

UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE INGENIERÍA ELÉCTRICA Y
ELECTRÓNICA
UNIDAD DE INVESTIGACIÓN



INFORME FINAL DEL PROYECTO DE INVESTIGACIÓN

**“USO DEL SOFTWARE R EN EL APRENDIZAJE DE
PROBABILIDADES EN ESTUDIANTES DE LA FACULTAD DE
INGENIERÍA ELÉCTRICA Y ELECTRÓNICA DE LA UNIVERSIDAD
NACIONAL DEL CALLAO, 2024”**

AUTOR: Dr. Lic. ADÁN ALMIRCAR, TEJADA CABANILLAS

(PERIODO DE EJECUCIÓN: Del 01 de marzo al 31 de diciembre de 2024)

(Resolución de aprobación N° 318-2024)

A handwritten signature in blue ink, reading "Adán Almircar Tejada Cabanillas".

A handwritten signature in blue ink, reading "Adán Almircar Tejada Cabanillas".

Bellavista, Callao

2024

A handwritten signature in blue ink, featuring a large, stylized initial 'S' or 'D' at the top, followed by several loops and a long horizontal stroke extending to the right.

Handwritten signature

DEDICATORIA

ESTE TRABAJO DE INVESTIGACION
LO DEDICO A TODOS LOS USUARIOS
QUE HACEN INVESTIGACION COMO
ALTERNATIVA PARA EL PROCESO DE
DATOS CON SOFTWARE R

A handwritten signature in blue ink, appearing to read 'Andrés Cepeda', is located at the bottom left of the page. The signature is fluid and cursive, with a large initial 'A'.

AGRADECIMIENTO

EN PRIMER LUGAR A DIOS POR DARME LA
LUCIDEZ Y SALUD

EN SEGUNDO LUGAR A MI FAMILIA POR
COMPRENDER Y ENTENDER EL USO DE MI
TIEMPO EN EL QUEHACER DE ESTA
INVESTIGACION.

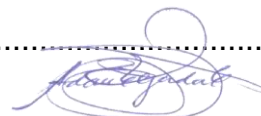
IGUALMENTE A TODOS MIS COLEGAS QUE ME
IMPULSARON HACER ESTE TRABAJO DE
INVESTIGACION

FINALMENTE AL VICERRECTORADO DE
INVESTIGACION POR PROGRAMAR CURSOS DE
CAPACITACION, EN ESTE CASO EL R

A handwritten signature in blue ink, appearing to read "Andrés Espinal", is located at the bottom left of the page. The signature is fluid and cursive, with a large initial 'A'.

ÍNDICE

DEDICATORIA.....	3
AGRADECIMIENTO.....	4
TABLAS DE CONTENIDO	3
RESUMEN	5
ABSTRACT	6
INTRODUCCIÓN.....	7
CAPITULO I. PLANTEAMIENTO DEL PROBLEMA	8
1.1. Descripción de la realidad problemática	8
1.2. Formulación del problema	10
1.2.1. Problema general	10
1.2.2. Problemas específicos.....	10
1.3. Objetivos	10
1.3.1. Objetivo general.....	10
1.3.2. Objetivos específicos.....	10
1.4. Limitantes de la investigación.....	11
1.4.1. Limitante teórica	11
1.4.2. Limitante espacial	11
1.4.3. Limitante temporal	11
CAPITULO II. MARCO TEÓRICO.....	13
2.1. Antecedentes	13
2.1.1. Antecedentes internacionales.....	13
2.1.2. Antecedentes nacionales.....	16
2.2. Marco	19
2.2.1. Teórico.....	19
2.2.2. Conceptual.....	29
2.3. Definición de términos básicos.....	31
CAPITULO III. HIPÓTESIS Y VARIABLES.....	32



3.1. Hipótesis	32
3.1.1. Hipótesis general	32
3.1.2. Hipótesis específicas	32
3.2. Definición conceptual de variables	32
3.3. Operacionalización de variable.....	33
CAPITULO IV. DISEÑO METODOLÓGICO.....	34
4.1. Tipo y diseño de investigación.....	34
4.2. Método de investigación.....	34
4.3. Población y muestra.....	34
4.4. Lugar de estudio y periodo desarrollado.....	36
4.5. Técnicas e instrumentos para la recolección de la información.....	36
4.6. Análisis y procesamiento de datos	36
CAPITULO V. RESULTADOS.....	38
5.1. Resultados descriptivos.....	38
5.2. Resultados inferenciales	45
CAPITULO VI. DISCUSIÓN DE RESULTADOS.....	50
6.1. Contrastación y demostración de la hipótesis con los resultados.....	50
6.2. Contrastación de los resultados con otros estudios similares.....	51
6.3. Responsabilidad ética	53
CONCLUSIONES.....	55
RECOMENDACIONES.....	56
REFERENCIAS BIBLIOGRÁFICAS	57
ANEXOS	60
ANEXO N° 02: INSTRUMENTO.....	62
ANEXO N.º 03: BASE DE DATOS	73



TABLAS DE CONTENIDO

ÍNDICE DE TABLAS

Tabla 1. Comparación de software Estadístico.....	23
Tabla 2. Operacionalización de variable.....	33
Tabla 3. Distribución de la población.....	34
Tabla 4. Distribución de la muestra	35
Tabla 5. Prueba en PRE TEST de distribuciones de probabilidad para variables discretas	38
Tabla 6. Prueba en POST TEST de distribuciones de probabilidad para variables discretas	39
Tabla 7. Prueba en PRE TEST de distribuciones de probabilidad para variables continuas	40
Tabla 8. Prueba en POST TEST de distribuciones de probabilidad para variables continuas	41
Tabla 9. Prueba en PRE TEST de Gráficos de distribuciones de probabilidad	42
Tabla 10. Prueba en POST TEST de Gráficos de distribuciones de probabilidad	43
Tabla 11. Comparativa del PRE y POST TEST.....	44
Tabla 12. Prueba de normalidad de Kolmogorov-Smirnov	45
Tabla 13. Prueba T Student para Hipótesis General	46
Tabla 14. Prueba T Student para Hipótesis Especifica 1.....	47
Tabla 15. Prueba T Student para Hipótesis Especifica 2.....	48
Tabla 16. Prueba T Student para Hipótesis Especifica 3.....	49



ÍNDICE DE GRÁFICOS

Figura 1. Software R.....	9
Figura 2. Prueba en PRE TEST de distribuciones de probabilidad para variables discretas	38
Figura 3. Prueba en POST TEST de distribuciones de probabilidad para variables discretas	39
Figura 4. Prueba en PRE TEST de distribuciones de probabilidad para variables continuas	40
Figura 5. Prueba en POST TEST de distribuciones de probabilidad para variables continuas	41
Figura 6. Prueba en PRE TEST de Gráficos de distribuciones de probabilidad.....	42
Figura 7. Prueba en POST TEST de Gráficos de distribuciones de probabilidad	43
Figura 8. Comparativa del PRE y POST TEST	44



RESUMEN

Objetivo: Determinar de qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Metodología: El tipo de investigación es descriptivo, el diseño de investigación es pre experimental y el método de investigación es cuantitativo. La población del estudio está constituida por los 80 alumnos inscritos en el curso de estadística descriptiva de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.

Resultados: En el pre-test, solo el 2% de los estudiantes alcanzó un nivel alto, mientras que el 94% se ubicó en un nivel medio y un 4% en el rango bajo. Esto indica que, antes de la intervención, la mayoría de los estudiantes tenía una comprensión básica del tema, pero un número considerable no lograba un rendimiento satisfactorio. En contraste, en el post-test, el porcentaje de estudiantes en el nivel alto aumentó al 26%, mientras que el 74% se mantuvo en el nivel medio y el 0% en el bajo. Este cambio significativo sugiere que el uso del software R ha mejorado la capacidad de los estudiantes para entender y aplicar los conceptos de probabilidad, aumentando su dominio del material. En cuanto a la hipótesis general se obtuvo el valor t de -25.531, junto con un valor p de 0,001, indica que existe una diferencia estadísticamente significativa entre las medias del pre-test y el post-test en el rendimiento de los estudiantes, permitiendo rechazar la hipótesis nula que afirmaba la ausencia de diferencia.

Conclusión: El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Palabras clave: software R, aprendizaje, probabilidades.



ABSTRACT

Objective: To determine how the use of R software improves probability learning in students of the Faculty of Electrical and Electronic Engineering of the National University of Callao, 2024.

Methodology: The type of research is descriptive, the research design is pre-experimental, and the research method is quantitative. The study population is made up of the 80 students enrolled in the descriptive statistics course at the Faculty of Electrical and Electronic Engineering of the National University of Callao.

Results: In the pre-test, only 2% of the students reached a high level, while 94% were at a medium level and 4% in the low range. This indicates that, before the intervention, most of the students had a basic understanding of the subject, but a considerable number did not achieve satisfactory performance. In contrast, in the post-test, the percentage of students at the high level increased to 26%, while 74% remained at the medium level and 0% at the low level. This significant change suggests that the use of R software has improved students' ability to understand and apply probability concepts, increasing their mastery of the material. Regarding the general hypothesis, the t-value of -25.531 was obtained, along with a p-value of 0.001, indicating that there is a statistically significant difference between the pre-test and post-test means in student performance, allowing us to reject the null hypothesis that stated the absence of a difference.

Conclusion: The use of R software improves the learning of probabilities in students of the Faculty of Electrical and Electronic Engineering of the National University of Callao, 2024.

Keywords: R software, learning, probabilities.



INTRODUCCIÓN

En el ámbito académico, el aprendizaje de probabilidades ha sido reconocido como fundamental para desarrollar el pensamiento lógico y analítico en disciplinas científicas y técnicas. En la formación de ingenieros, la capacidad para modelar incertidumbre y tomar decisiones basadas en datos resultó ser una competencia esencial. Sin embargo, los enfoques tradicionales en la enseñanza de probabilidades enfrentaron desafíos significativos, como la dificultad para conectar conceptos teóricos con aplicaciones prácticas y la escasa integración de herramientas tecnológicas en los procesos educativos.

Ante este panorama, el Software R demostró ser una herramienta poderosa, ampliamente utilizada en el análisis estadístico y la programación, al ofrecer un entorno dinámico y práctico que facilitó la enseñanza y el aprendizaje de probabilidades. Su capacidad para simular eventos aleatorios, visualizar distribuciones y resolver problemas complejos de manera interactiva lo posicionó como un recurso valioso para complementar la formación teórica con experiencias prácticas.

La presente investigación exploró el impacto del uso del Software R en el aprendizaje de probabilidades en estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao. Se evaluó cómo esta herramienta tecnológica contribuyó a mejorar la comprensión de conceptos probabilísticos y a fomentar habilidades analíticas en un contexto académico.

El estudio, desarrollado en el año 2024, tuvo como propósito identificar los beneficios pedagógicos del uso del Software R y ofrecer una propuesta metodológica innovadora para optimizar los procesos de enseñanza en la educación superior, específicamente en las carreras de ingeniería.



CAPITULO I. PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

A nivel mundial, la enseñanza y aprendizaje de conceptos estadísticos y probabilísticos en ingeniería enfrenta desafíos significativos debido a la tradicionalidad de los métodos de enseñanza, que suelen ser estáticos y poco interactivos, lo que dificulta la comprensión y aplicación práctica por parte de los estudiantes. Esta brecha entre teoría y práctica se acentúa en el campo de la ingeniería eléctrica y electrónica, donde los estudiantes deben tomar decisiones basadas en datos, pero carecen de habilidades y herramientas adecuadas. La falta de integración de tecnologías avanzadas en la enseñanza de probabilidades limita aún más la preparación de los estudiantes para el mundo laboral. Por lo tanto, es imperativo explorar estrategias pedagógicas innovadoras que incorporen herramientas como el software R para mejorar la comprensión y aplicación de conceptos estadísticos, preparando así a los estudiantes para los desafíos de un entorno laboral cada vez más tecnológico y data-driven.

En el contexto peruano, la enseñanza de conceptos estadísticos y probabilísticos en ingeniería enfrenta desafíos particulares derivados de las limitaciones en los recursos tecnológicos y la capacitación docente. La tradicionalidad en los métodos de enseñanza, enfocados en la teoría y con escasa aplicación práctica, dificulta la comprensión y el interés de los estudiantes. Además, la falta de acceso a herramientas tecnológicas especializadas, como el software R, limita las oportunidades de los estudiantes para desarrollar habilidades analíticas avanzadas y aplicar conceptos estadísticos de manera efectiva en el contexto de la ingeniería eléctrica y electrónica. Esta brecha entre la formación académica y las demandas del mercado laboral resalta la necesidad urgente de implementar estrategias pedagógicas innovadoras que integren el uso del software R para mejorar el aprendizaje de probabilidades, garantizando así una formación más sólida y relevante para los estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao en el año 2024.



En el ámbito específico de la Universidad Nacional del Callao, la enseñanza de probabilidades en la Facultad de Ingeniería Eléctrica y Electrónica enfrenta desafíos relacionados con la disponibilidad de recursos tecnológicos adecuados y la capacitación del cuerpo docente. La infraestructura limitada y la falta de acceso a herramientas especializadas, como el software R, dificultan la implementación de métodos de enseñanza más interactivos y orientados a la aplicación práctica de los conceptos estadísticos. Además, la formación de los profesores en el uso de estas tecnologías puede resultar insuficiente, lo que impacta en la calidad de la enseñanza y en la preparación de los estudiantes para enfrentar los retos del campo laboral, donde el manejo de herramientas analíticas es cada vez más valorado. Por lo tanto, es necesario abordar estas limitaciones mediante la adopción de estrategias que promuevan una mayor integración del software R en el proceso de aprendizaje de probabilidades, con el fin de mejorar la formación académica y profesional de los estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao en el año 2024.

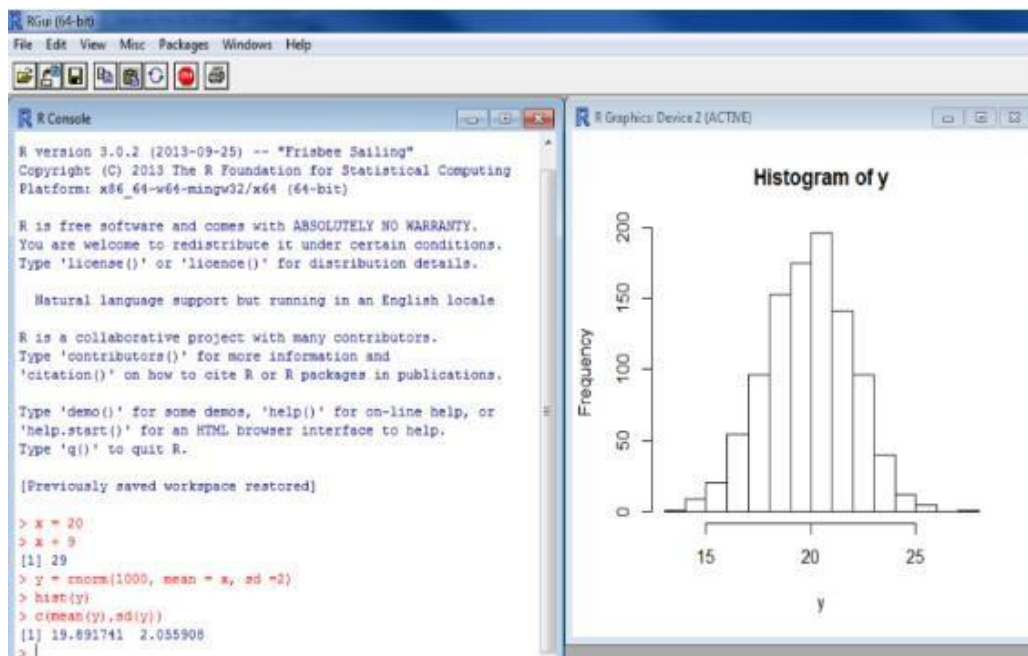


Figura 1. Software R

Por ello se plantea como objetivo, determinar de qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

1.2. Formulación del problema

1.2.1. Problema general

¿De qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?

