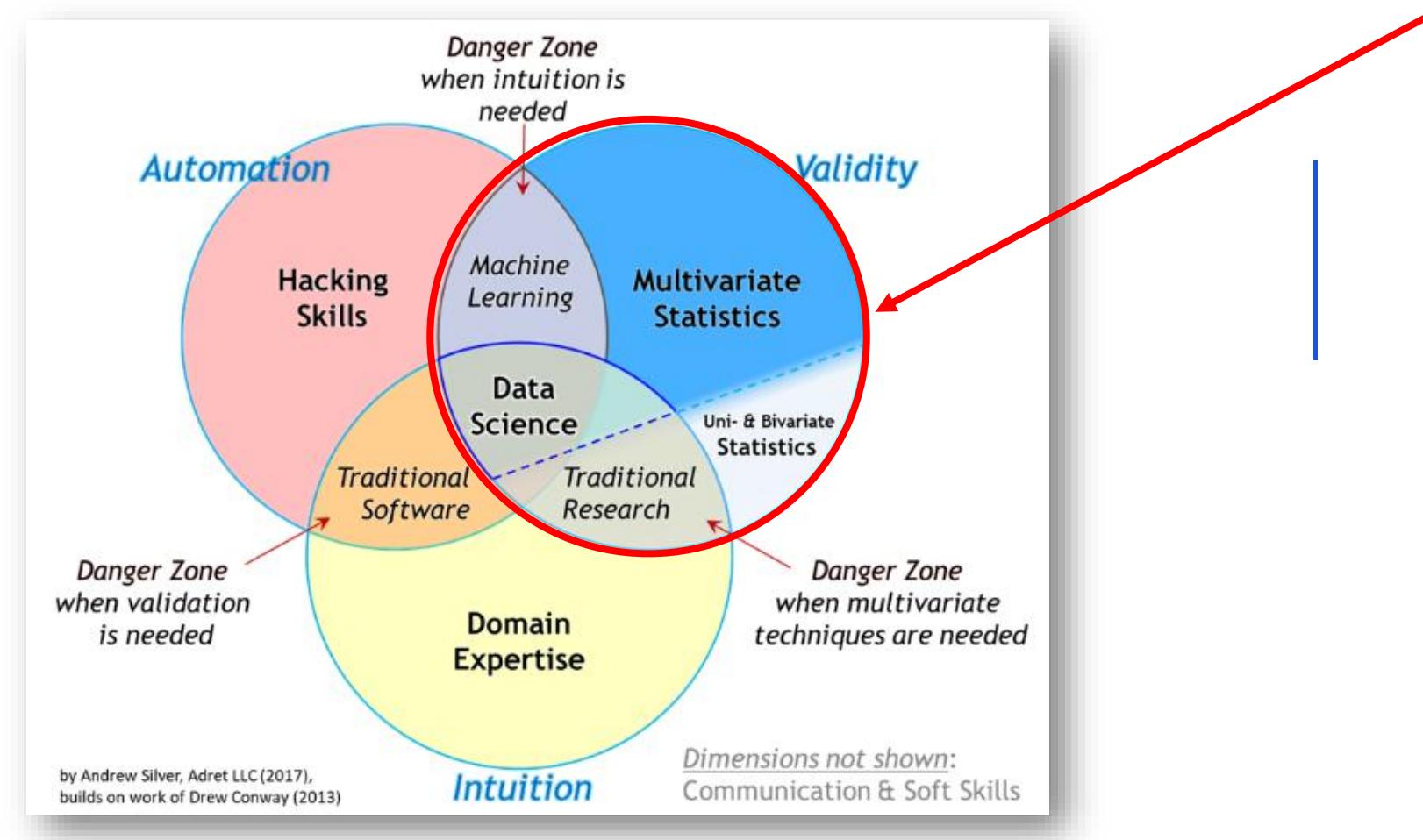


Estadística descriptiva y regresión lineal con Excel



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/>

Estadística Descriptiva y Regresión con Excel

OBJETIVO: Dominar el análisis de datos desde los fundamentos hasta técnicas avanzadas de regression

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

Evaluá 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.

