# Medidas de tendencia central y de dispersión – Asimetría y Curtosis

*Penna – Cobos – Vázquez Ferrero – Ulagnero*

# CONTEXTO DENTRO DEL PROCESO DE INVESTIGACIÓN (ETAPAS)

- **1ª**: Planeamiento del problema

¿Qué se necesita saber?

- **2ª**: Planeación

¿Qué recursos se requieren?
¿Qué actividades son necesarias?

- **3ª**: Recopilación de datos

¿Cómo se recogen los datos?
¿Con qué instrumento/s?

- **4ª**: Procesamiento de datos

¿Cómo reducir la información a unas cifras?
¿Qué riesgos se corren?

- **5ª**: Explicación e interpretación

¿Qué significan los resultados?

- **6ª**: Comunicación / decisión / solución

# MEDIDAS DE TENDENCIA CENTRAL (MTC)

- Son **valores típicos** alrededor de los cuales se agrupan los datos de la población o muestra, y que representan dicho conjunto. Estos valores se encuentran contenidos entre el mínimo y el máximo de la distribución.
- Son **índices** o **indicadores** que contienen información sustancial acerca de las propiedades de un grupo estudiado y expresan las característica del grupo. Su valor práctico reside en la reducción del conjunto inicial de observaciones.

En Ciencias Humanas, las más utilizadas son:

- $\bar{X}$ **= Media aritmética**
- $\tilde{X}$ **= Mediana**
- $\hat{X}$ **= Moda o Modo**

# Media Aritmética [$\bar{X}$ o $M(X)$]

- Medida que resulta de sumar todos los valores obtenidos, divididos por el número total de observaciones.
- Es la MTC más representativa.
- Se ve afectada por la presencia de valores extremos.
- Se calcula para variables con nivel de medición Intervalar o Métrico.
- Debe complementarse con alguna medida de dispersión.

$$\bar{X} = M(X) = \frac{x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

<u>*Propiedades*</u>: sean X e Y variables aleatorias; *a* y *b* constantes, entonces:

- $M(a) = a$
- $M(aX) = aM(X)$
- $M(aX \pm bY) = aM(X) \pm bM(Y)$

# MEDIANA $\left[\widetilde{\mathrm{X}} \text{ o } \mathbf{M_e}\right]$

- Dados *n* valores observados, ordenados de menor a mayor o de mayor a menor, la mediana es el valor de la distribución que ocupa la posición: $(\widetilde{\mathrm{X}})_0 = \frac{n}{2} + 0{,}5$. Y, por ende, dicho valor -observable o no- deja a su izquierda (o por debajo) el 50% de la distribución y a su derecha (o por encima) el 50% restante.
- No se ve afectada por la presencia de valores extremos (robusta).
- Puede calcularse en variables cuantitativas con cualquier nivel de medición.

# MODA $\left[\widehat{\mathrm{X}} \text{ o } \mathbf{M_o}\right]$

- Dados *n* valores observados, la moda es el valor (o valores) de la distribución que se presenta(n) con mayor frecuencia. La distribución puede no tener moda, o ser unimodal, bimodal o multimodal.
- Puede calcularse para cualquier tipo de variable con cualquier nivel de medición.

# MTC: *Ejemplo*

Los siguientes datos representan las edades de 10 estudiantes:

10 12 11 11 13 10 11 12 10 13

**Media aritmética**:

$$\bar{\mathbf{X}} = \frac{10 + 12 + 11 + 11 + 13 + 10 + 11 + 12 + 10 + 13}{10} = \frac{113}{10} = \mathbf{11{,}3}$$

$(\tilde{\mathbf{X}})_0 = \mathbf{5{,}5}$

**Mediana**: 10 10 10 11 11 11 12 12 13 13

La posición de la mediana es $(\tilde{\mathrm{X}})_0 = \frac{10}{2} + 0{,}5 = 5{,}5 \Rightarrow \tilde{\mathbf{X}} = \frac{11+11}{2} = \mathbf{11}$