1.2.2. Problemas específicos

- ¿De qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?
- ¿De qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?
- ¿De qué manera el uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?

1.3. Objetivos

1.3.1. Objetivo general

Determinar de qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

1.3.2. Objetivos específicos

- Determinar de qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.



- Determinar de qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.
- Determinar de qué manera el uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

1.4. Limitantes de la investigación

1.4.1. Limitante teórica

La delimitante teórica se centra en la necesidad de comprender y aplicar adecuadamente los conceptos probabilísticos utilizando herramientas tecnológicas avanzadas. A pesar de las ventajas que ofrece el software R en términos de análisis estadístico y visualización de datos, es fundamental que los estudiantes desarrollen una comprensión sólida de los fundamentos teóricos de las probabilidades para utilizar eficazmente esta herramienta en el contexto de la ingeniería eléctrica y electrónica. Por lo tanto, la delimitante teórica se enfoca en la integración armoniosa entre la teoría de probabilidades y la aplicación práctica del software R para garantizar una formación integral y efectiva en esta área.

1.4.2. Limitante espacial

La delimitante espacial establece que el proyecto se llevará a cabo exclusivamente dentro de las instalaciones de la Universidad Nacional del Callao. Todas las actividades relacionadas con la implementación del software R, la capacitación de estudiantes y docentes, la recolección de datos y el análisis de resultados se realizarán dentro del campus universitario, garantizando así un enfoque centrado en la comunidad académica de la facultad y asegurando la coherencia con los objetivos y recursos disponibles en este entorno específico.

1.4.3. Limitante temporal

El delimitante temporal establece un período máximo de 12 meses para la ejecución y evaluación del proyecto de investigación. Durante este lapso, se



llevarán a cabo todas las actividades planificadas, incluyendo la implementación del software R en el currículo académico, la realización de sesiones de capacitación para estudiantes y docentes, la recolección y análisis de datos, así como la elaboración de informes y conclusiones pertinentes. Esta delimitación temporal garantiza una gestión eficiente del proyecto y permite evaluar adecuadamente el impacto del uso del software R en el aprendizaje de probabilidades en el período especificado.



CAPITULO II. MARCO TEÓRICO

2.1. Antecedentes

2.1.1. Antecedentes internacionales

En la investigación realizada por Vargas, Arregocés, Solano y Peña (2021), titulada "Aprendizaje basado en proyectos apoyado en un diseño tecnopedagógico para la enseñanza de la estadística descriptiva", se propuso como propósito examinar cómo la incorporación de la estrategia educativa del aprendizaje basado en proyectos contribuye al fortalecimiento de la enseñanza de estadística descriptiva en el programa de Negocios Internacionales de la Universidad de La Guajira (Colombia). La metodología empleada incluyó observación directa a 21 estudiantes, análisis de contenido y evaluación estadística de los registros. Los hallazgos revelaron que las actividades apoyadas en tecnologías de la información y comunicación (TIC) facilitaron a los estudiantes un mayor manejo de datos estadísticos en contextos específicos. Asimismo, los foros virtuales promovieron la comunicación asincrónica entre los estudiantes, estimulando el aprendizaje colaborativo, mientras que los videos tutoriales dinamizaron el proceso educativo al fomentar una interacción bidireccional entre alumnos y profesores. Esta interacción, limitada en el modelo tradicional de enseñanza presencial, favoreció la resolución de dudas y amplió la comprensión de los temas tratados. En síntesis, la aplicación del aprendizaje basado en proyectos mostró ser efectiva para fortalecer la construcción del conocimiento en estadística descriptiva y mejorar las competencias de análisis estadístico entre los estudiantes del curso evaluado.

En la investigación realizada por Ibáñez, Maurandi y Castejón (2022), titulada "Enseñanza práctica virtual y adquisición de competencias en la formación estadística de maestros durante el confinamiento sanitario", se planteó como objetivo analizar la percepción de los estudiantes del Grado en Educación Primaria de una universidad pública respecto a la implementación de actividades prácticas virtuales durante el confinamiento por la pandemia de COVID-19, y su relación con el desarrollo de competencias matemático-estadísticas. El enfoque metodológico fue descriptivo, utilizando un diseño cuantitativo no experimental.



Para recopilar datos, se aplicó una encuesta elaborada específicamente para el estudio, compuesta por 10 preguntas evaluadas con una escala Likert de siete niveles. Los hallazgos reflejaron un alto nivel de aceptación por parte de los participantes sobre la pertinencia de estas actividades en su formación académica, además de expresar interés en aplicar métodos similares en su futuro desempeño profesional. También destacaron la posibilidad de adaptar dichas actividades para sus propios estudiantes. Los participantes señalaron que las actividades mejoraron su actitud hacia las matemáticas, y el análisis de los ítems específicos indicó que lograron desarrollar competencias clave relacionadas con el uso de recursos didácticos y la comunicación de ideas matemáticas. Sin embargo, se observó que aquellos estudiantes que no cursaron matemáticas durante la secundaria manifestaron una percepción menos favorable respecto al impacto de las actividades y al desarrollo de las competencias necesarias, lo que evidencia una limitada familiaridad con conceptos matemáticos, que representaba el principal objetivo de estas estrategias educativas.

El estudio realizado por Lázaro, Callejas y Griol (2022), titulado "Identificación de Factores Predictivos de Deserción Estudiantil en la Carrera de Ingeniería en Ciencias Informáticas mediante el Uso del Software SPSS", tuvo como propósito principal identificar los factores que inciden en la deserción estudiantil dentro de esta carrera, empleando el Paquete Estadístico para las Ciencias Sociales (SPSS). La investigación adoptó un enfoque cuantitativo y correlacional. Los resultados mostraron que, respecto a la hipótesis H1 sobre el género, aunque se evidenció una mayor cantidad de hombres en comparación con mujeres en la matrícula, esta diferencia no presentó una relación significativa con la deserción. Asimismo, la variable relacionada con la fuente de ingreso, donde los Institutos Preuniversitarios representaron la mayor proporción, tampoco evidenció una asociación relevante, ya que el 79% de los estudiantes provenientes de estos institutos continuaron en la carrera, mientras que el 84% de los casos de deserción también pertenecían a esta misma categoría. Del total de 485 estudiantes, 89 repitieron su primer año académico; de ellos, el 40% abandonó la carrera o cambió de disciplina durante el periodo 2014-2015, mientras que el



60% progresó al segundo año. Además, se confirmó la ausencia de una relación significativa entre estas variables, dado que el coeficiente de correlación de Spearman (r_s) fue de 0.052 con un valor p superior a 0.05. Finalmente, se concluyó que los enfoques de Ciencia, Tecnología y Sociedad (CTS) tienen un rol fundamental en el sistema educativo cubano. Este trabajo utilizó la tecnología como herramienta para analizar datos relacionados con fenómenos científicos y su impacto social, proporcionando una comprensión más amplia y generando nuevo conocimiento. También se destacó que la deserción estudiantil en la Universidad de Ciencias Informáticas (UCI) constituye un desafío que afecta tanto a la sociedad como al desarrollo científico y tecnológico del país.

El estudio realizado por Fernández (2020), titulado “Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales”, tuvo como propósito destacar las numerosas ventajas que ofrece el software R para el análisis y la visualización de datos en el ámbito de las Ciencias Sociales. La metodología empleada fue de carácter descriptivo. Los hallazgos indicaron que no es indispensable tener experiencia previa en programación ni ser programador para utilizar R. Sin embargo, una persona con experiencia en programación o conocimientos previos en este campo podrá adquirir habilidades en el manejo de R con mayor rapidez, ya que muchos lenguajes de programación comparten características similares. A pesar de ello, quienes carezcan de estos conocimientos también pueden aprender a utilizar R sin mayores dificultades. Aunque es necesario tener una base en estadística para emplear cualquier software especializado en esta disciplina, para tareas como crear gráficos de barras o calcular promedios no se requiere ser un “experto en estadística”. En conclusión, al igual que otros programas estadísticos, el uso de R implica un aprendizaje inicial relacionado con su sintaxis, lo cual puede parecer complejo al comienzo. Sin embargo, con la práctica, su manejo se vuelve más sencillo, destacándose como una herramienta versátil, poderosa, reproducible y completamente gratuita.

La investigación llevada a cabo por Sucasaire (2021), titulada "Estadística descriptiva para trabajos de investigación", tuvo como propósito proporcionar un enfoque al desarrollo de investigaciones centrándose en los aspectos clave de



la estadística descriptiva. La metodología utilizada fue descriptiva. Los resultados indicaron que programas estadísticos como R, Stata y SPSS cuentan con diversas herramientas, no solo para la estadística descriptiva, sino también para la inferencial. No obstante, R destaca por su codificación con una sintaxis específica, que es una de sus principales fortalezas; al dominarla, se facilita el manejo de los estadísticos descriptivos de manera sencilla y sin grandes limitaciones. R es una herramienta robusta y flexible que ofrece numerosas ventajas para el análisis de datos y la estadística, lo que la convierte en una opción ideal para investigadores, analistas y científicos de datos en distintas áreas. En conclusión, la integración de la estadística descriptiva con R no solo incrementa la precisión y eficiencia en el análisis de datos, sino que también potencia la capacidad de los investigadores para interpretar y presentar sus resultados. La combinación de herramientas avanzadas, visualización efectiva y un entorno flexible y reproducible, hace de R una alternativa perfecta para realizar análisis descriptivos en diversas disciplinas y situaciones. Esta investigación resalta la relevancia de adoptar R como una herramienta esencial en el análisis estadístico, destacando su capacidad para convertir los datos en información útil y aplicable.

2.1.2. Antecedentes nacionales

En el análisis realizado por Mamani y Díaz (2017), titulado "Percepciones sobre la estadística descriptiva y su impacto en el aprendizaje en estudiantes de quinto grado de secundaria de la I.E. 'Héroes del Cenepa' en el distrito de El Tambo, Huancayo", el objetivo principal fue examinar la relación entre las actitudes hacia el estudio de la estadística descriptiva y el rendimiento académico de los estudiantes de quinto grado de secundaria en dicho centro educativo. La metodología empleada fue explicativa y correlacional. Los resultados mostraron que la mayoría de los estudiantes encuestados tenían una edad promedio de 16 años, destacando un 34% (17 estudiantes) con 15 años, seguido por el 26% (13 estudiantes) con 17 años, uno (2%) con 18 años y otro (2%) con 14 años. Además, se encontró que el 48% de los encuestados estaba completamente de acuerdo en que consideran aburridos los conocimientos sobre estadística, mientras que el 50% expresó un notable temor hacia esta materia. También se



observó que el 58% de los participantes se mostró totalmente de acuerdo en que experimentan dificultades para pensar con claridad frente a problemas estadísticos, el 34% admitió sentirse muy nervioso al tratar con esta disciplina, y el 40% indicó sentirse incómodo con ella. Como conclusión principal, se determinó que existe una correlación positiva y significativa entre las actitudes hacia la estadística descriptiva y el rendimiento académico de los estudiantes, lo cual fue confirmado mediante el análisis de la prueba T de Student ($t_c=2,364$) con un nivel de significancia de $\alpha=0,05$ y 48 grados de libertad.

En el estudio realizado por Lindo (2020), titulado “El video tutorial YouTube en el aprendizaje de la estadística descriptiva en estudiantes de la Facultad de Contabilidad de la Universidad Nacional Agraria de la Selva, Tingo María 2018”, se propuso como objetivo analizar la influencia del video tutorial de YouTube en el aprendizaje de la Estadística Descriptiva en los estudiantes de dicha facultad. La investigación fue de nivel explicativo y experimental, utilizando un diseño cuasi experimental. Los resultados indicaron que el promedio final del grupo experimental fue de 17,38 puntos, lo que demuestra un logro satisfactorio en los aprendizajes. Por otro lado, el grupo de control obtuvo un promedio final de 12,82 puntos, lo que refleja dificultades en su aprendizaje. El nivel de desarrollo del aprendizaje de los estudiantes del grupo experimental, respecto al curso de Estadística Descriptiva al finalizar la investigación, mostró una mejora significativa gracias al uso del video tutorial de YouTube, en comparación con el grupo de control, que utilizó el método tradicional. En conclusión, se determinó que el video tutorial de YouTube tiene una influencia significativa en el aprendizaje de la Estadística Descriptiva en los estudiantes de la Facultad de Contabilidad de la Universidad Nacional Agraria de la Selva, ya que los puntajes obtenidos en el pos-test por el grupo experimental fueron superiores a los del grupo de control, con un nivel de significancia de 0,05.

En el estudio llevado a cabo por Benavides (2023), titulado “Modelo híbrido para la enseñanza de la estadística descriptiva e inferencial en un contexto de pandemia (COVID-19)”, el principal objetivo fue proponer un modelo híbrido para la enseñanza de ambas ramas de la estadística en el contexto de la pandemia (COVID-19) en la UNPRG-Lambayeque-Perú. La metodología empleada adoptó



un enfoque mixto dentro del paradigma positivo-interpretativo-socio-crítico. Los hallazgos indicaron que se diagnosticó la enseñanza de la estadística descriptiva e inferencial durante la pandemia. Se observó que el 100% de los docentes de estadística emplean algún tipo de software estadístico en la enseñanza de estas disciplinas, lo que evidencia un notable interés por esta herramienta. Además, el 83.3% de los profesores encuestados consideraron que la teoría es más relevante que la práctica, sugiriendo que la estadística debe ser enseñada con un enfoque más teórico que práctico. Un 75% de los docentes mencionaron que aplican estrategias para motivar a sus estudiantes durante las clases virtuales, mientras que el 91.7% de los docentes indicaron que los estudiantes participaban poco en dichas clases. Se concluyó, tras contrastar la hipótesis planteada con un 95% de nivel de confianza y utilizando una prueba paramétrica por cumplir con la normalidad de los datos, con un p-valor de 0.00000000000013, que existe una diferencia estadísticamente significativa en el rendimiento de los estudiantes en estadística descriptiva e inferencial entre ambos métodos.

En el trabajo de investigación de Ramos (2019), titulado "Los desafíos y oportunidades de la educación estadística en el nivel universitario", el objetivo principal fue analizar los factores que influyen en los procesos de enseñanza que impactan el aprendizaje de la estadística en el ámbito universitario. La investigación adoptó un enfoque descriptivo y un diseño no experimental. Los hallazgos revelaron que los modelos didácticos no surgen de manera natural, sino que son producto de las prácticas pedagógicas, la organización didáctica y los esfuerzos colaborativos. Se abordó la implementación de modelos culturales en la enseñanza de la estadística, los cuales se enfocan principalmente en la alfabetización estadística. Dichos modelos se basan en lo que ha sido validado culturalmente, lo que implica que el conocimiento adquirido se reproduce. Se estableció una diferencia entre los modelos interpretativos (MI), que proporcionan información procesada para que el estudiante la analice, y los modelos operativos (MO), que implican el análisis exploratorio y operativo de datos mediante estadígrafos de tendencia central para describir la información. Además, se introdujo el modelo productivo investigativo (MPI), que involucra la creación de información original a través de metodologías de investigación. La



conclusión subrayó que la estadística no solo se presenta como una herramienta para procesar datos, sino también como una forma de entender el mundo, promoviendo decisiones basadas en evidencia y la solución de problemas cotidianos, tanto a nivel individual como social.

