, si no existe diferencia significativa entre las medias.

De acuerdo con Vallecillos (2002), Thompson, Liu y Saldanha (2007) y Liu y Thompson (2009) comprender la lógica de las pruebas de hipótesis es uno de los principales desafíos que tienen los estudiantes y profesores al trabajar con las pruebas de hipótesis; además, si se tienen dificultades con la lógica de las pruebas de hipótesis se puede presentar un desafío mayor al comprender las inferencias producto de este tipo de pruebas.

- Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla. El estudiante puede no únicamente reconoce los datos con los que está trabajando y comprende de qué tratan las pruebas t-Student para una y dos muestras, además es capaz de establecer las relaciones entre ellos y el problema para decidir cuál es la prueba apropiada para resolver el problema.

5.2.2.2 Subnivel 2.2: Una aproximación a las pruebas con el estadístico t-Student

Una vez que el estudiante ha identificado la prueba, para dar respuesta al problema planteado debe poner en juego lo siguiente:

- Es capaz de identificar la hipótesis que se encuentra implícita en el problema.

  Como se observó en la historia, lo que hoy conocemos como hipótesis nula se encontraba implícita en el problema, muchas veces en forma de pregunta, desde que surge la prueba de bondad de ajuste con el estadístico χ^2 y se extendió a otras pruebas. También se puede encontrar esta forma de 'hipótesis de investigación' como primer componente en el ciclo de investigación estadística PPDAC (Wild y Pfannkuch, 1999). Diversos estudios han reportado las bondades de este tipo de hipótesis, a través de preguntas o conjeturas (e.g., Pfannkuch y Wild, 2004; Pfannkuch et al., 2016; Stohl, Angotti y Tarr, 2010).

- Conoce la distribución t-Student (e.g., es simétrica alrededor del valor cero, es más dispersa que la distribución normal y la relación grados de libertad-dispersión de la curva –conforme incrementan los grados de libertad, la dispersión de la curva t disminuye–) de acuerdo con la *propiedad/proposición* PP2 del $SP3_t$. Algunas investigaciones (e.g., Bakker y Gravemeijer, 2004; Reading y Reid, 2006; Rossman, 2008; Dinov, Palanimalai, Khare y Christou, 2018), se han apoyado con recursos tecnológicos para realizar actividades con simulaciones para promover la comprensión de la noción de distribución. En la Figura 5.9 podemos observar algunas de las características anteriormente señaladas y cómo se comporta la distribución t-Student cuando se incrementan los grados de libertad, lo cual va en relación con el siguiente indicador.

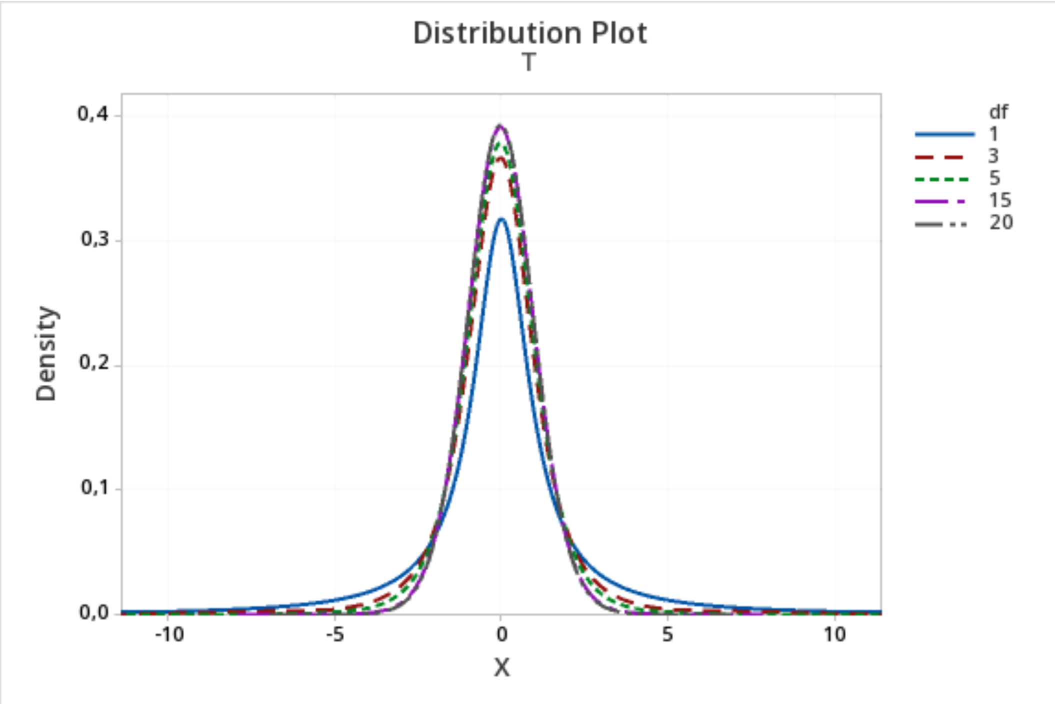


Figura 5.9. Distribución t-Student

- Comprende la relación entre la distribución normal y la t-Student (e.g., a medida que los grados de libertad tienden a infinito la curva de la t se aproxima más a la normal), de acuerdo con lo que se establece en las *propiedades/proposiciones* PP2 del $SP1_t$, PP2 del $SP2_t$ y PP2 del $SP3_t$. Puede auxiliarse de simulaciones tal como se hace para comprender la noción de distribución.

Además de los indicadores anteriores, compartidos por las pruebas t ($S1_t$ y $S3_t$), otros indicadores a considerar por prueba para este subnivel, se presentan en las secciones 5.2.2.2.1 y 5.2.2.2.2.

5.2.2.2.1 Prueba t-Student de la media de una muestra ($S1_t$)

El estudiante puede valorar si la media de una población es igual al valor μ_0, por medio del estadístico t-Student. Para hacer esta valoración, el estudiante:

- Calcula y comprende qué indica el estadístico t-Student, utilizando *conceptos/definiciones* tales como muestra –CD2–, población –CD3–, media –CD4– y desviación estándar –CD5–, del $SP2_t$, y propiedades tales como $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$, correspondientes al $SP3_t$.
- Calcula y comprende los grados de libertad, de acuerdo con la *propiedad/proposición* cinco del $SP3_t$ que indica que $gl = n - 1$.
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de la población se encuentre fuera del rango $\pm t$ (*concepto/definición* de probabilidad –CD7– y *propiedad* de distribución del estadístico t-Student –PP2– del $SP3_t$). Konold et al. (2011), realizan una crítica sobre la práctica de presentar a los estudiantes la probabilidad teórica y la experimental, e indican que primero es necesario fomentar un razonamiento probabilístico básico para que los estudiantes puedan realizar inferencias informales. En este sentido, consideramos que es necesario utilizar la probabilidad como una medida de certidumbre asociada a un evento previo a compararla con un límite preestablecido o un nivel de significancia.

Los elementos matemáticos que intervienen en estos indicadores pertenecen al $SP2_t$ y $SP3_t$.

Proceso de generalización intranivel

El nivel 2.2, hay un proceso de generalización en el uso del estadístico t-Student, al inicio se utiliza para una muestra y posteriormente podemos ver su uso para dos muestras que cuentan con tamaños de las muestras y las varianzas iguales; esto considera objetos matemáticos y procesos propios de los significados $SP3_t$, $SP7_t$ y $SP8_t$.

5.2.2.2.2 Pruebas t-Student para dos muestras ($S3_t$)

Una vez que el estudiante ha identificado que los tamaños de las muestras y las varianzas son iguales, puede valorar si las medias de dos poblaciones (distribuidas normalmente) son iguales. Para llevar a cabo esta valoración, los siguientes indicadores son necesarios:

- Comprende y es capaz de obtener la varianza y desviación estándar agrupada. Esto a partir de los objetos matemáticos primarios del $SP8_t$, tales como el *concepto/definición* de desviación estándar agrupada (como un método para calcular la desviación estándar agrupando las sumas de cuadrados de las dos muestras y dividiendo por el número total de grados de libertad) –CD6–; y la *propiedad/proposición* de la varianza agrupada –PP4–, entendida como:

$$s^2 = \frac{1}{(n_1 - 1)(n_2 - 1)}\left(\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2\right) = \frac{1}{2}\left(s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2\right)$$

  También el *concepto/definición* (del $SP7_t$) de tamaño muestral –CD10–, entendido como el número de sujetos que componen la muestra, es decir, el número de observaciones del subconjunto de la población.

- Comprende el error estándar, como la desviación estándar de la distribución de la diferencia entre las medias (*concepto/definición* once del $SP7_t$).
- Comprende qué indica el estadístico t-Student para dos muestras, a partir de *conceptos/definiciones* (del $SP2_t$) como muestra –CD2–, población –CD3– y media –CD4–; y la *propiedad/proposición* seis (del $SP8_t$) $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(2/n)}} = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{2/n}}$.
- Comprende y calcula los grados de libertad como $gl = 2n - 2$ (PP5 del $SP8_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la diferencia de las medias se encuentre fuera del rango $\pm t$, de acuerdo con los objetos matemáticos CD7 del $SP7_t$ y PP2 del SP3.

5.2.3 Nivel 3

Los indicadores que se encuentran en este nivel se pueden considerar preformales, pero con un mayor grado de formalidad que en el nivel 2, dado que se tiene una visión más amplia de las posibilidades para trabajar con las pruebas t-Student. Los aspectos que marcan cierto grado preformal son, por ejemplo, la forma de plantear las hipótesis (el lenguaje) y la forma de trabajar y comprender la significancia.

5.2.3.1 Subnivel 3.1: Restricciones de las pruebas t-Student

El estudiante reconoce las restricciones que se tienen al realizar pruebas de hipótesis con el estadístico t-Student.

- Comprende que puede trabajar con tamaños muestrales pequeños si las muestras provienen de distribuciones normales. Como vimos en el Capítulo 4, Gosset puntualizó que es crucial para las pruebas t-Student que las muestras provengan de poblaciones normalmente distribuidas.
- Identifica si las muestras son dependientes o independientes y comprende sus implicaciones.
  Trabajamos con muestras independientes cuando tenemos dos muestras aleatorias a partir de las dos poblaciones a ser comparadas y no se ha establecido relación alguna entre las unidades de las muestras; en cambio, diremos que se trata de muestras dependientes o pareadas si previo al análisis se han emparejado elementos (individuos) de una muestra con los elementos de la otra muestra, por ejemplo, cuando un grupo de individuos ha sido evaluado en dos ocasiones diferentes.
- Comprende las implicaciones cuando se tiene igualdad o desigualdad de varianzas y del tamaño de las muestras.
  Inicialmente, para el caso de las pruebas t-Student con dos muestras, se trabajaron con igualdad de varianzas y de tamaño muestral. Sin embargo, al expandir el uso de las pruebas t-Student, Fisher y Welch propusieron variaciones del estadístico t-Student (por ejemplo, la forma de calcular la desviación estándar y el error estándar) y de los grados de libertad, variaciones que permitieron optimizar los resultados de las pruebas t-Student para dichos casos.
- Es capaz de hacer explícita la hipótesis nula y alternativa en lenguaje natural.
  El estudiante además de identificar la hipótesis de investigación en el problema (criterio del nivel dos), en este nivel puede expresar en lenguaje natural la hipótesis nula y alternativa, lo cual constituye un segundo momento en la progresión de la formulación de las hipótesis estadísticas. Consideramos que expresar las hipótesis en lenguaje natural, previo a utilizar un lenguaje simbólico, puede ayudar al estudiante a comprender la lógica de las hipótesis estadísticas.

Además, en las secciones 5.2.3.1.1, 5.2.3.1.2, 5.2.3.1.3 y 5.2.3.1.4, se proponen algunos indicadores por prueba a considerar dentro de este subnivel.

5.2.3.1.1 Muestras dependientes ($S3_t$)

Una vez que el estudiante ha identificado que se trata de muestras dependientes, puede valorar si la media de las diferencias es igual al valor μ_0. Para ello:

- Calcula y comprende el estadístico t-Student como $t = \frac{\bar{x}_D - \mu_0}{s_D/\sqrt{n}}$ (a partir de lo expresado en el nivel previo y con las *propiedades/proposiciones* media de la diferencia –PP3–, desviaciones –PP4– y estadístico t –PP6–, del $SP9_t$).
- Calcula y comprende los grados de libertad como $gl = n - 1$ (PP5 del $SP3_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de las diferencias se encuentre fuera del rango $\pm t$ (definición de probabilidad del $SP9_t$).

> *Proceso de generalización internivel*
>
> El nivel 2.2, que considera objetos matemáticos y procesos propios del $SP2_t$ y $SP3_t$, se puede aplicar cuando se trabaja con una muestra; en cambio, bajos los elementos del nivel 3.1, $SP9_t$, se puede trabajar con dos muestras que sean dependientes o emparejadas.

5.2.3.1.2 Muestras independientes con $n_1 \neq n_2$ ($S3_t$)

Si el estudiante ya ha identificado que se trata de muestras independientes y que los tamaños de las muestras son desiguales, puede valorar si las medias de dos poblaciones son iguales:

- Comprende y es capaz de obtener la desviación estándar agrupada, de acuerdo con *conceptos/definiciones* tales como la desviación estándar agrupada –CD6 del $SP8_t$– y tamaño muestral –CD10 del $SP7_t$– y *propiedades/proposiciones* (propias del $SP7_t$) como las desviaciones al cuadrado $S_i = S(x - \bar{x})^2$ y la desviación estándar $s = \sqrt{\frac{(n_1-1)(s_{x_1}^2)+(n_2-1)(s_{x_2}^2)}{n_1+n_2-2}}$ (PP4 y PP5 respectivamente).

  Cuando tenemos dos muestras de poblaciones diferentes y las medias de las poblaciones pueden ser diferentes, estimamos el valor de la varianza agrupada bajo el supuesto de igualdad de varianzas poblacionales. La varianza agrupada puede darnos una mejor estimación que la estimación por varianza muestral (de cada muestra).

- Calcula y comprende el estadístico t-Student como $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ (propiedad del $SP7_t$).
- Comprende y calcula los grados de libertad como $gl = n_1 + n_2 - 2$ (propiedad del $SP7_t$).
- Comprende y es capaz de establecer relaciones con la prueba para dos muestras del nivel 2.2, como las relaciones del estadístico t-Student, los grados de libertad y la desviación estándar agrupada.

  La reflexión a la cual se pretende llevar al estudiante hace referencia a una integración de las pruebas como un sistema y no sólo como fórmulas o herramientas aisladas; esto en un sentido similar que en el razonamiento estadístico (e.g., Ben-Zvi y Garfield, 2004) donde se busca una integración de las grandes nociones de estadística para dar sentido a la información con la que se está trabajando, así como conectar dichas nociones. Esta propuesta va encaminada hacia una integración progresiva de los elementos que intervienen en las pruebas.

Proceso de generalización internivel

En el nivel 2.2, se trabaja con el estadístico t-Student ($SP8_t$), los grados de libertad y la desviación estándar agrupada, sólo cuando $\mathrm{n}_1 = \mathrm{n}_2$. Mientras que en el nivel 3.1, se puede trabajar sin la restricción de la igualdad de los tamaños muestrales, es decir cuando $\mathrm{n}_1 \neq \mathrm{n}_2$ ($SP7_t$).

5.2.3.1.3 Muestras independientes con $n_1 \neq n_2$ y $\sigma_1^2 \neq \sigma_2^2$ o $n_1 = n_2$ y $\sigma_1^2 \neq \sigma_2^2$ ($S3_t$)

El estudiante ha sido capaz de identificar que está trabajando con dos muestras que tienen diferentes tamaños y diferentes varianzas; ahora, para valorar si las medias de dos poblaciones son iguales es necesario que el estudiante refleje en su práctica los siguientes indicadores:

- Es capaz de estimar la varianza de las dos poblaciones por medio de las muestras y comprende por qué no se utiliza la varianza agrupada, como en las otras pruebas t-

Student para dos muestras. Lo anterior a partir del uso, principalmente, de *conceptos/definiciones* tales como desviación estándar (CD5 del $SP12_t$) y *propiedades/proposiciones* (del $SP12_t$) tales como la varianza estimada –PP6– y varianza de la muestra –PP7– $\sigma^2 = \frac{1}{n_1}s_1^2 + \frac{1}{n_2}s_2^2$, $s_i^2 = \frac{1}{n_i-1}\sum(x_a - \bar{x})^2$.

- Calcula y comprende el estadístico t-Student, para dos muestras independientes con desigualdad en las varianzas poblacionales, como $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{1}{n_1}s_1^2+\frac{1}{n_2}s_2^2}}$ (*propiedad/proposición* nueve del $SP12_t$).
- Comprende y calcula los grados de libertad como $gl = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)}+\frac{s_2^4}{n_2^2(n_2-1)}}$ (*propiedad/proposición* ocho del $SP12_t$).
- Comprende por qué este método también lo puede aplicar cuando se tienen tamaños muestrales iguales.

5.2.3.1.4 Muestras independientes con coeficientes de regresión ($S3_t$)

El estudiante puede comparar los coeficientes de regresión cuando las series de la variable independiente no son idénticas y valorar la diferencia de los coeficientes de regresión.

- Puede obtener el coeficiente b y comprende su significado, a partir del uso de *conceptos/definiciones* tales como variable independiente –CD13–, variable dependiente –CD14– y regresión –CD15–, y las *propiedades/proposiciones* función de regresión $Y = a + b(x - \bar{x})$ –PP4– y $b = \frac{S\{y(x-\bar{x})\}}{S(x-\bar{x})^2}$ –PP9– (objetos matemáticos primarios propios del $SP10_t$).
- Calcula y comprende qué indica el estadístico t-Student de acuerdo con $t = \frac{b\prime-b}{\sqrt{(s^2/S_A(x-\bar{x})^2)+(s^2/S_B(x-\bar{x})^2)}}$ (*propiedad/proposición* once del $SP11_t$).
- Comprende y calcula los grados de libertad como $gl = n_A + n_B - 2$ (PP8 del $SP11_t$).
- Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la diferencia de los coeficientes de regresión se encuentre fuera del rango $\pm t$. En un sentido similar a lo expresado en las otras pruebas para dos muestras. Esto, a partir del uso de objetos matemáticos como el CD7 del SP7 y de la PP2 del SP3.

- Reconoce que esta prueba sigue la misma lógica de la prueba de diferencia de medias de dos muestras.

5.2.3.2 Subnivel 3.2: Conexiones y argumentos

En este subnivel el estudiante trabaja con la significancia en una etapa previa a lo formal, así como se trabajó en las primeras décadas del siglo XX, antes de la metodología de inferencia. También hemos adaptado de la historia, la forma en que el estudiante puede rechazar o no la hipótesis nula, en las décadas mencionadas no hablaban de rechazar una hipótesis propiamente, pero sí hacían la relación entre la pregunta o suposición (Ho) y el límite como desviación significativa para dar respuesta al problema.

- Comprende la significancia como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria (*concepto/definición* ocho del SP2t).
  En 1925, cuando Fisher trabajó la significancia, buscaba evidencia en contra de la hipótesis nula, en este criterio se trabaja de dicha forma. Significancia en el sentido de un límite significativo a partir del cual se debemos considerar seriamente que si es probable que exista una diferencia. Esta forma de abordar la significancia puede considerarse como un momento previo o una versión intuitiva del nivel de significancia.
- Puede encontrar el valor del estadístico teórico, en la tabla de probabilidad de la distribución t-Student, con respecto a cierta P y n; y lo contrasta con el valor t calculado. Esta era una práctica que realizaba Fisher (1934) en algunas pruebas t-Student para fundamentar la respuesta al problema.
- Es capaz de rechazar o no rechazar la hipótesis nula (y comprende lo que significa) bajo un contraste con un límite preestablecido como desviación significativa –esto a partir de la regla de decisión con respecto al contraste del estadístico teórico y del estadístico calculado (PP7 del $SP3_t$), y la regla de decisión con respecto al contraste de la probabilidad obtenida a partir del estadístico calculado y un límite preestablecido (PP7 del $SP8_t$)–.
- Logra argumentar, con base en la significancia, por qué rechaza o no rechaza la hipótesis nula.
  En este indicador el estudiante puede recurrir a los dos anteriores, es decir, no basta con que el estudiante justifique con la regla de decisión. Es importante que el estudiante reconozca por qué se rechaza la hipótesis nula si el valor estadístico teórico es mayor que el valor del estadístico teórico (por mencionar uno de los casos).

- Es capaz de conectar los resultados de la prueba con el contexto del problema. Es necesario que se fomenten espacios para que los estudiantes puedan elaborar conclusiones en términos del problema inicial, y que proporcionen argumentos con base en los resultados de las pruebas, ya que de acuerdo con los ciclos de investigación estadística es necesario que una vez analizados los datos, estos puedan ser interpretados para elaborar conclusiones, como es el caso de lo expuesto en la dimensión uno denominada el ciclo de la investigación (PPDAC) del modelo propuesto por Wild y Pfannkuch (1999), como parte de la teoría de pensamiento estadístico.

Como hemos podido observar, en los indicadores de los niveles dos y tres se encuentran inmersos los tres principios clave (generalización, uso de los datos como evidencia y el empleo de lenguaje probabilístico) del marco de inferencia informal de Makar y Rubin (2009), aunque en diferente profundidad.

5.2.4 Nivel 4

Los indicadores que se presentan en este nivel corresponden a un razonamiento inferencial formal. Se espera que el estudiante pueda tomar decisiones basado en las técnicas estadísticas de la metodología de las pruebas de hipótesis, y sea capaz de fundamentarlas y validarlas estadísticamente.

5.2.4.1 Subnivel 4.1: Criterio para la toma de decisión

En este subnivel se espera que el estudiante comprenda y aplique los criterios para rechazar o no rechazar la hipótesis nula.

5.2.4.1.1 Valor-p

- Puede plantear las hipótesis nula y alternativa con lenguaje simbólico (e.g., $Ho: \mu = \mu_0$ y $Ha: \mu \neq \mu_0$, $Ho: \mu = \mu_0$ y $Ha: \mu > \mu_0$, y $Ho: \mu = \mu_0$ y $Ha: \mu < \mu_0$).
  Como señalamos anteriormente, el planteamiento de las hipótesis estadísticas es uno de los elementos claves de esta propuesta de niveles y se trabaja gradualmente en los niveles previos (2 y 3). En el nivel dos con la identificación de la hipótesis de investigación en el problema y en el nivel tres con el planteamiento de ambas hipótesis en lenguaje natural.
- Conoce y comprende los valores comunes del nivel de significancia, por ejemplo, 0.05 es muy común, también se utilizan 0.10 y 0.01; estos valores para el nivel de

significancia se han consensuado en la comunidad científica. La elección del nivel de significancia depende de la magnitud del error que se desea asumir.

Podemos interpretar el nivel de significancia como la probabilidad de equivocarnos que estamos dispuestos a asumir si rechazamos la Ho. Así, mientras más grande sea el nivel de significancia que utilizamos en una prueba es mayor el riesgo que estamos dispuestos a asumir cuando rechazamos la hipótesis nula, suponiendo que es verdadera.

- Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$.

Es importante que el estudiante comprenda de forma integral las nociones de inferencia para que pueda desarrollar e interpretar las inferencias; en este sentido, diversas investigaciones (e.g., Lane-Getaz, 2007; Lane-Getaz, 2013; Lyu, Peng y Hu, 2018) han trabajado (con propuestas de actividades, entrevistas, cuestionarios) entorno a la comprensión de un conjunto de nociones de forma holística, tal es el caso del nivel de confianza y de significancia, el valor-p y los errores tipo I y tipo II.

En la figura 5.10 podemos observar, mediante la representación gráfica, la relación entre el nivel de significancia y el de confianza, en este caso representado en un contraste bilateral.

-

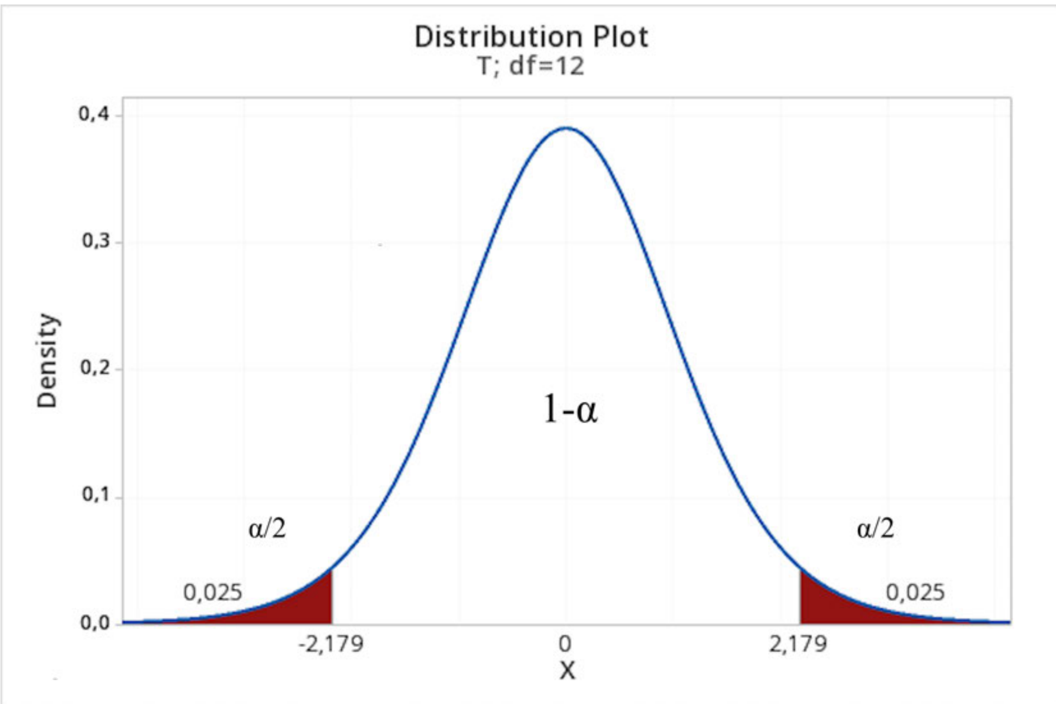


Figura 5.10. Representación gráfica del nivel de significancia y nivel de confianza

- Conoce, comprende y es capaz de aplicar la regla de decisión. Si el $valor - p < \alpha$ se rechaza H_0, (*propiedad/proposición* siete del $SP8_t$).

  El valor-p puede interpretarse como la probabilidad del error que cometeríamos si rechazamos la hipótesis nula con los datos que trabajamos. Así, cuando el valor-p es menor que el nivel de significancia, un riesgo que ya estábamos dispuestos a asumir, se rechaza la hipótesis nula; y cuando el valor-p es mayor que α significa que el riesgo es mayor al que estábamos dispuestos a asumir, por lo que entonces no se rechaza la hipótesis nula.

$$valor - p < \alpha \text{ entonces rechazamos } H_0$$

$$valor - p > \alpha \text{ entonces No rechazamos } H_0$$

El nivel de significancia o el valor-p no son un indicativo del tamaño de la diferencia.

Como se observó en la historia, y siguiendo con lo trabajado en los niveles, hemos abordado la prueba de hipótesis primero con el criterio *valor-p* para la toma de decisión y posteriormente con el estadístico teórico como valor crítico.

5.2.4.1.2 Valor crítico

- Es capaz de identificar el valor teórico del estadístico t-Student, de acuerdo con el nivel de significancia y los grados de libertad.

  Mientras que el estadístico de prueba o calculado depende de los valores observados en la muestra, el valor teórico del estadístico depende de la distribución teórica del estadístico cuando la hipótesis nula es verdadera.
- Puede representar gráficamente las regiones de rechazo y de no rechazo.

  En las figuras 5.11, 5.12 y 5.13 se puede observar las regiones de rechazo y no rechazo para los distintos contrastes con las pruebas t-Student. En la figura 5.11, se muestra los resultados de una prueba t-Student bilateral o de dos colas, las dos áreas sombreadas (en rojo) representan un nivel de significancia de $\alpha = 0.05$ (para este caso) con $\alpha/2$ en cada cola, el valor crítico es ± 2.179. Las Figuras 5.12 y 5.13 muestran resultados para contrastes unilaterales o de una cola; la Figura 5.12 para la cola derecha y la Figura 5.13 para la cola izquierda, en rojo también se encuentra la zona de rechazo con un $\alpha = 0.05$.

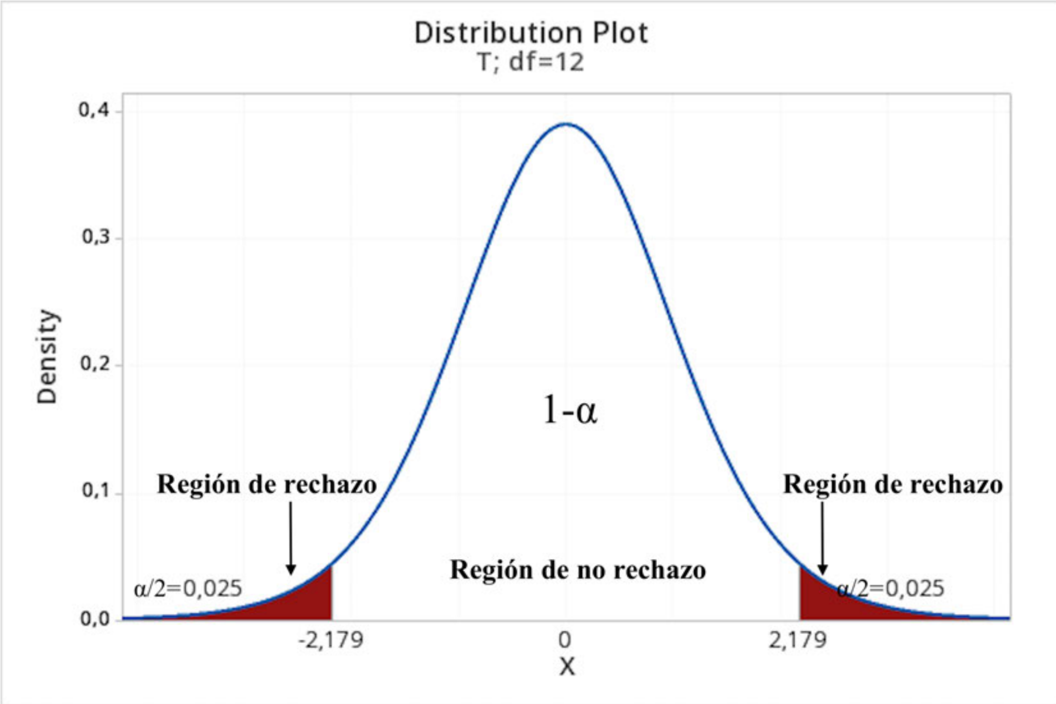


Figura 5.11. Regiones de rechazo y de no rechazo para prueba bilateral con $\alpha = 0.05$

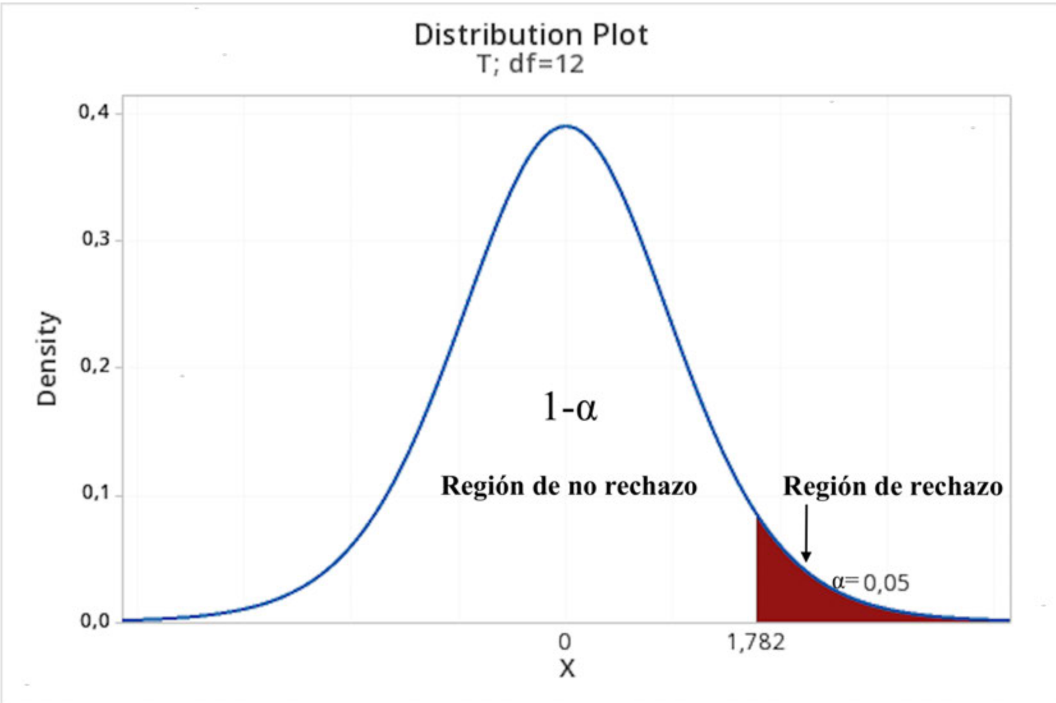


Figura 5.12. Regiones de rechazo y de no rechazo para prueba de cola derecha con $\alpha = 0.05$

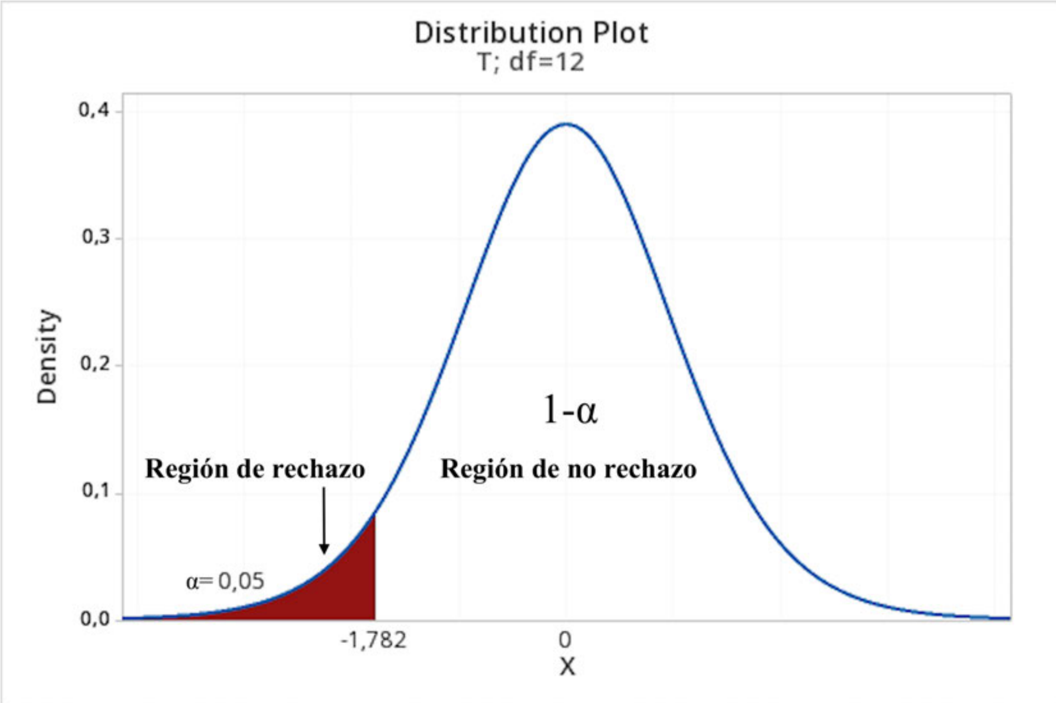


Figura 5.13. Regiones de rechazo y de no rechazo para prueba de cola izquierda con $\boldsymbol{\alpha = 0.05}$

- Conoce, comprende y es capaz de aplicar la regla de decisión (PP7 del SP8), si t se encuentra en región de rechazo (RR) se rechaza H_0 (e.g., RR de dos colas: $\mathrm{t} \geq t_{\alpha/2,gl}$ o $\mathrm{t} \leq -t_{\alpha/2,gl}$; RR cola izquierda: $\mathrm{t} \leq -t_{\alpha,gl}$; y RR cola derecha: $\mathrm{t} \geq t_{\alpha,gl}$).

  En la figura 5.14 se presenta un ejemplo donde se puede aplicar los dos criterios para la toma de decisión, podemos observar con una línea punteada el valor crítico definido por el estadístico teórico, el área sombreada en rojo inda el área del valor-p cuyo valor del estadístico calculado es de 1.33. Como 1.33 < 1.833 no se puede rechazar la hipótesis nula.

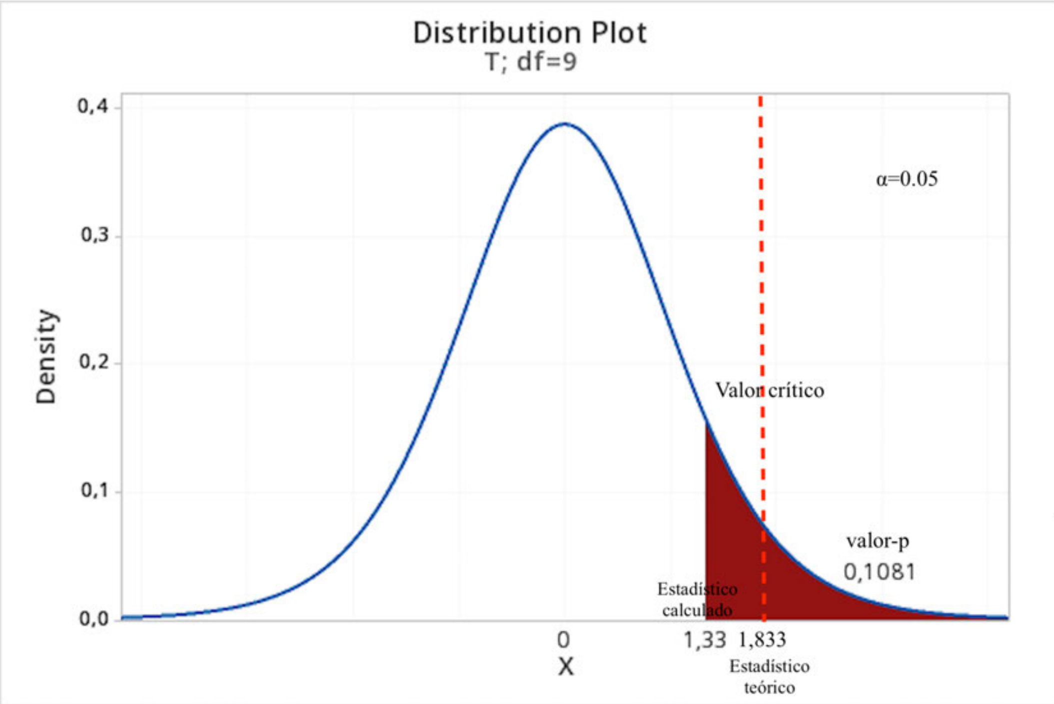


Figura 5.14. Representación gráfica sobre la aplicación de la regla de decisión con valor crítico

- Es capaz de brindar una respuesta al problema utilizando los resultados de la prueba y realizar procesos de argumentos con fundamentación estadística.

5.2.4.2 *Subnivel 4.2: Error tipo I y II, y Potencia de la prueba*

- Comprende cuándo se comete el Error tipo I y la probabilidad de cometerlo.

$$P\begin{bmatrix} Rechazar\ H_0 | H_0 \\ es\ verdadera \end{bmatrix} = \alpha$$

- Comprende cuándo se comete el Error tipo II y la probabilidad de cometerlo. También reconoce que existe un valor de β para cada posible valor del parámetro en la hipótesis alternativa, comprendiendo que se trata de una función.

$$P\begin{bmatrix} No\ rechazar\ H_0 | H_0 \\ es\ falsa \end{bmatrix} = \beta$$

- Es capaz de calcular la probabilidad de tomar la decisión correcta cuando H_0 es verdadera y cuando H_0 es falsa.

  En la Tabla 5.2 se muestran las probabilidades asociadas a la condición de la hipótesis nula (verdadera, falsa) y a la decisión (rechazar y no rechazar).

Tabla 5.2. *Decisiones en las pruebas de hipótesis*

	H_0 **es verdadera**	H_0 **es falsa**
Rechazar H_0	Error tipo I	Decisión correcta
	α	$1-\beta$
No rechazar H_0	Decisión correcta	Error tipo II
	$1-\alpha$	β

- Comprende las relaciones entre los dos tipos de error. Por ejemplo, puede auxiliarse de una representación gráfica y reconocer en ella que cuanto mayor es α, menor es β.
  Es importante que los estudiantes no confundan las probabilidades condicionales que intervienen en los errores tipo I y tipo II, el valor-p y el nivel de significancia, con probabilidades de un solo evento (Birnbaum, 1982; Falk, 1986; Shaughnessy y Dick, 1991; Sotos et al., 2007).
  La Figura 5.15 presenta gráficamente los dos tipos de error, al parecer (en este caso) α y β son pequeños lo cual es un buen indicativo para la prueba. La curva en rojo pertenece a la distribución del estadístico bajo la hipótesis nula mientras que la curva morada pertenece a la hipótesis alternativa.

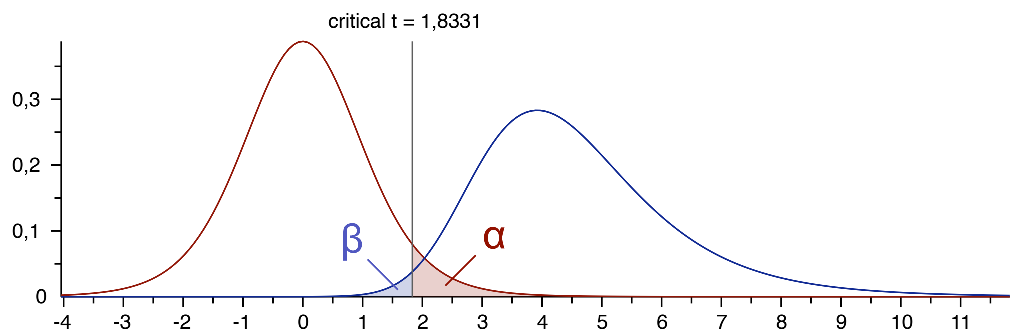


Figura 5.15. Representación gráfica de los errores tipo I y tipo II

- Comprende la potencia de la prueba y es capaz de calcularla. Por ejemplo, reconoce que la potencia es una función que para cada posible valor del parámetro le hace corresponder la probabilidad de que la hipótesis resulte rechazada. También que la potencia está relacionada con el tamaño de la muestra y el nivel de significancia, y que cuanto más lejana está H_a de H_0 la probabilidad de cometer el error tipo II es menor, y por lo tanto la potencia de la prueba es mayor. Podemos ver a la potencia como un índice de la validez de los resultados estadísticos (Cohen, 1992).

$$P\begin{bmatrix} Decidir\ H_1 | H_1 \\ es\ verdadera \end{bmatrix} = 1-\beta$$

De acuerdo con la teoría de la potencia estadística (Lipsey y Aiken, 1990) hay cuatro factores que determinan la potencia: la prueba estadística (en este caso t-Student), el nivel de significancia, tamaño de la muestra y tamaño del efecto.

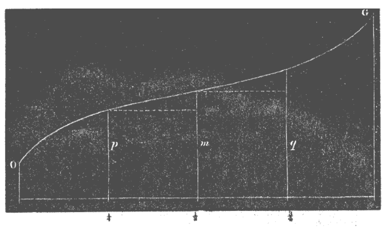
- Puede argumentar sobre la validez de la inferencia realizada.
  Es importante que el estudiante, así como utilizó los resultados de la prueba t-Student para dar respuesta al problema, pueda emplear los resultados de la potencia de la prueba para argumentar sobre la validez de la inferencia realizada.

