CAPITULO 3: MEDIDAS DE VARIABILIDAD Y ASIMETRÍA

1. INTRODUCCIÓN

- En el tema 1 veíamos que la distribución de frecuencias tiene tres propiedades: tendencia central, variabilidad y asimetría. Las medidas de tendencia central las hemos visto en el tema 2, ahora vamos a ver las medidas de las otras dos propiedades (variabilidad y asimetría).

2. MEDIDAS DE VARIABILIDAD

- La **variabilidad o dispersión** hace referencia al grado de variación que hay en un conjunto de puntuaciones (ejemplo de gráficos con más y menos dispersión – Pág. 103)
- Cuanto menor es la variabilidad en una distribución, más homogénea es la muestra de sujetos en la variable que estamos midiendo. El caso extremo de máxima homogeneidad es que todos los valores serían iguales entre sí y no habría variabilidad.

Para cuantificar la dispersión de los datos se distinguen 2 tipos de índices:

- \- Los que miden el grado en que las puntuaciones se asemejan o diferencian entre sí: **Amplitud total o rango y amplitud semi-intercuartil**
- \- Los que la dispersión de mide con respecto a alguna medida de tendencia central como la media aritmética: **Varianza y desviación típica.**

Amplitud total o rango

- La amplitud total, (AT) de un conjunto de puntuaciones es la distancia que hay en una escala numérica entre los valores que representan la puntuación máxima y la puntuación mínima. (ejemplo en página104, muy fácil. Hay que tener en cuenta lo que eran los límites exactos, superior e inferior, que vimos en el primer tema)

$$\mathbf{A_T = X_{max} - X_{min}}$$

- Sin embargo esta medida sólo aporta datos de los valores extremos, pero no nos dice la poca o mucha dispersión que pueda existir en el resto de valores.

Varianza y desviación típica

- La medida de variabilidad también se puede basar en la distancia observada entre las puntuaciones y la media aritmética.

Por lo tanto:

- - Una distribución con poca variabilidad es aquella en la que la mayoría de las puntuaciones están muy próximas a la media.
- - Una distribución con mucha variabilidad tiene sus puntuaciones muy alejadas del valor medio de la variable.

desviación media:

- **DM =** Σ I$\mathbf{X_i}$ **- X**I / n
- El libro dice que este índice se utiliza muy poco en la actualidad y que apenas existen técnicas estadísticas basadas en este índice. Sin embargo conviene que lo estudiemos en profundidad no vaya a ser que en los próximos años se convierta en una herramienta completamente fundamental y necesaria para desarrollar la profesión de psicólogo.

La **varianza**

La **varianza** de un conjunto n de puntuaciones en una variable X denotada por S^2_x, se define como el promedio de los cuadrados de las desviaciones de las puntuaciones con respecto a la media.

$S^2_x = \Sigma (X_i - \bar{X})^2 / n$

Esta otra fórmula sirve para lo mismo:

- $S^2_x = (\Sigma X^2_i / n) - \bar{X}^2$
- Cuando los datos se presentan en tablas de distribución de frecuencias, agrupados o sin agrupar en intervalos, la varianza se puede calcular con las siguientes fórmulas.

- $S^2_x = \Sigma n_i (X_i - \bar{X})^2 / n$ (para frecuencias absolutas)

- $S^2_x = (\Sigma n_i X^2_i / n) - \bar{X}^2$ (para frecuencias absolutas)

n = número total de observaciones (como siempre)

X_i = es el valor i de la variable X o el punto medio del intervalo

n_i = es la frecuencia absoluta del valor o intervalo i

Otra fórmula más:

- $S^2_x = \Sigma p_i X^2_i - \bar{X}^2$ (para frecuencias relativas)

p_i = Frecuencia relativa o proporción de observaciones del valor o del intervalo i

- La varianza es un número positivo que se expresa en las unidades de la variable al cuadrado. Si la variable X se mide en metros, la varianza vendrá expresada en metros al cuadrado. Por lo tanto para encontrar una medida de dispersión que tenga las mismas unidades que la variable, debemos hacer la raíz cuadrada de la varianza (para quitarle el cuadrado) y obtener un índice llamado **desviación típica.**
- La **desviación típica** de un conjunto de n puntuaciones, que se representa pos S_X, es la raíz cuadrada de la varianza.

