

Responde siempre razonadamente:

1. La siguiente tabla recoge las edades de los empleados que trabajan en una empresa:

Edad	[18, 28)	[28, 38)	[38, 48)	[48, 58)	[58, 68)
Nº empleados	19	23	25	15	10

- a) Construye la tabla de frecuencias. **(1 punto)**
b) Representar los datos por medio de un histograma. **(0.75 puntos)**
c) Calcula las medidas de centralización (media, moda y mediana). **(1 punto)**
d) Calcula las medidas de dispersión. **(1 punto)**
2. Se ha realizado una encuesta a 6 propietarios, preguntando el número de personas que habitan el hogar familiar (Y) y el número de habitaciones de la vivienda (X), obteniendo la tabla:

Nº de personas	3	5	4	6	5	4
Nº de habitaciones	2	4	3	5	5	4

- a) Calcula el coeficiente de correlación, indica los cálculos necesarios e interpreta su resultado.
b) Estima el número esperado de personas que viven en el hogar, si éste tiene 6 habitaciones. ¿Es buena la estimación?
(3,75 puntos)
3. De una variable bidimensional (X, Y) sabemos que:
- La recta de regresión de Y sobre X es $y = -5,6 + 10,4 \cdot x$
 - El coeficiente de correlación es $r = 0,99$
 - La covarianza es $\sigma_{xy} = 416$
- Calcula la desviación típica de la variable Y. **(2,5 puntos)**

1. La siguiente tabla recoge las edades de los empleados que trabajan en una empresa:

Edad	[18, 28)	[28, 38)	[38, 48)	[48, 58)	[58, 68)
Nº empleados	19	23	25	15	10

- a) Construye la tabla de frecuencias. (1 punto)

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

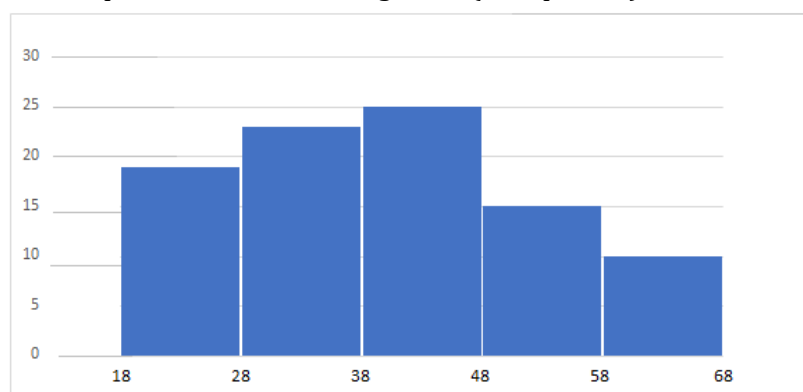
En la tabla de frecuencias indicaremos la marca de clase, las frecuencias absolutas y relativas y las frecuencias acumuladas.

intervalo	x_i	f_i	F_i	h_i	H_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[18, 28)	23	19	19	0,206521739	0,20652174	437	10051
[28, 38)	33	23	42	0,25	0,45652174	759	25047
[38, 48)	43	25	67	0,27173913	0,72826087	1075	46225
[48, 58)	53	15	82	0,163043478	0,89130435	795	42135
[58, 68)	63	10	92	0,108695652	1	630	39690
		92		1		3696	163148

Podemos observar que hay un total de 92 datos.

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

- b) Representar los datos por medio de un histograma. (0.75 puntos)



- c) Calcula las medidas de centralización (media, moda y mediana). (1 punto)

- La media de los datos se calcula de la siguiente forma: $\bar{x} = \frac{\sum_{i=1}^5 x_i \cdot f_i}{92} = \frac{3696}{92} = 40,17$
- 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 [38, 48) y la moda la marca de clase correspondiente $M_o = 43$
- 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 92 datos no hay un dato central, sino que la mediana es la media de los datos situados en las posiciones 46 y 47. Ambos corresponden a la marca de clase 43, por tanto, el intervalo mediano es [38, 48) y la mediana la marca de clase $Me = 43$.

