

INDICADORES ESTADÍSTICOS

Explorando los Datos



“Generalmente las decisiones se toman fundamentadas en indicadores resumen”

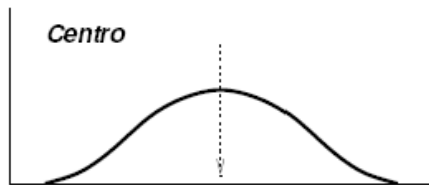
¿Cuánto tiempo tarda un estudiante en desplazarse desde su casa hasta la universidad?



Medidas de Resumen

*“En ocasiones la evaluación grafica puede ser una fuente de error por **percepción**, los indicadores contribuyen a corroborar lo observado”*

TENDENCIA CENTRAL

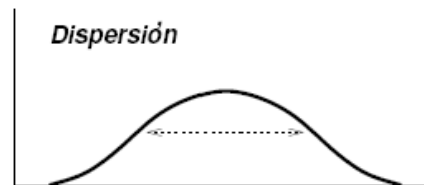


- MEDIA (PROMEDIO) → μ ó \bar{X}
- MEDIANA → Me
- MODA → Mo



Representan en un solo valor central, a un conjunto de datos.

DISPERSIÓN

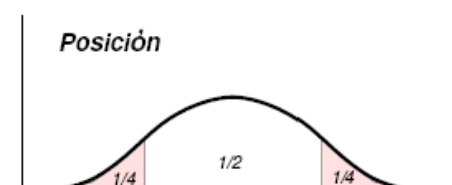


- RANGO → R
- DESVIACIÓN ESTANDAR → σ ó s
- VARIANZA → σ^2 ó s^2
- COEFICIENTE DE VARIACIÓN → CV

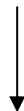


Indican sobre la variabilidad de los datos, con respecto a un valor central.

POSICIÓN



- Cuartiles → Q_1, Q_2, Q_3
- Percentiles → P_{25}, P_{50}, P_{75}



Medidas para resumir la distribución de los datos.

Indicadores de tendencia central

"La estadística es una ciencia en la que se demuestra que, si mi vecino tiene dos coches y yo ninguno, los dos tenemos uno"

George Bernard Shaw

Es increíble, lo importante que es conocer en qué punto se aglomeran los datos en una muestra. El valor alrededor del cual se concentran los datos ayuda a determinar cuál es la escala con la que se debe comparar si los valores de una variable aleatoria son "normales" o no.

Las fórmulas para hallar estos valores se llaman estadísticos de centralidad o más frecuentemente, indicadores de tendencia central

Indicadores de tendencia central

- MEDIA → μ ó \bar{X}

La **media** es el valor representativo de todos los datos. A la media se le suele llamar *promedio*.

Si las observaciones son x_1, x_2, \dots, x_n , su media es:

Media Aritmética

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + \dots + x_n}{n}$$

Datos puntuales

Media Ponderada:

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n}$$

Datos agrupados

- Es un valor que está en el centro o punto medio de un conjunto de datos.
- Tiene como objetivo resumir los datos en un valor típico o representativo del conjunto de valores.

Ejemplo:

Los siguientes datos se han obtenido al observar el número de chocolatinas defectuosas en una muestra de 10 cajas de un lote de producción:

2 3 2 2 2 3 1 3 0 4

¿Cuál es el número promedio de unidades defectuosas por caja?

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2 + 3 + 2 + 2 + 2 + 3 + 1 + 3 + 0 + 4}{10} = \frac{22}{10} = 2.2$$

El proceso presenta en promedio 2.2 unidades defectuosas por caja.

- MEDIA $\rightarrow \mu$ ó \bar{X}

Ejemplo

Calcula el numero promedio de unidades defectuosas por caja :

Unidades defectuosas	ni
0	3
1	11
2	13
3	8
4	4
5	1
Total	40

$$\bar{x} = \frac{\sum_{i=1}^6 (x_i n_i)}{40} = \frac{(0 * 3) + (1 * 11) + \dots + (5 * 1)}{40} = 2.05 \approx 2$$

Las cajas en promedio tiene 2 unidades defectuosas.

OBSERVACIONES:

- La media puede ser la medida de centralidad más utilizada y más informativa. No obstante la media tiene una desventaja, la cual radica en que es sensible a los valores extremos.

Ejemplo:

Sean los valores: 2, 4, 3, 6, 5

$$\bar{x} = \sum_{i=1}^5 \frac{x_i}{5} = \frac{2+4+3+6+5}{5} = \frac{20}{5} = 4$$

Sean los valores: 2, 4, 3, 6, 50

$$\bar{x} = \sum_{i=1}^5 \frac{x_i}{5} = \frac{2+4+3+6+50}{5} = \frac{65}{5} = 13$$

- MEDIA $\rightarrow \mu$ ó \bar{X}

Ejemplo:

Carlos quiere decidir entre dos ciudades donde puede trabajar. Sin embargo, considera que la temperatura de cada ciudad es un factor importante en su elección.

Las mediciones de la temperatura máxima (°C), en cada una de las ciudades durante los últimos ocho meses son las siguientes:

Ciudad 1: 25 24 26 27 28 27 26 27

Ciudad 2: 16 22 24 28 32 31 30 27

¿Qué le podemos recomendar con base a la temperatura máxima promedio?

$$\bar{x}_1 = 26.25 \quad y \quad \bar{x}_2 = 26.25$$

Las dos ciudades presentan la misma temperatura máxima promedio en los últimos ocho meses, que es de 26.25 °C.

Indicadores de tendencia central

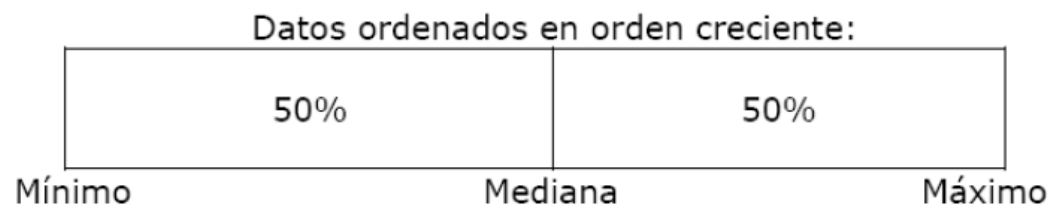
- MEDIANA → **Me**

La **mediana** (Me) representa el punto medio de las observaciones ordenadas; la mitad (50%) de las observaciones están por debajo de la mediana y la otra mitad (50%) están por encima.

Si x_1, x_2, \dots, x_n corresponde a un conjunto de datos ordenados de manera ascendente de una variable **X**, entonces la mediana puede ser calculada como:

$$Me = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ es impar} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{si } n \text{ es par} \end{cases}$$

- Es un valor que está en el centro o punto medio de un conjunto de datos.
- Tiene como objetivo resumir los datos en un valor típico o representativo del conjunto de valores.



Ejemplo:

Número de chocolatinas defectuosas en una muestra de 10 cajas de un lote de producción:

0 1 2 2 2 2 3 3 3 4

¿Cuál es la mediana del número de unidades defectuosas por caja?

$$Me = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} = \frac{X_{(5)} + X_{(6)}}{2} = \frac{2+2}{2} = 2$$

El 50% de las cajas presentan 2.0 o menos unidades defectuosas.

OBSERVACIONES:

- La mediana no es influenciado por los valores extremos.

Ejemplo:

Sean los valores: 2, 3, 4, 5, 6

$$\bar{x} = \sum_{i=1}^5 \frac{x_i}{5} = \frac{2+4+3+6+5}{5} = \frac{20}{5} = 4$$

$$Me = X_{\left(\frac{5+1}{2}\right)} = X_{(3)} = 4$$

Sean los valores: 2, 3, 4, 6, 50

$$\bar{x} = \sum_{i=1}^5 \frac{x_i}{5} = \frac{2+4+3+6+50}{5} = \frac{65}{5} = 13$$

$$Me = X_{\left(\frac{5+1}{2}\right)} = X_{(3)} = 4$$

- MEDIANA para Datos Agrupados

$$Me = LI_{clase} + \left(\frac{50 - FRA_{anterior}}{FR_{clase}} \right) * amplitud$$

Indicadores de tendencia central

- MODA → M_o

M_o : Es el valor que más se repite

- Cuando la variable de interés, es de naturaleza discreta, la moda (M_o) corresponde al dato de la muestra que tiene mayor frecuencia.
- Cuando se trata de una variable de naturaleza continua, la moda corresponde al(os) valor(es) alrededor del(os) cual(es) se produce una mayor concentración de datos, es decir a los puntos de mayor densidad de frecuencia.
- A diferencia de los otros indicadores este es el único que puede ser calculado cuando observamos variables cualitativas.

- Es un valor que está en el centro o punto medio de un conjunto de datos.
- Tiene como objetivo resumir los datos en un valor típico o representativo del conjunto de valores.

