

Estadística Técnica

Gráfico de caja y extensiones

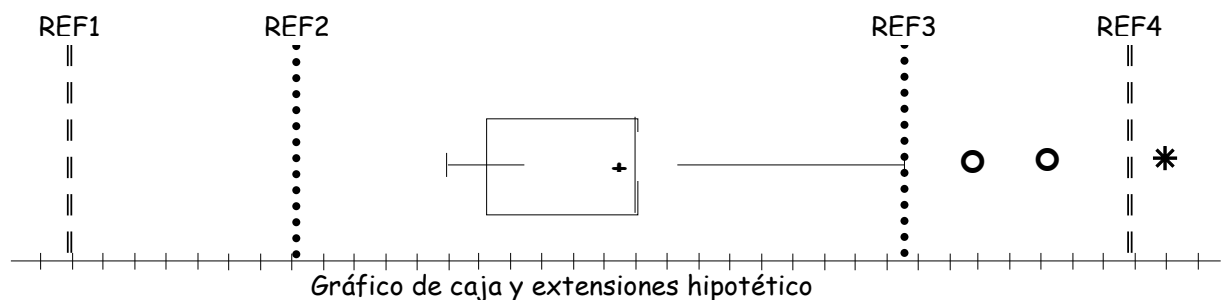
El gráfico de caja y extensiones fue descrito por Tukey, denominándolo 'box and whiskers'.

Para su construcción se utilizan cinco estadísticos de la distribución de frecuencias: el mínimo, el primer cuartil, la mediana, el tercer cuartil y el máximo.

Explicaremos su construcción paso a paso:

- Antes de comenzar la graficación debemos calcular algunos valores que serán necesarios para realizar el gráfico:
 - o Valor mínimo: x_{\min}
 - o Valor máximo: x_{\max}
 - o Mediana: \tilde{x}
 - o Media aritmética: \bar{x}
 - o Primer cuartil: Q_1
 - o Tercer cuartil: Q_3
 - o Rango intercuartílico: $RI = Q_3 - Q_1$
 - o $REF1 = Q_1 - 3.RI$
 - o $REF2 = Q_1 - 1,5.RI$
 - o $REF3 = Q_3 + 1,5.RI$
 - o $REF4 = Q_3 + 3.RI$
- Se traza una línea horizontal de longitud proporcional al recorrido de la variable, que llamaremos eje. Sobre el eje se señalarán las subdivisiones que se consideren necesarias, para representar los datos de la muestra.
- Paralelamente al eje se construye una caja rectangular con altura arbitraria y cuya base abarca desde el primer cuartil hasta el tercer cuartil. Como vemos, esta caja indica gráficamente el intervalo de variación de al menos el 50% de los valores centrales de la distribución.
- La caja se divide en dos partes, trazando una línea a la altura de la mediana. Cada una de estas partes indica, pues, el intervalo de variabilidad de al menos una cuarta parte de los datos.

- A la caja, así dibujada, se añaden dos guías paralelas al eje, que llamaremos **extensiones** o **bigotes**, una de cada lado, de la siguiente forma:
 - o el *primero* de estos segmentos se prolonga, hacia la izquierda, desde el primer cuartil (o sea, desde la caja) hasta el mínimo de la distribución o hasta el valor (observado en la muestra) igual o inmediato superior a la REF2, según cuál de estos valores sea mayor.
 - o el *segundo* de estos segmentos se prolonga, hacia la derecha, desde el tercer cuartil (o sea, desde la caja) hasta el máximo de la distribución o hasta el valor (observado en la muestra) igual o inmediato inferior a la REF3, según cuál de estos valores sea menor.
- Si alguno de los valores observados en la muestra queda fuera del intervalo cubierto por la caja y estas extensiones, se señala en el gráfico mediante un asterisco o cualquier otro símbolo que lo represente como un punto.
- Estos datos son los llamados *valores atípicos* ('outliers'), que son valores muy alejados de los valores centrales de la distribución. Lo indicaremos con ○.
- Si se observan valores menores que $Q_1 - 3.RI$ (REF1) o valores mayores que $Q_3 + 3.RI$ (REF4), éstos son los llamados *valores anómalos*, que son valores más alejados que los atípicos de los valores centrales de la distribución. Lo indicaremos con *.
- Finalmente, se indica con un signo + el valor de la *media aritmética*.



