

Análisis estadístico de datos y generación de gráficos

Breve descripción:

Este componente explora los conceptos fundamentales y las herramientas necesarias para el análisis estadístico, cubriendo desde las medidas de tendencia central y dispersión hasta las probabilidades y gráficos estadísticos. Abarca técnicas precisas para calcular medias, varianza y coeficientes, y profundiza en la interpretación de gráficos y la redacción de informes estadísticos efectivos.

Tabla de contenido

Introducción	1
1. Medidas de tendencia central.....	4
1.1. Media aritmética	4
1.2. Mediana	5
1.3. Moda	6
1.4. Media geométrica.....	6
1.5. Media ponderada	7
1.6. Media recortada	8
1.7. Trimedia.....	8
2. Medidas de dispersión	10
2.1. Rango	10
2.2. Desviación media.....	11
2.3. Varianza	12
2.4. Desviación estándar.....	13
2.5. Coeficiente de variación.....	13
2.6. Rango intercuartílico.....	14
2.7. Amplitud de variación.....	15
2.8. Coeficiente de variación.....	15

3.	Medidas de posición	16
3.1.	Cuartiles.....	16
3.2.	Deciles	17
3.3.	Percentiles	17
3.4.	Rango intercuartílico.....	18
4.	Probabilidades	19
4.1.	Cálculo de probabilidades.....	19
4.2.	Probabilidades marginales.....	19
4.3.	Regla de la adición	20
4.4.	Regla de la multiplicación	20
4.5.	Teorema de Bayes.....	21
	Distribución de probabilidades	21
4.6.	Distribución normal	22
4.7.	Distribución t-student	22
4.8.	Prueba de valor z	23
4.9.	Aplicaciones de la probabilidad en el análisis estadístico	24
5.	Gráficos estadísticos	25
5.1.	Tipos de gráficos	25
5.2.	Elección del gráfico adecuado según el tipo de datos	28

5.3.	Interpretación de gráficos estadísticos.....	29
6.	Informes estadísticos	30
6.1.	Estructura de un informe estadístico	30
6.2.	Presentación de datos en un informe	31
6.3.	Lenguaje claro y conciso	32
6.4.	Tipos de informes estadísticos	33
6.5.	Elaboración de conclusiones y recomendaciones	33
6.6.	Errores comunes en los informes estadísticos.....	34
	Síntesis	35
	Material complementario.....	37
	Glosario	38
	Referencias bibliográficas	42
	Créditos	44

Introducción

En el ámbito de la estadística aplicada, la comprensión y el uso de medidas estadísticas permiten describir, analizar y entender patrones dentro de los datos. Las medidas de tendencia central, dispersión y posición, junto con el cálculo de probabilidades, son fundamentales para interpretar correctamente la información y realizar inferencias válidas sobre poblaciones a partir de muestras.

¿Cómo se aplican estas herramientas para obtener resultados precisos y representativos en el análisis estadístico? Este componente formativo explora los principios y metodologías de las medidas estadísticas clave, abarcando desde los conceptos de media y varianza hasta las distribuciones de probabilidad y su interpretación en gráficos estadísticos. Además, se proporcionarán enfoques prácticos para seleccionar y presentar datos de manera clara y efectiva en informes estadísticos.

A lo largo de este componente, el aprendiz desarrollará habilidades para calcular y aplicar medidas estadísticas, interpretar distribuciones y utilizar gráficos para comunicar hallazgos. Asimismo, se destacarán las mejores prácticas para estructurar datos y elegir el tipo de gráfico adecuado, optimizando la precisión y la claridad en la presentación de resultados.

Las medidas estadísticas y su interpretación son pilares para cualquier análisis riguroso, ya que "la precisión de las conclusiones depende de la correcta aplicación de los principios estadísticos".

¡Bienvenido a este recorrido por el mundo de la estadística y la visualización efectiva de datos para el análisis!

Video 1. Análisis estadístico de datos y generación de gráficos



[Enlace de reproducción del video](#)

Síntesis del video: Análisis estadístico de datos y generación de gráficos

En el componente formativo «Análisis estadístico de datos y generación de gráficos» se profundiza en las metodologías para interpretar datos y representarlos visualmente, facilitando la toma de decisiones informada.

Durante el desarrollo de este componente, se busca que el aprendiz adquiera un dominio sólido de las medidas de tendencia central, dispersión y posición, que constituyen las bases para comprender la distribución de datos.

Se incluyen métodos de cálculo de probabilidades y aplicaciones prácticas de distribuciones como la normal y la t-student, esenciales para hacer inferencias sobre poblaciones a partir de muestras.

El componente también enseña cómo seleccionar el tipo de gráfico adecuado para cada tipo de variable, optimizando la claridad y efectividad en la presentación visual de los datos.

La generación de gráficos estadísticos es un paso fundamental en el análisis de datos, permitiendo identificar patrones, tendencias y posibles anomalías en un conjunto de datos.

Se abordan tanto gráficos básicos como histogramas y diagramas de dispersión, como representaciones más complejas, tales como polígonos de frecuencias y gráficos de boxplot.

El componente también destaca la importancia de la interpretación de gráficos para extraer insumos relevantes y comunicar hallazgos de forma comprensible para diferentes audiencias.

La estructura y claridad en los informes estadísticos se exploran, brindando pautas para elaborar conclusiones y recomendaciones basadas en datos.

A lo largo del componente, el aprendiz podrá asociar estos conceptos con aplicaciones reales de análisis y visualización, mejorando su capacidad para interpretar y comunicar datos de manera efectiva.

¡Bienvenidos al mundo del análisis estadístico y la generación de gráficos!

1. Medidas de tendencia central

Las medidas de tendencia central son valores que resumen un conjunto de datos, indicando el punto central o típico alrededor del cual los datos tienden a agruparse. Estas medidas proporcionan información sobre la distribución de los datos y permiten describir el comportamiento general de la muestra o población.

Figura 1. Medidas de tendencia central



Fuente. OIT, 2024.

1.1. Media aritmética

La **media aritmética**, también conocida simplemente como media, es la medida de tendencia central más utilizada. Se calcula sumando todos los valores de los datos y dividiendo el resultado entre el número total de observaciones. La media es especialmente útil cuando los datos están distribuidos de manera simétrica, ya que refleja el valor promedio del conjunto.

a) Fórmula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Donde:

- \bar{x} es la media aritmética.
- x_i representa cada uno de los valores del conjunto de datos.
- n es el número total de observaciones.

b) Propiedades:

- Es fácil de calcular y comprender.
- Afectada por valores atípicos (outliers), lo que puede distorsionar la interpretación.

