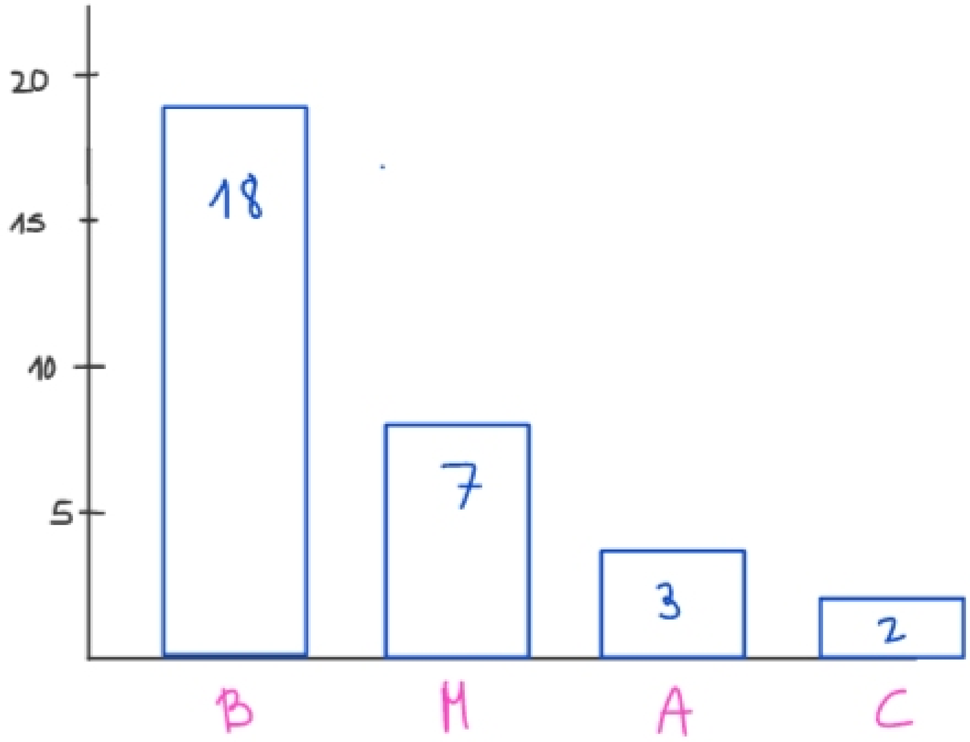
1. El nivel de recepción de spam en un servidor puede clasificarse en cuatro estados: bajo, medio, alto y crítico. Durante el mes de septiembre (n=30 días), se han registrado dichos niveles y los datos se recogen en la siguiente tabla de frecuencias.

   a) Completa la tabla de frecuencias.

	Fr. abs. n_i	Fr. rel. f_i	Fr. abs. acum. N_i	Fr. rel. acum. F_i
Bajo	18	18/30	18 $\left.\right\}$ 25−18=7	18/30
Medio	7	7/30	25	25/30
Alto	3	3/30	28	28/30
Crítico	2	2/30	30	30/30=1
	n=30	1		

   b) Representa el diagrama de barras



2. Del siguiente conjunto de datos de la variable X, $\{x_i, i = 1, \ldots, 6\} = \{\overset{x_1}{2}, \overset{x_2}{3}, \overset{x_3}{4}, \overset{x_4}{4}, \overset{x_5}{5}, \overset{x_6}{6}\}$.

a) ¿Podría ser la media menor que 2?

b) Calcula la media aritmética, la moda y la mediana.

c) Calcula la varianza, la desviación típica y el rango muestrales.

d) Calcula la media de los valores transformados $y_i = 3x_i - 2$.

a) No $\rightarrow$ $\min \leq \bar{X} \leq \text{máx}$ $\rightarrow$ $\boxed{\bar{X} \geq 2}$

b) Media: $\bar{X} = \frac{2+3+4+4+5+6}{6} = \boxed{4}$

Moda $\rightarrow$ Dato que más se repite

$Mo = \boxed{4}$

Mediana: $Me = \frac{X_3 + X_4}{2} = \frac{4+4}{2} = \boxed{4}$

c) Varianza: $s^2 = \frac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$

$$S^2 = \frac{(2-4)^2 + (3-4)^2 + (4-4)^2 + (4-4)^2 + (5-4)^2 + (6-4)^2}{6} = \frac{10}{6} = \boxed{\frac{5}{3}}$$

