

1. Se ha realizado un estudio sobre el consumo de gas (en m^3) en las viviendas de una urbanización durante el mes de enero, obteniéndose los datos que se muestran en la tabla.

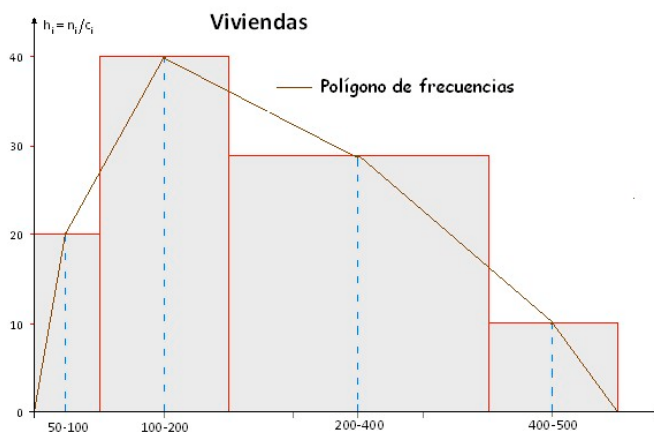
Consumo de gas (m^3)	Viviendas
50-100	10
100-200	40
200-400	60
400-500	10

- Represente el histograma de esta distribución.
- Calcule el consumo medio de gas de las viviendas. ¿El valor hallado es representativo de la distribución?
- Calcule el consumo más frecuente.
- Averigüe el valor del tercer cuartil de la distribución del consumo de gas y explique su significado
- Si la factura del gas consiste en una cantidad fija de 20€ más 0,5€ por cada m^3 consumido, calcule la factura media de las viviendas y determine si la factura es más dispersa que el consumo.

Solución:

a)

Consumo gas	amplitud c_i	n_i	densidad $h_i = \frac{n_i}{c_i}$	N_i	x_i	$x_i n_i$	$x_i^2 n_i$
50 - 100	50	10	0,2	10	75	750	56250
100 - 200	100	40	0,4	50	150	6000	900000
200 - 400	200	60	0,3	110	300	18000	5400000
400 - 500	100	10	0,1	120	450	4500	2025000
						29250	8381250



- b) El consumo medio de gas de las viviendas: $a_1 = \bar{x} = \frac{\sum_{i=1}^4 x_i n_i}{N} = \frac{29250}{120} = 243,75 \text{ m}^3$

$$a_2 = \frac{\sum_{i=1}^4 x_i^2 n_i}{N} = \frac{8381250}{120} = 69843,75 \quad s^2 = a_2 - a_1^2 = 69843,75 - (243,75)^2 = 10429,6875$$

$$s_x = \sqrt{10429,6875} = 102,1258 \text{ m}^3 \quad C.V = \frac{s_x}{\bar{x}} = \frac{102,1258}{243,75} = 0,42 \text{ (42\%)}$$

El consumo medio de gas de las viviendas es de 243,75 m³, con una dispersión del 42%. Con lo cual, el consumo medio de gas no es muy representativo.

- c) El consumo más frecuente se encuentra en el intervalo modal [100-200), puesto que es en el que se alcanza la máxima densidad de frecuencia.

$$M_d = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} c_i = 100 + \frac{0,4 - 0,2}{(0,4 - 0,2) + (0,4 - 0,3)} 100 = 166,67 \text{ m}^3$$

Adviértase que si la amplitud de los intervalos fuera constante: $M_d = L_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} c_i$

- d) El tercer cuartil: $\frac{3 \cdot 120}{4} = 90$, observando en la columna N_i , $Q_3 = P_{75} = L_i + \frac{\frac{3 \cdot N}{4} - N_{i-1}}{N_i - N_{i-1}} c_i$, de donde:

$$Q_3 = P_{75} = 200 + \frac{90 - 50}{110 - 50} 200 = 333,33 \text{ m}^3$$

El 75% de las viviendas que consumen menos, consumen como máximo 333,33 m³ de gas.