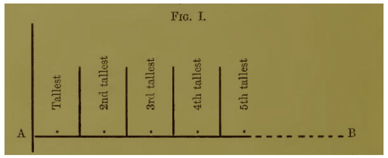
5.2.5 Sumario de los niveles de razonamiento inferencial para el estadístico t-Student

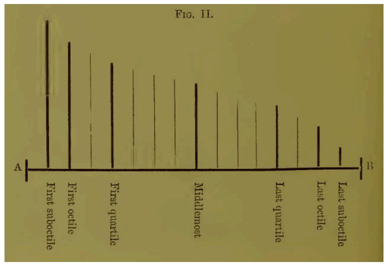
A partir de la propuesta de niveles de razonamiento inferencial para el estadístico t-Student que hemos descrito detalladamente a lo largo de la sección 5.2 realizamos la Tabla 5.3, la cual recoge de forma concisa la propuesta esta propuesta de niveles.

Tabla 5.3. *Propuesta de niveles de razonamiento inferencial para el estadístico t-Student*

RII → **RIF**

Nivel 1	**Nivel 2**	**Nivel 3**	**Nivel 4**
1.1 Visualización ($S1_t$) Elementos del $SP1_t$	2.1 Identificar la prueba paramétrica necesaria para analizar los datos	3.1 Restricciones de las pruebas t-Student	4.1 Criterio para la toma de decisión
			Valor-p
• Es capaz de identificar la variación interna de un conjunto de datos a través de observar e interpretar la forma y los cuartiles de la gráfica. 	• Reconoce el tipo de datos que está trabajando y que provienen de poblaciones normales • Comprende el problema a resolver • Comprende la lógica de las pruebas t-Student para 1 muestra ($S1_t$) y para dos muestras ($S3_t$). • Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla.	• Comprende que puede trabajar con tamaños muestrales pequeños si las muestras provienen de distribuciones normales. ($S1_t$, $S3_t$) • Identifica si las muestras son dependientes o independientes y comprende sus implicaciones. ($S3_t$) • Comprende las implicaciones cuando se tiene igualdad o desigualdad de varianzas y del tamaño de las muestras. ($S3_t$) • Es capaz de hacer explícita la hipótesis nula y alternativa en lenguaje natural.	• Puede plantear las hipótesis nula y alternativa con lenguaje simbólico. • Conoce y comprende los valores comunes del nivel de significancia. • Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$. • Conoce, comprende y es capaz de aplicar la regla de decisión: Si el $valor-p<\alpha$ se rechaza H_0 (PP7 del $SP8_t$)
		Muestras dependientes	Valor crítico

<table>
<tr>
<td>1.2 Trabajar con datos de una muestra
(S1$_t$) Elementos del SP1$_t$</td>
<td>2.2 Una aproximación a las pruebas con el estadístico t-Student</td>
<td>(S3$_t$) Elementos del SP9$_t$</td>
<td rowspan="2">• Es capaz de identificar y comprende el valor teórico del estadístico t-Student, de acuerdo con α y los gl.
• Puede representar gráficamente las regiones de rechazo y de no rechazo.
• Conoce, comprende y es capaz de aplicar la regla de decisión:
Sí t $\geq t_{\alpha,gl}$ se rechaza H_0 (PP7 del SP8$_t$)
• Es capaz de brindar una respuesta al problema utilizar los resultados de la prueba y realizando argumentos con fundamentación estadística.</td>
</tr>
<tr>
<td rowspan="4">• Puede utilizar el método de intercomparación, donde identifica cuartiles, octiles, deciles y percentiles; realiza la gráfica correspondiente y compara e interpreta la dispersión de los intervalos (formados por los cuartiles)
(CD1, CD2, CD3, CD4, CD5, PP3, PP4, PP5, PP6 y PP7 del SP1$_t$)



</td>
<td>(S1$_t$ y S3$_t$)
• Es capaz de identificar la hipótesis nula que se encuentra implícita en el problema.
• Conoce la distribución t-Student. (PP2 del SP3$_t$)
• Comprende la relación entre la distribución t-Student y la normal. (PP2 del SP1, PP2 del SP2$_t$ y PP2 del SP3$_t$)</td>
<td rowspan="2">Puede valorar si la media de las diferencias es igual al valor μ_0.
$$t = \frac{\bar{x}_D - \mu_0}{s_D/\sqrt{n}}$$
• Calcula y comprende el estadístico t-Student. (PP3, PP4 y PP6 del SP9$_t$)
• Calcula y comprende los grados de libertad $gl = n - 1$. (PP5 del SP3$_t$)
• Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de las diferencias se encuentre fuera del rango $\pm t$. (CD7 del SP9$_t$)</td>
</tr>
<tr>
<td>(S1$_t$)
Elementos del SP2$_t$ y SP3$_t$</td>
<td>4.2 Error tipo I y II, y Potencia de la prueba</td>
</tr>
<tr>
<td rowspan="2">Puede valorar si la media de una población es igual al valor μ_0.
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
• Calcula y comprende qué indica el estadístico t-Student. (CD2, CD3, CD4, CD5, CD6 del SP2$_t$ y PP6 del SP3$_t$)</td>
<td>Muestras independientes con $n_1 \neq n_2$
(S3$_t$) Elementos del SP7$_t$</td>
<td rowspan="2">• Comprende cuándo se comete el Error tipo I y la probabilidad de cometerlo.
$$P\left[\begin{matrix} Rechazar\ H_0 | H_0 \\ es\ verdadera \end{matrix}\right] = \alpha$$
• Comprende cuándo se comete el Error tipo II y la probabilidad de cometerlo.
$$P\left[\begin{matrix} No\ rechazar\ H_0 | H_0 \\ es\ falsa \end{matrix}\right] = \beta$$</td>
</tr>
<tr>
<td>Cuando los tamaños de las muestras son desiguales puede valorar si las medias de dos poblaciones son iguales.
• Comprende y es capaz de obtener la desviación estándar agrupada. (CD6 del SP8$_t$, CD10, PP4 y PP5 del SP7$_t$)</td>
</tr>
</table>

1.3 Trabajar con datos de dos muestras ($S2_t$) Elementos del $SP6_t$	• Calcula y comprende los grados de libertad $gl = n - 1$. (PP5 del $SP3_t$) • Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la media de la población se encuentre fuera del rango $\pm t$. (CD7 y PP2 del $SP3_t$) **Generalización** $SP2_t$ y SP3-N2.2 Prueba y estadístico para una muestra $SP9_t$ -N3.1 Prueba y estadístico para dos muestras pero que sean dependientes o emparejadas 1	$s = \sqrt{\dfrac{(n_1 - 1)(s_{\bar{x}_1}^2) + (n_2 - 1)(s_{\bar{x}_2}^2)}{n_1 + n_2 - 2}}$ • Calcula y comprende el estadístico t-Student. (PP7 del $SP7_t$) $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ • Comprende y calcula los grados de libertad $gl = n_1 + n_2 - 2$. (PP6 del $SP7_t$) • Comprende y es capaz de establecer relaciones del estadístico t-Student, los grados de libertad y la desviación estándar agrupada del nivel 2.3.	• Es capaz de calcular la probabilidad de tomar la decisión correcta cuando: * H_0 es verdadera * H_0 es falsa • Comprende las relaciones entre los dos tipos de error • Comprende la potencia de la prueba $P\begin{bmatrix} Decidir\ H_1 \| H_1 \\ es\ verdadera \end{bmatrix} = 1 - \beta$ y es capaz de calcularla • Puede argumentar sobre la validez de la inferencia realizada.
• Es capaz de analizar bajo el método de intercomparación cada una de las muestras y compara los valores representantes (e.g., cuartiles) de cada una de las muestras, puede realizar gráficos (e.g., diagrama de caja y bigote) y conjetura las semejanzas y/o diferencias entre las muestras. Nota: a manera de conexión con el nivel 1.2 • Puede analizar la variación interna de los datos por medio de elaborar gráficas y de la fluctuación $\left(2\frac{\sum e^2}{n}\right)$, donde n es la suma de las diferencias $(y_i - x_i)$ y $e = (y_i - x_i) - (\bar{y} - \bar{x})$ (CD2, CD3, CD4, PP4, PP6 del $SP6_t$)	($S3_t$) Elementos del $SP3_t$, $SP7_t$ y $SP8_t$ Cuando los tamaños de las muestras y las varianzas son iguales, puede valorar si las medias de dos poblaciones (distribuidas normalmente) son iguales.	Muestras independientes con $n_1 \neq n_2$ y $\sigma_1 \neq \sigma_2$ o $n_1 = n_2$ y $\sigma_1 \neq \sigma_2$ ($S3_t$) Elementos del $SP12_t$ Cuando se tienen dos muestras con diferentes tamaños muestrales y diferentes varianzas puede valorar si las medias de dos poblaciones son iguales. • Es capaz de estimar la varianza de las dos poblaciones por medio de las muestras y comprende por qué no se utiliza la varianza agrupada (como en las otras pruebas t para dos muestras). (CD5, PP6 y PP7 del	

	• Comprende y es capaz de obtener la varianza y desviación estándar agrupada. (CD6 y PP4 del SP8$_t$ y CD10 del SP7$_t$) $s^2 = \frac{1}{(n_1-1)(n_2-1)}\left(\sum(x_1-\bar{x}_1)+\sum(x_2-\bar{x}_2)\right) = \frac{1}{2}\left(s_{x_1}^2+s_{x_2}^2\right)$ • Comprende el error estándar. (CD11 del SP7$_t$) • Comprende qué indica el estadístico t-Student para dos muestras. (CD2, CD3, CD4, CD6 del SP2$_t$ y PP6 del SP8$_t$) $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{s^2\left(2/n\right)}} = \frac{\bar{x}_1-\bar{x}_2}{s\sqrt{2/n}}$ • Comprende y calcula los grados de libertad $gl = 2n - 2$. (PP5 del SP8$_t$) • Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida	SP12$_t$) $\sigma^2 = \frac{1}{n_1}s_1^2 + \frac{1}{n_2}s_2^2$, $s_i^2 = \frac{1}{n_i-1}\sum(x_i-\bar{x})^2$ • Calcula y comprende el estadístico t-Student para dos muestras con dichas características. (PP9 del SP12$_t$) $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{1}{n_1}s_1^2+\frac{1}{n_2}s_2^2}}$ • Comprende y calcula los grados de libertad. (PP8 del SP12) $gl = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)}+\frac{s_2^4}{n_2^2(n_2-1)}}$ • Comprende que este método también lo puede aplicar cuando se tienen tamaños muestrales iguales.	
		Muestras independientes con coeficientes de regresión (S3$_t$) Elementos del SP10$_t$ y SP11$_t$	
		Puede comparar los coeficientes de regresión cuando las series de la variable independiente no son idénticas y valorar la diferencia de los coeficientes de regresión. • Puede obtener el coeficiente b y comprende su significado. (CD13, CD14, CD15, PP4 y PP9 del SP10$_t$)	

	de ocurrencia de que la diferencia de las medias se encuentre fuera del rango $\pm t$. (CD7 del SP7$_t$ y PP2 del SP3$_t$) **Generalización** SP8$_t$ -N2.2 Estadístico t, grados de libertad y desviación estándar agrupada cuando $n_1 = n_2$ SP7$_t$ -N3.1 Estadístico t, grados de libertad y desviación estándar agrupada cuando $n_1 \neq n_2$	$b = \frac{S\{y(x-\bar{x})\}}{S(x-\bar{x})^2}$ • Comprende qué indica el estadístico t-Student. (PP11 del SP11) $t = \frac{b' - b}{\sqrt{(s^2/S_A(x-\bar{x})^2) + (s^2/S_B\,(x-\bar{x})^2)}}$ • Comprende y calcula los grados de libertad $gl = n_A + n_B - 2$. (PP8 del SP11$_t$) • Utiliza la tabla de probabilidad de la distribución t-Student para determinar la probabilidad y puede interpretarla como una medida de ocurrencia de que la diferencia de los coeficientes de regresión se encuentre fuera del rango $\pm t$. (CD7 del SP7$_t$ y PP2 del SP3$_t$) • Reconoce que esta prueba sigue la misma lógica de la prueba de diferencia de medias de dos muestras.	
		3.2 Conexiones y argumentos	
		• Comprende la significancia. (CD8 del SP2$_t$) • Puede encontrar el valor del estadístico teórico, en la tabla de probabilidad de la distribución t-Student, con respecto a cierta P y gl. Y lo compara contra el valor t calculado. • Es capaz de rechazar o no rechazar la hipótesis nula (y comprende lo que	

		significa) bajo un contraste con un límite preestablecido como desviación significativa. (PP7 del $SP3_t$ y PP7 del $SP8_t$) • Logra argumentar, con base en la significancia por qué rechaza o no rechaza la hipótesis nula. • Es capaz de conectar los resultados de la prueba con el contexto del problema.	

5.3 Reflexiones finales

En este capítulo hemos presentamos una propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. La propuesta de niveles se formuló a partir tanto de los aportes de la literatura científica sobre el razonamiento inferencial, como de los atributos matemáticos de los diversos significados parciales resultantes del estudio histórico-epistemológico sobre los estadísticos Chi-cuadrada y t-Student (Capítulos 3 y 4). En las secciones 5.1 y 5.2, describimos los cuatro niveles de razonamiento inferencial para el estadístico Chi-cuadrada y t-Student, respectivamente.

Como mencionamos anteriormente, para la propuesta de los niveles de razonamiento inferencial retomamos aportes de la literatura científica, de los cuales destacamos aspectos como la visualización, interpretar las conclusiones en el contexto del problema, la lógica de las pruebas de hipótesis, incertidumbre, planteamiento de las hipótesis nula y alternativa, distribución, estadístico-parámetro, valor-p, nivel de significancia, criterio de decisión, los errores tipo I y tipo II, concepción de asociación e independencia y la media. Lo anterior, fue consistente con las sugerencias de la literatura científica (ver capítulo 1).

Asimismo, destacamos atributos matemáticos que retomamos de los estudios históricos-epistemológicos para construir los niveles de razonamiento inferencial, primero señalaremos los correspondientes al estadístico Chi-cuadrada: (*Nivel 1*) el método de intercomparación y el método gráfico para para conjeturar si un grupo de datos sigue una distribución normal, coeficiente de asociación Q para conjeturar si existe asociación entre dos variables; (*Nivel 2*) estadístico Chi-cuadrada, grados de libertad, distribución Chi-cuadrada, la probabilidad para interpretar los valores obtenidos con el estadístico Chi-cuadrada, identificación de la hipótesis y planteamiento en lenguaje natural; (*Nivel 3*) factor de corrección de continuidad, significancia como un límite, estadístico teórico con respecto a cierta P y gl y regla de decisión intuitiva; (*Nivel 4*) planteamiento de las hipótesis en lenguaje simbólico y natural, regla de decisión, valor-p, valor crítico.

Por otra parte, destacamos algunos atributos correspondientes al estadístico t-Student: (*Nivel 1*) el método de intercomparación y de fluctuación para identificar la variación interna de uno y dos conjuntos de datos, respectivamente; (*Nivel 2*) planteamiento de la hipótesis en lenguaje natural, estadístico t-Student para una y dos muestras independientes, grados de libertad, muestras independientes, media y desviación estándar, la probabilidad para interpretar los valores obtenidos con el estadístico t-Student; (*Nivel 3*) planteamiento de la hipótesis nula y alternativa en lenguaje natural, muestras dependientes e independientes, desviación estándar agrupada, el error estándar, grados de libertad, estadístico t-Student para muestras independientes con $n_1 \neq n_2$ y $\sigma_1^2 \neq \sigma_2^2$ o $n_1 = n_2$ y $\sigma_1^2 \neq \sigma_2^2$, significancia como un límite, estadístico teórico con respecto a cierta P y gl y regla de decisión intuitiva; (*Nivel 4*) planteamiento de las hipótesis en lenguaje simbólico, regla de decisión, valor-p, valor crítico.

Además, en las Tablas 5.1 y 5.3 presentamos de forma sucinta la propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, para el estadístico Chi-cuadrada y t-Student, respectivamente.

CAPÍTULO 6

Niveles de Razonamiento Inferencial sobre los estadísticos Chi-cuadrada y t-Student de Profesores de Enseñanza Media

"Nuestro problema, desde el punto de vista de la psicología y desde el punto de vista de la epistemología genética, es explicar cómo la transición se hace de un menor nivel de conocimiento a un nivel que parece ser mayor". (Jean Piaget)

INTRODUCCIÓN

En este capítulo, en la primera sección, mostramos con más detalle aspectos del instrumento de indagación, los cuales fueron presentados en la sección 2.3.3 del Capítulo 2; esto debido a que ya se han caracterizado los significados para los estadísticos Chi-cuadrada y t-Student en los Capítulos 3 y 4, y se han propuesto los niveles de razonamiento inferencial para dichos estadísticos en el Capítulo 5.

En la segunda y tercera sección, presentamos la caracterización del razonamiento inferencial que evidencian profesores, que participaron en esta investigación, cuando realizan prácticas para resolver problemas sobre los estadísticos Chi-cuadrada y t-Student. Esta caracterización, tiene la finalidad de estudiar los elementos primarios (elementos lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos) que utilizan en sus prácticas los profesores y si dichos elementos permiten asociarlos a los distintos niveles de razonamiento inferencial propuestos en el capítulo anterior. Con esta caracterización se pretende explorar empíricamente si los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, propuestos a nivel teórico, son predictores del nivel de razonamiento inferencial asociado a las prácticas que desarrollan los profesores al resolver problemas sobre estos estadísticos.

Finalmente, en la cuarta sección presentamos una propuesta general de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial. La propuesta se formuló a partir de los puntos de encuentro entre los niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student, y los resultados de la caracterización de las prácticas desarrolladas por los profesores en ejercicio y en formación. En la última sección, realizamos algunas conclusiones a partir de los hallazgos del estudio experimental.

6.1 Instrumentos de indagación

Previamente, en el Capítulo 2, habíamos declarado que para realizar el diseño de las actividades sobre los estadísticos Chi-cuadrada y t-Student se consideraron, principalmente, dos criterios: 1) representatividad de los niveles de razonamiento inferencial, y 2) representatividad de los significados de los estadísticos mencionados. A continuación, presentamos con mayor detalle estos criterios para el diseño.

Representatividad de los niveles de razonamiento inferencial

Un primer criterio que se tuvo en consideración para el diseño, fue que la práctica matemática que realicen los profesores al resolver las actividades pudiera ser desarrollada en cualquiera de los niveles de razonamiento de inferencial, es decir, que las actividades se pueden resolver con prácticas que admiten rasgos de alguno de los cuatro niveles de razonamiento inferencial para el estadístico Chi-cuadrada o t-Student (la propuesta de niveles se desarrolló en el Capítulo 5). Para lograrlo, se diseñaron actividades que podrían resolverse tanto de forma intuitiva como formal. Además, la pregunta que se realiza en el planteamiento de la actividad es de forma general, es decir, no hemos incluido sub-preguntas que vayan guiando o parcializando la solución.

Por ejemplo, un primer tipo de prácticas, asociadas al nivel 1 de razonamiento inferencial sobre el estadístico Chi-cuadrada, podían ser las de carácter intuitivo, tales como realizar conjeturas sobre la problemática planteada utilizando gráficos (e.g., diagrama de barras) y/o medidas estadísticas como los cuartiles, la media y desviación estándar, así como reflexionar sobre la dispersión, simetría y forma de la gráfica. En prácticas asociadas al segundo nivel se podían indicar características de los datos a analizar, como si se trata de una muestra o población y si se clasifican de acuerdo con una o dos variables; también se podrían identificar la hipótesis implícita en el problema y hacerla explícita, algunas propiedades como el estadístico Chi-cuadrada y los grados de libertad, mientras que para los argumentos podrían utilizar la probabilidad en el contexto.

En las prácticas asociadas al tercer nivel podían expresarse ambas hipótesis en lenguaje natural, identificar si es necesario aplicar un factor de corrección de continuidad al estadístico Chi-cuadrada; también establecer un límite de desviación significativo y utilizarlo para argumentar su inferencia. Las prácticas asociadas al cuarto nivel, podían indicar y justificar un nivel de significancia, plantear las hipótesis en lenguaje simbólico, realizar la prueba con el estadístico Chi-cuadrada, tomar la decisión basándose en la prueba de hipótesis y argumentarla estadísticamente, así como analizar críticamente la eficiencia de su inferencia estadística.

Representatividad de los significados

El segundo criterio refiere a una visión integral de los estadísticos Chi-cuadrada y t-Student, esto debido a las particularidades de los campos de problemas o distintos tipos de pruebas en los que intervienen los estadísticos. Para ello se ha diseñado al menos una actividad para cada gran significado, por ejemplo, para el estadístico Chi-cuadrada tenemos actividades que corresponde a bondad de ajuste (S1), independencia (S2) y homogeneidad (S3), mientras que para el estadístico t-Student tenemos actividades que corresponden a la media de una muestra (S1) y a la media de dos muestras (S3). Los elementos del significado de distribución, S4 para el estadístico Chi-cuadrada y S2 para el estadístico t-Student corresponden a la distribución, se incorporan dentro de los otros significados. Los significados de los estadísticos Chi-cuadrada y t-Student los desarrollamos con profundidad en los Capítulos 3 y 4 de esta tesis.

6.1.1 Las actividades

A continuación, en la Tabla 6.1 presentamos una descripción de las actividades que conforman el instrumento de indagación, posteriormente mostramos algunas actividades o situaciones-problemas que forman parte de nuestro instrumento de indagación, así como las posibles prácticas matemáticas, que se desarrollarían en cada nivel de razonamiento inferencial, para dar solución a dichas actividades. El instrumento completo puede verse en anexos.

Tabla 6.1. *Descripción de las actividades del instrumento*

Actividad	Estadístico	Prueba	Contexto	Grupo de aplicación
1	Chi-cuadrada	Bondad de ajuste	Deportes: lanzamiento de tiro con arco	G1, G2, G4
2	Chi-cuadrada	Independencia	Vacuna de la viruela	G1, G2, G4
3	Chi-cuadrada	Homogeneidad	Cáncer de pulmón, género y raza	G1, G2, G4
4	t-Student	Una muestra	Fármaco para dormir	G2, G3
5	t-Student	Dos muestras dependientes	Fármacos para dormir	G2, G3
6	t-Student	Dos muestras independientes de diferente tamaño e igualdad de varianzas	Socavones y la resistencia a la tensión de un forro para tuberías	G2, G3
7	t-Student	Dos muestras independientes de mismo tamaño y con igualdad de varianzas	Anticuerpo del Covid-19	G2, G3

Para ejemplificar cómo las actividades (Tabla 6.1) admiten respuestas asociadas a cada nivel de razonamiento inferencial, presentamos las respuestas plausibles de una actividad sobre el estadístico Chi-cuadrada (Figura 6.1) y después de una actividad del estadístico t-Student (Figura 6.4).

6.1.1.1 Actividad sobre el estadístico Chi-cuadrada

En la Figura 6.1 se encuentra la segunda actividad sobre el estadístico Chi-cuadrada y es seguida por respuestas plausibles sobre la actividad, respuestas que se encuentran asociadas a cada nivel de razonamiento inferencial para el estadístico Chi-cuadrada.

Problema 2.
A continuación, se presentan los datos de una muestra que fueron recabados durante una epidemia de viruela en un pequeño pueblo de Inglaterra. Los datos de la tabla 2 muestran la presencia o ausencia de la cicatriz de la vacuna de viruela y si las personas que recibieron la vacuna se recuperaron o fallecieron. ¿Existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela? Explica con detalle tu respuesta

Tabla 2. *Presencia de cicatriz por vacuna y recuperados de viruela*

Cicatriz	Recuperados	Muertos	Total
Presente	35	1	36
Ausente	10	4	14
Total	45	5	50

Figura 6.1. Actividad 2

Ejemplo de respuesta asociada al Nivel 1:

Inicialmente, si vemos los datos de la tabla de contingencia de 2×2 podemos observar una amplia concentración en la intersección de los que tienen cicatriz y los que se recuperaron, lo que nos puede llevar a pensar que es posible que exista una asociación entre la recuperación y la presencia de la cicatriz. Si analizamos los datos en el mismo sentido que el coeficiente de correlación, pero ahora con variables discretas mediante el coeficiente de asociación, para probar si existe asociación entre las variables, se tiene que:

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)} = \frac{(35)(4) - (1)(10)}{(35)(4) + (1)(10)} = \frac{140 - 10}{140 + 10} = 0.8666$$

De acuerdo con el valor del coeficiente de asociación que es 0.8666, podemos decir que sí existe una relación entre las variables, es decir, existe una relación entre la presencia de cicatriz por vacuna y los recuperados de viruela.

Ejemplo de respuesta asociada al Nivel 2:

Se está trabajando con una muestra de 50 individuos que recibieron una vacuna para la viruela y se clasifican en dos variables (discretas). Una de las variables refiere a un efecto secundario de la vacuna que es una cicatriz, teniendo como atributos la “presencia o ausencia” de dicha cicatriz que deja la vacuna, mientras que los atributos de la segunda variable son “recuperado o muerto”. Por el tipo de datos presentados y dado que se desea conocer si existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela, la cual parece ser la hipótesis que se tiene, se puede aplicar una prueba de independencia con el estadístico Chi-cuadrada para probar si existe o no independencia entre las variables.

Si aplicamos la prueba de independencia podemos calcular las frecuencias esperadas bajo independencia para posteriormente aplicar el estadístico Chi-cuadrada, pero como se trata de una tabla de contingencia de 2×2 podemos aplicar la siguiente formula del estadístico:

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+c)(b+d)(a+b)(c+d)} = \frac{\left((35\times 4)-(1\times 10)\right)^2(50)}{(35+10)(1+4)(35+1)(10+4)}$$
$$= \frac{(140-10)^2(50)}{(45)(5)(36)(14)} = 7.45149$$

Ahora, procedemos a calcular los grados de libertad bajo la formula $k = (c-1)(r-1) = (2-1)(2-1) = 1$.

A partir de estos valores y mediante la distribución Chi-cuadrada, podemos obtener la probabilidad de que los datos de la tabla sean compatibles con las bases de independencia probabilística. Se tiene una probabilidad de ocurrencia de valores como los observados de 0.006338, es decir, sólo podríamos ver valores como estos 6.33 veces en 1000 casos si las variables fueran independientes.

A partir de los procedimientos realizados con la prueba de independencia, podríamos decir que no existe independencia entre la presencia de cicatriz por vacuna y los recuperados de viruela.

También se puede considerar la opción de que el profesor en ejercicio o en formación utilice un software estadístico para analizar los datos, es importante que sea capaz de identificar (en la pantalla de salida) e interpretar lo anteriormente descrito, incluso podría establecer las diferencias entre las frecuencias esperadas y las observadas como aspectos de interés. Es importante señalar que, en este caso, el profesor haría caso omiso de la advertencia generada por el Software por frecuencias observadas inferiores a 5.

Ejemplo de respuesta asociada al Nivel 3:

Para ejemplificar una respuesta de este nivel, consideremos lo esbozado en el ejemplo de respuesta asociado al nivel dos y lo siguiente:

Las hipótesis nula y alternativa son,

Hipótesis nula: Las variables son independientes (No existe asociación entre la cicatriz y la recuperación de la viruela).

Hipótesis alternativa: No existe independencia entre las variables (Existe asociación entre la cicatriz y la recuperación de la viruela).

Dos casillas de la tabla de contingencia de 2×2 presentan frecuencias observadas inferiores a 5, por lo cual debemos tomar precauciones. Esto se debe a que al trabajar la prueba de independencia con números pequeños podemos obtener una discrepancia porque la prueba se realiza con una distribución continua, mientras que la distribución que se intenta aproximar es discreta. Para disminuir esta discrepancia podemos utilizar el factor de corrección de continuidad de Yates. Entonces calculamos el estadístico bajo la siguiente fórmula:

$$\chi_c^2 = \frac{\left(\left(a-\frac{1}{2}\right)\left(d-\frac{1}{2}\right)-\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)\right)^2 N}{(a+c)(b+d)(a+b)(c+d)}$$

$$= \frac{\left(\left(35-\frac{1}{2}\right)\left(4-\frac{1}{2}\right)-\left(1+\frac{1}{2}\right)\left(10+\frac{1}{2}\right)\right)^2 50}{(35+10)(1+4)(35+1)(10+4)} = \frac{(120.75-15.75)^2(50)}{(45)(5)(36)(14)}$$

$$= 4.86111$$

Con 4.86111 como valor del estadístico y con un grado de libertad tenemos una probabilidad de ocurrencia de tener valores, bajo independencia, como los observados de 0.02746.

Si consideramos una probabilidad de 0.05 como un límite de desviación significativa, al ser menor la probabilidad obtenida tendríamos que decir que las variables no son independientes. Bajo este mismo límite el estadístico teórico es de 3.84145 y si comparamos el estadístico calculado es más grande que el estadístico teórico. Esto quiere decir que sobrepasa nuestro límite por lo que las desviaciones de las expectativas son claramente significativas.

Otra posibilidad que se encuentra dentro de este nivel, aunque en menor grado, sería que el profesor utilice un software para realizar la prueba y que, al ver la advertencia de la pantalla de salida –la cual en este caso indicaría que "dos celdas tienen un conteo inferior a 5"– como se resalta en la Figura 6.2, el profesor desarrolle una reflexión sobre cómo las frecuencias inferiores a cinco podrían afectar el valor del estadístico y por lo tanto de la probabilidad, y por qué. Sin embargo, el profesor no realizaría la prueba con el factor de corrección de continuidad.

Rows: Cicatriz Columns: Recuperación

	Recuperados	Muertos	All
Presente	35	1	36
	32,400	3,600	
Ausente	10	4	14
	12,600	1,400	
All	45	5	50

Cell Contents
Count
Expected

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	7,451	1	0,006
Likelihood Ratio	6,618	1	0,010

2 cell(s) with expected counts less than 5.

Figura 6.2. Pantalla de salida de la prueba de asociación en Minitab

Ejemplo de respuesta asociada al Nivel 4:

Además de lo expresado en la respuesta del nivel previo, una respuesta en este nivel podría considerar los siguientes aspectos:

El problema que plantea no indica el nivel de significancia, así que se utiliza el más común, un $\alpha = 0.05$.

Dado que tenemos un valor del estadístico con corrección $\chi_c^2 = 4.86111$ y con un valor-p de 0.02746, podemos ver que nuestro valor-p es menor que el alfa, por lo que se rechaza Ho. Otro criterio para tomar la decisión es contrastar el valor del estadístico que calculamos con el teórico y como $\chi_{.05,1}^2 = 3.84145$, tenemos entonces que $4.86111 > 3.84145$, por lo tanto, la diferencia es significativa, lo que nos lleva a rechazar Ho. En la Figura 6.3 podemos observar que en la gráfica de la izquierda se muestra las regiones críticas a partir del valor del estadístico teórico (según el nivel de significancia y los grados de libertad), mientras que en la gráfica de la derecha podemos ver el estadístico calculado y su probabilidad asociada, y cómo se encuentra dentro de la zona de rechazo.

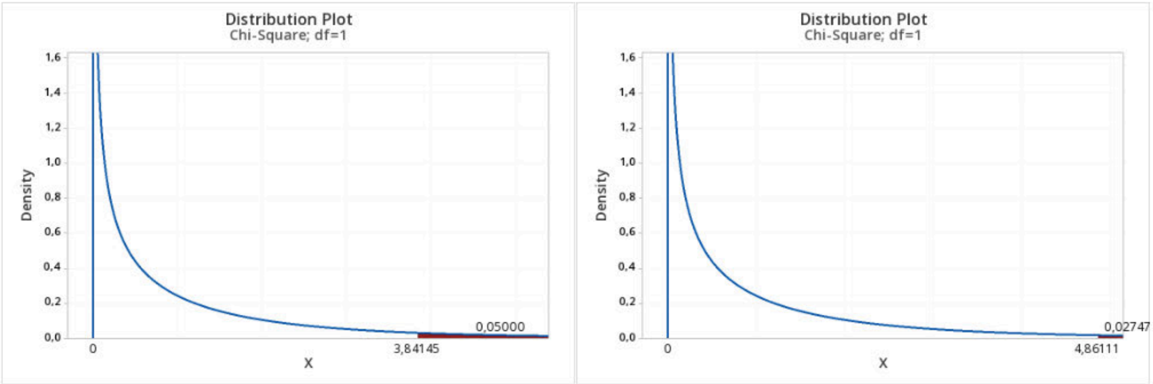


Figura 6.3. Gráficos de la distribución de Chi-cuadrada para valores del estadístico de 3.84145 y 4.86111

A partir de lo anterior, con un nivel de confianza del 95% podemos rechazar Ho y aceptar como probablemente cierta la Ha, es decir, que las variables no son independientes y por lo tanto existe asociación entre la cicatriz y la recuperación de la viruela.

La probabilidad que tenemos de cometer el error tipo I, ya que hemos rechazado Ho, es de 0.05, es decir de rechazar Ho cuando es realmente verdadera. En cuanto a la potencia de la prueba usualmente se pide un mínimo del 80%, y en este caso auxiliándonos de software para realizar el cálculo obtuvimos una probabilidad de 0.86275, esto indica que hay una prevalencia constante en la población. También podemos decir que la probabilidad de cometer el error tipo II en esta prueba es 0.13725.

6.1.1.2 Actividad sobre el estadístico t-Student

En la Figura 6.4 se muestra la primera actividad sobre el estadístico t-Student y posteriormente presentamos respuestas plausibles sobre dicha actividad, estas respuestas se encuentran asociadas a cada nivel de razonamiento inferencial para el estadístico t-Student.

Problema 4.
A continuación, se presentan los datos de una muestra que fueron recabados durante un experimento de un fármaco para dormir. Los datos de la Tabla 4 son sobre el aumento de horas de sueño cuando los pacientes han ingerido el fármaco. ¿Existe realmente un aumento en las horas de sueño al utilizar el fármaco? Explica con detalle tu respuesta

Tabla 4 – Horas adicionales de sueño ganadas por el uso del fármaco

Paciente	Horas adicionales
1	+ 0.7
2	− 1.6
3	− 0.2
4	− 1.2
5	− 0.1
6	+ 3.4
7	+ 3.7
8	+ 0.8
9	0.0
10	+ 2.0

Figura 6.4. Actividad 4

Ejemplo de respuesta asociada al Nivel 1

A partir de los datos de la tabla, podemos calcular los cuartiles y tenemos que $Q1 = -0.175$, $Q2 = 0.35$, $Q3 = 1.7$ y $Q4 = 3.7$, y elaborar los gráficos con los cuartiles (Figura 6.5).

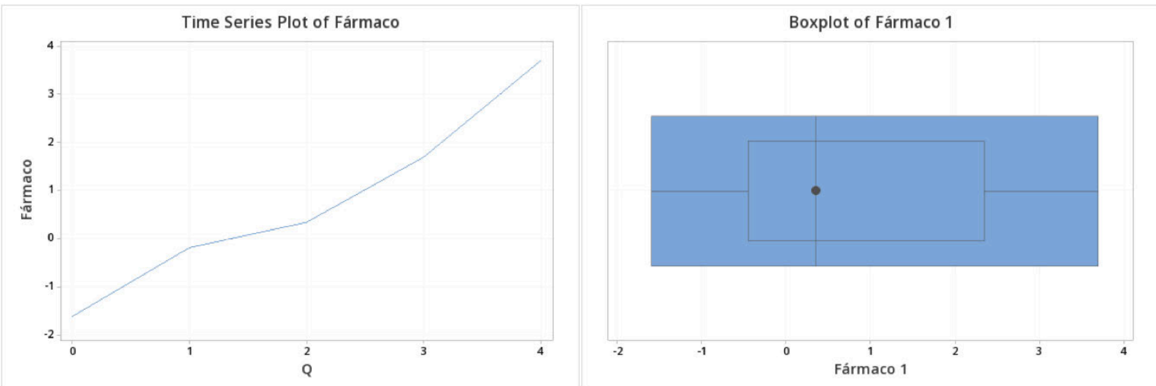


Figura 6.5. Gráficos que permiten visualizar la dispersión

Ahora podemos observar los cuartiles en las gráficas y vemos claramente que hay dispersión en el tamaño de los cuartiles, también podemos calcular cuánto es la amplitud de cada cuartil y tenemos que es 1.425, 0.525, 1.35 y 2. Así, podemos indicar que la serie de datos no es simétrica porque $Q2 - Q1 = 0.525$ y $Q3 - Q2 = 1.35$.

Consideramos que los análisis realizados hasta el momento no nos permiten asegurar si con el fármaco efectivamente se pueden incrementar las horas que se duerme, aunque algunos de los pacientes experimentaron una ganancia en las horas de sueño, otros pacientes incluso disminuyeron, la mediana (Q2) es cercana a cero y en este caso el cero indica que no hay un aumento en las horas de sueño.

Ejemplo de respuesta asociada al Nivel 2

Al tratarse de una muestra de tamaño pequeño, que proviene de una población normal, desconocemos el valor de la varianza poblacional y queremos saber si existe un incremento en las horas de sueño (la cual parece ser la hipótesis que se tiene), podemos utilizar una prueba t-Student para una muestra, la cual nos ayudará a determinar si la media de la población es igual a un valor μ_0, en este caso $\mu_0 = 0$.

Si aplicamos la prueba t-Student para una muestra, primero calculamos el valor del estadístico bajo la expresión

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.75 - 0}{1.79/\sqrt{10}} = \frac{0.75}{0.566} = 1.325$$

Ahora, procedemos a calcular los grados de libertad bajo la propiedad $gl = n - 1 = 9$. A partir del valor del estadístico t-Student y de los grados de libertad, buscamos la probabilidad en una tabla de distribución de probabilidad t-Student y resulta que para un valor del estadístico de 1.3 con 9 grados de libertad, la probabilidad es de 0.113; es decir, que podríamos ver 887 veces en 1000 experimentos, datos como estos cuando la media de la población es igual a 0. Esto quiere decir que no tenemos evidencia suficiente para concluir que con el fármaco se incrementan las horas de sueño.

Ejemplo de respuesta asociada al Nivel 3

Aunado a las prácticas del nivel previo también una respuesta en este nivel podría considerar los siguientes elementos:

> Hipótesis nula: No existe un incremento significativo en las horas de sueño con el uso del fármaco.

Hipótesis alternativa: Existe un incremento significativo en las horas de sueño con el uso del fármaco.

Ahora procedemos a calcular el estadístico t-Student, los grados de libertad y la probabilidad y obtenemos:

$$t = 1.325, gl = 9, P = 0.113$$

Si consideramos $P = 0.05$ como límite de desviación significativa, tenemos que el valor de la probabilidad obtenida es mayor que el límite, por lo que tendríamos que decir que no existe un incremento significativo en las horas de sueño con el uso del fármaco.

También podemos valorar las hipótesis bajo este mismo límite con el estadístico teórico $t_{.05,9} = 1.833$ y si lo comparamos con el estadístico calculado, tenemos que es menor que el estadístico calculado. Esto quiere decir que no sobrepasa nuestro límite, por lo que las desviaciones de las expectativas no son significativas.

Ejemplo de respuesta asociada al Nivel 4

Además de lo expresado en la respuesta del nivel previo, una respuesta en este nivel podría considera los siguientes aspectos:

La situación no indica el nivel de significancia, por lo que es posible utilizar el más común, un $\alpha = 0.05$.

Hipótesis nula $Ho: \mu = \mu_0$

Hipótesis alternativa $Ha: \mu > \mu_0$

Utilizando un software se obtiene la información de la Figura 6.6.

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for
10	0,750	1,789	0,566	-0,287

μ: population mean of Fármaco 1

Test

Null hypothesis	$H_0: \mu = 0$
Alternative hypothesis	$H_1: \mu > 0$

T-Value	P-Value
1,33	0,109

Figura 6.6. Prueba t-Student para una muestra

Dado que tenemos un valor del estadístico $t = 1.33$ y con un valor-p de 0.109, podemos ver que nuestro valor-p es mayor que el alfa, por lo que no tenemos evidencia suficiente para rechazar Ho. Otro criterio para tomar la decisión es contrastar el valor del estadístico que calculamos con el teórico $t_{.05,9} = 1.833$, tenemos entonces que $1.33 < 1.833$, por lo tanto, la diferencia no es significativa, lo que nos lleva a no rechazar Ho.

En la Figura 6.7 podemos observar que en la gráfica de la izquierda se muestra con rojo la región de rechazo a partir del valor del estadístico teórico (según el nivel de significancia y los grados de libertad), mientras que en la gráfica de la derecha podemos ver el estadístico calculado y su probabilidad asociada, y cómo se encuentra dentro de la zona de rechazo.

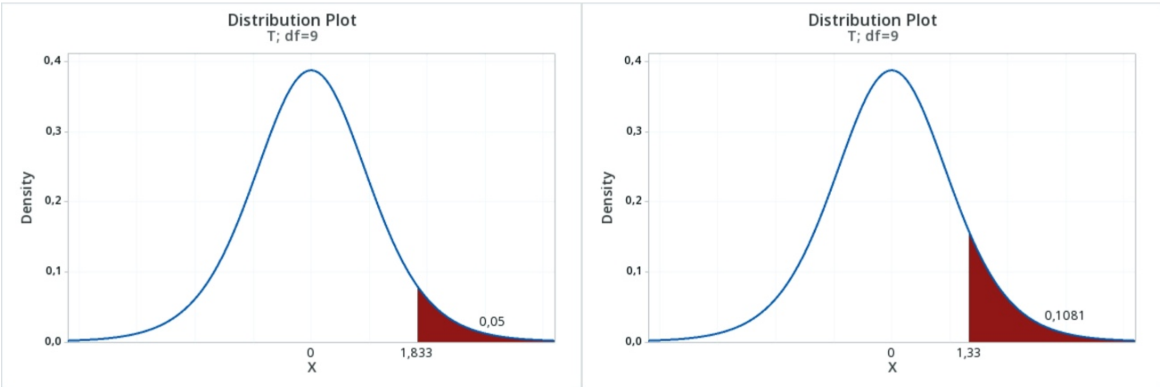


Figura 6.7. Distribución t con valor del estadístico teórico (izquierda) y calculado (derecha)

A partir de lo anterior, con un nivel de confianza del 95% no rechazamos Ho, es decir, no existe evidencia suficiente que nos lleve a rechazar Ho y podemos decir, con dicho nivel de confianza, que no hay diferencia significativa de la media de la población con $\mu_0 = 0$. Así, en términos del problema inicial, a través de esta prueba t-Student concluimos que no existe realmente un aumento en las horas de sueño al utilizar el fármaco.

En cuanto a la potencia de la prueba se tiene que es de 0.9862, entonces tenemos un muy buen valor de potencia de la prueba, esto nos indica que hay una prevalencia constante en la población, por lo que la conclusión obtenida con la prueba t-Student es muy buena. En la Figura 6.8 podemos ver dónde se ubica α y β, la probabilidad que tenemos de cometer el error tipo II (de no rechazar Ho cuando es falsa realmente), ya que no rechazamos Ho, es de $\beta = 0.0137$.

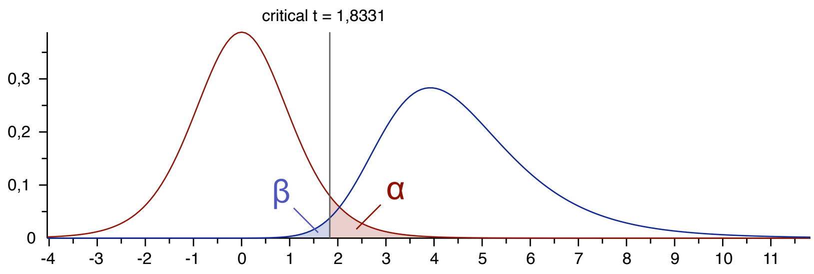


Figura 6.8. Errores tipo I y tipo II para prueba t de una muestra

6.1.2 Validación

Con relación a la validación de las actividades (ver Capítulo 2), esta se llevó a cabo por medio de la validación de contenido. Para ello, como se ejemplificó en el apartado anterior (6.1.1), se desarrollaron las prácticas que podrían realizar los profesores en cada nivel de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student, sobre las actividades dos y cuatro. De esta forma, se da cuenta de que cada actividad cubre en forma justa y completa el razonamiento inferencial, informal-preformal-formal, sobre los estadísticos Chi-cuadrada y t-Student.