- d) Calcula las medidas de dispersión (varianza, desviación típica, rango y coeficiente de variación). (1 punto)

- La varianza se calcula de la siguiente forma: $\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{92} - \bar{X}^2 = \frac{163148}{92} - 40,17^2 = 159,71$
- La desviación típica es la raíz cuadrada de la varianza, por tanto: $\sigma = \sqrt{159,71} = 12,63$
- El rango es la diferencia entre el mayor valor y el menor valor que toma la variable, en este caso, en datos agrupados por intervalos, es la diferencia entre el extremo superior y el inferior. Rango $68-18=50$
- El coeficiente de variación se calcula como $CV = \frac{\sigma}{\bar{x}} = \frac{12,63}{40,17} = 0,314$

2. Se ha realizado una encuesta a 6 propietarios, preguntando el número de personas que habitan el hogar familiar (Y) y el número de habitaciones de la vivienda (X), obteniendo la tabla:

Nº de personas	3	5	4	6	5	4
Nº de habitaciones	2	4	3	5	5	4

- a) Calcula el coeficiente de correlación, indica los cálculos necesarios e interpreta su resultado. (3,75 puntos)

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.

La forma de calcularlo es a partir de la covarianza y de las desviaciones típicas de las variables:

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

Se trata de datos de una variable bidimensional, que relaciona el número de habitaciones (X) y el número de personas que habitan en el hogar (Y) de 6 inquilinos.

Comenzamos calculando la media de habitaciones:

$$\bar{x} = \frac{2 + 4 + 3 + 5 + 5 + 4}{6} = \frac{23}{6} = 3,83 \approx 4 \text{ habitaciones}$$

La media de personas en el hogar es:

$$\bar{y} = \frac{3 + 5 + 4 + 6 + 5 + 4}{6} = \frac{27}{6} = 4,5$$

Calculamos la covarianza de X y de Y:

$$\begin{aligned} \sigma_{xy} &= \frac{\sum x_i \cdot y_i}{6} - \bar{x} \cdot \bar{y} = \\ &= \frac{2 \cdot 3 + 4 \cdot 5 + 3 \cdot 4 + 5 \cdot 6 + 5 \cdot 5 + 4 \cdot 4}{6} - 3,83 \cdot 4,5 = \frac{109}{6} - 3,83 \cdot 4,5 = 0,9317 \end{aligned}$$

Calculamos las desviaciones típicas. En primer lugar, calculamos la varianza de X y la varianza de Y:

$$\sigma_x^2 = \frac{\sum x_i^2 \cdot f_i}{6} - \bar{x}^2 = \frac{95}{6} - 3,83^2 = 1,1644 ; \text{ Luego, } \sigma_x = \sqrt{1,1644} = 1,07$$

$$\sigma_y^2 = \frac{\sum y_i^2 \cdot f_i}{6} - \bar{y}^2 = \frac{127}{6} - 4,5^2 = 0,917 ; \text{ Luego, } \sigma_y = \sqrt{0,917} = 0,9576$$

Por tanto, el coeficiente de correlación es: $r = \frac{0,9317}{1,07 \cdot 0,9576} = 0,91$ Hay una dependencia lineal fuerte y positiva entre ambas variables.

- b) Estima el número esperado de personas que viven en el hogar, si éste tiene 6 habitaciones. ¿Es buena la estimación?

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:

$$y = \frac{0,9317}{1,1644} \cdot x + 4,5 - \frac{0,9317}{1,1644} \cdot 3,83 = 0,8 \cdot x + 1,42$$

Así, para una casa con $x=6$ habitaciones, se esperan: $y = 0,8 \cdot 6 + 1,42 = 6,22 \simeq 6$ *personas*

Como el coeficiente de correlación es un valor cercano a 1, podemos concluir que la estimación es fiable.

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

- La recta de regresión de Y sobre X es $y = -5,6 + 10,4 \cdot x$
- El coeficiente de correlación es $r = 0,99$
- La covarianza es $\sigma_{xy} = 416$

Calcula la desviación típica de la variable Y. (2,5 puntos)

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

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

Y sabemos que, la ecuación es: $y = -5,6 + 10,4 \cdot x$

Por tanto, identificando los coeficientes:
$$\begin{cases} (1) & \frac{\sigma_{xy}}{\sigma_x^2} = 10,4 \\ (2) & \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \cdot \bar{x} = -5,6 \end{cases}$$

Como $\sigma_{xy} = 416$, de la igualdad (1) obtenemos el valor de la desviación típica de X:

$$\frac{416}{\sigma_x^2} = 10,4 \rightarrow \frac{416}{10,4} = \sigma_x^2 \rightarrow \sigma_x^2 = 40 \rightarrow \sigma_x = 6,32$$

El coeficiente de correlación es

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

$$\text{Por tanto: } 0,99 = \frac{416}{6,32 \cdot \sigma_y} \rightarrow \sigma_y = 66,49$$