Un enfoque teórico mediante simulaciones para la enseñanza de la prueba de hipótesis.

Jhan Carlos Jaimes Parra
Leidy Lizeth Figueroa Muñoz

Trabajo de grado presentado como requisito parcial para optar al título de Licenciados en Matemáticas

Director
PhD. en Ciencias Estadística Andrés Sebastián Ríos Gutiérrez

Universidad Industrial de Santander
Facultad de ciencias
Escuela de matemáticas
Bucaramanga
2026

Dedicatoria

A Dios, por darme vida, salud y fortaleza en cada paso. A mis padres, Ángel y María, por su amor incondicional y apoyo constante, siendo el pilar de mi formación. A mi familia, por su cariño y palabras de aliento, y a mis abuelos Pastora y Gerardo, cuyo recuerdo vive siempre en mí. A mis amigos, por su compañía en este camino, y especialmente a mi novia, por su amor, paciencia y apoyo incondicional, fundamentales para alcanzar este logro.

Jhan jaimes

Dedico este trabajo a Dios, por guiarme y brindarme salud y sabiduría a lo largo de mi vida. A mi madre, Benilda, por su amor y apoyo incondicional, siendo el pilar de mis logros. A mi hermano y abuelos, por su motivación, sabiduría e inspiración constante. A mis tíos y amigos, por sus consejos, compañía y momentos de ánimo. Y de manera muy especial, a mi novio, por su amor, apoyo y motivación, fundamentales en este camino.

Leidy Figueroa

Agradecimientos

Agradecemos en primer lugar a Dios por brindarnos la fortaleza, la constancia y la oportunidad de culminar esta etapa de mi formación académica.

A nuestras familias, por su amor, apoyo y motivación constante, los cuales fueron fundamentales para alcanzar este logro.

A nuestro director de trabajo de grado, el PhD Andrés Ríos, por su acompañamiento, orientación y apoyo durante el desarrollo de este proyecto, así como por su valiosa enseñanza, la cual contribuyó significativamente a nuestro crecimiento académico y profesional.

A la Universidad Industrial de Santander, por brindarnos el respaldo institucional y los recursos necesarios para el desarrollo de esta etapa de nuestra formación académica.

A nuestros profesores, quienes con su dedicación, compromiso y vocación por la enseñanza aportaron de manera significativa a nuestra formación académica.

A nuestros compañeros y amigos, por el acompañamiento, el apoyo y la disposición mostrada a lo largo de este proceso, haciendo más llevadero el camino recorrido.

Finalmente, agradecemos a todas aquellas personas que, de alguna manera, hicieron posible este proceso académico.

Este trabajo de grado de pregrado se enmarca dentro de los productos del Grupo de Investigación en Procesos Estocásticos de la Universidad Nacional de Colombia sede Bogotá, registrado en el GrupLAC de MinCiencias. Corresponde a un producto dirigido por el profesor Andrés Ríos, en el contexto de su Beca de Excelencia Doctoral del Bicentenario Corte I, perteneciente a la convocatoria del Fondo de Ciencia, Tecnología e Innovación del Sistema General de Regalías para la conformación de priorizados y aprobados por el OCAD.

Tabla de contenido

Lista de figuras

Lista de tablas

Lista de apéndices

Resumen

Título: Un enfoque teórico mediante simulaciones para la enseñanza de la prueba de hipótesis*

Autores: Jhan Carlos Jaimes Parra (1) y Leidy Lizeth Figueroa Muñoz (2) **

Palabras clave: Pruebas de hipótesis, simulaciones computacionales, enseñanza de la estadística, valor p.

Descripción:

La enseñanza de las pruebas de hipótesis en el nivel universitario presenta dificultades asociadas a la comprensión conceptual de elementos fundamentales como el valor p, el nivel de significancia y la toma de decisiones bajo incertidumbre, lo que conduce a un aprendizaje de carácter procedimental. En este contexto, el presente estudio tiene como objetivo desarrollar un enfoque didáctico que integre fundamentos teóricos y simulaciones computacionales para la enseñanza de las pruebas de hipótesis bajo el supuesto de normalidad.

La metodología adoptada corresponde a un enfoque cualitativo con componentes cuantitativos, en el que se diseñaron e implementaron simulaciones computacionales orientadas a representar de manera dinámica los conceptos teóricos, aplicadas a estudiantes universitarios. Asimismo, se evaluó el impacto de esta estrategia en la comprensión conceptual de los estudiantes.

Los resultados evidencian que el uso de simulaciones favorece la comprensión del significado del valor p, la interpretación de resultados y el desarrollo del razonamiento estadístico, en comparación con enfoques tradicionales centrados en procedimientos.

Se concluye que la incorporación de simulaciones computacionales constituye una estrategia didáctica efectiva para promover un aprendizaje significativo de las pruebas de hipótesis, contribuyendo al fortalecimiento del pensamiento estadístico en estudiantes universitarios.

*Trabajo de grado

**Facultad de Ciencias. Escuela de matemáticas. Director: Andrés Sebastián Ríos Gutiérrez, PhD en Ciencias Estadística

Abstract

Title: A Theoretical Approach to Teaching Hypothesis Testing Through Simulations*

Authors: Jhan Carlos Jaimes Parra (1) and Leidy Lizeth Figueroa Muñoz (2)**

Keywords: Hypothesis testing, computational simulations, statistics education, p-value.

Description:

The teaching of hypothesis testing at the university level presents difficulties associated with the conceptual understanding of fundamental elements such as the p-value, the significance level, and decision-making under uncertainty, which leads to a predominantly procedural type of learning. In this context, the present study aims to develop a didactic approach that integrates theoretical foundations and computational simulations for teaching hypothesis testing under the assumption of normality.

The methodology adopted corresponds to a qualitative approach with quantitative components, in which computational simulations were designed and implemented to dynamically represent theoretical concepts, and applied to university students. Likewise, the impact of this strategy on students' conceptual understanding was evaluated.

The results show that the use of simulations enhances the understanding of the meaning of the p-value, the interpretation of results, and the development of statistical reasoning, compared to traditional approaches focused on procedures.

It is concluded that the incorporation of computational simulations constitutes an effective didactic strategy to promote meaningful learning of hypothesis testing, contributing to the strengthening of statistical thinking in university students.

*Degree Work

**Faculty of Sciences. School of Mathematics. Advisor: Andrés Sebastián Ríos Gutiérrez, PhD in Statistical Sciences

Introducción

En los procesos de toma de decisiones fundamentados en datos, uno de los principales desafíos consiste en inferir con precisión las características de una población a partir de una muestra. En este contexto, la enseñanza de las pruebas de hipótesis se configura como uno de los contenidos más complejos dentro de la educación estadística, ya que no solo requiere la aplicación de procedimientos formales, sino también la comprensión de conceptos abstractos vinculados a la incertidumbre, la variabilidad y la inferencia. Para abordar estas dificultades, la inferencia estadística ha desarrollado herramientas como los intervalos de confianza y las pruebas de hipótesis. De acuerdo con Casella y Berger (2002), estas últimas representan un recurso fundamental en la toma de decisiones, puesto que permiten evaluar la evidencia muestral frente a una hipótesis nula bajo un control explícito de los posibles errores de decisión.

Por esta razón, los métodos de inferencia estadística, como los intervalos de confianza y las pruebas de hipótesis, han sido desarrollados con el propósito de posibilitar conclusiones más rigurosas acerca de la media de una población. La comprensión de estos conceptos resulta fundamental para que los estudiantes puedan interpretar información, formular conclusiones y tomar decisiones fundamentadas tanto en su vida personal como profesional (Reis et al., 2020).

Aunque las pruebas de hipótesis abarcan múltiples parámetros y contextos inferenciales, el presente documento se delimita específicamente al estudio de pruebas de hipótesis sobre la media bajo supuestos de normalidad. Esta delimitación responde tanto al alcance de la intervención didáctica como a la necesidad de abordar conceptos fundamentales de inferencia estadística, tales como el valor p, el nivel de significancia, los errores tipo I y tipo II, y la toma de decisiones bajo incertidumbre apoyados en simulaciones.

En una sociedad caracterizada por el acceso constante a la información, resulta fundamental que los ciudadanos desarrollen habilidades como la toma de decisiones, la resolución de problemas y la capacidad de realizar inferencias, competencias estrechamente vinculadas con el conocimiento estadístico (Reis et al., 2020). Asimismo, la educación estadística promueve el desarrollo del pensamiento crítico, la argumentación y la interpretación de la información, habilidades clave para la formación académica y profesional de los estudiantes.

En consecuencia, la inferencia estadística constituye uno de los contenidos más abordados en la educación superior y, al mismo tiempo, uno de los menos comprendidos por los estudiantes universitarios (Bernabeu, 2013). Una posible explicación de esta dificultad radica en que la enseñanza tradicional de estos temas suele privilegiar el uso de fórmulas y procedimientos matemáticos formales, lo que limita la comprensión conceptual y dificulta la interpretación de los resultados en contextos reales. En concordancia con ello, (Chick & Pierce, 2008) señalan que la enseñanza de la estadística tiende a centrarse en aspectos procedimentales, priorizando la correcta elaboración de gráficos y la aplicación de reglas de cálculo, mientras que la comprensión del origen de los datos y de sus implicaciones recibe menor atención (p. 1). Además, las autoras destacan que en muchos planes de clase faltan estrategias para articular habilidades y conceptos con una interpretación significativa de la información estadística (p. 5). Como resultado, los estudiantes suelen percibir los métodos estadísticos como un conjunto de técnicas aisladas, sin comprender plenamente su sentido ni su utilidad dentro del análisis de datos.

En este sentido, se han identificado dificultades persistentes en los estudiantes para comprender ideas fundamentales de la inferencia estadística, tales como el significado del valor-p y la relación entre la hipótesis nula y la hipótesis alternativa (Pfannkuch, 2011) . Esta problemática ha sido ampliamente documentada en la literatura especializada, donde se reconoce como uno de los principales desafíos en la enseñanza de la estadística en la educación superior (J. Garfield & Ben-Zvi, 2008b).En particular, conceptos como la prueba de hipótesis y el valor p suelen representar obstáculos conceptuales significativos, debido a que en torno a este último persisten diversos sesgos y concepciones erróneas que han afectado su comprensión y su correcta interpretación en la investigación científica.

Diversos estudios evidencian que muchos estudiantes no logran comprender el sentido de estos procedimientos y, en su lugar, se limitan a memorizar algoritmos de cálculo, sin alcanzar una comprensión conceptual sólida (J. B. Garfield, 2003). Frente a esta situación, la investigación en educación estadística ha destacado la importancia de implementar enfoques didácticos que favorezcan la comprensión conceptual por encima de la simple aplicación mecánica de fórmulas. En este marco, el uso de simulaciones y métodos computacionales brinda a los estudiantes la posibilidad de explorar el comportamiento de los datos, comprender la variabilidad muestral y construir el significado de los conceptos inferenciales a partir de la experiencia y la experimentación (Rossman & Chance, 2014).

En este contexto, el uso de técnicas de simulación ha emergido como una alternativa didáctica que permite introducir los conceptos fundamentales de la inferencia estadística de manera más intuitiva. A través de la simulación, es posible construir distribuciones muestrales, estimar probabilidades y comprender el significado de nociones como la variabilidad, la significación estadística y el valor p, sin depender inicialmente de modelos teóricos complejos. Este enfoque favorece que los estudiantes comprendan el proceso estadístico en su totalidad desde etapas tempranas, lo que facilita una comprensión más profunda de los fundamentos de la inferencia estadística (Montoro, 2019). En particular, la literatura destaca que las simulaciones contribuyen a superar las dificultades asociadas a la enseñanza tradicional, centrada en el cálculo algebraico, al proporcionar experiencias que permiten construir intuiciones sobre el azar y la variabilidad (J. Garfield & Ben-Zvi, 2005; Lane, 2014).

En línea con lo anterior, el uso de simulaciones como herramienta pedagógica ha adquirido relevancia en diversos campos, evidenciando su impacto positivo tanto en el aprendizaje como en la investigación. Por ejemplo, en el ámbito financiero, métodos como la simulación de Monte Carlo permiten modelar situaciones reales de incertidumbre y analizar el comportamiento de variables económicas, facilitando la comprensión de fenómenos complejos mediante la reproducción del comportamiento de sistemas reales (Hurtado et al., 2022)

En consecuencia, las simulaciones se han consolidado como una herramienta educativa que permite a los profesionales aplicar de manera efectiva sus conocimientos y enfrentarse a situaciones complejas propias de su campo. En este marco, han surgido enfoques pedagógicos orientados a articular la teoría con la práctica mediante estrategias activas de aprendizaje. Entre estas, las simulaciones destacan como un recurso didáctico que posibilita a los estudiantes enfrentarse a escenarios complejos, experimentar con modelos y construir significado a partir de la interacción con situaciones cercanas a la realidad (Koparan, 2022).

Teniendo en cuenta lo anterior, y considerando que las simulaciones favorecen la interpretación al promover la inferencia estadística mediante la generación y el análisis de múltiples escenarios controlados (Chance et al., 2007), estas se constituyen en una herramienta poderosa para observar el comportamiento de los métodos inferenciales y comprender sus implicaciones en la práctica estadística, incluso antes de abordar el análisis con datos reales. En este contexto, se plantea la siguiente pregunta:

¿Cómo las simulaciones contribuyen al fortalecimiento del razonamiento inferencial así, como la aplicación y comprensión de las pruebas de hipótesis en el análisis estadístico?

En este contexto, el presente documento tiene como propósito desarrollar un enfoque didáctico que integre fundamentos teóricos y una simulación computacional para la enseñanza de las pruebas de hipótesis bajo el supuesto de normalidad, con aplicaciones a situaciones reales. La propuesta busca promover que los estudiantes experimenten, exploren y reflexionen sobre conceptos como las pruebas de hipótesis y el valor p, en lugar de limitarse a la memorización de procedimientos matemáticos.

En cuanto a su organización, el documento se estructura de la siguiente manera: en primer lugar, se presentan los objetivos de la investigación. A continuación, en el Capítulo 3, se desarrolla la revisión bibliográfica, con el fin de fundamentar y justificar el estudio, estableciendo la relación entre las pruebas de hipótesis, las simulaciones en RStudio y su relevancia en la didáctica de la estadística.

Seguidamente, en el Capítulo 4, se expone el marco teórico junto con las bases conceptuales que sustentan la investigación. Posteriormente, en el Capítulo 5, se describe la metodología, enmarcada en un enfoque cualitativo con componentes cuantitativos propios de la estadística inferencial. Para la recolección de la información se emplearon dos instrumentos, denominados pretest y postest; además, se llevó a cabo una intervención didáctica orientada a fortalecer la comprensión de los conceptos relacionados con las pruebas de hipótesis mediante simulaciones en RStudio.

En el Capítulo 6, se presenta el análisis de los resultados, el cual considera la intervención de aula a partir del modelo RyA propuesto por Parada (2016), las producciones de los estudiantes y el análisis cuantitativo de los datos mediante la prueba de Wilcoxon. Finalmente, en el Capítulo 7 se exponen las conclusiones y, en el Capítulo 8, se plantean algunas recomendaciones derivadas del estudio.

Objetivos

Objetivo general

Desarrollar un enfoque didáctico que integre fundamentos teóricos y una simulación computacional para la enseñanza de las pruebas de hipótesis bajo el supuesto de normalidad, con aplicaciones a situaciones reales.

Objetivos específicos

- Diseñar y ejecutar simulaciones computacionales que ilustren de forma pedagógica las definiciones teóricas, facilitando su comprensión y su aplicación en contextos educativos.
- Implementar el uso de simulaciones computacionales como estrategia didáctica en un contexto educativo universitario para la enseñanza de las pruebas de hipótesis.

1. Revisión bibliográfica

Las pruebas de hipótesis constituyen uno de los parámetros fundamentales de la estadística inferencial, permitiendo contrastar parámetros poblacionales a través de información muestral, mediante la formulación de una hipótesis nula y una alternativa (Walpole et al., 2012).

En la práctica, las pruebas de hipótesis se utilizan como herramientas para tomar decisiones bajo incertidumbre más que como procedimientos destinados a establecer la verdad o falsedad de afirmaciones. Desde un enfoque teórico, rechazar H_0 no implica aceptar H_1, ni no rechazar H_0 supone creer que la hipótesis nula sea verdadera, sino únicamente adoptar una acción basada en la evidencia disponible (Casella & Berger, 2002). Desde esta perspectiva el objetivo de las pruebas de hipótesis es evaluar, a partir de los datos muéstrales, la compatibilidad entre un modelo probabilístico que representa la hipótesis nula y los resultados observados, con el fin de tomar decisiones.

A pesar de su relevancia, diversas investigaciones han evidenciado que los estudiantes universitarios tienden a utilizar las pruebas de hipótesis de manera mecánica, sin un análisis conceptual profundo ni verificación de supuestos (Rodríguez et al., 2009). Esto conduce a una aplicación rutinaria centrada en la obtención de valores p inferiores a un umbral arbitrario (generalmente 0.05), descuidando aspectos fundamentales como la magnitud del efecto, la variabilidad de los datos y el contexto de los resultados (Wasserstein & Lazar, 2016)

Desde el ámbito de la investigación en educación estadística, se ha evidenciado dificultades persistentes en la interpretación conceptual de las pruebas de hipótesis por parte de los estudiantes. En particular, la interpretación errónea del valor p como la probabilidad de que la hipótesis nula sea verdadera (Castro Sotos et al., 2007)), así como la confusión entre significación estadística y relevancia práctica. Asimismo, diversos estudios señalan la tendencia de los estudiantes a concebir el contraste de hipótesis como un algoritmo de pasos fijos, desvinculado del razonamiento probabilístico (Batanero, 2001a). Estas dificultades no solo afectan el aprendizaje en cursos de estadística, sino que también influyen en las prácticas de análisis poco reflexivas.

En particular, el estudio de Inzunsa Cazares y Jiménez Ramírez (2013) pone de manifiesto que los estudiantes universitarios tienden a abordar las pruebas de hipótesis desde una perspectiva fundamentalmente algorítmica, caracterizada por la aplicación mecánica de reglas y fórmulas, sin una comprensión profunda de los conceptos subyacentes. Este enfoque se traduce en errores frecuentes, tales como la confusión entre la hipótesis nula H_0 y la alternativa H_1, la interpretación incorrecta del valor p y la incomprensión del nivel de significancia como criterio de decisión bajo incertidumbre. Más allá de errores puntuales, estos hallazgos revelan una problemática estructural en la enseñanza, la cual es la desconexión entre los procedimientos estadísticos y su significado inferencial.

De manera complementaria, Estudios como el de delMas et al. (2007) evidencian que, incluso después de un curso introductorio de estadística, los estudiantes presentan dificultades significativas en la comprensión de conceptos de inferencia, especialmente en la interpretación del valor p y la lógica de las pruebas de hipótesis. Estos hallazgos sugieren que los enfoques tradicionales de enseñanza no son suficientes para promover una comprensión profunda, lo que justifica la incorporación de metodologías activas, como el uso de simulaciones, para favorecer el razonamiento estadístico.

Desde una perspectiva didáctica, esta situación puede interpretarse como una consecuencia de modelos de enseñanza tradicionales centrados en la transmisión de técnicas, en los cuales el énfasis recae en la resolución de ejercicios estandarizados más que en la construcción de significado. Esta interpretación es consistente con lo reportado por Inzunsa Cazares y Jiménez Ramírez (2013), quienes evidencian un predominio del razonamiento procedimental en estudiantes universitarios.

En respuesta a estas dificultades, diversas investigaciones en educación matemática y educación estadística han destacado el potencial de las tecnologías digitales como herramientas mediadoras en la construcción de conceptos estadísticos y probabilísticos. El uso de software educativo, simulaciones interactivas y herramientas computacionales permite representar fenómenos aleatorios y procesos inferenciales de manera dinámica, favoreciendo la exploración, la visualización y la interpretación de datos en distintos contextos (Batanero, 2001a; J. Garfield & Ben-Zvi, 2008a). Desde esta perspectiva, la tecnología trasciende su función instrumental y se convierte en un recurso didáctico que posibilita la comprensión de conceptos abstractos asociados a la variabilidad, el muestreo y la incertidumbre. Asimismo, diversos estudios señalan que los entornos tecnológicos promueven procesos de aprendizaje más activos y exploratorios, en los cuales los estudiantes pueden formular conjeturas, contrastar resultados, manipular variables y establecer relaciones entre representaciones gráficas, numéricas y simbólicas de los datos, fortaleciendo así el razonamiento estadístico y la comprensión conceptual (Chance et al., 2007).

En relación con lo anterior, investigaciones recientes sobre el uso de herramientas tecnológicas en la enseñanza de las matemáticas evidencian que la incorporación de plataformas digitales, simuladores interactivos y software educativo favorece la comprensión conceptual y el desarrollo del pensamiento lógico y matemático. En particular, el estudio de Cáceres-Mesa et al. (2025) señala que herramientas como GeoGebra, simuladores interactivos y plataformas virtuales promueven un aprendizaje más dinámico, visual e individualizado, permitiendo a los estudiantes explorar conceptos abstractos mediante representaciones gráficas y experimentación en tiempo real. Asimismo, los autores destacan que estas tecnologías fortalecen la motivación, la resolución de problemas y el aprendizaje autónomo, superando parcialmente las limitaciones de los enfoques tradicionales centrados en la memorización y repetición de procedimientos.

En esta misma línea, investigaciones sobre el uso de simulaciones en educación matemática destacan su potencial para fortalecer la comprensión conceptual y promover aprendizajes significativos. El estudio de Díaz Pinzón (2018), realizado con estudiantes de educación secundaria mediante el uso de simulaciones (Physics Education Technology) PhET, evidenció mejoras significativas en el rendimiento académico y en la motivación hacia el aprendizaje de las matemáticas. El autor señala que las simulaciones permiten representar procesos dinámicos, visualizar conceptos abstractos y favorecer la exploración interactiva, facilitando que los estudiantes construyan conocimiento a partir de la experimentación y el descubrimiento. Asimismo, resalta que estas herramientas promueven un aprendizaje activo y participativo, en el cual los estudiantes pueden formular hipótesis, manipular variables y analizar resultados en tiempo real.

En contraste, diversos autores proponen que la enseñanza de la estadística debe orientarse hacia el desarrollo del razonamiento estadístico y la alfabetización estadística. En este sentido, Rumsey (2002), plantea que la enseñanza de la estadística debería trascender el énfasis en el "qué" y el "cómo" de los procedimientos, para incorporar el "por qué" de los métodos estadísticos, promoviendo la capacidad de interpretar, evaluar y comunicar información basada en datos. En este sentido, la comprensión de conceptos fundamentales como el valor p, el nivel de significancia o los errores tipo I y II no puede limitarse a su definición formal, sino que debe desarrollarse a partir de contextos significativos que permitan a los estudiantes explicar, cuestionar y dar sentido a los resultados obtenidos.

Asimismo, en la educación estadística la interpretación de datos constituye uno de los pilares fundamentales. Esta habilidad es esencial para la toma de decisiones informadas, ya que permite evaluar afirmaciones a partir de información proveniente de diversos contextos (Dagnino et al., 2014). A través del estudio de la estadística, se busca proporcionar herramientas que permitan a los estudiantes extraer, analizar y comprender información cuantitativa del entorno que los rodea (J. Garfield & Ben-Zvi, 2008b). Teniendo en cuenta lo anterior, la prueba de hipótesis es un procedimiento estadístico que permite evaluar si la evidencia proporcionada por la muestra respalda o contradice la afirmación inicial sobre la población, conocida como hipótesis nula. La conclusión de una prueba de hipótesis es no rechazar o rechazar la hipótesis nula (Sanchez et al., 2024).

Por otro lado, según el Ministerio de Educación Nacional (1998) constituye un eje esencial para el desarrollo del pensamiento matemático al permitir la compresión de fenómenos caóticos regidos por el azar y la incertidumbre.

> "La probabilidad y la estadística son ramas de las matemáticas que desarrollan procedimientos para cuantificar, proponen leyes para controlar y elaboran modelos para explicar situaciones que por presentar múltiples variables y de efectos impredecibles son consideradas como regidas por el azar, y por tanto denominadas aleatorias. El carácter globalizante de la probabilidad y la estadística está en la presencia del pensamiento aleatorio para la comprensión de fenómenos de la vida cotidiana y de las ciencias."(de Educación Nacional, 1998, p. 98)

A pesar de la relevancia teórica, el pensamiento aleatorio recibe una atención limitada en el contexto educativo colombiano. Diversos estudios realizados señalan que "la enseñanza de la estadística no tiene mayor importancia en las prácticas de aula" (Flórez et al., 2020) y que la intensidad horaria destinada a este contenido es mínima en muchos colegios oficiales (Fernández & Sarmiento, 2020). De igual manera , Batanero, Gea y Álvarez-Arroyo (2023) destacan que, aunque la probabilidad se ha incorporado progresivamente desde los primeros niveles educativos, persisten dificultades conceptuales y errores sistemáticos en su aplicación a contextos cotidianos y en la toma de decisiones profesionales (Batanero et al., 2023), por lo tanto, se hace necesario replantear el enfoque pedagógico e incorporar estrategias didácticas que puedan promover un aprendizaje significativo.

De igual manera la UNESCO (2012) reconoce la importancia del pensamiento aleatorio y probabilístico dentro de la educación matemática, señalando que los estudiantes deben de familiarizarse con los diferentes modos de razonamiento estadístico y probabilístico los cuales le permiten comprender fenómenos caracterizados por la incertidumbre y el riesgo.

> "Es importante que el estudiante se familiarice gradualmente con los modos de razonamiento probabilístico y estadístico, necesarios para el pensamiento matemático, a fin de comprender los fenómenos que, tanto en la ciencia como en la vida social, implican incertidumbre y riesgo". (UNESCO, 2012)

Según Batanero (2001a), la educación estadística proviene de diversas áreas de conocimiento, adicional a la educación matemática. Comprender estos conceptos implica además de aplicar fórmulas, desarrollar una intuición que permita valorar la incertidumbre inherente a toda generalización basada en datos.

Sin embargo, La comprensión de las pruebas de hipótesis presenta numerosos retos para los estudiantes, incluso en niveles avanzados. según Suárez (2008), muchos estudiantes universitarios muestran una concepción incorrecta del intervalo, interpretándolo como un rango dentro del cual se encuentra la media muestral, o bien como una probabilidad de que el parámetro poblacional se encuentre en un intervalo específico en una muestra dada, lo que refleja una interpretación subjetiva o bayesiana inapropiada del nivel de confianza. Asimismo, se observa una falta de comprensión del papel que juegan el tamaño de la muestra, la variabilidad y el nivel de confianza en la construcción del intervalo. Estas dificultades no solo revelan una comprensión superficial del concepto, sino también una dependencia excesiva en procedimientos mecánicos, lo que limita la capacidad de los estudiantes para aplicar la inferencia estadística de forma crítica y contextualizada.

Como afirman J. Garfield y Ben-Zvi (2008b), la estadística es una disciplina emergente que ha evolucionado a partir de diversas ramas del conocimiento, y que en sus inicios estuvo orientada principalmente a la formación de especialistas. No obstante, este enfoque ha cambiado con el tiempo, reconociéndose hoy la necesidad de incorporar la enseñanza de la estadística en todos los niveles educativos, dada la creciente relevancia del análisis de datos en la vida cotidiana, la investigación científica y la toma de decisiones fundamentadas.

El desarrollo de las pruebas de hipótesis es el resultado de un proceso histórico que se remonta a los trabajos de Laplace (1814) y Gauss (1809) quienes sentaron las bases del estudio de la probabilidad y la distribución de los errores. En esta línea, Stigler (1999) en su libro The History of Statistics habla sobre el desarrollo de la estadística pero mas importante el teorema del limite central.

No obstante, un avance significativo se produjo con el trabajo de Gosset (1908) quien, bajo el seudónimo de "Student", desarrolló la prueba t de Student, diseñada para corregir las limitaciones de la distribución normal en muestras de tamaño reducido. posteriormente, (Fisher, 1935) quien formalizó el marco moderno de las pruebas de hipótesis en su obra titulada Statistical Methods for Research Workers (1925), introduciendo conceptos fundamentales como la hipótesis nula, el valor p y principios del diseño experimental, tales como la replicación y la aleatorización, orientados al control del error.

Años después Neyman y Pearson (1933) propusieron una nueva formalización, incorporando la hipótesis alternativa y el error tipo I y tipo II, además el concepto de potencia, estos enfoques son los que se utilizan en la actualidad.

Asimismo, en el desarrollo de la estadística, las simulaciones han jugado un papel importante como herramienta para comprender fenómenos complejos y evaluar la robustez de los métodos estadísticos (Hallgren, 2013). En el ámbito educativo, estas permiten representar de manera dinámica la variabilidad de los datos y facilitar la comprensión de conceptos abstractos. En este sentido, el uso de simulaciones constituye una estrategia didáctica pertinente para superar las limitaciones del enfoque tradicional, ya que favorece la comprensión conceptual y el desarrollo del razonamiento estadístico.

En síntesis, la revisión de la literatura evidencia que, aunque las pruebas de hipótesis constituyen un elemento central de la estadística inferencial (Casella & Berger, 2002; Walpole et al., 2012), su enseñanza en el ámbito universitario continúa enfrentando importantes dificultades. Estas problemáticas se asocian, principalmente, a enfoques tradicionales centrados en la aplicación mecánica de procedimientos, lo que limita el desarrollo de una comprensión conceptual profunda y del razonamiento estadístico en los estudiantes (Batanero, 2001a).

Asimismo, los estudios analizados muestran que los estudiantes presentan errores persistentes en la interpretación de conceptos clave, como el valor p, el nivel de significancia y la toma de decisiones bajo incertidumbre, lo que repercute en prácticas de análisis poco reflexivas y descontextualizadas. En particular, se ha documentado que muchos estudiantes interpretan el valor p como la probabilidad de que la hipótesis nula sea verdadera o como una medida directa de la importancia del resultado, lo que evidencia una comprensión inadecuada de la lógica inferencial (Castro Sotos et al., 2007; J. Garfield & Ben-Zvi, 2008a)

En este contexto, la literatura coincide en la necesidad de transformar las prácticas de enseñanza hacia enfoques que promuevan la comprensión conceptual, la interpretación de resultados y el desarrollo del pensamiento estadístico. En particular, se destaca el potencial de estrategias didácticas activas, como el uso de simulaciones, para favorecer la visualización de la variabilidad, la comprensión de la lógica inferencial y la toma de decisiones fundamentadas en datos.

Por tanto, se hace necesario replantear la enseñanza de las pruebas de hipótesis, incorporando metodologías que permitan a los estudiantes no solo aplicar procedimientos, sino también comprender su significado, interpretar sus resultados y utilizarlos en contextos reales. En coherencia con ello, el presente estudio propone el uso de simulaciones interactivas como una alternativa pedagógica orientada a fortalecer el aprendizaje significativo de la inferencia estadística.

2. Marco conceptual

La inferencia estadística constituye uno de los pilares fundamentales en la formación matemática y científica de los estudiantes, ya que permite establecer conclusiones sobre una población a partir de información parcial obtenida mediante una muestra. En este proceso intervienen conceptos esenciales como la noción de muestra aleatoria simple, la formulación de hipótesis sobre parámetros poblacionales y la aplicación de procedimientos formales orientados a evaluar la evidencia empírica proporcionada por los datos. La comprensión adecuada de estas nociones resulta indispensable para el desarrollo del pensamiento crítico y la toma de decisiones informadas, en la medida en que ofrecen un marco riguroso para analizar la incertidumbre inherente a los fenómenos reales.

En este contexto, se hace necesario introducir, desde una perspectiva formal, conceptos fundamentales como la región de rechazo, los tipos de error, el nivel de significancia, la potencia de una prueba, así como el uso de estadísticos de prueba y valores p, los cuales estructuran la metodología de las pruebas de hipótesis. Desde un enfoque didáctico, estas definiciones no solo permiten formalizar el procedimiento inferencial, sino también comprender sus alcances y limitaciones. Su presentación de manera clara y progresiva favorece que los estudiantes construyan una visión articulada del proceso de inferencia, entendiendo cómo se evalúan conjeturas sobre parámetros desconocidos y cómo la evidencia empírica se integra con la teoría para sustentar decisiones basadas en datos.

Definición 2.1. *Una muestra aleatoria simple de tamaño n, extraída de una población de tamaño N, se define como un subconjunto de elementos seleccionados de manera que todos los subconjuntos posibles de tamaño n tengan la misma probabilidad de ser elegidos. (Casella & Berger, 2002)*

$$P(\textit{Cualquier subconjunto específico de tamaño } n) = \frac{1}{\binom{N}{n}}$$

Cada elemento seleccionado se denota por $x_1, x_2, \ldots, x_n$, cuyos valores están en el rango de las variables aleatorias $X_i, i = 1, \cdots, n$ que corresponden a al valor de la población X para el i-ésimo elemento escogido. Esto quiere decir que:
$x_1, x_2, \ldots, x_n$ denotan los respectivos datos de las variables aleatorias:

$$X_1 = \textit{Valor de } X \textit{ para el primer elemento}, \ldots, X_n = \textit{Valor de } X \textit{ para el elemento } n,$$

Del concepto de muestra aleatoria simple se asume que:

$$X_1, X_2, \ldots, X_n \text{ son independientes}$$

Además de que son idénticamente distribuídas, es decir,

$$X_1, X_2, \ldots, X_n \sim \text{Distribución}(\Theta).$$

Donde Θ es el conjunto de parámetros dada la distribución. En el caso de la distribución normal $\Theta = \{\mu, \sigma^2\}$, que corresponden respectivamente a la media y a la desviación estándar.