Como mencionamos en el Capítulo 2, nuestros sujetos de estudio se componen por cuatro grupos y se aplicaron las actividades en el marco de talleres que se impartieron en modalidad virtual, esto debido al estado de contingencia sanitaria motivada por el Covid-19 que se tenía al momento de realizar este estudio. El esquema metodológico que se utilizó para aplicar las actividades fue similar en los grupos G1, G3 y G; en estos grupos, en las sesiones sincrónicas, los profesores formaron equipos para resolver las actividades y posteriormente regresaron a la sala general (de Zoom) para compartir sus estrategias de solución. Mientras que el G2 tuvo un esquema diferente, es decir, resolvieron las actividades de forma individual y la interacción fue por medio de foros y videos, esto se debió a la naturaleza del taller (asincrónica). Estos diferentes esquemas y nos permiten hablar sobre triangulación metodológica, pues se está utilizando el mismo método en diferentes ocasiones y diferentes métodos sobre el mismo objeto de estudio; por ejemplo, con los grupos G1 y G2, estos resolvieron las actividades 1, 2 y 3 sobre el estadístico Chi-cuadrada con diferente esquema y modalidad (Cohen, Manion y Morrison, 2011).

6.1.3 Dinámica de los talleres implementados

Como mencionamos en la sección 2.3.2, al abordar los sujetos de estudio, el taller de razonamiento estadístico con los grupos G1, G3 y G4, se llevó a cabo en modalidad sincrónica y para los grupos G3 y G4, también con modalidad asincrónica. En una primera parte de taller se estableció un dialogo con los profesores sobre aspectos como el razonamiento estadístico y la inferencia estadística; también se abrió un foro, posterior a la sesión sincrónica, para que los profesores de los grupos G3 y G4 pudieran comentar cómo consideraban que se podría promover este tipo de razonamiento en el aula de clases. Mientras que, en una segunda etapa, los formadores comentaron que los profesores participantes tendrían que resolver algunas actividades por equipos, los cuales se formaron aleatoriamente. Cada uno de los equipos tuvo una sala virtual (por medio de la herramienta de grupos pequeños de Zoom), los profesores contaron con 15 minutos para resolver la actividad. Se les pidió a los profesores participantes que una vez que estuvieran en sus salas comentaran las estrategias por medio de las cuales podrían solucionar el problema, que podían utilizar papel y lápiz, calculadora, software o el medio con el que se sintieran más cómodos para apoyarse para resolver el problema. Una vez que resolvían la actividad regresaban a la sala general para compartir con el grupo sus soluciones y las estrategias que siguieron (argumentos, procedimientos, representaciones, etc).

Antes de comenzar a trabajar en equipos, en la sala general, se les presentaba la actividad a los profesores y se les preguntaba si existían dudas sobre el problema propuesto; también se les mencionaba que no había una única forma de resolver la actividad, que podrían utilizar lo que ellos considerarán apropiado y que eso era precisamente lo que tenían que comentar con su equipo. Una vez que los equipos terminaron de realizar la actividad, regresaron a la sala general de Zoom, presentaron sus prácticas y comentaron porqué habían decido resolver la actividad siguiendo una determinada estrategia. Para finalizar con la actividad, los formadores en conjunto con los profesores participantes realizaron una nueva estrategia de solución para la actividad. Esta misma dinámica se realizó con cada una de las actividades.

En una tercera etapa del taller, al finalizar las actividades se procedió a presentar y comentar diversas prácticas matemáticas desarrolladas para resolver las actividades con distinto nivel de formalidad e identificar entre todos los rasgos de razonamiento inferencial observados en sus prácticas y en las prácticas presentadas por los formadores.

Por otra parte, por modalidad asincrónica del taller de razonamiento con el G2 la interacción fue por medio de videos y foros. En una primera etapa el formador del taller realizó y subió videos a una plataforma tipo Moodle, sobre temas como los desafíos que enfrenta la enseñanza de la estadística, el razonamiento estadístico y la inferencia estadística. Al igual que con los grupos G3 y G4, para el G2 también se abrió un foro para que los profesores comentaran cómo consideraban que se puede promover este tipo de razonamiento en el aula de clases.

En una segunda etapa del taller, mediante un video se les presentó a los profesores en formación las actividades y se les pidió que las resolvieran de forma individual; una vez que las resolvían y entregaban, debían ingresar a un foro de interacción sobre las actividades donde comentaban sus estrategias para resolverlas. Este espacio propició la interacción de los profesores participantes con el formador y sus pares, puesto que, además de plantear sus estrategias, plantearon las dudas y dificultades que habían tenido al resolver las actividades.

Para la tercera etapa, se utilizó un video, en el cual el formador mostraba y explicaba diversas estrategias de solución para las actividades, con distintos grados de formalidad. Estas estrategias de solución resultaron útiles para identificar en ellas rasgos de razonamiento inferencial.

6.2 Análisis de las prácticas de los profesores sobre el estadístico Chi-cuadrada

A continuación, analizamos las prácticas desarrolladas por los profesores en ejercicio (G4) y en formación (G1 y G2), a propósito de la solución de las actividades 1, 2 y 3. Para este análisis utilizamos la propuesta de niveles de razonamiento inferencial sobre el estadístico Chi-cuadrada (Capítulo 5) y la noción de práctica matemática y configuración de objetos y procesos, descritas en la sección de marco teórico.

Las figuras que presentamos en este capítulo, en las cuales se pueden observar las prácticas que realizaron los profesores, son capturas de pantalla de los videos, de las sesiones sincrónicas, y de los documentos (Word, Excel y PDF) que los profesores subieron a las plataformas como evidencia del desarrollo de las actividades sobre el estadístico Chi-cuadrada.

6.2.1 Prácticas asociadas a la actividad 1

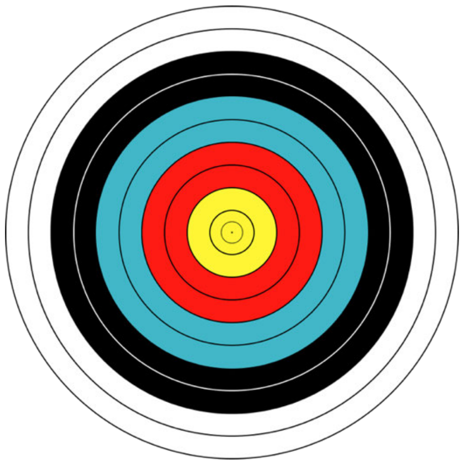
La actividad 1 trata de los lanzamientos de flechas en un torneo amateur de tiro con arco, donde interesa conocer si la forma en que se distribuyen las frecuencias observadas es conforme se espera (frecuencias esperadas), es decir, si las distancias siguen una distribución normal. A continuación, presentamos la actividad 1 y las prácticas que desarrollaron los profesores en formación y en ejercicio sobre esta actividad.

Actividad 1.

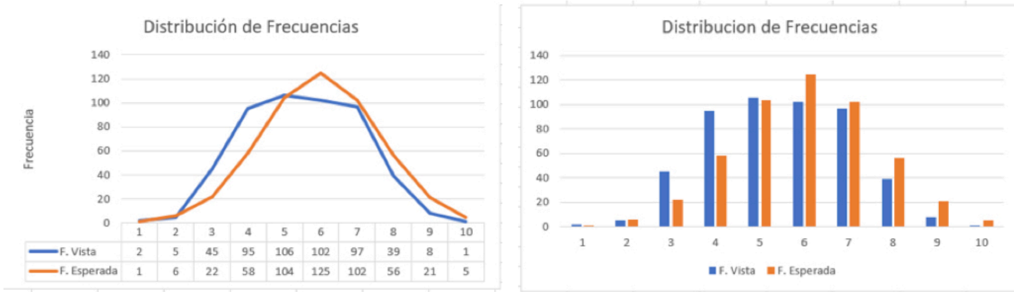
En el deporte de tiro con arco el objetivo es acertar lo más cerca posible del centro de la diana para obtener el máximo de puntos. En un torneo de amateurs de tiro con arco ninguna flecha ha dado en el punto central de la diana, sin embargo, se han contabilizado el número de flechas en cada uno de los anillos de la diana, siendo 1 el anillo más cercano al centro y por lo tanto con una desviación menor del punto central de la diana. Se cree que las desviaciones (distancias) del anillo de la diana donde se encuentra la flecha al centro de la diana siguen una distribución normal. En la tabla siguiente se encuentran contabilizadas las frecuencias que se observaron durante los tiros (frecuencia observada) y las frecuencias que se esperarían si las distancias siguen una distribución normal (frecuencia esperada). ¿Qué podrías comentar sobre la forma en que se distribuyen las frecuencias observadas? Explica detalladamente tu respuesta

Tabla 6.2. *Distribución de las frecuencias*

Anillo de la diana	Frecuencia Observada	Frecuencia Esperada
1	2	1
2	5	6
3	45	22
4	95	58
5	106	104
6	102	125
7	97	102
8	39	56
9	8	21
10	1	5
Total	500	500



El primer tipo de práctica característica para la actividad 1, se ejemplifica con los desarrollos del equipo 1 de los profesores en formación del primer grupo (Figura 6.9).



Conclusiones: Nos damos cuenta que los datos efectivamente siguen una distribución normal. Viendo la tabla de distribución de frecuencias nos damos cuenta que efectivamente hay diferencias en entre la distribución ideal y la observada, donde la observada tiene una mayor dispersión.

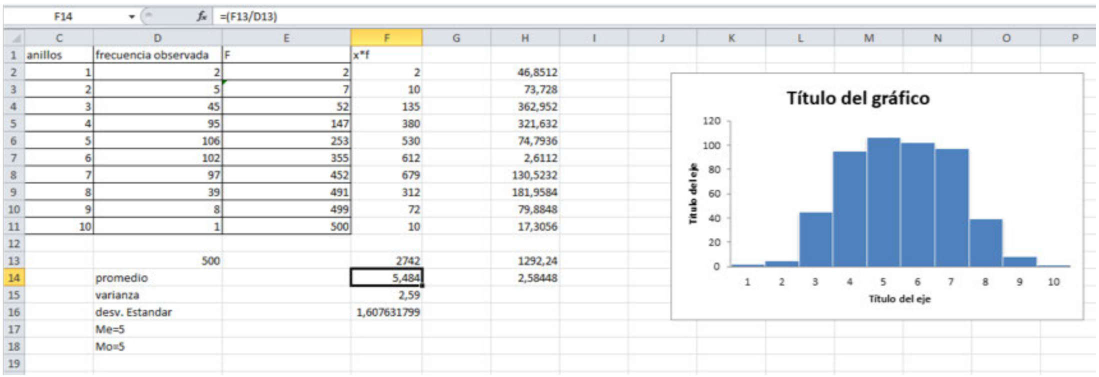
Figura 6.9. Práctica del equipo 1 de profesores en formación (G1) sobre la actividad 1

Los futuros profesores del equipo 1 señalaron que notaron que mientras más se acercaban al centro de la diana había menos flechas y que esto tenía sentido al ser un torneo de amateurs. También, que los anillos de la diana con menos flechas son el anillo 1 y el anillo 10, mientras que los anillos de las dianas con más flechas son los anillos 5 y 6. También indicaron que los datos que van desde el anillo 1 hasta el anillo 5 (en ese orden) tienen un comportamiento creciente, mientras que los datos que van del anillo 6 al anillo 10 (en ese orden) tienen un comportamiento decreciente. Además, realizaron dos gráficos de distribución de frecuencias, a partir de las frecuencias observadas (FO) y esperadas (FE) que se les proporcionó en el problema. A partir de las tendencias que ellos visualizaron en los gráficos, hicieron referencia a que "FO es parecida a la FE", "las formas de los gráficos son de tipo campana" y "parece que la FO tiene mayor dispersión que la FE". Finalmente, los profesores del equipo 1 conjeturaron que los datos observados efectivamente siguen una distribución normal.

En la práctica del equipo 1, identificamos objetos primarios como los *elementos lingüísticos* representaciones gráficas y lenguaje natural; algunos *conceptos/definiciones* como frecuencia observada y esperada, distribución de frecuencias y dispersión; mientras que la *propiedad/proposición* principal es la distribución normal. Los *procedimientos* que realizaron los futuros profesores fueron, realizar el gráfico de líneas y el gráfico de barras, basados en la información proporcionada en la tabla de la actividad 1, con el uso de Excel. Para *argumentar* sobre la conclusión que realizan, recurrieron principalmente a la forma de los gráficos, "parecida a una campana", y a la aparente mayor dispersión de los datos observados. Todos los objetos primarios que identificamos en la práctica matemática, y los procesos matemáticos que se encuentran asociados a ellos corresponden, al *Nivel 1* de razonamiento. Una de las particularidades del *Nivel 1*, en una primera etapa, tiene que ver con la visualización; entonces, en los indicadores que corresponden a un razonamiento inferencial informal se espera que los estudiantes o profesores puedan argumentar sus conjeturas en las características de los gráficos y, posteriormente, basarlas en un análisis los datos.

La práctica que realizó este primer equipo está relacionada con lo que se espera en la primera tarea para promover el RII del marco propuesto por Zieffler et al. (2008). Para una práctica con elementos más complejos del *Nivel 1*, los profesores podrían haber utilizado conceptos/definiciones como: ojiva, cuartiles y percentiles; y propiedades/proposiciones como: Q_2 en $\frac{1}{2}$, Q_1 en $\frac{1}{4}$, Q_3 en $\frac{3}{4}$, y las condiciones para una serie simétrica propias del método de intercomparación de Galton (1875). También podrían haber calculado y graficado las desviaciones, positivas y negativas, con respecto a la mediana, haciendo uso del concepto error medio y la propiedad error probable, y basar su conjetura en las características de los gráficos como la forma, pero también en la simetría, dispersión, amplitud de los cuartiles y/o en el sesgo.

Una segunda práctica que fue característica de la actividad 1, la realizó el equipo 3, este equipo estaba conformado por profesores en ejercicio. Los profesores se auxiliaron de Excel para realizar la actividad (Figura 6.10).



F14 =(F13/D13)

	C	D	E	F	G	H
1	anillos	frecuencia observada	F	x*f		
2	1	2	2	2		46,8512
3	2	5	7	10		73,728
4	3	45	52	135		362,952
5	4	95	147	380		321,632
6	5	106	253	530		74,7936
7	6	102	355	612		2,6112
8	7	97	452	679		130,5232
9	8	39	491	312		181,9584
10	9	8	499	72		79,8848
11	10	1	500	10		17,3056
12						
13		500		2742		1292,24
14		promedio		5,484		2,58448
15		varianza		2,59		
16		desv. Estandar		1,607631799		
17		Me=5				
18		Mo=5				
19						

Figura 6.10. Práctica del equipo 3 de profesores en ejercicio (G4) sobre la actividad 1

Los profesores del equipo 3 mencionaron que utilizaron Excel para determinar si las frecuencias observadas tenían un patrón de distribución normal y que lo primero que hicieron fue calcular el promedio, la varianza, desviación estándar, media y moda. Añadieron que, una de las características que tiene la distribución normal es que la media, moda y mediana tienen el mismo valor y que en el problema los valores de estas medidas corresponden a aproximadamente cinco. Después, señalaron que realizaron un histograma y que partir de éste apreciaron una simetría y que las dos colas eran muy parecidas (forma del gráfico).

Finalmente, concluyeron que los datos tienen una distribución normal y que otra estrategia podría ser utilizar un software estadístico para calcular algún estadístico de prueba.

En la práctica que realizaron los profesores en ejercicio del equipo 3 observamos *elementos lingüísticos* como lenguaje natural, representación tabular y gráfica, *conceptos/definiciones* como la frecuencia observada y esperada, y simetría, *propiedades/proposiciones* como la distribución normal, media, desviación estándar, mediana, moda y varianza. Los *procedimientos* que realizaron los profesores atienden a las propiedades indicadas con el uso de Excel; los *argumentos* estuvieron basados en la simetría y la forma del histograma, así como en la propiedad que tiene la distribución normal sobre que la media, moda y mediana coinciden. Estos objetos matemáticos primarios y los procesos que se encuentran asociados a ellos, pertenecen al *Nivel 1* de razonamiento inferencial sobre el estadístico Chi-cuadrada.

Para que la práctica de estos profesores progrese a una de *Nivel 2*, podrían iniciar por identificar la hipótesis que se encuentra implícita en el problema. De acuerdo con Pfannkuch y Wild (2004), Pfannkuch, et al. (2016) y Bakker, Ben-Zvi y Makar (2017), aproximarnos a las hipótesis estadísticas por medio de hipótesis en forma de preguntas o conjeturas puede favorecer la comprensión de las hipótesis. Además, los profesores podrían utilizar el estadístico Chi-cuadrada para valorar la bondad de ajuste del conjunto de datos y obtener la probabilidad asociada al valor del estadístico e interpretarla como una medida de ocurrencia de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema observado. De acuerdo con Stohl, Angotti y Tarr (2010), al utilizar la probabilidad de esta forma, los profesores estarán tomando decisiones de forma natural, tanto para mantener su hipótesis actual o para alterarla basándose en la probabilidad obtenida.

Un tercer tipo de práctica sobre la actividad 1, la ejemplificamos con la práctica que desarrolló el profesor en formación 21 (PF21) del grupo dos (Figura 6.11).

H_0: Los datos observados presentan una distribución normal.

H_A: $\neg H_0$

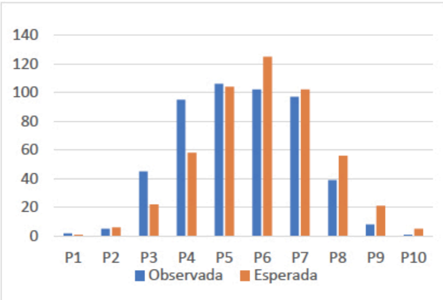
Se hizo el cálculo de chi-cuadrado y como resultado,

$$X^2 = 69{,}73..$$

Al no entender cómo interpretar este valor, utilicé una hoja de Excel para graficar los datos.

Noté que los datos que más aportan al X^2 son los que difieren más del dato esperado, por ejemplo, P3 y P4. Es decir, entre más cercano sea este valor valor a cero, los datos observados tendrán mayor similitud con los datos esperados.

Muestra	Observada	Esperada	X^2
P1	2	1	1
P2	5	6	0,166667
P3	45	22	24,04545
P4	95	58	23,60345
P5	106	104	0,038462
P6	102	125	4,232
P7	97	102	0,245098
P8	39	56	5,160714
P9	8	21	8,047619
P10	1	5	3,2
Suma	500	500	69,73946



Entonces, se entiende que los datos observados no corresponden a una distribución normal, de hecho, en la gráfica se nota que éstos representan una asimetría positiva.

Figura 6.11. Práctica del profesor en formación 21 (G2) sobre la actividad 1

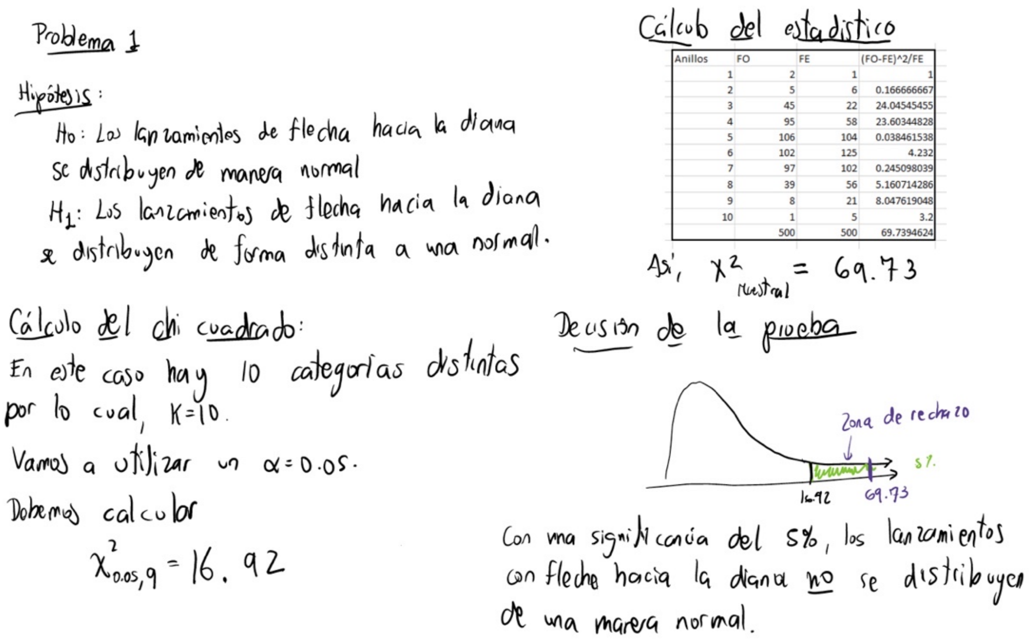
El PF21 planteó las hipótesis estadísticas y realizó el cálculo del estadístico Chi-cuadrada, sin embargo, no logró interpretar dicho valor del estadístico ni obtener el valor de la probabilidad, ya sea por medio de tablas de probabilidad de la distribución Chi-cuadrada, calculadora web o algún software (e.g., Excel, Minitab, R, SPSS); así que, a partir de la tabla que construyó para calcular el estadístico interpretó algunas contribuciones de cada categoría al estadístico Chi-cuadrada. En particular, distinguió que mientras dicha contribución sea más cercana a cero tendrá mayor similitud con las FE; también, realizó un gráfico de barras doble donde visualizó la asimetría de las FO. Estos aspectos le ayudaron a concluir que los datos observados no siguen una distribución normal.

En la práctica del PF21 distinguimos los *elementos lingüísticos* lenguaje natural y simbólico, y representaciones gráfica y tabular; algunos *conceptos/definiciones* como frecuencia observada y esperada, y simetría; mientras que la *propiedad/proposición* principal es el estadístico Chi-cuadrada. Los *procedimientos* que realizó el profesor fueron el cálculo del estadístico y el gráfico de barras con el uso de Excel. Para *argumentar* sobre la conclusión que realiza, recurrió principalmente a la asimetría que presentan los datos observados. En general, los objetos primarios que hemos identificado en la práctica matemática del PF21 corresponden al *Nivel 2* de razonamiento inferencial sobre el estadístico Chi-cuadrada.

Aunque la práctica que desarrolló este profesor es de *Nivel 2*, encontramos rasgos de *Nivel 1*, tales como el uso de las características observadas en el gráfico de barras para argumentar la conclusión, y de *Nivel 4*, como la forma en que se encuentran planteadas las hipótesis. Sin embargo, a pesar de que el PF21 reconoció el uso del estadístico Chi-cuadrada para resolver el problema, tuvo dificultades para concluir la prueba de hipótesis. Para robustecer esta práctica de *Nivel 2*, el PF21 podría haber obtenido el valor de la probabilidad para el estadístico que ha calculado e interpretarlo como lo hemos comentado en la práctica anterior. Para realizar esto es necesario que el profesor conozca la distribución Chi-cuadrada y comprenda sus propiedades, podría iniciar con applets de simulaciones para visualizar cómo se comporta esta distribución al variar su parámetro. De acuerdo con diversas investigaciones (e.g., Chance y Rossman, 2006; Tintle, Topliff, VanderStoep, Holmes y Swanson, 2012; Rossman y Chance, 2014), los recursos tecnológicos ayudan a los estudiantes y profesores a interactuar con las nociones estadísticas, lo cual favorece la comprensión de dichas nociones.

Para una práctica de *Nivel 3*, el profesor podría haber identificado que, como hay frecuencias esperadas inferiores a cinco, sería apropiado utilizar el estadístico Chi-cuadrada con factor de corrección de continuidad; también, podría emplear el nivel de significancia de forma preformal. En la propuesta de niveles de razonamiento inferencial (Capítulo 5), se sugiere trabajar con el concepto/definición de significancia como límite de desviación significativo; a partir de dicho límite, al que suele asociarse a un valor de probabilidad, se considera que sí es probable que exista una diferencia. Esta aproximación para trabajar la significancia fue aplicada por Fisher (1925a) y concuerda con la propuesta de Rossman (2008) para introducir algunas ideas sobre inferencia.

La siguiente práctica corresponde al profesor en formación cinco del grupo dos (Figura 6.12), y ejemplifica otra tipología de respuestas para la actividad.



Problema 1

Hipótesis:

Ho: Los lanzamientos de flecha hacia la diana se distribuyen de manera normal

H_1: Los lanzamientos de flecha hacia la diana se distribuyen de forma distinta a una normal.

Cálculo del chi cuadrado:

En este caso hay 10 categorías distintas por lo cual, K=10.

Vamos a utilizar un $\alpha = 0.05$.

Debemos calcular

$\chi^2_{0.05,9} = 16.92$

Cálcub del estadístico

Anillos	FO	FE	(FO-FE)^2/FE
1	2	1	1
2	5	6	0.166666667
3	45	22	24.04545455
4	95	58	23.60344828
5	106	104	0.038461538
6	102	125	4.232
7	97	102	0.245098039
8	39	56	5.160714286
9	8	21	8.047619048
10	1	5	3.2
	500	500	69.7394624

Así, $\chi^2_{muestral} = 69.73$

Decisión de la prueba

Con una significancia del 5%, los lanzamientos con flecha hacia la diana no se distribuyen de una manera normal.

Figura 6.12. Práctica del profesor en formación 5 (G2) sobre la actividad 1

El profesor en formación 5 (PF5) del segundo grupo, planteó las hipótesis nula y alternativa, posteriormente realizó el cálculo del estadístico Chi-cuadrada, definió un valor de alfa para obtener el valor crítico o del estadístico teórico y lo utilizó para decidir sí el conjunto de datos observados sigue una distribución normal, para lo cual realiza un gráfico de la distribución chi-cuadrada donde indicó la región de rechazo. A partir de esto, el profesor concluyó que puede afirmar con una significancia del 5% que los datos no se distribuyen como una normal.

En esta práctica podemos identificar que los *elementos lingüísticos* que utilizó el PF5 son lenguaje natural y simbólico, y representaciones tabular y gráfica; algunos *conceptos/definiciones* son frecuencia observada y esperada, significancia, hipótesis y zona de rechazo, mientras que las *propiedades/proposiciones* son el estadístico Chi-cuadrada, grados de libertad, alfa (nivel de significancia), regla de decisión con el estadístico y región crítica. Los *procedimientos* que realizó fueron calcular el estadístico Chi-cuadrada, los grados de libertad, determinar el nivel de significancia, obtener el valor crítico y compararlo con el valor del estadístico calculado mediante una representación gráfica para tomar la decisión de rechazar o no la hipótesis nula. Finalmente, *argumenta* su conclusión indicando que como el valor del estadístico Chi-cuadrada es mayor que el valor crítico, entonces los datos observados en el lanzamiento de tiro con arco no pueden seguir una distribución normal.

En síntesis, la práctica que desarrolla el PF5 es característica del *Nivel 4*. Sin embargo, al utilizar el estadístico Chi-cuadrada, tampoco se percata que es necesario utilizar un factor de corrección de continuidad debido a que se tienen frecuencias esperadas inferiores a cinco, y esto se debe a que esta prueba es aproximada; y es que χ es una distribución continua y la distribución que se intenta aproximar es discreta, entonces la precisión de la aproximación de la prueba depende de los valores de las observaciones y cuando se trabaja con números pequeños la discrepancia que se genera se hace notoria.

Para que el PF5 fortalezca su razonamiento inferencial formal del *Nivel 4*, podría explorar los errores tipo I y tipo II y comprender las relaciones entre dichos errores, el nivel de significancia y la potencia. De forma similar a la propuesta de Rossman (2008), para la comprensión de la distribución o el valor-p, para promover las propiedades/proposiciones errores tipo I y tipo II, nivel de significancia y la potencia de la prueba, así como las relaciones entre estas, se puede auxiliar de software, simulaciones y representaciones gráficas.

Las prácticas anteriores, son ejemplares del tipo de prácticas que realizaron tanto los profesores en formación como los profesores en ejercicio al resolver la actividad 1. A continuación, en la Figura 6.13 presentamos otra práctica que se destacó en el segundo grupo de profesores en formación.

Solución.

Se va a realizar una prueba de bondad de ajuste. Las hipótesis de prueba son las siguientes:

- H_0: Los lanzamientos de flecha hacia la diana siguen una distribución normal.
- H_1: Los lanzamientos de flecha hacia la diana no siguen una distribución normal.

Dado que hay 10 categorías distintas (en este caso, cada anillo de la diana), se sabe que $k = 10$, lo que implica que el grado de libertad es 9.

Se procede a usar un nivel de significancia $\alpha = 0.05$.

Continuación.

Ahora, calculando el p-valor (haciendo uso del software R)

$$x^2_{0.05,9} = 16.92$$

Luego, haciendo el cálculo del estadístico chi-cuadrado (haciendo uso de Excel), se tiene lo siguiente:

Continuación.

Anillo de la diana	Frecuencia Observada	Frecuencia Esperada
1	2	1
2	5	6
3	45	22
4	95	58
5	106	104
6	102	125
7	97	102
8	39	56
9	8	21
10	1	5
Total	500	500

Anillo de la diana	$\frac{(FO-FE)^2}{FE}$
1	1
2	0,16666667
3	24,0454545
4	23,6034483
5	0,03846154
6	4,232
7	0,24509804
8	5,16071429
9	8,04761905
10	3,2
Suma	69,7394624

por lo tanto

$$x^2_{\text{muestral}} = 69,74$$

Dado que el estadístico es mayor que el p-valor, se puede concluir que, con una significancia del 5 %, los lanzamientos de flecha hacia la diana no siguen una distribución normal. □

Figura 6.13. Práctica del profesor en formación 11 (G2) sobre la actividad 1

El profesor en formación 11 (PF11) del segundo grupo realizó una prueba de bondad de ajuste con el estadístico Chi-cuadrada, para llevarla a cabo planteó las hipótesis, identificó que los datos contaban con diez categorías, a partir de esto determinó los grados de libertad. Posteriormente, estableció un nivel de significancia de 0.05 e indicó que calculó el estadístico teórico con el uso del software R y con el uso de Excel calculó el valor del estadístico de prueba. El PF5 concluyó, con un nivel de significancia del 5%, que los lanzamientos de la flecha hacia la diana no siguen una distribución normal. Como podemos observar, la práctica que desarrolla el PF11 es muy similar a la desarrollada por el PF5; sin embargo, el PF11 parece comprender erróneamente el valor-p, pues realmente está trabajando con el valor crítico o estadístico teórico. Consideramos que no es una simple equivocación, pues este error, se observó en otras prácticas de las actividades 1 y 2, y también cuando se trabaja con el estadístico de prueba. Si el profesor realmente comprendiera esta noción, se percataría de que está trabajando con un valor de probabilidad y que no puede ser mayor a uno. La propuesta de niveles de razonamiento inferencial sobre el estadístico Chi-cuadrada promueve la comprensión progresiva de nociones como el valor-p. El profesor en formación podría iniciar con una aproximación informal al valor-p, auxiliándose de software para simulaciones o para el cálculo de la probabilidad del estadístico Chi-cuadrada (*Nivel 2*), y con ello apoyar su inferencia (Rossman, 2008; Rossman y Chance, 2014).

6.2.2 Prácticas asociadas a la actividad 2

La actividad 2 trata de los datos recabados durante una epidemia de viruela, donde interesa establecer si existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela. A continuación, presentamos la actividad 2 y las prácticas que realizaron los profesores en formación del grupo 1.

Actividad 2.

A continuación, se presentan los datos de una muestra que fueron recabados durante una epidemia de viruela en un pequeño pueblo de Inglaterra. Los datos de la tabla muestran la presencia o ausencia de la cicatriz de la vacuna de viruela y si las personas que recibieron la vacuna se recuperaron o fallecieron. ¿Existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela? Explica con detalle tu respuesta

Tabla 6.3. *Presencia de cicatriz por vacuna y recuperados de viruela*

Cicatriz	Recuperados	Muertos	Total
Presente	35	1	36
Ausente	10	4	14
Total	45	5	50

El primer tipo de práctica característica para esta actividad se ejemplifica con la desarrollada por el equipo 2 del primer grupo de profesores en formación (Figura 6.14).

Vemos que el 28.57% de las personas vacunadas sin cicatriz murieron, mientras que 2.77% de las personas que sí tienen cicatriz por la vacuna murieron. Lo anterior indica que tener la cicatriz por la vacuna de viruela pareciera tener una relación con no morir de esta enfermedad.

Figura 6.14. Práctica del equipo 2 de profesores en formación (G1) sobre la actividad 2

Los profesores del equipo dos, indicaron que en primera instancia les parecía muy atractivo decir que la muestra no es lo suficientemente grande como para asegurar una relación entre los muertos y que tuvieran o no la cicatriz de la vacuna, pero como equipo decidieron explorar los datos y realizaron gráficas de pastel sobre muertos versus sobrevivientes con y sin vacuna. Los profesores en formación indicaron que, una vez que realizaron los gráficos, interpretaron que sí existe una relación significativa entre la presencia de la cicatriz y la recuperación del paciente ya que podían notar que sólo el 2.78% de los enfermos con la cicatriz perecieron, mientras que el 28.57% de los enfermos que no tenían la cicatriz fallecieron.

En la práctica del equipo dos observamos que los *elementos lingüísticos* que utilizaron son lenguaje natural y representación gráfica, algunos *conceptos/definiciones* son muestra, variable, frecuencia y asociación (relación), mientras que los *procedimientos* que llevaron a cabo son realizar los gráficos de pastel mediante Excel. Finalmente, *argumentaron* su conclusión con los porcentajes de enfermos que fallecieron y que tenían la cicatriz, así como los que no tenían la cicatriz. Todos los objetos matemáticos primarios y sus procesos asociados que hemos identificado en esta práctica matemática corresponden al *Nivel 1* de razonamiento inferencial sobre el estadístico Chi-cuadrada. Además de lo desarrollado por los profesores en formación de este equipo, podrían haber analizado la asociación de las dos variables mediante el coeficiente de asociación (Yule, 1900), que se puede considerar un precedente a la prueba de independencia con el estadístico Chi-cuadrada; es por ello que en la propuesta de niveles de razonamiento inferencial se considera al coeficiente de asociación en el *Nivel 1*, como una versión intuitiva que puede apoyar las conjeturas o inferencias informales en los problemas de independencia.

En la Figura 6.15 podemos observar la práctica del equipo 5 de profesores en ejercicio, la cual es característica de esta actividad.

2. Sí es posible establecer una relación, en la medida que, si calculamos la probabilidad condicionada de que una persona recuperada presente la cicatriz de la viruela o no la presente, observaremos que

Sea A={Persona Recuperada} y B={Presenta cicatriz de la vacuna}

$$P(B|A) = 0{,}78$$

Por lo tanto, la probabilidad que una persona recuperada presente la cicatriz es de un 78% en comparación a un 22% de probabilidad que esa persona recuperada no presente dicha cicatriz. Cabe destacar que esta es la probabilidad medida para esta muestra, la cual nos permitiría someterla en el futuro a algún modelo o prueba estadística inferencial y así determinar, por ejemplo, algún intervalo de confianza.

Figura 6.15. Práctica del equipo 5 de profesores en ejercicio (G4) sobre la actividad 2

Los profesores del equipo cinco optaron por establecer si existe una relación entre haberse recuperado y la presencia de cicatriz mediante la probabilidad condicional. Para ello, establecieron los eventos A y B en términos del problema y calcularon la probabilidad de presentar cicatriz de la vacuna dado que se ha recuperado. Los profesores señalaron que existe una probabilidad de 78% de que quienes se recuperaron presenten cicatriz de la vacuna, en comparación con la probabilidad del 22% de no presentarla. Así mismo, que el 97% de los que presentan la cicatriz se recuperaron, mientras que el 71% de los que no la presentan se recuperaron. Por lo anterior, consideraron que sí existe una relación entre recuperarse y presentar cicatriz de la vacuna. Los profesores destacaron que las probabilidades que obtuvieron hacen referencia a esta única muestra, pero que se podría someter a algún tipo de técnica de inferencia estadística para poder determinar la existencia de asociación.

En la práctica que realizó el equipo 5, identificamos *conceptos/definiciones* tales como muestra, eventos de probabilidad, probabilidad y asociación (relación). La *propiedad/proposición* que destaca es probabilidad condicional, mientras que los *procedimientos* fueron calcular las probabilidades condicionales. Los profesores *argumentaron* su conclusión con las probabilidades obtenidas. La mayoría de los elementos primarios identificados en esta práctica corresponden al *Nivel 1*. Para robustecer la práctica de este nivel, que corresponde a un razonamiento inferencial informal, los profesores podrían obtener las distribuciones condicionales por renglón, por columnas y la distribución conjunta de toda la muestra. Además, también podría ayudar realizar gráficos de barras apiladas o de mosaico para visualizar las proporciones, tal y como lo sugieren Batanero y Borovcnik (2016) para aproximarse a una prueba de independencia con el estadístico Chi-cuadrada.

Un tercer tipo de práctica para la actividad dos, se ejemplifica con la desarrollada por el quipo 4 de profesores en formación del grupo uno (Figura 6.16).

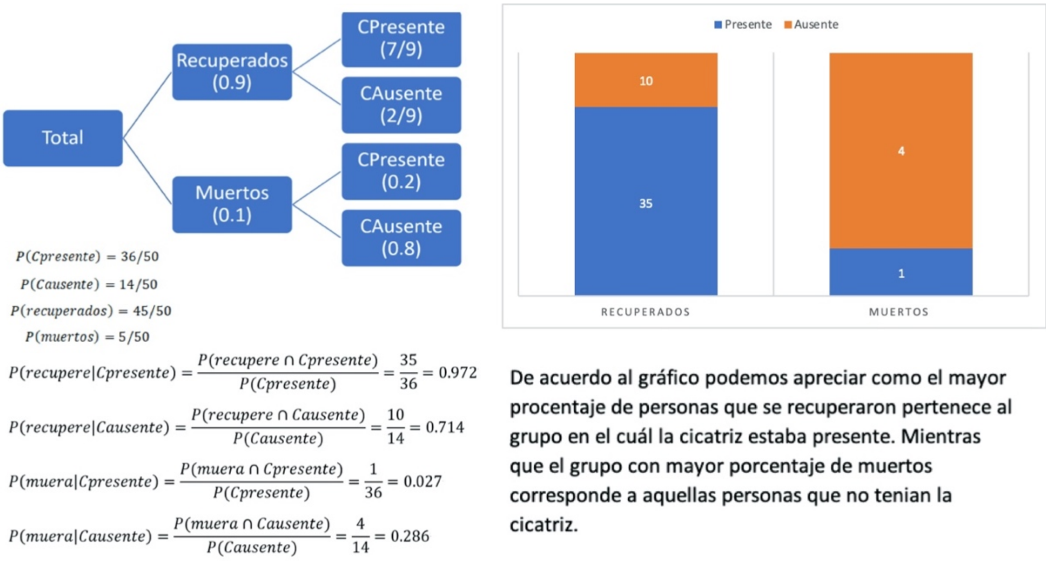


Figura 6.16. Práctica del equipo 4 de profesores en formación (G1) sobre la actividad 2

Los profesores en formación del equipo 4, mencionaron que primero realizaron un gráfico de barras apiladas con las frecuencias proporcionadas en la tabla de la actividad y después un diagrama de árbol con las posibilidades brindadas en la tabla, lo cual representa la distribución condicional por columnas, y que después consideraron calcular la distribución condicional por renglón. Además, a partir del gráfico y de las probabilidades calculadas decidieron que existe una relación entre recuperarse y la presencia de la cicatriz.

En la práctica del equipo 4, identificamos el uso del lenguaje natural y simbólico, así como de representaciones gráficas como *elementos lingüísticos*, algunos *conceptos/definiciones* como muestra, eventos de probabilidad, probabilidad y asociación (relación); también *propiedades/proposiciones* tales como probabilidad condicional, distribución condicional por renglón y por columna. Los *procedimientos* realizados por el equipo de profesores consisten en la elaboración del gráfico de barras apiladas, el cálculo de las probabilidades condicionales para las distribuciones condicionales y formación del diagrama de árbol. Los *argumentos* estuvieron basados en la forma del gráfico de barras apiladas y en los valores de probabilidad obtenidos. Los objetos matemáticos identificados en esta práctica corresponden al *Nivel 1* de razonamiento inferencial sobre el estadístico Chi-cuadrada.

Para que la práctica de los profesores progrese a una de *Nivel 2*, podrían preguntarse si la diferencia entre las probabilidades obtenidas será realmente evidencia suficiente para decidir si existe relación entre la presencia de cicatriz y los recuperados; y para responder los profesores podrían llevar a cabo una prueba de independencia con el estadístico Chi-cuadrada. Para ello, es necesario que los profesores primero reconozcan en el problema la hipótesis que se encuentra implícita, la identificación de la hipótesis en forma de pregunta es un rasgo clave en este nivel, además, esta forma presenta una primera aproximación a las hipótesis estadísticas formales que se trabajan en el *Nivel 4* (Capítulo 5). En este sentido, el ciclo de investigación estadística (PPDAC) de Wild y Pfannkuch (1999), tiene como primer componente la generación de una pregunta de investigación, la cual debe estar dada sobre un contexto particular, que es sobre el cual se desea indagar. En este nivel, es importante que el profesor conozca y comprenda la distribución Chi-cuadrada (e.g., es asimétrica positiva, tiene como único parámetro a los grados de libertad, a medida que incrementan los grados de libertad se aproxima a la curva normal y no puede tomar valores negativos), algunas investigaciones sugieren el uso de recursos tecnológicos para promover la comprensión de dicha noción (eg., Bakker y Gravemeijer, 2004; Reading y Reid, 2006; Rossman, 2008; Dinov, et al., 2018). Además, los profesores podrían utilizar la probabilidad como una medida de cuán lejos el sistema observado es o no compatible con las bases de independencia probabilística; en otras palabras, utilizar la probabilidad para realizar inferencias preformales (*Nivel 1*) y después podría explorar comparar la probabilidad con un límite preestablecido (*Nivel 3*) o un nivel de significancia (*Nivel 4*).

La Figura 6.17 ejemplifica el cuarto tipo de práctica característica de la actividad dos, esta práctica fue desarrollada por profesores en formación del segundo grupo.

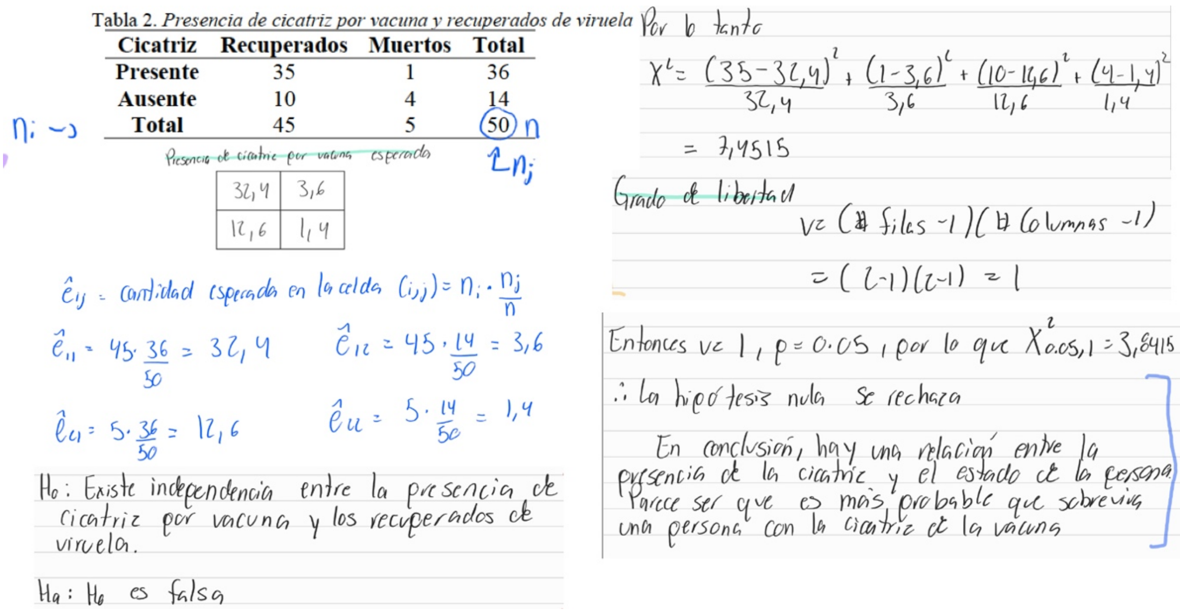


Figura 6.17. Práctica del profesor en formación 17 (G2) sobre la actividad 2

El profesor en formación 17 (PF17), señaló que decidió aplicar una prueba de independencia con el estadístico Chi-cuadrada porque se deseaba conocer si existía una relación entre las variables: cicatriz por la vacuna y sobrevivir. Para llevar a cabo la prueba de independencia calculó las frecuencias esperadas, planteó las hipótesis, calculó es estadístico Chi-cuadrada y los grados de libertad; también estableció una probabilidad de 0.05 para obtener el valor crítico, como un límite. Finalmente, el profesor rechaza la hipótesis nula debido a que le valor del estadístico Chi-cuadrada calculado excede el valor crítico, y concluye que sí hay una relación entre la presencia de cicatriz y el estado de la persona.