Ejemplo:

Número de chocolatinas defectuosas en una muestra de 10 cajas de un lote de producción:

0 1 2 2 2 2 3 3 3 4

¿Cuál es la moda del del número de unidades defectuosas por caja?

$$M_o = 2$$

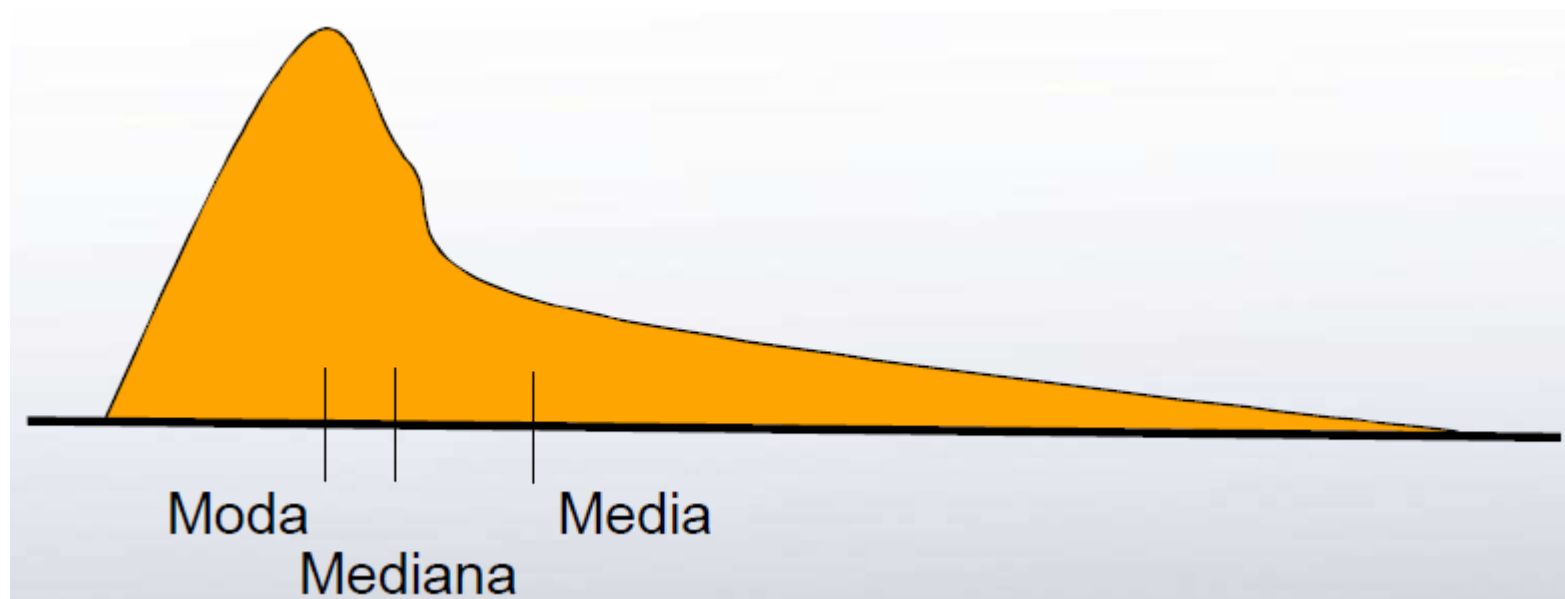
La mayoría de las cajas presentan 2.0 unidades defectuosas.

- MODA para Datos Agrupados

$$Mo = LI_{clase} + \left(\frac{FR_{clase} - FR_{anterior}}{2 * FR_{clase} - FR_{anterior} - FR_{posterior}} \right) * amplitud$$

Medidas de Tendencia Central

¿Cuál elegir?



Medidas de Tendencia Central

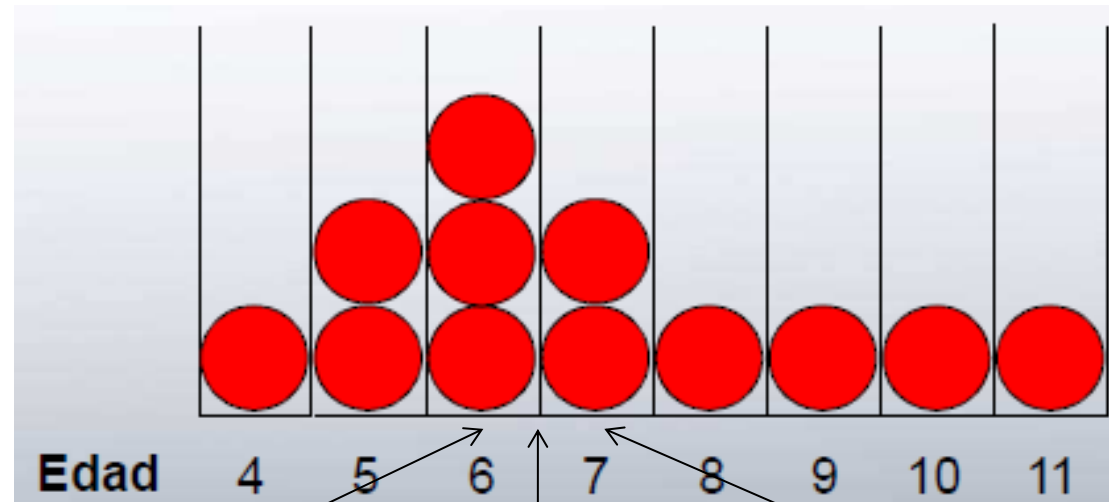
Consideremos la siguiente tabla de datos:

Individuo	Nombre	Edad
1	Juan	4
2	Alberto	5
3	Inés	6
4	Aurora	5
5	Rodrigo	6
6	Martha	7
7	Juana	8
8	Roberto	6
9	Silvia	7
10	Ana	11
11	Andres	10
12	Carlos	9

Medidas de Tendencia Central

Consideremos la siguiente tabla de datos:

Edad	4	5	6	7	8	9	10	11	Total
Frecuencia	1	2	3	2	1	1	1	1	12



Moda = 6

Mediana = 6.5

Media = 7

INDICADORES DE DISPERSION

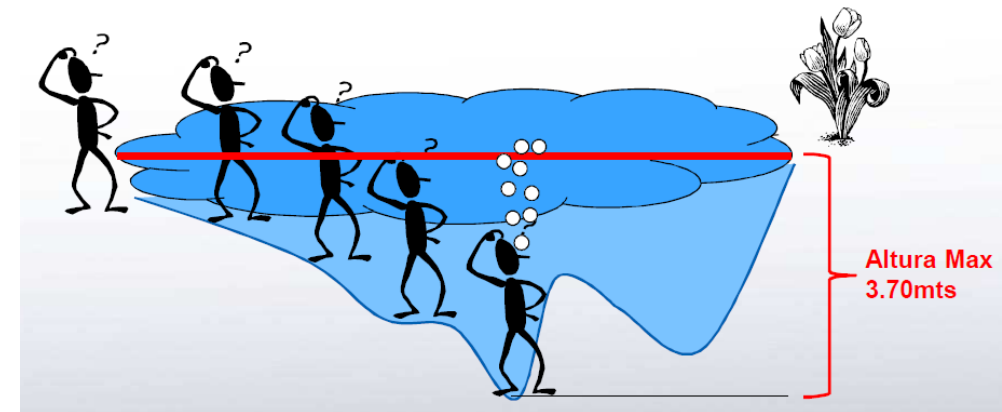
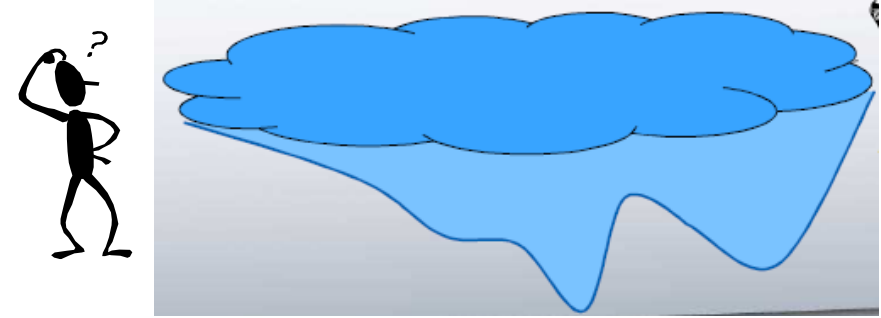
¿Por qué se estudia la dispersión?

Pregunta: ¿Un Guía turístico le dice a usted que cierto río tiene una profundidad los siguientes indicadores:

Media:	100 cms
Mediana:	50 cms
Moda:	30 cms

¿Con esta información cruzaría usted el río?

