

INSTITUTO SUPERIOR TECNOLÓGICO
SUDAMERICANO

Quito, Ecuador



GUÍA DE ESTUDIO PARA LA
ASIGNATURA ESTADÍSTICA

Juan Pablo Bustos Rivera

Diseño de carátula y edición:
D.I. Santa de la Caridad Ruiz Crespo

Dirección editorial:
Dr.C. Blas Yoel Juanes Giraud

ISBN: 978-9942-7209-9-3

Sobre la presente edición:
© YOL Editorial, 2024

Podrá reproducirse, de forma parcial o total el contenido de esta obra,
siempre que se haga de forma literal y se mencione a:

YOL Editorial

Pedro Vicente Maldonado y Vicente Andrade, 2-18, Quito, Ecuador.

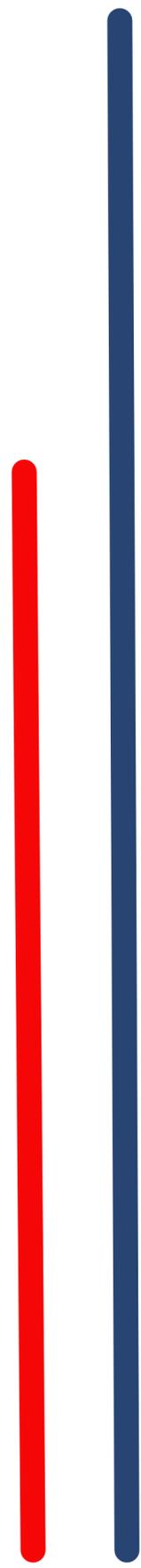
<http://www.yoeditorial.com>

E-mail: consejo@yoeditorial.com



GUÍA DE ESTUDIO PARA LA ASIGNATURA ESTADÍSTICA

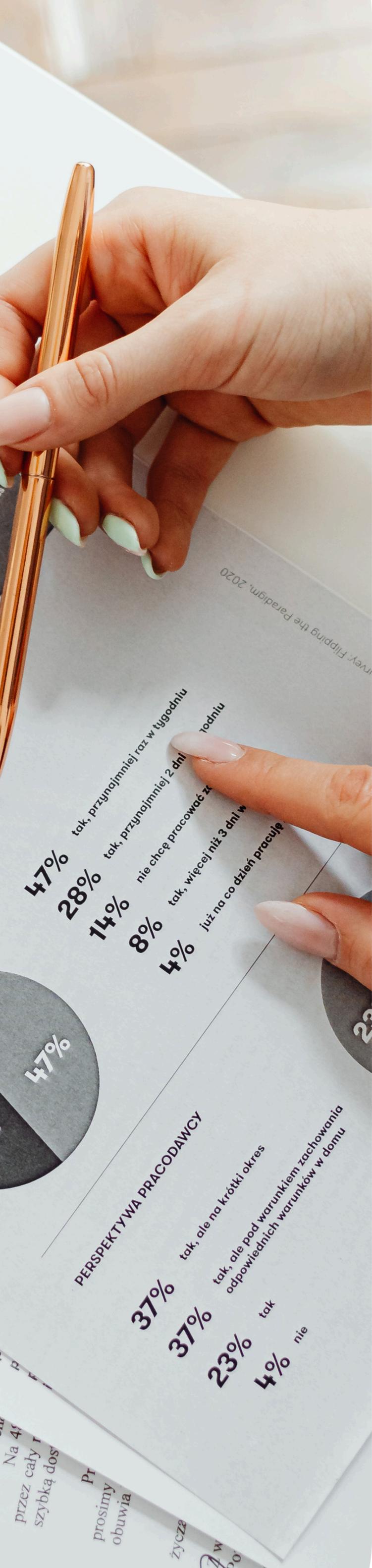
Juan Pablo Bustos Rivera



**TABLA DE
CON
TENIDO**

ÍNDICE

Sumilla de la Asignatura.....	7
Orientaciones generales.....	9
Unidad 1: Introducción a la estadística.....	10
Definición de estadística.....	10
Importancia de la estadística en diversas áreas.....	10
Tipos de datos.....	11
Métodos estadísticos descriptivos e inferenciales.....	12
Unidad 2: Estadística descriptiva.....	17
Medidas de tendencia central (media, mediana, moda).....	17
Medidas de dispersión (rango, desviación estándar, varianza).....	18
Representaciones gráficas (histogramas, diagramas de barras, diagramas de dispersión).....	19
Unidad 3: Probabilidad.....	25
Concepto de probabilidad.....	25
Eventos y espacio muestral.....	25
Reglas de probabilidad (suma, producto, complemento).....	26
Distribuciones de probabilidad (binomial, normal).....	29
Unidad 4: Población, Muestreo y distribuciones muestrales.....	35
Población.....	35
Métodos de muestreo (aleatorio, estratificado, por conglomerados).....	36
Estimación de parámetros poblacionales.....	38
Distribuciones muestrales (distribución muestral de la media, distribución muestral de la proporción).....	39
Solucionario de las Autoevaluaciones.....	44
Referencias en la guía.....	46
Bibliografía para la clase.....	47
Bibliografía básica.....	47
Bibliografía complementaria.....	47



SUMILLA DE LA ASIGNATURA

Sumilla de la asignatura

La estadística es una rama de la matemática y por ende un área fundamental y herramienta de apoyo indispensable para el desempeño de todo profesional y parte integral de la formación académica en diferentes áreas del saber. La estadística en general nos permite interpretar y entender fenómenos sociales, políticos, económicos aportando así en diferentes aristas del conocimiento del profesional.

En la materia de estadística se abordan temas fundamentales como la población, la muestra, variables y sus tipos que permiten identificar aspectos fundamentales para estudiar un fenómeno social, político, económico y técnico. Además, se abordan técnicas de interpretación y análisis de datos, que coadyuvan a explicar dichos fenómenos y plantear distintas soluciones o medidas a tomarse frente a estos o simplemente a conocer cuáles son las leyes que rigen estos.

Competencias

Competencias Generales

- Conocer definiciones y generalidades para estudiar un fenómeno social, político o económico.
- Aplicar diferentes métodos para el análisis e interpretación de datos.
- Plantear conclusiones, generalidades y posibles soluciones frente a un fenómeno social, económico y político.

Competencias Básicas

- Diferenciar términos generales usados en la estadística.
- Aplicar métodos y reglas estadísticas básicas.
- Comprender la información de acuerdo al contexto.
- Identificar variables, datos y preguntas en una situación planteada.

- Relacionar las variables para dar respuesta a las preguntas planteadas.
- Seleccionar el método apropiado al problema y dar solución adecuada.
- Utilizar diferentes estrategias para comprobar que la respuesta obtenida es apropiada al problema.

Competencias Específicas

- Manejar con mucha claridad los definiciones y métodos generales.
- Utilizar Métodos estadísticos en forma adecuada y pertinente.
- Entender y plantear posibles soluciones a problemas de índole económico, político o social.

Objetivo de la Asignatura

Interpretar y entender fenómenos sociales, políticos, económicos aportando así en diferentes aristas del conocimiento del profesional.

Logros del Aprendizaje

- Utiliza los diferentes métodos de análisis e interpretación de datos que respondan a un fenómeno.
- Desarrolla el razonamiento lógico y analítico.
- Utiliza Métodos estadísticos en forma adecuada y pertinente.
- Soluciona problemas de aplicación necesarios en otras asignaturas.

Orientaciones generales

Estimados estudiantes

¡Bienvenidos a la asignatura! Esta guía está estructurada de tal manera para que el contenido de cada unidad sea fácil de entender, para que tus estudios sean exitosos y alcances tus objetivos de aprendizaje. Con esta guía aprenderás a comprender y analizar la información concerniente a la materia desarrollada en clase y complementar la misma con trabajo autónomo de aprendizaje. Para ello te recomendamos lo siguiente:

- **Organiza tu tiempo:** Dedicar tiempo regular a estudiar y repasar la clase, estableciendo un horario fijo y cumpliéndolo.
- **Participa activamente:** Realiza ejercicios prácticos y participa en discusiones para reforzar tu comprensión. Consulta los textos de consulta en el PEA que lo encuentras en el Aula Virtual.
- **Practica el aprendizaje colaborativo:** Trabaja en equipo con compañeros de clase para discutir conceptos, resolver problemas y compartir conocimientos.
- **Consulta recursos adicionales:** Utiliza libros de texto, tutoriales en línea y otros recursos para complementar tu aprendizaje.
- **Consultas:** No dudes en pedir ayuda si tienes dudas o dificultades con algún tema. Utiliza recursos como el profesor o compañeros de clase para aclarar tus dudas.

Unidad 1: Introducción a la estadística

Definición de estadística

La estadística es una disciplina matemática que se encarga de recopilar, organizar, analizar e interpretar datos para obtener conclusiones significativas sobre fenómenos o poblaciones de interés (B., 1973). Utilizando métodos tanto descriptivos como inferenciales, la estadística permite resumir grandes cantidades de información en medidas comprensibles, identificar patrones, tendencias y relaciones en los datos, así como tomar decisiones informadas basadas en la incertidumbre inherente a cualquier proceso de recolección de datos. Es una herramienta fundamental en una amplia variedad de campos, desde la investigación científica hasta la toma de decisiones empresariales, contribuyendo al avance del conocimiento y la comprensión de nuestro mundo.

Importancia de la estadística en diversas áreas

La estadística despliega su importancia trascendental en una amplia gama de áreas, moldeando decisiones cruciales y delineando el camino hacia el progreso en la ciencia, la industria y la sociedad en general.

En el ámbito científico, la estadística actúa como el catalizador esencial para la investigación rigurosa. Desde la medicina hasta la biología, proporciona las herramientas necesarias para analizar datos experimentales, evaluar la significancia de los resultados y extraer conclusiones confiables. Sin la estadística, sería imposible determinar la eficacia de un nuevo medicamento, comprender la evolución de una especie o evaluar el impacto de un tratamiento médico.

En el mundo empresarial (Triola, 2004), la estadística se erige como el faro que guía las decisiones estratégicas. Desde el análisis de mercado hasta la gestión de recursos, brinda la capacidad de

identificar tendencias, comprender el comportamiento del consumidor y prever futuros escenarios. Empresas de todos los tamaños confían en la estadística para optimizar sus operaciones, mejorar la calidad de sus productos y servicios, y maximizar su rentabilidad.

En el ámbito gubernamental, la estadística es la columna vertebral de la formulación de políticas públicas basadas en evidencia. A través del análisis de datos demográficos, económicos y sociales, permite identificar necesidades, evaluar programas y medir el impacto de las intervenciones gubernamentales. Además, desempeña un papel crucial en la planificación urbana, la gestión de recursos naturales y la respuesta a crisis y emergencias.

En el campo académico, la estadística es el puente que conecta las teorías abstractas con la realidad concreta. Facilita la validación de hipótesis, la interpretación de resultados y el avance del conocimiento en todas las disciplinas. Ya sea en física, psicología, sociología o cualquier otra área, la estadística proporciona el marco metodológico necesario para llevar a cabo investigaciones significativas y publicar hallazgos relevantes.