1.2. Mediana

La **mediana** es el valor central de un conjunto de datos ordenado de menor a mayor. Divide el conjunto de datos en dos partes iguales, de tal manera que el 50 % de las observaciones son menores o iguales a la mediana, y el 50 % son mayores o iguales a ella. A diferencia de la media, la mediana no se ve afectada por valores atípicos, lo que la convierte en una medida más robusta en distribuciones asimétricas.

a) Cálculo:

- Si el número de observaciones es impar, la mediana es el valor que ocupa la posición central.
- Si el número de observaciones es par, la mediana es el promedio de los dos valores centrales.

b) Propiedades:

- No está afectada por valores extremos.
- Es útil para distribuciones sesgadas, ya que ofrece una representación más precisa del valor central.

1.3. Moda

La **moda** es el valor o los valores que más se repiten en un conjunto de datos. Es útil cuando se desea identificar el valor más frecuente en la distribución. Puede haber distribuciones **unimodales** (con una moda), **bimodales** (con dos modas) o **multimodales** (con más de dos modas).

a) Propiedades:

- Es la única medida de tendencia central que puede aplicarse a variables cualitativas.
- Puede no ser representativa del conjunto si los datos no tienen una moda clara o si hay varias modas.

1.4. Media geométrica

La **media geométrica** es una medida de tendencia central que se utiliza principalmente para conjuntos de datos positivos que están relacionados multiplicativamente, como tasas de crecimiento o porcentajes de cambio. Se calcula multiplicando todos los valores y extrayendo la raíz enésima del producto, donde n es el número total de observaciones.

a) Fórmula:

$$\bar{x}_g = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

Donde:

- \bar{x}_g es la media geométrica.
- x_i representa cada valor en el conjunto de datos.
- n es el número total de observaciones.

b) Propiedades:

- Es adecuada para datos que varían en proporción, como en el crecimiento de una población o el rendimiento de una inversión.
- Menos sensible a valores extremos que la media aritmética.

1.5. Media ponderada

La **media ponderada** es una variante de la media aritmética en la que cada valor del conjunto de datos se multiplica por un peso específico que refleja su importancia relativa. La media ponderada se usa cuando algunos valores tienen mayor relevancia o peso en el análisis que otros.

a) Fórmula:

$$\bar{x}_p = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

Donde:

- \bar{x}_p es la media ponderada.

- w_i es el peso asignado al valor x_i .
- n es el número total de observaciones.

b) Propiedades:

- Permite dar más relevancia a ciertos valores según su importancia o frecuencia.
- Es especialmente útil cuando se necesita reflejar la influencia desigual de los datos.

1.6. Media recortada

La **media recortada** es una versión modificada de la media aritmética, en la que se eliminan los valores más extremos (los mayores y menores) antes de calcular la media. Esto permite reducir el efecto de los valores atípicos en el cálculo.

a) Cálculo:

- Se eliminan un porcentaje predeterminado de los valores más altos y más bajos.
- Se calcula la media aritmética de los valores restantes.

b) Propiedades:

- Proporciona una medida de tendencia central menos sensible a valores atípicos.
- Es útil en situaciones donde los datos contienen valores extremos que podrían distorsionar la media.

1.7. Trimedia

La **trimedia** es una medida de tendencia central que toma en cuenta tanto la mediana como los cuartiles (primer y tercer cuartil) del conjunto de datos. Es una

medida robusta que combina información de diferentes puntos de la distribución para proporcionar un resumen más equilibrado de la tendencia central.

a) Fórmula:

$$Trimedia = \frac{Q_1 + 2\text{Mediana} + Q_3}{4}$$

Donde:

- Q_1 es el primer cuartil (25 % de los datos por debajo).
- Q_3 es el tercer cuartil (75 % de los datos por debajo).
- La mediana divide el conjunto de datos en dos mitades.

b) Propiedades:

- Proporciona una medida que tiene en cuenta tanto la dispersión como el centro de los datos.
- Es menos sensible a los valores atípicos que la media aritmética.

2. Medidas de dispersión

Las **medidas de dispersión** son herramientas estadísticas que describen qué tan dispersos o esparcidos están los datos en un conjunto. Estas medidas indican la variabilidad de los datos alrededor de una medida de tendencia central, como la media o la mediana. Cuanto mayor sea la dispersión, más lejos estarán los datos de la tendencia central, lo que refleja una mayor variabilidad en el conjunto de datos.

Figura 2. Medidas de dispersión.



Fuente. OIT, 2024.

2.1. Rango

El **rango** es la diferencia entre el valor más alto y el más bajo en un conjunto de datos. Es la medida de dispersión más simple y se usa para describir la amplitud total de los valores observados.

a) Fórmula:

$$R = x_{m\acute{a}x} - x_{m\acute{i}n}$$

Donde:

- $x_{m\acute{a}x}$ es el valor máximo del conjunto de datos.
- $x_{m\acute{i}n}$ es el valor mínimo del conjunto de datos.

b) Propiedades:

- Es fácil de calcular, pero es muy sensible a los valores atípicos, ya que solo depende de los extremos.
- No proporciona información sobre cómo se distribuyen los datos entre los valores más altos y bajos.

2.2. Desviación media

La **desviación media** es la media de las diferencias absolutas entre cada valor del conjunto de datos y la media aritmética. Proporciona una medida del promedio de la dispersión sin considerar la dirección de las desviaciones.

a) Fórmula:

$$Desviación\ media = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Donde:

- x_i es cada valor del conjunto de datos.
- \bar{x} es la media aritmética.
- n es el número total de observaciones.

b) Propiedades:

- Es una medida intuitiva de la dispersión, pero no es tan común como otras medidas como la varianza o la desviación estándar.
- Utiliza las desviaciones absolutas, lo que evita que las diferencias positivas y negativas se cancelen entre sí.

