BLOQUE 2 – TEMA 5
MEDIDAS DESCRIPTIVAS BÁSICAS Y REPRESENTACIONES BÁSICAS

De la organización a la descripción de los casos

- En Estadística, el concepto más importante es la **VARIABILIDAD**, es decir, el estudio de la dispersión de las puntuaciones.
- En Estadística, pretendemos estudiar la magnitud de la variabilidad y alguna otra característica ligada a la misma. Cuando efectuamos contraste de hipótesis, se tratará de encontrar posibles variables asociadas (que expliquen) la variabilidad de los datos.
- **Variables** → características o propiedades que pueden variar de unos sujetos a otros.
- **Constante** → contrario a variable. Cuando todos los sujetos tienen la misma puntuación, cuando no hay diferencias y por tanto, no hay variabilidad.
- Principales medidas en Estadística descriptiva**:**
    - medidas de **tendencia central.**
    - medidas de **variabilidad o dispersión**.
    - **representaciones gráficas** de las distribuciones de datos.

Medidas de tendencia central

1º) Primeramente se estudia la tendencia central del grupo. Es el caso de la **MEDIA ARITMÉTICA**. Este índice por sí solo no nos proporciona suficiente información para conocer el grupo.

- Solo nos indica que el grupo tiende hacia una puntuación media.

$$\bar{X} = \frac{suma\ \ de\ \ todas\ \ las\ \ puntuaciones}{número\ \ total\ \ de\ \ sujetos} = \frac{\sum Xi}{N}$$

- Para interpretar correctamente una x̄, es conveniente conocer la **puntuación mínima y máxima** de la escala de medida de la variable y situar la media dentro de ese recorrido. (De 0 a 10 de notas, no es lo mismo una x̄ de 2.5; que otra x̄ de 7.5).
- La x̄ solo puede utilizarse en variables cuyo **nivel de medida sea de intervalo o de razón**, aunque sí es posible hacerlo en **medidas nominales dicotómicas**.

2º) La **MODA** es el valor de frecuencia absoluta más alta, la puntuación que más se repite en la distribución.

- La moda no necesita ningún cálculo, es directamente la puntuación. Podemos encontrar distribuciones **bimodales**, con dos puntuaciones de moda; ó **plurimodales**, con más de dos puntuaciones que tienen la misma frecuencia alta.

3º) La **MEDIANA** es el valor de la distribución que deja por encima y por debajo de si al 50% de los sujetos de la muestra, es decir, el valor central de la distribución de frecuencias.

- Ordenadas las puntuaciones de menor a mayor, la puntuación que ocupa la **posición central**, esa es la mediana. Coincide con el **percentil 50**.
- Para calcular la mediana, observamos la columna de frecuencias acumuladas y comprobamos qué puntuación ocupa el lugar central (50%), es decir, **buscamos N/2**.
- **Cuando N sea par**, no tenemos una puntuación central, sino dos. En este caso, se recomienda hallar la media aritmética de las puntuaciones correspondientes a esa `posición central. (ver tablas 5.1; 5.2 y 5.3 págs. 98 y 99)
- **La x̄ es sensible a las puntuaciones extremas, pero no ocurre esto con la Md , ni con la Mo**. Por ello, la Md puede ser una medida preferible a la x̄ cuando las puntuaciones extremas pueden distorsionar la verdadera tendencia central del grupo.
- Cuando **el nivel de medida sea ordinal**, solo podremos utilizar la Md y la Mo. Con medidas nominales, solo podemos utilizar la Mo (excepto en el caso de dicotómicas).

Medidas de variabilidad o dispersión

El índice de tendencia central debe ir acompañado por un índice de dispersión o variabilidad que indique en qué grado las puntuaciones de los sujetos se dispersan o varían en torno a la x̄.

1º) la **DESVIACIÓN MEDIA** corresponde a una media aritmética de las desviaciones de las puntuaciones directas respecto de la x̄. Si la DM es alta, significa que las puntuaciones están alejadas o se desvían mucho de su x̄; mientras que si es pequeña, significa que las puntuaciones del grupo están próximas a la x̄.

$$\boxed{DM = \frac{\sum|(Xi - \bar{x})|}{N}}$$

2º) La **DESVIACIÓN TÍPICA** es el índice más conocido y utilizado, representándose por **s** (estadístico) ó por **σ** (parámetro). El mismo, pero elevado al cuadrado, recibe el nombre de **VARIANZA** (**s²** ó **σ²**)

$$\boxed{s = \sqrt{\frac{\sum(Xi - \bar{x})^2}{N}}} \qquad \boxed{s^2 = \frac{\sum(Xi - \bar{x})^2}{N}}$$

- Otras fórmulas: sin agrupar los datos (en PD)

$$\boxed{S^2 = \frac{\sum Xi^2}{N} - (\bar{x})^2}$$

- agrupados en distribución de frecuencias (en PD)

$$\boxed{\begin{array}{l} \bar{x} = \frac{\sum fiXi}{N} \\ S^2 = \frac{\sum fiXi^2}{N} - (\bar{x})^2 \end{array}}$$

