

Estadística para las ciencias sociales

Semana 3

Medidas de tendencia central: media, mediana y moda en datos agrupados y no agrupados

En el análisis de datos, las medidas de tendencia central son herramientas estadísticas clave para resumir y describir conjuntos de datos. Exploreemos las medidas de tendencia central más comunes: la media, la mediana y la moda, en datos agrupados y no agrupados. Estas medidas nos ayudan a comprender la ubicación y la dispersión de los datos, lo que nos permite obtener información valiosa sobre la distribución de los mismos.



La media: un promedio aritmético

En datos no agrupados

La media (promedio) se calcula sumando todos los valores y dividiendo por el número total de valores. Es útil para datos cuantitativos y proporciona una medida central de la distribución.

- Por ejemplo, supongamos que tenemos los siguientes datos: 5, 7, 9, 11, 13. Para calcular la media de estos datos, sumamos todos los valores ($5 + 7 + 9 + 11 + 13 = 45$), luego dividimos el resultado por el número de valores ($45 / 5 = 9$). Por lo tanto, la media de estos datos es 9.

En datos agrupados

La media (promedio) se calcula multiplicando el punto medio de cada intervalo por la frecuencia de ese intervalo, sumando estos productos y dividiendo por el total de frecuencias. La media en datos agrupados es una aproximación, ya que se asume que todos los datos en un intervalo están concentrados en su punto medio.

- Por ejemplo, supongamos que tenemos las edades de un grupo de personas agrupadas en intervalos de 10 años:

Rango de edad (años)	Frecuencia
0-9	5
10-19	10
20-29	15
30-39	20
40-49	10
50-59	5

Para calcular la media de estos datos, encontramos el punto medio de cada intervalo y lo multiplicamos por la frecuencia correspondiente. Luego, sumamos estos productos y dividimos por el total de frecuencias. Así, la media de las edades es aproximadamente de 29.88 años.

La media es una medida útil para resumir los datos, ya que proporciona una estimación del valor típico o central del conjunto. Sin embargo, también es sensible a los valores extremos o atípicos en los datos, lo que puede afectar su interpretación en ciertos casos.

La mediana: el valor central

En datos no agrupados

Mientras que la media se calcula como un promedio aritmético, la mediana es el valor medio de un conjunto de datos ordenados en orden ascendente o descendente. La mediana divide el conjunto de datos en dos mitades iguales, lo que significa que el 50% de los datos están por encima de la mediana y el 50% están por debajo de ella. Si hay un número impar de datos, es el valor central; si hay un número par, es el promedio de los dos valores centrales.

- Por ejemplo, consideremos los siguientes datos: 2, 5, 7, 9, 11, 13. Para calcular la mediana primero ordenamos los datos en orden ascendente: 2, 5, 7, 9, 11, 13. Como el número de datos es impar, la mediana es el valor en el centro de la lista ordenada, que en este caso es 7.

En datos agrupados

Se determina encontrando el intervalo que contiene la mitad de los datos (basado en las frecuencias acumuladas) y luego usando una fórmula para estimar el valor exacto dentro de ese intervalo.

- Por ejemplo, retomemos las edades de un grupo de personas agrupadas en intervalos de 10 años:

Rango de edad (años)	Frecuencia
0-9	5
10-19	10
20-29	15
30-39	20
40-49	10
50-59	5

Para calcular la mediana, primero determinamos la frecuencia acumulada para encontrar el intervalo que contiene la mediana. La mediana estará en el intervalo donde la frecuencia acumulada alcanza o supera la mitad del total de frecuencias. La mediana se estima en aproximadamente 31.25 años. Esto indica que la mitad de las personas en el grupo tienen una edad de 31.25 años o menos.

Se determinó este valor encontrando el intervalo mediano (30-39 años) y aplicando una fórmula para estimar la mediana dentro de ese intervalo.

La mediana es una medida de tendencia central robusta, lo que significa que no se ve afectada por los valores extremos o atípicos en los datos. Esto la hace especialmente útil cuando los datos contienen valores atípicos que podrían sesgar la media.

La moda: el valor más común

En datos no agrupados

La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. En otras palabras, es el valor más común. Un conjunto de datos puede tener una moda única o varias modas, si hay varios valores que se repiten con la misma frecuencia máxima.

Para identificar la moda, simplemente contamos cuántas veces aparece cada valor en el conjunto de datos y encontramos el valor con la frecuencia máxima.