Desviación típica: $s = +\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$ $S = +\sqrt{\frac{5}{3}}$

Rango muestral: $R = \text{máx}\{X_i\} - \min\{X_i\} = 6 - 2 = 4$

d) $y_i = 3X_i - 2$ $\bar{y} = \frac{4+7+10+10+13+16}{6} = 10$

ó: $\bar{y} = 3\bar{X} - 2 = 3 \cdot 4 - 2 = 10$

3. De una muestra de 4 datos conocemos su media $\overline{x} = 2$ y las diferencias de tres observaciones a la media: $(x_1 - \overline{x}) = -3$, $(x_2 - \overline{x}) = -1$ y $(x_3 - \overline{x}) = 3$. Calcula la varianza muestral y el rango.

$\overline{X} = 2$

$(X_1 - \overline{X}) = -3 \rightarrow X_1 = -1$

$(X_2 - \overline{X}) = -1 \rightarrow X_2 = 1$

$(X_3 - \overline{X}) = 3 \rightarrow X_3 = 5$

Varianza: $S^2 = \dfrac{(-3)^2 + (-1)^2 + (3)^2}{3} = \dfrac{19}{3}$

Rango: $5 - (-1) = 6$

4. Sea X =*edad* en un grupo de trabajadores de una empresa informática. ¿Cuáles de estas afirmaciones son posibles?

a) En una muestra de tamaño $n = 8$, si la mediana $Me = 23$ entonces $x_{(4)} = 23$, donde $x_{(i)}$ denota el dato ordenado i-ésimo.

b) El primer cuartil es $Q_1 = 20$ y la mediana es $Me = 18$.

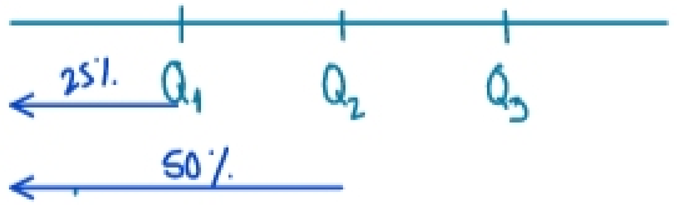
c) El decil $d_7 = 33$ y el tercer cuartil $Q_3 = 32$.

d) El decil $d_8 = 45$ y el percentil $p_{81} = 47$.

e) El rango es 25 y el rango intercuartílico es 10.

a) $M_e = \dfrac{X_4 + X_5}{2} \longrightarrow 23 = \dfrac{23 + X_5}{2}$ es posible si $X_5 = 23$

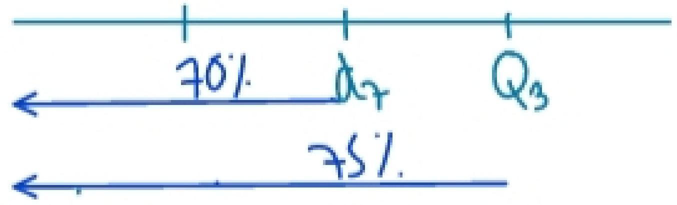
b)



$\boxed{Q_2 \geq Q_1}$ $18 \not\geq 20$

No es posible

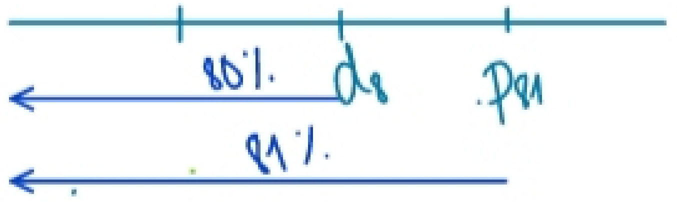
c)



$\boxed{Q_3 \geq d_7}$ $32 \not\geq 33$

No es posible

d)

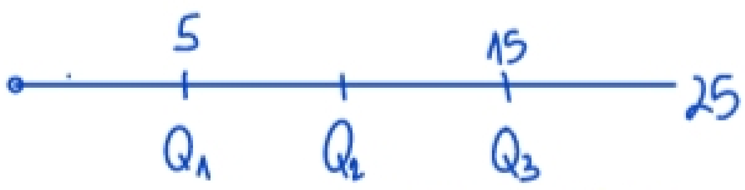


$\boxed{P_{81} \geq d_8}$ $47 \geq 45$

es posible

e) $\boxed{R = \max\{X_i\} - \min\{X_i\} = 25}$

$\boxed{RIQ = Q_3 - Q_1 = 10}$



es posible

5. Un profesor está haciendo a sus alumnos un examen que tiene una calificación máxima de 20 puntos. Su sistema de puntuación es tal que las notas deben ser necesariamente múltiplos de 5. Al final, una vez corregidos los exámenes, obtiene la siguiente distribución de notas:

Nota	f_i	Frec. abs n_i
5	0,1	2
10	0,2	4
15	0,3	6
20	0,4	8
		$n=20$

$$\overline{X} = X_1\cdot\frac{n_1}{n} + X_2\cdot\frac{n_2}{n} + X_3 + \frac{n_3}{n} \cdots = X_1 f_1 + X_2 f_2$$

a) Supón que hay 10 alumnos en la clase. ¿Puedes calcular la media y la desviación típica de la notas?

$$\overline{X} = \frac{5+10+10+15+15+15+20+20+20+20}{10} = 5\cdot\frac{1}{10} + 10\cdot\frac{2}{10} + 15\cdot\frac{3}{10} + 20\cdot\frac{4}{10} = \boxed{15}$$

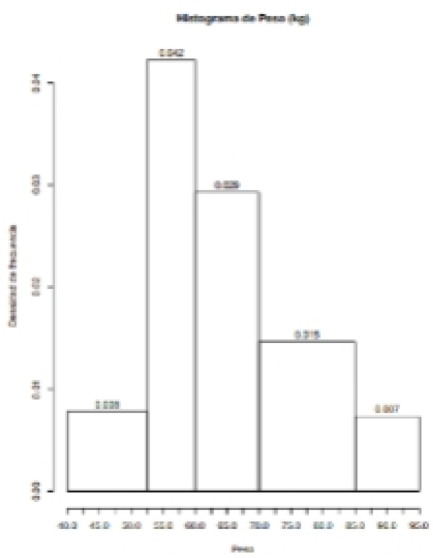
b) Supón que hay 20 alumnos en la clase. ¿Puedes calcular la media y la desviación típica de las notas?

c) Supón que no sabes cuántos alumnos hay en la clase. ¿Puedes calcular la media y la desviación típica de las notas?

6. El histograma de la Figura 1 (izquierda) resume la información recogida sobre el peso en kilogramos de un grupo de estudiantes de primer curso. Las barras verticales representan la densidad de frecuencia relativa, $h_i = f_i/a_i$.

a) Calcula a partir de estos datos la tabla de frecuencias, los tres cuartiles de la variable peso.

b) Se dice que una persona es delgada si su peso es inferior al del $90\,\%$ de las personas. ¿Cuál es el peso máximo de una persona considerada delgada?

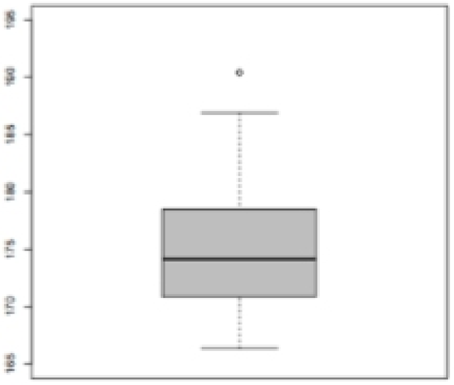


a) Tabla de frecuencias:

$\boxed{f_i = h_i \cdot a_i}$

Intervalos de clase	Amp. a_i	Alt h_i	Fr. relativa f_i	Frec. rel. acumulada
[40, 52'5)	12,5	0,008	0,1	0,1
[52'5, 60)	7,5	0,0042	0,315	0,415
[60, 70)	10	0,029	0,29	0,705
[70, 85]	15	0,015	0,229	0,93
[85, 95)	10	0,007	0,07	1
			1	

7. El diagrama de caja de la Figura 1 (derecha) se construyó a partir de las alturas de 100 personas de una cierta comunidad. A partir de la gráfica, obtén medidas descriptivas de localización y dispersión (aproximadas) e interpreta los resultados. ¿Podríamos afirmar que el $90\,\%$ de las personas medidas tienen una altura menor a 175?



