CONCEPCIONES QUE SOBRE LA DISTRIBUCIÓN NORMAL TIENEN ALGUNOS PROFESORES DE MATEMÁTICAS EN FORMACIÓN

LUZDARI RANGEL RUIZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2010

CONCEPCIONES QUE SOBRE LA DISTRIBUCIÓN NORMAL TIENEN ALGUNOS PROFESORES DE MATEMÁTICAS EN FORMACIÓN

LUZDARI RANGEL RUIZ

Trabajo de Grado para Optar al Título de

Licenciada en Matemáticas

Director

DR. GABRIEL YÁÑEZ CANAL

Especialidad en Matemática Educativa

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

BUCARAMANGA

2010

AGRADECIMIENTOS

Deseo expresar mi más sincero agradecimiento al Dr. Gabriel Yáñez Canal, director de esta tesis, por su amistad, respeto, paciencia y dedicación, por su permanente apoyo tanto a nivel académico como personal, por sus consejos, su constancia y por sus contribuciones incalculables, las cuales han sido esenciales para la adecuada realización de esta tesis.

A Úl, por su amor, constante apoyo, paciencia y por creer en mis capacidades.

A mi familia, en especial a mi madre, por todo en lo que de una u otra forma contribuyó a mi formación como persona.

A los estudiantes de Estadística II, por su paciencia y valiosa colaboración, sin la cual no hubiese sido posible la realización de esta investigación.

A los profesores, estudiantes y administrativos de la Escuela de Matemáticas, por su dedicación y entrega en la tarea de formación docente, tanto a nivel académico como personal. Quiero hacer reconocimiento especial a los profesores Luis H. Rodríguez, Rafael Aponte y Bernardo Mayorga por su apoyo académico y personal en todos estos años.

Al Departamento de Cultura Física y Deportes, en especial a la selección de Fútbol Sala, por permitirme compartir dos años de alegrías y triunfos, sin duda, perduraran en mis recuerdos.

A la Universidad Industrial de Santander (UIS), por ser mi segundo hogar, en el cual viví una de las mejores etapas de mi vida y en la cual conocí a seres maravillosos que sin duda se hicieron inolvidables.

A Dios por darme la vida, a la vida por permitirme conocerlos a ustedes, a los nombrados y no nombrados, no por olvido si no por la limitación del texto.

A todos les expreso mi más sentido agradecimiento.

Luzdari Rangel

TABLA DE CONTENIDO

LISTA DE FIGURAS

LISTA DE ANEXOS

1. **TÍTULO: CONCEPCIONES QUE SOBRE LA DISTRIBUCIÓN NORMAL TIENEN ALGUNOS PROFESORES DE MATEMÁTICAS EN FORMACIÓN***

2. **AUTOR:** Luzdari Rangel Ruiz†

3. **PALABRAS CLAVES:** Distribución Normal,Media, Desviación Estándar, Simetría, Teorema central del límite.

4. **RESUMEN:**

Esta investigación se orientó a analizar las concepciones que sobre la distribución normal tienen algunos profesores de matemáticas en formación. La metodología utilizada en este estudio fue la resolución de problemas. Para esto se realizaron dos cuestionarios en los que participaron catorce (14)estudiantes que estaban cursando la asignatura de Estadística IIofrecida por la Escuela de Matemáticas de la Universidad Industrial de Santander (UIS).

El primer cuestionario se aplicó luego de terminado el proceso de instrucción referente a la Distribución Normal y sus Propiedades, y se observó cierta dificultad en el momento de realizar el proceso de estandarización y de utilizar las propiedades de la normal (propiedad simétrica; Regla del 68-95-99,7) como estrategia en la resolución de los ítems propuestos.

El segundo cuestionario se aplicó luego de terminado el tema referente al Teorema Central del Límite (TCL), en este estudio se destacó la dificultad para reconocer los tipos de problemas para los cuales el TCLes una herramienta de análisis. En algunos estudiantes persistió la idea de que este teorema es aplicable solo a distribuciones normales.

Adicionalmente se observó que es importante poner especial atención a la simbología matemática usada por los docentes en formación, pues la distribución normal es un tema que requiere la combinación de muchos términos y nomenclatura matemática.

*Trabajo de Grado

† Facultad de Ciencias. Escuelade Matemáticas. Dr. Gabriel Yáñez Canal

1. **TITLE: CONCEPTIONS ABOUT NORMAL DISTRIBUTION THAT HAVE SOME MATHEMATICS UNDERGRADUATE STUDENTS**[*]

2. **AUTHOR:** Luzdari Rangel Ruiz[†]

3. **KEYWORDS:** Normal Distribution, Average, Standard Deviation, Symmetry, Limit Central Theorem

4. **ABSTRACT:**

This investigation was oriented to analyze the conceptions about the normal distribution that have some Mathematics undergraduate students. The methodology used in this study was problem resolution. We implemented for the study two questionnaires with the participation of fourteen (14) students cursing the subject Statics II offered forthe Mathematics Department of Universidad Industrial de Santander (UIS) located at Bucaramanga, Colombia.

The first questionnaire was applied just finish the Normal Distribution an its Properties lesson at the classroom, in this first experience was observed some difficulty at the moment of the standardization process and use of normal properties (Symmetry Property, Rule of 68-95-99.7) as solving strategy for the proposed items.

The second questionnaire was applied after finish the Central Limit Theorem (CLT) lesson at the classroom, in this study was noted the difficulty to recognize the type of problems in what Central Limit Theorem (CLT) is used as analysis tool. Some students had the persistent idea that this theorem is only applicable to normal distributions.

Additionally we observed that is important put special attention to the mathematic symbologyused for the students, since the normal distribution is a topic that requests the combination of a lot of mathematical terms and nomenclature.

[*]Undergraduate thesis

[†]Science Faculty, Mathematics Department.Advisor: Dr. Gabriel Yáñez Canal

INTRODUCCIÓN

La distribución normal es una de las distribuciones de probabilidad de variable continua que con más frecuencia describe fenómenos reales, lo cual hace que su uso sea de gran importancia.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y sicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, incluso si la distribución de la población de la cual se extrae la muestra no es normal, tal como lo expresa el teorema central del límite.

Por las razones anteriores este trabajo se orientó a responder la siguiente pregunta: ¿Cuál es el significado que sobre la distribución normal tienen los profesores en formación después de realizar un curso básico de estadística?

Para responder a la pregunta, esta investigación se interesó en analizar las estrategias usadas y las dificultades presentadas en los estudiantes en situaciones que requieren el uso de la distribución normal y sus propiedades durante y después del proceso de enseñanza.Para ello se aplicaron dos cuestionarios, el primero al finalizar la instrucción referente a la distribución normal y sus propiedades, y el segundo al finalizar la instrucción sobre el Teorema Central del Límite.

Este trabajo expone los resultados de esta investigación y está conformado por cinco capítulos. En el primer capítulo “Antecedentes”, se presentan algunas investigaciones previas relacionadas con la distribución normal y con el tema de estudio.

En el segundo capítulo “Marco Teórico” se describe en términos matemáticos y estadísticos la distribución normal y sus propiedades.

En el tercer capítulo “Diseño de la Investigación”, se presentan todos los aspectos relacionados con la metodología de la investigación, la justificación de cada ítem propuesto en los dos cuestionarios y los métodos empleados para el análisis de los datos obtenidos en la aplicación de éstos.

En el cuarto capítulo “Resultados y Análisis de la información”, se presentan los resultados de la investigación haciendo un análisis básicamente cualitativo, con algunos elementos cuantitativos asociados a las respuestas correctas e incorrectas de los cuestionarios, presentando las evidencias obtenidas y sus correspondientes análisis.

En el quinto capítulo “Conclusiones”, se presentan los resultados más importantes de nuestro estudio referidas estas a las dificultades y concepciones que los estudiantes presentaron cuando se les pidió resolver problemas que involucraban la distribución normal.

El presente estudio finaliza presentando las referencias utilizadas en esta investigación.

En los anexos se pueden apreciar los cuestionarios aplicados a los docentes en formación que participaron en la investigación.

Esperamos que este trabajo contribuya a una mejora en el proceso de formación de los estudiantes y docentes en ejercicio de su profesión.

1 ANTECEDENTES

A continuación se comentan algunas investigaciones previas relacionadas con la distribución normal, las cuales se asumen como referencias básicas para esta investigación.

Tauber (2001)

En su investigación "La Construcción del Significado de la Distribución Normal a Partir de Actividades de Análisis de Datos", Tauber (2001) se interesa por las dificultades que presenta este tema a los alumnos que realizan un curso de nivel introductorio a la estadística mediado con el uso de computadores y sustentado en el marco teórico referente al significado y comprensión de los objetos matemáticos (Godino y Batanero, 1994; 1998).

El objetivo principal de la investigación fue la enseñanza y aprendizaje de la distribución normal, más específicamente, se interesó en la problemática que presentaba este tema a los alumnos que realizaban un curso introductorio a la estadística con un enfoque basado en el uso de programas estadísticos (Tauber, 2001, p.17)

Durante la investigación se utilizó el programa Statgraphics para simular las situaciones planteadas a los estudiantes a partir de ficheros de datos y para analizar los datos obtenidos. La metodología que se utilizó en esta investigación fue la de resolución de problemas.

La investigación se llevó a cabo con estudiantes de primer año universitario que cursaban una asignatura de libre configuración de 9 créditosdesarrollada en la Facultad de Educación de la Universidad de Granada durante 1998-1999 y 1999-2000.

Algunas conclusiones generadas de esta investigación son:

- Una primera conclusión de este estudio es la complejidad del significado de la distribución normal, que se apoya en muchos otros conceptos previos, como los de variable aleatoria, distribución estadística, medidas de posición central y dispersión, simetría, curtosis, probabilidad, etc., dificultando así la comprensión del concepto de distribución normal.

- Los estudiantes se ayudaron de la representación gráfica para validar, comprobar y aplicar algunas propiedades de manera descriptiva, olvidando utilizar las debidas argumentaciones que subyacen a ellas.

- El uso del paquete estadístico Statgraphics facilitó el proceso de construcción del significado de la distribución normal, ya que al realizar simulaciones se estableció una relación más estrecha entre la teoría y la práctica.

- A los alumnos se les dificultó la interpretación de áreas en histogramas de frecuencia y problemas en el cálculo del área dentro de un intervalo. La no diferenciación entre el modelo teórico y los datos empíricos y dificultad en distinguir cuándo el programa de cálculo se refiere a una u otra distribución, así como la no discriminación entre estadísticos y parámetros. Además, una escasa capacidad de argumentación.

En esta investigación Tauber(2001), considera que el punto de partida de la estadística debería ser el encuentro de los alumnos con sistemas de datos reales e incluso construir un sistema de datos propio y analizarlo, que no es lo mismo que resolver un problema de cálculo rutinario tomado de un libro de texto. Por esto considera que la perspectiva del análisis exploratorio de datos, y el aprendizaje de un paquete estadístico, son dos puntos muy importantes en la enseñanza de la estadística, tanto en la enseñanza secundaria como en un curso introductorio de estadística a nivel universitario.

También destaca la importancia de que el profesor sea consciente de los diversos niveles de comprensión que pueden adquirir sus alumnos. Aconseja además que se propicie la comprensión intuitiva antes de conducir a los alumnos a un pensamiento más formal.

Alvarado y Batanero (2007)

En este artículo titulado Dificultades de Comprensión de la Aproximación Normal a la Distribución Binomial, se presenta un estudio de evaluación de la comprensión teórica y práctica de la aproximación normal a la distribución binomial, teniendo en cuenta que es un caso particular del Teorema Central del Límite.

En este trabajo se analizó la comprensión teórica y práctica de dicha aproximación alcanzada por un grupo de estudiantes de ingeniería después de un experimento de enseñanza apoyado en el uso de Excel.

La experiencia se llevó a cabo en el segundo año de estudios de ingeniería en la Universidad Católica de Concepción (Chile), los estudiantes habían tomado previamente un curso de cálculo de probabilidades en el primer año. Se contó con la participación de 123 estudiantes del curso de estadística.

Algunas conclusiones generadas de esta investigación son:

- El análisis de los datos obtenidos mediante evaluaciones y actividades reflejan la complejidad de la aproximación normal a la binomial y de su aplicación en la resolución de problemas.

- En el proceso de resolución de problemas se evidenció que los estudiantes son capaces de calcular y comparar probabilidades exactas y aproximadas utilizando la distribución normal para valores del número y proporción de éxitos y calculan el tamaño de la muestra necesario para obtener una precisión dada.

- Los estudiantes no diferenciaron entre variables aleatorias discretas y continuas. En algunos casos consideraron, en el cálculo de la probabilidad, la varianza para realizar el proceso de estandarización en lugar de la desviación estándar.

Alvarado y Batanero (2007), recomiendan reforzar la comprensión de la aproximación normal a la distribución binomial y relacionar más la convergencia de este caso particular con el enunciado general del Teorema Central del Límite.

Inzunsa (2006)

En este artículo titulado Students' Errors and DifficultiesforSolvingProblems of SamplingDistributionsbyMeans of ComputerSimulation, basado en la tesis doctoral del mismo autor titulada "Significados que estudiantes universitarios atribuyen a las distribuciones muestrales en un ambiente de simulación computacional y estadística dinámica", Inzunsa(2006), reporta los resultados de los principales errores y las dificultades de un grupo de once estudiantes universitarios al resolver problemas de distribuciones muestrales por medio de simulación por computador utilizando el software Fathom.

.Algunas conclusiones generadas por esta investigación son:

- Las principales dificultades, relatadas por el investigador, fueron la formulación del modelo, seguido de la definición de los estadísticos y la construcción de intervalos para calcular probabilidades.

