GRUPO BLR

UEB
UNIVERSIDAD ESTATAL DE BOLIVAR

ESTADÍSTICA APLICADA: HERRAMIENTA PARA LA INVESTIGACIÓN, LA EDUCACIÓN Y LA TOMA DE DECISIONES

Datos, estrategia, éxito digital.

Raúl Marcelo Chávez Benavides
Elsita Margoth Chávez García
Jhosselyn Briggeth Garcia Aldaz
Darwin Vladimir Rivera Piñaloza

ISBN: 978-9907-0-0423-6

2025

ESTADÍSTICA APLICADA: HERRAMIENTA PARA LA INVESTIGACIÓN, LA EDUCACIÓN Y LA TOMA DE DECISIONES

AUTORES:

RAÚL MARCELO CHÁVEZ BENAVIDES

ELSITA MARGOTH CHÁVEZ GARCÍA

JHOSSELYN BRIGGETH GARCIA ALDAZ

DARWIN VLADIMIR RIVERA PIÑALOZA

ISBN: 978-9907-0-0423-6

Este libro ha sido debidamente examinado y valorado en la modalidad doble par ciego con fin de garantizar la calidad científica.

©Grupo Editorial BLR
Universidad Estatal de Bolívar
Riobamba – Ecuador
Correo: publicaciones@grupoblr.com
https://grupoblr.com/libros-investigo
REPOSITORIO

Chávez, R., Chávez, E., García, J., Rivera, D. (2025) Estadística aplicada: herramienta para la investigación, la educación y la toma de decisiones. Grupo Editorial BLR.

© Raúl Marcelo Chávez Benavides
Elsita Margoth Chávez García
Jhosselyn Briggeth Garcia Aldaz
Darwin Vladimir Rivera Piñaloza

ISBN: 978-9907-0-0423-6

FILIACIONES DE LOS AUTORES

Raúl Marcelo Chávez Benavides

Universidad Estatal de Bolívar

Correo Electrónico: raul.chavez@ueb.edu.ec

ORCID: https://orcid.org/0009-0007-5323-2728

Elsita Margoth Chávez García

Universidad Estatal de Bolívar

Correo Electrónico: emchavez@ueb.edu.ec

ORCID: https://orcid.org/0000-0001-7290-162

Jhosselyn Briggeth García Aldaz

Universidad Estatal de Bolívar

Correo Electrónico: jhosselyn.garcia@ueb.edu.ec

ORCID: https://orcid.org/0009-0001-2210-376X

Darwin Vladimir Rivera Piñaloza

Universidad Estatal de Bolívar

Correo Electrónico: vrivera@ueb.edu.ec

ORCID: https://orcid.org/0000-0002-5695-9726

GRUPO BLR

UEB
UNIVERSIDAD
ESTATAL DE BOLIVAR

PRÓLOGO

Este libro es una invitación a mirar la estadística como una herramienta cercana, útil y necesaria en un mundo cada vez más digital. Los autores logran un equilibrio valioso: explicar conceptos fundamentales con claridad y, al mismo tiempo, mostrar cómo se aplican en situaciones reales vinculadas al marketing digital.

Cada capítulo muestra que los datos no deben entenderse únicamente como registros numéricos, sino como insumos esenciales para fundamentar decisiones. A lo largo del texto, la organización de la información y la interpretación de resultados se acompañan de ejemplos aplicados que vinculan los conceptos teóricos con la práctica profesional, lo que contribuye a comprender un campo que en ocasiones se percibe como exigente.

Estoy convencido de que este texto servirá de guía tanto para los estudiantes que se inician en la estadística como para quienes buscan reforzar su formación. Más allá de fórmulas y tablas, lo que se propone es aprender a pensar con criterio, a mirar la realidad desde los datos y a reconocer en ellos una oportunidad para mejorar lo que hacemos cada día.

Ing. Carlos Luis Cherres Quiroz

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

Este libro nace con el propósito de ofrecer un recurso que vincule la estadística con los desafíos actuales, especialmente aquellos relacionados con el marketing digital. A lo largo de sus páginas se busca mostrar que la estadística no es un conjunto de fórmulas abstractas, sino una herramienta práctica para comprender datos, analizar tendencias y tomar decisiones fundamentadas en escenarios cada vez más competitivos y cambiantes.

El primer capítulo desarrolla los fundamentos esenciales de la estadística y los procedimientos para organizar los datos. En él se explica la diferencia entre población y muestra, los tipos de variables y las escalas de medición, además de presentar las formas más utilizadas para representar la información. Este apartado constituye el punto de partida para que el lector comprenda cómo la adecuada clasificación y estructuración de los datos permite realizar análisis posteriores con mayor precisión.

El segundo capítulo desarrolla las medidas de tendencia central, dispersión y forma. Se explica cómo la media, la mediana, la moda y otras medidas aportan una visión resumida de los datos, mientras que la varianza, la desviación estándar y el rango y coeficiente de variación permiten comprender la variabilidad y la distribución de la información. Cada concepto se acompaña de ejemplos aplicados al marketing digital, como la segmentación de clientes o el análisis de campañas.

El Capítulo III se centra en las técnicas de muestreo, la confiabilidad de los datos y la probabilidad como herramienta de análisis. Aquí se

discuten los diferentes métodos de selección de muestras, la importancia de contar con datos representativos y la forma en que la probabilidad ayuda a estimar escenarios futuros. Los ejemplos prácticos muestran cómo estas herramientas resultan claves en estudios de mercado y encuestas digitales.

Finalmente, el Capítulo IV profundiza en la aplicación de la estadística descriptiva a la interpretación de resultados y su vínculo con la toma de decisiones. Se busca que el lector no solo domine los cálculos, sino que sea capaz de reflexionar sobre el sentido de los números en contextos reales, estableciendo conexiones entre teoría y práctica.

Metodológicamente, el libro combina explicaciones conceptuales con ejemplos prácticos y ejercicios que refuerzan cada tema. El propósito central es que los estudiantes y profesionales del marketing digital desarrollen un pensamiento analítico que les permita interpretar la información con criterio y confianza. El texto está dirigido principalmente a estudiantes universitarios, pero también a docentes y profesionales interesados en fortalecer su capacidad de análisis en un mundo en el que los datos marcan la diferencia.

CAPÍTULO I

1 PRINCIPIOS DE LA ESTADÍSTICA Y MÉTODOS DE ORGANIZACIÓN DE LA INFORMACIÓN

La estadística constituye una disciplina clave para transformar los datos en información significativa. En este capítulo se abordan los fundamentos que permiten comprender cómo se construye el conocimiento a partir de la evidencia cuantitativa. Se analizan conceptos esenciales, las formas en que se clasifican y describen los datos, así como los procedimientos más utilizados para organizarlos y presentarlos. La finalidad es proporcionar al lector una base sólida que facilite interpretar la realidad con rigor, aplicando herramientas estadísticas que resultan indispensables en distintos campos profesionales.

1.1 Introducción a la estadística

En el campo del marketing digital, la efectividad de las decisiones estratégicas depende directamente de la capacidad para analizar e interpretar los datos con criterio y precisión. La estadística descriptiva se convierte, en este sentido, en una aliada indispensable, ya que brinda las herramientas necesarias para recopilar información, organizarla con criterio y transformarla en conclusiones útiles para diseñar y ajustar estrategias.

El presente libro tiene como finalidad orientar a los estudiantes de la asignatura *Estadística Descriptiva* de la carrera de *Marketing Digital*, ofreciendo contenidos y ejemplos aplicados que faciliten su aprendizaje. Más allá de ser un texto académico, busca convertirse en un recurso

práctico que acompañe al futuro profesional en su formación y que le permita enfrentarse a problemas reales del sector con una base sólida de análisis.

El valor de la estadística descriptiva radica en su capacidad para dar sentido a grandes volúmenes de datos. Gracias a sus técnicas, es posible identificar tendencias del mercado, segmentar consumidores y evaluar la efectividad de campañas digitales, entre otras aplicaciones. En un entorno donde la información fluye en tiempo real, dominar estas herramientas es clave para generar decisiones ágiles y fundamentadas.

A lo largo de los distintos capítulos se desarrollan los temas esenciales de la estadística descriptiva, organizados de forma progresiva y vinculados con situaciones reales del marketing digital. El primer capítulo presenta los fundamentos conceptuales y las estrategias básicas para la organización y clasificación de los datos, estableciendo los cimientos del análisis estadístico. Por su parte, el segundo capítulo profundiza en las medidas de tendencia central, dispersión y forma, herramientas indispensables para sintetizar la información, evaluar su variabilidad y extraer conclusiones significativas a partir de los datos. El tercer capítulo aborda las técnicas de muestreo, la confiabilidad de los resultados y las bases de la probabilidad como herramienta de análisis. Finalmente, el último capítulo se centra en las pruebas de hipótesis de acuerdo a las características de la muestra.

Cada sección incorpora ejemplos y casos prácticos que ayudan a relacionar la teoría con la práctica profesional. Al finalizar la lectura, los estudiantes no solo habrán comprendido los principios fundamentales

de la estadística, sino que también desarrollarán las habilidades necesarias para aplicarla en la optimización de estrategias digitales, el análisis del comportamiento del consumidor y la toma de decisiones basada en datos.

1.2 Conceptualización de la estadística

La estadística se concibe como un conjunto de métodos y técnicas destinados a recopilar, organizar y analizar datos, con el propósito de convertirlos en información útil que facilite su interpretación y aplicación. Su valor radica en que facilita la interpretación de fenómenos complejos, ayudando a tomar decisiones mejor fundamentadas.

En el campo académico y profesional, la estadística no se limita únicamente a cálculos matemáticos. Se trata de una disciplina que ofrece un lenguaje común para describir la realidad, identificar patrones y anticipar escenarios. Desde el recuento de preferencias de los consumidores hasta la evaluación de resultados en una campaña publicitaria, su alcance se extiende a prácticamente cualquier área donde se generen datos.

1.2.1 Estadística en el ámbito del marketing digital

En el marketing digital, la estadística cobra una relevancia especial porque la mayor parte de las actividades se apoyan en datos: visitas a un sitio web, clics en anuncios, interacciones en redes sociales o niveles de conversión en una campaña. Analizar estos indicadores con criterio estadístico permite a las empresas comprender mejor a su público, detectar oportunidades y corregir estrategias a tiempo.

Por ejemplo, una empresa que lanza una campaña en redes sociales no se limita a contar los "me gusta" obtenidos. Mediante técnicas estadísticas puede segmentar a los usuarios según edad, ubicación o intereses, y evaluar cuál grupo respondió mejor al mensaje. Este análisis convierte simples números en información estratégica, lo que demuestra cómo la estadística descriptiva se convierte en una herramienta esencial en la gestión de proyectos digitales.

1.3 Clasificación: Estadística descriptiva e inferencial

Dentro de la disciplina estadística se reconocen dos vertientes fundamentales. La primera es la descriptiva, encargada de sistematizar y presentar la información; la segunda es la inferencial, cuyo propósito es realizar estimaciones y predicciones sobre la población a partir de una muestra. Aunque ambas se complementan, es importante reconocer sus diferencias para entender cómo se aplican en la práctica.

La estadística descriptiva reúne las técnicas destinadas a resumir y organizar los datos. Su objetivo es mostrar la información de forma clara, ya sea a través de tablas, gráficos o medidas numéricas como la media o la desviación estándar. En marketing digital, por ejemplo, la estadística descriptiva permite identificar la edad promedio de los clientes que compran en una tienda en línea o mostrar, mediante un gráfico de barras, cuáles son los productos más visitados en un catálogo virtual.

La estadística inferencial, por su parte, amplía el alcance del análisis descriptivo al permitir extraer conclusiones sobre una población a partir de la observación de una muestra. Este enfoque resulta fundamental

cuando no es viable estudiar a todos los individuos, como ocurre en el análisis del comportamiento de los clientes, donde se trabaja con grupos representativos que reflejan las características del conjunto total. En estos casos, se aplican técnicas que permiten hacer estimaciones o probar hipótesis. Por ejemplo, una empresa puede analizar la opinión de 500 usuarios sobre una campaña digital y, con base en esos resultados, inferir la percepción de todos sus clientes.

Como señalan Triola y colaboradores (2020), la estadística inferencial se apoya en la probabilidad para "generalizar conclusiones de una muestra hacia una población más amplia" . Esto convierte a la estadística en un puente entre los datos inmediatos y la toma de decisiones a gran escala (Triola , 2020).

En resumen, la estadística descriptiva se encarga de organizar y mostrar los datos de manera comprensible, mientras que la estadística inferencial permite proyectar conclusiones, identificar tendencias y respaldar decisiones a partir del análisis de una muestra representativa. Para quienes se forman en marketing digital, entender esta diferencia es esencial, ya que ambas perspectivas se aplican constantemente en la gestión de proyectos, la segmentación de audiencias y la medición de resultados de las campañas.

1.4 La estadística como herramienta fundamental para la toma de decisiones

En un mundo donde la información se multiplica a cada segundo, la estadística se ha convertido en una brújula que orienta las decisiones de personas, empresas e instituciones. Su valor radica en que transforma

los datos en conocimiento, lo que permite identificar patrones, anticipar escenarios y actuar con mayor seguridad.

En el caso del marketing digital, la importancia es aún mayor. Cada clic, cada visita a un sitio web o cada interacción en redes sociales genera datos que, al ser analizados con herramientas estadísticas, muestran mucho más que simples números. Revelan comportamientos de consumidores, preferencias de compra y tendencias que ayudan a diseñar campañas más efectivas y a invertir de manera más inteligente.

La estadística, como señalan Levin y Rubin (2017), se ha consolidado como "una herramienta fundamental para tomar decisiones en contextos de incertidumbre, ya que permite evaluar riesgos y reducir la posibilidad de error" . Esto explica por qué las grandes compañías tecnológicas, los negocios digitales y hasta los pequeños emprendimientos recurren cada vez más a su aplicación cotidiana (Levin & Rubin , 2017).

En la práctica académica, aprender estadística no solo ayuda a resolver ejercicios en el aula, sino que desarrolla la capacidad crítica y analítica del estudiante. Esta competencia resulta esencial para un futuro profesional, ya que lo prepara para enfrentar escenarios cambiantes, interpretar información con criterio y responder a las demandas del mercado con propuestas fundamentadas en evidencia.

En definitiva, la estadística no es un conjunto de fórmulas aisladas, sino un lenguaje que conecta la teoría con la realidad. Gracias a ella, las decisiones dejan de ser intuiciones y se convierten en estrategias fundamentadas.

1.5 Conceptos básicos

La estadística se concibe como una disciplina que integra métodos para recopilar, organizar, analizar e interpretar datos, con el propósito de convertirlos en información útil que oriente la toma de decisiones fundamentadas. Su importancia radica en que, en la vida real, los datos suelen estar dispersos y no siempre resultan fáciles de comprender. Gracias a la estadística, es posible darles orden y sentido, lo que permite que gobiernos, empresas e instituciones actúen con mayor claridad y fundamento.

En la actualidad, el desarrollo de herramientas informáticas ha fortalecido notablemente los procesos estadísticos, permitiendo analizar grandes volúmenes de datos con mayor rapidez y precisión. El uso de software especializado no solo optimiza el tratamiento de la información, sino que también facilita la elaboración de reportes comprensibles y visualmente claros. Gracias a ello, la estadística se ha consolidado como un recurso esencial en campos tan diversos como la economía, la educación y el marketing digital. Por ejemplo, un gobierno puede analizar el salario promedio de la población y, con base en ese dato, decidir sobre políticas salariales que beneficien a los trabajadores y evalúen su impacto en la economía nacional.

Desde una perspectiva académica, la estadística se clasifica en dos ramas principales. La primera es la estadística descriptiva, que se encarga de organizar, resumir y presentar los datos de manera informativa, sin realizar inferencias más allá de la información disponible (Triola M. F., 2020).

La segunda, denominada estadística inferencial**,** se centra en obtener conclusiones y estimaciones sobre una población utilizando la información proveniente de una muestra representativa. Con este enfoque, no es necesario estudiar a todos los individuos, sino trabajar con un grupo más pequeño que refleje las características esenciales del total (Levin & Rubin , 2017). Dentro de esta rama surgen dos conceptos fundamentales:

- **Población:** Entendida como el conjunto completo de elementos u observaciones que se desea estudiar, representada con la letra N.
- **Muestra:** Que corresponde a un subconjunto de la población, suficientemente representativo para realizar inferencias válidas, representada con la letra n.

De esta manera, la estadística no solo describe lo que ocurre en un conjunto de datos, sino que también ofrece la posibilidad de hacer estimaciones y predicciones que guían las decisiones en escenarios reales.

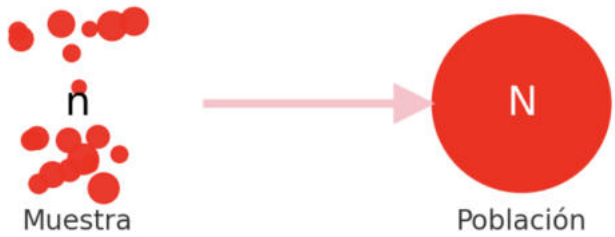

Figura 1. Población y muestra.

Una vez establecido que *N* representa la muestra, es necesario señalar que en la práctica resulta complejo acceder a la totalidad de la población.

Esto se debe a limitaciones relacionadas con el costo, el tiempo disponible, la inviabilidad de ciertos procedimientos y, en algunos casos, la naturaleza destructiva del estudio.

Ejemplo: Acceder al ingreso percibido por cada uno de los 12 millones de trabajadores resultaría complicado, se justifica entonces trabajar con datos más reducidos, es decir, con muestras (Hernández , Fernández, & Baptista, 2022).

