Joseline Jamileh Montiel Vázquez

Rosario Gómez Lujano

Medidas de tendencia central y de dispersión

Estadística

1er Cuatrimestre

Escolarizado

Pichucalco, Chiapas a 16 de octubre de 2020.

MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSION:

Las medias de tendencia central o posición nos indican donde se sitúa un dato dentro de una distribución de datos. Las medidas de dispersión, variabilidad o variación nos indican si esos datos están próximos entre sí o sí están dispersos, es decir, nos indican cuán esparcidos se encuentran los datos. Estas medidas de dispersión nos permiten apreciar la distancia que existe entre los datos a un cierto valor central e identificar la concentración de los mismos en un cierto sector de la distribución, es decir, permiten estimar cuán dispersas están dos o más distribuciones de datos. Estas medidas permiten evaluar la confiabilidad del valor del dato central de un conjunto de datos, siendo la media aritmética el dato central más utilizado. Cuando existe una dispersión pequeña se dice que los datos están dispersos o acumulados cercanamente respecto a un valor central, en este caso el dato central es un valor muy representativo. En el caso que la dispersión sea grande el valor central no es muy confiable. Cuando una distribución de datos tiene poca dispersión toma el nombre de distribución homogénea y si su dispersión es alta se llama heterogénea.

Las medidas de tendencia central se utilizan con bastante frecuencia para resumir un conjunto de cantidades o datos numéricos a fin de describir los datos cuantitativos que los forman.

- MEDIA ARITMÉTICA:

La media aritmética es lo que se conoce como media al uso. Sumamos todos los valores y lo dividimos entre la cantidad de observaciones.

Símbolo de la media aritmética

El símbolo de la media aritmética es una X con una barra encima. Por lo que quedaría así ↓

Símbolo de la media aritmética → x̄

Fórmula de la media aritmética

La fórmula de la media aritmética es la siguiente:

$$Media\ aritmética = \frac{\sum_1^N x_i}{N} = \frac{x_1 + \ x_2 + \ x_3 \ldots . + x_n}{N}$$

Si la leemos de izquierda a derecha, veremos que hay tres partes. La primera es el nombre, la segunda una fórmula pequeña y la tercera el desarrollo. La segunda parte de la fórmula se lee tal que así: Sumatorio desde 1 hasta N de la variable x dividido entre N. Adicionalmente, podríamos añadir un comentario que indicase: con i entre 1 y N. Vamos a ver ahora, qué significa todo esto.

- **Sumatorio:** El sumatorio nos indica que debemos sumar un conjunto de valores desde el primero, hasta el N. Así si existen 30 valores, deberemos sumar el primero, el segundo, el tercero, … , y el treintavo.
- **N:** Representa el número total de observaciones. Por ejemplo, si tenemos el peso de 10 manzanas individuales, N vale 10. Ya que tenemos 10 manzanas.
- **x:** La variable X es sobre la que calculamos la media aritmética. En este caso sería el peso de las manzanas.
- **i:** Representa la posición de cada observación. En este ejemplo, podríamos ponerle una etiqueta a cada manzana, la manzana 1, la manzana 2, etc.

- MEDIANA:

La mediana es un estadístico de posición central que parte la distribución en dos, es decir, deja la misma cantidad de valores a un lado que a otro. La mediana, junto con la media y la varianza es un estadístico muy ilustrativo de una distribución. Al contrario que la media que puede estar desplazada hacia un lado o a otro, según la distribución, la mediana siempre se sitúa en el centro de esta.

Fórmula de la mediana

Una vez definida la mediana vamos a pasar a calcularla. Para ello, necesitaremos una fórmula.

La fórmula no nos dará el valor de la mediana, lo que nos dará es la posición en la que está dentro del conjunto de datos. Debemos tener en cuenta, en este sentido, si el número total de datos u observaciones que tenemos (n) es par o impar. De tal forma que la fórmula de la mediana es:

Cuando el número de observaciones es par:

Mediana = (n+1) / 2 → Media de las observaciones

Cuando el número de observaciones es impar:

Mediana = (n+1) / 2 → Valor de la observación

Es decir, que si tenemos 50 datos ordenados preferiblemente de menor a mayor, la mediana estaría en la observación número 25,5. Esto es el resultado de aplicar la fórmula para un conjunto de datos par (50 es número par) y dividir entre 2. El resultado es 25,5 ya que dividimos entre 50+1. La mediana será la media entre la observación 25 y la 26.

