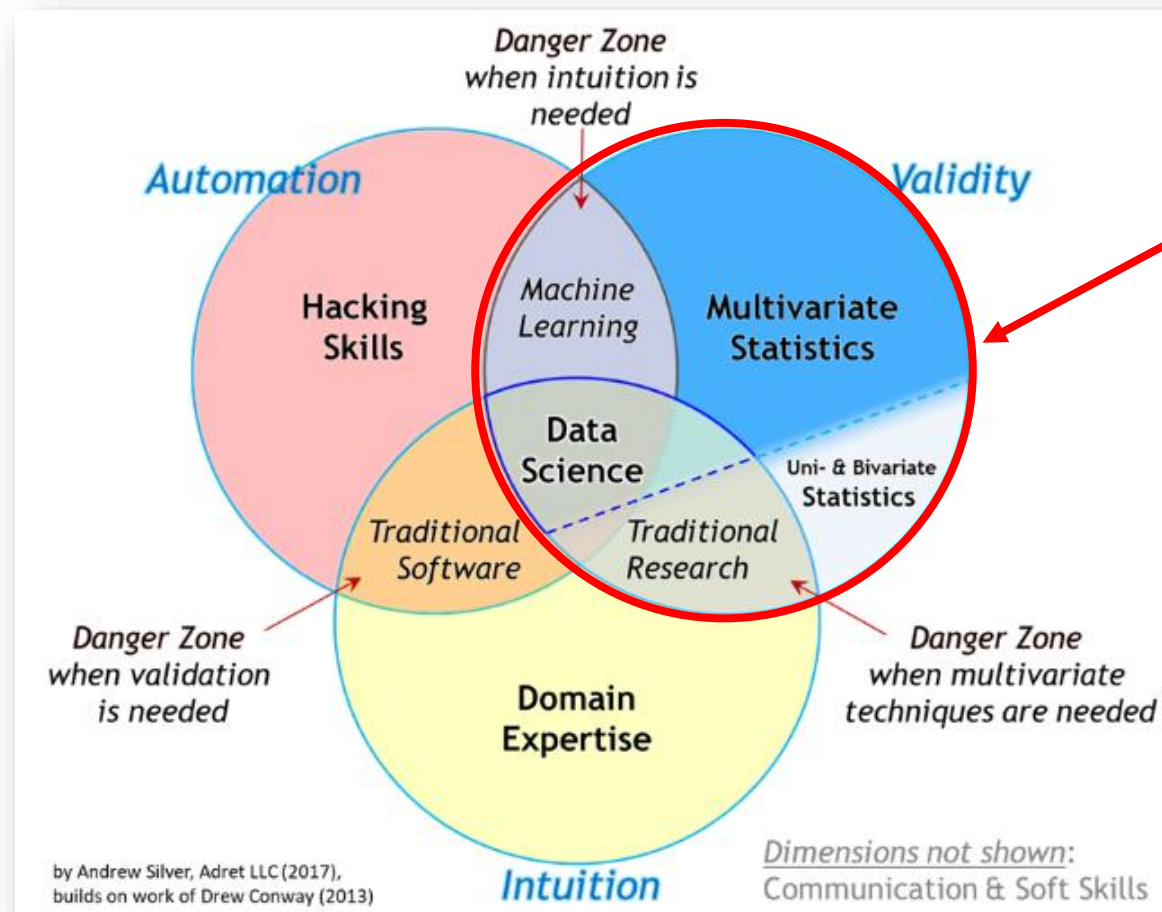


Estadística descriptiva y regresión lineal con Excel



Presentación Módulo AADD (IES Trassierra) © 2025 por Elena Fernández Chirino tiene licencia CC BY-NC-ND 4.0.
Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Hasta ahora hemos visto ...

01

Fundamentos Estadísticos

Domina los conceptos esenciales: introducción a la estadística, identificación de fuentes de datos confiables, comprensión de conceptos fundamentales y clasificación de tipos de variables.