- e) Según el enunciado del apartado, la factura del gas viene dada por la relación $Y = 20 + 0,5 \cdot X$, por tanto, hay un cambio de origen y de escala:

La factura media: $\bar{Y} = 20 + 0,5 \cdot \bar{X} = 20 + 0,5 \cdot 243,75 = 141,875 \text{ €}$

$$s_Y^2 = \text{Var}(20 + 0,5 \cdot X) = 0,5^2 \cdot s_X^2 \mapsto s_Y = 0,5 \cdot s_X = 0,5 \cdot 102,1258 = 51,063 \text{ €}$$

$$C.V = \frac{s_Y}{\bar{Y}} = \frac{51,063}{141,875} = 0,36 \text{ (36\%)}$$

La factura del gas está menos dispersa que el consumo.

CAMBIO DE ORIGEN Y DE ESCALA DE LA MEDIA Y VARIANZA: $y = a + bx$

$$\bar{y} = \frac{\sum_{i=1}^k y_i \cdot n_i}{N} = \frac{\sum_{i=1}^k (a + bx_i) \cdot n_i}{N} = \frac{a \sum_{i=1}^k n_i + b \sum_{i=1}^k x_i \cdot n_i}{N} = a \frac{\sum_{i=1}^k n_i}{N} + b \frac{\sum_{i=1}^k x_i \cdot n_i}{N} = a + b \bar{x}$$

$$\mapsto E(y) = E(a + bx) \equiv \bar{y} = a + b\bar{x}$$

La media se ve afectada por el mismo cambio de origen y de escala efectuada sobre la variable.

$$s_y^2 = \frac{\sum_{i=1}^k (y_i - \bar{y})^2 \cdot n_i}{N} = \frac{\sum_{i=1}^k [(a + b x_i - (a + b \bar{x}))^2] \cdot n_i}{N} = \frac{\sum_{i=1}^k (b x_i - b \bar{x})^2 \cdot n_i}{N} = b^2 \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{N} = b^2 s_x^2$$

$$\mapsto \text{Var}(a + b x) = b^2 \cdot s_x^2$$

La varianza no se ve afectada por el cambio de origen pero si por el cambio de escala efectuado sobre la variable.

2. De una distribución bidimensional (X,Y) se sabe que al aumentar los valores de X aumentan los de Y. Se ha obtenido la recta de regresión lineal mínimo cuadrática de Y sobre X y se ha comprobado que la varianza residual, S_{ry}^2 vale cero. Se tienen además los valores de los siguientes momentos respecto al origen:

$$a_{10} = 2 \quad a_{20} = 40 \quad a_{01} = 10 \quad a_{02} = 125$$

- Determine la varianza debida a la regresión en la recta de Y/X y el valor de la covarianza.
- Se hace un cambio de variable de la forma $X' = 2X$. Si se obtiene la nueva recta de regresión de Y/X', ¿será bueno el ajuste? Razone su respuesta.
- Se decide cambiar la función de ajuste de Y sobre X por una constante, $Y = c$. Utilizando el método de mínimos cuadrados, determine el valor de esta constante para nuestro caso.

Solución:

$$\text{a) Las varianzas de las variables X e Y, respectivamente, son: } \begin{cases} s_x^2 = a_{20} - a_{10}^2 = 40 - 2^2 = 36 \\ s_y^2 = a_{02} - a_{01}^2 = 125 - 10^2 = 25 \end{cases}$$

Siendo $s_{ry}^2 = s_y^2 (1 - R^2) = 0 \rightarrow 1 - R^2 = 0 \Rightarrow R^2 = 1$, existe una **dependencia funcional**, el ajuste es perfecto.