¿Por qué Excel para aprender los conceptos imprescindibles de Estadística?

- Funciones estadísticas integradas y fáciles de usar
- Visualización de datos instantánea y profesional
- Compatible con grandes volúmenes de información
- Herramientas de análisis de regresión incorporadas

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

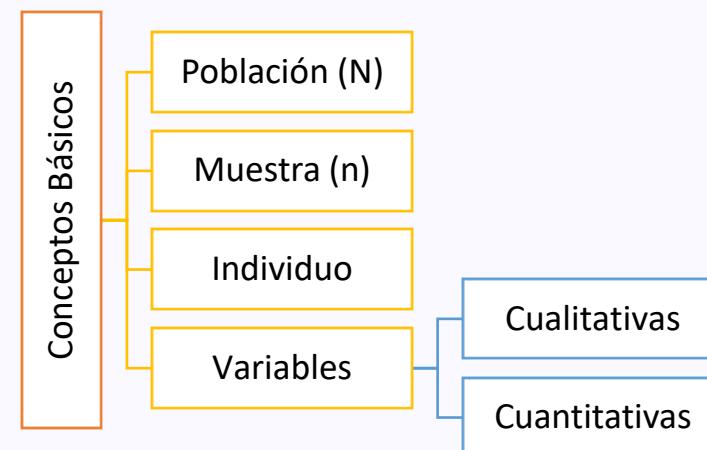
Visualización de Datos

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.



Indique de qué tipo son las siguientes variables, así como su escala de medición:

- Número de ingresos hospitalarios: **cuantitativa discreta – razón.**
- Diabetes Tipo I o Tipo II: **cualitativa dicotómica – nominal.**
- Hipertensión: grado I, grado II o grado III: **cualitativa policotómica – ordinal.**
- Número de piezas cariadas: **cuantitativa discreta – razón.**
- Niveles de urea en plasma (mg/dL): **cuantitativa continua – razón.**
- Tipo de dieta: hipocalórica, diabética, sin sal, hiperproteica: **cualitativa policotómica – nominal.**
- Nivel de estudios: sin estudios, primarios-secundarios, universitarios: **cualitativa policotómica – ordinal.**
- Nota de un examen de matemáticas (de 0 a 10): **cuantitativa discreta – razón.**
- Situación laboral (activo/parado/jubilado): **cualitativa policotómica – nominal.**
- Sugerencias de mejora: **cualitativa policotómica – nominal.**
- Consumo calórico (Kcal): **cuantitativa continua – razón.**
- Número de accidentes de trabajo: **cuantitativa discreta – razón.**
- Índice de masa corporal (Kg/m²): **cuantitativa continua – razón.**

Ahora vamos a ver el punto 04

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

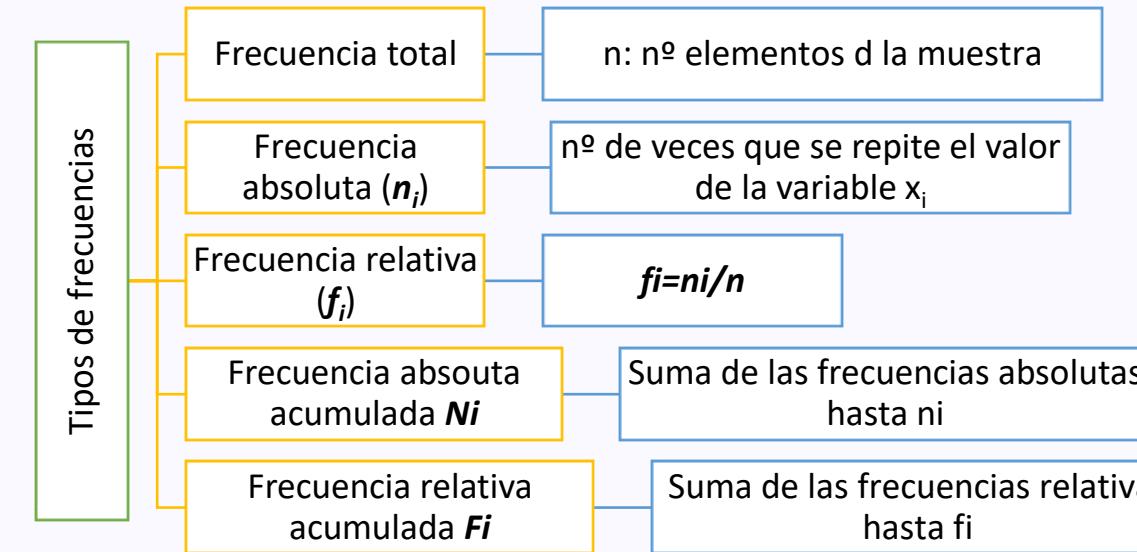
Ajusta curvas no lineales y aplica modelos exponenciales, logarítmicos y polinómicos para análisis sofisticados de datos complejos.