Existen diferentes tipos de muestreo; pero se los puede clasificar en dos grandes grupos: el muestreo probabilístico y el muestreo no probabilístico (Lind, Marchal, & Wathen , 2018).

- **Muestreo probabilístico:** Consiste en seleccionar una muestra de la población de manera que todos los elementos tengan la misma probabilidad de ser incluidos en el estudio. En otras palabras, la probabilidad de inclusión es conocida y diferente de cero ($pi \neq 0$).

- **Muestreo no probabilístico:** En este tipo de muestreo, no todos los integrantes de la población cuentan con la misma probabilidad de ser seleccionados para conformar la muestra. Esto puede generar sesgos y limitar la representatividad de los resultados.

Una empresa que desee realizar un estudio de mercados, deberá tener en conside-ración que para que el estudio de mercados tenga validez, la muestra obtenida en dicho estudio, será obtenida mediante un muestreo probabilístico (Lind, Marchal, & Wathen , 2018).

1.5.1 Variables: Cualitativas y cuantitativas

Cuando hablamos de una población o de una muestra, en realidad nos estamos refiriendo a la información que compone a esos conjuntos. Es en este punto donde aparece un concepto fundamental: el dato estadístico. Un dato es cada valor que se obtiene de la observación de una característica, y puede corresponder tanto a un número como a una cualidad.

El siguiente concepto clave es la variable, entendida como aquella característica de interés que puede medirse u observarse en los individuos de una población (Levin & Rubin , 2017).

De acuerdo con la forma en que se manifiestan, las variables pueden agruparse en dos categorías principales.

- **Variables cualitativas** que describen atributos o características que no pueden medirse con números. Se expresan en categorías, como el color de ojos, la profesión, el estado civil o la marca de un producto. Estas variables pueden ser:
- **Nominales:** Cuando no existe un orden natural entre las categorías (por ejemplo: género, nacionalidad o marcas comerciales).
- **Ordinales:** Cuando sí existe un orden o jerarquía entre las categorías (por ejemplo: nivel educativo o escalas de satisfacción en encuestas).

- **Variables cuantitativas** que expresan características numéricas y permiten realizar operaciones aritméticas. Estas pueden dividirse en:

- **Discretas:** Son variables que solo pueden adoptar un número finito o contable de valores. Ejemplos de este tipo son el número de hijos en un hogar o la cantidad de empleados que conforman una empresa.
- **Continuas:** Pueden adoptar una cantidad infinita de valores dentro de un intervalo determinado. Ejemplos de ello son la estatura de una persona, el peso de un producto o el tiempo que dura un viaje.

Además, los datos estadísticos pueden organizarse de acuerdo con su forma de recolección:

- **Datos de corte transversal**, que recogen información en un solo momento del tiempo para distintos individuos o unidades, como una encuesta a varias familias sobre su consumo mensual.
- **Datos temporales o series de tiempo,** que registran la evolución de una variable en intervalos regulares, como las ventas de una tienda a lo largo de un año.
- **Datos de panel,** que combinan los dos anteriores, observando a distintos individuos en varios momentos, muy usados en estudios de mercadotecnia para seguir el comportamiento de consumidores en temporadas específicas (Triola M. F., 2020).

En síntesis, comprender las variables y la naturaleza de los datos es indispensable para un análisis estadístico riguroso. Su correcta

identificación permite seleccionar las técnicas adecuadas, organizar mejor la información y, en consecuencia, obtener conclusiones más confiables.

1.5.2 Tipos de escalas de medición

El modo en que se recopila y clasifica la información determina los cálculos que pueden realizarse y las pruebas estadísticas que son aplicables. A este principio se le conoce como nivel de medición de los datos, y constituye la base para organizar y analizar la información con rigor (Triola , 2020).

En términos generales, las variables pueden medirse en cuatro niveles distintos: nominal, ordinal, de intervalo y de razón. Cada uno de ellos ofrece distintas posibilidades de análisis, desde la simple clasificación hasta operaciones matemáticas más complejas.

- **Nivel nominal:** Este nivel representa la forma más elemental de medición. Los datos se organizan en categorías que describen una característica o atributo, sin establecer ningún tipo de jerarquía u orden entre ellas. Un ejemplo son los colores, el género o la nacionalidad. En el ámbito cultural, cuando se registra la procedencia de los asistentes a un museo, los datos solo permiten agruparlos por ciudad o país, sin establecer jerarquías entre las categorías.

Este tipo de nivel resulta útil para clasificar y contabilizar, pero no permite realizar operaciones matemáticas como sumas o promedios. Su valor radica en que facilita la organización inicial de la información, lo

que constituye el primer paso para un análisis estadístico más profundo (Levin & Rubin , 2017).

En la siguiente tabla se muestran los datos recopilados y organizados sobre la asistencia al museo en la ciudad:

Tabla 1. Clasificación de datos en escala nominal.

Categoría	Total
Hombres	30
Mujeres	33

En este nivel los datos se organizan en categorías y se procede únicamente a su conteo, garantizando que cada valor pertenezca a una sola categoría, sin posibilidad de repetición.

- **Nivel ordinal**: se caracteriza porque los datos, además de estar clasificados en categorías, pueden ordenarse siguiendo una secuencia lógica. Esto significa que una categoría se reconoce como "mayor" o "más alta" que otra, estableciendo jerarquías entre los valores observados (Triola , 2020).

La siguiente tabla presenta un ejemplo de cómo puede clasificarse la información relacionada con el tipo de habitación disponible en un hotel.

Tabla 2. Datos.

Orden	Tipo de habitación (nominal)	Número de habitaciones
1	Suite	3
2	Triple	5
3	Doble	12
4	Simple	13

- **Nivel de intervalo** se distingue porque, además de clasificar y ordenar los datos, permite medir la distancia exacta entre los valores. En este nivel, las categorías son mutuamente excluyentes, exhaustivas y se organizan en un orden lógico, pero lo más relevante es que las diferencias entre los números asignados a cada categoría representan distancias iguales en la característica medida (Triola, 2020).

Un ejemplo representativo de escala de intervalo es la medición de la temperatura, expresada en grados Celsius o Fahrenheit, donde la distancia entre 20 °C y 30 °C equivale a la existente entre 30 °C y 40 °C. Sin embargo, en este nivel no existe un punto cero absoluto que denote la ausencia de la propiedad, lo que limita la posibilidad de realizar comparaciones de razón. Otro ejemplo lo constituye la clasificación de hoteles mediante estrellas.

La diferencia entre un hotel de tres estrellas y uno de cinco mantiene el mismo significado en distintas ciudades, ya sea en Cuenca o en Nueva York.

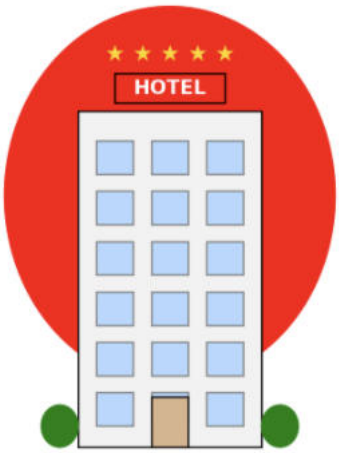

Figura 2. Diferencia de hoteles.

Fuente: *Adaptado de Triola (2018).*

- **Nivel de razón:** es el más completo de los cuatro niveles de medición, ya que incorpora todas las características de los anteriores: clasificación, orden y distancias iguales entre categorías. Lo que lo distingue es la presencia de un cero absoluto, entendido como la ausencia total de la característica medida. Esto permite no solo comparar diferencias, sino también establecer proporciones y realizar operaciones matemáticas más complejas (Triola, 2020).

Un ejemplo cotidiano se encuentra en la medición de los ingresos. Si un vendedor no percibe salario, el valor será cero, lo que representa la falta total de esa característica. En cambio, si un ejecutivo gana 2.000 dólares y otro 1.000, es posible afirmar que el primero percibe el doble de

ingresos que el segundo, algo que no puede afirmarse en escalas de intervalo.

Un caso ilustrativo lo constituyen los salarios de un vendedor o de un ejecutivo de ventas. En este tipo de medición, es posible que alguien no perciba ingresos, de modo que el valor cero representa una ausencia real de la variable.

Figura 3. Nivel de razón.

Fuente: *Adaptado de Triola (2018).*

1.6 Organización y presentación de datos

Una vez que se han definido los conceptos básicos y los niveles de medición, el siguiente paso consiste en organizar y presentar los datos de manera clara. Esta fase es crucial, porque los datos en bruto suelen carecer de significado; en cambio, cuando se estructuran en tablas o gráficos, se convierten en información comprensible y útil para la toma de decisiones (Anderson, Sweeney, & Williams , 2016).

La organización de los datos busca dar orden y coherencia a la información recolectada. A través de la tabulación, se agrupan los valores en categorías o intervalos que permiten observar patrones y tendencias. En el caso del marketing digital, esto puede significar clasificar a los usuarios de una tienda en línea según su edad, el tiempo de permanencia en la página o el número de compras realizadas.

La presentación de los datos se realiza mediante tablas de frecuencia y representaciones gráficas. Estas herramientas no solo ayudan a resumir la información, sino que también facilitan la comunicación de resultados a públicos diversos. Como señala Malhotra (2019), "una tabla o gráfico bien diseñado transmite la esencia de los datos de forma más rápida y efectiva que una explicación extensa" (p. 87) (Malhorta, 2019).

En la práctica, las tablas de frecuencia permiten identificar cuántas veces aparece un valor o categoría dentro del conjunto de datos. Por ejemplo, si una empresa analiza los resultados de una campaña de correo electrónico, puede construir una tabla que muestre la cantidad de clientes que abrió el mensaje, lo ignoró o lo marcó como spam. Esta organización facilita calcular proporciones y detectar comportamientos dominantes.

Las representaciones gráficas, por su parte, son especialmente útiles para visualizar la información. Los gráficos de barras, de pastel o los histogramas ofrecen un panorama inmediato de las tendencias. En el ámbito digital, un histograma puede ilustrar cómo se distribuyen las visitas a un sitio web a lo largo del día, mientras que un gráfico de barras

puede mostrar las redes sociales que generaron más interacción durante una campaña.

Además, el uso de gráficos responde a una necesidad práctica: captar la atención de los usuarios en entornos donde el tiempo es limitado. Un profesional del marketing digital que presenta un informe a su equipo de trabajo necesita comunicar resultados de forma sencilla y atractiva, y en ese sentido, las herramientas gráficas cumplen un papel fundamental (Hair, Wolfinbarger, Money, Samouel, & Page, 2015).

En conclusión, organizar y presentar los datos no es solo un paso técnico, sino también estratégico. De esta etapa depende que la información recolectada se convierta en conocimiento aplicable, capaz de guiar las decisiones en proyectos reales.

1.6.1 Distribuciones de frecuencia de datos: Cuantitativas y cualitativas

Una vez obtenidos los datos de una muestra, el paso siguiente consiste en organizarlos de forma que puedan interpretarse con facilidad y aporten información útil para el análisis. Una de las técnicas más empleadas es la distribución de frecuencias, que permite resumir la información y mostrarla de forma ordenada. Su objetivo es facilitar la interpretación de los resultados y evidenciar patrones que de otra manera permanecerían ocultos (Anderson, Sweeney, & Williams , 2016).

Una distribución de frecuencias consiste en una tabla que organiza los valores o categorías de una variable, mostrando cuántas veces se repite cada uno dentro del conjunto de datos. Este método puede aplicarse

tanto a variables cualitativas como a variables cuantitativas. Para datos cualitativos, las categorías pueden ser:

- **Nominales:** Cuando no hay un orden lógico entre ellas (por ejemplo, colores o marcas de productos).
- **Ordinales:** Cuando sí existe un orden o jerarquía (como el nivel educativo o una escala de satisfacción).

Para datos cuantitativos, las variables pueden ser:

- **Discretas:** Toman solo valores enteros, por ejemplo: número de hijos, cantidad de compras en línea.
- **Continuas,** que pueden asumir infinitos valores dentro de un rango (ejemplo: Estatura, tiempo de permanencia en un sitio web).

Asimismo, una tabla de frecuencias incorpora diversas medidas que permiten describir y analizar el comportamiento general de los datos.

- **Frecuencia absoluta (fi):** Representa el número de veces que un valor o categoría se repite dentro del conjunto de datos.
- **Frecuencia relativa (fi%):** Se obtiene al dividir la frecuencia absoluta entre el total de observaciones, y generalmente se expresa en forma porcentual.
- **Frecuencia absoluta acumulada (FAi):** Corresponde a la suma progresiva de las frecuencias absolutas hasta un determinado valor o categoría.
- **Frecuencia relativa acumulada (FAi%):** Resulta de dividir la frecuencia acumulada entre el total de datos, mostrando la proporción acumulada de casos observados.

Cuando la cantidad de datos es limitada, es posible elaborar una distribución unidimensional de frecuencias, en la que se presentan en una sola tabla los valores o categorías junto con la frecuencia con que aparecen. En el marketing digital, este tipo de tabla puede utilizarse para analizar, por ejemplo, cuántas veces los usuarios interactúan con una publicación en redes sociales según categorías como “me gusta”, “comentarios” y “compartidos”.

Como señala Malhotra (2019), la construcción de distribuciones de frecuencia es un paso esencial en la investigación, porque “permite simplificar grandes volúmenes de datos y presentarlos de forma clara para la toma de decisiones” (Malhorta, 2019). En otras palabras, se trata de una herramienta que transforma la información dispersa en conocimiento estructurado.

Cuando la cantidad de datos es reducida, también llamados datos simples, estos pueden organizarse en una tabla conocida como distribución unidimensional de frecuencias, como se muestra a continuación.

Tabla 3. Distribución unidimensional de frecuencias.

x_i	*fi*	*fi%*	*FAi*	*FAi%*
x_1				
x_2				
:				
x_3				

Primer ejemplo para datos cuantitativos

Considere el siguiente registro de horas trabajadas semanalmente durante los últimos dos meses: 52, 48, 37, 54, 48, 15, 42 y 12. Con base en estos valores, construya una tabla de distribución de frecuencias.

Tabla 4. Datos cuantitativos.

x_i	*fi*	*fi%*	*FAi*	*FAi%*
52	2	5,56%	2	5,56%
48	7	19,44%	9	25,00%
37	6	16,67%	15	41,67%
54	5	13,89%	20	55,56%
48	3	8,33%	23	63,89%
15	1	2,78%	24	66,67%
42	8	22,22%	32	88,89%
12	4	11,11%	36	100,00%
	Σ=36	Σ=100%		

Segundo ejemplo para datos cualitativos nominales

A continuación se realiza una encuesta a 50 personas para saber el color de sus ojos, de los cuales se obtienen los siguientes datos:

Tabla 5. Datos cualitativos nominales.

Color de ojos	fi	fi %	FA	FA%
Azules	20	40%	20	40%
Verdes	15	30%	35	70%
Castaños	15	30%	50	100%
Total	50	100%		

1.6.2 Distribución de frecuencia de datos cuantitativos: Discretas y continuos (serie con intervalos)

Cuando los conjuntos de datos son muy extensos, se hace necesario agrupar la información en clases o intervalos. Esta estrategia facilita la lectura de los resultados y permite detectar patrones que de otra manera pasarían desapercibidos. Sin embargo, es importante recordar que la agrupación también implica una cierta pérdida de detalle, ya que se resumen varios valores dentro de un mismo intervalo (Anderson, Sweeney, & Williams , 2016).

Una clase representa un rango de valores delimitado por un límite inferior (Li–1) y un límite superior (Li). Para construir estas clases es necesario seguir un procedimiento que incluye dos pasos principales:

- **Determinación del número de clases**

Existen varios métodos que ayudan a decidir en cuántas clases conviene agrupar los datos:

- Regla de Sturges:

$$k = 1 + 3.22 \log(n)$$

Donde n representa el número de observaciones.

Ejemplo: Si se tienen datos de índices de alfabetización en 57 países:

$$k = 1 + 3.22\log(57)$$

$$k = 1 + 3.22(1.7558)$$

$$k = 6.83 \approx 7$$

En este caso, se recomienda trabajar con 7 clases.

- Regla empírica de la raíz cuadrada:

$$k = \sqrt{n}$$

Este tipo de organización resulta especialmente conveniente cuando el número de observaciones es inferior a cien, ya que permite representar los datos de manera clara y comprensible.

- Regla empírica de potencias de 2:

Se elige el menor valor de k tal que $2^k \geq n$.

Ejemplo: Para 65 datos:

$$2^5 = 32,$$
$$2^6 = 64,$$
$$2^7 = 128.$$

En este caso, el valor adecuado sería $k = 7$.

- Método subjetivo:

Este método permite establecer el número de clases según el criterio del investigador, siempre que el resultado se mantenga dentro del rango de cinco a veinte intervalos.

- **Cálculo del ancho de clase**

El ancho de clase (ai o Ci) indica la amplitud de cada intervalo y se calcula con la fórmula:

$$ai = \frac{R_e}{k}$$

Donde $R_e$representa el recorrido o rango de los datos, calculado como la diferencia entre el valor máximo y el valor mínimo del conjunto ($R_e = V_{max} - V_{min}$).

- **Marcas de clase**

En la elaboración de distribuciones de frecuencias se emplean también las marcas de clase, conocidas como puntos medios. Estas representan el punto medio de cada intervalo y se determinan sumando los límites inferior y superior para luego dividir el resultado entre dos. Una vez obtenidas las marcas de clase, se continúa con la elaboración de la tabla de frecuencias como si se tratara de datos individuales.

En el ámbito del marketing digital, este procedimiento resulta especialmente valioso para estructurar grandes volúmenes de información, como el registro de visitas a un sitio web durante distintos días o la distribución de compras según rangos de edad. Agrupar los datos permite simplificar la información y obtener conclusiones prácticas sin perder la visión general de las tendencias.