Para calcular la covarianza s_{xy} tenemos en cuenta que

$$R^2 = b \cdot b' = \frac{s_{xy}}{s_x^2} \cdot \frac{s_{xy}}{s_y^2} = 1 \mapsto s_{xy}^2 = s_x^2 \cdot s_y^2 = 36 \cdot 25 = 900 \mapsto s_{xy} = \sqrt{900} = 30$$

- El coeficiente de determinación R^2 es invariante ante un cambio de origen y de escala, con lo que la bondad del ajuste será idéntico.
- $E(y) = E(c) \mapsto \bar{y} = c$

INVARIABILIDAD DEL COEFICIENTE DE CORRELACIÓN LINEAL R^2 : $X' = kX$

■ CAMBIO DE ORIGEN:

$$a'_{10} = \bar{X}' = E(X') = E(m + X) = m + E(X) = m + \bar{X}$$

$$a'_{11} = E(X'Y) = E[(m + X)Y] = E(mY) + E(XY) = m\bar{Y} + a_{11}$$

$$s_{x'y'} = a'_{11} - \bar{X}'\bar{Y} = m\bar{Y} + a_{11} - (m + \bar{X})\bar{Y} = a_{11} - \bar{X}\bar{Y} = s_{xy}$$

$$a'_{20} = s_{x'}^2 = \text{Var}(m + X) = s_x^2$$

$$c = \frac{s_{x'y'}}{s_{x'}^2} = \frac{s_{xy}}{s_x^2} \quad c' = \frac{s_{x'y'}}{s_y^2} = \frac{s_{xy}}{s_y^2}$$

$$R'^2 = c \cdot c' = \frac{s_{xy}}{s_x^2} \cdot \frac{s_{xy}}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = R^2$$

■ CAMBIO DE ESCALA:

$$a'_{10} = \bar{X}' = E(X') = E(kX) = kE(X) = k\bar{X}$$

$$a'_{11} = E(X'Y) = E(kXY) = kE(XY) = ka_{11}$$

$$s_{x'y'} = a'_{11} - \bar{X}'\bar{Y} = ka_{11} - k\bar{X}\bar{Y} = k(a_{11} - \bar{X}\bar{Y}) = ks_{xy}$$

$$a'_{20} = s_{x'}^2 = \text{Var}(kX) = k^2 \text{Var}(X) = k^2 s_x^2$$

$$c = \frac{s_{x'y'}}{s_{x'}^2} = \frac{ks_{xy}}{k^2 s_x^2} \quad c' = \frac{s_{x'y'}}{s_y^2} = \frac{ks_{xy}}{s_y^2}$$

$$R'^2 = c \cdot c' = \frac{ks_{xy}}{k^2 s_x^2} \cdot \frac{ks_{xy}}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = R^2$$

El coeficiente de determinación R^2 es invariante ante un cambio de origen y de escala

3. Abel Grandes Pistado preguntó a sus 31 compañeros de clase qué calificación obtuvieron en el último examen de estadística. Sólo recuerda que él aprobó con la nota mediana de 5,6667 y su tocayo Escasi Lopasa tuvo un 4,6 (una de las notas más frecuentes habidas). Y, haciendo memoria, ha podido completar los siguientes datos:

Nota de estadística	Número de alumnos
0 - 4	8
4 - 5	n_2
5 - 7	n_3
7 - 9	6
9 - 10	6

Calcule:

- ¿Qué proporción de alumnos ha obtenido una nota superior a 5? ¿Cómo es la distribución respecto a la moda?
- Estudie la dispersión relativa de las notas a partir del coeficiente de variación de Pearson. Interprete los resultados.
- ¿Cómo afecta a la homogeneidad de la distribución que este examen sea un 60 por ciento de la calificación final?
- Comente, con base estadística, el grado de concentración de las notas de este examen.

