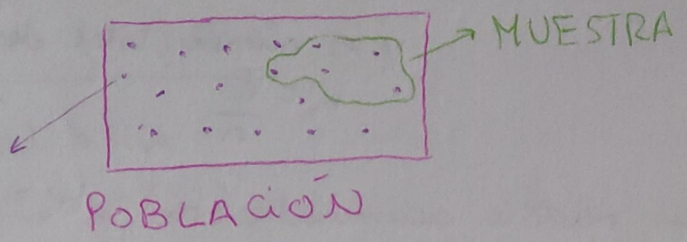
TEMA 1: Estadística descriptiva

Definiciones:

- Población: conjunto homogéneo y bien definido que es el objeto de estudio.
- Individuo (unidad estadística): cada elemento de la población.
- Muestra: subconjunto de la población.
- Muestreo: procedimiento mediante el que se obtienen muestras.



Variables estadísticas: carácter observado en un individuo de la población.

- Variable cualitativa: no es medible → no es numérica (Alto/Bajo; Sí/No; Blanco/Negro ...)
- Variable cuantitativa: es medible, numerable.
  - Discreta: procede de un recuento (número finito)
  - Continua: procede de realizar medidas (dentro de un intervalo puede tomar cualquier valor)

Tablas de frecuencias: estructuración y ordenación de los datos recogidos.

- Variable cualitativa:
  - Modalidad (A_i): categoría (no numérica)
  - Frecuencia absoluta (f_i): nº de veces que se ha observado cada modalidad.

Modalidad (A_i)	Frecuencia absoluta (f_i)
A_1	f_1
A_2	f_2
$\vdots$	$\vdots$
A_n	f_n

Profesional	f_i
Operario	45
Compras	13
Gerente	2

- Variable cuantitativa discreta:
  - Modalidad (x_i): cada valor que puede tomar la variable en ORDEN CRECIENTE.
  - Frecuencia absoluta (f_i): nº de veces que se ha observado cada modalidad.
  - Frecuencia absoluta acumulada (F_i): Suma de las f_i de los valores anteriores.
  - Frecuencia relativa (h_i): Cociente entre la frecuencia absoluta y el número total de observaciones (n):
    $$h_i = \frac{f_i}{n}$$
  - Frecuencia relativa acumulada (H_i): $H_i = \frac{F_i}{n}$

x_i	f_i	F_i	h_i	H_i
x_1	f_1	F_1	h_1	H_1
x_2	f_2	F_2	h_2	H_2
⋮	⋮	⋮	⋮	⋮
x_m	f_m	F_m	h_m	H_m

$n = 60$

$F_m = n$

$H_m = 1$

x_i	f_i	F_i	h_i	H_i
0	13	13	13/60	13/60
1	20	33	20/60	33/60
7	17	50	17/60	50/60
10	10	60 $= n$	10/60	60/60 $= 1$

- Variable cuantitativa continua: agrupar los datos en intervalos (clases), f_i corresponderá al nº de observaciones que caen dentro de él.

  1º- Obtener nº clases (K):
  - $K = \sqrt{n}$; $n = 60 \rightarrow K = \sqrt{60} = 7'6 \simeq 8$
  - Fórmula de Sturges: $K = 1 + 3'322 \log_{10} n$

  $* \; K \in \mathbb{Z} \rightarrow$ REDONDEAR

  2º- Rango de la variable estadística: resta: valor máximo - valor mínimo

  3º- Amplitud de intervalos: cociente: $\frac{\text{rango}}{K}$ (redondear hacia arriba ↑)

  4º- Intervalos:
  - el primero contiene al dato menor
  - el último contiene al dato mayor

  5º- Clases SEMIABIERTAS: $[l_i, l_{i+1})$

  6º- Marca de clase (x_i): punto medio del intervalo: $x_i = \frac{l_i + l_{i+1}}{2}$

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
$[1'57, 1'63)$	$1'60$	12	:	:	:
$[1'63, 1'69)$	$1'66$	14	:	:	:
...	...	...			

$n = 60$, $K = \sqrt{60} \simeq 8$

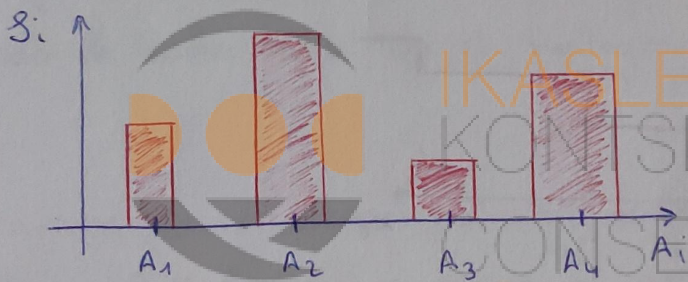
rango: $1'98 - 1'57 = 0'41$ m

amplitud: $\frac{0'41}{8} = 0'05 \simeq 0'06$

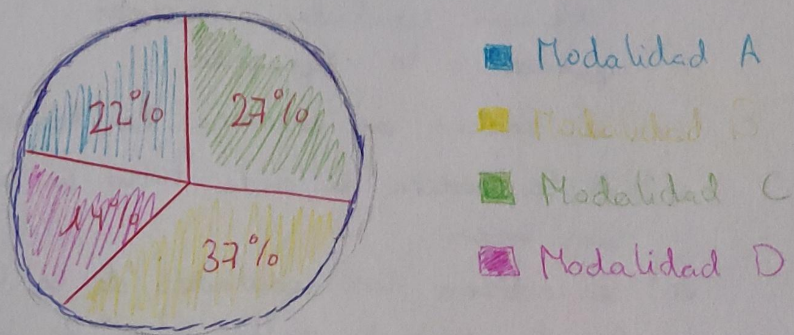
Del mismo modo que en V.C. Discreta.