2.3. Varianza

La **varianza** es una medida que refleja el promedio de los cuadrados de las desviaciones de cada valor respecto a la media. La varianza indica qué tan dispersos están los datos, y es una de las medidas más importantes en la estadística inferencial.

a) Fórmula:

$$Varianza = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Donde:

- x_i es cada valor del conjunto de datos.
- \bar{x} es la media aritmética.
- n es el número total de observaciones.

b) Propiedades:

- La varianza utiliza los cuadrados de las desviaciones, lo que magnifica el impacto de las observaciones más alejadas de la media.
- Es muy útil para cálculos estadísticos avanzados, pero su principal desventaja es que está en unidades cuadradas, lo que puede hacer que sea difícil de interpretar directamente.

2.4. Desviación estándar

La **desviación estándar** es la raíz cuadrada positiva de la varianza y es quizás la medida de dispersión más utilizada. A diferencia de la varianza, la desviación estándar está en las mismas unidades que los datos originales, lo que facilita su interpretación.

a) Fórmula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Donde:

- σ es la desviación estándar.
- x_i es cada valor del conjunto de datos.
- \bar{x} es la media aritmética.
- n es el número total de observaciones.

b) Propiedades:

- Indica qué tan dispersos están los valores con respecto a la media.
- Es fácil de interpretar porque está en las mismas unidades que los datos originales.
- Es sensible a los valores atípicos, ya que utiliza las desviaciones cuadradas.

2.5. Coeficiente de variación

El **coeficiente de variación (CV)** es una medida de dispersión relativa que expresa la desviación estándar como un porcentaje de la media. Es útil para comparar la

variabilidad entre diferentes conjuntos de datos que pueden tener diferentes unidades o medias.

a) Fórmula:

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

Donde:

- σ es la desviación estándar.
- \bar{x} es la media aritmética.

b) Propiedades:

- Permite comparar la dispersión de diferentes conjuntos de datos, incluso cuando tienen unidades o escalas diferentes.
- Un coeficiente de variación más alto indica una mayor dispersión en relación con la media.

2.6. Rango intercuartílico

El **rango intercuartílico (IQR)** mide la dispersión de los datos entre el primer cuartil (Q_1) y el tercer cuartil (Q_3), es decir, la dispersión del 50 % central de los datos. Es una medida robusta de dispersión, ya que no se ve afectada por valores atípicos.

a) Fórmula:

$$IQR = Q_3 - Q_1$$

Donde:

- Q_3 es el tercer cuartil (75 % de los datos están por debajo).
- Q_1 es el primer cuartil (25 % de los datos están por debajo).

b) Propiedades:

- El rango intercuartílico es más robusto que la varianza o la desviación estándar, ya que excluye los extremos.
- Es útil cuando se quiere analizar la dispersión central sin la influencia de valores extremos o atípicos.

2.7. Amplitud de variación

La **amplitud de variación** (rango total) es otra forma de medir la dispersión, y representa la diferencia entre el valor máximo y el mínimo de un conjunto de datos.

a) Fórmula:

$$\text{Amplitud de variación} = x_{\text{máx}} - x_{\text{mín}}$$

b) Propiedades:

- Es una medida simple que ofrece una indicación rápida de la variabilidad.
- No toma en cuenta cómo se distribuyen los datos entre el valor máximo y el mínimo.

2.8. Coeficiente de variación

El coeficiente de variación es una medida que estandariza la desviación estándar respecto a la media, expresando la dispersión en términos porcentuales.

3. Medidas de posición

Las **medidas de posición** son herramientas estadísticas que permiten dividir un conjunto de datos en partes iguales para identificar la ubicación de un valor en relación con el resto del conjunto. Estas medidas son fundamentales para describir la posición relativa de una observación dentro de una distribución. Las principales medidas de posición son los **cuartiles**, **deciles** y **percentiles**, que permiten dividir los datos en diferentes porcentajes, y también se utilizan para identificar valores atípicos.

3.1. Cuartiles

Los **cuartiles** dividen un conjunto de datos ordenados en cuatro partes iguales, de modo que cada cuartil contiene el 25 % de los datos. Los cuartiles son particularmente útiles para describir la dispersión de los datos y su distribución en el rango intercuartílico.

a) Fórmula:

- **Primer cuartil (Q_1):** Valor que deja por debajo al 25 % de los datos.
- **Segundo cuartil (Q_2):** Valor que deja por debajo al 50 % de los datos (es equivalente a la mediana).
- **Tercer cuartil (Q_3):** Valor que deja por debajo al 75 % de los datos.

b) Interpretación:

- Q_1 divide el conjunto en un 25 % de valores por debajo y un 75 % por encima.
- Q_2 , o mediana, divide el conjunto en dos partes iguales.
- Q_3 divide el conjunto en un 75 % de valores por debajo y un 25 % por encima.

3.2. Deciles

Los **deciles** dividen el conjunto de datos en diez partes iguales, de modo que cada decil contiene el 10 % de las observaciones. Son útiles para observar la posición de un valor dentro de una división más detallada del conjunto de datos que los cuartiles.

a) Fórmula:

- **Primer decil (D_1):** Valor que deja por debajo al 10 % de los datos.
- **Segundo decil (D_2):** Valor que deja por debajo al 20 % de los datos.
- ...
- **Noveno decil (D_9):** Valor que deja por debajo al 90 % de los datos.

b) Interpretación:

- Cada decil indica la posición del conjunto de datos en fracciones del 10 %, proporcionando una idea más granular de cómo se distribuyen los valores.

3.3. Percentiles

Los **percentiles** dividen un conjunto de datos en cien partes iguales, de modo que cada percentil contiene el 1 % de las observaciones. Los percentiles son ampliamente utilizados en muchos campos, como la educación y la salud, para determinar la posición relativa de un individuo o una observación dentro de una población.

a) Fórmula:

- **Primer percentil (P_1):** Valor que deja por debajo al 1 % de los datos.
- **Décimo percentil (P_{10}):** Valor que deja por debajo al 10 % de los datos.
- **Percentil 50 (P_{50}):** Es equivalente a la mediana, ya que deja por debajo al 50 % de los datos.

- **Percentil 90 (P_{90}):** Valor que deja por debajo al 90 % de los datos.

b) Interpretación:

- El percentil proporciona información sobre la posición de un valor respecto a un grupo más amplio. Por ejemplo, si una persona se encuentra en el percentil 80 de un examen, significa que su puntuación es superior a la del 80 % de los participantes.