Solución:

a)

$L_i - L_{i+1}$	amplitud c_i	n_i	$h_i = \frac{n_i}{c_i}$	N_i	$p_i = \frac{N_i}{N} \%$	x_i	$x_i \cdot n_i$	$U_i = \sum_{i=1}^m x_i \cdot n_i$	$x_i^2 \cdot n_i$	$q_i = \frac{U_i}{U_N} \%$	$p_i - q_i \%$
0 - 4	4	8	2	8	25	2	16	16	32	8,70	16,30
4 - 5	1	$n_2 = 6$	$h_2 = 6$	14	43,75	4,5	27	43	121,5	23,37	20,38
5 - 7	2	$n_3 = 6$	$h_3 = 3$	20	62,50	6	36	79	216	42,93	19,57
7 - 9	2	6	3	26	81,25	8	48	127	384	69,02	12,23
9 - 10	1	6	6	32	100	9,5	57	184	541,5	100	000
		32			212,5		184		1295		68,48

Sabemos que, $M_e = 5,6667$ y $M_d = 4,6$

Para hallar n_2 y n_3 , podemos recurrir a la moda o a la mediana, a saber.

La moda aproximada cuando existen distintas amplitudes: $M_d = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} c_i$

$$4,60 = 4 + \frac{h_3}{2 + h_3} 1 \mapsto h_3 = \frac{1,2}{0,4} = 3 \Rightarrow n_3 = h_3 \cdot c_3 = 3 \cdot 2 = 6$$

$$\text{siendo, } N = 32 = 8 + n_2 + 6 + 6 + 6 \Rightarrow n_2 = 32 - 26 = 6$$

$$\text{La mediana } M_e = L_i + \frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} c_i \rightarrow 5,6667 = 5 + \frac{\frac{32}{2} - (8 + n_2)}{(8 + n_2 + n_3) - (8 + n_2)} 2 \Rightarrow 0,6 = \frac{8 - n_2}{\underbrace{n_3}_{12 - n_2}} 2 \mapsto n_2 = 6$$

$$N = 32 = 20 + n_2 + n_3 \Rightarrow n_3 = 32 - 26 = 6$$

La proporción de alumnos que obtienen una nota superior a 5. La distribución respecto a la moda.

$$p\left(\frac{\sum x_i > 5}{N}\right) \cdot 100 = \frac{n_3 + n_4 + n_5}{32} \cdot 100 = \frac{6 + 6 + 6}{32} \cdot 100 = 56,25\%$$

La distribución es bimodal, puesto que $h_2 = h_5 = 6$

b) Dispersión relativa de las notas a partir del coeficiente de variación de Pearson. Interpretar los resultados.

$$a_1 = \bar{x} = \frac{\sum_{i=1}^5 x_i n_i}{N} = \frac{184}{32} = 5,75$$

$$a_2 = \frac{\sum_{i=1}^5 x_i^2 n_i}{N} = \frac{1295}{32} = 40,46875$$

$$s_x^2 = a_2 - a_1^2 = 40,46875 - 5,75^2 = 7,40625 \quad s_x = \sqrt{7,40625} = 2,72$$

$$C.V = \frac{s_x}{\bar{x}} = \frac{2,72}{5,75} = 0,4730 \text{ (47,30\%)}, \text{ la dispersión es del 47,30 \%, es decir, una dispersión media.}$$

c) La homogeneidad de la distribución, cuando el examen es un 60 % de la calificación final.

CAMBIO DE ESCALA DEL COEFICIENTE DE VARIACIÓN DE PEARSON C.V: $y = k \cdot x$

$$E(y) = E(k \cdot x) = k \cdot E(x) = k \cdot \bar{x}$$

$$\text{Var}(y) = \text{Var}(k \cdot x) = k^2 \cdot \text{Var}(x) = k^2 \cdot s_x^2 \quad \mapsto \quad s_y = \sqrt{k^2 \cdot s_x^2} = k \cdot s_x \quad \mapsto \quad C.V_y = \frac{k \cdot s_x}{k \cdot \bar{x}} = \frac{s_x}{\bar{x}} = C.V_x$$