**Moda**: 10 10 10 11 11 11 12 12 13 13

InfoStat

La distribución es bimodal ya que: $\hat{\mathbf{X}}_1 = \mathbf{10}$ y $\hat{\mathbf{X}}_2 = \mathbf{11}$

# MEDIDAS DE POSICIÓN: PERCENTILES

Dados $n$ valores observados, ordenados de menor a mayor, el percentil (o centil) i-ésimo, es el valor de la distribución que ocupa la posición:

$$(\mathrm{P_i})_0 = \frac{(\mathrm{i} \times n)}{100} + 0{,}5.$$

10 10 10 11 11 11 12 12 13 13

InfoStat

Queremos calcular el $\mathbf{P_{25}}$

La posición del $\mathrm{P_{25}}$ es $(\mathrm{P_{25}})_0 = \frac{(25 \times 10)}{100} + 0{,}5 = \mathbf{3} \Rightarrow \mathbf{P_{25}} = \mathbf{10}$

Queremos calcular el $\mathbf{P_{75}}$

La posición del $\mathrm{P_{75}}$ es $(\mathrm{P_{75}})_0 = \frac{(75 \times 10)}{100} + 0{,}5 = \mathbf{8} \Rightarrow \mathbf{P_{75}} = \mathbf{12}$

Queremos calcular el $\mathbf{P_{50}}$

La posición del $\mathrm{P_{50}}$ es $(\mathrm{P_{50}})_0 = \frac{(50 \times 10)}{100} + 0{,}5 = \mathbf{5{,}5} \Rightarrow \mathbf{P_{50}} = \frac{11+11}{2} = \mathbf{11} = \tilde{\mathrm{X}}$

# MEDIDAS DE DISPERSIÓN (MD)

Como vimos, la media y la mediana reducen la distribución de datos a un solo valor que será más o menos representativo dependiendo de la dispersión de los puntajes individuales. Las MD indican la variabilidad (concentración o dispersión) de los datos en torno a una MTC, y se calculan sólo para variables cuantitativas con nivel de medición intervalar o métrico. Siendo algunas de las MD, las siguientes:

- **R = Rango o Recorrido**
- **$R_I$ = Rango o Recorrido Intercuartílico**
- **$S^2$ = Varianza**
- **S = Desviación Estándar**
- **MAD**
- **CV = Coeficiente de Variación**

# Rango o Recorrido [R]

- Es la primera aproximación a la dispersión de los datos.
- Considera los “extremos” de la distribución sin tener en cuenta qué ocurre al interior de la misma.
- Se la interpreta como el número de unidades que contienen a toda la distribución.

$$\mathbf{R = x_{Max} - x_{Min}}$$

# Rango o Recorrido Intercuartílico [$R_I$]

- Es una medida adecuada cuando la distribución presenta valores extremos.
- Se la interpreta como el número de unidades que contienen el 50% central de la distribución.

$$\mathbf{R_I = P_{75} - P_{25}}$$

# VARIANZA $\left[S^2 \text{ o } V(X)\right]$

- Es la suma de los cuadrados de las desviaciones respecto de la media (desvíos reales), dividido por el número de observaciones menos 1, y se la puede pensar como un “promedio” de las distancias -al cuadrado- entre cada valor de la distribución y su media.
- Al trabajar con los cuadrados hace que tengan gran influencia, sobre el resultado final, aquellas desviaciones grandes.

$$\mathbf{S^2 = V(X) = \sum_{i=1}^{n} \frac{(x_i - \bar{X})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2}$$

# DESVIACIÓN ESTÁNDAR [S o DE]

- Es la raíz cuadrada de la Varianza.
- Mantiene las mismas unidades que los datos de la distribución.

$$\mathbf{S = DE = +\sqrt{S^2}}$$

# Desviación absoluta respecto a la Mediana [MAD]

- Frente a la presencia de valores extremos (outliers), es conveniente utilizar la mediana en lugar de la media aritmética; y una buena medida de dispersión para acompañarla es la MAD.
- Esta MD, es la mediana del valor absoluto de las distancia entre cada valor de la distribución y su mediana.

