

CREACIONES DE LAS DISTRIBUCIONES DE MUESTREO

CARLOS ALFREDO LÓPEZ GAITÁN

**UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS
BOGOTÁ, D.C.**

2018

CREACIONES DE LAS DISTRIBUCIONES DE MUESTREO

CARLOS ALFREDO LÓPEZ GAITÁN

Trabajo de grado de pregrado como requisito para obtener el título

Licenciado en Matemáticas

Director

FELIPE JORGE FERNÁNDEZ HERNÁNDEZ

Magister en Estadística

UNIVERSIDAD PEDAGÓGICA NACIONAL

FACULTAD DE CIENCIA Y TECNOLOGÍA

DEPARTAMENTO DE MATEMÁTICAS

LICENCIATURA EN MATEMÁTICAS

BOGOTÁ, D.C.

2018

Dedicatoria

Dedico este trabajo a mi madre Aracely Gaitán por su apoyo incondicional, y a mi hija Isabella López que estuvo acompañándome sentada en mi pierna izquierda.

Agradecimientos

Agradezco al profesor Felipe Jorge Fernández por su apoyo como director del presente trabajo de grado por compartirme sus experiencias, sus conocimientos y opiniones.

“Para todos los efectos, declaro que el presente trabajo es original y de mi total autoría; en aquellos casos en los cuales he requerido del trabajo de otros autores o investigadores, he dado los respectivos créditos”.



UNIVERSIDAD PEDAGÓGICA
NACIONAL
Alrededor de educadores

FACULTAD DE CIENCIA Y TECNOLOGÍA
DEPARTAMENTO DE MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS

ACTA DE EVALUACIÓN DE TRABAJO DE GRADO

Presentados y **aprobados** el documento escrito y la sustentación del Trabajo de Grado, en el tipo Monografía, titulado: "**Creaciones de las distribuciones de muestreo.**", elaborado por el estudiante:

Carlos Alfredo López Gaitán - código 2013140027 - cédula 1032479376

Como requisito parcial para optar al título de **Licenciado en Matemáticas**, el jurado evaluador asigna 41 puntos al mismo.

Sugerencia de Distinción: Ninguna Meritoria Laureada

En constancia se firma a los 23 días del mes de agosto de 2018.

Director del Trabajo: Profesor

FELIPE JORGE FERNÁNDEZ

Jurado:

Profesor

BENJAMÍN SARMIENTO LUGO

 UNIVERSIDAD PEDAGÓGICA NACIONAL <small>Universidad de maestros</small>	FORMATO
LICENCIA DE USO DEL TRABAJO Y/O TESIS DE GRADO A FAVOR DE LA UNIVERSIDAD PEDAGÓGICA NACIONAL	
CÓDIGO: FOR021GIB	Versión: 02
Fecha de Aprobación: 16-05-2013	Página 0 de 80

1. Información General

Tipo de documento	Trabajo de grado
Acceso al documento	Universidad Pedagógica Nacional. Biblioteca central
Título del documento	Creaciones de las distribuciones de muestreo.
Autor	López Gaitán, Carlos Alfredo
Director	Fernández Hernández, Felipe Jorge
Publicación	Bogotá. Universidad Pedagógica Nacional, 2018. 75 p.
Unidad Patrocinante	Universidad Pedagógica Nacional
Palabras claves	DEDUCCIÓN DE DISTRIBUCIONES, SIMULACIÓN, CONSTRUCCIÓN DE TABLAS, EDUCACIÓN ESTADÍSTICA, HISTORIA DE LA ESTADÍSTICA.

2. Descripción

Este trabajo de grado asume como objeto de estudio las siguientes distribuciones; normal, chi-cuadrado y t-Student; con el fin de suministrar información extra a profesores en formación o en ejercicio para una mejorar su comprensión, e insertar en el aula el poder de la historia de las matemáticas, justificación matemática, simulaciones en software, y aplicaciones.

3. Fuentes

- Bell E. (1985). *Historia de las matemáticas*. México: Fondo de Cultura Económica.
- Bernoulli, J. (1713). *Ars Conjectandi, opus posthumum*. Basilea, Suiza: Reprinted in Editions.
- Conde, A. G. (2015). La distribución normal una rápida revisión histórica. *Heurística, volumen XVII*, pp 59 – pp 65.

- Fisher R. A. (1925). Aplications of “Student’s” distribution. *Metrom* volumen V, pp 90 – pp 104.
- Gauss, C. F. (1809). *Theoria motvs corporvm coelestivm in sectionibvs conics solem ambientivm*. Hamburgo, Alemania: Petters et Besser.
- Gorroochurn, P. (2016). *Classic Topics on the History of Modern Mathematical Statistics: From Laplace to More Recent Times*. Estados Unidos: Wiley.
- Hald, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. Dinamarca: Wiley
- Hald, A. (2007). *A history of statistical inference from Bernoulli to Fisher, 1713 to 1935*. Dinamarca: Springer.
- Hernán, O. A. & Jiménez, F. N. (2006) La demostración elemento vivo en la didáctica de la matemática. *Scientia Et Technica*, volumen XII, pp 237- pp 240.
- Jinn, T. L. (1988). Approximating the Cumulative Chi-Square Distribution and its Inverse. *Journal of the royal statistical society. Series D. The statistician* volumen XXXVII, pp 3 – pp 5.
- Lehman, E. L. (1999). “Student” and small-sample theory. *Statistical Science*, volumen XIV, pp 418 – pp 426.
- Moivre, A. D. (1730). *Miscellanea Analytic*. Londres, Inglaterra: Tonson & Watts.
- Moivre, A. D. (1756). *The doctrine of chances*. Londres, Inglaterra: Millar.
- 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. *Revista de la Universidad de College*, volumen L, pp 157 – pp 175.
- Pearson, K. (1931) *Tables for Statisticians and Biometricalians*. Londres, Inglaterra: Universidad de College.
- Stahl, S. (2007). The evolution of the normal distribution. *Mathematics Magazine*, volumen LXXIX, pp 96 – pp 113.
- Stassen, B. (2007). *Psicología del desarrollo: infancia y adolescencia*. Bogotá, Colombia: Medica.
- Stewart, J. (2012). *Calculo de varias variables. Trascendentes tempranas 7E*. México: Cengage Learning.
- Stirling, J. (1730). *Methodus Differentialis: sive Tractatus de Summatione et Interpolatione Serierum Infini-tarum*. Londres, Inglaterra: Typis Gut.
- Student. (1908). The probable error of a mean. *Biometrika*, volumen XI, pp 1 – pp 25.
- Walpole, R. E. & Myers R. H. (2012). *Probabilidad y estadística para ingeniería y ciencia*. México: Pearson educación.

4. Contenidos

El presente documento está dividido seis capítulos, inicialmente se presenta la introducción, justificación y los objetivos del trabajo de grado. Exponiendo los motivos y propósitos que

llevaron a realizar este estudio, considerando que existe libros con la misma temática (distribuciones de muestro) pero con objetivos diferentes a los nuestros.

En el segundo capítulo, es presentada la distribución normal, comenzando con sus características como su forma acampanada y su función de densidad, es contada una breve historia de la probabilidad desde sus primeros inicios: porque sin ella no fuera posible construir las distribuciones, para después enfocarse en los documentos de Moivre (1756) sobre la aproximación de la distribución binomial. Luego se describe contribuciones de autores sobre la teoría de errores para finalizar con la deducción de la función de densidad teniendo en cuenta la justificación realizada por Gauss.

En las demás secciones es derivado varias peculiaridades de la distribución normal como los estadísticos descriptivos de la distribución por medio de la función generadora de momentos, construir la tabla de probabilidad por medio de la serie de Taylor, serie asintótica y el método de Hastings, simular una distribución normal en Excel mediante el método de la función inversa, y las aplicaciones centrándose en el comportamiento del coeficiente intelectual de la población.

En el tercer capítulo, se trata la distribución chi-cuadrado iniciando con las características de esta distribución como su forma y su relación con la distribución normal, luego se deduce matemáticamente la función de densidad teniendo en cuenta las aportaciones de Helmert (1876), su primer creador, y para seguir con el redescubrimiento de esta función en el documento de Pearson (1900). Similar al anterior capítulo, es derivado característica de la distribución como sus estadísticos por medio de la función generadora de momentos, construir la tabla de probabilidades, simular en Excel utilizando la relación con la normal y el método de la función inversa, y describir las aplicaciones dando hincapié al uso en la relación del rendimiento académico con los factores personales.

El cuarto capítulo, consiste en la distribución t-Student descrita similar a los dos anteriores capítulos: características, historia, deducción de la función de densidad, derivación de los estadísticos descriptivos, creación de tablas, simulación y aplicaciones.

Por último, en el quinto capítulo, se dan las conclusiones resultantes de una profunda interpretación de los textos investigados y conocimientos previos.

5. Metodología

Para el desarrollo de trabajo primero se buscó documentos que describieran el enfoque histórico, la deducción matemática, propiedades y aplicaciones de las distribuciones, luego de realizar una lectura detallada para interpretación y búsqueda de relaciones entre ellos. Lo cual implicó la traducción parcial de varios de ellos, además estudiar fuentes adicionales de conceptos matemáticos que se ameritaba para una mejor comprensión.

Al final organizar de manera adecuada la información mediante reseñas, redes temáticas e históricas, demostraciones y simulaciones electrónicas.

6. Conclusiones

El trabajo hace la presentación de una cronología del desarrollo histórico de tres de las distribuciones más relevantes de la inferencia estadística implicadas en proceso muestreo y aporta elementos de conocimiento de un profesor matemáticas sobre su desarrollo e importancia.

El estudio de la Historia de las distribuciones de muestreo permite al profesor de matemáticas estudiar cuidadosamente distintas ideas, esto lleva a una interpretación más amplia de estas distribuciones que en su momento serán de importancia y utilidad en la labor docente. Algunos hechos relevantes de la historia es el origen de la distribución normal está ligada a la astronomía para estimar los errores de las observaciones, luego al transcurrir varios años es aplicada en otras disciplinas; también las tres distribuciones fueron descubiertas de forma independientes por varios autores, difiriendo en pocos en el año de publicación, así con los conocimientos que se tenía de la época permite llegar a esta distribución independiente de la persona, debido que varios individuos pueden tener un razonamiento parecido como se evidencia en la curva normal que fue descubierta por Moivre, Laplace y Gauss; la Chi-cuadrado fue encontrada por Helbert y Pearson; y la t-Student por Luroth y Gosset.

El estudio realizado describe procesos de simulación asociados a cada una de las distribuciones consideradas, que pueden ser tomadas por el docente como recurso para introducir o presentar explicaciones ligadas al comportamiento estocástico del muestreo asociado a variables aleatorias.

En esta monografía el lector puede encontrar métodos para generar tablas de distribuciones de probabilidad con diferentes niveles de precisión tales como los basados en series de Taylor, la aproximación de Simpson, entre otras.

Existen diversas aplicaciones en diferentes ciencias para cada distribución, dos de ellas son, hacer inferencias sobre una población por medio de una muestra aleatoria y, describir el comportamiento de un estadístico.

Las tres distribuciones presentadas están relacionadas unas a otras, debido que para su deducción matemática es necesario tener en cuenta dos propiedades, la chi-cuadrado es el cuadrado de una normal, la t-Student es una normal sobre una raíz de una chi-cuadrado. Es decir, no se puede obtener estas dos distribuciones sin la distribución normal debido a sus relaciones. Las dos propiedades mencionadas anteriormente no solo permiten la deducción de las distribuciones sino también es de ayuda para crear las simulaciones electrónicas de estas.

Elaborado por:	López Gaitán, Carlos Alfredo		
Revisado por:	Fernández Hernández, Felipe Jorge		

Fecha de elaboración del Resumen:	7	15	2018
--	---	----	------

Contenido

1.	Presentación	1
	Introducción	1
	Justificación	1
	Objetivos	4
	General	4
	Específicos	4
2.	Distribución Normal.....	4
	Presentación	4
	Surgimiento y construcción	7
	Derivación de los estadísticos descriptivos	21
	Creación de tablas	24
	Simulación	29
	Aplicaciones.....	31
3.	Distribución χ^2	32
	Presentación	32
	Surgimiento y construcción	34
	Derivación de los estadísticos descriptivos	41
	Creación de tablas	43
	Simulación	46
	Aplicaciones.....	47
4.	Distribución t-Student	48
	Presentación	48
	Surgimiento y construcción	49
	Derivación de los estadísticos descriptivos	56

Creación de tablas	58
Simulación	60
Aplicaciones.....	61
5. Conclusiones	63
6. Bibliografía.....	65

Lista de tablas

Tabla 1. Presentación de la distribución normal con sus propiedades.....	6
Tabla 2. Interpretación del coeficiente de la curtosis.....	23
Tabla 3.Radio de convergencia de la serie de Maclaurin para la distribución normal.....	26
Tabla 4.Cuando la serie alterna es convergente dado la posición de sus términos.	26
Tabla 5.Coeficiente de los términos de la serie alterna para la aproximación de la distribución normal.....	27
Tabla 6.Comparación entre el valor dado de la serie asintótica y la función de Excel de la distribución normal.....	28
Tabla 7.Presentación de la distribución chi-cuadrado con sus propiedades.....	33
Tabla 8. Solución de integrales para deducir la distribución chi-cuadrado.....	38
Tabla 9.Regularidad de la solución de integrales para deducir la distribución chi-cuadrado.	
.....	39
Tabla 10. Comparación entre el valor dado del método de Simpson y la función de Excel de la distribución chi-cuadrado.	45
Tabla 11. Comparación entre el valor dado del método y la función de Excel de la distribución chi-cuadrado.	46
Tabla 12. Valores observados entre número de comidas y rendimiento académico.	47
Tabla 13. Valores esperados entre números de comidas y rendimiento académico.	47
Tabla 14. Presentación de la distribución t-Student con sus propiedades.	48
Tabla 15. Comparación entre el valor dado del método de Simpson y la función de Excel de la distribución t-Student.	60
Tabla 16. Comparación entre las simulaciones de las distribuciones normal y t-Student ...	61
Tabla 17. Falta el nombre de esta tabla	62

Lista de imágenes

Imagen 1. Gráficas de distribuciones normales.....	6
Imagen 2 Distribución triangular discreta.	13
Imagen 3 Distribución circular	13
Imagen 4 Simulación de la distribución uniforme.	30
Imagen 5 Simulación de la distribución normal.....	31
Imagen 6 Distribución del coeficiente intelectual	32
Imagen 7. Gráficas de distribuciones chi-cuadrado.....	34
Imagen 8 Simulación de la distribución chi-cuadrado.	46
Imagen 9 Gráficas de la distribución t-Student	49
Imagen 10 Valor observado vs valor teórico de la distribución t-Student.	55
Imagen 11 Simulación de la distribución t-Student.....	60

1. Presentación

Introducción

Las distribuciones normales, chi-cuadrado y t-Student, son funciones continuas importantes para el estudio de distintas poblaciones.

El orden de presentación de estas distribuciones está dado por la escala cronológica de su aparición, cada una de ellas describe como surgen en diferentes ámbitos, pues cada capítulo contara con la presentación de la distribución, como y quien la construyó en la historia, la justificación matemática, deducción matemática de los estadísticos, maneras de construir la tabla de probabilidades, simulación mediante Excel, y sus aplicaciones en diferentes ciencias.

En el inicio de los tres primeros capítulos se presenta una tabla en la cual se encuentran la función de densidad, media, varianza, asimetría y curtosis, de la distribución que se va a tratar. Las dos secciones siguientes se justifica las propiedades encontradas en la tabla anteriormente mencionada, en la segunda, la función de densidad es demostrada a partir de los documentos de sus creadores y los comentarios de Hald (2007) sobre su contexto histórico, y en la tercera sección la derivación de los estadísticos descritos en la tabla mencionada anteriormente por medio de los momentos absolutos y propiedades metamatemáticas. En las últimas tres secciones se realiza referente a la distribución del capítulo la construcción de la tabla de probabilidades, simulación en Excel y aplicaciones diferentes ciencias.

Justificación

Desde los documentos del Ministerio de Educación Nacional (2006), el desarrollo del Pensamiento Aleatorio tiene como objetivo ayudar a tomar decisiones en momentos de incertidumbre, de azar o de ambigüedad en donde hace falta información confiable. A su vez la toma de decisiones bajo incertidumbre se apoya en teorías de la probabilidad que considera distribuciones teóricas y de muestreo; las cuales se enseñan poco en la secundaria, en contraste con su aparición en currículos de carreras universitarias. En ambos ámbitos, procurar una mejor comprensión de este tema es un asunto que requiere atención; por

ejemplo, aunque en libros de estadística universitaria se presentan y se describen distribuciones como la normal, t-Student y chi-cuadrado, no es frecuente que se explique su origen y la manera como surgen desde los fenómenos estocásticos, sino que se enfatiza más en la exposición de interpretaciones y aplicaciones a la solución de problemas. En el mismo sentido en textos escolares no es frecuente que se expliciten conexiones, fenómenos y el origen histórico de distribuciones como la normal. Es de suma importancia el contexto histórico para la comprensión de objetos matemáticos como rectifica Bell (1985) “ningún tema pierde tanto cuando se le divorcia de su historia como las Matemáticas” (p.54).

A nivel universitario, autores como Hernán & Jiménez (2006) exponen la importancia de realizar las demostraciones en los libros estadísticos y como estas demostraciones ayudan al estudiante a dar validez a teorías científicas. Aunque hay teoremas como el del límite central que requieren conocimiento matemático muy exigente, Hernán y Jiménez (2006) aseguran que la demostración matemática sirve como herramienta al estudiante para la formación de un pensamiento lógico del individuo, cuando se proponen diferentes maneras de evidenciar un hecho matemático demostrable.

En ese sentido este trabajo de grado pretende recrear el origen histórico de las distribuciones normal, t-Student y Chi-cuadrado e ilustrar métodos mediante los cuales se genera tales distribuciones, más que presentar demostraciones formales que expliquen cómo llegar a éstas.

En este orden de ideas, el presente trabajo quiere aportar algunos elementos en los que se recurre a la historia, para que el lector se entere de cómo la distribución normal, t-Student y Chi-cuadrado surgieron, en qué contextos aparecieron, y cómo fue su apropiación y asimilación en la comunidad científica y así enriquecer el concepto de tales distribuciones más allá de su aplicación rutinaria. Por ejemplo, como se muestra en el documento de Student (1908), la distribución para la media en pequeñas muestras obtenidas de una población no se ajusta a la distribución normal, hallazgo que respondió al problema de no poder hacer juicios sobre la media en una muestra pequeña, pues se observó que el nivel de incertidumbre era alto cuando se utilizaba la normal. Al respecto, la historia documenta que este autor planteó una tabla de probabilidades modificada años después por el mismo Student como narra Lehman (1999), además Student, encuentra la función de densidad que describe el

comportamiento de la varianza conocida como la distribución chi-cuadrado, y junto con la distribución normal obtiene la t-Student, mediante el cociente entre una variable normal y raíz de una chi-cuadrado.

Actualmente, textos como Walpole (2012) muestran a partir de la función generadora de momentos de la distribución normal y el teorema de la unicidad, se obtiene la distribución de chi-cuadrado (ver colorario 7.1 de este texto).