Tipos de datos

La distinción entre datos cualitativos y cuantitativos es esencial para determinar qué tipo de análisis estadístico es apropiado para un conjunto de datos dado. Los tipos de datos se pueden clasificar en dos categorías principales: cualitativos y cuantitativos (Fernández, 2002).

Datos cualitativos:

- **Nominales:** Son datos que representan categorías o clases sin ningún orden inherente. Por ejemplo, el género (masculino, femenino), el estado civil (soltero, casado, divorciado) o el

color favorito (rojo, azul, verde).

- **Ordinales:** Son datos que representan categorías o clases con un orden específico. Aunque las categorías tienen un orden, la diferencia entre ellas no es necesariamente uniforme. Por ejemplo, la escala de calificación de una película (excelente, bueno, regular, malo) o la clasificación socioeconómica (bajo, medio, alto).

Datos cuantitativos:

- **Discretos:** Son datos que representan valores numéricos individuales que son contables y finitos. Estos valores generalmente son enteros y no pueden tomar valores intermedios. Por ejemplo, el número de hijos en una familia, el número de estudiantes en una clase o el número de veces que se lanza una moneda.
- **Continuos:** Son datos que representan valores numéricos que pueden tomar cualquier valor dentro de un rango específico. Estos valores pueden ser fraccionarios y tienen una cantidad infinita de posibilidades dentro del rango. Por ejemplo, la altura de una persona, el peso de un objeto o la temperatura.

Métodos estadísticos descriptivos e inferenciales

Los métodos estadísticos (Ross, 2018) se dividen principalmente en dos categorías: descriptivos e inferenciales:

Métodos estadísticos descriptivos:

- Los métodos descriptivos se centran en la descripción y resumen de los datos disponibles.
- Estos métodos incluyen técnicas para organizar, resumir y presentar datos de manera que sean comprensibles y significativos.
- Algunas de las técnicas descriptivas más comunes incluyen medidas de tendencia central (como la media, la mediana y la moda) y medidas de dispersión (como la desviación estándar y el rango).

Además, los métodos descriptivos pueden involucrar representaciones gráficas, como histogramas, diagramas de barras, diagramas de dispersión, entre otros, que permiten visualizar la distribución y las relaciones entre las variables.

Métodos estadísticos inferenciales:

- Los métodos inferenciales van más allá de la descripción de los datos para realizar inferencias o conclusiones sobre una población más amplia basándose en una muestra de datos.
- Estos métodos permiten generalizar los resultados obtenidos de la muestra a toda la población, asumiendo que la muestra es representativa y que se han seguido ciertos procedimientos estadísticos apropiados.
- Algunos ejemplos de técnicas inferenciales incluyen la estimación de parámetros poblacionales a partir de una muestra, pruebas de hipótesis para evaluar afirmaciones sobre la población, y la construcción de intervalos de confianza para estimar la precisión de nuestras conclusiones.
- Los métodos inferenciales también pueden implicar el uso de modelos estadísticos, como la regresión y el análisis de varianza, para examinar las relaciones entre variables y hacer predicciones sobre el comportamiento futuro de un sistema.

Autoevaluación

1. ¿Cuál de las siguientes opciones describe mejor la estadística?

- a) Un conjunto de métodos para organizar datos en tablas y gráficos.
- b) El estudio de fenómenos naturales utilizando modelos matemáticos.
- c) Una disciplina que se enfoca en recopilar, organizar, analizar e interpretar datos.
- d) La aplicación de la teoría de probabilidades en la toma de decisiones.

2. ¿Por qué la estadística es importante en diversas áreas?

- a) Porque solo se aplica en el ámbito científico.
- b) Porque permite tomar decisiones basadas en evidencia y comprender el mundo que nos rodea.
- c) Porque solo se utiliza en el campo de la economía.
- d) Porque proporciona métodos para realizar experimentos en laboratorios.

3. ¿Cuál de las siguientes opciones NO es un tipo de datos?

- a) Nominales.
- b) Ordinales.
- c) Continuos.
- d) Matemáticos.

4. ¿Cuál de las siguientes afirmaciones describe mejor los métodos estadísticos descriptivos?

- a) Son técnicas para hacer inferencias sobre una población a partir de una muestra.
- b) Se utilizan para organizar, resumir y presentar datos de manera comprensible.
- c) Incluyen pruebas de hipótesis y análisis de regresión.
- d) Son adecuados para realizar predicciones sobre el comportamiento futuro.

5. ¿Cuál es uno de los temas principales de la Introducción a la estadística?

- a) Análisis de regresión.
- b) Importancia de la probabilidad en la estadística.
- c) Métodos de muestreo.
- d) Definición de estadística.

6. ¿Cuál de las siguientes afirmaciones describe mejor la importancia de la estadística en diversas áreas?

- a) Ayuda a recopilar solo datos cualitativos.
- b) Facilita la toma de decisiones informadas y la comprensión de fenómenos complejos.
- c) Limita su aplicación al campo científico.
- d) Se centra únicamente en el análisis matemático de los datos.

7. ¿Cuál de los siguientes NO es un tipo de datos según se discute en el Capítulo 1?

- a) Nominales.
- b) Continuos.
- c) Temporales.
- d) Ordinales.

8. ¿Qué representan los métodos estadísticos descriptivos?

- a) Establecen relaciones entre variables en una población.
- b) Permiten hacer predicciones sobre el futuro.
- c) Organizan, resumen y presentan datos de manera comprensible.
- d) Realizan inferencias sobre una población a partir de una muestra.

9. ¿Cuál es el propósito principal del Capítulo 1: Introducción a la estadística?

- a) Introducir métodos avanzados de análisis estadístico.
- b) Explorar la historia de la estadística.
- c) Definir los conceptos básicos y la importancia de la estadística.
- d) Analizar casos de estudio específicos.

10. ¿Qué tipo de información se discute principalmente en el Capítulo 1: Introducción a la estadística?

- a) Métodos de muestreo.
- b) Análisis de regresión.

- c) Definiciones y conceptos fundamentales.
- d) Interpretación de intervalos de confianza.

Unidad 2: Estadística descriptiva

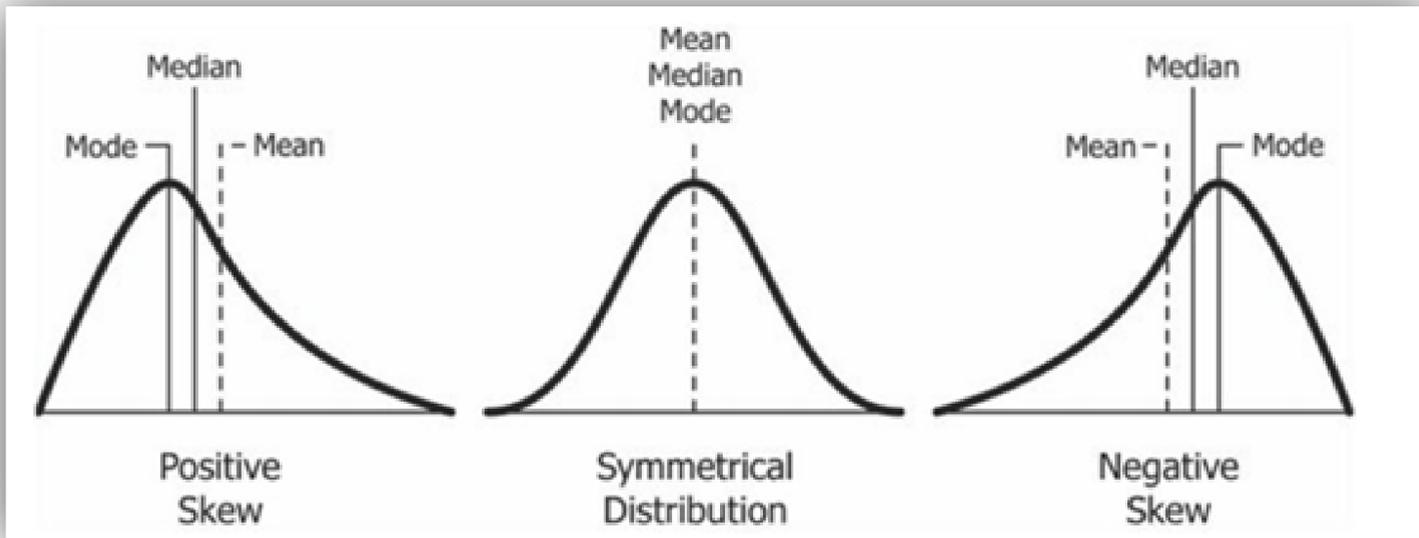
Medidas de tendencia central (media, mediana, moda)

Las medidas de tendencia central (Fernández, 2002) son estadísticas que se utilizan para describir la ubicación central o típica de un conjunto de datos. Estas medidas proporcionan un punto de referencia representativo que resume la distribución de los datos en torno a un valor central. Las tres medidas de tendencia central más comunes son:

- **Media:** También conocida como promedio, la media es la suma de todos los valores en un conjunto de datos dividida por el número total de valores. Es una medida útil para describir el valor típico o central de un conjunto de datos cuando los valores están distribuidos de manera relativamente uniforme.
- **Mediana:** La mediana es el valor que ocupa la posición central en un conjunto de datos ordenados de menor a mayor (o de mayor a menor). Si hay un número impar de observaciones, la mediana es simplemente el valor en el medio. Si hay un número par de observaciones, la mediana es el promedio de los dos valores centrales. La mediana es útil cuando los datos están sesgados o cuando hay valores atípicos que pueden distorsionar la media.
- **Moda:** La moda es el valor que ocurre con mayor frecuencia en un conjunto de datos. En otras palabras, es el valor que tiene la mayor frecuencia de ocurrencia. Puede haber una moda (unimodal) si un valor tiene la mayor frecuencia, dos modas (bimodal) si hay dos valores con la misma frecuencia máxima, o más de dos modas (multimodal) si hay varios valores con la misma frecuencia máxima. La moda es útil para describir la tendencia central en conjuntos de datos que tienen valores repetidos o categorías.