Análisis numérico: Descripción Numérica de Datos Univariantes

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/>

¿Por qué resumir datos?

La distribución de frecuencias ofrece información completa, pero a menudo es imposible interpretar toda esa información extensa. Las medidas descriptivas nos permiten:



Describir

Destacar propiedades importantes de la variable analizada



Comparar

Analizar diferentes conjuntos de datos mediante medidas numéricas



Interpretar

Evaluar la representatividad de los valores calculados



Tipos de Medidas Descriptivas

- .

Medidas de Posición

Indican la situación de la variable en su escala de medida

02

Medidas de Dispersion

Evalúan la representatividad y variabilidad de los datos



- 0

Medidas de Forma

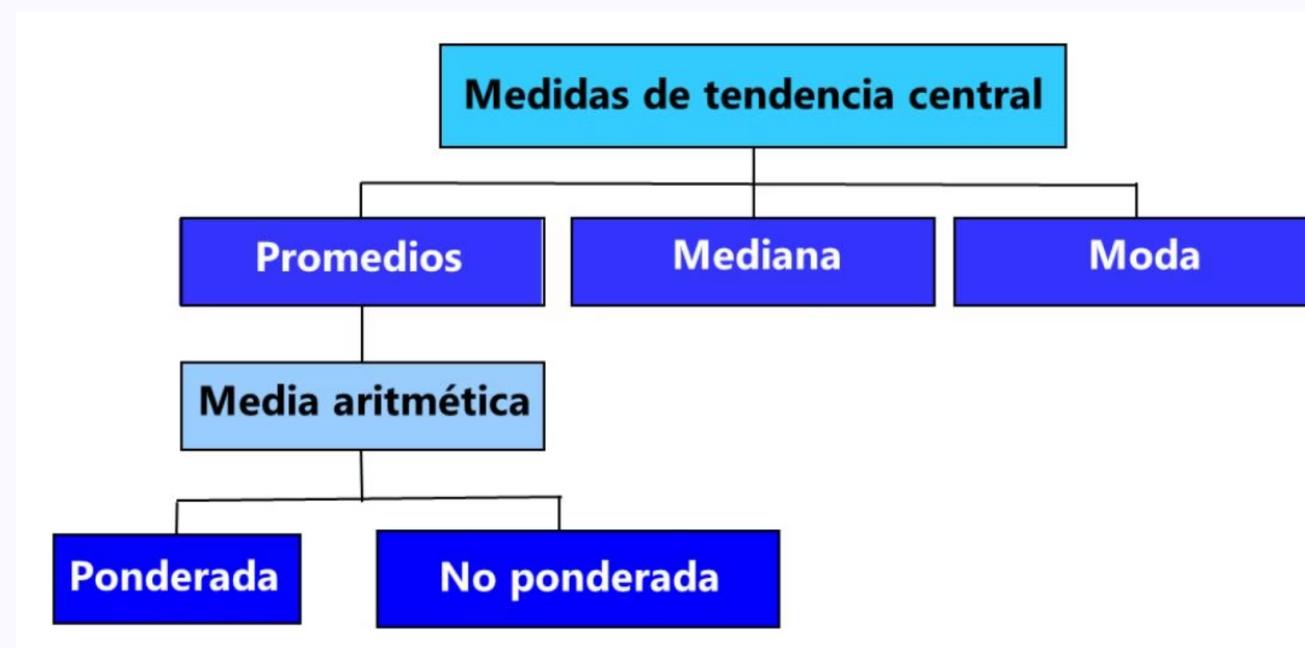
Caracterizan la simetría y apuntamiento de la distribución