En la investigación realizada por Fernández (2020) titulada "Beneficios de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales", se estableció como objetivo identificar las ventajas que brinda R para el análisis y la visualización de datos en el campo de las Ciencias Sociales. La metodología aplicada fue descriptiva y analítica, y se centró en los estudiantes de la Facultad de Ciencias Sociales y Humanidades de la Universidad Católica de Temuco. Los resultados mostraron que no es necesario ser programador ni tener experiencia previa en programación para usar R, aunque sí es importante contar con conocimientos básicos en estadística. Asimismo, se resaltó que, a diferencia de otros programas, en R no se requiere memorizar largas líneas de código, ya que la utilización de scripts facilita enormemente el proceso. En conclusión, se determinó que R es una herramienta potente y versátil para el análisis y la visualización de datos en Ciencias Sociales, destacándose por su capacidad, replicabilidad y accesibilidad, dado que es un software gratuito y de código abierto.

2.2. Marco

2.2.1. Teórico

2.2.1.1. Programa estadístico R

Según Sánchez (2019), se trata de un conjunto completo de aplicaciones diseñadas para realizar diversas operaciones relacionadas con la manipulación de datos, cálculos y representación gráfica. Entre sus atributos se incluyen la capacidad de almacenar y manipular datos, así como operadores para calcular variables específicas en matrices indexadas. Además, ofrece múltiples herramientas para el análisis de datos y la generación de gráficos. Se destaca por su lenguaje de programación bien estructurado, que incluye condicionales, bucles, funciones recursivas y capacidades de entrada y salida de manera efectiva y sencilla. (p. 10)



Como menciona Contento (2019), R es un entorno y lenguaje de programación especializado en análisis estadístico, reconocido como uno de los más ampliamente utilizados en la investigación científica. Es particularmente prominente en áreas como aprendizaje automático, minería de datos, investigación biomédica, bioinformática y matemáticas financieras. Este software se distingue por su versatilidad, potencia y nivel profesional, lo que lo convierte en una herramienta adecuada para una amplia gama de tareas estadísticas, desde las más básicas hasta las más complejas. Además, se beneficia de ser respaldado por una comunidad académica global y de estar disponible de manera gratuita, con un proceso de descarga e instalación accesible y fácil de realizar. (p. 20)

Historia y Desarrollo

Origen y Creadores

El software estadístico R fue concebido por Ross Ihaka y Robert Gentleman, dos profesores del Departamento de Estadística de la Universidad de Auckland, Nueva Zelanda, a principios de los años 90. Su objetivo principal era desarrollar una herramienta que facilitara tanto la enseñanza como la investigación en estadística, algo que las soluciones existentes en ese momento, como S y SAS, no lograban de manera completamente satisfactoria.

- Ross Ihaka: Profesor de estadística con interés en el desarrollo de software para análisis de datos. Se ha destacado por su capacidad para simplificar conceptos complejos y su enfoque en la accesibilidad del software para usuarios no expertos.
- Robert Gentleman: Estadístico y bioinformático, cuya experiencia en el análisis de datos biológicos ha influido significativamente en el desarrollo de R, particularmente en áreas como la bioestadística.

Primeras Versiones y Filosofía del Proyecto

El desarrollo de R comenzó en 1993, inspirado en el lenguaje de programación S, creado por John Chambers y sus colegas en Bell Laboratories. S había sido diseñado para la manipulación de datos y la implementación de métodos



estadísticos, pero era software propietario. Ihaka y Gentleman querían crear una versión gratuita y de código abierto, accesible a una comunidad más amplia.

- **Primeras Versiones:** Las primeras versiones de R fueron desarrolladas y distribuidas dentro de la Universidad de Auckland. Estas versiones iniciales ya incluían muchas de las características que han hecho de R un software popular, como la capacidad de manipulación de datos y la generación de gráficos de alta calidad.
- **Filosofía de Open Source:** Desde el principio, R fue concebido como un proyecto de código abierto. Esta filosofía permitió a los usuarios no solo utilizar el software de manera gratuita, sino también contribuir a su desarrollo y mejora continua.

Desarrollo y Crecimiento

El primer anuncio público de R se realizó en 1995 a través de la lista de correo de la comunidad estadística, lo que marcó el inicio de su distribución más amplia fuera de Nueva Zelanda. En 1997, se formó el R Core Group, un equipo de desarrolladores que se encargó de la dirección y el mantenimiento del proyecto.

- **R Core Group:** Este grupo está compuesto por desarrolladores voluntarios de diversas partes del mundo. El R Core Group ha sido crucial para la evolución del software, coordinando los esfuerzos de desarrollo, integrando nuevas funcionalidades y asegurando la calidad y estabilidad del software.
- **Adopción y Popularidad:** A medida que R comenzó a ganar popularidad en la academia, su uso se extendió rápidamente a otros sectores. Su adopción se vio impulsada por su capacidad de manejar datos complejos, su extensibilidad mediante paquetes, y su flexibilidad para integrarse con otros lenguajes de programación y herramientas de análisis.

Características Principales

Lenguaje de Programación



R es un lenguaje de programación de alto nivel, diseñado específicamente para el análisis estadístico y gráfico. Combina características de otros lenguajes como S y Scheme, permitiendo a los usuarios escribir scripts y programas para una amplia variedad de tareas estadísticas.

- **Sintaxis:** R utiliza una sintaxis sencilla y coherente, lo que facilita su aprendizaje y uso. Aunque tiene una curva de aprendizaje empinada para usuarios sin experiencia previa en programación, su diseño está orientado a facilitar la escritura de código legible y mantenible.
- **Funcionalidad:** R es un lenguaje funcional, lo que significa que los usuarios pueden crear funciones personalizadas para reutilizar código y simplificar análisis complejos. También soporta programación orientada a objetos, permitiendo un manejo más estructurado y modular del código.

Open Source

R es un software de código abierto, licenciado bajo la GNU General Public License. Esto implica que es gratuito para su uso y distribución, y su código fuente está disponible para que cualquiera pueda examinarlo, modificarlo y mejorarlo.

- **Accesibilidad:** La naturaleza de código abierto de R lo hace accesible a una amplia audiencia, desde estudiantes hasta investigadores y profesionales en diversos campos.
- **Comunidad:** La licencia de código abierto fomenta una comunidad activa y colaborativa que contribuye al desarrollo y expansión de R mediante la creación de paquetes, la mejora del código fuente y el soporte a otros usuarios.

Extensibilidad

Una de las características más destacadas de R es su extensibilidad. Los usuarios pueden ampliar sus capacidades mediante la instalación de paquetes adicionales desarrollados por la comunidad.



- Paquetes: CRAN (Comprehensive R Archive Network) alberga miles de paquetes que añaden funcionalidades específicas a R, desde métodos estadísticos avanzados hasta herramientas de visualización y manejo de datos.
- Desarrollo de Paquetes: La comunidad de R es altamente activa en el desarrollo de nuevos paquetes. Los usuarios pueden crear y compartir sus propios paquetes, contribuyendo a la continua expansión de las capacidades del software.

Análisis Estadístico

R ofrece un amplio conjunto de herramientas para realizar análisis estadísticos de diversos tipos, lo que lo hace ideal para investigadores y analistas de datos.

- Estadística Descriptiva: Funciones para calcular medidas descriptivas como medias, medianas, desviaciones estándar, y más. También incluye herramientas para crear tablas de frecuencia y resumen.
- Inferencia Estadística: Métodos para realizar pruebas de hipótesis, calcular intervalos de confianza y llevar a cabo análisis de varianza (ANOVA). R incluye implementaciones de pruebas t, chi-cuadrado, y muchas otras pruebas estadísticas estándar.
- Modelos Estadísticos: Capacidades para ajustar modelos de regresión lineal y no lineal, modelos lineales generalizados (GLM), y otros modelos avanzados como modelos de efectos mixtos y modelos de series temporales.

Comparación de software Estadísticos:

Tabla 1. Comparación de software Estadístico

Característica	R	Python	SAS	SPSS
Costo	Gratuito y de código abierto	Gratuito (principalmente) y de código abierto	Comercial (licencia costosa)	Comercial (licencia costosa)

Licencia	GNU General Public License (GPL)	Diferentes licencias de código abierto (BSD, MIT)	Licencia propietaria	Licencia propietaria
Popularidad en la Academia	Muy alta, especialmente en estadística y biología	Alta, especialmente en ciencias de datos y machine learning	Alta en sectores comerciales	Alta en ciencias sociales y salud
Lenguaje de Programación	Específicamente diseñado para análisis estadístico	General-purpose, con bibliotecas para análisis de datos	Propio lenguaje (SAS)	No es un lenguaje de programación completo
Facilidad de Uso	Moderada (requiere conocimiento de programación)	Moderada (requiere conocimiento de programación)	Alta para análisis predefinidos, baja para programación personalizada	Alta para análisis predefinidos, baja para programación personalizada
Extensibilidad	Muy alta (más de 15,000 paquetes en CRAN)	Muy alta (numerosas bibliotecas en PyPI)	Limitada a módulos adicionales comerciales	Limitada a módulos adicionales comerciales
Comunidad y Soporte	Grande y activa, con muchos recursos en línea	Grande y activa, especialmente en ciencia de datos	Amplio soporte comercial y comunidades especializadas	Amplio soporte comercial y comunidades especializadas
Capacidades de Análisis	Extensas en estadística y modelado avanzado	Extensas en estadística y machine learning	Extensas en análisis estadístico y minería de datos	Extensas en análisis estadístico y análisis descriptivo
Visualización de Datos	Muy buena (ggplot2, base graphics)	Muy buena (Matplotlib, Seaborn)	Básica y menos flexible	Básica y menos flexible
Manipulación de Datos	Muy buena (dplyr, tidyr)	Muy buena (Pandas)	Buena, pero con sintaxis más compleja	Básica y menos flexible



Integración con Otros Sistemas	Alta (SQL, Hadoop, Spark, etc.)	Alta (SQL, Hadoop, Spark, etc.)	Alta (SQL, bases de datos propietarias)	Limitada (principalmente con productos de IBM)
Curva de Aprendizaje	Pronunciada para principiantes	Pronunciada para principiantes	Moderada para uso básico, alta para personalización	Moderada para uso básico, alta para personalización
Rendimiento con Grandes Datos	Moderado (mejorado con data.table, parallel)	Moderado (mejorado con Dask, PySpark)	Alto (optimizado para grandes volúmenes de datos)	Bajo a moderado (no diseñado para grandes volúmenes de datos)

Fuente: Elaboración propia

2.2.1.2. Aprendizaje de probabilidades

Según Guerra, Aguilar y Leyva (2021), las probabilidades son una rama fundamental de la teoría matemática que se encarga de cuantificar la incertidumbre y la aleatoriedad en los eventos. Este concepto se basa en el análisis de la posibilidad de ocurrencia de diferentes resultados dentro de un conjunto de eventos posibles, utilizando modelos matemáticos y herramientas estadísticas para asignar valores numéricos que representen la probabilidad de que cada evento ocurra. Las probabilidades son utilizadas en una amplia variedad de campos, desde la toma de decisiones en la vida cotidiana hasta aplicaciones avanzadas en la ciencia, la ingeniería, la economía y muchas otras disciplinas, desempeñando un papel crucial en la comprensión y predicción de fenómenos aleatorios y la toma de decisiones fundamentada en datos. (p. 203)

Como menciona Álvarez y Barreda (2020), las distribuciones de probabilidad son modelos matemáticos que describen la probabilidad de ocurrencia de diferentes resultados en un experimento aleatorio o fenómeno. Estas distribuciones son fundamentales en la teoría de probabilidades y la estadística, ya que proporcionan un marco estructurado para entender y predecir el comportamiento de variables aleatorias en una amplia gama de contextos. Su importancia radica en que permiten cuantificar y visualizar la variabilidad de los datos, así como



calcular medidas de tendencia central y dispersión, lo que facilita la toma de decisiones informadas, la identificación de patrones y la inferencia sobre poblaciones. Además, las distribuciones de probabilidad son la base para la construcción de modelos probabilísticos y la realización de inferencias estadísticas, lo que las convierte en una herramienta indispensable en la investigación científica, la ingeniería, la economía y muchos otros campos donde se analizan datos y se toman decisiones basadas en evidencia. (p .41)

La probabilidad es una rama de la matemática que se ocupa del estudio de fenómenos aleatorios. Es fundamental para la estadística y otras áreas de investigación donde la incertidumbre y la variabilidad juegan un papel importante.

- Experimento aleatorio: Un experimento aleatorio es un proceso cuyo resultado no puede predecirse con certeza antes de que ocurra. Cada ejecución del experimento puede tener diferentes resultados, y se considera aleatorio debido a la impredecibilidad inherente. Ejemplo: Lanzar una moneda, donde los resultados posibles son "cara" o "cruz".
- Espacio Muestral (S): El espacio muestral es el conjunto de todos los posibles resultados de un experimento aleatorio. Cada posible resultado se denomina punto muestral. Ejemplo: En el lanzamiento de un dado, el espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$.
- Evento: Un evento es un subconjunto del espacio muestral. Puede consistir en uno o más puntos muestrales y representa un resultado específico que puede ocurrir en un experimento aleatorio. Ejemplo: Al lanzar un dado, un evento puede ser "obtener un número par", que corresponde al subconjunto $\{2, 4, 6\}$ del espacio muestral.

Probabilidad de un evento

La probabilidad es una medida de la posibilidad de que ocurra un evento. Las probabilidades se expresan como valores entre 0 y 1, donde 0 indica que el evento es imposible y 1 indica que el evento es seguro.

- Probabilidad Clásica: En situaciones donde todos los resultados son igualmente probables, la probabilidad de un evento A se calcula como el



número de resultados favorables dividido por el número total de posibles resultados en el espacio muestral.

Distribuciones de Probabilidad

Las distribuciones de probabilidad describen cómo se distribuyen las probabilidades entre todos los posibles resultados de una variable aleatoria.

- **Distribución Discreta:** Una distribución de probabilidad discreta se aplica a variables aleatorias que pueden tomar un número finito o contable de valores distintos. Cada valor tiene una probabilidad asociada.
- **Distribución Binomial:** Describe el número de éxitos en una serie de n ensayos de Bernoulli independientes y con la misma probabilidad de éxito p .
- **Distribución de Poisson:** Describe el número de eventos que ocurren en un intervalo fijo de tiempo o espacio, cuando estos eventos ocurren con una tasa promedio constante y de manera independiente.
- **Distribución Continua:** Una distribución de probabilidad continua se aplica a variables aleatorias que pueden tomar cualquier valor en un intervalo continuo. Las probabilidades se representan mediante una función de densidad de probabilidad.
- **Distribución Normal:** También conocida como distribución gaussiana, es una distribución continua que se caracteriza por su forma de campana y se describe por dos parámetros: la media (μ) y la desviación estándar (σ).
- **Distribución Exponencial:** Describe el tiempo entre eventos en un proceso de Poisson, con una tasa constante λ .

Métodos de Enseñanza y Aprendizaje

Enfoques Tradicionales

Los métodos tradicionales de enseñanza se centran en la transmisión de conocimientos de forma estructurada y sistemática. Estos enfoques son comunes en la educación formal y han demostrado ser efectivos en la enseñanza de conceptos teóricos básicos.