A continuación, un ejemplo:

La agencia de viajes nacional Moore ofrece tarifas especiales en ciertas travesías por el Caribe a ciudadanos de la tercera edad. El presidente de la agencia quiere información adicional sobre las edades de las personas que viajan. Una muestra aleatoria de 40 clientes

que hicieron un crucero el año pasado dio a conocer las siguientes edades:

Tabla 6. Ejemplo distribución de frecuencias.

Número de observaciones	Edad (según la muestra original)	Edad ordenada
1	77	18
2	18	26
3	63	34
4	84	36
5	38	38
6	54	41
7	50	43
8	59	44
9	54	45
10	56	50
11	36	50
12	26	51
13	50	52
14	34	52
15	44	53
16	41	53
17	58	54
18	58	54

19	53	56
20	51	58
21	62	58
22	43	58
23	52	59
24	53	60
25	63	60
26	62	61
27	62	61
28	65	62
29	61	62
30	52	62
31	60	63
32	60	63
33	45	63
34	66	65
35	83	66
36	71	71
37	63	71
38	58	77
39	61	83
40	71	84

Al trabajar con conjuntos de datos extensos, no resulta práctico presentar cada valor de manera individual. Por ello, se recomienda agrupar los datos en clases o intervalos, lo que permite simplificar la información y hacer más evidente su estructura.

Para determinar el número de clases puede aplicarse la regla de la raíz cuadrada, que consiste en calcular la raíz del número total de observaciones:

$$k = \sqrt{n}$$

En un ejemplo con 40 observaciones:

$$k = \sqrt{40} = 6.32$$

Al realizar el redondeo correspondiente, se determina que los datos pueden organizarse en seis clases. A continuación, se determina el recorrido (R_e), definido como la diferencia entre el valor máximo y el valor mínimo del conjunto de datos analizado.

$$Re = V_{max} - V_{min}$$

En este caso:

$$Re = 84 - 18 = 66$$

Con esta información se determina el ancho de clase (ai o Ci), que se obtiene dividiendo el recorrido entre el número de clases:

$$ai = \frac{Re}{k} = \frac{66}{6} = 11$$

El ancho de clase será entonces de 11 unidades. Cuando el resultado es un número decimal, se recomienda ajustar ligeramente el valor para que la presentación de la tabla de frecuencias sea más clara y práctica, evitando que los límites de los intervalos coincidan con los valores exactos de las observaciones (Levin & Rubin , 2017).

Finalmente, se recomienda que el límite inferior de la primera clase sea ligeramente menor que el valor mínimo del conjunto de datos, con el fin de garantizar que todas las observaciones queden comprendidas dentro de los intervalos establecidos. En este ejemplo, puede iniciarse en 15. De la misma manera, el límite superior de la última clase debe ser ligeramente mayor que el valor más alto, garantizando que el intervalo final abarque toda la información disponible (Triola , 2020).

Por tal motivo, las clases quedan definidas de la siguiente manera:

Clases	f	f%	Marca de clase	FA	FA%
15-26	2	5.0%	20.5	2	5.0%
26-37	2	5.0%	31.5	4	10.0%
37-48	5	12.5%	42.5	9	22.5%
48-59	14	35.0%	53.5	23	57.5%
59-70	12	30.0%	64.5	35	87.5%
70-81	3	7.5%	75.5	38	95.0%
81-92	2	5.0%	86.5	40	100.0%
	40	100.0%			

Ejemplo 2

Para comprender mejor cómo se construye una distribución de frecuencias agrupadas, veamos un ejemplo con datos cuantitativos discretos.

Imagina que se registra el número de llamadas recibidas por una central telefónica durante 20 días. Los valores obtenidos son los siguientes: 2, 3, 4, 4, 5, 6, 7, 7, 8, 9, 9, 9, 10, 10, 11, 12, 13, 13, 14, 15

El proceso de construcción de la distribución puede resumirse en varios pasos:

- **Cálculo del rango:** se determina restando el valor mínimo al valor máximo del conjunto de datos.

$$R = 15 - 2 = 13$$

- **Número de intervalos (k):** se determina aplicando la regla empírica de la raíz cuadrada, que consiste en calcular la raíz del número total de observaciones del conjunto de datos.

$$k = \sqrt{20} \approx 4.47 \Rightarrow 4 \text{ intervalos}$$

- **Ancho del intervalo:** se obtiene dividiendo el rango total de los datos entre el número de intervalos establecidos.

$$a_i = \frac{13}{4} = 3.25 \approx 3$$

- **Construcción de los intervalos:** Se parte del valor más pequeño (2) y se suman intervalos de tamaño 3. De esta forma, los intervalos quedan así:

2 – 4

5 – 7

8 – 10

11 – 15

- **Conteo de frecuencias:** Finalmente, se contabiliza cuántos valores caen en cada intervalo.

Tabla 7. Distribución de frecuencia con intervalos (clases).

Intervalos	f	f%	Marca de clase	FA	FA%
2 – 4	4	20%	3	4	20%
5-jul	4	20%	6	8	40%
8-oct	6	30%	9	14	70%
nov-15	6	30%	12	20	**100%**
Total	**20**	**100%**			

1.7 Representación gráfica de datos

Una vez que los datos han sido organizados en una tabla de frecuencias, el paso siguiente es presentarlos de forma gráfica. Las representaciones visuales permiten transmitir la información de manera más clara y rápida, haciendo que incluso un público no especializado pueda comprender los resultados.

Como señalan Anderson, Sweeney y Williams (2016), los gráficos son una herramienta esencial porque "comunican de forma directa lo que una tabla numérica podría tardar mucho más en explicar" (Anderson, Sweeney, Camm, & Cochran , 2019).

En estadística se emplean distintos tipos de representaciones gráficas, cada una con aplicaciones específicas:

- **Diagrama de barras:** Se representa sobre un plano cartesiano, ubicando en el eje horizontal (abscisas) las categorías o valores de la variable y en el eje vertical (ordenadas) las frecuencias correspondientes, ya sean absolutas o relativas.Es ideal para variables discretas, como el número de compras realizadas en una tienda virtual durante un día determinado.

Figura 4. Diagrama de barras.

- **Histograma:** El histograma constituye uno de los gráficos más representativos en el análisis estadístico. Aunque guarda cierta similitud con el diagrama de barras, se diferencia por estar destinado a variables continuas. En este tipo de gráfico, las clases o intervalos se disponen en el eje horizontal y las frecuencias, ya sean absolutas o relativas, se representan en el eje vertical. Un histograma puede mostrar, por ejemplo, la distribución de tiempos de permanencia de los usuarios en una página web.

Figura 5. Histograma.

- **Diagrama de sectores o gráfico circular**: Conocido también como "gráfico de pastel", se utiliza sobre todo con variables discretas. Cada sector representa la proporción o porcentaje de una categoría dentro del total. Es útil, por ejemplo, para mostrar qué porcentaje de clientes prefiere una determinada marca de teléfonos inteligentes frente a la competencia.

Figura 6. Diagrama de sectores o gráfico circular.

- **Pictograma.** Emplea dibujos o íconos relacionados con la variable para representar la frecuencia. Puede hacerse de dos

formas: repitiendo el ícono tantas veces como aparezca la variable o variando el tamaño de la figura de acuerdo con la frecuencia. Estos gráficos son más ilustrativos y suelen usarse en presentaciones dirigidas a un público general.

- **Polígono de frecuencias:** Se obtiene al unir mediante líneas los puntos medios de cada clase en una distribución agrupada. Cuando los datos no están agrupados, se trazan los pares ordenados que relacionan el valor de la variable con su frecuencia correspondiente. Este gráfico permite observar la forma de la distribución de manera más fluida que un histograma.

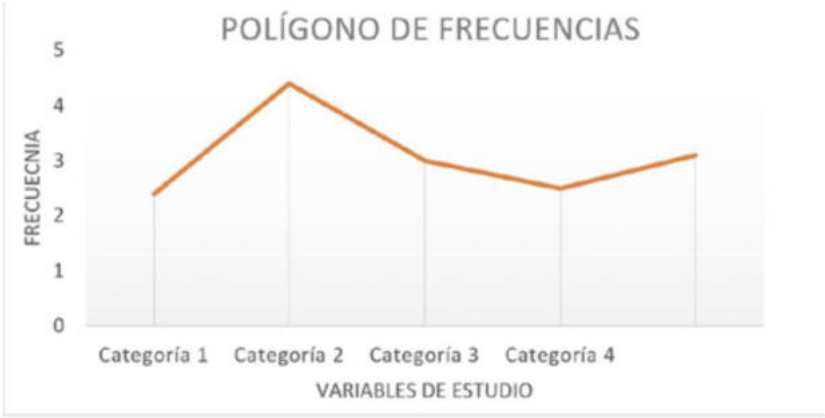

Figura 7. Polígono de frecuencia.

- **Ojiva.** Es un polígono de frecuencias construido a partir de las frecuencias acumuladas (absolutas o relativas). Puede elaborarse con la categoría “menor que” o “mayor que”. La ojiva es útil para identificar percentiles, cuartiles o el porcentaje de observaciones que no superan cierto valor (Triola , 2020).

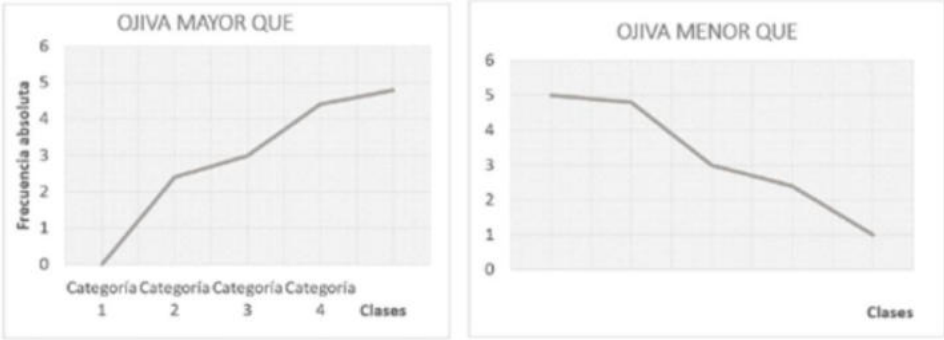

Figura 8. Ojiva.

Ejemplo práctico

Como ejemplo práctico, se puede retomar el conjunto de datos de la agencia de viajes Moore analizado en apartados anteriores. A partir de su distribución de frecuencias, es posible elaborar un histograma, un polígono de frecuencias y una ojiva del tipo “menor que”.

Tabla 8. Tabla de frecuencias.

Clases	f	f%	Marca de clase	FA	FA%
15-26	2	5.0%	20.5	2	5.0%
26-37	2	5.0%	31.5	4	10.0%
37-48	5	12.5%	42.5	9	22.5%
48-59	14	35.0%	53.5	23	57.5%
59-70	12	30.0%	64.5	35	87.5%
70-81	3	7.5%	75.5	38	95.0%
81-92	2	5.0%	86.5	40	100.0%
	40	100.0%			

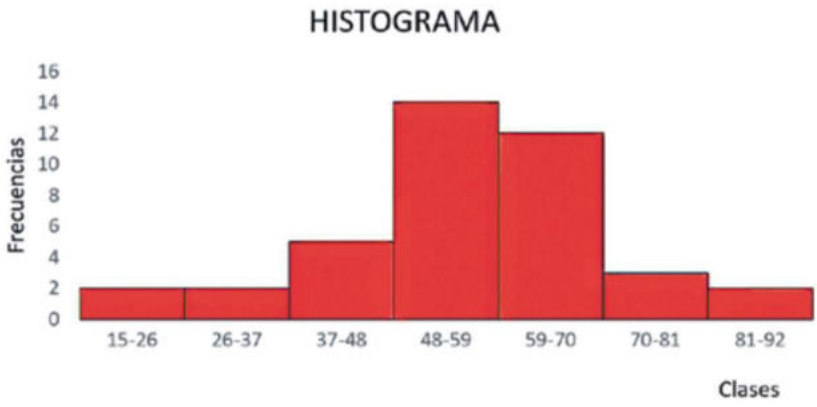

Figura 9. Histograma.

Figura 10. Polígono de frecuencia.

Para construir la ojiva, los datos se organizan inicialmente en una tabla de frecuencias acumuladas. Posteriormente, se grafican los puntos que relacionan el límite superior de cada clase con su frecuencia acumulada, y se unen mediante una línea continua, como se muestra en la tabla correspondiente a la categorización de los datos de la agencia de viajes Moore.

Tabla 9. Clases.

Clases	FA
Menos de 26	2
Menos de 37	4
Menos de 48	9
Menos de 59	23
Menos de 70	35
Menos de 81	38
Menos de 92	40

Figura 11. Ojiva.

De esta forma, las representaciones gráficas convierten a los números en imágenes fáciles de interpretar, favoreciendo la comunicación de resultados en ámbitos académicos y profesionales.

CAPÍTULO II

2 MEDIDAS DE CENTRALIZACIÓN, DISPERSIÓN Y FORMA

2.1 Medidas de centralización

2.1.1 Media aritmética

Las medidas de tendencia central permiten identificar el punto alrededor del cual se concentran los datos. Entre ellas, la más utilizada es la media aritmética, que corresponde al promedio obtenido al dividir la suma de todos los valores de la variable entre el número de observaciones (Triola , 2020).

Existen diferentes formas de calcular la media, según la naturaleza de los datos:

- Cuando se trabaja con una muestra, la media calculada se denomina estadístico. Se representa como:

$$\bar{X} = \frac{\sum x_i}{n}$$

donde x_i son los valores de la muestra y $n$es el tamaño de la muestra.

- Cuando se trabaja con una población, la media calculada se denomina parámetro, y se representa como:

$$\mu = \frac{\sum X_i}{N}$$

donde X_i son los valores de la población y N es el tamaño total de la población (Levin & Rubin , 2017).

- Cuando los datos están agrupados, se utilizan los puntos medios o marcas de clase (x_i) multiplicados por sus frecuencias (f_i), y se divide la suma entre el total de observaciones:

$$\bar{X} = \frac{\sum f_i x_i}{n}$$

donde n es el número total de datos (Anderson, Sweeney, & Williams , 2016).

Características de la media

La media tiene algunas propiedades importantes que conviene recordar:

- Todo conjunto de datos de nivel de intervalo o de razón posee un valor medio (Moore, McCabe, & Craig , 2017).
- Un conjunto solo puede tener una media (Moore, McCabe, & Craig , 2017).
- Incluye a todos los valores del conjunto en su cálculo (Moore, McCabe, & Craig , 2017).
- Es útil para comparar dos o más grupos de datos (Moore, McCabe, & Craig , 2017).
- Es la única medida de tendencia central en la que la suma de las desviaciones respecto a la media es siempre igual a cero (Moore, McCabe, & Craig , 2017).

- Puede verse afectada por valores extremos o atípicos, que pueden distorsionar la representatividad del promedio (Moore, McCabe, & Craig , 2017).

2.1.2 Media ponderada

En ocasiones, no todos los valores tienen la misma importancia dentro del análisis. Para estos casos se utiliza la media ponderada, que asigna a cada dato un peso o ponderación según su relevancia. La fórmula es:

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i}$$

donde $x_i$son los valores, $w_i$los pesos asignados, y la suma de los pesos se encuentra en el denominador.

Un ejemplo frecuente se encuentra en el ámbito académico, donde las calificaciones de distintas evaluaciones tienen pesos diferentes en la nota final. En el marketing digital, la media ponderada es útil para calcular el rendimiento de una campaña considerando que no todos los canales publicitarios tienen el mismo peso: por ejemplo, las conversiones obtenidas en redes sociales pueden representar el 60% de la estrategia, mientras que las de correo electrónico un 40%.

Como señalan Keller y Warrack (2018), la media ponderada "es especialmente útil cuando se busca dar mayor valor a ciertos elementos del conjunto de datos que son más representativos o relevantes para el análisis" (Keller & Warrack, 2016).

Ejemplo:

El Hotel Hilton registró la venta de 95 habitaciones destinadas a huéspedes selectos, a un costo regular de $400 cada una. Durante la temporada vacacional, el precio se redujo a $200 y se vendieron 126 habitaciones. Finalmente, en la promoción de fin de año, el valor disminuyó a $100 por habitación, logrando venderse 79. Con esta información, determine el precio promedio ponderado por habitación.

Tabla 10. Cálculo de la media ponderada.

Temporada	Precio (x_i)	Habitaciones (w_i)
Normal	400	95
Vacaciones	200	126
Fin de año	100	79
		300

$$X_w = \frac{\sum (f)_i X_i}{\sum (f)_i}$$

$$X_w = \frac{(95*400)+(126*200)+(79*100)}{95+126+79} = \$237$$

2.1.3 Media geométrica

La media geométrica (MG) constituye una medida de tendencia central particularmente apropiada para analizar datos expresados como tasas de crecimiento, porcentajes o rendimientos positivos.

A diferencia de la media aritmética, que suma los valores, la media geométrica los multiplica y luego extrae la raíz enésima según el número de observaciones (Keller & Warrack, 2016).

La fórmula general es:

$$MG = \sqrt[n]{X_1 \times X_2 \times X_3 \times \ldots \times X_n}$$

Un caso ilustrativo se presenta al analizar la variación en la ocupación hotelera durante la temporada baja en Cuenca, donde la media geométrica permite evaluar los cambios porcentuales de manera precisa. Supongamos que las variaciones fueron 5%, 6%, 8% y 2%. La media geométrica sería:

$$MG = \sqrt[4]{(5\% \times 6\% \times 8\% \times 2\%)} = 4.68\%$$

Esto significa que, en promedio, la tasa de ocupación creció un 4,68%, reflejando una medida más realista que la media aritmética para este tipo de datos.