03

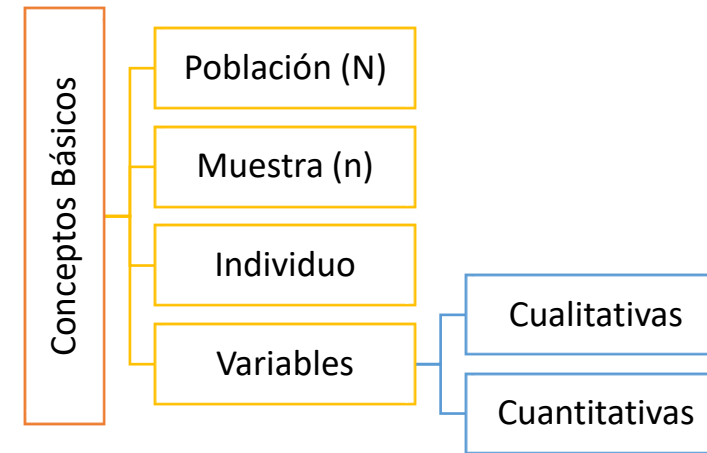
Visualización de Datos

Crea tablas de frecuencia profesionales y gráficos impactantes: barras, sectores e histogramas que comuniquen tus hallazgos con claridad.

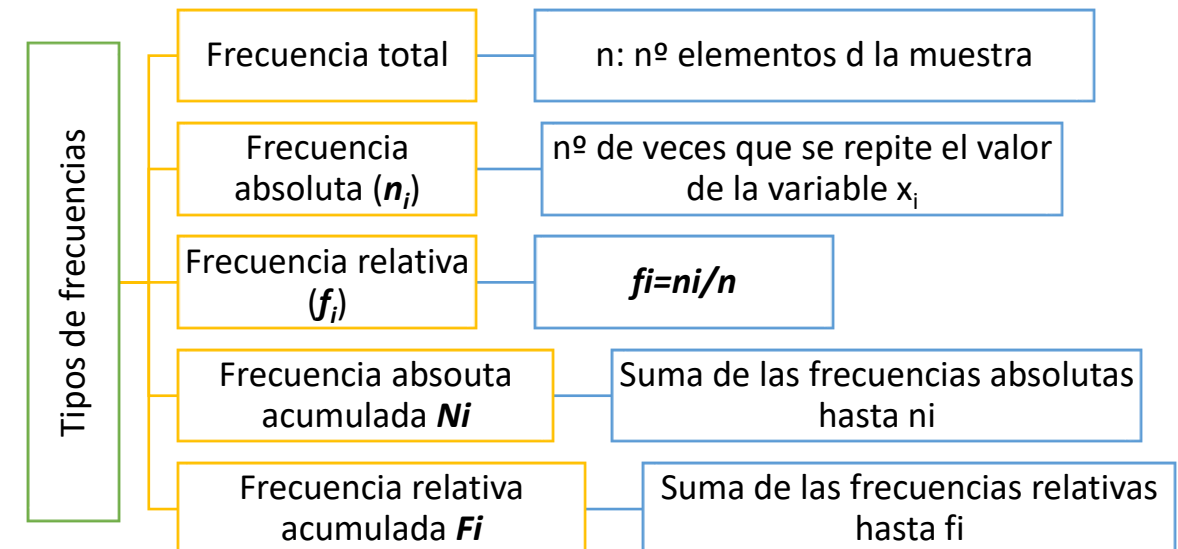
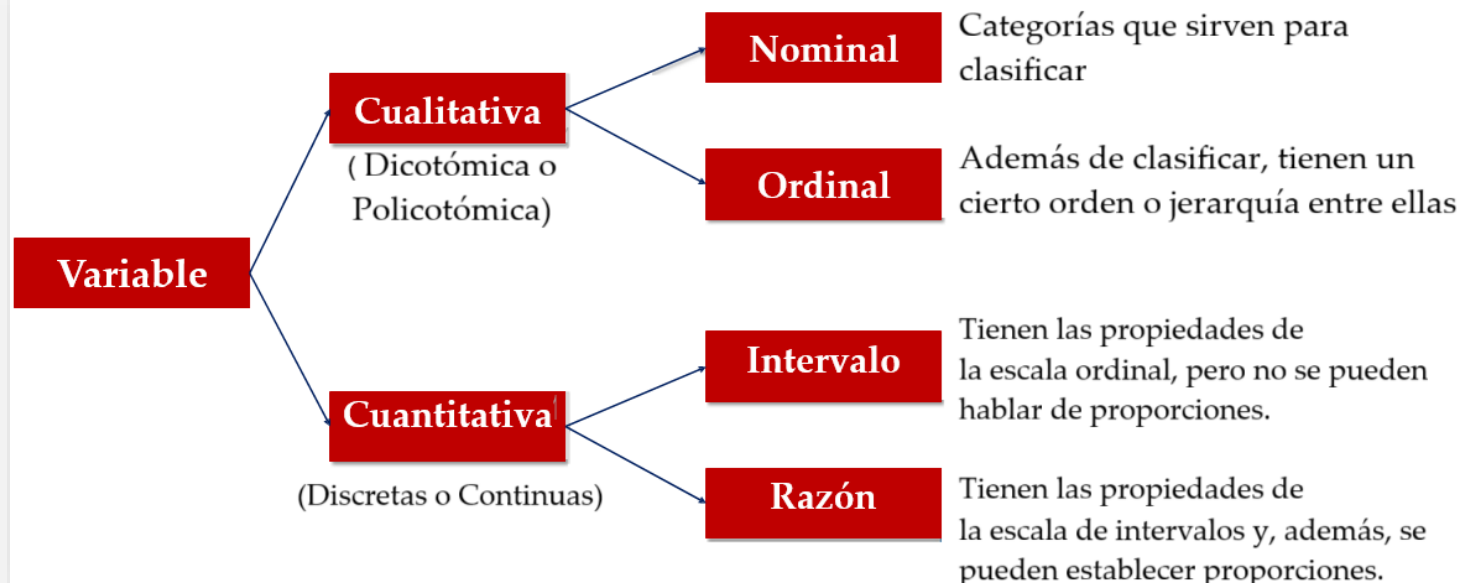
02