Por otra parte, el trabajo también quiere aportar una mirada a la manera como pueden aparecer estas distribuciones a partir de la generación aleatoria de números, tema el cual está relacionado con la simulación de distribuciones con métodos como el de Montecarlo¹. Por ejemplo, se mostrará cómo puede obtenerse una distribución normal con base en la utilización de generadores aleatorios en programas como Excel. Esto puede ser de especial utilidad para maestros que quieran mostrar ejemplos de simulaciones de distribuciones estadísticas con métodos complementarios a los que provee Excel a través de complementos como el de “Herramientas para análisis” o con base en fórmulas de generación de números pseudoaleatorios tales como: “=aleatorio ()”.

Finalmente, otro aspecto que suele llamar la atención por la forma en que se termina enseñando es cómo efectivamente calcular probabilidades con estas distribuciones utilizando tablas. Como es bien conocido, es común el uso de tablas para estas distribuciones, dada la complejidad de sus cálculos directos, porque sus expresiones analíticas no se tienen antiderivadas en términos elementales.

En este trabajo se quiere explicitar y aportar métodos que se utilizan en programación, para realizar tales cálculos; en particular, los resultados de estos métodos pueden ser comparados con los resultados que se pueden obtener con las fórmulas que tienen internamente en software de Excel.

¹ El método de Montecarlo es un método basado en la generación aleatoria de números, usado para aproximar distribución estadísticas complejas y difíciles de evaluar con exactitud. Su nombre se atribuye al Casino de Montecarlo que era “la capital de los juegos de azar”, tales como el juego de la ruleta.

Objetivos

General

- Elaborar una reseña comprensiva acerca del surgimiento de las distribuciones Normal, Chi-cuadrado y t-Student que esboce su recorrido histórico, presente ejemplos de su aplicación e ilustres métodos analíticos y de simulación que las generen.

Específicos

- Esbozar un recorrido histórico de cada distribución.
- Realizar procesos de simulación para generar cada distribución.
- Identificar y aplicar algoritmos de aproximación que permitan el cálculo de las funciones de estas distribuciones.
- Proponer ejemplos donde se apliquen las diferentes distribuciones.

2. Distribución Normal

Presentación

La distribución normal es la reina de las distribuciones por su importancia en diferentes áreas, también conocida como distribución de Gauss o de Laplace-Gauss. Es una de las distribuciones de probabilidad de variable continua que aparece con más frecuencia en la teoría y práctica de análisis de datos y probabilidad.

Esta distribución modela varios fenómenos naturales, sociales, físicos, económicos y psicológicos que provienen diferentes ciencias exactas, experimentales y sociales. Además, se constituye en una herramienta de análisis inferencial por su gran utilidad para las pruebas de hipótesis y la estimación por intervalos de diferentes tipos de parámetros, por ejemplo, los asociados a métodos de regresión.

Para calcular la probabilidad de una variable aleatoria X cuyo comportamiento sea normal, por ejemplo, x tome valores entre a y b , se debe encontrar la integral de la distribución normal entre dichos valores y resolverla. En términos simbólicos esto es:

$$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Como termina siendo conocido por los estudiosos de la estadística, resolver esta integral es imposible si se intenta encontrar una antiderivada de la función del integrando; teniendo en cuenta que hasta la fecha no se ha encontrado su función primitiva. Este hecho, hace necesario aproximar este tipo de integraciones por diversos métodos numéricos².

En la práctica, el resultado de aplicar dichos métodos tiene como producto la tabulación de las áreas asociadas o que representan probabilidades con base en tablas que se suelen presentar en los anexos de los libros. En el caso de la Normal, realizar este tipo de tablas para cada valor de μ y σ , no sería una tarea procedente, sin embargo, la solución de este problema es transformar todas las observaciones de la variable aleatoria normal X en un conjunto de observaciones de una nueva variable normal Z , su transformación es $Z = \frac{x-\mu}{\sigma}$, esta nueva variable Z tiene media $\mu = 0$ y desviación estándar $\sigma = 1$, y es llamada distribución normal estándar o normal tipificada.

En particular, si X toma un valor x_n , entonces el valor de Z será $z_n = \frac{x_n - \mu}{\sigma}$. Así, la probabilidad al sustituir en la integral es:

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz = P(z_1 < Z < z_2)$$

De esta manera solo se debe construir una tabla de la probabilidad de la distribución normal estándar para evaluar cualquier distribución normal.

Una de las principales características de un comportamiento normal, a veces reseñada como regla empírica, es contener el 68,26% de los datos en intervalo $(\mu - \sigma, \mu + \sigma)$, 95,44% de los datos en el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$ y el 99,74% en el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$. En otras palabras, es notar que el porcentaje de valores comprendidos en los intervalos de la

² En un apartado posterior es comentado algunos de estos métodos.

forma $(\mu - k\sigma, \mu + k\sigma)$, con $k \in \mathbb{R}^+$ se puede caracterizar de manera invariante para cualquier valor de μ y σ con esta distribución.

En la tabla 1 se resume sus principales características de centralidad, dispersión y forma:

Función densidad	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Esperanza	μ
Varianza	σ^2
Coeficiente de asimetría	0
Curtosis	0
Función generadora de momentos	$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Tabla 1. Presentación de la distribución normal con sus propiedades.

En cuanto a lo que se refiere al registro gráfico, la función de densidad de la distribución normal tiene forma acampanada y es simétrica respecto la media μ . Su curva es conocida como “Campana de Gauss”. Por ser una distribución simétrica en ella coinciden los valores de la media, mediana y moda, es decir, tiene $M_o = \mu$ y $M_e = \mu$. En la imagen 1 se muestran cuatro funciones de densidad con diferente media y desviación estándar, evidenciando su comportamiento simétrico y acampanado.

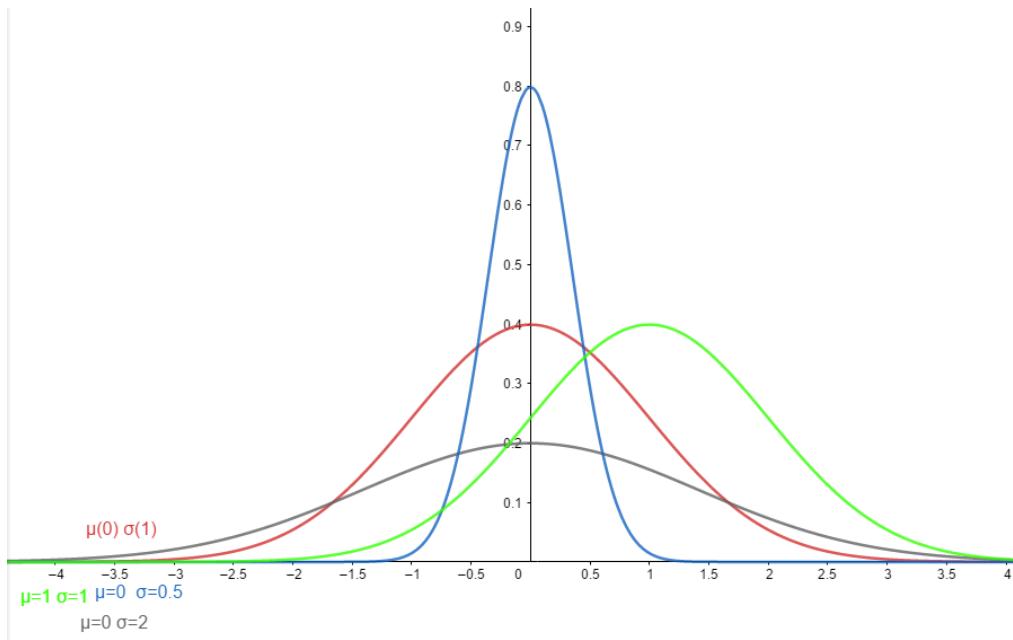


Imagen 1. Gráficas de distribuciones normales

Surgimiento y construcción

La teoría de probabilidades y la teoría de estimación de errores contribuyeron al origen de la distribución normal ya que estas dos teorías fueron utilizadas por Laplace y Gauss para llegar a esta distribución. En los siguientes párrafos se describe detalladamente algunos sucesos relevantes.

La historia de la probabilidad tiene sus primeros inicios en las civilizaciones antiguas, como en Mesopotamia que extraían y pulían el talón de los animales (astrágalo) para hacer un dado de cuatro caras. El imperio romano era muy dado a los juegos de azar; y también en la cultura árabe se han tratado estos juegos como lo sugiere el uso de la palabra azar, que tiene sus orígenes etimológicos en árabe donde “al-zar” significa “dado”.

El estudio de la probabilidad surge con Pacioli 1494 con su obra “Summa de arithmeticā, geometriā, proportioni et proportionalitā” (Suma de la aritmética, geometría, proporciones y proporcionalidad) en el cual da una solución sobre el problema de los tres puntos. Por otra parte, Cardano escribe “Liber de ludo aleae” (libro de los juegos de azar) alrededor de 1620 que no fue publicada si no hasta 1663, estos dos escritos tratan sobre problemas de azar. Pascal y Fermat en su correspondencia trataban de explicar estos mismos problemas como el de los tres puntos, estas cartas fueron conocidas por Christian Huygens que fueron publicadas en 1657 en su obra “Van rekeningh en spelen van gluck” (De los razonamientos en los juegos de azar).

En la época de Jacobo Bernoulli (1654-1705) se entendía la probabilidad como la cantidad de eventos favorables sobre el total de los eventos posibles (definición clásica). Bernoulli (1713) en “Ars Conjectandi” concluye que al realizar observaciones bajo las mismas condiciones y reportarlas en una tabla de frecuencias relativas, este número de la frecuencia relativa comienza a hacer estable con mayor cantidad de observaciones. Él concluye dos cosas importantes, primero, el modelo estadístico que representa la situación es la distribución binomial, y segundo, se puede utilizar la teoría de probabilidad en otras ciencias de manera que las frecuencias relativas sean estables.

Autores como Conde (2015) y Hald (2007) concuerdan que la primera aparición de una distribución normal fue en los documentos De Moivre (1667-1754), donde él descubre esta

distribución al aproximar la probabilidad de la media respecto un radio d con un comportamiento binomial y la cantidad de ensayos es grande. Al calcular la probabilidad por medio de la distribución binomial es una tarea complicada, pues al ser varios ensayos el número combinatorio es demasiado grande y, por el contrario, un número menor a uno elevado a un número grande su resultado es demasiado pequeño, y al multiplicarlos estos resultados debe dar un resultado modesto. Moivre llega a una solución a este problema, al aproximar la distribución binomial por medio de la distribución normal.

$$P_n(d) = P(|x - np| \leq d) = \sum_{x=np-d}^{np+d} \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{2}{\sqrt{2\pi}} \int_0^{\frac{d}{2\sqrt{npq}}} e^{-u^2} du$$

Moivre para encontrar la aproximación comienza con una propiedad de la distribución binomial, al tener cualquier valor de p y el número de ensayos n es “grande”, entonces la distribución binomial tiende a ser simétrica. Así que él descarta la aproximación para cualquier valor p , y se centra solo en la distribución simétrica, como se muestra a continuación.

$$b(x, n, p) \approx b\left(m + d, 2m, \frac{1}{2}\right)$$

$$b(m + d) = \binom{2m}{m + d} (2)^{-m}$$

El problema de aproximar la distribución binomial se centra en aproximar el número combinatorio y como el número combinatorio está conformado de factoriales, el problema se traduce en aproximar el número factorial; por lo cual Moivre (1730) presenta una estimación en su documento *Miscellanea Analytic* que es una agrupación de varios escritos entre 1721 y 1730.

$$k! \approx \sqrt{2\pi k} \cdot (k)^k e^{-k}$$

Hoy es conocida como la fórmula de Stirling, aunque Moivre para llegar a esta aproximación tuvo que hacer varios intentos y utilizar hechos de otros autores. Moivre (1730) describe la aproximación del máximo valor de la binomial utilizando la binomial simétrica y la serie de Newton del logaritmo, para concluir:

$$\ln b(m) \approx \left(2m - \frac{1}{2}\right) \ln(2m - 1) - 2m \ln(2m) + 0.7739$$

Elevando al exponencial, se obtiene la aproximación

$$b(m) = \binom{2m}{m} 2^{-m} \approx 2.168 \frac{\left(1 - \frac{1}{2m}\right)^{2m}}{\sqrt{2m - 1}}$$

Stirling en 1725 también trabaja con este problema y concluye que el valor verdadero de la aproximación de 2.168 que aparece en la anterior formula es $e\sqrt{2/\pi}$. En Stirling (1730) se encuentra la aproximación de $\ln n!$, en el cual él identifica la constante $\sqrt{2\pi}$ para cualquier valor n , esta es la razón por la cual la aproximación del número factorial lleva su nombre.

En 1733 Moivre realiza un pequeño artículo titulado “*Approximatio ad summan terminorum binomii $(a + b)^n$ in serien expansi*” solo hay nueve de ellos y su circulación es privada, afortunadamente en *The doctrine of chances* de 1738 en la segunda y 1756 en la tercera edición se encuentra una copia de este importante documento. Moivre (1730) obtiene otra aproximación del término máximo del binomio, utilizando propiedades de números grandes como $\left(1 - \frac{1}{2m}\right)^{2m} \approx e$, $\sqrt{2m - 1} \approx \sqrt{2m}$ y además utiliza la constante encontrada por Stirling en el cual él lo cita; por tanto, se concluye $b(m) \approx 1/\sqrt{\pi m}$. Este resultado también se observa con la fórmula de Stirling, al aproximar el número combinatorio.

$$\binom{2m}{m} 2^{-2m} = \frac{2^{-2m} (2m)!}{m! m!} \approx \frac{\sqrt{2\pi 2m} \cdot (2m)^{2m} e^{-2m} 2^{-2m}}{\sqrt{2\pi m} \cdot (m)^m e^{-m} \sqrt{2\pi m} \cdot (m)^m e^{-m}} = \frac{1}{\sqrt{\pi m}}$$

La aproximación de la distribución binomial mediante la expansión asintótica encontrada por Moivre en 1730, y la aproximación por series del logaritmo natural descubierta por Newton 1660 que hoy es llamada serie de Mercator-Newton, él concluye mediante estas dos series la aproximación del cociente de dos binomiales simétricas. Como se observa Moivre (1756) “Corolario 1... m o $\frac{1}{2}n$ es una cantidad infinitamente grande, entonces el logaritmo de la razón, que es un término distante del medio por el intervalo l , hasta el término medio, es $-\frac{2ll}{n}$,” (p.237). Escribiendo en notación del documento $2m = n$ y $ll = d^2$

$$\ln \frac{b(m+d)}{b(m)} \approx -\frac{d^2}{m}$$

Aunque Moivre no describe detalladamente la prueba en Hald (1990) demuestra el resultado utilizando la expansión de $\ln(1 + X)$, en el cual, es basada en los hallazgos de Moivre. Moivre en el corolario 1 no describe las condiciones de d , pero sí lo hace en el corolario 2, proporciona una relación entre d y \sqrt{n} , que es $\frac{d}{\sqrt{n}}$.

Con la aproximación obtenida del corolario 1 al despejar y aplicar la función exponencial en ambos lados, y considerando $b(m) \approx \frac{1}{\sqrt{m\pi}}$, se determina la aproximación de la binomial simétrica para un valor $m + d$.

$$b(m + d) \approx \frac{1}{\sqrt{m\pi}} e^{-\frac{d^2}{m}}$$

Al remplazar $2m = n$ y al sumar las probabilidades de una distancia d respecto al máximo valor, se obtiene en notación moderna:

$$\sum_{|x-n/2| \leq d} b\left(x, n, \frac{1}{2}\right) \approx \frac{4}{\sqrt{2\pi}} \int_0^{d/\sqrt{n}} e^{-2y^2/n}$$

En el corolario 8 de Moivre (1756) sin ninguna prueba determina el máximo valor para una distribución binomial:

$$b\left(\frac{n}{2}, n, p\right) = \frac{1}{\sqrt{2pqn\pi}}$$

Aunque Moivre no justificó esta parte, es posible argumentarla por medio de la fórmula de Stirling, como se observa a continuación.

$$\binom{n}{np} p^{np} q^{n-np} \approx \frac{\sqrt{2\pi n} \cdot n^n e^{-n}}{\sqrt{2\pi np} \sqrt{2\pi nq} np^{np} e^{-np} nq^{nq} e^{-nq}} = \frac{1}{\sqrt{2\pi npq}}$$

En el corolario 9 determina la aproximación del cociente entre dos binomiales, Moivre (1756) p.242 dice "Si las Probabilidades de pasar y no estar en cualquier proporción dada de

desigualdad, los problemas relacionados con la Suma de los términos del binomio ($a + b$) se resolverán con la misma facilidad que aquellos en los que las probabilidades de pasar y fallar están en una relación de igualdad" y continúa definiendo la siguiente propiedad omitiendo la prueba.

$$\ln \frac{b(np+d)}{b(np)} \approx -\frac{d^2}{2npq}$$

Solucionando la ecuación y sustituyendo a $b(np)$ por $\frac{1}{\sqrt{2\pi npq}}$, se obtiene el resultado esperado.

$$b(np+d, n, p) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{d^2}{2npq}}$$

Esta es la primera aparición de la distribución normal en su forma de aproximación de la distribución binomial. Moivre también deduce la suma de las probabilidades de varios eventos es calculada por medio de una integral definida.

$$P_n(d) = P(|x - np| \leq d) \approx \frac{2}{\sqrt{2\pi}} \int_0^{\frac{d}{\sqrt{npq}}} e^{-\frac{u^2}{2}} du$$

A continuación, se describe los sucesos que contribuyeron a la distribución normal por medio de la curva de errores. Galileo Galilei en su documento de 1632 llamado *Dialogue Concerning the two Chiefly System of the world-Ptolemaic and Copernican*, propone leyes sobre los errores en observación de cuerpos celestes como narra Hald (1990):

1. Solo una distancia que es verdadera
2. Hay error en los instrumentos de medición, observador y otras condiciones
3. Las observaciones están distribuidas simétricamente por el verdadero valor
4. Los errores pequeños son más frecuentes que los grandes errores

Estas propiedades son base para la construcción de la teoría del error, aunque no se conoce que las comunidades de astrónomos compartían estos hallazgos; en Gauss (1809) escribe

estos principios y agrega que el mejor estimador de la distancia verdadera es la media aritmética.

La estimación de una curva que describa el comportamiento de los errores fueron propuestas por Thomas Simpson (citado por Stahl, 2007) en el cual propone una distribución uniforme y otra triangular, en su documento de 1756 es una de las primeras discusiones sobre la curva que modela los errores, y describe sobre el mejor estimador.