Figura 1

Representación curva estadística



Nota: La figura muestra la representación de la curva estadística. Tomado de Jain (2018)

Medidas de dispersión (rango, desviación estándar, varianza)

Las medidas de dispersión (Fernández, 2002) son estadísticas que describen la variabilidad o dispersión de un conjunto de datos alrededor de su medida central. Estas medidas proporcionan información sobre la dispersión de los valores individuales en relación con la medida de tendencia central. Las tres medidas de dispersión más comunes son:

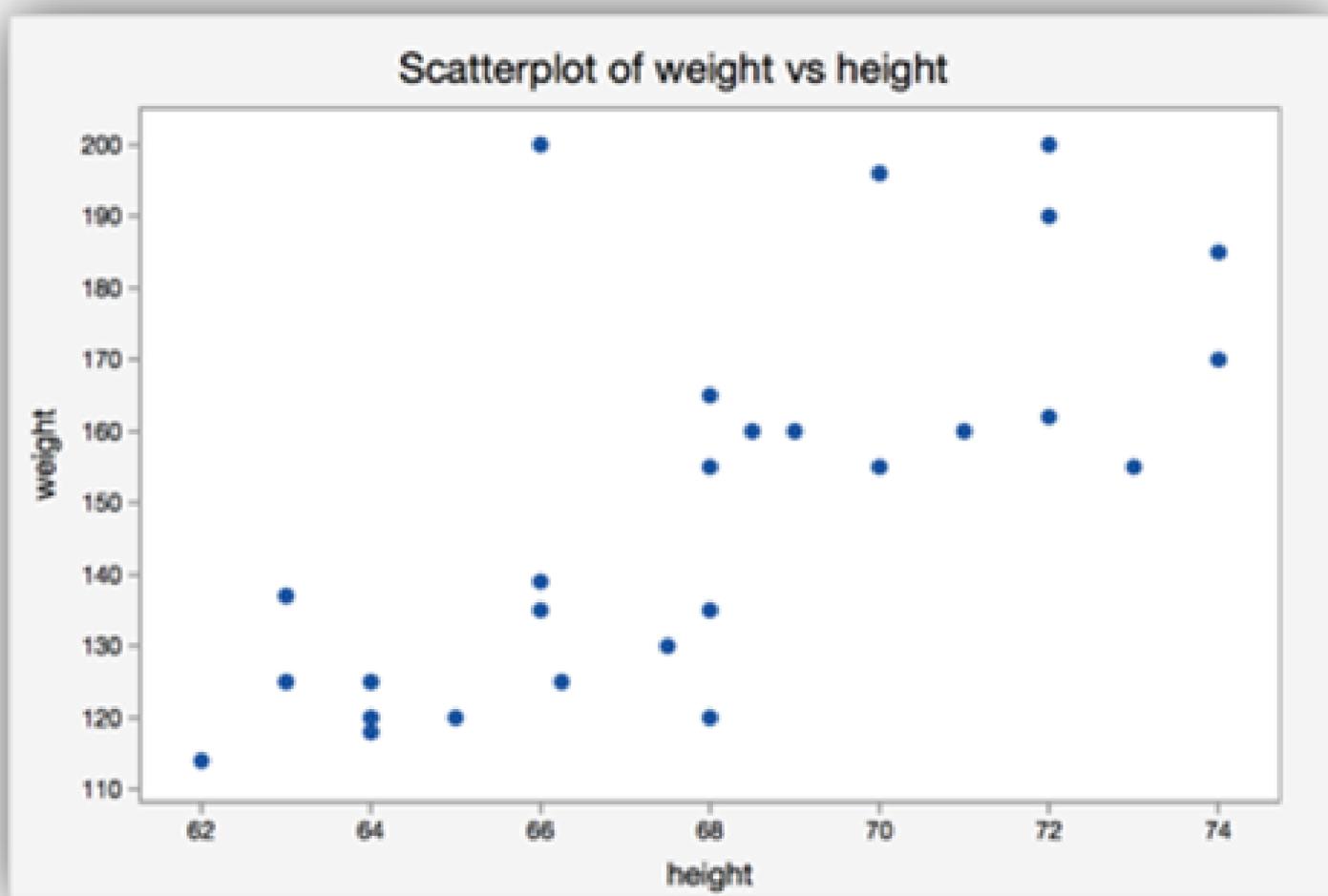
- **Rango:** El rango es la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos. Es la medida más simple de dispersión y proporciona una idea básica de la variabilidad de los datos. Sin embargo, el rango puede ser sensible a valores atípicos y no proporciona información sobre la dispersión de los valores intermedios.
- **Desviación estándar:** La desviación estándar es una medida de dispersión que indica cuánto se desvían los valores individuales del promedio (media) en un conjunto de datos. Se calcula como la raíz cuadrada de la varianza y se expresa en las mismas unidades que los datos originales. Una desviación estándar más alta indica una mayor dispersión de los datos alrededor de la media, mientras que una desviación estándar

más baja indica una menor dispersión.

Varianza: La varianza es otra medida de dispersión que calcula la media de las diferencias cuadráticas entre cada valor y la media del conjunto de datos. Es el cuadrado de la desviación estándar y proporciona una medida de la dispersión promedio de los datos en relación con la media. Al igual que la desviación estándar, una varianza más alta indica una mayor dispersión de los datos, mientras que una varianza más baja indica una menor dispersión.

Figura 2

Representación diagrama dispersión



Nota: La figura muestra un ejemplo de un diagrama de dispersión. Tomado de Pennstate University (s.f.)

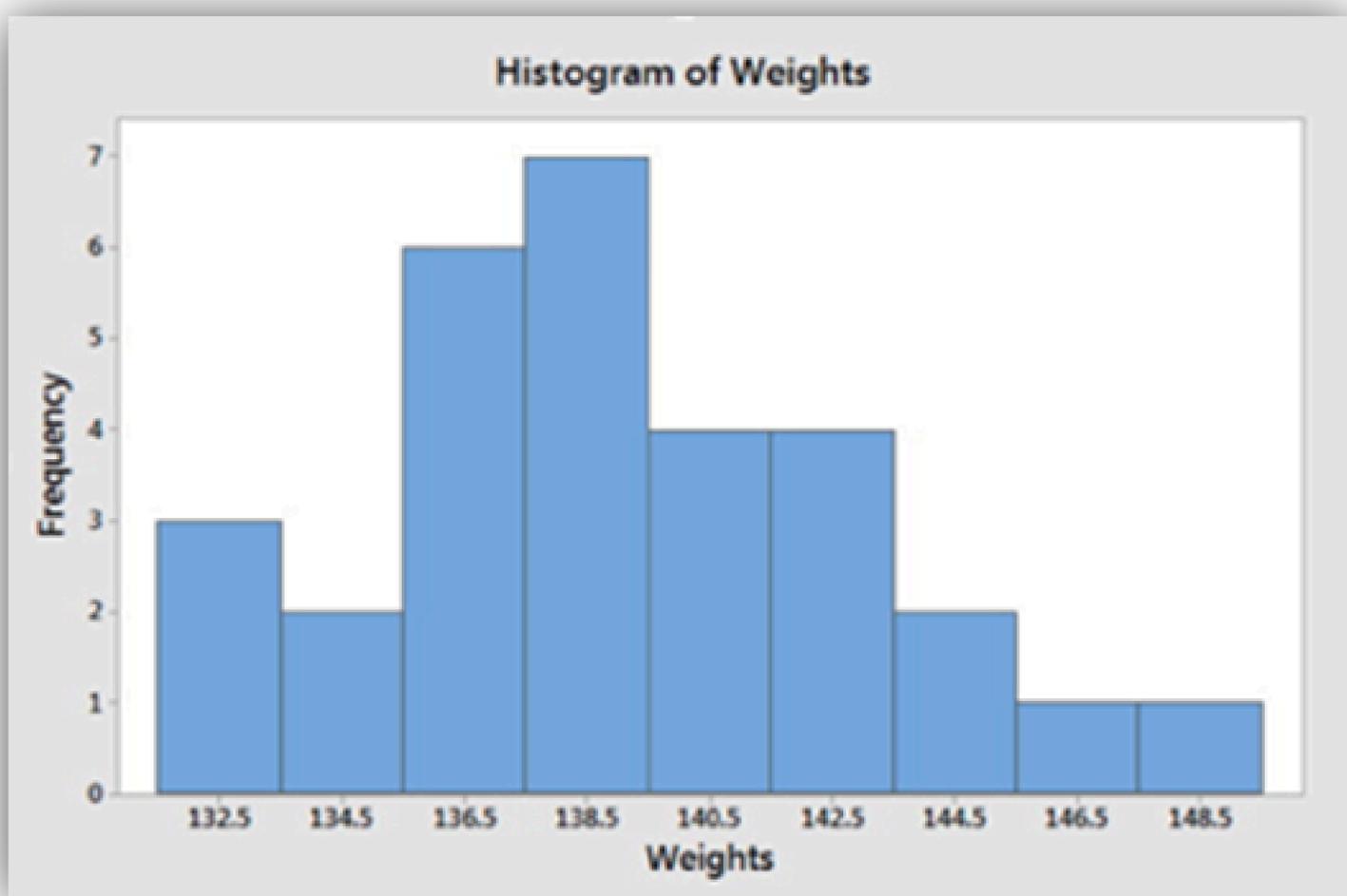
Representaciones gráficas (histogramas, diagramas de barras, diagramas de dispersión).

Las representaciones gráficas son herramientas visuales fundamentales en estadística para comunicar patrones, tendencias y relaciones en los datos.

- **Histogramas:** Los histogramas son gráficos de barras que muestran la distribución de frecuencia de un conjunto de datos numéricos continuos. En un histograma, el eje horizontal representa las diferentes categorías o intervalos de valores, mientras que el eje vertical muestra la frecuencia o la densidad de observaciones en cada intervalo. Los histogramas son útiles para visualizar la forma y la simetría de la distribución de los datos, así como para identificar cualquier sesgo o asimetría.

Figura 3

Representación de histograma



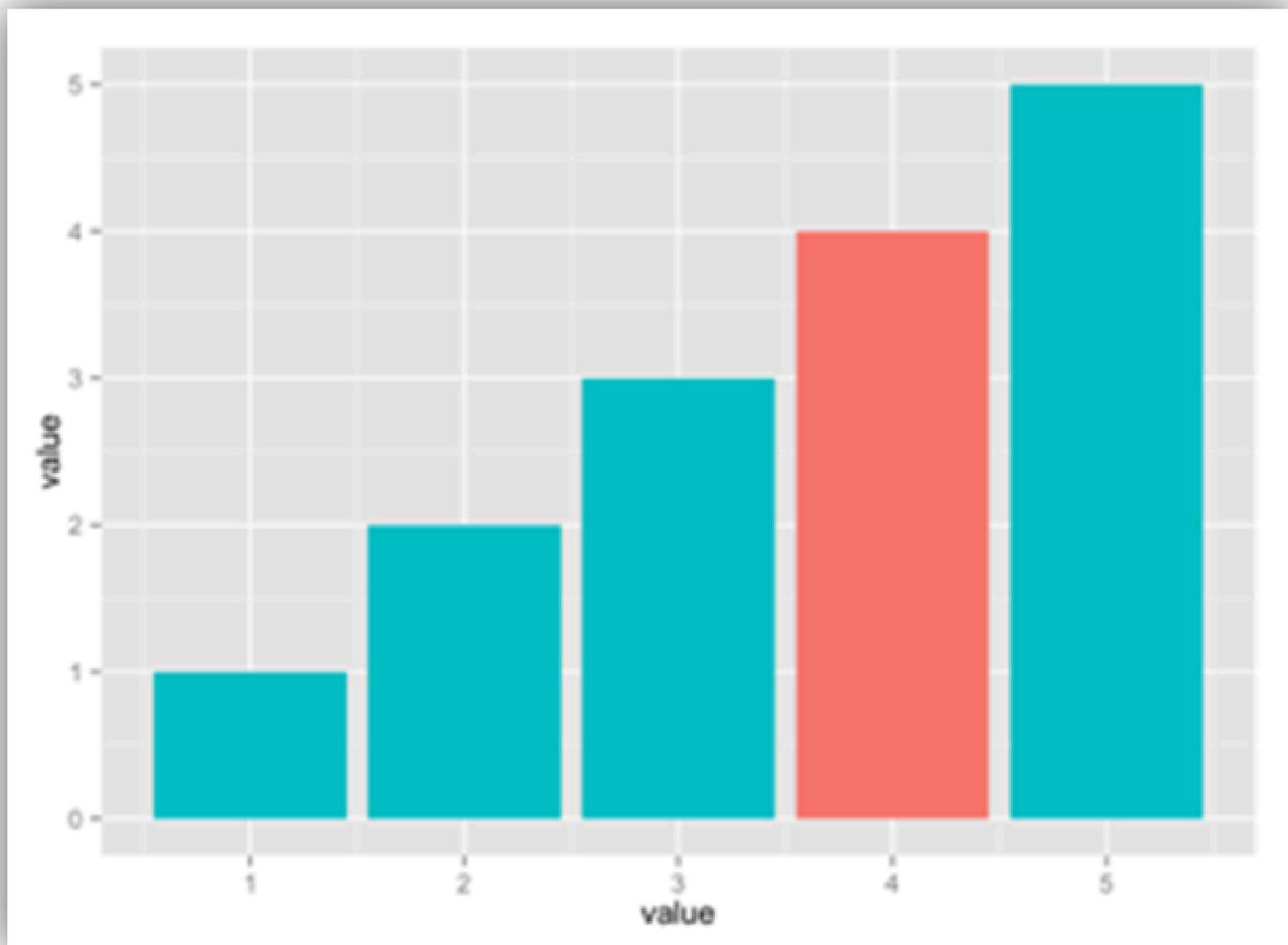
Nota: La figura muestra un ejemplo de un histograma. Tomado de Pennstate University (s.f.)