Métodos Didácticos:

- Clases Magistrales: El profesor presenta los conceptos y teorías fundamentales de la probabilidad de manera directa, utilizando presentaciones, pizarras y materiales impresos. Este método es útil para cubrir una gran cantidad de material en un período de tiempo limitado.
 - Ventajas: Permite una cobertura exhaustiva de los conceptos teóricos y asegura que todos los estudiantes reciben la misma información.
 - Desventajas: Puede ser pasivo y no siempre fomenta la participación activa de los estudiantes.
- Lecturas y Textos: Los estudiantes son asignados a leer capítulos de libros de texto y artículos académicos. Esto complementa las clases magistrales y proporciona una base sólida de conocimientos.
 - Ventajas: Refuerza el aprendizaje y proporciona recursos adicionales para el estudio independiente.
 - Desventajas: Requiere una disciplina de estudio autónomo y puede ser difícil para los estudiantes menos motivados.

Ejercicios de Problemas:

- Tareas y Deberes: Los estudiantes practican resolviendo problemas de probabilidad fuera del aula. Estos ejercicios refuerzan los conceptos aprendidos y desarrollan habilidades de resolución de problemas.
 - Ventajas: Fomenta la práctica y aplicación de conceptos teóricos.
 - Desventajas: Puede no proporcionar retroalimentación inmediata, lo que puede dificultar el aprendizaje para algunos estudiantes.
- Sesiones de Práctica en Clase: Se dedica tiempo en clase para que los estudiantes trabajen en problemas de probabilidad bajo la supervisión del profesor, quien puede ofrecer ayuda y retroalimentación inmediata.
 - Ventajas: Permite la identificación y corrección de errores en tiempo real.
 - Desventajas: Puede ser menos efectivo si no se gestiona adecuadamente el tiempo y la participación de los estudiantes.



Evaluación:


- Exámenes Escritos: Los exámenes miden la comprensión de los conceptos teóricos y la capacidad de los estudiantes para aplicar técnicas de resolución de problemas en situaciones controladas.
 - Ventajas: Proporciona una medida clara del rendimiento y la comprensión de los estudiantes.
 - Desventajas: Puede causar estrés y no siempre refleja la capacidad de los estudiantes para aplicar conocimientos en contextos prácticos.

2.2.2. Conceptual

Programa estadístico R

- Versatilidad: R es conocido por su versatilidad en términos de análisis estadístico y gráfico. Puede manejar una amplia variedad de técnicas estadísticas, desde análisis descriptivos básicos hasta modelos avanzados de regresión, series temporales, análisis de supervivencia y aprendizaje automático. Además, ofrece una amplia gama de herramientas gráficas para visualizar datos de manera efectiva.
- Comunidad y soporte: Una de las fortalezas de R es su gran comunidad de usuarios y desarrolladores. Esta comunidad activa contribuye constantemente con paquetes y recursos adicionales que amplían las capacidades del software y ofrecen soluciones a problemas específicos. Además, el soporte en línea, que incluye documentación detallada, foros de discusión y tutoriales, es abundante y accesible para los usuarios de todos los niveles de experiencia.
- Gratuidad y código abierto: R es un software de código abierto, lo que significa que es gratuito para su uso y distribución. Esta característica lo hace accesible para estudiantes, investigadores y profesionales de todo el mundo, sin importar su ubicación geográfica o situación económica. Además, el código abierto permite una mayor transparencia y flexibilidad, ya que los usuarios pueden modificar y adaptar el software según sus necesidades específicas.

Aprendizaje de probabilidades



- **Distribuciones de probabilidad para variables discretas:** Las distribuciones de probabilidad para variables discretas describen la probabilidad de ocurrencia de resultados específicos en experimentos donde los valores posibles son contables y distintos, como el número de veces que se obtiene cara al lanzar una moneda o el número de clientes que llegan a una tienda en una hora determinada. Estas distribuciones asignan probabilidades a cada posible resultado y proporcionan información sobre la variabilidad y la tendencia central de los datos, lo que permite realizar inferencias y tomar decisiones informadas en una amplia gama de contextos.
- **Distribuciones de probabilidad para variables continuas:** Las distribuciones de probabilidad para variables continuas modelan la probabilidad de ocurrencia de resultados en experimentos donde los valores posibles forman un conjunto infinito y no contable, como la altura de las personas o el tiempo que tarda en completarse una tarea. Estas distribuciones proporcionan una descripción de la densidad de probabilidad en diferentes rangos de valores, lo que permite analizar la variabilidad y calcular probabilidades acumulativas para intervalos específicos. Son fundamentales en el análisis estadístico y la inferencia, así como en la modelización de fenómenos naturales y sociales.
- **Graficar distribuciones de probabilidad:** Graficar distribuciones de probabilidad es una técnica fundamental para visualizar la distribución de datos y entender su comportamiento probabilístico. Las representaciones gráficas, como histogramas, gráficos de densidad y gráficos de funciones de distribución acumulativa, permiten identificar patrones, simetrías, sesgos y outliers en los datos, lo que facilita la interpretación y la comunicación de resultados. Estas visualizaciones son útiles tanto en el análisis exploratorio de datos como en la presentación de resultados en informes y presentaciones.



2.3. Definición de términos básicos

- Variable aleatoria: Una variable que puede tomar diferentes valores en función de los resultados de un fenómeno aleatorio.
- Distribución de probabilidad: La función que asigna probabilidades a los diferentes valores que puede tomar una variable aleatoria.
- Función de densidad de probabilidad: En el caso de variables continuas, esta función describe la probabilidad de que la variable aleatoria caiga dentro de un intervalo específico.
- Función de masa de probabilidad: En el caso de variables discretas, esta función describe la probabilidad de que la variable aleatoria tome un valor específico.
- Esperanza matemática (o valor esperado): La media ponderada de una variable aleatoria, que representa su valor promedio esperado.
- Varianza: Una medida de dispersión que indica qué tan dispersos están los valores de una variable aleatoria alrededor de su media.
- Desviación estándar: La raíz cuadrada de la varianza, que proporciona una medida de dispersión en la misma unidad que la variable aleatoria.
- Distribución normal: Una distribución de probabilidad continua con forma de campana, caracterizada por su media y desviación estándar.
- Distribución uniforme: Una distribución de probabilidad en la que todos los valores posibles de una variable aleatoria tienen la misma probabilidad de ocurrencia.
- Distribución binomial: Una distribución de probabilidad discreta que modela el número de éxitos en una secuencia de ensayos independientes, donde cada ensayo tiene solo dos resultados posibles (éxito o fracaso) con una probabilidad constante de éxito en cada ensayo.



CAPITULO III. HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.1.1. Hipótesis general

El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

3.1.2. Hipótesis específicas

- El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.
- El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.
- El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

3.2. Definición conceptual de variables

Software R

Es un entorno de programación y análisis estadístico de código abierto, utilizado para la manipulación de datos, cálculos estadísticos y generación de gráficos.

Aprendizaje de probabilidades

Es el proceso de adquirir conocimientos sobre el comportamiento de eventos aleatorios mediante la comprensión y aplicación de conceptos, reglas y modelos probabilísticos.



3.3. Operacionalización de variable

Tabla 2. Operacionalización de variable

Variable	Concepto	Dimensión	Indicador
Software R	R es un entorno y lenguaje de programación estadística de código abierto que proporciona herramientas poderosas para el análisis de datos y la generación de gráficos. Con una amplia gama de funciones y paquetes disponibles.	Versatilidad	Número de paquetes disponibles
		Comunidad y soporte	Tiempo de respuesta en foros de discusión
		Gratuidad y código abierto	Número de contribuyentes al desarrollo del software
Aprendizaje de probabilidades	Se refiere al proceso mediante el cual los individuos adquieren conocimientos y habilidades relacionados con la teoría de la probabilidad, que es fundamental en la comprensión y predicción de eventos aleatorios.	Distribuciones de probabilidad para variables discretas	Entendimiento de la distribución binomial
		Distribuciones de probabilidad para variables continuas	Interpretación de la distribución normal estándar
		Gráficos de distribuciones de probabilidad	Habilidad para crear histogramas y gráficos de densidad

Fuente: Elaboración propia del autor

CAPITULO IV. DISEÑO METODOLÓGICO

4.1. Tipo y diseño de investigación

El tipo de investigación es descriptivo dado que el objetivo principal de la investigación es describir el impacto del uso del software R en el aprendizaje de probabilidades, así como establecer relaciones entre variables como el rendimiento académico antes y después de la intervención, el enfoque descriptivo te permitirá recopilar datos numéricos a través de pre y post tests y analizarlos de manera cuantitativa para describir con precisión las características del fenómeno estudiado y las posibles relaciones entre variables.

El diseño de investigación sería pre experimental, específicamente un diseño de grupo único, ya que implica la medición de una variable (aprendizaje de probabilidades) antes y después de la intervención (uso del software R), pero no incluye un grupo de control para comparar los resultados con y sin la intervención.

4.2. Método de investigación

El método de investigación es cuantitativo debido a que se empleará una ficha de registro para recopilar datos numéricos durante el pre y post test, lo que permitirá una medición objetiva y cuantificable del aprendizaje de probabilidades en los estudiantes. Este enfoque cuantitativo es adecuado para analizar las diferencias estadísticas antes y después de la intervención del uso del software R, brindando así una evaluación más precisa y rigurosa de su impacto en el aprendizaje de los estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.

4.3. Población y muestra

La población del estudio estará constituida por los 80 alumnos inscritos en el curso de estadística descriptiva de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.

Tabla 3. Distribución de la población

Ingeniería Eléctrica	45
----------------------	----



Ingeniería Electrónica	35
TOTAL	80

Fuente: Elaboración propia

Para el cálculo de la muestra se usó la formula por proporciones:

$$n_0 = \frac{pq}{\left[\frac{ep}{Z}\right]^2} = \frac{(0.9)(0.1)}{\left[\frac{(0.05)(0.9)}{1.96}\right]^2} = 170.73$$

Si se cumple con la siguiente regla se realiza el ajuste de la muestra:

$$\text{Regla: } \frac{n_0}{N} > 0.05 \rightarrow \frac{170.73}{80} > 0.05 \rightarrow 2.1341 > 0.05$$

Dado que se cumple la regla hacemos el ajuste:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{170.73}{1 + \frac{170.73}{80}} = 54$$

Donde:

n_0 : Tamaño de muestra inicial

N : tamaño poblacional (80)

Z : valor de la tabla normal (1.96)

p : probabilidad de éxitos (0.9)

q : probabilidad de fracasos (0.1)

e : error relativo (0.05)

Tabla 4. Distribución de la muestra

Ingeniería Eléctrica	31
Ingeniería Electrónica	23
TOTAL	54

Fuente: Elaboración propia

La muestra será de 54 alumnos inscritos en el curso de estadística descriptiva de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao.



4.4. Lugar de estudio y periodo desarrollado

El estudio se desarrollará en la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao, situada en la dirección Av. Juan Pablo II 306, Bellavista – Callao.

4.5. Técnicas e instrumentos para la recolección de la información

Para la recolección de información en el estudio sobre el uso del software R en el aprendizaje de probabilidades en estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao, se emplearán las siguientes técnicas e instrumentos:

- Pre y post tests: Se aplicarán cuestionarios diseñados específicamente para evaluar el nivel de conocimiento y comprensión de probabilidades de los estudiantes antes y después de la intervención con el software R. Estos cuestionarios contendrán preguntas relacionadas con los conceptos fundamentales de probabilidades y su aplicación práctica.

Estas técnicas e instrumentos de recolección de información permitirán obtener datos cuantitativos relevantes para analizar el impacto del uso del software R en el aprendizaje de probabilidades de los estudiantes.

4.6. Análisis y procesamiento de datos

El análisis y procesamiento de datos se llevará a cabo mediante un enfoque cuantitativo, centrándose en los resultados obtenidos de los pre y post tests aplicados a los estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao. Los datos recopilados en ambos momentos se compararán utilizando técnicas estadísticas adecuadas, como pruebas de hipótesis y análisis de varianza, para evaluar las diferencias en el nivel de conocimiento y comprensión de probabilidades antes y después de la intervención con el software R. Además, se realizará un análisis descriptivo para examinar la distribución de los puntajes y calcular medidas de tendencia central y dispersión. Los resultados se presentarán de manera clara y concisa mediante tablas y gráficos que faciliten la interpretación de los hallazgos y permitan



identificar cualquier tendencia o patrón significativo en el aprendizaje de probabilidades de los estudiantes.

A handwritten signature in blue ink, appearing to read "Antonio J. García", located in the upper right quadrant of the page.

CAPITULO V. RESULTADOS

5.1. Resultados descriptivos

Tabla 5. Prueba en PRE TEST de distribuciones de probabilidad para variables discretas

Nivel	Estudiantes	Porcentaje
Alto	1	2%
Medio	45	83%
Bajo	8	15%
TOTAL	54	100%

Fuente: Elaboración propia del autor

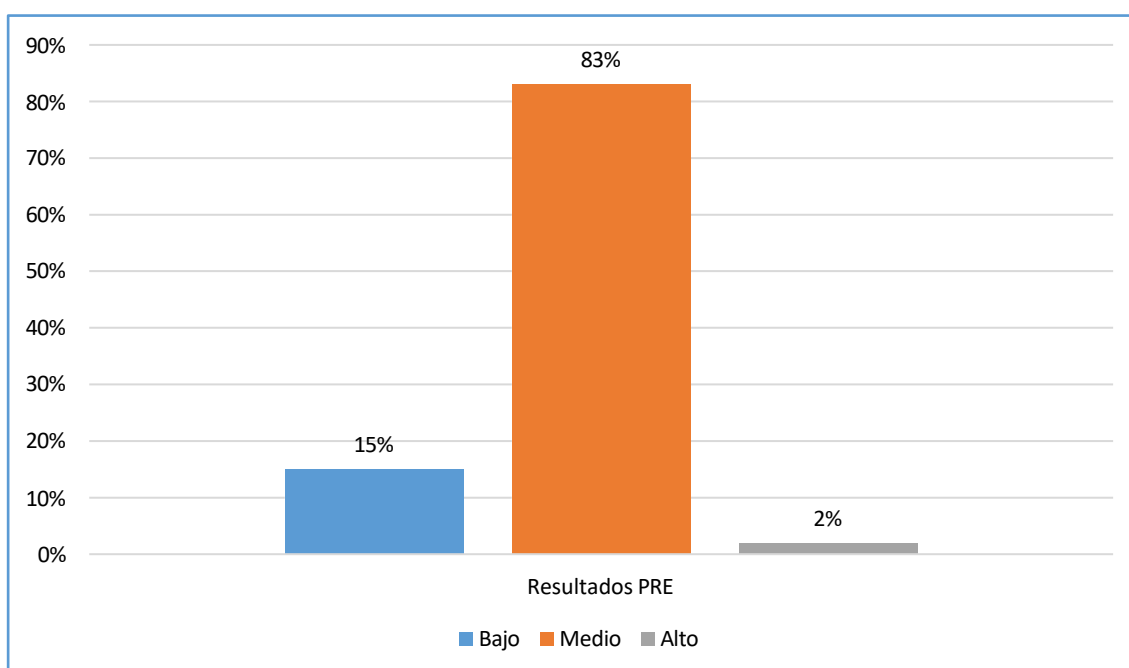


Figura 2. Prueba en PRE TEST de distribuciones de probabilidad para variables discretas

Los resultados del pre-test sobre distribuciones de probabilidad para variables discretas indican que el 83% de los estudiantes se ubicaron en un nivel medio, mientras que un 15% tuvo un rendimiento bajo y solo un 2% alcanzó un rendimiento alto. Esto sugiere una comprensión moderada del tema antes de la aplicación del software R, revelando la necesidad de intervenciones pedagógicas para reforzar conceptos fundamentales y brindar apoyo a aquellos con bajo rendimiento.