Representación gráfica:

- Variable cualitativa:
  - Diagrama de barras:
    eje de abscisas: modalidades de la variable
    eje de ordenadas: frecuencias absolutas



  - Diagrama de sectores: se representan sectores cuyo ángulo es proporcional a la frecuencia absoluta de la modalidad: $\alpha_i = 360° \cdot \frac{f_i}{n}$

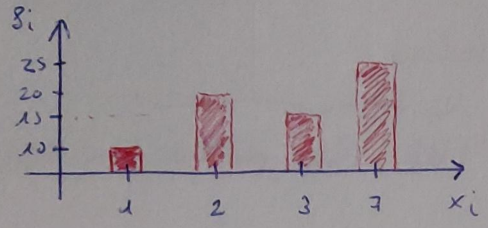


- Variable cuantitativa DISCRETA:

  - Gráficos de barras:

    eje de abscisas: valores de la ~~variable~~ variable de menor a mayor

    eje de ordenadas: frecuencias absolutas o relativas

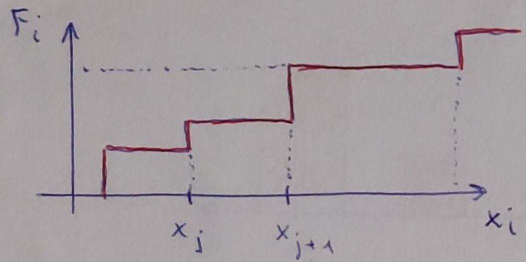


  - Gráfico de frecuencias absolutas acumulada:

    eje de abscisas: valores de la variable de menor a mayor

    eje de ordenadas: frecuencias absolutas acumuladas

    * La altura del eje de ordenadas será el tamaño de la muestra.



  - Gráfico de tallos y hojas:

    i) Se ordenan los valores de menor a mayor.

    ii) Se divide cada valor en Tallo / Hoja.

    iii) Se fija la posición del dígito que se tomará como hoja (décimas, unidades, decenas ...). Los tallos son los que quedan a la izquierda.

    iv) Se anotan en columna los tallos de menor a mayor

    v) A la derecha de cada tallo se anotan de forma creciente sus hojas.

    vi) Se indican las unidades de la hoja, y opcionalmente, la f_i de cada tallo (= nº de hojas).

Tallo	Hojas
2	1 1 2 3 4 4 7 7 7
5	3 3 8

Unidades de las hojas: 1

Presenta una mayor frecuencia absoluta (f_i)

• Variable cuantitativa CONTINUA:

- Histograma:

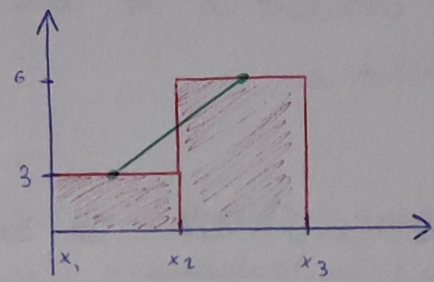
eje de abscisas: límites reales de la clase

eje de ordenadas: rectángulos cuya altura será:

i) Si la amplitud de clases es la misma: f_i

ii) Si es distinta: f_i/d_i = densidad = $\frac{\text{frecuencia absoluta}}{\text{amplitud de las clases}}$

clases	marca	f_i
$[x_1, x_2)$		3
$[x_2, x_3)$		6



- Polígono de frecuencias absolutas: se obtiene uniendo las marcas de cada clase en el histograma

- Histograma de frecuencias absolutas acumuladas y polígono: cada rectángulo será igual o mayor al anterior y el polígono nunca será descendente.

eje de abscisas: límites reales de la clase.

eje de ordenadas: F_i (frecuencia absoluta acumulada)

Parámetros y estadísticos:

- Parámetro: valores característico que se obtienen al realizar un análisis estadístico de TODA la POBLACIÓN cuando la variable es cuantitativa.

- Estadístico: valores característicos que se obtienen al realizar un análisis estadístico de una MUESTRA cuando la variable es cuantitativa.

Estadísticos descriptivos.

Se emplean para sintetizar y describir las características generales del conjunto de datos (observaciones de un fenómeno) que conforman la muestra.
Se diferencian cuatro grupos: medidas de TENDENCIA CENTRAL, medidas de DISPERSIÓN, medidas de POSICIÓN, medidas de FORMA.

- Notación:

  n: tamaño de la muestra

  l_i: límite inferior de la clase que incluye al estadístico.

  d_i: amplitud de la clase que incluye al estadístico.

  f_i: frecuencia absoluta de la clase que incluye al estadístico.

  f_{i-1} / f_{i+1}: frecuencia absoluta de la clase anterior/posterior a la que incluye al estadístico.

  F_{i-1}: frecuencia absoluta acumulada de la clase anterior a la que incluye al estadístico.

• Medidas de TENDENCIA CENTRAL: son medidas de posición que tratan de establecer un valor que pueda considerarse el centro de los datos en algún sentido.

- Moda (Mo): aquel valor o valores de la variable cuya frecuencia absoluta es máxima.

• Variables discretas: es el valor x_i que tiene una mayor frecuencia absoluta f_i.

• Variables continuas:

1) Determinar la clase modal; aquella que tiene una frecuencia absoluta o densidad máxima.