Escalas de Medición

Aprende a distinguir entre escalas nominal, ordinal, de intervalo y de razón. Conoce sus aplicaciones prácticas y ejemplos del mundo real.



Escalas de medición

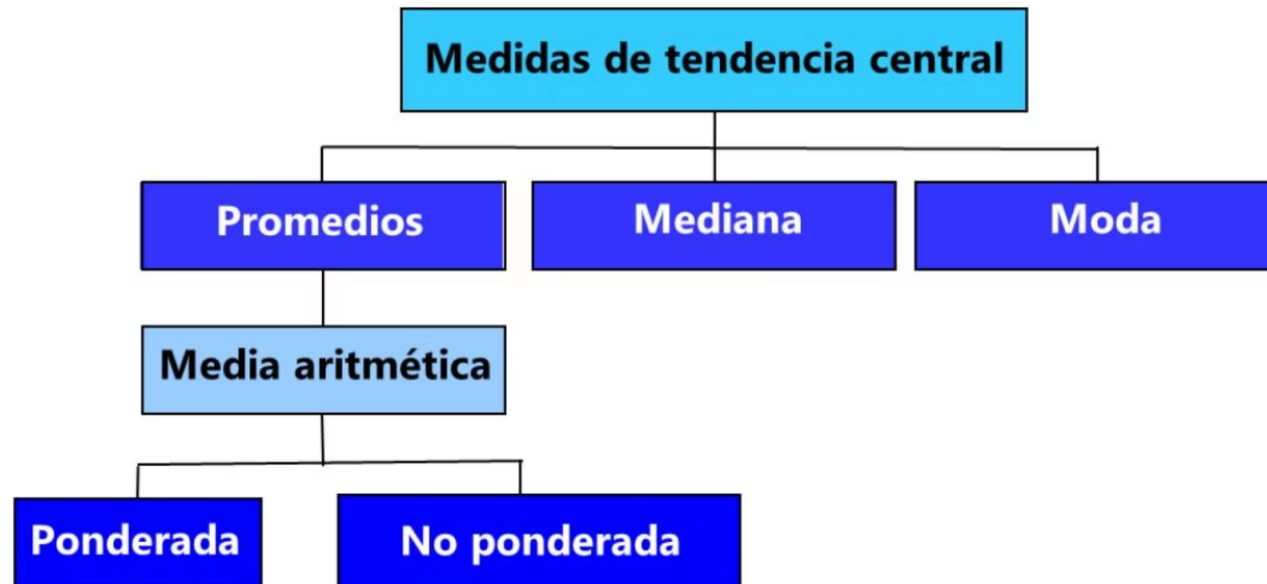


Medidas de Posición

Tendencia Central

Ubican el centro representativo de la distribución:

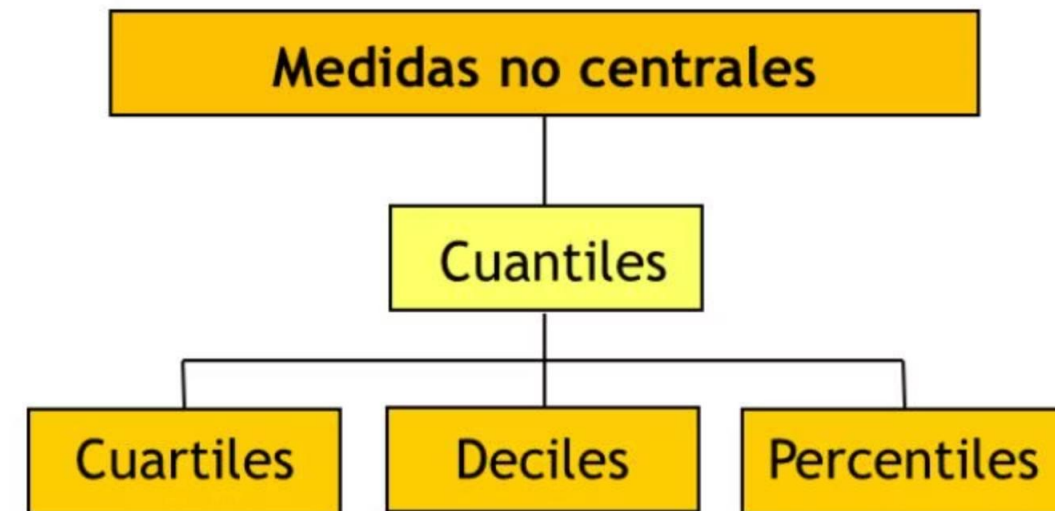
- **Media aritmética:** promedio de todos los valores
- **Mediana:** valor que divide la distribución en dos partes iguales
- **Moda:** valor más frecuente



Tendencia No Central

Identifican otros puntos característicos:

- **Cuartiles:** dividen en 4 partes
- **Deciles:** dividen en 10 partes
- **Percentiles:** dividen en 100 partes



Ahora vamos a ver el punto 05

01

Fundamentos Estadísticos

Domina los conceptos esenciales: introducción a la estadística, identificación de fuentes de datos confiables, comprensión de conceptos fundamentales y clasificación de tipos de variables.

03

Visualización de Datos

Crea tablas de frecuencia profesionales y gráficos impactantes: barras, sectores e histogramas que comuniquen tus hallazgos con claridad.

05

Dispersión y Variabilidad

Evalúa la variabilidad mediante rango, varianza, desviación típica y coeficiente de variación para entender la distribución completa de tus datos.

07

Regresión Lineal

Calcula la recta de regresión, determina el coeficiente de correlación e interpreta resultados para hacer predicciones fundamentadas.

02

Escalas de Medición

Aprende a distinguir entre escalas nominal, ordinal, de intervalo y de razón. Conoce sus aplicaciones prácticas y ejemplos del mundo real.

04

Análisis Numérico

Calcula e interpreta medidas de tendencia central (media, mediana, moda) y medidas de posición (cuartiles, percentiles) para comprender tus datos.

06

Análisis Bivariante

Explora relaciones entre variables con tablas de contingencia y técnicas de análisis conjunto para descubrir patrones ocultos.

08

Modelos Avanzados

Ajusta curvas no lineales y aplica modelos exponenciales, logarítmicos y polinómicos para análisis sofisticados de datos complejos.

Descripción Numérica de Datos: Medidas de Dispersión

Presentación Módulo AADD (IES Trassierra) © 2025 por Elena Fernández Chirino tiene licencia CC BY-NC-ND 4.0.
Para ver una copia de esta licencia, visite <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Tipos de Medidas Descriptivas

01

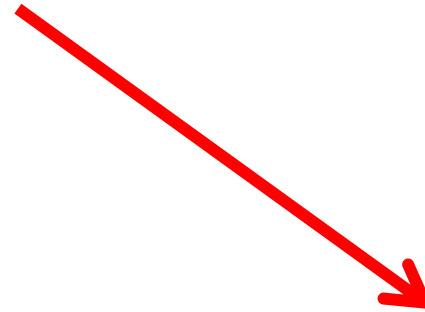
Medidas de Posición

Indican la situación de la variable en su escala de medida

02

Medidas de Dispersión

Evalúan la representatividad y variabilidad de los datos



03

Medidas de Forma

Caracterizan la simetría y apuntamiento de la distribución

¿Qué son las Medidas de Dispersión?