- A pesar de los errores y las dificultades descritas anteriormente, los estudiantes fueron capaces de resolver los problemas de las distribuciones de muestreo de distribuciones normales por medio de simulación por computador.

- Entre las ventajas de la solución por medio de la simulación fue el hecho de interpretar la probabilidad de los resultados del muestreo como la proporción de casos de interés en un número total de repeticiones posibles, es decir, el enfoque frecuencial de la probabilidad.

Inzunsa (2006), recomienda la utilización de Fathom, ya que al resolver problemas con la ayuda de dicho programa se evita llevar a cabo todos los procesos largos e inevitables del lápiz y el papel, tales como la normalización de la distribución muestral y la utilización de tablas de probabilidad que son la fuente de varios errores al resolver problemas.

Observaciones recogidas de los antecedentes que serán consideradas en la investigación

Teniendo en cuenta las dificultades que se presentan a la hora de resolver problemas que requieran el uso de la distribución normal y del teorema central del límite, y de acuerdo con los antecedentes mencionados previamente, se identificaron diversas dificultades e ideas importantes que sustentan la presente investigación.

- Los estudiantes tienden a apoyarse más en las representaciones gráficas que en las numéricas en el momento de dar solución a problemas referentes a la distribución normal (Tauber, 2001).

- Es preciso que el estudiante pueda apreciar la utilidad de la distribución normal y sus propiedades en el análisis de colecciones de datos amplios, preferiblemente, colecciones de datos recogidos por el mismo alumno o procedentes de investigaciones que el alumno pueda comprender (Inzunsa, 2006).

- La distribución normal se apoya en varios otros conceptos previos, como los de variable aleatoria continua y discreta, distribución estadística, medidas de posición central y dispersión, simetría, etc. El desconocimiento de estos conceptos dificulta el

- Uno de los teoremas más importantes en el cálculo de probabilidades es el teorema central del límite, por tal motivo, es de gran importancia que los estudiantes se apropien de su enunciado e implicaciones que son útiles en la aproximación de diferentes distribuciones mediante la normal. (Alvarado y Batanero 2007)

2 MARCO TEÓRICO

2.1 SIGNIFICADO Y COMPRENSIÓN DE LA DISTRIBUCIÓN NORMAL

Dado que el desarrollo de la investigación se centra en la distribución normal y en las concepciones que sobre dicha distribución tienen algunos profesores de matemáticas en formación, nuestra atención se enfocó en analizar las nociones construidas por dichos estudiantes, mediante el análisis de cuestionarios que evidenciaran las dificultades presentes en la resolución y argumentación de problemas.

Para esto es necesario tener en cuenta que en estadística y probabilidad se llama distribución normal a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

Hay varios modos de definir formalmente una distribución de probabilidad. La forma más visual es mediante su función de densidad.

Se dice que una variable aleatoria continua X sigue una distribución normal de parámetros μ y σ y se denota X~N (μ, σ) si su función de densidad está dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, x \in \mathbb{R}$$

Donde μ (mu) es la media y σ (sigma) es la desviación típica (σ^2 es la varianza).

En la Figura 1.se puede observar tres gráficas de la función de densidad con medias iguales y desviación estándar diferente, además, presentamos una gráfica en la cual el valor de la media es negativo.

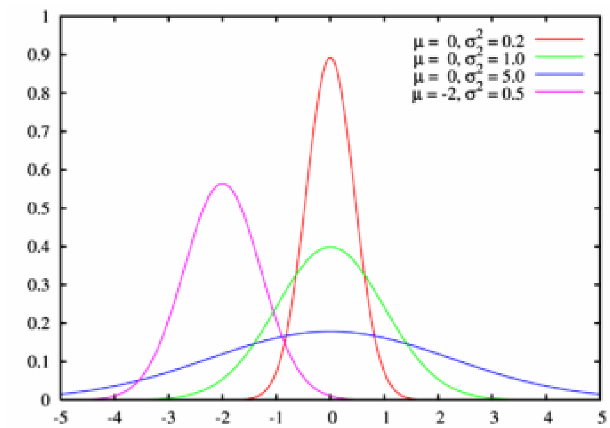


Figura 1. Gráficas de la función de densidad de la de la distribución normal

Dicha expresión tiene inmerso el siguiente componente:

- **Parámetros que definen la distribución normal:** todas las distribuciones normales tiene la misma forma general. La curva de densidad exacta de una distribución normal concreta se describe dando su media μ y su desviación estándar σ. La media se sitúa en el medio de la grafica, siendo este su parámetro de localización. La desviación estándar σ controla la dispersión de la curva normal, ya que esta es una medida natural de la dispersión de dicha distribución.

2.2 PROPIEDADES DE LA DISTRIBUCIÓN NORMAL

A continuación realizaremos una descripción de algunas propiedades de la distribución normal derivadas de análisis de la función de densidad de la distribución normal, las cuales denominaremos propiedades geométricas, y de otras que aparecen en Moore (1995), las cuales clasificaremos como estadísticas y algebraicas.

2.2.1 PROPIEDADES GEOMÉTRICAS

Las definimos como aquellas que se derivan de los análisis de la gráfica de la función de densidad, tales como simetría y concavidad.

1. La distribución normal es simétrica respecto de su media, μ.

   La simetría es importante en la distribución normal dado que simplifica el cálculo de áreas, es decir, como el área bajo la curva normal es 1, al realizar cálculos de probabilidades que hagan referencia al uso de una o dos colas se podrá observar con facilidad el área que se estaba pidiendo calcular.

   A continuación veremos algunas propiedades derivadas de la simetría.

   a. La distribución normal es simétrica respecto de su eje vertical donde tiene la media.

   b. La media, mediana y moda en las distribuciones simétricas coinciden en un mismo punto, por tanto en las distribuciones normales son iguales.

c. La media de una curva de densidad es el punto de balance, en el cual la curva podría balancearse si fuera un sólido.

2. La distribución normal presenta concavidad hacia arriba y concavidad hacia abajo.

   Esta propiedad es muy útil para ayudarnos a dibujar la gráfica, por lo cual hay que tener en cuenta las siguientes observaciones:

   a. La curva tiene sus puntos de inflexión en $x = \mu \pm \sigma$.

   b. La curva normal es cóncava hacia abajo si: $\mu - \sigma < X < \mu + \sigma$, y es cóncava hacia arriba en cualquier otro punto.

3. La curva normal se acerca al eje horizontal en forma asintótica en cualquiera de las dos direcciones, alejándose de la media.

En la Figura 2.se pueden observar las propiedades anteriormente mencionadas.

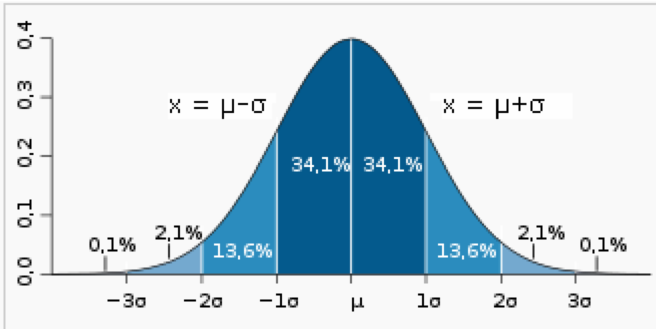


Figura 2. Distribución de probabilidad alrededor de la media en una distribución N (μ, σ)

2.2.2 PROPIEDADES ESTADÍSTICAS

Las definimos como aquellas que relacionan la distribución normal con la predicción de valores y el cálculo de probabilidades.

1. REGLA DEL 68-95-99,7

   En una distribución normal de media μ y desviación típica σ se cumple que:

   a. El 68% de las observaciones se encuentran entre $\mu \pm \sigma$.

   b. El 95% de las observaciones se encuentran entre $\mu \pm 1.96\sigma$.

c. El 99,7% de las observaciones se encuentra entre μ ± 3σ.

En la Figura 3.se puede apreciar gráficamente esta propiedad de la distribución normal.

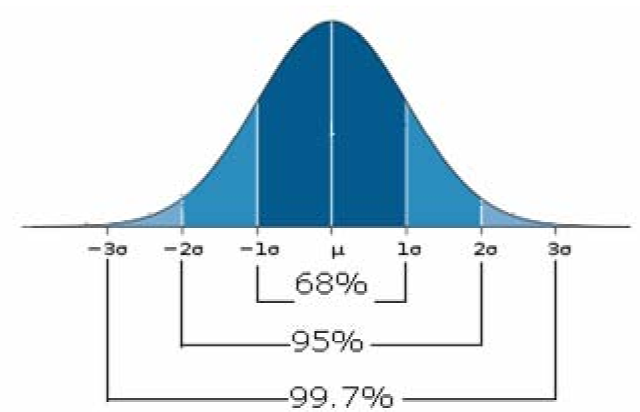


Figura 3. Regla del 68-95-99,7

Esta propiedad es de gran utilidad para el establecimiento de intervalos de confianza. Por otra parte, el hecho de que prácticamente la totalidad de la distribución se encuentre a tres desviaciones típicas de la media justifica los límites de las tablas empleadas habitualmente en la normal estándar.

2. Otra propiedad importante es el Teorema Central del Límite, ya que justifica el uso generalizado de la distribución normal.

   El teorema central del límite afirma que la distribución de las medias muestrales es aproximadamente normal, incluso si la distribución de la población de la cual se extrae la muestra no es normal.

La importancia práctica del Teorema del límite central es que la función de distribución de la normal puede usarse como aproximación de algunas otras funciones de distribución.

Por ejemplo: Una Distribución Binomial de parámetros n y p es aproximadamente normal para grandes valores de n, y p no demasiado cercano a 1 ó 0. La normal aproximada tiene parámetros $\mu = np$, $\sigma^2 = np(1 - p)$.

2.2.3 PROPIEDADES ALGEBRAICAS

Las definimos como aquellas que están relacionadas con el comportamiento de la función de densidad frente a transformaciones.

1. Una propiedad muy importante de las variables normales es que sus transformaciones lineales también siguen una distribución normal. La media de la variable vendrá afectada por la misma transformación lineal, esto es si X es normal con media μ_x y desviación típica σ_x, entonces la variable $y = ax + b \sim N(a\mu_x + b, a\sigma_x)$.

2. Si $X \sim N(\mu, \sigma^2)$ y a y b son números reales, entonces $(aX + b) \sim N(a\mu + b, a^2\sigma^2)$.

3. Si $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_x, \sigma_x^2)$ son variables aleatorias normales independientes, entonces:

a. Su suma está normalmente distribuida con $U = X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$. Recíprocamente, si dos variables aleatorias independientes tienen una suma normalmente distribuida, deben ser normales.

b. Su diferencia está normalmente distribuida con $V = X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$.

2.3 DISTRIBUCIÓN NORMAL “ESTÁNDAR”

Se llama **distribución normal "estándar"** a aquélla en la que sus parámetros toman los valores μ = 0 y σ = 1. En este caso la función de densidad tiene la siguiente expresión:

$$f(x) = f_{0,1}(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}, x \in \mathbb{R}$$

Este tipo de distribución normal conserva todas las propiedades enunciadas anteriormente.

Su gráfica se muestra a continuación y con frecuencia se usan tablas para el cálculo de los valores de su distribución.

En la Figura 4.se puede apreciar gráficamente la función de densidad de la distribución normal estándar.

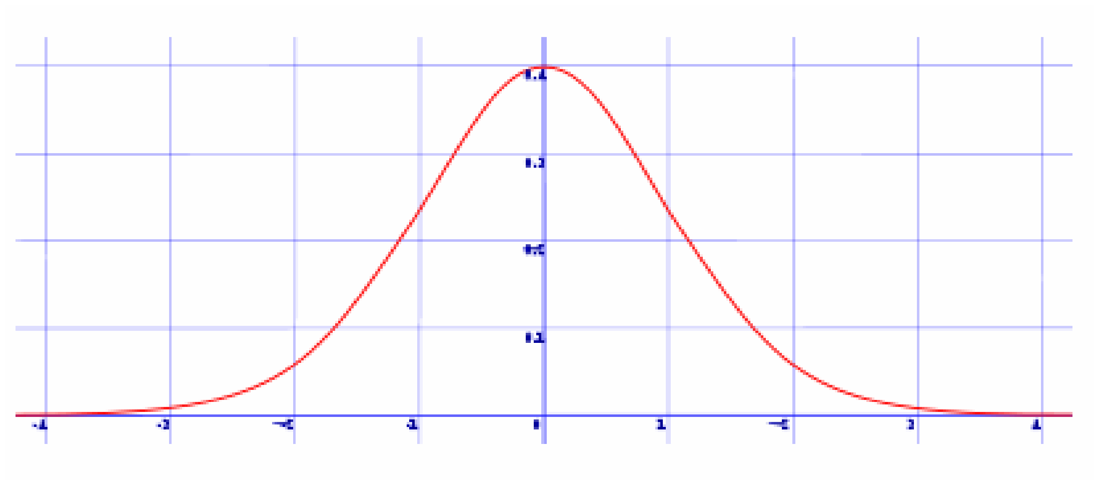


Figura 4. Gráfica de la distribución normal estándar N (0, 1)

2.3.1 ESTANDARIZACIÓN DE VARIABLES ALEATORIAS NORMALES

Una consecuencia de la primera propiedad algebraica, es que es posible relacionar todas las variables aleatorias normales con la distribución normal estándar.