2.1.4 Mediana

La mediana (Me) es el valor que divide un conjunto de datos ordenados en dos partes iguales: el 50% de los datos queda por debajo y el otro 50% por encima. (Moore, McCabe, & Craig , 2017). Para calcularla es necesario ordenar los datos de menor a mayor

Cuando la cantidad de observaciones es impar, la mediana coincide con el valor que ocupa la posición central dentro del conjunto de datos ordenados.

Ejemplo: Salarios semanales de cocina de un hotel → 45, 52, 56, 67, 67. La mediana es 56.

Si el número de observaciones es par, la mediana se obtiene promediando los dos valores centrales. **Ejemplo:** Salarios → 35, 45, 52, 56, 67, 67. La mediana es:

$$Me = \frac{52 + 56}{2} = 54$$

Cuando los datos se presentan agrupados, la mediana se determina aplicando la siguiente fórmula:

$$Me = L_{md} + (\frac{\frac{n}{2} - F}{f_{md}}) \times C_i$$

donde:

- $\boldsymbol{L_{md}}$**:** Representa el límite inferior del intervalo en el que se encuentra ubicada la mediana.
- $\boldsymbol{F}$: Corresponde a la frecuencia acumulada de todas las clases anteriores al intervalo que contiene la mediana.
- $\boldsymbol{f_{md}}$: Indica la frecuencia absoluta de la clase donde se localiza la mediana.
- $\boldsymbol{C_i}$: Hace referencia a la amplitud o ancho del intervalo de clase.

2.1.5 Moda

La **moda (Mo)** es el valor que se presenta con mayor frecuencia dentro de un conjunto de datos. Su uso es especialmente común en variables discretas. En el caso de datos agrupados, la moda corresponde al

intervalo o clase con la frecuencia absoluta más alta. La fórmula para calcularla es:

$$Mo = L_{mo} + (\frac{D_a}{D_a + D_b}) \times C_i$$

donde:

- **L_{mo}:** Corresponde al límite inferior del intervalo en el que se encuentra la moda.
- **D_a:** Expresa la diferencia entre la frecuencia de la clase modal y la frecuencia de la clase inmediata anterior.
- **D_b:** Indica la diferencia entre la frecuencia de la clase modal y la frecuencia de la clase siguiente.
- **C_i:** Representa la amplitud o el ancho del intervalo de clase.

Ejemplo: con datos de pasajeros de aerolíneas clasificados por clases, se identifica la clase modal. La fórmula permite determinar el valor exacto de la moda dentro de esa clase.

Tabla 11. Datos pasajeros.

Número de pasajeros (Clases)	Días frecuencia (f_i)	FA
50-59	3	3
60-69	7	10
70-79	18	28
80-89	12	40
90-99	8	48
100-109	2	50
	50	

Cálculo de la Me:

La posición de la mediana se determina al localizar la clase cuya frecuencia acumulada alcanza o sobrepasa la mitad del número total de observaciones.

FA ≥ n/2 = 25

En este caso, como 28 ≥ 25, la mediana se localiza en la tercera clase.

A continuación, se aplica la fórmula correspondiente:

$$\frac{n}{2} = \frac{50}{2} = 25$$

$$Me = 70 + \left[25 - \frac{10}{18}\right] * 10 = 78.33\ pasajeros$$

Cálculo de la moda:

La clase modal corresponde al intervalo con mayor frecuencia absoluta. En este caso, se identifica en la tercera clase, ya que presenta una frecuencia de 18.

$$Mo = 70 + \left[\frac{18 - 7}{(18 - 12) + (18 - 7)}\right] * 10 = 76.47\ pasajeros$$

2.2 Medidas de posición

Las medidas de posición permiten localizar un valor específico dentro de un conjunto de datos previamente ordenados. Entre las más utilizadas están los cuartiles, deciles y percentiles (Triola , 2020).

2.2.1 Cuartiles (Q)

Los cuartiles dividen el conjunto de datos en cuatro partes iguales, permitiendo identificar la posición relativa de los valores dentro de la distribución.

- $\boldsymbol{Q_1}$**:** Primer cuartil, delimita el 25 % inferior de los datos.
- $\boldsymbol{Q_2}$**:** Segundo cuartil, coincide con la **mediana** y representa el 50 % de los datos acumulados.
- $\boldsymbol{Q_3}$: Tercer cuartil, marca el punto por debajo del cual se encuentra el 75 % de las observaciones.

El rango intercuartílico (RIQ) se obtiene como:

$$RIQ = Q_3 - Q_1$$

y mide la dispersión de la mitad central de los datos.

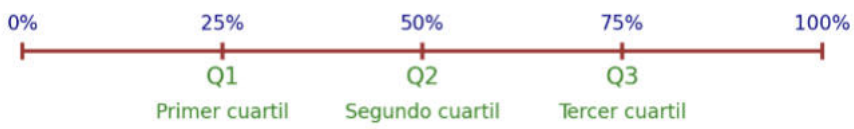

Figura 12. Cuartiles.

El primer cuartil (Q_1) indica el valor por debajo del cual se sitúa el 25 % de los datos, mientras que el tercer cuartil (Q_3) delimita el 25 % superior de la distribución. Entre ambos se concentra el 50 % central de las observaciones. La diferencia entre Q_3 y Q_1 recibe el nombre de rango intercuartílico**,** medida que refleja la dispersión de los datos centrales.

2.2.2 Deciles (D)

Dividen los datos en diez partes iguales, ubicando 9 puntos de referencia.

Figura 13. Deciles.

2.2.3 Percentiles (P):

Los percentiles dividen el conjunto de datos en cien partes iguales**,** estableciendo noventa y nueve puntos de referencia que permiten ubicar cada observación dentro de la distribución.La posición de un percentil en datos no agrupados se calcula con:

$$L_p = \frac{n \times P}{100}$$

- **Lp:** Representa la posición en la que se encuentra el percentil buscado dentro de una serie de datos previamente ordenada.
- **n:** Corresponde al número total de observaciones del conjunto.

- **P:** Indica el percentil específico que se desea calcular.

Gráficamente:

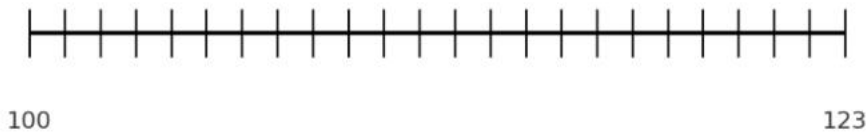

Figura 14. Percentiles.

Ejemplo: Si se registran las horas trabajadas por un mesero en un restaurante durante una semana con los valores (12, 15, 37, 42, 48, 52, 54), el percentil 40 (P_{40}) se localiza entre la tercera y la cuarta posición de la serie ordenada.

$$L_{40} = (8+1)\frac{40}{100} = 3.6$$

Esto significa que el percentil 40 (P40) se localiza entre la tercera y la cuarta posición de los datos ordenados, con una proporción de 60% de distancia entre ambos puntos.

$$P_{40} = 37 + 0.60(42 - 37) = 40$$

Esto indica que el 40% de los datos está por debajo de 40 horas trabajadas.

2.2.4 Cuartiles para datos agrupados en intervalos

Los cuartiles dividen la distribución de los datos en cuatro partes iguales**.** Su cálculo sigue un procedimiento similar al de la mediana, aunque cada cuartil se ubica en distintos puntos del porcentaje acumulado.

- **Primer cuartil (Q1):** Indica el valor por debajo del cual se encuentra el 25 % de los datos**,** marcando el límite del primer cuarto de la distribución.

- **Segundo cuartil (Q2):** Representa el punto que divide la muestra en dos mitades iguales, por lo que coincide con la mediana**.**

- **Tercer cuartil (Q3):** Corresponde al valor que deja por debajo al 75 % de las observaciones**,** delimitando la parte superior de la zona central de la distribución.

Fórmula del Primer Cuartil (Q1):

$$Q_1 = L_{Q_1} + \left(\frac{\frac{N}{4} - F_{Q_1-1}}{f_{Q_1}} \right) \times c$$

Donde los símbolos son los mismos que los utilizados para la mediana, adaptados al primer cuartil.

Interpretación: Los cuartiles permiten evaluar la dispersión y simetría de los datos, proporcionando una visión detallada de la distribución.

2.2.5 Deciles para Datos Agrupados en Intervalos

Los deciles dividen la distribución de los datos en diez partes iguales y se obtienen mediante un procedimiento semejante al utilizado para el cálculo de los cuartiles.

- **Décimo Decil (D1):** Valor que deja el 10% de los datos por debajo.
- **Quinto Decil (D5):** Corresponde a la mediana.
- **Noveno Decil (D9):** Valor que deja el 90% de los datos por debajo.

Fórmula del Primer Decil (D1):

$$D_1 = L_{D_1} + \left(\frac{\frac{N}{10} - F_{D_1 - 1}}{f_{D_1}} \right) \times c$$

Interpretación: Los deciles permiten un análisis más detallado de la posición relativa de los datos dentro de la distribución.

2.2.6 Percentiles para Datos Agrupados en Intervalos

Los percentiles dividen la distribución de los datos en cien secciones de igual tamaño, aplicando un procedimiento análogo al utilizado para el cálculo de cuartiles y deciles.

- **Percentil 25 (P_{25}):** Coincide con el primer cuartil (Q_1), e indica que el 25 % de las observaciones se encuentran por debajo de este valor.

- **Percentil 50 (P_{50}):** Equivale a la mediana o segundo cuartil (Q_2)**,** dividiendo la distribución en dos mitades iguales.

- **Percentil 75 (P_{75}):** Corresponde al tercer cuartil (Q_3)**,** lo que significa que el 75 % de los datos se ubica por debajo de este punto.

Fórmula del Percentil P_k para un valor k:

$$P_k = L_{P_k} + \left(\frac{\frac{kN}{100} - F_{P_k - 1}}{f_{P_k}} \right) \times c$$

Interpretación: los percentiles permiten identificar la posición relativa de un valor dentro de la distribución, facilitando la comparación entre diferentes conjuntos o grupos de datos.

Ejemplo completo

Un investigador ha registrado las edades de los empleados en una empresa, agrupadas en intervalos. La distribución de frecuencias es la siguiente:

Tabla 12. Datos ejercicio.

Intervalos de edades	f	Mc	fi * Mc
20 – 29	5	24.5	122.5
30 – 39	8	34.5	276
40 – 49	12	44.5	534

50 – 59	7	54.5	381.5
60 - 69	3	64.5	193.5
Total	35		1507.5

Cálculo de la Media (X)

Primero, calculamos el punto medio (mim_imi) de cada intervalo:

$$m_1 = \frac{20+29}{2} = 24.5$$
$$m_2 = \frac{30+39}{2} = 34.5$$
$$m_3 = \frac{40+49}{2} = 44.5$$
$$m_4 = \frac{50+59}{2} = 54.5$$
$$m_5 = \frac{60+69}{2} = 64.5$$

A continuación, se multiplica cada punto medio por su frecuencia respectiva y se obtiene la suma total de los productos resultantes.

$$\sum f_i m_i = (5 \times 24.5) + (8 \times 34.5) + (12 \times 44.5) + (7 \times 54.5) + (3 \times 64.5)$$
$$\sum f_i m_i = 122.5 + 276 + 534 + 381.5 + 193.5 = 1507.5$$

Por lo tanto, la media es:

$$\bar{x} = \frac{1507.5}{35} = 43.07 \text{ años}$$

Cálculo de la Mediana

Primero, identificamos el intervalo mediano. La posición de la mediana se encuentra en:

$$\frac{N}{2} = \frac{35}{2} = 17.5$$

La frecuencia acumulada hasta el intervalo de 40 - 49 es 25 (5 + 8 + 12), por lo que la mediana está en este intervalo.

Aplicamos la fórmula de la mediana:

$$\text{Mediana} = L_m + \left(\frac{\frac{N}{2} - F_{m-1}}{f_m} \right) \times c$$

Donde:

- $\mathbf{L_m}$ = 40 (límite inferior del intervalo mediano).
- $\mathbf{F_{m-1}}$ = 13 (frecuencia acumulada antes del intervalo mediano).
- $\mathbf{f_m}$ = 12 (frecuencia del intervalo mediano).
- **C** = 10 (amplitud del intervalo).

Sustituyendo los valores:

$$\text{Mediana} = 40 + \left(\frac{17.5 - 13}{12} \right) \times 10 = 40 + \left(\frac{4.5}{12} \right) \times 10 = 40 + 3.75 = 43.75 \text{ años}$$

Cálculo de la Moda

Identificamos el intervalo modal, que es el de 40-49, ya que tiene la mayor frecuencia ($f_m = 12$)

Aplicamos la fórmula de la moda:

$$\text{Moda} = L_m + \left(\frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \right) \times c$$

Donde:

- $\mathbf{L_m} = 40$
- $\mathbf{f_{m-1}} = 8$ (frecuencia del intervalo anterior).
- $\mathbf{f_{m+1}} = 7$ (frecuencia del intervalo siguiente).
- $\mathbf{c} = 10$

Sustituyendo los valores:

$$\text{Moda} = 40 + \left(\frac{12-8}{(12-8)+(12-7)}\right) * 10 = 40 + \left(\frac{4}{4+5}\right) * 10 = 44.44$$

La posición de los percentiles en distribuciones de datos agrupados se determina utilizando la siguiente expresión matemática:

$$P = Lp_{i-1} + \left[\frac{\%N - FA}{fp_i}\right] * c_i$$

- **P:** Representa el percentil que se desea calcular.

- $\boldsymbol{L_{p-1}}$**:** Corresponde al límite inferior del intervalo de clase en el cual se encuentra ubicado el percentil buscado.

- $(\boldsymbol{P/100})\ \boldsymbol{N}$**:** Indica la posición teórica del percentil dentro del total de observaciones N.

- **FA:** Expresa la frecuencia acumulada de todas las clases anteriores a aquella que contiene el percentil.

- $\boldsymbol{f_p}$**:** Es la frecuencia absoluta de la clase en la que se localiza el percentil.

- $\boldsymbol{c_i}$**:** Hace referencia a la amplitud o ancho del intervalo de clase.

Antes de calcular

- Ordena la tabla de frecuencias y obtén N (suma de todas las frecuencias).

- Calcula la posición: $L_{\text{pos}} = (P/100)\ N$.

- Identifica la clase percentílica: es la primera cuyo FA $\geq L_{\text{pos}}$.

- Anota L_p, FA, f_p y c_i de esa clase.

Estimación por interpolación lineal

- El valor del percentil se obtiene "avanzando" dentro de la clase percentílica una fracción del ancho c_i, proporcional a cuánto falta para alcanzar la posición teórica:

$$\text{Percentil } P \approx L_p + \frac{((P/100)N - FA_{\text{prev}})}{f_p} \times c_i$$

Ejemplo

Se dispone de una tabla de frecuencias correspondiente al número de turistas que visitaron Galápagos durante el último verano. Con base en dicha distribución, calcule los percentiles **P_{40}** y **P_{75}** siguiendo el procedimiento indicado.

Tabla 13. Visitantes a las Islas Galápagos.

Turistas visitantes	Número/frecuencia	FA
0-100	90	90
100-200	140	230
200-300	150	380
300-400	120	500
	N = 500	

Supongamos que deseamos encontrar el percentil 25 (P25), equivalente al primer cuartil (Q1).

1. **Determinación de la posición:** Se calcula la posición teórica del percentil mediante la expresión:

$$(25/100) \times 500 = 125$$

Esto indica que el P25 se localiza en la posición 125 dentro del conjunto ordenado.

2. **Identificación de la clase percentílica:** Al analizar la columna de frecuencias acumuladas, se identifica que el primer valor que alcanza o supera los 125 corresponde a 230**,** por lo que el percentil se ubica dentro de la segunda clase**.**

3. **Cálculo del percentil:** Sustituyendo los valores en la fórmula:

$$P_{25} = 100 + (\frac{125 - 90}{140}) \times 100 = 125$$

De esta manera, el P25 ≈ 125, lo cual significa que el 25% de los datos se ubica por debajo de este valor.

Ahora, si queremos determinar el decil 4, recordemos que este equivale al percentil 40, es decir, D4 = P40.

1. **Posición del percentil 40:**

$$(40/100) \times 500 = 200$$

Por tanto, la posición correspondiente es la **200**.

2. **Ubicación en la distribución:** La frecuencia acumulada más cercana que supera 200 es nuevamente 230, por lo que el valor buscado se localiza en la segunda clase.

3. **Cálculo del P40 (o D4):** Sustituyendo en la fórmula:

$$P_{40} = 100 + (\frac{200 - 90}{140}) \times 100 = 178.57$$

Esto indica que el 40% de las observaciones se sitúa por debajo de aproximadamente 178.57.

2.3 Medidas de dispersión

Las medidas de tendencia central, como la media, la mediana y la moda, permiten identificar el punto alrededor del cual se agrupan los datos. Sin embargo, por sí solas no muestran el grado de dispersión o concentración de los valores respecto a ese centro. Por eso, las medidas de dispersión resultan indispensables, ya que muestran la variabilidad de los datos y permiten evaluar la estabilidad de un fenómeno (Levin & Rubin , 2017).

En el marketing digital, la dispersión de los datos puede significar la diferencia entre una campaña predecible y una con resultados caóticos. Por ejemplo, si la media de visitas a un sitio web es 1.000 al día, la dispersión nos dirá si esas visitas se concentran cerca de ese valor (900, 950, 1.050) o si existen fluctuaciones grandes (200, 500, 2.000).