Tabla 6. Prueba en POST TEST de distribuciones de probabilidad para variables discretas

Nivel	Estudiantes	Porcentaje
Alto	27	50%
Medio	27	50%
Bajo	0	0%
TOTAL	54	100%

Fuente: Elaboración propia del autor

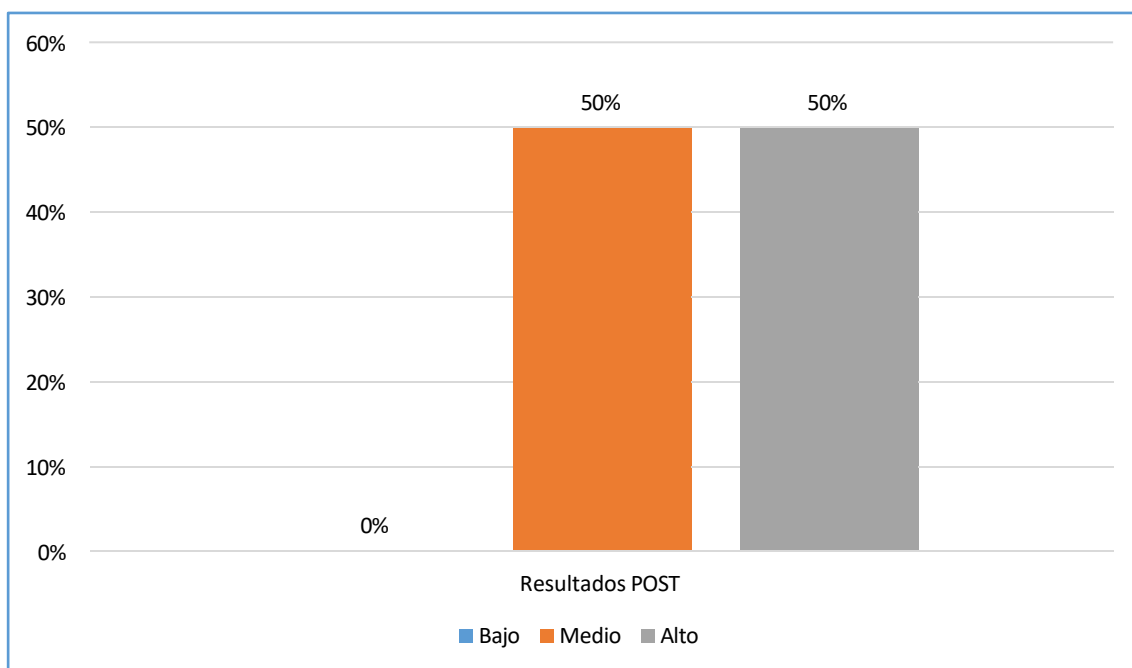


Figura 3. Prueba en POST TEST de distribuciones de probabilidad para variables discretas

Los resultados del post-test sobre distribuciones de probabilidad para variables discretas muestran un avance significativo en el rendimiento de los estudiantes, con un 50% alcanzando un nivel alto y otro 50% en un nivel medio, mientras que el 0% se ubicó en el rango bajo. Este cambio sugiere que la introducción del software R ha tenido un impacto positivo en la comprensión y aplicación de los conceptos, lo que indica una mejora considerable en el aprendizaje y dominio del tema.

Tabla 7. Prueba en PRE TEST de distribuciones de probabilidad para variables continuas

Nivel	Estudiantes	Porcentaje
Alto	1	2%
Medio	33	61%
Bajo	20	37%
TOTAL	54	100%

Fuente: Elaboración propia del autor

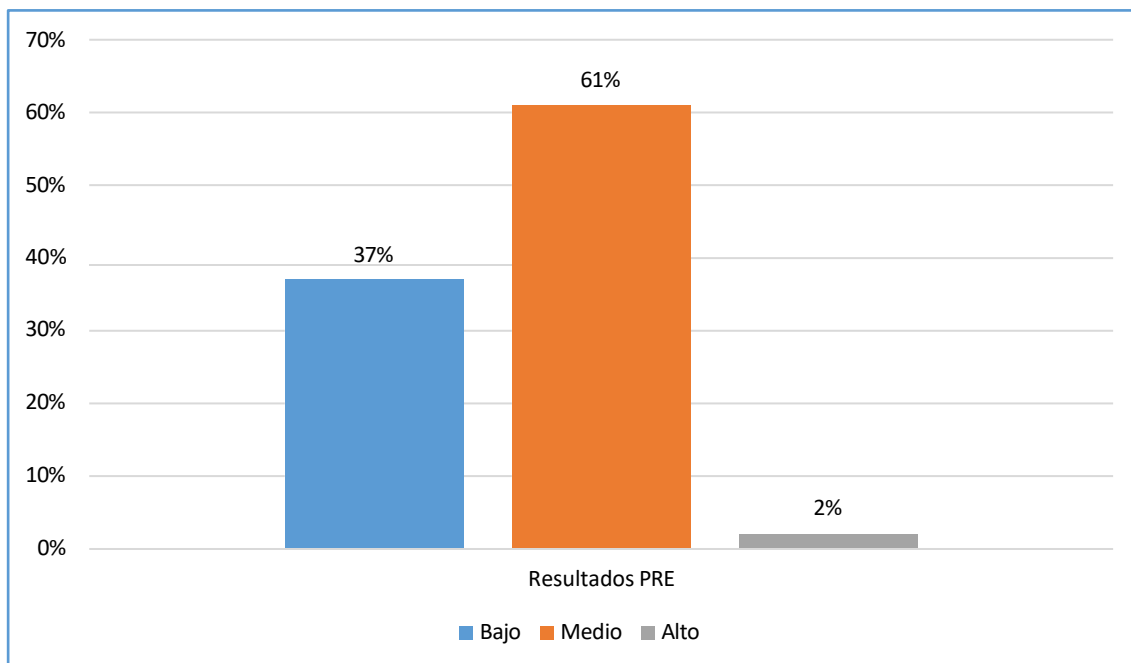


Figura 4. Prueba en PRE TEST de distribuciones de probabilidad para variables continuas

Los resultados del pre-test sobre distribuciones de probabilidad para variables continuas revelan que solo el 2% de los estudiantes alcanzó un rendimiento alto, mientras que el 61% se ubicó en un nivel medio y un 37% obtuvo un rendimiento bajo. Esta distribución de resultados indica una comprensión limitada del tema antes de la intervención, evidenciando que una proporción considerable de los estudiantes (37%) enfrenta dificultades significativas en la materia. La mayoría de los estudiantes se encuentran en el rango medio, lo que sugiere que, aunque han asimilado algunos conceptos, aún necesitan un fortalecimiento en su comprensión de las distribuciones de probabilidad para variables continuas.

Tabla 8. Prueba en POST TEST de distribuciones de probabilidad para variables continuas

Nivel	Estudiantes	Porcentaje
Alto	11	20%
Medio	40	74%
Bajo	3	6%
TOTAL	54	100%

Fuente: Elaboración propia del autor

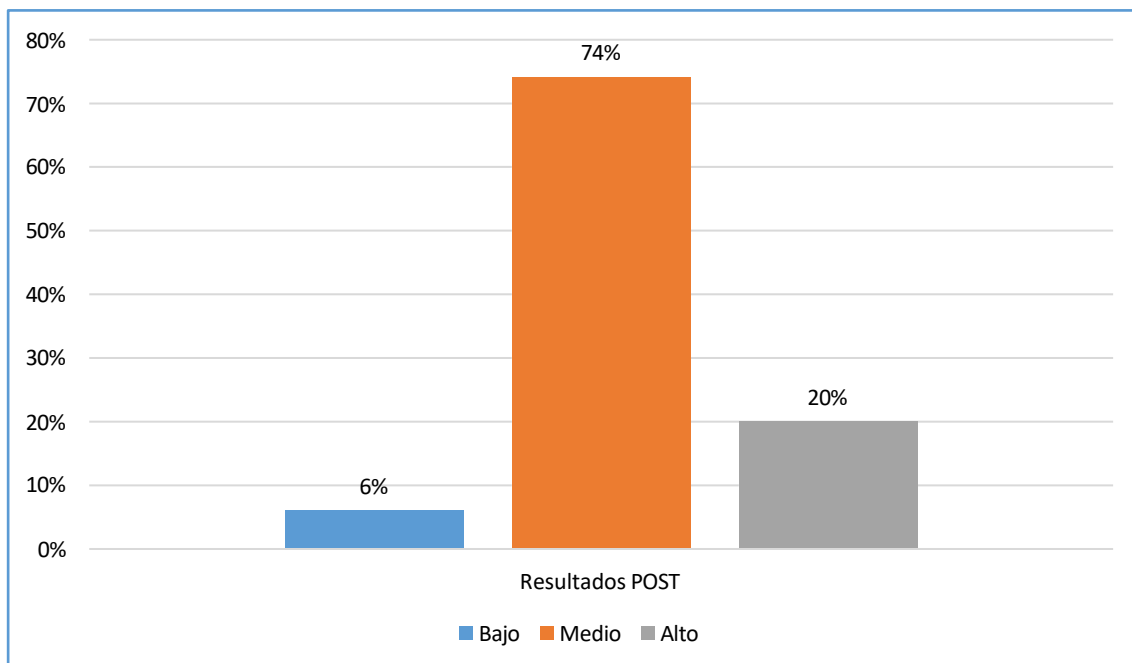


Figura 5. Prueba en POST TEST de distribuciones de probabilidad para variables continuas

Los resultados del post-test sobre distribuciones de probabilidad para variables continuas indican un progreso notable en el rendimiento de los estudiantes, con un 20% alcanzando un nivel alto, un 74% en un nivel medio y solo un 6% ubicándose en el rango bajo. Este cambio significativo en los resultados sugiere que, tras la intervención educativa, los estudiantes han mejorado considerablemente su comprensión y aplicación de los conceptos relacionados con las distribuciones de probabilidad para variables continuas. La reducción en el porcentaje de estudiantes con bajo rendimiento, combinado con el aumento en los niveles alto y medio, refleja la efectividad de las estrategias pedagógicas implementadas y la capacidad de los estudiantes para asimilar el contenido.

Tabla 9. Prueba en PRE TEST de Gráficos de distribuciones de probabilidad

Nivel	Estudiantes	Porcentaje
Alto	1	2%
Medio	48	89%
Bajo	5	9%
TOTAL	54	100%

Fuente: Elaboración propia del autor

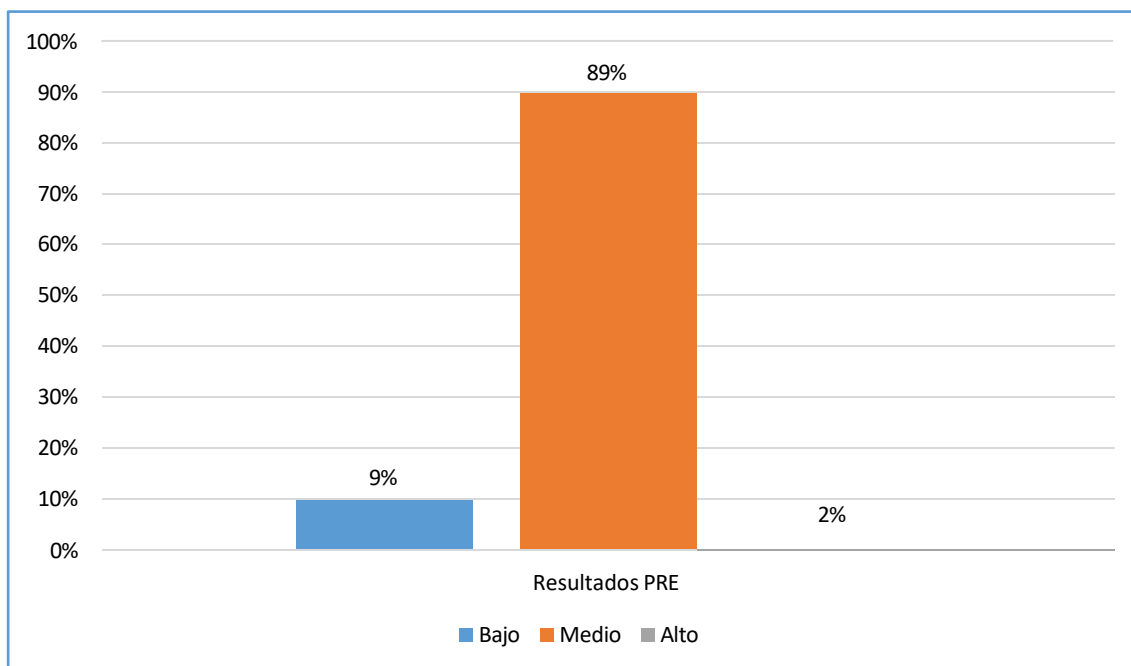


Figura 6. Prueba en PRE TEST de Gráficos de distribuciones de probabilidad

Los resultados del pre-test sobre gráficos de distribuciones de probabilidad muestran que solo el 2% de los estudiantes alcanzó un rendimiento alto, mientras que un 89% se ubicó en un nivel medio y un 9% obtuvo un rendimiento bajo. Esta distribución sugiere que, antes de la intervención, la mayoría de los estudiantes tiene una comprensión básica de los gráficos de distribuciones de probabilidad, pero un porcentaje considerable (9%) enfrenta dificultades que podrían obstaculizar su aprendizaje futuro.

Tabla 10. Prueba en POST TEST de Gráficos de distribuciones de probabilidad

Nivel	Estudiantes	Porcentaje
Alto	23	43%
Medio	31	57%
Bajo	0	0%
TOTAL	54	100%

Fuente: Elaboración propia del autor

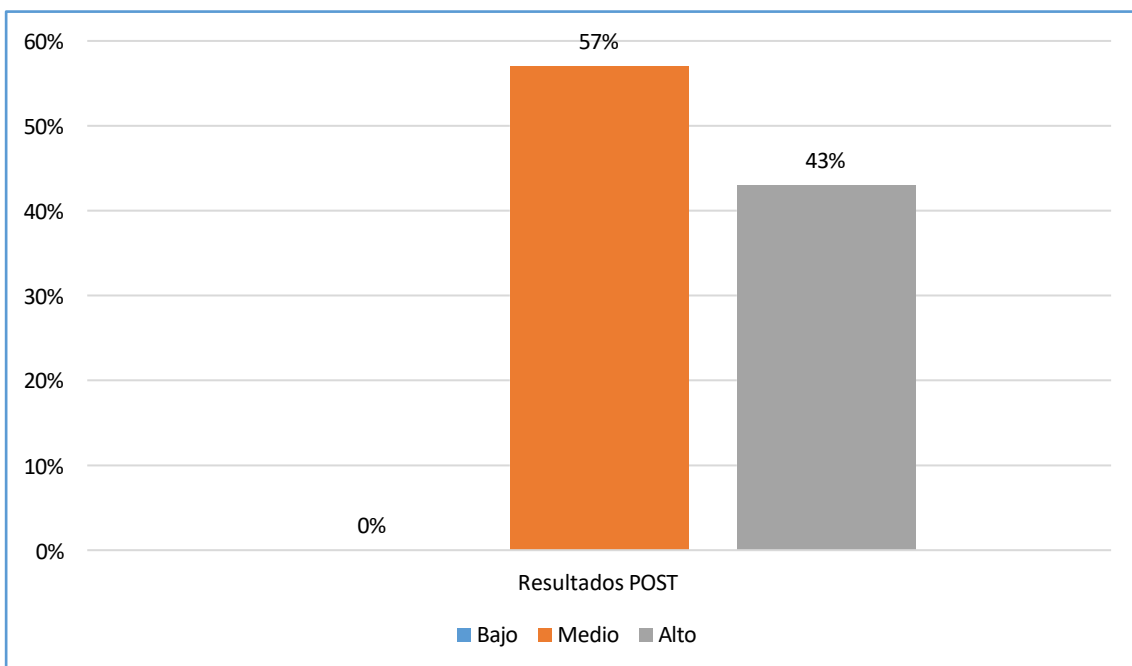


Figura 7. Prueba en POST TEST de Gráficos de distribuciones de probabilidad

Los resultados del post-test sobre gráficos de distribuciones de probabilidad indican un avance significativo en el rendimiento de los estudiantes, con un 43% alcanzando un nivel alto y un 57% en un nivel medio, mientras que el 0% se ubicó en el rango bajo. Este cambio positivo sugiere que la aplicación del software R ha tenido un impacto considerable en la comprensión y capacidad de los estudiantes para interpretar y crear gráficos de distribuciones de probabilidad. La ausencia de estudiantes en el nivel bajo refleja la efectividad de las estrategias pedagógicas implementadas, así como la mejora en la asimilación de los conceptos por parte de los estudiantes. Este progreso destaca la importancia de seguir utilizando herramientas tecnológicas y métodos interactivos en la enseñanza de la probabilidad, facilitando así un aprendizaje más profundo y efectivo en esta área.

