

Medidas de tendencia

UdeCataluña

Medidas de tendencia

1. Medidas de tendencia central

Una medida de tendencia central es un único valor que indica el centro de una serie de números a partir de los cuales se calcula. Implica entender el concepto de promedio: es un valor típico o representativo de un conjunto de datos; como tales valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud, los promedios se conocen también como medidas de centralización, las más comunes son la media aritmética o media, la mediana, la moda, la media geométrica y la media armónica.

La media (\bar{X}) también conocida como promedio o media aritmética, de un conjunto de datos o N mediciones (X_1, X_2, \dots, X_N) es el promedio de las mediciones. Se calcula de la siguiente forma.

$$(X) = \bar{x} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

Su principal ventaja es su fácil uso y su mayor desventaja es que por ser un punto de equilibrio, es muy sensible a valores extremos.

La mediana (Md)

La segunda medida de tendencia central de un conjunto de números es la mediana. Su característica principal es que divide un conjunto de valores ordenados en dos grupos iguales; la mitad de los números tendrá valores que son menores que la mediana y la otra mitad alcanzará valores mayores y se representa por las letras *Md*. Para conjuntos de datos asimétricos la mediana es la mejor medida de tendencia central.

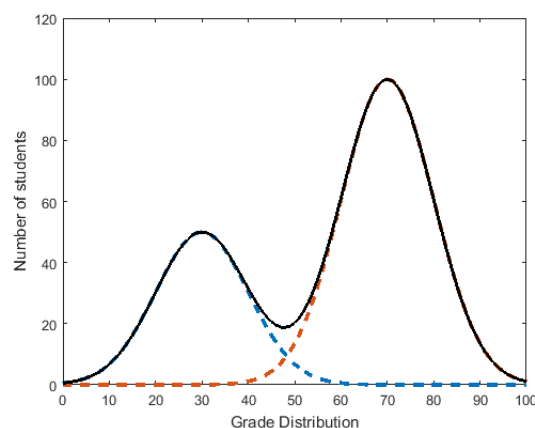
Para encontrar la mediana primero es necesario ordenar los valores (generalmente), de menor a mayor. Posteriormente se deberá separar la mitad de los valores para obtener la mediana. En resumen, podemos decir que el procedimiento para obtener la mediana es el siguiente:

- Ordenar o clasificar los valores de mayor a menor o de menor a mayor.
- Contar para saber si existe un número de valores par o impar.
- En caso de que se tenga un número impar de valores, la mediana es el valor central. En cambio, para un número par de valores, la mediana es el promedio de los valores centrales

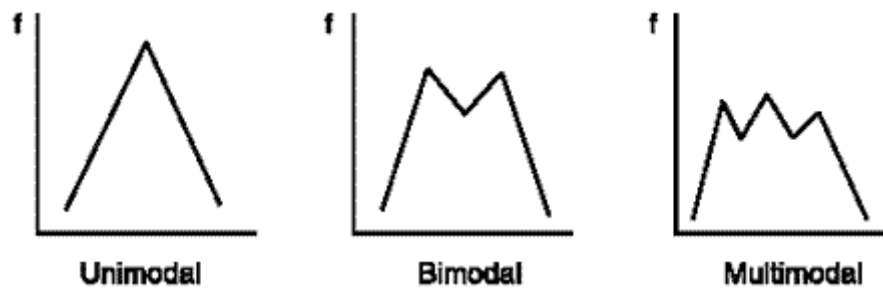
La moda (M_o)

La tercera medida de tendencia central, fácilmente obtenible, es la moda, ya que es el valor que se repite con mayor frecuencia en un conjunto de puntuaciones y se representa por M_o . Si el valor es único decimos que la distribución es unimodal. Si se tienen dos o más decimos que es bimodal, trimodal, etc.

Ejemplos:



Distribución bimodal. Tiene dos máximos, tiene dos modas.

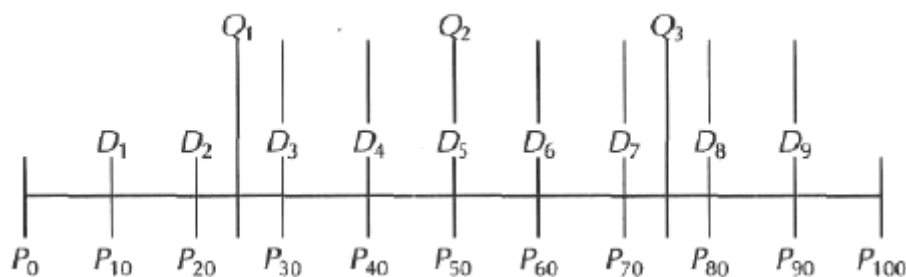


Diferentes distribuciones, f es la frecuencia absoluta para una determinada variable cuantitativa puesta en el eje x .