En esta práctica podemos identificar el uso de *elementos lingüísticos* como el lenguaje natural y simbólico, de *conceptos/definiciones* como muestra, variables, categorías, frecuencias (observada y esperada) e hipótesis; también *propiedades/proposiciones* tales como el estadístico Chi-cuadrada, frecuencia esperada, grados de libertad y criterio para la toma de decisión. Los *procedimientos* realizados consisten en el planteamiento de las hipótesis, cálculo de las frecuencias esperadas, el cálculo del estadístico Chi-cuadrada, de los grados de libertad, establece $P = 0.05$ como límite y obtiene el valor del estadístico teórico, compara este valor crítico con el estadístico calculado y, finalmente, decide si se rechaza la hipótesis nula. Los *argumentos* que realizó estuvieron basados en los criterios para la toma de decisión; es decir, como el valor del estadístico de prueba es mayor que el del estadístico teórico, entonces se rechaza la hipótesis nula. La mayoría de los objetos primarios que identificamos en la práctica del PF17 son del *Nivel 4*; sin embargo, nos llama la atención el uso que da a la probabilidad como un límite significativo, este aspecto es característico del *Nivel 3* y en nuestra propuesta de niveles se considera como una acepción preformal del nivel de significancia, donde a partir de este límite se puede considerar seriamente que sí es probable que exista una diferencia. Como mencionamos en la práctica del PF21 (Figura 6.11), esta aproximación al nivel de significancia concuerda con algunas propuestas para introducir nociones de inferencia.

Para que el PF17 fortalezca su razonamiento inferencial formal, podría profundizar en su comprensión del nivel de significancia y de confianza, así como trabajar con los errores tipo I y tipo II, y validar su inferencia haciendo uso de la potencia de la prueba. El profesor podría iniciar con una aproximación, Vera y Díaz (2013) sugieren iniciar la exploración de la idea de probabilidad condicional mediante applets, para visualizar las diversas posibilidades. Es importante que el profesor comprenda las relaciones que existen entre todas las nociones mencionadas y no las vea como conceptos aislados o fórmulas que le ayudan a realizar un algoritmo; en otras palabras, se requiere un enfoque holístico para que el profesor aprenda a razonar estadísticamente y comprenda profundamente estas y otras nociones estadísticas (Bakker y Derry, 2011; Makar y Ben-Zvi, 2011). Es por ello, que la propuesta de niveles de razonamiento inferencial sobre el estadístico Chi-cuadrada tiene un carácter progresivo, lo cual permiten explorar las diversas nociones en varios momentos (en los niveles) con diferente grado de complejidad, profundidad y formalidad.

6.2.3 Prácticas asociadas a la actividad 3

La actividad 3 trata de cómo se conforman de acuerdo con su raza, dos grupos (hombres y mujeres) que padecen cáncer de pulmón, donde interesa conocer si existe homogeneidad entre los dos grupos. A continuación, presentamos la actividad 3 y las prácticas que desarrollaron los profesores sobre esta actividad.

Actividad 3.

A partir de una encuesta realizada en el 2017 por el CDC (Centro para el Control y Prevención de Enfermedades de los Estados Unidos) a un grupo de hombres y un grupo de mujeres que padecen cáncer de pulmón, se obtuvo la siguiente información sobre las frecuencias de enfermos de cáncer por raza (casta o etnia). Para realizar un tratamiento adecuado de los datos es necesario conocer si existe homogeneidad entre los dos grupos respecto de las razas. ¿Consideras que existe homogeneidad entre los grupos (Hombres y Mujeres)? Argumenta tu respuesta

Tabla 6.4. *Cáncer de pulmón por género y raza*

	Caucásico	Afroamericano	Asiático	Hispano	Total
Hombres	95	12	3	8	118
Mujeres	92	10	3	6	111
Total	187	22	6	14	229

Un primer tipo de práctica característica para la actividad 3, se ejemplifica con los desarrollos del equipo 1 de los profesores en formación del primer grupo (Figura 6.18).

	Caucásico	Afroamericano	Asiático	Hispano
Hombres	$\frac{95}{118} \cdot 100 = 80{,}5\%$	$\frac{12}{118} \cdot 100 = 10{,}1\%$	$\frac{3}{118} \cdot 100 = 2{,}5\%$	$\frac{8}{118} \cdot 100 = 6{,}7\%$
Mujeres	$\frac{92}{111} \cdot 100 = 82{,}8\%$	$\frac{10}{111} \cdot 100 = 9\%$	$\frac{3}{111} \cdot 100 = 2{,}7\%$	$\frac{6}{111} \cdot 100 = 5{,}4\%$

Figura 6.18. Práctica del equipo 1 de profesores en ejercicio (G4) sobre la actividad 3

Los profesores del equipo 1 indicaron que, al calcular y comparar los porcentajes, los grupos (hombres y mujeres) tenían mucho en común, es decir, la probabilidad de caucásico (Hombres 80,5% - Mujeres 82,8%), afroamericano (Hombres 10,1% - Mujeres 9%), asiático (Hombres 2,5% - Mujeres 2,7%) e hispano (Hombre 6,7%- Mujeres 5,4%), son bastante similares, por lo cual consideraron que sí existe homogeneidad entre los grupos. Además, destacaron que para probarlo se deben de realizar pruebas estadísticas sobre diferencias de grupos y que creían que se podía resolver, siempre cuando se cumplan ciertos criterios, mediante una prueba t de Student. Finalmente, señalaron que para poder apreciar de manera informal se podía ver que las diferencias en los porcentajes no parecían ser significativas.

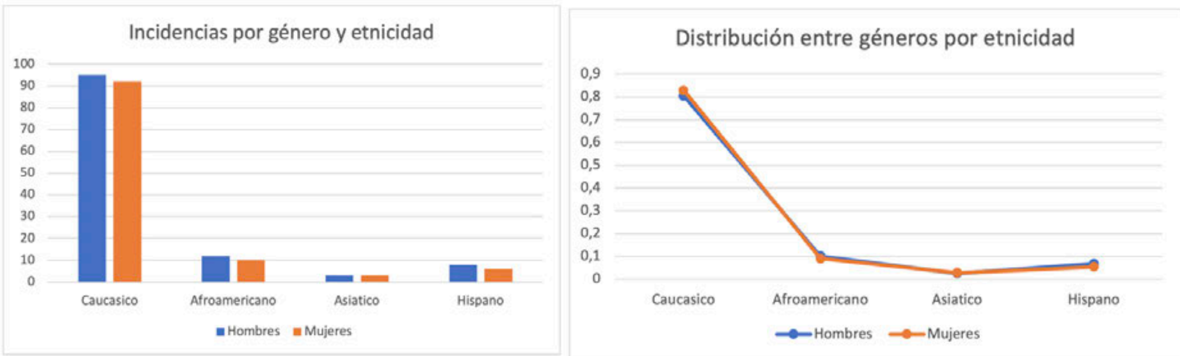
En la práctica del equipo uno observamos que los *elementos lingüísticos* que utilizaron son lenguaje natural y simbólico, algunos *conceptos/definiciones* tales como grupos, variable, frecuencia, porcentaje y significancia, mientras que la *propiedad/proposición* que destaca es la probabilidad condicional. Los *procedimientos* que llevaron a cabo son obtener la probabilidad condicional por renglón y las diferencias entre los grupos por categoría (raza), finalmente *argumentaron* su conclusión con la aparente no significancia de las diferencias de los porcentajes (probabilidad) entre los grupos por categoría de raza. Los elementos primarios identificados en esta práctica corresponden al *Nivel 1* de razonamiento inferencial para el estadístico Chi-cuadrada.

Para que el profesor desarrolle una práctica más consolidada de *Nivel 1*, se podrían atender la propuesta que se hizo para el equipo 5 de profesores en ejercicio (Figura 6.15), como obtener las distribuciones condicionales por columnas y la distribución conjunta; también podrían realizar gráficos de barras apiladas o mosaico y fundamentar en ello su inferencia. Es importante comentar que los profesores indicaron que este problema podría resolverse mediante una prueba t-Student, sin embargo, para el tipo de variable que se trabaja en este problema es factible recurrir a una prueba Chi-cuadrada de Homogeneidad. Entonces, consideramos que los profesores podrían iniciar con reconocer el tipo de datos con los que está trabajando (e.g., se trata de una muestra o de una población, son cualitativos o cuantitativos, se clasifican de acuerdo a una o dos variables, número de muestras), especialmente con las características de las variables.

En la figura 6.19 se ejemplifica un segundo tipo de práctica que resultó característica esta actividad.

	Caucasico	Afroameric	Asiatico	Hispano	Total
Hombres	95	12	3	8	118
Mujeres	92	10	3	6	111
Total	187	22	6	14	229

	Caucasico	Afroameric	Asiatico	Hispano	Total
Hombres	0,8050847	0,1016949	0,0254237	0,0677966	1
Mujeres	0,8288288	0,0900901	0,027027	0,0540541	1



Razón Hombres/Mujeres				
	Caucasico	Afroameric	Asiatico	Hispano
Ratio	1,0326087	1,2	1	1,3333333

Figura 6.19. Práctica del equipo 6 de profesores en formación (G1) sobre la actividad 3

El equipo 6 de profesores en formación, indicaron que a simple vista parecía que en los datos se mantiene una diferencia de 2 y 3 personas entre hombres y mujeres a excepción del caso de 3 que son los asiáticos, y que al hacer el cálculo de las medidas estadísticas se podía observar que sus compartimientos son muy parecidos tanto en hombres como en mujeres y las diferencias se mantienen en promedio de 2. Los profesores también señalaron que decidieron realizar los gráficos para observar si podían apreciar diferencias entre las muestras. Finalmente, decidieron obtener la razón por cada categoría y concluyeron que los datos son homogéneos porque tienen un comportamiento muy similar y sus diferencias ser muy cercanas.

En la práctica que realizaron los profesores en ejercicio del equipo 6 observamos elementos lingüísticos como lenguaje natural, representación tabular y gráfica, *conceptos/definiciones* como muestra, variable, categorías o cualidades, frecuencia; *propiedades/proposiciones* como probabilidad condicional y razón. Los *procedimientos* que realizaron los profesores son probabilidad condicional por renglón, razón y gráficos de barras dobles y de líneas con el uso de Excel; los *argumentos* estuvieron basados en las escasas diferencias que observaron tanto en los gráficos como en las probabilidades condicionales. Estos objetos matemáticos primarios y los procesos matemáticos que se encuentran asociados a ellos, pertenecen al *Nivel 1* de razonamiento inferencial.

Para que los profesores del equipo seis desarrollen una práctica de *Nivel 2*, sugerimos que es necesario que identifiquen que el problema se puede resolver mediante una prueba Chi-cuadrada, e inicien por reconocer en el problema la hipótesis nula y puedan expresarla en lenguaje natural, además de otros aspectos ya señalados en la propuesta para los profesores en ejercicio del equipo 3 (Figura 6.10).

La tercera práctica representativa de la actividad tres se muestra en la Figura 6.20 y fue desarrollada por el profesor en formación 14 (PF14).

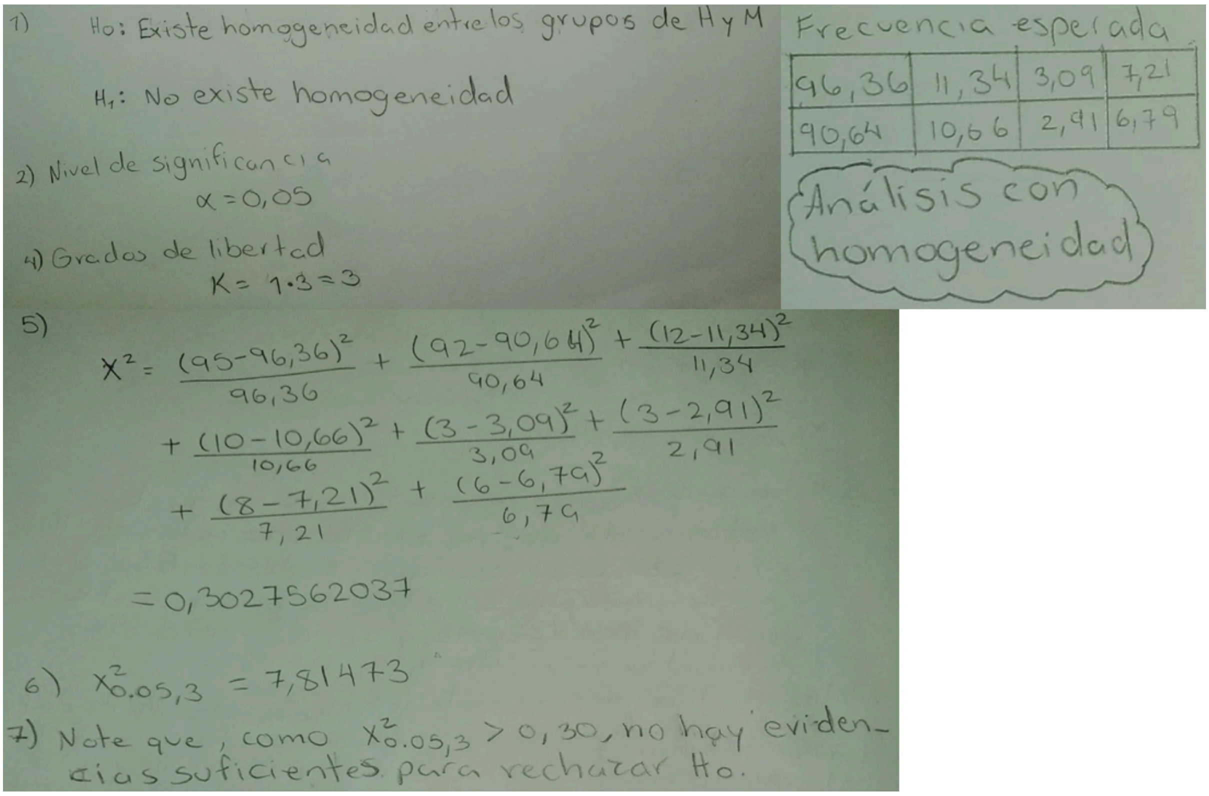


Figura 6.20. Práctica del profesor en formación 14 (G2) sobre la actividad 3

El PF14 partió indicando que, de acuerdo con los datos que brindaba el problema, se debía realizar un análisis con homogeneidad y que para llevarlo a cabo utilizaría una prueba Chi-cuadrada, e indicó que la hipótesis nula sería que "existe homogeneidad entre los grupos de hombres y mujeres", y que por lo tanto la hipótesis alterna sería que "no existe homogeneidad" ya que "debían ser complementarias". También, indicó que después estableció un nivel de significancia de 0.05 y que calculó los grados de libertad y el estadístico Chi-cuadrada de acuerdo con las formulas; pero que para calcular el estadístico primero tuvo que obtener la frecuencia esperada, ya que en el problema únicamente se encontraba la frecuencia observada. Finalmente, añadió que para resolver el problema comparó el valor que le dan las tablas de probabilidad de la distribución Chi-cuadrada cuando se tiene un alfa de 0.05 y tres grados de libertad con el valor del estadístico Chi-cuadrada que había calculado y que como el valor del estadístico que obtuvo de las tablas era más grande que el calculado no podía rechazar la hipótesis nula.

En esta práctica del PF14, identificamos el uso de *elementos lingüísticos* como el lenguaje natural y simbólico; *conceptos/definiciones* tales como muestra, variables, categorías, frecuencias observadas, frecuencias esperadas e hipótesis; *propiedades/proposiciones* como el estadístico Chi-cuadrada, frecuencia esperada, grados de libertad y criterio para la toma de decisión. Los *procedimientos* que realizó el PF14 consisten en plantear las hipótesis y el nivel de significancia, calcular las frecuencias esperadas, el estadístico y los grados de libertad, obtener el valor del estadístico teórico a partir de las tablas de probabilidad de la distribución Chi-cuadrada como valor crítico; así como comparar este valor crítico con el estadístico calculado y decidir si se rechaza o no la hipótesis nula. Los *argumentos* se basaron en el criterio del valor crítico para la toma de decisión, y como el valor crítico es menor que el valor del estadístico calculado, entonces no se rechaza la hipótesis nula. Los objetos primarios que se identificaron en la práctica del PF14 son del *Nivel 4*. Para fortalecer su razonamiento inferencial formal se le sugiere al PF14, así como en la práctica del PF5 (Figura 6.12), trabajar con el estadístico Chi-cuadrada con factor de corrección de continuidad y con los errores tipo I y tipo II, y validar su inferencia haciendo uso de la potencia de la prueba.

En la Figura 6.21 presentamos la práctica del profesor en formación 7, misma que se destacó en el segundo grupo de profesores en formación.

	Caucásico	Afroamericano	Asiático	Hispano	Total
Hombres	95	12	3	8	118
Mujeres	92	10	3	6	111
Total	187	22	6	14	229

Hipótesis	
H_0 =	Existe homogeneidad entre hombres y mujeres
H_1=	No existe homogeneidad entre hombres y mujeres

Se establece P	Calculo del Chi Cuadrado	Distribución chi cuadrado
0,05	0,301969166	0,959657709

Grado de Libertad
3

Calculo de las Frecuencias esperadas					
	Caucásico	\froamerican	Asiático	Hispano	Total
Hombres	96,3580786	11,3362445	3,09170306	7,2139738	118
Mujeres	90,6419214	10,6637555	2,90829694	6,7860262	111
Total	187	22	6	14	229

Figura 6.21. Práctica del profesor en formación 7 (G2) sobre la actividad 3

El Profesor en formación siete (PF7), comentó que decidió realizar una prueba Chi-cuadrada y que inició por plantear las hipótesis, establecer P y después realizar el cálculo de las frecuencias esperadas que necesitaba para calcular el estadístico Chi-cuadrada y que también calculó los grados de libertad y la probabilidad en la distribución Chi-cuadrada para el estadístico que había calculado mediante el uso de Excel. Finalmente, añadió que no existe homogeneidad entre hombres y mujeres ya que, según la información dada, se rechaza la hipótesis nula, ya que la distribución del Chi-cuadrada denota una probabilidad mayor a la significancia propuesta; además que los datos de la muestra reflejan más datos de un tipo de raza que de otra, por lo cual no se tiene una distribución más uniforme sobre la cantidad de personas por raza para lograr el cometido de un mejor tratamiento sobre cáncer de pulmón.

Como podemos observar, la práctica que desarrolla el PF7 es similar a la que desarrolló el PF14; no obstante, el PF7 parece comprender erróneamente el criterio de decisión del valor-p, pues cuando el profesor realiza la comparación del nivel de significancia con el valor-p, decide rechazar la hipótesis nula cuando el criterio indica que si el valor-p es mayor al nivel de significancia no se rechaza la hipótesis nula. La propuesta de niveles de razonamiento inferencial promueve la comprensión progresiva del criterio de decisión; en el *Nivel 2* utilizamos la probabilidad para tomar la decisión, mientras que el *Nivel 3* comparamos la probabilidad obtenida con un límite significativo que se estableció previo a la prueba, y en el *Nivel 4* tenemos el criterio del valor crítico o del valor-p.

6.2.4 Reflexiones y consideraciones sobre el estadístico Chi-cuadrada

Sobre las prácticas que desarrollaron los profesores en formación y en ejercicio podríamos decir que, en las asociadas al *Nivel 1* se destacó principalmente el uso de representaciones gráficas (e.g., barras dobles, líneas, barras apiladas y pastel), algunas medidas estadísticas (por ejemplo, la media y desviación estándar) y el uso de la probabilidad condicional. Cabe destacar que, aunque los profesores en formación del grupo 1 no habían recibido una instrucción sobre inferencia, fueron capaces de realizar conjeturas basadas en la visualización de las características de los gráficos y en las mediciones de probabilidad, mostrando así un razonamiento inferencial informal. Por su parte, en las prácticas asociadas al *Nivel 2* se enfatizó el uso del estadístico Chi-cuadrada y el reconocimiento de hipótesis; mientras que en las prácticas asociadas al *Nivel 4* encontramos el planteamiento de hipótesis, el estadístico y la distribución Chi-cuadrada, la regla del valor crítico para la toma de decisión, nivel de significancia, entre otros. Es preciso señalar, que no hubo evidencia de prácticas que utilizaran en profundidad elementos del *Nivel 3*; sin embargo, en las prácticas de los profesores encontramos rasgos del uso de elementos de este nivel, por ejemplo, el PF17 usó la probabilidad como límite significativo.

Además, destacamos que, en las prácticas de los profesores, no fue posible observar la aplicación del estadístico Chi-cuadrada con el factor de corrección de continuidad, esto también sucede en otros estudios que proponen actividades para trabajar con los estudiantes las pruebas Chi-cuadrada (e.g., Woodard, Lee y Woodard, 2020; Fellers y Kuiper, 2020; DePaolo, Robinson y Jacobs, 2016). Por ejemplo, Gibbs y Goossens (2013) hablan sobre la pertinencia del factor de corrección de continuidad y que es muy usual que no se utilice cuando se trabaja con softwares, porque algunos (e.g., Minitab) no tienen la opción de utilizar este factor. Entonces, es necesaria la comprensión no solo de esta propiedad/proposición, sino de lo que se encuentra de trasfondo de los procesos o procedimientos que realizan algunos softwares en el análisis de datos.

A partir de las prácticas puestas en juego por los profesores participantes en nuestro estudio, fue posible analizar y discutir cómo éstas podrían ser más robustas dentro de un mismo nivel o transitar a un nivel siguiente; es decir, se evidenció cómo los niveles de razonamiento inferencial para el estadístico Chi-cuadrada podrían ayudar con algunos errores frecuentes y dificultades que observamos en las prácticas de los profesores.

En general, a partir de los objetos matemáticos primarios identificados en las prácticas desarrolladas por los profesores participantes, fue posible constatar que éstos se relacionan con los elementos propuestos en los distintos niveles de razonamiento inferencial sobre la Chi-cuadrada y, con base en lo discutido en este capítulo, consideramos que los niveles de razonamiento inferencial son buenos predictores del razonamiento inferencial de profesores y estudiantes sobre este estadístico. Aunque las distribuciones de probabilidad condicional no formaban parte de la propuesta inicial de niveles, este análisis empírico nos permitió percatarnos de la importancia que tiene la probabilidad condicional en las prácticas que evidencian un razonamiento inferencial informal (*Nivel 1*), por lo cual, los criterios sobre probabilidad condicional se describen a continuación, y complementan lo ya planteado en la propuesta de niveles para la Chi-cuadrada (ver sección 5.1).

Subnivel 1.1: Visualización

Además, puede utilizar gráficos (e.g., barras dobles, barras apiladas y mosaico) para conjeturar y argumentar si existe asociación entre dos variables. La Figura 6.22 muestra algunos tipos de gráficos que se les pueden presentar a los estudiantes.

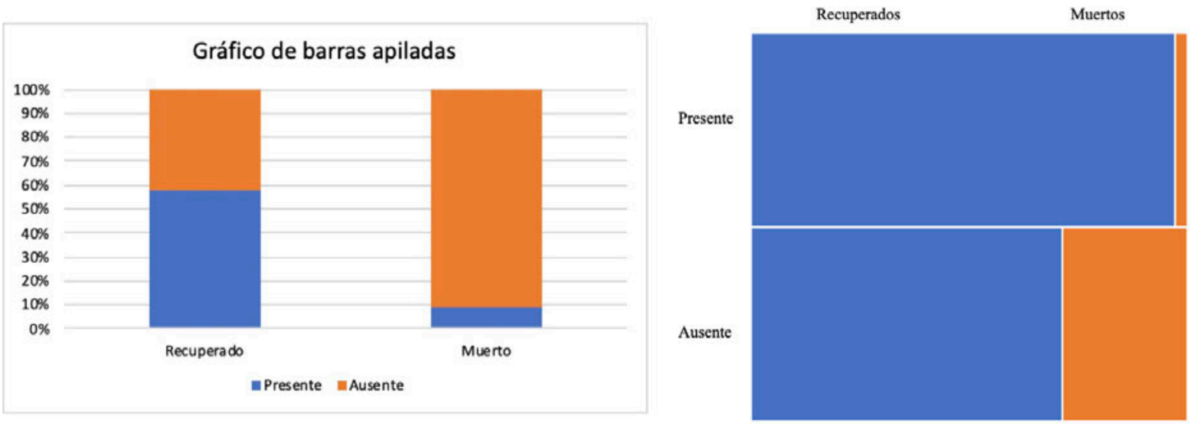


Figura 6.22. Gráficos de barras apiladas y de mosaico para visualizar la asociación (Fuente: Elaboración propia)

Subnivel 1.3: Asociación intuitiva

El estudiante puede utilizar métodos simples como la probabilidad condicional para establecer si existe asociación entre dos variables, tal como lo proponen Batanero y Borovcnik (2016), para lo cual el estudiante:

- Comprende conceptos básicos de probabilidad y sus propiedades –e.g., variable (CD1), espacio de probabilidad, evento, intersección, distribución marginal y teorema de Bayes–.
- Es capaz de calcular la distribución condicional por renglón (porcentajes por renglón), la distribución condicional por columna (porcentajes por columna) y la distribución conjunta (porcentajes de toda la muestra). Además, a partir de las distribuciones condicionales y conjunta puede elaborar gráficos como los mostrados en la Figura 6.22.
- Identifica los eventos y es capaz de calcular la probabilidad del evento y la probabilidad condicional. También, puede comparar las probabilidades condicionales, si existen diferencias, éstas pueden reflejar que los datos proveen evidencia empírica de la existencia de asociación entre las variables.
- Discrimina entre una relación de asociación y una relación de causa-efecto. Diversos investigadores han planteado que dicha discriminación es un aspecto clave para comprender la asociación, independencia y dependencia, en el sentido estadístico (e.g., Batanero, et al., 1996; Wijayatunga, 2016).

6.3 Análisis de las prácticas de los profesores sobre el estadístico t-Student

A continuación, analizamos las prácticas desarrolladas por los profesores en ejercicio y en formación, a propósito de la solución de las actividades 4, 5, 6 y 7. Destacamos que para este análisis utilizamos la propuesta de niveles de razonamiento inferencial sobre la t-Student (Capítulo 5) y la noción de práctica matemática y configuración de objetos y procesos, descrita en la sección de marco teórico.

Las figuras que presentamos sobre las prácticas que realizaron los profesores, al igual que en la sección 6.2, son capturas de pantalla de los videos, de las sesiones sincrónicas, y de los documentos (Word, Excel y PDF) que los profesores subieron a las plataformas como evidencia del desarrollo de las actividades sobre el estadístico t-Student.

6.3.1 Prácticas asociadas a la actividad 4

La actividad 4 trata de un experimento con un fármaco para dormir, donde interesa conocer si realmente existe un aumento en las horas de sueño cuando se utiliza el fármaco. A continuación, presentamos la actividad 4 y las prácticas que desarrollaron los profesores en ejercicio sobre esta actividad.

Actividad 4.

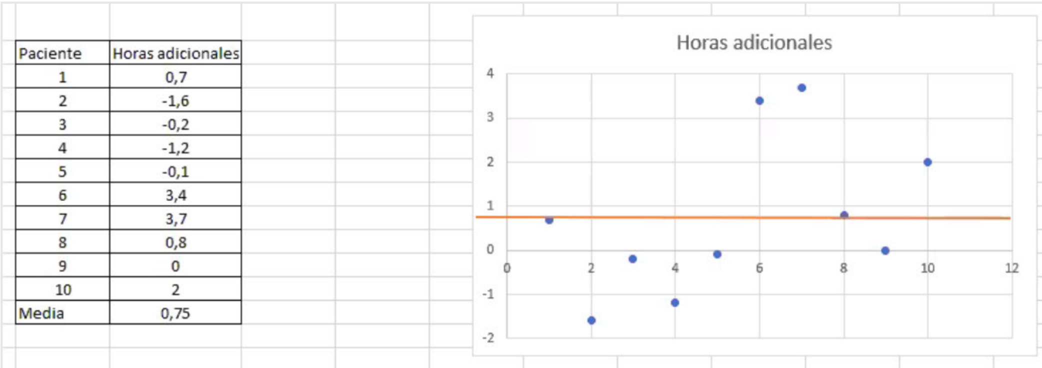
A continuación, se presentan los datos de una muestra que fueron recabados durante un experimento de un fármaco para dormir. Los datos de la tabla son sobre el aumento de horas de sueño cuando los pacientes han ingerido el fármaco. ¿Existe realmente un aumento en las horas de sueño al utilizar el fármaco? Explica con detalle tu respuesta

Tabla 6.5. *Horas adicionales de sueño ganadas por el uso del fármaco*

Paciente	Horas adicionales
1	$+ 0.7$
2	$- 1.6$
3	$- 0.2$
4	$- 1.2$
5	$- 0.1$
6	$+ 3.4$

7	+ 3.7
8	+ 0.8
9	0.0
10	+ 2.0

El primer tipo de práctica característica para la actividad 4 se ejemplifica con los desarrollos del equipo 4 (Figura 6.23).



Paciente	Horas adicionales
1	0,7
2	-1,6
3	-0,2
4	-1,2
5	-0,1
6	3,4
7	3,7
8	0,8
9	0
10	2
Media	0,75

Figura 6.23. Práctica del equipo 4 de profesores en ejercicio (G3) sobre la actividad 4

Los profesores de este equipo señalaron que graficaron los datos sobre el incremento de horas de sueño que se proporcionaban en el problema para observar cuan dispersos estaban los datos entre sí, también calcularon la media y la dibujaron en el mismo gráfico; ellos indicaron que a partir de lo anterior, podían ver qué tan dispersos están los datos en relación con la media. Además, visualizaron en el gráfico que sólo hay cuatro personas con incremento de hora de sueño por encima de la media, aunque una de ellas se encuentra casi en el valor de la media; a partir de esto, señalaron que en realidad sólo hay tres personas en la muestra a quienes les hace efecto de forma significativa el fármaco (por lo alejados que se encuentran de la media) y que entonces no existe un aumento en las horas de sueño con este fármaco.

En la práctica que realizaron los profesores podemos observar *conceptos/definiciones* como los de dispersión y significancia (aunque con una acepción más intuitiva), *propiedades/proposiciones* como la media; *procedimientos* tales como el cálculo de la media y elaboración del gráfico de puntos. Los *argumentos* estuvieron basados en el número de personas que ellos consideraron que tuvieron un aumento significativo. También observamos el papel central que tuvo la representación gráfica para que los profesores visualizaran los datos y tomaran una decisión. La mayoría de estos elementos pertenecen al *Nivel 1*, sin embargo, nos llama la atención el uso que dan a la significancia; para ellos “es significativo si se encuentra muy alejado de la media", pero cuando se les preguntó ¿qué tan alejado?, ellos indicaron que no lo cuantificaron, "simplemente que se pueda visualizar lejos de la media". Con respecto al concepto/definición significancia, como un límite significativo a partir de cierto valor de probabilidad, en los niveles de razonamiento inferencial sobre el estadístico t-Student se encuentra en el *Nivel 3*, aunque en las prácticas de los profesores no fue posible observar esta definición de significancia.

Podemos observar que la práctica que desarrolla el equipo 4 guarda relación con lo que se espera en la primera tarea para promover el RII del marco propuesto por Zieffler et al. (2008), pues los profesores realizaron conjeturas basados en las características del gráfico. Para una práctica con elementos más complejos del *Nivel 1*, los profesores podrían haber calculado, además de la media, otras medidas estadísticas como los cuartiles. Los cuartiles les podrían ayudar a analizar la variación interna del conjunto de datos, tal y como lo propuso Galton (1875), y basar su conclusión tanto en la visualización del gráfico como en la dispersión de los intervalos que ha calculado. Los aspectos anteriormente mencionados se abordan en el subnivel 1.2 de los niveles de razonamiento inferencial para el estadístico t-Student.

Una segunda práctica característica para esta actividad se ejemplifica con los desarrollos del equipo 1. Los profesores de este equipo se auxiliaron de Excel para resolver el problema, con el cual utilizaron algunas medidas estadísticas, que se pueden observar a la izquierda y al centro de la Figura 6.24, y un gráfico para representar el incremento de las horas de sueño de cada uno de los participantes.

Paciente	horas sueño
1	0,7
2	-1,6
3	-0,2
4	-1,2
5	-0,1
6	3,4
7	3,7
8	0,8
9	0
10	2
media	0,75
	1,78900966

Media	0,75
Error típico	0,565734527
Mediana	0,35
Moda	#N/D
Desviación estándar	1,789009658
Varianza de la muestra	3,200555556
Curtosis	-0,62982237
Coeficiente de asimetría	0,580920674
Rango	5,3
Mínimo	-1,6
Máximo	3,7
Suma	7,5
Cuenta	10
Nivel de confianza(95,0%)	1,279780414

Figura 6.24. Práctica del equipo 1 de profesores en ejercicio (G3) sobre la actividad 4

Los profesores en ejercicio del equipo 1, indicaron que lo primero en que pensaron para resolver el problema fue obtener la media y la desviación estándar, y después en realizar un gráfico para poder visualizar la dispersión. Posteriormente, también realizaron un análisis de los datos con diversas medidas enfocadas a estudiar la dispersión, de las cuales destacaron la curtosis; a partir de este análisis, argumentaron que “como lo observamos con la media, desviación y el gráfico, hay una gran dispersión de los datos y debido a dicha dispersión no existe realmente un aumento en las horas de sueño”. Podemos observar que los profesores aluden principalmente al *concepto* de dispersión, y *propiedades/proposiciones* tales como la media, desviación estándar, error típico o error estándar, varianza, curtosis, entre otros. Los *procedimientos* realizados atienden a las propiedades indicadas y otras, que se pueden ver en la figura 6.24, con el uso de Excel. En esta práctica logramos observar que, a pesar de obtener diversas medidas estadísticas, los profesores *argumentaron* sus conclusiones con la media, desviación estándar y la dispersión que lograron visualizar por medio del gráfico. Estos objetos matemáticos primarios y los procesos que se encuentran asociados a ellos pertenecen al *Nivel 1* de razonamiento inferencial sobre el estadístico t-Student.

Un aspecto a destacar en esta práctica, es que los profesores utilizan la fórmula de Excel para obtener un intervalo de confianza para la t-Student y lo etiquetan como “Nivel de Confianza (95,0%)”, pero no forman el intervalo. Lo anterior evidencia que identifican que se trata de una prueba t-Student pero no son capaces de realizar los procedimientos para resolver el problema por pruebas de hipótesis o por intervalos de confianza. Este aspecto es el único indicador de práctica del *Nivel 2* que observamos.

En esta práctica, a diferencia de la práctica desarrollada por el equipo 4, podemos observar una implementación más completa de los elementos que se consideran en el *Nivel 1*, pues su conjetura se basa tanto en la visualización del gráfico como en las diversas medidas estadísticas que han calculado para analizar la dispersión de los datos. Como hemos señalado, este equipo parece percatarse de que se puede realizar una prueba t-Student, no obstante, no la llevan a cabo. Para que la práctica de estos profesores progrese a una de *Nivel 2*, una vez que ellos comprendan las características de las diversas pruebas con el estadístico t-Student, podrían identificar la hipótesis que se encuentra implícita en el problema. La hipótesis es el primer componente del ciclo de investigación estadística (PPDAC) de Wild y Pfannkuch (1999), además diversos estudios han reportado las bondades de trabajar inicialmente con hipótesis a través de preguntas o conjeturas (e.g., Pfannkuch y Wild, 2004; Pfannkuch, et al., 2016; Bakker, Ben-Zvi y Makar, 2017). También podrían apoyarse en recursos tecnológicos para realizar actividades con simulaciones, que permitan promover la comprensión de la distribución t-Student, este tipo de actividades se han documentado en investigaciones como las de Bakker y Gravemeijer (2004), Reading y Reid (2006) y Rossman (2008). Además, de acuerdo con Konold et al. (2011), es necesario que primero se fomente un razonamiento probabilístico básico para que los estudiantes o profesores puedan realizar inferencias informales, en este caso preformales. En este sentido, en el *Nivel 2* de nuestra propuesta se propone utilizar la probabilidad como una medida de certidumbre asociada a un evento, antes de compararla con un límite preestablecido (*Nivel 3*) o un nivel de significancia (*Nivel 4*).

Un tercer tipo de práctica utilizada para resolver la actividad, se ejemplifica con la desarrollada por el equipo 6 de profesores en ejercicio del grupo tres. Los profesores establecieron ciertas estrategias las cuales enumeran en la parte izquierda de la Figura 6.25.

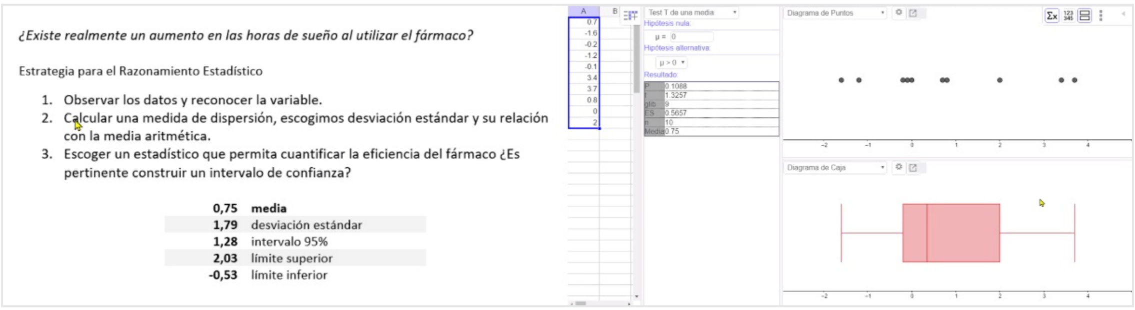


Figura 6.25. Práctica del equipo 6 de profesores en ejercicio (G3) sobre la actividad 4

Con respecto a la estrategia 1, los profesores detallaron que al observar los datos se percataron que se trataba de una variable continua que adquiere valores tanto positivos como negativos y que, por lo tanto, fue necesario definir qué significa el cero para esta variable; además, definir la variable cobra relevancia en el contexto del problema porque indica que no hay un aumento en las horas de sueño con el fármaco. Los profesores explicaron que realizaron un gráfico de puntos con papel y lápiz para observar qué tan lejos o cerca se encontraban los valores del cero, gráfico que posteriormente obtuvieron con el uso de Geogebra (Figura 6.25). También señalaron que podían utilizar un boxplot para observar cuánto se extienden los datos. Con la segunda estrategia buscaban observar la dispersión de los datos, para lo cual decidieron explorar la relación de la desviación estándar con la media aritmética, auxiliándose de Excel para realizar los cálculos. Los profesores indicaron que la parte que les generó más dificultad y discusión entre los miembros del equipo, fue la estrategia tres, pues debían escoger un estadístico que permitiera cuantificar la eficiencia del fármaco y pensaban que debía ser uno que utilizara la relación de la desviación estándar y la media aritmética; finalmente, indicaron que decidieron optar por el estadístico t-Student y aplicar una prueba t para una muestra (ver parte derecha de la figura 6.25) y construyeron un intervalo de confianza del 95% (parte de la izquierda de la figura 6.25). Una vez que obtuvieron los límites inferior y superior del intervalo, los profesores destacaron la importancia de haber reconocido a la variable y definido al cero como la no eficacia o no aumento de horas de sueño, e indicaron que "como el cero se encuentra dentro del intervalo no se tiene un verdadero aumento en las horas de sueño".

En la primera parte de la práctica (estrategia uno y dos) que realizaron los profesores podemos identificar su interés por observar el comportamiento de los datos y la visualización por medio de gráficos para realizar una conjetura, aspectos que son característicos del *Nivel 1*. Algunos de los objetos matemáticos primarios que se identificaron en la práctica son *conceptos/definiciones* como variable y dispersión; *propiedades/proposiciones* tales como media y desviación estándar; los *procedimientos* sobre la media, desviación estándar y los límites del intervalo (los cuales fueron realizados con el apoyo de Excel). Cabe destacar que estos objetos matemáticos primarios que caracterizan la práctica son propios del *Nivel 1* de razonamiento inferencial.

En la segunda parte de la práctica (estrategia tres) se identificaron elementos que son representativos del *Nivel 4*. Por ejemplo, *conceptos/definiciones* como nivel de confianza, muestra y probabilidad, *propiedades/proposiciones* tales como el estadístico t-Student, grados de libertad, error estándar, hipótesis nula y alternativa, límites inferior y superior y la regla de decisión; los *procedimientos* para llevar a cabo la prueba de hipótesis t-Student para una muestra los realizaron con el recurso tecnológico Geogebra y para formar el intervalo de confianza utilizaron las funciones de Excel. Por su parte, los *argumentos* estuvieron basados en que el cero se encuentra dentro del intervalo de confianza y por lo tanto "no se tendría un verdadero aumento en las horas de sueño".

Es importante destacar que, a diferencia de las prácticas anteriores, en la desarrollada por el equipo 6 los profesores se preocuparon por comprender el tipo de datos con los que estaban trabajando, por ejemplo, el tipo de variable. Además, dieron un significado al cero, aspecto que fue clave para la prueba de hipótesis, la formación del intervalo y para desarrollar el argumento sobre su inferencia. Sin embargo, cuando se les preguntó por qué optaron por utilizar un nivel de confianza del 95%, los profesores indicaron que era "el porcentaje que siempre utilizan", lo cual pone de manifiesto la falta de comprensión sobre algunas nociones que utilizaron. Es decir, los profesores pueden operar con ellas algorítmicamente, pero les falta comprender con mayor profundidad algunas nociones tales como el nivel de confianza y otras que se encuentran relacionadas con dicha noción (e.g., nivel de significancia y los tipos de error). Aunado a comprender las nociones anteriormente mencionadas, los profesores podrían trabajar con la potencia de la prueba para una movilización integral del *Nivel 4*, esto daría cuenta de un razonamiento inferencial formal. De acuerdo con Cohen (1992), podemos comprender a la potencia de la prueba como un índice de la validez de los resultados estadísticos; también es importante considerar a la prueba estadística (en este caso la t-Student), el nivel de significancia, el tamaño del efecto y el tamaño de la muestra como los factores que determinan a la potencia de la prueba (Lipsey y Aiken, 1990).

Con respecto a los profesores en formación, debemos señalar que las prácticas que desarrollaron para resolver la actividad 4 fueron similares a las que acabamos de analizar. No obstante, es de interés analizar algunas prácticas que se destacaron, como la que se muestra en la Figura 6.26.