Concepto Fundamental

Las medidas de dispersión evalúan la **mayor o menor variabilidad** existente en un conjunto de datos. Permiten **valorar el grado de representatividad de una medida de posición según la magnitud de la dispersión**.

Cuantifican lo separados que están los datos, bien entre sí o respecto del valor central que los representa.

Importancia

Las medidas de posición central sintetizan la información disponible. Una medida será más representativa según la proximidad de los datos a ella.

Las medidas de dispersión nos permiten conocer lo cerca o lejos que se encuentran los datos respecto a una medida de posición central.





Clasificación de las Medidas de Dispersión

Dispersión Absoluta

Sin referencia central:

- Recorridos muestral
- Intercuartílico
- Decil y percentil

Con referencia central:

- Varianza
- Desviación Típica

Dispersión Relativa

Sin referencia central:

- Recorridos relativo
- Semi-intercuartílico

Con referencia central:

- Coeficiente de Variación de Pearson

Recorridos: Medidas Simples de Dispersión

01

Recorrido Muestral

La forma más sencilla de medir dispersión: diferencia entre valor máximo y mínimo.

$$Re = x_k - x_1$$

Se ve muy afectado por observaciones anómalas o atípicas.

02

Recorrido Intercuartílico

Medida más fiable y menos sensible a datos atípicos.

$$R_I = C_3 - C_1$$

Comprende el 50% de los datos centrales.

03

Recorridos Generalizados

Para abarcar mayor porcentaje de datos:

$$R_D = D_9 - D_1$$

$$R_P = P_{99} - P_1$$

Varianza: La Medida Fundamental

Definición

La varianza mide la variabilidad de un conjunto de datos respecto de la media aritmética. Se denota por S^2 .

$$S^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{X})^2 \cdot n_i$$

Expresión equivalente para el cálculo:

$$S^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 \cdot n_i - \bar{X}^2$$

Propiedades Clave

- Es no negativa: $S^2 \geq 0$
- Vale cero solo si todos los valores coinciden
- Cuanto más próxima a 0, menor dispersión y mayor representatividad de la media
- Invariante ante cambios de origen, pero no de escala
- Inconveniente:
Se expresa en unidades cuadráticas

Desviación típica: $S = \sqrt{S^2}$

Concepto

La desviación típica es la raíz cuadrada de la varianza, expresada en las mismas unidades que los datos originales.

$$S = \sqrt{S^2} = \sqrt{\frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 \times n_i}$$

Ventaja Principal

Resuelve el inconveniente de la varianza al expresarse en las mismas unidades que la variable original, facilitando su interpretación.

Transformaciones

Para cambios lineales

$$Y = a + b \cdot X$$

$$S_Y = |b| \cdot S_X$$

Medidas de Dispersión Relativas

A veces queremos comparar cómo se reparten los datos en distintos grupos, aunque tengan unidades diferentes (por ejemplo, euros y kilos) o valores muy distintos. Para eso usamos **medidas de dispersión relativas**, que nos dicen **cuán dispersos están los datos en relación con su tamaño o posición**.



Recorrido Relativo y Semi-intercuartílico

Ambas son **medidas sin unidades** (adimensionales), lo que permite comparar distribuciones distintas. No cambian si se multiplica todo por un número, pero sí si se suma o resta. Es decir, no son invariantes ante cambios de origen, pero sí de escala.

Recorrido relativo: compara la diferencia entre el valor más alto y el más bajo, pero en proporción al valor máximo.

$$R_r = \frac{Re}{X_{\text{máx}}} = \frac{X_{\text{máx}} - X_{\text{min}}}{X_{\text{máx}}}$$

Recorrido semi-intercuartílico: mide la dispersión del 50% central de los datos, comparando los cuartiles Q1 y Q3.