Diagramas de barras: Los diagramas de barras son gráficos que muestran la relación entre variables categóricas mediante barras rectangulares de longitudes proporcionales a las frecuencias o porcentajes de cada categoría. Los diagramas de barras son útiles para comparar diferentes categorías entre sí y

para mostrar cambios en las frecuencias a lo largo del tiempo o en diferentes grupos. Son comunes en encuestas, estudios de mercado y análisis de datos cualitativos.

Figura 4

Representacion de diagrama de barra

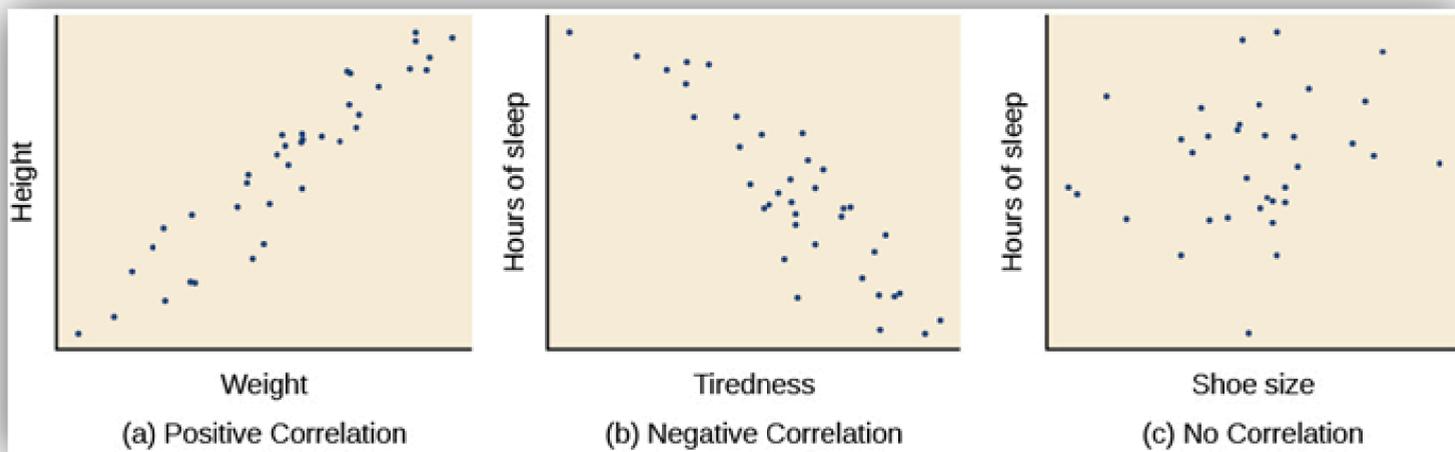


Nota: La figura muestra una representación del diagrama de barra. Tomado de Stackoverflow (2022)

Diagramas de dispersión: Los diagramas de dispersión son gráficos que muestran la relación entre dos variables numéricas mediante puntos en un plano cartesiano, donde cada punto representa una observación en el conjunto de datos. Los diagramas de dispersión son útiles para identificar patrones de asociación, tendencias o correlaciones entre las variables. También pueden ayudar a identificar valores atípicos o anomalías en los datos. Los diagramas de dispersión son comunes en campos como la ciencia, la ingeniería y la investigación social.

Figura 5

Representación de diagrama de dispersión



Nota: La figura muestra ejemplos de diagramas de dispersión.
Tomado de Pennstate University (s.f.)

Autoevaluación

1. ¿Qué medida de tendencia central es más sensible a valores atípicos en un conjunto de datos?

- a) Media
- b) Mediana
- c) Moda
- d) Varianza

2. ¿Qué medida de dispersión indica cuánto se desvían los valores individuales del promedio en un conjunto de datos?

- a) Rango
- b) Desviación estándar
- c) Varianza
- d) Media

3. ¿Cuál de las siguientes representaciones gráficas es más adecuada para datos numéricos continuos?

- a) Histogramas
- b) Diagramas de barras
- c) Diagramas de dispersión
- d) Moda

4. ¿Qué medida de tendencia central es el valor que ocurre con mayor frecuencia en un conjunto de datos?

- a) Media
- b) Mediana
- c) Moda
- d) Rango

5. ¿Cuál es una característica clave de la distribución normal?

- a) Es simétrica alrededor de su media
- b) Solo tiene dos posibles resultados
- c) Es sensible a valores atípicos
- d) No tiene parámetros

6. ¿Cuál de las siguientes medidas de tendencia central es más afectada por valores extremos en un conjunto de datos?

- a) Mediana
- b) Moda
- c) Media
- d) Rango

7. ¿Qué medida de dispersión es el cuadrado de la desviación estándar?

- a) Rango
- b) Varianza
- c) Media
- d) Moda

8. ¿Cuál de las siguientes representaciones gráficas es más adecuada para mostrar la relación entre dos variables numéricas?

- a) Histogramas
- b) Diagramas de barras
- c) Diagramas de dispersión
- d) Mediana

9. ¿Qué medida de tendencia central se utiliza para describir el valor típico o central de un conjunto de datos cuando los valores están distribuidos de manera relativamente uniforme?

- a) Media
- b) Mediana
- c) Moda
- d) Rango

10. ¿Cuál de las siguientes distribuciones de probabilidad es más adecuada para modelar experimentos de Bernoulli?

- a) Distribución normal
- b) Distribución binomial
- c) Distribución de Poisson
- d) Distribución uniforme

Unidad 3: Probabilidad

Concepto de probabilidad

Probabilidad, uno de los conceptos fundamentales que se aborda es el de probabilidad. La probabilidad es una medida numérica que describe la posibilidad de que ocurra un evento particular. Se expresa como un número entre 0 y 1, donde 0 indica que el evento es imposible de ocurrir, 1 indica que el evento es seguro de ocurrir, y los valores intermedios representan grados de incertidumbre sobre la ocurrencia del evento.

En términos más formales, la probabilidad se define como la proporción de veces que un evento ocurre en relación con el número total de posibles resultados en un experimento o situación. Se puede calcular utilizando diferentes enfoques, dependiendo de la naturaleza del experimento y de la información disponible. Por ejemplo, la probabilidad teórica se calcula utilizando principios matemáticos y modelos, mientras que la probabilidad empírica se calcula a partir de la frecuencia relativa de ocurrencia de un evento en un conjunto de datos observados.

La probabilidad se aplica en una amplia variedad de áreas, incluyendo la estadística, la ciencia, la ingeniería, la economía y muchos otros campos. Es fundamental para la toma de decisiones informadas, el diseño de experimentos, la predicción de resultados y la evaluación de riesgos en diversas situaciones.

Eventos y espacio muestral

En el contexto de la teoría de la probabilidad (Galindo de la Torre, 1999), es fundamental comprender los conceptos de eventos y espacio muestral.

- **Evento:**

Un evento es un resultado o conjunto de resultados observables que pueden ocurrir como resultado de un experimento aleatorio o

una situación probabilística. Puede ser cualquier resultado singular o una combinación de resultados.

Por ejemplo, si lanzamos un dado, los eventos posibles podrían ser obtener un 1, un 2, un 3, etc., o incluso obtener un número par. Los eventos se suelen denotar con letras mayúsculas, como A, B, C, etc.

- **Espacio muestral:**

El espacio muestral es el conjunto de todos los resultados posibles de un experimento aleatorio. Es decir, es el conjunto que contiene todos los eventos elementales o resultados individuales que podrían ocurrir.

Por ejemplo, si lanzamos un dado de seis caras, el espacio muestral sería $\{1, 2, 3, 4, 5, 6\}$, ya que estos son todos los posibles resultados de lanzar el dado. El espacio muestral se denota comúnmente con la letra griega Ω (omega) o con la letra S.

Es importante destacar que el espacio muestral abarca todos los resultados posibles, mientras que un evento es simplemente una selección específica de estos resultados. Además, un evento puede ser cualquier subconjunto del espacio muestral, incluido el espacio muestral completo y el conjunto vacío.

Comprender los conceptos de eventos y espacio muestral es esencial para analizar y calcular probabilidades en situaciones donde se llevan a cabo experimentos aleatorios. Estos conceptos proporcionan la base para el cálculo de probabilidades y la interpretación de resultados en el contexto de la teoría de la probabilidad.