Utilidades del gráfico de caja y extensiones

- El gráfico de caja y extensiones nos proporciona la posición relativa de la mediana, los cuartiles y extremos de una distribución.
- El gráfico de caja y extensiones nos proporciona información sobre los valores atípicos, sugiriendo la necesidad de utilizar (o no) determinados estadísticos.

- El gráfico de caja y extensiones nos informa de la simetría o asimetría de la distribución.
- El gráfico de caja y extensiones se puede utilizar para comparar la misma variable en dos muestras distintas.



Ejemplo:

A continuación, el siguiente conjunto de datos muestra las alturas de un grupo de hombres y de mujeres, por separado.

| Altura H | Altura M |
|----------|----------|
| 160 | 165 |
| 180 | 165 |
| 181 | 166 |
| 185 | 166 |
| 185 | 167 |
| 186 | 170 |
| 188 | 170 |
| 188 | 171 |
| 189 | 172 |
| 189 | 173 |
| 190 | 173 |
| 190 | 175 |
| 191 | 175 |
| 191 | 180 |
| 195 | 190 |

Gráfico realizado mediante software estadístico. Realice los cálculos y el gráfico y compárelo con su solución.

Resumen Estadístico

| | <i>Altura H</i> | <i>Altura M</i> |
|---------------------------|-----------------|-----------------|
| Recuento | 15 | 15 |
| Promedio | 185,867 | 171,867 |
| Mediana | 188,0 | 171,0 |
| Desviación Estándar | 8,14043 | 6,61024 |
| Coefficiente de Variación | 4,37972% | 3,84614% |
| Mínimo | 160,0 | 165,0 |
| Máximo | 195,0 | 190,0 |
| Rango | 35,0 | 25,0 |
| Cuartil Inferior | 185,0 | 166,0 |
| Cuartil Superior | 190,0 | 175,0 |
| Rango Intercuartílico | 5,0 | 9,0 |

Cuartiles para el grupo de varones

$$^{\circ}Q_1 = \frac{1(15 + 1)}{4} = ^{\circ}4 \rightarrow Q_1 = 185$$

$$^{\circ}Q_2 = \frac{2(15 + 1)}{4} = ^{\circ}8 \rightarrow Q_2 = 188$$

$$^{\circ}Q_3 = \frac{3(15 + 1)}{4} = ^{\circ}12 \rightarrow Q_3 = 190$$

Cuartiles para el grupo de mujeres

$$^{\circ}Q_1 = \frac{1(15 + 1)}{4} = ^{\circ}4 \rightarrow Q_1 = 166$$

$$^{\circ}Q_2 = \frac{2(15 + 1)}{4} = ^{\circ}8 \rightarrow Q_2 = 171$$

$$^{\circ}Q_3 = \frac{3(15 + 1)}{4} = ^{\circ}12 \rightarrow Q_3 = 175$$

Referencias para el grupo de varones

$$Ref1: 185 - 3 \cdot 5 = 170$$

$$Ref2: 185 - 1,5 \cdot 5 = 177,5$$

$$Ref3: 190 + 1,5 \cdot 5 = 197,5$$

$$Ref3: 190 + 3 \cdot 5 = 205$$

Referencias para el grupo de mujeres

$$Ref1: 166 - 3 \cdot 9 = 139$$

$$Ref2: 166 - 1,5 \cdot 9 = 152,5$$

$$Ref3: 175 + 1,5 \cdot 9 = 188,5$$

$$Ref3: 175 + 3 \cdot 9 = 202$$

Gráfico Caja y Bigotes