Si X ~ N (μ, σ^2), entonces $Z = \frac{X - \mu}{\sigma}$ es una variable aleatoria normal estándar de tal manera que $Z \sim N(0, 1)$

La transformación de una distribución $X \sim N(\mu, \sigma)$ en una $N(0, 1)$ se llama **normalización**, **estandarización** o **tipificación** de la variable X.

A la inversa, si Z es una distribución normal estándar $Z \sim N(0, 1)$, entonces $X = \sigma Z + \mu$ es una variable aleatoria normal de media μ y varianza σ^2.

La distribución normal estándar está tabulada (habitualmente en la forma del valor de la función de distribución Φ) y las otras distribuciones normales pueden obtenerse como transformaciones simples, como se describe más arriba, de la distribución estándar. De este modo se pueden usar los valores tabulados de la función de distribución normal estándar para encontrar valores de la función de distribución de cualquier otra distribución normal.

2.4 ESTIMACIÓN POR INTERVALOS DE CONFIANZA: Una aplicación del Teorema Central del Límite.

En algunos tipos de problemas, la finalidad primordial es predecir qué valor se obtendría para un estadístico si se conoce el valor del parámetro. Pero en la práctica estadística, lo que realmente interesa es estimar el valor del parámetro en la población cuando se conoce el valor del estadístico en la muestra.

Conocida la distribución muestral del estadístico, es posible la obtención de intervalos de confianza para el parámetro de interés. Por ejemplo, se sabe que en una distribución normal el 95% de los casos se encuentran a una distancia 2σ de la media. Si μ es la media de una población normal y σ su desviación típica, la media muestral $\bar{X}$ de la variable aleatoria X, sigue una distribución aproximadamente normal $N\left(\mu \pm \frac{2\sigma}{\sqrt{n}}\right)$, siendo n el tamaño de la muestra. Por ello, en el 95% de las muestras, la media muestral $\bar{x}$ estará a una distancia $\frac{2\sigma}{\sqrt{n}}$ de la verdadera media μ de la población. Y recíprocamente, se puede deducir que en el 95% de las muestras la media μ en la población estará dentro del intervalo $N\left(\text{ EMBED Equation.3 } \pm \frac{2\sigma}{\sqrt{n}}\right)$, siendo este el intervalo de confianza del 95%.

3 DISEÑO DE LA INVESTIGACIÓN

A continuación se presenta la metodología que sustenta la investigación. De igual manera se presenta el diseño y la descripción de los cuestionarios aplicados durante la investigación.

3.1 METODOLOGÍA DE LA INVESTIGACIÓN

Esta es una investigación básicamente cualitativa, sin embargo, haremos uso de algunos elementos cuantitativos asociados a las respuestas correctas e incorrectas de los cuestionarios.

3.1.1 FASES DE LA INVESTIGACIÓN

La investigación comprende dos aspectos fundamentales:

1. Diseño y aplicación de la parte experimental (instruccional), y recolección de la información.

   - La instrucción es realizada por el profesor de la asignatura. El proceso de instrucción está enmarcado en la filosofía de resolución de problemas. En la mayoría de los casos los datos provenían de investigaciones que el alumno podía comprender y relacionar con la cotidianidad. Contrario a lo propuesto por

- Durante el proceso instruccional se aplicaron dos cuestionarios sobre la distribución normal: el primero fue aplicado durante el proceso de instrucción y el segundo al finalizar el curso. Los cuestionarios contemplaron todos los aspectos relacionados con la distribución normal. Las respuestas dadas debían ser justificadas con el fin de evidenciar las posibles dificultades de los estudiantes así como las estrategias y representaciones que utilizaron para resolver los problemas.

- Adicionalmente, se realizaron entrevistas con algunos estudiantes del curso para conocer en detalle las argumentaciones que utilizaban para responder a los cuestionarios aplicados.

2. Análisis de la información.

Se analizaron e interpretaron las respuestas dadas por los estudiantes tanto en los cuestionarios escritos como en las entrevistas personales lo que permitió caracterizar el significado que adquirieron alrededor de la distribución normal. Para este análisis se adoptó parte de la metodología sugerida por Godino (1996) para caracterizar el significado personal de los estudiantes. Estos análisis también permitieron clasificar en orden de dificultad los conceptos y propiedades relacionados con la distribución normal.

3.1.1.1 Población y muestra

La población de interés en nuestro trabajo son los profesores de matemáticas en formación, más específicamente, los estudiantes de la carrera de Licenciatura en Matemáticas de la Universidad Industrial de Santander. La muestra estuvo formada por 14 estudiantes, quienes en el segundo semestre del 2009 cursaban la asignatura Estadística II. Los nombres de los estudiantes que participaron en esta investigación han sido cambiados para preservar su identidad.

3.1.1.2 Instrumentos de recogida de datos

Puesto que el interés de nuestro trabajo se centra en tratar de determinar las concepciones que se forman los estudiantes sobre la distribución normal, para ello en el diseño de la investigación se hace necesario utilizar varios instrumentos de recolección de datos, los cuales son mencionados a continuación:

- Aplicación de dos cuestionarios.
- Se filmaron dos entrevistas realizadas con algunos de los estudiantes participantes en la investigación con la finalidad de ganar mayor claridad respecto a las respuestas que dieron en los cuestionarios. Estas entrevistas fueron realizadas conjuntamente con el director de este trabajo quien era, precisamente, el profesor del curso de estadística mencionado.

3.1.1.3 Núcleos de interés

Como esta investigación es básicamente cualitativa, se analizaron las respuestas y argumentaciones ofrecidas por los estudiantes en cada ítem propuesto. De igual manera, se analizaron los procedimientos realizados por los estudiantes para resolver los problemas planteados en los cuestionarios.

Para el análisis de los datos se establecieron dos ejes de interés:

1. La distribución normal: parámetros y propiedades que la definen. Se analizaron las concepciones construidas y la comprensión de la distribución normal en situaciones que ameritan su uso. En particular, la capacidad de reconocer los parámetros que la definen y el uso de sus propiedades en la resolución y argumentación de problemas. Además, se establecieron las dificultades y errores presentes al trabajar con la distribución normal.

2. El teorema central del límite: implicaciones y uso en la resolución y argumentación de problemas. Dada la importancia de este teorema en la estadística y la probabilidad, nos apoyamos en las respuestas de los docentes para analizar las concepciones que se formaron y el uso que le dieron en la resolución y argumentación de problemas. Además, exponemos las dificultades que se evidenciaron en la resolución y argumentación de situaciones que ameritan su uso.

3.1.2 DISEÑO Y APLICACIÓN DE LOS CUESTIONARIOS

A continuación se presentan en detalle los dos cuestionarios utilizados explicando las razones que motivaron cada uno de los ítems propuestos. Los cuestionarios fueron presentados en forma escrita y debían ser resueltos individualmente.

3.1.2.1 Cuestionario 1. La distribución normal: parámetros y propiedades que la definen

La aplicación de este cuestionario tuvo una duración de dos horas. Este cuestionario esta conformado por tres ítems con los cuales se pretendió observar si el docente en formación tiene claro los parámetros que definen la distribución normal, las diferencias entre la distribución normal y la distribución normal estándar y el uso de las propiedades como apoyo en la argumentación y resolución de problemas.

A continuación presentamos lo ítems propuestos en el cuestionario y la justificación de los mismos.

Ítem 1*. La distribución de los coeficientes intelectuales de hombres entre 20 y 34 años tiene aproximadamente una distribución normal de media μ= 110 y una desviación estándar σ= 25.*

a. *De los hombres entre 20 y 34 años, ¿Qué porcentaje tiene un coeficiente intelectual superior a 110?*
b. *¿Qué porcentaje de estos hombres tiene un coeficiente intelectual superior a 160?*

Con este ítem se pretendió observar la capacidad del estudiante en aplicar algunas propiedades de la distribución normal, entre ellas la propiedad simétrica y la regla del 68-95-99,7; a su vez se quería conocer su habilidad en el uso del proceso de estandarización.

Ítem 2. *En las siguientes preguntas Responda Verdadero o Falso y justifique la respuesta.*

a. Para que una distribución normal quede completamente definida, basta con conocer la media y la desviación estándar.

b. El parámetro de localización de la normal es μ y puede tomar cualquier valor real.

c. En la curva normal la media es igual a la moda.

d. El celular de la normal es solo para "comunicarse" con la distribución normal estándar.

El objetivo principal de este ítem radica en analizar los conocimientos adquiridos por los estudiantes sobre los parámetros que definen la distribución normal, lo cual requiere de la comprensión de su función de densidad, y de la aplicabilidad de la regla del 68-95-99,7. A su vez se busca analizar la capacidad de argumentación del estudiante.

Ítem 3. *Considere un mercado de acciones idealizado. Solamente cuatro compañías negocian en este mercado: Compañía X, Compañía Y, Compañía V y compañía W. Su asegurador le dice que el precio actual de las acciones para cada una de éstas es $200 pero que tienen diferencias en la volatibilidad (variabilidad) de los precios en el tiempo.*

El precio de las acciones en un mes se puede considerar aproximadamente normal, $X \sim N(10, 1)$, $Y \sim N(10,3)$, $V \sim N(10,3)$ y $W \sim Normal(10, 0.5)$.

También indica que las acciones se comportan independientemente. Usted tiene $2,000.000 para invertir y existen tres portafolios posibles para realizar la inversión.

1. *El portafolio 1 consiste de 1000 acciones de W*
2. *El portafolio 2 consiste de 500 acciones de W y 500 acciones de V.*
3. *El portafolio 3 consiste de 250 acciones de cada una de los cuatro tipos.*

Sea Pi el valor del portafolio i para el próximo mes contado a partir de la fecha.

a. *Para cada uno de los tres portafolios calcule las siguientes probabilidades:*
    i. *Pr(Pi>2,4 millones)*
    ii. *Pr(Pi<1 millón)*
a. *¿Qué portafolio es el más seguro y cuál es el más riesgoso?*

b. *Un amigo requiere obtener una ganancia de 5 millones por su inversión. Sin calcular ninguna probabilidad, ¿qué portafolio debería escoger y por qué?*

Con este ejercicio se busca observar la capacidad de argumentación y resolución de problemas que requieren el uso de la distribución normal. Específicamente, el proceso para estimar la desviación estándar para la suma de variables aleatorias, el cálculo de probabilidades y análisis de los mismos. De otro lado, el problema pretende conocer la capacidad interpretativa que de los resultados obtenidos por procesos estadísticos poseen los estudiantes.

3.1.2.2 Cuestionario 2. Teorema central del límite

La aplicación de este cuestionario tuvo una duración de 2 horas. Este cuestionario está conformado por 8 ítems, el objetivo principal de este cuestionario consiste en observar las concepciones que se forman los docentes respecto al teorema central del límite y el uso que le dan para la resolución y la argumentación en los ítems propuestos. Adicionalmente, se proponen afirmaciones acerca de la distribución normal para que los estudiantes analicen su veracidad a la luz de su conocimiento y de los significados que le atribuyen a la distribución normal.

A continuación presentamos lo ítems propuestos en el cuestionario y la justificación de los mismos.

Ítem 1. *En una distribución normal, el 50% de las medidas caen por encima de la media.*

Con este ítem se pretendía verificar una hipótesis que junto al director del proyecto había surgido al discutir las respuestas de los estudiantes al primer cuestionario, como es que los estudiantes confunden la simetría de la función de densidad de probabilidad con el dominio de esta función. Es decir, la simetría respecto a la media conduce a afirmar que la medida de probabilidad se divide en dos: 50% antes de la media y 50% después de la media y que éstas mitades se relacionan con los valores posibles al afirmar que $P(X \leq \mu) = P(X \geq \mu) = 1/2$. En otras palabras cuando se afirma "50%" se hace con relación a la probabilidad y no de los valores posibles de la normal, al fin de cuentas en la recta real cualquier valor se puede considerar como la mitad en el sentido de que los dos pedazos son equipotentes.

Ítem 2. *Si una variable está distribuida normalmente, los casos extremos son poco frecuentes.*

Para este ítem se pretendía saber si los estudiantes tenían claro el concepto de casos extremos y si interpretaban adecuadamente la regla del 68-95-99,7.

Ítem 3. $\overline{X}_n$ *Tiene aproximadamente una distribución normal cuando la muestra es grande.*

El objetivo principal de este ítem consiste en observar si los docentes reconocen el enunciado simplificado del teorema central del límite.

En los ítems 4, 6 y 7, se pretende indagar por las concepciones acerca de las hipótesis básicas asociadas con el teorema central del límite y con algunas de sus implicaciones.

Ítem 4. *La forma de la distribución de* $\bar{X}$ *depende de la forma de la distribución de la población.*

Se trata con este ítem de contraponer dos resultados distintos dependiendo del origen de los datos y del tamaño de n, el tamaño muestral. Si la muestra se toma de una población con distribución normal, la media muestral tiene distribución normal independiente del tamaño muestral. Ahora, si la población no es normal, la distribución de $\bar{X}_n$ solamente es normal cuando el tamaño de n es suficientemente grande, en caso contrario, la forma de la distribución se parece más a la forma de la distribución original.

Ítem 5. *Para que la desviación estándar de la distribución de* $\bar{X}_n$ *se reduzca a la mitad basta que la muestra tenga el doble de tamaño.*

Con este ítem se pretende conocer si los estudiantes conocen e interpretan adecuadamente el hecho de que la desviación estándar de $\bar{X}_n$ es de la forma $\sigma/\sqrt{n}$.

Ítem 6. *El teorema central del límite dice que en muestras grandes, la media muestral* $\bar{X}$ *tiene que estar cerca de la media poblacional μ.*

En este ítem también la intención está asociada a estudiar la comprensión que tenían los estudiantes acerca del significado de la desviación estándar y, en particular, de la forma cómo ésta se afecta al aumentar el tamaño muestral al considerar la variable aleatoria $\bar{X}_n$.