3.4. Rango intercuartílico

El **rango intercuartílico (IQR)** es una medida de dispersión que se basa en los cuartiles y representa la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1). Es útil para medir la variabilidad de la parte central de los datos, eliminando el efecto de los valores extremos o atípicos.

a) Fórmula:

$$IQR = Q_3 - Q_1$$

b) Interpretación:

- El rango intercuartílico mide la dispersión del 50 % central de los datos, lo que proporciona una visión más clara de cómo están distribuidos los valores centrales del conjunto.
- Es menos sensible a los valores extremos que otras medidas de dispersión, como la desviación estándar.

4. Probabilidades

La **probabilidad** es una rama de la matemática que estudia la posibilidad de que ocurra un evento o fenómeno aleatorio. En estadística, la probabilidad permite cuantificar la incertidumbre asociada con los resultados de un experimento o evento, proporcionando una medida numérica entre 0 y 1, donde 0 indica que el evento no ocurrirá y 1 indica que el evento es seguro.

4.1. Cálculo de probabilidades

La **probabilidad** de que ocurra un evento AAA se define como la proporción de casos favorables entre el número total de casos posibles. La fórmula básica es:

a) **Fórmula:**

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

Donde:

- $P(A)$ es la probabilidad del evento A.
- El **número de casos favorables** son las situaciones que cumplen con la condición del evento.
- El **número total de casos posibles** incluye todas las situaciones que podrían ocurrir.

4.2. Probabilidades marginales

Las **probabilidades marginales** se refieren a la probabilidad de que ocurra un solo evento sin considerar ningún otro evento. Es la probabilidad simple de un evento particular.

- **Ejemplo:** en un lanzamiento de un dado de 6 caras, la probabilidad marginal de obtener un número par (eventos favorables: 2, 4, 6) es:

$$P(\text{Número par}) = \frac{3}{6} = 0.5$$

4.3. Regla de la adición

La **regla de la adición** se utiliza para calcular la probabilidad de que ocurra uno de varios eventos. Existen dos casos:

- **Eventos mutuamente excluyentes:** son aquellos eventos que no pueden ocurrir al mismo tiempo. La probabilidad de que ocurra AAA o BBB es la suma de las probabilidades de cada evento:

$$P(A \text{ o } B) = P(A) + P(B)$$

- **Eventos no mutuamente excluyentes:** son eventos que pueden ocurrir juntos. En este caso, es necesario restar la intersección de ambos eventos:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$

4.4. Regla de la multiplicación

La **regla de la multiplicación** se utiliza para calcular la probabilidad de que ocurran dos o más eventos de manera conjunta (intersección de eventos). Existen dos casos:

- **Eventos independientes:** si los eventos no afectan la probabilidad del otro, la probabilidad conjunta de que ocurra AAA y BBB es:

$$P(A \text{ y } B) = P(A) \times P(B)$$

- **Eventos dependientes:** si los eventos están relacionados, la probabilidad conjunta se calcula.

$$P(A \text{ y } B) = P(A) \times P(B|A)$$

Donde $P(B|A)$ es la probabilidad de que ocurra B dado que ya ocurrió A.

4.5. Teorema de Bayes

El **teorema de Bayes** es una fórmula que describe la probabilidad de que ocurra un evento AAA, dado que ya ha ocurrido otro evento BBB. Este teorema es útil cuando se tiene información condicional sobre la probabilidad de un evento y se desea actualizar esa probabilidad a la luz de nueva información.

a) Fórmula:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Donde:

- $P(A|B)$ es la probabilidad de A dado B.
- $P(B|A)$ es la probabilidad de B dado A.
- $P(A)$ y $P(B)$ son las probabilidades marginales de A y B, respectivamente.

Distribución de probabilidades

Una **distribución de probabilidades** describe cómo se distribuyen las probabilidades entre los diferentes resultados de un experimento aleatorio. Hay

diferentes tipos de distribuciones de probabilidad, según el tipo de variable (discreta o continua) y el contexto del problema.

- **Distribución de probabilidad discreta:** describe la probabilidad de que una variable aleatoria discreta tome ciertos valores. Ejemplos incluyen la distribución binomial o la distribución de Poisson.
- **Distribución de probabilidad continua:** describe la probabilidad de que una variable aleatoria continua caiga dentro de un rango de valores. Ejemplos incluyen la distribución normal y la distribución uniforme.

4.6. Distribución normal

La **distribución normal**, también conocida como campana de Gauss, es una de las distribuciones de probabilidad más importantes en estadística. Muchos fenómenos naturales y sociales siguen una distribución normal. Está completamente definida por dos parámetros: la media (μ) y la desviación estándar (σ).

a) Características:

- Es simétrica respecto a la media.
- La media, la mediana y la moda son iguales.
- Aproximadamente el 68 % de los datos se encuentra dentro de una desviación estándar de la media, el 95 % dentro de dos desviaciones estándar, y el 99.7 % dentro de tres desviaciones estándar (regla empírica).

4.7. Distribución t-student

La **distribución t-student** se utiliza cuando el tamaño de la muestra es pequeño y la desviación estándar de la población es desconocida. Es similar a la distribución

normal, pero con colas más gruesas, lo que refleja una mayor probabilidad de valores extremos.

a) Aplicaciones:

- La distribución t se utiliza principalmente en pruebas de hipótesis, como la prueba t, cuando se trabaja con muestras pequeñas (generalmente $n < 30$).

4.8. Prueba de valor z

La **prueba de valor z** es una técnica estadística que permite calcular la probabilidad de que una observación en particular esté alejada de la media de un conjunto de datos, medida en unidades de desviación estándar.

a) Fórmula:

$$z = \frac{\chi - \mu}{\sigma}$$

Donde:

- χ es el valor de la observación.
- μ es la media de la población.
- σ es la desviación estándar.

b) Interpretación:

- Un valor z positivo indica que el valor está por encima de la media.
- Un valor z negativo indica que el valor está por debajo de la media.

4.9. Aplicaciones de la probabilidad en el análisis estadístico

La probabilidad se utiliza en una amplia variedad de aplicaciones en el análisis estadístico, desde la estimación de probabilidades de eventos simples hasta la toma de decisiones basada en información incompleta o incierta. Algunos ejemplos de uso incluyen:

- **Pruebas de hipótesis:** se utilizan probabilidades para evaluar si una hipótesis es válida o no, con base en la evidencia de los datos muestrales.
- **Modelos predictivos:** la probabilidad es fundamental en modelos predictivos, como el análisis de regresión y las redes bayesianas, donde se calcula la probabilidad de resultados futuros con base en datos históricos.
- **Toma de decisiones:** en la teoría de la decisión, se usan probabilidades para evaluar el riesgo y la incertidumbre en situaciones donde se deben tomar decisiones, como en inversiones o análisis de riesgos.