2) La moda será el punto dentro de dicha clase al que le correspondería la frecuencia absoluta máxima.
Se obtiene mediante semejanza de triángulos, simplificando:

→ Si la amplitud de las clases es la misma: $\boxed{Mo = l_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot d_i}$

→ Si no es la misma: $\boxed{Mo = l_i + \frac{\frac{f_i}{d_i} - \frac{f_{i-1}}{d_{i-1}}}{\left(\frac{f_i}{d_i} - \frac{f_{i-1}}{d_{i-1}}\right) + \left(\frac{f_i}{d_i} - \frac{f_{i+1}}{d_{i+1}}\right)} \cdot d_i}$ d_i: amplitud de las clases.

- Mediana (Me): aquel valor que deja el mismo número de datos a su izquierda y a su derecha, cuando los valores de la variable están ordenados de forma creciente o decreciente.

• Variables discretas:

nº de datos IMPAR: $\boxed{me = x_i \,/\, F_{i-1} < \frac{n}{2} \;\wedge\; F_i > \frac{n}{2}}$

nº de datos PAR: $\boxed{Me = \frac{x_i + x_{i+1}}{2}}$ x_i: valor posición $\frac{n}{2}$; x_{i+1}: valor posición $\frac{n}{2}+1$

• Variables continuas: se determina la clase que contenga la mediana; es decir, que cumpla: $F_{i-1} \leq \frac{n}{2}$ y $F_i > \frac{n}{2}$

Por semejanza de triángulos, utilizando el histograma y simplificando:

$$\boxed{Me = l_i + \frac{n/2 - F_{i-1}}{f_i} \cdot d_i}$$

- Media aritmética ($\bar{x}$): representa el centro de gravedad de la distribución:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n}}$$

k: nº de modalidades o clases

Cuando los datos están agrupados en intervalos (V. C. Continua), x_i representa las marcas de clase.

- Media geométrica ($\bar{x}_G$):

$$\boxed{\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}}}$$

- Media armónica ($\bar{x}_H$):

$$\boxed{\bar{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \left[\sum_{i=1}^{k} \frac{f_i}{x_i}\right]^{-1}}$$

- Media cuadrática ($\bar{x}_Q$):

$$\boxed{\bar{x}_Q = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 = \frac{1}{n} \sum_{i=1}^{k} f_i \cdot x_i^2}$$

• Medidas de DISPERSIÓN: Son medidas que nos indican cuán dispersos están los datos en la muestra.

– Rango (R): la resta entre el máximo valor y el mínimo valor del conjunto de datos.

$$\boxed{R = \text{máx}(x_i) - \text{mín}(x_i)}$$

– Rango intercuartílico (RIC): la resta entre el tercer cuartil y el primer cuartil.

$$\boxed{RIC = Q_3 - Q_1}$$

– Varianza (s^2): es la medida aritmética de los cuadrados de las diferencias a la media. Siempre es POSITIVA.

$$\boxed{s^2 = Var(x) = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n} = \frac{\sum_{i=1}^{k} x_i^2 \cdot f_i}{n} - \bar{x}^2}$$

– Desviación típica (s): raíz cuadrada positiva de la varianza. Siempre es POSITIVA.

$$\boxed{s = \sqrt{s^2}}$$

– Cuasivarianza (S^2): producto de la varianza por $\frac{n}{n-1}$

$$\boxed{S^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n-1}}$$

– Cuasidesviación típica (S): raíz cuadrada positiva de la cuasivarianza

$$\boxed{S = \sqrt{S^2}}$$

* V.C. Continua; x_i representa la marca de clases.

- Coeficiente de variación (CV): coeficiente adimensional que se emplea para comparar distintas variables estadísticas. Compara la dispersión de cada variable estadística respecto a su media aritmética:

$$\boxed{CV = \frac{s}{|\bar{x}|}}$$

s: desviación típica
$|\bar{x}|$: valor absoluto de la media aritmética.

• Medidas de POSICIÓN: Cuantiles; valores que dividen la distribución en partes de igual frecuencia.

En general, un cuantil de orden α, deja a su izquierda un valor α de los datos:

$$\boxed{C_\alpha = l_i + \frac{\alpha n - F_{i-1}}{f_i} \cdot d_i}$$

Los casos particulares son los percentiles, deciles y cuartiles:

P_k: Percentil (100 partes iguales; 99 percentiles) 100 *

D_k: Decil (10 partes iguales; 9 deciles) 10 *

Q_k: Cuartil (4 partes iguales; 3 cuartiles) 4 *

$D_1 = P_{10}$; $D_2 = P_{20}$ ... $Q_1 = P_{25}$; $Q_2 = P_{50}$; $Q_3 = P_{75}$

• Variables discretas:

X_k es el valor x_i que cumple $F_{i-1} < \frac{k \cdot n}{*}$ y $F_i > \frac{k \cdot n}{*}$

Si se cumple: $F_i = \frac{k \cdot n}{*}$, entonces $X_k = \frac{x_i + x_{i+1}}{2}$

• Variables continuas:

Se determina la clase $[l_i, l_{i+1})$ que lo contiene, es decir, que cumpla $F_{i-1} \leq \frac{k \cdot n}{*}$ y $F_i > \frac{k \cdot n}{*}$;

y se aplica: $\boxed{X_k = l_i + \frac{\frac{k \cdot n}{*} - F_{i-1}}{f_i} \cdot d_i}$

- Medidas de FORMA: sirven para definir los momentos de una variable, que son los datos que obtienen al analizarla.