Ítem 7. *¿Qué le ocurre a* $\overline{X}$ *cuando la distribución poblacional no es normal?*

Este ítem comparte el objetivo del ítem 3.

Ítem 8*. La naturaleza de los casinos es tal que una gran cantidad de dinero fluye en ambas direcciones entre el casino y los jugadores. Los jugadores a menudo ganan lo suficiente para permanecer optimistas y continuar jugando. Lenta pero inexorablemente, sin embargo, el casino hace dinero. Los juegos están estructurados de tal forma que las posibilidades de una apuesta dada favorecen al casino más que al jugador, pero solamente por una pequeña cantidad. Como este es el caso, muchas personas se preguntan cómo es que los casinos funcionan para hacer grandes cantidades de dinero en el largo plazo. Todo se explica con las distribuciones muestrales de los promedios, como este ejercicio va a demostrar.*

Usaremos una apuesta simple en la ruleta como ilustración. La ruleta tiene 18 números negros, 18 números rojos y 2 números verdes (0, 00) el casino gana $ 1 si sale un número rojo o verde (20 chances de 38) y pierde $ 1 si sale un número negro (18 de 38). Así, las ganancias del casino en la i-ésima que es una variable aleatoria $\overline{X_i}$ *con función de probabilidad* $P(X = 1) = \frac{20}{38}$ *;* $P(X = -1) = \frac{18}{38}$

Los resultados de las diferentes apuestas son independientes.

a) *Encuentre la media y la desviación estándar de* X *.*

b) *(i) encuentre la media y la desviación estándar del total de ganancias del casino cuando se realizan 50 apuestas de $1.*

(ii) Haga lo mismo para las ganancias promedio del casino.

c) Escriba las respuestas correspondientes a las preguntas en (b) para un número n cualquiera.

d) ¿Qué proporción de jugadores de los que realizan 50 apuestas de $1 en el transcurso de una noche ganan dinero?

e) ¿Cuál es la probabilidad de que el casino pierda dinero con 1000 de estas apuestas?, ¿en 100.000?, ¿en 1,000.000?

f) Haga una tabla de intervalos de confianza de tres desviaciones estándar con 4 filas y 2 columnas. Intervalos en las primeras cuatro filas se refieren a conjuntos de 50, 1000, 100.000 y un millón de apuestas de $1, respectivamente. Coloque intervalos para el promedio de las ganancias del casino en la segunda columna.

Explique ahora porqué los casinos hacen tanto dinero, mientras que al mismo tiempo los jugadores ganan lo suficiente para seguir jugando.

El teorema central del límite facilita en gran medida la resolución de problemas relacionados con la cotidianidad, por lo tanto, con este ítem se pretende observar la habilidad de los estudiantes de reconocer la aplicabilidad de este teorema para dar respuesta al éxito económico de los casinos donde se proponen juegos de azar a las personas que los visitan. Este problema, además, guarda en su desarrollo la respuesta a la forma como los casinos mantienen el interés de las personas para que sigan jugando y cómo ellos, solamente en el largo plazo, pueden acumular jugosas ganancias. El problema parte de una forma no tradicional de proponer una situación binomial que exige recurrir a las definiciones generales para calcular su esperanza y su desviación estándar, y pretende que el estudiante sea capaz de utilizar el teorema central del límite para poder responder las preguntas planteadas.

En el cuarto capítulo de la investigación se presenta el análisis de los cuestionarios descritos previamente.

4 RESULTADOS Y ANÁLISIS DE LA INFORMACIÓN

4.1 CONCEPCIONES PERSONALES DE LOS ESTUDIANTES ALREDEDOR DE LA DISTRIBUCIÓN NORMAL

Para analizar las respuestas de los estudiantes a los cuestionarios, se definen cuatro categorías o núcleos de interés que se convierten en referencia de nuestro análisis. Estas categorías y sus aspectos más relevantes son las siguientes:

4.1.1 Categoría 1: Estandarización

En el cuestionario se evidenció que el proceso de estandarización es el que menos dificultad presenta.Sin embargo, los maestros en formación lo asumen solo como un proceso que indica a qué distancia de la media, en términos de desviaciones estándar, se encuentra un valor de la variable aleatoria X desconociendo que se trata en últimas, de transformar una variable con distribución normal cualquiera a una normal estándar. A continuación se presentan los apartes de una entrevista que alrededor del tema sostuvo el profesor de la materia con algunos estudiantes y que respaldan esta afirmación:

Profesor: ¿Qué es lo que dice el proceso de estandarización? Cuando estandarizo, el valor de z ¿qué me indica?

Laura: El proceso de estandarización me indica a cuántas desviaciones estándar se encuentra X de la media. Si se encuentra muy lejos X de la media, dicho valor es poco probable.

Profesor: ¿Qué es poco probable?

Laura: Es poco probable que ocurra el evento, que ocurra lo que quiero observar o cuantificar.

Aunque el profesor dejó escapar la oportunidad de preguntarle a Laura ¿cuándo es lejos para que sea poco probable?, no se puede evitar pensar que ella está pensando en la regla del 68-96-99,7 en términos de desviaciones estándar. Más adelante se corrobora la buena aplicación que los estudiantes hacen de esta regla.

Profesor: ¿Por qué estandarizamos?

Laura: Porque es más fácil trabajar con la distribución normal, que si yo observo el área bajo la curva es igual a 1. Que si yo trabajo con valores y hallo probabilidades, o sea es más fácil trabajar con probabilidades que con números.

Profesor: ¿Cómo así?

Laura: Pues la idea de cuando uno estandarizaba era $Z = \frac{X - \mu}{\sigma}$ entonces eso nos podía decir a nosotros cómo era el comportamiento de ese valor en la gráfica de la normal. Lo que decía, qué tan lejos está de la media.

No obstante lo poco claro de los dos últimos textos de Laura, parece que en últimas lo que está tratando de decir es lo que ya había dicho previamente: se trata de reducir todo a distancias respecto a la media. Su referencia a la gráfica de la normal, seguramente, es para ubicar el valor dado y tener una apreciación gráfica del área que se pretende buscar que ella asocia con probabilidades ya que "es más fácil trabajar con probabilidades que con números".

Profesor: ¿Y la X que menciona en la fórmula no tiene distribución normal?

Laura: eeee, la X, X es la variable.

Profesor: Es una variable aleatoria y ¿Qué distribución tiene?

Laura: Puede ser normal y puede que no sea normal.

Profesor: Asumamos que es normal, ¿Cuál es su media?

Laura: μ

Profesor: y ¿la desviación estándar que usted mencionó de cuál variable es?

Laura: deee…

Profesor: ¿De esa misma X o de otra variable?

Laura: De la función, la desviación estándar es σ.

Profesor:pero de quién es, ¿de qué distribución?

Laura: De la normal.

Profesor: ¿De cuál normal?, ¿de la que tiene X o de la estándar?

Laura: De la estándar.

Profesor: ¿Cuál es la desviación estándar de una normal estándar, cuánto vale?

Laura: σ.

Profesor: ¿Cuánto vale?

Laura: ...

Profesor: ¿Cuál es la normal estándar?

Laura: (0,1)

Profesor: Entonces, ¿cuánto vale la desviación estándar?

Laura: 1

Profesor: ¿De quién es la desviación estándar que pone en la fórmula?

Laura: De la variable X

Profesor: Entonces usted tiene una normal que no es estándar y estandariza ¿Qué estamos haciendo cuando estandarizamos?

Laura: Transformando a μ que sea cero y a σ que sea 1.

Profesor: O sea ¿está haciendo qué?

Laura: Estandarizando

Profesor: ¿Qué quiere decir eso en últimas?, ¿Qué significa estandarizar?

Laura: Estandarizar es, o sea cuando yo aplico el proceso estoy mirando a cuántas desviaciones estándar está ese valor.

Profesor: ¿Pero qué está haciendo? ¿En últimas qué es estandarizar? Tiene una normal que no es estándar y una estándar, ¿Qué hace?

Laura: Estoy estandarizándola.

Profesor: ¿Qué significa eso?

Laura: Hago que mi media se comporte como μ y...

Profesor: Escriba en el tablero las siguientes formulas: $X \sim N(\mu, \sigma)$ y $Z = \frac{X-\mu}{\sigma}$ $Z \approx N(0,1)$

Laura: Si, esa es la normalización de la variable.

Profesor: ¿Pero qué está haciendo, compare las distribuciones de X y Z?

Laura: Estoy es mirando el comportamiento de z, a qué distancia está de…

Profesor: Estoy mirando a X, X es lo que tengo.

Laura: Sí, o sea, por ejemplo, cuando yo tengo a X y la estandarizo yo estoy mirando qué tan cerca o tan lejos, estoy mirando a cuántas desviaciones estándar esta de mi μ.

Profesor: ¿De las desviaciones estándar de quién?

Laura: De X

A pesar de los variados intentos del profesor para que Laura aprecie que la estandarización es un proceso que transforma una normal cualquiera a una normal estándar, Laura insiste en pensar en el significado que le produce resultados prácticos: “a qué distancia está de…”.

Al final el profesor plantea la pregunta en forma directa:

Profesor: ¿Entonces se puede decir que transforma una normal no estándar en qué?

Laura: En una normal estándar

Profesor: ¿En una normal estándar qué significa el valor 2?

Laura: Que se encuentra a dos desviaciones estándar.

Profesor: Si ahora estoy en una normal no estándar y tengo el valor 17, ¿a cuántas desviaciones estándar está de mi media?

Laura: todavía no lo sé.

Este diálogo hace pensar que este proceso de estandarización debe presentarse utilizando variadas representaciones que permitan a los estudiantes comprender más ampliamente este proceso. Por ejemplo el valor de Z dado por $\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}$ debe desagregarse en la forma $Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$ lo que evidencia claramente que se trata de una transformación lineal (en el sentido de que se trata de una función lineal) que permite el cálculo de probabilidades de eventos definidos en términos de X a través de los equivalentes en términos de Z.

4.1.1.1 Cálculo de la probabilidad de un intervalo

También se pudo observar que el cálculo de probabilidades resulta complicado para la mayoría de los docentes en formación que conformónuestra muestra. Ya que después de realizar adecuadamente el proceso de estandarización, en el momento de hacer el cálculo de probabilidades se confunden al no tener en cuenta cuál es el intervalo (área) que se está pidiendo calcular. Carlos como otros estudiantes presentaron esta dificultad, lo anterior se puede observar en la Figura 5, la cual muestra la solución de Carlos al ítem 1 en la parte a del cuestionario 1:

- Carlos realizó bien el proceso de estandarización y dado que el valor de z es 0 no continuó con el cálculo de probabilidades; al parecer el estudiante se apoyó en un gráfico de la distribución normal estándar en el cual ubicó el cero en el centro; es como si el estudiante estuviera pensando que podía asociar este cero en la gráfica que tiene, que es la gráfica de la normal que está trabajando.

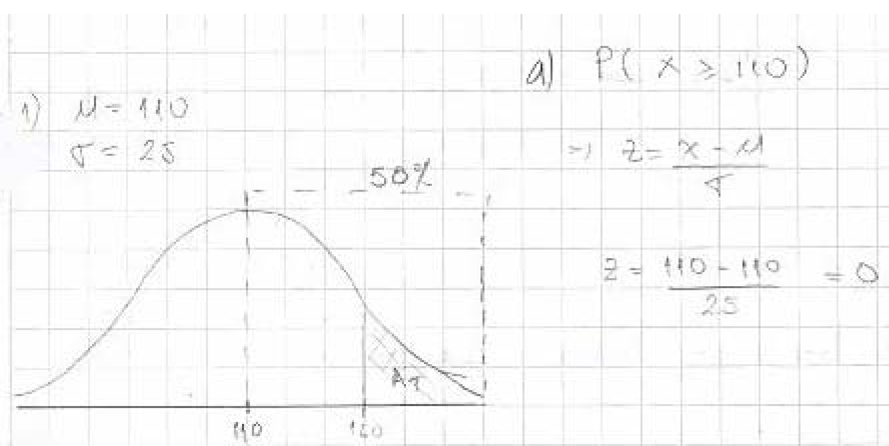


Figura 5. Recorte de la respuesta de Carlos en el ítem 1 parte a.

El problema no solo radica en el cálculo de la probabilidad de un intervalo, también en la comprensión del significado de los intervalos de confianza en sí mismos. Como este tema es de vital importancia nos permitimos transcribir algunos apartes de una entrevista realizada por el profesor del curso con algunos estudiantes.

Profesor: Un intervalo de confianza del 90% de confianza, significa que el 90% de los valores de esa variable aleatoria están en ese intervalo. Verdadero o falso.

Laura: Siii.

Mónica:Siii.

Alejandra: No es el 90% de los valores, si no, los que… los que tiene mayor probabilidad de… los que su probabilidad es mayor entonces aparecen en ese intervalo. Ahí yo puedo tener muchos datos dispersos, pero, y un intervalo del 90% de confianza es unos cuantos de ellos, no el 90% de los datos… ¿Eso depende es de que tan probable es el valor?

Profesor: ¿Entonces el nivel de confianza es la probabilidad de que la media poblacional esté en el intervalo?

Alejandra: Sí.

Laura: Sí.

Mónica: No le escuché bien profe…

Profesor: ¿El nivel de confianza es la probabilidad de que la media poblacional esté en el intervalo?

Mónica: Siii.

Profesor: Sergio, ¿qué opina usted?