- Los Indicadores de Tendencias central no trabajan solos, deben apoyarse con alguna medida de dispersión
- Un Índice de dispersión pequeño, indica baja variabilidad, por ende, el valor de tendencia central será más confiable.
- Un índice de dispersión grande, indica gran variabilidad, esto implica que el Indicador de tendencia central sea poco confiable

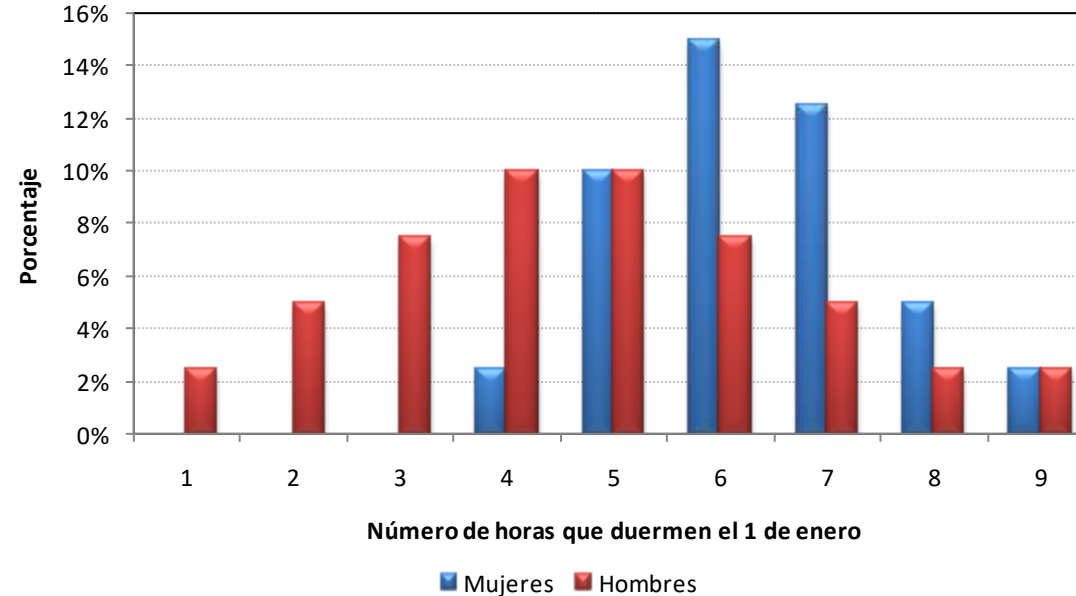


INDICADORES DE DISPERSION

Todo conjunto de datos tiene al menos dos características principales:

CENTRO Y DISPERSIÓN

Los gráficos de barra, histogramas, de puntos, entre otros, nos dan cierta idea sobre ellos.



Indicadores de Dispersión

- Varianza → s^2

La **varianza** representa el promedio de las desviaciones al cuadrado entre cada observación y la media.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

- Es siempre positiva.
- Si todos los valores son iguales, entonces la varianza es cero.
- Le afectan los valores extremos.
- La varianza no se interpreta.

=VAR(DATOS)

- Un índice de dispersión pequeño, indica baja variabilidad, por ende el valor de tendencia central será mas confiable.
- Un índice de dispersión grande, indica gran variabilidad, esto implica que el Indicador de tendencia central sea poco confiable.

- Desviación Estándar → s

La **desviación estándar**, indica en promedio cuanto se desvían las observaciones con respecto a la media.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- Mientras más alejadas están las observaciones del promedio, mayor será la desviación estándar.

=DESVEST(DATOS)

- Desviación Estándar para datos agrupados → s

$$s = \sqrt{\frac{\sum_{clases} (MC - \bar{X})^2 * FA}{n - 1}} = \sqrt{\frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}}$$

Ejemplo:

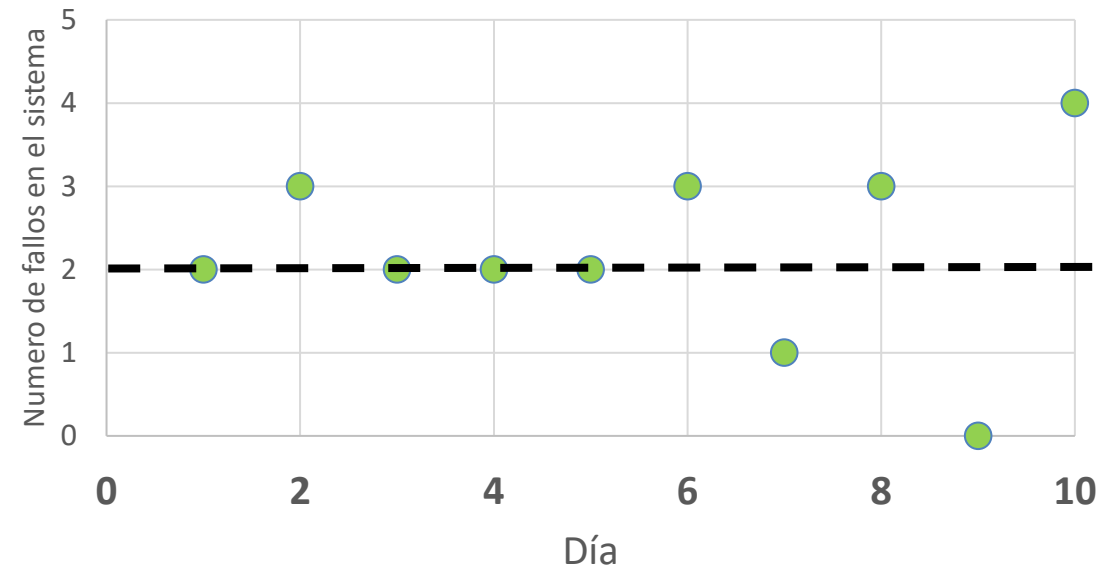
Los siguientes datos se han obtenido al observar el número de chocolatinas defectuosas en una muestra de 10 cajas de un lote de producción:

2 3 2 2 2 3 1 3 0 4

¿Qué tanto varía la cantidad de unidades defectuosas respecto al promedio?

$$s = 1.14$$

La cantidad de unidades defectuosas del proceso varia respecto a su promedio en 1.14 unidades.



Indicadores de Dispersión

Por la estructura de la varianza se sabe que cuando aumenta la dispersión el valor de la varianza aumenta, al igual que la desviación estándar.

pero, qué se respondería a las preguntas:

- ¿una desviación estándar de 200 metros indica que hay poca o mucha dispersión?
- ¿una desviación estándar de 100 kilogramos podría ser grande?



Depende de la magnitud de los datos.

Coeficiente de Variación

• Coeficiente de Variación → CV

Indica el grado de variabilidad porcentual de los datos con respecto a la media. Se suele interpretar en términos porcentuales.

$$CV = \frac{s}{\bar{X}}$$

- Útil para comparar la variabilidad relativa de una característica, en poblaciones que tiene diferente media.

Si **CV** < 30% → la variabilidad es baja.
Si 30% ≤ **CV** ≤ 80% → la variabilidad es moderada.
Si **CV** > 80% → la variabilidad es alta.

=DESVEST(DATOS)/PROMEDIO(DATOS)

$$CV = \frac{s}{\bar{x}} = \frac{1.14}{2.2} = 0.52 \approx 52\%$$

Ejercicio

El entrenador de un equipo de baloncesto duda entre seleccionar a Eva o a Tatiana. Los puntos conseguidos por cada una, en una semana de entrenamiento, fueron los siguientes:

Eva	18	23	22	24	19	25	16
Tatiana	18	26	18	28	22	17	18

Calcule e interprete:

a) ¿Cuál de las dos tiene mejor promedio de puntos?

Ambas tienen el mismo promedio de 21 puntos

b) ¿Cuál de las dos es más regular en sus puntos?

Es más regular Eva, porque la dispersión de los puntos es menor.

Desv. Est. (Eva) = 3.367 Desv. Est. (Tatiana) = 4.435

Ejercicio

Según la revista Chemical Engineering, una propiedad importante de una fibra es su absorción del agua. Se toma una muestra aleatoria de 20 pedazos de fibra de algodón y se mide la absorción de cada uno. Los valores de absorción son los siguientes:

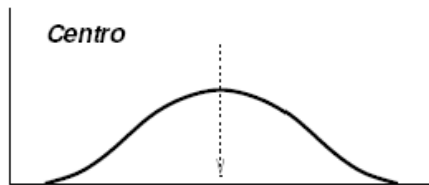
18.71	21.4	20.7	21.8	19.3	22.4	20.2
23.71	19.4	20.5	18.9	20.3	23.0	22.9
19.25	21.8	22.1	19.8	18.0	21.1	

- ¿Cuál es la variable de interés y como se clasifica?
- Hallar las medidas de centralidad (media, mediana).
- Hallar las medidas de dispersión (R, S y CV).
- ¿Qué puede decir sobre la absorción de la fibra de acuerdo con los indicadores anteriores?

