

## Una forma “rápida” de ordenar datos numéricos (Diagrama de Tallo y Hoja)

Los siguientes datos corresponden a los precios de la libra de cobre en la Bolsa de Metales de Londres en Enero de 2000.

Día	Precio		Día	Precio		Día	Precio
1			12	82.7		23	
2			13	84.2		24	84.9
3			14	83.8		25	84.1
4	83.1		15			26	83.6
5	82.5		16			27	82.5
6	83.1		17	83.7		28	83.5
7	83.1		18	83.7		29	
8			19	85.0		30	
9			20	86.1		31	82.2
10	83.0		21	85.6			
11	82.5		22				

Una forma de ordenar “rápidamente” estos datos es a través del método conocido como diagrama de tallo y hoja.

Cada dato se divide en dos partes, una conocida como **tallo**, que se pone en una primera columna, y la otra que se denomina **hoja**, que se pone en fila en frente al tallo correspondiente. Es importante notar que no existe una regla fija para determinar cuál es el tallo y cuál es la hoja de un dato. En nuestro caso, cada dato lo separamos en su parte entera (tallo) y su parte decimal (hoja). Por ejemplo:

$$83.1 = \boxed{83} \boxed{1} \quad 85.0 = \boxed{85} \boxed{0}$$

Tallo Hoja                      Tallo Hoja

El siguiente diagrama muestra los datos presentados en la tabla anterior, que llamaremos **diagrama de tallo y hoja**.

Tallos	Hojas									
82.	2	5	5	5	7					
83.	0	1	1	1	5	5	6	7	7	8
84.	1	2	3							
85.	0	6								
86.	1									

## Una forma visual de analizar datos (**Diagrama de Caja**)

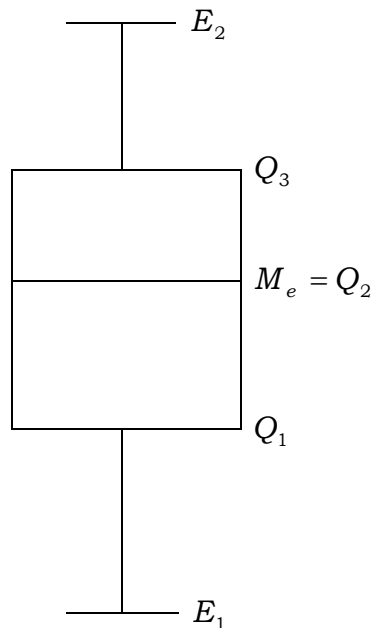
Tanto la media como la desviación estándar podrían no ser medidas adecuadas para representar un conjunto de valores, pues

- a) son afectadas, de forma exagerada, por valores extremos,
- b) sólo con estos dos valores no podemos darnos una idea acerca de la asimetría de la distribución de los datos.

Para tratar de superar estas dificultades y como una forma de resumir gráficamente los datos, se sugiere confeccionar un diagrama, conocido como **diagrama de caja**, que contenga la siguiente información:

- i) La Mediana,  $M_e = Q_2$ ,
- ii) Los Extremos,  $E_1$  y  $E_2$ , (mínimo y máximo respectivamente)
- iii) Los Cuartiles,  $Q_1$  y  $Q_3$ .

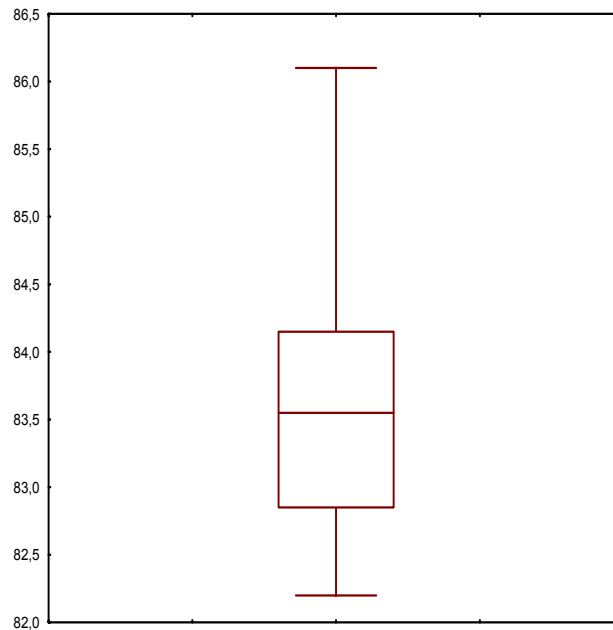
El diagrama de caja consiste en graficar estos 5 valores en la siguiente forma: La mitad central de los datos, que va desde el primer hasta el tercer cuartil, se representa mediante un rectángulo. La mediana se identifica mediante una barra horizontal dentro de esta caja. Una línea vertical se extiende desde el tercer cuartil hasta el valor máximo y otra se extiende desde el primer cuartil hasta el mínimo. Las dos líneas horizontales dibujadas a ambos extremos de la caja, se suelen denominar “**bigotes**”. La figura siguiente ilustra este diagrama, que también es conocido como “**diagrama de caja y bigote**”.