$$Sx = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}}$$

Propiedades de la varianza y la desviación típica

- - Para calcularlas las dos requieren la utilización de todas las puntuaciones de la distribución.
- - Las dos miden la variabilidad de los datos con respecto a la media aritmética, por lo que sólo deben aplicarse si estamos utilizando la media como medida central
- - Siempre son iguales o mayores que 0.
- - Si a una variable **X** se le suma o resta una constante **a**, la varianza y desviación típica de la variable original no se ven afectadas y siguen siendo las mismas. En cambio, cuando multiplicamos los valores de **X** por una constante **b**, la varianza queda multiplicada por la constante al cuadrado y la desviación típica por el valor absoluto de dicha constante.
- Un primo hermano de la varianza y que se utiliza en inferencia estadística es la **cuasivarianza:**

$S^2_{n-1} = \Sigma (\mathbf{X_i} - \mathbf{X})^2 / n - 1$

La cuasidesviación típica es la raíz cuadrada de la cuasivarianza

Coeficiente de variación

Se trata de un índice de variabilidad relativa que no depende de las unidades de medida.

$$CV = (S_X/\bar{X})\ 100$$

- Cuando comparamos dos conjuntos de puntuaciones obtenidas de la misma variable, también es necesario el coeficiente de variación para comparar la dispersión de ambas distribuciones

Amplitud intercuartil (rango intercuartil)

Como hemos dicho antes, este índice se utiliza cuando la distribución es muy asimétrica. se define como la distancia entre el tercer y el primer cuartil. No informa de la variabilidad del conjunto de puntuaciones sino del 50% de las mismas comprendidas entre el percentil 25 y el 75 de la distribución.

- $\mathbf{A_{IQ} = Q_3 - Q_1 = P_{75} - P_{25}}$

INDICE DE ASIMETRÍA DE PEARSON

La **asimetría** es una propiedad de la distribución de frecuencias que nos indica el grado en el que las puntuaciones de los sujetos se reparten por debajo y por encima de la medida de tendencia central. **El índice de Pearson** es un índice numérico que cuantifica el grado de asimetría de una distribución.

$$\mathbf{A_p = \bar{X} - Mo / S_x}$$

- Este índice es **adimensional** (no tiene unidades de medida) y se aplica a distribuciones unimodales.
- Cuando la distribución es **simétrica**, la media y la moda coinciden, por lo que el numerador se anula y el valor del índice (Ap) es = 0.
- En distribuciones con **asimetría positiva**, la media es mayor que la moda, por lo tanto Ap será mayor que 0. (recordamos que asimetría positiva se produce cuando la mayor cantidad de puntuaciones se concentran en la parte baja de la tabla)
- En distribuciones con **asimetría negativa**, la media es menor que la moda, por lo tanto Ap será menor que 0. (la asimetría negativa se produce cuando la mayor cantidad de puntuaciones se sitúa en la parte alta de la tabla)

Medidas de Forma: Asimetría

Coeficiente de Asimetría de Fisher:

- *No es de fácil cálculo, pero si su interpretación.*

$$ASF = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^3}{ns^3}$$ *Datos NO agrupados*

$$ASF = \frac{\sum_{i=1}^{k}\left(M_i - \overline{x}\right)^3 \times f_i}{ns^3}$$ *Datos Agrupados*

Si **As = 0** la distribución es **simétrica**.

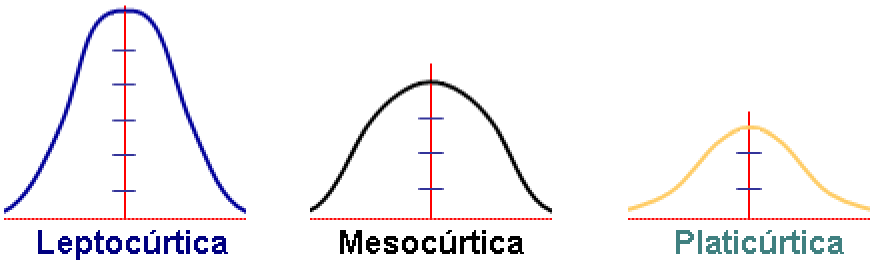
Si **As > 0** La distribución es asimétricamente positiva.