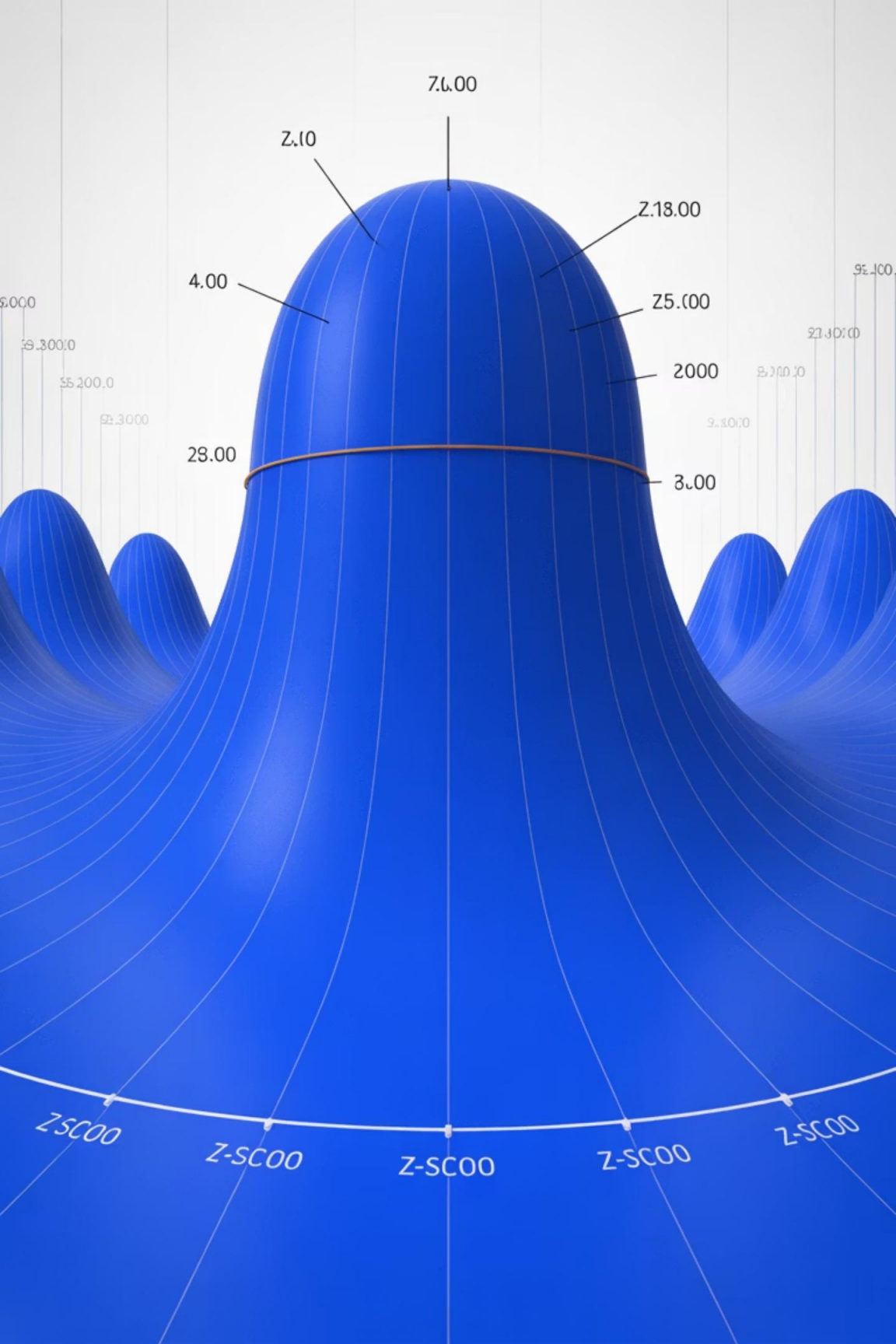
$$R_{SI} = \frac{(C_3 - C_1)}{(C_3 + C_1)}$$

Coeficiente de Variación de Pearson: $CVP = \frac{s}{\bar{X}}$

Versión relativa de la varianza, medida adimensional **fundamental para comparar distribuciones**.

- $CVP < 0.2$ (20%) → dispersión relativa baja y la media es representativa
- $CVP \cong 0$ → menor dispersión relativa y mayor homogeneidad
- **IMPORTANTE:** No usar si $\bar{X} = 0$





Tipificación de Variables

Tipificar una variable es transformar los datos para que todos estén en la misma escala, sin importar sus unidades (como euros, kilos o centímetros).

◆ ¿Cómo se hace?

A cada dato se le resta la media (el valor promedio) y se divide por la desviación típica (una medida de cuánto se separan los datos entre sí).