Reglas de probabilidad (suma, producto, complemento)

Las reglas de probabilidad son principios fundamentales que rigen

el cálculo y la interpretación de probabilidades en el contexto de la teoría de la probabilidad. Entre las reglas más importantes se encuentran las siguientes:

Regla de la suma (o regla de la adición):

Esta regla establece que la probabilidad de que ocurra al menos uno de dos eventos mutuamente excluyentes es igual a la suma de las probabilidades individuales de cada evento. Matemáticamente, para dos eventos A y B mutuamente excluyentes, la regla de la suma se expresa como:

Adición

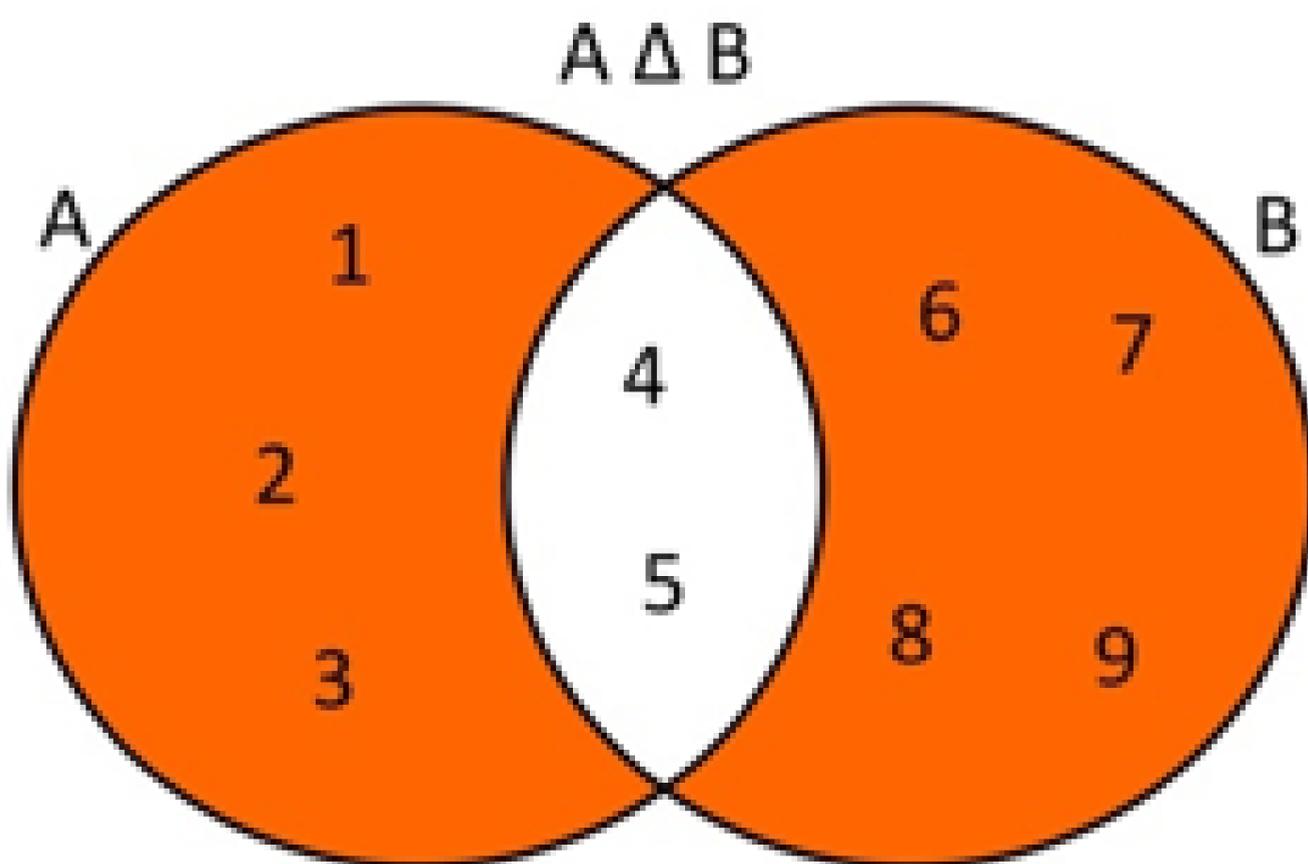
$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

Donde:

- $\Pr(A \cup B)$ es la probabilidad de que ocurra al menos uno de los eventos A o B.
- $\Pr(A)$ es la probabilidad del evento A.
- $\Pr(B)$ es la probabilidad del evento B.

Figura 6

Representación de diagrama de regla de suma



Producto

$$\Pr(A \cap B) = \Pr(A) * \Pr(B)$$

Regla del producto:

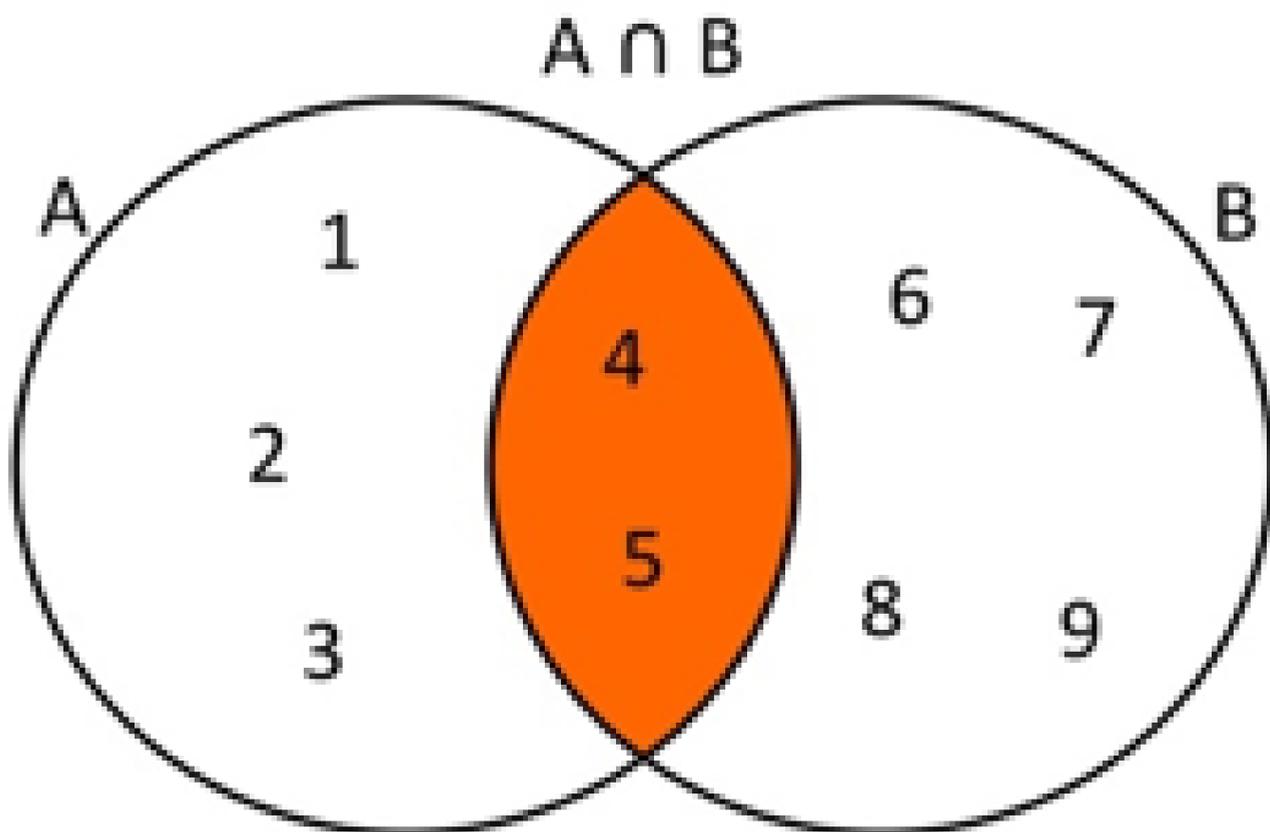
Esta regla se aplica cuando dos eventos son independientes entre sí. Establece que la probabilidad de que ocurran ambos eventos es igual al producto de sus probabilidades individuales. Matemáticamente, para dos eventos independientes A y B, la regla del producto se expresa como:

Donde:

- $Pr(A \cap B)$ es la probabilidad de que ocurran ambos eventos A y B.
- $Pr(A)$ es la probabilidad del evento A.
- $Pr(B)$ es la probabilidad del evento B.

Figura 7

Representación de diagrama de regla de producto



Nota: La figura muestra la representación de un diagrama para la regla de producto.

Regla del complemento:

Esta regla establece que la probabilidad del evento complementario de un evento A es igual a la resta de 1 menos la probabilidad del evento A. Es decir, la probabilidad de que no

ocurra el evento A. Matemáticamente, para un evento A, la regla del complemento se expresa como:

$$\text{Complemento}$$

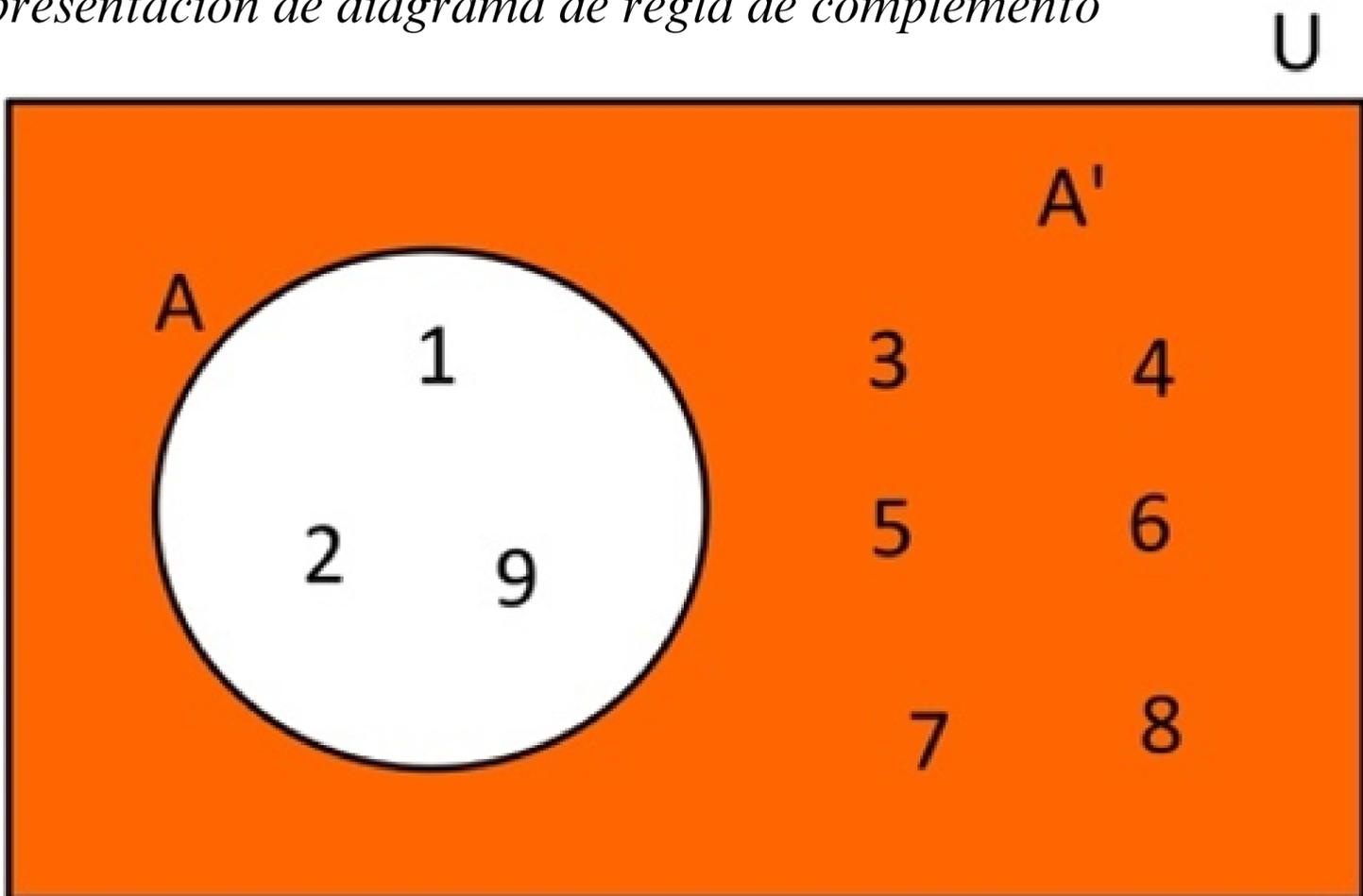
$$\Pr(\bar{A}) = 1 - \Pr(A)$$

Donde:

- $\Pr(A)$ es la probabilidad del evento complementario de A.
- $\Pr(\bar{A})$ es la probabilidad del evento A.