El Coeficiente de Variación de Pearson es invariante ante un cambio de escala.

$$C.V_{\text{final}} = \frac{s_{\text{final}}}{\bar{x}_{\text{final}}} = \frac{2 \cdot s_x}{2 \cdot \bar{x}} = 0,4730 \text{ (47,30\%)}$$

d) Grado de concentración de las notas de este examen.

$$\text{El índice de concentración de Gini: } I_G = \frac{\sum_{i=1}^{5-1} (p_i - q_i)}{\sum_{i=1}^{5-1} p_i} = \frac{68,48}{212,5} = 0,32 \text{ (32\%)}$$

La concentración es medio-baja.

4. Se han obtenido las siguientes expresiones para las rectas de regresión mínimo cuadráticas de una variable bidimensional (X,Y), donde X es el gasto mensual en ocio e Y el gasto mensual en transporte de un grupo de amigos:

$$\begin{cases} Y = 4X + 2 \\ Y = 2X + 10 \end{cases}$$

Sabiendo además que la covarianza entre ambas variables $s_{xy} = 60$. Se pide:

- Identifique cuál es la recta de regresión de Y/X y de X/Y.
- Interprete los coeficientes de las rectas de regresión.
- Porcentaje de variabilidad explicada y no explicada por la recta.
- Calcule la varianza residual en la regresión Y/X. ¿Coincidirá con la varianza residual en la regresión X/Y? Justifique su respuesta.

Solución:

- a) Recta de regresión Y/X: $Y = 2X + 10$, pendiente $b = 2$

$$\text{Recta de regresión X/Y: } Y = 4X + 2 \quad \mapsto \quad 4X = Y - 2 \quad \mapsto \quad X = \frac{1}{4}Y - \frac{1}{2}, \text{ pendiente } b' = \frac{1}{4}$$

La otra opción no puede ocurrir:

$$\text{Recta de regresión Y/X: } Y = 4X + 2$$

$$\text{Recta de regresión X/Y: } Y = 2X + 10 \quad \mapsto \quad 2X = Y - 10 \quad \mapsto \quad X = \frac{1}{2}Y - 5$$

puesto que $R^2 = b \cdot b' = 4 \cdot \frac{1}{2} = 2$ cuando se sabe que $0 \leq R^2 \leq 1$

- b) Como las dos pendientes son positivas (2 y 1/4), la recta de regresión de Y/X tiene mayor pendiente en valor absoluto que la de X/Y
- c) El coeficiente de determinación lineal $R^2 = b \cdot b' = 2 \cdot \frac{1}{4} = \frac{1}{2} = 0,5$

La recta de regresión de Y sobre X explica el 50% de la variabilidad de la variable dependiente y el otro 50% es no explicado.

$$d) \begin{cases} b = \frac{s_{xy}}{s_x^2} \Rightarrow 2 = \frac{60}{s_x^2} \mapsto s_x^2 = 30 \\ b' = \frac{s_{xy}}{s_y^2} \Rightarrow \frac{1}{4} = \frac{60}{s_y^2} \mapsto s_y^2 = 240 \end{cases}$$

Las varianzas residuales: $\begin{cases} s_{ry}^2 = s_x^2 \cdot (1 - R^2) \mapsto s_{ry}^2 = 30 \cdot (1 - 0,5) = 15 \\ s_{rx}^2 = s_y^2 \cdot (1 - R^2) \mapsto s_{rx}^2 = 240 \cdot (1 - 0,5) = 120 \end{cases}$

5. Sabiendo que $\bar{x} = 3$, $s_x^2 = 6$, $s_y^2 = 8$ y que la recta de regresión de Y sobre X es $y = 4 - 0,667 \cdot x$. Obtener la recta de regresión de X sobre Y.