La muestra aleatoria simple se entiende como un procedimiento de selección en el que cada elemento de la población tiene la misma probabilidad de ser elegido. Este método busca garantizar que la elección de los participantes no esté influenciada por criterios subjetivos o decisiones del investigador que puedan introducir sesgos. Al asegurar igualdad de oportunidades en el proceso de selección, se promueve una representación más equilibrada de la población.

De esta manera, la muestra aleatoria simple se convierte en un recurso esencial para realizar análisis estadísticos confiables, ya que permite obtener conclusiones válidas sin necesidad de estudiar a todos los individuos que conforman el grupo de interés.

Definición 2.2. *(Casella & Berger, 2002)[p 373]. Una hipótesis es una aseveración o conjetura acerca de un parámetro poblacional.*

Las dos hipótesis en una prueba de hipótesis se conocen como la *nula* y la *alternativa*. Se denotan por H_0 y H_1 (o H_a), respectivamente. Si θ denota un parámetro poblacional sobre el cual se desea plantear una hipótesis, se tiene las hipótesis están dadas por

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c \tag{2.1}$$

tal que Θ_0 es un subconjunto del espacio de parámetros tal que $\theta \in \Theta$.

Definición 2.3. *El subconjunto del espacio muestral para el cual H_0 será rechazada se llama* ***región de rechazo o región crítica****. El complemento de la región de rechazo se llama región de confianza.*

En la hipótesis 2.1 se tiene que la región de rechazo es Θ_0^c y la región de confianza es Θ_0.

Según, Casella y Berger, 2002, un procedimiento de prueba de hipótesis o una prueba de hipótesis es una regla que especifica:

1. Para cuales valores muéstrales, la decisión no se rechaza H_0.
2. Para cuales valores muéstrales H_0 es rechazada y H_1 no es rechazada.

Definición 2.4. *Región de rechazo (o región crítica). Es el conjunto de valores del estadístico de prueba que, según los valores críticos determinados por el nivel de significancia α, se consideran lo suficientemente extremos como para ser incompatibles con la hipótesis nula H_0.*

Cuando el estadístico obtenido cae dentro de esta zona, se concluye que la evidencia es estadísticamente significativa y se procede a rechazar H_0 (Pearson, 2025).

Definición 2.5. *Región de no rechazo (o región de aceptación) comprende los valores del estadístico de prueba que no son lo suficientemente extremos como para justificar el rechazo de la hipótesis nula H_0. Estos valores corresponden a resultados que se consideran compatibles con el comportamiento esperado bajo H_0. Si el estadístico cae en esta zona, se concluye que no existe evidencia estadística suficiente para rechazar la hipótesis nula, y esta región cubre la probabilidad complementaria $1 - \alpha$ (Pearson, 2025).*

En términos probabilísticos, $1 - \alpha$ representa la porción de la distribución donde se espera que se encuentren la mayoría de los valores si H_0 es verdadera. Dado que la región de rechazo ocupa únicamente una fracción α de la distribución ,la parte más extrema y menos probable bajo H_0, la región restante corresponde a la zona donde los resultados se consideran habituales o no sorprendentes para la hipótesis nula. Por ello, cuando el estadístico calculado se sitúa dentro de este intervalo, se interpreta que el resultado es coherente con H_0 y, en consecuencia, no se rechaza.

Nota 2.6. *Hay dos estructuras posibles para Θ_0: Si Θ_0 consiste de solo un punto, se dice que la hipótesis nula es simple. Cuando Θ_0 contiene más de un punto, se dice que la hipótesis nula es compuesta.*

Definición 2.7. ***Evaluación de Tests*** *El desempeño de una prueba es medido por la frecuencia con que se hacen juicios correctos al no rechazarla (ver Bickel y Doksum, 1997 [p 166]): Hay dos tipos de error que se pueden cometer cuando se utiliza un prueba.*

1. Que se rechace la hipótesis H_0 cuando en realidad es verdadera, o
2. Que no se rechace la hipótesis H_0 cuando en realidad es falsa.

Al error presentado en el ítem 1 se le llama *error tipo I* y al error presentado en el ítem 2 se le llama *error tipo II* Bickel y Doksum, 2015, como se muestra en la tabla **??**.

Tabla 1: *Tipos de error*

Decisión	H_0 es verdadera	H_1 es verdadera
No rechazar H_0	No hay error	Error tipo II
Rechazar H_0	Error tipo I	No hay error

Cuando se realiza una prueba de hipótesis, la decisión tomada a partir de los datos puede equivocarse de dos maneras.

El error tipo I ocurre cuando se rechaza la hipótesis nula H_0 siendo esta verdadera. Es decir, la prueba concluye que existe un efecto o diferencia cuando en realidad no la hay.

Error tipo II. El error tipo II ocurre cuando no se rechaza la hipótesis nula H_0 a pesar de que esta es falsa. En este caso, la prueba falla al detectar un efecto real que sí está presente.

Definición 2.8. *(Casella & Berger, 2002) [p 383] La* ***función de potencia*** *de una prueba de hipótesis con región de rechazo R se define por:*

$$\beta(\theta) = P_\theta(X \in R).$$

Una función de potencia ideal corresponde a una prueba que toma siempre la decisión correcta: no rechaza la hipótesis nula cuando esta es verdadera y la rechaza de manera sistemática cuando es falsa. Esto implica que en la región nula Θ_0 la probabilidad máxima de rechazar H_0 es cero, evitando por completo el error tipo I, mientras que en la región alternativa Θ_1 la probabilidad máxima de rechazar H_0 es uno, de modo que el test detecta perfectamente cualquier desviación respecto de la hipótesis nula, sin incurrir en error tipo II.

Nota 2.9. *Una función de potencia* ***ideal*** *sería que tome el valor 0 para todo $\theta \in \Theta_0$ y tome el valor de 1 para todo $\theta \in \Theta_1$.*

Por tanto una función de potencia ideal sería aquella que toma decisiones perfectas:

- Nunca rechaza la hipótesis nula cuando es verdadera.

- Siempre la rechaza cuando es falsa.

En términos del supremo, esto se expresa como:

$$\sup_{\theta \in \Theta_0} \beta(\theta) = 0 \quad \text{y} \quad \sup_{\theta \in \Theta_1} \beta(\theta) = 1$$

Esto quiere decir:

- En la región nula, la máxima probabilidad de rechazar H_0 es 0 (sin error tipo I).
- En la región alternativa, la máxima probabilidad de rechazar H_0 es 1 (sin error tipo II).

Este es el ideal teórico, pero en la práctica no se puede lograr porque las distribuciones se traslapan,porque los valores posibles bajo la hipótesis nula y la alternativa no son completamente distintos. Esto significa que algunos resultados son compatibles con ambas hipótesis, lo que hace imposible diseñar una prueba que nunca se equivoque. Por eso, la función de potencia ideal es solo un concepto teórico.

Definición 2.10. *(Casella & Berger, 2002) . Sea* $0 \leq \alpha \leq 1$*. Un test con función de potencia* $\beta(\theta)$ *es un* ***test de tamaño*** α *si*

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

donde:

- $\beta(\theta)$ es la **función de potencia**, que indica la probabilidad de rechazar H_0 cuando el verdadero valor del parámetro es θ.
- Θ_0 es el conjunto de valores del parámetro bajo la hipótesis nula H_0.
- El **tamaño del test** α se define como el peor caso del error tipo I:

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta(\text{rechazar } H_0).$$

- Se usa el **supremo** porque la probabilidad de rechazar H_0 puede variar dentro de Θ_0. El tamaño es el máximo posible.

En otros términos, un test de tamaño α es un procedimiento estadístico construido para controlar el riesgo de rechazar la hipótesis nula H_0 cuando esta es verdadera. El tamaño del test se define como el máximo valor que puede alcanzar la probabilidad de cometer un error tipo I, considerando todos los valores del parámetro que satisfacen la hipótesis nula. Por esta razón, el tamaño se expresa mediante el supremo de dicha probabilidad, ya que la probabilidad de rechazo puede variar dentro del conjunto Θ_0. De este modo, incluso en el escenario más desfavorable para H_0, la probabilidad de rechazarla no excede el nivel α. Este enfoque garantiza que el procedimiento de prueba actúe de manera conservadora, permitiendo rechazar la hipótesis nula únicamente cuando la evidencia en su contra es suficientemente fuerte y sin sobrepasar el nivel máximo de error establecido(Lehmann & Romano, 2005).

Nota 2.11. *(Bickel & Doksum, 1997) La* ***potencia*** *de una prueba es la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es verdadera.*

Formalmente se expresa como:

$$Potencia = P(rechazar\ H_0 \mid H_1\ \ es\ verdadera)$$

Por tanto, la potencia es igual a $1-\beta$, donde β representa la probabilidad de cometer un error de tipo II (no rechazar H_0 cuando en realidad es falsa).

En otras palabras, la potencia es una función de θ sobre Θ^c. Note que la potencia y la probabilidad de cometer error tipo I están contenidas en la función de potencia. En general, se denota a la probabilidad de cometer error tipo II por β, luego la potencia será $1-\beta$. Ver cuadro 1.

Definición 2.12. ***estadístico de prueba*** *según (Casella & Berger, 2002) de una prueba de hipótesis se especifica en términos de un estadístico de prueba, la cual es una función de la muestra la cual se utiliza para tomarla decisión de rechazar o no rechazar la hipótesis nula.*

$$W(X_1, \ldots, X_n) = W(X),$$

Por ejemplo, una prueba podría especificar que H_0 debe rechazarse si $\bar{X}$, la media muestral, es mayor que 3. En este caso,

$$W(X) = \bar{X}$$

es el estadístico de prueba y la región de rechazo es

$$\{(x_1, \ldots, x_n) : \bar{x} > 3\}.$$

En el contraste de hipótesis, el *p-valor* se interpreta como una medida del grado de compatibilidad entre los datos observados y la hipótesis nula H_0. Desde una perspectiva intuitiva, este valor indica qué tan sorprendentes resultan los datos bajo la suposición de que H_0 es verdadera. Un p-valor pequeño sugiere que obtener resultados tan extremos como los observados sería poco probable si H_0 fuera cierta, lo que refleja una baja concordancia entre los datos y dicha hipótesis y constituye evidencia en su contra, favoreciendo a la hipótesis alternativa H_1. Por el contrario, un p-valor grande indica que los resultados observados no son inusuales bajo H_0, por lo que no se dispone de evidencia estadística suficiente para rechazarla. De este modo, el p-valor permite evaluar de forma gradual la evidencia empírica aportada por la muestra en contra de la hipótesis nula.

Definición 2.13. *según (Casella & Berger, 2002)Un* ***valor p*** *$p(X)$, es una estadística de prueba que satisface $0 \le p(x) \le 1$ para cada punto muestral x. Los valores pequeños de $p(X)$ proporcionan evidencia de no rechazar H_1. Un valor p es válido si, para todo $\theta \in \Theta_0$ y todo $0 \le \alpha \le 1$,*

$$P_\theta\big(p(X) \le \alpha\big) \le \alpha.$$

No obstante, esta definición contrasta con lo que dice (Bickel & Doksum, 2015) [p 222]. el *p-valor* o tamaño observado del test es un estadístico que se define como el nivel de insignificancia más pequeño α, al cual un experimentador usando el estadístico $T(\boldsymbol{X})$ rechazaría la hipótesis nula, basándose en los resultados observados $\boldsymbol{x}$, (Bickel & Doksum, 2015)[p.221]. Por tanto, para un α dado por el investigador, si $\alpha < p$-valor la hipótesis nula no se rechaza y si $\alpha > p$-valor la hipótesis nula se rechaza

El *p-valor* es una medida que indica qué tan compatibles son los datos observados con la hipótesis nula H_0. Un p-valor pequeño significa que, si H_0 fuera verdadera, sería poco probable obtener resultados como los observados; esto sugiere que los datos no concuerdan con H_0 y aportan evidencia a favor de la hipótesis alternativa H_1. En cambio, un p-valor grande indica que los resultados observados son razonables bajo la suposición de que H_0 es cierta, por lo que no se cuenta con evidencia suficiente para rechazarla. De esta manera, el p-valor permite evaluar de forma gradual la evidencia en contra de H_0: cuanto menor es su valor, mayor es el apoyo de los datos hacia H_1.

Nota 2.14. *Asimismo (Bickel, 2021), define el* valor p *de la siguiente manera:*

> *La probabilidad, bajo la hipótesis nula H_0, de obtener un resultado tan extremo o más extremo que el observado en la muestra.*

Matemáticamente:

$$p = P(T \geq t_{obs} \mid H_0)$$

donde:

- *T es el estadístico de prueba bajo la distribución asumida por H_0.*
- *t_{obs} es el valor observado del estadístico en la muestra.*

El valor p no representa la probabilidad de que la hipótesis nula sea verdadera. Más bien, indica cuán compatibles son los datos con H_0. Un valor p pequeño (por ejemplo, menor que $\alpha = 0{,}05$) sugiere que los datos son poco probables bajo H_0, lo que proporciona evidencia para rechazarla. Por el contrario, un valor p grande indica que los datos son consistentes con H_0, por lo que no se rechaza.

Regla de decisión:

$$\text{Si } p \leq \alpha, \text{ se rechaza } H_0; \quad \text{si } p > \alpha, \text{ no se rechaza } H_0.$$

El software *RStudio* es un entorno de desarrollo integrado diseñado para apoyar el trabajo con el lenguaje de programación *R*, proporcionando en una misma interfaz herramientas para escribir y ejecutar código, generar gráficos, depurar funciones y organizar proyectos orientados al análisis estadístico y la ciencia de datos. Su desarrollo estuvo inicialmente a cargo de *RStudio Inc.*, entidad que posteriormente, adoptó el nombre *Posit Software PBC*, y desde su primera versión pública en 2011 se ha consolidado como uno de los entornos más utilizados por investigadores y profesionales, debido a su capacidad para estructurar flujos de trabajo reproducibles y facilitar la integración de diversos lenguajes y tecnologías empleadas en el análisis moderno, como *Python*, *Quarto* y *R Markdown* (Posit Software, PBC, 2025).

A continuación, se presentan ejemplos de pruebas de hipótesis para la media, considerando tanto el caso de varianza poblacional conocida como el de varianza desconocida, con el fin de ilustrar su implementación y análisis mediante simulaciones computacionales en *RStudio*.

2.1. Pruebas de hipótesis para la media en una distribución normal con varianza conocida

Modelo probabilístico

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple tal que:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n,$$

donde $\mu \in \mathbb{R}$ es desconocido y $\sigma^2 > 0$ es conocida. Se asume además que las variables $X_1, \ldots, X_n$ son independientes e idénticamente distribuidas.

En este modelo probabilístico se asume que las observaciones $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria simple proveniente de una población que sigue una distribución normal con media μ y varianza conocida σ^2. La media poblacional μ es desconocida y representa el parámetro de interés sobre el cual se realizarán las inferencias estadísticas. Asimismo, se supone que las variables son independientes e idénticamente distribuidas, lo cual garantiza que cada observación aporta información equivalente y permite derivar de forma exacta la distribución de la media muestral. Bajo estos supuestos, es posible construir pruebas de hipótesis para la media poblacional basadas en la distribución normal estándar, controlando el error tipo I y analizando el comportamiento del test desde un enfoque teórico riguroso (Casella & Berger, 2002).

Distribución de la media muestral

La media muestral se define como el promedio de las observaciones que conforman la muestra y se expresa como

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Este estadístico resume la información contenida en los datos muestrales y constituye el estimador natural de la media poblacional μ, ya que representa el valor central de la muestra y es insesgado bajo supuestos de muestreo aleatorio (Casella & Berger, 2002)

La media muestral sigue una distribución normal, la cual se expresa como

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Este resultado indica que, cuando las observaciones provienen de una población con distribución normal, el promedio de los datos también se distribuye normalmente, con la misma media poblacional μ y una varianza que disminuye en función del tamaño de la muestra. En particular, al aumentar el número de observaciones, la media muestral presenta menor variabilidad y se aproxima con mayor precisión al valor real de la media poblacional. Esta propiedad es fundamental en inferencia estadística, ya que permite utilizar la media muestral como base para la construcción de pruebas de hipótesis sobre la media poblacional (Mood et al., 1974).

Para justificar formalmente la distribución de la media muestral, se utiliza la función generadora de momentos, la cual permite caracterizar de manera única la distribución de una variable aleatoria (Casella & Berger, 2002). En este caso, dado que cada observación X_i sigue una distribución normal con media μ y varianza σ^2, su función generadora de momentos está dada por

$$M_{X_i}(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

A partir de esta expresión y considerando que las variables son independientes, es posible obtener la función generadora de momentos de la media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ como el producto de las funciones generadoras de momentos individuales evaluadas en t/n. Esto conduce a la expresión

$$M_{\bar{X}}(t) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right),$$

lo que permite demostrar que la media muestral también sigue una distribución normal. En términos interpretativos, este resultado indica que, aunque los datos individuales presentan variabilidad, el promedio de estos tiende a ser más estable a medida que aumenta el tamaño de la muestra, lo cual es fundamental para la toma de decisiones en inferencia estadística.

Entonces, se obtiene:

$$M_{\bar{X}}(t) = \left[\exp\left(\mu\frac{t}{n} + \frac{1}{2}\sigma^2\frac{t^2}{n^2}\right)\right]^n = \exp\left(\mu t + \frac{1}{2}\frac{\sigma^2}{n}t^2\right),$$

lo cual corresponde a la función generadora de momentos de una distribución normal $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

En efecto, esta igualdad se obtiene aplicando la propiedad $(\exp(a))^n = \exp(na)$ y simplificando término a término. Además, dado que la función generadora de momentos caracteriza de manera única la distribución de una variable aleatoria, se concluye que la media muestral también sigue una distribución normal con varianza reducida (Winkelbauer, 2012)(Goulet & Dutang, 2023).

Formulación del contraste de hipótesis

Se desea contrastar si la media poblacional coincide con un valor de referencia μ_0, a partir de la información obtenida en una muestra aleatoria. Este tipo de análisis permite evaluar si las diferencias observadas entre la media muestral y el valor propuesto pueden atribuirse al azar o si, por el contrario, son lo suficientemente significativas como para cuestionar la validez de la hipótesis planteada.

En este sentido, se formulan las siguientes hipótesis:

$$H_0 : \ \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \ \mu \neq \mu_0$$

donde la hipótesis nula H_0 establece que la media poblacional es igual al valor de referencia, mientras que la hipótesis alternativa H_1 plantea que dicha media es diferente. Este contraste de tipo bilateral permite detectar desviaciones en ambas direcciones, ya sea que la media sea mayor o menor que μ_0.

Estadístico de prueba

Se define el estadístico de prueba con el fin de estandarizar la diferencia entre la media muestral y el valor propuesto en la hipótesis nula, teniendo en cuenta la variabilidad de los datos:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Este estadístico permite medir qué tan alejada se encuentra la media muestral del valor de referencia μ_0 en términos de unidades de desviación estándar.

Bajo la hipótesis nula H_0, se cumple que:

$$Z \sim \mathcal{N}(0,1).$$

teoricamente se determina que Bajo H_0, se tiene que $\mu = \mu_0$, por lo que la media muestral sigue una distribución normal dada por:

$$\bar{X} \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right).$$

Por propiedades de la distribución normal, al estandarizar la variable aleatoria $\bar{X}$, se obtiene:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

Región crítica y tamaño del test

Sea $\alpha \in (0,1)$ el nivel de significancia, el cual representa la probabilidad máxima de cometer un error tipo I, es decir, rechazar la hipótesis nula cuando esta es verdadera. El tamaño del test se define como dicha probabilidad bajo la hipótesis nula y constituye un criterio fundamental en la construcción de pruebas de hipótesis (Casella & Berger, 2002; Lehmann & Romano, 2005).

Prueba bilateral. Para el contraste de hipótesis

$$H_0: \ \mu = \mu_0 \qquad \text{vs} \qquad H_1: \ \mu \neq \mu_0,$$

la región crítica está dada por

$$R = \{\, x : |Z| > z_{\alpha/2} \,\}.$$

Esta región contiene los valores extremos del estadístico de prueba que conducen al rechazo de la hipótesis nula. El test definido es de tamaño α, ya que bajo la hipótesis nula se cumple

$$P(\text{rechazar } H_0) = P(|Z| > z_{\alpha/2}) = \alpha,$$

lo cual garantiza el control del error tipo I en ambas colas de la distribución normal estándar (Casella & Berger, 2002; Devore, 2020).

Prueba unilateral a la derecha. Para el contraste

$$H_0: \ \mu \leq \mu_0 \qquad \text{vs} \qquad H_1: \ \mu > \mu_0,$$

la región crítica está dada por

$$R = \{\, x : Z > z_\alpha \,\}.$$

En este caso, el rechazo de la hipótesis nula ocurre cuando el estadístico de prueba toma valores suficientemente grandes. El test es de tamaño α, puesto que

$$P(\text{rechazar } H_0) = P(Z > z_\alpha) = \alpha,$$

lo que asegura que la probabilidad de cometer un error tipo I no exceda el nivel de significancia fijado (Devore, 2020; Montgomery & Runger, 2018).

Prueba unilateral a la izquierda. Para el contraste

$$H_0 : \ \mu \geq \mu_0 \qquad \text{vs} \qquad H_1 : \ \mu < \mu_0,$$

la región crítica está dada por

$$R = \{ \, x : Z < -z_\alpha \, \}.$$

En este caso, el rechazo de la hipótesis nula se produce cuando el estadístico toma valores suficientemente pequeños. El test es igualmente de tamaño α, ya que

$$P(\text{rechazar } H_0) = P(Z < -z_\alpha) = \alpha,$$

garantizando el control del error tipo I en pruebas unilaterales a la izquierda (Devore, 2020; Lehmann & Romano, 2005).

Región crítica y tamaño del test

Sea $\alpha \in (0, 1)$ el nivel de significancia, el cual representa la probabilidad máxima de cometer un error tipo I, es decir, rechazar la hipótesis nula siendo verdadera (Casella & Berger, 2002) . Para una prueba bilateral, la región crítica está dada por:

$$R = \{x : |Z| > z_{\alpha/2}\}.$$

Esta región contiene los valores extremos del estadístico de prueba que conducen al rechazo de la hipótesis nula.

El test definido anteriormente es de tamaño α, lo que implica que la probabilidad de rechazar H_0 cuando esta es verdadera es igual al nivel de significancia fijado.

Demostración. Bajo la hipótesis nula H_0, se tiene:

$$P(\text{rechazar } H_0) = P(|Z| > z_{\alpha/2}),$$

donde $P(\cdot)$ denota probabilidad y $Z \sim \mathcal{N}(0, 1)$ es una variable aleatoria con distribución normal estándar. La expresión $P(|Z| > z_{\alpha/2})$ representa la probabilidad de que el estadístico de prueba tome valores mayores que $z_{\alpha/2}$ o menores que $-z_{\alpha/2}$, es decir, que caiga en las colas de la distribución.

Como $Z \sim \mathcal{N}(0, 1)$, se cumple que:

$$P(|Z| > z_{\alpha/2}) = 2P(Z > z_{\alpha/2}) = \alpha.$$

Potencia de una prueba

La potencia de una prueba de hipótesis mide la capacidad del procedimiento estadístico para rechazar la hipótesis nula cuando esta es falsa. En términos del error tipo II, la potencia se define como $1 - \beta$, donde β representa la probabilidad de no rechazar la hipótesis nula cuando la hipótesis alternativa es verdadera. Una prueba con mayor potencia tiene una mayor capacidad para detectar diferencias reales respecto al valor planteado en la hipótesis nula, y depende del nivel de significancia, del tamaño de la muestra y de la distancia entre el valor real del parámetro y el valor especificado bajo la hipótesis nula (Lehmann & Romano, 2005).

La potencia de una prueba se define como:

$$1 - \beta = 1 - P(\text{Error tipo II}) = P(\text{rechazar } H_0 \mid H_a \text{ es verdadera})$$

donde β representa la probabilidad de cometer un error de tipo II, es decir, no rechazar la hipótesis nula cuando esta es falsa.

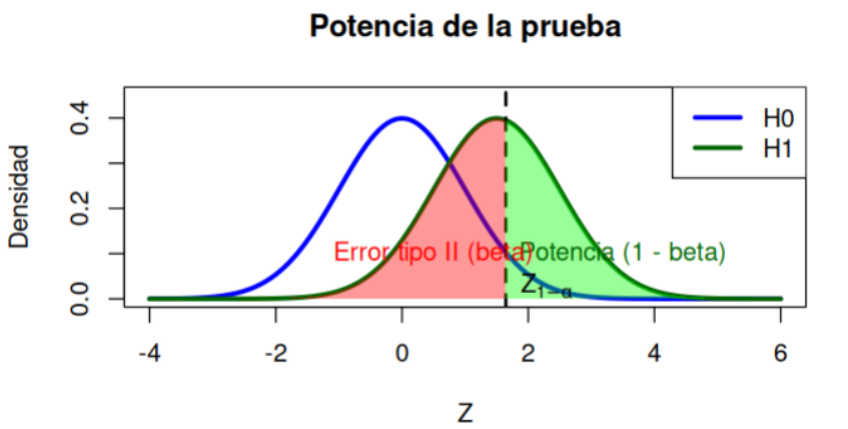
Caso unilateral (cola derecha)

La elección de una prueba unilateral de cola derecha se fundamenta en la naturaleza de la hipótesis alternativa, la cual plantea un interés específico en detectar incrementos en el parámetro poblacional. En este caso, la evidencia en contra de la hipótesis nula se concentra en valores grandes del estadístico de prueba, lo que implica que la región crítica se ubica en el extremo derecho de su distribución (Casella & Berger, 2002)

Sea la hipótesis:

$$H_0 : \mu = \mu_0$$
$$H_a : \mu > \mu_0$$

Figura 1: *Representación de la potencia de una prueba unilateral para la media poblacional*



Nota. Elaboración propia.

En la Figura anterios se muestran las distribuciones del estadístico bajo la hipótesis nula H_0 y la hipótesis alternativa H_1. La línea vertical punteada indica el punto crítico $z_{1-\alpha}$ que delimita la región de rechazo de la hipótesis nula. El área sombreada en rojo corresponde al error tipo II (β), mientras que el área sombreada en verde representa la potencia de la prueba $(1 - \beta)$, es decir, la probabilidad de rechazar correctamente la hipótesis nula cuando la hipótesis alternativa es verdadera.

Un caso particular es considerar un valor específico $\mu_a > \mu_0$:

$$H_0 : \mu = \mu_0, \quad H_a : \mu = \mu_a$$

El estadístico de prueba es:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

La región crítica está dada por:

$$Z \geq z_{1-\alpha}$$

Equivalente a:

$$\bar{X} \geq \mu_0 + \frac{\sigma z_{1-\alpha}}{\sqrt{n}}$$

La potencia se calcula como:

$$1 - \beta = P\left(\bar{X} \geq \mu_0 + \frac{\sigma z_{1-\alpha}}{\sqrt{n}} \mid \mu = \mu_a\right)$$

Dado que:

$$\bar{X} \sim N\left(\mu_a, \frac{\sigma}{\sqrt{n}}\right)$$

Entonces:

$$1 - \beta = 1 - P\left(Y \leq \mu_0 + \frac{\sigma z_{1-\alpha}}{\sqrt{n}}\right)$$

donde $Y \sim N\left(\mu_a, \frac{\sigma}{\sqrt{n}}\right)$.

En este caso particular se considera una alternativa específica $\mu_a > \mu_0$, lo que define un contraste unilateral a la derecha. La región crítica se establece en la cola superior de la distribución del estadístico de prueba, rechazando la hipótesis nula cuando la media muestral es suficientemente mayor que μ_0 (Casella & Berger, 2002).

La potencia de la prueba se interpreta como la probabilidad de que la media muestral supere el umbral crítico cuando la verdadera media poblacional es μ_a. Dado que, bajo esta condición, la media muestral sigue una distribución normal con media μ_a y varianza σ^2/n, la potencia se obtiene como el complemento de la probabilidad de que la media muestral no alcance dicho umbral (Devore, 2020).

Esto refleja que, a medida que μ_a se aleja de μ_0 hacia valores mayores, aumenta la probabilidad de rechazar correctamente la hipótesis nula, incrementando así la potencia de la prueba (Casella & Berger, 2002).

Caso bilateral

En el caso de contrastes bilaterales, la hipótesis alternativa plantea la existencia de diferencias en ambas direcciones, es decir,

Sea:

$$H_0 : \mu = \mu_0, \quad H_a : \mu \neq \mu_0$$

En consecuencia, la región crítica se distribuye en los dos extremos de la distribución del estadístico de prueba, asignando un nivel de significancia $\alpha/2$ a cada cola.

Región crítica:

$$|Z| \geq z_{1-\alpha/2}$$

Equivalente a:

$$\bar{X} \geq \mu_0 + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \quad \text{o} \quad \bar{X} \leq \mu_0 - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}$$

La región crítica puede expresarse en términos de la media muestral, lo que permite una interpretación más directa. En particular, se rechaza la hipótesis nula cuando la media muestral se desvía significativamente del valor hipotético μ_0, superando un umbral determinado por el nivel de significancia, la variabilidad poblacional y el tamaño de muestra. Esta condición define dos regiones críticas simétricas y es equivalente a verificar si μ_0 pertenece al intervalo de confianza bilateral para la media. (Casella & Berger, 2002; Montgomery & Runger, 2018)

La potencia de la prueba bilateral se define como la probabilidad de que la media muestral caiga en cualquiera de las regiones críticas cuando la media poblacional es $\mu_a \neq \mu_0$. Dado que la distribución de la media muestral bajo la hipótesis alternativa es normal con media μ_a y varianza σ^2/n la potencia se expresa como la suma de las probabilidades correspondientes a ambas colas de la distribución:

La potencia es:

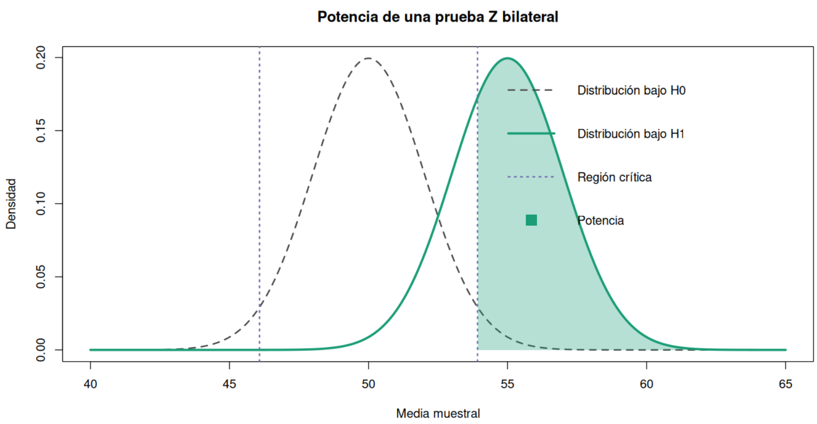
$$1 - \beta = P\left(\bar{X} \geq \mu_0 + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}\right) + P\left(\bar{X} \leq \mu_0 - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}\right)$$

Esta formulación refleja que la evidencia contra la hipótesis nula puede surgir tanto por desviaciones positivas como negativas respecto al valor hipotético.

con:

$$\bar{X} \sim N\left(\mu_a, \frac{\sigma}{\sqrt{n}}\right)$$

Figura 2: *Representación gráfica de la potencia de una prueba bilateral para la media poblacional*



Nota. Elaboración propia.

La Figura 2 representa la distribución de la media muestral bajo dos escenarios: la hipótesis nula H_0 y la hipótesis alternativa H_1 en el contexto de una prueba bilateral para la media poblacional.

Las líneas verticales punteadas indican los valores críticos que delimitan la región de rechazo. En una prueba bilateral, dicha región se ubica en ambos extremos de la distribución, de manera que se rechaza la hipótesis nula tanto para valores suficientemente grandes como suficientemente pequeños de la media muestral. Esta construcción se fundamenta en la distribución del estadístico de prueba y en el nivel de significancia , el cual se distribuye equitativamente en ambas colas como $\alpha/2$ (Pando et al., 1992).

Desde el punto de vista de la potencia de la prueba, la figura permite identificar que la probabilidad de rechazar H_0 cuando la hipótesis alternativa es verdadera corresponde al área bajo la curva de H_1 que cae dentro de las regiones críticas (Casella & Berger, 2002).

Caso con varianza desconocida

Cuando la varianza poblacional σ^2 es desconocida, se reemplaza por la varianza muestral S^2, lo que conduce al uso de la distribución t de Student. Esta distribución permite incorporar la incertidumbre adicional que surge al estimar la variabilidad a partir de la muestra.

En este contexto, el estadístico de prueba se define como:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}.$$

Este estadístico mide la diferencia entre la media muestral $\bar{X}$ y el valor de referencia μ_0, en unidades de error estándar estimado. A diferencia del caso en que la varianza es conocida, aquí se utiliza S en lugar de σ, lo que introduce mayor variabilidad y justifica el uso de la distribución t con $n-1$ grados de libertad.

