

Diagrama de caja

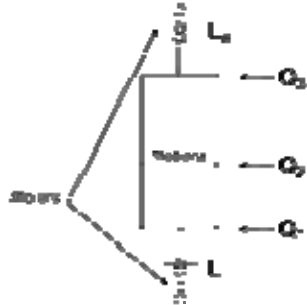


Diagrama de caja (Box-Plot).

Un **diagrama de caja** es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un rectángulo, la "caja", y dos brazos, los "bigotes".

Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos, a saber: mediana, primer cuartil, tercer cuartil, valor máximo y valor mínimo.

Esta presentación visual, asocia las cinco medidas que suelen trabajarse de forma individual. Presenta al mismo tiempo, información sobre la tendencia central, dispersión y simetría de los datos de estudio. Además, permite identificar con claridad y de forma individual, observaciones que se alejan de manera poco usual del resto de los datos. A estas observaciones se les conoce como valores atípicos.

Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la simetría de la distribución.

1.-Límite superior: Es el extremo superior del bigote. Las opiniones por encima de este límite se consideran atípicas.

2.-Tercer cuartil (Q₃): Por debajo de este valor se encuentran como máximo el 75% de las opiniones de los estudiantes.

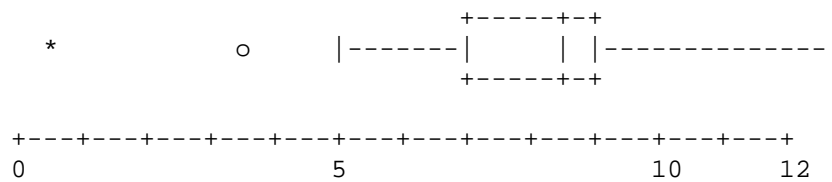
3.-Mediana: Coincide con el segundo cuartil. Divide a la distribución en dos partes iguales. De este modo, 50% de las observaciones están por debajo de la mediana y 50% está por encima.

4.-Primer cuartil (Q₁): Por debajo de este valor se encuentra como máximo el 25% de las opiniones de los estudiantes

5.-Límite inferior: Es el extremo inferior del bigote. Las opiniones por debajo de este valor se consideran atípicas.

6.-Valores atípicos: Opiniones que están apartadas del cuerpo principal de datos. Pueden representar efectos de causas extrañas, opiniones extremas o en el caso de la tabulación manual, errores de medición o registro.

Cómo expresarlo gráficamente



- Ordenar los datos y obtener el valor mínimo, el máximo, los cuartiles Q1, Q2 y Q3 y el Rango Inter Cuartilico (RIC)

En el ejemplo:

- Valor 7: es el Q1 (25% de los datos)
- Valor 8.5: es el Q2 o mediana (el 50% de los datos)
- Valor 9: es el Q3 (75% de los datos)
- Rango Inter Cuartilico RIC (Q3-Q1)=2
- Dibujar un rectángulo con Q1 y Q3 como extremos e indicar la posición de la mediana (Q2) mediante una línea.
- Para dibujar los bigotes, las líneas que se extienden desde la caja, hay que calcular los límites superior e inferior, LI y LS, que identifiquen a los valores atípicos.

Para ello se calcula cuándo se consideran atípicos los valores. Son aquellos inferiores a $Q1 - 1.5 \times RIC$ o superiores a $Q3 + 1.5 \times RIC$.

En el ejemplo:

- inferior: $7 - (1.5 \times 2) = 4$
- superior: $9 + (1.5 \times 2) = 12$

Ahora se buscan los últimos valores que **NO** son atípicos, que serán los extremos de los bigotes.

- En el ejemplo: 5 y 10
- Marcar como atípicos todos los datos que están fuera del intervalo (LI, LS).

En el ejemplo: 0.5 y 3.5

- Además, se pueden considerar valores extremadamente atípicos aquellos que exceden $Q1 - 3 \times RIC$ ó $Q3 + 3 \times RIC$.

De modo que, en el ejemplo:

- inferior: $7 - (3 \times 2) = 1$
- superior: $9 + (3 \times 2) = 15$

El valor 0.5 sería atípico extremo y el 3.5 sería atípico

Utilidad

- Proporcionan una visión general de la simetría de la distribución de los datos; si la mediana no está en el centro del rectángulo, la distribución no es simétrica.
- Son útiles para ver la presencia de valores atípicos también llamados outliers.