Es bien sabido por su Señoría que el método practicado por los astrónomos, para disminuir los errores que surgen de las imperfecciones de los instrumentos y de los órganos de los sentidos, tomando la Media de varias observaciones, no ha sido generalmente recibido, pero que algunas personas, de nota considerable, han sido de opinión, e incluso se han mantenido públicamente, que una sola observación, tomada con el debido cuidado, era tan confiable como la Media de un gran número. (p.101)

El define que hay un límite respecto a las medias de las observaciones e identifica que es proporcional a los errores respecto a la media, cuando son más distantes de la media disminuye la probabilidad, cuando se acerca a la media tiene mayor probabilidad, teniendo que las observaciones fueron realizadas con las mismas condiciones. Aunque, no graficó la proporción se puede interpretar la distribución del error como:

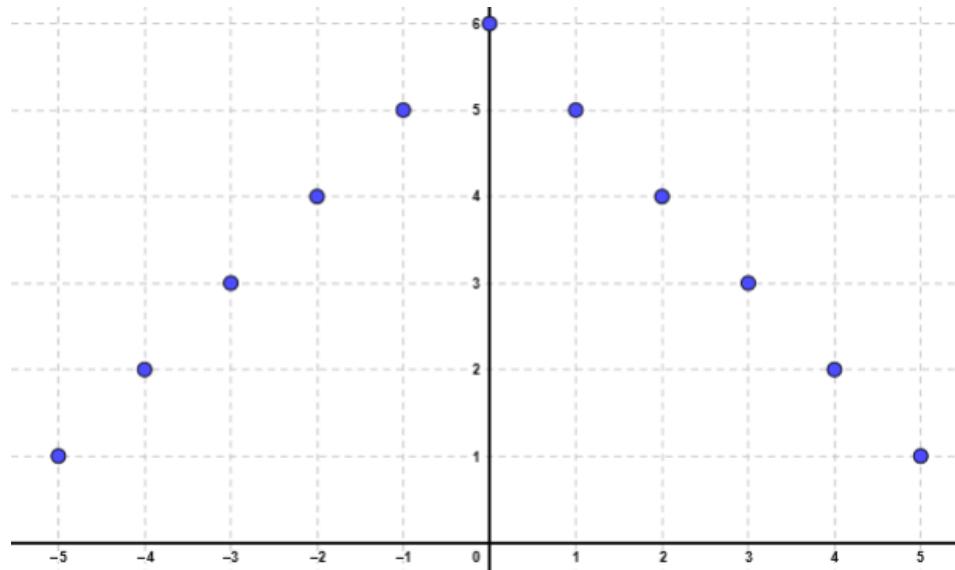


Imagen 2 Distribución triangular discreta.

Luego, como narra Stahl (2007) Simpson un año después, determinó esta distribución para el intervalo continuo, teniendo el mismo comportamiento de la imagen 2.

Daniel Bernoulli (1700,1782) estableció que es universalmente aceptado la media aritmética para estimación de una observación, el aceptó que la curva del error debería tener un límite, el máximo punto la tangente debe ser horizontal y la tangente es casi vertical en los extremos del dominio, así que él introdujo la curva del error como una semielipse y luego lo remplazó por una semicircunferencia.

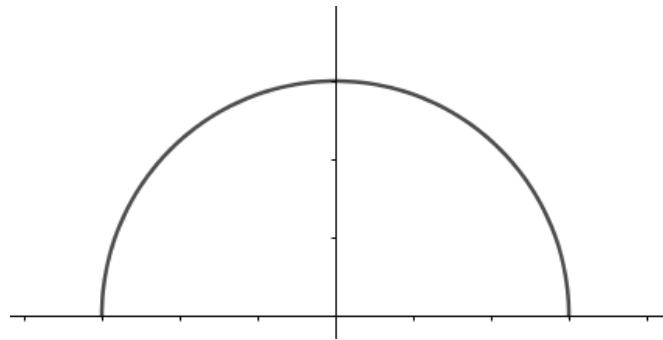


Imagen 3 Distribución circular

Laplace demuestra y generaliza el teorema sobre la aproximación de la distribución binomial de Moivre en 1812 “*Théorie analytique des probabilités*” (*Teoría analítica de probabilidades*) pero no compara estos resultados con la curva de los errores, por el contrario, propone curvas de errores.

En la época de Laplace era común que la media aritmética era el estimador del verdadero valor, pero a un faltaba la formulación con argumentos, uno de los intereses de él era argumentar este hecho. En 1774 Laplace introduce dos curvas de errores, la primera es la doble exponencial y la segunda utiliza el logaritmo natural.

$$\phi(x) = \frac{m}{2} e^{-m|x-\theta|}$$

Y

$$y = \frac{1}{2a} \ln \frac{a}{|x|} \quad -a \leq x \leq a$$

Como narra Stahl (2007) esta ecuación al tratar de utilizarlas tiene muchas operaciones complejas y solo fue utilizada para estimar la órbita de Venus.

En la antigüedad se conocía la existencia de los planetas Mercurio, Venus, Tierra, Marte, Júpiter y Saturno; en 1766 Daniel Titius plantea una hipótesis que relaciona la distancia de los planetas con el sol y en 1781 fue descubierto Urano rectificando esta ley. Pero según esta ley se encontraba un planeta entre Marte y Júpiter pero no se había encontrado hasta que el 1 de enero de 1801 el astrónomo Giuseppe Piazzi descubre un nuevo planeta llamado Ceres (en la actualidad es llamado planeta enano), luego de seis semanas se esconde atrás del sol no permitiendo su observación, varios científicos de la época tratan de predecir donde y cuando volverá a poder ser observado, el que logró la predicción más exacta fue Gauss calculando la órbita del planeta, él utilizó el método de los mínimos cuadrados y la curva de errores. El interés de Gauss de obtener la distribución normal no surge por la aproximación de la distribución binomial, sino para estimar el error en las mediciones astronómicas y justificar el método de los mínimos cuadrados.

En varios documentos se discuten sobre el primero en llegar al método de los mínimos cuadrados; si fue Gauss o Lagrange, porque Lagrange utiliza este método para determinar las cinco medidas de los arcos meridianos que componen el total del arco a través de París, y se usa para determinar el medidor estándar, o Gauss para determinar las órbitas de los planetas.

Gauss publica su revolucionario libro de “Teoría del movimiento de los cuerpos celestes” en 1809 en la cual expone la demostración de la ley de errores: la probabilidad de obtener un

error ε es dado por la función $\phi(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-(h\varepsilon)^2}$. También Laplace demostró en “Teoría analítica de probabilidades” de 1811 la probabilidad de los errores $\pm t$ donde un elemento es sensible proporcional a la exponencial e^{-ht} , pero en si le esta distribución es atribuida a Gauss.

Gauss utiliza el método de máxima verosimilitud para determinar la función ϕ , en el cual este método era conocido por matemáticos de la época, aunque en ocasiones se suele confundir que el primero en desarrollarlo fue Fisher, pero en realidad él publicó su definición formal.

Los dos postulados esenciales para la deducción matemática de la distribución normal, son formulados por Gauss (1809), que tiene una gran parecido a las leyes de Galileo de 1632 descritas en párrafos preliminares.

- Primer postulado: la probabilidad de obtener un error de magnitud ε es proporcional a la magnitud del error. Así existe una función que describe este comportamiento.
- Segundo postulado: a magnitud del error tiene la misma probabilidad de caer en ambos costados.

Gauss adoptó otra manera para decir el segundo postulado, pero él con el tiempo decidió el anterior mencionado. A continuación, se presenta el postulado, y la explicación de su equivalencia.

Obtenidos por observación n valores distintos de una magnitud, siempre que todos hayan sido obtenidos en las mismas circunstancias y con el mismo esmero, el valor más probable de la magnitud es la media aritmética de los valores observados.

La equivalencia de los dos postulados es justificada en los posteriores párrafos comenzando con el segundo postulado que el error tiene la misma probabilidad de caer en ambos costados, otra manera equivalente de decir esta afirmación es que la suma de los errores observados es cero, o en lenguaje matemático, si $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ valores observados, entonces:

$$\sum_{i=1}^n \varepsilon_i = 0$$

Se tiene l_1, l_2, \dots, l_n las magnitudes de medir a l , que se han observado con el mismo método y las mismas condiciones. El error se define como $\varepsilon_i = l_i - l$ donde $i \in \mathbb{Z}^+$

Se suma todos los valores

$$\sum_{i=1}^n l_i - l = nl - \sum_{i=1}^n l_i$$

Como consecuencia del segundo postulado $\sum \varepsilon = 0$, se tiene

$$nl - \sum_{i=1}^n l_i = 0$$

$$l = \frac{\sum_{i=1}^n l_i}{n}$$

Como el error se define con la magnitud verdadera y nunca se podrá saber este valor, entonces se utiliza el segundo postulado; diciendo que la media aritmética es el mejor estadístico para estimar el valor verdadero.

Así los errores observados respecto a los errores verdaderos poseen el mismo comportamiento, aunque esto carece de rigurosidad matemática, es un hecho que en la experimentación se cumple.

Gauss define $\phi(\varepsilon)$ como la función que modela la probabilidad del error, es diferenciable, es simétrica y con único máximo. La probabilidad de obtener dos errores será el producto de las probabilidades, porque los errores son independientes, esto se puede decir por la experimentación; en otras palabras, lo toma como una verdad.

$$\phi(\varepsilon_1 \cap \varepsilon_2) = \phi(\varepsilon_1) \cdot \phi(\varepsilon_2)$$

Al querer obtener la probabilidad del error en todo el intervalo, Gauss (1809) utiliza un procedimiento análogo al de máxima verosimilitud, comenzando con multiplicar los errores obtenidos de n observaciones.

$$g = \phi(\varepsilon_1 \cap \varepsilon_2 \cap \dots \cap \varepsilon_n) = \prod_{i=1}^n \phi(l_i - l)$$

Como no se sabe el error verdadero de l es utilizado la media aritmética, se deriva la función g respecto a l utilizando la derivada de un producto de varias funciones y se evalúa en la media aritmética:

$$\frac{\partial g}{\partial l} \Big|_{l=\bar{l}} = \sum_{i=1}^n - \left(\phi'(l_i - \bar{l}) \prod_{\substack{j \neq i \\ j=1}}^n \phi(l_j - \bar{l}) \right)$$

Un truco utilizado comúnmente es la multiplicación por 1, que en este caso $1 = \frac{\phi(l_i - \bar{l})}{\phi(l_i - \bar{l})}$

$$\frac{\partial g}{\partial l} \Big|_{l=\bar{l}} = - \prod_{i=1}^n \phi(l_i - \bar{l}) \sum_{i=1}^n \frac{\phi'(l_i - \bar{l})}{\phi(l_i - \bar{l})}$$

Como se quiere determinar el máximo valor, se evalúa la derivada en cero y $\prod_{i=1}^n \phi(l_i - \bar{l})$ es diferente de cero, pues son las probabilidades de cualquier error y es diferente de cero, se concluye:

$$\sum_{i=1}^n \frac{\phi'(l_i - \bar{l})}{\phi(l_i - \bar{l})} = 0$$

Gauss (1809) supone que los errores que comete en una observación son iguales excepto uno, en notación matemática: l_1 y $l_2 = l_3 = \dots = l_n$, en el cual existe un número real N , talque, $l_n = l_1 - nN$, al hacer esta suposición la media aritmética se convierte a

$$\bar{l} = \frac{l_1 + (n-1)(l_1 - nN)}{n} = l_1 - (n-1)N$$

Al sustituir \bar{l} en $\sum_{i=1}^n \frac{\phi'(l_i - \bar{l})}{\phi(l_i - \bar{l})}$ se obtiene

$$\frac{\phi'((n-1)N)}{\phi((n-1)N)} + \frac{(n-1)\phi'(-N)}{\phi(-N)} = 0$$

Al despejar.

$$\frac{\phi'((n-1)N)}{\phi((n-1)N)} = - \frac{(n-1)\phi'(-N)}{\phi(-N)}$$

Como $\phi(\varepsilon)$ es simétrica es lo mismo que una función par, además la derivada de una función par es una función impar, y al utilizar estos dos hechos se obtiene la siguiente igualdad.

$$\frac{\phi'((n-1)N)}{\phi((n-1)N)} = (n-1) \frac{\phi'(N)}{\phi(N)}$$

Por la igualdad obtenida se afirma que el cociente de ϕ' y ϕ es una función homogénea, pues una función homogénea es $f(ax) = a^t f(x)$ en el cual t representa el grado de homogeneidad, en este caso es de grado uno.

Como es función homogénea de grado uno debe ser una función lineal y además es una función continua por ser $\phi(\varepsilon)$ diferenciable, y $\phi(\varepsilon) \neq 0$

$$\frac{\phi'(\varepsilon)}{\phi(\varepsilon)} = k\varepsilon$$

En el cual $k \in \mathbb{R}$

Se integra en ambos lados de la igualdad

$$\int \frac{\phi'(\varepsilon)}{\phi(\varepsilon)} \cdot d\varepsilon = \int k\varepsilon \cdot d\varepsilon$$

$$\ln(\phi(\varepsilon)) = \frac{k \cdot \varepsilon^2}{2} + D$$

$$\phi(\varepsilon) = e^{\frac{k \cdot \varepsilon^2}{2} + D}$$

$$\phi(\varepsilon) = C \cdot e^{-(h\varepsilon)^2}$$

La sustitución de k por h , para escribir mejor la expresión y el k debe ser negativo porque los errores grandes son cada vez menos probables que suceda. Donde $C = e^D$ y $k = -2h^2$

Para determinar el valor de la constante C es encontrada con la característica de una función de distribución que la suma de las probabilidades es igual a 1

$$\int_{-\infty}^{\infty} C \cdot e^{-(h\varepsilon)^2} d\varepsilon = 1$$

Solucionando la integral

$$C = \frac{h}{\sqrt{\pi}}$$

$$\phi(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-(h\varepsilon)^2}$$

Gauss deriva la distribución de h suponiendo $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ ser independientes y las variables normalmente distribuidas con media cero y precisión h , y sea h distribuida uniformemente en $[0, \infty)$... concluye que el más probable valor de h es $\hat{h} = \sqrt{\frac{n}{2\varepsilon\varepsilon}}$ y $\sigma = \frac{1}{h\sqrt{2}}$.

Además, Hald (2017) dice que este método es equivalente al método de máxima verosimilitud, por tanto, se estima el parámetro h por este método.

Las observaciones l_1, l_2, \dots, l_n en el cual se comete los siguientes errores respectivamente $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$; la probabilidad de obtener todos los errores será multiplicarlos, recordando que son independientes los errores

$$\phi' = \prod_{i=1}^n \phi(\varepsilon_i) = \prod_{i=1}^n \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon_i^2} = \frac{h^n}{(\sqrt{\pi})^n} e^{\sum_{i=1}^n (-h\varepsilon_i)^2} = \frac{h^n}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i^2}$$

Se quiere encontrar el máximo valor de los errores obtenidos, para determinarlo se deriva la función respecto a h , en el cual, es la derivada de un producto.

$$\begin{aligned} \frac{d\phi^*}{dh} &= \frac{hh^{n-1}}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} + \left(\frac{h^n}{(\sqrt{\pi})^n} \right) \cdot \left(e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} \right) \cdot \left(-2h \sum_{i=1}^n \varepsilon_i \right) \\ \frac{d\phi^*}{dh} &= \frac{nh^{n-1}}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} + \left(\frac{-2h^{n+1} \sum_{i=1}^n \varepsilon_i}{(\sqrt{\pi})^n} \right) \cdot \left(e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} \right) \\ &= \frac{h^{n-1}}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} \cdot \left(n - 2h^2 \sum_{i=1}^n \varepsilon_i \right) \end{aligned}$$

Igualando a cero

$$\frac{h^{n-1}}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i^2} \cdot \left(n - 2h^2 \sum_{i=1}^n \varepsilon_i \right) = 0$$

Por ser los números reales con las operaciones usuales es un dominio de integridad, se tiene dos soluciones

Primera

$$\frac{h^{n-1}}{(\sqrt{\pi})^n} e^{-h^2 \sum_{i=1}^n \varepsilon_i} = 0 ;$$

Como un número elevado a la e no puede dar cero, entonces $\frac{h^{n-1}}{(\sqrt{\pi})^n} = 0$ lo cual se concluye que $h = 0$ y se descarta porque cancelaría la función de densidad.

Segundo

$$n - 2h^2 \sum_{i=1}^n \varepsilon_i = 0$$

Despejando h

$$h = \sqrt{\frac{n}{2 \sum_{i=1}^n \varepsilon_i}}$$

En notación moderna, se escribe en términos de la desviación estándar

$$h = \frac{1}{\sqrt{2} \sigma}$$

Encontrando la distribución normal.

$$\phi(\varepsilon) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{\varepsilon}{\sigma \sqrt{2}}\right)^2}$$

Laplace en 1810 publica la demostración del teorema del límite central, resaltando su relación con el documento Gauss (1809).

La distribución normal solo fue utilizada por los astrónomos hasta que Gosset aplica la distribución normal no para errores de observaciones astronómicos, sino para otros contextos como sociales, políticos y económicos; además el utilizo el método de mínimos cuadrados y

nombró el método de regresión lineal. Gosset es un autor relevante en este escrito porque es crucial para sentar las bases de la estadística inferencial, como es mostrado en el siguiente capítulo.

Derivación de los estadísticos descriptivos

Para determinar los estadísticos descriptivos (media, varianza, asimetría y curtosis) de la distribución normal es utilizada su función generadora de momentos, por lo cual es lo primero que se derivara. Además, se interpreta estos estadísticos con el comportamiento de la distribución

La función generadora de momentos se define como la esperanza de e^{xt} .

$$E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Se aplica propiedades algebraicas.

$$\begin{aligned} &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x\mu+\mu^2+xt}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x\mu-2\sigma^2xt+\mu^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x(\mu+\sigma^2t)+\mu^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x^2-(\mu+\sigma^2t))^2-(\mu+\sigma^2t)^2+\mu^2}{2\sigma^2}} dx \\ &= e^{-\frac{-(\mu+\sigma^2t)^2+u^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x^2-(\mu+\sigma^2t))^2}{2\sigma^2}} dx \end{aligned}$$

La integral corresponde a una distribución normal, por lo tanto, su resultado es 1 y simplificando es conseguida la función generadora de momentos.

$$= e^{-\frac{-(\mu+\sigma^2t)^2+u^2}{2\sigma^2}} \cdot (1) = e^{-\frac{-\mu^2-2\mu\sigma^2t-\sigma^4t^2+\mu^2}{2\sigma^2}} = e^{\frac{2\mu\sigma^2t+\sigma^4t^2}{2\sigma^2}} = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

La media de la distribución normal es el primer momento, es decir, la primera derivada de la función generadora de momento e igualada $t = 0$.

$$M_x^1(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)$$

$$M_x^1(0) = e^0(\mu + 0) = \mu$$

La varianza es conseguida con el segundo momento y la propiedad de la varianza relacionada con los dos primeros momentos $\sigma^2 = E(x^2) - E(x)^2$.