Para una prueba unilateral a la derecha, la región crítica está dada por:

$$T \geq t_{n-1,\,1-\alpha}.$$

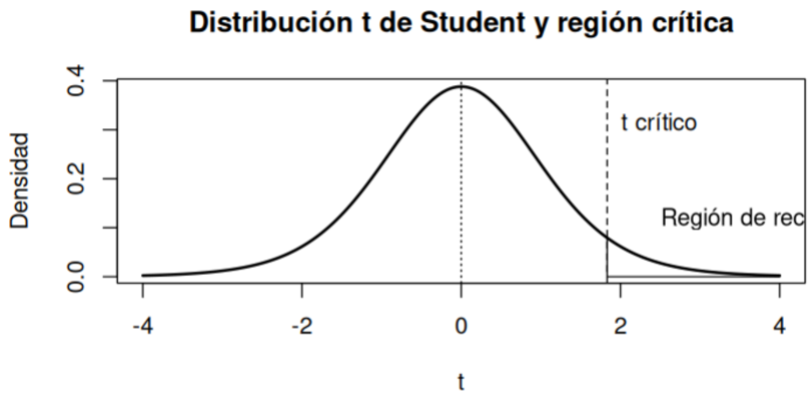
Esto significa que se rechaza la hipótesis nula cuando el valor observado del estadístico de prueba es suficientemente grande.

De manera equivalente, esta condición puede expresarse en términos de la media muestral como:

$$\bar{X} \geq \mu_0 + \frac{S}{\sqrt{n}}\, t_{n-1,\,1-\alpha}.$$

Esta expresión indica que se rechaza la hipótesis nula cuando la media muestral supera un umbral determinado por el valor de referencia, la variabilidad de los datos y el nivel de significancia.

Figura 3: *Distribución t de Student y región crítica para una prueba unilateral*



Nota. Elaboración propia.

La Figura 3 muestra la distribución t de Student con $n-1$ grados de libertad. La región sombreada corresponde a la zona de rechazo, donde el estadístico de prueba toma valores extremos. Si el valor observado de T se ubica en esta región, se rechaza la hipótesis nula.

Potencia en cola izquierda

Sea el contraste de hipótesis

$$H_0: \ \mu = \mu_0 \qquad \text{vs} \qquad H_a: \ \mu < \mu_0.$$

El estadístico de prueba está dado por

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

el cual sigue una distribución normal estándar bajo la hipótesis nula.

De manera equivalente, la condición de rechazo puede expresarse en términos de la media muestral como

$$\bar{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}.$$

Esta expresión permite interpretar la regla de decisión directamente sobre el valor observado de la media muestral. En particular, la hipótesis nula se rechaza cuando $\bar{X}$ toma valores suficientemente menores que el valor de referencia μ_0, teniendo en cuenta la variabilidad poblacional y el tamaño de la muestra. El término $\frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ actúa como un umbral que depende del nivel de significancia fijado y controla la probabilidad de cometer un error tipo I(Devore, 2020).

La probabilidad de cometer un error tipo II está dada por

$$\beta = P\left(Y \geq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right),$$

donde

$$Y \sim \mathcal{N}\left(\mu_a, \frac{\sigma}{\sqrt{n}}\right), \quad \mu_a < \mu_0.$$

El error tipo II corresponde a la probabilidad de no rechazar la hipótesis nula cuando esta es falsa, es decir, cuando el valor real de la media poblacional es $\mu_a < \mu_0$. Esta probabilidad se calcula bajo la distribución de la media muestral asociada a la hipótesis alternativa y depende del tamaño de la muestra, de la variabilidad poblacional y de la distancia entre μ_a y μ_0 (Montgomery & Runger, 2018).

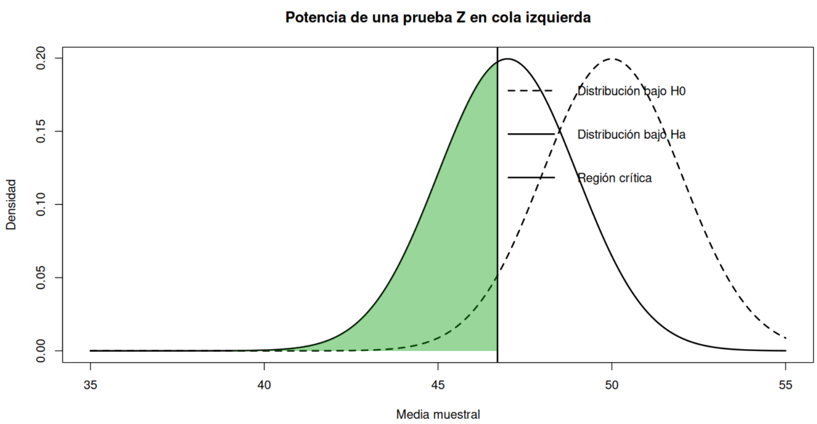
Por lo tanto, la potencia de la prueba se expresa como

$$1 - \beta = P\left(Y \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right),$$

la cual representa la probabilidad de rechazar correctamente la hipótesis nula cuando el valor real de la media poblacional es menor que μ_0.

La potencia de la prueba cuantifica la capacidad del procedimiento estadístico para detectar correctamente una diferencia real en la población. En el caso de una prueba unilateral a la izquierda, esta probabilidad corresponde al área bajo la distribución de la media muestral asociada a la hipótesis alternativa que cae dentro de la región crítica (Devore, 2020).

Figura 4: *Potencia de una prueba Z en cola izquierda*



Nota. El área sombreada corresponde a la probabilidad de rechazar correctamente la hipótesis nula cuando la media poblacional es menor que el valor de referencia.

La Figura 4 ilustra gráficamente la potencia de una prueba Z unilateral en cola izquierda. La región sombreada bajo la distribución correspondiente a la hipótesis alternativa representa la probabilidad de rechazar correctamente la hipótesis nula cuando el valor real de la media poblacional es menor que el valor de referencia. Esta representación permite visualizar cómo la potencia depende del desplazamiento entre μ_a y μ_0, así como del tamaño de la muestra, y muestra que, a medida que la media alternativa se aleja de μ_0 o que aumenta n, la capacidad del test para detectar diferencias reales se incrementa (Devore, 2020; Montgomery & Runger, 2018) .

Error tipo II

El error tipo II está dado por:

$$\beta_{II}(\mu) = P(\text{no rechazar } H_0)$$

El error tipo II se presenta cuando no se rechaza la hipótesis nula H_0 siendo esta falsa. Es decir, ocurre cuando el procedimiento estadístico no logra detectar un efecto o diferencia que realmente existe en la población. Representándose mediante la probabilidad:

Valor p

Sea

$$z_{\text{obs}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

El cual corresponde a la estandarización de la media muestral bajo el supuesto de normalidad y varianza poblacional conocida. Este estadístico sigue una distribución normal estándar bajo la hipótesis nula H_0 (Walpole et al., 2012)

$$p = 2P(Z \geq |z_{\text{obs}}|).$$

Demostración. El valor p se define como la probabilidad, bajo la hipótesis nula, de observar un valor del estadístico de prueba al menos tan extremo como el observado. En el caso de una prueba bilateral, esto implica considerar ambos extremos de la distribución, por lo que se expresa como:

$$p = P(|Z| \geq |z_{\text{obs}}| \mid H_0).$$

Esta definición refleja que el valor p mide la evidencia en contra de la hipótesis nula, cuantificando cuán compatible es el resultado observado con lo que se esperaría si H_0 fuera verdadera (Casella & Berger, 2002).

Debido a la simetría de la distribución normal estándar, la probabilidad de observar valores extremos en ambos lados de la distribución es igual. Por esta razón, el valor p en una prueba bilateral puede calcularse como:

$$p = 2P(Z \geq |z_{\text{obs}}|).$$

□

La optimalidad del test basado en el estadístico Z se fundamenta en el Teorema de Neyman–Pearson, el cual establece que, para contrastes simples de hipótesis de la forma $H_0 : \mu = \mu_0$ frente a $H_1 : \mu = \mu_1$, el test basado en la razón de verosimilitud es el más potente entre todos los tests que tienen el mismo nivel de significancia α. En el modelo normal con varianza conocida, la función de verosimilitud puede expresarse, salvo constantes, como

$$L(\mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{X} - \mu)^2\right),$$

lo que muestra que la evidencia a favor de un valor del parámetro μ depende únicamente de su cercanía con la media muestral $\bar{X}$. Al considerar la razón de verosimilitud entre μ_1 y μ_0, se obtiene una función monótona de la media muestral, lo que implica que la evidencia a favor de la hipótesis alternativa aumenta a medida que $\bar{X}$ toma valores más extremos. En consecuencia, de acuerdo con el Teorema de Neyman–Pearson, el criterio óptimo de decisión consiste en rechazar la hipótesis nula para valores extremos de $\bar{X}$, lo cual es equivalente a rechazar H_0 para valores grandes del estadístico estandarizado Z. Este resultado garantiza que el test basado en Z es uniformemente más potente para contrastes simples bajo el modelo normal con varianza conocida (Lehmann & Romano, 2005).

Ejemplo 1: Prueba Z para la media con varianza conocida (prueba bilateral)

Tabla 2: *Prueba Z para la media poblacional con varianza conocida*

```
# PRUEBA DE HIPÓTESIS PARA LA MEDIA
# X ~ N(μ, σ^2) con varianza conocida
# Se usa la prueba Z para una muestra

# Hipótesis
# H0 :  μ = 4
# H1 :  μ ≠ 4

horas_estudio <- c(4.8, 3.9, 4.5, 5.1, 4.2, 4.7, 5.0, 4.3,
   4.6, 4.9, 3.8, 4.4, 4.7, 5.2, 4.1, 4.6,
   4.9, 4.3, 5.0, 4.5, 4.6, 4.8, 4.2, 4.7,
   5.1, 4.4, 4.6, 4.9, 4.3, 4.8, 4.7, 5.0,
   4.5, 4.6, 4.9, 4.8)

x_bar <- mean(horas_estudio)
sigma <- 1
n <- length(horas_estudio)

Z <- (x_bar - 4)/(sigma/sqrt(n))
p_valor <- 2 * (1 - pnorm(abs(Z)))
```

$\bar{x} = 4{,}62$

$\sigma = 1 \qquad n = 36$

$Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 3{,}72$

Valor $p = 2P(Z \geq |3{,}72|) = 0{,}0002$

Decisión: Se rechaza H_0 con una confianza del 95 %

La tabla anterior muestra la aplicación de una prueba de hipótesis para la media poblacional bajo el supuesto de normalidad y varianza conocida, utilizando la prueba Z para una muestra. A partir de los datos observados se obtiene una media muestral que difiere del valor propuesto en la hipótesis nula, lo cual se cuantifica mediante el estadístico de prueba Z. El valor p asociado al estadístico es menor que el nivel de significancia $\alpha = 0{,}05$, lo que conduce al rechazo de la hipótesis nula con una confianza del 95 %. En consecuencia, existe evidencia estadísticamente significativa para afirmar que la media poblacional es diferente del valor de referencia, de acuerdo con el criterio establecido para la toma de decisiones en pruebas de hipótesis.

Ejemplo 2: Prueba Z para la media con varianza conocida (prueba unilateral a la derecha)

Tabla 3: *Prueba Z unilateral para la media poblacional con varianza conocida*

```
# PRUEBA DE HIPÓTESIS PARA LA MEDIA
# X ~ N(μ, σ²) con varianza conocida
# Se usa la prueba Z para una muestra (unilateral)

# Hipótesis
# H0 :  μ ≥ 4
# H1 :  μ < 4

horas_estudio <- c(4.8, 3.9, 4.5, 5.1, 4.2, 4.7, 5.0, 4.3,
   4.6, 4.9, 3.8, 4.4, 4.7, 5.2, 4.1, 4.6,
   4.9, 4.3, 5.0, 4.5, 4.6, 4.8, 4.2, 4.7,
   5.1, 4.4, 4.6, 4.9, 4.3, 4.8, 4.7, 5.0,
   4.5, 4.6, 4.9, 4.8)

x_bar <- mean(horas_estudio)
sigma <- 1
n <- length(horas_estudio)

Z <- (x_bar - 4)/(sigma/sqrt(n))
p_valor <- pnorm(Z)
```

$\bar{x} = 4{,}62$

$\sigma = 1 \qquad n = 36$

$Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 3{,}72$

Valor $p = P(Z \leq 3{,}72) = 0{,}9999$

Decisión: No se rechaza H_0 con una confianza del 95 %

De acuerdo con los resultados presentados en la tabla anterior, el estadístico de prueba no proporciona evidencia suficiente para rechazar la hipótesis nula. En particular, el valor p obtenido es mayor que el nivel de significancia $\alpha = 0{,}05$, lo que indica que no se dispone de evidencia estadísticamente significativa en contra de H_0. Por tanto, con un nivel de confianza del 95 %, no se rechaza la hipótesis nula.

Ejemplo 3: Prueba Z para la media con varianza conocida (prueba unilateral a la izquierda)

Tabla 4: *Prueba Z unilateral para la media poblacional con varianza conocida*

```
# PRUEBA DE HIPÓTESIS PARA LA MEDIA
# X ~ N(μ, σ²) con varianza conocida
# Se usa la prueba Z para una muestra (unilateral)

# Hipótesis
# H0 :  μ < 4
# H1 :  μ > 4

horas_estudio <- c(4.8, 3.9, 4.5, 5.1, 4.2, 4.7, 5.0, 4.3,
   4.6, 4.9, 3.8, 4.4, 4.7, 5.2, 4.1, 4.6,
   4.9, 4.3, 5.0, 4.5, 4.6, 4.8, 4.2, 4.7,
   5.1, 4.4, 4.6, 4.9, 4.3, 4.8, 4.7, 5.0,
   4.5, 4.6, 4.9, 4.8)

x_bar <- mean(horas_estudio)
sigma <- 1
n <- length(horas_estudio)

Z <- (x_bar - 4)/(sigma/sqrt(n))
p_valor <- 1 - pnorm(Z)
```

$\bar{x} = 4{,}62$

$\sigma = 1 \qquad n = 36$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 3{,}72$$

Valor $p = P(Z \geq 3{,}72) = 0{,}0001$

Decisión: Se rechaza H_0 con un nivel de confianza del 95 %

Con base en los resultados presentados en la tabla anterior, el estadístico de prueba proporciona evidencia suficiente para rechazar la hipótesis nula. En particular, el valor p obtenido es menor que el nivel de significancia $\alpha = 0{,}05$, lo que indica evidencia estadísticamente significativa a favor de la hipótesis alternativa. En consecuencia, con un nivel de confianza del 95 %, se rechaza la hipótesis nula y se concluye que la media poblacional es mayor que el valor propuesto.

2.2. Pruebas de hipótesis para la media en una distribución normal con varianza desconocida

El modelo normal con varianza conocida permite desarrollar pruebas de hipótesis exactas basadas en la distribución normal estándar. Este marco constituye un caso fundamental en la teoría estadística, ya que permite derivar explícitamente las propiedades del test, incluyendo su tamaño, potencia y valor p.

Se presenta un ejemplo basado en una muestra aleatoria simple proveniente de una distribución normal, en el que se realiza una prueba de hipótesis para la media, considerando el estadístico de prueba, la región de rechazo y la potencia de la prueba.

Tabla 5: *Prueba de hipótesis para la media poblacional mediante la prueba t de Student*

```
# PRUEBA DE HIPÓTESIS PARA LA MEDIA
# X ~ N(μ = 3, σ²) con varianza desconocida
# Se usa la prueba t de Student para una muestra

# Hipótesis
# H0 :  μ = 3
# H1 :  μ ≠ 3

y <- rnorm(100, mean = 3, sd = 1)
t.test(y, mu = 3)

t = 0.13426,  gl = 99,  p-value = 0.8935
IC 95%:  [2.79905,  3.23015]
Media muestral = 3.014587
```

A partir de la tabla anterior se presenta el resultado de una prueba de hipótesis para la media poblacional, asumiendo normalidad y varianza desconocida, por lo cual se empleó la prueba t de Student para una muestra. Se contrastó la hipótesis nula $H_0 : \mu = 3$ frente a la alternativa $H_1 : \mu \neq 3$ utilizando una muestra aleatoria de tamaño $n = 100$.

El valor p obtenido es mayor que el nivel de significancia $\alpha = 0{,}05$, por lo que no se rechaza la hipótesis nula. En consecuencia, con una confianza del 95 %, no existe evidencia estadísticamente significativa para afirmar que la media poblacional sea diferente de 3.

Una vez ilustrada la prueba de hipótesis para una sola muestra, se procede a analizar el comportamiento del valor p cuando el experimento se repite múltiples veces mediante simulación. En particular, se estudia la frecuencia con la que se rechaza la hipótesis nula cuando esta es verdadera, lo cual permite estimar el error tipo I y verificar empíricamente que dicha proporción es cercana al nivel de significancia establecido, $\alpha = 0{,}05$.

Tabla 6: *Simulación del error tipo I para la prueba t de Student*

```
# SIMULACIÓN DEL ERROR TIPO I
# H0 :  μ = 3 (verdadera)
# Nivel de significancia α = 0,05

Y <- list()
for (j in 1:1000) {
    Y[[j]] <- rnorm(100, mean = 3, sd = 1)
}

valoresp <- NULL
for (i in 1:1000) {
    valoresp[i] <- t.test(Y[[i]], mu = 3)$p.value
}

Proporción de rechazo (error tipo I): 0.039
```

A partir de la simulación realizada, se obtuvo una proporción de rechazo de la hipótesis nula igual a 0,039, la cual corresponde a una estimación empírica del error tipo I. Este valor es cercano al nivel de significancia teórico $\alpha = 0{,}05$, lo que indica que la prueba estadística presenta un comportamiento consistente con lo esperado desde el punto de vista teórico. En consecuencia, cuando la hipótesis nula es verdadera, la probabilidad de rechazarla incorrectamente es aproximadamente del 5 %, tal como lo establece la definición del nivel de significancia.

Para llevar a cabo esta estimación, se generaron 1000 muestras aleatorias independientes, cada una de tamaño $n = 100$, provenientes de una distribución normal con media $\mu = 3$ y desviación estándar igual a 1. Para cada muestra se aplicó la prueba t de Student con el fin de contrastar la hipótesis nula $H_0 : \mu = 3$.

Posteriormente, se registraron los valores p obtenidos en cada una de las pruebas y se calculó la proporción de casos en los cuales dichos valores fueron inferiores al nivel de significancia $\alpha = 0{,}05$. Esta proporción representa la frecuencia con la que se rechaza la hipótesis nula siendo verdadera, lo que permite obtener una estimación empírica del error tipo I de la prueba.

Tabla 7: *Simulación del error tipo II para la prueba t de Student*

```
# SIMULACIÓN DEL ERROR TIPO II
# H0 :  μ = 3,5 (falsa)
# H1 :  μ ≠ 3,5
# Nivel de significancia α = 0,05

valoresp1 <- NULL
for (i in 1:1000) {
    valoresp1[i] <- t.test(Y[[i]], mu = 3.5)$p.value
}

Proporción de no rechazo (error tipo II β): 0.001
```

En esta sección se analiza el error tipo II, el cual ocurre cuando no se rechaza una hipótesis nula que es falsa. Para ello, se considera la hipótesis nula $H_0 : \mu = 3{,}5$, valor que difiere de la media real utilizada para generar las muestras, $\mu = 3$.

A partir de las 1000 muestras aleatorias generadas previamente, se aplica la prueba t de Student para cada muestra y se calculan los respectivos valores p. El error tipo II (β) se estima como la proporción de veces en las que no se rechaza la hipótesis nula, es decir, cuando el valor p es mayor que el nivel de significancia $\alpha = 0{,}05$, a pesar de que H_0 es falsa.

En esta simulación se obtiene un valor $\beta = 0{,}001$, lo cual indica que la probabilidad de no rechazar una hipótesis nula falsa es muy baja. A partir de este resultado se calcula la potencia de la prueba como $1 - \beta$, la cual es igual a 0,999. Esto evidencia que la prueba tiene una alta capacidad para detectar diferencias reales respecto al valor hipotético planteado.

Este comportamiento es coherente con la teoría, ya que cuando la hipótesis nula es falsa los valores p tienden a concentrarse cerca de cero, lo que incrementa la probabilidad de rechazar correctamente la hipótesis nula.

Tabla 8: *Regla de decisión basada en el valor p*

```
# Interpretación
Si p-valor < α, se rechaza H0
Si p-valor > α, no se rechaza H0
```

Aplicando la regla de decisión basada en el valor p presentada en la tabla anterior, se observa que el valor p obtenido en la prueba es mayor que el nivel de significancia $\alpha = 0{,}05$. En consecuencia, no se rechaza la hipótesis nula H_0. Por lo tanto, con una confianza del 95 %, no existe evidencia estadísticamente significativa para afirmar que la media poblacional sea diferente del valor propuesto en la hipótesis nula.

A partir del ejemplo basado en la distribución normal se observa que el nivel de significancia α se fija previamente y se utiliza como criterio para contrastar la hipótesis nula mediante el estadístico de prueba. Este nivel controla la probabilidad de cometer un error tipo I, es decir, rechazar una hipótesis nula cuando esta es verdadera. En contraste, el error tipo II no se fija de antemano, debido a que su valor no depende únicamente de una decisión del investigador, sino de varios factores inherentes al problema estadístico, tales como el tamaño de la muestra, la variabilidad de la población y la distancia entre el valor real del parámetro y el valor especificado en la hipótesis nula (Casella & Berger, 2002).

En particular, el error tipo II corresponde a la probabilidad de no rechazar la hipótesis nula cuando esta es falsa, lo que implica que su magnitud varía según el escenario bajo el cual se evalúe la hipótesis alternativa. A diferencia del error tipo I, que se controla explícitamente al seleccionar un nivel de significancia, el error tipo II depende del comportamiento del estadístico de prueba bajo valores alternativos del parámetro poblacional, los cuales no son únicos ni conocidos a priori. Por esta razón, no es posible fijar un único valor del error tipo II antes de realizar el análisis, sino que este debe calcularse para valores específicos de la hipótesis alternativa, teniendo en cuenta el tamaño de la muestra y la variabilidad de la población (Devore, 2020).

A partir del cálculo del error tipo II se define la potencia de la prueba, expresada como $1 - \beta$, la cual representa la probabilidad de rechazar correctamente una hipótesis nula falsa. En este contexto, un error tipo II elevado indica una baja capacidad del test para detectar diferencias reales, mientras que una potencia alta refleja un mejor desempeño del procedimiento estadístico. En consecuencia, el análisis conjunto del error tipo II y de la potencia resulta fundamental para evaluar la efectividad de una prueba de hipótesis, complementando la información proporcionada por el nivel de significancia y permitiendo comparar distintos procedimientos estadísticos bajo un mismo criterio (Montgomery & Runger, 2018)

En la práctica, cuando se dispone de varios estadísticos de prueba para contrastar una misma hipótesis bajo un nivel de significancia fijo, se prefiere aquella prueba que presente un menor error tipo II o, de manera equivalente, una mayor potencia para las alternativas consideradas. Este criterio está respaldado por la teoría clásica de pruebas de hipótesis, particularmente por el enfoque de Neyman–Pearson, el cual establece que, entre pruebas de igual nivel, es preferible aquella que maximiza la probabilidad de detectar correctamente una hipótesis alternativa verdadera (Casella & Berger, 2002; Lehmann & Romano, 2005).

2.3. Uso de simulaciones y tecnologías en la enseñanza de la estadística

El uso de simulaciones y herramientas tecnológicas en la enseñanza de la estadística ha adquirido una gran relevancia dentro de los procesos educativos, debido a que estas permiten representar fenómenos aleatorios, analizar datos y visualizar conceptos estadísticos de manera dinámica e interactiva. De acuerdo con (Batanero, 2001b), las tecnologías favorecen la comprensión de conceptos estadísticos al posibilitar que los estudiantes exploren situaciones contextualizadas, interpreten información y analicen diferentes comportamientos de los datos mediante representaciones gráficas y simulaciones. En este sentido, el uso de herramientas tecnológicas dentro del aula permite superar algunas dificultades asociadas a la enseñanza tradicional de la estadística, especialmente en temas donde los estudiantes presentan problemas para comprender conceptos abstractos relacionados con probabilidad, variabilidad e inferencia estadística.

Asimismo, Batanero, Sánchez y Tauber (Batanero et al., 2004) señalan que las simulaciones constituyen una estrategia didáctica que favorece el desarrollo del razonamiento estadístico, debido a que permiten observar múltiples resultados a partir de una misma situación aleatoria, facilitando procesos de comparación, interpretación y análisis de la variabilidad presente en los datos. De esta manera, los estudiantes pueden comprender que los resultados estadísticos no son completamente determinísticos, sino que se encuentran influenciados por el azar y las características de las muestras analizadas.

En relación con lo anterior, el uso de simulaciones resulta especialmente pertinente en la enseñanza de conceptos asociados a inferencia estadística y pruebas de hipótesis, ya que permite representar de manera visual e interactiva procesos relacionados con distribución muestral, comportamiento del valor p, variabilidad de las muestras y toma de decisiones estadísticas. A través de estas representaciones, los estudiantes pueden observar cómo cambian los resultados al modificar el tamaño de muestra o repetir un procedimiento estadístico varias veces, favoreciendo una mejor comprensión de conceptos que, desde una explicación únicamente teórica, suelen generar dificultades o interpretaciones erróneas.

De igual manera, las herramientas tecnológicas promueven procesos de aprendizaje más activos y participativos, debido a que los estudiantes no se limitan únicamente a realizar cálculos o aplicar fórmulas, sino que también interactúan con los datos, ejecutan simulaciones y analizan el significado de los resultados obtenidos (Batanero, 2011). Esto favorece procesos de argumentación, reflexión y toma de decisiones fundamentadas en evidencia estadística, permitiendo que los estudiantes cuestionen sus ideas iniciales y construyan nuevas interpretaciones a partir de la experimentación y el análisis de situaciones contextualizadas.

Por otra parte, el uso de software estadístico dentro del aula facilita la visualización de fenómenos relacionados con la aleatoriedad y la variabilidad, aspectos fundamentales dentro de la inferencia estadística. En este sentido, las simulaciones permiten evidenciar que resultados diferentes pueden obtenerse incluso bajo condiciones similares, favoreciendo la comprensión de que el comportamiento de los datos depende de procesos aleatorios inherentes al muestreo y a la naturaleza probabilística de los fenómenos estudiados (Chance et al., 2007; J. Garfield & Ben-Zvi, 2008a). Además, estas herramientas contribuyen a disminuir algunas concepciones erróneas frecuentes en los estudiantes, especialmente aquellas relacionadas con la interpretación del valor p, el rechazo de hipótesis y la influencia del tamaño de muestra en las pruebas estadísticas.

En consecuencia, el uso de simulaciones y tecnologías dentro de la educación estadística constituye una estrategia didáctica que fortalece los procesos de razonamiento estadístico y favorece la construcción significativa del conocimiento. De esta manera, las herramientas tecnológicas no solo facilitan la comprensión de conceptos complejos relacionados con inferencia estadística y pruebas de hipótesis, sino que también permiten que los estudiantes desarrollen procesos de interpretación, análisis y reflexión sobre los resultados obtenidos durante el desarrollo de las actividades propuestas.

"En concordancia con los planteamientos sobre el uso de tecnologías en la educación estadística, en la presente investigación se empleó el software RStudio como herramienta de apoyo para el desarrollo de simulaciones relacionadas con las pruebas de hipótesis. El uso de este entorno permitió que los estudiantes interactuaran de manera dinámica con conceptos como el valor-p, el nivel de significancia, los errores tipo I y tipo II, así como con la variabilidad muestral presente en los procesos inferenciales. De igual manera, las simulaciones realizadas en RStudio favorecieron la visualización y exploración de los conceptos estadísticos, permitiendo que los estudiantes contrastaran sus ideas iniciales y reflexionaran sobre el comportamiento de los resultados obtenidos en diferentes situaciones. Asimismo, el uso de representaciones gráficas y repeticiones simuladas facilitó una comprensión más profunda de las pruebas de hipótesis, promoviendo procesos de razonamiento, interpretación y argumentación estadística durante el desarrollo de la intervención didáctica (Batanero, 2001b)."

2.4. Prueba no paramétrica de Wilcoxon

La prueba de rangos con signo de Wilcoxon corresponde a una prueba estadística no paramétrica utilizada para comparar dos muestras relacionadas o dependientes, especialmente cuando los datos no cumplen con el supuesto de normalidad (Siegel & Castellan, 1995). Esta prueba permite analizar las diferencias existentes entre dos mediciones realizadas sobre un mismo grupo de individuos, con el propósito de determinar si los cambios observados son estadísticamente significativos. Por esta razón, suele emplearse en investigaciones educativas y sociales donde se comparan resultados obtenidos antes y después de una intervención o estrategia aplicada (Daniel, 2005).

Asimismo, Daniel (2005) señala que el procedimiento de la prueba consiste en calcular las diferencias entre cada par de observaciones, ordenarlas de acuerdo con su magnitud y asignarles rangos acompañados de signos positivos o negativos. Posteriormente, dichos rangos son utilizados para obtener un estadístico de prueba que permita determinar si existe evidencia suficiente para aceptar o rechazar la hipótesis planteada en la investigación (Conover, 1999).

En este sentido, la prueba de Wilcoxon se fundamenta en el planteamiento de dos hipótesis estadísticas. La hipótesis nula (H_0) establece que no existen diferencias significativas entre las mediciones realizadas, indicando que la intervención aplicada no produjo cambios relevantes en los resultados obtenidos (Siegel & Castellan, 1995). Por otro lado, la hipótesis alternativa (H_1) plantea que sí existen diferencias significativas entre las mediciones, evidenciando posibles efectos o cambios derivados de la intervención desarrollada (Daniel, 2005). Según Conover (1999), la decisión de aceptar o rechazar estas hipótesis depende del análisis del valor de significancia o p-valor obtenido durante la aplicación de la prueba estadística.

De igual manera, la prueba de rangos con signo de Wilcoxon representa una alternativa adecuada a la prueba t de Student para muestras relacionadas, particularmente cuando se trabaja con tamaños de muestra pequeños o con datos que no presentan distribución normal (Conover, 1999). Debido a ello, esta prueba ha adquirido gran relevancia dentro de investigaciones experimentales y educativas, ya que permite evaluar de manera confiable los cambios producidos en un mismo grupo y determinar si dichos cambios son consecuencia de la estrategia o intervención aplicada y no simplemente producto del azar (Daniel, 2005).

2.5. Modeo RyA

El modelo de Reflexión y Acción (RyA), propuesto por Parada (2011), surge como una propuesta orientada al fortalecimiento de los procesos de enseñanza y aprendizaje de las matemáticas a partir de la reflexión constante sobre la práctica educativa. Este modelo reconoce que el aprendizaje matemático no depende únicamente de la adquisición de procedimientos o contenidos, sino también de la capacidad que tienen los estudiantes y docentes para analizar, interpretar y comprender las acciones desarrolladas dentro del aula.

De acuerdo con (Parada, 2011), la reflexión docente constituye un proceso que requiere orientación y herramientas que permitan al profesor centrar su atención en aspectos específicos de su acción pedagógica. En este sentido, el modelo RyA busca fortalecer el pensamiento reflexivo del docente a partir del análisis de la actividad matemática que promueve en el aula. Para ello, el modelo propone tres momentos fundamentales de reflexión: la reflexión para la acción, la reflexión en la acción y la reflexión sobre la acción.

La **reflexión para la acción** se desarrolla antes de la clase y está asociada a la planificación de las actividades de enseñanza, la anticipación de dificultades y la organización de los recursos y estrategias didácticas. La **reflexión en la acción** ocurre durante la práctica docente y permite al profesor analizar las situaciones emergentes en el aula, tomar decisiones pedagógicas y adaptar sus estrategias de acuerdo con las respuestas de los estudiantes. Finalmente, la **reflexión sobre la acción** se realiza después de la clase, cuando el docente evalúa críticamente la actividad matemática desarrollada, compara lo planeado con lo logrado e identifica fortalezas y aspectos susceptibles de mejora.

Asimismo, este modelo plantea que la construcción del conocimiento matemático se desarrolla mediante un proceso continuo de acción y reflexión. Por ello, la reflexión no se limita únicamente al resultado obtenido, sino que también involucra el análisis de los procedimientos, estrategias y razonamientos utilizados durante el desarrollo de la actividad matemática (Parada, 2011). De esta manera, el estudiante adquiere un papel más participativo dentro de su proceso de aprendizaje, permitiéndole comprender con mayor profundidad los conceptos trabajados.

Por otra parte, Parada (2011) señala que la interacción entre los participantes favorece el intercambio de ideas y experiencias, contribuyendo al fortalecimiento de procesos de comprensión matemática y pensamiento crítico. Debido a ello, el modelo de Reflexión y Acción ha sido empleado en diferentes investigaciones educativas, especialmente en el área de matemáticas, donde se busca promover aprendizajes significativos y mejorar las prácticas pedagógicas mediante procesos permanentes de análisis y reflexión sobre la acción desarrollada en el aula.

2.6. El razonamiento estadístico en la educación matemática

Razonamiento estadístico

El razonamiento estadístico constituye una de las competencias fundamentales dentro del aprendizaje de la estadística, debido a que permite interpretar, analizar y comprender información a partir de datos, gráficos y situaciones relacionadas con el contexto. Según J. Garfield y Gal (1999), el razonamiento estadístico hace referencia a la manera en que las personas utilizan conceptos e ideas estadísticas para interpretar información, establecer conclusiones y tomar decisiones fundamentadas a partir de los datos analizados. En este sentido, el razonamiento estadístico no se limita únicamente a la aplicación de procedimientos matemáticos o fórmulas, sino que involucra procesos de interpretación, argumentación, análisis y comprensión de la información estadística.