$$\mathbf{MAD} = \mathbf{Mediana}\left\{\left|\mathbf{x_1} - \tilde{\mathbf{X}}\right|, \left|\mathbf{x_2} - \tilde{\mathbf{X}}\right|, \dots, \left|\mathbf{x}_{\boldsymbol{n-1}} - \tilde{\mathbf{X}}\right|, \left|\mathbf{x}_{\boldsymbol{n}} - \tilde{\mathbf{X}}\right|\right\}$$

# Coeficiente de Variación: [CV]

- Es una medida de dispersión relativa y surge del cociente entre la DE y $|\overline{X}|$, multiplicado por 100.
- Es un número adimensional y por ello al multiplicarlo por 100, se expresa en porcentaje.
- Determina cuan representativa es la media aritmética de un conjunto de datos: si el CV ≤ 20%, la media es representativa.

$$\mathbf{CV} = \frac{\mathbf{S}}{|\overline{\mathbf{X}}|} \times \mathbf{100}$$

# MD: *Ejemplo*

Continuando con los datos de las edades de los 10 estudiantes:

10 12 11 11 13 10 11 12 10 13

**Rango o Recorrido**:

$$\mathbf{R} = \mathrm{x}_{\mathrm{Max}} - \mathrm{x}_{\mathrm{Min}} = 13 - 10 = \mathbf{3}$$

**Rango o Recorrido Intercuartílico**:

Ordenando los datos: 10 10 10 11 11 11 12 12 13 13

Como sabemos que: $P_{25} = 10$ y $P_{75} = 12$ $\Rightarrow$ $\mathbf{R_I} = P_{75} - P_{25} = 12 - 10 = \mathbf{2}$

InfoStat

**Varianza**: ya que $\overline{\mathrm{X}} = 11{,}3$

$$\mathbf{S^2} = \sum_{\mathrm{i}=1}^{10} \frac{(\mathrm{x_i} - \overline{\mathrm{X}})^2}{n-1} = \frac{(10-11{,}3)^2 + (12-11{,}3)^2 + \cdots + (10-11{,}3)^2 + (13-11{,}3)^2}{9} =$$

$$= \frac{1{,}69 + 0{,}49 + \cdots + 1{,}69 + 2{,}89}{9} \cong \mathbf{1{,}34}$$

# MD: $\boldsymbol{Ejemplo\ (cont.)}$

**Desviación Estándar**:

$$\text{Como } S^2 = 1{,}34 \ \Rightarrow \mathbf{S} = +\sqrt{1{,}34} = \mathbf{1{,}16}$$

**MAD**:

$$\text{Siendo } \tilde{X} = 11 \Rightarrow$$

$$\mathbf{MAD} = \text{Mediana}\,\{|10-11|, |12-11|, \ldots, |10-11|, |13-11|\} =$$

$$= \text{Mediana}\,\{1, 1, \ldots, 1, 2\} = \mathbf{1}$$

**CV**:

$$\text{Ya que } \bar{X} = 11{,}3 \text{ y } S = 1{,}16 \Rightarrow$$

$$\mathbf{CV} = \frac{1{,}16}{|11{,}3|} \times 100 \cong \mathbf{10{,}27\%}$$

InfoStat

# MOMENTOS

Los momentos de una variable aleatoria X son los valores esperados de ciertas funciones de X y forman una colección de medidas descriptivas que pueden emplearse para caracterizar la distribución de probabilidad de X y especificarlas (siempre que todos los momentos de X sean conocidos). Además determinan que: *dos distribuciones de probabilidad son iguales, si tienen todos sus momentos iguales*. Donde:

- El primer momento alrededor del cero, es la media o valor esperado de la variable aleatoria y se denota con μ.
- El segundo momento centrado alrededor de la media, recibe el nombre de varianza de la variable aleatoria y se denota con $\sigma^2$.
- El tercer momento centrado, está relacionado con la asimetría de la distribución de probabilidad de X; la cual es llamada sesgo.
- El cuarto momento centrado, es una medida de cuan concentrada es la distribución de probabilidad y está asociado con la curtosis (o apuntamiento).