5. Gráficos estadísticos

Los gráficos estadísticos son representaciones visuales de datos que permiten observar tendencias, patrones y relaciones dentro de un conjunto de datos. Son herramientas poderosas para simplificar y comunicar resultados complejos de manera clara y concisa.

5.1. Tipos de gráficos

Existen varios tipos de gráficos que se pueden utilizar según el tipo de datos y el objetivo del análisis.

Tabla 1. Tipos de gráficos

Gráfico	Descripción	Características	Aplicaciones
Histograma.	Representación gráfica de la distribución de una variable cuantitativa, donde la altura de cada barra representa la frecuencia en intervalos o clases.	<ul style="list-style-type: none"> • Para variables cuantitativas (discretas o continuas). • Muestra la forma de la distribución (simétrica, sesgada, multimodal). • Permite visualizar la concentración de datos en ciertos rangos. 	<ul style="list-style-type: none"> • Distribución de edades en una población. • Distribución de ingresos en una empresa.
Gráfico de barras.	Representa datos cualitativos o categóricos mediante barras donde la altura indica la frecuencia o porcentaje de cada categoría.	<ul style="list-style-type: none"> • Para variables cualitativas (nominales u ordinales). • Las barras no se tocan entre sí. 	<ul style="list-style-type: none"> • Preferencias de consumidores para marcas. • Cantidad de personas en diferentes grupos demográficos.

Gráfico	Descripción	Características	Aplicaciones
		<ul style="list-style-type: none"> • Ideal para comparar varias categorías. 	
Diagrama de tallos y hojas.	<p>El diagrama de tallos y hojas es una representación gráfica de datos numéricos que muestra la distribución y la forma de los datos, dividiéndolos en "tallos" (dígitos principales) y "hojas" (dígitos restantes). Esto permite ver cada valor individual al tiempo que se muestra la estructura general de la distribución.</p>	<ul style="list-style-type: none"> • Se utiliza principalmente para datos numéricos, especialmente con valores de dos o más dígitos. • Cada "tallo" representa una categoría o rango de valores principales. • Cada "hoja" representa la parte final de cada dato, permitiendo visualizar valores individuales. • Es una herramienta útil para ver la dispersión y concentración de datos, similar a un histograma pero sin perder el detalle de cada valor. 	<ul style="list-style-type: none"> • Análisis exploratorio de datos en pequeñas muestras. • Visualización de distribuciones en contextos educativos y estadísticos básicos. • Comparación rápida de dos conjuntos de datos en un mismo diagrama (usando dos columnas de hojas en cada lado de un tallo común). • Útil para detectar patrones, tendencias y valores atípicos en datos numéricos.
Gráfico circular.	<p>También llamado gráfico de pastel, muestra la proporción o porcentaje de cada categoría respecto al total mediante sectores proporcionales.</p>	<ul style="list-style-type: none"> • Para datos cualitativos nominales. • Ideal para proporciones o porcentajes. • No adecuado para categorías con 	<ul style="list-style-type: none"> • Participación de mercado de empresas. • Distribución de votos en una elección.

Gráfico	Descripción	Características	Aplicaciones
		pequeñas diferencias en proporciones.	
Polígono de frecuencias.	Gráfico de líneas que conecta los puntos medios de los intervalos de un histograma, útil para observar la forma de la distribución de los datos de manera continua y comparar diferentes conjuntos de datos.	<ul style="list-style-type: none"> • Para variables cuantitativas continuas. • Conecta puntos medios de intervalos de clase. • Puede superponerse con otros polígonos para comparar distribuciones. 	<ul style="list-style-type: none"> • Comparar puntuaciones en exámenes. • Evolución de la temperatura en diferentes horas del día.
Ojiva.	Representación gráfica de la frecuencia acumulada o relativa acumulada de los datos, mostrando cómo se acumulan las frecuencias al aumentar los valores de la variable.	<ul style="list-style-type: none"> • Para variables cuantitativas. • Los puntos representan frecuencias acumuladas conectadas por líneas. • Permite calcular percentiles y cuartiles visualmente. 	<ul style="list-style-type: none"> • Proporción acumulada de empleados con ciertos ingresos. • Estudiantes con calificaciones bajo cierto valor.
Diagrama de dispersión.	Muestra la relación entre dos variables cuantitativas mediante puntos (x, y), útil para observar correlaciones o tendencias entre las variables.	<ul style="list-style-type: none"> • Para relaciones entre dos variables cuantitativas. • Permite identificar tendencias lineales o no lineales. 	<ul style="list-style-type: none"> • Relación entre peso y altura. • Relación entre tiempo de estudio y rendimiento académico.

Gráfico	Descripción	Características	Aplicaciones
		<ul style="list-style-type: none"> Útil para detectar relaciones directas, inversas o la ausencia de relación. 	
Boxplot (Diagrama de caja y bigotes).	Gráfico que representa la dispersión de un conjunto de datos, mostrando el rango intercuartílico, la mediana y los valores atípicos mediante una caja y "bigotes".	<ul style="list-style-type: none"> Resume la distribución de datos y muestra valores atípicos. Muestra cuartiles y mediana, indicando simetría o asimetría. Permite comparar distribuciones entre grupos. 	<ul style="list-style-type: none"> Comparar ingresos entre departamentos. Evaluar la dispersión de puntuaciones en un examen.

Fuente. OIT, 2024.

5.2. Elección del gráfico adecuado según el tipo de datos

La elección del gráfico adecuado depende del tipo de datos y del objetivo del análisis. Algunas pautas generales para seleccionar el gráfico correcto son:

- **Datos cualitativos:** gráfico de barras o gráfico circular.
- **Datos cuantitativos continuos:** histograma, polígono de frecuencias, ojiva.
- **Comparación de distribuciones:** boxplot, histograma.
- **Relación entre dos variables:** diagrama de dispersión.