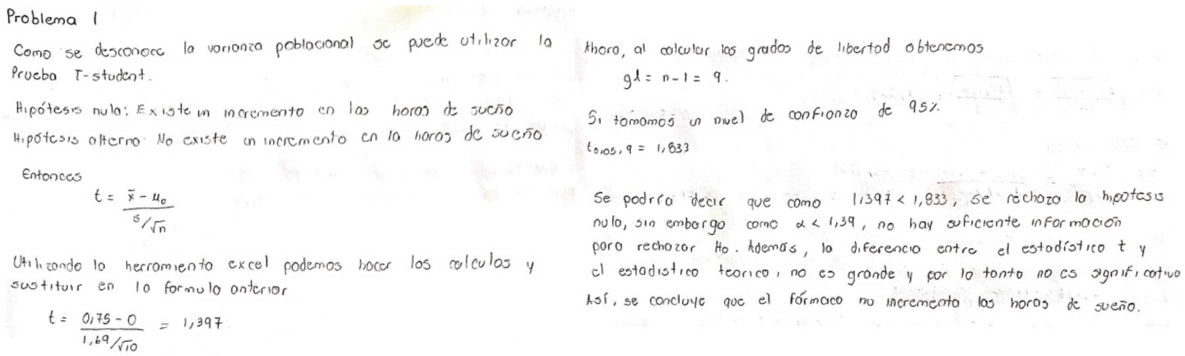


Figura 6.26. Práctica del profesor en formación 8 (G2) sobre la actividad 4

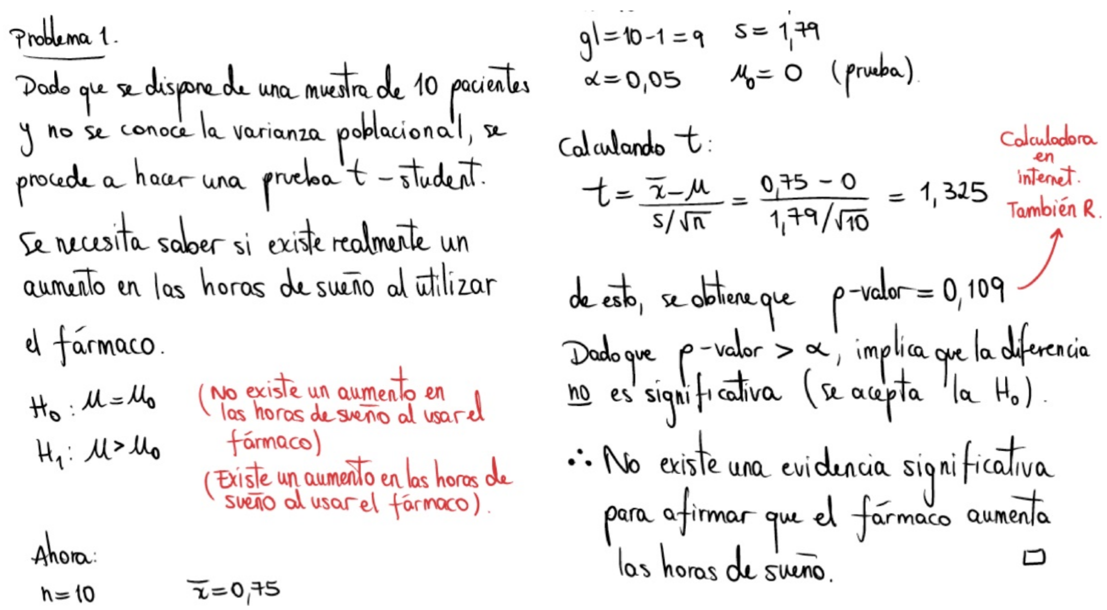
El profesor en formación etiquetado con el número 8 (PF8), señaló que decidió aplicar una prueba t-Student debido a que con la información proporcionada se desconoce la varianza poblacional; también planteó las hipótesis nula y alternativa en lenguaje natural, y aplicó la prueba, para lo cual decidió utilizar un nivel de confianza de 95%.

En esta práctica podemos identificar el uso de *conceptos/definiciones* como varianza poblacional, media poblacional, población, muestra, hipótesis y nivel de confianza; también *propiedades/proposiciones* tales como la media de la muestra, desviación estándar, tamaño muestral, estadístico t-Student, grados de libertad, estadístico teórico y las reglas para la toma de decisión. Los *procedimientos* realizados por el futuro profesor consisten en el planteamiento de las hipótesis, el nivel de significancia, el cálculo del estadístico t-Student, de los grados de libertad, obtener el valor del estadístico teórico y del valor de la probabilidad para el estadístico de prueba y, finalmente, decidir si se rechaza la hipótesis nula. Los *argumentos* que realizó estuvieron basados en los criterios para la toma de decisión; es decir, como el valor del estadístico de prueba es menor que el del estadístico teórico, entonces no se rechaza la hipótesis nula. Los *elementos lingüísticos* que utiliza son lenguaje natural y simbólico.

En la práctica que realizó PF8, se encuentran presentes elementos o criterios característicos del *Nivel 3*, como el realizar el planteamiento de las hipótesis en lenguaje natural, y cuatro, como el uso del nivel de significancia.

Queremos resaltar el siguiente argumento que desarrolla el profesor, el cual se puede observar en la Figura 6.26, "la diferencia entre el estadístico t y el estadístico teórico, no es grande y por lo tanto no es significativo", pues aunque el profesor ha realizado adecuadamente los procedimientos, este argumento deja de manifiesto que tal vez algunas de las propiedades/proposiciones o conceptos/definiciones que está utilizando (e.g., nivel de significancia y nivel de confianza) necesitan ser comprendidas con mayor profundidad y de manera integral. Podría ayudar a dicha comprensión sí estableciera las regiones de Rechazo y No Rechazo, y se formulara preguntas como ¿qué significa el valor crítico que he establecido a partir del estadístico teórico?, ¿cómo en estas regiones están presentes los niveles de significancia y de confianza, y la relación que existe entre ellos?, y ¿cómo podría interpretar el nivel de confianza en términos de la hipótesis nula? Reflexionar sobre estos aspectos le permitiría al PF8 que su práctica transite al *Nivel 4*.

También podemos observar en la práctica del PF8, que comete un error cuando va a trabajar con el criterio de decisión del valor-p, por consiguiente, podría iniciar la comprensión del valor-p con una aproximación informal, auxiliándose de software para simulaciones o para el cálculo de la probabilidad del estadístico t-Student, y con ello apoyar su inferencia (Rossman, 2008).



Problema 1.

Dado que se dispone de una muestra de 10 pacientes y no se conoce la varianza poblacional, se procede a hacer una prueba t-student.

Se necesita saber si existe realmente un aumento en las horas de sueño al utilizar el fármaco.

$H_0: \mu = \mu_0$ (No existe un aumento en las horas de sueño al usar el fármaco)

$H_1: \mu > \mu_0$ (Existe un aumento en las horas de sueño al usar el fármaco).

Ahora:

$n = 10$ $\quad \bar{x} = 0{,}75$

$gl = 10 - 1 = 9$ $\quad s = 1{,}79$

$\alpha = 0{,}05$ $\quad \mu_0 = 0$ (prueba)

Calculando t:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{0{,}75 - 0}{1{,}79/\sqrt{10}} = 1{,}325$$

Calculadora en internet. También R.

de esto, se obtiene que p-valor $= 0{,}109$

Dado que p-valor $> \alpha$, implica que la diferencia no es significativa (se acepta la H_0).

$\therefore$ No existe una evidencia significativa para afirmar que el fármaco aumenta las horas de sueño. $\square$

Figura 6.27. Práctica del profesor en formación 2 (G2) sobre la actividad 4

En la Figura 6.27 podemos observar la práctica que realizó el profesor en formación 2 (PF2) respecto a la actividad 4. Inicialmente, el profesor indicó que debido a que se trata de una muestra de 10 pacientes (muestra pequeña) y se desconoce la varianza poblacional se puede realizar una prueba t-Student; además, estableció las hipótesis nula y alternativa con lenguaje simbólico y las explicó en lenguaje natural, calculó el estadístico t-Student, obtuvo el p-valor y concluyó en términos del problema. En esta práctica podemos identificar que los *elementos lingüísticos* que utilizó son lenguaje natural y lenguaje simbólico, algunos *conceptos/definiciones* son población, varianza poblacional, media poblacional, muestra e hipótesis, mientras que las *propiedades/proposiciones* son la media de la muestra, desviación estándar, tamaño muestral, estadístico t-Student, grados de libertad, alfa y la regla para la toma de decisión. Los *procedimientos* que realizó fueron calcular el estadístico t-Student, obtener el p-valor y compararlo con el valor de alfa para tomar la decisión de rechazar o no la hipótesis nula. Finalmente, *argumenta* su conclusión indicando que como el p-valor es mayor que el alfa, la diferencia no es significativa. En síntesis, la práctica que desarrolla es característica del *Nivel 4*.

Para que el PF2 desarrolle una práctica de *Nivel 4* más robusta, podría trabajar con los errores tipo I y tipo II y comprender las relaciones entre dichos errores. Para promover dicha comprensión puede auxiliarse de software y representaciones gráficas de forma similar a lo propuesto por Rossman (2008) para la distribución o el valor-p. Además, el PF2 podría validar su inferencia haciendo uso de la potencia de la prueba, tal y como sugerimos al PF8.

En este sentido, para aproximarse a los errores tipo I y tipo II, y la potencia de la prueba, Vera y Díaz (2013) proponen primero explorar con profundidad la idea de probabilidad condicional mediante applets, pues esto les permitirá a los profesores experimentar diferentes posibilidades de forma visual.

Los recursos que, en general, utilizaron los profesores en formación fueron Excel, R, Geogebra, calculadoras en línea, lápiz y papel para cálculos manuales, y tablas de probabilidad.

6.3.2 Prácticas asociadas a la actividad 5

Como se puede observar a continuación, la actividad 5 trata sobre un experimento con dos fármacos para dormir, y se desea conocer si existe diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2.

Actividad 5.

Una farmacéutica está realizando experimentos sobre la efectividad de dos fármacos para dormir, para lo cual ha seleccionado aleatoriamente a diez pacientes quienes han tomado ambos fármacos. La Tabla 6.6 muestra el aumento de horas de sueño cuando los pacientes han ingerido cada uno de los fármacos. ¿Existe realmente una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2?

Tabla 6.6. *Horas ganadas de sueño por participante con ambos fármacos*

Paciente	Fármaco 1	Fármaco 2
1	$+0.7$	$+1.9$
2	-1.6	$+1.8$
3	-0.2	$+1.1$
4	-1.2	$+0.1$
5	-0.1	-0.1
6	$+3.4$	$+4.4$
7	$+3.7$	$+5.5$
8	$+0.8$	$+1.6$
9	0.0	$+4.6$

10	+ 2.0	+ 3.4

A continuación, presentamos las prácticas referentes a la actividad 5 y en la Figura 6.28 se encuentra el primer tipo de práctica que desarrollaron los profesores sobre esta actividad.

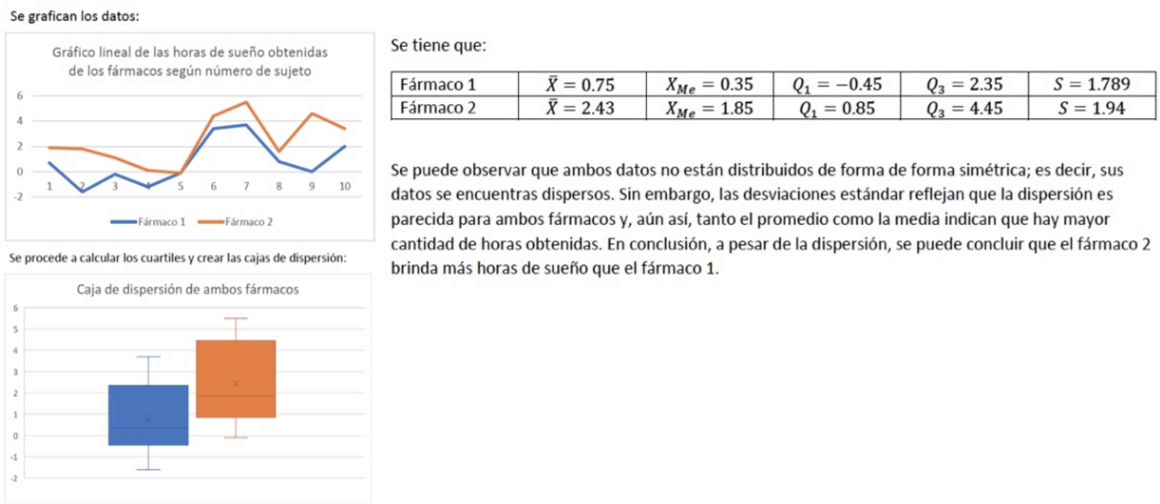


Figura 6.28. Práctica del profesor en formación 12 (G2) sobre la actividad 5

Como se puede apreciar en la Figura 6.28, el profesor en formación doce (PF12) primeramente realizó un gráfico de líneas sobre las horas de sueño ganadas por cada paciente; después realizó boxplots para observar la dispersión en ambas muestras y finalmente concluyó que el fármaco dos brinda más horas de sueño que el fármaco uno, basándose en la media, los cuartiles y la visualización en los gráficos.

En la práctica del PF12 podemos identificar objetos matemáticos primarios, dentro de los que distinguimos los *elementos lingüísticos* representaciones gráficas, lenguaje natural y simbólico; algunos *conceptos/definiciones* como muestra, simetría y dispersión; mientras que las *propiedades/proposiciones* son media, desviación estándar, mediana y cuartiles uno y tres. Los *procedimientos* que realizó fueron el gráfico de líneas basado en la información proporcionada en la tabla de la actividad y posteriormente el cálculo de los cuartiles para construir dos boxplots con el uso de Excel, también calculó la media y desviación estándar. Para *argumentar* sobre la conclusión que realiza, recurre principalmente a la dispersión, media y mediana. Todos los objetos primarios que hemos identificado en la práctica matemática del PF12 corresponden al *Nivel 1* de razonamiento inferencial sobre el estadístico t-Student.

Además de la práctica desarrollada por el PF12 para esta actividad, el profesor podría analizar la variación interna de ambas muestras mediante el método de fluctuación (Edgeworth, 1885), que se entiende como una medida de la variación de una serie de datos con respecto a otra serie. De acuerdo con Stigler (2017), la idea de la comparación en términos de la variación interna de los datos fue retomada por Gosset y Fisher para proponer nuevos métodos, es por ello que en nuestra propuesta de niveles de razonamiento inferencial se considera este método en el *Nivel 1*, como una versión intuitiva, de este tipo de prueba, que pueda apoyar las conjeturas o inferencias informales.

En la Figura 6.29 podemos observar la práctica del profesor en formación nueve (PF9), del grupo dos, la cual es característica, principalmente, del *Nivel 2*.

Note que este problema puede abarcarse por medio de una prueba de hipótesis para dos promedios, asumiendo que los datos siguen una distribución normal, además, no se cuenta con una desviación estándar poblacional, por lo cual se utiliza la distribución T como base para resolver el ejercicio. Tome como hipótesis nula e hipótesis alternativa los siguientes hechos:

H_0: no existe una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2.

H_1: existe una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2.

Además, de los datos dados en el ejercicio, se obtiene la siguiente información:

$$n_1 = 10$$
$$n_2 = 10$$
$$\bar{x}_1 = 0.75$$
$$\bar{x}_2 = 2.43$$
$$s_1 = 1.78901$$
$$s_2 = 1.941391$$
$$S_p^2 = \frac{(10-1)(1.78901)^2 + (10-1)(1.941391)^2}{10+10-2} = 3.484777897$$

Dado que por lo definido en H_0 no existe una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2, con los valores obtenidos y asumiendo que σ_1 y σ_2 son iguales, se tiene que:

$$Valor\ P = P\left(T \leq \frac{0.75 - 2.43}{\sqrt{\frac{3.484777897}{10} + \frac{3.484777897}{10}}}\right)$$
$$= P(T \leq -2.012365152)$$

También, note que los grados de libertad en este problema vienen dados por $v = 10 + 10 - 2$, es decir $v = 18$, así:

$$Valor\ P = P(T \geq 2.01)$$
$$= \in\ ]0.025, 0.040[$$

Como el $Valor\ P < 0.04 < 0.05$ entonces se rechaza H_0, es decir, existe una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2.

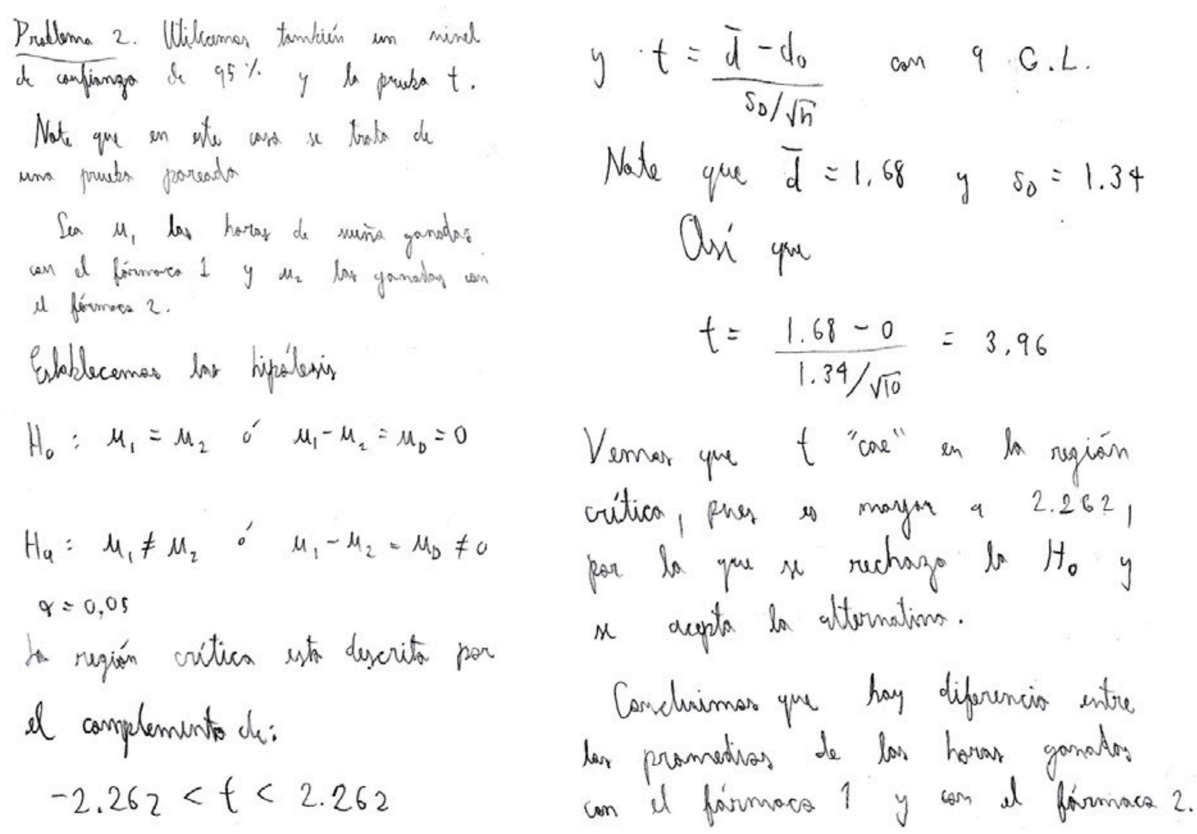
Figura 6.29. Práctica del profesor en formación 9 (G2) sobre la actividad 5

El PF9 declaró en su práctica que el problema puede ser abordado mediante una prueba de hipótesis para dos medias (promedios) y que se puede utilizar la distribución t-Student, debido a que asumió que los datos siguen una distribución normal e identificó que se desconoce la desviación estándar de la población. Planteó las hipótesis en lenguaje natural, aunque utilizó la simbología para indicar cuál es la nula y alternativa. También identificó el tamaño de las muestras y obtuvo la media y desviación estándar de cada muestra. A partir de las desviaciones estándar, calculó la varianza agrupada, utilizando esta propiedad/proposición para cuando se trabaja con dos muestras independientes y con desigualdad del tamaño muestral. Además, indicó que asumió que las desviaciones poblacionales son iguales, y para realizar la prueba utilizó el estadístico t-Student y el valor-p como la probabilidad a partir de la distribución t, para lo cual recurrió a los grados de libertad; asumió un nivel de significancia e identificó que se trata de una prueba con dos colas iguales.

En la práctica que realizó el PF9 identificamos *conceptos/definiciones* tales como hipótesis, muestra y población. Las *propiedades/proposiciones* que destacan son distribución normal, distribución t-Student, tamaño muestral, media, desviación estándar, varianza agrupada, estadístico t-Student para muestras independientes con diferente tamaño muestral, grados de libertad y valor-p. Los *procedimientos* que aplicó fueron, calcular las medias y desviaciones estándar a partir de los datos de la actividad, varianza agrupada, estadístico t-Student, grados de libertad, establecer el nivel de significancia y encontrar la probabilidad mediante la distribución t-Student. Finalmente, concluye que existe una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2, y lo *argumenta* con base en la comparación que realiza del valor-p con respecto del nivel de significancia.

Es importante destacar que, aunque el PF9 identifica que puede aplicar una prueba t-Student para resolver el problema, a partir de las características previamente descritas, no identifica que se trata de dos muestras dependientes y utiliza las propiedades/proposiciones del estadístico t-Student y varianza agrupada para cuando se está trabajando con muestras independientes con diferente tamaño muestral. Otro aspecto que resalta en su práctica, es que asume que las desviaciones poblacionales son iguales, al parecer como justificación para su uso de la varianza agrupada en el estadístico; esto y los otros aspectos de su práctica que hemos caracterizado nos permite percatarnos de que el profesor tiene algunas nociones sobre el estadístico t-Student, su distribución y de las características de las pruebas de hipótesis. Sin embargo, presenta errores y dificultades al resolver la actividad, que inician con la identificación del tipo de datos que está trabajando y la selección apropiada del estadístico t-Student.

Para que el PF9 logre desarrollar una práctica de *Nivel 3* a partir de la práctica que ha realizado, es necesario que primeramente reconozca el tipo de datos con los que se encuentra trabajando, así como la lógica y restricciones de las pruebas t-Student. En este caso, es esencial que comprenda las implicaciones que tiene en el estadístico t-Student al estar trabajando con muestras dependientes, estos aspectos se pueden observar en el *Nivel 3* de la propuesta de niveles de razonamiento inferencial. También, el profesor podría abordar el nivel de significancia de forma preformal, es decir, trabajar con el concepto/definición de significancia como límite de desviación significativo, a partir de este límite se considera seriamente que sí es probable que exista una diferencia; esta aproximación fue trabajada por Fisher (1925a) y concuerda con la propuesta de Rossman (2008) para introducir algunas ideas sobre inferencia.



Problema 2. Utilizamos también un nivel de confianza de 95% y la prueba t.

Note que en este caso se trata de una prueba pareada

Sea μ_1 las horas de sueño ganadas con el fármaco 1 y μ_2 las ganadas con el fármaco 2.

Establecemos las hipótesis

$H_0: \mu_1 = \mu_2$ ó $\mu_1 - \mu_2 = \mu_D = 0$

$H_a: \mu_1 \neq \mu_2$ ó $\mu_1 - \mu_2 = \mu_D \neq 0$

$\alpha = 0{,}05$

La región crítica está descrita por el complemento de:

$-2.262 < t < 2.262$

y $t = \dfrac{\bar{d} - d_0}{s_D/\sqrt{n}}$ con 9 G.L.

Note que $\bar{d} = 1{,}68$ y $s_D = 1.34$

Así que

$t = \dfrac{1.68 - 0}{1.34/\sqrt{10}} = 3{,}96$

Vemos que t "cae" en la región crítica, pues es mayor a 2.262, por lo que se rechaza la H_0 y se acepta la alternativa.

Concluimos que hay diferencia entre los promedios de las horas ganadas con el fármaco 1 y con el fármaco 2.

Figura 6.30. Práctica del profesor en formación 20 (G2) sobre la actividad 5

Un tercer tipo de prácticas interesante de analizar, se ejemplifica con la respuesta del profesor en formación 20 (PF20), la cual se presenta en la Figura 6.30. En dicha práctica, podemos ver que el futuro profesor identificó que se tienen muestras pareadas y puede aplicar una prueba t-Student para este tipo de muestras. Indicó que μ_1 y μ_2 son las horas de sueño ganadas con el fármaco 1 y 2, respectivamente, a partir de esto estableció las hipótesis estadísticas en lenguaje simbólico. También, estableció un valor de alfa, obtuvo la región crítica, indicó la propiedad/proposición sobre el estadístico t-Student y lo calculó. Para concluir si existe una diferencia entre las horas de sueño que se ganan con ambos fármacos, comparó el valor que obtuvo del estadístico con la región crítica y señaló que como "cae en la región crítica" rechaza la hipótesis nula y acepta la alternativa. Lo anterior lo llevó a concluir que "hay diferencia entre los promedios de las horas ganadas con el fármaco 1 y con el fármaco 2".

La práctica de este profesor es característica del *Nivel 4*, debido a que podemos observar diversos criterios de dicho nivel que se encuentran vinculados con los siguientes objetos primarios: los *elementos lingüísticos* presentes son lenguaje natural y simbólico, algunos *conceptos/definiciones* son muestra, población, muestras pareadas, nivel de confianza, nivel de significancia, hipótesis y región crítica. También observamos *propiedades/proposiciones* tales como estadístico t-Student para muestras pareadas, media y desviación estándar de la diferencia de las muestras, regla de decisión con el estadístico y región crítica. Los *procedimientos* que realizó el profesor se enfocaron en la obtención de los grados de libertad, la región crítica (dos colas iguales) mediante la distribución t-Student para nueve grados de libertad y un nivel de confianza de 0.05, el cálculo del estadístico t-Student y en aplicar la regla de decisión para concluir. Los *argumentos* que formuló versaron sobre que el estadístico calculado 'cae en la región crítica' por lo cual se rechaza la hipótesis nula, en otras palabras, argumentó basándose en la regla o criterio para la toma de decisión sobre el valor crítico. A partir de los objetos matemáticos primarios y los procesos (e.g., comunicación, definición, enunciación, algoritmización, argumentación y generalización) que se identificaron en la práctica del PF20, podemos percatarnos que para su desarrollo se involucra un razonamiento inferencial de tipo formal. Este tipo de razonamiento le permitió al PF20 tomar decisiones basado en técnicas de inferencia estadística, concretamente, en la metodología de las pruebas de hipótesis. Al igual que en las prácticas del PF2 y los profesores del equipo 6 para la actividad 4, para que el PF20 fortalezca su razonamiento inferencial formal podría profundizar en su comprensión del nivel de significancia y de confianza, así como abordar los errores tipo I y tipo II, y la potencia de la prueba.

Las prácticas anteriores, son ejemplares del tipo de prácticas que realizaron tanto los profesores en formación como los profesores en ejercicio para resolver la actividad 5. A continuación, la Figura 6.31 presenta otra práctica que se destacó en el grupo 3 de profesores en ejercicio.

5. Existe una diferencia estadísticamente significativa entre uno y otro medicamento, en primer lugar 9 de los 10 los pacientes manifestaron una diferencia positiva de horas de sueño bajo el Fármaco 2

Paciente	Fármaco 1	Fármaco 2	Diferencia
1	0,7	1,9	1,2
2	-1,6	1,8	3,4
3	-0,2	1,1	1,3
4	-1,2	0,1	1,3
5	-0,1	-0,1	0
6	3,4	4,4	1
7	3,7	5,5	1,8
8	0,8	1,6	0,8
9	0	4,6	4,6
10	2	3,4	1,4

Al calcular algunos descriptores sobre la Diferencia, podemos argumentar esta diferencia:

Media	Desviación Estándar	Intervalo confianza (95%)	Límite inferior	Límite superior
1,680	1,340	0,958	2,638	0,722

Y por lo cual podemos concluir que con un nivel de confianza del 95%, suponiendo una distribución t, las diferencias entre el tiempo de sueño entre los dos fármacos se encuentra en un intervalo entre 2,638 y 0,722 horas más de sueño para el Fármaco 2 y, en consecuencia, poseen una diferencia significativa.

Figura 6.31. Práctica del equipo 6 de profesores en ejercicio (G3) sobre la actividad 5

Los profesores en ejercicio del equipo 6 indicaron en la clase sincrónica que primero buscaron observar la diferencia entre las horas ganadas entre el fármaco 2 y 1, y determinaron que al parecer existía una diferencia positiva de horas de sueño con el fármaco 2. También comentaron que, al observar los datos proporcionados en la actividad, advirtieron que se trataba de una variable continua que admite valores positivos y negativos. Otro aspecto que resaltaron es el significado que tiene el cero en esta variable y su importancia para determinar si existe diferencia alguna entre el aumento de las horas de sueño con los fármacos. Posteriormente, señalaron que podían resolver el problema con un intervalo de confianza del 95% basado en la t-Student. Una vez que obtuvieron los límites inferior y superior del intervalo, los profesores destacaron la importancia que tiene el significado del cero, como sin diferencia entre las horas de sueño que se ganan con ambos fármacos, para establecer una conclusión en términos del problema. Finalmente, indicaron que como el cero no se encuentra dentro del intervalo, entonces existe una diferencia estadísticamente significativa entre las horas de sueño que se ganan con los medicamentos.

Podemos separar la práctica del equipo 6 en dos partes, en la primera observamos que los profesores utilizan un razonamiento inferencial informal para realizar un análisis de los datos basado en las diferencias de las horas de sueño que se observó en cada uno de los pacientes con cada fármaco; este tipo de práctica se puede caracterizar en el *Nivel 1* debido a que la conjetura que realizan se basa en un análisis de la variación de los datos de ambas muestras, a partir de este análisis obtuvieron que 9 de los 10 pacientes presentaron un aumento mayor de las horas de sueño con el fármaco 2. En cambio, la segunda parte de la práctica se considera de *Nivel 4*, debido a que podemos identificar criterios representativos de dicho nivel. Los mencionados criterios se encuentran relacionados con los siguientes objetos primarios que observamos en la práctica de los profesores. Los *elementos lingüísticos* identificados son representación tabular y lenguaje natural y simbólico, mientras que algunos *conceptos/definiciones* son muestras dependientes, intervalo de confianza y nivel de confianza; *propiedades/proposiciones* tales como media y desviación estándar de la diferencia de las muestras, estadístico t-Student, límite inferior y superior del intervalo de confianza y distribución t-Student. Los *procedimientos* para llevar a cabo el intervalo de confianza los realizaron con los recursos tecnológicos Excel; y los *argumentos* estuvieron basados en que el cero se encuentra fuera del intervalo de confianza y por lo tanto se tendría una diferencia significativa en el aumento de las horas de sueño de ambos fármacos.

Como podemos observar, existe un salto abrupto entre las dos partes de la práctica de los profesores del equipo 6. Por un lado, en la primera parte, los profesores llevan a cabo un análisis de las diferencias de horas de sueño y a partir de este análisis realizan una conjetura; en esta parte como hemos visto, por el tipo de análisis e inferencia (conjetura), se pone en juego un razonamiento inferencial informal. Por otro lado, en la segunda parte, podemos percatarnos que los mismos profesores desarrollan una práctica que se asocia a un razonamiento inferencial formal, esto es debido a que los profesores utilizaron objetos matemáticos primarios, y sus procesos asociados, que son propios de inferencia estadística, específicamente, por intervalos de confianza. Si bien, los intervalos de confianza no se encuentran contemplados en los niveles de razonamiento inferencial sobre el estadístico t-Student, los elementos que han surgido en la práctica de estos profesores podrían asociarse a un *Nivel 4* considerando la propuesta de Pfannkuch, Arnold y Wild (2015), quienes han desarrollado una propuesta para trabajar con los intervalos de confianza desde ideas intuitivas hasta formales. En nuestra propuesta de niveles se ha conceptualizado una progresión en el razonamiento inferencial desde lo informal a lo formal con las pruebas de hipótesis; sin embargo, el resultado del tipo práctica desarrollada por estos profesores podría constituir un punto de partida para analizar la pertinencia de inclusión de los intervalos de confianza en la propuesta de niveles, aspecto que no abordamos en nuestra investigación.

En las prácticas asociadas al *Nivel 4* de los profesores en ejercicio, a diferencia de las prácticas de los profesores en formación, no se plantean las hipótesis nula y alternativa en lenguaje natural o simbólico, en vez de ello, las hipótesis parecieran estar implícitas pues realizan directamente la prueba de hipótesis o el intervalo de confianza y centran su conclusión en términos del problema. Diversas investigaciones (e.g., Pfannkuch y Wild, 2004; Stohl, Angotti y Tarr, 2010) han reconocido la importancia del planteamiento de hipótesis y sugieren la introducción de una hipótesis en forma de pregunta y después ir evolucionando como se ha planteado en la propuesta de niveles de razonamiento inferencial.

6.3.3 Prácticas asociadas a la actividad 6

Como podemos observar, la siguiente actividad trata sobre dos propuestas para rehabilitar las tuberías existentes con un forro flexible, una de ellas utiliza un proceso de fusión y la otra no, hay quienes piensan que el proceso de fusión incrementa la resistencia a la tensión promedio; se desea conocer si realmente existe un incremento en la resistencia a la tensión promedio cuando se utiliza dicho proceso de fusión.

Actividad 6.

En época de lluvias, es común ver en las noticias los socavones que se generan en diversas ciudades del mundo, una de las principales causas el deterioro de las tuberías (que es un problema creciente). Este deterioro en las tuberías genera el reblandecimiento del suelo por la humedad y con el paso de los automóviles, el peso de las construcciones, los camiones, entre otros, provocan que la parte superficial empiece a vibrar y la resistencia de la parte superior ya no es suficiente y colapsa, provocando así los socavones y con ellos lamentables accidentes. Una posible solución, que no implica el cambio de la red de tuberías, es rehabilitar las tuberías existentes utilizando un forro flexible. Sin embargo, existen dos propuestas utilizar un proceso de fusión y no utilizar el proceso en el forro. Quienes están a favor del proceso de fusión consideran que el proceso incrementa la resistencia a la tensión promedio. Se han reportado los siguientes datos de resistencias a la tensión $(lb/pulg^2)$ de especímenes, cuando se utilizó el proceso de fusión en el forro y cuando este proceso no se utilizó. ¿Realmente existe un incremento en la resistencia a la tensión promedio cuando se utiliza el proceso de fusión? Explica con detalle tu respuesta.

Tabla 6.7. *Resistencias a la tensión* $(\boldsymbol{lb/pulg^2})$ *de especímenes*

Sin proceso de fusión	**Con proceso de fusión**
2748	3027
2700	3356
2655	3359
2822	3297
2511	3125
3149	2910
3257	2889
3213	2902
3220	
2753	

El primer tipo de práctica característica para esta actividad se ejemplifica con la práctica que desarrolló el equipo 5 de profesores en ejercicio del grupo tres (Figura 6.32).

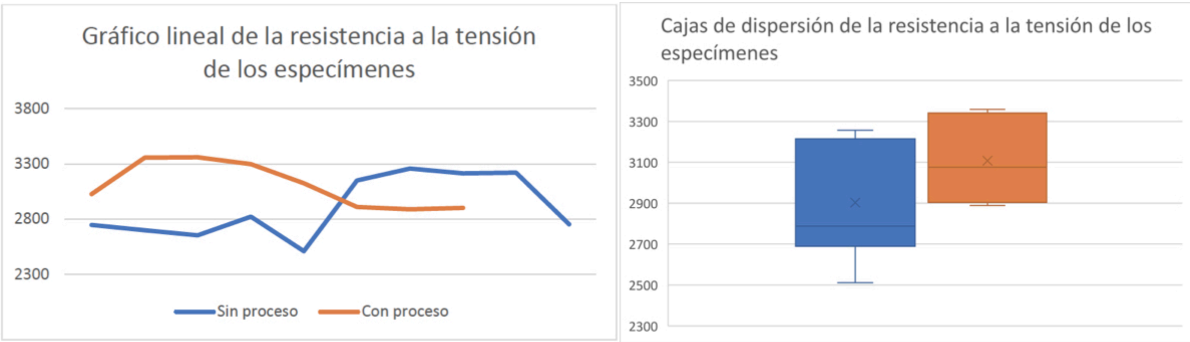
	Sin proceso	*Con proceso*
Media	2902.8	3108.125
Error típico	87.6789345	72.7859969
Mediana	2787.5	3076
Moda	#N/D	#N/D
Desviación estándar	277.2651358	205.869888
Cuenta	10	8

Figura 6.32. Práctica del equipo 5 de profesores en ejercicio (G3) sobre la actividad 6

Los profesores en ejercicio del equipo 5 mencionaron que decidieron realizar algunos cálculos sobre los datos de ambas muestras para compararlos, por ejemplo, calcularon la media, el error típico y la desviación estándar. También indicaron que compararon los valores promedios de las muestras y observaron que, con el proceso de fusión el valor promedio es aproximadamente 3108 $lb/pulg^2$ y sin proceso de fusión es 2902 $lb/pulg^2$, y que se percataron de diferencias evidentes en el promedio de con y sin proceso de fusión, y que por lo tanto, concluyen que sí existe en promedio una diferencia de incremento, cuando de utiliza el proceso de fusión de 206 $lb/pulg^2$, en comparación con cuando no se utiliza el proceso de fusión; aunque no descartan que la comparación pueda ser errónea por no ser la misma cantidad de datos en ambas muestras.

En la práctica desarrollada por el equipo 5 de profesores en ejercicio (G3), se identificaron *elementos lingüísticos* como lenguaje natural y representación tabular, *conceptos/definiciones*, tales como, muestra; así como las *propiedades/proposiciones* media, mediana, desviación estándar, error típico y moda. Los *procedimientos* que realizaron los profesores son el cálculo de la media, desviación estándar, mediana, error típico, las diferencias entre las medias de las muestras. Los argumentos que utilizaron estuvieron centrados en las diferencias que obtuvieron entre las medias de las muestras, sin embargo, se preguntaron si esto era válido debido a que las muestras eran de diferente tamaño. Los objetos matemáticos identificados en esta práctica corresponden al *Nivel 1* de razonamiento inferencial sobre el estadístico t-Student. Para que la práctica de estos profesores involucrara más elementos del *Nivel 1*, podrían considerar las comparaciones de las otras medidas estadísticas que obtuvieron (e.g., mediana, desviación estándar), también podrían calcular los cuartiles y realizar gráficos de caja y bigote, lo cual les podría ayudar a analizar la variación interna de cada muestra, como lo propone Galton (1875), y visualizar el comportamiento de los datos.

Una segunda práctica característica para esta actividad se ejemplifica en la Figura 6.33, con los desarrollos del profesor en formación 13 (PF13).



Sin proceso	$\bar{X} = 2902.8$	$X_{Me} = 2787.5$	$Q_1 = 2688.75$	$Q_3 = 3214.75$	$S = 277.265$
Con proceso	$\bar{X} = 3108.13$	$X_{Me} = 3076$	$Q_1 = 2904$	$Q_3 = 3341.25$	$S = 205.869$

Figura 6.33. Práctica del profesor en formación 13 (G2) sobre la actividad 6

El PF13 indicó que, para responder este problema, inicialmente realizó una gráfica comparativa de líneas para los datos de las muestras y que, al observarla, le pareció que no había diferencias significativas entre llevar a cabo el proceso de fusión y no, para forrar las tuberías. Pero que después, decidió analizar por medio de un gráfico de caja y bigote, para lo cual tuvo que calcular los cuartiles y que también había calculado la media y desviación estándar. A partir de estos cálculos y el gráfico de caja y bigote, señaló que pudo observar que los datos recogidos de la muestra donde se llevó a cabo el proceso de fusión se encuentran más centralizados y que los datos de la muestra sin proceso de fusión se encuentran más dispersos, por lo que la media de esta última muestra se puede ver afectada. A partir de ello, el profesor en formación concluyó que, le parecía que los datos indicaban que sí hay un incremento en la resistencia a los especímenes cuando se realiza el proceso de fusión para recubrir las tuberías.

En la práctica del PF13 identificamos objetos matemáticos primarios, dentro de los que distinguimos los *elementos lingüísticos* representaciones gráficas, lenguaje natural y simbólico; *conceptos/definiciones* tales como muestra, simetría y dispersión; algunas *propiedades/proposiciones* son media, desviación estándar, mediana (cuartil dos) y cuartiles uno y tres. Los *procedimientos* fueron realizar el gráfico de líneas y de caja bigote, calcular los cuartiles, media y desviación estándar. Para *argumentar,* hace uso principalmente de la dispersión que visualiza en los diagramas de caja y bigote, así como en media. Los objetos primarios que hemos identificado en esta práctica matemática corresponden al *Nivel 1* de razonamiento inferencial. En esta práctica el profesor también podría realizar un análisis de la variación interna de ambas muestras mediante el método de fluctuación (Edgeworth, 1885), tal como se propuso en la práctica del PF12 (Figura 6.28).

La práctica de la Figura 6.34, corresponde al profesor en formación 1 (PF1) del grupo dos y resultó característica de las prácticas desarrolladas en la actividad 6.

Si μ_1 representa el promedio de la tensión **cuando se utiliza el proceso de fusión** y μ_2 representa el promedio de la tensión **cuando no** se utiliza. Entonces la interrogante es si $\mu_1 > \mu_2$.

con proceso	$n_1 = 8$	$\bar{x}_1 = 3108.125$	$s_1 = 205.8699$
sin proceso	$n_2 = 10$	$\bar{x}_2 = 2902.8$	$s_2 = 277.2651$

Primero se pondrá a prueba el supuesto de varianzas iguales mediante una prueba de hipótesis con $\alpha = 0.05$.
Ensayo de hipótesis:

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ Grados de libertad del numerador: $n_2 - 1 = 9$

$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ Grados de libertad del denominador: $n_1 - 1 = 7$

Note que $s_2 > s_1$ $F_{0.05/2;9,7} = F_{0.025;9,7} = 0.2383$

$F_{0.975;9,7} = 4.8232$

Regla de decisión:
Si $0.2383 \leq F \leq 4.197$ No se rechaza H_0,
Si $F < 0.2383$ o $4.197 < F$ Se rechaza H_0,

$$F = \frac{s_2^2}{s_1^2} = \frac{277.2651^2}{205.8699^2} = 1.8139$$

Decisión: Como $0.2383 \leq 1.8139 \leq 4.197$ No se rechaza H_0
Se concluye con un 95% de confianza que existe suficiente evidencia para decir que las varianzas de las poblaciones son iguales.
Se procede a comparar las medias con una prueba para varianzas desconocidas pero iguales.

Se utiliza una prueba para varianzas desconocidas pero iguales, con un nivel de significancia de 0.05.

$H_0: \mu_1 = \mu_2$. Esta hipótesis implica que $\mu_1 - \mu_2 = 0$
$H_1: \mu_1 > \mu_2$. Esta hipótesis implica que $\mu_1 - \mu_2 > 0$

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{205.8699^2(8-1) + 277.2651^2(10-1)}{8 + 10 - 2}$$
$$= 61785.0207$$
$$s_p = 248.5659$$

$$t = \frac{(3108.125 - 2902.8) - 0}{248.5659\sqrt{1/8 + 1/10}} = 1.741442$$
$$t_{0.05;n_1+n_2-2} = t_{0.05;10+8-2} = t_{0.05;16} = 1.746$$

Si $t \leq 1.746$ No se rechaza la hipótesis nula
Si $t > 1.746$ se rechaza la hipótesis nula
Como $1.7414 \leq 1.746$ entonces No se rechaza la hipótesis nula.
Se concluye que no existe un incremento significativo en la resistencia a la tensión promedio cuando se utiliza el proceso de fusión.

Figura 6.34. Práctica del profesor en formación 1 (G2) sobre la actividad 6

El PF1 indicó que, la interrogante es, si realmente existe un incremento en la media de la resistencia a la tensión cuando se utiliza el proceso de fusión o, en otras palabras, si el promedio de la resistencia a la tensión cuando se utiliza el proceso es mayor que el promedio de la resistencia a la tensión cuando no se utiliza. Para responder a esa interrogante, se planteó realizar una prueba t, pero que antes de hacerlo era necesario poner a prueba el supuesto de varianzas iguales mediante una prueba de hipótesis con un nivel de significancia de 0.05 para lo cual recurría a una prueba F. Y una vez que, mediante la prueba F, concluyó que las varianzas son iguales, realizó una prueba t-Student para varianzas desconocidas pero iguales, con un nivel de significancia de 0.05. Finalmente, indicó que esta prueba le permitió concluir que no existe un incremento significativo en la resistencia a la tensión promedio cuando se utiliza el proceso de fusión.