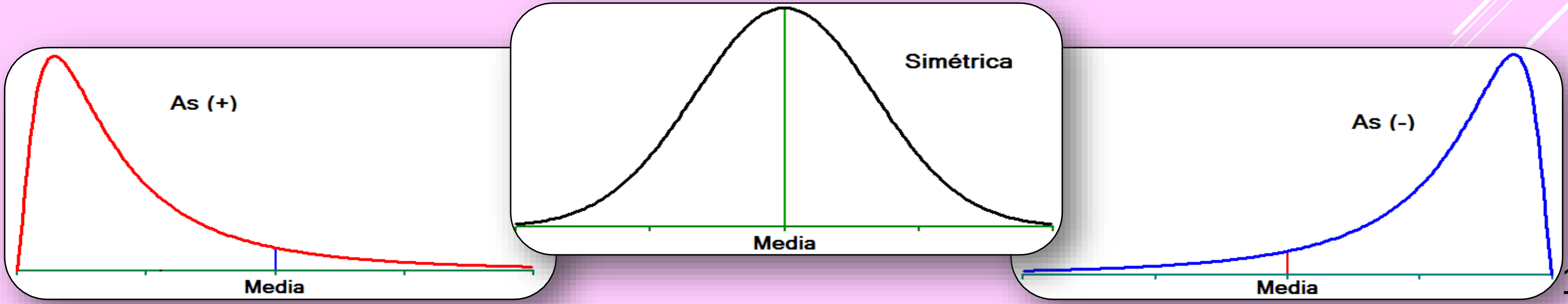
Siendo: $\boldsymbol{m_r} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^r}{n}$ el momento centrado de orden r.

# COEFICIENTE DE ASIMETRÍA DE FISHER $[A_S]$

$$\mathbf{A_S} \cong \frac{\boldsymbol{m_3}}{\boldsymbol{S^3}}$$

Comportamiento de los datos alrededor de la media.

- Si $\mathbf{A_S} > \mathbf{0,3} \Rightarrow$ la asimetría es <u>positiva</u>, $A_S(+)$, y presenta mayor concentración de valores a la izquierda de la media.
- Si $-\mathbf{0,3} \leq \mathbf{A_S} \leq \mathbf{0,3} \Rightarrow$ Si $A_S = 0$, la distribución es perfectamente <u>simétrica</u> y si $A_S \cong 0$ será aproximadamente simétrica o levemente asimétrica, alrededor de la media.
- Si $\mathbf{A_S} < -\mathbf{0,3} \Rightarrow$ la asimetría es <u>negativa</u>, $A_S(-)$, y presenta mayor concentración de valores a la derecha de la media.

