TEMA 3

psicologia.isipedia.com

Medidas de variabilidad y asimetría

1. MEDIDAS DE VARIABILIDAD

La variabilidad o dispersión hace referencia al grado de variación que hay en un conjunto de puntuaciones. Por ejemplo: "entre dos distribuciones que presentan la misma media aritmética, difieren en la variabilidad de sus puntuaciones". Así, cuanto menor es la variabilidad, más homogénea es la muestra de sujetos en la variable. En el caso de máxima homogeneidad, todos los valores de la variable serán iguales. De otro modo, cuanto más o menos dispersión en los datos, la muestra es más o menos heterogénea y las puntuaciones difieren entre sí.

Para cuantificar la dispersión de los datos, se pueden distinguir dos tipos de índices: los que miden el grado de semejanza y diferencia de las puntuaciones entre sí (amplitud total o rango y la amplitud semi-intercuartil), y los que la dispersión se mide a alguna medida de tendencia central como la media aritmética (varianza y la desviación típica).

1.1. Amplitud total o rango:

La amplitud total o rango (A_T), de un conjunto de puntuaciones es la distancia que hay en la escala numérica entre los valores que representan la puntuación máxima y la puntuación mínima.

$$A_T = X_{máx} - X_{mín}$$

Uno de los inconvenientes de la amplitud total es su limitación al utilizar únicamente los valores extremos de la distribución; de esta forma, no recoge la poca o mucha dispersión que pueda existir entre los restantes valores, que son la mayoría de las puntuaciones. Aún así se recomienda incluir éste valor como complementario de otras medidas de dispersión.

1.2. Varianza y desviación típica:

La variabilidad se puede basar en la distancia observada entre las puntuaciones y un valor central de la distribución como la media aritmética. De modo que, una distribución con poca variabilidad es en la que la mayoría de las puntuaciones están próximas a la media, mientras que con mucha variabilidad, las puntuaciones se alejan del valor medio de la variable.

Un primer índice podría ser el promedio de las desviaciones o diferencias de cada puntuación con su media.

$$\bar{X}_d = \frac{\sum d_i}{n} = \frac{\sum (X_i - \bar{X})}{n}$$

El problema de este índice es que el sumatorio del numerador $\sum (X_i - \bar{X})$, siempre es igual a cero; para ello se han propuesto dos soluciones. La primera consiste en calcular el valor absoluto de cada desviación antes de realizar la suma ***"desviación suma":***

$$DM = \frac{|X_1-\bar{X}|+|X_2-\bar{X}|+...+|X_n-\bar{X}|}{n} = \frac{\sum (X_i - X)}{n}$$

La desviación media se utiliza muy poco, ya que resulta poco manejable por el uso del valor absoluto.

Una segunda opción se basa en el cuadrado de las diferencias y así obtenemos la varianza.

$$S^2_x = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + ... + (X_3 - \bar{X})^2}{n} = \frac{\sum(X_i - \bar{X})^2}{n}$$

Otra forma derivada de la anterior y que simplifica los cálculos:

$$S^2_x = \frac{\sum X_i^2}{n} - \bar{X}^2$$

Para el cálculo de la varianza, primero se calcula el cuadrado de las diferencias y luego se obtiene el promedio de las desviaciones al cuadrado.

Cuando los datos se presentan en tablas de distribuciones de frecuencias agrupados o sin agrupar en intervalos, la varianza se puede calcular por medio de:

Cálculo de la varianza en tablas de distribución de frecuencias con datos agrupados o no en intervalos

Varianza a partir de una distribución de frecuencias absolutas:

$$S^2_x = \frac{\sum n_i(X_i - \bar{X})^2}{\sum n_i} = \frac{\sum n_i(X_i - \bar{X})^2}{n}$$

ó

$$S^2_x = \frac{\sum n_i X_i^2}{\sum n_i} - \bar{X}^2 = \frac{\sum n_i X_i^2}{n} - \bar{X}^2$$

donde:

n es el número total de observaciones.

X_i es el valor i en la variable X ó el punto medio del intervalo.

n_i es la frecuencia absoluta del valor o intervalo i.

Varianza a partir de una distribución de frecuencias relativas:

$$S^2_x = \sum p_i X_i^2 - \bar{X}^2$$

donde:

p_i es la frecuencia relativa o proporción de observaciones del valor o del intervalo i.

Hay que tener en cuenta que al elevar la variable al cuadrado, las unidades se elevan al cuadrado también. Para lograr una medida de dispersión en las mismas unidades que la variable y que sea más fácil interpretar, se calcula la raíz cuadrada de la varianza y se obtiene un índice que se llama desviación típica.

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

La varianza y la desviación son índices muy útiles en la estadística inferencial siendo la base de numerosas técnicas estadísticas. Ambos índices presentan una serie de propiedades ha destacar:

1. El cálculo de la varianza y la desviación típica, requieren el uso de todas las puntuaciones observadas en la distribución.