- MODA:

Se define como el número que está representado más veces dentro de esos datos, es decir, aquel número que presenta una mayor frecuencia absoluta dentro de la muestra. La moda puede ser calculada tanto para variables cuantitativas como para variables cualitativas. Podemos distinguir distintos tipos de moda, en función del número de números que se repitan una misma cantidad de veces, siendo ese número de repeticiones el máximo del conjunto.

1. Moda unimodal : cuando el máximo número de repeticiones se da para un solo número.

Ejemplo conjunto de datos: [3, 5, 5, 6, 8]

La moda del conjunto es 5 porque se repite en dos ocasiones, mientras que el resto de números se repiten únicamente una vez.

2. Moda bimodal: cuando el máximo número de repeticiones se da para dos números.

Ejemplo conjunto de datos: [3, 5, 5, 6, 8, 8]

La moda del conjunto sería 5 y 8 porque ambos números se repiten en dos ocasiones, mientras que el resto de números se repiten únicamente una vez.

3. Moda multimodal: cuando el máximo número de repeticiones se da para tres o más números.

Ejemplo conjunto de datos: [3, 3, 5, 5, 6, 8, 8]

La moda del conjunto en este caso serían tres números, porque los tres se repiten el mismo número de veces: 3, 5, 8.

- Cuartiles:

Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales.

Q1, Q2 y Q3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos.

Q2 coincide con la mediana.

- Deciles:

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D5 coincide con la mediana.

- Percentil:

Los percentiles son los 99 valores que dividen la serie de datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P50 coincide con la mediana.

P50 coincide con D5.

- RANGO:

Señala la amplitud de la variación de un fenómeno entre su límite menor y uno claramente mayor. El rango estadístico, por lo tanto, es el intervalo que contiene dichos datos y que puede calcularse a partir de restar el valor mínimo al valor máximo considerado.

- VARIANZA:

La varianza (S2) mide la dispersión de los datos de una muestra (X1,X2,…,XN) respecto a la media (x), calculando la media de los cuadrados de las distancias de todos los datos.

$$S_X^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{x})^2}{N-1}$$

siendo $(X_1, X_2, ..., X_N)$ un conjunto de datos y $\overline{x}$ la media

Al elevar las diferencias al cuadrado se garantiza que las diferencias absolutas respecto a la media no se anulan entre si. Además, resaltan los valores alejados.

Siempre se cumple que la varianza es mayor o igual que cero (S2 ≥ 0). Ésta es cero cuando todos los datos son el mismo (ejemplo: {1,1,1,1,1}).

Si en vez de tratarse de una muestra, la varianza se refiere a la población, el denominador será N.

- DESVIACION ESTANDAR:

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos.

El símbolo σ (sigma) se utiliza frecuentemente para representar la desviación estándar de una población, mientras que s se utiliza para representar la desviación estándar de una muestra. La variación que es aleatoria o natural de un proceso se conoce comúnmente como ruido.

La desviación estándar se puede utilizar para establecer un valor de referencia para estimar la variación general de un proceso.

- COEFICIENTE DE VARIACION:

El coeficiente de variación, también denominado como coeficiente de variación de Pearson, es una medida estadística que nos informa acerca de la dispersión relativa de un conjunto de datos.

Fórmula del coeficiente de variación

Su cálculo se obtiene de dividir la desviación típica entre el valor absoluto de la media del conjunto y por lo general se expresa en porcentaje para su mejor comprensión.

$$CV = \frac{\sigma_x}{|\bar{X}|}$$

X: variable sobre la que se pretenden calcular la varianza

σx: Desviación típica de la variable X.

$| \bar{x} |$: Es la media de la variable X en valor absoluto con $\bar{x} \neq 0$

El coeficiente de variación se puede ver expresado con las letras CV o r, dependiendo del manual o la fuente utilizada. Su fórmula es la siguiente:

El coeficiente de variación se utiliza para comparar conjuntos de datos pertenecientes a poblaciones distintas. Si atendemos a su fórmula, vemos que este tiene en cuenta el valor de la media. Por lo tanto, el coeficiente de variación nos permite tener una medida de dispersión que elimine las posibles distorsiones de las medias de dos o más poblaciones.