  - Momento respecto al origen (a_r): el momento de orden r respecto al origen se define como:

$$\boxed{a_r = \frac{\sum_{i=1}^{k} x_i^r \cdot f_i}{n}}$$

Por tanto: $a_1 = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n} = \bar{x}$

  - Momentos centrales (m_r): el momento central de orden r se define como:

$$\boxed{m_r = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^r \cdot f_i}{n}}$$

Por tanto: $m_2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot f_i}{n} = S^2$

  - Coeficiente de asimetría de Fisher:

$$\boxed{g_1 = \frac{m_3}{S^3} = \frac{1}{n \cdot S^3} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^3}$$

$\Rightarrow$
- $g_1 > 0$ Sesgo positivo, cola a la derecha
- $g_1 = 0$ Distribución simétrica
- $g_1 < 0$ Sesgo negativo, cola a la izquierda.

Indica si los resultados de los datos son o no simétricos.

  - Coeficiente de asimetría de Pearson: Solo para distribuciones UNIMODALES.

$$\boxed{P = \frac{\bar{x} - Mo}{S} \approx \frac{3(\bar{x} - Me)}{S}}$$

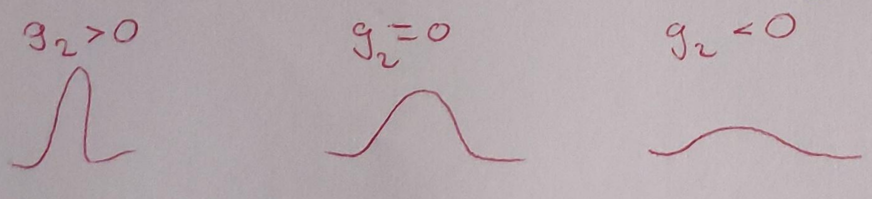
$\Rightarrow$
- $P > 0$ Asimetría o cola a la derecha
- $P = 0$ Distribución simétrica
- $P < 0$ Asimetría o cola a la izquierda.

  - Coeficiente de Curtosis: Solo para distribuciones UNIMODALES.

Compara la distribución con una campana de Gauss.

$$\boxed{g_2 = \frac{m_4}{S^4} - 3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n \cdot S^4} - 3}$$

$\Rightarrow$
- $g_2 > 0$ LEPTOCÚRTICA
- $g_2 = 0$ MESOCÚRTICA
- $g_2 < 0$ PLATICÚRTICA



Cambio de escala y de origen: Son transformaciones que se realizan a los datos recogidos en la muestra. Las más usuales son las "lineales"; es decir, el cambio de escala y el cambio de origen.

- Cambio de escala: los valores de la nueva variable se obtienen multiplicando los datos primitivos por una constante K.

$$\boxed{y = K \cdot x}$$

- Medidas de Tendencia Central:
  - Media aritmética $(\bar{y})$: $\bar{y} = K \cdot \bar{x}$
  - Moda (Mo): $Mo(y) = K \cdot Mo(x)$
  - Mediana (Me): $Me(y) = K \cdot Me(x)$

- Medidas de Dispersión:
  - Varianza (s^2): $s^2(y) = K^2 \cdot s^2(x)$ } siempre positivos
  - D. Típica (s): $s(y) = |K| \cdot s(x)$

- Cambio de origen: los valores de la nueva variable se obtienen sumando una constante a a los datos primitivos.

$$\boxed{y = x + a}$$

- Medidas de tendencia central:
  - Media aritmética $(\bar{y})$: $\bar{y} = \bar{x} + a$
  - Moda (Mo): $Mo(y) = Mo(x) + a$
  - Mediana (Me): $Me(y) = Me(x) + a$

- Medidas de Dispersión: permanecen invariantes.

SE PUEDEN REALIZAR AMBOS CAMBIOS SIMULTÁNEAMENTE. (Cambio de escala y origen)

Variables tipificadas: Se emplean para realizar una comparación relativa entre variables de diferentes elementos.

Son ADIMENSIONALES.

- Cómo se obtienen:

1) Cambio de origen: $X - \bar{x}$ Conseguimos una variable centrada, dado que la media aritmética de "$x-\bar{x}$" es 0.

2) Cambio de escala: $\frac{x-\bar{x}}{s(x)}$ La nueva variable se denomina "Variable tipificada" (estandarizada). Su media aritmética es 0 y su desviación típica es 1.

$$\boxed{Z = \frac{x-\bar{x}}{s(x)}} \qquad \bar{Z} = 0 \qquad s(Z) = 1$$

- Ejemplo: Sea la variable estadística X que designa la longitud de los recién nacidos en un nuevo hospital siendo su media aritmética 50 cm y su desviación típica 1'6 cm. Por otro lado, se considera la variable estadística Y que designa la altura de varones de entre 18 y 28 años cuya media aritmética es de 175 cm y desviación típica 6'7 cm.

Se seleccionan un recién nacido de 54 cm y un varón de 190 cm. ¿Cuál es relativamente más alto (respecto a su grupo)?