En la práctica desarrollada por el PF1 se identificaron *elementos lingüísticos* como lenguaje natural y simbólico, *conceptos/definiciones* tales como muestra, muestras independientes, nivel de significancia, nivel de confianza e hipótesis. También observamos *propiedades/proposiciones* como el estadístico t-Student para muestras independientes con igualdad varianzas, desviación estándar agrupada, regla de decisión con el estadístico t-Student, estadístico F, grados de libertad, media y desviación estándar. Los *procedimientos* que realizó el profesor corresponden a dos pruebas de hipótesis, la primera de ella para comprobar si las varianzas de las muestras son iguales y con base en ello decidir qué estadístico t-Student debe seleccionar para responder a la problemática de las tuberías, es decir, primero realizó una prueba con el estadístico F y después una prueba t-Student. Los *argumentos* se apoyaron principalmente en el criterio de decisión del valor crítico. De acuerdo con los objetos matemáticos primarios que identificamos en la práctica del PF1, podemos indicar que en esta práctica se involucra un razonamiento inferencial de tipo formal, mismo que es característico del *Nivel 4* de razonamiento inferencial. Para que el PF1 fortalezca su razonamiento inferencial formal, podría trabajar los errores tipo I y tipo II, y la potencia de la prueba. Es importante destacar que, en este caso, el profesor no asumió igualdad de varianzas, sino que fue más allá, realizó una prueba de hipótesis para probar si realmente no había evidencia suficiente como para decir que las varianzas no eran iguales.

6.3.4 Prácticas asociadas a la actividad 7

La actividad 7 trata de un estudio experimental con pacientes ambulatorios con Covid-19 leve o moderado, en esta actividad nos interesa conocer si realmente existe una diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y cuando se recibe el placebo. A continuación, presentamos la actividad 7 y las prácticas que desarrollaron los profesores en ejercicio sobre esta actividad.

Actividad 7.

Como sabemos, el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) causado por la enfermedad Covid-19, con mayor frecuencia es leve, pero puede ser grave y potencialmente mortal. Ante este escenario los investigadores de BLAZE 1, han realizado un estudio sobre el anticuerpo neutralizante LY-CoV555 del SARS-CoV-2 en pacientes ambulatorios con Covid-19. Prevén que los anticuerpos monoclonales neutralizantes de virus reduzcan la carga viral, mejoren los síntomas y eviten la hospitalización. Este estudio se llevó a cabo en la fase dos, con pacientes ambulatorios con Covid-19 leve o moderado que fueron diagnosticados recientemente. Los pacientes fueron asignados al azar para recibir una única dosis intravenosa del anticuerpo neutralizante LY-CoV555 (de 2800 mg) o de placebo, los resultados obtenidos sobre la carga del virus se resumen en la siguiente tabla. ¿Existe realmente una diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y cuando se recibe el placebo?

Tabla 6.8. *Disminución media en la carga viral desde la baseline al día 11*

	Media	**Desviación estándar**
LY-CoV555 (n=20)	-4.00	0.737
Placebo (n=20)	-3.47	0.669

Nota: La carga viral media o baseline al iniciar el estudio fue de 23.9

En la figura 6.35, se encuentra la primera práctica característica para esta actividad, ejemplificada con los desarrollos del equipo 3 de profesores en ejercicio (G3).

Respuesta:

Para esta pregunta que se nos realiza la actividad, consideramos que **NO** hay diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y cuando se recibe el placebo, ya que las medias y las desviaciones estándar son muy similares. Aunque si se deseara saber y probar, creo que sería bueno llevar a cabo una prueba estadística, como la T de Student por ejemplo.

Figura 6.35. Práctica del equipo 3 de profesores en ejercicio (G3) sobre la actividad 7

Cuando los profesores del equipo tres comentaron lo que habían realizado para resolver la actividad 7, destacaron que como solo contaban con algunos datos que resumen o describen las muestras, como son la media, desviación estándar y el tamaño de la muestra, consideraron que podían comparar las medias; entonces notaron que, en promedio, a un paciente se le disminuye la carga viral -4.00 cuando recibe la dosis del anticuerpo y que esta disminución, de la carga viral, se desvía de la media 0.737. Mientras que en promedio, a un paciente se le disminuye la carga viral -3.47 cuando recibe el placebo y que esta disminución de la carga viral se desvía de la media 0.669. A partir de esto, concluyeron que realmente no existe una diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y del placebo porque la disminución de la carga viral promedio y la desviación con la media son muy similares. Finalmente, indicaron que tal vez, se podría realizar una prueba T para probar estadísticamente su conclusión, es decir, si realmente no existe una diferencia significativa entre la disminución de la carga viral promedio de ambas muestras.

En la práctica matemática que desarrolló el equipo 3 de profesores en ejercicio podemos identificar objetos matemáticos primarios, el *elemento lingüístico* identificado es lenguaje natural, algunos *conceptos/definiciones* son muestra, desviación y significancia, mientras que las *propiedades/proposiciones* son media y desviación estándar. Los *procedimientos* que realizaron los profesores corresponden a una comparación de las muestras mediante la media y desviación estándar. El *argumento* que utilizaron los profesores para fundamentar su conclusión fue que, las medias y desviaciones estándar de las muestras son muy similares. Los elementos identificados en esta práctica corresponden al *Nivel 1* de razonamiento inferencial sobre el estadístico t-Student.

Para que los profesores realicen una práctica de *Nivel 2*, como bien señalaron ellos mismos, podrían realizar una prueba t-Student para dos muestras, plantear en lenguaje natural la hipótesis que se encuentra implícita en la actividad e interpretar la probabilidad como una medida de ocurrencia de que la diferencia de medias se encuentre fuera del rango $\pm t$. Esta forma de interpretar la probabilidad es la primera de tres fases que se proponen en los niveles de razonamiento inferencial para comprender el valor-p.

La segunda práctica característica de la actividad 7 se ejemplifica con la realizada por el profesor en formación 22 del segundo grupo (Figura 6.36).

H_0: No existe una diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y cuando se recibe el placebo.

H_A: $\neg H_0$

		$\bar{x}$	S
1	CoV555	-4	0,737
2	Placebo	-3,47	0,669

Nótese que, gráficamente, de la información brindada se puede realizar un boceto de las distribuciones en ambos casos (No necesariamente normales). Lo que sirve para interpretar que sí existe una disminución, sin embargo, no es tan significativa ya que los datos de x_1 están más dispersos que los de x_2.

Además, la diferencia entre los promedios es relativamente baja por lo que no hay una diferencia significativa.

Considero que se puede rechazar la hipótesis nula pero no sé como probarlo.

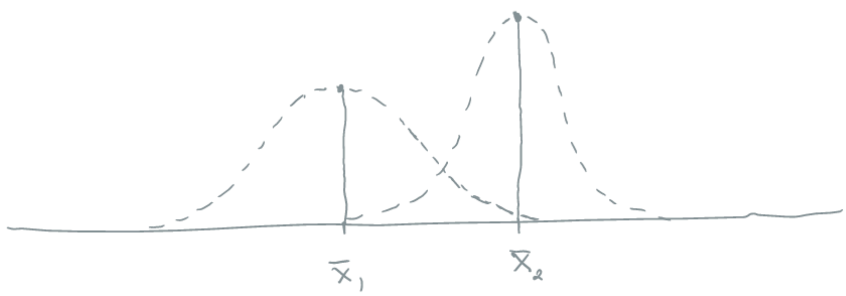


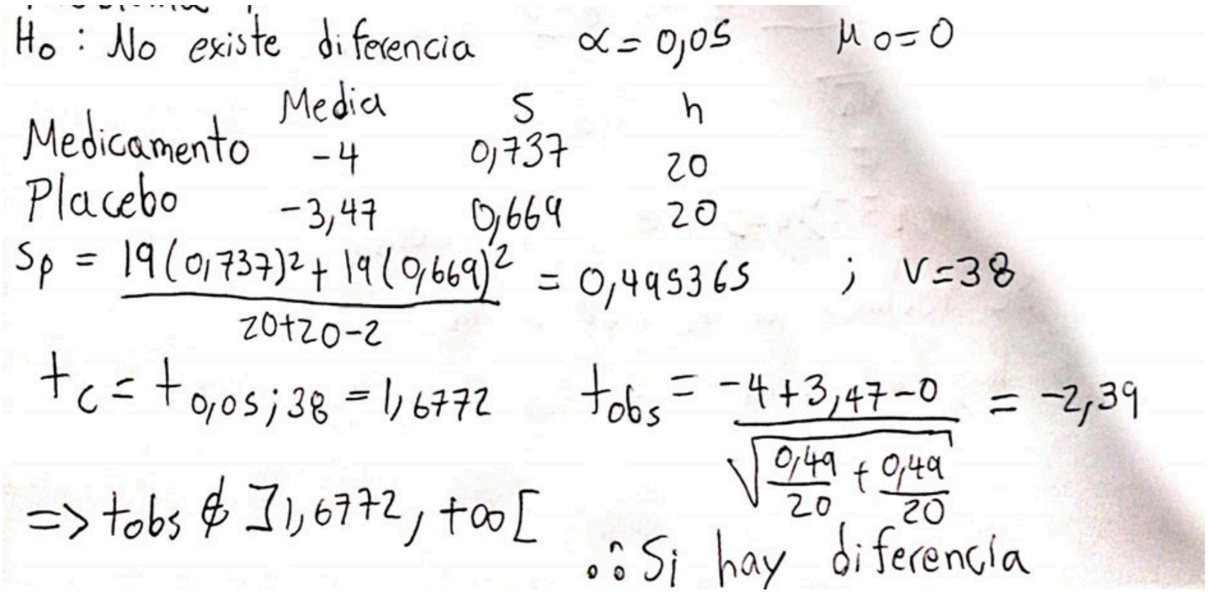
Figura 6.36. Práctica del profesor en formación 22 (G2) sobre la actividad 7

El profesor en formación 22 (PF22), inició planteando las hipótesis y comentó que para analizar los datos de las muestras decidió realizar gráficas que le permitieran compararlas, a modo de bocetos de las distribuciones de las muestras. Señaló que, a partir de las gráficas logró interpretar que sí existe una disminución, pero se preguntó ¿qué tan significativa es dicha disminución?, debido a que los datos de la muestra uno se encuentran más dispersos. También indicó que la diferencia entre los promedios es muy pequeña y que eso lo llevó a pensar que no hay una diferencia significativa. Finalmente, concluyó que considera que se debe rechazar la hipótesis nula, sin embargo, aclara que no sabe cómo probarlo más allá de las comparaciones que ha realizado.

En la práctica que realizó el PF22 a propósito de la actividad 7, se identificaron *elementos* lingüísticos como lenguaje natural y simbólico y representaciones gráficas; algunos *conceptos/definiciones* fueron muestra, distribución, significancia e hipótesis; *propiedades/proposiciones* tales como media y desviación estándar. Los *procedimientos* que realizó el profesor son plantear las hipótesis con lenguaje natural y simbólico, representar gráficamente la distribución de las muestras y realizar una comparación de las gráficas y los valores de las medias y desviaciones estándar de las muestras. Los *argumentos* que empleó se centran en la disminución que observó en la gráfica. La mayoría de los objetos matemáticos primarios y sus procesos matemáticos, corresponden al *Nivel 1* de razonamiento inferencial.

Para que la práctica del PF22 transite a una de *Nivel 3*, el profesor podría, además de lo propuesto al equipo tres de profesores en ejercicio (Figura 6.35) para transitar a una de *Nivel 2*, plantear la hipótesis alternativa en lenguaje natural antes de realizar la prueba con el estadístico t-Student, para dos muestras independientes y con igualdad del tamaño muestral. También podría establecer un valor de probabilidad como límite significativo, a partir del cual se va a considerar seriamente que existe realmente una diferencia en la disminución de la carga viral cuando se recibe la dosis del anticuerpo y cuando se recibe el placebo. A partir de esto, el PF22 podría comparar el valor de la probabilidad obtenido y el límite que ha establecido, también podría encontrar el valor del estadístico teórico que corresponde al valor de la probabilidad preestablecido y los grados de libertad (límite de significancia) y compararlo con el valor del estadístico calculado. En este nivel se introduce la noción de significancia, únicamente como un límite que le permite al estudiante tener un punto crítico para la toma de decisiones. Para realizar estas comparaciones y comprender las nociones involucradas, podría auxiliase de representaciones gráficas o de applet, tal y como sugiere Rossman (2008) para la comprensión del valor-p.

Un tercer tipo de práctica que es característica de la actividad 7 se ejemplifica con la desarrollada por el profesor en formación 7 (Figura 6.37).



H_0 : No existe diferencia $\quad \alpha = 0,05 \quad \mu_0 = 0$

	Media	s	n
Medicamento	-4	0,737	20
Placebo	-3,47	0,669	20

$$S_p = \frac{19(0,737)^2 + 19(0,669)^2}{20+20-2} = 0,495365 \quad ; \quad v = 38$$

$$t_c = t_{0,05;38} = 1,6772 \qquad t_{obs} = \frac{-4+3,47-0}{\sqrt{\frac{0,49}{20} + \frac{0,49}{20}}} = -2,39$$

$$\Rightarrow t_{obs} \notin \left]1,6772, +\infty\right[$$

$\therefore$ Si hay diferencia

Figura 6.37. Práctica del profesor en formación 7 (G2) sobre la actividad 7

El profesor en formación 7 (PF7) destacó que realizó una prueba t para dos muestras independientes, con la intención de comparar las muestras y definir si existían diferencias significativas entre las medias. El profesor también indicó que primeramente planteó la hipótesis nula, estableció el nivel de significancia y el valor de μ_0, que eran necesarios para realizar la prueba. El PF7 finalizó comentando que, sí hay diferencia entre las medias de las muestras, debido a que la t observada se encuentra dentro de la región de rechazo.

En la práctica desarrollada por el PF7 profesores podemos observar *elementos lingüísticos* como lenguaje natural y simbólico, *conceptos/definiciones* como muestra, muestras independientes, nivel de significancia e hipótesis; *propiedades/proposiciones* tales como media y desviación estándar de la muestra, varianza agrupada, estadístico t-Student y región crítica. Los *procedimientos* que realizó el PF7 se centraron en establecer la hipótesis nula, calcular la varianza agrupada, los grados de libertad, el estadístico t-Student y la región crítica (dos colas iguales) mediante la distribución t-Student y aplicar la regla de decisión para concluir. El *argumento* que utilizó el profesor fue que el valor del estadístico t-Student observado se encuentra dentro de la región de rechazo. La mayoría de los objetos matemáticos primarios identificados en la práctica que desarrolló el PF7 pertenecen al *Nivel 4* de razonamiento inferencial para el estadístico t-Student. La práctica que desarrolló el PF7 da cuenta de un razonamiento inferencial formal, sin embargo, el profesor podría trabajar en establecer ambas hipótesis primero en lenguaje natural (*Nivel 3*) y después en lenguaje simbólico y con los errores tipo I y tipo II, así como con la potencia de la prueba (*Nivel 4*).

Las prácticas anteriores, son ejemplares del tipo de prácticas que realizaron tanto los profesores en formación como los profesores en ejercicio al resolver la actividad 7. A continuación, en las figuras 6.38 y 6.39 presentamos otras prácticas que se destacaron en el segundo grupo de profesores en formación. En estas prácticas se identifican objetos matemáticos primarios, en su mayoría, del *Nivel 4* y son muy similares a la desarrollada por el PF7 (Figura 6.37), no obstante, destacamos las siguientes particularidades:

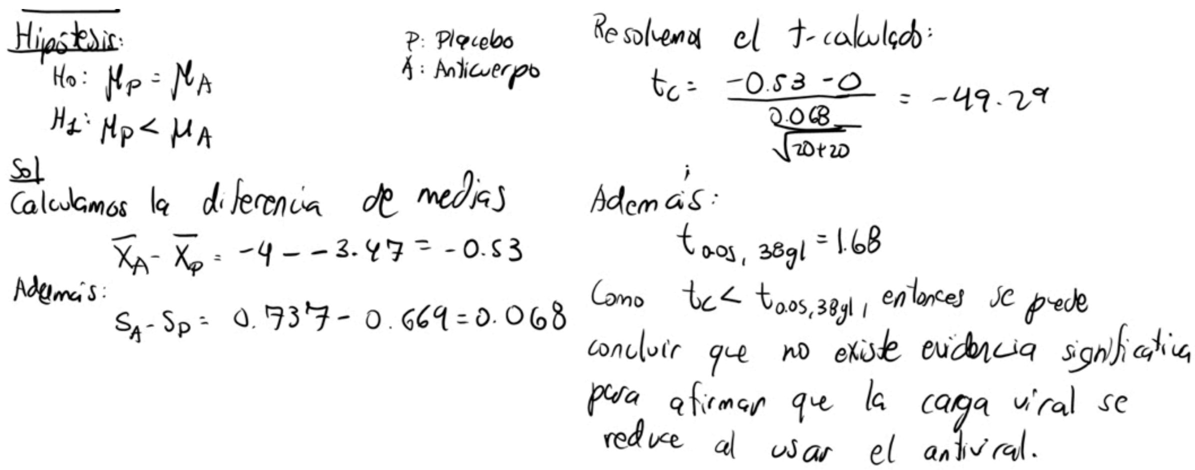
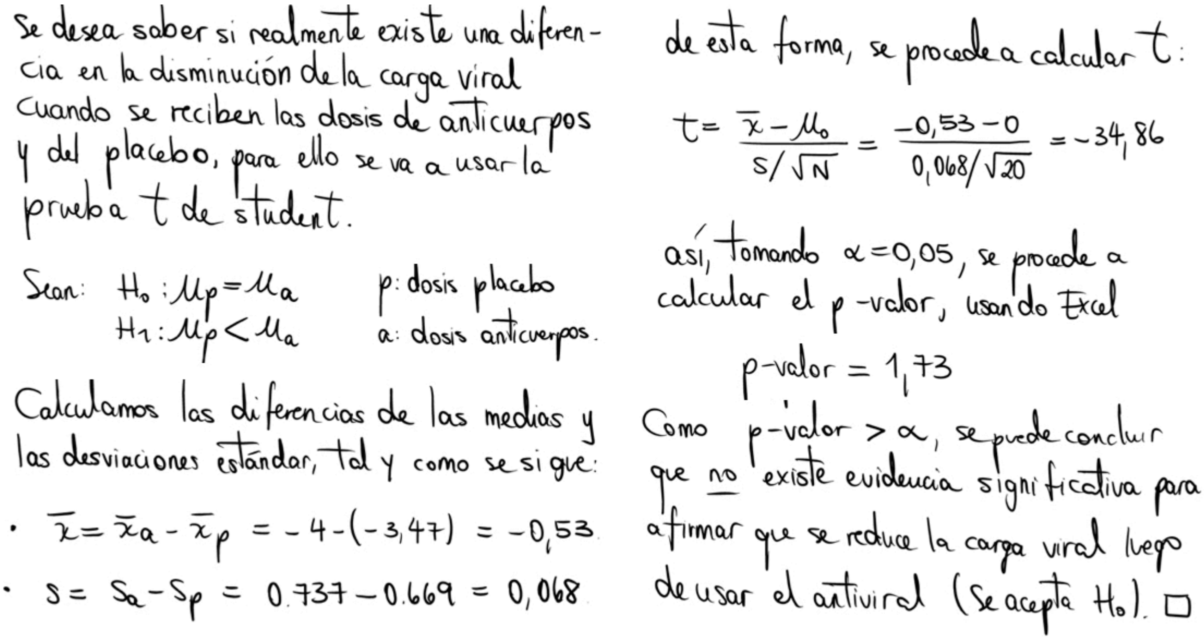


Figura 6.38. Práctica del profesor en formación 5 (G2) sobre la actividad 7

En la práctica del profesor en formación 5 (PF5), identificamos que el profesor calcula una diferencia de medias y de desviaciones estándar para poder aplicar el estadístico t-Student para muestras pareadas, pero incluye un cambio en el tamaño de las muestras. La interrogante de la actividad refiere a una diferencia y al plantear la hipótesis alternativa el profesor hace referencia únicamente a un contraste unilateral, aunque podría considerarse comprensible por el contexto del problema. También, en el mismo sentido del contrataste unilateral de la cola izquierda, al profesor se le dificulta aplicar la regla para la toma de decisión con el valor crítico; al momento de obtener el valor crítico es importante que recuerde que está trabajando con la cola izquierda, por lo que el valor es -1.68, y para compararlo con el estadístico que ha obtenido podría apoyarse con una representación gráfica de la distribución t-Student, en la cual podría identificar dónde se encuentra el valor crítico, la región de rechazo y el valor del estadístico calculado, para tomar la decisión. Esto se ha propuesto en los niveles de razonamiento inferencial.



Se desea saber si realmente existe una diferencia en la disminución de la carga viral cuando se reciben las dosis de anticuerpos y del placebo, para ello se va a usar la prueba t de student.

Sean: $H_0: \mu_p = \mu_a$ p: dosis placebo
$H_1: \mu_p < \mu_a$ a: dosis anticuerpos.

Calculamos las diferencias de las medias y las desviaciones estándar, tal y como se sigue:

- $\bar{x} = \bar{x}_a - \bar{x}_p = -4-(-3{,}47) = -0{,}53$
- $s = S_a - S_p = 0.737 - 0.669 = 0{,}068$

de esta forma, se procede a calcular t:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{N}} = \frac{-0{,}53-0}{0{,}068/\sqrt{20}} = -34{,}86$$

así, tomando $\alpha = 0{,}05$, se procede a calcular el p-valor, usando Excel

p-valor = 1,73

Como p-valor > α, se puede concluir que no existe evidencia significativa para afirmar que se reduce la carga viral luego de usar el antiviral (Se acepta H_0). □

Figura 6.39. Práctica del profesor en formación 2 (G2) sobre la actividad 7

En la práctica del PF2, se observa que para obtener la desviación estándar agrupada el profesor calcula la diferencia entre las desviaciones estándar de las muestras. También utiliza el estadístico t-Student para cuando se tienen muestras dependientes o pareadas, mientras que las muestras de este problema son independientes. Además, si realizamos una prueba con el estadístico F para comprobar si las varianzas de las muestras son iguales, podría identificar que el problema se puede abordar con el estadístico t-Student para muestras independientes, con igualdad del tamaño muestral y de varianzas. También, destacamos que el profesor al calcular el valor-p obtiene que es de 1.73 y no se percata que ese valor no es posible, esto es debido a que podemos representar el valor-p como el área bajo la curva de la distribución y esta área tiene una probabilidad de 1. En la propuesta de niveles de razonamiento inferencial se sugieren tres momentos (niveles 2, 3 y 4) para la compresión del valor-p, además en la propuesta se recomienda una introducción progresiva de las pruebas con este estadístico, atendiendo así, a las diversas especificaciones en las que el estadístico t-Student, con las variaciones correspondientes, se puede aplicar.

6.3.5 Reflexiones y consideraciones sobre el estadístico t-Student

En las prácticas desarrolladas por los profesores, a propósito de las actividades sobre el estadístico t-Student, encontramos que tanto los profesores en ejercicio como los profesores en formación desarrollaron prácticas similares, en las cuales se identificaron objetos matemáticos primarios (representaciones, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos) que se asocian a los diversos niveles de razonamiento inferencial sobre el estadístico t-Student, tal como se ha logrado evidenciar en el análisis de las prácticas. Además, hemos discutido cómo la práctica de los profesores podría ser más robusta dentro del mismo nivel o transitar al siguiente nivel.

Respecto a las prácticas de cada nivel que desarrollaron los profesores, podríamos decir que en las asociadas al *Nivel 1* se destacó el uso de representaciones gráficas (e.g., de puntos, de líneas y de caja y bigote) y de diversas medidas estadísticas como la media y desviación estándar para interpretar la dispersión de los datos. Mientras que en las prácticas de los *Niveles 2 y 3* se enfatizó el uso del planteamiento de las hipótesis en lenguaje natural, el uso del estadístico y la distribución t-Student. En las prácticas que dan cuenta del *Nivel 4* encontramos el planteamiento de las hipótesis con lenguaje simbólico, el estadístico y la distribución t-Student, reglas para la toma de decisión, nivel de confianza, nivel de significancia, muestras dependientes, muestras independientes, varianza agrupada, entre otros. Sin embargo, no fue posible observar en las prácticas algunos elementos de los niveles de razonamiento inferencial, como aquellos relacionados con los criterios sobre el error tipo I y II, y la potencia de la prueba del *Nivel 4*, la acepción de significancia del *Nivel 3*, el uso de la probabilidad del *Nivel 2* y la intercomparación y fluctuación del *Nivel 1*.

6.4 Propuesta niveles de razonamiento inferencial

En el Capítulo 5 presentamos una propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, sobre los estadísticos Chi-cuadrada (ver sección 5.1) y t-Student (ver sección 5.2), a partir de la riqueza matemática recuperada de los estudios histórico-epistemológicos sobre estos dos estadísticos y de la literatura de Educación Estadística sobre razonamiento inferencial. A partir de ellos, y de la práctica desarrollada por profesores para resolver problemas sobre estos estadísticos, realizamos la siguiente propuesta de niveles de razonamiento inferencial en términos más generales.

La propuesta de niveles de razonamiento inferencial, al igual que los niveles para los estadísticos Chi-cuadrada y t-Student, consta de cuatro niveles progresivos. Los indicadores del primer nivel están estrechamente vinculados con un RII, mientras que los del cuarto nivel con el RIF; entonces, tanto el segundo como el tercer nivel, que hemos denominado preformales, presentan rasgos de inferencia informal como de inferencia formal, esto se da en distinta gradualidad. Por lo que, en los cuatro niveles se presentan 'indicadores' graduables en procesos de generalidad y formalización. Consideramos que esta propuesta nos permite dar cuenta sobre una transición continua desde un RII a un RIF, en otras palabras, podría brindar un camino para primero promover un RII en los estudiantes y posteriormente que sobre ese razonamiento los estudiantes puedan construir un RIF.

A continuación, indicaremos a groso modo algunos puntos de encuentro o elementos clave que podemos observar tanto en los niveles para la Chi-cuadrada como para la t-Student, que nos permitieron desarrollar una propuesta más general.

Nivel 1

Por un lado, en este nivel tenemos el uso de la visualización como elemento clave para los indicadores del primer nivel. A partir de la visualización de las características de las gráficas, principalmente, (e.g., forma, dispersión, amplitud de los cuartiles, mediana, sesgo, proporción) se realizan las conjeturas, ya sea que los gráficos sean proporcionados como la única forma de acceder a los datos o que se les proporcionen datos a los estudiantes y ellos realicen los gráficos.

Por otro lado, en las estrategias de solución del *Nivel 1*, se destaca el uso de nociones de estadística descriptiva y probabilidad, principalmente. La situación problema es abordada de una forma intuitiva a partir del conocimiento previo del estudiante, utilizando nociones como la variación interna de un conjunto de datos, las medidas de tendencia central, de dispersión y de posición y, la probabilidad condicional.

En general, los indicadores de este nivel corresponden a un razonamiento inferencial informal y van más allá de simples interpretaciones de los gráficos o cálculos de medidas estadísticas y de probabilidad, pues dichos indicadores están enfocados en que el estudiante razone inferencialmente y realice conjeturas a partir de los datos.

Nivel 2

Un primer aspecto, que destaca en el *Nivel 2*, es el reconocimiento de la hipótesis de investigación en el problema y hacerla explícita por medio del lenguaje natural. Anteriormente ya hemos destacado las bondades de trabajar de forma intuitiva con las hipótesis, esta forma de trabajar con la hipótesis corresponde a un primer momento, de tres, para promover la comprensión progresiva de las hipótesis estadísticas (e.g. lógica, propiedades y características).

El segundo aspecto que resaltamos es el uso de la probabilidad como una medida de ocurrencia en términos del contexto, por ejemplo, en cuántos casos sería posible observar un conjunto de datos como los proporcionados en la situación problema y que las desviaciones de lo esperado se deban al azar. Esta concepción y uso que se le da a la probabilidad, está enmarcada en un primer momento para promover la comprensión progresiva del valor-p y otras nociones relacionadas como la significancia.

El tercer aspecto a destacar del *Nivel 2*, refiere a que el estudiante se encuentra por primera vez con el estadístico y los contextos en que se utiliza. No basta con conocer los estadísticos, sino que es primordial que el estudiante pueda reconocer el tipo de datos que le proporciona el problema y la lógica de cada prueba con el estadístico, para que el estudiante pueda identificar qué prueba es adecuada para analizar los datos.

En síntesis, en este nivel, se encuentran indicadores que corresponden a un nivel de tipo pre-formal, donde se pueden identificar rasgos del RII tales como la forma en que se utiliza (define) la probabilidad, como una medida de certidumbre asociada a un evento, sin compararla con un nivel de significancia o un límite preestablecido, y cómo se hace explícita la hipótesis del problema en lenguaje natural.

Nivel 3

El primer elemento clave en el *Nivel 3* corresponde al planteamiento de las hipótesis (nula y alternativa) en lenguaje natural, detrás de esto se encuentra la teoría de Neyman y Pearson, aunque el estudiante incursiona primero de una forma más accesible con esta forma de planteamiento de las hipótesis. En este caso, se encuentra en el segundo momento para promover la comprensión progresiva de las hipótesis estadísticas.

Un segundo elemento clave, son las restricciones que tienen las pruebas que realizamos con el estadístico que estamos trabajando (Chi-cuadrada o t-Student). Por ejemplo, cuándo se aplica un factor de corrección de continuidad, por qué se aplica y qué pasa si no lo aplicamos, o qué implicaciones tiene si no utilizamos la varianza agrupada, si contamos con muestras dependientes o independientes, muestras con diferentes tamaños e incluso desigualdad de varianzas.

El tercer elemento clave en el *Nivel 3* es la introducción de la significancia, como un límite, y la introducción de la regla de decisión intuitiva, para rechazar o no rechazar la hipótesis nula. Este límite incide directamente en la toma de decisión y lo utiliza para justificar la respuesta al problema. En este caso, podemos notar que considerar la significancia como un límite puede ser visto como una versión intuitiva del nivel de significancia (*Nivel 4*). También es importante resaltar que la significancia es un concepto clave en la Inferencia Estadística, en este nivel se introduce únicamente como un límite que le permite al estudiante tener un punto crítico para la toma de decisiones. Entonces, en este nivel se está trabajando en un primer momento (de dos) para comprender el nivel de significancia.

Consecuentemente, los indicadores que se encuentran en este nivel se consideran pre-formales, pero con un mayor grado de formalidad que en el nivel anterior, dado que se tiene una visión más amplia de las posibilidades para trabajar con las pruebas Chi-cuadrada y t-Student, hacia la generalización de estas pruebas y las restricciones que implican. Los aspectos que marcan cierto grado preformal son, por ejemplo, el lenguaje utilizado para plantear las hipótesis y la forma de trabajar y comprender la significancia.

Nivel 4

El primer elemento a destacar, es el planteamiento de las hipótesis nula y alternativa con lenguaje estadístico, esto se debe a que las hipótesis son uno de los elementos claves en los niveles de razonamiento inferencial, esta forma de plantear las hipótesis –con lenguaje simbólico y natural, cuando corresponda– es el tercer momento, el primero lo abordamos en el *Nivel 2*, con la identificación de la hipótesis del problema y hacerla explícita en lenguaje natural, y el segundo en el *Nivel 3*, con el planteamiento de las hipótesis nula y alternativa en lenguaje natural.

El segundo elemento que se destaca en este nivel, es el valor-p y valor crítico, las reglas de decisión, el nivel de significancia y el nivel de confianza, y cómo son utilizados para responder el problema. Como podemos notar, en este caso recurrimos a varias nociones, esto se debe a que se encuentran relacionadas y que, para realmente comprenderlas, sugerimos que se promuevan de forma integral.

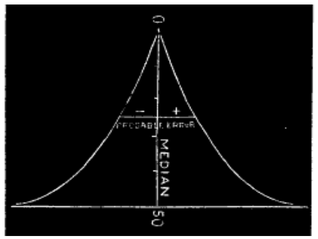
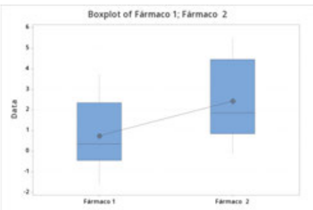
En el *Nivel 4*, se trabaja con un segundo momento del nivel de significancia, es importante que el estudiante comprenda el valor del nivel de significancia como la probabilidad de concluir que existe una desviación o diferencia cuando en realidad no existe, en otras palabras, es la probabilidad de equivocarnos que estamos dispuestos a asumir, si rechazamos la Ho. Mientras que el nivel de confianza, se puede entender como la probabilidad de que el verdadero valor del parámetro se encuentre dentro de nuestra estimación. Entonces, es importante recordar que tenemos a 1 como la probabilidad de certeza, (α) como el nivel de significancia y (1-α) como el nivel de confianza. Del mismo modo, podemos interpretar al valor-p como una medida de probabilidad de que el valor del estadístico calculado sea posible dada la hipótesis nula.

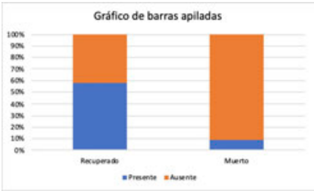
Mientras que el tercer elemento, corresponde al uso del error tipo I y tipo II, así como de la potencia de la prueba. En este nivel el estudiante puede auxiliase de gráficos para comprender las relaciones entre los errores tipo I y tipo II, por ejemplo, que cuanto mayor es α, menor es β. Además de calcular la potencia de la prueba, se promueve que el estudiante reconozca que la potencia está relacionada con el tamaño de la muestra y el nivel de significancia; y pueda utilizarla para argumentar sobre la prevalencia de su estimación en la población.

Los indicadores que se presentan en este nivel corresponden a un razonamiento inferencial formal. Se espera que el estudiante pueda tomar decisiones basado en las técnicas estadísticas de la metodología de las pruebas de hipótesis y sea capaz de analizar críticamente sus desarrollos estadísticos. Cabe destacar, que más allá de los aspectos técnicos, consideramos esencial el razonamiento que hay detrás de las pruebas de hipótesis.

La Tabla 6.9 recoge la propuesta general de niveles progresivos de razonamiento inferencial, como señalamos anteriormente, esta propuesta se realiza tanto a partir de los puntos de encuentro entre las propuestas de niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student, como de la práctica desarrollada por profesores para resolver problemas sobre estos estadísticos.

Tabla 6.9. *Propuesta general de niveles de razonamiento inferencial*

Nivel 1	Nivel 2	Nivel 3	Nivel 4
1.1 Visualización	2.1 Identificar la prueba necesaria para analizar los datos	3.1 Restricciones de las pruebas	4.1 Criterio para la toma de decisión
• A través de las características de las gráficas (e.g., forma, dispersión, amplitud de los cuartiles, mediana, sesgo, proporción) puede establecer conjeturas y dar respuesta al problema planteado. Un ejemplo del tipo de gráficas:   	• Reconoce el tipo de datos que está trabajando (e.g., se trata de una muestra o de una población, son cualitativos o cuantitativos, se clasifican de acuerdo a una o dos variables, tipo de muestreo y número de muestras y tamaño de la muestra). • Comprende el problema a resolver. • Comprende la lógica de las pruebas a aplicar. • Es capaz de seleccionar la prueba adecuada, aunque aún no puede desarrollarla.	• Conoce y comprende las implicaciones de las limitaciones y/o restricciones de las pruebas. • Es capaz de hacer explicita la hipótesis nula y alternativa en lenguaje natural. • Comprende y es capaz de aplicar las especificidades de las pruebas (como aplicar un factor de corrección al estadístico o calcularlo con una desviación estándar agrupada).	<u>Valor-p</u> • Puede plantear la hipótesis nula y alternativa con lenguaje simbólico. • Conoce y comprende los valores comunes del nivel de significancia. • Comprende la relación entre el nivel de significancia (α) y el nivel de confianza $(1-\alpha)$. • Conoce, comprende y es capaz de aplicar la regla de decisión. Si el $valor-p<\alpha$ se rechaza H_0 <u>Valor crítico</u> • Es capaz de identificar y comprende el valor teórico del estadístico, de acuerdo con α y gl. • Puede representar gráficamente las regiones de aceptación y de rechazo, y comprende sus relaciones con el nivel de confianza y la hipótesis nula.

<table>
<tr>
<td>
</td>
<td></td>
<td></td>
<td>• Conoce, comprende y es capaz de aplicar la regla de decisión.
• Es capaz de brindar una respuesta al problema utilizando los resultados de la prueba y realiza argumentos con fundamentación estadística.</td>
</tr>
<tr>
<td>1.2 Trabajar con datos</td>
<td>2.2 Una aproximación a las pruebas de hipótesis</td>
<td>3.2 Conexiones y argumentos</td>
<td>4.2 Error tipo I y II, y Potencia de la prueba</td>
</tr>
<tr>
<td>• Para realizar conjeturas (inferencias informales) a partir de uno o dos grupos de datos, el estudiante es capaz de analizarlos con las herramientas de estadística descriptiva y de probabilidad.
• Comprende conceptos básicos de estadística descriptiva y probabilidad.</td>
<td>• Es capaz de identificar la hipótesis nula que se encuentra implícita en el problema.
• Conoce la distribución con la que va a trabajar.
• Calcula y comprende qué indica el estadístico.
• Calcula y comprende los grados de libertad.
• Utiliza la tabla de probabilidad de dicha distribución para determinar la probabilidad y puede interpretarla como una medida de ocurrencia, por medio de la cual da solución al problema.</td>
<td>• Comprende la significancia.
• Es capaz de identificar el valor teórico del estadístico, con respecto a cierta P y gl. Y lo compara contra el valor del estadístico calculado.
• Es capaz de aceptar o rechazar la hipótesis nula (y comprende lo que significa) bajo el contraste con un límite preestablecido como desviación significativa.
• Logra argumentar, con base en la significancia por qué acepta o rechaza la hipótesis nula.
• Es capaz de conectar los resultados de la prueba con el contexto del problema.</td>
<td>• Comprende cuándo se comete el error tipo I y el tipo II y la probabilidad de cometerlos
$P\begin{bmatrix} Rechazar\ H_0 \vert H_0 \\ es\ verdadera \end{bmatrix} = \alpha$
$P\begin{bmatrix} No\ rechazar\ H_0 \vert H_0 \\ es\ falsa \end{bmatrix} = \beta$
• Es capaz de calcular la probabilidad de tomar la decisión correcta cuando H_0 es verdadera y cuando H_0 es falsa.
• Comprende las relaciones entre el Error tipo I y el Error tipo II.
• Comprende la potencia de la prueba y es capaz de calcularla $P\begin{bmatrix} Decidir\ H_1 \vert H_1 \\ es\ verdadera \end{bmatrix} = 1 - \beta.$
• Puede argumentar sobre la validez de la inferencia realizada.</td>
</tr>
</table>

PARTICULAR GENERAL

6.5 Reflexiones finales

En este capítulo, realizamos una caracterización del razonamiento inferencial que evidencian profesores de matemáticas, cuando realizan prácticas para resolver problemas sobre los estadísticos Chi-cuadrada y t-Student, tanto para identificar los objetos matemáticos primarios (elementos lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos) que los profesores utilizan en sus prácticas, y si éstos se pueden asociar a los distintos niveles de razonamiento inferencial, así como para ilustrar empíricamente si los niveles de razonamiento inferencial sobre sobre los estadísticos Chi-cuadrada y t-Student, propuestos a nivel teórico, se pueden considerar como predictores del razonamiento inferencial que los profesores de matemáticas exhiben en sus prácticas sobre dichos estadísticos.

Como pudimos observar en las prácticas que desarrollaron los profesores para resolver problemas sobre los estadísticos Chi-cuadrada y t-Student, se identificaron objetos matemáticos primarios, y sus procesos matemáticos, que se encuentran asociados a los criterios o elementos de los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student. Además, queremos destacar aspectos que hemos observado en las prácticas de los profesores y que se han reportado en la literatura de educación estadística como errores frecuentes y dificultades en inferencia estadística, y cómo los niveles de razonamiento inferencial propuestos podrían ayudar con dichas dificultades.

En las prácticas de los profesores en ejercicio (G3 yG4) y de los profesores en formación del G2, sobre los estadísticos Chi-cuadrada y t-Student, observamos que no plantearon las hipótesis nula y alternativa, aspecto que es fundamental en las pruebas de hipótesis y que se ha reportado como una de las principales dificultades que presentan los estudiantes y profesores en inferencia estadística (e.g., Vallecillos, 1997; Sotos et al., 2007; Batanero, 2013; López-Martín, Batanero y Gea, 2019). Es por ello, que en los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y estadístico t-Student (Capítulo 5) se plantea trabajar en tres momentos con las hipótesis, en los *Niveles 2, 3 y 4* respectivamente: (1) en un primer momento, se identifica la hipótesis que se encuentra implícita en el problema y se hace explicita por medio de un lenguaje natural; (2) en un segundo momento, se plantean las hipótesis nula y alternativa en lenguaje natural; (3) y en un tercer momento, se trabaja con las hipótesis nula y alternativa en lenguaje simbólico, que es como solemos encontrarlas en los estudios estadísticos.

Transitar por estos tres momentos de las hipótesis podría ayudar a los estudiantes y profesores a comprender la naturaleza y características de las hipótesis de una forma progresiva.

Respecto a las hipótesis, también resaltamos que en las prácticas de los profesores en formación sobre la t-Student se identificaron ambas hipótesis planteadas en lenguaje natural (e.g., PF8 y PF9), en dichas prácticas se identificaron dificultades que presentan los profesores, como en la interpretación de los resultados, el nivel de significancia o el reconocimiento del tipo de datos. Estas dificultades brindan una pauta para percatarnos tanto de la necesidad de una comprensión más profunda y holística de los diversos conceptos y propiedades que se encuentran involucradas en las pruebas de hipótesis, como de la presencia de un nivel de razonamiento preformal.

Por otro lado, en las prácticas de los profesores sobre el estadístico Chi-cuadrada, usualmente utilizaron el criterio de decisión del valor crítico (*Nivel 4*), pero evidenciaron dificultad con el valor-p; esta dificultad la ejemplificamos con la práctica del PF11 (Figura 6.13), quien indica que calculará el valor-p con la ayuda del software R, pero obtiene el valor del estadístico teórico (valor crítico), y no se percata de que este valor corresponde a un punto y no al área bajo la curva. Asimismo, en las prácticas sobre el estadístico t-Student también identificamos dificultades para trabajar con el valor-p (e.g., PF8 y PF2). Cuando el PF8 (Figura 6.26) realiza la conclusión de la actividad, indica que como alfa menor que 1.39 no puede rechazar la hipótesis nula. Como podemos ver, en estos casos con los que hemos ejemplificado, los profesores parecen comprender erróneamente al valor-p; consideramos que no son simples equivocaciones, pues si los profesores realmente comprendieran esta noción entenderían que están trabajando con un valor de probabilidad (que no puede ser mayor a uno), lo podrían asociar con el área bajo la curva de la distribución y no con un punto.

No obstante, el cálculo e interpretación del valor-p es otra de las principales dificultades en inferencia que se ha reportado en diversas investigaciones (e.g., Inzunza y Jiménez, 2013; Biehler, Frischemeier y Podworny, 2015; López-Martín, Batanero y Gea, 2019). En este sentido, en nuestra propuesta de niveles de razonamiento inferencial se propone trabajar progresivamente el valor-p en tres momentos, en los *Niveles 2, 3 y 4* respectivamente: (1)

trabajar con la acepción tradicional de probabilidad, (2) relacionar la noción de significancia con la probabilidad y una regla intuitiva para la toma de decisión, y (3) usar el criterio de decisión con el nivel de significancia y el valor-p. Esta propuesta de progresión podría ayudar en la comprensión tanto del valor-p como de las nociones relacionadas, por ejemplo, el nivel de significancia.