De igual manera, J. Garfield (2002) señala que el razonamiento estadístico implica la comprensión de conceptos relacionados con variabilidad, probabilidad, distribución, muestreo, incertidumbre e inferencia estadística. Estos elementos permiten que los estudiantes desarrollen habilidades para interpretar datos, analizar representaciones gráficas y formular conclusiones sustentadas en evidencia estadística. Asimismo, el razonamiento estadístico favorece procesos de análisis crítico y toma de decisiones, los cuales resultan fundamentales dentro de la educación matemática y estadística.

En el contexto de las pruebas de hipótesis, el razonamiento estadístico implica comprender la lógica inferencial que sustenta la toma de decisiones a partir de información muestral. Esto involucra habilidades relacionadas con la formulación de hipótesis estadísticas, la interpretación del valor p, la comprensión del nivel de significancia, el reconocimiento de los errores tipo I y tipo II y la interpretación de las conclusiones inferenciales. Sin embargo, J. Garfield y Ahlgren (1988) afirman que muchos estudiantes presentan dificultades en la comprensión de conceptos estadísticos, especialmente en temas relacionados con probabilidad, interpretación de gráficos, análisis de muestras, variabilidad y comprensión de datos.

En consecuencia, diversas investigaciones han resaltado la necesidad de implementar estrategias didácticas orientadas al fortalecimiento del razonamiento estadístico mediante actividades centradas en el análisis, la interpretación y la reflexión sobre la información trabajada en el aula (J. Garfield, 2002). En particular, la enseñanza de las pruebas de hipótesis requiere metodologías que favorezcan la comprensión de la variabilidad y la incertidumbre, promoviendo procesos de interpretación, argumentación y toma de decisiones fundamentadas en evidencia estadística (J. Garfield & Ben-Zvi, 2008a). Asimismo, el uso de herramientas tecnológicas y simulaciones interactivas ha demostrado ser una estrategia pertinente para facilitar la comprensión conceptual de estos procesos inferenciales, al permitir la visualización dinámica de fenómenos aleatorios y fortalecer el razonamiento estadístico de los estudiantes (Chance et al., 2007)

Statistical Reasoning Assessment (SRA)

En relación con lo anterior, J. Garfield (2003) desarrolló el instrumento denominado *Statistical Reasoning Assessment* (SRA), el cual fue diseñado con el propósito de evaluar el razonamiento estadístico de los estudiantes a partir de situaciones relacionadas con probabilidad y estadística. Este instrumento permite identificar la forma en que los estudiantes interpretan la información estadística y los razonamientos utilizados durante la resolución de diferentes situaciones problema. En este sentido, el análisis no se centra únicamente en determinar si una respuesta es correcta o incorrecta, sino también en comprender los procesos de pensamiento y argumentación utilizados por los estudiantes.

A diferencia de las evaluaciones tradicionales centradas en la aplicación de fórmulas o procedimientos, el SRA busca analizar cómo los estudiantes interpretan conceptos relacionados con la variabilidad, el muestreo, la incertidumbre y la inferencia estadística. Asimismo, el modelo planteado en el SRA organiza las respuestas de los estudiantes mediante categorías de razonamiento estadístico correcto e incorrecto (J. B. Garfield, 2003). Entre los razonamientos correctos se encuentran aspectos relacionados con la comprensión de la variabilidad, el reconocimiento de la importancia del tamaño de muestra, la interpretación adecuada de probabilidades y el análisis de distribuciones estadísticas.

Por otro lado, los razonamientos incorrectos permiten identificar dificultades o concepciones erróneas presentes en los estudiantes, especialmente en temas relacionados con causalidad, representatividad, interpretación de promedios y comprensión de probabilidades. De esta manera, el modelo de razonamiento estadístico propuesto por Garfield permite establecer categorías, matrices y criterios de análisis para interpretar las respuestas obtenidas en las pruebas aplicadas a los estudiantes. A partir de estas categorías es posible identificar patrones de razonamiento, secuencias de pensamiento, dificultades conceptuales y formas de interpretación desarrolladas durante la resolución de las actividades estadísticas.

No obstante, dado que el instrumento original no se encuentra específicamente orientado al estudio de las pruebas de hipótesis, en la presente investigación se realizó una adaptación de sus principios teóricos con el propósito de analizar habilidades de razonamiento inferencial asociadas a dicho contenido. Esta adaptación permitió establecer categorías de análisis relacionadas con la identificación de hipótesis estadísticas, la comprensión de errores inferenciales, la interpretación del valor p, el nivel de significancia y la toma de decisiones bajo incertidumbre.

En este sentido, el modelo de razonamiento estadístico propuesto por Garfield permite establecer categorías, matrices y criterios de análisis para interpretar las respuestas obtenidas en las pruebas aplicadas a los estudiantes. A partir de estas categorías es posible identificar patrones de razonamiento, secuencias de pensamiento, dificultades conceptuales y formas de interpretación desarrolladas durante la resolución de las actividades estadísticas. Por ello, este enfoque resulta pertinente para la presente investigación, debido a que posibilita analizar no solo los resultados obtenidos por los estudiantes, sino también los procesos de razonamiento estadístico evidenciados en sus respuestas y argumentaciones (J. Garfield & Chance, 2000). Asimismo, permite interpretar las dificultades conceptuales presentes en la comprensión de las pruebas de hipótesis y analizar el impacto de la propuesta didáctica basada en simulaciones interactivas sobre el desarrollo del razonamiento estadístico.

2.6.1. Habilidades de razonamiento en pruebas de hipótesis

En coherencia con los fundamentos del razonamiento estadístico y con la adaptación realizada del SRA, la presente investigación define un conjunto de habilidades de razonamiento asociadas a las pruebas de hipótesis. Estas habilidades tienen como propósito analizar la comprensión conceptual de los estudiantes frente a los procesos inferenciales.

Las habilidades consideradas se fundamentan en las dificultades reportadas en la literatura sobre educación estadística, particularmente en relación con la interpretación del valor p, la comprensión del nivel de significancia, la diferenciación entre hipótesis estadísticas y la toma de decisiones bajo incertidumbre (Castro Sotos et al., 2007; J. Garfield & Ben-Zvi, 2008a). En este sentido, se busca identificar cómo los estudiantes interpretan los conceptos inferenciales y de qué manera justifican sus decisiones estadísticas en situaciones contextualizadas.

Entre las habilidades consideradas se encuentran la identificación de la hipótesis nula y alternativa, la comprensión de los errores tipo I y tipo II, la interpretación del valor p y de la significancia práctica, la comprensión del nivel de significancia , la aplicación de pruebas de hipótesis en contexto y la interpretación de la decisión inferencial.

Estas habilidades orientaron tanto el diseño del pretest como el análisis de las respuestas de los estudiantes durante la implementación de la propuesta didáctica basada en simulaciones interactivas. En particular, permitieron identificar dificultades conceptuales relacionadas con la inferencia estadística y analizar la manera en que las simulaciones favorecieron procesos de interpretación, argumentación y razonamiento estadístico en torno a las pruebas de hipótesis.

Por consiguiente, las habilidades de razonamiento definidas en esta investigación constituyen una herramienta de análisis pertinente para comprender cómo los estudiantes construyen significado alrededor de los conceptos inferenciales y cómo interpretan la toma de decisiones estadísticas en contextos de incertidumbre.

3. Metodología

El presente estudio se inscribe en un enfoque cualitativo con componentes cuantitativos, dado que su objetivo principal es diseñar e implementar una propuesta didáctica que integre fundamentos teóricos y simulaciones computacionales para la enseñanza de las pruebas de hipótesis bajo el supuesto de normalidad, con aplicaciones a situaciones contextualizadas. Como lo señala Calixto et al. (2023), este tipo de investigación se desarrolla en el entorno habitual de los participantes, lo que favorece el análisis de los procesos de aprendizaje en contextos auténticos. El diseño metodológico corresponde a un estudio cuasi-experimental basado en la implementación de una secuencia didáctica, cuyo propósito es analizar cómo los estudiantes construyen significados en torno a las pruebas de hipótesis.

El análisis de los datos se fundamenta en el modelo de Reflexión-y-Acción (RyA) propuesto por Parada, el cual orienta la interpretación de los procesos de construcción de significado a partir de la articulación entre las acciones desarrolladas por los estudiantes y los razonamientos que emergen de dichas acciones en contextos educativos auténticos (Parada, 2016; Parada y Fiallo, 2011). Este modelo resulta especialmente pertinente para la presente investigación, ya que permite caracterizar de manera sistemática cómo los estudiantes interpretan, argumentan y toman decisiones al interactuar con tareas de inferencia estadística mediadas por simulaciones, favoreciendo una comprensión profunda de sus procesos cognitivos y del desarrollo del razonamiento estadístico (J. Garfield y Ben-Zvi, 2008a).

Teniendo en cuenta lo anterior en esta sección se describe la metodología que se seguirá en la investigación la cual se validara con un formulario pretest cuyo propósito es identificar a la luz del marco teórico los conocimientos iniciales de los participantes para posteriormente, valorar su progreso después de la intervención con un formulario postest. La metodología de divide en cuatro fases las cuales son: fase 1: Revisión bibliográfica, fase 2: contexto y participantes, fase 3: Diseño y valoración de los instrumentos de recolección de datos, fase 4: Técnicas e instrumentos de recolección de datos, fase 5: valoración del instrumento, fase 6: procedimiento y valoración de la intervención.

Fase 1: revisión bibliográfica

Según Gómez-Luna et al. (2014) la fase de la revisión bibliográfica constituye una parte fundamental en el proyecto de investigación donde se obtiene la información relevante del objeto de estudio. Es por esto que, en este documento la revisión bibliográfica se concibió como un proceso de análisis y síntesis de información referentes a las pruebas de hipótesis donde los resultados se organizaron de la siguiente forma: En primer lugar la relevancia de la interpretación de la prueba de hipótesis luego, se revisaron los avances de la didáctica en la inferencia estadística, los cuales se presentaron en la sección 3.

Fase 2: Contexto y participantes

El estudio se llevó a cabo en la Universidad Industrial de Santander (UIS), específicamente en el curso Tópicos en Estadística. se eligió este curso ya que se considera pertinente dado que en dicho espacio se abordan contenidos relacionados con pruebas de hipótesis, lo cual resulta coherente con los propósitos de la investigación, orientada a analizar la experiencia de aprendizaje de la prueba de hipótesis mediante el uso de simulaciones computacionales.

La selección de los participantes se realizó mediante un muestreo intencional, dado que, como señala Patton (2015), este tipo de muestreo es característico de los estudios cualitativos, ya que permite seleccionar participantes que poseen experiencias significativas relacionadas con el objeto de investigación. En este sentido, se consideró pertinente trabajar con estudiantes que se encuentran en un proceso de formación inicial y que, dentro del curso, abordan contenidos propios de la inferencia estadística.

Se contó con un grupo de participantes de 8 estudiantes, con quienes se realizó un análisis detallado de las producciones, interacciones y significados construidos en torno al aprendizaje de las pruebas de hipótesis. Este número de participantes se considera adecuado para el tipo de estudio, ya que permite una aproximación analítica profunda, privilegiando la riqueza interpretativa de la información recolectada y la comprensión de los significados construidos por los participantes en relación con el objeto de estudio (Creswell, 2013).

Fase 3: Diseño y validación de los instrumentos de recolección de datos

Adaptación del Statistical Reasoning Assessment y fundamentos didácticos

La propuesta didáctica se sustenta en el enfoque del razonamiento estadístico, entendido como la capacidad de los estudiantes para interpretar, relacionar y justificar información estadística en contextos de incertidumbre, más allá de la aplicación mecánica de procedimientos. Desde esta perspectiva, el aprendizaje de la estadística implica reconocer y analizar las concepciones erróneas que los estudiantes construyen en torno a la inferencia estadística, tales como la formulación de hipótesis, la interpretación del valor p, el significado del nivel de significancia y la distinción entre los errores tipo I y tipo II, aspectos que han sido ampliamente documentados en la literatura en educación estadística (J. Garfield & Ben-Zvi, 2008b) .

En este marco, el Statistical Reasoning Assessment (SRA) propuesto por Joan Garfield es un instrumento diseñado para evaluar el razonamiento estadístico de los estudiantes, en el contexto de la investigación en educación estadística. Su objetivo principal es identificar tanto niveles adecuados de comprensión como concepciones erróneas frecuentes en temas fundamentales de la estadística. fue adaptado con el propósito de orientar tanto el diagnóstico como el diseño de la intervención didáctica. Dicha adaptación no se concibe como una modificación meramente técnica del instrumento original, sino como una decisión metodológica fundamentada en criterios didácticos. En particular, se realizó una clasificación de concepciones erróneas, ver cuadro 11, relevantes para el contexto de la propuesta, lo que permitió identificar dificultades específicas en el razonamiento inferencial de los estudiantes y definir parámetros de análisis propios, acordes con los objetivos del estudio.

La adaptación del SRA posibilitó, además, establecer una relación entre las concepciones erróneas detectadas en el diagnóstico inicial, las actividades diseñadas durante la intervención didáctica y la valoración posterior del aprendizaje, ver cuadro 12 . De este modo, el instrumento cumple una doble función: por un lado, permite caracterizar el nivel de razonamiento estadístico de los estudiantes; por otro, orienta el diseño de situaciones didácticas dirigidas a promover el tránsito desde interpretaciones intuitivas o incorrectas hacia formas más coherentes y fundamentadas de razonamiento estadístico. Esta articulación entre diagnóstico, intervención y evaluación es coherente con enfoques actuales de la educación estadística centrados en la comprensión conceptual y la toma de decisiones informadas (J. Garfield & Ben-Zvi, 2008b; Zieffler et al., 2008).

Instrumentos de recolección de datos

Para la recolección de datos se diseñaron dos cuestionarios, llamados pretest y postest, estructurados con el propósito de evaluar la interpretación de los estudiantes sobre las pruebas de hipótesis, particularmente en contextos apoyados en simulaciones. El instrumento se orientó a identificar tanto el nivel de comprensión conceptual como las dificultades en la interpretación de resultados estadísticos.

El instrumento pretest y postest fueron diseñados tomando como referencia distintos aportes de la literatura en educación estadística. En este sentido, se ha evidenciado que la confusión entre la hipótesis nula y la hipótesis alternativa constituye una de las dificultades más frecuentes en la comprensión de la prueba de hipótesis, así como en la interpretación adecuada de sus resultados (Vallecillos y Batanero, 1996, citado en (delMas et al., 2007) . Asimismo, conceptos como el error tipo I y el error tipo II suelen resultar complejos para los estudiantes, debido a las exigencias de razonamiento que implica su correcta interpretación dentro del proceso inferencial (Batanero, 2001a). Teniendo en cuenta estos antecedentes, dichos contenidos fueron considerados en la elaboración del instrumento, con el propósito de identificar posibles dificultades conceptuales en los estudiantes antes de la intervención didáctica, así como su progreso luego de la intervención.

Los cuestionarios estuvieron conformados por 7 preguntas de selección múltiple (ver anexo 6 y 6), organizadas en cuatro dimensiones: (1) interpretación conceptual de los tipos de errores, (2) interpretación de resultados en contextos estadísticos, (3) interpretación conceptual del valor p y (4) interpretación conceptual de las pruebas de hipótesis. Esta estructura permitió recoger información objetiva y sistemática sobre el nivel de comprensión de los estudiantes en relación con estos conceptos fundamentales de la inferencia estadística.

Diseño de la intervención didáctica

La secuencia didáctica se diseñó en torno al uso de simulaciones computacionales para favorecer la construcción conceptual de distribuciones muestrales y pruebas de hipótesis mediante procedimientos de remuestreo, dado que estos enfoques permiten a los estudiantes explorar empíricamente la variabilidad y comprender la inferencia como un proceso probabilístico basado en modelos, más que como la aplicación mecánica de algoritmos (Waters, 2025). En concecuencia, se ha mostrado que la enseñanza basada en simulaciones mejora la comprensión conceptual del valor p, la variabilidad muestral y la toma de decisiones inferenciales, en comparación con enfoques tradicionales centrados en fórmulas (J. Garfield & Ben-Zvi, 2008a).

El uso de simulaciones computacionales y de entornos interactivos para la enseñanza de la inferencia estadística permite que los estudiantes exploren la variabilidad, construyan modelos empíricos y desarrollen interpretaciones basadas en datos, en lugar de limitarse a la aplicación mecánica de procedimientos formales (Erickson et al., 2019). Esta organización favorece la emergencia de acciones observables y razonamientos verbalizados por parte de los estudiantes, lo cual resulta coherente con el modelo de Reflexión-y-Acción (RyA), que propone analizar el aprendizaje a partir de la articulación entre las prácticas realizadas y los significados construidos por los participantes (Parada, 2016; Parada & Fiallo, 2011).

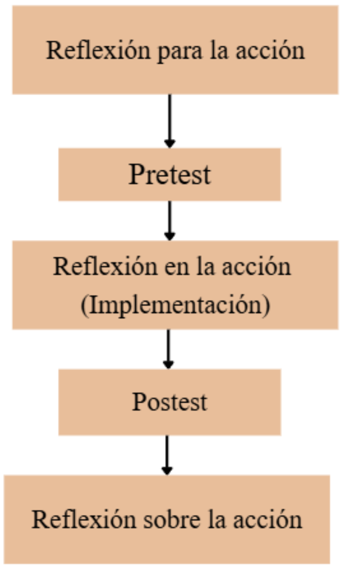
Es por esto que la intervención didáctica se realizo con el propósito de fortalecer la interpretación de los conceptos fundamentales para el aprendizaje de las pruebas de hipótesis, específicamente el muestreo aleatorio simple, las pruebas de hipótesis, los errores tipo I y tipo II, el nivel de significancia y el valor p.

La propuesta se estructuró mediante una secuencia didáctica compuesta por una sesión de dos horas, en la cual se abordaron de forma progresiva los contenidos mencionados anteriormente. En una primera fase, se introdujo el concepto de muestreo aleatorio simple, enfatizando su importancia en la obtención de muestras representativas.

Posteriormente, se desarrollaron actividades orientadas a la comprensión de las pruebas de hipótesis, incluyendo la formulación de hipótesis nula y alternativa, así como la interpretación de los errores tipo I y tipo II y el nivel de significancia.

La intervención incluyó el uso de ejemplos contextualizados y simulaciones en R Studio, con el fin de facilitar la comprensión de los conceptos y fomentar el aprendizaje activo. Con el fin de representar la estructura de la intervención didáctica, se presenta a continuación un esquema basado en el modelo de Reflexión y Acción.

Figura 5: *Secuencia de la intervención didáctica basada en el modelo de Reflexión y Acción.*



Fase 4: Técnicas e instrumentos de recolección de datos

Para la recolección de datos se emplearon diferentes técnicas e instrumentos orientados a documentar tanto el proceso de aprendizaje como los resultados obtenidos por los estudiantes durante la secuencia didáctica. En primer lugar, se consideró la observación del desarrollo de las actividades propuestas, con el fin de identificar las acciones, interacciones y formas de razonamiento que emergieron durante la implementación de la secuencia didáctica, especialmente durante el trabajo con simulaciones y el uso del software RStudio. Asimismo, se analizaron las respuestas obtenidas en el pretest y el postest, ya que estas permitieron reconocer dificultades, interpretaciones y posibles avances en relación con la comprensión de las pruebas de hipótesis.

Como parte de los instrumentos de recolección de información, se diseñó e implementó una secuencia didáctica centrada en el uso de simulaciones computacionales para abordar el aprendizaje del muestreo aleatorio simple, las distribuciones muestrales y las pruebas de hipótesis. Además, se aplicó un pretest y un postest con el propósito de identificar posibles cambios en el razonamiento estadístico de los estudiantes antes y después de la intervención.

De igual manera, se tomó como base una adaptación del Statistical Reasoning Assessment (SRA), por tratarse de un instrumento reconocido para evaluar habilidades de razonamiento estadístico e identificar concepciones erróneas frecuentes en estudiantes universitarios (J. B. Garfield, 2003). A partir de esta adaptación, se construyó un instrumento conformado por 7 preguntas de opción múltiple, orientadas a explorar la comprensión de conceptos asociados con la inferencia estadística y las pruebas de hipótesis.

En este sentido, el pretest y el postest fueron concebidos como instrumentos equivalentes, diseñados con fines comparativos para identificar tanto los conocimientos iniciales como los posibles avances de los estudiantes tras la implementación de la secuencia didáctica. A continuación, se presentan las tablas que describen la estructura y organización general de los instrumentos aplicados.

Con el fin de organizar el análisis de los resultados obtenidos en el pretest y el postest, se establecieron diferentes categorías de razonamiento correctas y concepciones erróneas a partir de la adaptación del Statistical Reasoning Assessment (SRA). Estas categorías permitieron clasificar las respuestas de los estudiantes de acuerdo con los conceptos fundamentales de las pruebas de hipótesis, facilitando así la comparación entre el desempeño inicial y final de los participantes.

Tabla 9: *Parámetros de análisis del pretest (adaptación del SRA)*

Parámetro de análisis (habilidad de razonamiento)	**Descripción**	**Ítems del pretest y respuesta correcta**
Identificación de la hipótesis nula y alternativa	Reconoce la formulación adecuada de H_0 y H_1.	1(c)
Comprensión del error tipo I	Identifica que el error tipo I ocurre cuando se rechaza H_0 siendo verdadera.	2(B)
Comprensión del error tipo II	Identifica que el error tipo II ocurre cuando no se rechaza H_0 siendo falsa.	3(A)
Aplicación de hipótesis en contexto	Evalúa la aplicación de las hipótesis en la interpretación de situaciones y la toma de decisiones inferenciales.	6(A)
Interpretación del valor-p y significancia práctica	Distingue entre significancia estadística y relevancia práctica, reconociendo que un valor-p pequeño no implica necesariamente un efecto importante.	4(A)
Comprensión del nivel de significancia α	Reconoce α como la probabilidad máxima de cometer un error tipo I.	5(A)
Interpretación de la decisión inferencial	Comprende que "no rechazar H_0" significa ausencia de evidencia suficiente para rechazarla, no aceptación de su veracidad.	7(A)

La tabla anterior presenta las habilidades de razonamiento estadístico consideradas en el pretest, organizadas a partir de la adaptación del SRA. Su propósito fue identificar el nivel inicial de comprensión de los estudiantes frente a conceptos fundamentales de las pruebas de hipótesis, tales como la formulación de hipótesis, la interpretación del valor-p, la comprensión de los errores tipo I y II, el nivel de significancia y la toma de decisiones inferenciales.

Tabla 10: *Parámetros de análisis del postest (adaptación del SRA)*

Parámetro de análisis (habilidad de razonamiento)	**Descripción**	**Ítems del postest y respuesta correcta**
Identificación de la hipótesis nula y alternativa	Evalúa de manera indirecta la coherencia en la toma de decisiones inferenciales.	4(B)
Comprensión del error tipo I	Identifica que el error tipo I ocurre cuando se rechaza H_0 siendo verdadera.	2(B)
Comprensión del error tipo II	Identifica que el error tipo II ocurre cuando no se rechaza H_0 siendo falsa.	3(B)
Aplicación de hipótesis en contexto	Evalúa la aplicación de las hipótesis en la interpretación de situaciones y la toma de decisiones inferenciales.	1(B)
Interpretación del valor-p	Reconoce el valor-p como la probabilidad de obtener un resultado tan extremo como el observado, o más, suponiendo que H_0 es verdadera.	5(B)
Comprensión del nivel de significancia α	Relaciona el nivel de significancia con la frecuencia esperada de rechazo de H_0 cuando esta es verdadera.	6(B)
Interpretación de la decisión inferencial	Comprende que "no rechazar H_0" implica falta de evidencia suficiente para rechazarla, no que H_0 haya sido aceptada.	7(B)

La tabla anterior resume las habilidades de razonamiento estadístico evaluadas en el postest, manteniendo criterios comparables con el pretest para favorecer el análisis de los posibles avances conceptuales de los estudiantes. Aunque algunos ítems del postest profundizan en aspectos más avanzados, como la potencia de una prueba y el carácter aleatorio del valor-p, las categorías se conservaron en función de su equivalencia conceptual con las habilidades inicialmente evaluadas.

Además de las habilidades de razonamiento correctas, se definieron categorías asociadas a las principales concepciones erróneas relacionadas con las pruebas de hipótesis, con el fin de identificar las dificultades conceptuales presentes en las respuestas de los estudiantes antes y después de la intervención.

Tabla 11: *Concepciones erróneas comunes para la triangulación (adaptación del SRA)*

Concepción errónea	**Descripción**	**Evidencia en el pretest**	**Evidencia en el postest**
Formular incorrectamente las hipótesis	Se ubica la desigualdad en H_0, se invierte la dirección de la prueba o se usa una relación inadecuada entre H_0 y H_1.	1(B), 1(C), 1(D), 6(B), 6(C), 6(D)	4(C)
Confundir error tipo I con error tipo II	No se distingue entre rechazar H_0 siendo verdadera y no rechazarla siendo falsa.	2(A), 2(C), 2(D), 3(B), 3(C), 3(D)	2(A), 2(C), 2(D), 3(A), 3(C), 3(D)
Interpretar erróneamente el valor-p	Se asume como probabilidad de que H_0 sea verdadera, como evidencia absoluta o como valor fijo.	4(B), 4(C), 4(D)	4(A), 4(C), 4(D), 5(A), 5(C), 5(D)
Confundir el nivel de significancia α con otros conceptos	Se interpreta α como valor-p, nivel de confianza, tamaño muestral o relevancia práctica.	5(B), 5(C), 5(D)	6(A), 6(C), 6(D)
Interpretar "no rechazar H_0" como "aceptar H_0"	Se asume que la hipótesis nula ha sido demostrada como verdadera o que H_1 es falsa.	7(B), 7(C), 7(D)	7(A), 7(C), 7(D)
Desconocer el carácter aleatorio del valor-p	Se cree que el valor-p es fijo, independiente de la muestra o siempre indicador de significancia.	5(B)	1(A), 1(C), 1(D), 5(A), 5(C), 5(D)

La tabla anterior presenta las principales concepciones erróneas asociadas a la comprensión de las pruebas de hipótesis, identificadas a partir de la adaptación del instrumento SRA. Estas categorías se utilizaron como referencia para analizar las respuestas de los estudiantes tanto en el pretest como en el postest, permitiendo reconocer las dificultades conceptuales iniciales y examinar su posible permanencia, disminución o transformación después de la implementación de la secuencia didáctica.

Fase 5: Valoración del instrumento

La valoración de instrumento se realizo con el fin de garantizar la fiabilidad de su contenido y su relevancia para medir la dificultades conceptuales de los estudiantes en inferencia estadística. Los instrumentos de recolección de datos fueron sometidos a un proceso de valoración con el fin de garantizar su validez y pertinencia en relación con los objetivos de la investigación.

En cuanto a la validez de contenido, el pretest y el postest fueron diseñados considerando los conceptos fundamentales de la inferencia estadística, tales como el muestreo, las pruebas de hipótesis, el nivel de significancia y el valor p, asegurando la correspondencia entre los ítems y los contenidos a evaluar.

Adicionalmente, los instrumentos fueron revisados por un experto en estadística, quien evaluó la pertinencia, claridad y coherencia de los ítems en relación con los conceptos de inferencia estadística. A partir de las observaciones realizadas, se efectuaron ajustes que permitieron mejorar la calidad y precisión de los instrumentos.

Fase 6: Procedimiento e implementación de la intervención didáctica

El desarrollo de la investigación se llevó a cabo mediante un procedimiento estructurado en tres fases: diagnóstico inicial (pretest), implementación de la intervención didáctica y evaluación final de los aprendizajes (postest).

En la primera fase, se aplicó un pretest (ver anexo 6) con el propósito de identificar los conocimientos previos de los estudiantes en relación con los conceptos de inferencia estadística, tales como el muestreo aleatorio simple, las pruebas de hipótesis, el nivel de significancia, los errores tipo I y tipo II y el valor p.

En la segunda fase, se desarrolló la implementación de la intervención didáctica, la cual se llevó a cabo en un grupo de estudiantes de Tópicos en estadística. Esta se desarrolló a lo largo de una sesión, con una duración aproximada de dos horas.

Durante esta fase, se implementó la secuencia didáctica previamente diseñada, orientada a la enseñanza de la inferencia estadística mediante el uso de simulaciones. En primer lugar, se abordó el concepto de Muestra Aleatoria Simple (MAS) desde una perspectiva conceptual, con el fin de establecer una base teórica sólida en los estudiantes. Posteriormente, se desarrolló una simulación (ver anexo 6) utilizando el software RStudio, lo que permitió a los estudiantes relacionar el concepto teórico con su aplicación práctica. De esta manera, se favoreció la comprensión del proceso de muestreo a través de la experimentación y el análisis de resultados. De igual manera, se abordaron los conceptos de pruebas de hipótesis, tipos de errores, nivel de significancia y valor p, los cuales fueron desarrollados tanto desde una perspectiva conceptual como mediante actividades prácticas apoyadas en simulaciones en RStudio (ver anexo 6)

Las actividades propuestas permitieron a los estudiantes participar activamente en la generación de datos y la interpretación de resultados, favoreciendo la comprensión de los conceptos abordados.

Finalmente, en la tercera fase se aplico la prueba postest (ver anexo 6), con el objetivo de evaluar los aprendizajes alcanzados por los estudiantes tras la intervención didáctica. Los resultados obtenidos fueron organizados y analizados, permitiendo establecer comparaciones entre el desempeño inicial y final, así como identificar cambios en los procesos de razonamiento de los estudiantes.

Fase 7: Procedimiento para el análisis de la intervención

Para el análisis de los resultados se realizó un enfoque mixto, integrando procedimientos cuantitativos y cualitativos con el fin de obtener una amplia interpretación de los resultados de la intervención didáctica.

En el componente cuantitativo, se llevó a cabo una comparación entre los resultados del pretest y el postest, considerando que ambos instrumentos fueron aplicados al mismo grupo de estudiantes en dos momentos distintos. En este sentido, los datos fueron tratados como muestras relacionadas o pareadas.

Teniendo en cuenta que el instrumento estuvo conformado por 7 preguntas de opción múltiple, los puntajes correspondieron a un conteo discreto de aciertos entre 0 y 7. En consecuencia, se empleó la prueba de rangos con signo de Wilcoxon, al tratarse de una prueba no paramétrica adecuada para comparar dos mediciones relacionadas (Hollander et al., 2013).

A partir de este análisis, se contrastó la hipótesis nula, que plantea que no existen diferencias estadísticamente significativas entre los resultados del pretest y el postest, frente a la hipótesis alternativa, que sostiene que sí existen diferencias estadísticamente significativas entre ambas mediciones. Este procedimiento permitió valorar si la intervención didáctica produjo cambios en el desempeño de los estudiantes en relación con el razonamiento estadístico asociado a las pruebas de hipótesis.

En el componente cualitativo, se analizaron los procesos de razonamiento de los estudiantes a partir de sus respuestas en el pretest y el postest, clasificándolos en categorías de razonamiento correcto y concepciones erróneas. Este análisis permitió identificar si hubo cambios en la comprensión conceptual de los estudiantes, especialmente en relación con la interpretación del valor p, el nivel de significancia y la toma de decisiones en pruebas de hipótesis.

En el marco del modelo de Reflexión y Acción (RyA) propuesto por Parada, 2016, para el análisis de los datos se consideró la fase de reflexión en la acción y sobre la acción como elementos clave para la interpretación de los resultados. En esta fase se examinó la participación de los estudiantes durante la intervención en el aula, asimismo, de manera posterior y sistemática las respuestas de los estudiantes en el pretest y el postest, identificando los cambios en sus procesos de razonamiento, así como la presencia de errores conceptuales y avances en la comprensión de los conceptos.

4. Resultado y análisis

4.1. Transformación del razonamiento en pruebas de hipótesis a partir de la simulación

En relación con la evolución de la comprensión de las pruebas de hipótesis, se evidenció que en el pretest los estudiantes presentaban dificultades en la interpretación de conceptos fundamentales como el error tipo I, el error tipo II, el valor p y la formulación de hipótesis. En particular, se observó confusión entre los tipos de error, así como dificultades para diferenciar entre "no rechazar H_0" y "aceptar H_0", lo que refleja la presencia de concepciones erróneas y una comprensión inicial de carácter más mecánico que conceptual.

Durante el desarrollo de las actividades apoyadas en simulaciones, se trabajó con un código implementado en *RStudio* (ver Anexo 6), el cual fue desarrollado de manera guiada, explicando paso a paso cada uno de sus componentes. Cabe resaltar que los estudiantes demostraron un buen manejo del software *RStudio*, lo cual facilitó el desarrollo de las actividades, permitió una mayor participación en la clase y favoreció la discusión colectiva de los resultados obtenidos.

El uso de la simulación permitió evidenciar la naturaleza del error tipo I y del error tipo II en contextos de muestreo repetido, así como cuestionar la interpretación común del valor p como una probabilidad directa de la hipótesis nula. A través de la observación de múltiples ejecuciones, los estudiantes pudieron reconocer que el valor p depende de la muestra y que los resultados de una prueba de hipótesis pueden variar debido al carácter aleatorio del proceso inferencial.

A partir del ejemplo basado en la distribución normal, se observó que el nivel de significancia se asume previamente y se utiliza como criterio para contrastar la hipótesis mediante el estadístico de prueba. El error tipo II no se fija de manera anticipada, sino que se calcula a partir del nivel de significancia seleccionado, lo que permite determinar la potencia de la prueba, definida como $1-\beta$. Este análisis posibilitó comprender que una prueba con mayor potencia tiene una mayor capacidad para detectar diferencias reales cuando la hipótesis nula es falsa.

