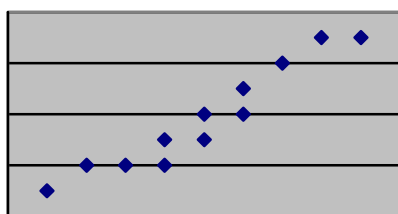
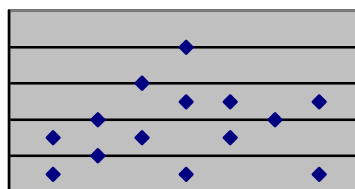


ESTADÍSTICA

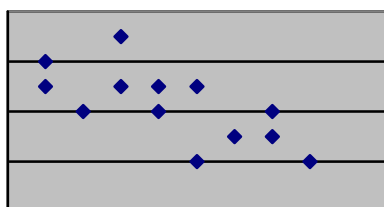
1.- Los números 0'1, 0'99, 0'6 y 0'89 son los valores absolutos del coeficiente de correlación de las distribuciones bidimensionales cuyas nubes de puntos representamos en las siguientes figuras. Asigna razonadamente a cada diagrama su coeficiente de correlación cambiando el signo cuando sea necesario. (1,5 puntos)



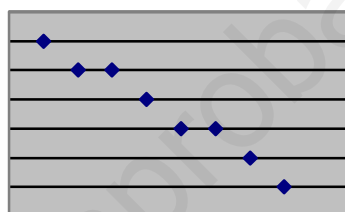
a)



b)



c)



d)

2.- Del ayuntamiento de un pueblo se han obtenido los siguientes datos sobre el número de fincas agrícolas en relación con su superficie:

Superficie (Ha)	[0,5)	[5,10)	[10,15)	[15,20)	[20,25]
Número de fincas	2	5	9	4	2

- Halla, a partir de la tabla, la media, y la desviación típica. (1 punto)
- Halla la mediana y moda de esta distribución. (1,5 puntos)
- Halla el porcentaje de fincas que tienen 15 o más Hectáreas. (1 punto)
- Halla el tercer cuartil. (1 punto)

3.- Se han estudiado los errores cometidos por un grupo de 30 personas en una prueba de ortografía (x) y en otra de cálculo numérico (y). Los resultados están recogidos en la tabla adjunta:

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

- Calcula el coeficiente de correlación lineal entre x e y. (2 puntos)
- Utiliza la recta de regresión de x sobre y para estimar el número de errores ortográficos que tendría alguien con 5 errores de cálculo. (1,5 puntos)
- ¿Es fiable el resultado obtenido en el apartado anterior? ¿Por qué? (0,5 puntos)

SOLUCIONES

1.- La correlación en la gráfica a) es positiva y fuerte, pero más fuerte es en la d), aunque ésta es negativa, la c) es media y negativa y la b) es baja y podría ser positiva o negativa. Luego, la solución sería:

- a) 0'89
- b) 0'1
- c) -0'6
- d) -0'99

2.-

Intervalo	x_i	f_i	F_i	$x_i f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
[0,5)	2'5	2	2	5	95'4529	190'9058
[5,10)	7'5	5	7	37'5	22'7529	113'7645
[10,15)	12'5	9	16	112'5	0'0529	0'4761
[15,20)	17'5	4	20	70	27'3529	109'4116
[20,25)	22'5	2	22	45	104'6529	209'3058
		N=22		270		623'8638

$$a) \bar{x} = \frac{270}{22} = 12'27 \text{ Ha} \quad \sigma = \sqrt{\frac{623'8638}{22}} = 5'325 \text{ Ha}$$

b) Intervalo modal: [10,15) Intervalo mediano: [10,15)

$$M = L_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot c = 10 + \frac{11 - 7}{9} \cdot 5 = 12'2 \text{ Ha}$$

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot c = 10 + \frac{9 - 5}{4 + 5} \cdot 5 = 12'2 \text{ Ha}$$

c) Las fincas que tienen 15 o más Ha son: 9+4+2=15 de 22 en total, lo que supone un porcentaje de 68'18%

d) El tercer cuartil está en el intervalo [15,20) ($\frac{3n}{4} = 16'5$)

$$Q_3 = L_i + \frac{\frac{3n}{4} - F_{i-1}}{f_i} \cdot c = 15 + \frac{16'5 - 16}{4} \cdot 5 = 15'625 \text{ Ha}$$

3.- 3.- a) Las distribuciones marginales son:

X	0	1	2	3	4
f_i	6	6	8	7	3

Y	0	1	2	3	4
f_i	4	7	7	6	6

Podemos hallar las medias y las desviaciones típicas con la calculadora y tenemos que:

$$\bar{x} = 1,8\bar{3} \text{ errores} \quad \sigma_x = 1,267 \text{ errores} \quad \bar{y} = 2,1 \text{ errores} \quad \sigma_y = 1,325 \text{ errores}$$

b) Para hallar la covarianza $\sigma_{xy} = \frac{\sum x_i y_i f_i}{N} - \bar{x} \cdot \bar{y} = \frac{154}{30} - 1,83 \cdot 2,1 = 1,28334$

$$\sum x_i y_i f_i = 1 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot 2 + 2 \cdot 1 \cdot 3 + 2 \cdot 2 \cdot 2 + 3 \cdot 2 \cdot 1 + 2 \cdot 3 \cdot 3 + 3 \cdot 3 \cdot 2 + 4 \cdot 3 \cdot 1 + 3 \cdot 4 \cdot 4 + 4 \cdot 4 \cdot 2 = 154$$

El coeficiente de correlación lineal será $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1,28334}{1,267 \cdot 1,325} = 0,76$

b) Recta de regresión de x sobre y: $x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y})$

$$\text{por tanto, } x = 1,8333 + \frac{1,28334}{1,325^2} (y - 2,1) \Rightarrow x = 1,8333 + 0,731y - 1,535$$

por lo que la ecuación de la recta es: $x = 0,731y + 0,298$

Alguien con 5 errores de cálculo tendría, por tanto $x = 0,731 \cdot 5 + 0,298 = 3,953$

Aproximadamente 4 errores ortográficos.

c) No es muy fiable ya que el coeficiente de correlación no es demasiado alto y además el valor que damos (5) está fuera del rango de los valores estudiados.