Se determina el segundo momento por medio la segunda derivada de la función generador e igualada $t = 0$

$$M_x^2(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)^2 + e^{\mu t + \frac{\sigma^2 t^2}{2}} (\sigma^2) = e^{\mu t + \frac{\sigma^2 t^2}{2}} ((\mu + t\sigma^2)^2 + \sigma^2)$$

$$M_x^2(0) = e^0(\mu^2 + \sigma^2) = \mu^2 + \sigma^2$$

Aplicando la propiedad mencionada es obtenida la varianza

$$E(x^2) - E(x)^2 = \mu^2 + \sigma^2 - \mu = \sigma^2$$

El coeficiente de asimetría de Fisher es definido $\gamma_1 = \frac{\mu_3}{\sigma^3}$, donde μ_3 es el momento centrado y σ^3 es la desviación estándar elevada al cubo. Para hallar el momento centrado primero se necesita tener el momento absoluto que es la tercera derivada de la función generadora de momentos e igualada $t = 0$.

$$\begin{aligned} M_x^3(t) &= e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)((\mu + t\sigma^2)^2 + \sigma^2) + e^{\mu t + \frac{\sigma^2 t^2}{2}} (2(\mu + t\sigma^2)\sigma^2) \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)((\mu + t\sigma^2)^2 + \sigma^2 + 2\sigma^2) \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)((\mu + t\sigma^2)^2 + 3\sigma^2) \end{aligned}$$

Al evaluarlo en cero se obtiene el tercer momento absoluto.

$$M_x^3(0) = \mu (\mu^2 + 3\sigma^2)$$

El momento centrado se define como $\mu_3 = E(x - \mu)^3$, al aplicar propiedades algebraicas, propiedades de la esperanza y sustituir los valores de los momentos encontrados anteriormente es comprobado el valor de μ_3

$$\begin{aligned} \mu_3 &= E(x^3) - 3E(x^2)\mu + 3\mu^2 E(x) - \mu^3 = \mu (\mu^2 + 3\sigma^2) - 3\mu(\mu^2 + \sigma^2) + 3\mu^3 - \mu^3 \\ &= \mu^3 + 3\mu\sigma^2 - 3\mu^3 - 3\mu\sigma^2 + 2\mu^3 = 0 \end{aligned}$$

Al remplazar este valor en el coeficiente de asimetría es cero, comprobando la simetría de la distribución normal.

$$\gamma_1 = \frac{0}{\sigma^3} = 0$$

La curtosis describe la concentración alrededor de su media, la siguiente tabla representa como se interpreta sus valores.

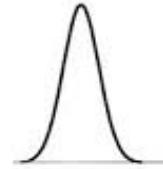
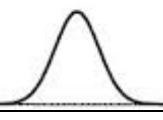
Leptocúrtica, $\gamma_2 < 0$ la concentración es baja	
Mesocúrtica, $\gamma_2 = 0$ la concentración es normal	
Platicúrtica, $\gamma_2 > 0$ la concentración es alta	

Tabla 2. Interpretación del coeficiente de la curtosis.

La curtosis se define matemáticamente $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$, donde μ_4 es el momento central y σ^4 es la desviación elevada a la cuatro. Para calcular la curtosis de la distribución normal es similar a los anteriores procedimientos, por medio de la función generadora de momentos.

$$M_x^4(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} ((\mu + t\sigma^2)^2 + \sigma^2)((\mu + t\sigma^2)^2 + 3\sigma^2) \\ + e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + t\sigma^2)(2\sigma^2(\mu + t\sigma^2))$$

Cuando $t = 0$, es obtenido el cuarto momento.

$$M_x^4(0) = e^0(\mu^2 + \sigma^2)(\mu^2 + 3\sigma^2) + e^0\mu(2\sigma^2\mu) = \mu^4 + 4\mu^2\sigma^2 + 3\sigma^4 + 2\sigma^2\mu^2 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

El cuarto momento central se define $\mu_4 = E(x - \mu)^4$ y se determina utilizando propiedades algebraicas y estadísticas.

$$\begin{aligned}
\mu_4 &= E(x^4) - 4\mu E(x^3) + 6\mu^2 E(x^2) - 4\mu^3 E(x) + \mu^4 \\
&= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - 4\mu^2(\mu^2 + 3\sigma^2) + 6\mu^2(\mu^2 + \sigma^2) - 4\mu^4 + \mu^4 \\
&= \mu^4 + 6\mu^2\sigma^2 + \sigma^4 - 4\mu^4 - 12\mu^2\sigma^2 + 6\mu^4 + 6\mu^2\sigma^2 - 4\mu^4 + \mu^4 = 3\sigma^4
\end{aligned}$$

Se comprueba que la curtosis de la distribución normal es cero, debido que esta distribución es de referencia para el estadístico.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0$$

Creación de tablas

Al generar la tabla de probabilidades de la función de distribución normal estándar es necesario utilizar métodos de aproximación porque no se ha encontrado esta función en términos elementales, por lo cual, en esta sección es presentado tres maneras para construir esta tabla de probabilidad. Los dos primeros métodos (serie Taylor y serie asintótica) son justificados y el tercero no, debido a su complejidad, aunque los tres métodos serán comparados con los valores arrojados por Excel con la función “=distr.norm.estand.n(Z,verdadero)”, con el fin de determinar el error del método o el error de la función de Excel.

El primer método para construcción la tabla de la distribución normal estándar es por medio de aproximación polinómica conocida como serie de Taylor, el primer paso es tener un polinomio que se aproxima a la función de densidad, luego obtener la aproximación de la función acumulada al integrar este polinomio y, por último, evaluar en el resultado de la integral.

Para determinar los coeficientes del polinomio que aproxima a la función de densidad es utilizado la derivada de la función, en este caso la función de densidad está conformada por un exponencial que será eje central para la aproximación, ya que tiene varias propiedades importantes.

Para la función $f(x) = e^x$ es derivable infinitamente y definida en \mathbb{R} , pues $\frac{de^x}{dx} = e^x$.

La serie polinómica de aproximación centrado en un valor a de la función exponencial es:

$$e^x = \sum_{n=0}^{\infty} \frac{e^a}{n!} (x-a)^n$$

Se centra la serie en cero es llamada serie de Maclaurin y es de la forma:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Si se desea una aproximación en los extremos de la distribución es necesario utilizar la serie de Taylor para centrarla en otro valor. En este caso se centró en cero y los valores extremos tienen mayor error.

Como la función de densidad de la distribución normal es $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, al remplazar se obtiene su aproximación polinómica:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^n n!}$$

Para encontrar la aproximación de la función acumulada se considera la integral del normal estándar desde su media a cualquier valor z y se evalúa la integral.

$$\int_0^z \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^n n!} dx = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n! \sqrt{2\pi}} \left(\frac{z^{2n+1}}{2n+1} \right)$$

Antes de evaluar la aproximación se comprueba su convergencia y se determina su radio de convergencia teniendo en cuenta que es una serie alterna. Para realizar esta tarea se utiliza el criterio de la serie alterna: $\sum_{n=0}^{\infty} (-1)^n a_n$ es convergente si el $\lim_{n \rightarrow \infty} a_n = 0 \wedge a_n > a_{n+1}$.

La primera condición del criterio es inmediata:

$$\lim_{n \rightarrow \infty} \frac{z^{2n+1}}{2^n n! \sqrt{2\pi} (2n+1)} = 0$$

Para la segunda condición es definida la desigualdad

$$\frac{z^{2n+1}}{\sqrt{2\pi} 2^n n! (2n+1)} > \frac{z^{2n+3}}{\sqrt{2\pi} 2^{n+1} (n+1)! (2n+3)}$$

Como $z \geq 0 \wedge n \geq 0$ la desigualdad no se cumple para todos los casos, por lo cual, aplicando propiedades para despeje se obtiene los casos en que se cumple.

$$\frac{1}{(2n+1)} > \frac{z^2}{2(n+1)(2n+3)}$$

$$z < \sqrt{\frac{2(n+1)(2n+3)}{2n+1}}$$

A partir de esta desigualdad se concluye que a mayor sea n el radio de convergencia es mayor, la serie comienza a converger teniendo en cuenta el valor de z y el termino del polinomio n , los únicos valores que comienza a converger desde el primer término son $z < \sqrt{6} \approx 2.449489743$, en el cual se logra al remplazar en la desigualdad anterior por $n = 0$. Este hecho también se puede observar en la siguiente tabla (Tabla 3).

z	2,4	2,5
Primer término	0,95746147	0,9973557
Segundo término	-0,91916301	-1,03891219

Tabla 3. Radio de convergencia de la serie de Maclaurin para la distribución normal.

La distribución normal en valores mayores a 4.5 es muy poca su probabilidad y muy rara vez se considera más allá de ese valor, es así solo se necesita nueve términos para que comienza a converger entre estos valores. Para determinar cuando comienza la serie a converger se remplaza la cantidad de términos en la desigualdad anterior como se muestra en la siguiente tabla (Tabla 2).

n	valores
0	$z < 2.449489743$
1	$z < 2.581988897$
2	$z < 2.898275349$
3	$z < 3.207134903$
4	$z < 3.496029494$
5	$z < 3.765875487$
6	$z < 4.019184762$
7	$z < 4.258325179$
8	$z < 4.48527001$

Tabla 4. Cuando la serie alterna es convergente dado la posición de sus términos.

Para determinar el error de una serie alternante es aplicado solo cuando es convergente, por lo cual, se evalúa el error cuando $z \in (0, \sqrt{6})$ pues converge desde el primer término. El teorema de estimación para la serie alternante se define:

$$|R_n| = |\hat{p} - p_n| \leq b_{n+1}$$

Donde R_n es el error, \hat{p} es el valor verdadero, p_n es el valor aproximado y b_{n+1} es el $n + 1$ término de la serie. Es decir que el error es aproximadamente el siguiente coeficiente de la serie. La siguiente tabla (Tabla 3) ejemplifica la afirmación anterior sobre la serie alterna cuando $z \in (0, \sqrt{6})$

Z	0.1	0.3	0.5	0.7	0.9
b_0	0.03989423	0.11968268	0.19947114	0.2792596	0.35904805
b_1	-6.649E-05	-0.00179524	-0.0083113	-0.0228062	-0.04847149
b_2	9.9736E-08	2.4236E-05	0.00031167	0.00167626	0.00588929
b_3	-1.1873E-10	-2.5967E-07	-9.276E-06	-9.7782E-05	-0.0005679
b_4	1.1543E-13	2.2721E-09	2.2546E-07	4.6582E-06	4.4722E-05
b_5	-9.4447E-17	-1.6731E-11	-4.6116E-09	-1.8675E-07	-2.9638E-06
b_6	0	1.0618E-13	8.1295E-11	6.4525E-09	1.6928E-07
b_7	-4.1227E-23	-5.9156E-16	-1.2581E-12	-1.9573E-10	-8.4882E-09

Tabla 5. Coeficiente de los términos de la serie alterna para la aproximación de la distribución normal

Como la serie alterna conseguida para aproximar las probabilidades fue centrada en cero, cada vez que los valores de z se alejen de él aumentara su error, la tabla 5 de los coeficientes de los polinomios muestra el comportamiento mencionado.

La aproximación por serie asintótica de la función distribución de la normal sirve para los valores que se encuentra en sus extremos. Para obtenerla es utilizado el método por integración por partes.

$$\int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_z^\infty \frac{1}{x} xe^{-\frac{x^2}{2}} dx$$

$$\text{Sustitución: } u = \frac{1}{x} \quad du = -\frac{1}{x^2} \quad dv = xe^{-\frac{x^2}{2}} \quad v = -e^{-\frac{x^2}{2}}$$

$$= \frac{1}{\sqrt{2\pi}} \left[-\frac{1}{x} e^{-\frac{x^2}{2}} \right]_z^\infty + \frac{1}{\sqrt{2\pi}} \int_z^\infty \frac{1}{x^2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left(\frac{e^{-\frac{z^2}{2}}}{z} - \int_z^\infty \frac{e^{-\frac{x^2}{2}}}{x^2} dx \right)$$

$$\int_z^\infty \frac{e^{-\frac{x^2}{2}}}{x^2} dx = \int_z^\infty \frac{1}{x^3} xe^{-\frac{x^2}{2}} dx = \left[-\frac{1}{x^3} e^{-\frac{x^2}{2}} \right]_z^\infty - 3 \int_z^\infty \frac{e^{-\frac{x^2}{2}}}{x^4} dx$$

$$\int_z^\infty \frac{e^{-\frac{x^2}{2}}}{x^{2(n-1)}} dx = \int_z^\infty \frac{x}{x^{2n-1}} e^{-\frac{x^2}{2}} dx = \left[-\frac{1}{x^{2n-1}} e^{-\frac{x^2}{2}} \right]_z^\infty - (2n-1) \int_z^\infty \frac{e^{-\frac{x^2}{2}}}{x^{2n}} dx$$

Al reiterar este procedimiento y encontrando la regularidad, es determinada la siguiente serie:

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \left(\frac{1}{z} - \frac{1}{z^3} + \frac{1 \cdot 3}{z^5} - \frac{1 \cdot 3 \cdot 5}{z^7} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{z^9} - \dots \right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \sum_{n=0}^{\infty} \frac{(-1)^n (1 \cdot 3 \cdot 5 \dots (2n-1))}{z^{2n+1}} \end{aligned}$$

En la siguiente tabla (Tabla 6) se muestra la aproximación de la probabilidad con la serie asintótica utilizando 14 términos y su diferencia con la función de Excel. Al mirar los valores cerca de cero son demasiado lejos de su valor verdadero y si es alejado de cero se consigue una buena aproximación, por lo cual este método es importante para aproxima la probabilidad en los extremos de la curva normal.

Valor	Excel	Asintótica	Diferencia
2,4	-0,008197535924596160	-7,81870760001919E+00	7,810510064094600000
2,6	-0,004661188023718730	-5,25400571568335E-01	0,520739383544616000
2,8	-0,002555130330427980	-3,78745372040865E-02	0,035319406873658600
3	-0,001349898031630100	-2,04770634563037E-03	0,000697808314000268
3,2	-0,000687137937915860	3,78062277661267E-04	-0,001065200215577130
3,4	-0,000336929265676855	3,06930552600139E-04	-0,000643859818276994
3,6	-0,000159108590157553	1,56038891763255E-04	-0,000315147481920807
3,8	-0,000072348043925086	7,20201744579101E-05	-0,000144368218382996
4	-0,000031671241833120	3,16349968510836E-05	-0,000063306238684204
4,2	-0,000013345749015903	1,33416317745196E-05	-0,000026687380790422
4,4	-0,000005412543907735	5,41206628459866E-06	-0,000010824610192333
4,5	-0,000003397673124739	3,39750944139531E-06	-0,000006795182566134

Tabla 6. Comparación entre el valor dado de la serie asintótica y la función de Excel de la distribución normal

Tenga en cuenta si se agrega demasiados términos a la serie asintótica comienza a dar una mala aproximación.

El método de Hastings utiliza la siguiente formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} (b_1 t + b_2 t^2 + b_3 t^4 + b_4 t^5 + b_5 t^5)$$

En la cual: $t = \frac{1}{1+pz}$, $p = 0.2316419$ $b_1 = 0.319381530$

$$b_2 = -0.356563782 \quad b_3 = 1.781477937 \quad b_4 = -1.821255978 \quad b_5 = 1.330274429$$

Un ejemplo con $z = 1.3$, el valor dado por la aproximación es 0.9031994506, mientras Excel da 0.90319951541439 concluyendo que es una buena aproximación. Este método es bastante sencillo, porque solo es remplazar los valores de la ecuación sin la necesidad de sumar demasiados términos.

Simulación

Para simular datos que tengan una distribución normal, se puede recurrir a diferentes métodos como en recrear el teorema del límite central, obteniendo muestras de tamaño mayor a 30 y calcular su media muestral, al realizar este procedimiento varias veces escribiendo las frecuencias de las medias en una tabla o un gráfico se consigue un comportamiento similar a la distribución normal. En este documento se realiza de otra manera, por medio de la función inversa de la distribución normal acumulada y números “aleatorios” dados por el software Excel.

Lo primero es simular una distribución uniforme con ayuda de la función “=aleatorio” de Excel, los valores arrojado están entre [0,1] que es interpretando como una probabilidad. Estos valores no son completamente escogidos al azar, al existir una función que los determina incumpliendo la definición de aleatorio, sin embargo, dice (los compas) no son aleatorios, pero simulan que lo son, por la manera en que los generan definiéndolos como pseudoaleatorios.

Al realizar el histograma de los 500 valores dados por la función “=aleatorio” se comporta de manera uniforme (todos los valores tienen la misma probabilidad de salir), esta conclusión se justifica por medio de oprimir la tecla F9 que actualiza nuevos números pseudoaleatorios, en todas las ocasiones el histograma mostraba el mismo comportamiento uniforme. La siguiente imagen fue escogida al azar para ejemplificar lo mencionado.

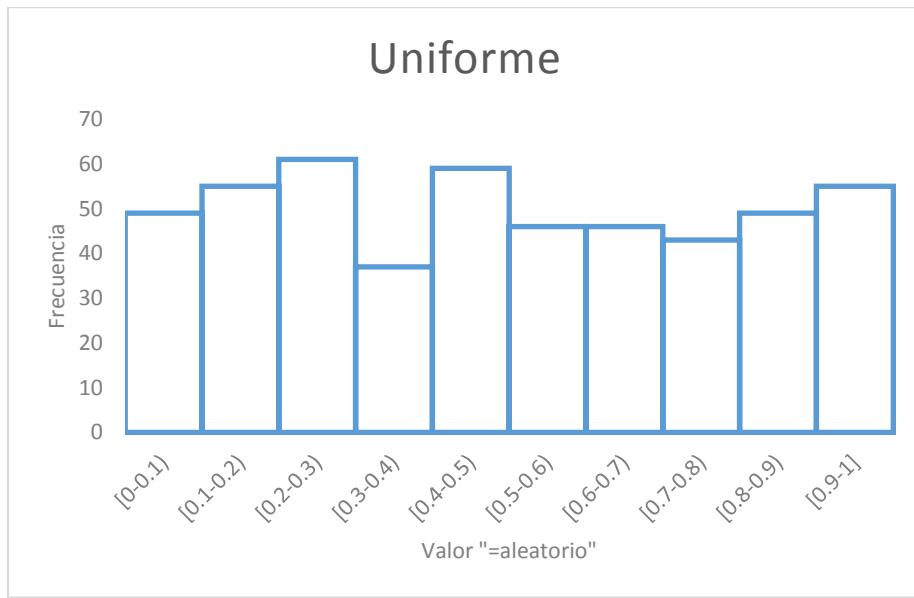


Imagen 4 Simulación de la distribución uniforme.

La función de distribución de la normal estándar inversa acumulada $F^{-1}(z)$ determinar valores de z , teniendo en cuenta que el software utiliza una aproximación de esta función pues no se conoce en términos elementales, como se manifestó en la sección anterior para encontrar $F(z)$; así la función y su inversa se debe obtener por medio de aproximaciones.

El comportamiento acampanado (Imagen 5) lo garantiza la función inversa gracias a su naturaleza, como se representa en el siguiente histograma, para la construcción de este se utilizó los datos de la distribución uniforme de la parte de arriba.

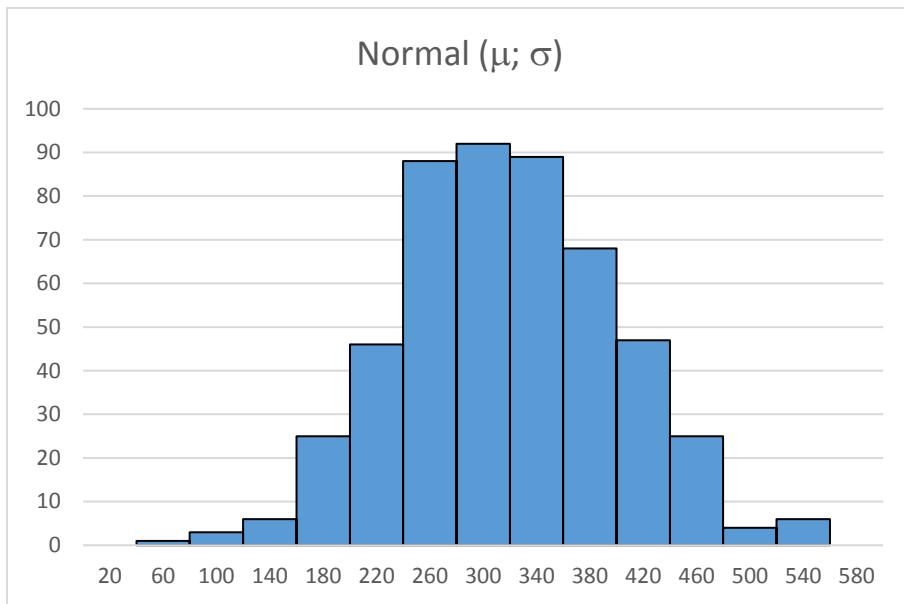


Imagen 5 Simulación de la distribución normal.