5.3. Interpretación de gráficos estadísticos

La correcta **interpretación** de los gráficos estadísticos sirve para extraer conclusiones válidas de los datos. Al interpretar gráficos, es importante prestar atención a:

- **Distribución:** observar si los datos son simétricos, sesgados o multimodales.
- **Tendencias:** identificar patrones ascendentes, descendentes o cíclicos en los datos.
- **Relaciones:** en diagramas de dispersión, buscar correlaciones o asociaciones entre las variables.
- **Valores atípicos:** detectar puntos que se alejan significativamente del resto del conjunto de datos (particularmente en boxplots).

6. Informes estadísticos

Un informe estadístico es un documento que presenta los resultados de un análisis estadístico de manera estructurada y comprensible. Su objetivo es comunicar información cuantitativa y cualitativa basada en datos, apoyada en gráficos, tablas y descripciones que faciliten la interpretación de los hallazgos. Los informes estadísticos son fundamentales en la toma de decisiones en diversos campos como la ciencia, economía, salud, educación y más.

6.1. Estructura de un informe estadístico

Un informe estadístico debe seguir una estructura clara y organizada que permita a los lectores comprender fácilmente el análisis realizado. A continuación, se presenta una estructura común para estos informes:

- a) **Portada:** contiene el título del informe, el nombre de los autores, la fecha de elaboración y la entidad para la que se realiza el estudio.
- b) **Resumen ejecutivo:** es una breve descripción del objetivo del análisis, la metodología empleada y las principales conclusiones. Aunque aparece al inicio, es recomendable redactarlo al final, cuando el informe esté completo.
- c) **Introducción:** expone el contexto del estudio, los objetivos específicos del análisis y la importancia de los resultados. También puede incluir la definición de palabras clave.
- d) **Metodología:** describe los métodos estadísticos y técnicas de recolección de datos utilizados en el estudio. Incluye información sobre la muestra, el tipo de datos recolectados, las fuentes de información y el enfoque de análisis.

- e) **Resultados:** presenta los resultados obtenidos a través de tablas, gráficos y descripciones detalladas. Los resultados deben ser claros y estar apoyados por las medidas estadísticas correspondientes (medidas de tendencia central, dispersión, frecuencias, etc.).
- f) **Análisis e interpretación:** aquí se interpreta la información presentada en los resultados. Se deben destacar los patrones importantes, las relaciones entre variables y cualquier conclusión relevante que se pueda extraer del análisis. Es fundamental discutir el significado de los hallazgos y relacionarlos con los objetivos planteados.
- g) **Conclusiones:** resume los hallazgos más importantes del informe y sugiere posibles acciones o recomendaciones basadas en los datos. También se pueden mencionar las limitaciones del estudio y posibles áreas de mejora para investigaciones futuras.
- h) **Anexos:** incluyen tablas, gráficos adicionales o cualquier información complementaria que no sea esencial en el cuerpo principal del informe pero que sea útil para el lector interesado en más detalles.

6.2. Presentación de datos en un informe

La **presentación de datos** es uno de los aspectos más importantes de un informe estadístico. Debe ser clara y comprensible, utilizando tablas y gráficos que faciliten la interpretación de los resultados. Los datos deben estar organizados de manera coherente y apoyados por explicaciones claras.

a) Tablas

- Las tablas deben estar numeradas y acompañadas de títulos descriptivos.

- Se debe incluir una breve explicación de cada tabla para facilitar su interpretación.
- Los datos en las tablas deben estar alineados correctamente y ser fáciles de leer.

b) Gráficos

- Los gráficos utilizados deben ser apropiados para el tipo de datos (histogramas para distribuciones, gráficos de barras para categorías, etc.).
- Cada gráfico debe tener un título claro y etiquetas en los ejes.
- Deben ser visualmente simples, evitando el uso excesivo de colores o elementos distractores.

Fuente. OIT, 2024.

6.3. Lenguaje claro y conciso

El lenguaje utilizado en un informe estadístico debe ser claro, directo y objetivo. Los términos técnicos deben estar explicados de manera que cualquier lector, independientemente de su nivel de conocimiento estadístico, pueda comprender el informe. Es importante evitar el uso excesivo de jerga técnica y centrarse en comunicar las ideas de manera accesible.

- **Evitar el uso excesivo de tecnicismos:** si es necesario incluir términos estadísticos, asegúrate de proporcionar una breve explicación o definición.
- **Claridad:** el informe debe ser fácil de leer y entender. Se deben evitar frases ambiguas o complicadas.
- **Precisión:** los resultados deben presentarse de manera precisa y respaldada por los datos.

6.4. Tipos de informes estadísticos

Dependiendo del contexto y los objetivos, existen diferentes tipos de informes estadísticos:

- **Informes descriptivos:** se centran en describir los datos mediante el uso de estadísticas descriptivas como la media, mediana, moda, desviación estándar y gráficas que representen visualmente los datos.
- **Informes comparativos:** en estos informes, se comparan diferentes grupos o condiciones para identificar diferencias significativas o patrones en los datos.
- **Informes predictivos:** utilizan modelos estadísticos para hacer predicciones sobre tendencias futuras o resultados probables basados en los datos actuales.

6.5. Elaboración de conclusiones y recomendaciones

Las **conclusiones** deben estar directamente relacionadas con los objetivos planteados al inicio del informe. Las conclusiones sean objetivas y se basen en los resultados presentados. Además, las recomendaciones deben ofrecer sugerencias prácticas o acciones que se pueden tomar en base a los hallazgos del informe.

- **Conclusiones:** resumen de los hallazgos más importantes del análisis. Se deben mencionar los principales aspectos que respondan a las preguntas de investigación.
- **Recomendaciones:** sugerencias sobre cómo aplicar los resultados en un contexto práctico. Pueden incluir propuestas de cambios, mejoras, o recomendaciones para futuras investigaciones.