Figura 8

Representación de diagrama de regla de complemento



Nota: La figura muestra la representación de un diagrama para la regla de complemento.

Estas reglas son esenciales para calcular y comprender las probabilidades de diferentes eventos y para resolver problemas de probabilidad en una amplia variedad de contextos.

Distribuciones de probabilidad (binomial, normal)

Las distribuciones de probabilidad son modelos matemáticos que describen la probabilidad de ocurrencia de cada resultado posible en un experimento aleatorio o proceso estocástico. Dos de las distribuciones de probabilidad más importantes y ampliamente

utilizadas son la distribución binomial y la distribución normal:

- **Distribución binomial:**

La distribución binomial modela la probabilidad de obtener un número específico de éxitos en un número fijo de ensayos independientes, donde cada ensayo tiene solo dos posibles resultados: éxito o fracaso. Las características clave de la distribución binomial son:

- Cada ensayo es independiente. Solo hay dos resultados posibles en cada ensayo (éxito o fracaso).
- La probabilidad de éxito (p) es constante en todos los ensayos.
- El número total de ensayos (n) es fijo.

La función de masa de probabilidad (PMF) de la distribución binomial se define por la fórmula:

Distribución Binomial

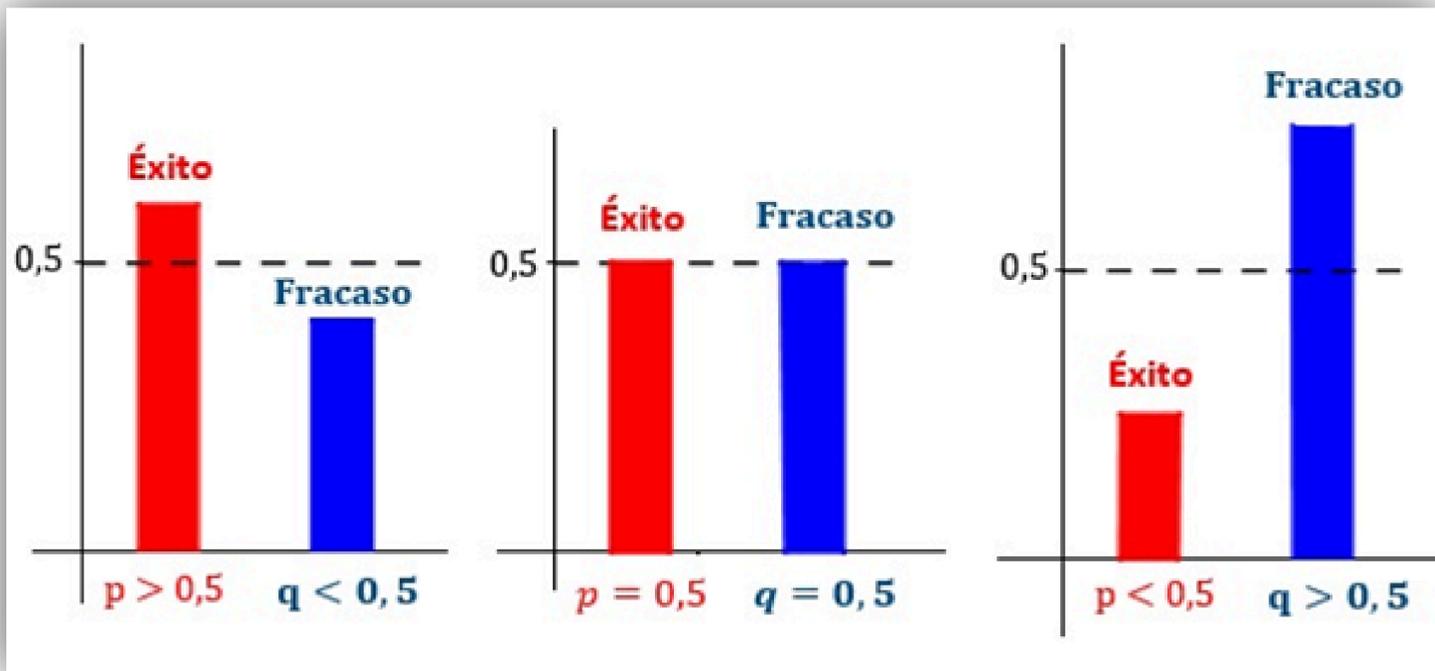
$$\Pr(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$$

Donde:

- $\Pr(X = k)$ es la probabilidad de exactamente k éxitos en n ensayos.
- $\binom{n}{k}$ es el coeficiente binomial que representa el número de formas de elegir k éxitos de n ensayos.
- p es la probabilidad de éxito en un solo ensayo.
- $1-p$ es la probabilidad de fracaso en un solo ensayo.
- n es el número total de ensayos.

Figura 9

Representación de distribución binomial.



Nota: La figura muestra la representación de un diagrama con distribución binomial.

Distribución normal (o gaussiana):

La distribución normal (Alvarado Martínez, 2018) es una de las distribuciones más importantes en estadística. Se caracteriza por ser una distribución de campana simétrica en forma de una curva suave y continua. Las características clave de la distribución normal son:

- Es simétrica alrededor de su media.
- La mayor parte de los datos se concentran cerca de la media, y la dispersión de los datos disminuye a medida que nos alejamos de la media.
- Está completamente determinada por dos parámetros: la media (μ) y la desviación estándar (σ).

La función de densidad de probabilidad (PDF) de la distribución normal se define por la fórmula:

$$f(x) = \frac{1}{\sqrt{2 * \pi * \sigma^2}} * e^{-\frac{(x-\mu)}{2\sigma^2}}$$

Donde:

- $f(x)$ es la densidad de probabilidad en el punto x . μ es la media de la distribución.
- σ es la desviación estándar de la distribución.
- e es la base del logaritmo natural (aproximadamente 2.71828).

La distribución binomial es apropiada para modelar experimentos de Bernoulli, mientras que la distribución normal es útil para describir fenómenos donde los datos tienden a distribuirse simétricamente alrededor de una media y donde la variabilidad es continua y suavizada. Ambas distribuciones son fundamentales en estadística y tienen una amplia gama de aplicaciones en diversos campos.

Autoevaluación

1. ¿Cómo se define la probabilidad en el contexto de la teoría de la probabilidad?

- a) Como la suma de los eventos mutuamente excluyentes.
- b) Como la proporción de veces que un evento ocurre en relación con el número total de posibles resultados.
- c) Como el complemento de un evento.
- d) Como el producto de dos eventos independientes.

2. ¿Qué es un evento en el contexto de la teoría de la probabilidad?

- a) Un conjunto de resultados observables.
- b) La media de un conjunto de datos.
- c) El espacio muestral.
- d) Una medida de dispersión.

3. ¿Cuál de las siguientes reglas de probabilidad establece que la probabilidad de que ocurran ambos eventos es igual al producto de sus probabilidades individuales?

- a) Regla de la suma.
- b) Regla del producto.
- c) Regla del complemento.
- d) Regla de la adición.

4. ¿Cuál es una característica clave de la distribución normal?

- a) Es simétrica alrededor de su media.
- b) Solo tiene dos posibles resultados.
- c) Es sensible a valores atípicos.
- d) Es una distribución binomial.

5. ¿Cuál es la función de masa de probabilidad (PMF) de la distribución binomial?

a) $\Pr(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

b) $\Pr(\bar{A}) = 1 - \Pr(A)$

c) $f(x) = \frac{1}{\sqrt{2*\pi*\sigma^2}} * e^{-\frac{(x-\mu)}{2\sigma^2}}$

d) $\Pr(A \cap B) = \Pr(A) * \Pr(B)$

6. ¿Cuál es la definición de probabilidad en términos formales en el contexto de la teoría de la probabilidad?

- a) La suma de los eventos mutuamente excluyentes.
- b) La proporción de veces que un evento ocurre en relación con el número total de posibles resultados.
- c) El complemento de un evento.
- d) El producto de dos eventos independientes.

7. ¿Cuál es la función de densidad de probabilidad (PDF) de la distribución normal?

a) $\Pr(X = k) = \binom{n}{k} * p^k * (1 - p)^{n-k}$

b) $\Pr(\bar{A}) = 1 - \Pr(A)$

c) $f(x) = \frac{1}{\sqrt{2*\pi*\sigma^2}} * e^{-\frac{(x-\mu)}{2\sigma^2}}$

d) $\Pr(A \cap B) = \Pr(A) * \Pr(B)$

8. ¿Cuál de las siguientes reglas de probabilidad establece que la probabilidad del evento complementario de A es igual a 1 menos la probabilidad de A?

- a) Regla de la suma.
- b) Regla del producto.
- c) Regla del complemento.
- d) Regla de la adición.

9. ¿Qué es un espacio muestral en el contexto de la teoría de la probabilidad?

- a) Un conjunto de resultados observables.
- b) La media de un conjunto de datos.
- c) El conjunto de todos los resultados posibles de un experimento aleatorio.
- d) Una medida de dispersión.

10. ¿Cuál de las siguientes distribuciones de probabilidad es adecuada para modelar la probabilidad de obtener un número específico de éxitos en un número fijo de ensayos independientes?