$x = 54$ cm
$\bar{x} = 50$ cm
$s(x) = 1'6$ cm

$$Z_x = \frac{x-\bar{x}}{s(x)} = \frac{54-50}{1'6} = \frac{4}{1'6} = 2'5$$

$y = 190$ cm
$\bar{y} = 175$ cm
$s(y) = 6'7$ cm

$$Z_y = \frac{y-\bar{y}}{s(y)} = \frac{190-175}{6'7} = \frac{15}{6'7} = 2'24$$

Un recién nacido (x) es relativamente más alto que un varón (y):

$Z_x > Z_y$

Ley empírica:

cuando la distribución es unimodal y campaniforme (más o menos Campana de Gauss), según la ley empírica:

- $68'27\%$ de los datos se encuentra en el intervalo:

$$(\bar{x}-s\ ,\ \bar{x}+s)$$

- $95'45\%$ de los datos se encuentra en el intervalo:

$$(\bar{x}-2s\ ,\ \bar{x}+2s)$$

- $99'73\%$ de los datos se encuentra en el intervalo:

$$(\bar{x}-3s\ ,\ \bar{x}+3s)$$

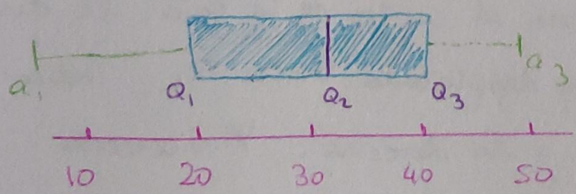
Diagrama de caja y valores atípicos:

es una representación gráfica para distribuciones estadísticas unidimensionales, creado por Tukey en 1977.

"La caja" comprende el 50% de los valores centrales de los datos, extendiéndose entre el primer y tercer cuartil.

La línea de dentro de la caja corresponde a la MEDIANA.

"Los bigotes" se extienden desde el menor al mayor de los valores observados y considerados normales (no atípicos).



- Pasos a seguir:

1) Calcular los cuartiles y el rango intercuartílico:

Q_1 ; $Q_2 = Me$; Q_3 ; $RIC = Q_3 - Q_1$

2) Límites interiores, m_1 y m_3:

$m_1 = Q_1 - 1'5 RIC$; $m_3 = Q_3 + 1'5 RIC$

3) Límites de los bigotes, a_1 y a_3:

a_1: primer valor del conjunto de datos que sea igual o superior a m_1.

a_3: primer valor del conjunto de datos que sea igual o inferior a m_3.

4) Límites exteriores, M_1 y M_3:

$M_1 = Q_1 - 3 RIC$; $M_3 = Q_3 + 3 RIC$

5) Al valor de la muestra que se encuentre en los intervalos (M_1, m_1) ó (m_3, M_3) se le denomina "valor ATÍPICO" (outlier) y se denota "o".

Si se encuentra fuera de las barreras exteriores, "valor EXTREMADAMENTE ATÍPICO", denotándose "*".

- Observaciones:

Si el ancho de la caja no es muy amplio, la dispersión de los datos es baja dado que el RIC es pequeño.

Si la mediana coincide con el centro de la caja, la distribución es simétrica.

Si la mediana está muy desplazada hacia Q_1, indica asimetría positiva; es decir, hay una cola a la derecha. Y viceversa.

ejemplo: construir gráfica de tallos y hojas

58	62	75	67	79	71	65	72	70
78	81	89	91	61	100	95	85	89

1. Ordenar los valores de menor a mayor:

x_i	f_i
58	1
61	1
62	1
65	1
67	1
70	1
71	1
72	1
75	1
78	1
79	1
81	1
85	1
89	2
91	1
95	1
100	1

2. Dividir cada valor en → Tallo / → Hoja

3. Se fijan las unidades de las hojas: 1

4. Se anotan en columna los tallos de menor a mayor y hojas:

Tallos	Hojas
5	8
6	1 2 5 7
7	0 1 2 5 8 9
8	1 5 9 9
9	1 5
10	0

Presenta una mayor frecuencia absoluta.

ejemplo: histograma y polígono de frecuencias absolutas

$n = 27$; 2.1 3.3 4.4 3.0 4.0 5.0 2.7 2.6 4.8

L/100 km 4.7 2.8 4.8 3.9 2.3 3.8 2.8 3.0 3.7

3.3 4.4 3.1 4.0 3.7 2.5 2.7 5.1 4.7

1. Definir la variable estadística: V. Cuantitativa Continua

2. Tabla de frecuencias:

   a) obtener las clases: $K = \sqrt{n}$ $K = \sqrt{27}$ $K = 3\sqrt{3} \simeq 5'2 \simeq 5$

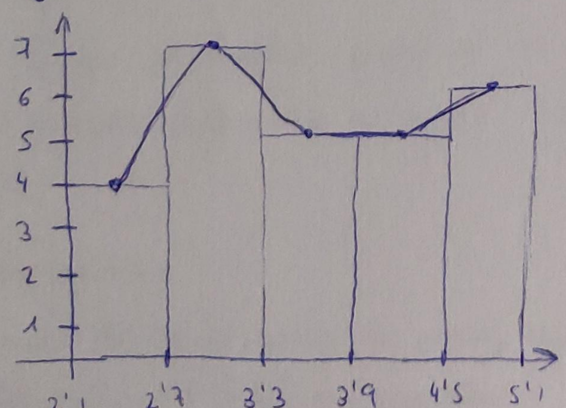
   b) Rango de la variable estadística: máx − mín : $5'1 - 2'1 = 3$

   c) Amplitud de los intervalos: $\frac{\text{rango}}{n} = \frac{3}{5} = 0'6$

$[l_i, l_{i+1})$	x_i (marca)	f_i	F_i	h_i	H_i
[2'1, 2'7)	2'4	4	4	4/27	4/27
[2'7, 3'3)	3'0	7	11	7/27	11/27
[3'3, 3'9)	3'6	5	16	5/27	16/27
[3'9, 4'5)	4'2	5	21	5/27	21/27
[4'5, 5'1]	4'8	6	27	6/27	27/27

