

Responde siempre razonadamente:

1. La siguiente tabla recoge las edades de las personas que han subido a un avión Madrid - Barcelona.

Edad	[0, 18)	[18, 36)	[36, 54)	[54, 72)	[72, 90)
Nº personas	15	30	20	12	3

- a) Construye la tabla de frecuencias. (0,75 puntos)
- b) Representa los datos en un histograma. (0,5 puntos)
- c) Calcula las medidas de centralización. (1 punto)
- d) Calcula las medidas de dispersión, el primer cuartil y el coeficiente de variación. (1,25 puntos)
2. Demuestra y razona: (1,5 puntos)
- a) Iván y Jorge han realizado 3 exámenes de matemáticas en la primera evaluación. Jorge ha obtenido 2 puntos que más que Iván en cada uno de ellos. Sabiendo que la nota media de Iván es 5,2 y que la varianza es 1,7, ¿cuál es la media y cuál es la varianza de las notas de Jorge?
- b) Un atleta entrena, de lunes a viernes, recorriendo 30, 20, 20, 40 y 30 km respectivamente. Si el sábado también entrena, ¿cuántos km debería entrenar para aumentar la media 3 km más?
3. Durante los últimos 5 años, una empresa ha estudiado la relación entre el número de ventas anual (Y) y los gastos en publicidad (X):

Ventas (miles de €)	100	150	200	210	300
Gastos en publicidad (miles de €)	10	14	21	22	28

- a) Calcula el coeficiente de correlación e interpreta su valor. (1,25 puntos)
- b) Calcula la recta de regresión de Y sobre X. (1,25 puntos)
- c) Si un año, los gastos en publicidad se elevan a 25 mil euros, ¿cuál será el valor esperado de las ventas en ese año? (1 punto)
4. De una variable bidimensional (X, Y) sabemos que:
- La recta de regresión de X sobre Y es $x = 0,6 - 0,44 \cdot y$
 - $r = -0,7$; $\sigma_y = 1,91$
- Calcula la desviación típica de la variable X.

1. La siguiente tabla recoge las edades de las personas que han subido a un avión Madrid - Barcelona.

Edad	[0, 18)	[18, 36)	[36, 54)	[54, 72)	[72, 90)
Nº personas	15	30	20	12	3

- a) Construye la tabla de frecuencias. (0,75 puntos)

Al estar los datos agrupados en intervalos, calculamos la marca de clase x_i como el punto medio de cada uno de ellos.

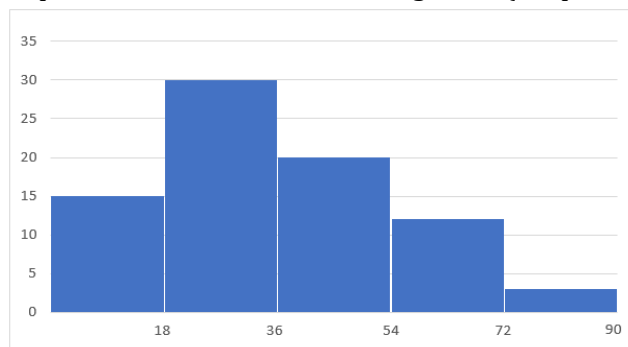
En la tabla de frecuencias indicaremos la marca de clase, las frecuencias absolutas y relativas y las frecuencias acumuladas.

intervalo	x_i	f_i	F_i	h_i	H_i	%	% acumulado	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[0, 18)	9	15	15	0,1875	0,1875	18,75	18,75	135	1215
[18, 36)	27	30	45	0,375	0,5625	37,5	56,25	810	21870
[36, 54)	45	20	65	0,25	0,8125	25	81,25	900	40500
[54, 72)	63	12	77	0,15	0,9625	15	96,25	756	47628
[72, 90)	81	3	80	0,0375	1	3,75	100	243	19683
		80		1		100		2844	130896

Podemos observar que hay un total de 80 datos.

Las dos últimas columnas nos sirven para calcular las medidas de centralización y de dispersión de los siguientes apartados.