Medidas de Resumen

*“En ocasiones la evaluación grafica puede ser una fuente de error por **percepción**, los indicadores contribuyen a corroborar lo observado”*

TENDENCIA CENTRAL

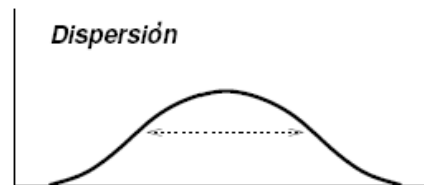


- MEDIA (PROMEDIO) → μ ó \bar{X}
- MEDIANA → Me
- MODA → Mo



Representan en un solo valor central, a un conjunto de datos.

DISPERCIÓN

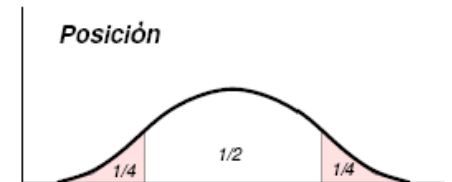


- RANGO → R
- DESVAIACIÓN ESTANDAR → σ ó s
- VARIANZA → σ^2 ó s^2
- COEFICIENTE DE VARIACIÓN → CV

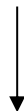


Indican sobre la variabilidad de los datos, con respecto a un valor central.

POSICIÓN



- Cuartiles → Q_1, Q_2, Q_3
- Percentiles → P_{25}, P_{50}, P_{75}



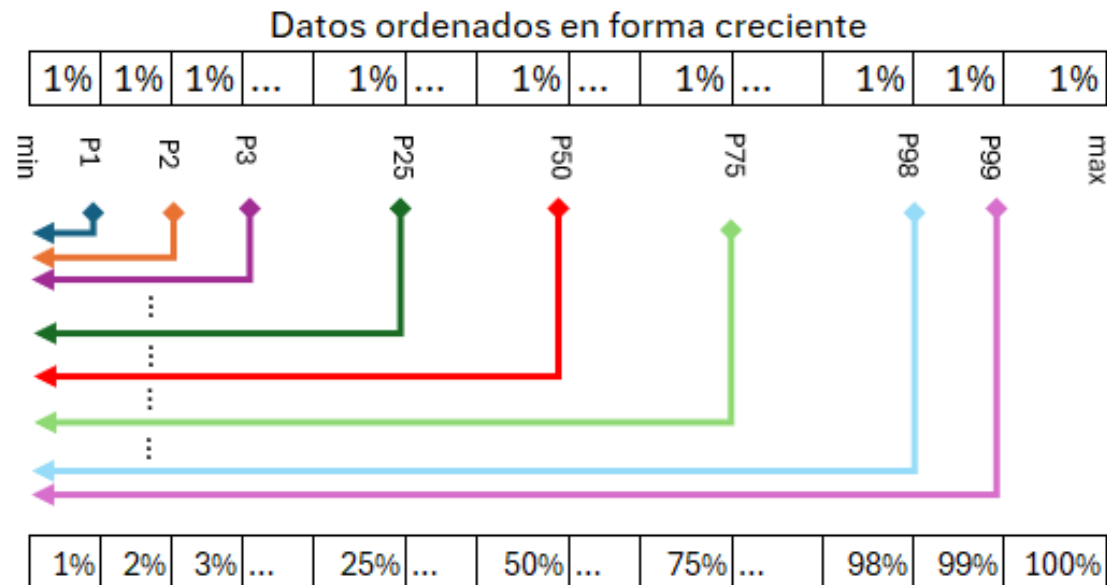
Medidas para resumir la distribución de los datos.

Indicadores de Posición

Percentiles:

P1, P2, P3, ..., P99

=PERCENTIL.EXC(DATOS;#)

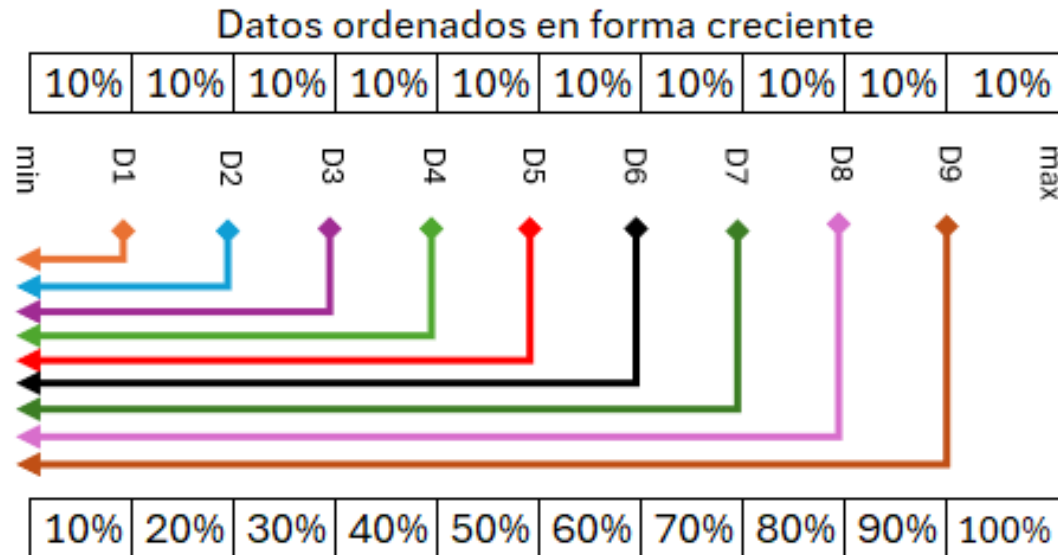


Son 99 valores (P1 , P 2 , ...,P 99) que dividen la serie de datos en 100 partes iguales del 1% de los datos cada una

Indicadores de Posición

Deciles:

D1, D2, D3, D4, ..., D9



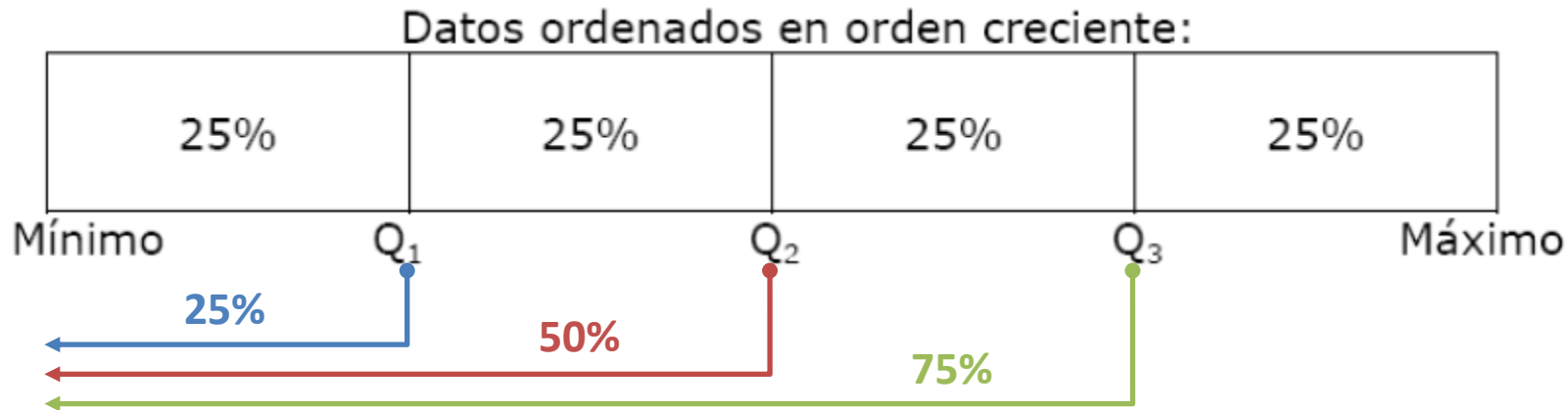
Son 9 valores (D1 , D2 , ...,D9) que dividen la serie de datos en 10 partes iguales del 10% de los datos cada una

Indicadores de Posición

Cuartiles:

Q_1, Q_2, Q_3

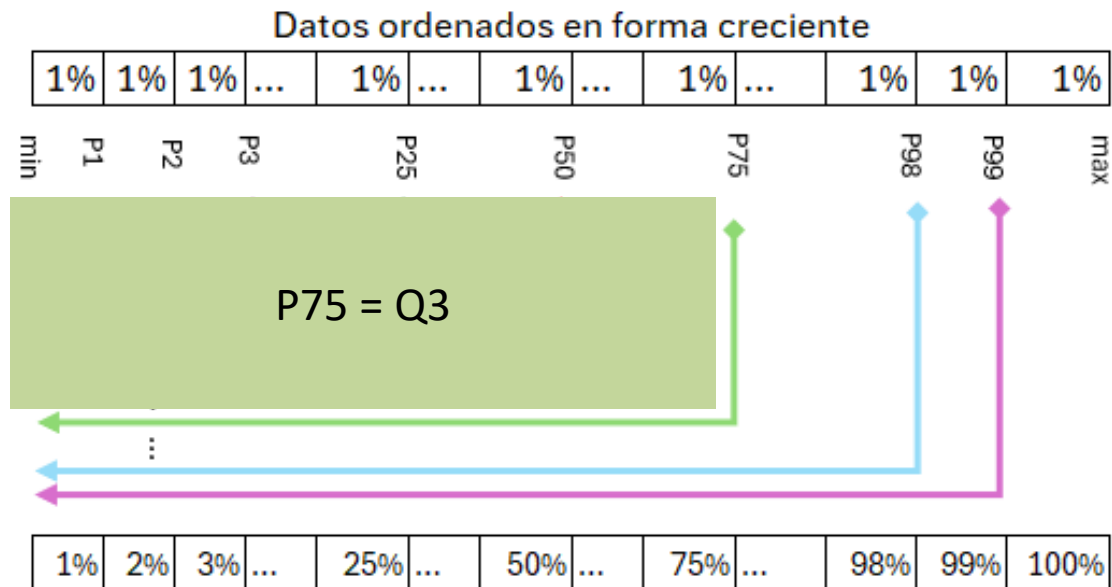
`=CUARTIL(DATOS;#)`



Los cuartiles, son tres valores Q_1 , Q_2 y Q_3 , que dividen a las observaciones de forma ordenada, en cuatro partes que contienen aproximadamente el mismo número de datos. Estos tres indicadores junto al **mínimo (mín)** y el **máximo (máx)** conforman los 5 números resumen.

«La representación grafica de los Cuartiles es el diagrama de cajas y alambres»

Indicadores de Posición



Indicadores de Posición

Cálculo del p-ésimo percentil:

Paso 1: Ordene los datos de manera ascendente.

Paso 2: Calcule: $L_p = \frac{P}{100} \times (n + 1)$ Donde **P** es el percentil de interés y **n** es la cantidad de observaciones.

Paso 3:

- a) Si **Lp** no es entero se interpola un valor a partir de las dos posiciones encontradas.
- b) Si **Lp** es entero el valor del percentil corresponde a la posición encontrada.

Cálculo del p-ésimo percentil

Los siguientes datos corresponden a las edades de 14 personas seleccionadas al azar, entre cierta clase de empleados de la población objetivo de estudio:

25, 38, 29, 42, 39, 54, 23, 33, 45, 45, 26, 34, 30, 31.

¿Cómo determino los cuartiles?

1. Ordenar los datos de menor a mayor

23	25	26	29	30	31	33	34	38	39	42	45	45	54
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)

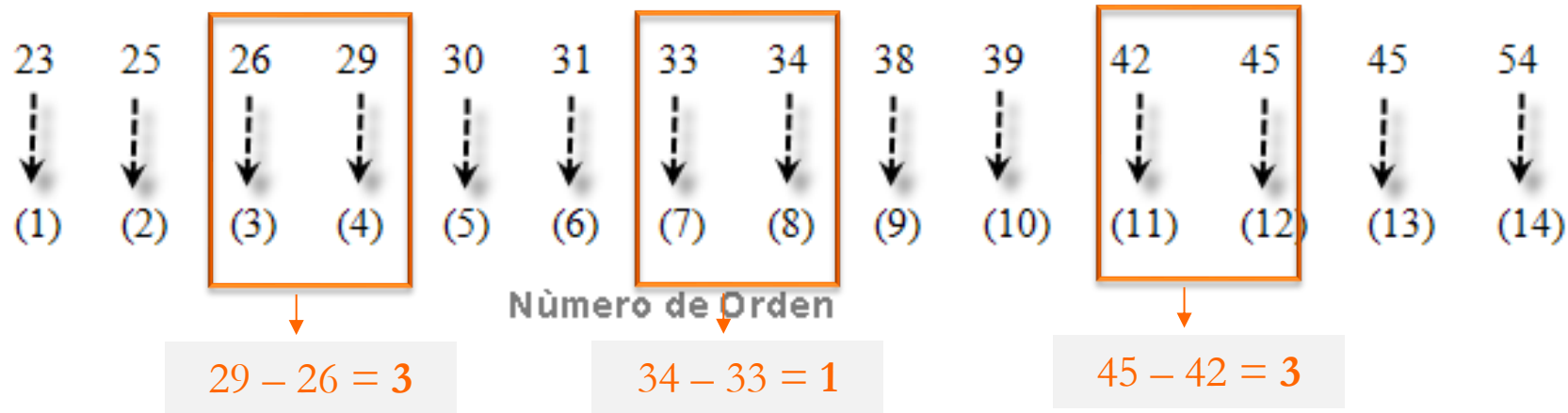
Número de Orden

Q1=P25= Valor que es superior al **25%** de las observaciones

Q2=P50= Valor que es superior al **50%** de las observaciones

Q3 =P75= Valor que es superior al **75%** de las observaciones

Cálculo del p-ésimo percentil



Primer cuartil

$$L_{25} = \frac{25 * (14 + 1)}{100} = 3.75 \rightarrow 0.75$$

$$P_{25} = Q_1 = 26 + (3 * 0.75) = 28.25$$

Segundo cuartil

$$L_{50} = \frac{50 * (14 + 1)}{100} = 7.5 \rightarrow 0.50$$

$$P_{50} = Q_2 = Me = 33 + (1 * 0.5) = 33.5$$

Tercer cuartil

$$L_{75} = \frac{75 * (14 + 1)}{100} = 11.25 \rightarrow 0.25$$

$$P_{75} = Q_3 = 42 + (3 * 0.25) = 42.75$$

Ejemplo

La resistencia a la tensión del caucho de silicio se considera una función de la temperatura de vulcanizado. Se llevó a cabo un estudio en el que se prepararon muestras de 12 especímenes del caucho utilizando temperaturas de vulcanizado de 20°C y 45°C. Los siguientes datos presentan los valores de resistencia a la tensión en megapascuales:

20°C

2.07	2.14	2.22	2.03	2.21	2.03
2.05	2.18	2.09	2.14	2.11	2.02

45°C

2.52	2.15	2.49	2.03	2.37	2.05
1.99	2.42	2.08	2.42	2.29	2.01

¿Determinar las cinco medidas resumen de la resistencia a la tensión para cada temperatura?

Indicador	20°C	45°C
Mínimo	2,02	1,99
Cuartil 1	2,04	2,04
Cuartil 2	2,10	2,22
Cuartil 3	2,17	2,42
Máximo	2,22	2,52

«La representación grafica de los Cuartiles es el diagrama de cajas y alambres»

Boxplot (Diagrama de cajas y alambres)

Permiten hacerse una idea acerca de la forma de la distribución de una variable y su dispersión.

Q1= Valor que es superior al **25%** de las observaciones

Q2= Valor que es superior al **50%** de las observaciones

Q3 = Valor que es superior al **75%** de las observaciones

- **Rango Inter Cuartilico (RIC):**
Se define como la diferencia entre el 3er y 1er cuartil

$$\text{RIC} = Q3 - Q1$$



Contiene el 50% de las observaciones

- La caja se construye entre los cuartiles Q1 y Q3 con un ancho arbitrario. Dentro de la caja se marca Q2, con un trazo.
- Los alambres que salen de Q1 y Q3, van hasta el dato más próximo al cerco inferior y superior (sin cruzarlos).