6.6. Errores comunes en los informes estadísticos

Al elaborar un informe estadístico, es importante evitar errores comunes que pueden afectar la claridad y precisión del documento. Algunos de estos errores incluyen:

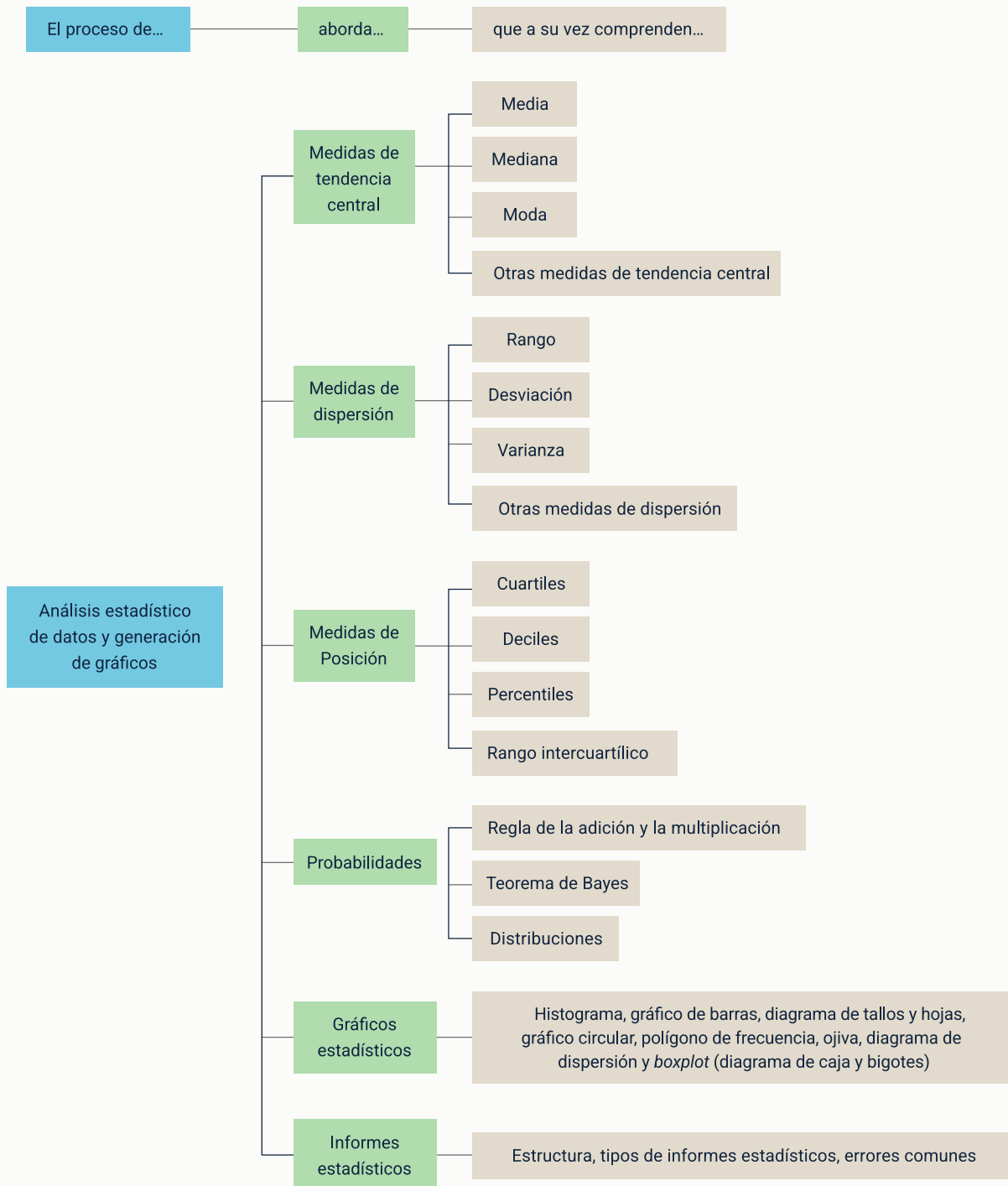
- **Mala selección de gráficos:** utilizar gráficos que no representen correctamente los datos puede llevar a malinterpretaciones.
- **Exceso de detalles técnicos:** incluir demasiada información técnica puede abrumar al lector y dificultar la comprensión de los resultados.
- **Interpretaciones erróneas:** sacar conclusiones que no están respaldadas por los datos o exagerar la importancia de ciertos resultados.
- **Falta de coherencia:** no seguir una estructura clara o incluir secciones que no estén bien conectadas entre sí.

Síntesis

El siguiente diagrama proporciona una visión general de los principales temas abordados en este componente, centrado en el uso de medidas estadísticas y probabilísticas para el análisis de datos. Este mapa conceptual está diseñado para facilitar al lector la visualización de las interrelaciones entre las diferentes áreas que componen el proceso de análisis estadístico, contribuyendo a una comprensión estructurada y práctica de los conceptos clave.

En el origen del diagrama se encuentra el concepto principal de "Análisis estadístico de datos y generación de gráficos", del cual se derivan los temas esenciales: Medidas de tendencia central, Medidas de dispersión, Medidas de posición, Probabilidades y Gráficos estadísticos. Cada área se subdivide en conceptos específicos, como media, varianza, cuartiles y distribución de probabilidades, los cuales reflejan la organización del contenido del componente.

Este diagrama actúa como una guía visual que acompaña al lector en la exploración de los conceptos presentados, permitiéndole captar de manera rápida la estructura y el alcance de los temas tratados, así como sus conexiones. Al consultar este mapa, el lector podrá apreciar cómo los distintos aspectos de la estadística se integran para formar un proceso coherente de análisis y presentación de datos. Se invita a utilizar este diagrama como un complemento al contenido detallado del componente, sirviendo como una referencia visual y un recordatorio de los principios esenciales en el análisis estadístico y la visualización de datos.



Fuente. OIT, 2024.

Material complementario

Tema	Referencia	Tipo de material	Enlace del recurso
Introducción a la estadística	Ecosistema de Recursos Educativos Digitales SENA. (2023c, septiembre 20). Introducción a la estadística.	Video	https://www.youtube.com/watch?v=wMCDkknpuVw
Introducción a la probabilidad	Ecosistema de Recursos Educativos Digitales SENA. (2023a, marzo 24). Introducción a la probabilidad.	Video	https://www.youtube.com/watch?v=rmx4jxS81vY
Muestreo, distribución de medias muestrales	Ecosistema de Recursos Educativos Digitales SENA. (2022, 11 mayo). Muestreo, distribución de medias muestrales.	Video	https://www.youtube.com/watch?v=5VK7hvyj760
Principales elementos de la estadística	Ecosistema de Recursos Educativos Digitales SENA. (2022b, octubre 26). Principales elementos de la estadística.	Video	https://www.youtube.com/watch?v=Ad5gxB9PhKQ
Principios básicos en el análisis de datos	Ecosistema de Recursos Educativos Digitales SENA. (2023b, marzo 24). Principios básicos en el análisis de datos.	Video	https://www.youtube.com/watch?v=bprDNbxNyo4
Teoría de probabilidades	Ecosistema de Recursos Educativos Digitales SENA. (2024, 18 junio). Teoría de probabilidades.	Video	https://www.youtube.com/watch?v=m-UUs_SwA9k

Glosario

Amplitud de variación: diferencia entre el valor máximo y mínimo de un conjunto de datos, utilizada como medida básica de dispersión.

