

Tema 4 Medidas de centralidad y dispersión

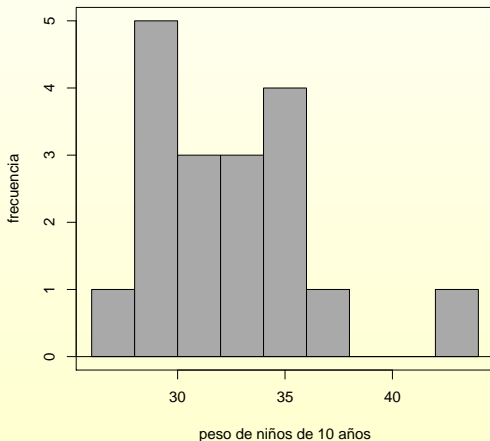
Alicia Nieto Reyes

BIOESTADÍSTICA

Medidas de Centralidad

Dado un conjunto de números podemos querer saber entorno a que valor se agrupan dichos números

Esto lo podíamos intuir con el diagrama de barras y el histograma



Medidas de Centralidad

El valor entorno al que se agrupa un conjunto de números se puede calcular con:

- Media
- Mediana
- Moda

El conjunto de números puede referirse tanto a una población como a una muestra

Vamos a aprender a calcularlas en el caso en que el conjunto sea finito:

Ejemplo 1: 3 4 4 5 2

- **Media:** Sumamos los datos y lo dividimos por el número de datos
$$\frac{3+2+4+5+2}{5} = 3,6$$
- **Mediana:** Ordenamos los datos 2 3 4 4 5 y tomamos el que se encuentre en el centro
- **Moda:** Es el valor que más veces aparece en los datos de que disponemos, 4

Medidas de Centralidad

Ejemplo 2:

29,43 35,18 30,81 33,66 29,58 29,48 35,49 34,53 32,69
26,15 31,12 29,07 34,21 37,39 33,34 31,07 42,03 28,59

Son los datos que vimos en el Tema 3 con el dato medido en libras ahora en kg y sin los dos datos que consideramos como outliers. Por eso hemos pasado de tener un tamaño muestral 20 a 18

- **Media** = $\frac{29,43+35,18+\dots+28,59}{18} = 32,43$
- **Mediana:** Ordenamos los datos

26,15 28,59 29,07 29,43 29,48 29,58 30,81 31,07 31,12
32,69 33,34 33,66 34,21 34,53 35,18 35,49 37,39 42,03

Como hay dos datos en el centro, tomamos como mediana la media de estos dos datos

$$\text{Mediana} = \frac{31,12+32,69}{2} = 31,91$$

- **Moda:** Todos los valores aparecen una sola vez; pero podríamos agrupar los valores en clases como hacíamos para calcular los diagramas de barras e histogramas

Comparación entre Medidas de Centralidad

Ejemplo 3: 3 1 1 4 5 2

Ejemplo 4: 3 1 1 4 20 2

- **Media:** Cálculo razonable y facil

Pero

- No se puede calcular con datos medidos en una escala ordinal
- se ve afectada por las observaciones extremas cuando el n° de observaciones es reducido

En Ejemplo 3: media=2,67

En Ejemplo 4: media=5,17

- **Mediana:** Cuando los datos vienen agrupados en intervalos no es facil calcularla
- **Moda:** Puede estar muy alejada del centro de los datos por lo que no es buena medida de centralidad
En Ejemplos 3 y 4: moda=1

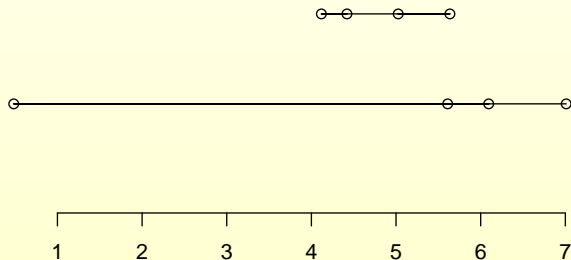
Medidas de Dispersión

Con las medidas de centralidad sabemos alrededor de que punto se encuentran los datos

Con las medidas de dispersión sabemos como de agrupados se encuentran los datos alrededor del punto central

Ejemplo 5: 4,42 4,12 5,64 5,03 **Ejemplo 6:** 7,01 0,48 6,09 5,61

La media en ambos ejemplos es 4,8



Ejemplo 5: 4,42 4,12 5,64 5,03

- **Rango:** Valor mayor menos valor menor, $5,64-4,12=1,52$
- **Varianza:**
 - 1 A cada uno de los datos se les resta la media
 - 2 Los números resultantes se elevan al cuadrado y se suman
 - 3 El resultado se dividen por el nº de datos menos 1

$$\frac{(4,42-4,8)^2+(4,12-4,8)^2+(5,64-4,8)^2+(5,03-4,8)^2}{4-1} = 0,46$$

- **Desviación estandar:** Raíz cuadrada positiva de la Varianza, **0,68**

Los más utilizados son la varianza o su equivalente, la desviación estandar

Ejemplo 3: 3 1 1 4 5 2

Ejemplo 4: 3 1 1 4 20 2

- **Media:**

En Ejemplo 3: 2,67

En Ejemplo 4: 5,17

- **Mediana:** 2,5

- **Varianza con respecto a la media:**

En Ejemplo 3: 2,67

En Ejemplo 4: 54,17

- **Varianza con respecto a la mediana:**

- 1 A cada uno de los datos se les resta la mediana

- 2 Los números resultantes se elevan al cuadrado y se suman

- 3 El resultado se dividen por el n° de datos menos 1

- **Desviación estandar con respecto a la media:**

En Ejemplo 3: 7,36

En Ejemplo 4: 1,63