Si **As < 0** La distribución es asimétricamente negativa

Índice de curtosis

Se refiere al grado de apuntamiento de los datos en la distribución de frecuencias

Leptocúrtica. Platicúrtica y mesocúrtica



$$C_r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^4 / n}{S_x^4} - 3$$

$$K = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^4 \cdot n_i}{N}}{\sigma_x^4} - 3$$

Si la distribución es normal (mesocúrtica), el índice vale 0

Si la distribución es leptocúrtica, el índice es superior a 0

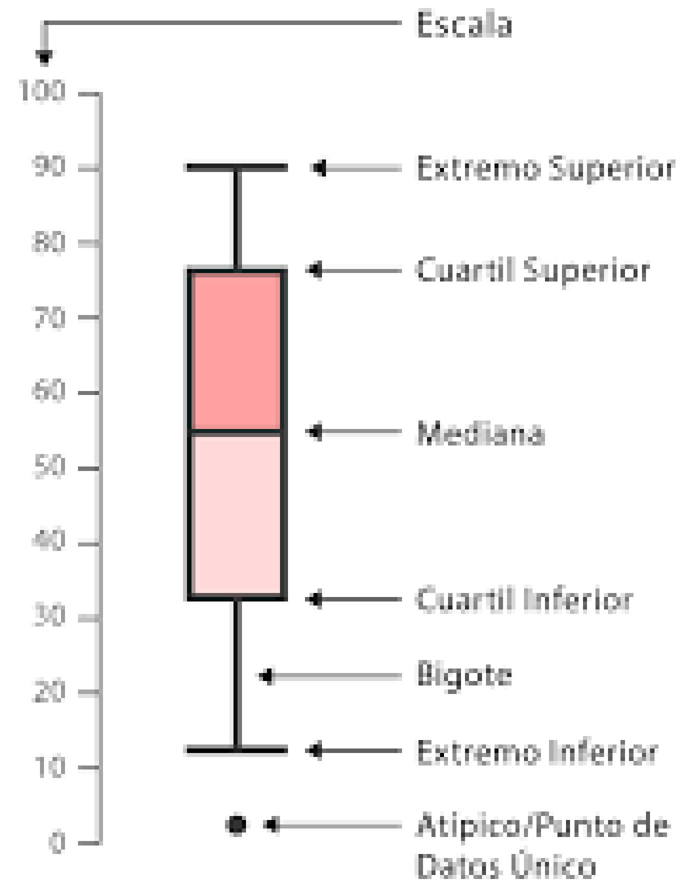
Si la distribución es platicúrtica, el índice es inferior a 0

Diagrama de caja

- Propuesta por Tukey, útil para estudiar visualmente la asimetría de una variable cuantitativa y detectar valores atípicos en la distribución de frecuencias sin agrupar en intervalos.
- Valores atípicos a partir de:

$$L_s = Q_3 + A_{IQ}.1,5$$

$$L_i = Q_1 - A_{IQ}.1,5$$



PUNTUACIONES TÍPICAS

Hasta ahora hemos visto **puntuaciones directas** (nota de un sujeto en un test), sin embargo estas puntuaciones nos dan poca información ya que no sabemos si se trata de un valor alto o bajo, ya que esto depende del promedio del grupo.

- Si a una puntuación directa Xi le restamos la media de su grupo obtenemos una **puntuación diferencial (xi)**

$$\mathbf{xi = Xi - \bar{X}}$$

Propiedades de las puntuaciones diferenciales

- - Su media es cero:
- $\bar{x} = 0$
- - La varianza de las puntuaciones diferenciales es igual a la varianza de las
- puntuaciones directas:
- $S^2_x = S^2_X$
- Sin embargo, dos puntuaciones diferenciales idénticas pueden tener un significado muy diferente en función de la media y de la varianza de las distribuciones de las que proceden.
- Para solucionar este problema tenemos las **puntuaciones típicas** que nos permiten no sólo comparar las puntuaciones de un sujeto en dos variables distintas sino también comparar dos sujetos distintos en dos pruebas o variables distintas.

$$z_x = x / S_x = X - \bar{X} / S_x$$

- Al proceso de obtener puntuaciones típicas se llama **tipificación**. Por ello estas puntuaciones también se llaman **puntuaciones tipificadas.**

Propiedades de las puntuaciones típicas

- - Su media es cero:

$$\bar{z}_x = 0/ns_x = 0$$

- - Su varianza es igual a 1

$$S^2_{zx} = 1$$