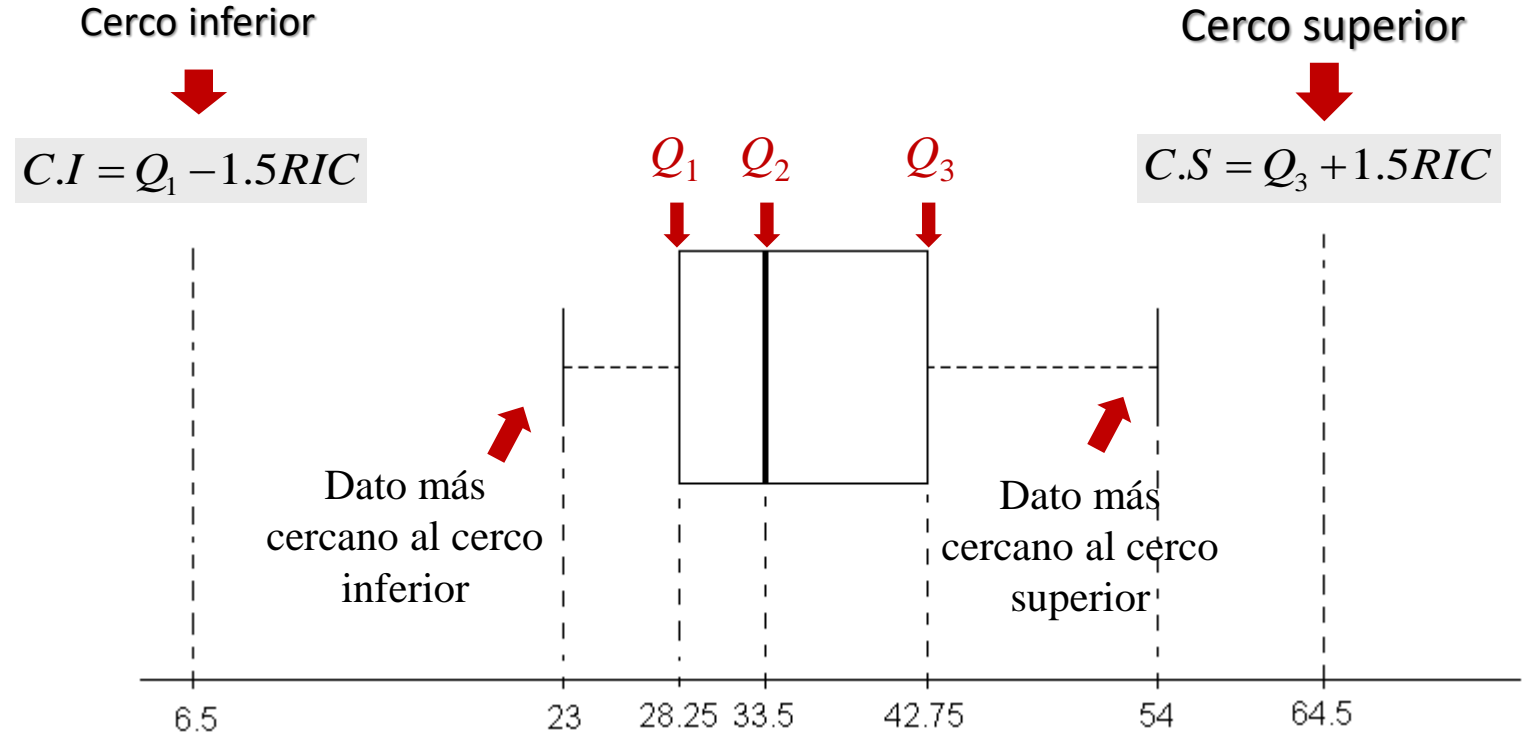
$$CI = Q1 - (1.5 * RIC) \quad (\text{Cerco Inferior})$$

$$CS = Q3 + (1.5 * RIC) \quad (\text{Cerco Superior})$$

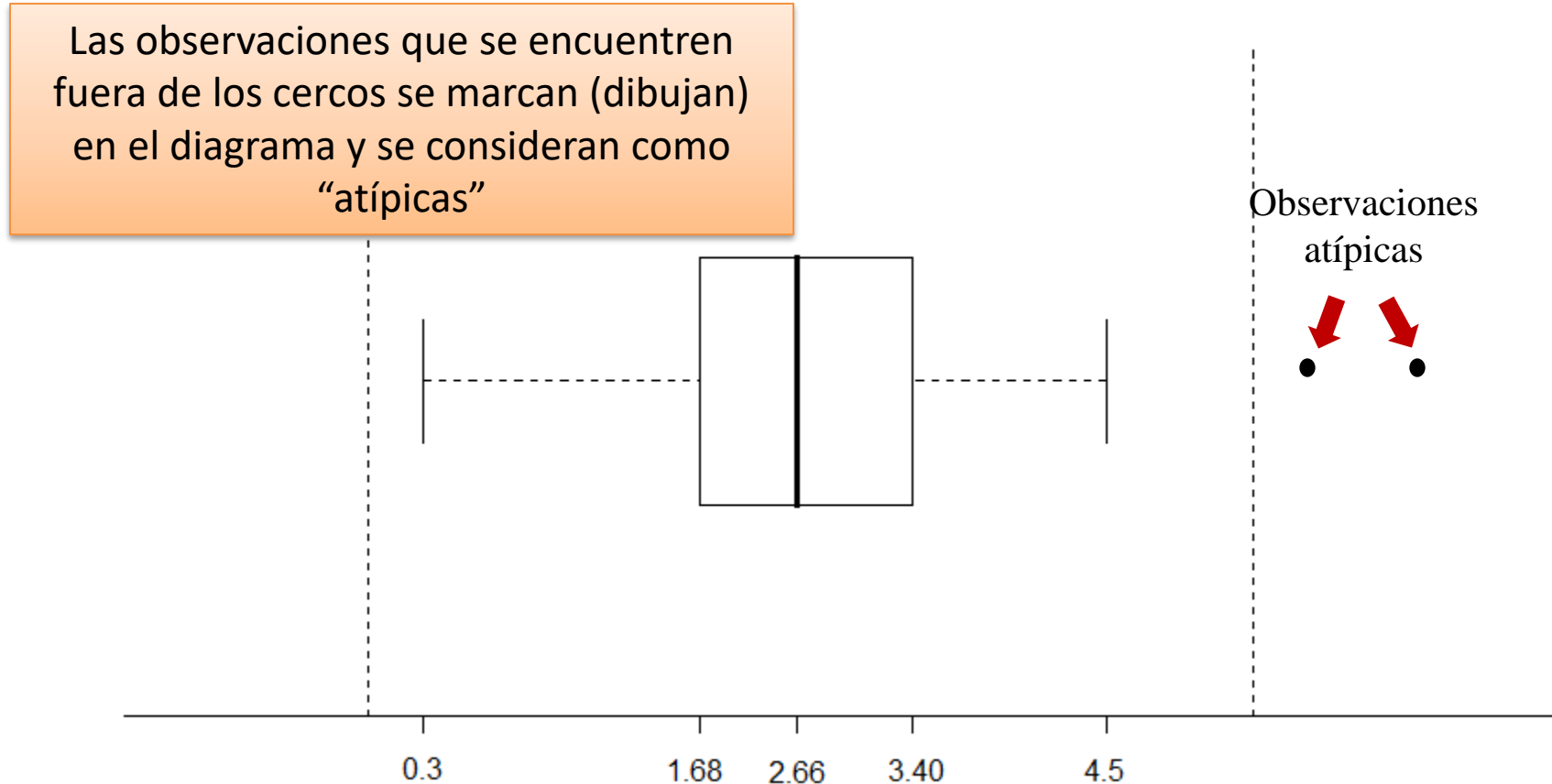
- Los valores que salen de los cercos son marcados sobre el gráfico como puntos **(puntos atípicos)**.

- $\text{Min}(x_i) = 23$
- $\text{Max}(x_i) = 54$
- $Q_1 = 28.25$
- $Q_2 = 33.5$
- $Q_3 = 42.75$
- $\text{RIC} = 14.5$
- $\text{CI} = 28.25 - 1.5(14.5) = 6.5$
- $\text{CS} = 42.75 + 1.5(14.5) = 64.5$

Boxplot (Diagrama de cajas y alambres)



Boxplot (Diagrama de cajas y alambres)

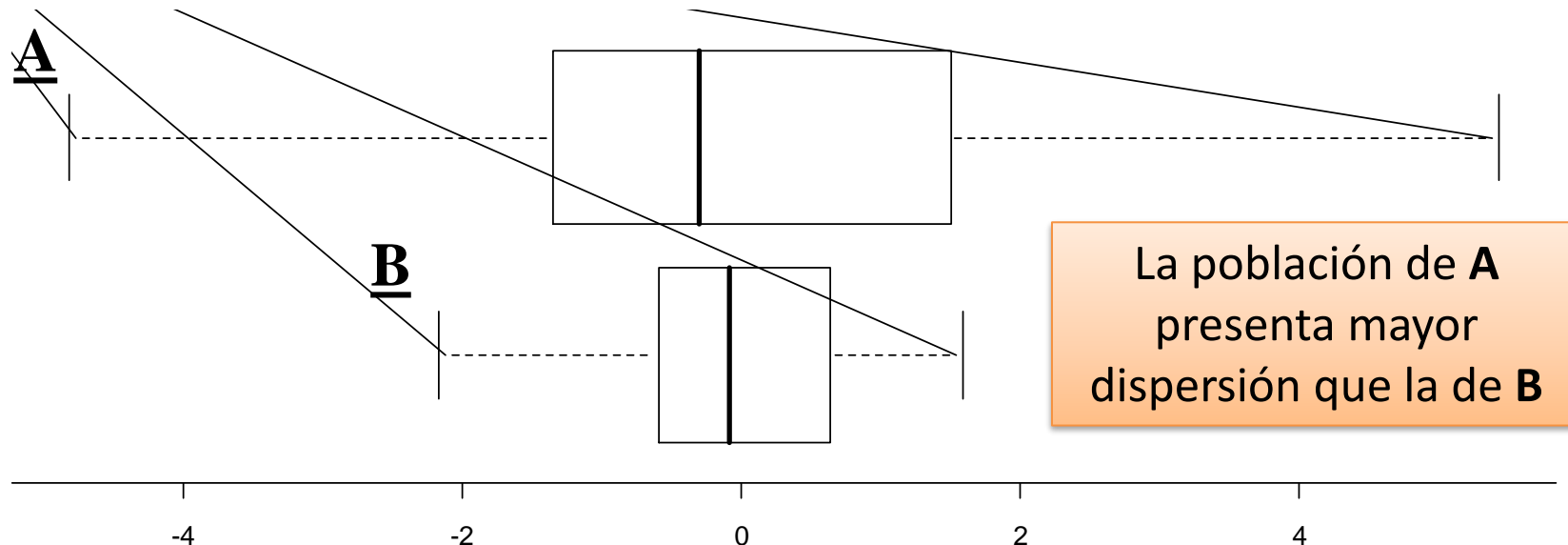


Las observaciones atípicas, son datos que tienen magnitudes “raras” con respecto al conjunto de datos.