3. Histograma:

4. Polígono



ejemplo: diagrama de caja

X: edades ; $n = 40$

x_i	f_i	F_i
15	2	2
16	2	4
17	1	5
18	2	7
20	2	9
23	2	11
24	3	14
25	4	18
26	2	20
27	3	23
28	1	24
29	3	27
38	5	32
45	1	33
47	1	34
50	1	35
54	1	36
59	1	37
63	1	38
66	1	39
85	1	40

1) Q_K: $\begin{cases} \frac{k\cdot n}{4} > F_{i-1} \\ \frac{k\cdot n}{4} < F_i \end{cases}$

Q_1: $\frac{1\cdot 40}{4} = \frac{25\cdot 40}{100} = 10$

Q_2: $Me = \frac{50\cdot 40}{100} = 20$

Q_3: $\frac{3\cdot 40}{4} = \frac{75\cdot 40}{100} = 30$

$RIC = Q_3 - Q_1 = 38 - 23 = 15$

Q_1: $F_{i-1} < 10$; $F_i > 10$; $Q_1 = 23$

2) Límites interiores:

$m_1 = Q_1 - 1'5\, RIC = 23 - 1'5 \cdot 15 = 0'5$ años

$m_3 = Q_3 + 1'5\, RIC = 38 + 1'5 \cdot 15 = 60'5$ años

3) Bigotes:

$a_1 = 15$ (primer valor $>= m_1$) años

$a_3 = 59$ (primer valor $<= m_3$) años

Q_2: $F_i = 20$; $Q_2 = \frac{26+27}{2} = 26'5$

4) Límites exteriores:

$M_1 = Q_1 - 3RIC = 23 - 3\cdot 15 = -22$ años

$M_3 = Q_3 + 3RIC = 38 + 3\cdot 15 = 83$ años

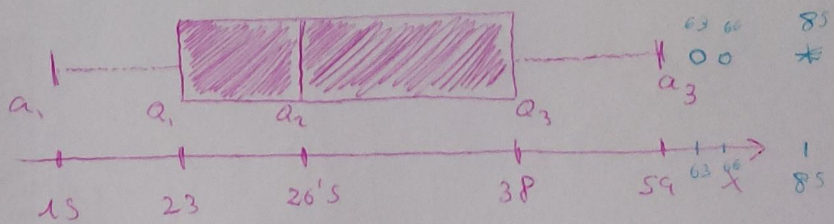
Q_3: $F_{i-1} < 30$; $F_i > 30$; $Q_3 = 38$

5) Valores atípicos:

$x = 63$; $x = 66$

extremadamente atípicos:

$x = 85$



Ejercicio 1:

Nº defectos	f_i (frecuencia absoluta)	F_i	h_i	H_i
0	600	600	$600/1000$	$600/1000$
1	310	910	$310/1000$	$910/1000$
2	75	985	$75/1000$	$985/1000$
3	13	998	$13/1000$	$998/1000$
4	2	1000	$2/1000$	$1000/1000$
TOTAL	100			

Media aritmética ($\bar{x}$): $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$

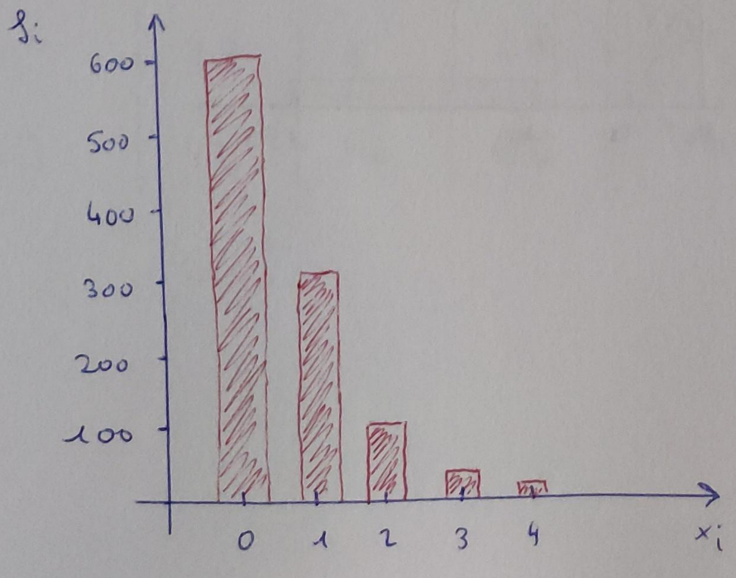
$$\bar{x} = \frac{0 \cdot 600 + 1 \cdot 310 + 2 \cdot 75 + 3 \cdot 13 + 4 \cdot 2}{1000} = 0'507 \text{ defectos / pieza.}$$

Varianza (s^2): $s^2 = \dfrac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot f_i}{n} = \dfrac{\sum_{i=1}^{k} x_i^2 \cdot f_i}{n} - \bar{x}^2$

$$s^2 = \frac{0^2 \cdot 600 + 1^2 \cdot 310 + 2^2 \cdot 75 + 3^2 \cdot 13 + 4^2 \cdot 2}{1000} - 0'507^2 = 0'502 \left(\frac{\text{defectos}}{\text{pieza}}\right)^2$$

Desviación típica (s): $s = \sqrt{s^2}$ $\quad s = \sqrt{0'502} = 0'70 \dfrac{\text{defectos}}{\text{pieza}}$

Gráfico de barras:



Ejercicio 4:

Tabla de frecuencias: Variable: Km/día (V. C. Continua)