Tabla 11. Comparativa del PRE y POST TEST

Nivel	Estudiantes	Porcentaje	Estudiantes	Porcentaje
	PRE		POST	
Alto	1	2%	14	26%
Medio	51	94%	40	74%
Bajo	2	4%	0	0%
TOTAL	54	100%	54	100%

Fuente: Elaboración propia del autor

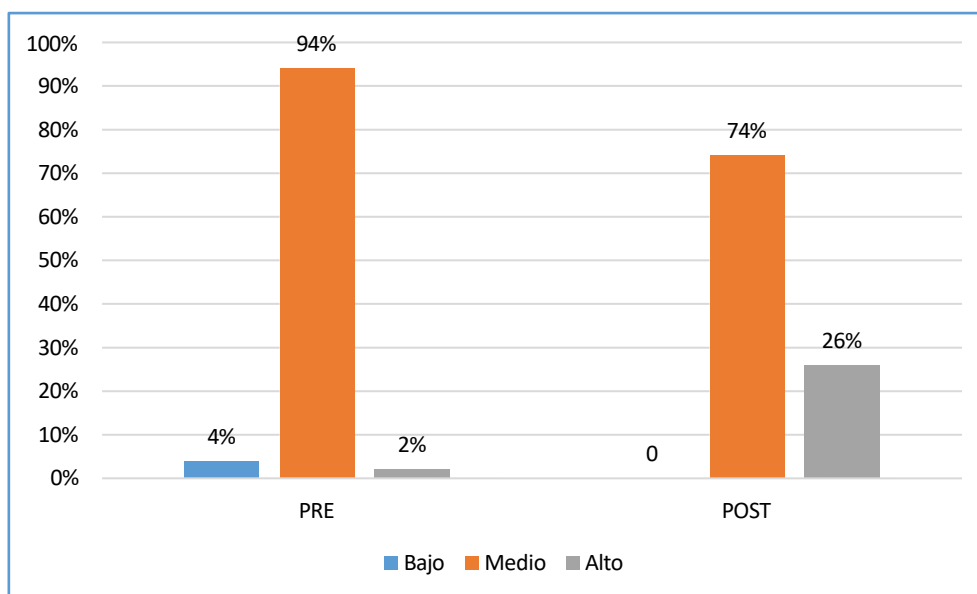


Figura 8. Comparativa del PRE y POST TEST

La comparación de los resultados entre el pre-test y el post-test del curso de probabilidades en estadística revela un progreso significativo en el rendimiento de los estudiantes tras la introducción del software R. En el pre-test, solo el 2% de los estudiantes alcanzó un nivel alto, mientras que el 94% se ubicó en un nivel medio y un 4% en el rango bajo. Esto indica que, antes de la intervención, la mayoría de los estudiantes tenía una comprensión básica del tema, pero un número considerable no lograba un rendimiento satisfactorio. En contraste, en el post-test, el porcentaje de estudiantes en el nivel alto aumentó al 26%, mientras que el 74% se mantuvo en el nivel medio y el 0% en el bajo. Este cambio significativo sugiere que el uso del software R ha mejorado la capacidad de los estudiantes para entender y aplicar los conceptos de probabilidad, aumentando su dominio del material.

5.2. Resultados inferenciales

Prueba de normalidad

Dado que se maneja una muestra de 54 siendo esta superior a 50 la prueba de normalidad que aplicaremos será la de Kolmogorov-Smirnov para ello plantearemos 2 hipótesis:

- Sig. < 0.05 acepta una distribución no normal
- Sig. \geq 0.05 acepta una distribución normal

Tabla 12. Prueba de normalidad de Kolmogorov-Smirnov

	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
T_PRE	,229	54	,000
T_POST	,148	54	,005

Fuente: Elaboración propia del autor

El estadístico de prueba fue de 0,229 con un valor de significancia (p) de 0,000. Dado que el p-valor es menor que 0,05, se rechaza la hipótesis nula de normalidad, lo que implica que los datos del pre-test no siguen una distribución normal. Este resultado sugiere que la distribución de los puntajes del pre-test está sesgada o presenta características que impiden su ajuste a la normalidad.

El estadístico de prueba fue de 0,148 con un valor de significancia (p) de 0,005. Al igual que en el pre-test, el p-valor es menor que 0,05, lo que también lleva a rechazar la hipótesis nula de normalidad. Esto indica que los datos del post-test tampoco se distribuyen de manera normal.

En conjunto, los resultados de la prueba de normalidad para ambos tests sugieren que las distribuciones de los puntajes no son normales, lo que podría tener implicaciones en el tipo de análisis estadístico a realizar en el futuro. Es importante considerar el uso de métodos no paramétricos para el análisis inferencial, ya que estos métodos no asumen la normalidad de los datos y son más adecuados para este tipo de distribución. Además, esta falta de normalidad podría influir en la interpretación de los resultados, destacando la necesidad de

abordar las características específicas de los datos al momento de realizar comparaciones entre los grupos de pre-test y post-test.

Prueba T-Student para comprobación de Hipótesis

Hipótesis General

H₁: El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

H₀: El uso del software R no mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Tabla 13. Prueba T Student para Hipótesis General

	Media	Prueba de T-Student		
		t	gl	Sig. (bilateral)
T_PRE	9.78	-25.531	53	<0,001
T_POST	13.50			

Fuente: Elaboración propia del autor

El valor t de -25.531, junto con un valor p de 0,001, indica que existe una diferencia estadísticamente significativa entre las medias del pre-test y el post-test en el rendimiento de los estudiantes, permitiendo rechazar la hipótesis nula que afirmaba la ausencia de diferencia. Este resultado sugiere que el uso del software R en la enseñanza de probabilidades ha tenido un impacto positivo y notable en la comprensión del tema, reflejando una mejora considerable en el rendimiento académico de los estudiantes. Por lo tanto, se concluye que la intervención educativa con el software R es efectiva.

Por lo tanto: El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Hipótesis específica 1

H₁: El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

H₀: El uso del software R no mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Tabla 14. Prueba T Student para Hipótesis Específica 1

	Media	Prueba de T-Student		
		t	gl	Sig. (bilateral)
PRE_D1	10.13	-21.407	53	<0,001
POST_D1	14.35			

Fuente: Elaboración propia del autor

El valor t de -21.407, junto con un p-valor de 0,001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de distribuciones de probabilidad para variables discretas entre los estudiantes que utilizaron el software R y aquellos que no lo hicieron, permitiendo rechazar la hipótesis nula que postulaba la ausencia de mejora. Este resultado sugiere que el uso del software R ha tenido un impacto positivo en la comprensión de las distribuciones de probabilidad, reflejando una mejora notable en el aprendizaje de los estudiantes.

Por lo tanto, se concluye que el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao, 2024.



Hipótesis específica 2

H₁: El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

H₀: El uso del software R no mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Tabla 15. Prueba T Student para Hipótesis Especifica 2

	Media	Prueba de T-Student		
		t	gl	Sig. (bilateral)
PRE_D2	8.76	-13.165	53	<0,001
POST_D2	12.43			

Fuente: Elaboración propia del autor

El valor t de -13.165, junto con un p-valor de 0,001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de distribuciones de probabilidad para variables continuas entre los estudiantes que utilizaron el software R y aquellos que no lo hicieron, permitiendo rechazar la hipótesis nula que afirmaba la ausencia de mejora. Este resultado sugiere que el uso del software R ha tenido un impacto positivo en la comprensión de las distribuciones de probabilidad para variables continuas, reflejando una mejora considerable en el aprendizaje de los estudiantes.

Por lo tanto, se concluye que el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao, 2024.



Hipótesis específica 3

H₁: El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

H₀: El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Tabla 16. Prueba T Student para Hipótesis Específica 3

	Media	Prueba de T-Student		
		t	gl	Sig. (bilateral)
PRE_D3	10.48	-14.299	53	<0,001
POST_D3	13.70			

Fuente: Elaboración propia del autor

El valor t de -14.299, junto con un p-valor de 0,001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de graficar distribuciones de probabilidad entre los estudiantes que utilizaron el software R y aquellos que no lo hicieron, permitiendo rechazar la hipótesis nula que postulaba la ausencia de mejora. Este resultado sugiere que el uso del software R ha tenido un impacto positivo en la habilidad de los estudiantes para graficar distribuciones de probabilidad, reflejando una mejora notable en su comprensión y aplicación del tema.

Por lo tanto, se concluye que el uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao, 2024.



CAPITULO VI. DISCUSIÓN DE RESULTADOS

6.1. Contrastación y demostración de la hipótesis con los resultados

Hipótesis general

La hipótesis general “El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024” se demuestra debido a los resultados obtenidos donde obtuvimos mediante la prueba T-Student donde el valor t de -25.531, junto con un valor p de 0,001, indica que existe una diferencia estadísticamente significativa entre las medias del pre-test y el post-test en el rendimiento de los estudiantes, esto a su vez se ve plasmado al haber obtenido en el pre-test una media de 9.78 en las calificaciones del curso de probabilidades y en el caso del post-test una media de 13.5. Este resultado sugiere que el uso del software R en la enseñanza de probabilidades ha tenido un impacto positivo y notable en la comprensión del tema.

Hipótesis específica 1

En la hipótesis específica 1 “El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024” se demuestra debido a los resultados obtenidos donde obtuvimos mediante la prueba T-Student un valor t de -21.407, junto con un valor p de 0.001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de distribuciones de probabilidad para variables discretas, esto a su vez se ve reflejado al haber obtenido en el pre-test una media de 10.13 en las calificaciones del tema de distribuciones de probabilidad para variables discretas y en el caso del post-test una media de 14.35. Este resultado sugiere que el uso del software R ha tenido un impacto positivo en la comprensión de las distribuciones de probabilidad, reflejando una mejora notable en el aprendizaje de los estudiantes.

Hipótesis específica 2

En la hipótesis específica 2 “El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del



callao, 2024” se demuestra debido a los resultados obtenidos donde obtuvimos mediante la prueba T-Student donde el valor t de -13.165, junto con un valor p de 0,001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de distribuciones de probabilidad para variables continuas, esto a su vez se ve reflejado al haber obtenido en el pre-test una media de 8.76 en las calificaciones de tema de distribuciones de probabilidad para variables continuas y en el caso del post-test una media de 12.43. Este resultado sugiere que el uso del software R en la enseñanza de probabilidades ha tenido un impacto positivo y notable en la comprensión del tema.

Hipótesis específica 3

En la hipótesis específica 3 “El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024” se demuestra debido a los resultados obtenidos donde obtuvimos mediante la prueba T-Student donde el valor t de -14.299, junto con un valor p de 0,001, indica que existe una diferencia estadísticamente significativa en el aprendizaje de graficar distribuciones de probabilidad, esto a su vez se ve reflejado al haber obtenido en el pre-test una media de 10.48 en las calificaciones de tema de graficar distribuciones de probabilidad y en el caso del post-test una media de 13.70. Este resultado sugiere que el uso del software R ha tenido un impacto positivo en la habilidad de los estudiantes para graficar distribuciones de probabilidad, reflejando una mejora notable en su comprensión y aplicación del tema.

6.2. Contrastación de los resultados con otros estudios similares

La investigación llevada a cabo por Fernández (2020), tuvo como objetivo presentar las amplias ventajas de R como herramienta para el análisis y visualización de datos en Ciencias Sociales. En nuestro estudio, se encontró que el uso de R mejora significativamente el aprendizaje en temas de distribuciones de probabilidad y su representación gráfica, lo cual refleja su capacidad para facilitar la comprensión de conceptos complejos sin requerir conocimientos previos de programación, alineándose con lo reportado por Fernández. Además, Fernández destacó que, aunque inicialmente la sintaxis



puede parecer compleja, la práctica continua con R lo convierte en una herramienta accesible y poderosa para los estudiantes, lo que también se evidencia en nuestra investigación al observar una mejora significativa en el rendimiento de los estudiantes tras su uso. Por lo tanto, ambos estudios concluyen que R es una herramienta versátil y efectiva, no solo para la visualización y análisis de datos, sino también para mejorar la comprensión y el aprendizaje de conceptos estadísticos.

En la investigación realizada por Lindo (2020) planteó como objetivo determinar la influencia del video tutorial YouTube en el aprendizaje de la Estadística Descriptiva en estudiantes de la Facultad de Contabilidad de la Universidad Nacional Agraria de la selva, Tingo María. Los resultados de esta investigación, en la que se demuestra que el uso del software R mejora significativamente el aprendizaje de distribuciones de probabilidad y su representación gráfica, guardan una clara relación con los hallazgos de Lindo (2020), quien evidenció que el uso de herramientas tecnológicas, como los tutoriales de YouTube, influye de manera significativa en el aprendizaje de conceptos estadísticos en comparación con los métodos tradicionales. En ambos estudios, se observa que la integración de recursos tecnológicos modernos facilita la comprensión de temas complejos y eleva el rendimiento académico. Mientras que Lindo mostró una mejora en el aprendizaje de Estadística Descriptiva, los resultados de nuestra investigación confirman un impacto positivo similar en la enseñanza de distribuciones de probabilidad mediante el software R, sugiriendo que el uso de herramientas tecnológicas no solo moderniza la enseñanza, sino que también optimiza el proceso de aprendizaje, reflejándose en puntajes significativamente más altos tras la intervención educativa.

