

Respecto de las preguntas de clase...

veamos

¿En qué difieren un gráfico de barras y un histograma?

- ▶ El gráfico de barras representa el porcentaje en la altura de la barra. Mientras que en un histograma el porcentaje se representa en el área de la barra.
- ▶ En el gráfico de barras, las barras se representan separadas para indicar que no hay continuidad entre las categorías. En un histograma barras adyacentes deben estar en contacto indicando que la variable es continua o discreta.

¿Qué miramos en un histograma?

- ▶ Rango de variación de los datos (Mínimo y Máximo).
- ▶ Intervalos más frecuentes
- ▶ ¿La distribución es unimodal o hay más de una moda (pico)?
- ▶ ¿La distribución es simétrica?
- ▶ Si es asimétrica, ¿la asimetría es a derecha o a izquierda?
- ▶ ¿En torno a qué valor están aproximadamente centrados los datos?
- ▶ ¿Cuán dispersos en torno a este centro están los datos ?
- ▶ ¿Hay datos atípico en relación a la mayoría de los datos?

Variables cuantitativas: Medidas de resumen

Resumimos la información de los datos mediante medidas de fácil interpretación que reflejen sus características más relevantes.

Las medidas de resumen son útiles para comparar conjuntos de datos y para presentar los resultados de un estudio.

Podemos agruparlas en:

Medidas de posición o localización: describen un valor alrededor del cual se encuentran las observaciones.

Medidas de dispersión o escala: expresan la variabilidad presente en un conjunto de datos.

Mediana y Media α -podada Muestrales

- ▶ n datos: x_1, x_2, \dots, x_n
- ▶ datos ordenados $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. En R: `sort(datos)`

Mediana Muestral

$$\tilde{x} = \begin{cases} x_{(k+1)} & \text{si } n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2} & \text{si } n = 2k \end{cases}$$

Mediana y Media α –podada Muestrales

- ▶ n datos: x_1, x_2, \dots, x_n
- ▶ datos ordenados $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. En R: `sort(datos)`

Mediana Muestral

$$\tilde{x} = \begin{cases} x_{(k+1)} & \text{si } n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2} & \text{si } n = 2k \end{cases}$$

Media α –podada Muestral

$$\bar{x}_\alpha = \frac{x_{([n\alpha])} + \dots + x_{(n-[n\alpha])}}{(n - 2[n\alpha])}$$

Ejemplo de Juguete

- ▶ $n = 10$ datos ordenados: 2, 5, 8, 10, 14, 17, 21, 25, 28, 40

- ▶ Media Muestral:

$$\bar{x} = \frac{170}{10} = 17$$

- ▶ Mediana Muestral:

$$\tilde{x} = \frac{x_{(5)} + x_{(6)}}{2} = 15.5$$

Media 10%podada Muestral

$$\bar{x}_{10} = \frac{x_{(2)} + \dots + x_{(9)}}{8} = 16$$

#¿Qué pasa si reemplazamos a 40 por 400?

Ejemplo de Juguete

- ▶ $n = 10$ datos ordenados: 2, 5, 8, 10, 14, 17, 21, 25, 28, 40

- ▶ Media Muestral:

$$\bar{x} = \frac{170}{10} = 17$$

- ▶ Mediana Muestral:

$$\tilde{x} = \frac{x_{(5)} + x_{(6)}}{2} = 15.5$$

Media 10%podada Muestral

$$\bar{x}_{10} = \frac{x_{(2)} + \dots + x_{(9)}}{8} = 16$$

#¿Qué pasa si reemplazamos a 40 por 400?

$$\bar{x} = 53, \tilde{x} = 15.5, \bar{x}_{10} = 16$$

Medidas de resumen: Percentiles

- ▶ n datos: x_1, x_2, \dots, x_n
- ▶ datos ordenados $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. En R: `sort(datos)`

El $\alpha \times 100\%$ percentil de la muestra es el valor debajo del cual se halla el $\alpha \times 100\%$ de los datos ordenados.

Hay distintas formas de calcularlos. Una de ellas es la siguiente:

- ▶ Ordenamos los n datos de menor a mayor.
- ▶ Hallamos el dato que ocupa la posición $\alpha(n+1)$ en la muestra ordenada. Si este valor no es entero, pueden interpolarse los datos que están en las dos posiciones adyacentes.

Percentiles Muestrales

Ejemplo: Tenemos 19 datos que ordenados son

1 1 2 2 3 4 4 5 5 6 7 7 8 8 9 9 10 10 11



Percentil	Posición	Valor	
10%	$0.10 (19+1) = 2$	1	
25%	$0.25 (19+1) = 5$	3	Cuartil Inferior
50%	$0.50 (19+1) = 10$	6	Mediana
75%	$0.75 (19+1) = 15$	9	Cuartil Superior
95%	$0.95 (19+1) = 19$	11	

- Notemos que el percentil 50% o segundo cuartil coincide con la mediana. Denotaremos Q_1 al primer cuartil (25%) y Q_3 al tercer cuartil (75%).
- Los cuartiles y la mediana dividen a la muestra en cuatro partes igualmente pobladas: 25% de la muestra en cada una de ellas.
- Entre Q_1 y Q_3 se halla el 50% central de los datos.

Pasos para hacer un boxplot

- ▶ Representamos una escala vertical u horizontal.
- ▶ Dibujamos una caja cuyos extremos son los cuartiles y dentro de ella un segmento que corresponde a la mediana.
- ▶ A partir de cada extremo dibujamos un segmento hasta el dato más alejado que está a lo sumo $1.5 d_I$ del extremo de la caja. Estos segmentos se llaman bigotes.
- ▶ Marcamos aquellos datos que están a más de $1.5d_I$ de cada extremo de la caja.

¿Qué vemos en un boxplot?

- ▶ Posición
- ▶ Dispersión
- ▶ Asimetría
- ▶ Longitud de las colas
- ▶ Puntos anómalos (outliers)