2. La varianza ya la desviación típica miden la variabilidad de los datos con la media aritmética.

3. Siempre son iguales o mayores que cero. Únicamente es a cero si todas las puntuaciones son iguales entre sí.

4. Si a las puntuaciones de la variable X les aplicamos la siguiente transformación lineal: $Y_i = bX_i + a$, las nuevas puntuaciones serán $S^2_y = b^2S^2_x$ y la desviación típica igual a $S_y = |b|S_x$. Si a una variable X se le suma o resta una constante a, la varianza y desviación típica no se verán afectadas. Sin embargo, cuando multiplicamos los valores de X por b, la varianza queda multiplicada por el cuadrado de la constante y la desviación típica por el valor absoluto de dicha constante.

Otro índice de variabilidad relacionado con la varianza es la cuasivarianza:

$$S^2_{n-1} = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

La cuasidesviación típica se define como la raíz cuadrada de la cuasivarianza:

$$S_{n-1} = \sqrt{S^2_{n-1}} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

1.3. Coeficiente de variación:

Es frecuente la comparación del grado de variabilidad o dispersión entre dos conjuntos de puntuaciones en una misma o distintas variables. Es necesario definir un índice de variabilidad relativa que no dependa de las unidades de medida. Para ello está el coeficiente de variación.

$$CV = \frac{S_x}{\bar{X}} \cdot 100$$

donde:

$\bar{X} > 0$ y se acompaña siempre de la medida y desviación típica de la distribución de donde se ha calculado.

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

1.4. Amplitud semi-intercuartil:

La amplitud semi-intercuartil “Q”, o rango semi-intercuartil es la distancia media entre el tercer y el primer cuartil. Se emplea como otra alternativa cuando debido a la asimetría de la distribución, no es aconsejable el uso de estos índices.

$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}$$

En Psicología la amplitud intercuartil $(P_{75} - P_{25})$, se ha usado en procedimientos de selección de ítems para la valoración de jueces o expertos en la materia.

2. INDICE DE ASIMETRÍA DE PEARSON

La asimetría de una distribución 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. De entre numerosos indicadores se ha elegido el índice de asimetría de Pearson que se basa en relación entre la media y la moda.

$$A_s = \frac{\bar{X} - Mo}{S_x}$$

Este índice no tiene unidades de medida “adimensional”. Cuando la distribución es simétrica, la media y la moda coinciden; por lo que $A_s = 0$. Cuando la asimetría positiva $A_s > 0$; y cuando es negativa $A_s < 0$.

3. PUNTUACIONES TÍPICAS

Todas las puntuaciones vistas hasta el momento han sido directas (de un test, …). Éstas nos ofrecen muy poca información, no sabemos si se trata de un valor alto o bajo por que esto depende del promedio del grupo.

Si a una puntuación directa X_i le restamos la media de su grupo, obtendremos una puntuación diferencial o de diferencia “x_i”.

$$x_i = X_i - \bar{X}$$

Las puntuaciones diferenciales nos indican si la puntuación coincide con la media de su grupo, si es mayor o menor. Presentan las siguientes propiedades:

a) Su media es cero: $\bar{x} = 0$.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum (X_i - \bar{X})}{n} = \frac{\sum X_i - \sum \bar{X}}{n} = \frac{\sum X_i}{n} - \frac{\sum X}{n} = \bar{X} - \bar{X} = 0$$

b) La varianza de las puntuaciones es igual a la varianza de las puntuaciones directas:

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} = \frac{\sum (X_i - \bar{X})^2}{n} = Sx^2$$

Las puntuaciones típicas nos permiten no sólo las puntuaciones de un sujeto en dos variables distintas sino dos sujetos distintos en dos pruebas o variables distintas.

$$z_x = \frac{x}{S_x} = \frac{X - X}{S_x}$$

A este proceso de obtener puntuaciones típicas se denomina tipificación, y a esas puntuaciones tipificadas.

Las puntuaciones típicas se caracterizan:

a) Su media es cero:

$$\bar{z}_x = \frac{\sum z_x}{n} = \frac{\sum \left(\frac{x_i}{s_x} \right)}{n} = \frac{\frac{1}{s_x} \sum x_i}{n} = \frac{\sum x_i}{n s_x} = \frac{0}{n s_x} = 0$$

b) Su varianza es igual a 1:

$$S^2_{zx} = \frac{\sum (z_x - \bar{z}_x)^2}{n} = \frac{\sum (z_x)^2}{n} = \frac{\sum \left(\frac{x}{s_x} \right)^2}{n} = \frac{\frac{1}{S^2_x} \sum x^2}{n} = \frac{1}{S^2_x} \frac{\sum x^2}{n} = \frac{1}{S^2_x} S^2_x = 1$$