Sergio: Es que… el nivel de confianza, lo que me dice es, eee, pues eee, en este caso en especial… pues igual en el intervalo de confianza yo estoy considerando la variabilidad, ¿sí?, y yo estoy considerando la media, yo creo que sí debería... por ejemplo el 90%, el 90% debería estar ahí.

Profesor: ¿El 90% de qué?

Sergio: Digo, un nivel de confianza del 90%, osea, el 90% de los datos está en ese intervalo.

El profesor nota que los estudiantes no tienen claro qué es el nivel de confianza de un intervalo en términos estadísticos, por tal motivo procede a plantearles preguntas con la finalidad de darles claridad. Luego continuo planteando preguntas a Laura para así poder dar claridad a los estudiantes las cuales se aprecian a continuación.

Profesor: Laura ¿El 90% de qué?

Laura: El 90% de las veces se va a encontrar la media en ese intervalo

Profesor: ¿Cuál es el significado del nivel de confianza?

Laura: Que el 90% de las veces.

Profesor: ¿De cuáles veces?

Laura: De las veces en que usted… jejeje… osea el 90% de las veces en que uno… eeehh… halla las muestras, se va a encontrar la media muestral metida.

Profesor: ¿Le suena Mónica?

Mónica: No, es que las veces…

Profesor: De cada muestra se genera un intervalo, si toma 100 muestras se tienen 100 intervalos. ¿Entonces qué quiere decir el nivel de confianza del 90%?

Laura: Que de esas 100 muestras que usted tomó, en 90 se va a encontrar la media muestral.

4.1.2 Categoría 2: Parámetros de la distribución normal

A pesar del análisis que se le hizo a la expresión de la función de densidad de la distribución normal durante el proceso de instrucción, en algunos estudiantes persistió el hecho de desconocer a μ y σ como los parámetros que definen la distribución normal, en muchos casos se tiene la creencia de que es necesaria más información. Un ejemplo de esta concepción se observa en la respuesta de Harley al ítem 2a. que se muestra en la Figura 13.

Harley en el ítem 2a: "Para que una distribución normal quede completamente definida, basta con conocer la media y la desviación estándar", afirma que "también debe conocerse la X para poderla estandarizar, es decir, para hallar Z". Al parecer, Harley está pensando más en calcular la probabilidad asociada a un evento particular que en caracterizar la función de densidad. Parece ser una mala interpretación de la afirmación.

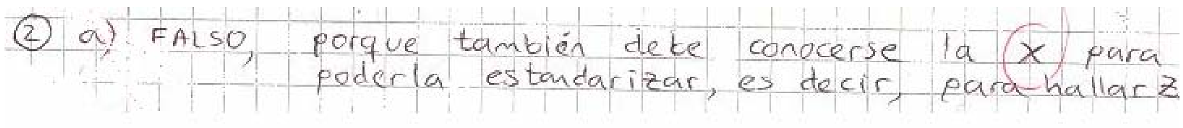

2 a) FALSO, porque también debe conocerse la X para poderla estandarizar, es decir, para hallar Z

Figura 6. Respuesta de Harley en el ítem 2 parte a.

En la entrevista realizada con algunos estudiantes se observó que los estudiantes no asocian μ y σ con las que aparecen en la expresión de la función de densidad de la distribución normal, lo cual se debe al parecer porque no recuerdan esta expresión. A continuación se muestra la evidencia de este hecho y en ella se puede observar cómo el profesor debe encaminar a los estudiantes para que perciban que μ y σ determinan la distribución normal.

Profesor: En general, en una variable aleatoria continua ¿cómo determino su distribución?

Andrea: Pues porque yo tengo que tomar todo un intervalo, no puedo tomar solo un punto porque eso me daría cero, pues la probabilidad en un punto es cero. Entonces para la continua…

Andrea evidencia lo que se había comentado antes a raíz de la argumentación de Harley, es decir, que al preguntar por la forma como se determina o caracteriza la distribución de una variable aleatoria continua se está indagando por los eventos posibles a los cuales se les puede calcular probabilidad

Profesor: ¿Cómo calculamos probabilidades de eventos en un intervalo de variable continua?, ¿cómo calculamos la probabilidad de que la variable X este en el intervalo (a, b)? Mejor dicho, ¿qué se necesita para calcular la probabilidad de una cierta variable aleatoria en un intervalo (a, b)?, es decir, ¿cómo se calcula la Pr (x Є (a, b)) siendo X una variable continua?

Mónica: Hallamos el área.

Profesor: Y para eso ¿qué se necesita?

Laura: La función que describe esa curva.

Profesor: Y esa curva ¿cómo se llama?

Andrea: Función de densidad.

Profesor: ¿Cuál es la expresión de la función de densidad de la distribución normal?

Andrea: Esa era la que teníamos que demostrar.

Mónica: Yo nunca me la aprendí.

Profesor: la expresión es $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, \quad x \in \mathbb{R}$ ¿Qué necesito para determinar completamente esta fórmula?, entre otras cosas, ¿qué valores puede tomar X en una distribución normal?

Mónica: Cualquier valor real.

Profesor: ¿Qué necesito conocer para determinar completamente esta fórmula?

Andrea: μ y σ

4.1.2.1 ¿Qué valores pueden tomar μ y σ?

Continuando con el estudio de las concepciones de los estudiantes alrededor de los parámetros de la distribución normal, se analizan ahora las respuestas que los estudiantes dieron a la afirmación: El parámetro de la localización de la normal es μ y puede tomar cualquier valor real.

Iván dice en su respuesta (ver la Figura 14) que μ solamente puede tomar valores reales positivos, desafortunadamente no da razones para ello. Se podría especular que o bien la está confundiendo con la desviación estándar o que, simplemente, en su experiencia resolviendo problemas relacionados con la distribución normal nunca se encontró con un problema donde la media tomara valor negativo. En este último caso, se impondría la práctica sobre los razonamientos teóricos, aspecto que nunca pueden olvidar los profesores y que deben evitar dando a sus estudiantes problemas que impiden generalizaciones de este tipo. De todas maneras, una cosa son los razonamiento matemáticos puros y otra cosa muy distinta, es la práctica estadística del manejo de datos.

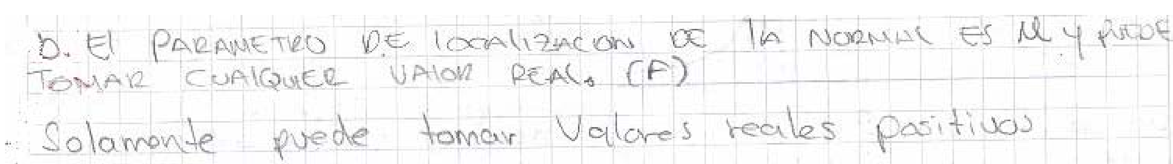

b. El PARAMETRO DE localizacion DE la NORMAL ES μ y puede TOMAR CUALQUIER VALOR REAL. (F)
Solamente puede tomar Valores reales positivos

Figura 7. Respuesta de Iván en el ítem 2 b.

Carlos responde: “El único valor que puede tomar la normal es aquel que vuelve simétrica la distribución” (Figura 8). Carlos, en su particular argumentación, cree que la simetría de la distribución depende de la ubicación de parámetro μ y no se da cuenta que la simetría se puede lograr en cualquier valor real. De hecho, John no se compromete con ningún valor posible.

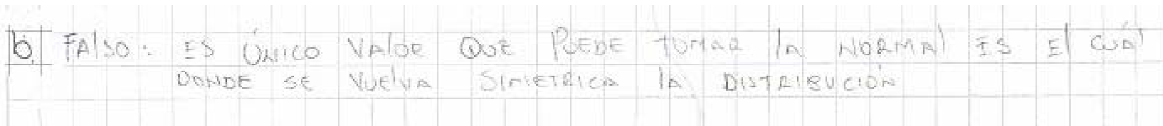

b FALSO: ES ÚNICO VALOR QUE PUEDE TOMAR LA NORMAL ES EL CUAL
DONDE SE VUELVA SIMETRICA LA DISTRIBUCIÓN

Figura 8. Respuesta de Carlos al ítem 2b.

En los apartes de la entrevista que a continuación se muestra, se puede observar cómo ahora los estudiantes al analizar la forma de la función de densidad de la normal adoptan que el parámetro σ asociado con la desviación estándar puede tomar cualquier valor real. Ya que de todas maneras la función de densidad sigue bien definida en el sentido de que tiene imagen para cualquier valor real x, olvidando sí, que la función densidad por definición es no negativa.

Profesor: ¿Qué valores puede tomar μ?

Laura: μ puede tomar cualquier valor real.

Profesor: ¿y σ?

Laura: σ tiene que ser distinta de cero, o sea todos los reales menos el cero.

Profesor: ¿Qué opina Mónica?

Mónica: Estoy de acuerdo con Laura.

Profesor: Y ¿usted Pili?

Andrea: No sé si lo que voy a decir es una locura, pero a mí me parece que no… porque σ me está diciendo la distancia que hay de un dato observado a la media. …para mi debería ser, esto, partir de uno en adelante.

Profesor: ¿0.5 no sirve?

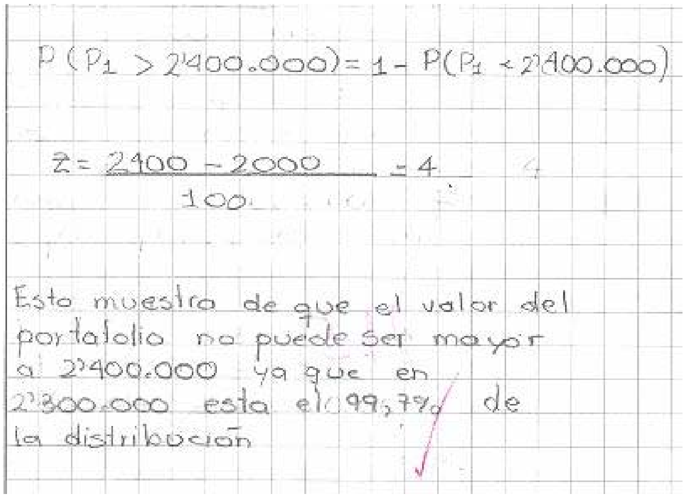
Andrea: Ah no, sí. Mayores que cero.

La argumentación de Andrea en la cual recurre a la idea de que las desviaciones estándar se utilizan como medidas de distancia (seguramente pensando en el proceso de estandarización) para concluir que necesariamente debe tomar valores positivos puede ser una argumentación no ortodoxa, pero es correcta.

4.1.2.2 Estimación de la desviación estándar para la suma de variables aleatorias independientes

Las respuestas de los estudiantes al ítem 3, el problema de los portafolios de acciones, evidenciaron la dificultad que representa para ellos el cálculo de la desviación estándar de la suma de variables aleatorias. En este problema el 46.1% de los docentes presentaron dificultad en este proceso, el 23.1% no supo qué cálculos realizar, el 15.3% hizo el cálculo de probabilidades para el primer portafolio y no continuó, el 7.6% no supo calcular ni la media ni la desviación estándar para variables aleatorias y el 7.9% restante realizó correctamente los cálculos. En la Figura 9, la **¡Error! No se encuentra el origen de la referencia.**y la Figura 10, se pueden apreciar algunas respuestas de los estudiantes.

Diana hizo el proceso de estandarización, indicó la probabilidad que debía calcular y se apoyó en la regla del 68-95-99,7 de la distribución normal para argumentar.



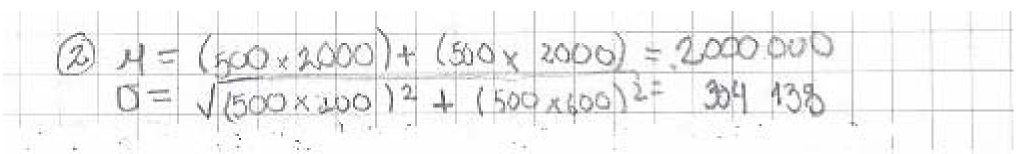
$P(P_1 > 2'400.000) = 1 - P(P_1 < 2'400.000)$

$z = \frac{2400 - 2000}{100} = 4$

Esto muestra de que el valor del portafolio no puede ser mayor a 2'400.000 ya que en 2'300.000 esta el 99,7% de la distribución

Figura 9. Respuesta de Diana utilizando la regla del 68-95-99,7 como estrategia

Alejandra por el contrario utilizó las fórmulas para realizar la estimación de la desviación estándar para el portafolio 2 y de igual manera con los demás portafolios. En la siguiente figura se aprecia los cálculos realizados por el estudiante.



(2) $\mu = (500 \times 2000) + (500 \times 2000) = 2.000.000$

$\sigma = \sqrt{(500 \times 200)^2 + (500 \times 400)^2} = 304\ 138$

Figura 10. Respuesta de Alejandra en el ítem 3

En los apartes de la entrevista que a continuación se muestra, se puede observar cómo los estudiantes asumen que no es necesario hacer el cálculo de la desviación estándar para cada portafolio.Es evidente que no se desconoce la fórmula para hacer este cálculo, sin embargo, al momento de estandarizar notan que deben hacerlo, pero en el momento no recuerdan la fórmula y se confunden con el hecho de que están calculando probabilidades.

Profesor: ¿En el ítem 3a había que calcular la desviación estándar para variables independientes para cada portafolio?

Andrea: No, ya sabemos la media y la desviación estándar para cada portafolio y no es necesario hacer más cálculos.

Profesor: En el portafolio 1 hay 1000 acciones de W, conocemos la desviación de W pero ¿conocemos la desviación estándar de las 1000 acciones? y el portafolio 2 consiste de 500 acciones de W y 500 acciones de V. ¿en este caso cuál desviación, la de W o la de V?, ¿Cómo se calcula la desviación estándar de una suma de variables aleatorias independientes?