1

Variable Original $X \rightarrow \begin{cases} \text{Con media } \bar{X} \\ \text{Con Desviación típica: } S_X \end{cases}$

2

Transformación: $Z = \frac{X - \bar{X}}{S_X}$

3

Variable Tipificada
 $\bar{Z} = 0$ y $S_Z = 1$

◆ ¿Para qué sirve?

- Para comparar datos distintos en una misma escala.
- Para detectar valores atípicos (los que se alejan mucho del resto).
- Para trabajar con variables muy diferentes sin que las unidades influyan.

◆ Ejemplo:

Si un alumno tiene una nota muy alta en una asignatura y muy baja en otra, al tipificar podemos ver en cuál se aleja más del promedio de su grupo.

Medidas de forma

Las **medidas de forma** nos ayudan a entender **cómo se distribuyen los datos** más allá de si están más o menos dispersos. Nos dicen si los datos están **equilibrados o desviados respecto a la posición central** (asimetría de Fisher) y si están **muy concentrados o muy repartidos** (curtosis).

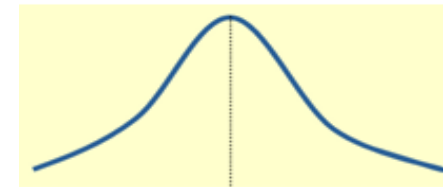
Coeficiente de **Asimetría de Fisher**

$$\varphi_1 = \frac{\sum_{i=1}^k (x_i - \bar{x})^3 n_i}{ns^3}$$

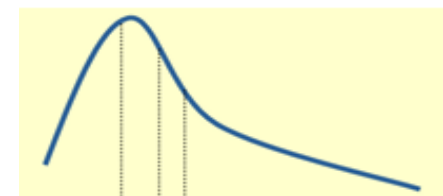
$$\begin{cases} \varphi_1 < 0 \text{ asimétrica a izq} \\ \varphi_1 = 0 \text{ simétrica} \\ \varphi_1 > 0 \text{ asimétrica a der} \end{cases}$$



Asimetría a Izquierda



Simetría

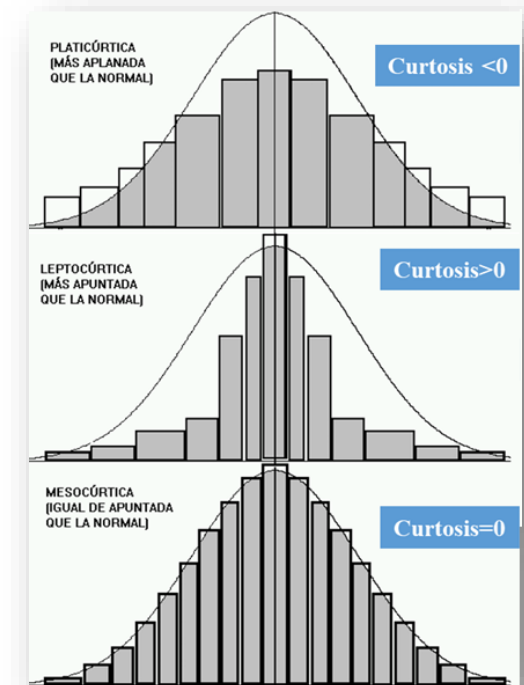


Asimetría a Derecha

Medidas de **Curtosis** o Apuntamiento

$$\varphi_2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^4 n_i}{ns^4} - 3$$

$$\begin{cases} \varphi_2 < 0 \text{ platicúrtica} \\ \varphi_2 = 0 \text{ mesocúrtica} \\ \varphi_2 > 0 \text{ leptocúrtica} \end{cases}$$



ÍNDICES ESTADÍSTICOS PARA V. A. CUANTITATIVAS

Medidas de
Tendencia Central

Media aritmética: $\bar{X} = \frac{\sum_{i=1}^k x_i n_i}{n}$

Media ponderada: $\frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i}$

Mediana: Me: Deja la mitad de la distribución a ambos lados

Moda: Mo: Valor más repetido (el de mayor frecuencia absoluta)

Medidas de
Dispersión

Miden el grado de
variabilidad de los
datos respecto al centro

Desviación Media (DM) = $\frac{\sum_{i=1}^k |x_i - \bar{x}| n_i}{n}$

Varianza Muestral (s^2) = $\frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n} \rightarrow \text{Unidades}^2 \text{ v. a.}$