- Por ejemplo, consideremos los siguientes datos: 2, 5, 5, 7, 9, 9, 9, 11, 13. Para calcular la moda, el valor 9 aparece tres veces, más que cualquier otro valor en el conjunto. Por lo tanto, la moda de estos datos es 9.

En datos agrupados

Se identifica como el intervalo con la mayor frecuencia. La moda en datos agrupados es menos precisa, ya que sólo indica el intervalo más común, no un valor específico.

- Por ejemplo, retomemos las edades de un grupo de personas agrupadas en intervalos de 10 años:

Rango de edad (años)	Frecuencia
0-9	5
10-19	10
20-29	15
30-39	20
40-49	10
50-59	5

Para calcular la moda, observamos cuál intervalo tiene la mayor frecuencia. En este caso, el intervalo con la frecuencia es de 30 a 39 años. Esto significa que la mayoría de las personas en el grupo se encuentran dentro de este rango de edad, ya que este intervalo tiene la mayor frecuencia.

La moda es una medida de tendencia central útil para identificar los valores más comunes en un conjunto de datos. A diferencia de la media y la mediana, la moda puede usarse con datos cualitativos, como categorías o atributos.

Relación entre media, mediana y moda

Mientras que la media ofrece un promedio, la mediana proporciona un punto medio y la moda indica el valor o intervalo más común. Estas medidas ofrecen diferentes perspectivas sobre la "centralidad" de un conjunto de datos, y su aplicabilidad y precisión pueden variar según la naturaleza de los datos y su distribución. Su relación puede proporcionar información adicional sobre la forma y la distribución de los datos. Cuando los datos están distribuidos simétricamente, la media, mediana y moda tienden a ser aproximadamente iguales. Esto ocurre en distribuciones normales o en forma de campana.

Cuando los datos están sesgados hacia la derecha, la media tiende a ser mayor que la mediana, y la mediana tiende a ser mayor que la moda. Esto se

debe a la influencia de los valores extremos o atípicos en el extremo derecho de la distribución.

Cuando los datos están sesgados hacia la izquierda, la media tiende a ser menor que la mediana, y la mediana tiende a ser menor que la moda. Esto se debe a la influencia de los valores extremos o atípicos en el extremo izquierdo de la distribución.

Es importante tener en cuenta que la relación entre media, mediana y moda puede variar dependiendo de la distribución y la forma de los datos. Por lo tanto, es recomendable calcular y analizar estas medidas en conjunto para obtener una comprensión más completa de los datos.

Cuartiles y percentiles: medidas adicionales de tendencia central

Además de la media, mediana y moda, existen otras medidas de tendencia central que se utilizan para analizar conjuntos de datos más detalladamente. Estas medidas incluyen los cuartiles y los percentiles.

Los **cuartiles** dividen un conjunto de datos ordenados en cuatro partes iguales. El primer cuartil (Q1) es el valor que divide el conjunto en el 25% inferior, el segundo cuartil (Q2) es la mediana que divide el conjunto en el 50% inferior y el 50% superior, y el tercer cuartil (Q3) es el valor que divide el conjunto en el 75% inferior.

Los **percentiles** son medidas similares a los cuartiles, pero dividen un conjunto de datos en 100 partes iguales. Por ejemplo, el percentil 25 (P25) divide el conjunto en el 25% inferior, el percentil 50 (P50) es la mediana que divide el conjunto en el 50% inferior y el 50% superior, y el percentil 75 (P75) divide el conjunto en el 75% inferior.

Estas medidas adicionales de tendencia central nos permiten comprender la distribución de los datos de manera más detallada. Por ejemplo, los cuartiles y los percentiles son especialmente útiles para identificar valores atípicos o para comparar la posición relativa de un valor con relación al resto del conjunto.

Las medidas de tendencia central nos ayudan a obtener una comprensión más completa de la ubicación y la dispersión de los datos. Además, los cuartiles y los percentiles proporcionan medidas adicionales que nos permiten analizar los datos de manera más detallada. Al utilizar estas medidas de tendencia central, podemos obtener información valiosa sobre la distribución de los datos y tomar decisiones informadas basadas en el análisis estadístico.

Referencias

INITE (2012). **Medidas de tendencia central y dispersión.**
https://apps.utel.edu.mx/syllabus/cloud/visor.php?container=L1C115_925_505_28250_0&object=A1437_R2660.pdf

Rojas, J. (2008). **Formulario de estadística.**
https://apps.utel.edu.mx/syllabus/cloud/visor.php?container=L1C115_925_505_36927_0&object=Formulario%20De%20Estad%C3%ADstica.pdf