Laura: Si conocemos el valor de la desviación estándar de X_1 y de X_2 entonces la desviación estándar de la suma de variables aleatorias es $Sd(X_1 + X_2) = \sqrt{Sd(x_1)^2 + Sd(x_2)^2}$, donde Sd es una manera de llamar a la desviación estándar.

Profesor: ¿Cuál es el portafolio más seguro? y ¿cuál es el más riesgoso? ¿Qué tenemos que mirar para responder a esta pregunta?

Laura: La desviación estándar de cada portafolio.

Profesor: Laura, si tienes claro como se calcula la desviación estándar de una suma de variables aleatorias independientes, ¿por qué no hizo estos cálculos en el cuestionario?

Laura: Porque me enfoque en hallar la solución, cuando fui a estandarizar y note que debía calcular la desviación no tuve presente la fórmula. Además como estábamos trabajando con probabilidades me perdí y no supe que hacer.

4.1.3 Categoría 3: Estrategias para resolver problemas relacionados con la distribución normal

A continuación se describen las estrategias utilizadas por los estudiantes para resolver algunos de los ítems propuestos.

4.1.3.1 Uso de la gráfica de la distribución normal como apoyo en la argumentación y resolución de problemas

El 84.6% de los docentes en formación hizo uso de la gráfica de densidad de la distribución normal para apoyarse en la resolución de los diferentes ítems del cuestionario, en estas gráficas indicaron de manera correcta la media y el área que se les pedía calcular.

En la Figura 11y Figura 12 se pueden observar diferentes modelos de la gráfica de densidad utilizados por los estudiantes para apoyar sus respuestas.

Diana, en el ítem 1a. y 1b. del cuestionario 1 utiliza la siguiente gráfica, en la cual indica la ubicación de la media y el área que se le pide calcular; además, determina correctamente la escala de valores en términos de desviaciones estándar.

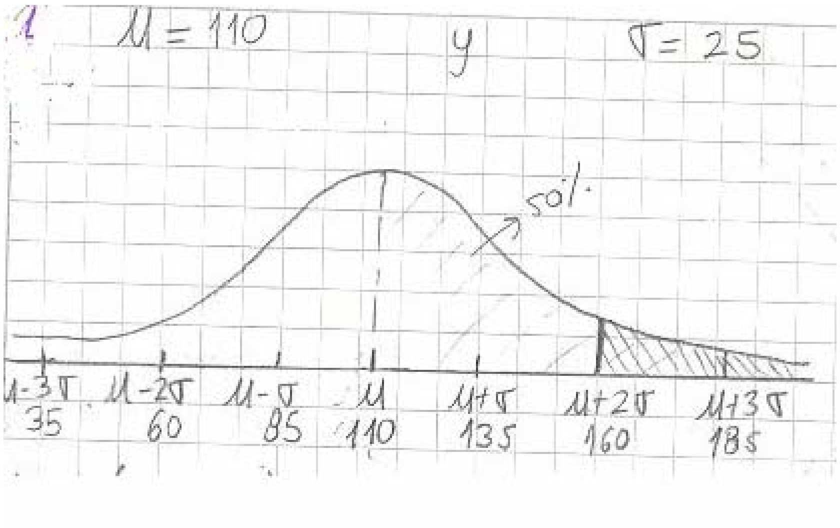


Figura 11. Respuesta de Diana al ítem 1 parte a y b

Ivonne utiliza la siguiente gráfica para apoyar la respuesta del ítem 2d.

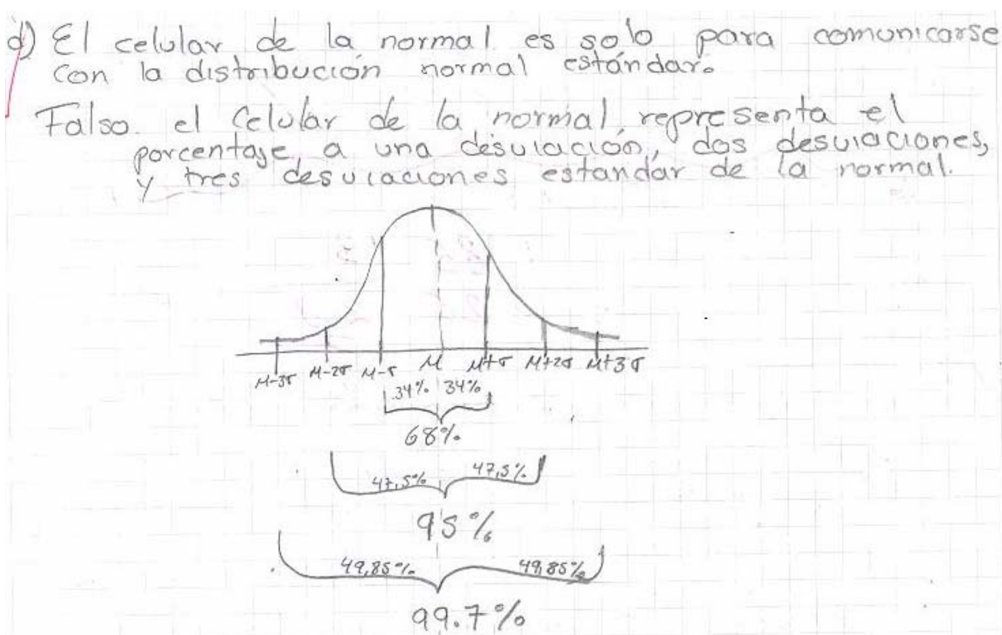


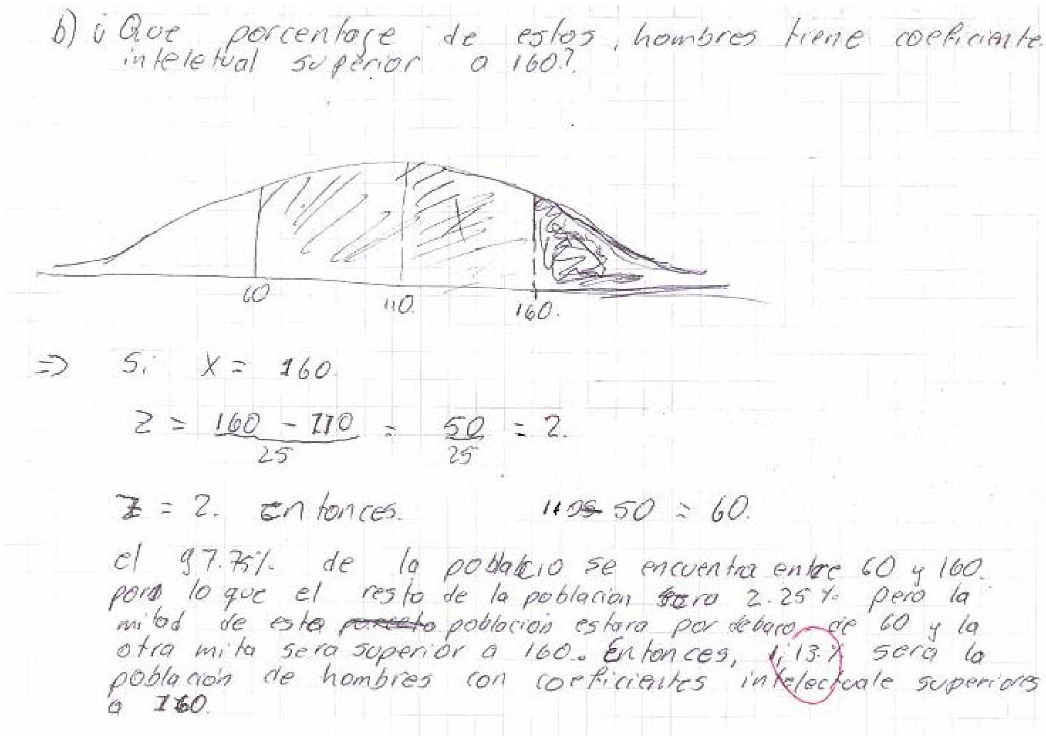
Figura 12. Respuesta de Ivonne al ítem 2 parte d

El 15.4% restante no se apoyó en la gráfica de la distribución normal pero sus respuestas y cálculos fueron acertados.

4.1.3.2 Manejo de la Tabla de Probabilidades de la Normal Estándar

El cálculo de probabilidades de eventos asociados a una variable continua se realiza con el cálculo de áreas bajo una curva, su función de densidad, en nuestro caso la curva de la distribución normal, lo cual implica, entre otras cosas, que los estudiantes manejen la tabla de probabilidades de la Distribución Normal Estándar, aspecto éste que parece no es tan evidente como se pueda pensar a priori. En la Figura 13 se muestra la respuesta de Julio al ítem 1 parte b, lo cual es evidencia de lo descrito anteriormente.

Julio realizó el proceso de estandarización de manera correcta. Para calcular la probabilidad hace un análisis apoyándose en la gráfica, en la cual dibuja el área que va a calcular. Al buscar el valor de z en la tabla nota que el porcentaje restante es el que él necesita, sin embargo, malinterpreta el valor de probabilidad dado por la Tabla, ya que Julio lo asume como si fuera la probabilidad de caer en el intervalo (60, 160), error que después lo lleva a dividir por 2.


b) ¿Que porcentaje de estos, hombres tiene coeficiante inteletual superior a 160?

60 110 160.

=> Si X = 160.

$Z = \frac{160 - 110}{25} = \frac{50}{25} = 2.$

Z = 2. entonces. ~~110~~ 110 − 50 = 60.

el 97.75% de la poblacio se encuentra entre 60 y 160. para lo que el resto de la poblacion sera 2.25% pero la mitad de esta ~~porcento~~ poblacion estara por debajo de 60 y la otra mita sera superior a 160. Entonces, 1.13% sera la poblacion de hombres con coeficientes intelectuale superiores a 160.

Figura 13. Respuesta de Julio al ítem 1 parte b

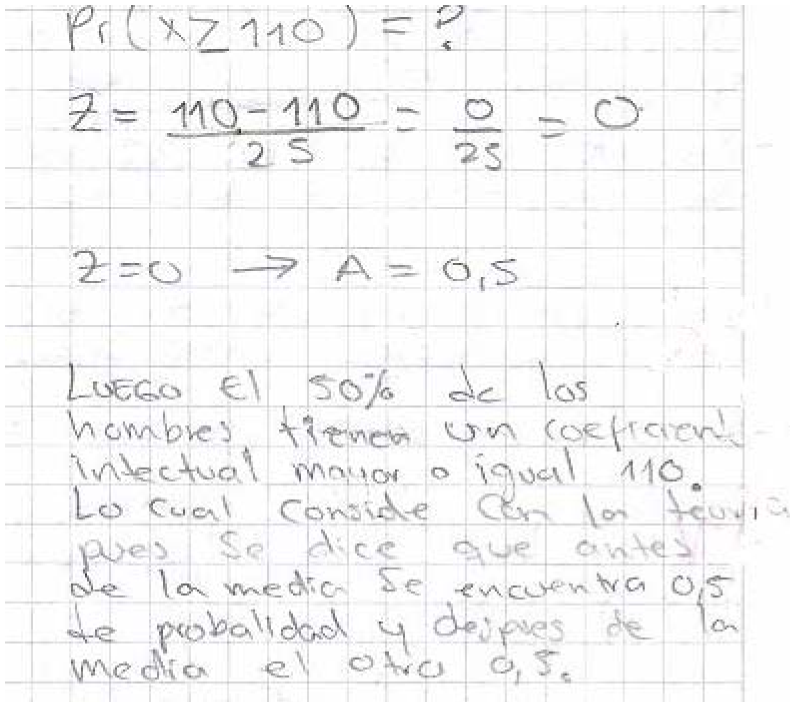
4.1.3.3 Uso de las propiedades de la distribución normal

Los docentes en formación hicieron uso de algunas propiedades de la distribución normal para su argumentación.

4.1.3.3.1 Propiedad simétrica

La propiedad simétrica es de uso frecuente por los estudiantes en la resolución de problemas y argumentación de las soluciones propuestas. En la Figura 14 se muestra la respuesta de Iván al ítem 1a.

Iván se apoya en la propiedad simétrica para sustentar su respuesta obtenida mediante el proceso de estandarización.



$Pr(X \geq 110) = ?$

$Z = \frac{110-110}{25} = \frac{0}{25} = 0$

$Z=0 \rightarrow A = 0,5$

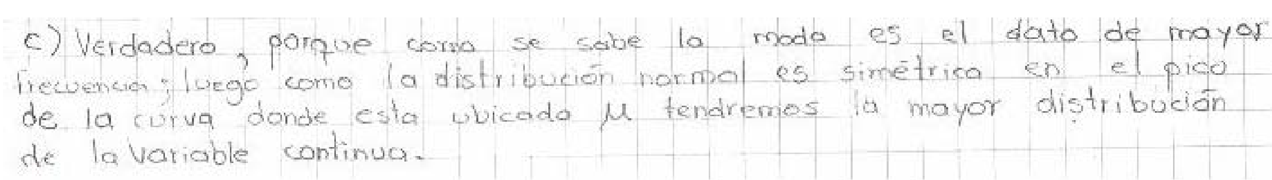
Luego el 50% de los hombres tienen un coeficiente intectual mayor o igual 110. Lo cual conside con la teoria pues se dice que antes de la media se encuentra 0,5 de probalidad y despues de la media el otro 0,5.

Figura 14. Respuesta de Iván al ítem 1 parte a.

Ahora no siempre la propiedad simétrica fue bien utilizada por los estudiantes, no es sino recordar su uso para intentar probar que la media y la moda coinciden en la distribución normal.