- a) Distribución normal
- b) Distribución binomial
- c) Distribución de Poisson
- d) Distribución uniforme

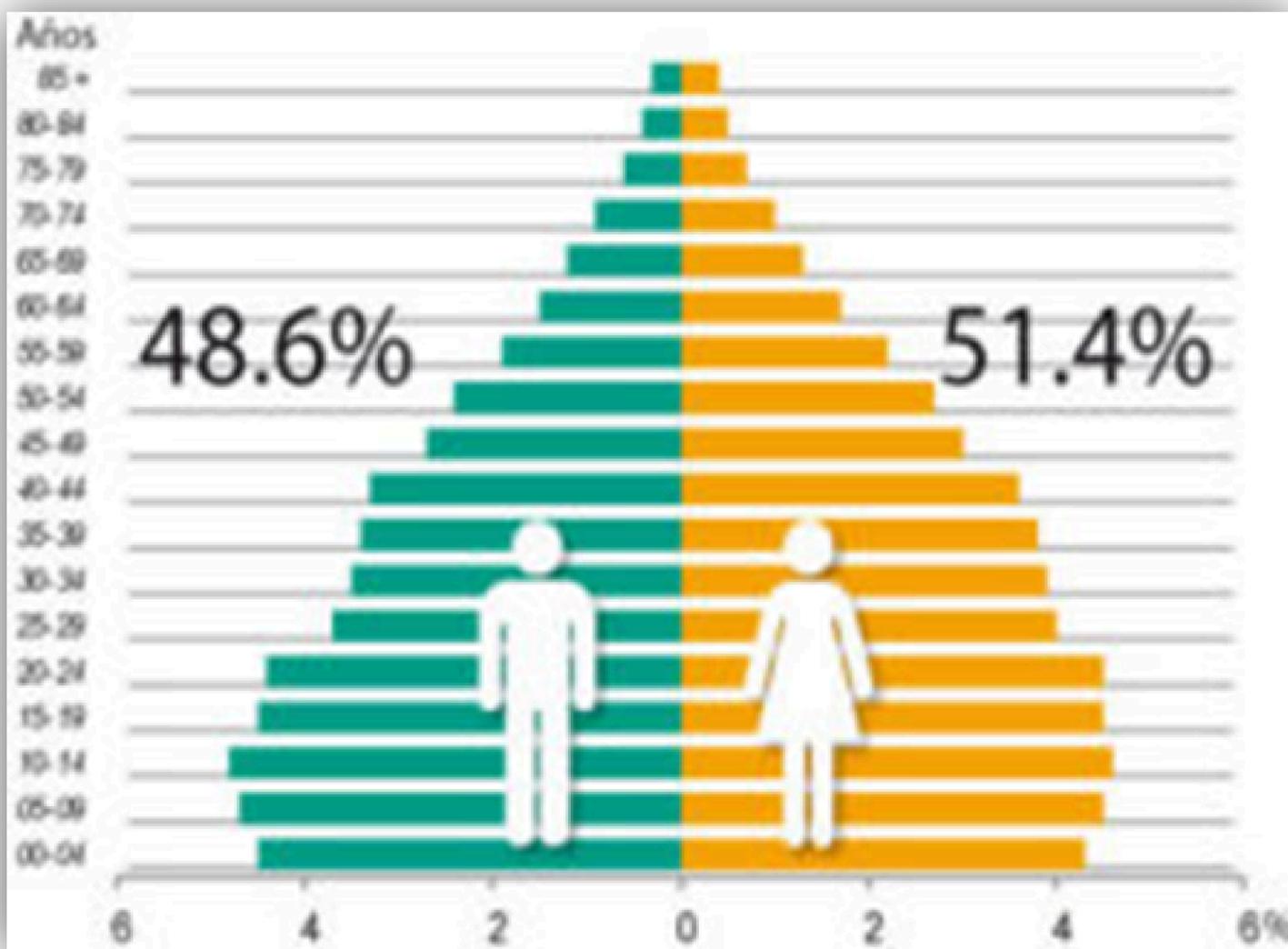
Unidad 4: Población, Muestreo y distribuciones muestrales

Población

En estadística, el término "población" se refiere al conjunto completo de elementos o individuos que comparten una característica específica y sobre los cuales se desea realizar inferencias o estudiar. Esta característica puede ser cualquier atributo o cualidad que se esté investigando. La población puede ser finita o infinita en tamaño y puede variar en su composición dependiendo del contexto del estudio. En el análisis estadístico, la población sirve como la base teórica para la cual se intenta generalizar los resultados obtenidos a través de la recolección y análisis de datos de una muestra, que es un subconjunto representativo de la población.

Figura 10

Representación de pirámide poblacional



Nota: La figura muestra un ejemplo que representa la pirámide poblacional de México. Tomado de Palacio Nava (2015).

En Ecuador el Instituto Nacional de Estadísticas y Censos es el organismo competente para facilitar los datos poblaciones nacionales y otras estadísticas oficiales considerando otros organismos del sector privado, público y social como fuentes de información. Como antecedentes de población tenemos registros de censos de los años 1950, 1962, 1974, 1982, 1990, 2001, 2010, y 2022.

El último Censo 2022 proyecta una población para Ecuador de 17`677.303 personas [1], así también, reveló la caída de la Tasa Global de Fecundidad (TGF) desde 1950 a 2022 con una caída de 6,83 a 1,86 respectivamente. En ese sentido también se revela un aplazamiento de la fecundidad, lo que significa que actualmente las mujeres tienen su primer hijo aproximadamente a partir de los 28 años (antes 21 años[2]).

Este fenómeno de contracción de la población revela elementos claves para Ecuador como: menos matrimonio, aumento de hogares sin hijos o unipersonales, mayor autonomía femenina, equidad de género en el mercado laboral, mayor nivel de escolaridad, incremento de urbanización entre otros elementos claves del último censo 2022.

Métodos de muestreo (aleatorio, estratificado, por conglomerados)

Los métodos de muestreo son técnicas utilizadas en estadística para seleccionar una muestra representativa de una población más grande. Cada método de muestreo tiene sus propias características y se elige en función de la naturaleza de la población y el objetivo del estudio.

[1] Proyecciones de población y omisión censal revisión 2023, INEC Febrero 2024.

[2] Tasa específica de fecundidad 2010 era de 24 años, Censo 2010, INEC.

Muestreo aleatorio:

En el muestreo aleatorio, cada elemento de la población tiene la misma probabilidad de ser seleccionado para formar parte de la muestra. Esto se logra a través de un proceso de selección aleatoria, donde cada unidad de la población tiene una oportunidad igual de ser incluida en la muestra.

Este método es especialmente útil cuando se desea evitar el sesgo y garantizar que la muestra sea representativa de la población en su conjunto. El muestreo aleatorio puede realizarse con o sin reemplazo, dependiendo de si los elementos seleccionados se devuelven a la población después de ser seleccionados.

Muestreo estratificado:

El muestreo estratificado implica dividir la población en subgrupos homogéneos o estratos basados en ciertas características relevantes. Luego, se selecciona una muestra aleatoria de cada estrato de manera independiente.

Este método es útil cuando la población presenta una gran variabilidad y se desea asegurar que cada subgrupo esté representado adecuadamente en la muestra final. Al estratificar la población, se puede mejorar la precisión de las estimaciones al reducir la variabilidad dentro de cada estrato y garantizar una representación proporcional de cada segmento de interés.

Muestreo por conglomerados:

En el muestreo por conglomerados, la población se divide en grupos o conglomerados naturales o predefinidos, como ciudades, escuelas o empresas. Luego, se seleccionan aleatoriamente algunos de estos conglomerados y se toman muestras de todos los elementos dentro de los conglomerados seleccionados. Este método es útil cuando no es factible o práctico seleccionar una

muestra directamente de la población completa debido a su dispersión geográfica o a la falta de una lista completa de elementos. El muestreo por conglomerados puede reducir los costos y el tiempo de recolección de datos al agrupar elementossimilares en unidades más grandes.

Cada método de muestreo tiene sus ventajas y limitaciones, y es importante seleccionar el método más adecuado según las características específicas de la población y los objetivos del estudio.

Estimación de parámetros poblacionales

La estimación de parámetros poblacionales es un proceso fundamental en estadística que implica inferir o predecir características desconocidas de una población más amplia basándose en la información recopilada de una muestra representativa de esa población. Los parámetros poblacionales son medidas numéricas que describen alguna característica específica de interés en la población en su conjunto.

El objetivo de la estimación de parámetros poblacionales es obtener una idea lo más precisa posible de los valores verdaderos de estos parámetros a partir de la información limitada proporcionada por la muestra. Para lograr esto, se utilizan diferentes técnicas estadísticas, como la estimación puntual y la estimación por intervalos.

Estimación puntual:

En la estimación puntual, se calcula un único valor que se considera una buena aproximación del parámetro poblacional de interés. Este valor se basa en los datos observados en la muestra y puede tomar la forma de la media muestral, la proporción muestral, la mediana muestral u otras medidas estadísticas. Sin

embargo, la estimación puntual no proporciona información sobre la precisión o incertidumbre asociada con el valor estimado.

Estimación por intervalos:

La estimación por intervalos, también conocida como estimación intervalar o estimación de intervalo de confianza, proporciona un rango de valores dentro del cual se espera que se encuentre el verdadero valor del parámetro poblacional con cierto nivel de confianza. Este intervalo se construye utilizando la información de la muestra y tiene en cuenta tanto la estimación puntual como la variabilidad de los datos. El nivel de confianza suele expresarse como un porcentaje, como el 95% o el 99%.

La elección entre la estimación puntual y la estimación por intervalos depende del grado de precisión y certeza que se requiere en la inferencia sobre el parámetro poblacional. Mientras que la estimación puntual proporciona un valor único y fácilmente interpretable, la estimación por intervalos ofrece una medida de la incertidumbre asociada con la estimación y permite tomar decisiones más informadas sobre la población en cuestión.

La estimación de parámetros poblacionales es un proceso esencial en la estadística que permite hacer inferencias sobre una población más amplia basándose en la información limitada proporcionada por una muestra representativa. La elección de la técnica de estimación adecuada depende de los objetivos del estudio y del nivel de precisión requerido en la inferencia estadística.

Distribuciones muestrales (distribución muestral de la media, distribución muestral de la proporción)

Las distribuciones muestrales son un concepto fundamental en estadística que se refiere a la distribución de una estadística de

interés (como la media o la proporción) calculada a partir de múltiples muestras independientes tomadas de una población en particular. Dos de las distribuciones muestrales más comunes son la distribución muestral de la media y la distribución muestral de la proporción.

Distribución muestral de la media:

La distribución muestral de la media describe la distribución de todas las posibles medias muestrales que podrían obtenerse al tomar muestras de tamaño (n) de una población. La forma de esta distribución muestral depende del tamaño de la muestra (n) y de la distribución de la población subyacente. Según el Teorema del Límite Central, cuando el tamaño de la muestra es lo suficientemente grande (generalmente $n \geq 30$) y las muestras se toman de forma aleatoria y independiente, la distribución muestral de la media se aproxima a una distribución normal, independientemente de la forma de la población original.

La media de esta distribución muestral es igual a la media poblacional (μ), y la desviación estándar (σ) de esta distribución muestral, conocida como error estándar de la media, se calcula como $\frac{\sigma}{n}$.