Por medio de la función inversa es muy rápido simular una distribución normal, en comparación en otros métodos que son muy elaborados.

Aplicaciones

Una de las aplicaciones más importantes de distribución normal es descrita en los siguientes dos capítulos en los cuales se deduce las distribuciones chi-cuadrado y t-Student, porque, χ^2 esa suma de variables normales al cuadrado y t es una variable normal sobre la raíz de una chi-cuadrado. Además, existe otras aplicaciones en diferentes ciencias como la psicología, eugenésia, física, astronomía y economía.

En la psicología el coeficiente intelectual era medido con el test de Binet, en el cual determinaba la edad mental del individuo y se dividía por la edad cronológica multiplicada por cien, pero insatisfecho Wechsler crea un nuevo test por medio de puntuaciones y observa al graficar varias de estas puntuaciones tiene un comportamiento normal, él determina el coeficiente intelectual se distribuye con una distribución normal con media 100 puntos y una desviación estándar de 15 puntos. Se define varios rangos que determina la clasificación de las puntuaciones del coeficiente intelectual, en la primera desviación respecto a su media se encuentra la mayor parte de la población el 68,26% de los datos, para los demás rangos se resume en la imagen 6 tomada e editada de Stassen (2007).

Distribución teórica de las puntuaciones de C.I.

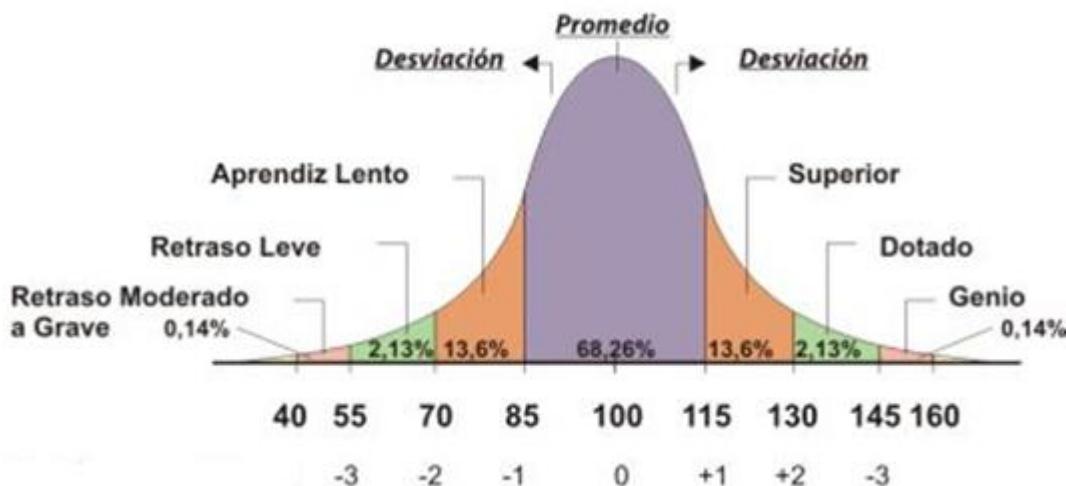


Imagen 6 Distribución del coeficiente intelectual

Al tener una puntuación de una persona es comparado con las puntuaciones dadas por la distribución normal, y se clasifica en su rango. La distribución del CI varía durante los años, pues cada década aumenta en su media 3 puntos, este fenómeno es llamado el efecto Flynn.

Otras aplicaciones de esta distribución en la estadística inferencia es la aproximación de una distribución binomial como se narró en secciones anteriores cuando Moivre llega a la primera distribución normal, en la regresión lineal es utilizada para estimar los parámetros de la función y, es crucial para hallar el intervalo de confianza de la media aritmética.

3. Distribución χ^2

Presentación

La distribución chi-cuadrado es una familia de funciones determinados por sus grados de libertad, como se evidencia en su función de densidad. Esta distribución describe el comportamiento de la suma de variables normales independientes elevadas al cuadrado.

Una de sus aplicaciones es la estimación de una varianza poblacional por medio de intervalos de confianza y prueba de hipótesis, teniendo en cuenta que el éxito de la prueba parte de tener la normalidad en la muestra. Otra aplicación de esta distribución es el test chi-cuadrado,

creado por Pearson (1900) en el cual se compone de prueba de independencia, bondad y ajuste, y homogeneidad, estas pruebas tienen en común el uso del parámetro $\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$ (o_i es observado y e_i el esperado) y la comparación de este parámetro con el valor arrojado de la distribución chi-cuadrado teniendo en cuenta su grado de libertad y su significancia, para así aceptar o rechazar una hipótesis.

A continuación, en la Tabla 7 se presenta la función de densidad de la chi-cuadrado y sus estadísticos que describe su comportamiento.

Función de densidad	$y = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} \cdot x^{\frac{v}{2}-1} e^{-\frac{x}{2}}$ Para $x > 0$
Valor esperado	$u = v$
Varianza	$\sigma^2 = 2v$
Asimetría	$\gamma_1 = \sqrt{8/v}$
Curtosis	$\gamma_2 = \frac{12}{v}$
Función generadora de momentos	$M_x(t) = (1 - 2t)^{-\frac{v}{2}}$ Para $2t < 1$

Tabla 7. Presentación de la distribución chi-cuadrado con sus propiedades

Al aumentar los grados de libertad en la función de densidad se acerca a una distribución normal, como es evidenciado en el gráfico.

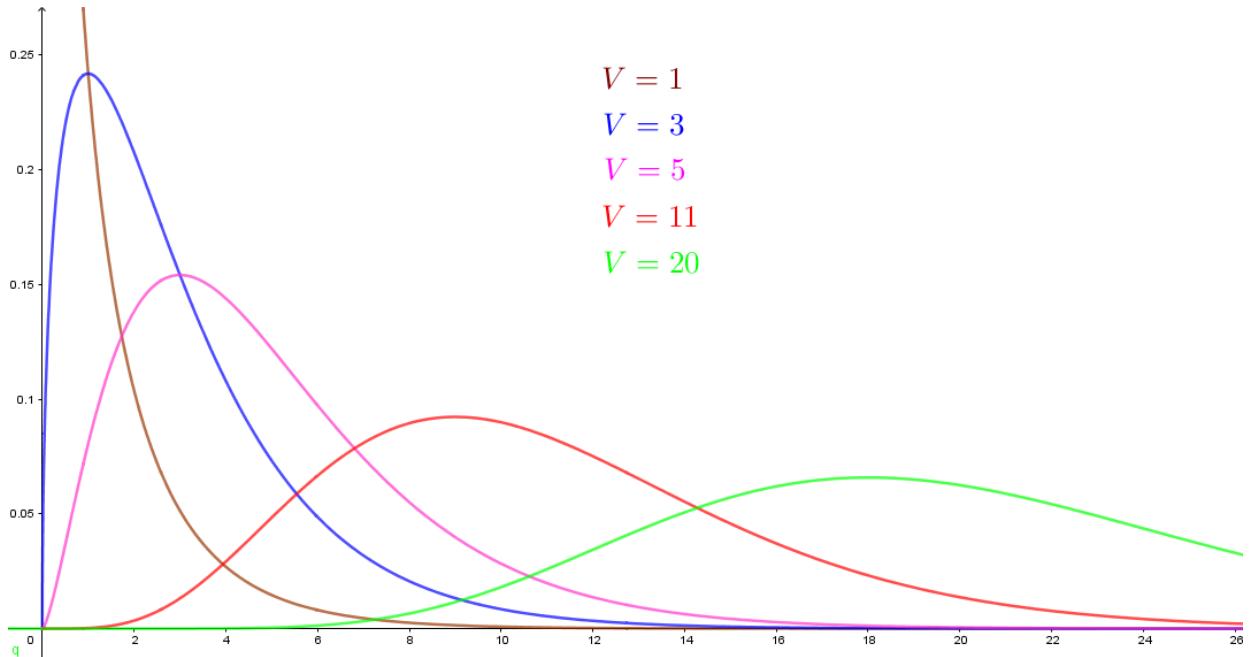


Imagen 7. Gráficas de distribuciones chi-cuadrado.

Surgimiento y construcción

El primero en descubrir la distribución chi-cuadrado fue Helmert en 1876, por medio de encontrar la distribución de la suma de errores elevados al cuadrado, que es exactamente la distribución chi cuadrado, después en el mismo año en su segunda edición menciona $s^2 = \frac{e^2}{n-1}$ y es un estimador de σ^2 . Concluye que $y = e^2 = (n - 1)s^2$, esta distribuido como $\sigma^2\chi^2$ con $v = n - 1$ grados de libertad. A continuación, es presentado la deducción de la distribución chi-cuadrado teniendo en cuenta las narraciones de Hald (2007) y Gorroochurn (2016) sobre la demostración que realizó Helmert.

Un teorema importante en estadística inferencial es: la distribución chi-cuadrado es la suma de variables aleatorias normales independientes; en el cual va a ser esencial para toda demostración. Primero se tiene la función de densidad del error, y con ella determinar la distribución del error al cuadrado.

Sea $y = \varepsilon^2$ donde $\varepsilon = x - \mu$

$$\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{\varepsilon^2}{2\sigma^2}\right)}, \text{ para } \varepsilon \in \mathbb{R}$$

Al sustituir $\varepsilon^2 = y$, y obtener su inversa $\varepsilon = \pm\sqrt{y}$ se logra dos eventos que cumple esta condición y además tienen la misma probabilidad. Para solucionar este problema se multiplica por dos la función de densidad y se tiene en cuenta solo los valores positivos del error. Al realizar esta sustitución y determinar su diferencial, es derivado la distribución chi-cuadrado con un grado de libertad.

$$d\varepsilon = \frac{1}{2\sqrt{y}} dy$$

$$g_1(y) = 2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{y}{2\sigma^2}\right)} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\left(\frac{y}{2\sigma^2}\right)}$$

Definida para $y > 0$

En el caso que

$$y = \varepsilon_1^2 + \varepsilon_2^2 = y_1 + y_2$$

Como se quiere evaluar la probabilidad de obtener un error y , se evalua la integral en términos de y_1 pues $y_2 = y - y_1$. Como son variables independientes se multiplican.

El límite de la integral se define por su dominio teniendo en cuenta la ecuación $Y = Y_1 + Y_2$

$$p(y_1, y_2) = p(y_1) \cdot p(y_2)$$

Si $y = y_1 + y_2$ es una transformación representando la suma de dos variables independientes, al calcular su inversa $y_2 = y - y_1$, por lo cual hacemos la sustitución con su función inversa. Por el teorema 7.4 de Walpole sobre la sustitución de variables en distribuciones continuas.

Al sustituir $y_2 = y - y_1$, el jacobiano es $\frac{\delta(y_1, y_2)}{\delta(y_1, y)} \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$

$$p(y_1, y - y_1) = p(y_1) \cdot p(y - y_1) \cdot 1$$

Al determinar la marginal de y , debido al contexto de la situación el máximo valor que toma los límites de la integral entre 0 , y

$$p(y) = \int_0^y p(y_1) \cdot p(y - y_1) dy_1$$

El teorema de la convolución consiste en hallar una distribución de Y , en el cual esta variable es la suma de dos variables aleatorias independientes $Y = Y_1 + Y_2$. La función de densidad es dada por la integral:

$$h(y) = \int_0^y p_1(y_1) \cdot p_2(y - y_1) dy_1$$

Al utilizar la convolución para determinar la suma de dos variables chi cuadrado con un grado de libertad.

$$\begin{aligned} g_2(y) &= \int_0^y \frac{1}{\sqrt{2\pi\sigma^2}} y_1^{-\left(\frac{1}{2}\right)} e^{-\left(\frac{y_1}{2\sigma^2}\right)} \frac{1}{\sqrt{2\pi\sigma^2}} (y - y_1)^{-\left(\frac{1}{2}\right)} e^{-\left(\frac{(y-y_1)}{2\sigma^2}\right)} \\ &= \frac{1}{2\pi\sigma^2} e^{-\left(\frac{y}{2\sigma^2}\right)} \int_0^y (y - y_1)^{-\left(\frac{1}{2}\right)} (y_1)^{-\left(\frac{1}{2}\right)} dy_1 \end{aligned}$$

Dejaremos la constante aparte, se centra en la solución de la integral

$$I = \int_0^y \frac{1}{\sqrt{y - y_1} \sqrt{y_1}} dy_1$$

Al sustituir $u = \frac{\sqrt{y_1}}{\sqrt{y}}$, y su diferencial $du = \frac{1}{2\sqrt{y}\sqrt{y_1}} dy_1$ o $dy_1 = 2\sqrt{y}\sqrt{y_1} du$

$$\int_0^y \frac{2\sqrt{y}\sqrt{y_1}}{\sqrt{y - yu^2}\sqrt{y_1}} du$$

$$2 \int_0^1 \frac{1}{1 - u^2} du = 2 \operatorname{arcsen}(u)$$

$$2 \operatorname{arcsen} \left(\frac{\sqrt{y_1}}{\sqrt{y}} \right) \Big|_0^y = 2 \operatorname{arcsen}(1) = 2 \frac{\pi}{2} = \pi$$

Por tanto, la suma de dos variables independientes chi-cuadrado con un grado de libertad es la misma distribución, pero con dos grados de libertad.

$$g_2(y) = \frac{1}{2\sigma^2} e^{-\left(\frac{y}{2\sigma^2}\right)}$$

En el caso de sumar tres errores para obtener la función de distribución es necesario realizar la siguiente sustitución, $y = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 = y_1 + y_2 + y_3 = x + y_3$ siendo $x = y_1 + y_2$.

Utilizando el teorema de convolución, similar al caso de anterior suponiendo que x tiene una distribución chi-cuadrado con dos grados de libertad y, $y - x$ tiene una distribución chi-cuadrado con un grado de libertad.

$$\begin{aligned} g_3(y) &= \int_0^y g_2(x)g_1(y-x) dx \\ \int_0^y \left(\frac{1}{2\sigma^2} e^{-\frac{x}{2\sigma^2}} \right) \left(\frac{1}{\sqrt{2\pi\sigma^2}} (y-x)^{-\frac{1}{2}} e^{-\frac{y-x}{2\sigma^2}} \right) dx &= \frac{1}{2\sigma^3\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} \int_0^y (y-x)^{-\left(\frac{1}{2}\right)} dx \\ &= -\frac{1}{2\sigma^3\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} 2\sqrt{(y-x)} \Big|_0^y = -\frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} \sqrt{(y-x)} \Big|_0^y \\ &= \frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} \sqrt{y} \end{aligned}$$

Una manera sencilla de obtener las distribuciones chi-cuadrado es utilizando $g_2(y)$, por ejemplo, para obtener la distribución de $y = y_1 + y_2 + y_3 + y_4 + y_5$, se sustituye $x = y_1 + y_2 + y_3$; $t = y_1 + y_2$. Utilizando el teorema de la convolución y la función de densidad de t es $g_2(t)$, pues es la suma de dos errores al cuadrado independientes, y la función de densidad de x es $g_3(x)$ ya que representa la suma de tres errores al cuadrado independientes.

$$g_5(y) = \int_0^y g_3(x)g_2(y-x)dx$$

Al remplazar

$$\int_0^y \left(\frac{1}{\sigma^3 2^{\frac{1}{2}} \sqrt{\pi}} e^{-\frac{x}{2\sigma^2}} \sqrt{x} \right) \left(\frac{1}{2\sigma^2} e^{-\frac{y-x}{2\sigma^2}} \right) dx = \frac{1}{\sigma^5 2^{\frac{5}{2}} \sqrt{\pi}} e^{-\frac{y}{2\sigma^2}} \int_0^y x^{\frac{1}{2}} dx = \frac{1}{\sigma^5 2^{\frac{3}{2}} \sqrt{\pi}} e^{-\frac{y}{2\sigma^2}} \frac{2y^{\frac{3}{2}}}{3}$$

Es la distribución chi-cuadrado con cinco grados de libertad. Para obtener las demás distribuciones, es utilizado la siguiente convolución

$$g_n(y) = \int_0^y g_{n-2}(x)g_2(y-x)dx$$

Donde $y = y_1 + y_2 + \dots + y_n$; $x = y_1 + y_2 + \dots + y_{n-2}$ y $y - x = y_{n-1} + y_n$

Tenga en cuenta al momento de determinar la distribución solo se integra la variable x , y el resto es una constante. La siguiente tabla (Tabla 7) muestra la integral respecto a x omitiendo las constantes.

Grados de libertad	Par	Grados de libertad	Impar
6	$\int_0^y x dx = \frac{1}{2}y^2$	7	$\frac{2}{3} \int_0^y x^{\frac{3}{2}} dx = \frac{2}{3} \cdot \frac{2y^{\frac{5}{2}}}{5}$
8	$\frac{1}{2} \int_0^y x^2 dx = \frac{1}{2} \cdot \frac{1}{3} y^3$	9	$\frac{2}{3} \cdot \frac{2}{5} \int_0^y x^{\frac{5}{2}} dx = \frac{2}{3} \cdot \frac{2}{5} \cdot \frac{2}{7} y^{\frac{7}{2}}$
10	$\frac{1}{2 \cdot 3} \int_0^y x^3 dx = \frac{1}{2 \cdot 3 \cdot 4} y^4$	11	$\frac{2}{3} \cdot \frac{2}{5} \cdot \frac{2}{7} \int_0^y x^{\frac{7}{2}} dx = \frac{2}{3} \cdot \frac{2}{5} \cdot \frac{2}{7} \cdot \frac{2}{9} y^{\frac{9}{2}}$
12	$\frac{1}{2 \cdot 3 \cdot 4} \int_0^y x^4 dx = \frac{1}{2 \cdot 3 \cdot 4 \cdot 5} y^5$	13	$\frac{2}{3} \cdot \frac{2}{5} \cdot \frac{2}{7} \cdot \frac{2}{9} \int_0^y x^{\frac{9}{2}} dx = \frac{2}{3} \cdot \frac{2}{5} \cdot \frac{2}{7} \cdot \frac{2}{9} \cdot \frac{2}{11} y^{\frac{11}{2}}$
\vdots	\vdots	\vdots	\vdots
n	$\frac{1}{(\frac{n}{2}-2)!} \int_0^y x^{\frac{n}{2}-2} dx = \frac{1}{(\frac{n}{2}-1)!} y^{\frac{n}{2}-1}$	n	$\left(\frac{2}{3}\right)\left(\frac{2}{5}\right)\dots\left(\frac{2}{n-4}\right) \int_0^y x^{\frac{n}{2}-2} dx = \frac{2}{3} \frac{2}{5} \frac{2}{7} \dots \frac{2}{(n-2)} y^{\frac{n}{2}-1}$

Tabla 8. Solución de integrales para deducir la distribución chi-cuadrado

La constante que surge a partir de dar solución a la integral, es posible escribirla como una función Gamma, para ellos se define la función y las propiedades que cumple.