Aunque en el postest se evidenció una mejora en la comprensión de estos conceptos, este aspecto será analizado con mayor detalle en la sección correspondiente. No obstante, los resultados obtenidos en esta fase permiten identificar un proceso de transición desde concepciones erróneas iniciales hacia interpretaciones más acordes con la lógica de la inferencia estadística. En este sentido, se refuerza la idea de que, en la práctica, cuando se dispone de varios estadísticos de prueba para contrastar una muestra, se prefiere aquella prueba que sea uniformemente más potente, es decir, la que presenta mayor potencia bajo las alternativas consideradas.

4.2. Análisis de la intervención didáctica desde el modelo RyA

Teniendo en cuenta que el análisis de la intervención didáctica se realizó a la luz del modelo de Reflexión y Acción (RyA) propuesto por Parada (2011), el cual se orienta hacia la reflexión del docente sobre su práctica pedagógica antes, durante y después de la intervención, se analizaron las decisiones didácticas implementadas y su incidencia en el desarrollo del razonamiento estadístico de los estudiantes. En este sentido, el modelo RyA permitió interpretar cómo los procesos de planificación, acción y reflexión docente favorecieron la comprensión de conceptos asociados a las pruebas de hipótesis a través del uso de simulaciones interactivas y actividades orientadas al análisis e interpretación de resultados estadísticos.

Reflexión antes de la acción

En un primer momento, previo al desarrollo de la intervención didáctica, se realizó un proceso de planificación y diseño de las actividades teniendo en cuenta las dificultades y concepciones erróneas identificadas en los estudiantes a partir del pretest. Entre las principales dificultades evidenciadas se encontraron errores relacionados con la interpretación del valor p, comprensión de pruebas de hipótesis, variabilidad muestral, toma de decisiones estadísticas.

Con base en lo anterior, se diseñó una secuencia de actividades orientada al fortalecimiento del razonamiento estadístico mediante explicaciones teóricas y simulaciones realizadas en Rstudio. Asimismo, se planearon diferentes códigos para que los estudiantes pudieran ejecutarlos, analizarlos y observar el comportamiento de distintos conceptos estadísticos abordados durante la intervención. Estas simulaciones permitieron representar situaciones relacionadas con el valor p , error tipo I y tipo II, tamaño de muestra y pruebas de hipótesis, favoreciendo una comprensión más visual e interactiva de los conceptos trabajados.

De igual manera, las actividades fueron estructuradas con el propósito de promover espacios de participación, análisis y reflexión, permitiendo que los estudiantes confrontaran las ideas y razonamientos manifestados inicialmente en el pretest. En este sentido, la intervención buscó aclarar algunas concepciones erróneas identificadas previamente, especialmente aquellas relacionadas con la interpretación del rechazo de la hipótesis nula, el significado del valor p y los diferentes tipos de errores.

Reflexión en la acción

Durante la fase de reflexión en la acción, se explicó a los estudiantes cada concepto de forma teórica, contrastándolo posteriormente, con las simulaciones desarrolladas en Rstudio. Se pudo evidenciar que, al interactuar con las simulaciones luego de estudiar la parte conceptual, los estudiantes comenzaban a cuestionar algunas de las afirmaciones realizadas inicialmente en el pretest.

Desde el momento en que se explicó el concepto de muestra aleatoria simple hasta el estudio del valor p, los estudiantes se mostraron activos y participativos durante las clases, realizando preguntas como: “¿Entre más grande la muestra, siempre es mejor?” o “¿Rechazar la hipótesis quiere decir que es falsa?”. Asimismo, durante el desarrollo de las simulaciones cuestionaban situaciones como: “¿Por qué en la simulación salen resultados diferentes cada vez?”

Teniendo en cuenta lo anterior, se evidencian procesos de reflexión en la acción, ya que los estudiantes analizaban y cuestionaban aspectos relacionados con la inferencia estadística y la variabilidad presente en los datos. Esto permitió abordar la naturaleza aleatoria del muestreo y la variabilidad inherente a los datos, favoreciendo la comprensión de que los resultados pueden cambiar debido al azar de la muestra, incluso bajo las mismas condiciones.

Durante la clase se llevó a cabo la siguiente ruta cognitiva.

Figura 6: *Ruta cognitiva*

Nota. Elaboración propia.

Reflexión sobre la acción

En la fase de reflexión sobre la acción, se evidenció una evolución significativa en los procesos de razonamiento estadístico de los estudiantes, manifestada en su capacidad para interpretar adecuadamente el nivel de significancia y emplear el valor p como criterio para la toma de decisiones en pruebas de hipótesis.

Asimismo, se observó una disminución en los errores conceptuales identificados inicialmente en el pretest, especialmente en aspectos relacionados con la interpretación de hipótesis, error tipo I y tipo II y comprensión de resultados estadísticos, lo que da cuenta de un proceso de consolidación de los aprendizajes promovidos durante la intervención didáctica. De igual manera, los estudiantes mostraron una mayor capacidad para justificar sus respuestas y argumentar sus decisiones a partir de los resultados obtenidos en las actividades y simulaciones desarrolladas durante la intervención.

No obstante, desde esta fase de reflexión también se identificaron algunos aspectos susceptibles de fortalecimiento en futuras implementaciones de la propuesta. En particular, se considera pertinente desarrollar la intervención con un grupo más amplio de estudiantes, ya que ello permitiría obtener una mayor diversidad de producciones y razonamientos estadísticos, favoreciendo un análisis más robusto de los procesos de aprendizaje y del impacto de las simulaciones en la interpretación de las pruebas de hipótesis.

Figura 7: *Modelo RyA*

Nota. Elaboración propia.

4.3. Resultados del instrumento y producciones de los estudiantes

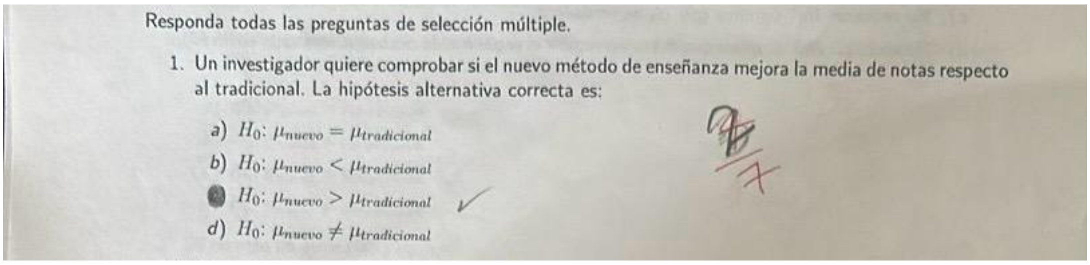
En esta sección se presentan y analizan los resultados obtenidos a partir de la aplicación del pretest y el postest, considerando la adaptación del Statistical Reasoning Assessment (SRA), así como las producciones elaboradas por los estudiantes durante la intervención didáctica. El análisis se orienta a identificar habilidades de razonamiento estadístico, dificultades conceptuales y concepciones erróneas relacionadas con las pruebas de hipótesis, especialmente en torno a la formulación de hipótesis, la interpretación del valor p, la comprensión del nivel de significancia, los errores tipo I y tipo II, y la toma de decisiones inferenciales.

Asimismo, se establece una comparación entre las ideas previas evidenciadas en el pretest y las respuestas obtenidas en el postest, con el propósito de analizar posibles transformaciones en el razonamiento estadístico de los estudiantes después de la implementación de la propuesta didáctica basada en simulaciones interactivas.

Además se realizó un análisis de algunas producciones escritas de los estudiantes con el propósito de identificar formas de razonamiento, dificultades conceptuales y concepciones erróneas relacionadas con las pruebas de hipótesis.

En la primera producción analizada la cual corresponde al estudiante 3 de la figura 15 se observa que el estudiante respondió correctamente la pregunta relacionada con la formulación de hipótesis estadísticas (ver figura **??**), identificando adecuadamente que la hipótesis alternativa debía representar una mejora en el nuevo método de enseñanza. Asimismo, logró interpretar correctamente la diferencia entre significancia estadística y relevancia práctica en la pregunta relacionada con el tamaño del efecto.(ver figura 10)

Figura 8: *Producción estudiante 3 pregunta 1*

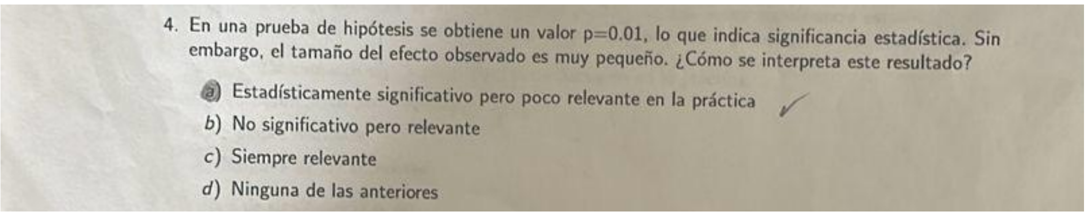

Responda todas las preguntas de selección múltiple.

1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis alternativa correcta es:
   a) H_0: $\mu_{nuevo} = \mu_{tradicional}$
   b) H_0: $\mu_{nuevo} < \mu_{tradicional}$
   c) H_0: $\mu_{nuevo} > \mu_{tradicional}$
   d) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$

Nota. Elaboración propia.

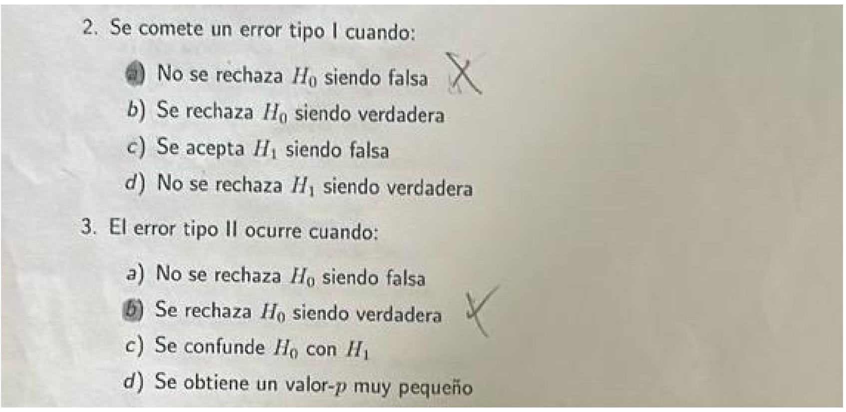
Figura 9: */*

Producción estudiante 3 pregunta 4


4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?
   a) Estadísticamente significativo pero poco relevante en la práctica
   b) No significativo pero relevante
   c) Siempre relevante
   d) Ninguna de las anteriores

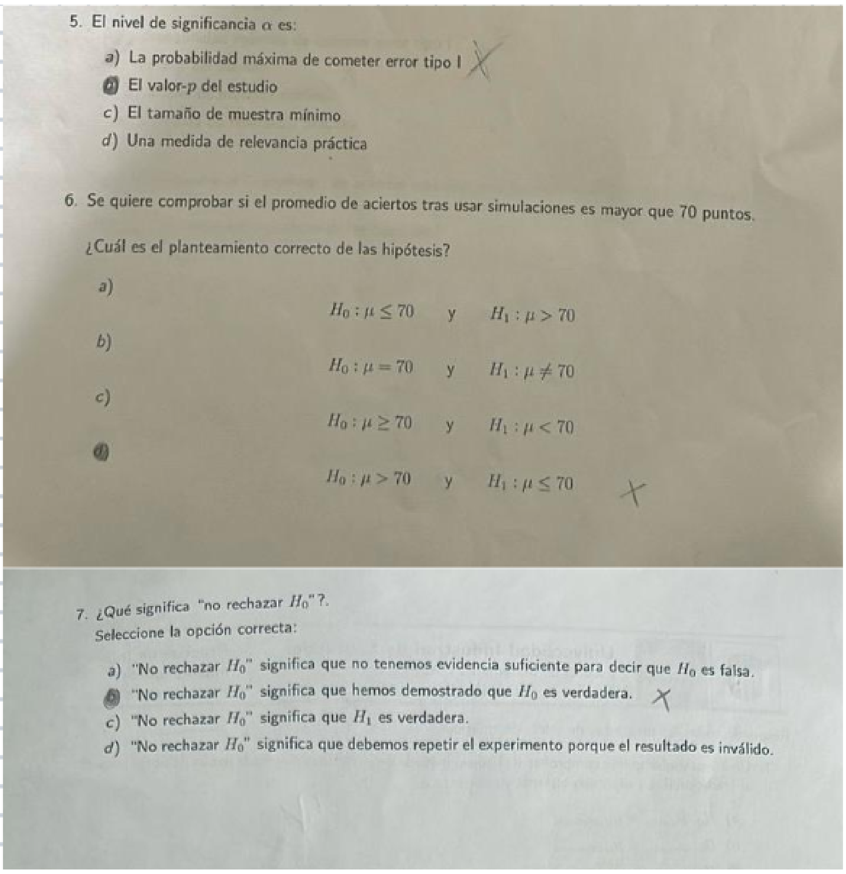
No obstante, como se puede evidenciar en la figura 9 el estudiante presentó dificultades importantes en conceptos inferenciales asociados a los errores tipo I y tipo II, evidenciando confusión entre las condiciones bajo las cuales ocurre cada uno de estos errores.

Figura 10: *Producción estudiante 3 pregunta 2 y 3.*


2. Se comete un error tipo I cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. El error tipo II ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se confunde H_0 con H_1
   d) Se obtiene un valor-p muy pequeño

De igual manera, se identificó una interpretación incorrecta del nivel de significancia α, al asociarlo con el valor p del estudio y no con la probabilidad máxima de cometer un error tipo I. Asimismo, en la formulación de hipótesis contextualizadas el estudiante presentó dificultades para establecer correctamente la relación entre H_0 y H_1 (ver figura 11)

Figura 11: *Producción estudiante 3 pregunta 5, 6, 7.*



5. El nivel de significancia α es:

a) La probabilidad máxima de cometer error tipo I
b) El valor-p del estudio
c) El tamaño de muestra mínimo
d) Una medida de relevancia práctica

6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

¿Cuál es el planteamiento correcto de las hipótesis?

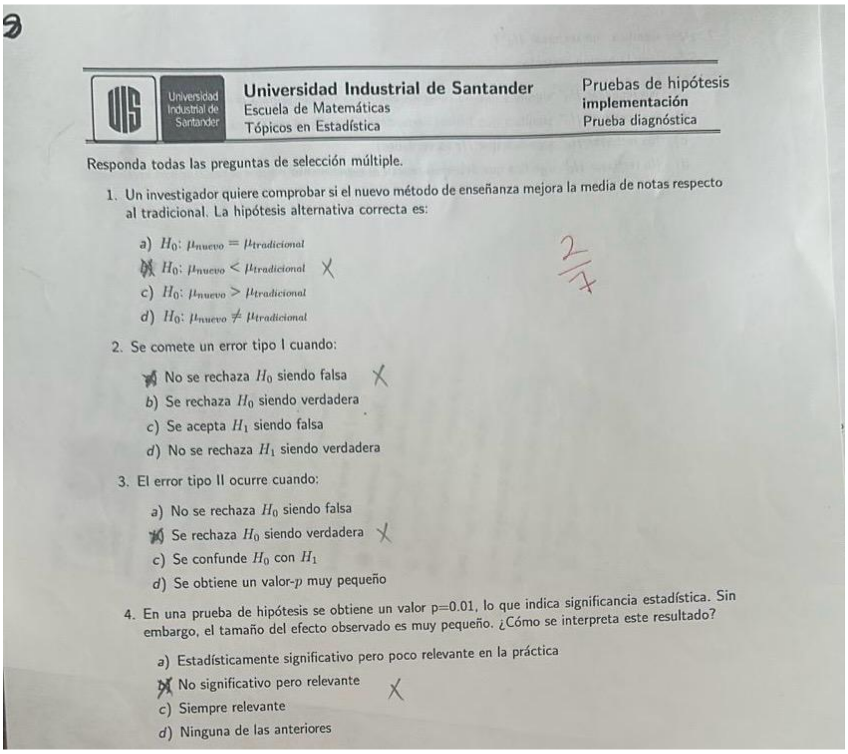
a) $H_0 : \mu \leq 70$ y $H_1 : \mu > 70$
b) $H_0 : \mu = 70$ y $H_1 : \mu \neq 70$
c) $H_0 : \mu \geq 70$ y $H_1 : \mu < 70$
d) $H_0 : \mu > 70$ y $H_1 : \mu \leq 70$

7. ¿Qué significa "no rechazar H_0"?.
Seleccione la opción correcta:

a) "No rechazar H_0" significa que no tenemos evidencia suficiente para decir que H_0 es falsa.
b) "No rechazar H_0" significa que hemos demostrado que H_0 es verdadera.
c) "No rechazar H_0" significa que H_1 es verdadera.
d) "No rechazar H_0" significa que debemos repetir el experimento porque el resultado es inválido.

Por otra parte, la segunda producción la cual corresponde al estudiante 8 evidencia concepciones erróneas similares. El estudiante presentó dificultades en la comprensión de los errores inferenciales, confundiendo nuevamente el error tipo I con el error tipo II. Además, en la pregunta relacionada con la interpretación del valor p, el estudiante no logró diferenciar entre significancia estadística y relevancia práctica, lo cual coincide con las dificultades reportadas en la literatura sobre educación estadística (Castro Sotos et al., 2007)

Figura 12: *Producción estudiante 8 pregunta 1, 2, 3, 4.*



9

Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Prueba diagnóstica

Responda todas las preguntas de selección múltiple.

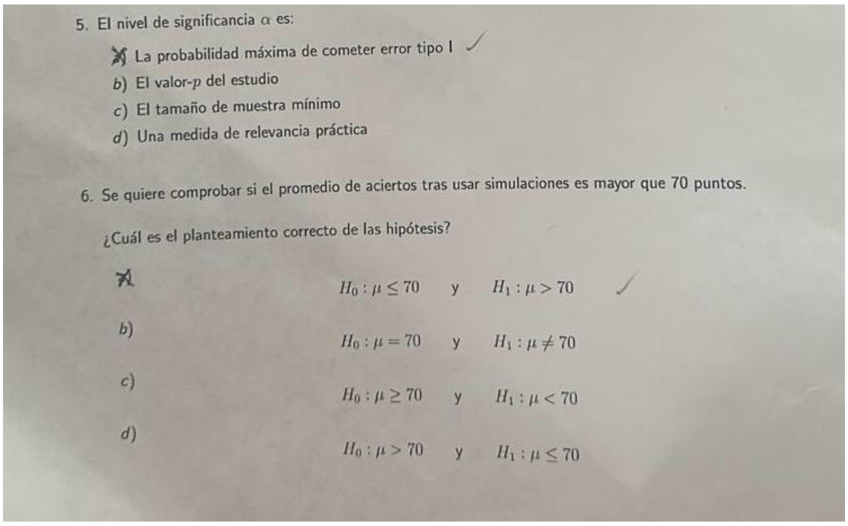
1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis alternativa correcta es:
   a) H_0: $\mu_{nuevo} = \mu_{tradicional}$
   b) H_0: $\mu_{nuevo} < \mu_{tradicional}$ X
   c) H_0: $\mu_{nuevo} > \mu_{tradicional}$
   d) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$

2/7

2. Se comete un error tipo I cuando:
   a) No se rechaza H_0 siendo falsa X
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. El error tipo II ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera X
   c) Se confunde H_0 con H_1
   d) Se obtiene un valor-p muy pequeño
4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?
   a) Estadísticamente significativo pero poco relevante en la práctica
   b) No significativo pero relevante X
   c) Siempre relevante
   d) Ninguna de las anteriores

Sin embargo, esta producción también muestra algunos elementos de razonamiento parcialmente correctos. Por ejemplo, el estudiante identificó adecuadamente el nivel de significancia α como la probabilidad máxima de cometer un error tipo I y logró formular correctamente las hipótesis estadísticas en una situación contextualizada. Esto sugiere que, aunque existían dificultades conceptuales importantes, algunos estudiantes comenzaban a establecer relaciones adecuadas entre ciertos conceptos inferenciales. (ver figura 13).

Figura 13: *Producción estudiante 8 pregunta 5,6.*


5. El nivel de significancia α es:

a) La probabilidad máxima de cometer error tipo I

b) El valor-p del estudio

c) El tamaño de muestra mínimo

d) Una medida de relevancia práctica

6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

¿Cuál es el planteamiento correcto de las hipótesis?

a) $H_0 : \mu \leq 70$ y $H_1 : \mu > 70$

b) $H_0 : \mu = 70$ y $H_1 : \mu \neq 70$

c) $H_0 : \mu \geq 70$ y $H_1 : \mu < 70$

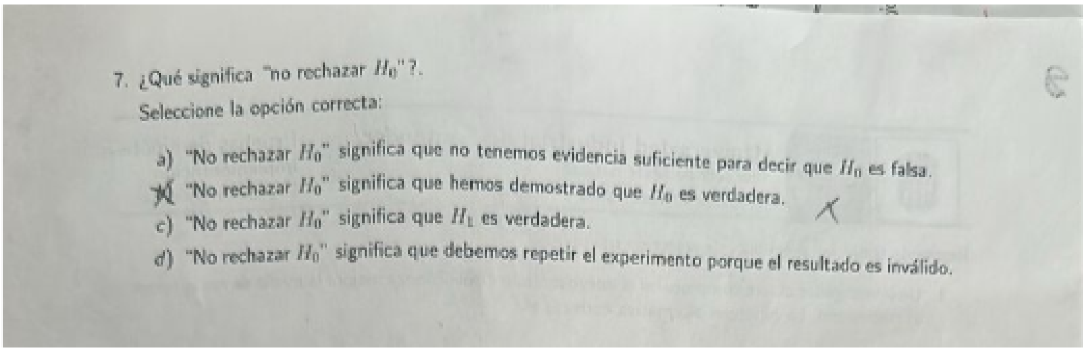
d) $H_0 : \mu > 70$ y $H_1 : \mu \leq 70$

.

Además, en ambas producciones se identificó una dificultad relacionada con la interpretación de la decisión inferencial asociada a "no rechazar H_0". En la pregunta 7, los estudiantes seleccionaron la opción que interpretaba "no rechazar la hipótesis nula" como una demostración de que H_0 era verdadera, en lugar de reconocer que esta decisión únicamente indica que no existe evidencia estadística suficiente para rechazarla. (ver figura 15)

Figura 14: *Producción estudiante 8 pregunta 7*

[Producción estudiante 8 pregunta 7]


7. ¿Qué significa "no rechazar H_0"?.

Seleccione la opción correcta:

a) "No rechazar H_0" significa que no tenemos evidencia suficiente para decir que H_0 es falsa.

b) "No rechazar H_0" significa que hemos demostrado que H_0 es verdadera.

c) "No rechazar H_0" significa que H_1 es verdadera.

d) "No rechazar H_0" significa que debemos repetir el experimento porque el resultado es inválido.

.

Estos hallazgos son coherentes con lo reportado por Batanero (2001b), J. Garfield y Ben-Zvi (2008a) y Inzunsa Cazares y Jiménez Ramírez (2013), quienes señalan que muchos estudiantes abordan las pruebas de hipótesis desde una perspectiva algorítmica y mecánica, sin comprender plenamente la lógica inferencial subyacente.

En general los resultados del pretest evidenciaron que los estudiantes presentaban dificultades importantes en la comprensión de conceptos fundamentales asociados a la inferencia estadística y a las pruebas de hipótesis. En particular, se identificaron concepciones erróneas relacionadas con la formulación de hipótesis, la comprensión de los errores tipo I y tipo II, la interpretación del valor p, el significado del nivel de significancia α y la interpretación de la decisión de "no rechazar H_0" Estas dificultades coinciden con las reportadas en la literatura sobre educación estadística, en donde se señala que muchos estudiantes tienden a interpretar las pruebas de hipótesis de manera algorítmica, sin comprender plenamente la lógica inferencial subyacente.

En relación con la habilidad de razonamiento asociada a la identificación de la hipótesis nula y alternativa, evaluada en las preguntas 1 y 6, se esperaba que los estudiantes reconocieran la formulación adecuada de las hipótesis estadísticas en contextos inferenciales. En la pregunta 1, relacionada con la comparación entre un método de enseñanza tradicional y uno nuevo, únicamente cerca de la mitad de los estudiantes respondió correctamente que la hipótesis nula debía expresar igualdad entre las medias poblacionales (H_1: μ nuevo $= \mu$ tradicional). De manera similar, en la pregunta 6, orientada a la formulación de hipótesis en una prueba unilateral, varios estudiantes presentaron dificultades para establecer correctamente la relación entre H_0 y H_1.

A partir de las respuestas obtenidas, se identificó como principal dificultad la formulación incorrecta de las hipótesis, particularmente la tendencia a invertir el sentido de la prueba o a considerar que la hipótesis nula debía representar directamente la afirmación que se deseaba demostrar. Algunos estudiantes seleccionaron opciones en las cuales H_0 se formulaba como "mayor que" o "menor que", evidenciando una comprensión limitada del papel que desempeña la hipótesis nula dentro del razonamiento inferencial.

Respecto a la comprensión de los errores tipo I y tipo II, evaluada en las preguntas 2 y 3, se observó que varios estudiantes tendían a confundir ambos conceptos. Aunque algunos reconocieron correctamente que el error tipo I ocurre cuando se rechaza H_0 siendo verdadera, persistieron respuestas que evidencian confusión entre rechazar y no rechazar la hipótesis nula. Del mismo modo, en la pregunta relacionada con el error tipo II, algunos estudiantes no lograron identificar que este ocurre cuando no se rechaza H_0 siendo falsa.

Estas respuestas permiten identificar una dificultad asociada a la interpretación de las decisiones inferenciales bajo incertidumbre. En particular, los estudiantes mostraron problemas para diferenciar adecuadamente las condiciones bajo las cuales se producen los errores tipo I y tipo II, lo cual evidencia una comprensión parcial de la lógica probabilística implicada en las pruebas de hipótesis.

En cuanto a la interpretación del valor p y la significancia práctica, evaluada mediante la pregunta 4, se esperaba que los estudiantes reconocieran que un resultado estadísticamente significativo no implica necesariamente una relevancia práctica importante. Sin embargo, algunos estudiantes interpretaron que un valor p pequeño garantiza automáticamente la importancia o magnitud del efecto observado.

Esta situación evidencia una interpretación errónea del valor p, asociada a la idea de que la significancia estadística equivale a relevancia práctica. Asimismo, durante el análisis de las respuestas se identificaron dificultades relacionadas con el carácter aleatorio del valor p, ya que algunos estudiantes parecían concebirlo como una medida fija o determinista, sin reconocer que este depende de la variabilidad muestral y del comportamiento aleatorio de los datos.

En relación con la comprensión del nivel de significancia α, evaluada en la pregunta 5, únicamente dos de los ocho estudiantes seleccionaron correctamente que el nivel de significancia corresponde a la probabilidad máxima de cometer un error tipo I. La mayoría de los estudiantes confundió α con otros conceptos estadísticos, particularmente con el valor p.

Esta dificultad evidencia una comprensión inadecuada del significado probabilístico del nivel de significancia. En particular, los estudiantes mostraron dificultades para interpretar α como un criterio previo de decisión asociado al riesgo de rechazar incorrectamente la hipótesis nula.

Por otra parte, en la pregunta 7, relacionada con la interpretación de la decisión inferencial, se esperaba que los estudiantes reconocieran que "no rechazar H_0" significa únicamente que no existe evidencia suficiente para rechazarla. No obstante, varios estudiantes asumieron incorrectamente que no rechazar la hipótesis nula implica aceptar que H_0 es verdadera o concluir automáticamente que la hipótesis alternativa es falsa.

Esta situación evidencia una de las concepciones erróneas más frecuentes en la enseñanza de las pruebas de hipótesis: interpretar la decisión inferencial desde una lógica determinista y no probabilística. En consecuencia, los estudiantes presentaron dificultades para comprender que las conclusiones inferenciales están sujetas a incertidumbre y dependen de la evidencia proporcionada por la muestra.

En conjunto, los resultados del pretest permitieron identificar diversas dificultades conceptuales y habilidades de razonamiento parcialmente desarrolladas en torno a las pruebas de hipótesis. Estas dificultades se relacionan principalmente con la formulación incorrecta de hipótesis, la confusión entre errores tipo I y tipo II, la interpretación errónea del valor p, la confusión del nivel de significancia α con otros conceptos y la interpretación de "no rechazar H_0" como "aceptar H_0". Dichos resultados constituyen un punto de partida relevante para el desarrollo de la intervención didáctica basada en simulaciones interactivas.

Figura 15: *Resultados del pretest*

[Resultados del pretest]

Estudiante	Pregunta1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta5	Pregunta6	Pregunta7
Estudiante 1	C	B	A	B	D	A	B
Estudiante 2	C	B	A	B	B	D	B
Estudiante 3	C	A	B	A	B	D	B
Estudiante 4	C	B	A	A	D	D	A
Estudiante 5	B	C	A	A	A	A	B
Estudiante 6	B	B	A	A	B	D	B
Estudiante 7	B	B	C	A	B	A	B
Estudiante 8	B	A	B	B	A	A	B

Nota: Las celdas en color verde representan respuestas correctas, mientras que las celdas en color rojo indican respuestas incorrectas.. .

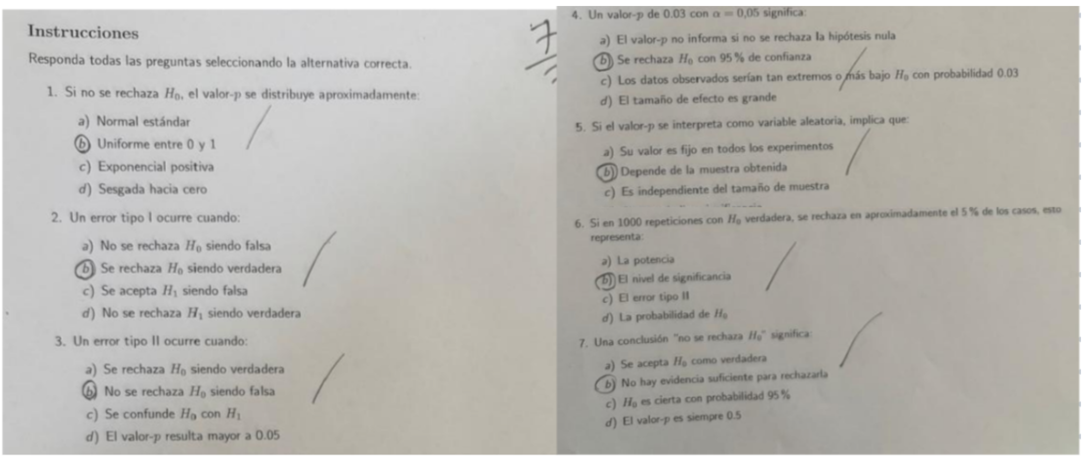
Luego, con el propósito de analizar el desarrollo de las habilidades de razonamiento estadístico asociadas a las pruebas de hipótesis, se aplicó un postest posterior a la implementación de la propuesta didáctica basada en simulaciones interactivas. Este instrumento permitió identificar los cambios en la comprensión conceptual de los estudiantes respecto a elementos inferenciales como el valor p, el nivel de significancia, los errores tipo I y II, la formulación de hipótesis y la toma de decisiones estadísticas bajo incertidumbre.

Se realizó un análisis en particular de los estudiantes 3 y 8 con el propósito de identificar avances en los procesos de razonamiento estadístico después de la implementación de la propuesta didáctica basada en simulaciones interactivas. Este análisis permitió evidenciar cambios en la comprensión conceptual de las pruebas de hipótesis, particularmente en relación con el valor p, los errores inferenciales, el nivel de significancia y la interpretación de las decisiones estadísticas.

En la primera producción analizada la cual corresponde al estudiante 3 se observa un desempeño favorable en la mayoría de las habilidades de razonamiento evaluadas. El estudiante identificó correctamente que, bajo la hipótesis nula, el valor p se distribuye aproximadamente de manera uniforme entre 0 y 1, lo cual evidencia comprensión del carácter aleatorio del valor p. Asimismo, logró diferenciar adecuadamente el error tipo I y el error tipo II, reconociendo las condiciones bajo las cuales ocurre cada uno de estos errores inferenciales.

De igual manera, el estudiante interpretó correctamente un valor p =0.03 con α =0.05, identificando que dicho valor representa la probabilidad de obtener datos tan extremos o más extremos que los observados bajo el supuesto de que H_0 es verdadera. Este resultado evidencia avances en la comprensión de la lógica inferencial asociada al valor p, superando interpretaciones erróneas frecuentes identificadas en el pretest. Además, el estudiante reconoció correctamente que el valor p, al interpretarse como variable aleatoria, depende de la muestra obtenida, mostrando comprensión de la variabilidad inherente al muestreo y al razonamiento estadístico. Asimismo, en la pregunta relacionada con simulaciones repetidas bajo H_0, identificó adecuadamente que rechazar aproximadamente el 5 % de las veces corresponde al nivel de significancia α, evidenciando comprensión de la frecuencia esperada del error tipo I.

Figura 16: *Producción estudiante 3 postest*



Instrucciones

Responda todas las preguntas seleccionando la alternativa correcta.