Distribución muestral de la proporción:

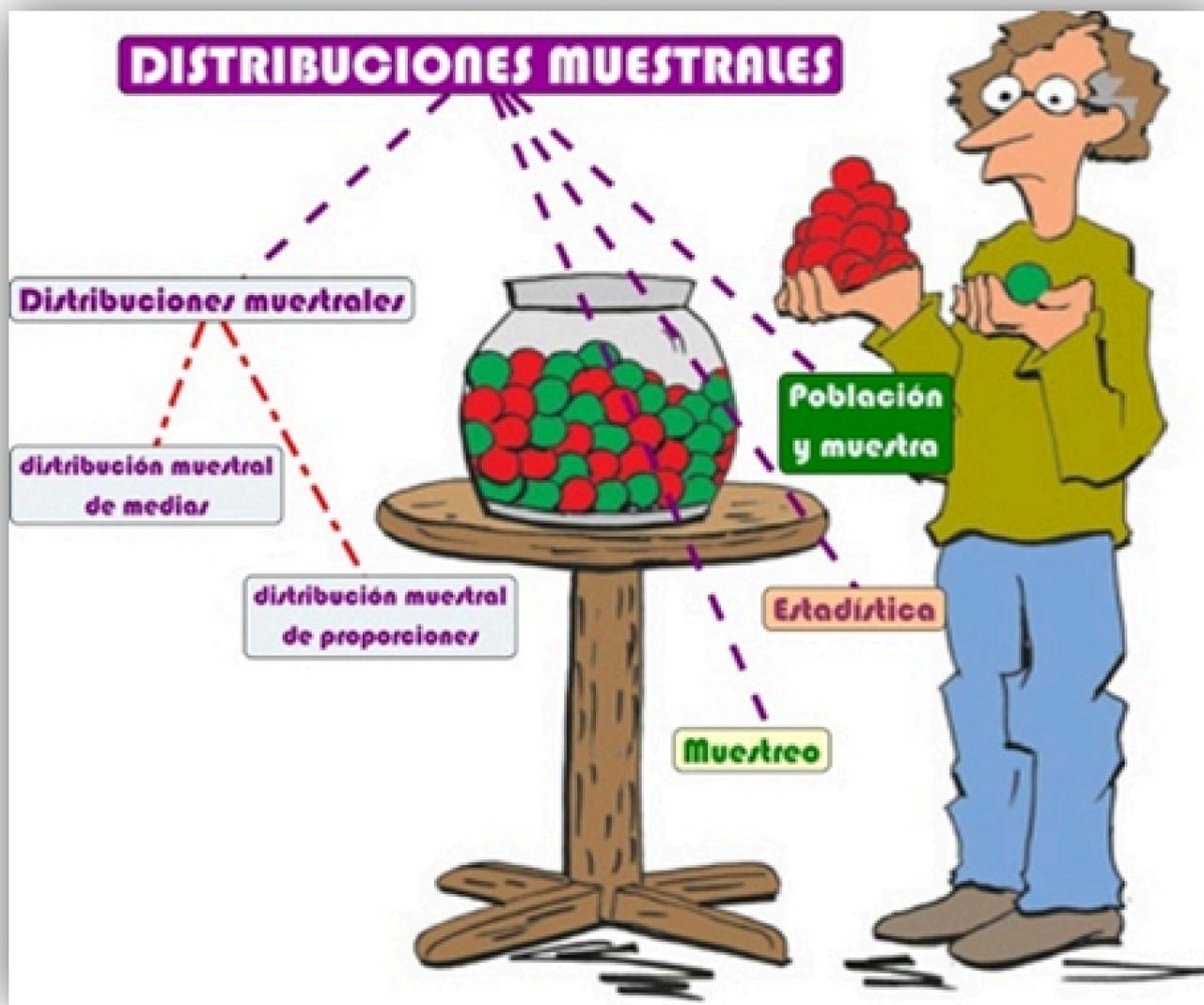
La distribución muestral de la proporción describe la distribución de todas las posibles proporciones muestrales que podrían obtenerse al tomar muestras de tamaño (n) de una población binaria (es decir, una población con dos resultados posibles, como éxito o fracaso).

Esta distribución muestral se utiliza comúnmente en el contexto de pruebas de proporciones, como pruebas de hipótesis para comparar proporciones entre dos grupos. Cuando el tamaño de la

muestra es lo suficientemente grande y las condiciones de independencia y aleatoriedad se cumplen, la distribución muestral de la proporción se aproxima a una distribución normal, con una media igual a la proporción poblacional (p) y una desviación estándar $\sqrt{\frac{p*(1-p)}{n}}$.

Figura 11

Representación de distribución muestral



Nota: La figura muestra un ejemplo de distribuciones muestrales. Tomado de Victormat (s.f.)

Estas distribuciones muestrales son fundamentales en inferencia estadística, ya que permiten realizar inferencias sobre los parámetros poblacionales basadas en las muestras observadas. Además, proporcionan herramientas importantes para la construcción de intervalos de confianza y la realización de pruebas de hipótesis, lo que permite tomar decisiones informadas en una amplia variedad de contextos de investigación y análisis de datos.

Autoevaluación

1. ¿Qué método de muestreo implica dividir la población en subgrupos homogéneos y luego seleccionar muestras aleatorias de cada subgrupo?

- a) Muestreo aleatorio
- b) Muestreo estratificado
- c) Muestreo por conglomerados
- d) Muestreo secuencial

2. ¿Cuál de los siguientes métodos de muestreo selecciona aleatoriamente algunos conglomerados y luego toma muestras de todos los elementos dentro de los conglomerados seleccionados?

- a) Muestreo aleatorio
- b) Muestreo estratificado
- c) Muestreo por conglomerados
- d) Muestreo secuencial

3. En la distribución muestral de la media, ¿cuál es la relación entre el tamaño de la muestra (n) y la desviación estándar de la distribución muestral (σ)?

- a) σ/n
- b) $\sigma*n$
- c) $\sigma-n$
- d) $\sigma+n$

4. ¿Cuál es la distribución que describe la distribución de todas las posibles proporciones muestrales que podrían obtenerse al tomar muestras de tamaño (n) de una población binaria?

- a) Distribución normal
- b) Distribución de Poisson
- c) Distribución t de Student
- d) Distribución muestral de la proporción

5. ¿Qué técnica de estimación proporciona un rango de valores dentro del cual se espera que se encuentre el verdadero valor del parámetro poblacional con cierto nivel de confianza?

- a) Estimación puntual
- b) Estimación por intervalos
- c) Estimación secuencial
- d) Estimación acumulativa

6. ¿En qué método de muestreo se seleccionan muestras de cada estrato de manera independiente?

- a) Muestreo aleatorio
- b) Muestreo estratificado
- c) Muestreo por conglomerados
- d) Muestreo secuencial

7. ¿Qué distribución describe la distribución de todas las posibles medias muestrales que podrían obtenerse al tomar muestras de tamaño (n) de una población?

- a) Distribución normal
- b) Distribución binomial
- c) Distribución t de Student
- d) Distribución muestral de la media

8. ¿Qué método de estimación proporciona un único valor considerado una buena aproximación del parámetro poblacional de interés?

- a) Estimación puntual
- b) Estimación por intervalos
- c) Estimación secuencial
- d) Estimación acumulativa

9. En la distribución muestral de la proporción, ¿cuál es la desviación estándar?

- a) $\sqrt{p*(1-p)}$
- b) $\sqrt{p/n}$
- c) σ/n
- d) \sqrt{n}

10. ¿Qué método de muestreo implica seleccionar muestras de cada estrato de manera independiente?

- a) Muestreo aleatorio
- b) Muestreo estratificado
- c) Muestreo por conglomerados
- d) Muestreo secuencial

Solucionario de las Autoevaluaciones.

UNIDAD 1	
Pregunta	Respuesta
1)	c
2)	b
3)	d
4)	b
5)	d
6)	b
7)	c
8)	c
9)	c
10)	c

UNIDAD 3	
Pregunta	Respuesta
1)	b
2)	a
3)	b
4)	a
5)	a
6)	b
7)	c
8)	c
9)	c
10)	b



UNIDAD 2

Pregunta	Respuesta
1)	b
2)	b
3)	a
4)	c
5)	a
6)	c
7)	b
8)	c
9)	a
10)	b

UNIDAD 4

Pregunta	Respuesta
1)	b
2)	c
3)	a
4)	d
5)	b
6)	b
7)	d
8)	a
9)	a
10)	b

Referencias en la guía

Alvarado Martínez, H. A. (2018). Evaluación del aprendizaje de la estadística orientada a proyectos en estudiantes de ingeniería. Educación matemática, 30(3), 151-183.

B., O. (1973). Estadística Aplicada. Limusa.

Fernández, S. F. (2002). Estadística descriptiva. Esic Editorial.

Galindo de la Torre, E. (1999). Estadística para la Administración y la Ingeniería.

Jain, D. (23 de agosto de 2018). File:Relationship between mean and median under different skewness.png. Obtenido de https://en.m.wikipedia.org/wiki/File:Relationship_between_mean_and_median_under_different_skewness.png

Palacios Nava, M. E. (2015). Population Pyramid Mexico 2015. Obtenido de https://www.researchgate.net/figure/Population-Pyramid-Mexico-2015_fig1_327354521

Pennstate University. (s.f.). Linear Relationships. Obtenido de <https://online.stat.psu.edu/stat500/book/export/html/587>

Ross, S. M. (2018). Introducción a la estadística. Reverté.

Stackoverflow. (junio de 2022). Pandas/matplotlib bar chart with colors defined by column. Obtenido de <https://stackoverflow.com/questions/26793165/pandas-matplotlib-bar-chart-with-colors-defined-by-column>

Triola, M. (2004). Estadística. Person.

Victormat. (s.f.). Distribuciones muestrales. Obtenido de <https://www.victormat.es/mcs2/Tema12-DistribucionesMuestrales/index.html>

Bibliografía para la clase

Bibliografía básica

B., O. (1973). Estadística Aplicada. Limusa.

Triola, M. (2004). Estadística. Person.

Fernández, S. F. (2002). Estadística descriptiva. Esic Editorial.

Ross, S. M. (2018). Introducción a la estadística. Reverté.

Galindo de la Torre, E. (1999). Estadística para la Administración y la Ingeniería.

Alvarado Martínez, H. A., Galindo Illanes, M. K., & Retamal Pérez, M. L. (2018). Evaluación del aprendizaje de la estadística orientada a proyectos en estudiantes de ingeniería. Educación matemática, 30(3), 151-183.
B., O. (1973). Estadística Aplicada. Limusa.

Triola, M. (2004). Estadística. Person.

Fernández, S. F. (2002). Estadística descriptiva. Esic Editorial.

Ross, S. M. (2018). Introducción a la estadística. Reverté.

Galindo de la Torre, E. (1999). Estadística para la Administración y la Ingeniería.

Alvarado Martínez, H. A., Galindo Illanes, M. K., & Retamal Pérez, M. L. (2018). Evaluación del aprendizaje de la estadística orientada a proyectos en estudiantes de ingeniería. Educación matemática, 30(3), 151-183.

Bibliografía complementaria

*Proyecciones de población y omisión censal revisión 2023, INEC
Febrero 2024.*

Detalle	Nombre	Firma
Elaborado por:	Ing. Juan Pablo Bustos Rivera	
Revisado por:	Ing. Edison Torres	
Aprobado por:	Mtr. Oscar Toscano	

INSTITUTO SUPERIOR TECNOLÓGICO **SUDAMERICANO**

Quito, Ecuador



ISBN: 978-9942-7209-9-3