2.3.1 Rango

El rango (R) constituye la medida más simple para evaluar la dispersión de un conjunto de datos. Se calcula como la diferencia entre el valor más alto y el valor más bajo observados.

$$R = X_{máx} - X_{mín}$$

Aunque es fácil de calcular y de interpretar, su principal limitación es que depende únicamente de dos valores extremos, lo que puede generar distorsiones si existen outliers.

Ejemplo: Si los clics en una campaña de anuncios oscilaron entre 200 y 1.500, el rango sería:

$$R = 1.500 - 200 = 1.300$$

2.3.2 Varianza

La varianza (σ^2 o s^2) expresa el grado promedio de dispersión de los datos con respecto a la media. Para ello, se eleva al cuadrado cada desviación individual, evitando así que las diferencias positivas y negativas se compensen.

Para una muestra:

$$s^2 = \frac{\sum(x_i - \bar{X})^2}{n - 1}$$

Para una población:

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$$

La varianza resulta útil para el análisis estadístico, aunque su interpretación práctica es limitada porque se expresa en unidades al cuadrado.

2.3.3 Desviación estándar

La desviación estándar (σ o s) se obtiene como la raíz cuadrada de la varianza. Representa el grado promedio de dispersión de los datos respecto a la media y constituye una de las medidas más empleadas en el análisis estadístico. Tiene la ventaja de expresarse en las mismas unidades que los datos originales, lo que la hace más intuitiva (Moore, McCabe, & Craig , 2017).

La desviación estándar, aplicada tanto a datos agrupados como no agrupados, cuantifica el grado de dispersión de los valores con respecto a la media, mostrando qué tan alejados se encuentran los datos del promedio general. En el caso poblacional, se denota por δ y se calcula aplicando la fórmula correspondiente.

Desviación estándar poblacional (δ):

La desviación estándar en el contexto poblacional se simboliza con la letra griega δ. Esta medida expresa cuán dispersos están los valores de una población con respecto a su media. Desde el punto de vista matemático, se calcula como la raíz cuadrada del promedio de las desviaciones al cuadrado de cada valor respecto a la media poblacional.

$$\delta = \sqrt{\left[\frac{\sum (X_i - \mu)^2}{N}\right]}$$

Desviación muestral (s):

Cuando no se cuenta con la información de toda la población y solo se dispone de una muestra, la medida utilizada para representar la dispersión de los datos es la desviación estándar muestral, simbolizada con la letra *s*. Su cálculo sigue un procedimiento similar al de la población, aunque en este caso se divide entre *(n − 1)* en lugar de *N*, con el fin de corregir el sesgo en la estimación.

$$S = \sqrt{\left[\frac{\sum (X_i - \bar{X})^2}{n-1}\right]}$$

Cuando los datos se presentan agrupados, la varianza muestral se determina tomando en cuenta las frecuencias de cada clase o intervalo. Luego, al calcular la raíz cuadrada de dicha varianza, se obtiene la desviación estándar, que refleja de manera clara la variabilidad de los valores en torno al promedio.

$$S^2 = \frac{\sum f Xi^2 - n\bar{X}^2}{n-1}$$

Por lo tanto:

$$S = \sqrt{\frac{\sum f Xi^2 - n\bar{X}^2}{n-1}}$$

2.3.4 Coeficiente de Variación

El coeficiente de variación (CV) representa una medida relativa de dispersión. Se calcula dividiendo la desviación estándar entre la media aritmética y multiplicando el resultado por 100, con el fin de expresar la variabilidad en términos porcentuales.

$$CV = \frac{s}{\bar{X}} \times 100$$

El coeficiente de variación es especialmente valioso para comparar la dispersión relativa entre distintos conjuntos de datos, incluso cuando se expresan en unidades o escalas de medida diferentes.

Ejemplo: Si una campaña en Facebook obtiene un CV del 10% y otra en Instagram un CV del 25%, esto indica que los resultados de la primera son más consistentes y estables, mientras que en la segunda existe mayor variabilidad.

Ejercicios completos

1. Se cuenta con el siguiente conjunto de datos correspondiente a las horas trabajadas por semana durante los dos últimos meses: 52, 48, 37, 54, 48, 15, 42 y 12. A partir de esta información, calcule el rango, la varianza, la desviación estándar y el coeficiente de variación.

Tabla 14. Datos.

Observaciones	X_i (horas trabajadas)	$(X_i - \bar{x})^2$
1	12	702.25
2	15	552.25
3	37	2.25
4	42	14.0625
5	48	95.0625
6	48	95.0625
7	52	182.25
8	54	240.25
n = 8	Σxi = 308	Σ=1.883.7375

Solución:

$$\overline{x} = \frac{\sum fx_i}{n} = \frac{393.5}{50} = 78.7$$

$$S^2 = \frac{316902.5 - 50*(78.7)^2}{49} = 174.31$$

$$S = 12.14$$

2. Durante el mes más reciente, los pasajeros de la aerolínea A&A fueron clasificados en distintas categorías de acuerdo con su tipo de viaje. Con base en estos datos agrupados, calcule la media, la varianza y la desviación estándar.

Tabla 15. Datos.

Número de pasajeros	fi (días)	Xi (Punto medio)	f*Xi	Xi²	f*Xi²
50-59	3	54.5	163.5	2970.25	8910.75
60-69	7	64.5	451.5	4160.25	29121.75
70-79	18	74.5	1341.0	5550.25	99904.5
80-89	12	84.5	1014.0	7140.25	85683.0
90-99	8	94.5	756.0	8930.25	71442.0
100-109	2	104.5	209.0	10920.25	21840.5
	N = 50		Σ=393.5		Σ=316902.5

$$\overline{x} = \frac{\sum f x_i}{n} = \frac{393.5}{50} = 78.7$$

$$S^2 = \frac{316902.5 - 50 * (78.7)^2}{49} = 174.31$$

$$S = 12.14$$

Las medidas de dispersión permiten a los analistas evaluar el riesgo y la estabilidad de sus decisiones. En el mundo empresarial y digital, no basta con conocer el promedio: es fundamental identificar si los datos se comportan de manera homogénea o si presentan variaciones extremas que puedan afectar los resultados de una estrategia (Anderson, Sweeney, & Williams , 2016).

CAPÍTULO III

3 MUESTREO, CONFIABILIDAD Y PROBABILIDAD

La estadística no solo se ocupa de organizar y describir los datos, sino también de inferir conclusiones sobre una población a partir del estudio de una muestra, determinando además el nivel de confianza asociado a dichas inferencias. Este capítulo introduce al estudiante en tres grandes áreas: el muestreo, la confiabilidad de las estimaciones y la probabilidad como fundamento de la inferencia estadística.

Como señalan Walpole, Myers, Myers y Ye (2012), el análisis estadístico moderno se apoya en la probabilidad y en la teoría del muestreo, ya que "no es posible observar a toda la población, pero sí se puede obtener información confiable mediante un subconjunto representativo" (Walpole , Myers, Myers, & Ye, 2012).

En el contexto empresarial, no es necesario evaluar a todos los clientes para conocer su nivel de satisfacción; basta con analizar una muestra representativa que refleje las percepciones generales. Con una muestra representativa es posible obtener conclusiones válidas y generalizables. De igual manera, la probabilidad permite prever situaciones como la posibilidad de que un cliente vuelva a comprar o que una campaña publicitaria alcance cierto nivel de interacción.

3.1 Muestreo

3.1.1 Concepto de población y muestra

En estadística, se denomina *población* al conjunto completo de individuos, elementos u observaciones que comparten una o más características que son objeto de estudio. Puede estar conformada por personas, empresas, productos, visitas a una página web o cualquier unidad que se desee analizar. En la práctica, suele ser demasiado grande o difícil de estudiar en su totalidad (Triola , 2020).

Una muestra constituye un subconjunto de la población que se selecciona de forma representativa, con el propósito de obtener información que permita realizar inferencias o conclusiones válidas sobre el conjunto total. Como explican Walpole, Myers, Myers y Ye (2012), la clave del muestreo está en que "los resultados obtenidos de una muestra bien diseñada pueden generalizarse a toda la población, siempre que se cumplan ciertos supuestos de aleatoriedad y representatividad" (Walpole , Myers, Myers, & Ye, 2012).

En el marketing digital, la diferencia entre población y muestra es evidente. Una empresa de comercio electrónico puede considerar como población a los 50.000 clientes que han comprado en su tienda durante un año. Sin embargo, analizar a todos sería costoso e innecesario. En lugar de eso, puede seleccionar una muestra de 500 clientes y, a partir de ella, estudiar patrones de consumo, niveles de satisfacción o intención de recompra.

El proceso de muestreo permite optimizar recursos, reducir el tiempo de recolección y análisis de datos, y obtener al mismo tiempo información confiable que respalde la toma de decisiones estratégicas.

3.1.2 Tipos de muestreo probabilístico

El muestreo probabilístico se caracteriza porque cada elemento que forma parte de la población posee una probabilidad conocida y diferente de cero de ser incluido en la muestra. Este tipo de muestreo se considera más riguroso y confiable porque evita sesgos y permite generalizar los resultados con un margen de error calculable (Levin & Rubin , 2017).

A continuación, se presentan los principales tipos:

a) Muestreo aleatorio simple

Este método, considerado el más elemental, consiste en seleccionar de manera completamente aleatoria y con igual probabilidad a los elementos que integran la población, sin que un mismo individuo pueda ser elegido más de una vez.

Ejemplo: Una agencia de publicidad quiere conocer la satisfacción de sus clientes. Tiene 2.000 registros en su base de datos y selecciona al azar 200 de ellos para aplicar una encuesta.

b) Muestreo estratificado

Se aplica cuando la población es heterogénea y puede dividirse en estratos o subgrupos con características comunes (edad, género, nivel de ingresos). De cada estrato se extrae una muestra proporcional o

equitativa.

Ejemplo: Un e-commerce desea analizar hábitos de compra. Divide su base de clientes en tres estratos: jóvenes (18–25), adultos (26–45) y mayores de 45 años, y selecciona proporcionalmente casos de cada grupo.

c) Muestreo sistemático

Consiste en elegir a los individuos a intervalos regulares, a partir de un punto de inicio seleccionado al azar.

Ejemplo: Una empresa quiere analizar la experiencia de usuarios en su página web. Decide revisar a cada décimo visitante registrado en su servidor hasta completar la muestra.

d) Muestreo por conglomerados

Se aplica principalmente cuando la población es muy numerosa o se encuentra ampliamente distribuida en diferentes zonas geográficas. En lugar de seleccionar individuos, se eligen conglomerados (grupos naturales) y luego se estudian todos los elementos dentro de esos grupos o se selecciona una muestra interna.

Ejemplo: Una marca de delivery desea conocer la satisfacción de restaurantes aliados. En vez de encuestar a todos los locales de la ciudad, selecciona 5 barrios (conglomerados) y entrevista a todos los restaurantes de esas zonas.

En el ámbito del marketing digital, los muestreos probabilísticos permiten obtener información fiable sobre clientes, campañas y

mercados sin necesidad de analizar a toda la población. Como indican Anderson, Sweeney y Williams (2016), el muestreo es "una herramienta clave porque equilibra la precisión estadística con la eficiencia en costos y tiempos" (Anderson, Sweeney, & Williams , 2016).

3.1.3 Tipos de muestreo no probabilísticos

El muestreo no probabilístico se caracteriza porque los elementos de la población no poseen la misma probabilidad de ser seleccionados, ya que la elección depende del criterio del investigador o de circunstancias específicas del estudio. En este tipo de diseño, la elección depende de criterios del investigador o de la accesibilidad de los datos. Aunque no ofrece la misma precisión estadística que el muestreo probabilístico, resulta muy utilizado en estudios exploratorios, donde se busca rapidez y bajos costos (Malhorta, 2019).

Dentro de esta categoría encontramos los siguientes métodos:

- **Muestreo por conveniencia:** Se seleccionan los elementos que están más a mano o disponibles. Aunque puede introducir sesgos, es útil en investigaciones preliminares.

Ejemplo: una empresa de marketing digital aplica encuestas solo a los clientes que visitan su página en una semana específica, por ser más fácil contactarlos.

- **Muestreo intencional o por juicio:** En este tipo de muestreo, el investigador selecciona de manera deliberada a las personas o

elementos que, según su criterio, representan mejor las características de la población en estudio.

Ejemplo: un community manager selecciona a los 20 usuarios más activos de una red social para conocer sus percepciones sobre una nueva campaña.

- **Muestreo por bola de nieve:** Se aplica principalmente en poblaciones de difícil acceso o identificación. El proceso inicia con un grupo reducido de participantes que, posteriormente, recomiendan o contactan a otras personas con características similares.

Ejemplo: Un estudio sobre creadores de contenido en TikTok empieza con un grupo inicial, que luego recomienda a otros influencers para participar en la investigación.

Aunque el muestreo no probabilístico limita la generalización de resultados, su aplicación en áreas como el marketing digital permite obtener información rápida para orientar decisiones estratégicas, sobre todo cuando se busca explorar un fenómeno emergente o segmentar nichos de mercado (Hair, Wolfinbarger, Money, Samouel, & Page, 2015).

3.2 Cálculo de la muestra y ejercicios

El cálculo del tamaño muestral constituye una etapa fundamental en el diseño de una investigación estadística, pues determina la cantidad de individuos o unidades que deben incluirse para garantizar resultados

válidos y representativos de la población. Un tamaño de muestra adecuado permite realizar inferencias precisas sobre la población, minimizando tanto los costos como los errores.

Uno de los aspectos más importantes en cualquier investigación estadística es determinar cuántos elementos deben incluirse en la muestra. Si el número es demasiado pequeño, los resultados podrían no reflejar la realidad de la población; si es muy grande, se incrementan los costos y el tiempo de recolección de datos sin necesidad (Walpole , Myers, Myers, & Ye, 2012).

3.2.1 Elementos determinantes en la estimación del tamaño muestral

El nivel de confianza (Z) expresa el grado de certeza con el que se espera que los resultados obtenidos a partir de una muestra reflejen con precisión las características de la población. Los valores más utilizados en investigación son 90%, 95% y 99%, cuyos valores críticos aproximados son 1.645, 1.96 y 2.576, respectivamente.

- **Margen de Error (E):** Es la precisión requerida para las estimaciones. Un margen de error pequeño implica un mayor tamaño de muestra.
- **Variabilidad o Proporción Estimada (p): La** variabilidad dentro de la población, a menudo expresada como una proporción o desviación estándar. Si no se conoce, se asume p = 0.5 para máxima variabilidad.

- **Tamaño de la Población (N):** Es relevante en muestras muy grandes o cuando la población es pequeña y finita.
- **Efecto del Diseño:** Ajuste adicional cuando se utilizan diseños complejos como el muestreo estratificado o por conglomerados.

3.2.2 Métodos de estimación del tamaño de una muestra

Las fórmulas utilizadas para determinar el tamaño de la muestra varían según la naturaleza del estudio y el tipo de dato que se analice, ya sea una proporción o una media.

Tamaño de la muestra para proporciones (Muestreo Aleatorio Simple)

Cuando se estima una proporción en la población, la fórmula es la siguiente:

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{E^2}$$

Donde:

- **n:** Representa el tamaño de la muestra que se requiere seleccionar.
- **Z:** Corresponde al valor crítico de la distribución normal, determinado por el nivel de confianza elegido.
- **p:** Indica la proporción estimada del rasgo o característica de interés; si no se dispone de información previa, se recomienda usar $p = 0,5$ para reflejar la máxima variabilidad.

- **E:** Simboliza el margen máximo de error que se acepta al realizar la estimación.

Ejemplo

Si deseas estimar la proporción de estudiantes que prefieren estudiar por la mañana con un 95% de confianza y un margen de error de 5%, ¿cuántos estudiantes debes encuestar?

- **Nivel de confianza:** 95% (valor crítico Z = 1.96).
- **Margen de error:** 5% (E = 0.05).
- **Proporción esperada:** 0.5, utilizada para reflejar el escenario de máxima variabilidad cuando no se dispone de información previa.

$$n = \frac{1.96^2 \cdot 0.5 \cdot (1 - 0.5)}{0.05^2} = \frac{3.8416 \cdot 0.25}{0.0025} = 384.16$$

Por lo tanto, se necesita encuestar al menos 385 estudiantes (redondeando al siguiente número entero).

Tamaño de la Muestra para Medias (Muestreo Aleatorio Simple)

Cuando se estima la media de una población, la fórmula es:

$$n = \frac{Z^2 \cdot \sigma^2}{E^2}$$

Donde:

- **n:** número de observaciones que conformarán la muestra.
- **Z:** valor crítico asociado al nivel de confianza seleccionado.
- **σ:** estimación de la desviación estándar poblacional.
- **E:** margen de error máximo aceptado en la estimación.

Ejemplo: Una empresa desea estimar el tiempo promedio de producción de un artículo con un nivel de confianza del 95% y un margen de error de 2 minutos**.** Se sabe que la desviación estándar estimada (σ) del tiempo de producción es de 10 minutos**.**

Se requiere determinar el tamaño mínimo de la muestra necesaria.

- Z = 1.96 (95% de confianza)
- E = 2 minutos
- σ = 10 minutos

$$n = \frac{1.96^2 \cdot 10^2}{2^2} = \frac{3.8416 \cdot 100}{4} = 96.04$$

Por lo tanto, se debe incluir al menos 97 mediciones.

Ajuste para poblaciones finitas

Si la población es finita y conocida, se ajusta el tamaño de la muestra con la fórmula:

$$n_{ajustado} = \frac{n}{1 + \frac{n-1}{N}}$$

Donde:

- $n_{ajustado}$ = Tamaño de la muestra ajustado.
- n = Tamaño de la muestra calculado sin ajuste.
- N = Tamaño de la población.