Ante la afirmación de que en una curva normal la media es igual a la moda, todos los estudiantes respondieron que era verdadera. La argumentación generalizada fue la propiedad simétrica que posee la distribución. Algunos de los argumentos de los estudiantes se presentan en seguida.

Andrea, intenta relacionar los valores de la moda y la media a través de la propiedad simétrica diciendo: “…la moda es el dato de mayor frecuencia; luego como la distribución normal es simétrica en el pico de la curva donde está ubicada μ tendremos la mayor distribución de la variable continua" (ver la Figura 9). En esta argumentación aparecen varios elementos que traen a la luz los significados que la estudiante posee acerca de varios conceptos estadísticos y de la función de densidad de la distribución normal. En primer lugar, extrapola el concepto que se tiene de la moda en una muestra a una variable aleatoria continua cuando afirma que es "el dato de mayor frecuencia". Luego, por simetría, infiere que en el pico donde está ubicada la media se tiene la mayor distribución. Aunque la estudiante no lo hace explícito, tal vez su invocación de la simetría debe relacionarse con el hecho de que la función viene creciendo y en μ está el eje de simetría, la curva desciende de ahí en adelante. Esto permite afirmar que en ese pico se tiene el valor máximo de la función, valor que es precisamente la moda de la distribución. En su conclusión, cuando afirma "tendremos la mayor distribución de la variable continua", Andrea evidencia que su interpretación de la función de densidad de probabilidad aún está muy atada a los análisis de muestras y a los histogramas de valores. A propósito, no estaría de más realizar una investigación acerca de las concepciones que los estudiantes universitarios podrían tener acerca de las funciones de densidad de las variables aleatorias continuas.



c) Verdadero, porque como se sabe la moda es el dato de mayor frecuencia; luego como la distribución normal es simétrica en el pico de la curva donde esta ubicado μ tendremos la mayor distribución de la Variable continua.

Figura 15. Respuesta de Andrea al ítem 2 c.

Por el contrario otros estudiantes asumieron que la media y la moda en la distribución normal no son iguales basados simplemente en sus definiciones. La argumentación de Mónica, que se muestra en la Figura

17, es un ejemplo de esta forma de argumentación: “ya que la moda es el valor que más se repite, y la media es una medida de centro la cual me divide los datos en dos proporciones iguales”.

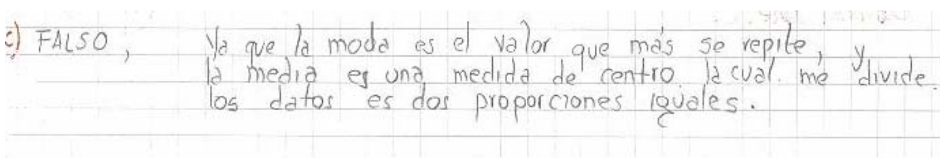


Figura 16. Recorte de la respuesta de Mónica en el ítem 2 parte c.

En la entrevista que se le realizó, Mónica da mayores detalles de su argumentación.

Profesor: ¿En la distribución normal la media es igual a la moda?

Mónica: Yo respondí que era falso.

Profesor: ¿Por qué?

Mónica: Porque la verdad no me acordaba muy bien de… esto… la concepción de la moda en una distribución normal…. No la tenía muy clara.

Profesor: ¿Qué pensaba que era la moda en la distribución normal?

Mónica: O sea, nooo… nunca o sea… no la… como que no la había trabajado ese punto, que la moda era igual a la media.

Profesor: ¿Qué es para usted la moda?

Mónica: Para mí la moda es el valor que más se repite, pero como en la distribución normal la variable es continua, entonces no puedo hablar del dato que más se repite.

Aquí, Mónica, al contrario de Andrea, refleja el problema que representa el contar casos en una variable aleatoria continua cuando se está tratando en términos teóricos.

Profesor: ¿Qué será la moda en una variable aleatoria continua?

Mónica: O sea, hmmm… no sé.

El profesor al notar que los estudiantes entrevistados no tenían claro el concepto de moda para variables aleatorias continuas procedió hacer varias gráficas en el tablero con la intención de que las estudiantes ubicaran la moda en dichos gráficos y de esta manera clarificaran el concepto de moda para este tipo de variables.

4.1.3.3.2 Regla del 68-95-99,7

En el ítem 2d se pudo observar que todos los encuestados reconocen el enunciado de la regla del 68-95-99,7, o el celular de la normal como lo conocen los estudiantes.Sin embargo, no siempre la interpretan de forma adecuada, tal como lo muestra la siguiente entrevista.

Profesor: Mónica, cómo entiende la siguiente afirmación y qué quiere decir: El celular de la normal es solo para “comunicarse” con la distribución normal estándar. ¿Usted que entendió cuando leyó el término “comunicarse”?

Mónica: Pues, ¿si funciona solo para la distribución normal estándar?

Profesor: ¿Está de acuerdo con esta afirmación?

Mónica: Sí.

Profesor: ¿Por qué?

Mónica: Pues lo asocié con la estandarizada porque la mayoría de veces resolvíamos ejercicios de ese estilo.

Profesor: ¿Siempre la han utilizado para eso?

Mónica:Hmm

Profesor: Laura, ¿está usted de acuerdo con la afirmación?

Laura: Sí.

Profesor: ¿Por qué?

Laura: Pues porque el celular de la normal nos indica que a 1 desviación estándar de la media esta el 68% de los datos, a 2 desviaciones estándar de la media el 95% de los datos y a 3 desviaciones estándar de la media esta el 97% de los datos siendo en la curva de la distribución normal.

Profesor: ¿El celular de la normal se aplica para cualquier normal?

Laura: Es que es más fácil verla en la distribución normal estándar que en cualquier distribución normal.

Profesor: ¿Por qué?

Laura: Porque cuando estandarizamos ya sabemos de qué…se puede utilizar para la distribución normal estándar pero también para cualquier normal y es más fácil verla en la distribución normal estándar que en cualquier distribución normal.

Profesor: ¿No acabo de entender qué es más fácil de ver?

Laura: Por que cuando yo estandarice dije bueno, μ se encuentra en cero y tengo a una desviación estándar el 68% a dos desviaciones estándar el 95% y a tres desviaciones estándar esta el 99.7%.

Profesor: ¿Cuál es la diferencia entre una desviación normal estándar y una no estándar? ¿Tiene que ver que σ valga 7, 10, 20, 150?,

Laura: Pues no porque en la distribución normal estándar σ vale uno.

Profesor: ¿Entonces como se traduce el celular de la normal en la normal estándar?

Laura: Que en el intervalo entre -1 y 1 encontramos el 68% de los datos, en el intervalo de -2 a 2 encontramos el 95% de los datos y en el intervalo de -3 a3 encontramos el 99.7% de los datos.

Profesor: Ahora traduzca el celular en cualquier distribución normal.

Laura: Estaría diciendo que entre en el intervalo $(X \pm \sigma)$ se encuentra el 68% de los datos, en el intervalo $(X \pm 2\sigma)$ se encuentra el 95% de los datos y en el intervalo $(X \pm 3\sigma)$ se encuentra el 99.7% de los datos.

El diálogo anterior permite apreciar cómo los estudiantes creen que la regla 6895997 propia de cualquier normal es exclusiva de la normal estándar. Sus razones parecen estar basadas en la práctica, como bien lo afirma Mónica: "Pues lo asocié con la estandarizada porque la mayoría de veces resolvíamos ejercicios de ese estilo". Nuevamente se observa que los estudiantes no lograron asimilar del todo la "equivalencia" que existe entre cualquier distribución normal y la normal estándar y que se logra gracias a la transformación lineal expresada en el proceso conocido como "estandarización". Esta equivalencia permite transformar cualquier evento o intervalo asociado a una normal cualquiera del que se quiere conocer su probabilidad a uno asociado con la normal estándar que tiene exactamente el mismo valor de probabilidad.

4.1.4 Categoría 4. El teorema central del límite: Implicaciones y uso en la resolución de problemas

Abordamos este apartado con una explicación breve sobre el teorema central del límite y su conveniencia en la resolución de problemas, para luego analizar las estrategias y dificultades presentadas por los estudiantes al momento de resolver los ítems propuestos en el cuestionario 2.

4.1.4.1 Teorema Central del Límite

El Teorema Central del Límite (TCL) indica que, en condiciones muy generales, la distribución de la media muestral tiende a una distribución normal cuando la cantidad de variables es muy grande. Este teorema tiene una gran aplicación en inferencia estadística, pues muchos parámetros de diferentes distribuciones de probabilidad, como la media, pueden expresarse en función de una suma de variables. Permite también aproximar muchas distribuciones de uso frecuente: binomial, poisson, chi cuadrado, t-student, etc., cuando sus parámetros crecen y el cálculo se hace difícil (Moore 1995, Pág. 303). Por otro lado, la suma de variables aleatorias aparece en forma natural en la descripción de fenómenos reales, ya que permite, por ejemplo, inferir la media de una población a partir de una muestra.

El teorema se apoya y relaciona entre sí con otros conceptos y procedimientos básicos en estadística, como los de variable aleatoria y sus transformaciones, distribución muestral, estandarización, cálculo de probabilidades, etc., lo que contribuye a dificultar su comprensión dificultando el aprendizaje de los estudiantes.

4.1.4.2 Enunciado e implicaciones del teorema central del límite

Reconocer el enunciado del TCL es de gran importancia para resolver problemas. El cuestionario evidenció que los docentes en formación reconocen el enunciado del teorema central del límite. Sin embargo, no perciben algunas de sus implicaciones como se muestra a continuación.

El 80% de los estudiantes reconocen el teorema central del límite para la suma de variables independientes e idénticamente distribuidas. A continuación se presenta la evidencia que apoya esta afirmación:

Alejandra, en su respuesta al ítem 3, "$\bar{X}$ tiene aproximadamente una distribución normal cuando la muestra es grande", reconoce el enunciado del teorema central del límite y a su vez identifica los parámetros de la distribución normal asociada a la población en estudio. En la siguiente figura se observa dicha respuesta.

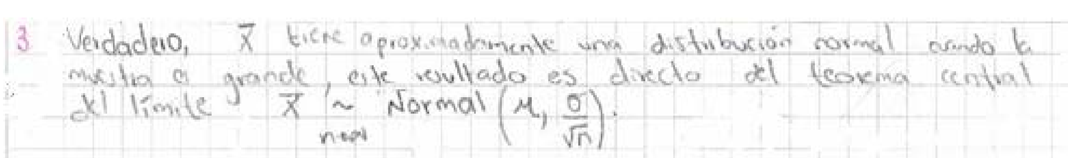
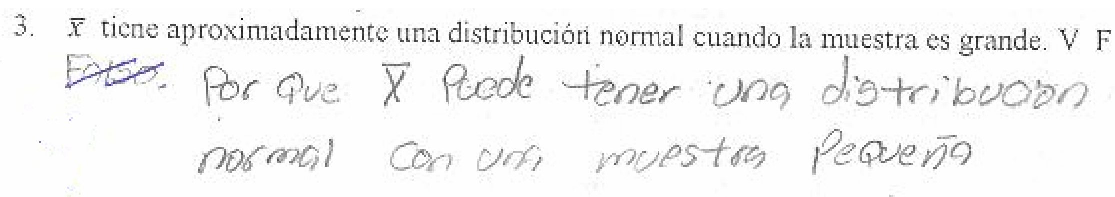


Figura 17. Respuesta de Alejandra al ítem 3

Sin embargo, existen otros estudiantes, como Nicolas, que no identificaron el enunciado del TCL en la afirmación del ítem 3.

En la siguiente figura se puede apreciar la respuesta de Nicolas:



3. $\bar{x}$ tiene aproximadamente una distribución normal cuando la muestra es grande. V F

Falso. Por que $\bar{X}$ Puede tener una distribucion normal con una muestra Pequeña

Figura 18. Respuesta de Nicolas al ítem 3

Al preguntarle personalmente al estudiante el porqué de su respuesta, comentó que fue debido a que en el momento no asoció la afirmación con el enunciado del TCL. Al preguntarle por el enunciado del teorema, respondió que en el momento no se acordaba.

Se presume que la afirmación de Nicolas puede ser debida a que en el caso particular de que la muestra sea tomada de una población con distribución normal, la media muestral también es normal sin importar su tamaño.

Sergio en su respuesta al ítem 3 (Figura 19), asume que $\bar{X}$es un valor estimado y a su vez explica que si $\bar{X}$ fuera una variable aleatoria, entonces sería el enunciado del TCL, resaltando un aspecto que ha sido reportado en otros trabajos (Tauber, 2001; Insunza, 2006) como es la confusión que muchas veces se presenta entre estimador y valor estimado, siendo el primero una variable aleatoria en tanto que el segundo es una realización concreta de esa misma variable aleatoria ante una muestra.

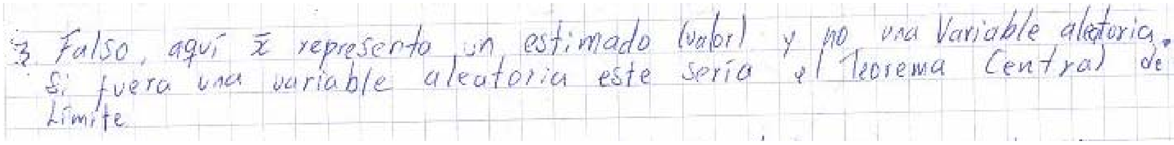

3. Falso, aquí $\bar{x}$ representa un estimado (valor) y no una Variable aleatoria. Si fuera una variable aleatoria este sería el Teorema Central de Límite

Figura 19. **Repuesta de Sergio al ítem 3**

4.1.4.3 El Teorema central del límite como estrategia en la resolución de problemas

El uso del TCL para la argumentación no fue utilizado como se esperaba inicialmente, esto se observó en la respuestas ofrecidas por los estudiantes al ítem 4, en el cual se afirmaba que la forma de la distribución de $\bar{X}$ depende de la forma de la distribución de la población.