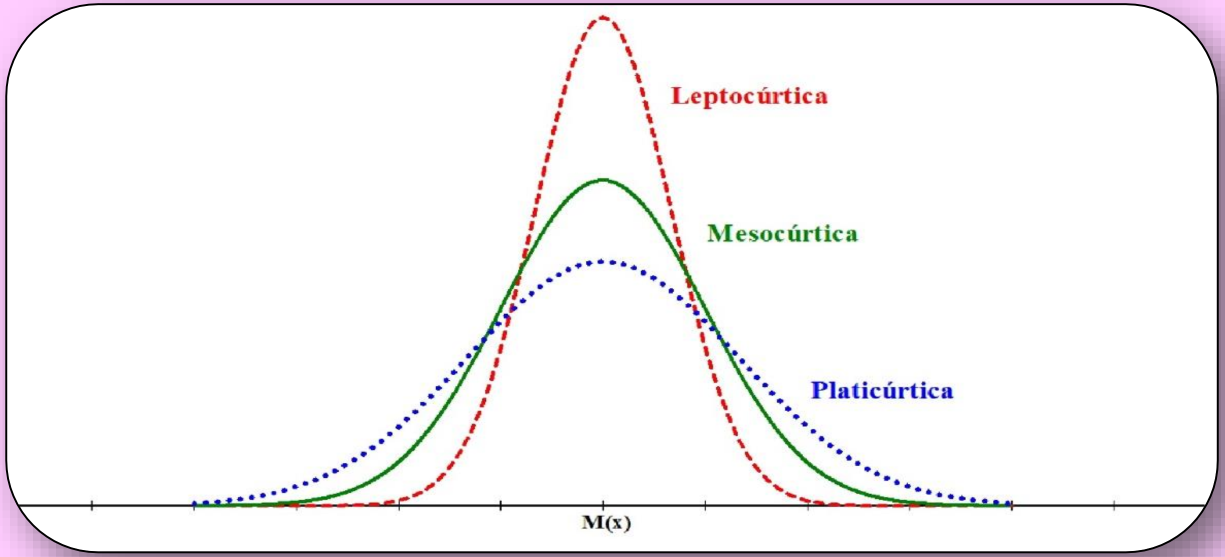


# COEFICIENTE DE CURTOSIS $[\mathrm{C_U}]$

$$\mathbf{C_U} \cong \frac{\boldsymbol{m_4}}{\boldsymbol{S^4}} - \mathbf{3}$$

Grado en que las observaciones están agrupadas en torno a la media.

- Si $\mathbf{C_U} > \mathbf{0,3} \Rightarrow$ la distribución es <u>leptocúrtica</u>: más homogénea respecto a la media.
- Si $-\mathbf{0,3} \leq \mathbf{C_U} \leq \mathbf{0,3} \Rightarrow$ la distribución es <u>mesocúrtica</u>: se distribuye uniformemente respecto a la media.
- Si $\mathbf{C_U} < -\mathbf{0,3} \Rightarrow$ la distribución es <u>platicúrtica</u>: más heterogénea respecto a la media.



# Asimetría y Curtosis: *Ejemplo*

Tomando las edades de los alumnos y recordando que: $\overline{X} = 11{,}3$ y $S = 1{,}16$:

10 12 11 11 13 10 11 12 10 13

**Asimetría**:

$$\boldsymbol{m_3} = \frac{\sum_{i=1}^{10}(x_i - 11{,}3)^3}{10} = \frac{(10-11{,}3)^3+(12-11{,}3)^3+\cdots+(10-11{,}3)^3+(13-11{,}3)^3}{10} \cong 0{,}384$$

$$\Rightarrow \mathbf{A_S} \cong \frac{m_3}{(1{,}16)^3} \cong \frac{0{,}384}{1{,}561} \cong \mathbf{0{,}246}$$

**Curtosis**:

$$\boldsymbol{m_4} = \frac{\sum_{i=1}^{10}(x_i - 11{,}3)^4}{10} = \frac{(10-11{,}3)^4+(12-11{,}3)^4+\cdots+(10-11{,}3)^4+(13-11{,}3)^4}{10} \cong 2{,}578$$

InfoStat

$$\Rightarrow \mathbf{C_U} \cong \frac{m_4}{(1{,}16)^4} - 3 \cong \frac{2{,}578}{1{,}811} - 3 \cong \mathbf{-1{,}576}$$

# REFERENCIAS BIBLIOGRÁFICAS


- Bologna, E. (2011). *Estadística para Psicología y Educación*. Córdoba: Brujas.
- Gorgas García, J., Cardiel López, N. & Zamorano Calvo, J. (2009). *Estadística Básica para Estudiantes de Ciencias*. Madrid: Departamento de Astrofísica y Ciencias de la Atmósfera. Facultad de Ciencias Físicas. Universidad Complutense de Madrid.
- Penna, F.O., Esteva, G.C., Cobos, O.H. & Ulagnero, C.A. (2018). *Fórmulas y Tablas III (para cursos de Estadística básica)* (2ª ed.). San Luis: Nueva Editorial Universitaria.
- Sabulsky, J. (2000) *Investigación científica en salud-enfermedad*. (3ª ed.). Córdoba: Ed. Kosmos.
- Triola M. (2018). *Estadística* (12ª ed.). México: Pearson Educación.