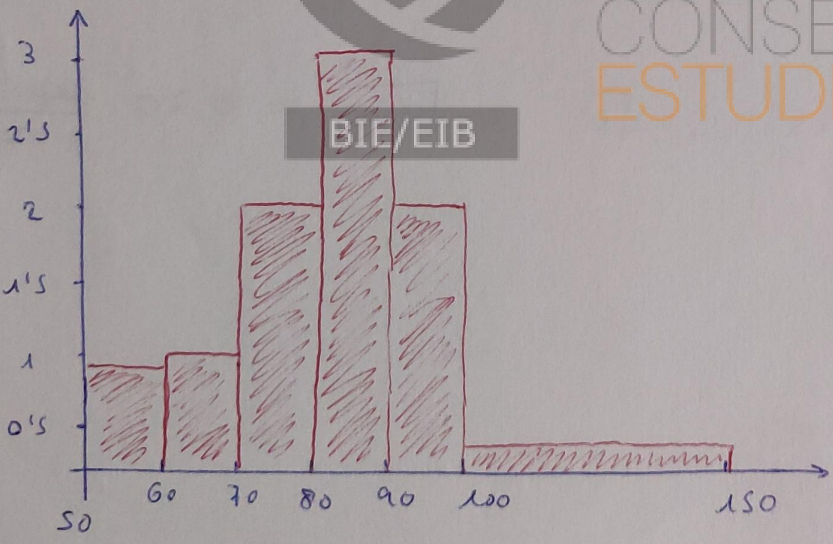
	Clases	x_i	f_i	F_i	h_i	H_i
1.	[50, 60)	55	8	8	8/100	8/100
2.	[60, 70)	65	10	18	10/100	18/100
3.	[70, 80)	75	20	38	20/100	38/100
4.	[80, 90)	85	30	68	30/100	68/100
5.	[90, 100)	95	20	88	20/100	88/100
6.	[100, 150)	125	12	100	12/100	100/100

Histograma: la amplitud de las clases es diferente (d_i), por lo que emplearemos la densidad.

$f_1/d_1 = 8/10 = 0'8$ $f_4/d_4 = 30/10 = 3$

$f_2/d_2 = 10/10 = 1$ $f_5/d_5 = 20/10 = 2$

$f_3/d_3 = 20/10 = 2$ $f_6/d_6 = 12/50 = 0'24$



② $n = 40$

33	21	32	44	35	22	40	36	22	37
20	37	42	31	23	44	32	30	44	44
42	35	40	36	32	31	37	43	24	46
25	30	26	35	33	41	25	44	36	27

a) variable estadística: torsiones necesarios para cortar barras de aleación foriada.

Variable cuantitativa Continua : tabla de frecuencias

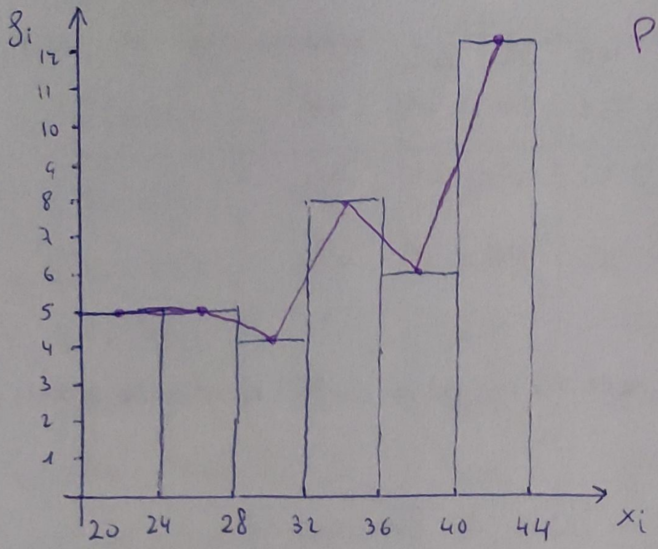
1.- Obtener nº clases (k) : $k = \sqrt{n} \rightarrow k = \sqrt{40} = 2\sqrt{10} = 6'32 \simeq 6$ clases

2.- Rango de la variable estadística: $V_{máx} - V_{mín}$: $44 - 20 = 24$

3.- Amplitud de intervalos: $\dfrac{\text{rango}}{k} = \dfrac{24}{6} = 4$

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
[20, 24)	22	5	5	5/40	5/40
[24, 28)	26	5	10	5/40	10/40
[28, 32)	30	4	14	4/40	14/40
[32, 36)	34	8	22	8/40	22/40
[36, 40)	38	6	28	6/40	28/40
[40, 44]	42	12	40	12/40	40/40

b) Histograma:



Polígono de frecuencias absolutas

c) Media aritmética: $\bar{x} = \dfrac{\sum_{i=1}^{k} x_i \cdot f_i}{n}$; x_i = marca de clase

$$\bar{x} = \frac{22 \cdot 5 + 26 \cdot 5 + 30 \cdot 4 + 34 \cdot 8 + 38 \cdot 6 + 42 \cdot 12}{40} = 34'1 \text{ torsiones}$$

Moda: 1. Clase modal (f_i máxima) $\rightarrow [40, 44]$

$$Mo = l_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot d_i$$ amplitud de la clase

$$Mo = 40 + \frac{12 - 6}{12-6 + 12-0} \cdot 4 = 40 + \frac{6}{18} \cdot 4 = 41'\overline{33} \text{ torsiones}$$