Laura, al igual que otros estudiantes, no tiene en cuenta el TCL; en la Figura 20 se puede observar su respuesta.

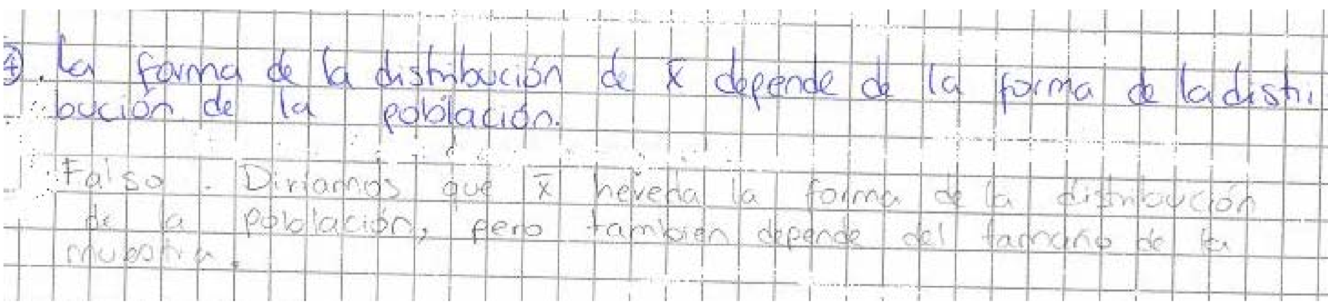

4. La forma de la distribución de $\bar{x}$ depende de la forma de la distribución de la población.
Falso. Diríamos que $\bar{x}$ hereda la forma de la distribución de la población, pero también depende del tamaño de la muestra.

Figura 20. Respuesta de Laura en el ítem 4

A continuación se presentan los detalles del diálogo que se dio entre el profesor y Laura alrededor de esta pregunta.

Profesor: Entonces, ¿La forma de la distribución de $\bar{X}$ depende de la forma de la distribución de la población?

Laura: Yo respondí que es falso, diríamos que $\bar{X}$ hereda la forma de la distribución de la población, pero también depende del tamaño de la muestra…

Profesor: Es decir…

Laura: O sea… es que… para este punto yo me acordé de un problema que veíamos en el libro y tenía un tamaño muestral… eeee… tenia la distribución de… no me acuerdo en este momento, pero era una serie de puntos.Entonces a medida que iban aumentando el tamaño, la distribución de $\bar{X}$ se iba… se iba más o menos acercando a la poblacional. Entonces yo acudí a eso para responder este punto.

Profesor: ¿Se acerca a la distribución poblacional con muchos datos o pocos datos?

Laura: Con muchos datos.

Profesor: ¿y entonces el teorema central del límite?

Laura: …

Profesor: Si los datos son obtenidos de una distribución normal ¿Qué distribución tiene $\bar{X}$?

Laura: La distribución normal.

Profesor: Y si los datos son obtenidos de una población cuya distribución no es normal ¿Qué pasa con la distribución de $\bar{X}$?

Laura: Pues a medida que se va aumentando el tamaño muestral, se aproxima a la normal.

Profesor: ¿Entonces para valores pequeños?

Laura: Para valores pequeños pues… no. O sea… no puedo decir que no es normal.

Profesor: ¿Pero la distribución de $\bar{X}$ se parece a quién?

Laura: A la de la población… sí a la poblacional. Hmmm

En este diálogo se aprecia bien cómo Laura posee la idea de que la forma de la distribución de $\bar{X}$ depende del tamaño muestral: “con pocos datos se parece a la forma de la distribución poblacional y a medida que se aumentan su forma se hace más parecida a la distribución normal”.Sin embargo, a pesar de que Laura posee esta idea no la uso de para dar respuesta a este ítem.

5 CONCLUSIONES

Para finalizar este trabajo presentamos en este capítulo las conclusiones más importantes de nuestro estudio.

1. Se pudo observar que el cálculo de probabilidades asociado a intervalos de variables aleatorias normales resulta complicado para la mayoría de los docentes en formación que conformaron nuestra muestra, ya que después de realizar adecuadamente el proceso de estandarización, en el momento de hacer el cálculo de probabilidades se confunden al no tener en cuenta cuál es el intervalo (área) que se está pidiendo calcular.

2. Nuestro análisis muestra que el proceso de estandarización es el que menos dificultad presenta, sin embargo, es asumido solo como un proceso que indica a qué distancia de la media en términos de desviaciones estándar se encuentra un valor de la variable aleatoria X, desconociendo que es un proceso que transforma una variable normal a una normal estándar.

3. Algunas propiedades de la distribución normal, como la simétrica y la regla del 68-95-99,7, son utilizadas frecuentemente como

4. Algunos estudiantes no asumen a μ y a σ como los parámetros que definen la distribución normal. Ellos consideran que es necesaria más información, olvidando que en la función de densidad los únicos términos desconocidos son precisamente μ y σ, además se observó que los estudiantes no tienen claro qué valores pueden tomar dichos parámetros.

5. Los alumnos prefieren, en general, las representaciones gráficas a las numéricas. La gráfica como herramienta y estrategia para resolver problemas siempre fue utilizada adecuadamente. En pocos casos se usan procesos algebraicos, sin embargo, cuando se utilizaron produjeron buenas respuestas.

6. En algunos estudiantes persiste la dificultad para reconocer los tipos de problemas para los cuales el Teorema Central del Límite (TCL) es una herramienta de análisis.

7. Algunos estudiantes asumen que el TCL es aplicable sólo a poblaciones específicamente normales, olvidando que la principal aplicación del TCL es que es aplicable a cualquier población, sin importar su distribución, en este caso el cuidado que se debe tener es con el tamaño muestral, lo cual no es del todo asumido por los estudiantes.

8. Los estudiantes confunden la simetría de la función de densidad de probabilidad con el dominio de esta función. Es decir, la simetría respecto a la media conduce a afirmar que la medida de probabilidad se divide en dos: 50% antes de la media y 50% después de la media y que éstas mitades se relacionan con los valores posibles al afirmar que $P(X \leq \mu) = P(X \geq \mu) = {}^{1}/_{2}$. En otras palabras cuando se afirma "50%" se hace con relación a la probabilidad y no a la de los valores posibles de la normal.

9. Adicionalmente se presentaron dificultades en el cálculo de intervalos de confianza y en la comprensión del nivel de confianza que los acompaña, la noción de esperanza, en términos estadísticos, también presentó dificultad en el momento de responder los ítems que requerían de su comprensión.

10. Se recomienda para estudios futuros analizar el lenguaje utilizado por los estudiantes. Además de poner especial atención

6 REFERENCIAS

Alvarado, H y Batanero, M. C. (2007). Dificultades de Comprensión de la Aproximación Normal a la Distribución Binomial. Ideas y Recursos para el Aula Nº. 67, 2007, Chile.

Alvarado, H. (2007). Significados Institucionales y Personales del Teorema Central del Límite en la enseñanza de estadística en ingeniería. Tesis Doctoral no publicada, Universidad de Granada, Granada, España

Godino, J. (1996) Significado y comprensión de los objetos matemáticos. En: L.Puig, y A. Gutierrez (Eds.), Proceedings of the 20th Conference of the Internacional Group for the Psychology of Mathematics Education (v.2, pp. 417-424). Universidad de Valencia.

Inzunsa, S. (2006).Students' Errors and Difficulties for Solving Problems of Sampling Distributions by Means of Computer Simulation.International Congress of Teaching Statistics, ICOTS 7, Salvador (Bahía), Brasil.

Moore (1995).Estadística Aplicada Básica. Antoni Bosch, Editor. Barcelona España.

Tauber, L.M. (2001). La construcción del significado de la distribución normal a partir de actividades de análisis de datos. Tesis Doctoral no publicada. Universidad de Sevilla y Granada, Sevilla (España)

ANEXOS

ANEXO 1

CUESTIONARIO 1. Distribución normal: propiedades y parámetros que la definen

1) La distribución de los coeficientes intelectuales de hombres entre 20 y 34 años tiene aproximadamente una distribución normal de media $\mu = 110$ y una desviación estándar $\sigma = 25$.

   c. De los hombres entre 20 y 34 años, ¿Qué porcentaje tiene un coeficiente intelectual superior a 110?

   d. ¿Qué porcentaje de estos hombres tiene un coeficiente intelectual superior a 160?

2) En las siguientes preguntas Responda Verdadero o Falso y justifique la respuesta.

   e. Para que una distribución normal quede completamente definida, basta con conocer la media y la desviación estándar.

   f. El parámetro de localización de la normal es μ y puede tomar cualquier valor real.

   g. En la curva normal la media es igual a la moda.

   h. El celular de la normal es solo para “comunicarse” con la distribución normal estándar.

3) Considere un mercado de acciones idealizado. Solamente cuatro compañías negocian en este mercado: Compañía X, Compañía Y, Compañía V y compañía W. Su asegurador le dice que el precio actual de las acciones para cada una de éstas es $200 pero que tienen diferencias en la volatibilidad (variabilidad) de los precios en el tiempo. El precio de las acciones en un mes se puede considerar aproximadamente normal, $X \sim N(10, 1)$, $Y \sim N(10,3)$, $V \sim N(10,3)$ y $W \sim N(10, 0.5)$.
También indica que las acciones se comportan independientemente. Usted tiene $2,000.000 para invertir y existen tres portafolios posibles para realizar la inversión.
   1. El portafolio 1 consiste de 1000 acciones de W
   2. El portafolio 2 consiste de 500 acciones de W y 500 acciones de V.
   3. El portafolio 3 consiste de 250 acciones de cada una de los cuatro tipos.

Sea Pi el valor del portafolio i para el próximo mes contado a partir de la fecha.

c. Para cada uno de los tres portafolios calcule las siguientes probabilidades:
   iii. Pr(Pi>2,4 millones)
   iv. Pr(Pi<1 millón)
d. ¿Qué portafolio es el más seguro y cuál es el más riesgoso’
e. Un amigo requiere obtener una ganancia de 5 millones por su inversión. Sin calcular ninguna probabilidad, ¿qué portafolio debería escoger y por qué?

ANEXO 2

CUESTIONARIO 2. Teorema Central del Límite

Responda falso o verdadero y justifique su respuesta

1. En una distribución normal, el 50% de las medidas caen por encima de la media.

2. Si una variable está distribuida normalmente, los casos extremos son poco frecuentes.

3. $\bar{X}$ tiene aproximadamente una distribución normal cuando la muestra es grande.

4. La forma de la distribución de $\bar{X}$ depende de la forma de la distribución de la población.

5. Para que la desviación estándar de la distribución de $\bar{X}$ se reduzca a la mitad basta que la muestra tenga el doble de tamaño.

6. El teorema central del límite dice que en muestras grandes, la media muestral $\bar{X}$ tiene que estar cerca de la media poblacional μ .

7. ¿Qué le ocurre a $\bar{X}$ cuando la distribución poblacional no es normal?

8. La naturaleza de los casinos es tal que una gran cantidad de dinero fluye en ambas direcciones entre el casino y los jugadores. Los jugadores a menudo ganan lo suficiente para permanecer optimistas y continuar jugando. Lenta pero inexorablemente, sin embargo, el casino hace dinero. Los juegos están estructurados de tal forma que las posibilidades de una apuesta dada favorecen al casino más que al jugador, pero solamente por una pequeña cantidad. como este es el caso, muchas personas se preguntan cómo es que los casinos funcionan para hacer grandes cantidades de dinero en el largo plazo. Todo se explica con las distribuciones muestrales de los promedios, como este ejercicio va a demostrar.

Usaremos una apuesta simple en la ruleta como ilustración. La ruleta tiene 18 números negros, 18 números rojos y 2 números verdes (0, 00) el casino gana \$ 1 si sale un numero rojo o verde (20 chances de 38) y pierde \$ 1 si sale un numero negro (18 de 38). Así, las ganancias del casino en la i-ésima es una variable aleatoria $\overline{X_i}$ con función de probabilidad $P(X = 1) = \frac{20}{38}$; $P(X = -1) = \frac{18}{38}$

Los resultados de las diferentes apuestas son independientes.

g) Encuentre la media y la desviación estándar de X .

h) (i) encuentre la media y la desviación estándar del total de ganancias del casino cuando se realizan 50 apuestas de \$1.

(ii) Haga lo mismo para las ganancias promedio del casino.

i) Escriba las respuestas correspondientes a las preguntas en (b) para un número n cualquiera.

j) ¿Qué proporción de jugadores de los que realizan 50 apuestas de $1 en el transcurso de una noche ganan dinero?

k) ¿Cuál es la probabilidad de que el casino pierda dinero con 1000 de estas apuestas?, ¿en 100.000?,¿en 1,000.000?

l) Haga una tabla de intervalos de confianza de tres desviaciones estándar con 4 filas y 2 columnas. Intervalos en las primeras cuatro filas se refieren a conjuntos de 50, 1000, 100.000 y un millón de apuestas de $1, respectivamente. Coloque intervalos para el promedio de las ganancias del casino en la segunda columna.

Explique ahora porqué los casinos hacen tanto dinero, mientras que al mismo tiempo los jugadores ganan lo suficiente para permanecerse jugando.