7/7

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

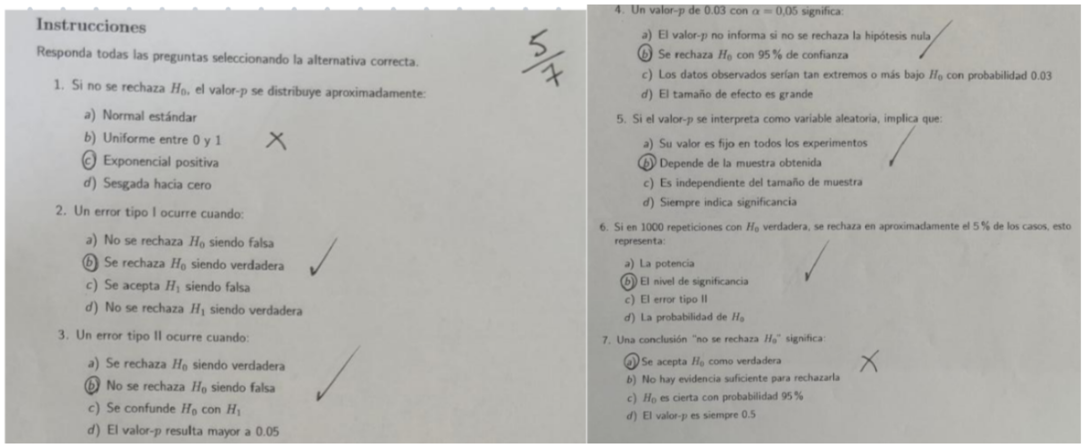
Nota: Elaboración propia.

Por otra parte, la producción del estudiante 8 evidencia también avances importantes en comparación con los resultados obtenidos en el pretest. El estudiante logró identificar correctamente el error tipo I y el error tipo II, mostrando una mejor comprensión de las decisiones inferenciales asociadas a las pruebas de hipótesis. Asimismo, reconoció adecuadamente que el valor p depende de la muestra obtenida, lo que evidencia avances en la comprensión de la variabilidad inherente al muestreo y del carácter aleatorio del valor p. De igual manera, comprendió correctamente el significado del nivel de significancia α a partir de situaciones de simulación repetida, relacionándolo con la frecuencia esperada de cometer un error tipo I cuando la hipótesis nula es verdadera.

Sin embargo, el estudiante presentó dificultades en la interpretación de la conclusión inferencial "no rechazar H_0", seleccionando nuevamente la opción asociada a aceptar la hipótesis nula como verdadera. Asimismo, aunque reconoció parcialmente la variabilidad del valor p, todavía se evidencian dificultades para comprender completamente sus implicaciones dentro de la lógica inferencial. Estas respuestas sugieren que, pese a los avances logrados durante la intervención, algunos estudiantes continúan interpretando las decisiones estadísticas desde una perspectiva determinista, atribuyendo a las pruebas de hipótesis un carácter de verificación absoluta más que de toma de decisiones bajo incertidumbre.ver figura 18.

Figura 17: *Producción estudiante 8 postest*

[Producción estudiante 8 postest]


Instrucciones

Responda todas las preguntas seleccionando la alternativa correcta.

5/7

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1 X
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera ✓
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa ✓
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza ✓
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida ✓
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia ✓
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera X
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

*Nota:*Elaboración propia. .

La Figura 19 presenta las respuestas obtenidas por los estudiantes en el postest. Las celdas en color verde representan respuestas correctas, mientras que las celdas en color rojo indican respuestas incorrectas. En términos generales, los resultados evidencian un aumento significativo en el número de respuestas correctas en comparación con el pretest, lo cual sugiere avances en la comprensión de conceptos relacionados con las pruebas de hipótesis y una disminución de las concepciones erróneas identificadas inicialmente.

En relación con la comprensión de los errores inferenciales, los resultados obtenidos en las preguntas 2 y 3 muestran que la mayoría de los estudiantes logró diferenciar adecuadamente el error tipo I del error tipo II. En particular, los estudiantes reconocieron que un error tipo I ocurre cuando se rechaza H_0 siendo verdadera, mientras que el error tipo II ocurre cuando no se rechaza H_0 siendo falsa. Este avance resulta significativo si se compara con los resultados del pretest, donde se evidenciaba una marcada confusión entre ambos conceptos.

Por otra parte, las preguntas 4 y 5 permitieron analizar la comprensión del valor p y su carácter aleatorio. En la pregunta 4, la mayoría de los estudiantes logró identificar que un valor p = 0.03 representa la probabilidad de obtener resultados tan extremos como los observados, o más extremos, bajo el supuesto de que H_0 es verdadera. Este resultado evidencia una disminución de la interpretación errónea del valor p como la probabilidad de que la hipótesis nula sea verdadera, dificultad ampliamente documentada en investigaciones sobre educación estadística (Castro Sotos et al., 2007).

Asimismo, en la pregunta 5 los estudiantes identificaron que el valor p depende de la muestra obtenida y, por tanto, puede variar entre diferentes repeticiones de un experimento. Esta comprensión resulta relevante desde el razonamiento inferencial, debido a que permite reconocer el carácter aleatorio y variable de los resultados estadísticos. Sin embargo, aunque la mayoría de los estudiantes respondió correctamente, aún persisten algunas dificultades en torno a la comprensión profunda de la distribución del valor p, especialmente en situaciones relacionadas con su interpretación probabilística.

En cuanto al nivel de significancia, los resultados de la pregunta 6 muestran que los estudiantes lograron asociar correctamente el rechazo de H_0 en aproximadamente el 5 % de los casos, cuando H_0 es verdadera, con el nivel de significancia α. Este resultado sugiere un avance desde una comprensión exclusivamente memorística del concepto hacia una interpretación frecuentista relacionada con la repetición de experimentos y el análisis de la variabilidad muestral. Este aspecto fue trabajado durante la intervención mediante simulaciones repetidas, en las cuales los estudiantes observaron el comportamiento de las decisiones inferenciales en múltiples muestras.

Por otra parte, la pregunta 7 permitió analizar la interpretación de la decisión inferencial "no rechazar H_0". Los resultados evidencian que la mayoría de los estudiantes comprendió que no rechazar la hipótesis nula no implica aceptarla como verdadera, sino únicamente reconocer que no existe evidencia suficiente para rechazarla. Este resultado representa un avance importante respecto al pretest, donde varios estudiantes asociaban incorrectamente "no rechazar" con "aceptar" H_0. No obstante, algunos estudiantes continuaron presentando dificultades en este aspecto, lo que evidencia que la lógica inferencial asociada a la toma de decisiones estadísticas sigue siendo compleja.

De manera particular, los estudiantes 6, 7 y 8 presentaron respuestas incorrectas en las preguntas 1 y 7. En el caso de la pregunta 1, relacionada con la distribución del valor p cuando no se rechaza H_0, las respuestas incorrectas sugieren dificultades para comprender el carácter aleatorio y probabilístico del valor P. De igual manera, las respuestas obtenidas en la pregunta 7 muestran que aún persisten algunas concepciones erróneas relacionadas con la interpretación de la decisión inferencial.

Figura 18: *Resultados del postest.*

Estudiante	Pregunta1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta5	Pregunta6	Pregunta7
Estudiante 1	B	B	B	B	B	B	B
Estudiante 2	B	B	B	B	B	B	B
Estudiante 3	B	B	B	B	B	B	B
Estudiante 4	B	B	B	B	B	B	B
Estudiante 5	B	B	B	B	B	B	B
Estudiante 6	C	B	B	B	B	B	A
Estudiante 7	D	B	B	B	B	B	A
Estudiante 8	D	B	B	B	B	B	A

*Nota:*Las celdas en color verde representan respuestas correctas, mientras que las celdas en color rojo indican respuestas incorrectas. .

Desde la perspectiva del Statistical Reasoning Assessment (SRA), estos resultados permiten identificar una evolución en las formas de razonamiento estadístico desarrolladas por los estudiantes. En comparación con el pretest, se observa una transición desde razonamientos intuitivos y procedimentales hacia formas de razonamiento más estructuradas y conceptualmente fundamentadas. En particular, se evidencia una mejora en habilidades relacionadas con la interpretación de evidencia estadística, la comprensión de la variabilidad y la toma de decisiones inferenciales bajo incertidumbre.

Asimismo, desde el modelo Reflexión y Acción (RyA), los cambios observados pueden interpretarse como resultado de la articulación entre los momentos de acción y reflexión desarrollados durante la intervención didáctica. Las simulaciones interactivas permitieron a los estudiantes experimentar con diferentes escenarios inferenciales, mientras que los espacios de discusión y análisis favorecieron procesos de reflexión sobre el significado de los resultados obtenidos y sobre los conceptos estadísticos involucrados.

En este sentido, el análisis del postest evidencia que la propuesta didáctica basada en simulaciones interactivas contribuyó al fortalecimiento de habilidades de razonamiento estadístico asociadas a las pruebas de hipótesis, especialmente en relación con la comprensión del valor p, los errores inferenciales, el nivel de significancia y la toma de decisiones bajo incertidumbre.

Con el fin de realizar una interpretación más integral del proceso de aprendizaje desarrollado por los estudiantes, se llevó a cabo una triangulación entre los resultados del pretest, las evidencias obtenidas durante la intervención didáctica y los resultados del postest. Este proceso permitió analizar no solo el desempeño final de los estudiantes, sino también la evolución de las habilidades de razonamiento estadístico asociadas a las pruebas de hipótesis como se evidencia en la tabla 12.

Tabla 12: *Matriz de triangulación entre pretest, intervención didáctica y postest (adaptación del SRA)*

Parámetro de análisis	**Evidencia en el pretest**	**Evidencia en la intervención didáctica**	**Evidencia en el postest**	**Propósito de la triangulación**
Identificación de H_0 y H_1	Ítems 1	Actividades de formulación de hipótesis, discusión de pruebas unilaterales y bilaterales, uso de simulaciones y contextualización de problemas.	Ítem 4	Identificar si las dificultades iniciales en la formulación de hipótesis fueron abordadas durante la intervención, aunque no se evalúen explícitamente en el postest.
Comprensión del error tipo I	Ítem 2	Simulaciones sobre rechazo de H_0 cuando es verdadera, discusión de decisiones erróneas y análisis de ejemplos.	Ítem 2	Comparar si disminuye la confusión entre el error tipo I y otras decisiones inferenciales.
Comprensión del error tipo II	Ítem 3	Simulaciones y actividades donde se analiza el no rechazo de H_0 siendo falsa.	Ítem 3	Evaluar si mejora la distinción entre error tipo II y error tipo I.
Aplicación de hipótesis en contexto	Ítem 6	Actividades de análisis de situaciones problemáticas, toma de decisiones y discusión de resultados en contextos aplicados.	Ítems 1	Evaluar si los estudiantes logran aplicar correctamente la formulación de hipótesis en situaciones contextualizadas y en la toma de decisiones estadísticas.
Interpretación del valor-p	Ítem 4	Actividades de simulación, análisis de resultados, discusión sobre evidencia estadística y comparación con la relevancia práctica.	Ítems 5	Analizar si los estudiantes logran una interpretación más rigurosa del valor-p y su carácter aleatorio.
Comprensión del nivel de significancia α	Ítem 5	Trabajo con región de rechazo, simulaciones repetidas y discusión sobre frecuencia de errores.	Ítem 6	Contrastar si los estudiantes pasan de una visión declarativa de α a una comprensión frecuentista más sólida.

Parámetro de análisis	**Evidencia en el pretest**	**Evidencia en la intervención didáctica**	**Evidencia en el postest**	**Propósito de la triangulación**
Interpretación de la decisión inferencial (“no rechazar H_0”)	Ítem 7	Discusión de conclusiones estadísticas, análisis de reportes y contraste entre “aceptar” y “no rechazar”.	Ítem 7	Verificar si disminuye la concepción errónea de equiparar “no rechazar” con “aceptar”.

La triangulación se fundamenta en las categorías de análisis definidas a partir de las habilidades de razonamiento estadístico propuestas en esta investigación y en los principios teóricos del Statistical Reasoning Assessment (SRA), los cuales permiten identificar formas de razonamiento correcto e incorrecto presentes en las respuestas de los estudiantes (J. B. Garfield, 2003). Asimismo, la triangulación se articula con el modelo Reflexión y Acción (RyA), debido a que relaciona los momentos de acción desarrollados mediante las simulaciones con los espacios de reflexión generados durante la intervención.

Como se observa en la Tabla 12, las categorías de análisis consideradas fueron: identificación de H_0 y H_1, comprensión del error tipo I, comprensión del error tipo II, aplicación de hipótesis en contexto, interpretación del valor p, comprensión del nivel de significancia α e interpretación de la decisión inferencial. Para cada una de estas categorías se establecieron evidencias provenientes del pretest, de las actividades desarrolladas durante la intervención y de los resultados obtenidos en el postest. Asimismo, la organización de las categorías de análisis posibilita una comparación entre el pretest y el postest en términos de habilidades de razonamiento correctas y concepciones erróneas.

En relación con la identificación de las hipótesis estadísticas, los resultados del pretest evidenciaron dificultades en la formulación correcta de H_0 Y H_1, especialmente en situaciones contextualizadas. Sin embargo, durante la intervención los estudiantes participaron en actividades de formulación de hipótesis apoyadas en simulaciones y discusión, lo cual favoreció una mejor comprensión de la lógica inferencial. Aunque esta habilidad no fue evaluada explícitamente en el postest, Las discusiones realizadas durante las actividades mostraron avances en la correcta formulación e interpretación de hipótesis estadísticas.

Respecto a la comprensión de los errores tipo I y tipo II, la triangulación evidencia una disminución de la confusión entre ambos conceptos. En el pretest, varios estudiantes asociaban incorrectamente el rechazo de H_0 con el error tipo II o confundían las condiciones bajo las cuales ocurren estos errores. No obstante, las simulaciones desarrolladas durante la intervención permitieron analizar repetidamente escenarios de rechazo y no rechazo de H_0, favoreciendo la comprensión de las decisiones. Estos avances se reflejan en los resultados del postest, donde la mayoría de los estudiantes respondió correctamente las preguntas relacionadas con los errores tipo I y tipo II.

Por otra parte, la triangulación de resultados relacionada con la interpretación del valor p muestra avances importantes en la comprensión de este concepto. Inicialmente, en el pretest se evidenciaron interpretaciones erróneas del valor p, particularmente su asociación con la probabilidad de que la hipótesis nula fuera verdadera o con una medida directa de relevancia práctica. Durante la intervención, las actividades de simulación y análisis de resultados permitieron que los estudiantes visualizaran el comportamiento del valor p en diferentes muestras y comprendieran su relación con la variabilidad y la evidencia estadística. En consecuencia, en el postest la mayoría de los estudiantes logró interpretar correctamente el valor p y reconocer su carácter aleatorio.

De igual manera, la triangulación evidencia avances en la comprensión del nivel de significancia α. En el pretest, varios estudiantes interpretaban α únicamente como un valor numérico asociado a la prueba, sin relacionarlo con la probabilidad de cometer un error tipo I. Sin embargo, las simulaciones repetidas y el análisis de frecuencias de rechazo desarrollados durante la intervención favorecieron una comprensión del concepto. Esto se reflejó en el postest, donde los estudiantes lograron asociar correctamente el nivel de significancia con la frecuencia esperada de rechazo de H_0 cuando esta es verdadera.

En síntesis, la triangulación de resultados permitió identificar que la intervención didáctica basada en simulaciones interactivas favoreció el desarrollo de habilidades de razonamiento estadístico asociadas a las pruebas de hipótesis.

4.4. Análisis cuantitativo de los resultados

El análisis cuantitativo de los resultados permitirá examinar de manera objetiva las variaciones observadas entre el pretest y el postest aplicados a los estudiantes. En este apartado se presentan, en primer lugar, los resultados globales obtenidos en ambos instrumentos; posteriormente, se realiza una comparación por cada pregunta. Para ello, se utiliza la prueba exacta basada en la distribución binomial, debido al tamaño reducido de la muestra ($n = 8$ estudiantes) (Storer & Kim, 1990) . A partir de este análisis se identifican las principales concepciones erróneas evidenciadas en las respuestas.

Finalmente, con el fin de determinar si las diferencias observadas entre los resultados totales del pretest y del postest son estadísticamente significativas, se aplica la prueba de rangos con signo de Wilcoxon(Hollander et al., 2013). Esta prueba no paramétrica se basa en el análisis de las diferencias pareadas entre ambas mediciones y permite evaluar si la mediana de dichas diferencias es significativamente distinta de cero. Su uso es adecuado en este estudio debido a que las mediciones corresponden a los mismos estudiantes evaluados en dos momentos distintos, lo que da lugar a un diseño de datos pareados, al tamaño reducido de la muestra y a que no se asume normalidad en las diferencias. De esta manera, la prueba de Wilcoxon permite determinar de forma robusta si los resultados del postest tienden a ser mayores que los del pretest.

Resultados globales del pretest y postest

En este apartado se presentaran los resultados globales obtenidos por los estudiantes en el pretest y el postest, con el propósito de identificar si hubo mejora en cada ítem antes y después de la implementación. Para ello, se comparan el total de respuestas correctas, el promedio de aciertos por estudiante y el porcentaje general alcanzado en ambos instrumentos.

Dado que participaron ocho estudiantes y cada prueba estuvo conformado por siete preguntas , el total de respuestas correctas en el pretest y el postest fue de 56.

Tabla 13: *Resultados globales del pretest y postest*

Instrumento	Total de respuestas correctas	Promedio	% de aciertos
Pretest	26	3.25	46.43 %
Postest	50	6.25	89.29 %

Los resultados globales evidencian una mejora notable en el desempeño general de los estudiantes entre la aplicación del pretest y el postest. En el pretest se obtuvo un total de 26 respuestas correctas de 56 posibles, lo que corresponde a un 46.43 % de aciertos y un promedio de 3.25 respuestas correctas por estudiante. Por su parte, en el postest se alcanzaron 50 respuestas correctas de 56 posibles, equivalentes al 89.29 % de aciertos y un promedio de 6.25 respuestas correctas por estudiante. En conjunto, estos resultados muestran un incremento importante en la comprensión de los conceptos abordados, lo que sugiere un avance favorable tras la implementación de la intervención didáctica. No obstante, esta mejora descriptiva será complementada posteriormente, mediante la prueba de rangos de Wilcoxon para muestras pareadas, por tratarse de un mismo conjunto de estudiantes que responden ambas encuestas.

Comparación por pregunta

En este apartado se realizara la comparación de los resultados obtenidos en cada una de las preguntas del pretest y el postest, con el propósito de identificar en cuáles ítems se evidenciaron mayores avances y en cuáles persistieron algunas dificultades. Para ello, se analiza el número de respuestas correctas por pregunta y su correspondiente porcentaje de acierto en ambas pruebas.

Tabla 14: *Comparación de resultados por pregunta en el pretest y el postest*

Pregunta	Aciertos Pretest	% Pretest	Aciertos Postest	% Postest	Diferencia (p.p.)
P1	2	25.00 %	8	100.00 %	+75.00
P2	3	37.50 %	8	100.00 %	+62.50
P3	3	37.50 %	7	87.50 %	+50.00
P4	6	75.00 %	8	100.00 %	+25.00
P5	3	37.50 %	7	87.50 %	+50.00
P6	6	75.00 %	6	75.00 %	0.00
P7	3	37.50 %	6	75.00 %	+37.50

La comparación de resultados por pregunta permite evidenciar una mejora general en la mayoría de los ítems evaluados. En particular, las preguntas 1 y 2 mostraron los avances más notorios, al pasar de 25.00 % y 37.50 % de aciertos en el pretest a 100.00 % en el postest, respectivamente. De igual manera, las preguntas 3 y 5 presentaron incrementos importantes, alcanzando un 87.50 % de respuestas correctas en el postest. La pregunta 4, que ya mostraba un desempeño relativamente alto en el pretest (75.00 %), también evidenció una mejora al alcanzar el 100.00 % de aciertos en el postest. Por su parte, la pregunta 7 presentó un avance moderado, al pasar de 37.50 % a 75.00 %. En contraste, la pregunta 6 mantuvo el mismo porcentaje de aciertos en ambos instrumentos (75.00 %), lo que sugiere que, aunque no hubo retroceso, tampoco se evidenció un avance cuantitativo en este ítem. En conjunto, estos resultados muestran que la intervención didáctica favoreció especialmente la comprensión de varios de los contenidos evaluados, aunque algunos aspectos aún requieren un análisis más detallado en relación con las concepciones erróneas persistentes.

Con el fin de profundizar en el análisis de los resultados por pregunta, se utilizó la prueba binomial exacta (ver Storer y Kim, 1990) implementada mediante la función `binom.test` en el software R, tomando como referencia la proporción observada en el pretest y evaluando si la proporción del postest es significativamente mayor. En este contexto, para cada pregunta se contrastaron las hipótesis :

$$H_0 : p_{\text{postest}} \leq p_{\text{pretest}}$$

$$H_a : p_{\text{postest}} > p_{\text{pretest}}$$

donde p_{pretest} y p_{postest} representan las proporciones de aciertos en cada momento de evaluación, para cada pregunta del pre y postest.

La función `binom.test` permite comparar las proporcioes de cada pregunta del pretest y el postest, en particular, para determinar si la del postest es significativamente mayor que la otra proporción. Este procedimiento es especialmente adecuado en muestras pequeñas, ya que no depende de aproximaciones asintóticas y proporciona valores p exactos, lo que garantiza una mayor precisión en la toma de decisiones estadísticas Storer y Kim, 1990.

Tabla 15: *Código en R para la prueba binomial exacta y valores p obtenidos*

```
# Datos
pre <- c(2, 3, 3, 6, 3, 6, 3)
post <- c(8, 8, 7, 8, 7, 6, 6)
n <- 8
# Vector para guardar p-valores
pvalores <- numeric(length(pre))
# Test binomial por pregunta
for(i in 1:length(pre)) {
   p_pre <- pre[i] / n
   prueba <- binom.test(post[i], n, p = p_pre, alternative =
"greater")
   pvalores[i] <- prueba$p.value
}
# Resultados
pvalores
[1] 0.000015   0.000391   0.005605   0.100113   0.005605
0.678543   0.036022
```

Con base en los resultados obtenidos, se observa que los ítems evaluados 1, 2, 3, 5 y 7 presentan mejoras estadísticamente significativas, mientras que las preguntas 4 y 6 no evidencian cambios significativos. Estos resultados son coherentes con el análisis descriptivo previo, en el cual se identificó que la pregunta 4 ya presentaba un alto porcentaje de aciertos desde el pretest, mientras que la pregunta 6 no mostró variación entre ambos momentos de evaluación.

Tabla 16: *Análisis por pregunta mediante prueba binomial exacta*

Pregunta	p-valor	Decisión	Interpretación
P1	0.000015	No rechaza H_a	Mejora significativa
P2	0.000391	No rechaza H_a	Mejora significativa
P3	0.005605	No rechaza H_a	Mejora significativa
P4	0.100113	Rechaza H_a	No evidencia mejora
P5	0.005605	No rechaza H_a	Mejora significativa
P6	0.678543	Rechaza H_a	No evidencia mejora
P7	0.036022	No rechaza H_a	Mejora significativa

A partir de los resultados obtenidos y los criterios de análisis del Cuadro 12 , se observa que las preguntas 1, 2, 3, 5 y 7 presentan mejoras estadísticamente significativas con una confianza del 95 % , mientras que las preguntas 4 y 6 no evidencian cambios significativos . Estos resultados son consistentes con el análisis descriptivo previo, en el cual se identificó que la pregunta 4 ya presentaba un alto porcentaje de aciertos desde el pretest y que la pregunta 6 no mostró variación entre ambos momentos de evaluación.

Teniendo en cuenta los anterior y que la pregunta 4 se refiere a la interpretación del valor p, según lo que se ha documentado en la literatura estadística, el valor p constituye uno de los conceptos más difíciles de comprender para los estudiantes (J. Garfield & Ben-Zvi, 2008b). Entre las principales dificultades se encuentra la tendencia a interpretar erróneamente el valor p como la probabilidad de que la hipótesis nula sea verdadera, en lugar de entenderlo como la probabilidad de observar datos tan extremos como los obtenidos bajo la suposición de que dicha hipótesis es cierta (Gigerenzer, 2004).

Por lo que a pesar de las ventajas de las simulaciones computacionales para la enseñanza del valor p, pueden persistir sus dificultades en la interpretación. Esto se debe, en parte, a la resistencia de las concepciones previas de los estudiantes, las cuales no son fácilmente reemplazadas únicamente mediante la exposición a representaciones visuales o experimentales (J. Garfield & Ben-Zvi, 2008b). Adicionalmente, la complejidad inherente del valor p, que implica razonamiento probabilístico condicional y comprensión de la variabilidad muestral, continúa representando un desafío cognitivo significativo (Gigerenzer, 2004). En consecuencia, el cambio conceptual en torno al valor p requiere intervenciones didácticas más prolongadas, estructuradas y centradas en la comprensión del concepto en la toma de decisiones.

Luego, en la pregunta 6 los resultados obtenidos sugieren que no se evidenció una mejora significativa en la interpretación del nivel de significancia. Este hallazgo puede explicarse a la luz de la literatura en educación estadística, la cual señala que los estudiantes suelen desarrollar habilidades procedimentales sin lograr una comprensión conceptual profunda que les permita interpretar los resultados en contextos reales (J. Garfield & Ben-Zvi, 2008b) . En particular, la interpretación en contexto requiere la articulación entre resultados estadísticos y situaciones del mundo real.

Tabla 17: *Frecuencia de concepciones erróneas identificadas en el pretest y el postest*

Concepción errónea identificada	**Pretest**	**Postest**
Confunde el planteamiento de la hipótesis nula, asociándola con afirmaciones de superioridad en lugar de igualdad o referencia	X	X
No distingue correctamente el significado del error tipo I, confundiéndolo con otras decisiones en la prueba de hipótesis	X	–
Presenta dificultades para diferenciar el error tipo II de otras situaciones relacionadas con la decisión estadística	X	–
Interpreta incorrectamente el nivel de significancia (α), asociándolo con el valor-p o con la relevancia práctica	X	–
Toma decisiones erróneas al interpretar el valor-p, sin relacionarlo adecuadamente con la evidencia contra H_0	X	–
Interpreta de manera incorrecta el significado de "no rechazar H_0", asociándolo con la veracidad de la hipótesis nula	X	X

El análisis de las concepciones erróneas identificadas en las respuestas de los estudiantes permite evidenciar que, en el pretest, se presentaban dificultades relacionadas con la formulación de la hipótesis nula, la interpretación del error tipo I y del error tipo II, la comprensión del nivel de significancia y la toma de decisiones a partir del valor-p. Estas dificultades se manifestaron en varios ítems del instrumento, lo que indica que, al inicio de la intervención, los estudiantes presentaban vacíos conceptuales relevantes en torno a nociones fundamentales de la inferencia estadística.

En el postest se observa una reducción importante en la frecuencia de estas concepciones erróneas. En particular, dejaron de evidenciarse dificultades asociadas a la interpretación del error tipo I, del error tipo II, del nivel de significancia y de la toma de decisiones basadas en el valor-p, lo que sugiere un avance notable en la comprensión de los conceptos abordados durante la intervención. No obstante, persisten algunas dificultades puntuales relacionadas con la formulación adecuada de la hipótesis nula y con la interpretación del significado de “no rechazar H_0”, aunque con menor incidencia en comparación con el diagnóstico inicial. En conjunto, los resultados reflejan un progreso sólido y pedagógicamente relevante en el aprendizaje de los estudiantes, al evidenciar una disminución clara de errores conceptuales en la mayoría de los contenidos evaluados.

Análisis de diferencias entre el pretest y el postest mediante la prueba de rangos con signo de Wilcoxon

Dado que se comparan los puntajes obtenidos por los mismos estudiantes antes y después de la intervención, y considerando el tamaño reducido de la muestra (n = 8), se empleó la prueba no paramétrica de rangos con signo de Wilcoxon para muestras relacionadas. Esta prueba permite determinar si existen diferencias significativas entre dos mediciones pareadas cuando no se asume normalidad en los datos.

El análisis estadístico se realizó utilizando el software RStudio. La salida completa del procedimiento se presenta en la siguiente tabla.

Tabla 18: *Prueba de Wilcoxon para muestras relacionadas*

```
# Datos reales
pretest <- c(3, 4, 3, 1, 2, 4, 4, 4)
postest <- c(7, 7, 7, 7, 7, 5, 5, 5)
# Prueba de Wilcoxon (una cola: postest >pretest)
resultado <- wilcox.test(postest, pretest,
   paired = TRUE,
   alternative = "greater",
   exact = FALSE)
# Mostrar resultado
resultado
Wilcoxon signed rank test with continuity correction
data: postest and pretest
V = 36, p-value = 0.006838
alternative hypothesis: true location shift is greater than 0
```

Planteamiento de hipótesis

Para contrastar los resultados del pretest y del postest se utiliza la prueba de rangos con signo de Wilcoxon, considerando las diferencias pareadas $D_i = \text{Postest}_i - \text{Pretest}_i$. Las hipótesis planteadas son:

$$H_0 : \ \text{Mediana}(D) \leq 0$$

$$H_1 : \text{ Mediana}(D) > 0$$

donde la hipótesis alternativa indica que los resultados del postest son mayores que los del pretest.

Resultados de la prueba La hipótesis nula establece que no hubo mejoras entre el puntaje total del pretest y del postest. Al aplicar la prueba de rangos con signo de Wilcoxon se obtuvo un valor $p = 0{,}006838$. Dado que este valor es menor que el nivel de significancia $\alpha = 0{,}05$, se rechaza la hipótesis nula con una confianza del 95 %. Por lo tanto, existe evidencia estadísticamente significativa para afirmar que la implementación de la clase basada en simulaciones mejora la comprensión de los conceptos asociados a las pruebas de hipótesis.

Conclusión Los resultados evidencian diferencias estadísticamente significativas entre el pretest y el postest ($p = 0{,}006838$) con una confianza del 95 %, lo que sugiere una mejora en el desempeño de los estudiantes tras la intervención. Este hallazgo constituye un indicio favorable sobre el aporte de la estrategia implementada en el fortalecimiento de los aprendizajes evaluados.

5. Conclusiones

A partir del desarrollo e implementación de la secuencia didáctica apoyada en simulaciones en *RStudio*, se concluye que la intervención favoreció procesos de comprensión en torno a conceptos fundamentales de la inferencia estadística, particularmente en relación con el muestreo, la formulación de hipótesis, el nivel de significancia, el valor p y la toma de decisiones en pruebas de hipótesis.

Desde la perspectiva del modelo de Reflexión y Acción (RyA) propuesto por Parada, 2016, fue posible evidenciar que, durante la intervención, los estudiantes no solo participaron activamente en la resolución de las actividades, sino que también cuestionaron sus ideas iniciales, contrastaron sus respuestas previas con los resultados obtenidos en las simulaciones y avanzaron hacia interpretaciones más consistentes de los conceptos abordados. En este sentido, la articulación entre la explicación teórica y la experimentación mediante simulaciones permitió promover procesos de reflexión en la acción y reflexión sobre la acción, favoreciendo una comprensión más significativa de la variabilidad, el azar y la lógica inferencial.

En relación con el análisis comparativo entre el pretest y el postest, los resultados muestran una mejora en el desempeño de los estudiantes después de la intervención. En particular, se observó una disminución en las concepciones erróneas asociadas a la interpretación del error tipo I, el nivel de significancia y la toma de decisiones a partir del valor p. Aunque persistieron algunas dificultades puntuales, especialmente en la formulación de hipótesis y en la interpretación del no rechazo de H_0, estas se presentaron con menor frecuencia y de manera más delimitada en comparación con el diagnóstico inicial.

Asimismo, la aplicación de la prueba de rangos con signo de Wilcoxon permitió identificar diferencias estadísticamente significativas entre los resultados del pretest y el postest ($p = 0,006838$), lo que sugiere un avance en el desempeño de los estudiantes tras la implementación de la propuesta. Este hallazgo constituye un indicio favorable sobre el aporte de la estrategia didáctica desarrollada.

En términos pedagógicos, se concluye que el uso de simulaciones en *RStudio*, integrado a una secuencia de actividades orientadas desde la reflexión y la acción, representa una alternativa pertinente para la enseñanza de la inferencia estadística en la educación, al facilitar la visualización de procesos aleatorios, la interpretación de resultados y la construcción de significados asociados a conceptos tradicionalmente abstractos.

Finalmente, se reconoce que el alcance de los resultados debe interpretarse considerando el tamaño reducido de la muestra y el carácter contextualizado de la intervención. No obstante, los hallazgos obtenidos permiten valorar positivamente la experiencia y abren la posibilidad de continuar explorando el uso de estrategias didácticas mediadas por simulación para fortalecer la enseñanza y el aprendizaje de la estadística inferencial.

6. Recomendaciones

A partir de los resultados obtenidos en esta investigación, se plantean las siguientes recomendaciones orientadas al fortalecimiento de futuras propuestas de enseñanza en torno a la inferencia estadística:

- Implementar estrategias didácticas que integren simulaciones computacionales, ya que estas favorecen la visualización de procesos aleatorios, la comprensión de la variabilidad muestral y la interpretación de conceptos abstractos como el nivel de significancia, el valor p y la toma de decisiones en pruebas de hipótesis.
- Promover secuencias de enseñanza que articulen la explicación teórica con espacios de exploración, discusión y reflexión, de manera que los estudiantes no solo apliquen procedimientos, sino que también construyan significados más sólidos sobre los conceptos de inferencia estadística.
- Fortalecer, en futuras intervenciones, aquellos aspectos conceptuales en los que persistieron dificultades, especialmente la formulación de hipótesis estadísticas y la interpretación del no rechazo de la hipótesis nula, dado que estos continúan siendo puntos sensibles en el aprendizaje de los estudiantes.
- Diseñar actividades que permitan un mayor tiempo de interacción con las simulaciones en *RStudio*, de modo que los estudiantes puedan contrastar múltiples resultados, reconocer patrones y profundizar en la comprensión del carácter aleatorio de los fenómenos estadísticos.
- Replicar esta propuesta en grupos con un mayor número de estudiantes y en otros contextos educativos, con el fin de ampliar el alcance de los resultados y analizar la consistencia de los hallazgos en diferentes poblaciones escolares.
- Complementar futuras investigaciones con instrumentos de recolección de información adicionales, como entrevistas, diarios de campo o grabaciones de clase, que permitan profundizar en el análisis de los procesos de razonamiento y en las transformaciones conceptuales de los estudiantes durante la intervención.
- Continuar explorando el uso pedagógico de herramientas tecnológicas como *RStudio* en la enseñanza de la estadística, no solo como recurso de cálculo, sino como mediador didáctico para favorecer procesos de interpretación, argumentación y toma de decisiones fundamentadas.
- Se sugiere que en futuras implementaciones se tomen muestras de mayor tamaño, ya que el tamaño muestral influye directamente en la estabilidad de los resultados y en la capacidad de las pruebas de hipótesis para detectar diferencias significativas. Un mayor número de observaciones favorece una mejor aproximación a la distribución teórica.