Ejemplo: Si en el ejemplo anterior la población total de empleados es de 300, ajusta el tamaño de la muestra calculado para la población finita.

$$n_{ajustado} = \frac{97}{1 + \frac{97-1}{300}} = \frac{97}{1 + \frac{96}{300}} = \frac{97}{1.32} \approx 73.48$$

Por lo tanto, se necesita una muestra de al menos 74 empleados.

En síntesis, el tamaño de muestra no debe seleccionarse de manera arbitraria, sino considerando criterios estadísticos y el contexto del estudio. En marketing digital, calcular un tamaño adecuado garantiza que los resultados de las encuestas o estudios de mercado puedan respaldar decisiones estratégicas con un nivel alto de confiabilidad (Levin & Rubin , 2017).

3.3 Confiabilidad de las estimaciones

Cuando trabajamos con una muestra, los resultados que obtenemos no son idénticos a los que se obtendrían si analizáramos a toda la población. Siempre existirá un margen de error. Por eso, en estadística se habla de

confiabilidad de las estimaciones, es decir, del grado de certeza que podemos tener al usar una muestra para generalizar conclusiones (Newbold, Carlson, & Thorme, 2013).

En términos simples, se trata de responder a dos preguntas:

- ¿Qué tan cerca están los resultados de la muestra de los verdaderos valores de la población?
- ¿Con qué grado de seguridad podemos afirmarlo?

Los intervalos de confianza son la herramienta que nos permite dar respuesta.

3.4 Intervalos de confianza

Un intervalo de confianza es un rango estimado de valores dentro del cual, con una probabilidad determinada, se espera que se ubique el verdadero parámetro de la población (Anderson, Sweeney, Camm, & Cochran , 2019). El nivel de confianza más común es el 95%, lo que significa que, si repitiéramos el estudio muchas veces, en 95 de cada 100 ocasiones el parámetro real estaría dentro del intervalo (Anderson, Sweeney, Camm, & Cochran , 2019).

Ejercicio 1: Una plataforma de e-commerce quiere estimar la proporción de clientes satisfechos. Se encuestan $n = 400$usuarios y 320 declaran estar satisfechos.

Datos:
$\hat{p} = 320/400 = 0.80$; $n = 400$; nivel de confianza 95% $\Rightarrow Z_{0.975} = 1.96$.

Error estándar (EE) de la proporción

$$EE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.80 \times 0.20}{400}} = \sqrt{\frac{0.16}{400}} = \sqrt{0.0004} = 0.02$$

Margen de error (E)

$$E = Z \times EE = 1.96 \times 0.02 = 0.0392 \approx 0.039$$

Intervalo de confianza (IC 95%)

$$\hat{p} \pm E = 0.80 \pm 0.039 \Rightarrow [0.761,\ 0.839]$$

Interpretación: Con 95% de confianza, entre 76.1% y 83.9% de los clientes están satisfechos.

Ejercicio 2: Se desea estimar el tiempo medio de permanencia (en minutos) de los usuarios en el sitio. Se toma una muestra de $n =$ 25sesiones, con media muestral $\bar{X} = 8.4$y desviación estándar muestral $s = 2.5$.

Datos:
$\bar{X} = 8.4$; $s = 2.5$; $n = 25$; gl $= 24$; 95% $\Rightarrow t_{0.975,24} \approx 2.064$.

Error estándar de la media (EEx)

$$EE_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{2.5}{\sqrt{25}} = \frac{2.5}{5} = 0.5$$

Margen de error (E)

$$E = t \times EE_{\bar{X}} = 2.064 \times 0.5 = 1.032$$

Intervalo de confianza (IC 95%)

$$\bar{X} \pm E = 8.4 \pm 1.032 \Rightarrow [7.368,\ 9.432]$$

Interpretación: Con 95% de confianza, el tiempo medio de permanencia está entre 7.37 y 9.43 minutos.

3.5 Nivel de confianza y error muestral

El nivel de confianza representa la probabilidad de que el intervalo calculado a partir de una muestra contenga el valor real del parámetro de la población. A medida que aumenta este nivel de confianza, el intervalo tiende a ampliarse. Usualmente se trabaja con 90%, 95% o 99% (Hernández Sampieri , Fernández, & Baptista, 2014).

El error muestral representa la diferencia máxima admisible entre el valor estimado a partir de la muestra y el valor verdadero del parámetro poblacional. Un error del 5% significa que, si la muestra reporta un 60% de satisfacción, el valor real podría oscilar entre 55% y 65%.

En el ámbito empresarial y digital, comprender el margen de error es vital. No es lo mismo decir “el 60% de los clientes está satisfecho” que afirmar “el nivel de satisfacción está entre 55% y 65% con un 95% de confianza”. La segunda forma comunica transparencia y rigor, y facilita la toma de decisiones responsables (Gujarati & Porter, 2010).

En definitiva, los intervalos de confianza y el error muestral permiten entender la incertidumbre de nuestras conclusiones. En marketing digital, esto se traduce en diseñar campañas basadas en estimaciones sólidas y realistas. No se trata de adivinar, sino de cuantificar el grado de certeza con el que hablamos de los clientes y sus comportamientos.

3.6 Probabilidad

La probabilidad es el lenguaje matemático que nos permite medir la incertidumbre. En palabras simples, indica qué tan posible es que ocurra un evento dentro de un conjunto de resultados posibles (Ross, 2014). En estadística, constituye la base de la inferencia, ya que gracias a la probabilidad podemos pasar de los datos observados en una muestra a conclusiones sobre toda la población.

En el mundo del marketing digital, hablar de probabilidad es cotidiano. Una empresa puede preguntarse: ¿cuál es la probabilidad de que un cliente que ya compró una vez vuelva a hacerlo? o ¿qué probabilidad hay de que un anuncio sea visto por más de mil personas en una semana? Estas preguntas se responden aplicando reglas de probabilidad.

3.7 Concepto y reglas básicas de la probabilidad

En un experimento aleatorio, el espacio muestral (S) corresponde al conjunto de todos los resultados que pueden obtenerse. Un evento, por su parte, es cualquier grupo de resultados contenidos en ese espacio. La probabilidad de un evento se calcula como:

$$P(A) = \frac{\text{número de casos favorables}}{\text{número total de casos posibles}}$$

si todos los resultados son igualmente probables (Rice, 2007).

Ejemplo: Supongamos que una tienda online registra 1.000 visitas en un día y 250 terminan en una compra. La probabilidad de que un visitante compre es:

$$P(\text{Compra}) = \frac{250}{1000} = 0.25$$

Es decir, 25%.

Reglas básicas

- $0 \leq P(A) \leq 1$. La probabilidad nunca puede ser negativa ni superar el 100%.
- $P(S) = 1$. El espacio muestral tiene probabilidad 1 porque siempre ocurre "algo".
- Para eventos mutuamente excluyentes:

$$P(A \cup B) = P(A) + P(B)$$

- Para el complemento de un evento:

$$P(A^c) = 1 - P(A)$$

3.8 Probabilidad condicional e independencia

La probabilidad condicional se define como la probabilidad de que ocurra un evento A, dado que previamente ha tenido lugar otro evento B.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Esto es clave en marketing digital. Por ejemplo, si de 500 clientes que visitan un sitio web, 200 añaden un producto al carrito y de esos 200, 50 concretan la compra:

$$P(\text{Compra} \mid \text{Carrito}) = \frac{50}{200} = 0.25$$

Es decir, un 25% de quienes agregaron productos finalmente compraron.

Cuando dos eventos son independientes, significa que el hecho de que ocurra uno no afecta la probabilidad del otro:

$$P(A \cap B) = P(A) \times P(B)$$

3.9 Teorema de Bayes

El teorema de Bayes permite actualizar probabilidades cuando aparece nueva información. Se expresa como:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Ejemplo en marketing: Una empresa sabe que el 40% de sus clientes llega por anuncios en redes sociales (evento A) y el 60% por búsqueda orgánica (evento A'). La probabilidad de compra es 20% para quienes vienen de anuncios y 10% para quienes llegan por orgánico. Si un cliente compró, ¿cuál es la probabilidad de que haya llegado por un anuncio?

- Calcular P(B):

$$P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$$
$$= (0.20)(0.40) + (0.10)(0.60) = 0.08 + 0.06 = 0.14$$

- Aplicar Bayes:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{0.20 \times 0.40}{0.14} = \frac{0.08}{0.14} \approx 0.571$$

3.9.1 Distribución probabilística binomial

Una distribución se considera binomial cuando un experimento se realiza varias veces bajo las mismas condiciones, manteniendo constante la probabilidad de éxito en cada repetición. Este proceso, descrito en los ensayos de Bernoulli, cumple con las siguientes condiciones:

1. Solo existen dos resultados posibles: éxito (π) y fracaso ($1 - \pi$).

2. La probabilidad de éxito y de fracaso se mantiene invariable en cada repetición del experimento.

3. Cada ensayo es independiente de los demás.

4. El experimento puede realizarse un número finito de veces.

Si se conoce la probabilidad de éxito en un único ensayo, se puede calcular el número esperado de éxitos en un conjunto de *n* repeticiones mediante la aplicación de la fórmula de la distribución binomial.

$$P(x)= \frac{n!}{x!(n-x)!\,\pi(1-\pi)^{n-x}}$$

O también puede calcularse como:

$$P(x) = {}_nC_x\,(\pi)^x(1-\pi)^{n-x}$$

Los valores de P(x), calculados para diferentes combinaciones de los parámetros π, n y x, suelen encontrarse organizados en tablas estadísticas que se presentan habitualmente en los apéndices de libros o manuales especializados en estadística.

Ejemplo:

En este caso, se busca determinar la probabilidad de que, dentro de un grupo de 20 pasajes aéreos seleccionados aleatoriamente, exactamente cinco no hayan sido pagados en su totalidad. La aerolínea ha identificado que, en promedio, el 10 % de los pasajeros no cancela el valor completo del boleto durante un mes determinado. Desde el enfoque de la estadística inferencial**,** esta situación puede representarse mediante una distribución binomial**,** donde cada pasaje se considera un ensayo

independiente con dos posibles resultados: el pago completo o el pago incompleto del boleto.

$$\pi = 10\%;\ n = 20\ y\ x = 5, \text{es igual a } 0.0319 \text{ o } 3.19\%.$$

$$P(x) = {}_{20}C_5\ (10\%)^5\ (1-\ 10\%)^{20-5}$$

$$P(x) = 3.19\%$$

Dado que existe una probabilidad del 10% de que un pasaje no sea cancelado por completo, se puede concluir que la probabilidad de que exactamente cinco de los veinte pasajes elegidos al azar presenten pago completo es del 3,19%**.**

a. Media y varianza de las distribuciones binomiales

En una distribución binomial, los resultados posibles se reducen a dos escenarios: éxito o fracaso. A partir de esta estructura, tanto la media como la varianza pueden obtenerse mediante fórmulas específicas que vinculan la probabilidad de éxito con el número total de ensayos realizados.

$$E(X) = \mu = n\ \pi$$

$$\sigma 2\ = n\ \pi(1-\pi)$$

3.9.2 Distribución hipergeométrica

Este tipo de distribución se presenta cuando la probabilidad de éxito cambia con cada selección, situación que suele darse en poblaciones

reducidas donde el muestreo se realiza sin reemplazo. En estos casos, la posibilidad de obtener un determinado resultado varía progresivamente a medida que se extraen los elementos del conjunto.

La función de probabilidad de la distribución hipergeométrica se utiliza para calcular la probabilidad de obtener exactamente x éxitos al seleccionar una muestra de tamaño n, tomada sin reemplazo de una población de N elementos que contiene K éxitos.

$$P(x) = \frac{rCx \ N\text{-}rCn\text{-}x}{NCn}$$

En donde:

- **N:** tamaño total de la población.
- **r:** número total de éxitos presentes en la población.
- **n:** tamaño de la muestra seleccionada.
- **x:** número de éxitos observados en la muestra.

La distribución hipergeométrica resulta apropiada cuando se extrae una muestra sin reemplazo de una población finita y conocida, especialmente si la muestra representa una fracción considerable de dicha población.

Ejemplo:

En un grupo de 15 docentes del Programa de Turismo y Gastronomía, 8 ya cuentan con formación previa en un nuevo enfoque de turismo. Se seleccionarán 12 docentes sin reemplazo para una estancia de estudio en Japón.

¿Cuál es la probabilidad de que exactamente 5 de los seleccionados tengan esa formación previa?

Modelo y cálculo

Usamos la distribución hipergeométrica**:**

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Con $N = 15$, $r = 8$, $n = 12$, $x = 5$:

$$P(X = 5) = \frac{\binom{8}{5}\binom{7}{7}}{\binom{15}{12}} = \frac{56 \cdot 1}{455} = \frac{8}{65} \approx 0.1231$$

La probabilidad de que exactamente cinco de los doce docentes seleccionados ya tengan conocimiento del enfoque es 8/65 ≈ 12.31%

3.9.3 Distribución probabilística de Poisson

Cuando las probabilidades de éxito son moderadas y el tamaño de la muestra es pequeño, la distribución binomial ofrece resultados adecuados. No obstante, si la probabilidad de éxito (π) es muy baja y el

número de observaciones (n) es considerablemente alto, se recurre a la distribución de Poisson, también llamada ley de los eventos poco frecuentes**,** la cual permite modelar la ocurrencia de sucesos que se presentan de manera aislada o esporádica. Esta distribución discreta se aplica al conteo de eventos cuya ocurrencia es poco probable en un intervalo definido.

Entre sus usos más frecuentes se encuentran los servicios, como el número de clientes que esperan atención en un restaurante o la cantidad de personas en fila para ingresar a un centro recreativo.

$$P(x) = \frac{\mu^{x} \; e^{-\mu}}{X!}$$

Donde:

- **μ (mu):** Corresponde al promedio o media aritmética de ocurrencias esperadas en un intervalo de tiempo o espacio determinado.
- **e:** Es la constante matemática base de los logaritmos naturales, cuyo valor aproximado es 2.71828.
- **x:** Número de eventos u ocurrencias que se desea analizar.
- **P(x):** Representa la probabilidad de que ocurran exactamente *x*eventos en el intervalo considerado

a. Media y varianza de las distribuciones de Poisson

- La **media** en una distribución de Poisson se obtiene multiplicando el número de ensayos (n) por la probabilidad de éxito (π). Su expresión matemática es:

$$\mu = n\pi$$

- La **varianza** de esta distribución también resulta igual al producto $n\pi$, lo que implica que en Poisson la media y la varianza tienen el mismo valor.

Ejemplo:

Una aerolínea enfrenta dificultades recurrentes con el manejo del equipaje. En una revisión aleatoria de 5.000 maletas, se observó que la mayoría no contenía ningún objeto cortopunzante, mientras que otras incluían uno, y unas pocas, dos o más. Al analizar la frecuencia de estos hallazgos, se determinó que la distribución del número de objetos cortopunzantes por equipaje se ajusta al modelo de distribución de Poisson**.** En total, el personal de control identificó 3.500 objetos cortopunzantes dentro del conjunto examinado.

Se desea calcular la probabilidad de que, al escoger una maleta al azar, esta no contenga ningún objeto cortopunzante.

Para resolverlo, se aplica la distribución de Poisson, pues se trata de un conteo de eventos poco frecuentes en un conjunto grande de observaciones.

Primero se determina la media o parámetro λ, que representa el número promedio de ocurrencias (armas) por equipaje:

$$\lambda = \frac{3500}{5000} = 0.7$$

Esto indica que, en promedio, cada equipaje contiene 0.7 armas cortopunzantes.

A continuación, se calcula la probabilidad de que un equipaje no contenga ninguna arma ($x = 0$):

$$P(x = 0) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Reemplazando los valores conocidos:

$$P(0) = \frac{e^{-0.7}(0.7)^{0}}{0!} = e^{-0.7} = 0.4966$$

Por consiguiente, se puede estimar que la probabilidad de que una maleta elegida al azar no contenga ningún tipo de arma cortopunzante es de aproximadamente 0,4966**,** es decir, un 49,66%**.**

El resultado indica que casi la mitad (49,66%) de los equipajes no contiene armas cortopunzantes**.** Este hallazgo permite inferir que, aunque existen casos de equipajes con presencia de armas, la mayoría no presenta este tipo de elementos. En consecuencia, la aerolínea puede enfocar sus controles en el porcentaje restante, correspondiente a los equipajes con uno o más objetos cortopunzantes, optimizando los procedimientos de revisión y seguridad.

3.9.4 La distribución exponencial

La distribución de Poisson se utiliza para modelar situaciones en las que se cuenta cuántas veces ocurre un determinado evento dentro de un periodo de tiempo o en una zona específica. Por su parte, la distribución exponencial se aplica cuando se busca analizar el tiempo que transcurre entre un suceso y el siguiente, ya que trabaja con datos continuos.

Por ejemplo:

- Mediante una distribución de Poisson, es posible representar cuántos clientes llegan a una agencia bancaria durante una hora determinada.
- Por otro lado, la distribución exponencial permite analizar el tiempo promedio que transcurre entre la atención de un cliente y el siguiente.

Desde el punto de vista matemático, la función de probabilidad acumulada permite calcular la posibilidad de que el tiempo de espera sea igual o inferior a un valor determinado x, y se expresa mediante la siguiente ecuación:

$$P(T \leq x) = 1 - e^{-\lambda x}$$

donde:

- T= tiempo hasta que ocurra el evento,
- λ= parámetro de la distribución (tasa promedio de ocurrencia),
- x= tiempo específico de interés.