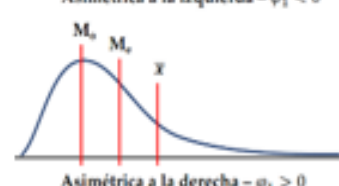
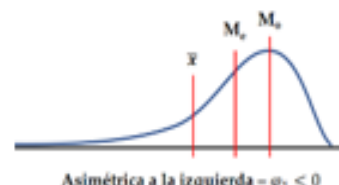
Desviación Estándar o Típica (s) = $\sqrt{s^2} \rightarrow \text{Unidades v. a.}$

Coeficiente de Variación (COV) = $\frac{s}{\bar{x}} 100 \rightarrow \text{Comparar v. a.} \neq \text{unidades}$

Rango (Amplitud): $R = x_n - x_1$

Rango Intercurtílico (RQ) = $Q_3 - Q_1$

Error Estándar de la Media (E.E) = $\sqrt{s^2/n} = s/\sqrt{n}$



Medidas **de Forma**
(Sin unidades)

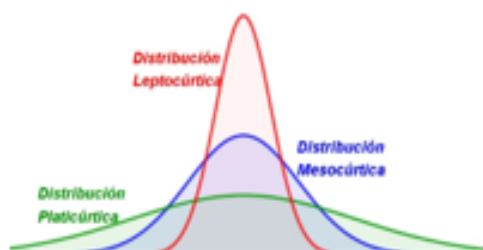
Coeficientes de o
asimetría o sesgo

Asimetría de **Fisher**: $\varphi_1 = \frac{\sum_{i=1}^k (x_i - \bar{x})^3 n_i}{ns^3} = \frac{m^3}{s^3} \begin{cases} \varphi_1 < 0 \text{ asimétrica a izq} \\ \varphi_1 = 0 \text{ simétrica} \\ \varphi_1 > 0 \text{ asimétrica a der} \end{cases}$

Asimetría de **Pearson**: $A_p = \frac{\bar{x} - Mo}{s}$

Curtosis, apuntamiento o deformación: $\varphi_2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^4 n_i}{ns^4} - 3 = \frac{m^4}{s^4} \begin{cases} \varphi_2 < 0 \text{ platicúrtica} \\ \varphi_2 = 0 \text{ mesocúrtica} \\ \varphi_2 > 0 \text{ leptocúrtica} \end{cases}$

Medidas **de Posición** (Cuantiles): Dividen la distribución en k partes iguales: $\begin{cases} k = 3 \rightarrow \text{terciles} (T_r) & k = 4 \rightarrow \text{Cuartiles} (Q_r) \\ k = 10 \rightarrow \text{deciles} (D_r) & k = 100 \rightarrow \text{percentiles} (P_r) \end{cases}$



Interpretación de los valores

Interpreta el significado de las principales **medidas estadísticas** y explicación de cómo las **medidas de dispersión** afectan la **representatividad de las medidas centrales**:

Medida Estadística	Significado	Influencia de la Dispersión
Media aritmética	Promedio de todos los valores.	Muy sensible a valores extremos. Alta dispersión reduce su representatividad.
Media ponderada	Promedio que considera el peso o frecuencia de cada valor.	Si hay alta dispersión en los pesos, puede sesgar el resultado hacia valores más frecuentes.
Mediana	Valor central de la distribución ordenada.	Poco afectada por la dispersión. Representativa incluso con datos muy dispersos.
Moda	Valor que más se repite.	Puede no existir o haber varias. En distribuciones muy dispersas, pierde representatividad.
Desviación estándar (s)	Promedio de las diferencias respecto a la media.	Cuanto mayor sea, más dispersos están los datos. Reduce la fiabilidad de la media como resumen.
Varianza (s²)	Cuadrado de la desviación estándar.	Mide la dispersión total. Alta varianza indica baja representatividad de la media.
Rango (R)	Diferencia entre el valor máximo y mínimo.	Muy sensible a valores extremos. No refleja bien la dispersión interna.
Rango intercuartílico (RQ)	Diferencia entre el tercer y primer cuartil.	Mide la dispersión del 50% central. Menos sensible a valores extremos.
Coeficiente de variación (CV)	Relación entre la desviación estándar y la media (en %).	Permite comparar dispersión entre variables con distintas unidades. $CVP < 0.2$ (20%) → dispersión relativa baja
Error estándar de la media (EEM)	Estima la precisión de la media como representante de la población.	Menor EEM indica mayor fiabilidad de la media.