Medidas de Posición

Tendencia Central

Ubican el centro representativo de la distribución:

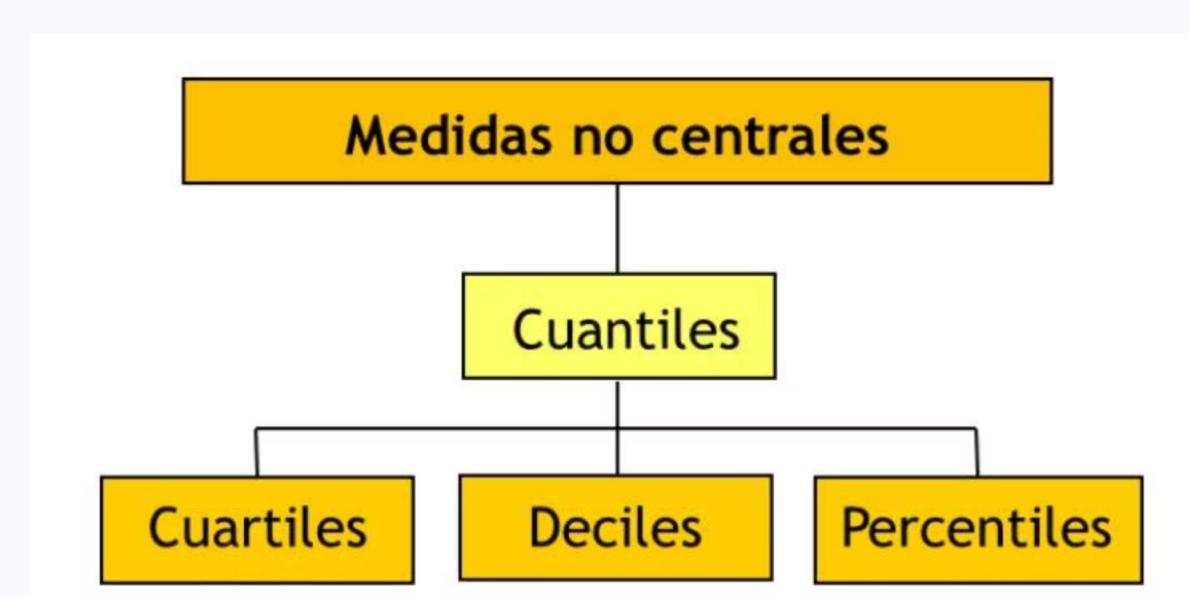
- **Media aritmética:** promedio de todos los valores
- **Mediana:** valor que divide la distribución en dos partes iguales
- **Moda:** valor más frecuente



Tendencia No Central

Identifican otros puntos característicos:

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



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)