1. Medidas de posición

Un cuantil es un punto en una escala numérica que se supone abarca una serie de observaciones dividiéndola en dos grupos, cuyas respectivas proporciones se conocen. A medida que se divide la serie en más partes aparecen los cuartiles, quintiles, deciles y percentiles. Una serie donde sus valores se encuentren ordenados de forma creciente o decreciente, aquellos tres términos que dividen a la serie en cuatro partes o grupos de números iguales, se llaman “cuartiles”. El primero de éstos recibe el nombre del primer cuartil, el siguiente segundo cuartil y el último tercer cuartil.

Análogamente, los términos que dividen a la serie en 10 partes iguales, se llaman “deciles” y se representan por D_1, D_2, \dots, D_{10} , y los términos que dividen a la serie ordenada en 100 partes iguales, se llaman “percentiles” y se representan por los símbolos $P\%$. Para emplear los percentiles se deben tener más de 100 elementos.



2. Medidas de dispersión

Las medidas de tendencia central se refieren a la concentración de puntuaciones de una determinada escala numérica en un grupo. Una medida particular de tendencia central da lugar a una puntuación o valor que en cierto sentido "representa" a todas las puntuaciones del grupo. Este proceso pasa por alto las diferencias entre las puntuaciones en sí. Asimismo, para evaluar la variación de las puntuaciones se necesitan estadígrafos descriptivos que miden la heterogeneidad, dispersión o esparcimiento de puntuaciones. Para describir en forma adecuada un conjunto de datos, son necesarios dos tipos de medidas de resumen: las de tendencia central y las de dispersión. Además, para obtener información respecto a la parte media de un conjunto de números, también es conveniente tener un método para expresar la cantidad de dispersión o difusión que hay entre los datos. Por ejemplo, las medidas de dispersión indican si los valores están relativamente cercanos uno de otro o si se encuentran dispersos.

Rango

La amplitud o rango de un grupo de números es generalmente la medida más sencilla de calcular y comprender. Se concentra en el número mayor y el menor del grupo, es decir, los puntos extremos. Dicha medida se puede expresar en dos formas:

- ✓ Como la diferencia entre los valores mayor y menor de la muestra.
- ✓ Como los valores mayor y menor del grupo.

Rango intercuantil

El rango intercuartil (RQ) es la diferencia entre el primer y tercer cuartil, o sea:

$$RQ = Q_3 - Q_1$$

Esta medida, para propósitos descriptivos, es definitivamente superior al rango, se dice que es superior porque se calcula entre el primer y tercer cuartil, eliminando dos extremos de un cuarto de la muestra, considerando que la mayoría de las puntuaciones se agruparán en la proporción restante, excepto en cuanto a simplicidad de cálculos.

Varianza (s^2)

La varianza de una muestra o conjunto de datos es la desviación promedio de valores obtenidos a partir de la media elevada al cuadrado y dividida entre $n-1$. Cuando hablamos de datos que reflejan únicamente una parte de la población, es decir, una muestra, utilizamos n ; cuando nos referimos a toda una población, es decir, los datos de la muestra son iguales al total de la población estudiada, pero como eso sucede en pocos casos casi siempre se utiliza $n-1$.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Desviación estándar (s)

Vale la pena recordar que se llama desviación o desvío a la diferencia entre un valor individual x_j y la media \bar{x} . La varianza es una medida de dispersión en la que hallamos las desviaciones al cuadrado. Esto indica que la unidad de varianza se expresa en unidades al cuadrado. Para superar esta insuficiencia y disponer de una

medida de la dispersión de las puntuaciones que se exprese en unidades, que no sean al cuadrado, se calcula la raíz cuadrada de la varianza conocida como desviación estándar.

$$s = \sqrt{s^2}$$

Referencias

1. <https://www.rdocumentation.org/>
2. Estadística Descriptiva, Salvador Monroy Saldívar. 2008. Instituto Politécnico Nacional. México. Capítulos 2.3 y 2.4. Disponible en la biblioteca del curso.

UdeCataluña

© U de Cataluña, 2020

Todos los derechos reservados. Prohibida la reproducción total o parcial sin permiso o autorización de la Universidad, Bogotá - Colombia.