La distancia entre  $E_1$  y  $Q_2$  es la diferencia  $Q_2 - E_1$ , que se denomina **dispersión inferior**. La distancia entre  $Q_2$  y  $E_2$  es la diferencia  $E_2 - Q_2$ , que se denomina **dispersión superior**. Recordemos también que la distancia entre  $Q_1$  y  $Q_3$ , esto es,  $Q_3 - Q_1$ , se denomina **rango intercuartil**.

La comparación de estas distancias entrega información sobre la dispersión de los datos y también sobre la forma de su distribución. Por ejemplo, si la distribución fuese relativamente simétrica, entonces la dispersión inferior debería ser aproximadamente igual a la dispersión superior. También,  $Q_2 - Q_1$  debería ser similar a  $Q_3 - Q_2$ , y  $Q_1 - E_1$  similar a  $E_2 - Q_3$ .

A continuación se presenta el **diagrama de caja** de los precios de la libra de cobre en la Bolsa de Metales de Londres en Enero de 2000.



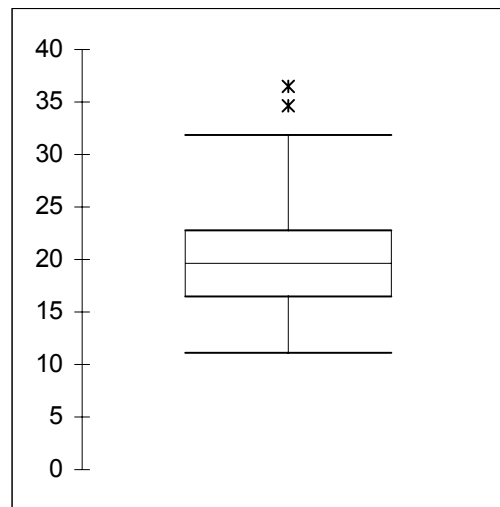
**Observación:** Generalmente un (\*) indica un punto extremo. Un punto extremo es un valor que es inconsistente con el resto de los datos. La definición estándar de un punto extremo, es un valor que es mayor que  $Q_3$  o menor que  $Q_1$  en más de 1.5 veces el rango intercuartil. En caso de existir observaciones extremas, los bigotes del diagrama representan a la observación mayor y menor dentro de los límites.

**Ejemplo:** Los siguientes datos corresponden al consumo de combustible (millas/galón) de 25 vehículos.

18,90	22,12	21,50	17,80	13,27
20,00	<b>34,70</b>	19,70	23,54	23,90
18,25	30,40	17,80	21,47	19,73
20,07	16,50	14,39	16,59	13,90
11,20	<b>36,50</b>	14,89	31,90	16,50

<b>Estadígrafo</b>	<b>Valor</b>
$Q_1$	16.50
$M_e$	19.70
$Q_3$	22.83
$RIQ$	6.33

La información la podemos visualizar en el siguiente diagrama de caja:



Note que

$$Q_1 - 1.5RIQ = 16.5 - 1.5 \cdot 6.33 = 7.005 \text{ (No hay observaciones menores)}$$

$$Q_3 + 1.5RIQ = 22.83 + 1.5 \cdot 6.33 = 32.325 \text{ (Hay dos observaciones mayores (*))}$$