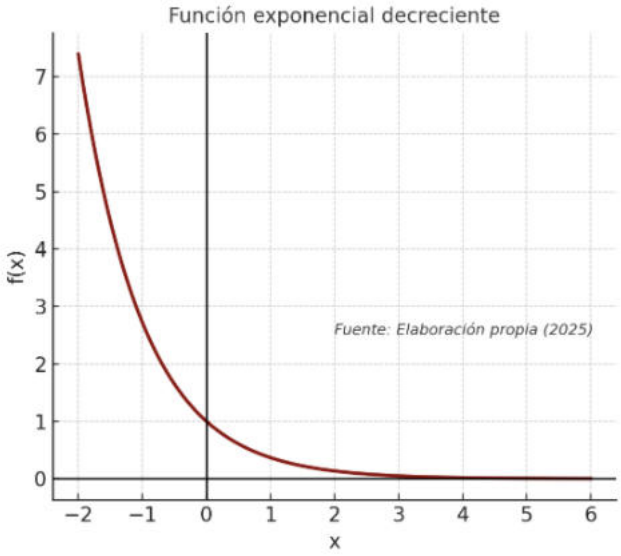

Figura 15. Generar modelos de regresión lineal y no lineal simple con una base de datos reales.

La figura representa el comportamiento de una función exponencial decreciente**,** la cual se caracteriza por tomar valores elevados cuando $x$es pequeño y disminuir de manera progresiva conforme $x$aumenta. Este tipo de funciones aparece con frecuencia en procesos donde la magnitud de un fenómeno disminuye con rapidez al inicio y luego se estabiliza, como ocurre en la depreciación de activos, la disminución de audiencias digitales con el paso del tiempo, o en ciertos modelos de probabilidad asociados a la supervivencia.

Ejemplo

El Hotel Marriott coordina el servicio de transporte hacia el aeropuerto considerando que los taxis arriban según una distribución de Poisson, con una tasa promedio de 12 llegadas por hora. Si un pasajero acaba de

aterrizar y desea estimar la probabilidad de que un taxi llegue en un lapso igual o inferior a cinco minutos, el fenómeno puede describirse mediante una distribución exponencial, dado que esta permite modelar el tiempo de espera entre una llegada y la siguiente.

$$P(X < 5) = 1 - e^{-12(1/12)} = 1 - e^{-1}$$

$$P(X < 5) = 1 - 0.3679 = 0.6321$$

En consecuencia, existe aproximadamente un 63,21 % de probabilidad de que llegue un taxi en cinco minutos o menos.

3.9.5 La distribución uniforme

La distribución uniforme se describe como un modelo de probabilidad en el que todos los eventos posibles tienen la misma posibilidad de presentarse**.** En otras palabras, no existe preferencia ni sesgo hacia ningún resultado: cada uno tiene idéntica probabilidad.Esto significa que no existe preferencia por un valor en particular dentro del intervalo considerado; todos son igualmente probables.

Un ejemplo cotidiano sería el lanzamiento de un dado justo de seis caras: cada número del 1 al 6 tiene la misma probabilidad de aparecer, es decir, $\frac{1}{6}$. Su fórmula es la siguiente:

$$E(x) = \mu = \frac{a+b}{2}$$

La distribución uniforme se distingue porque cada valor comprendido dentro de un intervalo específico tiene exactamente la misma probabilidad de presentarse. Esto significa que ningún resultado es más probable que otro, lo que refleja una situación de total equilibrio o ausencia de sesgo. Matemáticamente, se representa en un rango entre ay b, donde:

- a: Es el valor más bajo posible.
- b: Es el valor más alto posible.

La función de probabilidad es constante en todo el intervalo, lo que significa que cualquier valor entre ay b es igualmente probable.

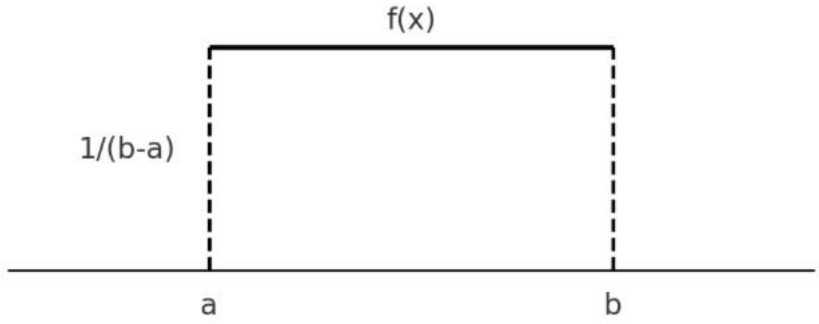

Figura 16. Distribución uniforme.

En cambio, la varianza está definida por:

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Dado que la distribución uniforme adopta la forma de un rectángulo, la probabilidad de que una variable aleatoria asuma un valor dentro de un intervalo determinado se calcula a partir del área bajo su curva.

Matemáticamente, si la variable Xsigue una distribución uniforme en el intervalo $[a, b]$, la función de densidad está dada por:

$$f(x) = \frac{1}{b-a}, a \leq x \leq b$$

En la distribución uniforme continua, el rango de valores posibles está delimitado por el intervalo $[a, b]$. La anchura del intervalo, representada por $b - a$, constituye la base del rectángulo que define esta distribución.

La probabilidad de que una observación se ubique entre dos valores x_1**y** x_2, dentro del intervalo $[a, b]$, se obtiene dividiendo la longitud del subintervalo entre el rango total:

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b-a}, a \leq x_1 < x_2 \leq b$$

Ejemplo

Imaginemos que los equipajes permitidos por Continental Airlines tienen un peso que varía de 14.5 kg a 17.5 kg**,** distribuyéndose de manera uniforme en ese rango. La empresa desea estimar la probabilidad de que

un equipaje escogido al azar tenga un peso comprendido entre 16 kg y 17.2 kg. En este caso, como todos los valores dentro del intervalo son equiprobables, la probabilidad se determina calculando la proporción del subintervalo (16 a 17.2 kg) respecto al intervalo total (14.5 a 17.5 kg).

Sabemos que la probabilidad en una distribución uniforme se calcula con la siguiente expresión:

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b - a}$$

Donde:

- $a = 14.5$representa el valor mínimo,
- $b = 17.5$representa el valor máximo,
- $x_1 = 16$y $x_2 = 17.2$son los límites del intervalo de interés.

Sustituyendo los valores:

$$P(16 \leq X \leq 17.2) = \frac{17.2 - 16}{17.5 - 14.5} = \frac{1.2}{3} = 0.4$$

Por lo tanto, la probabilidad es del 40%, lo que significa que 4 de cada 10 equipajes aproximadamente pesan entre 16 y 17.2 kg.

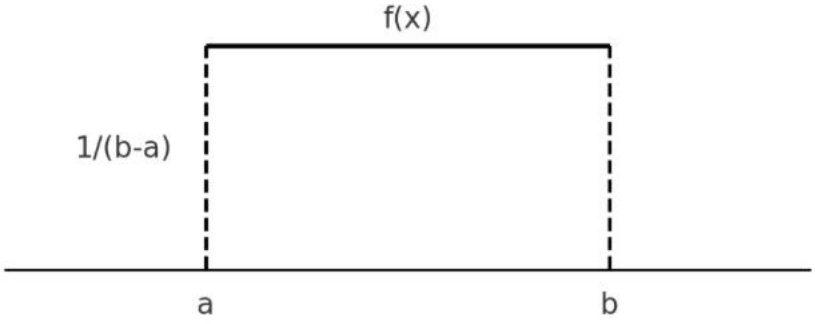

Figura 17. Distribución uniforme.

3.9.6 Distribución probabilística normal

La distribución normal ocupa un papel central en la estadística por su capacidad para modelar fenómenos naturales, sociales y económicos. Entre sus características principales se destacan:

- Es simétrica con respecto a la media.
- Presenta una forma de campana, donde la mayoría de los valores se concentran alrededor del centro.
- Describe variables continuas que pueden tomar un número infinito de valores dentro de un rango.

Una de sus aplicaciones más conocidas es la regla empírica, que permite interpretar cómo se dispersan los datos en torno a la media utilizando la desviación estándar (σ). De acuerdo con esta regla:

- El 68% de los datos se ubica dentro de una desviación estándar por encima y por debajo de la media.

Esta propiedad resulta muy útil para analizar variaciones naturales, por ejemplo, en los tiempos de atención al cliente, el peso de productos o el rendimiento académico, donde los valores extremos son poco frecuentes.

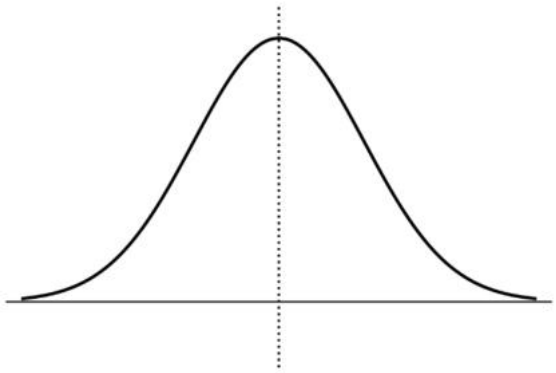

Figura 18. Distribución probabilística normal.

La distribución normal se define completamente a partir de dos parámetros esenciales:

- La media (μ), que determina el punto central o el eje de simetría de la curva.
- La desviación estándar (σ), que refleja el grado de variabilidad o dispersión de los datos respecto a la media.

La forma de la curva cambia según estos parámetros: cuando la desviación estándar es pequeña, la curva se vuelve más alta y estrecha; en cambio, si la desviación aumenta, la curva se aplana y se ensancha.

Además, el área total bajo la curva equivale al 100% de la probabilidad, lo que significa que todos los valores posibles de la variable se encuentran dentro de ese espacio. En este sentido, la probabilidad de que

ocurra un resultado específico se interpreta como el área comprendida bajo la curva dentro de un intervalo determinado.

a. Distribución probabilística normal estándar

Existen infinitas distribuciones normales, y cada una se distingue por los valores específicos de su media (μ) y su desviación estándar (σ). Sin embargo, trabajar con tantas posibilidades puede resultar complejo.

Para facilitar su análisis, todas estas distribuciones pueden transformarse en una forma común llamada distribución normal estándar. Este proceso recibe el nombre de estandarización y se logra mediante la fórmula Z, que convierte cualquier valor en una medida relativa respecto a la media y la desviación estándar:

$$Z = \frac{X - \mu}{\sigma}$$

Donde:

- $\boldsymbol{X}$**:** Corresponde al valor observado.
- $\boldsymbol{\mu}$**:** Corresponde a la media de la distribución.
- $\boldsymbol{\sigma}$: Corresponde a la desviación estándar.

El valor resultante, conocido como puntuación Z, indica cuántas desviaciones estándar se encuentra un dato por encima o por debajo de la media.

En esta transformación, el valor Z indica cuántas desviaciones estándar se encuentra una observación por encima o por debajo de la media. A su vez, X representa un valor concreto de la variable aleatoria original antes de ser estandarizada. Esta conversión permite comparar distintos conjuntos de datos bajo una misma escala, facilitando el análisis y la interpretación de probabilidades en la distribución normal estándar.

Una vez realizada la conversión, la distribución resultante adquiere características únicas:

- La media se convierte en 0.
- La desviación estándar pasa a ser 1.

De este modo, todas las distribuciones normales, sin importar su forma original, pueden representarse bajo un mismo modelo denominado distribución normal estándar.

El gráfico resultante conserva la forma característica de campana, aunque ahora está centrado en el valor 0, lo que simplifica tanto la interpretación como el cálculo de las probabilidades, ya que la media es cero y la desviación estándar es uno en la distribución normal estandarizada.

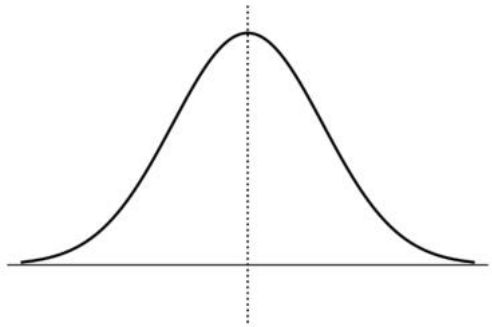

Figura 19. Distribución probabilística normal.

El proceso de estandarización de una distribución normal simplifica la estimación de probabilidades, ya que permite calcular el área bajo la curva comprendida entre un valor observado y la media. Este área corresponde directamente a la probabilidad de que ocurra un determinado evento.

Para realizar estos cálculos se emplean las tablas de la distribución normal estándar, también llamadas tablas Z, las cuales muestran el área acumulada bajo la curva para distintos valores de Z. Dichas tablas, comúnmente incluidas en los manuales de estadística, son una herramienta esencial para la resolución de problemas aplicados y la interpretación de resultados en contextos reales.

Ejemplo

El Ministerio de Turismo llevó a cabo un estudio sobre los destinos más frecuentados en la provincia del Azuay, concluyendo que el tiempo promedio de estadía de los turistas se ajusta a una distribución normal, con una media (μ) de 2,2 días y una desviación estándar (σ) de 0,8 días.

El objetivo es determinar la probabilidad de que un visitante permanezca más de 3,3 días durante su estancia en la temporada vacacional. Para ello, se procede a estandarizar el valor observado, transformándolo a su equivalente en la distribución normal Z, y posteriormente se calcula el área bajo la curva que representa dicha probabilidad.

Datos del problema:

$$X = 3.3,\ \mu = 2.2,\ \sigma = 0.8$$

Primero, estandarizamos el valor utilizando la fórmula:

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{3.3 - 2.2}{0.8} = 1.375$$

El valor obtenido ($Z = 1.38$) representa cuántas desviaciones estándar se encuentra el valor 3.3 por encima de la media.

Buscamos el área asociada en la tabla Z, donde:

$$P(Z < 1.38) = 0.9162$$

Como el ejercicio solicita la probabilidad de hospedarse más de 3.3 días, necesitamos el área a la derecha de Z = 1.38:

$$P(Z > 1.38) = 1 - 0.9162 = 0.0838$$

Por consiguiente, la probabilidad de que un turista permanezca más de 3,3 días en su estadía es de 0,0838, lo que equivale a un 8,38 %.

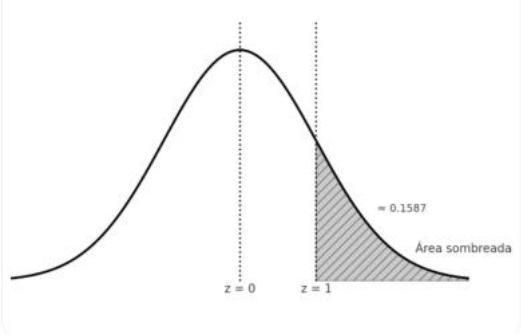

Figura 20. Área Z.

En el gráfico de la distribución normal, el área sombreada indica la probabilidad buscada, que en este caso es 0,0838 o 8,38 %. Dicho valor se obtiene al restar del 0,5 (que representa el área total a la derecha de la media) el área comprendida entre la media y el valor $Z = 1{,}38$, la cual, según la tabla de la distribución normal estándar, equivale a 0,4162. De esta manera, el área resultante bajo la curva simboliza la proporción de turistas que permanecen en el destino más de 3,3 días.

$$P(Z > 1.38) = 0.5 - 0.4162 = 0.0838$$

De esta forma, se demuestra que el área bajo la curva refleja directamente la probabilidad de ocurrencia del evento.

Asimismo, el proceso puede invertirse. Es decir, si se conoce una probabilidad, es posible determinar el valor correspondiente de X a partir del valor de Z. Para ello, se despeja la variable en la fórmula de estandarización, obteniendo:

$$X = Z\sigma + \mu$$

Esta expresión permite transformar nuevamente un valor estándar Zen su equivalente real dentro de la distribución original.

Por ejemplo, si se conoce el valor $Z = 1.38$, la media $\mu = 2.2$y la desviación estándar $\sigma = 0.8$, se puede calcular:

$$X = 1.38(0.8) + 2.2 = 1.104 + 2.2 = 3.304$$

Este resultado confirma que el valor real asociado a $Z = 1.38$es 3.3 días, coincidiendo con el caso analizado.

CAPÍTULO IV

4 PRUEBA DE HIPÓTESIS Y TOMA DE DECISIONES ESTADÍSTICAS

Hasta este punto hemos aprendido a organizar datos, calcular medidas de resumen y usar probabilidades para describir la incertidumbre. Sin embargo, en el análisis estadístico no basta con describir: muchas veces queremos comprobar una afirmación. Aquí es donde aparecen las pruebas de hipótesis, que nos permiten tomar decisiones con respaldo estadístico.

Una hipótesis se entiende como una proposición o afirmación preliminar acerca de una característica o parámetro de una población. Por ejemplo: "El 60% de los clientes de una tienda online está satisfecho con el servicio". Con una muestra, podemos contrastar esa afirmación y decidir si la aceptamos o la rechazamos.

Las pruebas de hipótesis constituyen herramientas fundamentales en la estadística aplicada, ya que orientan la toma de decisiones cuando existen elementos de incertidumbre. A través de su uso, es posible evaluar la validez de los resultados y reducir la probabilidad de interpretar erróneamente la información obtenida en un estudio (Lind, Marchal, & Wathen , 2018).

En el marketing digital, esta lógica se aplica a diario: comparar si un anuncio A genera más clics que un anuncio B, analizar si la tasa de conversión supera un valor esperado o verificar si un nuevo diseño de página mejora la permanencia de los usuarios.

4.1 Conceptos básicos de hipótesis

Antes de aplicar fórmulas o pruebas estadísticas, es necesario comprender la lógica que hay detrás de las hipótesis. En estadística, una hipótesis es una proposición sobre una característica de una población. Puede ser algo tan simple como afirmar que "el tiempo medio que los clientes pasan en una página web es de 10 minutos".