La función Gamma se define como: $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$

Las propiedades usuales de esta función son:

1. $\Gamma(n) = (n-1)!$ Si n es un entero no negativo
2. $\Gamma(x) = (x-1)\Gamma(x-1) = (x-1)(x-2)\Gamma(x-2) = \dots$
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

A partir de estas tres propiedades se deduce otras que van a hacer de ayuda para escribir las constantes. En el caso que n sea par, por la propiedad 1 se tiene: $\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right)!$

Para n impar, se obtiene por la propiedad 2 y 3, $\Gamma\left(\frac{n}{2}\right) = \sqrt{\pi} \left(\frac{1}{2}\right) \left(\frac{3}{2}\right) \dots \left(\frac{n-2}{2}\right)$

Escribiendo el resultado obtenido en la tabla X y agregando sus demás constantes, se obtiene:

Par	Impar
$\frac{1}{2^{\frac{n}{2}} \sigma^n \left(\frac{n}{2} - 1\right)!} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}$	$\frac{1}{2^{\frac{n-1}{2}} \sqrt{\pi}} \frac{2}{3} \frac{2}{5} \frac{2}{7} \dots \frac{2}{(n-2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}$

Tabla 9. Regularidad de la solución de integrales para deducir la distribución chi-cuadrado.

Utilizando las propiedades de la función Gamma mencionadas anteriormente, se obtiene la distribución esperada.

$$\frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \sigma^n} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}$$

En otro documento Helmert estudia nuevamente la distribución de ε^2 y llega a la misma distribución obtenida anteriormente pero con $n - 1$, el al llegado a tratar de decir los grados de libertad, y llega que $y = \varepsilon^2 = s^2(n - 1)$ pues este hecho es verdadero porque $s^2 = \frac{(x-\mu)^2}{n-1}$ y $\varepsilon^2 = (x - \mu)^2$ en el cual $s^2(n - 1)$ esta distribuido como $\sigma^2 \chi^2$ con $v = n - 1$ grados de libertad. Helmert utilizo transformaciones matrices y su transformación, esta afirmación se puede observar mediante la siguiente sustitución de la probabilidad obtenida $g_n(y)$ por $\sigma^2 \chi^2 = y$; $d\chi^2 = \frac{dy}{\sigma^2}$

$$p(\chi^2) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) \sigma^v} (\sigma^2 \chi^2)^{\frac{v}{2}-1} e^{-\frac{\sigma^2 \chi^2}{2\sigma^2}} \cdot \sigma^2 = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} (\chi^2)^{\frac{v}{2}-1} e^{-\frac{\chi^2}{2}}$$

Este importante descubriendo no fue popularizado en la comunidad científica y otros autores independientemente redescubrieron esta distribución, como Karl Pearson en 1900.

El redescubrimiento por parte de Pearson se remonta a su colega a Francis Galton familia de Charles Darwin, en el cual compartían el mismo abuelo, los dos estaban interesados en la evolución, debido que Galton leyó “El origen de las especies”, él publica varios libros sobre este tema y es considerado el creador de la eugenésia (el mejoramiento de los rasgos hereditarios humanos).

El interés de Galton de fue el estudio sobre las leyes de la herencia por métodos estadísticos, además de realizar los estudios aporto a la estadística el termino desviación para decir el error para variación biológicas, aplico la distribución normal a diferentes ciencias, introdujo y utilizo los términos regresión y correlación. En 1901 Galton, Pearson y Weldon fueron los fundadores de la revista científica Biometrika, en el cual, tiene diversas publicaciones revolucionarias para la estadística. Galton ayudo financieramente a la revista, por eso Pearson hizo una biografía de Galton después de su muerte.

Pearson también estaba interesado en la teoría de la evolución, porque publico varios artículos, por ejemplo, en 1898 “frequency constans” consolido la teoría de muestra de la estimación de una distribución multivariable, en el cual lo presento con ejemplos de herencia y biología; también publica una familia de distribuciones debido que en la biología no todo tiene un comportamiento normal, invento un método de estas distribuciones para poder ajustarlas por medio de los momentos y por último, construye un procedimiento para comprobar si los datos siguen una distribución, el famoso ajuste chi-cuadrado.

Pearson (1931) escribe “desviaciones de las muestras de una gran población normal parece haber sido descubierta en siete ocasiones generales, sin duda de forma independiente, y por autores posteriores atribuidos a varios investigadores”, en otro documento sugiere que se llamaría ecuación de Helmert pues fue su primer descubridor.

En 1900 Pearson publica el test de la chi-cuadrado que se le atribuye, él comienza con una distribución multinomial y la transforma en una distribución chi cuadrado, determina el comportamiento de χ^2 en las primeras páginas luego utiliza las desviaciones de una distribución multinomial y otros hechos para llegar al comportamiento del estadístico:

$$\chi^2 = \sum_{i=1}^k \frac{o_i - e_i}{e_i}$$

Es la base para del test en las pruebas de bondad y ajuste: independencia y homogeneidad.

Fisher realiza una corrección sobre su test, debido que no se utiliza en la distribución la cantidad de la muestra sino la cantidad de grados de libertad.

Gosset o Student publica una construcción de la distribución chi-cuadrado por medio de los momentos y la familia de distribuciones de Pearson, en su documento revolucionario sobre la construcción de distribución t-Student.

Derivación de los estadísticos descriptivos

Para determinar los estadísticos se realiza de manera similar al anterior capítulo por medio de la función generadora de momentos, por lo cual se omitirá algunas aclaraciones.

Función generadora de momentos

$$E(e^{xt}) = \int_0^\infty e^{xt} \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} \int_0^\infty e^{-x(1-t)} x^{\frac{n}{2}-1} dx$$

Se realiza la sustitución de $u = x\left(\frac{1}{2} - t\right)$; $du = \frac{1}{2} - t$ en la integral

$$E(e^{xt}) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} \int_0^\infty \frac{e^{-u} 2^{\frac{n}{2}-1} u^{\frac{n}{2}-1}}{(1-2t)^{\frac{n}{2}-1}} \frac{2}{1-2t} du$$

$$E(e^{xt}) = \frac{1}{\Gamma\left(\frac{n}{2}\right) (1-2t)^{\frac{n}{2}}} \int_0^\infty u^{\frac{n}{2}-1} e^{-u} du$$

La integral definida es $\Gamma\left(\frac{n}{2}\right)$, por lo cual se sustituye y se cancela con su equivalente, encontrando la función generadora de momentos de la distribución chi-cuadrado.

$$E(e^{xt}) = \frac{1}{(1-2t)^{\frac{n}{2}}}$$

Media es derivada con el primer momento.

$$M_x^1(t) = -\frac{\frac{n}{2}(1-2t)^{\frac{n}{2}-1} \cdot (-2)}{(1-2t)^{\frac{n}{2}+1}} = \frac{n}{(1-2t)^{\frac{n}{2}+1}}$$

$$t = 0$$

$$M_x(0) = n$$

Se concluye que los grados de libertad es su media.

Varianza es determinada por la propiedad de la varianza y los dos primeros momentos.

$$M_x^2(t) = \frac{(-n)\left(\frac{n}{2} + 1\right)(1-2t)^{\frac{n}{2}} \cdot (-2)}{(1-2t)^{n+2}} = \frac{n(n+2)}{(1-2t)^{\frac{n}{2}+2}}$$

El segundo momento

$$M_x^2(0) = n(n+2)$$

La varianza es doble de sus grados de libertad, $\sigma^2 = E(x^2) - E(x)^2 = n(n+2) - n^2 = 2n$

El coeficiente de asimetría es determinado con la función generadora de momentos y el momento central.

$$M_x^3(t) = -\frac{n(n+2)\left(\frac{n}{2} + 2\right)(1-2t)^{\frac{n}{2}+1}(-2)}{(1-2t)^{n+4}} = \frac{n(n+2)(n+4)}{(1-2t)^{\frac{n}{2}+3}}$$

$$M_x^3(0) = n(n+2)(n+4)$$

Coeficiente de asimetría de Fisher

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

$$\begin{aligned} \mu_3 &= E(x - \mu)^3 = E(x^3) - 3E(x^2)\mu + 3\mu^2E(x) - \mu^3 \\ &= n(n+2)(n+4) - 3(n(n+2))n + 3n^3 - n^3 \\ &= n^3 + 6n^2 + 8n - 3n^3 - 6n^2 + 2n^3 = 8n \end{aligned}$$

$$\gamma_1 = \frac{8n}{(\sqrt{2n})^3} = \frac{4}{\sqrt{2n}} = \sqrt{\frac{8}{n}}$$

Como los grados de libertad se encuentran en el denominador cada vez que sea más grande tenderá a ser simétrica.

La curtosis es determinada por el cuarto momento central y la función generadora de momentos.

$$M_x^4(t) = -\frac{n(n+2)(n+4)\left(\frac{n}{2}+3\right)(1-2t)^{\frac{n}{2}+2}(-2)}{(1-2t)^{n+6}} = \frac{n(n+2)(n+4)(n+6)}{(1-2t)^{\frac{n}{2}+4}}$$

$$M_x^4(0) = n(n+2)(n+4)(n+6)$$

$$\begin{aligned}\mu_4 &= E(x^4) - 4\mu E(x^3) + 6\mu^2 E(x^2) - 4\mu^3 E(x) + \mu^4 \\ &= n(n+2)(n+4)(n+6) - 4n(n(n+2)(n+4)) + 6n^2(n(n+2)) \\ &\quad - 4n^3(n) + n^4 \\ &= n^4 + 12n^3 + 44n^2 + 48n - (4n^4 + 24n^3 + 32n^2) + 6n^4 + 12n^3 \\ &\quad - 4n^4 + n^4 = 12n(n+4)\end{aligned}$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{12n(n+4)}{(\sqrt{2n})^4} - 3 = \frac{(3n+12)}{n} - 3 = \frac{12}{n}$$

Como los grados de libertad se encuentran en el denominador cada vez que sea más grande tenderá a ser mesocúrtica.

Creación de tablas

La función de distribución chi-cuadrado tampoco es encontrada en términos elementales, por esto, es necesario utilizar aproximaciones. Por ejemplo, la regla de Simpson que permite resolver una integral por método numérico en un intervalo comprendido por a y b , en la cual su fórmula es:

$$\int_a^b f(x)dx \approx \frac{b-a}{3n} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + 4f(x_{n-1}) + f(x_n)] \text{ donde } n \in \mathbb{N}$$

La función por aproximar es la función de distribución chi-cuadrado:

$$\int_0^b f(x) dx = \int_0^b \frac{x^{k/2-1} e^{-x/2}}{2^{\frac{k}{2}} \Gamma(k/2)} dx$$

En este caso la regla de Simpson se convierte:

$$\frac{b}{3n} \left[f(0) + 4f\left(\frac{b}{n}\right) + 2f\left(\frac{2b}{n}\right) + 4f\left(\frac{3b}{n}\right) + \dots + 4f\left(\frac{(n-1)b}{n}\right) + f(b) \right]$$

Como $f(x)$ tiene una constante de una función gamma y es necesario evaluarla para obtener la aproximación, por lo cual, es utilizada la siguiente propiedad que fue justificada en secciones anteriores.

$$\Gamma\left(\frac{k}{2}\right) = \begin{cases} k \text{ es par} & \left(\frac{k}{2} - 1\right)! \\ k \text{ es impar} & \sqrt{\pi} \left(\frac{1}{2}\right) \left(\frac{3}{2}\right) \dots \left(\frac{k}{2} - 1\right) \end{cases}$$

Por lo tanto, la fórmula para generar los valores de la tabla se divide en dos casos:

Caso de k sea par

$$\begin{aligned} & \int_0^b \frac{x^{k/2-1} e^{-x/2}}{2^{\frac{k}{2}} \Gamma(k/2)} dx \\ & \approx \frac{b}{3n} \left[0 + 4 \left(\frac{\frac{b^{\frac{k}{2}-1}}{n} e^{-\frac{b}{2n}}}{2^{\frac{k}{2}} \left(\frac{k}{2} - 1\right)!} \right) + 2 \left(\frac{\frac{2b^{\frac{k}{2}-1}}{n} e^{-\frac{2b}{2n}}}{2^{\frac{k}{2}} \left(\frac{k}{2} - 1\right)!} \right) + \dots \right. \\ & \quad \left. + 4 \left(\frac{\frac{(n-1)b^{\frac{k}{2}-1}}{n} e^{-\frac{(n-1)b}{2n}}}{2^{\frac{k}{2}} \left(\frac{k}{2} - 1\right)!} \right) + \left(\frac{b^{\frac{k}{2}-1} e^{-\frac{b}{2}}}{2^{\frac{k}{2}} \left(\frac{k}{2} - 1\right)!} \right) \right] \end{aligned}$$

Caso de k se impár

$$\begin{aligned}
& \int_0^b \frac{x^{k/2-1} e^{-x/2}}{\frac{k}{2^2 \Gamma(k/2)}} dx \\
& \approx \frac{b}{3n} \left[0 + 4 \left(\frac{\frac{b^{\frac{k}{2}-1}}{n} e^{-\frac{b}{2n}}}{\frac{k}{2^2 \sqrt{\pi}} \left(\frac{1}{2} \right) \left(\frac{3}{2} \right) \dots \left(\frac{k}{2} - 1 \right)} \right) + 2 \left(\frac{\frac{2b^{\frac{k}{2}-1}}{n} e^{-\frac{2b}{2n}}}{\frac{k}{2^2 \sqrt{\pi}} \left(\frac{1}{2} \right) \left(\frac{3}{2} \right) \dots \left(\frac{k}{2} - 1 \right)} \right) \right. \\
& \quad \left. + \dots + 4 \left(\frac{\frac{(n-1)b^{\frac{k}{2}-1}}{n} e^{-\frac{(n-1)b}{2n}}}{\frac{k}{2^2 \sqrt{\pi}} \left(\frac{1}{2} \right) \left(\frac{3}{2} \right) \dots \left(\frac{k}{2} - 1 \right)} \right) + \left(\frac{b^{\frac{k}{2}-1} e^{-\frac{b}{2}}}{\frac{k}{2^2 \sqrt{\pi}} \left(\frac{1}{2} \right) \left(\frac{3}{2} \right) \dots \left(\frac{k}{2} - 1 \right)} \right) \right]
\end{aligned}$$

Se introduce esta función en Excel para generar valores de la función acumulada de la distribución chi-cuadrado. En la siguiente tabla muestra la aproximación con 44 términos (si se desea una mejor aproximación debe agregar más términos) y la comparación con el valor de la función de Excel.

Grados de libertad	Valor de x	Probabilidad por método de Simpson	Probabilidad de Excel	Diferencia entre las dos
8	7	0.4633671229651020	0.4633673320992150	0.0000002091341129
8	10	0.7349731891096840	0.7349740847026380	0.0000008955929541
11	10	0.469612873865362	0.469612848998959	0.0000000248664025
11	13	0.706674641132914	0.706674590583821	0.0000000505490930
16	8	0.0511336080181193	0.0511336157928474	0.0000000077747281
17	19	0.671467919771206	0.671467835695571	0.0000000840756346

Tabla 10. Comparación entre el valor dado del método de Simpson y la función de Excel de la distribución chi-cuadrado.

Para aproximar la probabilidad de una variable chi-cuadrado con un número grande en los grados de libertad, se puede utilizar la siguiente formula, que fue encontrada en Jinn (1988):

$$\int_b^\infty \frac{x^{k/2-1} e^{-x/2}}{\frac{k}{2^2 \Gamma(k/2)}} dx \approx \begin{cases} 1 - \frac{1}{2} e^{0.8055(h) + (-0.9911)h^2} & \text{para } h \leq 0 \\ \frac{1}{2} e^{-1.2451(h) + (-0.6763)h^2} & \text{para } h > 0 \end{cases}$$

En el cual $h = \sqrt{b} - \sqrt{k}$. (p.3)

La Tabla 10 compara el valor que da la fórmula con la función de Excel, en el cual se observa que esta fórmula tiene una precisión de dos decimales.

Grados de libertad	Valor de b	Probabilidad por el método	Probabilidad de Excel	Diferencia entre las dos probabilidades
35	30	0.70989814	0.708130951	0.001767189
42	47	0.28507199	0.207614373	0.077457617
54	48	0.70082392	0.703819382	0.002995462
63	70	0.25862031	0.254367305	0.004253005

Tabla 11. Comparación entre el valor dado del método y la función de Excel de la distribución chi-cuadrado.

Simulación

El teorema principal para deducir la distribución chi-cuadrado matemáticamente también ayudara a simular esta distribución, en el cual, consiste en la suma de ν variables normales elevadas al cuadrado tienen un comportamiento chi-cuadrado con ν grados de libertad.

Lo primero es simular un comportamiento normal, esto ya se realizó en el primer capítulo por medio de la función inversa, luego con los valores obtenidos se elevan al cuadrado y se suman de grupos de a ν , en este caso se toma $\nu = 11$, al graficar el histograma 350 variables se obtiene el comportamiento deseado.

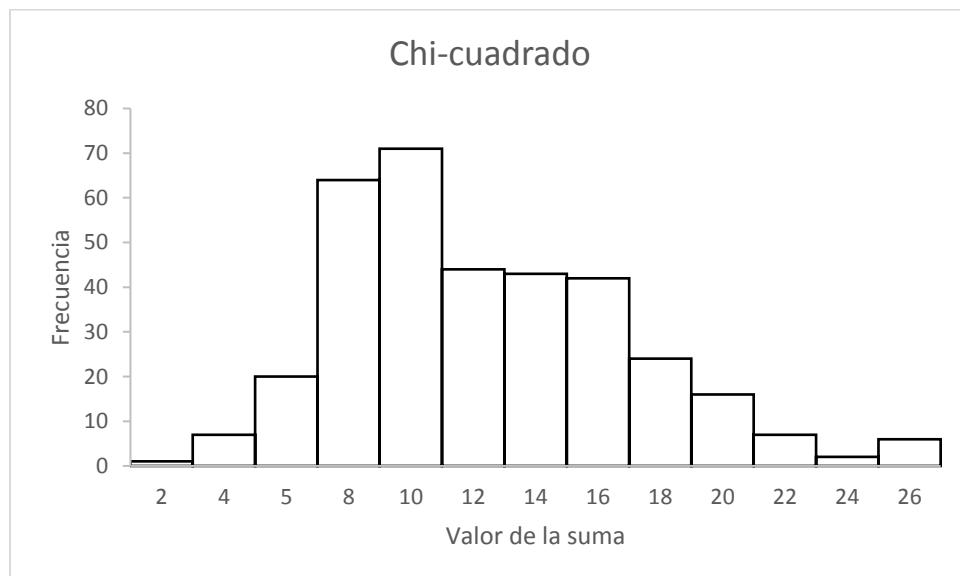


Imagen 8 Simulación de la distribución chi-cuadrado.

Aplicaciones

Las aplicaciones de la distribución chi-cuadrado son: prueba de hipótesis e intervalos de confianza con el fin de estimar la varianza de una población, la prueba de bondad de ajuste consiste en determinar si unos datos tienen una distribución dada y las pruebas de contingencia para inferir si dos variables son independientes.