Mediana: 1. Clase mediana $(F_{i-1} \leq \frac{n}{2}) \wedge (F_i > \frac{n}{2}) \rightarrow [32, 36)$

$$Me = l_i + \frac{n/2 - F_{i-1}}{f_i} \cdot d_i$$

$$Me = 32 + \frac{20 - 14}{8} \cdot 4 = 32 + \frac{6}{8} \cdot 4 = 32 + 3 = 35 \text{ torsiones}$$

Varianza: $S^2 = \dfrac{\sum_{i=1}^{k} x_i^2 \cdot f_i}{n} - (\bar{x})^2$

$$S^2 = \frac{22^2 \cdot 5 + 26^2 \cdot 5 + 30^2 \cdot 4 + 34^2 \cdot 8 + 38^2 \cdot 6 + 42^2 \cdot 12}{40} - (34'1)^2 = 49'19 \text{ torsiones}^2$$

Desviación típica: $S = \sqrt{S^2}$

$$S = \sqrt{49'19} = 7'0136 \text{ torsiones.}$$

③ 51 55 42 53 46 60 29 56 20 52 51

33 61 57 55 59 38 56 41 47 68 24

67 52 64 69 43 47 42 65 96 21 48

47 25 82 37 66 12 77 56 92 28 45

63 28 45 63 28 52 60 51 61 62 52

97 73 45 69 67 29 75 63 30 17 65

68 74 16 83 47 16

a) Tabla de frecuencias

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
[10'5, 21'5)	16	6	6	6/72	6/72
[21'5, 32'5)	27	8	14	8/72	14/72
[32'5, 43'5)	38	7	21	7/72	21/72
[43'5, 54'5)	49	17	38	17/72	38/72
[54'5, 65'5)	60	18	56	18/72	56/72
[65'5, 76'5)	71	10	66	10/72	66/72
[76'5, 87'5)	82	3	69	3/72	69/72
[87'5, 98'5)	93	3	72	3/72	72/72

b) Histograma:

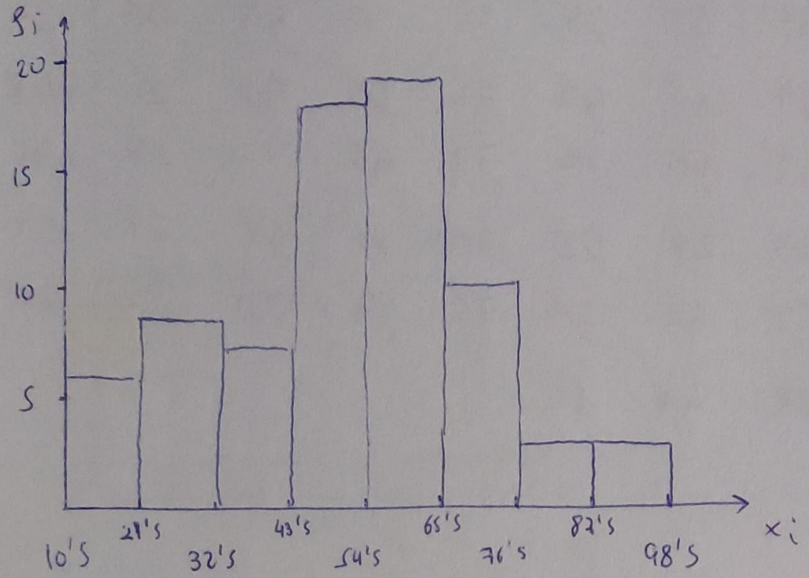


Gráfico de tallos y hojas: unidades de las hojas . 1

tallo	hojas
1	2 6 6 7
2	0 1 4 5 8 8 9 9
3	0 3 7 8
4	1 2 2 3 5 5 5 6 7 7 7 7 8
	36 37
5	1 1 1 2 2 2 2 3 5 5 6 6 6 7 9
6	0 0 0 1 1 2 3 3 3 4 5 7 7 8 8 9 9 9
7	3 4 5 7
8	2 3
9	6 7 7

c) Agrupados:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n} = \frac{16\cdot 6 + 27\cdot 8 + 38\cdot 7 + 49\cdot 17 + 60\cdot 18 + 71\cdot 10 + 82\cdot 3 + 93\cdot 3}{72}$$

$\bar{x} = 51'75$.

Me: 1.- Clase mediana : $F_{i-1} \leq \frac{n}{2}$ $\wedge$ $F_i > \frac{n}{2}$: $[43'5, 54'5)$

$$Me = l_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot d_i = 43'5 + \frac{36 - 21}{17} \cdot 11 = 53'2059$$

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 \cdot f_i}{n} - \bar{x}^2 = 372'2431$$

$$S = \sqrt{S^2} = \sqrt{372'2431} = 19'2936$$

Sin agrupar:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12 + 16 + 16 + 17 + 20 + 21 + 24 + 25 + 28\cdot 2 + 29\cdot 2 \cdots}{72}$$

$$\bar{x} = \frac{3737}{72} = 51'9027$$

Me: nº datos PAR (72) $\rightarrow Me = \frac{x_i + x_{i+1}}{2}$; $x_i = \frac{n}{2}$; $x_{i+1} = \frac{n}{2} + 1$

$x_i = 52$

$x_{i+1} = 52$

$$Me = \frac{52 + 52}{2} = 52$$

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 \cdot f_i}{n} - \bar{x}^2 = \frac{12^2 + 16^2 + 16^2 + 17^2 + 20^2 + \cdots}{72} - 50'9027$$

$S^2 = 370'7822$

$$S = \sqrt{S^2} = \sqrt{370'7822} = 19'2557$$