Media Aritmética y Mediana

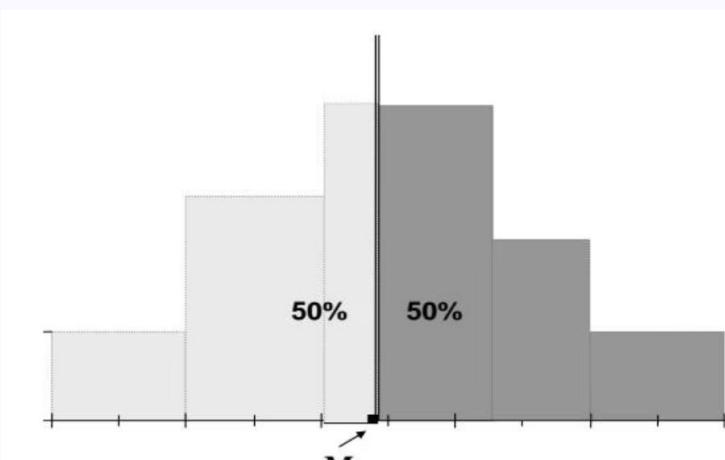
Media Aritmética (\bar{x})

Suma de todos los valores dividida por el número total de observaciones.

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

- Medida única y sencilla
- Utiliza todos los valores
- Centro de gravedad de la distribución
- No robusta ante valores extremos

Valor que divide la distribución en dos partes iguales, dejando 50% de datos a cada lado.



- Más robusta que la media
- No utiliza todos los datos
- Ideal para distribuciones asimétricas

Mediana (Me)

a) En datos no agrupados –repetidos ó no– se ordenan de forma creciente y se identifica el valor o dato que ocupa la posición central. Formalmente, se detecta el entero m que cumple $m - 1 < N/2 \leq m$

- Si N es impar, $Me = x_m$, correspondiendo con el valor central
- Si N es par, puede observarse que hay dos valores centrales, x_m y x_{m+1} y, si la variable es cuantitativa, se tomará como mediana la media de ambos:

$$Me = \frac{x_m + x_{m+1}}{2}$$

Si N es par y la variable no es cuantitativa, se sigue la regla anterior.

b) Para datos agrupados en intervalos determinamos, en primer lugar, el **intervalo mediano**, que es aquél donde la frecuencia acumulada alcanza el 50%. Esto es:

$$N_{m-1} < \frac{N}{2} \leq N_m \quad \text{o} \quad F_{m-1} < 0,5 \leq F_m$$

El valor concreto se puede determinar dentro de dicho intervalo por interpolación lineal:

$$Me = L_{m-1} + \frac{N/2 - N_{m-1}}{n_m} \cdot a_m \quad \text{o} \quad Me = L_{m-1} + \frac{0.5 - F_{m-1}}{f_m} \cdot a_m$$

La Moda: Valor Más Frecuente

La moda (**Mo**) es el valor que más se repite en la distribución, correspondiente a la modalidad con mayor frecuencia.

1

Datos No Agrupados

Identificar la modalidad con mayor frecuencia:

$$Mo = x_j \quad si \quad n_j = \max\{n_i\}$$

2

Datos Agrupados

Determinar el intervalo modal (mayor densidad) y calcular usando

$$\text{formula: } Mo = x_m = \frac{L_{m-1} + L_m}{2}$$

Pueden darse **varias modas** y, dependiendo del número de modas, las distribuciones se clasifican como:

- unimodales
- bimodales
- multimodales

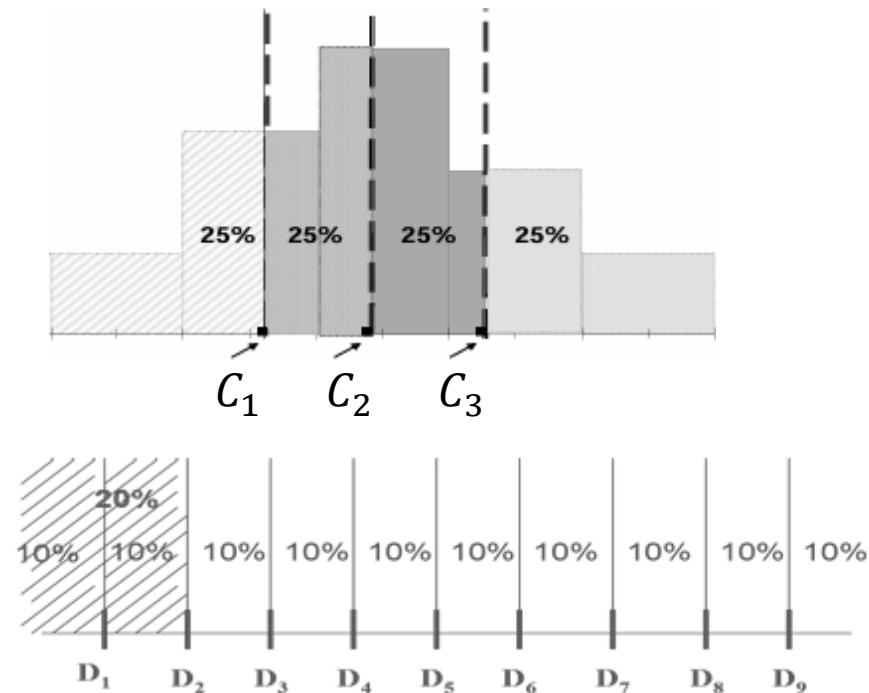
En caso de no ser única, la Moda pierde representatividad.



Propiedades clave: Fácil de interpretar, no afectada por valores extremos, aplicable a cualquier tipo de variable. Pueden existir distribuciones unimodales, bimodales o multimodales.

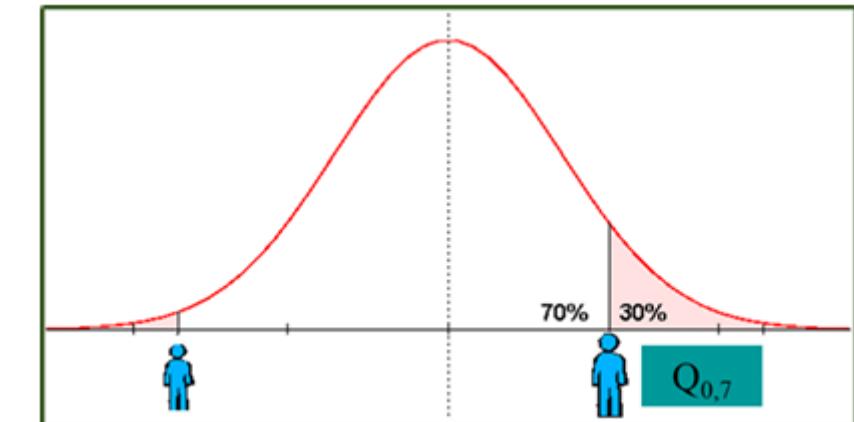
Medidas de Posición (Cuantiles)

$$\left\{ \begin{array}{l} k = 3 \rightarrow \text{terciles} \\ k = 4 \rightarrow \text{Cuartiles} \\ k = 10 \rightarrow \text{deciles} \\ k = 100 \rightarrow \text{percentiles} \end{array} \right.$$



Se define cuantil de orden p , Q_p , como el valor de la variable que divide la distribución en dos partes cuyas frecuencias son p y $1 - p$, es decir, hay una frecuencia igual a p por debajo de él y una frecuencia $1 - p$, por encima de él.

Es una extensión del concepto de mediana



Dividen la distribución en k partes iguales

Cálculo de Cuantiles

Calcular frecuencias acumuladas

Obtener Ni o Fi para cada valor o intervalo

Identificar posición m

tal que $F_{m-1} < 100 \times p\% \leq F_m$

$$N_{m-1} < N \times p \leq N_m$$

Aplicar fórmula según tipo

Si la variable es discreta

- Si $F_m > 100 \times p\%:$ $Q_p = x_m$
- Si $F_m = 100 \times p\%:$ $Q_p = \frac{x_m + x_{m+1}}{2}$

Si la variable es continua **agrupada** en intervalos:

$$Q_p = L_{m-1} + \frac{N \cdot p - N_{m-1}}{n_m} \cdot a_m$$

$$Q_p = L_{m-1} + \frac{p - F_{m-1}}{f_m} \cdot a_m$$