Medidas de localización -> central -> Mediana $\simeq$ 174
-> no central -> Q1 $\simeq$ 173 ; Q3 $\simeq$ 177
-> máx{Xi} $\simeq$ 191 ; mín{Xi} $\simeq$ 166

Medidas de dispersión -> Abs -> RIQ = Q3 - Q1 $\simeq$ 177 - 171 = 6
-> R = máx{Xi} - mín{Xi} $\simeq$ 191 - 166 = 25
-> Rel, no se pueden calcular porque no tenemos la media

$\boxed{\text{Simetría : Q3 - Q2 = Q2 - Q1}}$ A ojo podría haber simetría por el diagrama de caja

$d_9 = 175?$ $\quad Q_3 \leq d_9$

$177 \not\leq 175$

No es posible

8. Panasonic y KTS fabrican un tipo de pilas que se utilizan en la BIOS de los ordenadores. La primera de las marcas garantiza una duración media de 1600 días, con una desviación típica de 200 días, mientras que KTS nos da una media de 1800 días, con desviación típica de 300 días. Si tenemos dos pilas de ambas marcas que han durado 1500 días, ¿cuál de ellas presenta un mejor rendimiento relativo? ¿Y si tuviésemos dos pilas que durasen 2000 y 2100 días, respectivamente? (*Aplicación de tipificación de datos*). ¿Qué marca ofrece una menor variación? (*Aplicación del coeficiente de variación*)

a)

① Panasonic	② KTS
$\bar{X}_1 = 1600$	$\bar{X}_2 = 1800$
$S_1 = 200$	$S_2 = 300$
$X_1 = 1500$	$X_2 = 1500$

Tipificamos para poder comparar

$$\left\{\begin{array}{l} Z_1 = \dfrac{X_1 - \bar{X}_1}{S_1} = \dfrac{1500-1600}{200} = -0{,}5 \\[2ex] Z_2 = \dfrac{X_2 - \bar{X}_2}{S_2} = \dfrac{1500-1800}{300} = -1 \end{array}\right\} Z_2 < Z_1 \rightarrow Z_2 \text{ es peor que } Z_1$$

b)

$\bar{X}_1 = 1600$	$\bar{X}_2 = 1800$
$S_1 = 200$	$S_2 = 300$
$X_1 = 2000$	$X_2 = 1500$

Tipificamos datos

$$\left\{\begin{array}{l} Z_1 = \dfrac{2000-1600}{200} = 2 \\[2ex] Z_2 = \dfrac{2100-1800}{300} = 1 \end{array}\right\} Z_2 < Z_1$$

$$\left.\begin{array}{l} CV_1 = \dfrac{S_1}{\bar{X}_1} = \dfrac{200}{1600} = \dfrac{1}{8} \\[2ex] CV_2 = \dfrac{S_2}{\bar{X}_2} = \dfrac{300}{1800} = \dfrac{1}{6} \end{array}\right\} CV_1 < CV_2 \rightarrow \text{① Panasonic presenta menor variabilidad.}$$

10. Queremos ejecutar un programa que requiere de alta computación en un servidor cloud. En función del precio pagado nos dan acceso a más recursos computacionales. Tras hacer unas pruebas iniciales hemos registrado los siguientes tiempos de resolución Y en función del dinero pagado X. Los datos son los siguientes:

Dinero	5	6	7	8	9
Tiempo (horas)	40	35	25	12	10

a) ¿Existe relación lineal entre el dinero pagado y el tiempo de ejecución?

b) Calcula una recta de regresión que permita describir el tiempo de ejecución en función del dinero pagado.

c) Evalúa el ajuste del modelo.

d) Si invirtiésemos $5{,}5$, ¿cuál sería el tiempo estimado de ejecución?

e) ¿Cuánto dinero tendríamos que invertir para que el programa tardase 15 horas?

a) $\bar{X} = \frac{1}{n}(X_i + \dots + X_n) = \frac{1}{5}(5+6+\dots+9) = 7$

$\bar{y} = \frac{1}{5}(40+35+\dots+10) = 24{,}4$

$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{5}\left[(5-7)^2 + \dots + (9-7)^2\right] = \frac{1}{5}(4+1+0+1+4) = 2$

$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}) = \frac{1}{5}\left[40^2 + \dots + 10^2\right] - 24{,}4^2 = 143{,}44$

