TEMA-2

psicologia.isipedia.com

MEDIDAS DE TENDENCIA CENTRAL Y POSICIÓN

1. MEDIDAS DE TENDENCIA CENTRAL:

De la tendencia central de la distribución, nos interesa calcular un valor central que actúe como resumen numérico para representar al conjunto de datos. Estos valores son las medidas o índices de tendencia central. Los índices de tendencia central permiten representar la distribución con un único valor y facilitan la comparación de otros conjuntos de puntuaciones de una variable.
Las medidas más utilizadas en el análisis de datos son:

1.1. ***La media aritmética***; también promedio o media es la medida de tendencia central más conocida y usada en la práctica, por su sencillez de cálculo y es el fundamento de un gran número de técnicas estadísticas. Indica la tendencia general de una distribución de frecuencias de una variable y es el valor central de las observaciones "centro de gravedad". Sin embargo se limita para calcular variables cuantitativas. La media aritmética de una variable X, denotada por $\bar{X}$, se define:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum X_i}{n}$$

donde:
X_i es el valor que toma la variable u observación del sujeto i.
n es el número total de observaciones.
Cuando el número de observaciones es elevado, los datos se presentan en tablas de distribución de frecuencias agrupados o no en intervalos; en este caso se puede calcular a partir de:

1. Media aritmética a partir de una distribución de frecuencias absolutas:

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{\sum n_i X_i}{n}$$

donde:
n es el número total de observaciones.
X_i es e valor i en la variable X_i o el punto medio del intervalo.
n_i es la frecuencia absoluta del valor o intervalo i.

2. Media aritmética a partir de una distribución de frecuencias relativas:

$$\bar{X} = \sum p_i X_i$$

donde:

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

La media aritmética aprovecha toda la información disponible en los datos, pues requiere de todas las puntuaciones de los sujetos. Así, la media aritmética presenta las siguientes propiedades:

1. En una distribución, la suma de las desviaciones de cada valor con respecto a su media es igual a cero.

$$\sum_{i=1}^{n}(X_i - \bar{X}) = 0$$

2. Si a los valores de la variable X les aplicamos la siguiente transformación lineal $Y_i = bX_i + a$, la media de los nuevo valores Y será $\bar{Y} = b\bar{X} + a$.

A la hora de utilizar la media aritmética conviene tener en cuenta las siguientes limitaciones:

a) Cuando los datos están agrupados en intervalos, la media no se puede calcular si el intervalo máximo no tiene límite superior y/o el mínimo no lo tiene inferior.
b) En distribuciones asimétricas no es recomendable su uso debido a que afecta a su representatividad como valor central de la distribución. Estos valores extremos pueden ser producto de errores en la recogida o grabación de datos, o valores que aportan información relevante de la variable. En el primer caso se eliminan estas observaciones, volviendo la distribución más simétrica. Y en el caso dos, se recomienda aplicar otros índices de tendencia central menos sensibles a los valores extremos como la mediana.

1.2. ***La mediana:*** es el índice empleado cuando la distribución es asimétrica y no es posible aplicar la media aritmética. La mediana no se ve afectada por los valores extremos que en su cálculo ocupan las posiciones centrales. Por lo tanto, la mediana es un valor apropiado para representar la tendencia central de la distribución y se puede obtener todo tipo de variables excepto en variables cualitativas. La mediana de una variable X, representada por Md, se define como el valor de la variable que divide la distribución de frecuencias en dos partes iguales, conteniendo cada una el 50% de las observaciones.

Para el cálculo de la mediana con pocos casos se procede de la siguiente manera:

1. Se ordenan las n puntuaciones de menor a mayor.
2. Se observa si el número de observaciones n es impar o par. Si es impar, el valor de la mediana es el de la observación que ocupa la posición central, dentro de ese conjunto de observaciones ya ordenadas. Si es par, la mediana es la media aritmética de los dos valores centrales de la distribución.

Lo normal es que el número n de observaciones no sea pequeño; incluso aparecerán repetidos y, por ello, los datos se presentarán en tablas de distribución de frecuencias agrupados o no en intervalos. Cuando los datos están en intervalos se denomina intervalo crítico y se corresponde con el que la frecuencia absoluta acumulada n_a es igual o superior a n/2. La mediana se obtiene así:

$$Md = L_i + \left(\frac{\frac{n}{2} - n_d}{n_c}\right) . I$$

donde:

L_i = Límite exacto inferior del intervalo.
n = Número de observaciones.
n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico.
n_c = Frecuencia absoluta del intervalo crítico.
I = Amplitud del intervalo crítico.