Igualmente, destacamos que no fue posible observar en las prácticas de los profesores participantes la probabilidad de cometer los errores tipo I y tipo II y, la potencia de la prueba. Yates (1951), ya nos venía advirtiendo que esto sucedía con los científicos y en los artículos de revistas especializadas, pues, así como en las prácticas de los profesores no nos fue posible identificar objetos primarios que den cuenta del uso de dichas nociones, en aquella época los científicos se centraban en el resultado de las pruebas de hipótesis y dejaban de lado la magnitud del efecto, la potencia de la prueba y los errores. La progresividad sugerida dentro del *Nivel 4*, manifiesta que los criterios sobre los errores y potencia de prueba presentan una complejidad mayor a la toma de decisión basada en el criterio de decisión (valor-p o valor crítico), esta complejidad puede ser la causa de que en las prácticas analizadas no se hayan identificado elementos relacionados con ellas. Sin embargo, no podemos dejar de lado la potencia de la prueba o la probabilidad de cometer los errores tipo I y tipo II, pues hacerlo, nos llevaría a una visión determinista de la inferencia o de las pruebas de hipótesis que estamos realizando (Batanero, 2018).

En definitiva, con base en los análisis y lo que hemos discutido a lo largo de este Capítulo, consideramos que los niveles de razonamiento inferencial pueden ser utilizados como predictores del razonamiento inferencial en diversos momentos de la enseñanza.

Asimismo, una vez que presentamos las actividades del instrumento de indagación y mostramos y analizamos el tipo de prácticas que realizaron los profesores para dar solución a las actividades, podemos hablar de confiabilidad, en el sentido de estabilidad de las observaciones, es decir, nos referimos a la consistencia que obtuvimos en las prácticas que desarrollaron los profesores para resolver las actividades sobre los estadísticos Chi-cuadrada y t-Student; y también de confiabilidad inter-evaluador ya que varios investigadores llegamos a las mismas interpretaciones sobre dichas prácticas.

También, en este capítulo, presentamos una propuesta general de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial, formulada a partir de los puntos de encuentro entre los niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student y la caracterización de las prácticas desarrolladas por los profesores en ejercicio y en formación. Consideramos que esta propuesta de niveles de razonamiento inferencial, que se realizó en términos más generales, es una primera aproximación a una propuesta de niveles de razonamiento inferencial, los cuales podrán ser ampliados o complementados, por ejemplo, con el estudio de otros estadísticos como el z y F y de las pruebas de hipótesis a los intervalos de confianza, así como de otras nociones clave de la inferencia.

CAPÍTULO 7

Conclusiones

"Todo el mundo sabe que en la investigación no hay respuestas finales, solo ideas que permiten formular nuevas preguntas". (Salvador Luria)

INTRODUCCIÓN

En este Capítulo presentamos los resultados más relevantes que hemos obtenido a lo largo de esta investigación. Asimismo, estos resultados nos permitirán evidenciar la respuesta a la pregunta de investigación y el cumplimiento del objetivo general. En la primera sección, realizamos una síntesis de la problemática en la cual se enmarca la investigación, en la segunda sección, presentamos los objetivos específicos y cómo es que se ha dado cumplimiento a cada uno de ellos; mientras que, en la tercera sección, se resumen las principales contribuciones de esta investigación. En la cuarta sección, exponemos las limitaciones y posibles líneas de continuidad para futuros estudios.

7.1 Resumen del problema de investigación

Como hemos señalado anteriormente, principalmente en el Capítulo 1, una problemática que ha preocupado a la comunidad de Educación Estadística, refiere a los procesos de enseñanza y aprendizaje de la inferencia estadística. En diversas investigaciones se han identificado dificultades que presentan tanto estudiantes como profesores al trabajar inferencia estadística, por ejemplo, sobre la comprensión del nivel de significancia, los errores tipo I y tipo II, el planteamiento de las hipótesis estadísticas, las distribuciones muestrales, la relación entre el estadístico y el parámetro, y realizar interpretaciones incorrectas de los resultados (e.g., Watson, 2004; Bakker y Gravemeijer, 2004; Carver, 2006; Sotos, Vanhoof, Van den Noortgate y Onghena, 2007; Garfield y Ben-Zvi, 2008; Harradine, Batanero y Rossman, 2011; Batanero, Vera y Díaz, 2012).

Para afrontar estas dificultades, han surgido propuestas sobre cómo promover el razonamiento inferencial (RI); por una parte, se encuentran aquellas propuestas que se enfocan en cómo aproximarnos a la inferencia desde edades tempranas bajo una perspectiva informal, denominada razonamiento inferencial informal -RII- (e.g., Zieffler, delMas, Garfield y Reading, 2008; Makar y Rubin, 2009; Doerr, delMas y Makar, 2017), y por otra parte, para promover un razonamiento inferencial formal sobre las bases del RII (e.g., Jacob y Doerr, 2014; Pfannkuch, Arnold y Wild, 2015; Makar y Rubin, 2018).

Los resultados y reflexiones de las investigaciones de Educación Estadística, apuntan al desarrollo temprano del razonamiento inferencial informal de los estudiantes y la necesidad de promover progresivamente un razonamiento inferencial formal. No obstante, a pesar de los avances sobre cómo promover el RII y sobre cómo promover progresivamente el RIF, aún es necesario contar con propuestas que permitan explorar y desarrollar progresivamente (del RII al RIF) el razonamiento inferencial de estudiantes y profesores. Consideramos que en esta búsqueda de la integración del RII al RIF, los estadístico Chi-cuadrada y t-Student nos permitirán brindar una trayectoria desde lo intuitivo a lo formal. Iniciamos cuestionándonos ¿cuáles son los significados de los estadísticos Chi-cuadrada y t-Student?, ¿tales significados de los estadísticos Chi-cuadrada y t-Student están relacionados con niveles progresivos de razonamiento inferencial?, ¿existen vínculos entre los diversos niveles sobre el estadístico Chi-cuadrada y t-Student?, ¿cuáles son los elementos de los niveles que los profesores y estudiantes utilizan en sus prácticas matemáticas?

Conscientes de la problemática, con esta tesis nos planteamos realizar una propuesta de niveles progresivos de razonamiento inferencial, de lo informal a lo formal, sobre los estadísticos Chi-cuadrada y t-Student, tomando como base tanto la riqueza matemática que se recupera de un estudio histórico-epistemológico sobre estos estadísticos, como los aportes de la literatura de Educación Estadística sobre el razonamiento inferencial. En este sentido, a partir de la problemática, formulamos la siguiente pregunta de investigación, la cual dio origen y orientó a esta investigación (sección 2.2):

> ***¿Cómo los significados parciales de los estadísticos Chi-cuadrada y t-Student, y sus articulaciones, pueden contribuir a la propuesta de niveles de razonamiento inferencial, de lo informal (RII) a lo formal (RIF), sobre dichos estadísticos?***

Por su parte, esta pregunta da origen al planteamiento del siguiente objetivo general de investigación (sección 2.2):

> ***Caracterizar niveles progresivos de razonamiento inferencial (de lo informal a lo formal) sobre los estadísticos Chi-cuadrada y t-Student, a partir de la riqueza matemática subyacente a los significados parciales que se le han conferido a tales estadísticos en su evolución histórico-epistemológica.***

Para logar el objetivo general de la investigación y en consecuencia dar respuesta a la pregunta de investigación fue necesario plantearnos una serie de objetivos específicos (OE). En la siguiente sección presentamos los resultados de cada uno de los objetivos específicos.

7.2 Objetivos específicos y resultados de la investigación

En esta sección presentamos los objetivos específicos y los resultados más relevantes de nuestra investigación, para dar cuenta de cómo es que se ha dado cumplimiento a cada uno de dichos objetivos.

7.2.1 Objetivo Específico 1

El primer objetivo específico (OE1) que nos propusimos atendía a la necesidad de conocer a los objetos matemáticos con los cuales trabajamos en esta investigación, es decir, entender ¿qué es el estadístico Chi-cuadrada?, y ¿qué es el estadístico t-Student?

> ***OE1: Reconstruir el significado holístico de los estadísticos Chi-cuadrada y t-Student, con base en un estudio histórico-epistemológico, determinando los pares <Prácticas matemáticas/estadísticas, Configuraciones de objetos matemáticos primarios>.***

Para lograr este objetivo y como primera fase de nuestra investigación, realizamos un estudio de tipo histórico-epistemológico sobre el estadístico Chi-cuadrada (Capítulo 3) y sobre el estadístico t-Student (Capítulo 4). Para determinar el significado holístico o global de cada uno de estos estadísticos, primeramente, se revisaron principalmente fuentes primarias, en las cuales identificamos las grandes problemáticas (situaciones/problema) que resultaron claves para el surgimiento y desarrollo del objeto matemático (Chi-cuadrada y t-Student) y cómo estas problemáticas se solucionaron.

Una vez identificadas las problemáticas, sobre los estadísticos Chi-cuadrada y t-Student, procedimos a realizar la descripción de las problemáticas y de las prácticas matemáticas desarrolladas para dar solución a éstas, estos desarrollos se pueden ver en las secciones 3.1 del Capítulo 3 y 4.1 del Capítulo 4. Para realizar esta descripción recurrimos al primer nivel de análisis del EOS con la noción de *práctica matemática* (ver sección 2.1).

Posteriormente, por medio de la noción de *configuración ontosemiótica* (epistémica), la cual corresponde a un segundo nivel de análisis, realizamos una caracterización de los los objetos matemáticos primarios –elementos lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos– que se activaron en las prácticas desarrolldas; esto se puede encontrar en las secciones 3.2 y 4.2. De esta forma, cada configuración nos ayudó a determinar significados parciales y, a su vez, estos significados parciales conforman el significado global para los estadísticos Chi-cuadrada y t-Student.

En este sentido, para el estadístico Chi-cuadrada se identificaron cuatro grandes significados y doce significados parciales, los cuales conforman el significado global de este estadístico (sección 3.2). Mientras que, para el estadístico t-Student el significado global lo conforman los tres grandes significados y doce significados parciales (sección 4.2). Cada uno de estos significados parciales, lleva implícita una configuración epistémica de objetos matemáticos primarios, y procesos asociados. Las Figuras 7.1 y 7.2 dan cuenta de las relaciones entre los grandes significados y los significados parciales para ambos estadísticos (también puede ver las síntesis de los significados en los capítulos 3 y 4).

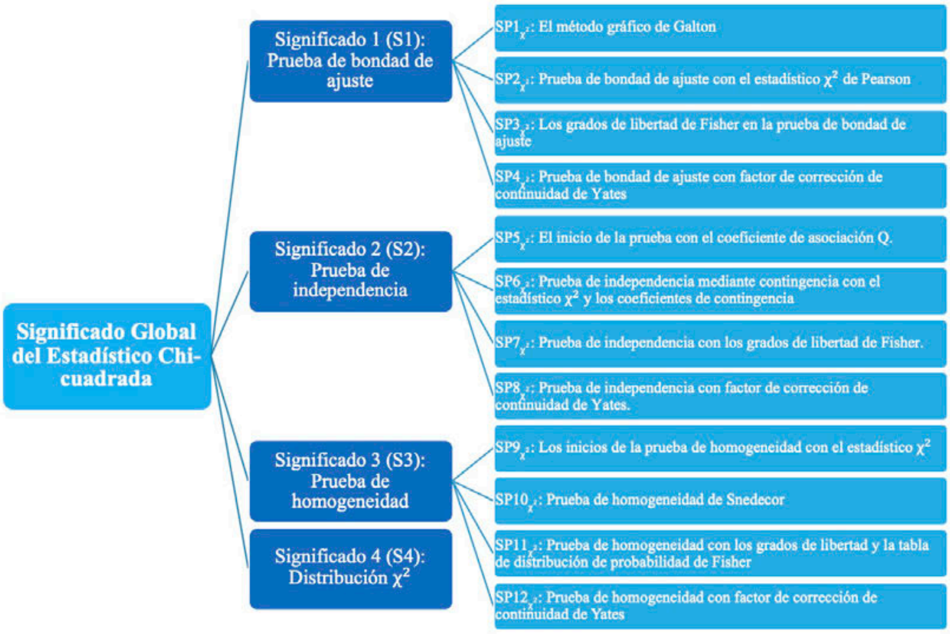


Figura 7.1. Significado global del Estadístico Chi-cuadrada

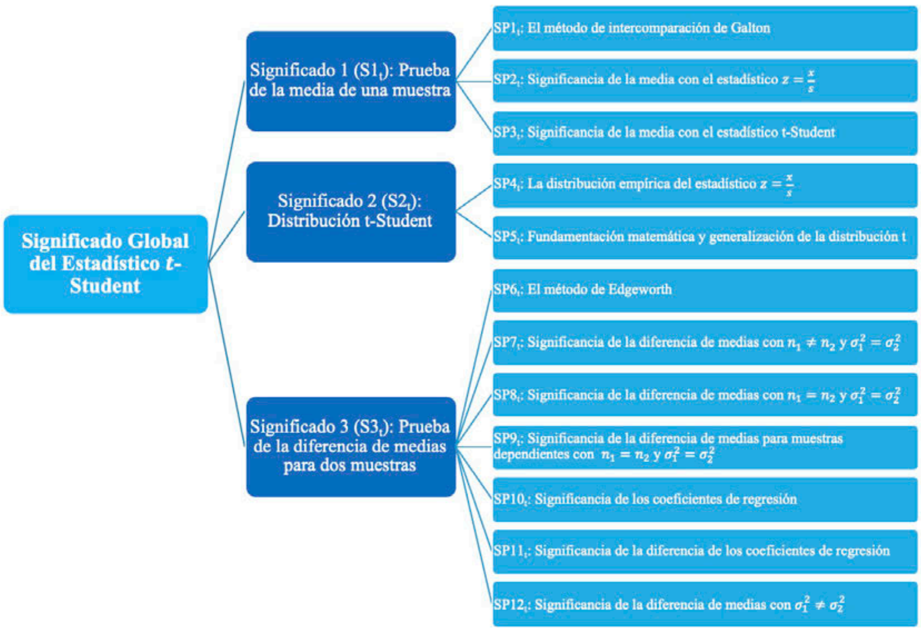


Figura 7.2. Significado global del Estadístico t-Student

El cumplimiento del primer objetivo especifico era crucial para nuestra investigación, pues por medio del estudio histórico-epistemológico logramos acceder a la riqueza matemática subyacente de ambos estadísticos.

7.2.2 Objetivo Específico 2

El segundo objetivo específico (OE2) que nos propusimos en esta investigación, atendía a la necesidad de conocer las relaciones entre los significados parciales de los estadísticos Chi-cuadrada y t-Student y, la literatura científica sobre estos estadísticos, el razonamiento inferencial informal y formal.

> ***OE2: Establecer las posibles relaciones entre los significados parciales, esto es entre los pares <Prácticas matemáticas/estadísticas, Configuraciones de objetos matemáticos primarios>, de los estadísticos Chi-cuadrada y t-Student, para determinar los elementos que podrían constituir una propuesta de niveles de razonamiento inferencial***

Para lograr este objetivo y como segunda fase de esta investigación, en primer lugar, buscamos relaciones entre los significados parciales de los estadísticos, por ejemplo, identificamos relación entre los SP1χ^2 y SP1t, lo cual se encuentra plasmado en el desarrollo de los Capítulos 3 y 4, en las configuraciones ontosemióticas epistémicas. Además, se identificaron objetos matemáticos primarios que podrían constituir posibles 'caminos o trayectorias' para cada significado. Si tomamos como ejemplo la prueba de bondad de ajuste (S1χ^2) del estadístico Chi-cuadrada, a partir del estudio histórico-epistemológico notamos que este tipo de ajuste podría valorarse en un inicio con métodos más intuitivos (primer etapa de la trayectoria), como lo es el método de intercomparación y el método gráfico (Galton, 1875; 1885) del SP1χ^2, de este significado parcial destacamos los objetos matemáticos primarios CD2, CD3, CD4, CD6, CD7, PP2, PP3, PP4, PP5, PP6 y PP7. Siguiendo con la trayectoria que nos indica el estudio histórico, en una segunda etapa se podría valorar la bondad de ajuste mediante la interpretación de la probabilidad, obtenida a partir del valor estadístico Chi-cuadrada calculado, como una medida de ocurrencia de un complejo sistema de n errores que ocurren con una frecuencia tan grande o más grande que la del sistema observado, además se podría identificar y plantear con lenguaje natural la hipótesis que se encuentra implícita en el problema; en este caso intervienen objetos matemáticos primarios del SP2χ^2 y SP3χ^2, de los cuales se destacan CD3, CD4, CD5, CD6, PP3, PP4 y PP7 del SP2χ^2; CD7, CD8, PP3 y PP6 del SP3χ^2. En una tercera etapa, se podría introducir la limitación que tiene el uso del estadístico Chi-cuadrada cuando la muestra que analizamos tiene frecuencias esperadas inferiores a cinco y el uso que tiene entonces el estadístico Chi-cuadrada con factor de continuidad; además, se podría abordar la significancia como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria y aplicar una regla de decisión para concluir sobre si los datos observados se ajustan a la distribución teórica esperada. Para llevar a cabo lo planteado en esta tercera etapa destacamos la relevancia de los objetos matemáticos primarios CD3, CD4, CD5, CD6, PP3 y PP4 del SP2χ^2; CD7, CD8, PP3 y PP6 del SP3χ^2; y CD6, CD8 y PP6 del SP4χ^2. En la cuarta etapa de la trayectoria, se podría valorar la bondad de ajuste con base en las técnicas estadísticas de la metodología de las pruebas de hipótesis, para lo cual podría plantear las hipótesis estadísticas, utilizar el estadístico Chi-cuadrada, con el factor de corrección de continuidad en caso de ser necesario, encontrar el valor de la probabilidad asociado al valor del estadístico y utilizar la regla de decisión (valor-p o valor crítico) para poder concluir si las frecuencias observadas se distribuyen de acuerdo a lo esperado; en esta etapa de la trayectoria destacamos los objetos matemáticos primarios CD3, CD4, CD5, CD6, PP3 y PP4 del SP2χ^2; CD7, CD8, PP3 y PP6 del SP3χ^2; y CD6, CD8, PP6 y PP7 del SP4χ^2 (ver sección 3.2).

Lo que llamamos caminos o trayectorias y que recientemente ejemplificamos con la prueba de bondad de ajuste del estadístico Chi-cuadrada, constituyeron un insumo significativo de cara a la construcción de nuestra propuesta de niveles para el estadístico Chi-cuadrada y t-Student; de hecho, se utilizan en el Capítulo 5 para el logro del Objetivo Específico 3.

En segundo lugar, se realizó una revisión de la literatura de científica de Educación Estadística sobre RII, RIF y sobre los estadísticos Chi-cuadrada y t-Student. En esta revisión se idenctificaron diferentes posturas sobre el RII y RIF, así como errores y dificultades, y propuestas de actividades que se han reportado sobre inferencia estadística y también sobre dichos estadísticos. Los resultados de esta revisión se pueden ver en el Capítulo 1.

Además, recurrimos a diversas posturas que existen en Educación Estadística sobre razonamiento para tener un antecedente (ver Capítulo 1) y partimos de la postura que se tiene en el EOS sobre razonamiento para intentar dar una definición. En concordancia con los planteamientos de Aké (2013), para el algebra, y de Molina (2019), para la geometría, recurrimos a una visión pragmatista sobre la construcción del conocimiento matemático y asumimos el razonamiento como un “macro proceso social y epistémico”, que involucra poner en juego tanto los objetos matemáticos primarios como los procesos matemáticos para la solución de una situación-problema. Mayores detalles sobre nuestra postura del razonamiento se encuentran en el Capítulo 2. Entonces, estos caminos o trayectorias identificadas, la riqueza matemática que surge en la historia, las sugerencias de la literatura y la concepción sobre razonamiento que logramos adoptar, nos aportaron los elementos necesarios para la construcción de los niveles.

Con base es todo lo anterior, se logró evidenciar que, para efectos de nuestra propuesta, el razonamiento inferencial formal e informal no se diferían en términos de los contenidos (Estadística Descriptiva o Estadística Inferencial), sino más bien en el grado de formalidad con la que se estudian las nociones que permiten la conceptualización/comprensión de los estadísticos Chi-cuadrada y t-Student (i.e., de lo intuitivo a lo formal).

7.2.3 Objetivo Específico 3

El tercer objetivo específico (OE3) atendió a un aspecto esencial en esta investigación, nos referimos a realizar una propuesta de niveles progresivos de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student.

> ***OE3: Determinar niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student, con base en la postura adoptada en este estudio sobre el razonamiento inferencial (informal y formal), y las relaciones entre los significados parciales de ambos estadísticos y los aportes sobre el tema de la literatura científica.***

A partir de la caracterización de los significados de los estadísticos Chi-cuadrada y t-Student, un aspecto relevante era identificar elementos o características para una transición continua del RII al RIF, entonces, para la contrucción de los niveles de razonamiento inferencial, tomamos como punto de partida las trayectorias identificadas para cada significado y vinculamos progresivamente el desarrollo de los significados de los estadísticos Chi-cuadrada y t-Student.

En este sentido, con base en los criterios epistémicos, identificados con el estudio de tipo histórico-epistemológico sobre estos estadísticos, y en la literatura de investigación desarrollada sobre razonamiento inferencial (informal, formal y de progresión), resultado del OE2, construimos una propuesta de niveles progresivos, de lo informal a lo formal, de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. Las propuestas de niveles desarrolladas a lo largo del Capítulo 5, dan cuenta del logro de este objetivo específico. Estas propuestas cuentan con cuatro niveles, el primero de ellos vinculado con el razonamiento inferencial informal, los niveles dos y tres con un razonamiento inferencial preformal, mientras que el cuarto nivel se vincula con el razonamiento inferencial formal.

Asimismo, en el Capítulo 6 (sección 6.2.4) se encuentran indicadores adicionales, producto de nuevos hallazgos a partir del estudio empírico; en la siguiente sección daremos más detalles de este estudio. Adicionalmente, se identificaron los puntos de encuentro o conexiones entre los niveles para la Chi-cuadrada y para la t-Student, y a partir de ellos, la posibilidad de una propuesta de niveles de forma general. Realizamos una primera aproximación a la propuesta general de niveles de razonamiento inferencial, la cual se refinó después de la caracterización de las prácticas de profesores sobre los estadísticos Chi-cuadrada y t-Student.

A continuación, presentamos las características clave de los niveles de razonamiento inferencial.

Nivel 1

Los indicadores de este nivel corresponden a un razonamiento inferencial informal y van más allá de simples interpretaciones de los gráficos o cálculos de medidas estadísticas y de probabilidad, pues dichos indicadores están enfocados en que el estudiante razone inferencialmente y realice conjeturas a partir de la visualización de las características de las gráficas (e.g., forma, dispersión, amplitud de los cuartiles, mediana, sesgo, proporción) y/o a partir de la variación interna de un conjunto de datos, las medidas de tendencia central, de dispersión y de posición y, la probabilidad condicional.

Nivel 2

En este nivel, se encuentran indicadores que dan cuenta de un razonamiento de tipo preformal, y se pueden identificar rasgos del RII tales como la forma en que se utiliza (define) la probabilidad, como una medida de certidumbre asociada a un evento, sin compararla con un nivel de significancia o un límite preestablecido, y el reconocimiento de la hipótesis de investigación en el problema y hacerla explícita por medio del lenguaje natural. Además, en este caso, el estudiante se encuentra por primera vez con el estadístico y los contextos en que se utiliza.

Nivel 3

Los indicadores del tercer nivel se consideran preformales, pero con un mayor grado de formalidad que en el segundo nivel, esto se debe a que se tiene una visión más amplia de las posibilidades para trabajar con las pruebas Chi-cuadrada y t-Student, hacia la generalización de estas pruebas y las restricciones que implican. Los aspectos que marcan cierto grado preformal son, por ejemplo, que las hipótesis nula y alternativa son planteadas en lenguaje natural, y la introducción de la significancia, como un límite significativo, y la introducción de la regla de decisión intuitiva, para rechazar o no rechazar la hipótesis nula.

Nivel 4

Mientras que el tercer elemento, corresponde al uso del error tipo I y tipo II, así como de la potencia de la prueba. En este nivel el estudiante puede auxiliase de gráficos para comprender las relaciones entre los errores tipo I y tipo II, por ejemplo, que cuanto mayor es α, menor es β. Además de calcular la potencia de la prueba, se promueve que el estudiante reconozca que la potencia está relacionada con el tamaño de la muestra y el nivel de significancia; y pueda utilizarla para argumentar sobre la prevalencia de su estimación en la población.

Los indicadores de este nivel corresponden a un razonamiento inferencial formal. En este nivel, se espera que el estudiante pueda tomar decisiones basado en las técnicas estadísticas de la metodología de las pruebas de hipótesis y sea capaz de analizar críticamente sus desarrollos estadísticos. Algunos aspectos que se destacan en este nivel son: (1) el planteamiento de las hipótesis nula y alternativa con lenguaje estadístico; (2) el valor-p y valor crítico, las reglas de decisión, el nivel de significancia y el nivel de confianza, y cómo son utilizados para responder el problema; (3) el uso del error tipo I y tipo II, así como de la potencia de la prueba. Cabe destacar, que mas allá de los aspectos técnicos, consideramos esencial el razonamiento que hay detrás de cada uno de estos aspectos de las pruebas de hipótesis.

7.2.4 Objetivo Específico 4

Una vez que realizamos las propuestas ‘teóricas’ de niveles sobre los estadísticos Chi-cuadrada y t-Student, era necesario realizar un estudio que nos permitiera ‘probar empíricamente’ los niveles, para ello se propuso el siguiente objetivo específico:

> ***OE4: Caracterizar las prácticas matemáticas de profesores y futuros profesores de educación media, determinando los niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student que movilizan con tales prácticas.***

Para lograr este objetivo específico y como última fase de esta investigación, buscamos caracterizar las prácticas de profesores en formación y en ejercicio de matemáticas. Para realizar dicha caracterización, primeramente, diseñamos una serie de actividades sobre los estadísticos Chi-cuadrada y t-Student que atendieran a los siguientes criterios:

1. Las actividades se pueden resolver con prácticas que admiten rasgos de alguno de los cuatro niveles de razonamiento inferencial para el estadístico Chi-cuadrada o t-Student.
2. Considerar una visión integral de los estadísticos Chi-cuadrada y t-Student, esto se debe a las particularidades de los campos de problemas o distintos tipos de pruebas en los que intervienen los estadísticos.
3. Aspectos generales como el uso de contextos accesibles e interesantes y uso de diversas representaciones (lenguaje natural, tablas, figuras).

Una vez que contamos con una serie de actividades para los estadísticos Chi-cuadrada y t-Student, procedimos a validarlas. Para la validación, primero recurrimos a una validación por contenido; en la sección 6.1.2 del Capítulo 6 mostramos un ejemplo de cómo las actividades cubren de forma justa y completa el razonamiento inferencial, informal-preformal-formal, sobre los estadísticos Chi-cuadrada y t-Student. Posteriormente, realizamos una validación de metodología ya que, como se muestra a lo largo del Capítulo 6, fue posible aplicar las actividades sobre estos estadísticos a cuatro grupos de profesores utilizando esquemas metodológicos similares en tres grupos (G1, G3 y G4) y con un esquema diferente en uno de los grupos (G2), por la particularidad del taller.

Una vez que, en el marco de talleres de razonamiento estadístico, aplicamos las actividades a profesores en formación y en ejercicio, y seleccionamos algunas prácticas que realizaron los profesores, pues las consideramos representativas de todas las prácticas realizadas a propósito de la resolución de cada una de las actividades; estás prácticas matemáticas se pueden observar en los desarrollos del Capítulo 6. También, en el Capítulo 6 (secciones 6.2 y 6.3), describimos dichas prácticas matemáticas e identificamos los objetos matemáticos primarios (elementos lingüísticos, conceptos/definiciones, propiedades/proposiciones, procedimientos y argumentos) que utilizan en sus prácticas los profesores y cómo dichos objetos matemáticos permitieron asociarlos a distintos niveles de razonamiento inferencial. Además, realizamos una discusión, después de cada práctica, de cómo éstas podrían ser más robustas dentro del mismo nivel o transitar al siguiente nivel, y en la sección de reflexiones finales del Capítulo 6 (sección 6.4) discutimos cómo los niveles de razonamiento inferencial podrían ayudar con algunos errores y dificultades que observamos en distintas actividades de ambos estadísticos, por ejemplo, con la comprensión de las hipótesis estadísticas, el valor-p, el nivel de significancia. A continuación brindaremos detalles sobre cómo la propuesta de niveles de razonamiento inferencial podría ayudar con éstas dificultades reportadas en la literatura científica y observadas en las prácticas de los estudiantes.

Planteamiento de las hipótesis

Nuestra propuesta de niveles de razonamiento inferencial sugiere trabajar en tres momentos con las hipótesis, en los *Niveles 2, 3 y 4* respectivamente: (1) en un primer momento, se identifica la hipótesis que se encuentra implícita en el problema y se hace explicita por medio de un lenguaje natural; (2) en un segundo momento, se plantean las hipótesis nula y alternativa en lenguaje natural; (3) y en un tercer momento, se trabaja con las hipótesis nula y alternativa en lenguaje simbólico. Consideramos que transitar por estos tres momentos de las hipótesis podría ayudar a los estudiantes a comprender la naturaleza de las hipótesis.

Significancia

En la propuesta de niveles se plantea trabajar la noción de significancia en dos momentos, en los *Niveles 3 y 4* respectivamente: (1) en el nivel tres, se trabaja con la significancia como indicativo del nivel en el que la posibilidad del efecto debe recibir una consideración seria, es decir, como un límite significativo; (2) mientras que en el nivel 4, se propone trabajar con el nivel de significancia, el cual, además de proporcionarle un punto crítico o 'límite' para la toma de decisión, se puede interpretar como probabilidad de concluir que existe una desviación o diferencia cuando en realidad no existe, en otras palabras, es la probabilidad de equivocarnos que estamos dispuestos a asumir si rechazamos la hipótesis nula.

Valor-p

En la propuesta de niveles de razonamiento inferencial se propone trabajar progresivamente el valor-p en tres momentos, en los *Niveles 2, 3 y 4* respectivamente: (1) en un primer momento, trabajar con la probabilidad como una medida de ocurrencia en términos del contexto, por ejemplo, en cuantos casos sería posible observar un conjunto de datos como los proporcionados en la situación problema y que las desviaciones de lo esperado se deban al azar; (2) en un segundo momento, se sugiere relacionar la noción de probabilidad con la de significancia (como límite) y una regla intuitiva para la toma de decisión; (3) en el tercer momento, se destaca el valor-p como como una medida (probabilidad) de que el valor del estadístico calculado sea posible dada la hipótesis nula y el uso que se le otorga junto con el nivel de significancia en la regla para la toma de decisión. Esta propuesta de progresión podría ayudar en la comprensión tanto del valor-p como de las nociones y aspectos relacionados, por ejemplo, el nivel de significancia y la regla para la toma de decisión.

Al mismo tiempo, a partir del estudio empírico nos percatamos que los niveles de razonamiento inferencial sobre sobre los estadísticos Chi-cuadrada y t-Student, propuestos a nivel teórico, efectivamente se pueden considerar como predictores del razonamiento inferencial que los profesores de matemáticas exhiben en sus prácticas sobre dichos estadísticos. Además, la experimentación permitió observar la importancia que tiene la probabilidad condicional en las prácticas que evidencian un razonamiento inferencial informal (nivel uno), por lo cual, en la sección de reflexiones y consideraciones sobre el estadístico Chi-cuadrada del Capítulo 6 (sección 6.2.4), se describen los criterios sobre probabilidad condicional que se incorporan a la propuesta de niveles de razonamiento inferencial sobre el estadístico Chi-cuadrada para fortalecer la propuesta de niveles. Es decir, que el estudio empírico permitió hacer algunos ajustes y refinamientos a la propuesta teórica de niveles de razonamiento inferencial.

Finalmente, en dicho Capítulo presentamos una propuesta general de niveles progresivos (Tabla 6.9), de lo informal a lo formal, de razonamiento inferencial. Esta propuesta general, se formuló a partir de los puntos de encuentro entre los niveles de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student y de la caracterización de las prácticas desarrolladas por los profesores.

7.3 Principales contribuciones de la tesis

Los resultados más relevantes de esta investigación son las *propuestas de niveles de razonamiento inferencial sobre los estadísticos Chi-cuadrada y t-Student*. En el Capítulo 5 presentamos las propuestas de niveles progresivos de razonamiento inferencial para los estadísticos Chi-cuadrada y t-Student. Estas propuestas están conformadas por cuatro niveles que van de lo informal a lo formal y en cada nivel se encuentran diversos criterios que vinculan los elementos epistémicos, en términos de objetos primarios y procesos, de los significados que se identificaron con los estudios de tipo histórico-epistemológico sobre estos estadísticos, y la investigación desarrollada sobre razonamiento inferencial (informal, formal y de progresión). Estas propuestas de niveles de razonamiento inferencial podrían constituir directrices para la gestión del aprendizaje de los estudiantes, el diseño de actividades y el análisis del currículo.

Para validar las propuestas teóricas de niveles de razonamiento para los estadísticos Chi-cuadrad y t-Student, así como complementarla, se llevó a cabo un estudio de corte empírico donde participaron profesores de Latinoamérica; este estudio empírico lo desarrollamos detalladamente en el Capítulo 6. A partir de las propuestas teóricas de niveles de razonamiento y de la caracterización de las practicas, en términos de objetos matemáticos primarios y niveles de razonamiento, desarrolladas por los profesores en ejercicio y en formación fue posible realizar el segundo aporte principal de esta investigación, la *propuesta general de niveles de razonamiento inferencial*. Esta propuesta general, no pretende ser definitiva, más bien es una primera aproximación que puede ser ampliada a los intervalos de confianza con investigaciones que recuperen la riqueza matemática subyacente a los significados parciales por medio de estudios histórico-epistemológicos. Consideramos que esta propuesta general de niveles de razonamiento inferencial podría ser utilizada para otras nociones, siempre y cuando se consideren las adecuaciones necesarias (matices epistemológicos) relativas a la noción que se pretenda estudiar.

Otra contribución la constituye el estudio de los significados de los estadísticos Chi-cuadrada y t-Student. Para realizar las propuestas de niveles de razonamiento inferencial, primeramente, fue necesario desarrollar los estudios histórico-epistemológicos de estos estadísticos para conocer cuáles eran sus significados. En este sentido, en los Capítulos 3 y 4 se describieron las prácticas que permitieron la emergencia y evolución de estas nociones y, además, se identificaron y caracterizaron los objetos matemáticos primarios que se activaron en dichas prácticas, dando cuenta así de las configuraciones epistémicas.

Entonces, los estudios histórico-epistemológico realizados, permiten ver la complejidad y riqueza matemática de los estadísticos Chi-cuadrada y t-Student. Esta complejidad, no se limita a una visión formalista, lo cual solemos encontrarnos en los primeros cursos universitarios, sino que también da cuenta de aspectos intuitivos y de la evolución de estos estadísticos. La riqueza matemática de los significados globales de estos estadísticos, recogida en los Capítulos 3 y 4, podría ser utilizada para desarrollar secuencias didácticas y análisis del currículo sobre estas nociones. Sin embargo, en esta investigación la utilizamos para la propuesta de niveles de razonamiento inferencial, la cual incluye los elementos de los significados.

7.4 Limitaciones y líneas de continuidad

Las limitaciones de esta investigación se enmarcan en el contexto sanitario, que desde inicios del año 2020 ha impactado al sector educativo, con el cierre de colegios y universidades. Nuestra investigación sufrió modificaciones a raíz del cierre de los establecimientos educativos, motivo por el cual se decidió realizar talleres de razonamiento estadístico en modalidad virtual, enfocados a profesores en ejercicio y en formación, en los cuales pudiéramos aplicar las actividades sobre los estadísticos Chi-cuadrada y t-Student y así, acceder a las prácticas que desarrollan los profesores cuando resuelven problemas sobre estos estadísticos. Sin embargo, este cambio, también trajo como ventaja contar con la participación de profesores de diversos países de Latinoamérica, concretamente de Argentina, Chile, Colombia, Costa Rica, Guatemala, México y Perú.

Desarrollar la aplicación de las actividades en modalidad virtual tuvo como limitante, por una parte, el tiempo acotado de las sesiones que fue previamente establecido por los organizadores de los eventos y, por otra parte, la interacción y reflexiones que se puede generar bajo las condiciones de los medios virtuales (modalidades sincrónica y asincrónica). Aunque, para esta investigación funcionaron las estrategias de interacción que establecimos bajo las modalidades sincrónicas y asincrónica, según fue el caso del taller, consideramos que para investigaciones orientadas al desarrollo del razomiento inferencial podría causar mayores dificultades.

A pesar de las contribuciones que resultaron de nuestra investigación, somos conscientes de que aún sigue abierta la discusión sobre cómo promover el razonamiento inferencial de forma progresiva y que apenas iniciamos un largo trayecto para aportar con propuestas que permitan explorar y desarrollar progresivamente (del RII al RIF) el razonamiento inferencial de estudiantes y profesores. A continuación, presentamos posibles líneas de continuidad de nuestra investigación:

- Determinar niveles de razonamiento inferencial para otros estadísticos, por ejemplo, los estadísticos z y F, a partir de la riqueza matemática que se recupera a través de estudios histórico-epistemológicos sobre dichos estadísticos y, la literatura científica de educación estadística. Estas propuestas también ayudarán a mejorar nuestra propuesta general de niveles de razonamiento inferencial.

- Los niveles de razonamiento inferencial se podrían ampliar de las pruebas de hipótesis a los intervalos de confianza para estos mismos estadísticos, sobre las bases de estudios similares al presentado en esta investigación.
- Diseño de secuencias didácticas basadas en los niveles de razonamiento inferencial para promover el desarrollo del razonamiento inferencial en los estudiantes de forma progresiva, desde elementos intuitivos hasta los formales.
- Estudio sobre las prácticas matemáticas propuestas en el curriculo <entendido como la dupla planes y programas y libros de texto> de enseñanza básica, media y universitario para determinar los niveles de razonamiento inferencial que se pretende promover en cada nivel educativo.

7.5 Algunas contribuciones a la comunidad académica

Lugo-Armenta, J. G. & Pino-Fan L. R. (2021). Niveles de Razonamiento Inferencial para el Estadístico t-Student. *Bolema: Boletim de Educação Matemática*. (En Prensa).

Lugo-Armenta, J. G., Pino-Fan, L. R., & Ruiz, B. R. (2021). Chi-square reference meanings: a historical-epistemological overview. *Revemop - Revista de Educação Matemática de Ouro Preto*. *3*, 1-33. https://doi.org/10.33532/revemop.e202108

Castro, W. F., Pino-Fan, L. R., Lugo-Armenta, J. G., Toro, J. A., & Retamal, S. (2020). A Mathematics Education research agenda in Latin America motivated by coronavirus pandemic. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(12), em1919. https://doi.org/10.29333/ejmste/9277

Carrera, B., Pino-Fan, L. R., Alvarado, H., Lugo-Armenta, J. G. (2020). Practices on the discrete random variable proposed in the mathematics chilean curriculum of secondary education. In M. Inprasitha, N. Changsri, N. Boonsena (Eds.), *Interim Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (pp. 72-80). Khon Kaen, Thailand: PME.

Lugo, J. G., García-García J., Ruíz, B. R., (2019) Desarrollo de la Investigación en Razonamiento Inferencial Informal. *Revista Digital: Matemática, Educación E Internet*, *20*(1). https://doi.org/10.18845/rdmei.v20i1.4589

Lugo-Armenta, J. G. & Pino-Fan, L. R. (2021). An approach to inferential reasoning levels. *Educational Studies in Mathematics*. (En Revisión).

Lugo-Armenta, J. G. & Pino-Fan, L. R. (2021). Razonamiento Inferencial de Profesores de Matemáticas de Enseñanza Media sobre el Estadístico T-Student. *AIEM: Avances de Investigación en Educación Matemática*. (En Revisión).

Lugo-Armenta, J. G. & Pino-Fan, L. R. (2021). Inferential Reasoning of high school Mathematics Teachers about Chi-squared Statistic. *Journal of Mathematics Teacher Education*. (En Revisión).

Lugo-Armenta, J. G. & Pino-Fan, L. R. (2021). Development of inferential statistical reasoning: experiences in virtual contexts generated by the Covid-19 pandemic. *Education Sciences*. (En Revisión).

Lugo-Armenta, J. G. & Pino-Fan, L. R. (2021). Razonamiento Inferencial de Profesores de Enseñanza Media. *Reunión Latinoamericana de Matemática Educativa 34*. Quetzaltenango, Guatemala: RELME. (Ponencia Aceptada).

Lugo-Armenta, J. G. (2019). Una aproximación al Razonamiento Inferencial sobre los estadísticos Chi-cuadrada y t-Student. Conferencia especial presentada en las *XXIII Jornadas Nacionales de Educación Matemática* (p. 15). Puerto Montt, Chile: JNEM

Lugo-Armenta, J. G., Pino-Fan L. R. y Ruiz, B.R. (2019). Desarrollo Histórico del estadístico Chi-cuadrada. Ponencia presentada en las *XXIII Jornadas Nacionales de Educación Matemática* (pp. 44-45). Puerto Montt, Chile: JNEM

Lugo-Armenta, J. G., Carrera, B., y Pino-Fan, L. R. (2020). Razonamiento Estadístico en la Enseñanza Media. Taller invitado presentado en la 2da Jornadas Regionales sobre Educación Matemática

Lugo-Armenta, J. G. (2020). Razonamiento Inferencial: Directrices para su desarrollo. Taller invitado presentado en la *V Escuela de Verano Encuentro sobre Didáctica de la Estadística, la Probabilidad y el Análisis de Datos*. Cartago, Costa Rica: EDEPA.

Lugo-Armenta, J. G. (2019). Diseño de actividades de aprendizaje para el desarrollo del Razonamiento Inferencial Informal. Taller invitado presentado en la *IV Escuela de Verano Encuentro sobre Didáctica de la Estadística, la Probabilidad y el Análisis de Datos*. Turrialba, Costa Rica: EDEPA.

REFERENCIAS

Aké, L. P. (2013). *Evaluación y desarrollo del razonamiento algebraico elemental en maestros en formación* (Tesis de Doctorado). Universidad de Granada, España.

Alvarado, H. (2007). *Significados del teorema central del límite en la enseñanza de la estadística en ingeniería* (Tesis de Doctorado). Universidad de Granada, España.

Aridor, K., & Ben-Zvi, D. (2017). The co-emergence of aggregate and modelling reasoning. *Statistics Education Research Journal*, *16*(2).

Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: from hands-on to “movies”. *Journal of Statistics Education*, *19*(2).

Bakker, A., Ben-Zvi, D., & Makar, K. (2017). An inferentialist perspective on the coordination of actions and reasons involved in making a statistical inference. *Mathematics Education Research Journal*, *29*(4), 455-470. https://doi.org/10.1007/s13394-016-0187-x

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*(1-2), 5-26. https://doi.org/10.1080/10986065.2011.538293

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht: Kluwer Academic Publishers.

Bakker, A., Kent, P., Derry, J., Noss, R. y Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal*, *7*(2), 130-145.

Barnard, G.A. (1992). Introduction to Pearson (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. En S. Kotz and N.L. Johnson eds., *Breakthroughs in Statistics*, *Vol. II* (pp. 1-10). New York: Springer.

Batanero, C. (2004). Los retos de la cultura estadística. *Yupana. Revista de Educación Matemática de la UNL*, *1*, 27-36.

Batanero, C. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática*, *11*, 277-291.

Batanero, C. (2018). Treinta años de investigación didáctica sobre el análisis inferencial de datos. En A. Avila. (Ed.) *Rutas de la Educación Matemática* (pp. 186-199). México: Sociedad Mexicana de Investigación y divulgación de la Educación Matemática. Doi: 10.24844/EM

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam, The Netherlands: Sense Publishers. Doi: 10.1007/978-94-6300-624-8

Batanero, C., & Díaz, C. (2015). Aproximación informal al contraste de hipótesis. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, (2), 135-144.

Batanero, C., Díaz, C., Contreras, J. M., & Roa, R. (2013). El sentido estadístico y su desarrollo. *Números. Revista de Didáctica de las Matemáticas*, *83*, 7-18.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, *27*(2), 151-169.