Cálculo de probabilidades: métodos y fórmulas utilizados para determinar la probabilidad de ocurrencia de un evento en un espacio muestral.

Coefficiente de variación: medida de dispersión relativa que permite comparar la variabilidad de dos o más conjuntos de datos con unidades diferentes.

Cuartiles: valores que dividen un conjunto de datos ordenado en cuatro partes iguales, cada una representando el 25 % de la distribución.

Deciles: valores que dividen un conjunto de datos en diez partes iguales, cada una representando el 10 % de la distribución.

Desviación estándar: medida de dispersión que indica cuánto se alejan los datos en promedio de la media; calcula la variabilidad en una distribución.

Desviación media: promedio de las diferencias absolutas de cada valor con respecto a la media, utilizada para medir la dispersión de los datos.

Distribución de probabilidades: distribución que describe cómo se asignan las probabilidades a los diferentes valores de una variable aleatoria.

Distribución normal: distribución de probabilidad simétrica en forma de campana que describe cómo se distribuyen los datos alrededor de la media.

Distribución t-student: distribución de probabilidad utilizada en inferencia estadística, especialmente cuando el tamaño de la muestra es pequeño.

Elección del gráfico adecuado: proceso de seleccionar el tipo de gráfico más apropiado según el tipo de datos y el objetivo del análisis.

Informe estadístico: documento que presenta los resultados de un análisis estadístico de forma organizada, con conclusiones y recomendaciones.

Interpretación de gráficos: proceso de analizar gráficos estadísticos para extraer información y comprender patrones o tendencias en los datos.

Media aritmética: promedio de un conjunto de datos, calculado sumando todos los valores y dividiendo entre la cantidad de observaciones.

Media geométrica: media calculada multiplicando todos los valores y extrayendo la raíz enésima (según el número de valores), utilizada en datos con crecimiento proporcional.

Media ponderada: promedio que considera la importancia o peso de cada valor en el conjunto de datos, otorgando mayor relevancia a ciertos valores.

Media recortada: media calculada excluyendo los valores extremos (superiores e inferiores) de un conjunto de datos para reducir el impacto de los valores atípicos.

Mediana: valor central de un conjunto de datos ordenado; divide la distribución en dos partes iguales.

Moda: valor o categoría que aparece con mayor frecuencia en un conjunto de datos; útil en datos cualitativos y cuantitativos.

Percentiles: valores que dividen un conjunto de datos en 100 partes iguales; indican el porcentaje de datos que se encuentra por debajo de un valor específico.

Probabilidades marginales: probabilidades de ocurrencia de un solo evento en un espacio muestral, sin considerar la ocurrencia de otros eventos.

Prueba de valor z: prueba estadística para evaluar si una muestra pertenece a una población conocida o para comparar dos medias de poblaciones cuando se conoce la desviación estándar.

Rango: diferencia entre el valor máximo y mínimo en un conjunto de datos; medida sencilla de dispersión.

Rango intercuartílico: diferencia entre el tercer y el primer cuartil de un conjunto de datos; mide la dispersión en la parte central de la distribución.

Regla de la adición: regla de probabilidad que permite calcular la probabilidad de que ocurra al menos uno de varios eventos mutuamente excluyentes.

Regla de la multiplicación: regla de probabilidad que permite calcular la probabilidad conjunta de que ocurran dos o más eventos independientes.

Teorema de Bayes: teorema que permite actualizar la probabilidad de un evento con base en nueva evidencia o información; utilizado en inferencia estadística.

Tipos de gráficos: diferentes formas de representación visual de datos, como histogramas, gráficos de barras y gráficos circulares, cada uno adecuado para tipos específicos de datos.

Trimedia: medida de tendencia central calculada a partir de los cuartiles, que combina la mediana con otros valores para una representación robusta de los datos.

Varianza: medida de dispersión que indica la variabilidad de los datos respecto a la media; es el cuadrado de la desviación estándar.

Visualización de datos: representación gráfica de los datos para facilitar la interpretación de patrones y tendencias en los análisis estadísticos.

Referencias bibliográficas

Batanero, C. (2001). Didáctica de la estadística. Granada: Universidad de Granada.

Cochran, W. G. (1980). Técnicas de muestreo (3.ª ed.). México: CECSA.

Escobar Mercado, M., Fernández Macías, E., & Bernardi, F. (2012). Análisis de datos con Stata (2ª ed.). Centro de Investigaciones Sociológicas. Recuperado de Academia.

Hernández Martín, Z. (2018). Métodos de análisis de datos: apuntes. Universidad de La Rioja. Recuperado de Universidad de la Rioja.

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

Martínez, J. (2004). Muestreo estadístico. Madrid: Alianza Editorial.

Montgomery, D. C., & Runger, G. C. (2015). Probabilidad y estadística aplicada a la ingeniería (5.ª ed.). México: McGraw-Hill.

Obando-Bastidas, J. A., & Castellanos Sánchez, M. T. (2021). Gráficos estadísticos: guía práctica para estadística descriptiva. Universidad Cooperativa de Colombia. Recuperado de Repositorio UCC.

Panteleeva, O. V. (2005). Fundamentos de Probabilidad y Estadística. UAEM.

Ramos Calcina, A., Villasante Saravia, F. H., & Álvarez Rozas, T. P. (2023). Estadística descriptiva con R (2ª ed.). Universidad Nacional del Altiplano. Recuperado de Repositorio UNAP.

Scheaffer, R. L., Mendenhall, W., & Ott, R. L. (2007). Elementos de muestreo (6.^a ed.). México: Thomson.

Triola, M. F. (2018). Estadística (12.^a ed.). México: Pearson Educación.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). Probabilidad y estadística para ingenieros (9.^a ed.). México: Pearson Educación.

Créditos

Elaborado por:



**Organización
Internacional
del Trabajo**