Cuando los datos no están agrupados en intervalos, el cálculo es un caso particular de la fórmula anterior en la que la amplitud de los intervalos es igual a uno (I = 1).
La mediana se puede calcular en cualquier distribución excepto cuando los datos están agrupados y existe uno abierto en el que se encuentra la mediana.

1.3. La moda: la moda es otro índice de tendencia central que se puede obtener tanto en variables cualitativas como en cuantitativas. Se representa por Mo, y se define como el valor o categoría de la variable con mayor frecuencia absoluta.

Si se da en una variable cuantitativa con los datos no agrupados en intervalos, la moda es el valor con la mayor frecuencia absoluta.
Si se da en una distribución de una variable cuantitativa en intervalos, se localiza el intervalo modal que es el intervalo con la frecuencia máxima y la moda es el medio de dicho intervalo.
Una frecuencia es unimodal cuando existe un único valor con la frecuencia máxima. Si presenta varios valores con la frecuencia más alta, ésta será bimodal, trimodal, …
Sus principales características son:

a) es un índice de cálculo sencillo y de fácil interpretación.
b) Es el único que, además de aplicarse a variables cuantitativas, se puede calcular en variables cualitativas.
c) Sólo se excluye su cálculo en el supuesto de que coincida con el intervalo abierto.

1.4. La elección de una medida de tendencia central: para seleccionar un valor que resuma adecuadamente la tendencia central de la distribución de frecuencias,se recomienda como primera opción la media aritmética, por que en ella basan su importancia muchos estadísticos. Únicamente se desaconseja su uso con pocos valores extremos, cuando el nivel de medida de la variable es nominal u ordinal y/o en datos agrupados en los que existen intervalos abiertos en los extremos de la distribución.

Cuando la media no es aplicable, se recomienda la mediana. Y por último, si ésta no es posible, se aconseja tomar la moda. Resaltar que si una variable cuantitativa es simétrica y unimodal, coinciden todos los valores de la media, mediana y moda.

2. MEDIDAS DE POSICIÓN:

Las medidas o índices de posición, también llamados cuantiles, informan acerca de la posición relativa de un sujeto con respecto a su grupo de referencia, dentro de la distribución de frecuencias de la variable. Para ello debemos dividir la distribución en un número de partes o secciones iguales entre sí en cuanto al número de observaciones. Trataremos tres medidas de posición o cuantiles:

2.1. Percentiles: los también llamados centiles, son los 99 valores de la variable que dividen en 100 la distribución de frecuencias. Representado por P_k , es un valor de la variable de interés que deja por debajo de sí un porcentaje k de sujetos, donde k = 1,2....99-
El cálculo de los percentiles se realiza utilizando una extensión del método expuesto para la mediana. La diferencia está en que para la mediana se traba de localizar la posición de n/2; mientras los percentiles se hace en base al número n.k/100 . Este número es igual a n/2 cundo calculamos el percentil 50. entonces si k = 50; n.50/100 = n/2.
El percentil k se obtiene de la siguiente manera:

$$P_k = L_i + \left(\frac{\frac{n.k}{100} - n_d}{n_c} \right) . I$$

donde:

n_d = Frecuencia absoluta acumulada por debajo del intervalo crítico.

n_c = Frecuencia absoluta del intervalo crítico.
L_i = Límite inferior exacto del intervalo crítico.
I = Amplitud del intervalo.

Al igual que con la mediana, cuando en la distribución de frecuencias los datos no están en intervalos, se aplica I = 1.
Sin embargo lo que queremos calcular es qué percentil ocupa un valor X_i debemos de despejar k para X_i:

$$K = \left(\frac{\frac{(P_k - L_i).n_c}{I} + n_d}{n} \right) . 100$$

A veces el resultado puede darnos un valor con decimales; en este caso, tomamos la cantidad entera más próxima.

2.2. Cuartiles y deciles: son medidas de posición en las que las secciones son muchas menos que en los percentiles. Los cuartiles son tres valores de la distribución que se dividen en cuatro partes. El primer cuartil se representa por Q_1, deja por debajo de sí el 25%, correspondiendo con el percentil 25. El segundo cuartil Q_2, deja por debajo el 50%, equivalente al percentil 50 y a la mediana de la distribución. El tercero Q_3, deja por debajo de sí el 75%, equivalente al percentil 75.
Para calcular los cuartiles emplearemos las mismas fórmulas que para los percentiles en su equivalencia.
Los deciles son nueve valores que dividen en diez partes iguales a la distribución. Se representa por D_i, donde i = 1,2,...9.
El primer decil, deja por debajo de sí al 10% de los sujetos, el segundo el 20% y así hasta el 90%; pudiendo emplear nuevamente los percentiles correspondientes para su cálculo.