Batanero, C., Vera, O.D., & Díaz, C. (2012). Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis. *Números. Revista de Didáctica de las Matemáticas*, *80*, 91-101.

Ben-Zvi, D. (2004). Reasoning about data analysis. En D. Ben-Zvi, J. Garfield (Eds). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 121-145). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2278-6_6

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July, 2006), Salvador, Bahia, Brazil.

Ben-Zvi, D., & Aridor-Berger, K. (2016). Children's wonder how to wander between data and context. En D. Ben-Zvi K. Makar (Eds.), *The teaching and learning of statistics* (pp. 25-36). Springer, Cham. https://doi.org/10.1007/978-3-319-23470-0_3

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - International Journal on Mathematics Education*, *44*(7), 913-925.

Ben-Zvi, D., & Garfield, J. B. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. En D. Ben-Zvi, J. Garfield (Eds). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3-16). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2278-6_1

Ben-Zvi, D., & Gil, E. (2010). The role of context in the development of students' informal inferential reasoning. *Proceedings of the Eighth International Conference on Teaching Statistics* (ICOTS8, July, 2010), Lijbljana, Slovenia.

Biehler, R., Frischemeier, D., & Podworny, S. (2015). Preservice teachers reasoning about uncertainty in the context of randomization tests. En A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning*, (pp. 129-162). Minneapolis, Minnesota: Catalyst Press.

Braham, H., & Ben-Zvi, D. (2017). Students' emergent articulations of statistical models and modeling in making informal statistical inferences. *Statistics Education Research Journal*, *16*(2).

Burrill, G., & Camden, M. (2005). *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable*. Voorburg, The Netherlands: International Statistical Institute. http://www. stat. auckland. ac. nz/~ iase/publications.

Büscher, C., & Schnell, S. (2017). Students'emergent modelling of statistical measures–a case study. *Statistics Education Research Journal*, *16*(2), 144-162.

Callingham, R., & Watson, J. M. (2017). The Development of Statistical Literacy at School. *Statistics Education Research Journal*, *16*(1), 181-201.

Cañadas, G., Batanero, C., Contreras, J. M., & Arteaga, P. (2011). Estrategias en el estudio de la asociación en tablas de contingencia por estudiantes de psicología. *Educación matemática*, *23*(2), 5-31.

Cañadas, G., Batanero, C., Díaz, C., & Gea, M. M. (2012). Comprensión del test chi-cuadrado por estudiantes de Psicología. En A. Estepa, Á. Contreras, J. Deulofeu, M. C. Penalva, F. J. García y L. Ordóñez (Eds.), *Investigación en Educación Matemática XVI* (pp. 153 - 163). Jaén: SEIEM.

Carver, R. (2006). Ambiguity intolerance: An impediment to inferential reasoning. En *ASA Section on Statistical Education*, (pp. 2248-2245).

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910677

Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi, J. Garfield (Eds). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295-323). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2278-6_13

Chance, B., & Rossman, A. (2006, July). Using simulation to teach and learn statistics. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July, 2006), Salvador, Bahia, Brazil.

Cohen, J. (1992). Cosas que he aprendido (hasta ahora). *Anales de Psicología/Annals of Psychology*, *8*(1-2), 3-18.

Cohen, J. (1994). The earth is round (p<. 05). *American psychologist*, *49*(12), 997-1003. https://doi.org/10.1037/0003-066X.49.12.997

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (6th Ed.). New York, NY: Routledge.

Cumming, G. (2006). Understanding replication: Confidence intervals, p-values, and what's likely to happen next time. *Developing a statistically literate society*. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July, 2006), Salvador, Bahia, Brazil.

Crisóstomo, E. (2012). *Idoneidad de procesos de estudio del cálculo integral en la formación de profesores de matemáticas: una aproximación desde la investigación en didáctica del cálculo y el conocimiento profesional* (Tesis de Doctorado). Universidad de Granada, España.

Cumming, G. (2015). Confidence intervals. En G. Ritzer (Ed.). *The Blackwell encyclopedia of sociology*. https://doi.org/10.1002/9781405165518.wbeosc087.pub2

Cumming, G., Williams, J., & Fidler, F. (2004). Replication and researchers' understanding of confidence intervals and standard error bars. *Understanding statistics*, *3*(4), 299-311.

delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79–95). Dordrecht: Kluwer Academic Publishers.

Denzin, N. & Lincoln, Y. (1994). *Handbook of Qualitative Research* (2da Ed.). Thousands Oaks, CA: Sage

DePaolo, C. A., Robinson, D. F. & Jacobs, A. (2016) Café Data 2.0: New Data From a New and Improved Café. *Journal of Statistics Education*, *24*(2), 85-103. https://doi.org/10.1080/10691898.2016.1196064

de Vetten, A., Schoonenboom, J., Keijzer, R., & van Oers, B. (2019). Pre-service primary school teachers' knowledge of informal statistical inference. *Journal of Mathematics Teacher Education*, *22*(6), 639-661. https://doi.org/10.1007/s10857-018-9403-9

Devore, J. L. (2008). *Probability & Statistics for Engineering and the Sciences* (7th Ed.). Mexico: Cengage Learning.

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning*, *13*(1-2), 132-151. https://doi.org/10.1080/10986065.2011.538294

Dinov, I. D., Palanimalai, S., Khare, A., & Christou, N. (2018). Randomization-based statistical inference: A resampling and simulation infrastructure. *Teaching Statistics*, *40*(2), 64-73. https://doi.org/10.1111/test.12156

Doerr, H. M., delMas, R. & Makar, K. (2017). A modeling approach to the development of students' informal inferential reasoning. *Statistics Education Research Journal, 16*(2), 86-115.

Dolor, J., & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal*, *14*(1), 60-89.

Edgeworth, F.Y. (1885). On Methods of Ascertaining Variations in the rate of Births, Deaths and Marriages", *Journal of the Statistical Society*, *48*, 628-649.

Eisenhart, C. (1979). On the Transition from "Student's" z to "Student's" t. *The American Statistician*, *33*(1), 6-12.

English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*(1), 15-30.

English, L. D., & Watson, J. (2018). Modelling with authentic data in sixth grade. *ZDM - International Journal on Mathematics Education*, *50*(1-2), 103-115.

Falk, R. (1986). Conditional probabilities: insights and difficulties. En *Proceedings of the Second International Conference on Teaching Statistics* (ICOTS2, 1986), (pp. 292-297), Victoria, Canada.

Fellers, P. S. & Kuiper, S. (2020). Introducing Undergraduates to Concepts of Survey Data Analysis. *Journal of Statistics Education*, *28*(1), 18-24. https://doi.org/10.1080/10691898.2020.1720552

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507-521. https://doi.org/10.2307/2331838

Fisher, R. A. (1922a). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, *85*(1), 87-94.

Fisher, R. A. (1922b). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, *85*(4), 597-612.

Fisher, R. A. (1925a). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Fisher, R. A. (1925b). Applications of 'Student's' Distribution. *Metron*, 5, 90-104, Collected Papers, 2, no. 43.

Fisher, R. A. (1932). *Statistical methods for research workers* (4th Ed.). Edinburgh: Oliver and Boyd.

Fisher, R. A. (1934). *Statistical methods for research workers* (5th Ed.). Edinburgh: Oliver and Boyd.

Fisher Box, J. (1980). Gosset, Fisher, and the t distribution. *The American Statistician*, *35*(2), 61-66.

Font, V., Godino, J. D., & Gallardo, J. (2013). The emergence of objects from mathematical practices. *Educational Studies in Mathematics*, *82*(1), 97-124. https://doi.org/10.1007/s10649-012-9411-0

Font, V., & Rubio, N. V. (2017). Procesos matemáticos en el enfoque ontosemiótico. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*.

GAISE (2005). Guidelines for assessment and instruction in statistics education. *PreK-12 and College report*. Alexandria, VA: American Statistical Association.

GAISE (2016). Guidelines for assessment and instruction in statistics education. *College report*. Alexandria, VA: American Statistical Association.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*, 1-51.

Galton, F. (1875). IV. Statistics by intercomparison, with remarks on the law of frequency of error. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *49*(322), 33-46.

Galton, F. (1880). Statistics of mental imagery. *Mind*, *5*(19), 301-318.

Galton, F. (1885). The Application of a Graphic Method to Fallible Measures. *Journal of the Statistical Society of London*, 262-271.

Galton, F. (1889). *Natural inheritance*. London: Macmillan and Company.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910676

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22-38. Online: www.stat.auckland.ac.nz/~iase/serj/

Garfield, J. & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht: Springer. DOI: 10.1007/978-1-4020-8383-9

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics*, *88*(3), 327-342.

Gibbs, A. L. & Goossens, E. T. (2013) The Evidence for Efficacy of HPV Vaccines: Investigations in Categorical Data Analysis. *Journal of Statistics Education*, *21*(3), doi: 10.1080/10691898.2013.11889688

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning*, *13*(1-2), 87-108. https://doi.org/10.1080/10986065.2011.538295

Godino, J.D., y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, *14*(3), 325-355.

Godino, J. D., & Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area for research in mathematics education. En Sierpinska A., Kilpatrick J. (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 177-195). Springer, Dordrecht. https://doi.org/10.1007/978-94-011-5470-3_12

Godino, J.D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education*, *39*(1-2), 127-135. https://doi.org/10.1007/s11858-006-0004-1

Godino, J.D., Batanero, C., & Font, V. (2019). The onto-semiotic approach: implications for the prescriptive character of didactics. *For the Learning of Mathematics*, *39*(1), 38-43

Godino, J.D., Font, V., Wilhelmi, M.R., & Lurduy, O. (2011). Why is the learning of elementary arithmetic concepts difficult? Semiotic tools for understanding the nature of mathematical objects. *Educational Studies in Mathematics*, *77*, 247–265. https://doi.org/10.1007/s10649-010-9278-x

Gordillo, W., & Pino-Fan, L. (2016). Una propuesta de reconstrucción del significado holístico de la antiderivada. *Bolema: Boletim de Educação Matemática*, *30*(55), 535-558. doi: http://dx.doi.org/10.1590/1980-4415v30n55a12

Greenwood, M., & Yule, G. U. (1915). The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine*, *8*, 113-194

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. Springer Science & Business Media.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research*, *7*(1), 1-20.

Harradine A., Batanero C., Rossman A. (2011). Students and Teachers' Knowledge of Sampling and Inference. En C. Batanero, G. Burrill, C. Reading (Eds.) *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-1131-0_24

Hartzell, F. Z. (1929). Application of one of Pearson's probability theorems and some special probability equations to entomological data. *Journal of Economic Entomology*, *22*(1), 202-209.

Henriques, A., & Oliveira, H. (2016). Students' expressions of uncertainty in making informal inference when engaged in a statistical investigation using tinkerplots. *Statistics Education Research Journal*, *15*(2), 62-80.

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). *Metodología de la investigación* (Vol. 4). México: McGraw-Hill.

Heyde, C. C., & Seneta, E. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*. New York: Springer.

Hoekstra, R. (2015). Risk as an Explanatory Factor for Researchers' Inferential Interpretations. *The Mathematics Enthusiast*, *12*(1), 103–112.

Inzunza, S., & Jiménez, J.V. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista latinoamericana de investigación en matemática educativa*, *16*(2), 179-211. https://doi.org/10.12802/relime.13.1622

Irwin, M. R. (1929). The inheritance of resistance to the Danysz bacillus in the rat. *Genetics*, *14*(4), 337-365.

Jacob, B.L., & Doerr, H.M. (2014). Statistical Reasoning with the sampling distribution. En K. Makar, B. de Sousa, R. Gould (Eds.), *Sustainability in statistics education*. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagstaff, Arizona, USA.

Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2004). Models of development in statistical reasoning. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 97–117). Dordrecht: Kluwer Academic Publishers.

Juárez, J. A., & Inzunza, S. (2014). Comprensión y razonamiento de profesores de Matemáticas de bachillerato sobre conceptos estadísticos básicos. *Perfiles Educativos*, *36*(146), 14-29.

Kaplan, J. J. (2009). Effect of belief bias on the development of undergraduate students' reasoning about inference. *Journal of Statistics Education*, *17*(1). https://doi.org/10.1080/10691898.2009.11889501

Kazak, S., Fujita, T., & Wegerif, R. (2016). Students' informal inference about the binomial distribution of "bunny hops": A dialogic perspective. *Statistics Education Research Journal*, *15*(2), 46-61.

Kazak, S., & Pratt, D. (2017). Pre-service mathematics teachers' use of probability models in making informal inferences about a chance game. *Statistics Education Research Journal*, *16*(2).

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N.J., & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, *13*(1-2), 68-86. https://doi.org/10.1080/10986065.2011.538299

Lancaster, H. O., (1966). Forerunners of the Pearson χ^2. *Australian Journal of Statistics*, *8*(3), 117-126.

Lane-Getaz, S. (2007). *Development and validation of a research-based assessment: reasoning about P-values and statistical significance* (Tesis de Doctorado). University of Minnesota, USA.

Lane-Getaz, S. J. (2013). Development of a reliable measure of students' inferential reasoning ability. *Statistics Education Research Journal*, *12*(1), 20-47.

Lane-Getaz, S. J. (2017). Is the p-value really dead? Assessing inference learning outcomes for social science students in an introductory statistics course. *Statistics Education Research Journal*, *16*(1), 357-399.

Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two?. *Journal of the American statistical Association*, *88*(424), 1242-1249.

Leigh, M. & Dowling A. D. (2010). Water Taste Test Data. *Journal of Statistics Education*, *18*(1), doi: 10.1080/10691898.2010.11889484

Liljedahl P., Santos-Trigo M., Malaspina U., Bruder R. (2016) *Problem Solving in Mathematics Education. Springer*, Cham. https://doi.org/10.1007/978-3-319-40730-2_1

Lipsey, M. W., & Aiken, L. S. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). London: SAGE.

Lipson, A. (2003). The role of the sampling distribution in understanding statistical inference. *Mathematical Educational Research Journal*, *15*(3), 270-287.

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies: An International Journal*, *4*(2), 126-138.

López-Martín, M.D.M., Batanero, C., & Gea, M.M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? *Bolema: Boletim de Educação Matemática*, *33*(64), 672-693. http://dx.doi.org/10.1590/1980-4415v33n64a11

Lovett, M. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. En S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 347-384). Hillsdale, NJ: Erlbaum.

Lyu, Z., Peng, K., & Hu, C. P. (2018). P-value, confidence intervals, and statistical inference: A new dataset of misinterpretation. *Frontiers in Psychology*, *9*(868). https://doi.org/10.3389/fpsyg.2018.00868

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. Mathematical *Thinking and Learning*, *13*(1-2), 109-131. https://doi.org/10.1080/10986065.2011.539078

Magnello, M. E. (1993). *Karl Pearson: evolutionary biology and the emergence of a modern theory of statistics* (Tesis de Doctorado). University of Oxford, USA.

Magnello, M. E. (2005). Karl Pearson, paper on the chi square goodness of fit test (1900). En *Landmark Writings in Western Mathematics* 1640-1940 (pp. 724-731). Elsevier Science.

Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics*, *86*(1), 61-78. https://doi.org/10.1007/s10649-013-9526-y

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, *18*(1), 1-24.

Makar, K., Bakker, A. y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1-2), 152-173. https://doi.org/10.1080/10986065.2011.538301

Makar, K., & Ben-Zvi, D. (2011). The Role of Context in Developing Reasoning about Informal Statistical Inference. *Mathematical Thinking and Learning*. *13*(1-2), 1-4. https://doi.org/10.1080/10986065.2011.538291

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. En D. Ben-Zvi, K. Makar, J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261-294). Switzerland: Springer International. https://doi.org/10.1007/978-3-319-66195-7_8

Ministerio de Educación de Chile [Mineduc]. (2019). *Bases Curriculares 3º y 4º medio*. Santiago de Chile: Unidad de Currículum y Evaluación.

Molina, O. J. (2019). *Sistema de normas que influyen en procesos de argumentación: un curso de geometría del espacio como escenario de investigación* (Tesis de Doctorado). Universidad de Los Lagos, Chile.

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, *4*(1), 23-63.

Moore, D. S. (2000). *Estadística aplicada básica* (2da ed.). Barcelona, España: Antoni Bosch.

Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–256). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Newton, J., Dietiker, L., & Horvath, A. (2011). Statistics education in the United States: Statistical reasoning and the statistical process. En C. Batanero, G. Burril y C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 9-13). Dordrecht: Springer.

Noll, J., Gebresenbet, M., & Glover, E. D. (2016). A modeling and simulation approach to informal inference: Successes and challenges. En D. Ben-Zvi K. Makar (Eds.), *The teaching and learning of statistics* (pp. 139-150). Springer, Cham. https://doi.org/10.1007/978-3-319-23470-0_19

Noll, J., & Hancock, S. (2015). Proper and paradigmatic metonymy as a lens for characterizing student conceptions of distributions and sampling. *Educational Studies in Mathematics*, *88*(3), 361-383. https://doi.org/10.1007/s10649-014-9547-1

Olivo, E. (2008). *Significados de los intervalos de confianza para los estudiantes de ingeniería en México* (Tesis de Doctorado). Universidad de Granada, España.

Otte, M. (2003). Does mathematics have objects? In what sense? *Synthese*, *134*, 181-216.

Park, J. (2012). *Developing and validating an instrument to measure college students' inferential reasoning in statistics: an argument-based approach to validation* (Tesis de Doctorado). University of Minnesota, USA.

Pearce, L. E. (1992). The cowrie shell in Virginia: a critical evaluation of potential archaeological significance. *College of William & Mary*. https://dx.doi.org/doi:10.21220/s2-e12a-8115

Pearson, K. (1895). X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London*, *A*(186), 343-414.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

Pearson, K. (1904). On the theory of contingency and its relation to association and normal correlation. *Draper's Company Research Memoirs Biometric Series I*.

Pearson, K. (1907). On the influence of past experience on future expectation. *Philosophical Magazine*, *Series 6* (13), 365-378.

Pearson, K. (1911). On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika*, *8*(1/2), 250-254.

Pearson, K. (1914). *Tables for Statisticians and Biomnetricians*. Cambridge: University Press.

Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? En G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267–294). Dordrecht, The Netherlands: Kluwer Academic.

Pfannkuch, M. (2006). Informal inferential reasoning. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July, 2006), Salvador, Bahia, Brazil.

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematical Education*. *2*(3), 149-167.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1-2), 27-46. https://doi.org/10.1080/10986065.2011.538302

Pfannkuch, M. (2018). Reimagining curriculum approaches. En D. Ben-Zvi, K. Makar, J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387-413). Switzerland: Springer International. https://doi.org/10.1007/978-3-319-66195-7_12

Pfannkuch, M., Arnold, P., & Wild, C. J. (2015). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics*, *88*(3), 343-360. https://doi.org/10.1007/s10649-014-9539-1

Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., & Ziedins, I. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal*, *15*(2), 11–37.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Dordrecht: Kluwer Academic Publishers.

Pino-Fan, L. (2013). *Evaluación de la faceta epistémica del conocimiento didáctico-matemático de futuros profesores de bachillerato sobre la derivada* (Tesis de Doctorado), Universidad de Granada, España.

Pino-Fan, L. (2017). Contribución del Enfoque Ontosemiótico a las investigaciones sobre didáctica del cálculo. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M, Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponible en: http://enfoqueontosemiotico.ugr.es/civeos/pino-fan.pdf

Pino-Fan, L., Font, V., Gordillo, W., Larios, V., & Breda, A. (2017). Analysis of the meanings of the antiderivative used by students of the first engineering courses. *International Journal of Science and Mathematics Education*, *16*(6), 1091-1113. doi: 10.1007/s10763-017-9826-2

Pino-Fan, L. R., Godino, J. D., & Font, V. (2011). Faceta epistémica del conocimiento didáctico-matemático sobre la derivada. *Educação Matemática Pesquisa*, *13*(1), 141-178.

Pino-Fan, L. R., Godino, D. J., & Font, V. (2015). Una propuesta para el análisis de las prácticas matemáticas de futuros profesores sobre derivadas. *Bolema: Boletim de Educação Matemática*, *29*(51), 60-89. http://dx.doi.org/10.1590/1980-4415v29n51a04

Polya, G. (1989). Cómo plantear y resolver problemas. México D.F., México: Editorial Trillas.

Presmeg, N. (2014) Semiotics in Mathematics Education. En S. Lerman (Eds.), *Encyclopedia of Mathematics Education*. Cham, Switzerland: Springer.

Prodromou, T. (2013). Informal inferential reasoning using a modelling approach within a computer-based simulation. *International Scholarly and Scientific Research & Innovation*, *7*(6), 1810-1815.

Ramsey, F. & Schafer, D. (2012). *The statistical sleuth: a course in methods of data analysis (3era Ed)*. Estados Unidos de America: Cengage Learning.

Reading, C. (2009). Cognitive development of informal inferential reasoning. *57th Session*

of the International Statistical Institute. Durban, South Africa.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, *5*(2), 46-68.

Riemer, W. & Seebach, G. (2014). Rolling pencils - inferential statistics in the pencil case. En *Understanding more mathematics with GeoGebra* (pp. 91-105). Heidelberg: Springer Spektrum.

Rochowicz, J. A. (2010). Bootstrapping analysis, inferential statistics and EXCEL. *Spreadsheets in Education (eJSiE), 4*(3), 1-23.

Rondero, C., & Font, V. (2015). Articulación de la complejidad matemática de la media aritmética. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, *33*(2), 29-49.

Rossman, A.J. (2008). Reasoning about Informal Statistical Inference: One Statistician's View. *Statistics Education Research Journal*, *7*(2), 5-19.

Rossman, A. J., & Chance, B. L. (2014). Using simulation-based inference for learning introductory statistics. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(4), 211-221. https://doi.org/10.1002/wics.1302

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. En *Proceedings of the Seventh International Conference on Teaching Statistics* (ICOTS7, July, 2006), Salvador, Bahia, Brazil.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3).

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, *17*(4), 688-690. https://doi.org/10.1093/beheco/ark016

Schield, M. (2010). Assessing Statistical Literacy: Take CARE. En P. Bidgood, N. Hunt y F. Jolliffe (Eds.) *Assessment Methods in Statistical Education: An International Perspective*, (pp. 133-152). John Wiley & Sons Inc: Chischester, Great Britain.

Schoenfeld, A. H. (1982). Some thoughts on problem-solving research and mathematics education. En F. K. Lester & J. Garofalo (Eds.), *Mathematical problem solving: Issues in research* (pp. 27–37). Philadelphia: Franklin Institute Press.

Seier, E. (2014). An Early Start on Inference. En K. Makar, B. de Sousa, R. Gould (Eds.), *Sustainability in statistics education*. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagstaff, Arizona, USA.

Sevilla, M. J. (1993). Teoría de errores de observación. *Física de la Tierra*, (37), 133-166.

Shaughnessy, J. M., & Dick, T. (1991). Monty's dilemma: should you stick or switch?. *The Mathematics Teacher*, *84*(4), 252-256.

Snedecor, G. W. (1930). A statistical test of experimental technique. *Proceedings of the Iowa Academy of Science*, *37*(1), 279-287.

Snedecor, G., & Irwin, M. R. (1933). On the chi-square test for homogeneity. *Iowa State Coll. J. Sci*, *8*(1), 75-81.

Sotos, A.E.C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98-113.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Stigler, S. M. (1995). Galton and identification by fingerprints. *Genetics*, *140*(3), 857.

Stigler, S. M. (2017). *Los siete pilares de la sabiduría estadística*. México: Grano de Sal.

Stohl Lee, H., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, *9*(1), 68-96.

Student. (1908a). The probable error of a mean. *Biometrika*, *6*(1,)1-25. https://doi.org/10.2307/2331554

Student. (1908b). Probable error of a correlation coefficient. *Biometrika*, *6*(2/3) ,302-310. https://doi.org/10.2307/2331474

Student. (1917). Tables for Estimating the Probability that the Mean of a Unique Sample of Observations Lies Between-∞ and Any Given Distance of the Mean of the Population from Which the Sample is Drawn. *Biometrika*, *11*(4), 414-417. https://doi.org/10.2307/2331831

Student (1925). New table for testing the significance of observations. *Metron*, *5*, 105-108.

Tarlow, K. R. (2016). Teaching principles of inference with ANOVA. *Teaching Statistics*, *38*(1), 16-21. https://doi.org/10.1111/test.12085

Taylor, L., & Doehler, K. (2015). Reinforcing sampling distributions through a randomization-based activity for introducing ANOVA. *Journal of Statistics Education*, *23*(3). https://doi.org/10.1080/10691898.2015.11889750

Thompson, P. W., Liu, Y., & Saldanha, L. A. (2007). Intricacies of statistical inference and teachers' understandings of them. *Thinking with data*, 207-231.

Tintle, N. L., Topliff, K., VanderStoep, J., Holmes, V. L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, *11*(1), 21-40.

Trumpower, D. L. (2013). Formative use of intuitive analysis of variance. *Mathematical Thinking and Learning*, *15*(4), 291-313.

Trumpower, D. L. (2015). Aspects of first year statistics students' reasoning when performing intuitive analysis of variance: effects of within-and between-group variability. *Educational Studies in Mathematics, 88*(1), 115-136.

Vallecillos, A (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios* (Tesis de Doctorado). Universidad de Granada, España.

Vallecillos, A. (1997). El papel de las hipótesis estadísticas en los contrastes: Concepciones y dificultades de aprendizaje. *Educación Matemática*, *9*(2), 5-20.

Vallecillos, A. (2002). Empirical evidence about understanding of the level of significance concept in hypotheses testing by university students. *Themes in Education*, *3*(2), 183-198.

Vallecillos, A., & Batanero, C. (1997). Análisis del aprendizaje de conceptos clave en el contraste de hipótesis estadísticas mediante el estudio de casos. *Recherches en Didactique des Mathématiques*, *17*(1), 29-48.

Vera, O. D., & Díaz, C. (2013). Dificultades de estudiantes de psicología en relación al contraste de hipótesis. En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 197-203). Granada, Departamento de Didáctica de la Matemática de la Universidad de Granada.

Vera, O. D., Díaz, C., & Batenero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *UNIÓN. Revista Iberoamericana de Educación Matemática*, *27*, 41-61.

Watson, J. M. (1997). Assessing statistical literacy using the media. En I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam: ISO Press and the International Statistical Institute.

Watson, J. M. (2004). Developing reasoning about samples. En *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Watson, J. M. (2006). *Statistical literacy at school*. Mahwah, NJ: Lawrence Erlbaum.

Watson, J. M. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal*, *7*(2).

Watson, J. M. (2012). Resampling with TinkerPlots. *Teaching Statistics: an international journal for teachers*, *35*(1), 32-36.

Watson, J. M., & English, L. (2017). Reaction time in grade 5: Data collection within the practice of statistics. *Statistics Education Research Journal*, *16*(1), 262-293.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*(2), 145-168.

Weinberg, A., Wiesner, E., & Pfaff, T. J. (2010). Using informal inferential reasoning to develop formal concepts: Analyzing an activity. *Journal of Statistics Education*, *18*(2). Doi: 10.1080/10691898.2010.11889494

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*(3/4), 350-362. https://doi.org/10.2307/2332010

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28-35. https://doi.org/10.2307/2332510

Wijayatunga, P. (2016). '*A geometric view on Pearson's correlation coefficient and a generalization of it to non-linear dependencies*', *Ratio Mathematica*, 30, pp. 3–21. doi:10.23755/rm.v30i1.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 247-295.

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics?. En D. Ben-Zvi y K. Makar (Eds.), *International Handbook of Research in Statistics Education. Springer International Handbooks of Education* (pp. 5-36). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_1

Wilkerson, M., & Olson, M. R. (1997). Misconceptions about sample size, statistical significance, and treatment effect. *The Journal of Psychology*, *131*(6), 627-631. https://doi.org/10.1080/00223989709603844

Williams, A. M. (1999). Novice students' conceptual knowledge of statistical hypothesis testing. En J.M. Truran & K.M. Truran (Eds.), *Making the difference: Proceedings of the twenty-second annual conference of the mathematics education research group of Australasia* (pp. 554-560). Adelaide, South Australia.

Woodard, V., Lee, H. & Woodard, R. (2020). Writing Assignments to Assess Statistical Thinking, *Journal of Statistics Education*, *28*(1), 32-44. https://doi.org/10.1080/10691898.2019.1696257

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, *1*(2), 217-235.

Yates, F. (1951). The influence of “Statistical methods for research workers” on the development of the science of statistics. *Journal of the American Statistical Association*, *46*, 19-34.

Yule, G. U. (1900). On the association of attributes in statistics: with illustrations from the material of the childhood society, &c. *Philosophical Transactions of the Royal Society of London. Series A* 194, 257-319.

Yule, G. U. (1922). On the application of the χ 2 method to association and contingency tables, with experimental illustrations. *Journal of the Royal Statistical Society*, *85*(1), 95-104. https://doi.org/10.2307/2340522

Zabell, S. L. (2008). On student's 1908 article “the probable error of a mean”. *Journal of the American Statistical Association*, *103*(481), 1-7. https://doi.org/10.1198/016214508000000030

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40–58.

ANEXOS

A. ACTIVIDADES DE LOS GRUPOS 1 Y 4 (ESTADÍSTICO CHI-CUADRADA)

Taller de Razonamiento Estadístico

POSTGRADOS EN EDUCACIÓN MATEMÁTICA ULAGOS

La actividad 1 se presenta en el grupo y se forman los equipos

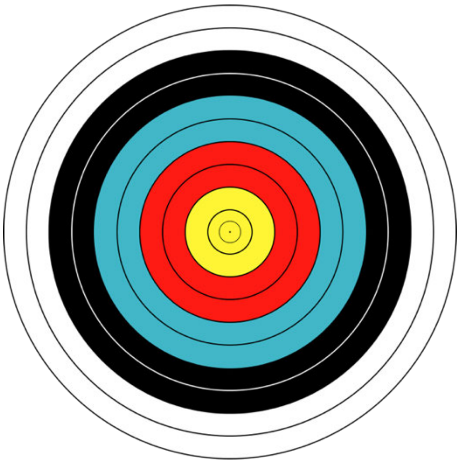
Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 1.
En el deporte de tiro con arco el objetivo es acertar lo más cerca posible del centro de la diana para obtener el máximo de puntos. En un torneo de amateurs de tiro con arco ninguna flecha ha dado en el punto central de la diana, sin embargo, se han contabilizado el número de flechas en cada uno de los anillos de la diana, siendo 1 el anillo más cercano al centro y por lo tanto con una desviación menor del punto central de la diana. Se cree que las desviaciones (distancias) del anillo de la diana donde se encuentra la flecha al centro de la diana siguen una distribución normal. En la tabla siguiente se encuentran contabilizadas las frecuencias que se observaron durante los tiros (frecuencia observada) y las frecuencias que se esperarían si las distancias siguen una distribución normal (frecuencia esperada). ¿Qué podrías comentar sobre la forma en que se distribuyen las frecuencias observadas? Explica detalladamente tu respuesta

Tabla 1. *Distribución de las frecuencias*

Anillo de la diana	Frecuencia Observada	Frecuencia Esperada
1	2	1
2	5	6
3	45	22
4	95	58
5	106	104
6	102	125
7	97	102
8	39	56
9	8	21
10	1	5
Total	500	500



Interacción en el Grupo con los equipos

La actividad 2 se presenta en el grupo y se forman los equipos

Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 2.
A continuación, se presentan los datos de una muestra que fueron recabados durante una epidemia de viruela en un pequeño pueblo de Inglaterra. Los datos de la tabla 2 muestran la presencia o ausencia de la cicatriz de la vacuna de viruela y si las personas que recibieron la vacuna se recuperaron o fallecieron. ¿Existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela? Explica con detalle tu respuesta

Tabla 2. *Presencia de cicatriz por vacuna y recuperados de viruela*

Cicatriz	Recuperados	Muertos	Total
Presente	35	1	36
Ausente	10	4	14
Total	45	5	50

Interacción en el Grupo con los equipos

La actividad 3 se presenta en el grupo y se forman los equipos

Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 3.
A partir de una encuesta realizada en el 2017 por el CDC (Centro para el Control y Prevención de Enfermedades de los Estados Unidos) a un grupo de hombres y un grupo de mujeres que padecen cáncer de pulmón, se obtuvo la siguiente información sobre las frecuencias de enfermos de cáncer por raza (casta o etnia). Para realizar un tratamiento adecuado de los datos es necesario conocer si existe homogeneidad entre los dos grupos respecto a las razas. ¿Consideras que existe homogeneidad entre los grupos (Hombres y Mujeres)? Argumenta tu respuesta

Tabla 3. *Cáncer de pulmón por género y raza*

	Caucásico	**Afroamericano**	**Asiático**	**Hispano**	**Total**
Hombres	95	12	3	8	118
Mujeres	92	10	3	6	111
Total	187	22	6	14	229

Interacción en el Grupo con los equipos

B. ACTIVIDADES DEL GRUPO 2 (ESTADÍSTICO CHI-CUADRADA)

Taller de Razonamiento Estadístico

POSTGRADOS EN EDUCACIÓN MATEMÁTICA ULAGOS

Indicaciones:

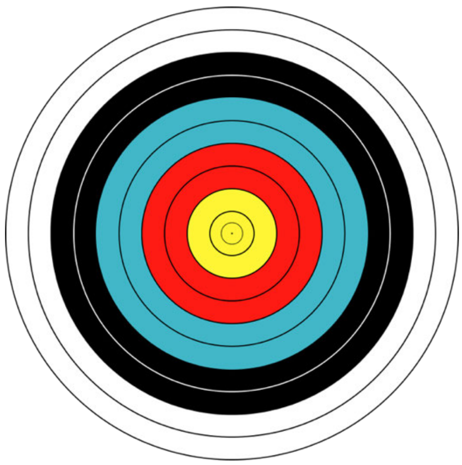
- Lee con atención cada una de las actividades que se presentan a continuación.
- Para resolver las actividades puedes utilizar papel y lápiz, calculadora, software o el medio con el que te sientas más cómodo.
- Explica la estrategia de solución que seguiste.
- No olvides subir a la plataforma tus soluciones (incluye fotografías o documentos).

Actividad 1.

En el deporte de tiro con arco el objetivo es acertar lo más cerca posible del centro de la diana para obtener el máximo de puntos. En un torneo de amateurs de tiro con arco ninguna flecha ha dado en el punto central de la diana, sin embargo, se han contabilizado el número de flechas en cada uno de los anillos de la diana, siendo 1 el anillo más cercano al centro y por lo tanto con una desviación menor del punto central de la diana. Se cree que las desviaciones (distancias) del anillo de la diana donde se encuentra la flecha al centro de la diana siguen una distribución normal. En la tabla siguiente se encuentran contabilizadas las frecuencias que se observaron durante los tiros (frecuencia observada) y las frecuencias que se esperarían si las distancias siguen una distribución normal (frecuencia esperada). ¿Qué podrías comentar sobre la forma en que se distribuyen las frecuencias observadas? Explica detalladamente tu respuesta

Tabla 4. *Distribución de las frecuencias*

Anillo de la diana	Frecuencia Observada	Frecuencia Esperada
1	2	1
2	5	6
3	45	22
4	95	58
5	106	104
6	102	125
7	97	102
8	39	56
9	8	21
10	1	5
Total	500	500



Actividad 2.

A continuación, se presentan los datos de una muestra que fueron recabados durante una epidemia de viruela en un pequeño pueblo de Inglaterra. Los datos de la tabla 2 muestran la presencia o ausencia de la cicatriz de la vacuna de viruela y si las personas que recibieron la vacuna se recuperaron o fallecieron. ¿Existe alguna relación entre la presencia de cicatriz por vacuna y los recuperados de viruela? Explica con detalle tu respuesta

Tabla 5. *Presencia de cicatriz por vacuna y recuperados de viruela*

Cicatriz	Recuperados	Muertos	Total
Presente	35	1	36
Ausente	10	4	14
Total	45	5	50

Actividad 3.

A partir de una encuesta realizada en el 2017 por el CDC (Centro para el Control y Prevención de Enfermedades de los Estados Unidos) a un grupo de hombres y un grupo de mujeres que padecen cáncer de pulmón, se obtuvo la siguiente información sobre las frecuencias de enfermos de cáncer por raza (casta o etnia). Para realizar un tratamiento adecuado de los datos es necesario conocer si existe homogeneidad entre los dos grupos respecto a las razas. ¿Consideras que existe homogeneidad entre los grupos (Hombres y Mujeres)? Argumenta tu respuesta

Tabla 6. *Cáncer de pulmón por género y raza*

	Caucásico	**Afroamericano**	**Asiático**	**Hispano**	**Total**
Hombres	95	12	3	8	118
Mujeres	92	10	3	6	111
Total	187	22	6	14	229

C. ACTIVIDADES DEL GRUPO 2 (ESTADÍSTICO T-STUDENT)

Taller de Razonamiento Estadístico

POSTGRADOS EN EDUCACIÓN MATEMÁTICA ULAGOS

Indicaciones:

- Lee con atención cada una de las actividades que se presentan a continuación.
- Para resolver las actividades puedes utilizar papel y lápiz, calculadora, software o el medio con el que te sientas más cómodo.
- Explica la estrategia de solución que seguiste.
- No olvides subir a la plataforma tus soluciones (incluye fotografías o documentos).

Actividad 1.

A continuación, se presentan los datos de una muestra que fueron recabados durante un experimento de un fármaco para dormir. Los datos de la Tabla 4 son sobre el aumento de horas de sueño cuando los pacientes han ingerido el fármaco. ¿Existe realmente un aumento en las horas de sueño al utilizar el fármaco? Explica con detalle tu respuesta

Tabla 7 – Horas adicionales de sueño ganadas por el uso del fármaco

Paciente	Horas adicionales
1	$+0.7$
2	-1.6
3	-0.2
4	-1.2
5	-0.1
6	$+3.4$
7	$+3.7$
8	$+0.8$
9	0.0
10	$+2.0$

Actividad 2.

Una farmacéutica está realizando experimentos sobre la efectividad de dos fármacos para dormir, para lo cual ha seleccionado aleatoriamente a diez pacientes quienes han tomado ambos fármacos. La Tabla 5 muestra el aumento de horas de sueño cuando los pacientes han ingerido cada uno de los fármacos. ¿Existe realmente una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2?

Tabla 8 – Horas ganadas de sueño por participante con ambos fármacos

Paciente	Fármaco 1	Fármaco 2
1	$+0.7$	$+1.9$
2	-1.6	$+1.8$
3	-0.2	$+1.1$
4	-1.2	$+0.1$
5	-0.1	-0.1
6	$+3.4$	$+4.4$
7	$+3.7$	$+5.5$
8	$+0.8$	$+1.6$
9	0.0	$+4.6$
10	$+2.0$	$+3.4$

Actividad 3.

En época de lluvias, es común ver en las noticias los socavones que se generan en diversas ciudades del mundo, una de las principales causas el deterioro de las tuberías (que es un problema creciente). Este deterioro en las tuberías genera el reblandecimiento del suelo por la humedad y con el paso de los automóviles, el peso de las construcciones, los camiones, entre otros, provocan que la parte superficial empiece a vibrar y la resistencia de la parte superior ya no es suficiente y colapsa, provocando así los socavones y con ellos lamentables accidentes. Una posible solución, que no implica el cambio de la red de tuberías, es rehabilitar las tuberías existentes utilizando un forro flexible. Sin embargo, existen dos propuestas utilizar un proceso de fusión y no utilizar el proceso en el forro. Quienes están a favor del proceso de fusión consideran que el proceso incrementa la resistencia a la tensión promedio. Se han reportado los siguientes datos de resistencias a la tensión $(lb/pulg^2)$ de especímenes, cuando se utilizó el proceso de fusión en el forro y cuando este proceso no se utilizó. ¿Realmente existe un incremento en la resistencia a la tensión promedio cuando se utiliza el proceso de fusión? Explica con detalle tu respuesta.

Tabla 9. *Resistencias a la tensión* $(\boldsymbol{lb/pulg^2})$ *de especímenes*

Sin proceso de fusión	**Con proceso de fusión**
2748	3027
2700	3356
2655	3359
2822	3297
2511	3125
3149	2910
3257	2889
3213	2902
3220	
2753	

D. ACTIVIDADES DEL GRUPO 3 (ESTADÍSTICO T-STUDENT)

Taller de Razonamiento Estadístico POSTGRADOS EN EDUCACIÓN MATEMÁTICA ULAGOS

La actividad 1 se presenta en el grupo y se forman los equipos

Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 1.

A continuación, se presentan los datos de una muestra que fueron recabados durante un experimento de un fármaco para dormir. Los datos de la Tabla 4 son sobre el aumento de horas de sueño cuando los pacientes han ingerido el fármaco. ¿Existe realmente un aumento en las horas de sueño al utilizar el fármaco? Explica con detalle tu respuesta

Tabla 10 – Horas adicionales de sueño ganadas por el uso del fármaco

Paciente	Horas adicionales
1	$+0.7$
2	-1.6
3	-0.2
4	-1.2
5	-0.1
6	$+3.4$
7	$+3.7$
8	$+0.8$
9	0.0
10	$+2.0$

Interacción en el Grupo con los equipos

La actividad 2 se presenta en el grupo y se forman los equipos

Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 2.

Una farmacéutica está realizando experimentos sobre la efectividad de dos fármacos para dormir, para lo cual ha seleccionado aleatoriamente a diez pacientes quienes han tomado ambos fármacos. La Tabla 5 muestra el aumento de horas de sueño cuando los pacientes han ingerido cada uno de los fármacos. ¿Existe realmente una diferencia entre las horas de sueño que se ganan con el fármaco 1 y 2?

Tabla 11 – Horas ganadas de sueño por participante con ambos fármacos

Paciente	Fármaco 1	Fármaco 2
1	$+0.7$	$+1.9$
2	-1.6	$+1.8$
3	-0.2	$+1.1$
4	-1.2	$+0.1$
5	-0.1	-0.1
6	$+3.4$	$+4.4$
7	$+3.7$	$+5.5$
8	$+0.8$	$+1.6$
9	0.0	$+4.6$
10	$+2.0$	$+3.4$

Interacción en el Grupo con los equipos

La actividad 3 se presenta en el grupo y se forman los equipos

Indicaciones:

- En el equipo lean con atención nuevamente la actividad y comenten estrategias por medio de las cuales podrían solucionar el problema.
- Utilicen papel y lápiz, calculadora, software o el medio con el que se sientan más cómodos para apoyarse para resolver el problema.
- Una vez que resuelvan la actividad regresen a la sala general para compartir con el grupo sus soluciones y cuáles fueron las estrategias que siguieron.

Actividad 3.

En época de lluvias, es común ver en las noticias los socavones que se generan en diversas ciudades del mundo, una de las principales causas el deterioro de las tuberías (que es un problema creciente). Este deterioro en las tuberías genera el reblandecimiento del suelo por la humedad y con el paso de los automóviles, el peso de las construcciones, los camiones, entre otros, provocan que la parte superficial empiece a vibrar y la resistencia de la parte superior ya no es suficiente y colapsa, provocando así los socavones y con ellos lamentables accidentes. Una posible solución, que no implica el cambio de la red de tuberías, es rehabilitar las tuberías existentes utilizando un forro flexible. Sin embargo, existen dos propuestas utilizar un proceso de fusión y no utilizar el proceso en el forro. Quienes están a favor del proceso de fusión consideran que el proceso incrementa la resistencia a la tensión promedio. Se han reportado los siguientes datos de resistencias a la tensión $(lb/pulg^2)$ de especímenes, cuando se utilizó el proceso de fusión en el forro y cuando este proceso no se utilizó. ¿Realmente existe un incremento en la resistencia a la tensión promedio cuando se utiliza el proceso de fusión? Explica con detalle tu respuesta.

Tabla 12. *Resistencias a la tensión* $(\boldsymbol{lb/pulg^2})$ *de especímenes*

Sin proceso de fusión	**Con proceso de fusión**
2748	3027
2700	3356
2655	3359

2822	3297
2511	3125
3149	2910
3257	2889
3213	2902
3220	
2753	

Interacción en el Grupo con los equipos