Boxplot (Diagrama de cajas y alambres)

Los diagramas de cajas y alambres son útiles, entre otros para los siguientes propósitos:

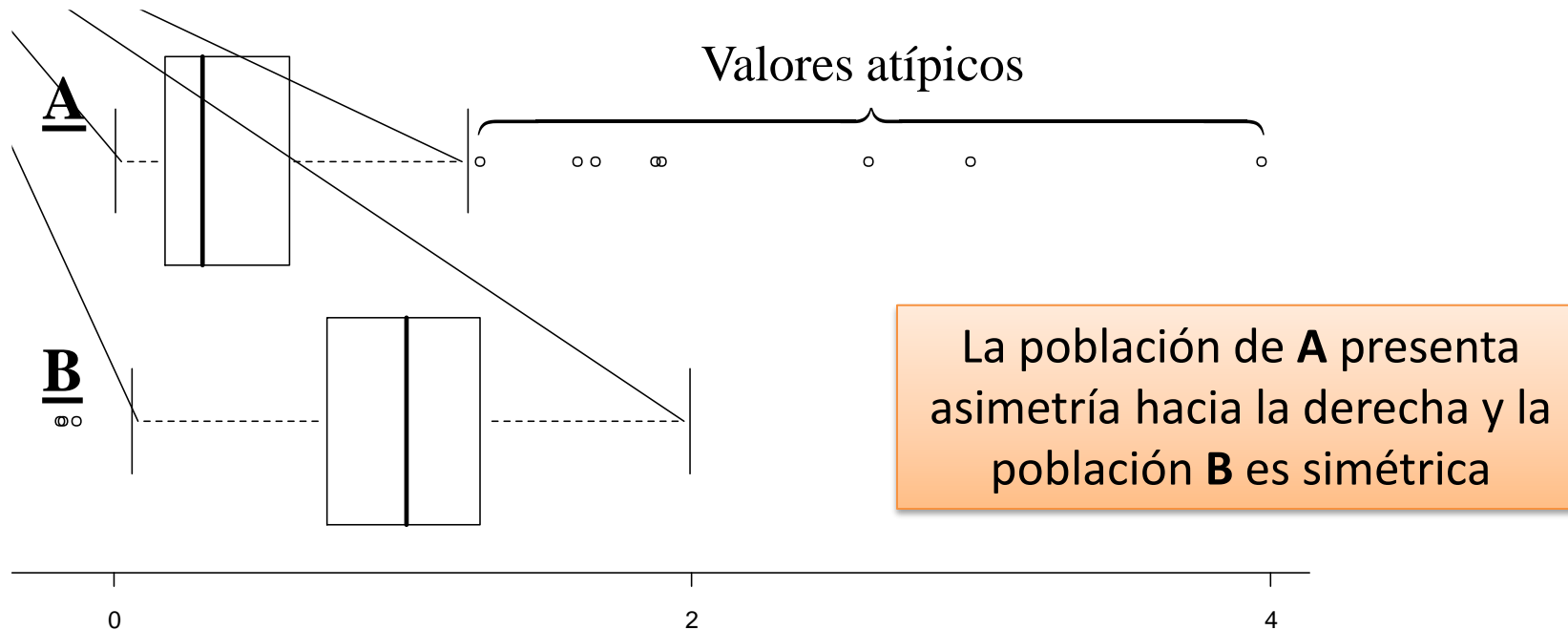
1. Para identificar la localización de los datos alrededor de la mediana.
2. Para hacerse una buena idea de la dispersión de los datos, basándose en la longitud de la caja. Además, se aprecia el rango de los datos



Boxplot (Diagrama de cajas y alambres)

Los diagramas de cajas y alambres son útiles, entre otros para los siguientes propósitos:

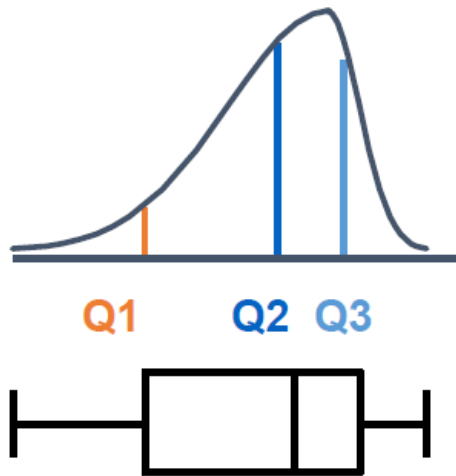
3. Permite observar el grado de asimetría de una distribución, comparando las proporciones de la caja que queda a los lados de la mediana.
4. Útil para identificar posibles valores atípicos (fuera de los cercos)



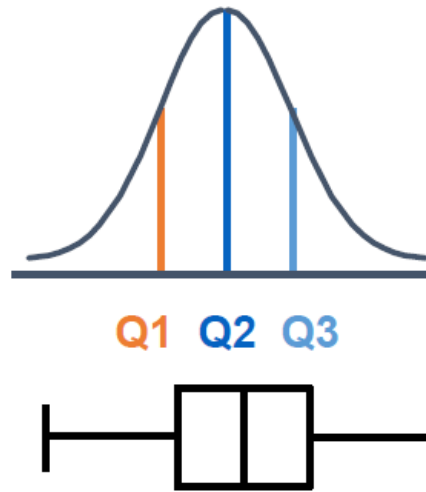
Boxplot (Diagrama de cajas y alambres)

Forma de la distribución y el diagrama de Caja y alambres

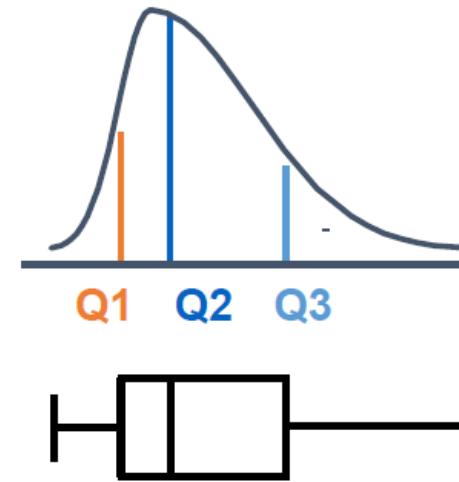
Sesgo a la izquierda ó
sesgo negativo



Simétrica



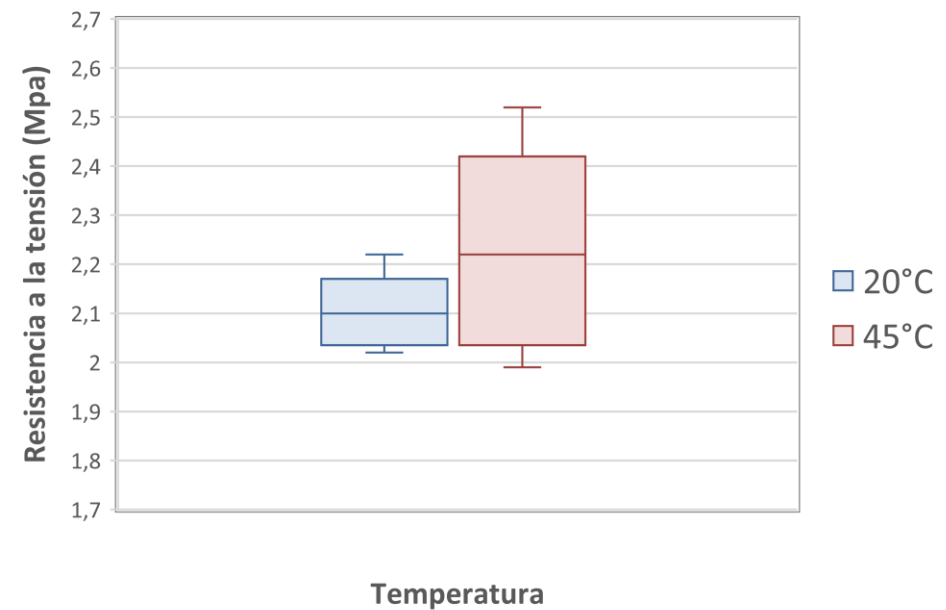
Sesgo a la derecha ó
sesgo positivo



Ejemplo

La resistencia a la tensión del caucho de silicio se considera una función de la temperatura de vulcanizado. Se llevó a cabo un estudio en el que se prepararon muestras de 12 especímenes del caucho utilizando temperaturas de vulcanizado de 20°C y 45°C. Los siguientes datos presentan los valores de resistencia a la tensión en megapascuales.