En la investigación realizada por Benavides (2023) la cual plantea como objetivo principal proponer un modelo híbrido para la enseñanza de la estadística descriptiva e inferencial, en un contexto de pandemia (COVID-19) en la UNPRG-Lambayeque-Perú. Los resultados de esta investigación, que muestran la mejora significativa en el aprendizaje de distribuciones de probabilidad y su representación gráfica con el uso del software R, pueden compararse con los hallazgos de Benavides (2023). Mientras Benavides



propuso un modelo híbrido para la enseñanza de estadística durante la pandemia, destacando el uso de software estadístico por el 100% de los docentes y la importancia teórica sobre la práctica, en nuestro estudio se evidencia que una herramienta práctica como R no solo facilita la enseñanza, sino que también tiene un impacto significativo en la comprensión y rendimiento de los estudiantes. Al igual que Benavides, quien concluyó que el uso de software y estrategias mixtas mejora los resultados de los estudiantes, nuestros resultados refuerzan que el uso de tecnología en la enseñanza estadística, particularmente en el contexto de distribuciones de probabilidad, optimiza el aprendizaje y permite una comprensión más profunda de conceptos complejos.

En el estudio llevado a cabo por Mamani y Díaz (2017) tuvo como propósito principal determinar la naturaleza de la relación entre las actitudes hacia el estudio de la estadística descriptiva y el rendimiento académico en alumnos pertenecientes al quinto grado de secundaria en el mencionado centro educativo. Los resultados obtenidos en nuestra investigación, que demuestran una mejora significativa en el aprendizaje de distribuciones de probabilidad con el uso del software R, contrastan con el estudio de Mamani y Díaz (2017), en el cual se encontró una actitud negativa y temor hacia la estadística descriptiva en estudiantes de secundaria, lo que afectaba su rendimiento académico. Mientras en ese estudio las percepciones de los estudiantes eran mayormente negativas y conllevaban dificultades en el aprendizaje de la estadística, nuestros resultados indican que el uso de herramientas tecnológicas, como el software R, no solo mejora el rendimiento académico, sino que probablemente mitiga esos sentimientos de incapacidad o nerviosismo frente a la estadística. Esto sugiere que la implementación de software especializado puede transformar las percepciones hacia la estadística y favorecer una actitud más positiva y receptiva hacia esta disciplina, mejorando así el rendimiento académico.

6.3. Responsabilidad ética

Se garantizó el cumplimiento de todas las normativas éticas vigentes, tanto nacionales como institucionales, relacionadas con la investigación científica.



El estudio se centró exclusivamente en el análisis de los efectos del uso del software R en el aprendizaje de los estudiantes, sin involucrar información personal sensible ni afectar los derechos fundamentales de los participantes. Los datos obtenidos se manejaron de manera confidencial, respetando el anonimato de los estudiantes en todo momento. La participación fue completamente voluntaria, con la previa obtención de consentimiento informado, en el cual se explicó a los estudiantes los objetivos y alcances del estudio, así como su derecho a retirarse en cualquier momento sin consecuencias.

Asimismo, se garantizó la transparencia en el uso de los resultados, sin distorsionar los datos ni manipular los hallazgos para favorecer ninguna conclusión. Los resultados se presentaron de manera objetiva y rigurosa, contribuyendo al avance del conocimiento en el área educativa y fomentando el uso de herramientas tecnológicas como el software R para la mejora en el aprendizaje de la estadística y las probabilidades.

Por último, el uso de recursos tecnológicos, como el software R, se ajustó a las normativas de propiedad intelectual y licenciamiento, promoviendo el acceso ético y responsable a dichas herramientas para los fines educativos establecidos.



CONCLUSIONES

Primera conclusión:

El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Segunda conclusión:

El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Tercera conclusión:

El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.

Cuarta conclusión:

El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.



RECOMENDACIONES

- Dado que el uso del software R mostró mejoras significativas en el aprendizaje de probabilidades, se recomienda a la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Callao que considere la inclusión formal de este software en los cursos de estadística y probabilidades. Esto proporcionará a los estudiantes una herramienta moderna y eficiente para el análisis de datos, que les permitirá enfrentar mejor los retos académicos y profesionales.
- Para potenciar los beneficios del uso de software como R, es necesario que los docentes reciban capacitación continua sobre su manejo y sobre las nuevas tendencias en enseñanza estadística. Esta formación permitirá mejorar la calidad de la enseñanza y asegurar que los estudiantes adquieran las competencias necesarias para el uso de herramientas tecnológicas avanzadas en su aprendizaje.
- Se recomienda realizar estudios adicionales que evalúen el impacto del uso de software R en otras áreas del conocimiento dentro de la facultad, tales como el análisis de datos en ingeniería eléctrica o en áreas de simulación de sistemas. De esta manera, se podría ampliar el espectro de aplicación de este tipo de herramientas en el proceso de enseñanza-aprendizaje.
- Para profundizar en los hallazgos, sería útil realizar investigaciones comparativas entre el software R y otros programas estadísticos, como Python o MATLAB, para determinar cuál de estas herramientas ofrece un mayor impacto en el aprendizaje y la retención de conceptos estadísticos entre los estudiantes.
- Con base en la mejora en el aprendizaje observada con el uso de software R, se sugiere implementar metodologías mixtas (presenciales y virtuales) que combinen la enseñanza tradicional con el uso de tecnologías interactivas. Esto podría incluir la incorporación de tutoriales y videos educativos, que refuercen el contenido visto en clase y faciliten el aprendizaje autónomo de los estudiantes.



REFERENCIAS BIBLIOGRÁFICAS

LÁZARO Alvarez, Niurys, CALLEJAS Carrión, Zoraida y GRIOL Barres, David. Utilización del software SPSS para identificar factores predictivos de deserción estudiantil. Revista Luz [en línea]. enero-marzo 2022, vol. 21, no. 1. [Fecha de consulta: 1 de julio del 2023]. Disponible en http://scielo.sld.cu/scielo.php?pid=S1814-151X2022000100038&script=sci_arttext&tlng=en

MAMANI Rosas, Luciano y DÍAZ Palacios, Ronald. Actitudes hacia la estadística descriptiva y el aprendizaje en alumnos del quinto grado de educación secundaria de la I.E. “Héroes del Cenepa” del distrito de el tambo, Huancayo. Tesis de Licenciatura en Pedagogía y Humanidades, Universidad Nacional del Centro del Perú, Facultad de Educación, 2017. 120 pp.

FERNÁNDEZ Lizana, M. Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales. Revista Científica de la UCSA [en línea]. agosto – octubre 2020, vol. 7, n. ° 2. [Fecha de consulta: 31 de marzo del 2024]. Disponible en: http://scielo.iics.una.py/scielo.php?pid=S2409-87522020000200097&script=sci_arttext

SUCASAIRE Pilco, Jorge. Estadística descriptiva para trabajos de investigación. 1er Ed. Biblioteca Nacional del Perú. 2021.

LINDO Pizarro, César. El video tutorial YOU TUBE en el aprendizaje de la estadística descriptiva en estudiantes de la Facultad de Contabilidad de la Universidad Nacional Agraria de la Selva, Tingo María 2018. Tesis de doctorado en Ciencias de la educación, Universidad Nacional Hermilio Valdizán, Escuela de Posgrado, 2020. 90 pp.

BENAVIDES Campos, Grimaldo. Modelo híbrido para la enseñanza de la estadística descriptiva e inferencial en un contexto de pandemia (COVID-19). Tesis de doctorado en ciencias de la educación, Universidad Nacional Pedro Ruiz Gallo, Unidad de Posgrado, 2023. 104 pp.

VARGAS, Javier; ARREGOCÉS, Isabel; SOLANO, Andrés y PEÑA, Katia. Aprendizaje basado en proyectos soportado en un diseño tecno-pedagógico



para la enseñanza de la estadística descriptiva. Revista Formación Universitaria [en línea]. marzo-mayo 2021, vol. 14, n. °6. [Fecha de consulta: 4 de marzo del 2024]. Disponible en: <http://dx.doi.org/10.4067/S0718-50062021000600077>

IBÁÑEZ López, Francisco; MAURANDI López, Antonio y CASTEJÓN Mochón, José. Docencia práctica virtual y adquisición de competencias en la formación estadística de maestros durante el confinamiento sanitario. Revista de Investigación en Didáctica de la Matemática [en línea]. Febrero – abril 2022, vol. 16, n. °2. [Fecha de consulta: 4 de marzo del 2024]. Disponible en: <https://doi.org/10.30827/pna.v16i2.21364>

RAMOS Vargas, Luis. La educación estadística en el nivel universitario: retos y oportunidades. Revista Digital de Investigación en Docencia Universitaria [en línea]. julio – diciembre 2019, vol. 13, n. °2. [Fecha de consulta: 4 de marzo del 2024]. Disponible en <http://dx.doi.org/10.19083/ridu.2019.1081>

FERNÁNDEZ Lizana, Miguel. Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales. Revista Científica de la UCSA [en línea]. Agosto 2020, vol. 7, n. °2. [Fecha de consulta: 4 de marzo del 2024]. Disponible en <https://doi.org/10.18004/ucsa/2409-8752/2020.007.02.097>

SÁNCHEZ Villena, Andy. Uso de programas estadísticos libres para el análisis de datos: Jamovi, Jasp y R. Revis Perspectiva [en línea]. diciembre – marzo 2019, vol. 20, n. °1. [Fecha de consulta: 4 de marzo del 2024]. Disponible en <https://doi.org/10.33198/rp.v20i1.00026>

CONTENTO Rubio, Manuel. Estadística con aplicaciones en R [en línea]. Bogotá: Universidad de Bogotá Jorge Tadeo Lozano, 2019. [Fecha de consulta: 4 de marzo del 2024]. Disponible en <https://expeditiorepositorio.utadeo.edu.co/handle/20.500.12010/21660>. ISBN: 978-958-725-273-6

GUERRA Véliz, Yusimí; AGUILAR García, Andier y LEYVA Haza, Julio. Aprendizaje de la estadística descriptiva en secundaria básica con datos provenientes del consumo de energía. Revista Horizonte de la ciencia [en línea].



abril – agosto 2021, vol. 11, n. °21. [Fecha de consulta: 4 de marzo del 2024].
Disponibile en <https://doi.org/10.26490/uncp.horizonteciencia.2021.21.906>

ÁLVAREZ Pardo, Elen y BARREDA Jorge, Liset. La estadística descriptiva en la formación investigativa del instructor de arte. Revista Conrado [en línea]. abril – junio 2020, vol. 16, n. °73. [Fecha de consulta: 4 de marzo del 2024]. Disponible en <https://acortar.link/FyilRt>

A handwritten signature in blue ink, appearing to read 'Álvarez Pardo'.

ANEXOS



ANEXO N.º 01: MATRIZ DE CONSISTENCIA

Problema	Objetivo	Hipótesis	Variables	Dimensiones	Indicadores	Metodología
General:	General:	Principal:	V.I. Software R	Versatilidad	Número de paquetes disponibles	TIPO DE INVESTIGACIÓN: Descriptiva DISEÑO DE INVESTIGACIÓN: Pre experimental MÉTODO DE INVESTIGACIÓN: Cuantitativo POBLACIÓN: La población estará conformada por los estudiantes de la FIEE de la UNAC que estén llevando el curso de estadística descriptiva MUESTRA: La muestra será un conjunto representativo de los estudiantes de la FIEE de la UNAC que estén llevando el curso de estadística descriptiva
¿De qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?	Determinar de qué manera el uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	El uso del software R mejora el aprendizaje de probabilidades en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.		Comunidad y soporte	Tiempo de respuesta en foros de discusión	
Específicos:	Específicos:	Secundarias		Gratuidad y código abierto	Número de contribuyentes al desarrollo del software	
¿De qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?	Determinar de qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables discretas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.		V.D. Aprendizaje de probabilidades	Distribuciones de probabilidad para variables discretas	
¿De qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?	Determinar de qué manera el uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	El uso del software R mejora el aprendizaje de distribuciones de probabilidad para variables continuas en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	Distribuciones de probabilidad para variables continuas		Interpretación de la distribución normal estándar	
¿De qué manera el uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024?	Determinar de qué manera el uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	El uso del software R mejora el aprendizaje de graficar distribuciones de probabilidad en estudiantes de la facultad de ingeniería eléctrica y electrónica de la universidad nacional del callao, 2024.	Gráficos de distribuciones de probabilidad		Habilidad para crear histogramas y gráficos de densidad	

ANEXO N° 02: INSTRUMENTO
Cuestionario de preguntas para evaluar la

Distribuciones de Probabilidad para Variables Discretas

1. Defina qué es una variable aleatoria discreta.

Respuesta: Una variable aleatoria discreta es aquella que puede tomar un número finito o infinito numerable de valores específicos. Ejemplos incluyen el número de veces que aparece una cara en lanzamientos de una moneda o el número de autos vendidos en un día.

2. Explique la diferencia entre una distribución de probabilidad discreta y una continua. Proporcione ejemplos.

Respuesta: Una distribución discreta se aplica a variables que solo pueden tomar valores específicos y separados, como el número de fallos en un proceso. En cambio, una distribución continua se aplica a variables que pueden tomar cualquier valor dentro de un intervalo, como el peso de una persona. Ejemplos incluyen la distribución binomial (discreta) y la normal (continua).

3. Describa las propiedades que debe cumplir una función de probabilidad para una variable discreta.

Respuesta:

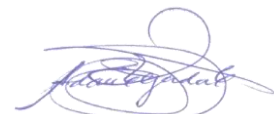
- Todas las probabilidades deben ser mayores o iguales a cero:

$$P(X = \alpha) \geq 0 \text{ para todo } \alpha$$

- La suma de las probabilidades de todos los posibles valores debe ser igual a 1:

$$\sum P(X = \alpha) = 1$$

4. La siguiente tabla representa la función de probabilidad de una variable aleatoria discreta X. Completa los valores faltantes y verifica si cumple con las propiedades de una función de probabilidad:



X	1	2	3	4
P(X)	0.2	0.3	0.4	?

Respuesta: El valor faltante para $P(X = 4)$ es 0.1. Para verificar:

$$0.2+0.3+0.4+0.1=1$$

Se cumple que la suma de las probabilidades es 1, por lo que es una función de probabilidad válida.

Verificar si cumple con las propiedades de probabilidad

```
P_X <- c(0.2, 0.3, 0.4, 1 - (0.2 + 0.3 + 0.4))
```

```
suma_P <- sum(P_X)
```

```
suma_P # Debe ser 1
```

5. En una distribución binomial con parámetros $n = 5$ y $p = 0.6$, calcula la probabilidad de obtener exactamente 3 éxitos.

Respuesta: Usamos la fórmula de la distribución binomial:

$$P(X = 3) = \binom{5}{3} (0.6)^3 (0.4)^2 = 10 \times (0.216) \times (0.16) = 0.3456$$

Probabilidad de obtener exactamente 3 éxitos


```
prob_3_exitos <- dbinom(3, size=5, prob=0.6)
```

```
prob_3_exitos
```

6. Dado un experimento donde tiras un dado equilibrado, ¿cuál es la probabilidad de obtener un número mayor a 4? Explica cómo se calcula.

Respuesta: Los números mayores a 4 en un dado son 5 y 6, por lo que la probabilidad es:

$$P(X > 4) = P(5) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$



Probabilidad de obtener un número mayor a 4 en un dado

```
prob_mayor_4 <- 2 / 6
```

```
prob_mayor_4
```

7. Explique el concepto de esperanza matemática (media esperada) para una variable aleatoria discreta.

Respuesta: La esperanza matemática o valor esperado de una variable aleatoria discreta X es la suma ponderada de todos los valores posibles de la variable, donde los pesos son las probabilidades asociadas a cada valor. Se calcula como:

$$E(X) = \sum_i x_i P(X = x_i)$$

8. Calcule la media esperada (esperanza matemática) para la siguiente distribución de probabilidad:

X	1	2	3
P(X)	0.2	0.5	0.3

Respuesta:

$$E(X) = 1(0.2) + 2(0.5) + 3(0.3) = 0.2 + 1 + 0.9 = 2.1$$

Calcular la media esperada

```
X <- c(1, 2, 3)
```

```
P_X <- c(0.2, 0.5, 0.3)
```

```
esperanza <- sum(X * P_X)
```

```
esperanza
```

9. Para una variable aleatoria discreta con función de probabilidad

$P(X = x) = \frac{1}{6}$ para $x=1,2,3,4,5,6$, ¿Cómo calcularías la varianza?