- b) Representa los datos en un histograma. (0,5 puntos)



- c) Calcula las medidas de centralización. (1 punto)

La media de los datos se calcula de la siguiente forma: $\bar{x} = \frac{\sum_{i=1}^{80} x_i \cdot f_i}{80} = \frac{2844}{80} = 35,55$

La moda es el dato que más se repite, es decir, aquel con mayor frecuencia absoluta. En este caso, observando la tabla de frecuencias podemos concluir que el intervalo modal sería [18, 36) y la moda la marca de clase correspondiente $M_o = 27$

La mediana es el dato central, es decir, el que deja por encima y por debajo el 50% de los datos. En este caso, al existir 80 datos no hay un dato central, sino que la mediana es la media de los datos situados en las posiciones 40 y 41. Ambos corresponden a la marca de clase 27, por tanto, el intervalo mediano es [18, 36) y la mediana la marca de clase $Me = 27$.

- d) Calcula las medidas de dispersión, el primer cuartil y el coeficiente de variación. (1,25 puntos)

El rango es $R = 90 - 0 = 90$

La desviación típica es la raíz cuadrada de la varianza, por tanto, calculamos en primer lugar esta medida de dispersión.

La varianza se calcula de la siguiente forma: $\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{80} - \bar{x}^2 = \frac{130896}{80} - 35,55^2 = 372,3975$

De forma que la desviación típica es: $\sigma = \sqrt{372,3975} = 19,2976$

El primer cuartil (Q_1), deja por debajo el 25% de los datos, por tanto: $25\% \cdot 80 = 20$. Como $F_2 = 45 > 20 \rightarrow Q_1$ es el dato correspondiente a dicha frecuencia, es decir, $Q_1 = 27$.

El coeficiente de variación se calcula como $CV = \frac{\sigma}{\bar{x}} = \frac{19,2976}{35,55} = 0,54283$.

2. Demuestra y razona: (1,5 puntos)

- a) Iván y Jorge han realizado 3 exámenes de matemáticas en la primera evaluación. Jorge ha obtenido 2 puntos que más que Iván en cada uno de ellos. Sabiendo que la nota media de Iván es 5,2 y que la varianza es 1,7, ¿cuál es la media y cuál es la varianza de las notas de Jorge?

Llamamos: a, b, c a las tres notas de Iván. Entonces:

$$m = 5,2 = \frac{a + b + c}{3} ; \sigma^2 = 1,7 = \frac{a^2 + b^2 + c^2}{3} - m^2$$

Las notas de Jorge entonces son: $a + 2, b + 2, c + 2$

La media es:

$$\frac{a + 2 + b + 2 + c + 2}{3} = \frac{a + b + c}{3} + 2 = m + 2 = 5,2 + 2 = 7,2$$

La varianza es:

$$\begin{aligned} & \frac{(a + 2)^2 + (b + 2)^2 + (c + 2)^2}{3} - (m + 2)^2 = \\ & = \frac{a^2 + 4a + 4 + b^2 + 4b + 4 + c^2 + 4c + 4}{3} - m^2 - 4m - 4 = \\ & = \frac{a^2 + b^2 + c^2}{3} + \frac{4a + 4b + 4c}{3} + \frac{12}{3} - m^2 - 4m - 4 = \\ & = \frac{a^2 + b^2 + c^2}{3} - m^2 + 4m - 4m + 4 - 4 = \frac{a^2 + b^2 + c^2}{3} - m^2 = \sigma^2 = 1,7 \end{aligned}$$

- b) Un atleta entrena, de lunes a viernes, recorriendo 30, 20, 20, 40 y 30 km respectivamente. Si el sábado también entrena, ¿cuántos km debería entrenar para aumentar la media 3 km más?