Referencias

Batanero, C. (2001a). Didáctica de la Estadística. *Granada: Universidad de Granada.*

Batanero, C. (2001b). Didáctica de la Estadística. *Universidad de Granada.*

Batanero, C. (2011). El currículo de estadística en la enseñanza obligatoria. *UNO Revista de Didáctica de las Matemáticas.*

Batanero, C., Gea, M. M., & Álvarez-Arroyo, R. (2023). La educación del razonamiento probabilístico. *Educação Matemática Pesquisa*, *25*(2), 127-144. https://www.researchgate.net/publication/374528346_La_educacion_del_razonamiento_probabilistico

Batanero, C., Sánchez, V., & Tauber, L. (2004). Diseño, implementación y análisis de una secuencia de enseñanza de la distribución normal en un curso universitario. *Revista EMA*, *9*(2), 82-105.

Bernabeu, C. B. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática*, 277-291.

Bickel, P. J. (2021). *Mathematical Statistics: Basic Ideas and Selected Topics*. Springer.

Bickel, P. J., & Doksum, K. A. (1997). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume 1* (Second). Prentice Hall.

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman; Hall/CRC.

Cáceres-Mesa, M. L., Ríos-Ramírez, P., Veytia-Bucheli, M. G., & García-Robelo, O. (2025). Análisis de herramientas tecnológicas para la enseñanza de las matemáticas y su impacto en el aprendizaje de los estudiantes del CETIS91. *Revista Metropolitana de Ciencias Aplicadas*, *8*(2), 177-186.

Calixto, N. J. C., Alvarado, W. P., & Rolón, Á. J. C. (2023). *Conceptos y enfoques de metodología de la investigación.* Editorial [indicar si aparece].

Casella, G., & Berger, R. (2002). *Statistical Inference* (2nd). Duxbury Press.

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98-113. https://doi.org/10.1016/j.edurev.2007.04.001

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1).

Chick, H. L., & Pierce, R. U. (2008). Teaching Statistics at the Primary School Level: Beliefs, Affordances, and Pedagogical Content Knowledge [Joint ICMI/IASE Study]. En C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Teaching Statistics in School Mathematics: Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. ICMI/IASE.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (3rd). SAGE Publications.

Dagnino, J., et al. (2014). Elección de una prueba de hipótesis. *Rev Chil Anest*, *43*, 139-142.

Daniel, W. W. (2005). *Bioestadística: Base para el análisis de las ciencias de la salud.* Limusa Wiley.

de Educación Nacional, M. (1998). *Lineamientos Curriculares: Matemáticas.*

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics: The Statistical Reasoning Assessment. *Statistics Education Research Journal*, *6*(2), 28-58. https://iase-web.org/documents/SERJ/SERJ6(2)_DelMas.pdf

Devore, J. L. (2020). *Probability and Statistics for Engineering and the Sciences* (10.ª ed.). Cengage Learning.

Díaz Pinzón, J. E. (2018). Aprendizaje de las matemáticas con el uso de simulación. *Sophia-Educación*, *14*(1), 22-30. https://doi.org/10.18634/sophiaj.14v.1i.519

Erickson, T., Wilkerson, M., Finzer, W., & Reis, M. (2019). Data Moves. *Technology Innovations in Statistics Education*, *12*(1). https://escholarship.org/uc/item/0mg8m7g6

Fernández, F., & Sarmiento, B. (2020). Acerca del contexto escolar de la enseñanza de lo estocástico en colegios oficiales de Bogotá [Disponible en FUNES: Repositorio de Educación Matemática]. https://funes.uniandes.edu.co/

Fisher, R. A. (1935). *The Design of Experiments*. Oliver; Boyd.

Flórez, D., Camargo, A., & Rodríguez, L. (2020). Las actividades orientadoras de enseñanza como estrategia para enseñar la probabilidad en primaria. *Trilogía Ciencia Tecnología Sociedad*, *12*(22), 131-146. https://portal.amelica.org/ameli/jatsRepo/156/156876003/

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, *10*(3).

Garfield, J. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal*, *2*(1), 22-38.

Garfield, J., & Ahlgren, A. (1988). Difficulties in Learning Basic Concepts in Statistics: Implications for Research. *Journal for Research in Mathematics Education*, *19*(1), 44-63.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, *4*(1), 92-99.

Garfield, J., & Ben-Zvi, D. (2008a). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Springer.

Garfield, J., & Ben-Zvi, D. (2008b). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.

Garfield, J., & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematical Thinking and Learning*, *2*(1-2), 99-125.

Garfield, J., & Gal, I. (1999). Teaching and Assessing Statistical Reasoning. En *Developing Mathematical Reasoning in Grades K-12* (pp. 207-219). National Council of Teachers of Mathematics.

Garfield, J. B. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal*, *2*(1), 22-38. https://doi.org/10.52041/SERJ.v2i1.557

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Perthes et Besser.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of socio-economics*, *33*(5), 587-606.

Gómez-Luna, E., Fernando-Navas, D., Aponte-Mayor, G., & Betancourt-Buitrago, L. A. (2014). Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización. *Dyna*, *81*(184), 158-163.

Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1-25.

Goulet, V., & Dutang, C. (2023). *Moments and Moment Generating Function of the Normal Distribution* [R package documentation].

Hallgren, K. A. (2013). Conducting simulation studies in the R programming environment. *Tutorials in quantitative methods for psychology*, *9*(2), 43.

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric Statistical Methods* (3.ª ed.). John Wiley & Sons.

Hurtado, O. G., Ch, R. P., & Moncada, J. (2022). The Montecarlo Method In The Teaching Of Financial Mathematics. *Journal of Language and Linguistic Studies*, *18*(4).

Inzunsa Cazares, S., & Jiménez Ramírez, J. V. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista latinoamericana de investigación en matemática educativa*, *16*(2), 179-211.

Koparan, T. (2022). How does simulation contribute to prospective mathematics teachers' learning experiences and results? *Education Sciences*, *12*(9), 624.

Lane, D. (2014). Simulations and statistical understanding. En D. Ben-Zvi & K. Makar (Eds.), *The Teaching and Learning of Statistics* (pp. 147-155). Springer.

Laplace, P.-S. (1814). *Essai philosophique sur les probabilités*. Courcier.

Lehmann, E. L., & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3.ª ed.). Springer.

Montgomery, D. C., & Runger, G. C. (2018). *Applied Statistics and Probability for Engineers* (7.ª ed.). Wiley.

Montoro, A. G. (2019). Contrastes de hipótesis mediante técnicas de simulación. *Revista de Educación Matemática*, *34*(1).

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3.ª ed.). McGraw-Hill.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, *231*, 289-337.

Pando, J. C., Berenson, M. L., & Levine, D. M. (1992). *Estadística básica en administración*. Prentice Hall.

Parada, S. E. (2011). *Reflexión y acción en comunidades de práctica: Un modelo de desarrollo profesional* [Tesis doctoral, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional].

Parada, S. E. (2016). *Comunidades de práctica y desarrollo profesional docente en matemáticas*. Universidad Industrial de Santander.

Parada, S. E., & Fiallo, J. E. (2011). Modelo de Reflexión-y-Acción en una comunidad de práctica de educadores matemáticos. *Revista Latinoamericana de Investigación en Matemática Educativa*, *14*(3), 353-380.

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th). SAGE Publications.

Pearson. (2025). Critical values and rejection regions [Pearson Statistics Learning Channel].

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1-2), 27-46.

Posit Software, PBC. (2025). *Download RStudio IDE* [Documentación técnica y descripción oficial del IDE]. https://posit.co/downloads/

Reis, S. R. d., Müller, I., Moreira Junior, F. d. J., & Ansuj, A. P. (2020). Educação estatística – uma proposta de ensino. *Ciência e Natura*, *42*(Edição Comemorativa: Estatística), e16. https://doi.org/10.5902/2179460X40457

Rodríguez, M. I., Agnelli, H., & Albert, A. (2009). Pruebas de hipótesis estadísticas: algunas consideraciones para la práctica docente. *Revista de Educación en Ciencias*, *13*(2), 45-62.

Rossman, A., & Chance, B. (2014). *Workshop Statistics: Discovery with Data and Simulation*. Wiley.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of statistics education*, *10*(3).

Sanchez, M., Marin, G., & Quintero, I. (2024). La importancia de la prueba de hipótesis. *Revista Semilla Científica*, (5), 211-216.

Siegel, S., & Castellan, N. J. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta*. Trillas.

Stigler, S. M. (1999). On the history of statistics. En *Statistics on the Table: The History of Statistical Concepts and Methods* (pp. 157-174). Harvard University Press.

Storer, B. E., & Kim, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, *85*(409), 146-155.

Suárez, E. O. (2008). *Significado de los intervalos de confianza para los estudiantes de ingeniería en México*. Universidad de Granada.

UNESCO. (2012). *Challenges in Basic Mathematics Education*. United Nations Educational, Scientific; Cultural Organization. https://personales.unican.es/reciot/UNIA%20TUMP/UNESCO-math-education.pdf

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (9.ª ed.). Pearson.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose.

Waters, S. (2025). *Investigating Undergraduate Students' Statistical Literacy in Media Contexts* [Tesis doctoral, University of Northern Colorado].

Winkelbauer, A. (2012). Moments and Absolute Moments of the Normal Distribution. *arXiv preprint arXiv:1209.4340*.

Zieffler, A., Garfield, J., DelMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40-58.

Apéndices

Apéndice A: Códigos utilizados en RStudio

Código de clase sobre muestreo aleatorio simple

Tabla 19: *Muestreo aleatorio simple sin reemplazo*

```
# Muestreo aleatorio simple SIN reemplazo
# Definir la población
poblacion <- 1:100
# Fijar semilla
set.seed(123)
# Tomar muestra
muestra_sin_reemplazo <- sample(poblacion, size = 10, replace =
FALSE)
# Mostrar resultado
print(muestra_sin_reemplazo)
```

Tabla 20: *Muestreo aleatorio simple con reemplazo*

```
# Muestreo aleatorio simple CON reemplazo
# Definir la población
poblacion <- 1:100
# Fijar semilla
set.seed(123)
# Tomar muestra
muestra_con_reemplazo <- sample(poblacion, size = 10, replace =
TRUE)
# Mostrar resultado
print(muestra_con_reemplazo)
```

Tabla 21: *Muestreo con datos simulados*

```
# Generar población simulada
set.seed(123)
estaturas <- rnorm(50, mean = 1.70, sd = 0.10)
# Muestra sin reemplazo
muestra1 <- sample(estaturas, size = 10, replace = FALSE)
# Muestra con reemplazo
muestra2 <- sample(estaturas, size = 10, replace = TRUE)
# Mostrar valores
muestra1[1]
muestra1[2]
# Mostrar muestras completas
print(muestra1)
print(muestra2)
```

Tabla 22: *Prueba t de Student y estimación de errores tipo I (α) y tipo II (β) mediante simulación*

```
# PRUEBA DE HIPÓTESIS PARA LA MEDIA
# H0 : μ = 3
# Ha : μ ≠ 3
y <- rnorm(100, mean = 3, sd = 1)
t.test(y, mu = 3)
t = 0.54973,  gl = 99,  p-value = 0.5837
IC 95%: [2.880033,  3.211915]
Media muestral = 3.045974
```

```
# Simulación del error tipo I
Y <- list()
for (j in 1:1000) {
   Y[[j]] <- rnorm(100, mean = 3, sd = 1)
}
valoresp <- NULL
for (i in 1:1000) {
   valoresp[i] <- t.test(Y[[i]], mu = 3)[[3]]
}
Proporción de rechazo (α): 0.039
```

```
# Error tipo II
# H0 : μ = 3,5 (falsa)
# Ha : μ ≠ 3,5
valoresp1 <- NULL
for (i in 1:1000) {
   valoresp1[i] <- t.test(Y[[i]], mu = 3.5)[[3]]
}
Proporción de no rechazo (β): 0.001
```

```
# Interpretación
Si p-valor < α, se rechaza H0
Si p-valor > α, no se rechaza H0
```

Apéndice B: instrumento de recolección de datos.

Pretest

Propósito Detectar dificultades frecuentes en intervalos de confianza, pruebas de hipótesis y conceptos asociados.

Tiempo estimado

10 minutos.

Instrucciones

Responda todas las preguntas de selecion múltiple y en las preguntas abiertas explique con sus palabras de manera clara.

1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis nula correcta es:

   A) H_0: $\mu_{nuevo} = \mu_{tradicional}$

   B) H_0: $\mu_{nuevo} < \mu_{tradicional}$

   C) H_0: $\mu_{nuevo} > \mu_{tradicional}$

   D) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$

2. Se comete un error tipo I cuando:

   A) No se rechaza H_0 siendo falsa

   B) Se rechaza H_0 siendo verdadera

   C) Se acepta H_1 siendo falsa

   D) No se rechaza H_1 siendo verdadera

3. El error tipo II ocurre cuando:

   A) No se rechaza H_0 siendo falsa

   B) Se rechaza H_0 siendo verdadera

   C) Se confunde H_0 con H_1

   D) Se obtiene un valor-p muy pequeño

4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?

   A) Estadísticamente significativo pero poco relevante en la práctica

   B) No significativo pero relevante

   C) Siempre relevante

   D) Ninguna de las anteriores

5. El nivel de significancia α es:

A) La probabilidad máxima de cometer error tipo I
B) El valor-p del estudio
C) El tamaño de muestra mínimo
D) Una medida de relevancia práctica

6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

   ¿Cuál es el planteamiento correcto de las hipótesis?

   A) $$H_0 : \mu \leq 70 \quad \text{y} \quad H_1 : \mu > 70$$
   B) $$H_0 : \mu = 70 \quad \text{y} \quad H_1 : \mu \neq 70$$
   C) $$H_0 : \mu \geq 70 \quad \text{y} \quad H_1 : \mu < 70$$
   D) $$H_0 : \mu > 70 \quad \text{y} \quad H_1 : \mu \leq 70$$

7. ¿Qué significa “no rechazar H_0”?.

   Seleccione la opción correcta:

   A) “No rechazar H_0” significa que no tenemos evidencia suficiente para decir que H_0 es falsa.
   B) “No rechazar H_0” significa que hemos demostrado que H_0 es verdadera.
   C) “No rechazar H_0” significa que H_1 es verdadera.
   D) “No rechazar H_0” significa que debemos repetir el experimento porque el resultado es inválido.

Apéndice C: instrumento de recolección de datos.

Propósito

Evaluar si los estudiantes pudieron superar las dificultades detectadas en el diagnostico inicial,comprobando su capacidad para interpretar correctamente intervalos de confianza, plantear y contrastar hipótesis estadísticas identificando H_0 y H_1, tomar decisiones coherentes con base en el valor p o la región de rechazo, y relacionar conceptos como error tipo I, error tipo II y nivel de significancia en la interpretación de resultados.

Tiempo estimado

20 minutos.

Postest de comprensión en inferencia estadística

Objetivo

Evaluar la comprensión de los estudiantes respecto a conceptos avanzados de pruebas de hipótesis, errores tipo I y II, valor-p y potencia de una prueba, luego de la intervención didáctica.

Instrucciones

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:

   A) Normal estándar

   B) Uniforme entre 0 y 1

   C) Exponencial positiva

   D) Sesgada hacia cero

2. Un error tipo I ocurre cuando:

   A) No se rechaza H_0 siendo falsa

   B) Se rechaza H_0 siendo verdadera

   C) Se acepta H_1 siendo falsa

   D) No se rechaza H_1 siendo verdadera

3. Un error tipo II ocurre cuando:

   A) Se rechaza H_0 siendo verdadera

   B) No se rechaza H_0 siendo falsa

   C) Se confunde H_0 con H_1

   D) El valor-p resulta mayor a 0.05

4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:

A) El valor-p no informa si no se rechaza la hipótesis nula

B) Se rechaza H_0 con 95 % de confianza

C) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03

D) El tamaño de efecto es grande

5. Si el valor-p se interpreta como variable aleatoria, implica que:

   A) Su valor es fijo en todos los experimentos

   B) Depende de la muestra obtenida

   C) Es independiente del tamaño de muestra

   D) Siempre indica significancia

6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:

   A) La potencia

   B) El nivel de significancia

   C) El error tipo II

   D) La probabilidad de H_0

7. Una conclusión "no se rechaza H_0" significa:

   A) Se acepta H_0 como verdadera

   B) No hay evidencia suficiente para rechazarla

   C) H_0 es cierta con probabilidad 95 %

   D) El valor-p es siempre 0.5

Apéndice D: Producciones de los estudiantes

Figura 19: *Comparación entre el pretest y postest del estudiante 1*

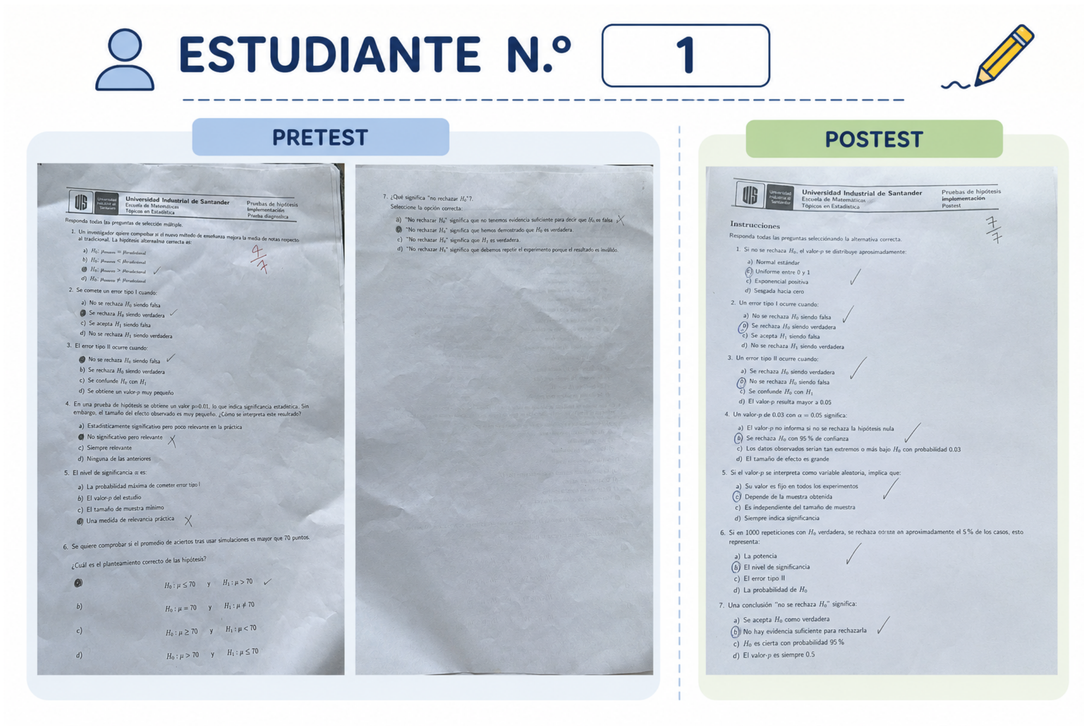


Figura 20: *Comparación entre el pretest y postest del estudiante 2*

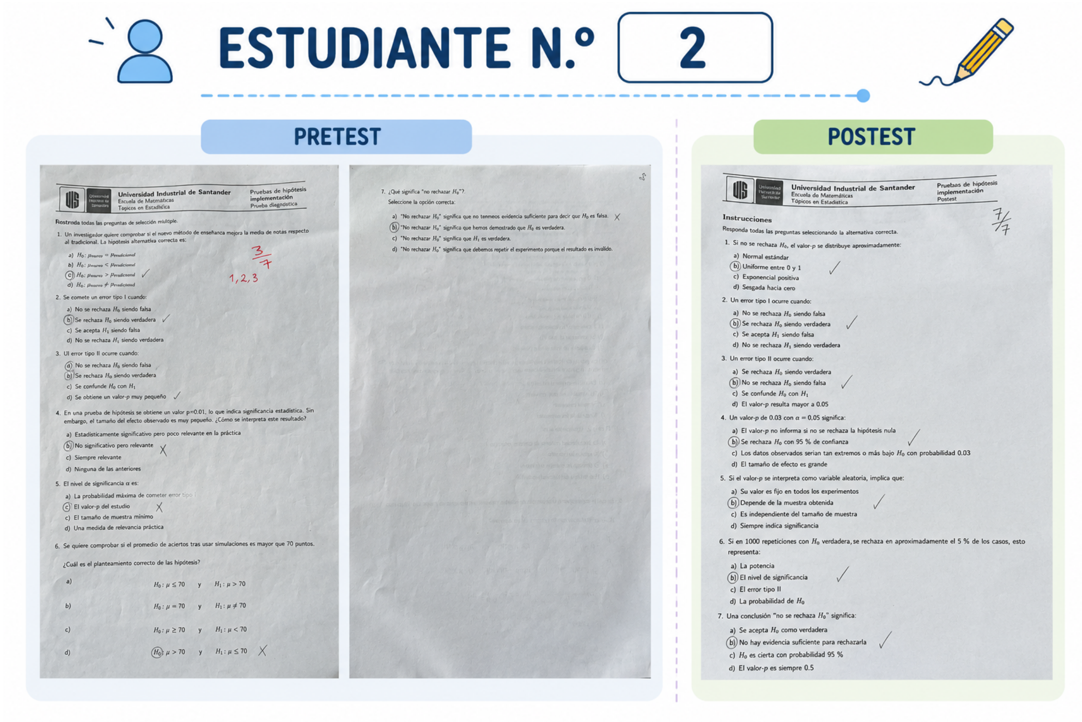


Figura 21: *Comparación entre el pretest y postest del estudiante 3*

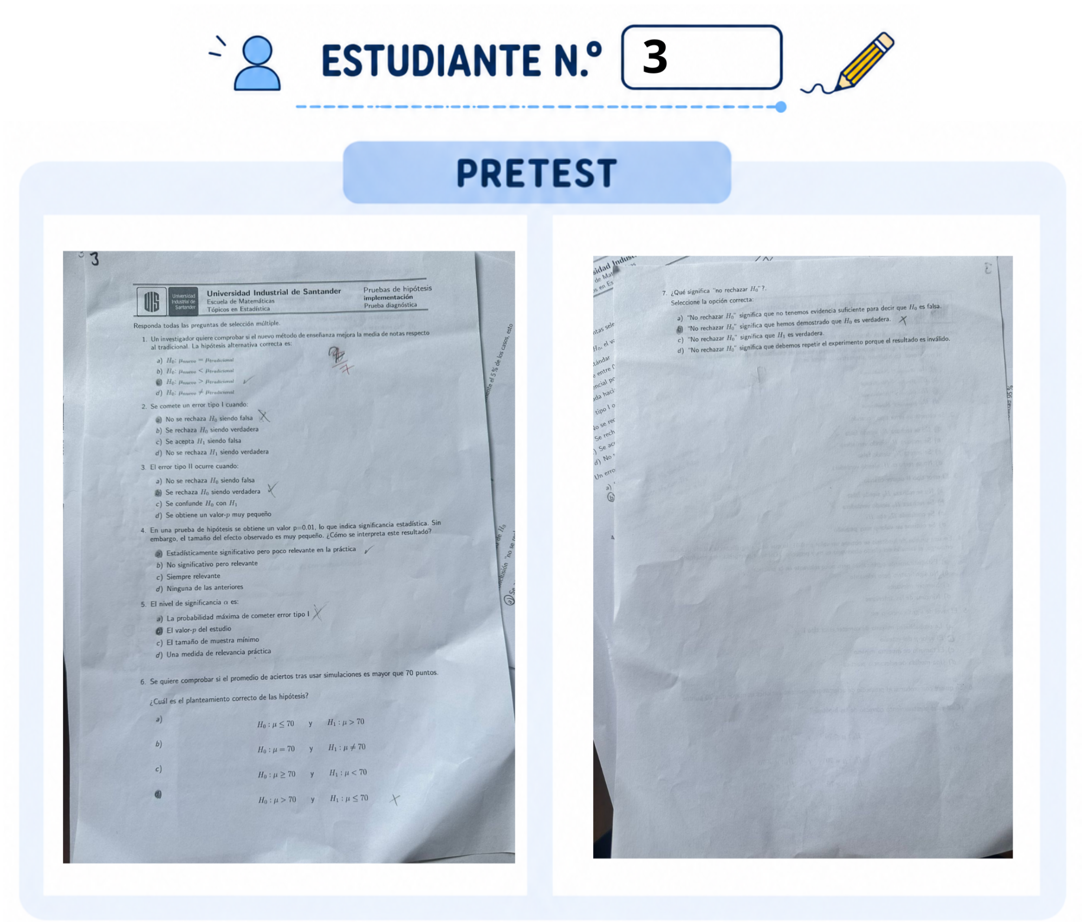
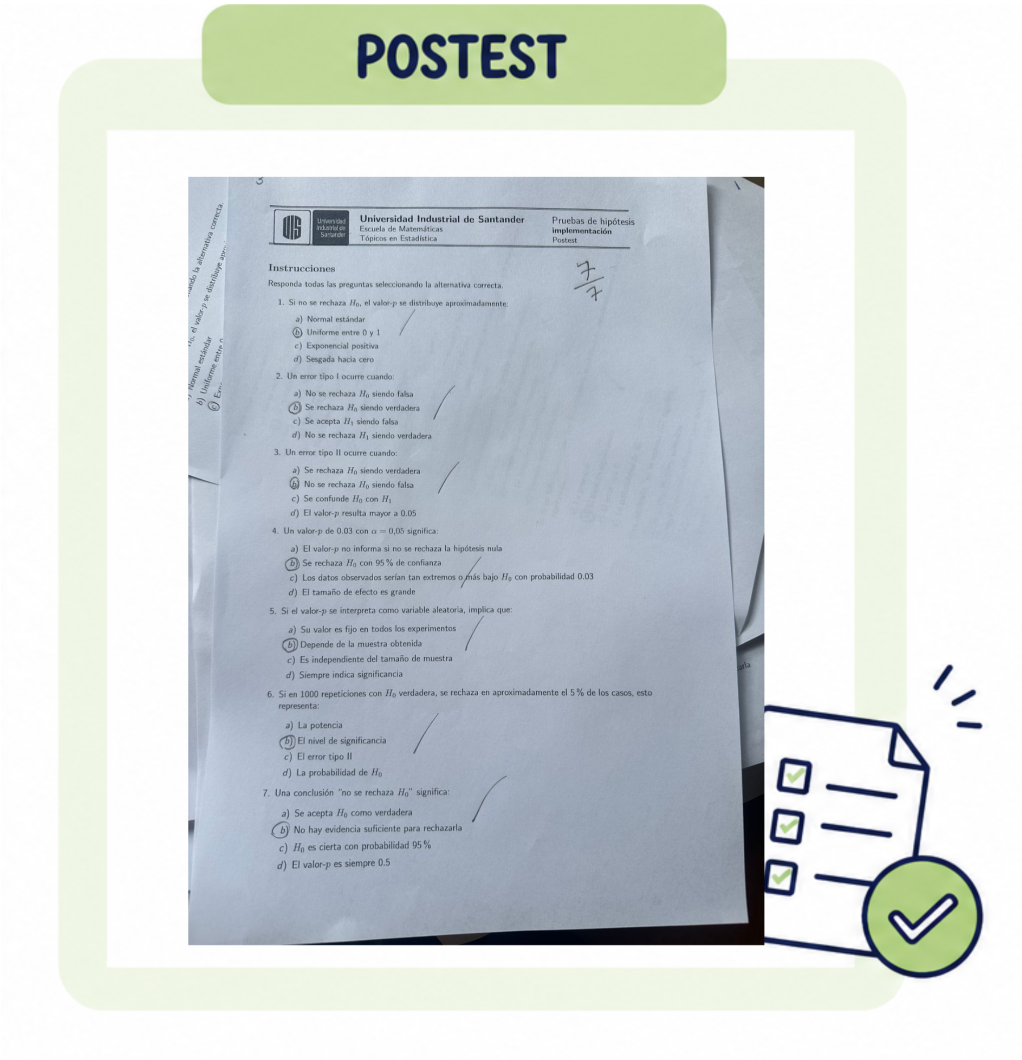


Figura 22: *Comparación postest del estudiante 3*



Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Postest

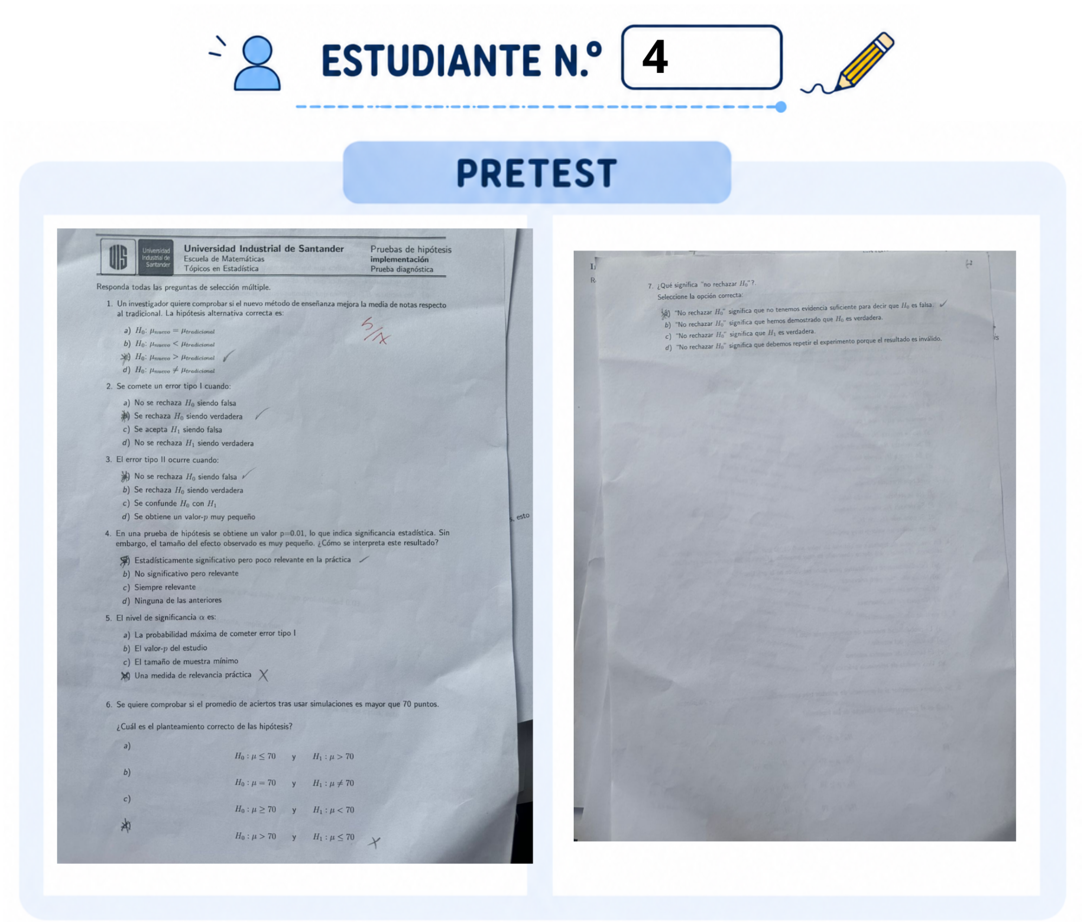
Instrucciones

7/7

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

Figura 23: *Comparación entre el pretest y postest del estudiante 4*



Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística
Pruebas de hipótesis
implementación
Prueba diagnóstica

Responda todas las preguntas de selección múltiple.