Respuesta: Primero, se calcula la media:



$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{1}{6}(21) = 3.5$$

Luego, la varianza se calcula como:

$$Var(X) = E(X^2) - (E(X))^2$$

$$\text{Donde: } E(X^2) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{1}{6}(91) = 15.17$$

Entonces, la varianza es:

$$Var(X) = 15.17 - (3.5)^2 = 15.17 - 12.25 = 2.92$$

Calcular la varianza

```
X <- c(1, 2, 3, 4, 5, 6)
```

```
P_X <- rep(1/6, 6)
```

```
media_X <- sum(X * P_X)
```

```
varianza_X <- sum((X - media_X)^2 * P_X)
```

```
varianza_X
```

10. Una tienda de autos vende, en promedio, 2 autos por día. ¿Qué distribución discreta puede modelar este fenómeno? Explica tu elección y describe sus parámetros.

Respuesta: La distribución que puede modelar este fenómeno es la distribución de Poisson, que se utiliza para modelar la cantidad de eventos que ocurren en un intervalo de tiempo. El parámetro de la distribución es $\lambda = 2$, que es la tasa promedio de ventas por día.


Simular una distribución de Poisson con lambda = 2

```
lambda <- 2
```

```
rpois(10, lambda) # Generar 10 valores aleatorios
```

Cuestionario de preguntas para evaluar la

Distribuciones de Probabilidad para Variables Continuas



1. Defina qué es una variable aleatoria continua

Respuesta: Una variable aleatoria continua es aquella que puede tomar cualquier valor dentro de un intervalo, es decir, sus posibles valores forman un conjunto no numerable. Ejemplos incluyen la altura de una persona o el tiempo que tarda en completarse un proceso.

2. Mencione dos ejemplos de distribuciones de probabilidad para variables continuas y explique brevemente cuándo se aplican.

Respuesta:

- **Distribución Normal:** Se aplica cuando los datos tienden a acumularse alrededor de la media, como las alturas o los pesos de una población.
- **Distribución Exponencial:** Se aplica en situaciones que modelan tiempos entre eventos, como el tiempo de espera entre llegadas de clientes a un servicio.

3. Describa la diferencia entre una función de densidad de probabilidad (PDF) y una función de distribución acumulativa (CDF) para variables continuas.

Respuesta:

- **PDF (Función de Densidad de Probabilidad):** Describe la probabilidad relativa de que una variable aleatoria continua tome un valor en un intervalo pequeño. El área bajo la curva de la PDF en un intervalo específico representa la probabilidad de que la variable tome un valor dentro de ese intervalo.
 - **CDF (Función de Distribución Acumulativa):** Describe la probabilidad acumulada de que una variable aleatoria continua sea menor o igual a un valor específico.
4. ¿Cuál es la probabilidad de que una variable aleatoria continua tome un valor exacto? Explica tu respuesta.

Respuesta: La probabilidad de que una variable aleatoria continua tome un valor exacto es siempre 0. Esto se debe a que una variable continua puede tomar



infinitos valores posibles en cualquier intervalo, por lo que la probabilidad de un solo valor específico es infinitesimal.

5. Sea $X \sim N(0,1)$, ¿Cuál es la probabilidad de que X esté entre -1 y 1?

Respuesta: Para una distribución normal estándar $N(0,1)$, la probabilidad de que X esté entre -1 y 1 corresponde a aproximadamente el 68.27% del área bajo la curva (propiedad de la distribución normal). Esto se puede verificar usando tablas de la distribución normal o con el software R.

Calcular la probabilidad de que X esté entre -1 y 1

```
prob <- pnorm(1, mean = 0, sd = 1) - pnorm(-1, mean = 0, sd = 1)
```

```
prob
```

6. Explique qué es la media y la varianza en una distribución normal.

Respuesta:

- **Media:** Es el valor central de la distribución, donde se concentran la mayoría de los datos. Para una distribución normal $N(\mu, \sigma^2)$, la media es μ .
 - **Varianza:** Mide la dispersión o variabilidad de los datos respecto a la media. Para una distribución normal, la varianza es σ^2 , y su raíz cuadrada σ es la desviación estándar.
7. Si se sabe que la altura de las personas sigue una distribución normal con media 170 cm y desviación estándar de 10 cm, ¿cuál es la probabilidad de que una persona mida más de 180 cm?

Respuesta: Primero, se estandariza el valor de 180 cm usando la fórmula:

$$Z = \frac{180 - 170}{10} = 1$$

Luego, usando una tabla de la distribución normal, la probabilidad de que $Z > 1$ es aproximadamente 0.1587, por lo que la probabilidad de que una persona mida más de 180 cm es 15.87%.

Calcular la probabilidad de medir más de 180 cm



```
prob_mayor_180 <- 1 - pnorm(180, mean = 170, sd = 10)
```

```
prob_mayor_180
```

8. Describa el concepto de percentil y cómo se relaciona con la función de distribución acumulativa (CDF).

Respuesta: Un percentil es un valor tal que un porcentaje específico de los datos cae por debajo de ese valor. Por ejemplo, el percentil 90 indica que el 90% de los valores son menores o iguales a ese valor. En términos de la CDF, el percentil p corresponde al valor x tal que $F(x) = p$, donde $F(x)$ es la CDF.

9. Explique qué es una distribución exponencial y proporcione un ejemplo de un fenómeno que se modele con esta distribución.

Respuesta: La distribución exponencial modela el tiempo entre eventos que ocurren de manera continua e independiente a una tasa constante. Un ejemplo típico es el tiempo de espera hasta que ocurre el próximo evento en un proceso de llegada de clientes a una cola o sistema.

10. En una distribución uniforme continua en el intervalo $[a, b]$, ¿cómo se calcula la probabilidad de que la variable aleatoria tome un valor dentro de un subintervalo $[c, d]$, donde $a \leq c < d \leq b$?

Respuesta: Para una distribución uniforme continua, la probabilidad de que la variable aleatoria tome un valor dentro de un subintervalo $[c, d]$ se calcula como:

$$P(c \leq X \leq d) = \frac{d - c}{b - a}$$

donde a y b son los límites del intervalo total de la distribución.

Parámetros de la distribución uniforme

```
a <- 0
```

```
b <- 10
```

```
c <- 3
```

```
d <- 7
```



```
# Calcular la probabilidad  
prob_uniforme <- (d - c) / (b - a)  
prob_uniforme
```

Cuestionario de preguntas para evaluar

Gráficos de Distribuciones de Probabilidad

1. ¿Qué es un histograma y para qué se utiliza en el análisis de distribuciones de probabilidad?

Respuesta: Un histograma es un gráfico de barras que muestra la distribución de una variable dividiendo el rango de datos en intervalos (o "bins"). Se utiliza para visualizar la frecuencia de ocurrencia de valores dentro de esos intervalos, lo que permite observar la forma de la distribución de la variable (simétrica, asimétrica, multimodal, etc.).

```
# Crear un histograma en R
```

```
data <- rnorm(100) # Generar datos de ejemplo con distribución normal  
hist(data, main="Histograma de distribución normal", xlab="Valores",  
col="lightblue")
```

2. Explique cómo se construye una función de densidad de probabilidad (PDF) para una variable continua y cómo se representa gráficamente.

Respuesta: La PDF se construye mediante una función que describe la probabilidad relativa de que una variable continua tome un valor en un intervalo pequeño. Gráficamente, se representa como una curva suave donde el área bajo la curva en un intervalo específico corresponde a la probabilidad de que la variable tome un valor dentro de ese intervalo.

```
# Crear y graficar la función de densidad de probabilidad en R
```

```
data <- rnorm(100) # Generar datos de ejemplo  
plot(density(data), main="Función de densidad de probabilidad", xlab="Valores")
```



3. Describa qué es un diagrama de caja y bigotes (boxplot) y qué información proporciona sobre una distribución.

Respuesta: Un diagrama de caja y bigotes (boxplot) es una representación gráfica de la distribución de un conjunto de datos que muestra los cuartiles, la mediana y los valores atípicos. Proporciona información sobre la dispersión, la simetría y la presencia de valores atípicos.

Crear un diagrama de caja y bigotes en R

```
data <- rnorm(100)
```

```
boxplot(data, main="Boxplot de la distribución", horizontal=TRUE)
```

4. ¿Qué tipo de gráfico utilizaría para representar una distribución de probabilidad discreta y por qué?

Respuesta: Se utiliza un gráfico de barras, ya que las variables discretas solo pueden tomar ciertos valores específicos. Las barras muestran la frecuencia o la probabilidad de cada uno de esos valores, permitiendo visualizar la distribución.

Gráfico de barras para una distribución discreta en R

```
data <- sample(1:6, 100, replace=TRUE) # Simular lanzamiento de un dado
```

```
barplot(table(data), main="Distribución de probabilidad discreta", xlab="Valores", ylab="Frecuencia")
```

5. Explique cómo se utiliza un gráfico de la función de distribución acumulativa (CDF) y qué información proporciona.

Respuesta: El gráfico de la CDF muestra la probabilidad acumulada de que una variable aleatoria sea menor o igual a un valor específico. Proporciona información sobre la distribución de los datos y facilita la identificación de percentiles y probabilidades acumuladas.

Graficar la función de distribución acumulativa (CDF) en R

```
data <- rnorm(100)
```



```
plot(ecdf(data), main="Función de distribución acumulativa (CDF)",
      xlab="Valores", ylab="Probabilidad acumulada")
```

6. En un gráfico de distribución normal, ¿qué representan la media, la desviación estándar y cómo se reflejan visualmente?

Respuesta: En una distribución normal, la media representa el valor central, donde la curva es más alta, mientras que la desviación estándar mide la dispersión de los datos alrededor de la media. Visualmente, la desviación estándar determina el ancho de la campana: una desviación estándar pequeña produce una curva estrecha, mientras que una grande genera una curva más ancha.

Graficar una distribución normal con la media y desviación estándar

```
data <- rnorm(1000, mean=0, sd=1)
```

```
hist(data, prob=TRUE, main="Distribución normal", xlab="Valores",
      col="lightblue")
```

```
curve(dnorm(x, mean=mean(data), sd=sd(data)), add=TRUE, col="red")
```

```
abline(v=mean(data), col="blue", lwd=2) # Línea de la media
```

7. ¿Cómo se puede identificar la presencia de asimetría (sesgo) en una distribución utilizando un histograma?

Respuesta: La asimetría o sesgo se puede identificar observando la forma del histograma. Si la cola de la distribución se extiende más hacia la derecha, la distribución tiene sesgo positivo (asimetría a la derecha). Si la cola se extiende más hacia la izquierda, tiene sesgo negativo (asimetría a la izquierda). Si el histograma es simétrico, no hay asimetría.

Crear un histograma para visualizar asimetría

```
data <- rlnorm(100) # Generar datos con sesgo positivo
```

```
hist(data, main="Histograma con sesgo positivo", xlab="Valores",
      col="lightgreen")
```



8. Describa qué es un gráfico Q-Q (cuantil-cuantil) y para qué se utiliza en el análisis de distribuciones.

Respuesta: Un gráfico Q-Q (cuantil-cuantil) compara los cuantiles de dos distribuciones. Se utiliza para evaluar si una distribución de datos sigue una distribución teórica específica (por ejemplo, normal). Si los puntos del gráfico siguen una línea recta, las distribuciones son similares.

Crear un gráfico Q-Q en R

```
data <- rnorm(100)
```

```
qqnorm(data)
```

```
qqline(data, col="red")
```

9. Explique la diferencia entre un gráfico de barras y un histograma.

Respuesta: Un gráfico de barras se utiliza para representar variables categóricas o discretas, con espacios entre las barras que indican que los valores no están relacionados. Un histograma, en cambio, se usa para variables continuas o discretas, y las barras están adyacentes, ya que representan la frecuencia de los datos en intervalos consecutivos.

10. ¿Qué información proporciona un gráfico de densidad comparado con un histograma y cuándo sería recomendable utilizarlo?

Respuesta: Un gráfico de densidad suaviza los datos y muestra una estimación continua de la distribución de la variable, mientras que un histograma muestra las frecuencias en intervalos discretos. El gráfico de densidad es recomendable cuando se desea observar la forma general de la distribución sin depender del número y tamaño de los intervalos (bins) en un histograma.

Comparar histograma con gráfico de densidad en R

```
data <- rnorm(100)
```

```
hist(data, prob=TRUE, col="lightblue", main="Histograma y gráfico de densidad")
```

```
lines(density(data), col="red", lwd=2)
```



ANEXO N.º 03: BASE DE DATOS

Nº	PROBABILIDADES						TOTAL	TOTAL
	Distribuciones de probabilidad para variables discretas		Distribuciones de probabilidad para variables continuas		Gráficos de distribuciones de probabilidad			
	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST	PRE	POST
1	12	14	6	11	8	12	9	12
2	9	15	9	11	9	14	9	13
3	10	16	5	8	11	13	9	12
4	9	13	6	12	7	11	7	12
5	14	17	9	14	13	15	12	15
6	11	15	13	16	7	8	10	13
7	14	19	8	10	7	9	10	13
8	15	18	15	17	15	15	15	17
9	8	14	7	8	12	15	9	12
10	14	16	7	14	12	17	11	16
11	13	18	5	6	8	9	9	11
12	7	12	13	20	12	14	11	15
13	11	16	13	15	11	17	12	16
14	11	16	9	13	10	13	10	14
15	8	12	9	11	14	15	10	13
16	7	12	13	14	12	15	11	14
17	7	10	8	13	10	12	8	12
18	10	16	8	10	8	9	9	12
19	8	9	9	10	10	11	9	10
20	7	12	10	14	13	17	10	14
21	12	15	7	14	12	15	10	15
22	8	14	12	19	10	13	10	15
23	10	13	13	17	7	10	10	13
24	10	16	8	13	12	14	10	14
25	10	12	6	12	13	17	10	14
26	9	14	12	16	14	19	12	16
27	8	10	11	13	8	13	9	12
28	14	17	5	9	7	10	9	12
29	11	15	7	11	11	14	10	13
30	14	17	9	11	11	16	11	15
31	9	13	10	13	9	10	9	12
32	9	15	12	13	10	14	10	14
33	7	12	7	13	8	14	7	13
34	12	13	7	14	14	16	11	14
35	12	16	5	8	9	15	9	13
36	8	11	10	11	14	20	11	14
37	10	14	9	11	10	12	10	12

38	7	12	9	15	9	11	8	13
39	7	9	7	8	10	15	8	11
40	10	16	5	12	13	18	9	15
41	11	15	10	13	10	15	10	14
42	10	14	11	18	9	13	10	15
43	13	18	8	12	8	10	10	13
44	12	17	6	10	13	17	10	15
45	9	14	9	11	13	19	10	15
46	12	15	11	14	11	12	11	14
47	10	15	6	7	9	10	8	11
48	10	12	7	9	11	14	9	12
49	7	13	11	17	11	17	10	16
50	14	20	5	7	8	10	9	12
51	8	14	10	11	12	16	10	14
52	10	16	8	13	10	12	9	14
53	9	13	11	17	8	12	9	14
54	10	15	7	12	13	16	10	14