En primer lugar, calculamos la media de km de entrenamiento de lunes a viernes.

$$\bar{x} = \frac{30 + 20 + 20 + 40 + 30}{5} = \frac{140}{5} = 28$$

Llamamos x al número de km de entrenamiento del sábado. De forma que, si la media debe incrementarse 3 km más, debe ser 31 km:

$$\frac{30 + 20 + 20 + 40 + 30 + x}{6} = 31 \rightarrow \frac{140 + x}{6} = 31 \rightarrow 140 + x = 186 \rightarrow x = 46 \text{ km}$$

3. Durante los últimos 5 años, una empresa ha estudiado la relación entre el número de ventas anual (Y) y los gastos en publicidad (X):

Ventas (miles de €)	100	150	200	210	300
Gastos en publicidad (miles de €)	10	14	21	22	28

- a) Calcula el coeficiente de correlación e interpreta su valor. (1,25 puntos)

En primer lugar, calculamos las medias porque se van a necesitar:

$$\bar{x} = \frac{10 + 14 + 21 + 22 + 28}{5} = 19 \text{ mil euros de media en gastos en publicidad}$$

La nota media de física será:

$$\bar{y} = \frac{100 + 150 + 200 + 210 + 300}{5} = 192 \text{ mil euros en ventas anuales}$$

El coeficiente de correlación es una medida que indica el grado de dependencia lineal que existe entre ambas variables y toma siempre un valor comprendido entre -1 y 1. Se calcula:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Calculamos la covarianza y las desviaciones típicas. Para ello, calculamos la varianza de X y la varianza de Y:

$$\sigma_{xy} = \frac{\sum x_i \cdot y_i}{9} - \bar{x} \cdot \bar{y} = \frac{20320}{5} - 19 \cdot 192 = 416$$

$$\sigma_x^2 = \frac{\sum x_i^2 \cdot f_i}{5} - \bar{x}^2 = \frac{100 + 196 + 441 + 484 + 784}{5} - 19^2 = 40$$

$$\text{Luego, } \sigma_x = \sqrt{40} = 6,32$$

$$\sigma_y^2 = \frac{\sum y_i^2 \cdot f_i}{9} - \bar{y}^2 = \frac{10000 + 22500 + 40000 + 44100 + 90000}{5} - 192^2 = 4456$$

$$\text{Luego, } \sigma_y = \sqrt{4456} = 66,75$$

Por tanto, el coeficiente de correlación es: $r = \frac{416}{6,32 \cdot 66,75} = 0,98 \approx 1$ Hay una dependencia lineal fuerte y positiva entre ambas variables. A medida que aumentan los gastos en publicidad, aumentan las ventas anuales.

b) Calcula la recta de regresión de Y sobre X. (1,25 puntos)

La ecuación de la recta de regresión de Y sobre X es la siguiente:

$$y = \frac{\sigma_{xy}}{\sigma_x^2} \cdot x + \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x}$$

Sustituyendo:

$$y = \frac{416}{40} \cdot x + 192 - \frac{416}{40} \cdot 19 = 10,4 \cdot x - 5,6$$

c) Si un año, los gastos en publicidad se elevan a 25 mil euros, ¿cuál será el valor esperado de las ventas en ese año? (1 punto)

Se trata de estimar el valor de la variable Y sabiendo que X=25. Para ello, sustituimos el valor de x en la recta de regresión anterior:

$$y = 10,4 \cdot 25 - 5,6 = 254,4 \text{ mil euros}$$

A la vista del valor del coeficiente de correlación calculado anteriormente, podemos concluir que al existir una dependencia lineal fuerte entre ambas variables, se trata de una buena estimación y es bastante fiable.

4. De una variable bidimensional (X, Y) sabemos que:

- La recta de regresión de X sobre Y es $x = 0,6 - 0,44 \cdot y$
- $r = -0,7$; $\sigma_y = 1,91$

Calcula la desviación típica de la variable X.

La ecuación de la recta de regresión de X sobre Y es la siguiente:

$$x = \frac{\sigma_{xy}}{\sigma_y^2} \cdot y + \bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \cdot \bar{y}$$

Y sabemos que, la ecuación es: $x = 0,6 - 0,44 \cdot y$

Por tanto, identificando el primer coeficiente:

$$\frac{\sigma_{xy}}{\sigma_y^2} = -0,44 \rightarrow \frac{\sigma_{xy}}{1,91^2} = -0,44 \rightarrow \sigma_{xy} = -1,61$$

Por otro lado, $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$:

$$-0,7 = \frac{-1,61}{1,91 \cdot \sigma_x} = \sigma_x \rightarrow \frac{-1,61}{-0,7 \cdot 1,91} = 1,2$$