Xi	fiXi	Xi²	fiXi²
	$\sum$		$\sum$

- Para un solo grupo, los valores de referencia son entre cero, s mínima; y el valor máximo, que corresponde al valor de $\frac{puntuaci\acute{o}n\ mayor\ -\ puntuaci\acute{o}n\ menor}{2}$
- Cuando la muestra es pequeña (N ≤ 30) ó queremos estimar la **σ**, utilizamos el cálculo del valor insesgado. La **σ** debe ser mayor que **s** por lógica, pues al haber más sujetos, habrán mayores diferencias entre los sujetos.

$$\boxed{s = \sqrt{\frac{\sum(Xi - \bar{x})^2}{N-1}}} \qquad \boxed{s^2 = \frac{\sum(Xi - \bar{x})^2}{N-1}}$$

3º) La **AMPLITUD o RECORRIDO** de una variable se calcula como la diferencia entre la puntuación mayor y la menor, más uno:

$$\boxed{A = (Xi)mayor - (Xi)memor + 1}$$

* Esta medida se utiliza sobre todo para organizar los datos de una distribución de frecuencias en intervalos y realizar gráficos.
* También nos sirve para interpretar el valor de la **s**. Cuando tenemos s = 2, en una A = 5, es una **s** muy alta. Sin embargo, con s = 2 y A =100, entonces es una s muy pequeña.

* La **A** se utiliza como medida de dispersión solo cuando no se puede calcular otra ó como complemento de la Mo, en niveles de medida nominales.

4º) La **DESVIACIÓN SEMI-INTERCUARTÍLICA**, representada por **Q_2,** indica la dispersión en el 50% central de la distribución.

* Es adecuada cuando el nivel de medida es ordinal (coincidiendo con la Md)
* También es adecuada cuando existan puntuaciones extremas que pueden distorsionar en exceso la S, puesto que en el valor de Q se prescinde del 25% inferior y del 25% superior, calculando el valor del Q_{50} entre los percentiles 25 y 75

$$Q_2 = Q_{50} = \frac{Q_{75} - Q_{25}}{2} = \frac{Q_3 - Q_1}{2}$$

5º) Otro índice de variabilidad es el **CORFICIENTE DE VARIACIÓN (V)**, que nos permite comparar la variabilidad entre variables que tienen distinta amplitud. Dado que dos S procedentes de instrumentos con distinta A, ó diferente escala de medida no son directamente comparables, calculamos V en porcentajes:

$$V = \left(\frac{S}{\bar{x}}\right) 100$$

* A mayor valor de V, más heterogéneo es el grupo y más variabilidad tienen sus medidas.

<u>Media y desviación típica para variables dicotómicas</u>

Como tienen solo dos valores, ceros para noes o respuestas incorrectas, y unos para síes o respuestas correctas, la $\bar{x}$ será la proporción de unos.

- La $\bar{x}$ indica la proporción de sujetos que responden con 1 y se representa por **p**. Con N = 50 sujetos y de ellos 30 responden que sí, entonces : p = 30/50 = 0,6
- La proporción de noes ó ceros, será (1-**P**), en este caso 0,4, y se representa por **q**. Entonces **p+q=1**
- Como la media nos indica el porcentaje de sujetos que contesta acertadamente a una pregunta de examen, se denomina además **índice de dificultad del ítem**.
- La desviación típica es fácil de calcular:

$$S = \sqrt{p \cdot q} \qquad S^2 = p \cdot q$$

<u>Asimetría y apuntamiento: relación con la curva normal</u>

La asimetría y el apuntamiento son dos características relativas a la forma gráfica de la distribución de frecuencias. El modelo de comparación es la CURVA NORMAL, que carece de asimetría, es decir, es simétrica o tiene un índice de asimetría igual a cero.

- ver fig. 5.1 pág.106
- La asimetría **positiva** indica que la mayoría de los sujetos se concentran en la parte baja de las puntuaciones y que la cola de la distribución se sitúa a la derecha. En general, podemos afirmar que en estos casos, Mo ≤ Md ≤ $\bar{x}$
- La asimetría **negativa** indica que los sujetos tienden a agruparse en torno a las puntuaciones altas de la distribución. En general, encontramos que $\bar{x}$ ≤ Md ≤ Mo
- El hablar de positiva ó negativa se debe al signo del cálculo del índice, denominado **índice de Pearson**:

$$As = \frac{\bar{x} - Mo}{S}$$

Si As > 0 → asimetría positiva
Si As < 0 → asimetría negativa
Si As = 0 → distribución simétrica

- Otras fórmulas:

$$As = \frac{3(\bar{x} - Md)}{S}$$

$$As = \frac{(Q3 - Q2) - (Q2 - Q1)}{(Q3 - Q1)}$$

El **apuntamiento ó curtosis** indica el grado en el que la distribución es más o menos picuda; o más o menos abierta o dispersa respecto a su media.

- más chata ó aplanada → **platicúrtica** → $g_2 < 0{,}263$
- más apuntada y estrecha → **leptocúrtica** → $g_2 > 0{,}263$
- curva nornal → **mesocúrtica** → $g_2 = 0{,}263$

ver fig. 5.2 pág. 108

$$g_2 = \left(\frac{1}{N} \cdot \frac{\sum fi\,(Xi - \bar{x})^4}{\sigma^4}\right) - 3$$

<u>Represetnaciones gráficas</u>

Ver páginas 108 a 114 (a.i.)