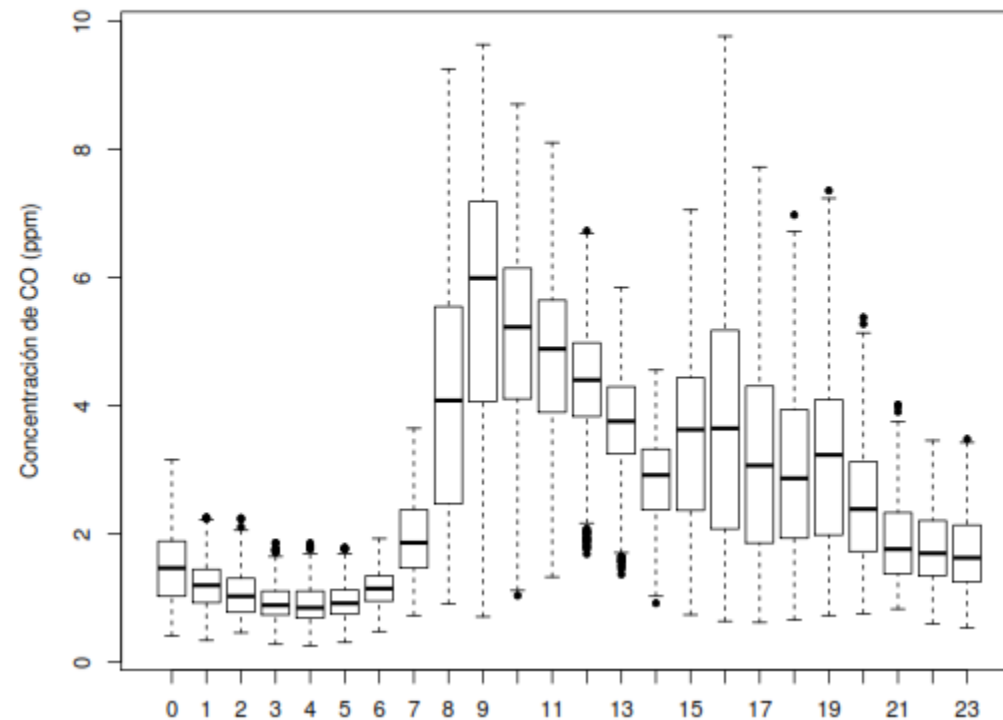
Indicador	20°C	45°C
Mínimo	2,02	1,99
Cuartil 1	2,04	2,04
Cuartil 2	2,10	2,22
Cuartil 3	2,17	2,42
Máximo	2,22	2,52
RIC	0,11	0,38
CI	1,86	1,43
CS	2,38	3,08



Boxplot (Diagrama de cajas y alambres)

En el siguiente gráfico se observa el comportamiento de los niveles de monóxido de carbono (CO) durante un día ordinario (lunes a viernes).

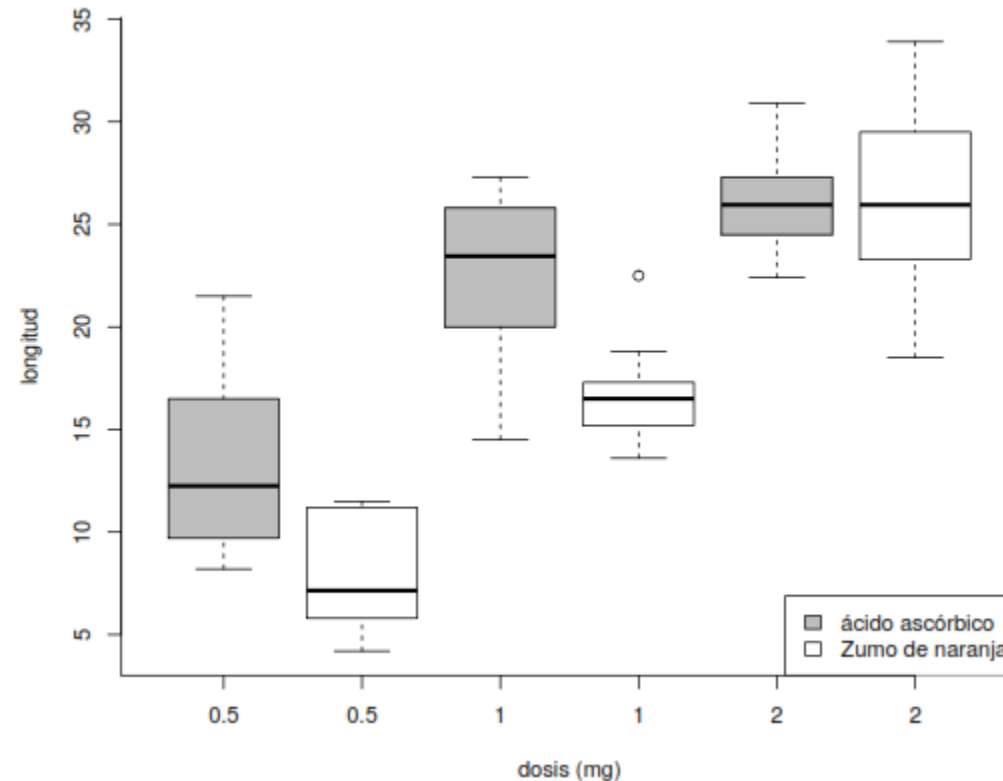
Figura: Diagrama de cajas y alambres de la concentración de CO por hora



Boxplot (Diagrama de cajas y alambres)

En un experimento se observó la longitud de los dientes de conejillos de indias para dos tipos de administración (zumo de naranja o ácido ascórbico) y tres niveles de dosis de vitamina C (0.5, 1 y 2).

Figura: Longitud del diente según dosis y tipo de administración



Ejemplo:

Una entidad encargada del control de contaminación de cierto río lleva registros sobre el oxígeno disuelto (x), expresado en mg/l; éstos se presentan a continuación:

Intervalos	x'_i	n_i	f_i	N_i	F_i
[0.3 , 1.0]	0,65	4	0,08	4	0,08
(1.0 , 1.7]	1,35	6	0,12	10	0,20
(1.7 , 2.4]	2,05	12	0,24	22	0,44
(2.4 , 3.1]	2,75	13	0,26	35	0,70
(3.1 , 3.8]	3,45	9	0,18	44	0,88
(3.8 , 4.5]	4,4	6	0,12	50	1,00
	Total	50	1.0		

A partir de la tabla de frecuencias: **Encuentre Q1, Q2 y Q3.**

$$Q_i = L_{i-1} + \left[(P_i - F(L_{i-1})) * \frac{C_i}{f_i} \right]$$

Ejercicio

Según la revista Chemical Engineering, una propiedad importante de una fibra es su absorción del agua. Se toma una muestra aleatoria de 20 pedazos de fibra de algodón y se mide la absorción de cada uno. Los valores de absorción son los siguientes:

1	2	3	4	5	6	7	8	9	10
18.0	18.71	18.9	19.25	19.3	19.4	19.8	20.2	20.3	20.5
11	12	13	14	15	16	17	18	19	20
20.7	21.1	21.4	21.8	21.8	22.1	22.4	22.9	23.0	23.71

Ejercicio

Según la revista Chemical Engineering, una propiedad importante de una fibra es su absorción del agua. Se toma una muestra aleatoria de 20 pedazos de fibra de algodón y se mide la absorción de cada uno. Los valores de absorción son los siguientes:

a. ¿Cuál es la variable de interés?

R/ absorción del agua

b. Hallar las medidas de centralidad con sus respectivas interpretaciones.

Promedio	20.77	Las fibras de algodón absorben en promedio 20.77
Mediana	20.61	El 50% de las fibras absorben 20.61
Moda	Sin moda	No se identifica moda en las mediciones.

c. Hallar las medidas de dispersión con sus respectivas interpretaciones.

Varianza	2.53	No es interpretable
Desviación	1.59	Las mediciones de absorción de las fibras varían en 1.59 respecto de la absorción promedio.
CV	7.7%	Las mediciones de absorción presentan poca variabilidad.

d. Determinar e interpretar las cinco medidas resumen y realizar el diagrama de cajas y alambres.

Min	18.04	La fibra de algodón presento una absorción mínima de 18.04
Q1	19.40	El 25% de las fibras presentaron una absorción de 19.40
Q2	20.61	El 50% de las fibras presentaron una absorción de 20.61
Q3	21.89	El 75% de las fibras presentaron una absorción de 21.89
Max	23.71	La fibra de algodón presento una absorción maxima de 23.71