5/7

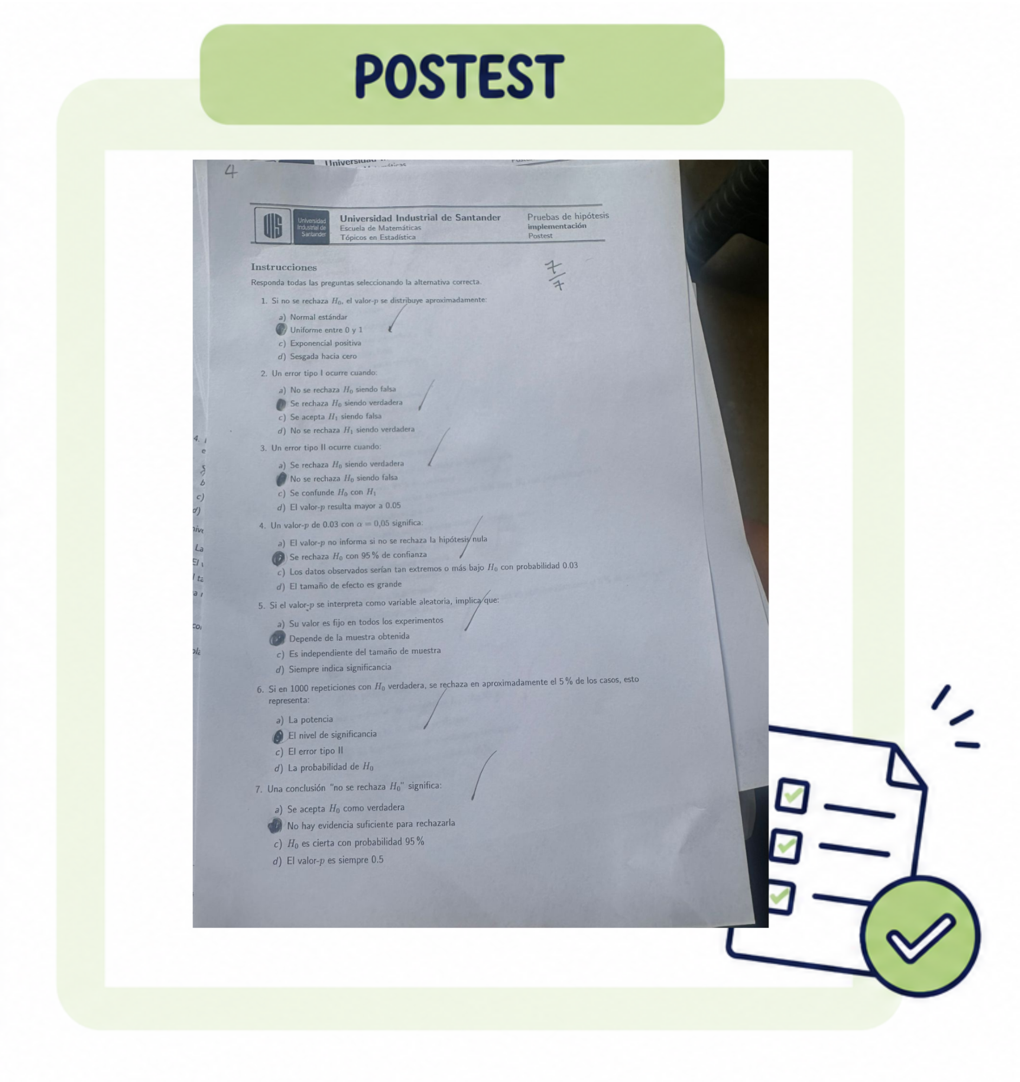
1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis alternativa correcta es:
   a) H_0: $\mu_{nuevo} = \mu_{tradicional}$
   b) H_0: $\mu_{nuevo} < \mu_{tradicional}$
   c) H_0: $\mu_{nuevo} > \mu_{tradicional}$ ✓
   d) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$
2. Se comete un error tipo I cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera ✓
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. El error tipo II ocurre cuando:
   a) No se rechaza H_0 siendo falsa ✓
   b) Se rechaza H_0 siendo verdadera
   c) Se confunde H_0 con H_1
   d) Se obtiene un valor-p muy pequeño
4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?
   a) Estadísticamente significativo pero poco relevante en la práctica ✓
   b) No significativo pero relevante
   c) Siempre relevante
   d) Ninguna de las anteriores
5. El nivel de significancia α es:
   a) La probabilidad máxima de cometer error tipo I
   b) El valor-p del estudio
   c) El tamaño de muestra mínimo
   d) Una medida de relevancia práctica X
6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

   ¿Cuál es el planteamiento correcto de las hipótesis?
   a) $H_0 : \mu \leq 70$ y $H_1 : \mu > 70$
   b) $H_0 : \mu = 70$ y $H_1 : \mu \neq 70$
   c) $H_0 : \mu \geq 70$ y $H_1 : \mu < 70$
   d) $H_0 : \mu > 70$ y $H_1 : \mu \leq 70$ X

7. ¿Qué significa "no rechazar H_0"?.

   Seleccione la opción correcta:
   a) "No rechazar H_0" significa que no tenemos evidencia suficiente para decir que H_0 es falsa. ✓
   b) "No rechazar H_0" significa que hemos demostrado que H_0 es verdadera.
   c) "No rechazar H_0" significa que H_1 es verdadera.
   d) "No rechazar H_0" significa que debemos repetir el experimento porque el resultado es inválido.

Figura 24: *Comparación postest del estudiante 4*


POSTEST

4

Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Postest

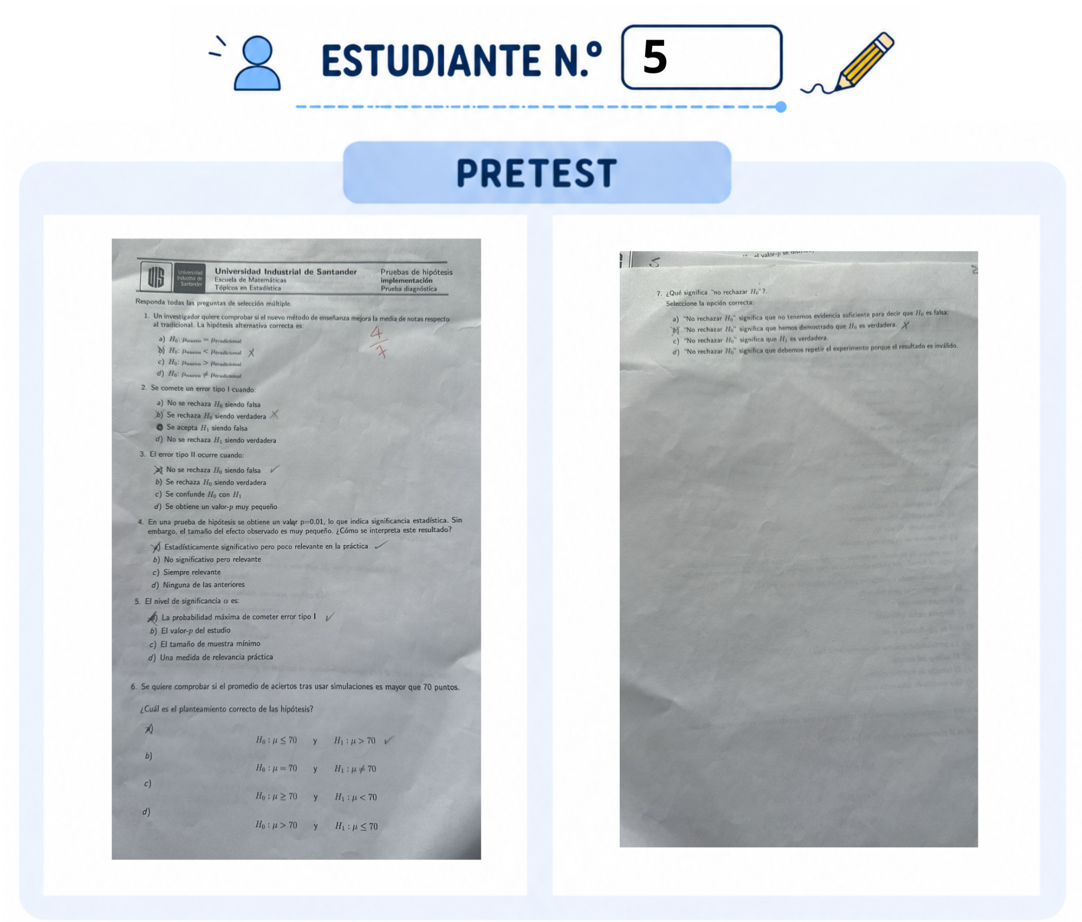
Instrucciones

7/7

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

Figura 25: *Comparación entre el pretest y postest del estudiante 5*



Universidad Industrial de Santander — Escuela de Matemáticas — Tópicos en Estadística
Pruebas de hipótesis — **implementación** — Prueba diagnóstica

Responda todas las preguntas de selección múltiple.

4/7

1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis alternativa correcta es:
   a) H_0: $\mu_{nuevo} = \mu_{tradicional}$
   b) H_0: $\mu_{nuevo} < \mu_{tradicional}$ ✗
   c) H_0: $\mu_{nuevo} > \mu_{tradicional}$
   d) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$
2. Se comete un error tipo I cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera ✗
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. El error tipo II ocurre cuando:
   a) No se rechaza H_0 siendo falsa ✓
   b) Se rechaza H_0 siendo verdadera
   c) Se confunde H_0 con H_1
   d) Se obtiene un valor-p muy pequeño
4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?
   a) Estadísticamente significativo pero poco relevante en la práctica ✓
   b) No significativo pero relevante
   c) Siempre relevante
   d) Ninguna de las anteriores
5. El nivel de significancia α es:
   a) La probabilidad máxima de cometer error tipo I ✓
   b) El valor-p del estudio
   c) El tamaño de muestra mínimo
   d) Una medida de relevancia práctica
6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

   ¿Cuál es el planteamiento correcto de las hipótesis?
   a) $H_0 : \mu \leq 70$ y $H_1 : \mu > 70$ ✓
   b) $H_0 : \mu = 70$ y $H_1 : \mu \neq 70$
   c) $H_0 : \mu \geq 70$ y $H_1 : \mu < 70$
   d) $H_0 : \mu > 70$ y $H_1 : \mu \leq 70$

5

7. ¿Qué significa "no rechazar H_0"?.
   Seleccione la opción correcta:
   a) "No rechazar H_0" significa que no tenemos evidencia suficiente para decir que H_0 es falsa.
   b) "No rechazar H_0" significa que hemos demostrado que H_0 es verdadera. ✗
   c) "No rechazar H_0" significa que H_1 es verdadera.
   d) "No rechazar H_0" significa que debemos repetir el experimento porque el resultado es inválido.

Figura 26: *Comparación postest del estudiante 5*

POSTEST

a) Normal estándar

5

Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Postest

7/7

Instrucciones

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

Figura 27: *Comparación entre el pretest y postest del estudiante 6*

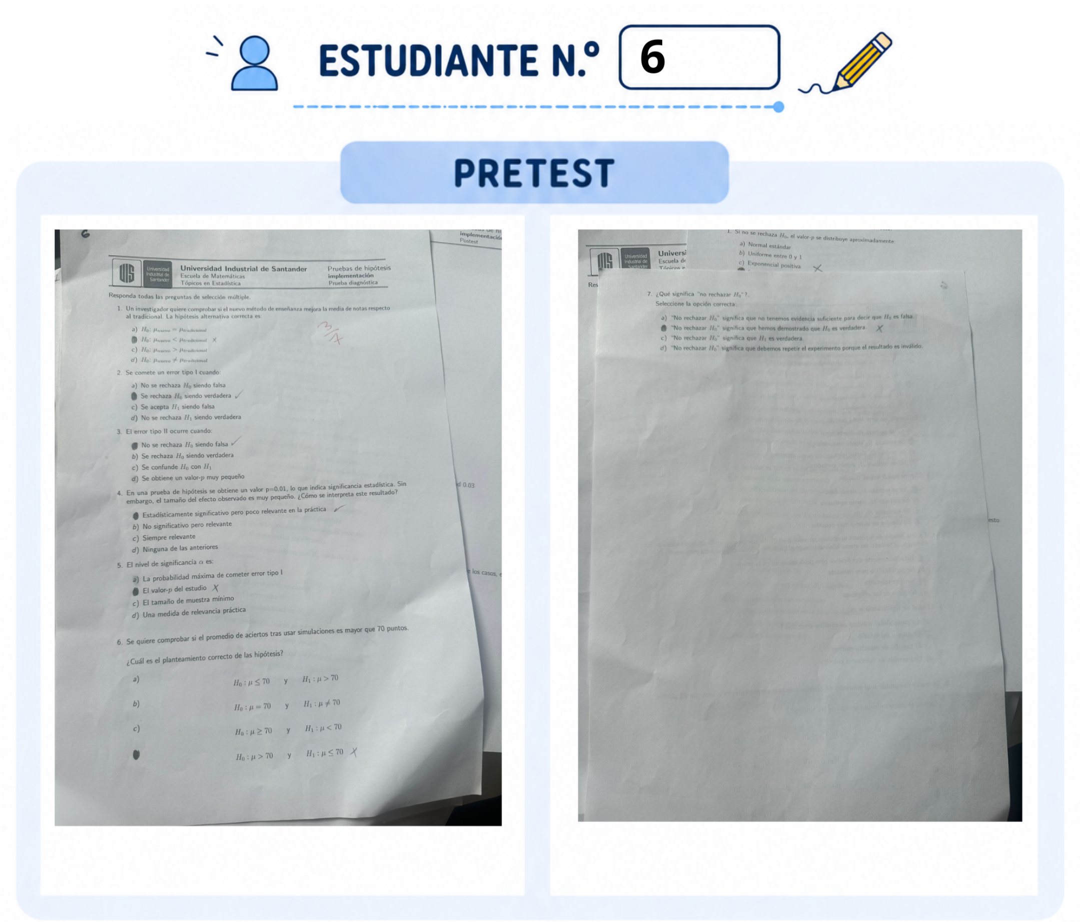
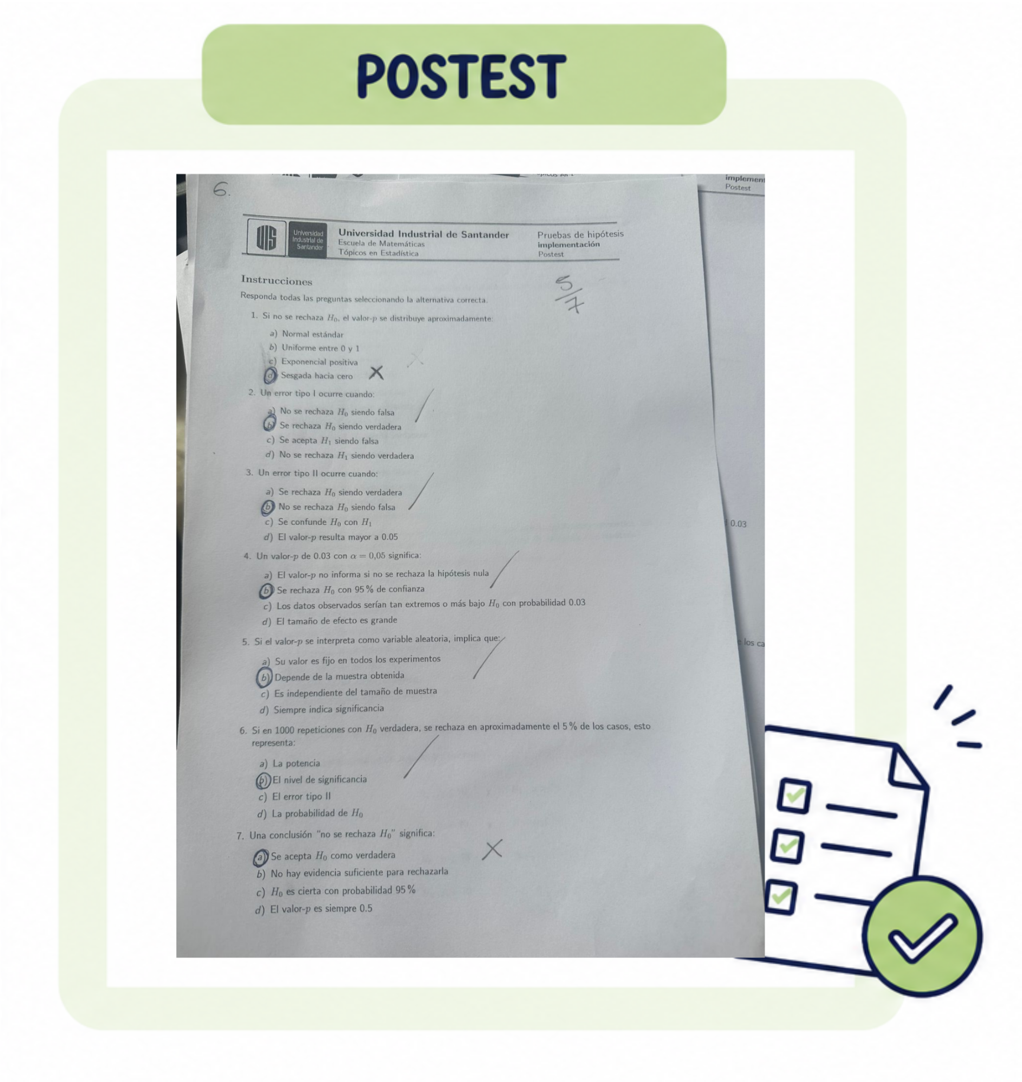


Figura 28: *Comparación postest del estudiante 6*


POSTEST

6.

Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Postest

Instrucciones

5/7

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

Figura 29: *Comparación entre el prestest y postest del estudiante 7*

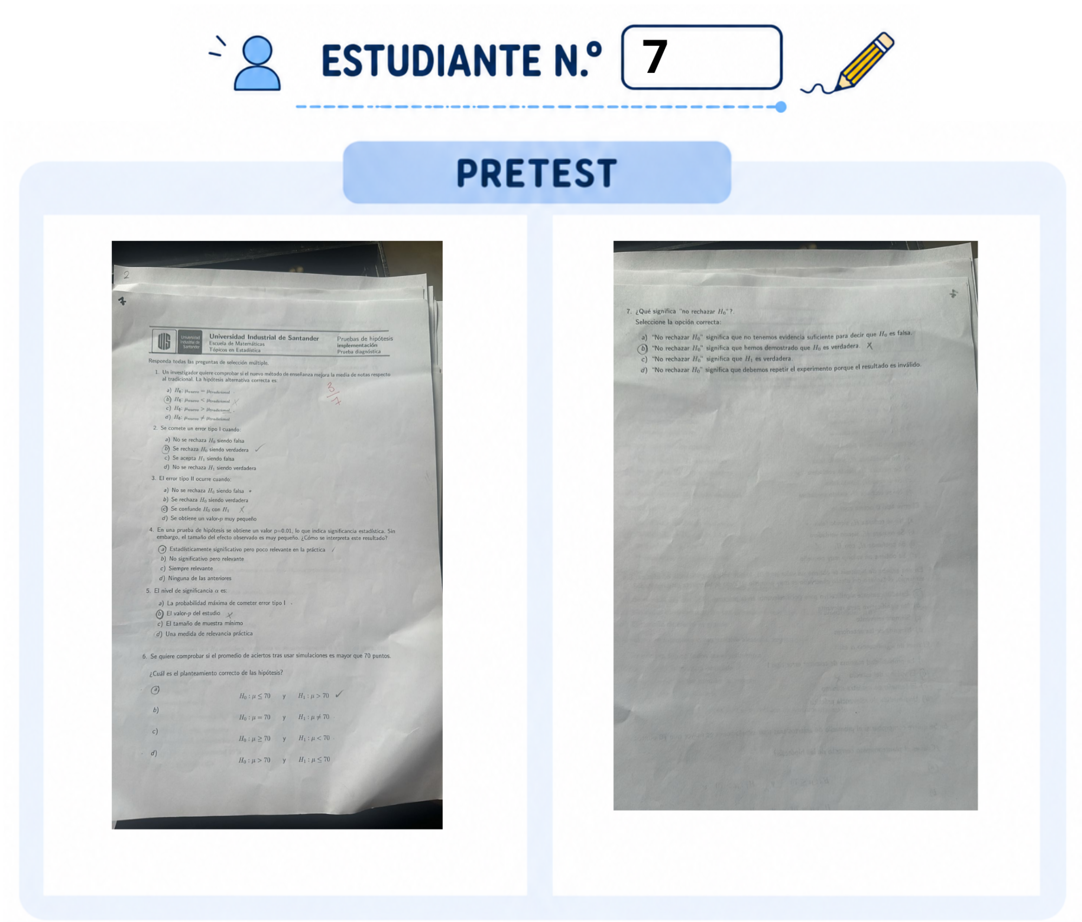


Figura 30: *Comparación postest del estudiante 7*

Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

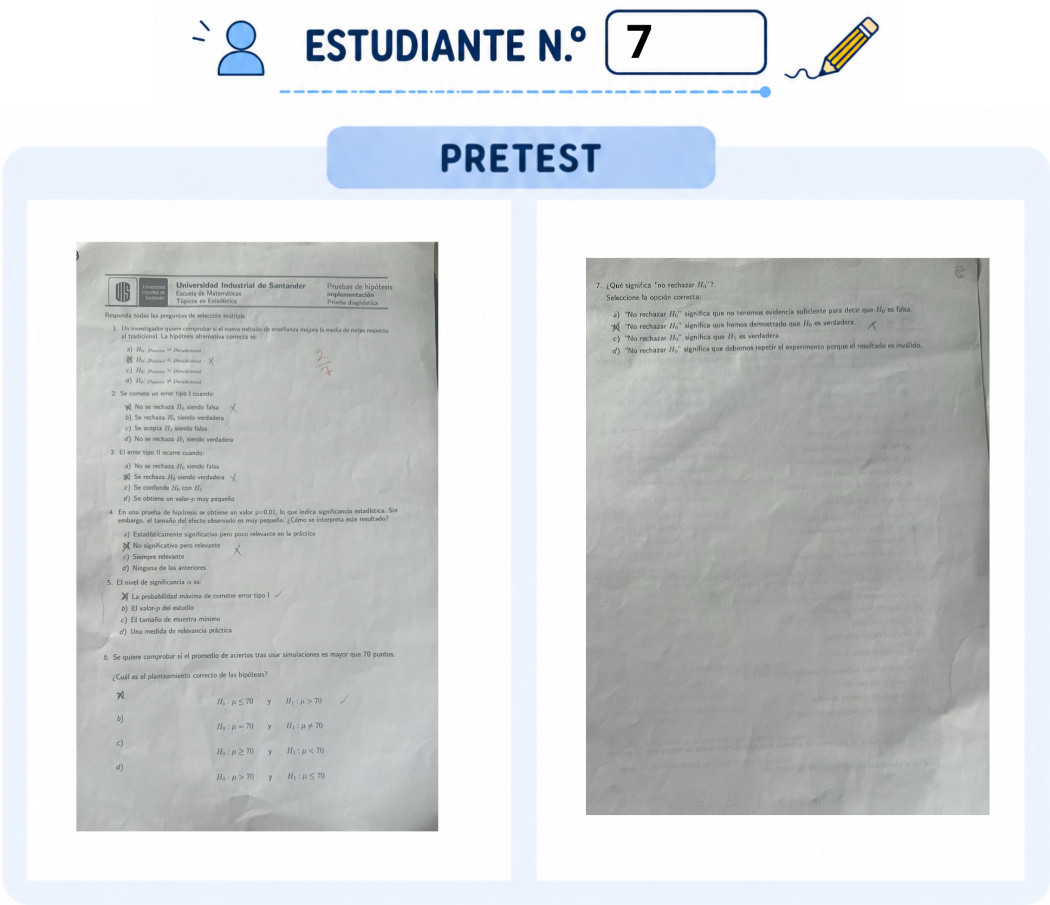
Pruebas de hipótesis
implementación
Postest

Instrucciones

Responda todas las preguntas seleccionando la alternativa correcta.

1. Si no se rechaza H_0, el valor-p se distribuye aproximadamente:
   a) Normal estándar
   b) Uniforme entre 0 y 1
   c) Exponencial positiva
   d) Sesgada hacia cero
2. Un error tipo I ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. Un error tipo II ocurre cuando:
   a) Se rechaza H_0 siendo verdadera
   b) No se rechaza H_0 siendo falsa
   c) Se confunde H_0 con H_1
   d) El valor-p resulta mayor a 0.05
4. Un valor-p de 0.03 con $\alpha = 0{,}05$ significa:
   a) El valor-p no informa si no se rechaza la hipótesis nula
   b) Se rechaza H_0 con 95 % de confianza
   c) Los datos observados serían tan extremos o más bajo H_0 con probabilidad 0.03
   d) El tamaño de efecto es grande
5. Si el valor-p se interpreta como variable aleatoria, implica que:
   a) Su valor es fijo en todos los experimentos
   b) Depende de la muestra obtenida
   c) Es independiente del tamaño de muestra
   d) Siempre indica significancia
6. Si en 1000 repeticiones con H_0 verdadera, se rechaza en aproximadamente el 5 % de los casos, esto representa:
   a) La potencia
   b) El nivel de significancia
   c) El error tipo II
   d) La probabilidad de H_0
7. Una conclusión "no se rechaza H_0" significa:
   a) Se acepta H_0 como verdadera
   b) No hay evidencia suficiente para rechazarla
   c) H_0 es cierta con probabilidad 95 %
   d) El valor-p es siempre 0.5

Figura 31: *Comparación entre el pretest y postest del estudiante 8*



Universidad Industrial de Santander
Escuela de Matemáticas
Tópicos en Estadística

Pruebas de hipótesis
implementación
Prueba diagnóstica

Responda todas las preguntas de selección múltiple.

1. Un investigador quiere comprobar si el nuevo método de enseñanza mejora la media de notas respecto al tradicional. La hipótesis alternativa correcta es:
   a) H_0: $\mu_{nuevo} = \mu_{tradicional}$
   b) H_0: $\mu_{nuevo} < \mu_{tradicional}$ X
   c) H_0: $\mu_{nuevo} > \mu_{tradicional}$
   d) H_0: $\mu_{nuevo} \neq \mu_{tradicional}$

2/7

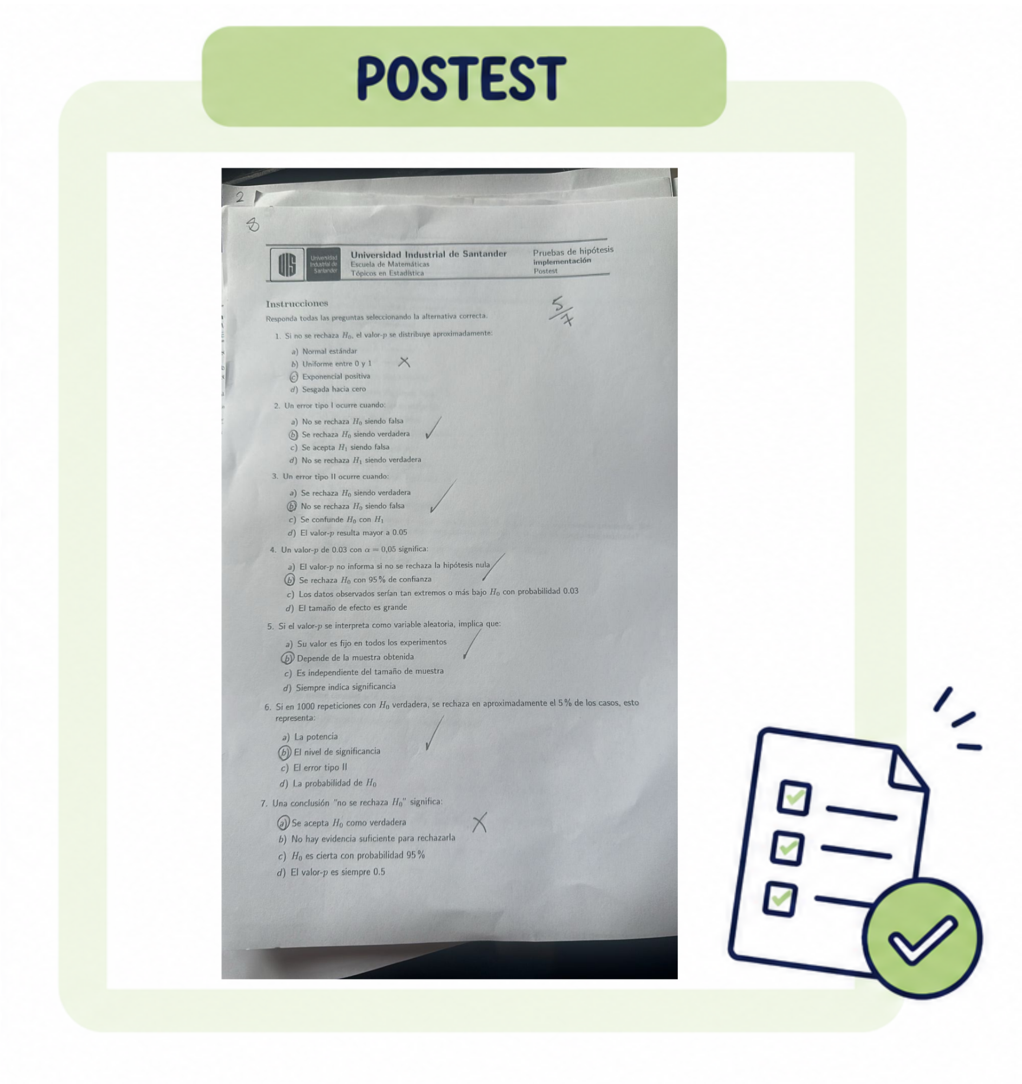
2. Se comete un error tipo I cuando:
   a) No se rechaza H_0 siendo falsa X
   b) Se rechaza H_0 siendo verdadera
   c) Se acepta H_1 siendo falsa
   d) No se rechaza H_1 siendo verdadera
3. El error tipo II ocurre cuando:
   a) No se rechaza H_0 siendo falsa
   b) Se rechaza H_0 siendo verdadera X
   c) Se confunde H_0 con H_1
   d) Se obtiene un valor-p muy pequeño
4. En una prueba de hipótesis se obtiene un valor p=0.01, lo que indica significancia estadística. Sin embargo, el tamaño del efecto observado es muy pequeño. ¿Cómo se interpreta este resultado?
   a) Estadísticamente significativo pero poco relevante en la práctica
   b) No significativo pero relevante X
   c) Siempre relevante
   d) Ninguna de las anteriores
5. El nivel de significancia α es:
   a) La probabilidad máxima de cometer error tipo I ✓
   b) El valor-p del estudio
   c) El tamaño de muestra mínimo
   d) Una medida de relevancia práctica
6. Se quiere comprobar si el promedio de aciertos tras usar simulaciones es mayor que 70 puntos.

   ¿Cuál es el planteamiento correcto de las hipótesis?
   a) $H_0 : \mu \leq 70$ y $H_1 : \mu > 70$ ✓
   b) $H_0 : \mu = 70$ y $H_1 : \mu \neq 70$
   c) $H_0 : \mu \geq 70$ y $H_1 : \mu < 70$
   d) $H_0 : \mu > 70$ y $H_1 : \mu \leq 70$

7. ¿Qué significa "no rechazar H_0"?.

   Seleccione la opción correcta:
   a) "No rechazar H_0" significa que no tenemos evidencia suficiente para decir que H_0 es falsa.
   b) "No rechazar H_0" significa que hemos demostrado que H_0 es verdadera. X
   c) "No rechazar H_0" significa que H_1 es verdadera.
   d) "No rechazar H_0" significa que debemos repetir el experimento porque el resultado es inválido.

Figura 32: *Comparación postest del estudiante 8*



Apéndice E: intervención didáctica

Descripción de la intervención didáctica

La intervención didáctica desarrollada en esta investigación tuvo como propósito fortalecer el razonamiento estadístico de los estudiantes en torno a las pruebas de hipótesis mediante el uso de simulaciones en RStudio. La implementación se realizó en tres fases:

Aplicación del pretest

En esta fase se aplicó el instrumento diagnóstico (ver anexo 6) con el objetivo de identificar los conocimientos previos y las dificultades iniciales de los estudiantes relacionadas con las pruebas de hipótesis, el valor p, el nivel de significancia y los errores tipo I y tipo II.

Desarrollo de la intervención mediada por simulaciones

Posteriormente, se desarrolló la intervención didáctica mediante sesiones teórico prácticas apoyadas en diapositivas explicativas y simulaciones en RStudio. Inicialmente, se trabajó el concepto de muestra aleatoria simple, explicando cómo una muestra puede representar a una población y cómo interviene el azar en los procesos de selección. A partir de ello, los estudiantes analizaron ejemplos relacionados con la probabilidad de seleccionar subconjuntos de una población y la importancia del muestreo en los procesos estadísticos.

Seguidamente, se abordó el tema de las pruebas de hipótesis, explicando la diferencia entre hipótesis nula e hipótesis alternativa, así como la estructura general de una prueba estadística. Durante esta etapa, los estudiantes analizaron diferentes situaciones para identificar correctamente las hipótesis y comprender su función dentro de la toma de decisiones.

Posteriormente, se trabajaron los errores tipo I y tipo II mediante ejemplos y representaciones gráficas que permitieron comprender las consecuencias de rechazar o no rechazar la hipótesis nula de manera incorrecta. Asimismo, se explicó el nivel de significancia como la probabilidad máxima aceptada para cometer un error tipo I, favoreciendo la comprensión de su relación con la toma de decisiones estadísticas.

La intervención también incluyó el estudio del error estándar y del estadístico de prueba, utilizando simulaciones para observar cómo cambian los resultados dependiendo del tamaño de muestra y la variabilidad de los datos. A través de estas actividades, los estudiantes pudieron visualizar el comportamiento de las distribuciones muestrales y comprender el procedimiento utilizado para contrastar hipótesis.

Finalmente, se trabajó la interpretación del valor p mediante simulaciones y ejemplos prácticos. Los estudiantes analizaron cómo este valor permite determinar si existe suficiente evidencia para rechazar la hipótesis nula, fortaleciendo así la interpretación y argumentación de resultados estadísticos.

Aplicación del postest

Finalmente, se aplicó un postest (ver anexo 6) con características similares al pretest, con el propósito de identificar los avances alcanzados por los estudiantes después de la implementación de la intervención didáctica y analizar los cambios en sus habilidades de razonamiento estadístico.