$S_{xy} = \sum x_i \cdot y_i - \bar{x}\bar{y} = \frac{1}{5}(5 \cdot 40 + \dots + 9 \cdot 10) - 7 \cdot 24{,}4 = -16{,}6$

b) Recta de regresión: $\boxed{y = a + bx}$

$b = \frac{S_{xy}}{S_x^2} = \frac{-16{,}6}{2} = -8{,}3$

$a = \bar{y} - b\bar{x} = 24{,}4 - (-8{,}3) \cdot 7 = 82{,}5$

$\boxed{y = 82{,}5 - 8{,}3x}$

c) Coef de correlación lineal $r = \frac{S_{xy}}{S_x S_y} = \frac{-16{,}6}{\sqrt{2}\sqrt{143{,}44}} = -0{,}98 \leftarrow \simeq -1 \rightarrow$ relación inversa

Ajuste $\rightarrow r^2 = (-0{,}98)^2 = 0{,}96 \simeq$ valores cercanos a 1 indican un buen ajuste

d) $y = 82{,}5 - 8{,}3x \rightarrow \hat{y} = 82{,}5 - 8{,}3 \cdot 5{,}5 = \underline{36{,}85h}$ debería tardar

11. Sobre la recta de regresión:

a) Si la recta de regresión de Y sobre X es $y = 3{,}74 + 0{,}625x$ y la recta de regresión de X sobre Y es $x = 4{,}2 + 0{,}4y$, calcula las medias de ambas variables y el coeficiente de correlación entre X e Y. ¿Cuál es el porcentaje de variabilidad explicada por cada modelo?

b) Si la recta de regresión de Y (variable respuesta) sobre X (variables explicativa) es creciente, ¿puede ser la recta de regresión de X sobre Y decreciente? Razona la respuesta.

y sobre x $\rightarrow y = 3'74 + 0'625x \rightarrow \hat{b}_x = \frac{S_{xy}}{S_x^2} : 0'625 = \frac{S_{xy}}{S_x^2} ; 3'74 = \bar{y} - 0{,}625\bar{x}$

x sobre y $\rightarrow x = 4'2 + 0'4y \rightarrow \hat{b}_y = \frac{S_{xy}}{S_y^2} : 0{,}4 = \frac{S_{xy}}{S_y^2} ; 4{,}2 = \bar{x} - 0{,}4\bar{y}$

$\boxed{\hat{b} = \frac{S_{xy}}{S_x^2}}$ $\boxed{\hat{a} = \bar{y} - \hat{b}\bar{x}}$

$\bar{y} = 3'74 + 0'625\bar{x} = 5'99$

$4'2 = \bar{x} - 0{,}4(3'74 + 0'625\bar{x}) \rightarrow 4'2 + 1'496 = \bar{x} - 0{,}25\bar{x} \rightarrow \bar{x} = \frac{2'7}{0'75} = 3'6$

$S_{xy} = 0'625 \cdot S_x^2 = 0'4 \cdot S_y^2$

b) No, si la recta es creciente indica que la relación entre x e y es creciente, por lo que la recta de y sobre x...

12. De una muestra de 100 personas se obtiene la talla X en centímetros y el peso Y en kilogramos. Tras un análisis descriptivo, hemos obtenido las siguientes medidas:

$$\bar{x} = 167\ cm, \quad \bar{y} = 70\ kg, \quad s_X = 10\ cm, \quad s_Y = 7\ kg, \quad r = 0{,}8.$$

a) ¿Se puede afirmar que las personas con mayor peso tienen también mayor talla, o viceversa? Si, ya que $r > 0$, por lo que la relación es acuerda o directa ($\uparrow x \rightarrow \uparrow y$)

b) Calcula la recta de regresión de Y sobre X (es decir, la que explica el peso en función de la talla).

c) ¿Qué altura se puede predecir para una persona con 65 kg de peso?

d) ¿Qué peso se puede predecir para una persona de 180 cm de altura?

b) $y = \hat{a} + \hat{b}x = -23{,}50 + 0{,}56x$

$\hat{a} = \bar{y} - \hat{b}\bar{x} = 70 - 167 \cdot 0'56 = -23{,}52\ Kg$

$\hat{b} = \frac{S_{xy}}{S_x^2} = \frac{56}{10^2} = 0'56\ Kg/cm$

$r_{xy} = \frac{S_{xy}}{S_x S_y} \rightarrow S_{xy} = 0'8 \cdot 10 \cdot 7 = 56\ Kg \cdot cm$

c) $\cancel{65 = 0'56x - 23'52 \rightarrow x = 88}$ hay que hacer X sobre y.

d) $y = -23'52 + 0'56 \cdot 180 = 77'28\ Kg$