ESTADÍSTICA

1.- En una fábrica de lámparas se hace un estudio de las bombillas que hay en el mercado, para lo que se toma una muestra de 500 bombillas y los resultados son:

Vida (horas)	[300-500)	[500- 700)	[700- 900)	[900- 1100)	[1100-1300)
Nº bombillas	50	150	200	75	25

a) Completa la tabla de frecuencias. (1 punto)
b) Usando la tabla de frecuencias halla la mediana y la moda. (1 punto)
c) ¿Cuál es la vida media de las bombillas? ¿Y la desviación típica?(1 punto)
d) ¿Cuántas horas duran el 25% de las bombillas menos duraderas? (1 punto)

2.- La nota media de 100 estudiantes de 2º de Bachillerato en Selectividad es de 5,75 con una desviación típica de 1,25. La recta de regresión de la nota en Selectividad (y) respecto a la nota en el Instituto (x) es $y = 0{,}55 + 0{,}8x$

a) ¿Cuál es la nota media en el Instituto de esos estudiantes? (0,75 puntos)
b) ¿Cuál es el signo del coeficiente de correlación entre la nota en el Instituto y la nota en Selectividad? ¿Por qué? (0,75 puntos)

3.- Se ha realizado una encuesta preguntando por el número de personas que habitan el hogar familiar (x) y el número de dormitorios que tiene la casa (y). La tabla siguiente recoge la información obtenida:

y \ x	1	2	3	4	6
1	2	1	0	0	0
2	0	3	1	0	0
3	0	0	4	4	1
4	0	0	0	2	2

a) Halla la media y la desviación típica del número de personas del hogar familiar y del número de dormitorios (distribuciones marginales). (1 punto)
b) Calcula el coeficiente de correlación lineal entre ambas variables e interprétalo.(1,5 puntos)
c) Obtén la recta de regresión de y sobre x. (1 punto)
d) Si en una casa hay 5 personas ¿qué número de dormitorios podemos estimar que hay en la misma? ¿Esta estimación es buena?¿Por qué? (1 punto)

SOLUCIONES

1.- a)

Vida(h)	x_i	f_i	F_i	h_i	H_i
[300-500)	400	50	50	0,1	0,1
[500-700)	600	150	200	0,3	0,4
[700-900)	800	200	400	0,4	0,8
[900-1100)	1000	75	475	0,15	0,95
[1100-1300)	1200	25	500	0,05	1

b) La mediana estará en el intervalo [700,900) y la moda en el mismo. Calculemos:

$$M = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i}.c = 700 + \frac{250-200}{200}.200 = 750 \text{ horas}$$

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}.c = 700 + \frac{200-150}{50+125} \cdot 200 = 757{,}14 \text{ horas}$$

c)

Vida(h)	x_i	f_i	F_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
[300-500)	400	50	50	20000	122500	6125000
[500-700)	600	150	200	90000	22500	3375000
[700-900)	800	200	400	160000	2500	500000
[900-1100)	1000	75	475	75000	62500	4687500
[1100-1300)	1200	25	500	30000	202500	5062500

$\bar{x} = \frac{375000}{500} = 750$ horas de vida media $\quad \sigma = \sqrt{\frac{19750000}{500}} = 198{,}74$ horas

d) El 25% de las bombillas que menos duran es el percentil 25 o el primer cuartil, es decir, está en el intervalo [500,700) (primero que pasa de N/4=125)

$$Q_3 = L_i + \frac{\frac{n}{4} - F_{i-1}}{f_i}.c = 500 + \frac{125-50}{150} \cdot 200 = 600 \text{ horas o menos duran el 25\% peores.}$$

2.- $y = 0{,}55 + 0{,}8x$ recta de regresión, sabemos que $\bar{y} = 5{,}75$

a) Como la recta de regresión siempre pasa por el centro de gravedad $(\bar{x}, \bar{y})$, tenemos que al sustituir tiene que verificarse la ecuación, es decir: $5{,}75 = 0{,}55 + 0{,}8x$, de donde tendremos que

$0{,}8x = 5{,}75 - 0{,}55 \Rightarrow x = \frac{5{,}2}{0{,}8} = 6{,}5 = \bar{x}$ por lo tanto, la nota media en el Instituto ha sido de 6,5.

b) La pendiente de la recta de regresión es positiva (0,8), de donde podemos deducir que el coeficiente de correlación también lo es, así como la covarianza.

3.- a) Las distribuciones marginales son:

X	1	2	3	4	6
f_i	2	4	5	6	3

Y	1	2	3	4
f_i	3	4	9	4

Podemos hallar las medias y las desviaciones típicas con la calculadora y tenemos que:
$\bar{x} = 3{,}35$ personas $\quad \sigma_x = 1{,}459$ per. $\qquad \bar{y} = 2{,}7$ dormitorios $\quad \sigma_y = 0{,}954$ dorm.

b) Para hallar la covarianza $\sigma_{xy} = \frac{\sum x_i y_i f_i}{N} - \bar{x} \cdot \bar{y} = \frac{204}{20} - 3{,}35 \cdot 2{,}7 = 1{,}155$

$\sum x_i y_i f_i = 1 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot 1 + 2 \cdot 2 \cdot 3 + 3 \cdot 2 \cdot 1 + 3 \cdot 3 \cdot 4 + 4 \cdot 3 \cdot 4 + 4 \cdot 4 \cdot 2 + 6 \cdot 3 \cdot 1 + 6 \cdot 4 \cdot 2 =$
$= 204$

El coeficiente de correlación lineal será $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1{,}155}{1{,}459 \cdot 0{,}954} = 0{,}83$ que indica

que la correlación es fuerte y positiva, es decir que a más personas mas dormitorios.

c) La fórmula de la recta de regresión de y sobre x es: $y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$

por tanto, $y = 2{,}7 + \frac{1{,}155}{1{,}459^2}(x - 3{,}35) \Rightarrow y = 2{,}7 + 0{,}543x - 1{,}818$

por lo que la ecuación de la recta pedida es: $y = 0{,}543x + 0{,}882$

d) Si hay 5 personas, usando la recta de regresión tendremos que:
$y = 0{,}543 \cdot 5 + 0{,}882 = 3{,}597$ o sea que la casa tendrá 3 o 4 dormitorios.

La estimación es bastante aceptable ya que el coeficiente de correlación lineal es bastante próximo a 1 (0,83) y el número de personas (5) está dentro del intervalo en que se ha trabajado.