Las pruebas de hipótesis buscan confirmar o rechazar estas afirmaciones usando datos de una muestra. Lo valioso es que nos permiten tomar decisiones basadas en evidencia y no en intuiciones.

4.1.1 Definición de hipótesis nula y alternativa

Cuando realizamos un contraste de hipótesis, siempre planteamos dos escenarios:

- Hipótesis nula (H_0): corresponde al supuesto base que se somete a verificación. Expresa que no existe diferencia significativa o efecto real entre las variables analizadas, sino que cualquier variación observada se debe al azar o a fluctuaciones muestrales. Ejemplo: "La tasa de clics en el anuncio A es igual a la del anuncio B".
- Hipótesis alternativa (H_1). Es la afirmación que contradice a la hipótesis nula. Generalmente expresa diferencia o cambio.

Ejemplo: "La tasa de clics en el anuncio A es mayor que la del anuncio B".

En la práctica, nunca demostramos que una hipótesis sea verdadera de forma absoluta; lo que hacemos es evaluar si los datos son consistentes con H_0 o si nos llevan a rechazarla en favor de H_1 (Agresti, 2018).

4.1.2 Errores tipo I y tipo II

Toda decisión conlleva un riesgo, y en estadística esos riesgos se resumen en dos tipos de error:

- **Error tipo I (α).** Ocurre cuando rechazamos H_0 siendo cierta. Es como decir que un nuevo diseño de página mejora la tasa de conversión, cuando en realidad no lo hace.
- **Error tipo II (β).** Aparece cuando no rechazamos H_0 siendo falsa. Es como mantener una estrategia de anuncios ineficiente porque los datos de la muestra no fueron suficientes para detectar la diferencia real.

Lo importante es entender que ninguna prueba es infalible. Lo que hacemos es controlar estos riesgos con criterios estadísticos (Cochran, 1977).

4.1.3 Poder de la prueba y nivel de significancia

El nivel de significancia (α) representa el límite de tolerancia que el investigador establece respecto al riesgo de cometer un error tipo I, es decir, rechazar la hipótesis nula cuando en realidad es verdadera. En el contexto del marketing y la investigación de mercados, se suele utilizar un valor de $\alpha = 0.05$, lo que implica aceptar un 5% de probabilidad de error al concluir que existe un cambio o diferencia

cuando, en realidad, no la hay. Este valor proporciona un equilibrio razonable entre el rigor estadístico y la posibilidad de detectar efectos relevantes en los datos analizados.

El poder de la prueba $(1 - \beta)$ indica la capacidad de detectar un efecto real cuando este existe. Un poder bajo puede hacer que una campaña exitosa pase desapercibida porque la muestra no tuvo la fuerza estadística suficiente. Por eso, planificar bien el diseño de la investigación y el tamaño de la muestra es clave (Lohr, 2021).

4.2 Procedimiento general para una prueba de hipótesis

Hacer una prueba de hipótesis no es un ritual matemático: es un proceso lógico que sigue pasos ordenados. La idea central es evaluar, con ayuda de los datos, si la evidencia es suficiente para rechazar una afirmación inicial (H_0) en favor de otra alternativa (H_1).

De forma general, todo contraste de hipótesis sigue estas etapas (Hogg, Tanis, & Zimmerman, 2015):

1. Plantear H_0 y H_1

El primer paso es formular claramente las hipótesis.

- H_0: Expresa igualdad o no diferencia.
- H_1: Plantea cambio, desigualdad o diferencia.

Ejemplo: "La tasa de clics en el nuevo anuncio es igual a la del actual" (H_0) frente a "la tasa de clics en el nuevo anuncio es mayor" (H_1).

2. Seleccionar el nivel de significancia (α)

Se fija el riesgo máximo de cometer un error tipo I. Lo usual es $\alpha = 0.05$, aunque en estudios más exigentes se usa $\alpha = 0.01$.

3. Elegir la prueba estadística adecuada

Depende de:

- El tipo de variable (cuantitativa o cualitativa),
- El tamaño de la muestra,
- Si la varianza es conocida o no,
- Y si se trata de una o dos poblaciones.

Ejemplo: para comparar tasas de clics (proporciones) usamos la prueba z para proporciones; para tiempos de permanencia (medias) usamos z o t, según el caso.

4. Cálculo del estadístico de prueba

Es el valor que resume la evidencia de los datos frente a H_0. Según el caso, puede ser z, t, χ^2, entre otros.

Ejemplo

Supongamos que una campaña digital tiene como referencia que la tasa de clics es del 12%. En una muestra de 500 visitas, 75 dieron clic.

$$\hat{p} = \frac{75}{500} = 0.15$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.15 - 0.12}{\sqrt{0.12 \times 0.88/500}}$$

$$Z = \frac{0.03}{\sqrt{0.0002112}} = \frac{0.03}{0.0145} = 2.07$$

5. Regla de decisión

Si |Z| > Z crítico (1.96 en el caso de 95%), rechazamos H_0.

Si |Z| ≤ Z crítico, no rechazamos H_0.

En este caso: 2.07 > 1.96 → se rechaza H_0.

4.3 Pruebas para una población

Cuando se dispone de información de una muestra, una de las aplicaciones más comunes es comprobar hipótesis sobre un único parámetro poblacional. Estas pruebas permiten responder preguntas como:

- ¿Existe evidencia estadística que indique que el tiempo promedio de permanencia en una página web difiere del valor esperado?
- ¿Proporción de clientes satisfechos alcanza el nivel fijado como meta?

A continuación se explican los dos casos más usuales: pruebas para la media y pruebas para la proporción.

4.3.1 Prueba de hipótesis para la media

a) Caso con σ conocida (Z)

Se aplica en situaciones donde se dispone del valor de la desviación estándar poblacional o cuando la muestra es lo suficientemente amplia (más de 30 observaciones).

Ejemplo.

Una empresa digital afirma que los usuarios pasan en promedio 10 minutos en su portal. Se toma una muestra de 36 usuarios: $\bar{X} = 9.2, \sigma = 1.8$.

1. Hipótesis

H_0: $\mu = 10$

H_1: $\mu \neq 10$

2. Estadístico de prueba

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{9.2 - 10}{1.8/\sqrt{36}} = \frac{-0.8}{0.3} = -2.67$$

3. Decisión

Al nivel 5% ($\alpha = 0.05$), el valor crítico es ±1.96. Como $-2.67 < -1.96$, se rechaza H_0.

Conclusión: El tiempo promedio de permanencia es menor a 10 minutos, lo que sugiere revisar la experiencia de usuario.

b) Caso con σ desconocida (t de Student)

Se aplica cuando la muestra es reducida (menos de 30 observaciones) y no se dispone del valor de la desviación estándar poblacional.

Ejemplo.
Se desea saber si el tiempo medio de permanencia en un blog es de 8 minutos. Con una muestra de 25 usuarios:

$\bar{X} = 8.4, s = 2.5.$

1. Hipótesis

H_0: $\mu = 8$

H_1: $\mu \neq 8$

2. Estadístico de prueba

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{8.4 - 8}{2.5/\sqrt{25}} = \frac{0.4}{0.5} = 0.8$$

3. Decisión

Con gl = 24 y α = 0.05, el valor crítico es ± 2.064. Como 0.8 está dentro del rango, no se rechaza H_0.

Conclusión: No hay evidencia suficiente para afirmar que el tiempo medio sea distinto a 8 minutos.

4.3.2 Prueba de hipótesis para una proporción

Sirve para comprobar afirmaciones sobre la proporción de individuos que cumplen cierta característica.

Ejemplo.
Un community manager sostiene que el 60% de los seguidores interactúa con las publicaciones. Se toma una muestra de 100 usuarios y 54 interactúan.

1. Hipótesis

H_0: $p = 0.60$

H_1: $p \neq 0.60$

2. Estadístico de prueba

$$\hat{p} = \frac{54}{100} = 0.54$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.54 - 0.60}{\sqrt{0.60 \times 0.40/100}} = \frac{-0.06}{0.049} = -1.22$$

3. Decisión

Con $\alpha = 0.05$, el valor crítico es ± 1.96. Como - 1.22 está dentro del rango, no se rechaza H_0.

Conclusión: La proporción de usuarios que interactúan no es estadísticamente diferente al 60%.

4.4 Pruebas para dos poblaciones

En la práctica muchas veces no nos interesa estudiar un solo grupo, sino comparar dos poblaciones.

Por ejemplo:

- ¿Los clientes de Facebook reaccionan más a una campaña que los de Instagram?
- ¿El tiempo de permanencia en la versión antigua de un sitio web es distinto al de la nueva versión?

Las pruebas de hipótesis para dos poblaciones permiten responder a estas preguntas de manera objetiva.

- **Comparación de medias independientes**

Se aplica cuando tenemos dos grupos distintos de datos y queremos contrastar sus medias.

Ejemplo.
Una empresa digital quiere comparar el tiempo medio de permanencia en dos diseños de página web:

- Diseño A: $n_1 = 30$, $\bar{X}_1 = 9.5$minutos, $s_1 = 1.2$.
- Diseño B: $n_2 = 28$, $\bar{X}_2 = 8.7$minutos, $s_2 = 1.5$.

1. Hipótesis

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

2. Estadístico de prueba (t de dos muestras):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Sustituyendo:

$$t = \frac{9.5 - 8.7}{\sqrt{\frac{1.44}{30} + \frac{2.25}{28}}} = \frac{0.8}{\sqrt{0.048 + 0.080}} = \frac{0.8}{\sqrt{0.128}} = \frac{0.8}{0.357} = 2.24$$

3. Decisión

Al 5% (gl ≈ 56), el valor crítico es ±2.00. Como 2.24 > 2.00, se rechaza H_0.

Conclusión: El tiempo medio de permanencia es significativamente mayor en el diseño A.

4.4.1 Comparación de proporciones independientes

Se utiliza cuando comparamos proporciones de dos grupos distintos.

Ejemplo:

Dos anuncios publicitarios obtuvieron estos resultados:

- Anuncio A: 600 impresiones, 120 clics → $\hat{p}_1 = 0.20$.
- Anuncio B: 700 impresiones, 180 clics → $\hat{p}_2 = 0.257$.

1. Hipótesis

H_0: $p_1 = p_2$

H_1: $p_1 \neq p_2$

2. Proporción combinada

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 180}{600 + 700} = \frac{300}{1300} = 0.231$$

3. **Estadístico de prueba**

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$Z = \frac{0.20 - 0.257}{\sqrt{0.231 \times 0.769(\frac{1}{600} + \frac{1}{700})}} = \frac{-0.057}{\sqrt{0.1777 \times 0.003095}}$$

$$= \frac{-0.057}{\sqrt{0.00055}} = \frac{-0.057}{0.0235} = -2.42$$

4. **Decisión**

Al 5%, el valor crítico es ±1.96. Como -2.42 < -1.96, se rechaza H_0.

Conclusión: El Anuncio B tiene una tasa de clics significativamente mayor que el A.

4.4.2 Comparación de muestras relacionadas (antes y después)

Se aplica cuando el mismo grupo se mide en dos momentos distintos.

Ejemplo:
Se mide la satisfacción de 8 usuarios antes y después de una actualización en la app:

- Antes: 6, 7, 5, 6, 7, 6, 5, 6
- Después: 7, 8, 6, 7, 8, 7, 6, 7

1. Calcular diferencias (d)

1, 1, 1, 1, 1, 1, 1, 1 → todas las diferencias = 1.

2. Media de diferencias

$\bar{d} = 1.$

3. Estadístico de prueba (t apareada)

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

En este caso, la varianza de d es cero (todas las diferencias iguales). El valor t es infinito → evidencia contundente.

Conclusión: La satisfacción aumentó significativamente tras la actualización.

BIBLIOGRAFÍA

Agresti, A. (2018). *Statistical Methods for the Social Sciences (5th ed.).* Nueva York.: Pearson.

Anderson, D. R., Sweeney, D. J., & Williams , T. A. (2016). *Estadística para los negocios y la economía (12.ª ed.).* Ciudad de México.: Cengage Learning.

Anderson, D. R., Sweeney, D. J., Camm, J. D., & Cochran , J. J. (2019). *Estadística para administración y economía (14.ª ed.).* Ciudad de México.: Cengage Learning.

Cochran, W. G. (1977). *Sampling Techniques (3rd ed.).* Nueva York.: John Wiley & Sons.

Gujarati, D. N., & Porter, D. C. (2010). *Econometría (5.ª ed.).* Ciudad de México.: McGraw-Hill.

Hair, J. F., Wolfinbarger, M., Money, A. H., Samouel, P., & Page, M. J. (2015). *Fundamentos de investigación de mercados (3.ª ed.).* Ciudad de México.: MLacGraw-Hill Education.

Hernández , R., Fernández, C., & Baptista, P. (2022). *Metodología de la investigación (7.ª ed.).* Ciudad de México: McGraw-Hill.

Hernández Sampieri , R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación (6.ª ed.).* México: McGraw-Hill.

Keller, G., & Warrack, B. (2016). *Estadística aplicada a los negocios y la economía (9.ª ed.).* Ciudad de México.: Cengage Learning.

Levin , R., & Rubin , D. (2017). *Estadística para administración y economía (8.ª ed.).* Ciudad de México: Pearson Educación.

Lind, D. A., Marchal, W. G., & Wathen , S. A. (2018). *Estadística aplicada a los negocios y la economía (16.ª ed.).* México: McGraw-Hill.

Lohr, S. L. (2021). *Sampling: Design and Analysis (2nd ed.).* Boca Ratón.: Chapman & Hall/CRC.

Malhorta, N. K. (2019). *Investigación de mercados (7.ª ed.).* Ciudad de México.: Pearson Educación.

Moore, D. S., McCabe, G. P., & Craig , B. A. (2017). *Introducción a la práctica de la estadística (8.ª ed.).* Nueva York.: H. Freeman.

Newbold, P., Carlson, W., & Thorme, B. (2013). *Estadística para administración y economía (8.ª ed.).* Madrid: Pearson Educación.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis (3rd ed.).* Belmont: Cengage Learning.

Triola , M. F. (2020). *Estadística (13.ª ed.).* Ciudad de México: Pearson Educación.

Walpole , R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *robabilidad y estadística para ingeniería y ciencias (9.ª ed.).* Pearson Educación.: Ciudad de México.

Estadística aplicada: herramienta para la investigación, la educación y la toma de decisiones, se publicó en el mes de diciembre de 2025.

ISBN: 978-9907-0-0423-6

Grupo Editorial BLR
Ecuador
Cel: +593 98 320 4362
https://grupoblr.com/
publicaciones@grupoblr.com

BIOGRAFÍA DE LOS AUTORES

Raúl Marcelo Chávez Benavides:
Economista por la Pontificia Universidad Católica del Ecuador; Máster en Análisis Económico por la Universidad Oberta de Catalunya (UOC), y actualmente cursando la maestría en Estadística Aplicada en la Universidad Politécnica del Carchi (UPEC). Fue docente del Sistema Nacional de Nivelación y Admisión Actualmente, es docente universitario.

Elsita Margoth Chávez García:
Elsita Margoth Chávez García, graduada de la Universidad de Las Américas como Ingeniera Comercial con mención en Negocios Internacionales, Magister en Gestión de Marketing y Servicio al Cliente en la Escuela Superior Politécnica del Chimborazo, Máster en Dirección Estratégica con mención en Gerencia por la Universidad Iberoamericana de Puerto Rico. Doctora en Gerencia (PhD) en la Universidad Central de Venezuela. Doctor Honoris Causa por la Universidad Del Norte – Chile.

Jhosselyn Briggeth Garcia Aldaz:
Psicóloga Clínica y Máster en Salud, Integración y Discapacidad por la Universidad Complutense de Madrid. Docente investigadora en la Universidad Estatal de Bolívar. Autora de publicaciones en educación inclusiva y desarrollo socioemocional, con experiencia en proyectos internacionales vinculados a derechos humanos, salud mental e inserción social..

Darwin Vladimir Rivera Piñaloza:
Darwin Vladimir Rivera Piñaloza, Doctor en Contabilidad y Auditoría, especialista en Auditoría Integral. Docente titular de la Universidad Estatal de Bolívar. Actualmente cursa estudios Doctorales en Administración Gerencial en la Universidad Benito Juárez de la ciudad de México. Cargos ocupados: Docente Universitario, Director Provincial de la Contraloría de Tungurahua, entre otros.

ESTADÍSTICA APLICADA: HERRAMIENTA PARA LA INVESTIGACIÓN, LA EDUCACIÓN Y LA TOMA DE DECISIONES

Estimado lector, este libro invita a ver la estadística como una herramienta cercana, útil y necesaria en el mundo digital, logrando un valioso equilibrio entre la claridad de los conceptos fundamentales y su aplicación práctica en el marketing digital.

La obra guía al lector desde la organización hasta la interpretación de los datos, enfatizando que estos no son simples registros, sino la base para tomar decisiones más inteligentes y estratégicas en la práctica profesional. Mediante ejemplos concretos, el texto facilita la comprensión de un campo complejo.

El libro está diseñado como una guía para estudiantes que se inician y para profesionales que buscan reforzar su formación. Su principal propuesta es enseñar a pensar con criterio y a mirar la realidad desde los datos, reconociéndolos como una oportunidad continua para mejorar las acciones diarias, más allá del simple uso de fórmulas.

Agradecemos a todos los lectores que se acercan a esta obra con ánimo de aprender, aplicar y transformar.

GRUPO BLR
UEB UNIVERSIDAD ESTATAL DE BOLIVAR
Grupo Editorial BLR
Ecuador
Cel: +593 98 320 4362
https://grupoblr.com/
publicaciones@grupoblr.com
ISBN: 978-9907-0-0423-6