A continuación, se muestra un ejemplo de la prueba de contingencia sobre el rendimiento académico de estudiantes de Licenciatura de Matemáticas de la Universidad Pedagógica Nacional que cursaron Aritmética en el semestre 2017-2, teniendo en cuenta las cantidades de comidas que consumen al día.

H_0 : el rendimiento académico es independiente con la cantidad de comidas que consume al día el estudiante

H_1 : el rendimiento académico es dependiente con la cantidad de comidas que consume al día el estudiante

Valores observados

Nº comidas\Nota del curso	(0-2.5)	[2.5-35)	[35-50]	Total
1 a 2	1	3	1	5
3	2	5	18	25
4 a 6	3	4	3	10
Total	6	12	22	40

Tabla 12. Valores observados entre número de comidas y rendimiento académico.

Valores esperados

Nº comidas\Nota del curso	(0-2.5)	[2.5-35)	[35-50]
1 a 2	0.7	1.5	2.7
3	3.7	7.5	13.75
4 a 6	1.5	3	5.5

Tabla 13. Valores esperados entre números de comidas y rendimiento académico.

$$\begin{aligned} \chi^2 = & \frac{(1 - 0.7)^2}{0.7} + \frac{(3 - 1.5)^2}{1.5} + \frac{(1 - 2.7)^2}{2.7} + \frac{(3.7 - 2)^2}{3.7} + \frac{(7.5 - 5)^2}{7.5} + \frac{(13.75 - 18)^2}{13.75} \\ & + \frac{(1.5 - 3)^2}{1.5} + \frac{(3 - 4)^2}{3} + \frac{(5.5 - 3)^2}{5.5} = 8.596689546 \end{aligned}$$

El estadístico $\chi^2_{0.1} = 7,77$ con 4 grados de libertad con un nivel de confianza del 90%, lo cual se acepta la hipótesis alternativa, el rendimiento académico es dependiente con la cantidad de comidas que consume al día el estudiante. Se atreve a decir, las personas que consume tres comidas diarias presentan mejor desempeño académico en el curso de Aritmética.

4. Distribución t-Student

Presentación

La distribución t es simétrica respecto a su media, y tiene un comportamiento acampanado como la distribución normal, una de sus aplicaciones son estimar la media de una población normal dado una muestra pequeña, en el caso la muestra se grande la distribución t es similar a la normal. Una de las razones de este hecho es porque esta distribución nace por estimar la media aritmética teniendo en cuenta la distribución de la desviación estándar. En algunos documentos definen la distribución t como una normal sobre la raíz cuadrada de un chi-cuadrado. Otras aplicaciones de esta son en las pruebas de hipótesis y determinación de intervalos de confianza para diferencias de dos medias de la población o los parámetros de regresión lineal.

La distribución t no es solo una, por el contrario, es una familia de distribuciones debido que posee un parámetro llamado grados de libertad, en la cual es representada con la letra v en la Tabla X que describe su función de densidad y algunas propiedades de esta.

Función de densidad	$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)\left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}}$
Valor esperado	0 para $v > 1$, indefinida para otro valor
Varianza	$\frac{v}{v-2}$ para $v > 2$, indefinida para otro valor
Asimetría	0 para $v > 3$
Curtosis	$\frac{6}{v-4}$ para $v > 4$
Función generadora de momentos	No definida

Tabla 14. Presentación de la distribución t-Student con sus propiedades.

Al aumentar los grados de libertad de la distribución t se asemeja cada vez más a la distribución normal estándar, además, una de sus diferencias es al disminuir los grados de libertad de la distribución t el ancho de sus colas aumentan en comparación a la normal estándar que sus colas son más delgadas. Estas dos afirmaciones son observadas en el siguiente gráfico que representa varias distribuciones t con diferentes grados de libertad.

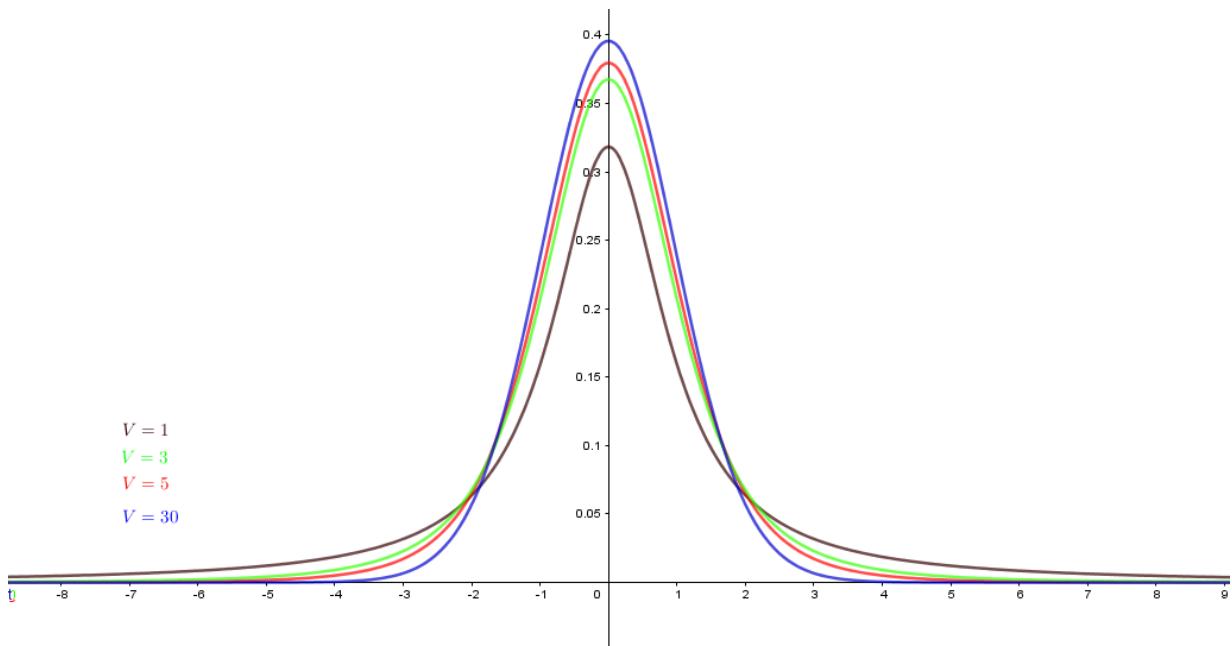


Imagen 9 Gráficas de la distribución t-Student

Surgimiento y construcción

La distribución t-Student aparece en el documento “The probable error of a mean” publicado por William Sealy Gosset en 1908 con su seudónimo Student, su motivación para llegar a esta distribución fue por su trabajó, consistía en mejorar el sabor de la cerveza para la compañía Guinness Bredwery.

En la preparación de la cerveza existe momentos cruciales para determinar el sabor y color, esto lo tenía en cuenta Gosset para variar la cantidad de enzimas, tiempo de tostado, y agregar otros almidones; para luego tomar medidas, pero él no estaba seguro si tenía éxito en mejorar el sabor de la cerveza, debido a las muestras eran tan pequeñas.

Antes de Gosset se había descubierto la distribución t como dice Hald (2007) por Jacob Lüroth en 1876 y Edgeworth en 1883. Luroth siguió el trabajo de Gauss sobre la regresión lineal múltiple, y determinó la distribución de los coeficientes de la recta, llegando a una distribución t, aunque concluye que no difiere tanto de una distribución normal. Edgeworth no conocía el escrito de Luroth, él se planteó una pregunta sobre cuál es los límites del intervalo de la media aritmética con una desviación estándar desconocida, en el cual llega a la distribución t.

El documento antes mencionado por Student (1908) comienza con el problema de estimar la media de poblacional teniendo una muestra pequeña, Gosset estaba enterado del método con la distribución normal para muestras grandes, luego siguen 10 secciones describiendo, justificando y aplicando la distribución t-Student. En la primera sección deduce la distribución chi cuadrado, diciendo que es una curva de Pearson tipo III, esta curva es conocida en la actualidad como una distribución Gamma, que en un caso especial es una distribución chi-cuadrado.

El método para determinar una distribución por medio de las curvas de Pearson que utilizó Gosset consiste en encontrar: media, varianza, asimetría y curtosis, pues son primordiales para determinar los parámetros, sustituirlos y solucionar una ecuación diferencial. Student encontró los momentos de la desviación estándar con el fin de determinar los estadísticos mencionados, y dedujo que probablemente era una curva tipo III, cerró esta sección con la función de probabilidad de la desviación estándar.

Teniendo en cuenta los comentarios y argumentos de Fisher (1925) sobre la distribución t se escribe en notación moderna la varianza muestral y su distribución.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Si s^2 es la varianza muestral de una muestra aleatoria de tamaño n de una población con comportamiento normal y varianza σ^2 , entonces el siguiente estadístico se distribuye como una distribución chi cuadrado:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

La función de probabilidad del estadístico χ^2 es:

$$G(\chi^2) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \int_0^a (\chi^2)^{\frac{\nu}{2}-1} e^{-\left(\frac{\chi^2}{2\sigma^2}\right)} d(\chi^2)$$

Al sustituir χ^2 a términos de s^2 , es decir

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\nu s^2}{\sigma^2}$$

con $\nu = (n - 1)$ que se define como grados de libertad

$$d(\chi^2) = \frac{\nu}{\sigma^2} d(s^2)$$

$$G(s^2) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \left(\frac{\nu}{\sigma^2}\right)^{\frac{\nu}{2}} \int_a^b (s^2)^{\frac{\nu}{2}-\frac{1}{2}} e^{-\left(\frac{(s^2)(\nu)}{2\sigma^2}\right)} d(s^2)$$

Siguiendo con el documento de Student (1908), en la segunda sección argumenta que la media y la varianza no están correlacionados, y con esto supone la independencia de estos dos estadísticos, pero como menciona en Fisher (1925) esta no es una prueba rigurosa. Esta afirmación de la independencia de \bar{x} y s^2 es verdadera como muestra Walpole justificando, “Al obtener muestras de poblaciones normales se puede demostrar \bar{x} y s^2 son independientes y, en consecuencia, también lo son Z y V” (p.247). Helmert un autor importante en el capítulo por ser el primero en deducir la distribución chi-cuadrado, también argumentó la independencia de la media y varianza muestrales en el mismo documento que dedujo la distribución, este trabajo no era conocido por Gosset pues si lo hubiera conocido, no escribiría las dos primeras secciones que se están narrando.

El comportamiento de la media aritmética se asemeja cada vez más a la distribución normal cuando su muestra crece, es el teorema del límite central. La función de densidad que representa el comportamiento de la media aritmética es la distribución normal con media μ y desviación $\frac{\sigma}{\sqrt{n}}$; su expresión algebraica:

$$P(c < \bar{x} < d) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_d^c e^{-\left(\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)} d\bar{x}$$

Al momento de estimar la media aritmética es utiliza la distribución normal suponiendo que la desviación estándar es conocida, cuando no se conoce la desviación estándar se utiliza el estimador s . En muestras pequeñas muy rara vez concuerda la desviación estándar poblacional con la desviación estándar muestral, además no utilizar el valor verdadero de σ sino su estimador s , no asegurara que la media aritmética siga distribuyendo con un comportamiento normal. La nueva variable para estimar la media aritmética es representada con una la letra t :

$$t = (\bar{x} - \mu) \cdot \frac{\sqrt{n}}{s}$$

Al multiplicar por σ y v se obtiene:

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s} = \frac{\frac{(\bar{x} - \mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{vs^2}{v\sigma^2}}}$$

Por lo cual, la distribución de t es el cociente entre una distribución normal y raíz cuadrada de una distribución chi-cuadrado:

$$T = \frac{Z}{\sqrt{\chi^2/v}}$$

Estas dos distribuciones son las dos implicadas para determinar la distribución t , por lo cual, es determinado una función de densidad probabilidad $p(\bar{x}, s^2)$ con los estadísticos implicados. Anteriormente se mencionado la independencia de \bar{x} y s^2 , la definición de Stewart (2012) para variables independientes “[...] si su función de densidad conjunta es el producto de sus funciones de densidad individuales” (p.1010), es decir:

$$p(\bar{x}, s^2) = (\bar{x})g(s^2)$$

Para evaluar la probabilidad de un evento dado los intervalos de la media y la varianza, se determina por medio de una integral doble.

$$P(\bar{x}, s^2) = \iint_A f(\bar{x})g(s^2)d(s^2)d\bar{x}$$

Al sustituir por $t = (\bar{x} - \mu) \cdot \frac{\sqrt{n}}{\sqrt{s^2}}$ y ; $w = s^2$, para la sustitución en integrales dobles se determina el diferencial del volumen $dv = ds^2 d\bar{x}$ por medio del jacobiano.

$$\frac{\delta(\bar{x}, s^2)}{\delta(t, w)} = \begin{vmatrix} \sqrt{\frac{w}{n}} & \frac{t}{2\sqrt{nw}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{w}{n}}$$

por lo cual,

$$P(t, w) = \frac{1}{\sigma\sqrt{2\pi}} \iint_B we^{-\left(\frac{t^2w^2}{2\sigma^2}\right)} \cdot \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2\sigma^2}\right)^{\frac{v}{2}} (w)^{\frac{v}{2}-1} e^{-\left(\frac{ws^2}{\sigma^2}\right)} d(w) dt$$

Como $P(t, w)$ es su función de distribución, y $p(t, w)$ es su función de densidad.

$$p(t, w) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{t^2w^2}{2\sigma^2}\right)} \cdot \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2\sigma^2}\right)^{\frac{v}{2}} (w)^{\frac{v}{2}-1} e^{-\left(\frac{ws^2}{\sigma^2}\right)}$$

Para obtener la probabilidad de t dado que ocurrió el valor de w , se debe encontrar la marginal de t . La distribución marginal se define como $h(t) = \int_{-\infty}^{\infty} h(t, w) dw$, teniendo en cuenta que la varianza es positiva se modifica los límites de la integral y se procede a resolver la integral.

$$\begin{aligned} h(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\infty} se^{-\left(\frac{t^2w^2}{2\sigma^2}\right)} \cdot \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2\sigma^2}\right)^{\frac{v}{2}} (w)^{\frac{v}{2}-1} e^{-\left(\frac{ws^2}{\sigma^2}\right)} d(w) \\ &= \frac{1}{\sqrt{2\pi} \sigma \Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2\sigma^2}\right)^{\frac{v}{2}} \int_0^{\infty} e^{-\left(\frac{w(v+t^2)}{2\sigma^2}\right)} (w)^{\frac{v-1}{2}} d(w) \end{aligned}$$

Al multiplicar por 1 convenientemente con el fin de completar la forma de una función Gamma.

$$\begin{aligned}
h(t) &= \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}} \left(\frac{v}{2\sigma^2}\right)^{\frac{v+1}{2}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{\left(\frac{v+1}{2}\right)} \int_0^\infty (w)^{\frac{v-1}{2}} e^{-\left(\frac{s^2(t^2+v)}{2\sigma^2}\right)} d(w) \\
&= \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \left(\frac{(v+t^2)}{2\sigma^2}\right)^{\left(\frac{v+1}{2}\right)} \int_0^\infty (w)^{\frac{v-1}{2}} e^{-\left(\frac{w(t^2+v)}{2\sigma^2}\right)} d(w) \\
&= \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \left(\frac{v+t^2}{2\sigma^2}\right) \int_0^\infty \left(\frac{w(t^2+v)}{2\sigma^2}\right)^{\frac{v-1}{2}} e^{-\left(\frac{w(t^2+v)}{2\sigma^2}\right)} d(w)
\end{aligned}$$

Al tener la forma de una función gamma, se realiza otra sustitución para el diferencial.

$$\begin{aligned}
u &= \frac{w(t^2 + v)}{2\sigma^2} \quad ; \quad du = \frac{t^2 + v}{2\sigma^2} d(w) \\
h(t) &= \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \int_0^\infty u^{\frac{v-1}{2}} e^{-(u)} d(u) \\
&= \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \int_0^\infty u^{\frac{v+1}{2}-1} e^{-(u)} d(u)
\end{aligned}$$

La integral es $\Gamma\left(\frac{v+1}{2}\right)$

$$h(t) \frac{1}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \Gamma\left(\frac{v+1}{2}\right)$$

Al aplicar propiedad commutativa de la multiplicación, se logra la distribución t-Student de su forma típica.

$$h(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$$

La esencia de la demostración es considerar la distribución del estadístico t, como la coalición entre las distribuciones normal y chi-cuadrado. Al multiplicar las dos funciones por ser

independientes se realiza una sustitución por el parámetro t y por último se llega a la distribución marginal de t .

En las siguientes secciones se discute las dos curvas halladas chi cuadrado y t-Student, y realiza una tabla para la distribución t . Una parte de resaltar de estas secciones es como obtiene inicialmente la distribución antes deducir analíticamente, realiza experimentos extrayendo muchas muestras de tamaño 5 de una población normal, luego las dividía por su desviación estándar, obteniendo el comportamiento de una distribución t , como se muestra en la imagen de Student (1908).

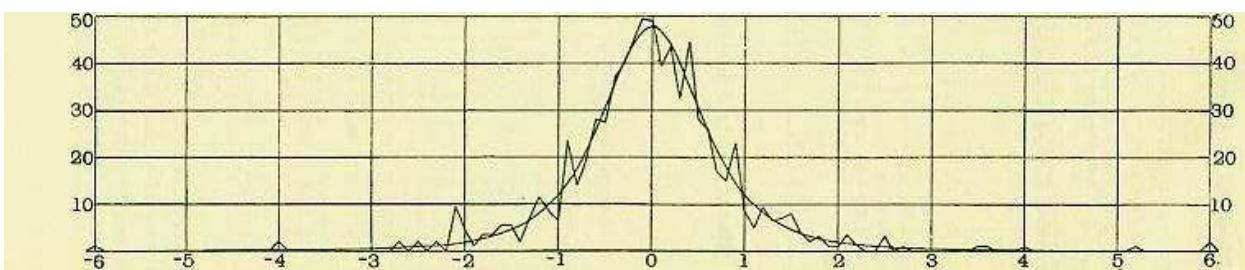


Imagen 10 Valor observado vs valor teórico de la distribución t-Student.

Al final del documento Student (1908) agradece a Pearson “[...] expreso mis agradecimientos al profesor Karl Pearson sin cuyo constante consejo y crítica este documento no podría haber escrito”. Estos dos protagonistas de la estadística moderna junto con Fisher se conocieron porque estaban en la misma universidad de Londres, es aquí donde adquirieron los saberes para trabajar en la estadística moderna. Teniendo en cuenta que en ese momento en la historia solo existía en el planeta el departamento de estadística dirigido por Pearson.

Gosset y Fisher estuvieron hablando varios años, se vio reflejado en el documento Fisher (1925) “Aplicaciones de la distribución de Student” en cual realiza comentarios sobre el documento de Gosset en las justificaciones matemáticas, en otras secciones presentaba aplicaciones de esta distribución para la diferencia de medias y los coeficientes de regresión. Tiempo después estos dos personajes publican una nueva tabla de la t-distribución utilizando expansiones de la integral derivadas por Fisher. Además, Fisher inspirado por Gosset planteo las bases para el análisis de varianza y covarianza.

Derivación de los estadísticos descriptivos

En comparación con los dos capítulos anteriores sobre la deducción de los estadísticos en la distribución t-Student no se realiza por medio de la función generadora de momentos, porque no está definida para determinarlos se utiliza propiedades de las funciones, y las definiciones de momento absoluto y centrado.

Por la naturaleza de la distribución t descrita en secciones anteriores es válido afirmar que el máximo valor que toma la función de densidad de esta distribución es su media, entonces utilizando el método de encontrar máximo por medio de derivar la función de densidad e igualarla a cero.

La siguiente expresión es la derivada de la función de densidad

$$f'(x) = -\left(\frac{v+1}{2}\right) \frac{\Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+3}{2}\right)} \left(\frac{2x}{v}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}}$$

En el caso de igualarla a cero se obtiene como solución cuando $x = 0$ y es un máximo, por tanto, $E(x) = 0$.

Para determinar la varianza es utilizado la definición de segundo momento y dos propiedades importantes de la función Beta. Las dos propiedades son:

$$B(x, y) = \int_0^\infty \frac{u^{x-1}}{(1+u)^{x+y}} du$$

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Es definido el segundo momento y sustituido la su función de densidad.

$$E(x^2) = \int_{-\infty}^\infty x^2 f(x) dx = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \int_{-\infty}^\infty x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx$$

Como la función de densidad de la distribución t-Student es par, se multiplica por 2 y cambia sus límites de integración.

$$E(x^2) = \frac{2\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \int_0^\infty x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx$$

Al sustituir por $u = \frac{x^2}{n}$ $du = \frac{2x}{n} dx$ $c = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}$

$$E(x^2) = c \frac{2}{2} \int_0^\infty u^{\frac{1}{2}} n^{\frac{3}{2}} (1+u)^{-\frac{n+1}{2}} du$$

Se aplican propiedades algebraicas para transformarla en una función beta.

$$E(x^2) = cn^{\frac{3}{2}} \int_0^\infty u^{\frac{3}{2}-1} (1+u)^{\frac{-2-1-n+2}{2}} du = cn^{\frac{3}{2}} \int_0^\infty u^{\frac{3}{2}-1} (1+u)^{-\frac{(2+1)}{2}-\frac{n-2}{2}} du$$

Como es una función beta es aplicado las dos propiedades mencionadas anteriormente.

$$E(x^2) = cn^{\frac{3}{2}} B\left(\frac{3}{2}, \frac{n-2}{2}\right) = \frac{n^{\frac{3}{2}} \Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma(n/2)} \frac{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{3}{2} + \frac{n-2}{2}\right)}$$

Con las propiedades de la función Gamma $\Gamma(1+z) = z\Gamma(z)$; $\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$ y cancelando términos semejantes, es encontrado el segundo momento.

$$E(x^2) = \frac{n\Gamma\left(\frac{n+1}{2}\right) \frac{\sqrt{\pi}}{2} \Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-2}{2} + 1\right) \Gamma\left(\frac{n+1}{2}\right)} = \frac{n}{2\left(\frac{n-2}{2}\right)} = \frac{n}{n-2}$$

La varianza se determina mediante la relación del segundo momento y la esperanza.

$$\sigma^2 = E(x^2) - E(x)^2 = \frac{n}{n-2} - 0 = \frac{n}{n-2}$$

El coeficiente de asimetría de la distribución t es cero porque es simétrica, para justificarlo se utiliza la definición de función par y se tiene en cuenta que la variable x es elevada al cuadrado.

$$\frac{\Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{(-x)^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} = \frac{\Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}}$$

$$f(-x) = f(x)$$

Como la función de densidad es par es decir es simétrica, entonces su coeficiente de asimetría es $\gamma_1 = 0$.

La curtosis es encontrada por medio del cuarto momento, este procedimiento es análogo como se obtuvo el segundo momento. Definiendo el cuarto momento y utilizando las propiedades de la función beta, es determinado $E(x^4)$

$$E(x^4) = \frac{n^2 \Gamma\left(\frac{5}{2}\right) \Gamma\left(\frac{n-4}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{\pi}} = \frac{n^2 \frac{3\sqrt{\pi}}{4} \Gamma\left(\frac{n-4}{2}\right)}{\left(\frac{n-2}{2}\right) \left(\frac{n-4}{2}\right) \Gamma\left(\frac{n-4}{2}\right)} = \frac{3n^2}{(n-2)(n-4)}$$

Es definido el momento central como $\mu_4 = E((x - E(x))^4) = E(x^4)$, porque, $E(x) = 0$. Al escribir la expresión de la curtosis, sustituir por sus equivalencias y simplificarlo, es obtenido su valor.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\frac{3n^2}{(n-2)(n-4)}}{\left(\frac{n}{n-2}\right)^2} - 3 = \frac{3(n-2)}{(n-4)} - 3 = \frac{3n-6-3n+12}{(n-4)} = \frac{6}{n-4}$$

Creación de tablas

El método de Simpson también es utilizado para resolver la integral de la distribución t, similar como se obtuvo la aproximación de la distribución chi-cuadrado.

La función aproximar es:

$$\int_0^b \frac{\Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} dx$$

Como están involucrada dos funciones gamma en la aproximación y es necesario evaluarlas, se define $h(v)$:

$$h(v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}$$

$h(v)$ se compone de dos funciones gamma, para obtener su valor en términos elementales es utilizada propiedades mencionadas en el capítulo anterior, como:

$$\Gamma(x) = (x - 1)\Gamma(x - 1) = (x - 1)(x - 2)\Gamma(x - 2) = \dots$$

Al aplicar esta propiedad $h(v)$ se transforma:

$$h(v) = \frac{v-1}{v-2} \cdot \frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v-2}{2}\right)}$$

Valida desde $v > 3$, en el cual $h(1) = \frac{1}{\sqrt{\pi}}$; $h(2) = \frac{\sqrt{\pi}}{2}$

Cuando v es par

$$h(v) = \frac{v-1}{v-2} \cdot \dots \cdot \frac{7}{6} \cdot \frac{5}{4} \cdot \frac{3}{2} \cdot \frac{\sqrt{\pi}}{2}$$

Cuando v es impar

$$h(v) = \frac{v-1}{v-2} \cdot \dots \cdot \frac{8}{7} \cdot \frac{6}{5} \cdot \frac{4}{3} \cdot \frac{2}{1} \cdot \frac{1}{\sqrt{\pi}}$$

Por tanto, la solución de la integral por medio de la regla de Simpson es:

$$\begin{aligned} & \int_0^b \frac{\Gamma\left(\frac{v+1}{2}\right) \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} dx \\ & \approx \frac{b}{3n} \left[g_v + 4g_v\left(\frac{b}{n}\right) + 2g_v\left(\frac{2b}{n}\right) + 4g_v\left(\frac{3b}{n}\right) + \dots + 4g_v\left(\frac{(n-1)b}{n}\right) \right. \\ & \quad \left. + g_v(b) \right] \end{aligned}$$

En el cual

$$g_v(x) = h(v) \cdot \frac{\left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}}{\sqrt{v\pi}}$$

La Tabla 13 muestra varios ejemplos de la anterior aproximación utilizando 44 termino y comparados con la función de Excel.

Grados libertad	Valor b	Probabilidad método de Simpson	Probabilidad Excel	Diferencia
4	3	0.499937005674111	0.480029015964141	0.019907989709970
4	2.5	0.479345575659162	0.466616727594006	0.012728848065157
15	1.8	0.516284737174721	0.453998940897359	0.062285796277362
15	1.8	0.469375756724456	0.453998940897359	0.015376815827097

Tabla 15. Comparación entre el valor dado del método de Simpson y la función de Excel de la distribución t-Student.

Simulación

La simulación de la t-Student es por medio de la función inversa y los números pseudoaleatorios similar al realizado en el capítulo 1. Al tener los 500 valores distribuidos de forma uniforme se aplica la función inversa acumulada de la t Student con 5 grados de libertad, Excel no arroja valores negativos porque su función inversa acumulada es calculada respecto a su media y da el valor t de la cola derecha. La solución de este problema es escribir en la formula “=distr.t.inv(a15,\$h\$4)*si(aleatorio.entre(0,1)=0,-1,1)” en el cual convierte el signo pseudoaleatorios. El histograma (Imagen X) de los datos obtenido tiene comportamiento acampanado como se muestra a continuación.

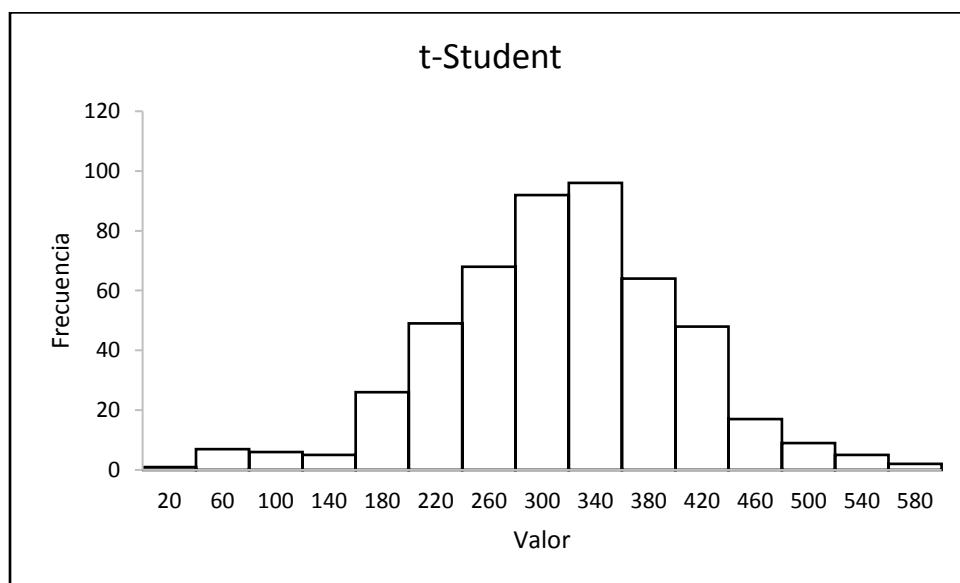


Imagen 11 Simulación de la distribución t-Student.

Al comparar las simulaciones de las distribuciones normal y t-Student, los extremos de la distribución t se encuentra con mayor frecuencia, esto se puede corroborar con las siguientes dos tablas de frecuencias de dos simulaciones.

Normal		t-Student	
Interv	Fre	Interv	Fre
20	0	20	1
60	1	60	7
100	2	100	6
140	8	140	5
180	34	180	26
220	35	220	49
260	79	260	68
300	83	300	92
340	112	340	96
380	80	380	64
420	40	420	48
460	18	460	17
500	6	500	9
540	0	540	5
580	1	580	2

Tabla 16. Comparación entre las simulaciones de las distribuciones normal y t-Student

Otra manera de simular la distribución t-Student es obteniendo muestras de tamaño menor a 30 y calcular su media muestral divida por su desviación estándar muestral, al realizar este procedimiento varias veces y escribiendo las frecuencias de las medias en una tabla o un gráfico se consigue un comportamiento similar a la distribución t-Student, este método fue descrito en la segunda sección del presente capítulo.

Aplicaciones

Las aplicaciones de la distribución t-Student son en la estimación de la media y diferencias de medias para pequeñas muestras, el comportamiento del coeficiente de correlación y en los parámetros de regresión.

El siguiente ejemplo determina la distancia media en metros en que ha sido atrapada la regla por el autor de este documento. El experimento consistió en atrapar entre el índice y la pulgar

una regla marca Faber Castell de una longitud de 30 cm mientras estaba cayendo, cada ensayo contó con la colaboración de dos integrantes, uno que soltara la regla y otro que la atrapara.

Dado que el procedimiento se repitió 15 veces despreciando las medidas más dispersas, se asumirá que la distribución de los datos obedece una distribución de T- Student, con parámetro k . Los datos obtenidos se representan en la Tabla 17.

0,294	0,216	0,218	0,104	0,163
0,25	0,19	0,222	0,197	0,212
0,231	0,184	0,271	0,135	0,223

Tabla 17. Falta el nombre de esta tabla

El promedio de la distancia recorrida es

$$\bar{Y} = 0,207333 \text{ m} \approx 0,207 \text{ m}$$

Evaluación de incertidumbre tipo A de la altura

$$s(y) = \sqrt{\frac{1}{15-1} \sum_{i=1}^{15} (y_i - 0,20733)^2} = 0,04887545 \text{ m} \approx 0,049 \text{ m}$$

$$\mu_A(y) = \frac{s(y)}{\sqrt{n}} = \frac{0,049}{\sqrt{15,0}} = 0,01261958693 \text{ m} \approx 0,013 \text{ m}$$

Evaluación de incertidumbre tipo B de la altura:

La evaluación de incertidumbre tipo B asociada a la longitud de la regla es medio centímetro (0,5 cm) lo que corresponde en metros a 0,005, luego:

$$\mu_B(Y) = 0,005 \text{ m}$$

Evaluación de incertidumbre combinada de la altura:

$$\mu_c(y) = \sqrt{\mu_A(y)^2 + \mu_B(y)^2}$$

$$\mu_c(y) = \sqrt{0,0126196^2 + 0,005^2}$$

$$\mu_c(y) = 0,013574 \text{ m} \approx 0,013 \text{ m}$$

Evaluación de incertidumbre expandida de la altura, está dada por:

$$U_p = k * \mu_c(y)$$

Donde $\mu_c(y)$ es la incertidumbre combinada de la altura y k puede ser el factor de cobertura o el coeficiente de t-Student, dependiendo de la distribución que determine el experimento, en este caso como $n = 15 < 30$, es utilizada una distribución t-Student y calculamos k a partir de:

Grados de libertad: $v = n - 1$; $v = 15 - 1 = 14$

Y un nivel de confianza de 95 % asignado por el experimento, de aquí que:

$$1 - \alpha = 0,95$$

$$\alpha = 0,05$$

$$\alpha/2 = 0,025$$

$$1 - \frac{\alpha}{2} = 0,975$$

Al buscar los valores correspondientes en la tabla de la distribución T-Student se encuentra que el factor de expansión es $k = 2,145$. Con lo que la incertidumbre expandida es:

$$U_p = k * \mu_c(y) = (2,145 * 0,013) = 0,0291162 \approx 0,03 \text{ m}$$

Luego la distancia promedio del autor para atrapar la regla es de $(0,207 \pm 0,03) \text{ m}$ con un nivel de confianza del 95% basado en una distribución de t-Student.

5. Conclusiones

El trabajo hace la presentación de una cronología del desarrollo histórico de tres de las distribuciones más relevantes de la inferencia estadística implicadas en proceso muestreo y aporta elementos de conocimiento de un profesor matemáticas sobre su desarrollo e importancia.

El estudio de la Historia de las distribuciones de muestreo permite al profesor de matemáticas estudiar cuidadosamente distintas ideas, esto lleva a una interpretación más amplia de estas distribuciones que en su momento serán de importancia y utilidad en la labor docente. Algunos hechos relevantes de la historia es el origen de la distribución normal está ligada a la astronomía para estimar los errores de las observaciones, luego al transcurrir varios años es aplicada en otras disciplinas; también las tres distribuciones fueron descubiertas de forma

independientes por varios autores, difiriendo en pocos en el año de publicación, así con los conocimientos que se tenía de la época permite llegar a esta distribución independiente de la persona, debido que varios individuos pueden tener un razonamiento parecido como se evidencia en la curva normal que fue descubierta por Moivre, Laplace y Gauss; la Chi-cuadrado fue encontrada por Helbert y Pearson; y la t-Student por Luroth y Gosset.

El estudio realizado describe procesos de simulación asociados a cada una de las distribuciones consideradas, que pueden ser tomadas por el docente como recurso para introducir o presentar explicaciones ligadas al comportamiento estocástico del muestreo asociado a variables aleatorias.

En esta monografía el lector puede encontrar métodos para generar tablas de distribuciones de probabilidad con diferentes niveles de precisión tales como los basados en series de Taylor, la aproximación de Simpson, entre otras.

Existen diversas aplicaciones en diferentes ciencias para cada distribución, dos de ellas son, hacer inferencias sobre una población por medio de una muestra aleatoria y, describir el comportamiento de un estadístico.

Las tres distribuciones presentadas están relacionadas unas a otras, debido que para su deducción matemática es necesario tener en cuenta dos propiedades, la chi-cuadrado es el cuadrado de una normal, la t-Student es una normal sobre una raíz de una chi-cuadrado. Es decir, no se puede obtener estas dos distribuciones sin la distribución normal debido a sus relaciones. Las dos propiedades mencionadas anteriormente no solo permiten la deducción de las distribuciones sino también es de ayuda para crear las simulaciones electrónicas de estas.

6. Bibliografía

- Bell E. (1985). *Historia de las matemáticas*. México: Fondo de Cultura Económica.
- Bernoulli, J. (1713). *Ars Conjectandi, opus posthumum*. Basilea, Suiza: Reprinted in Editions.
- Conde, A. G. (2015). La distribución normal una rápida revisión histórica. *Heurística, volumen XVII*, pp 59 – pp 65.
- Fisher R. A. (1925). Applications of “Student’s” distribution. *Metrom volumen V*, pp 90 – pp 104.
- Gauss, C. F. (1809). *Theoria motus corporvm coelestivm in sectionibvs conics solem ambientivm*. Hamburgo, Alemania: Petters et Besser.
- Gorroochurn, P. (2016). *Classic Topics on the History of Modern Mathematical Statistics: From Laplace to More Recent Times*. Estados Unidos: Wiley.
- Hald, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. Dinamarca: Wiley
- Hald, A. (2007). *A history of statistical inference from Bernoulli to Fisher, 1713 to 1935*. Dinamarca: Springer.
- Hernán, O. A. & Jiménez, F. N. (2006) La demostración elemento vivo en la didáctica de la matemática. *Scientia Et Technica, volumen XII*, pp 237- pp 240.
- Jinn, T. L. (1988). Approximating the Cumulative Chi-Square Distribution and its Inverse. *Journal of the royal statistical society. Series D. The statistician volumen XXXVII*, pp 3 – pp 5.
- Lehman, E. L. (1999). “Student” and small-sample theory. *Statistical Science, volumen XIV*, pp 418 – pp 426.
- Moivre, A. D. (1730). *Miscellanea Analytic*. Londres, Inglaterra: Tonson & Watts.

- Moivre, A. D. (1756). *The doctrine of chances*. Londres, Inglaterra: Millar.
- 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. *Revista de la Universidad de College, volumen L*, pp 157 – pp 175.
- Pearson, K. (1931) *Tables for Statisticians and Biometricalians*. Londres, Inglaterra: Universidad de College.
- Stahl, S. (2007). The evolution of the normal distribution. *Mathematics Magazine, volumen LXXIX*, pp 96 – pp 113.
- Stassen, B. (2007). *Psicología del desarrollo: infancia y adolescencia*. Bogotá, Colombia: Medica.
- Stewart, J. (2012). *Calculo de varias variables. Trascendentes tempranas 7E*. México: Cengage Learning.
- Stirling, J. (1730). *Methodus Differentialis: sive Tractatus de Summatione et Interpolatione Serierum Infini-tarum*. Londres, Inglaterra: Typis Gut.
- Student. (1908). The probable error of a mean. *Biometrika, volumen XI*, pp 1 – pp 25.
- Walpole, R. E. & Myers R. H. (2012). *Probabilidad y estadística para ingeniería y ciencia*. México: Pearson educación.