Solución:

$$Y/X: y = 4 - 0,667 \cdot x \mapsto \begin{cases} \bar{y} = 4 - 0,667 \cdot \bar{x} = 4 - 0,667 \cdot 3 \approx 2 \\ b = -0,667 = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{6} \mapsto s_{xy} = -0,667 \cdot 6 \approx -4 \end{cases}$$

$$X/Y: x = a' + b'y \mapsto \begin{cases} b' = \frac{s_{xy}}{s_y^2} = \frac{-4}{8} = -0,5 \\ \bar{x} = a' + b'\bar{y} \mapsto 3 = a' - 0,5 \cdot 2 \mapsto a' = 4 \end{cases} \Rightarrow x = 4 - 0,5 \cdot y$$

6. Hallar la recta de regresión de Y sobre X sabiendo que $\bar{x} = 4,1$, $\bar{y} = 2,3$ y la recta pasa por el punto (5,9, 3,5)

Solución:

$$Y/X: y = a + b \cdot x \mapsto \begin{cases} \bar{y} = a + b \bar{x} \mapsto 2,3 = a + 4,1 \cdot b \\ \text{por pasar por } (5,9, 3,5) \mapsto 3,5 = a + 5,9 \cdot b \end{cases}$$

$$\begin{cases} a + 4,1 \cdot b = 2,3 \\ a + 5,9 \cdot b = 3,5 \end{cases} \mapsto \begin{cases} b = \frac{-1,2}{-1,8} = 0,667 \\ a = 2,3 - 4,1 \cdot 0,667 = -0,435 \end{cases} \mapsto y = -0,435 + 0,667 \cdot x$$

7. La tabla muestra la comprensión lectora (X) de dos grupos de individuos educados en niveles socioculturales altos (A) y bajos (B).

Intervalos	n_A	n_B
0 - 6	4	4
7 - 13	6	7
14 - 20	9	9
21 - 27	12	8
28 - 34	9	2

Si a partir de la puntuación $X = 19$ se considera una comprensión lectora buena. Se pide:

- Porcentaje de personas en cada grupo con una buena comprensión lectora.
- ¿Entre qué valores de comprensión lectora estará la quinta parte central del Grupo A?
- ¿Entre qué valores de comprensión del Grupo B se encuentran los 12 centrales?
- ¿Cuál de los dos grupos presenta mayor variabilidad?

Solución:

- a) Advuértase que los intervalos son cerrados, se deben expresar abiertos a la derecha con extremos reales:

Intervalos	x	c_i	n_A	N_A	$x \cdot n_A$	$x^2 \cdot n_A$	n_B	N_B	$x \cdot n_B$	$x^2 \cdot n_B$
- 0,5 - 6,5	3	7	4	4	12	36	4	4	12	36
6,5 - 13,5	10	7	6	10	60	600	7	11	70	700
13,5 - 20,5	17	7	9	19	153	2601	9	20	153	2601
20,5 - 27,5	24	7	12	31	288	6912	8	28	192	4608
27,5 - 34,5	31	7	9	40	279	8649	2	30	62	1922
					792	18798			489	9867

Se calcula el orden k del percentil que es igual a 19.

Este da el porcentaje de las personas que tienen menos de 19 puntos. La respuesta será su diferencia hasta 100.

- En el Grupo A:

$$P_k = 19 = 13,5 + \frac{\frac{k \cdot 40}{100} - 10}{19 - 10} \cdot 7 \mapsto 19 = 13,5 + \frac{7 \cdot (0,4 \cdot k - 10)}{9} \leftrightarrow 49,5 = 2,8 \cdot k - 70$$

$$k = 119,5 / 2,8 = 42,68$$

El 57,32% $[100 - 42,68 = 57,32]$ tiene una buena comprensión lectora en el Grupo A.

- En el Grupo B:

$$P_k = 19 = 13,5 + \frac{\frac{k \cdot 30}{100} - 11}{20 - 11} \cdot 7 \mapsto 19 = 13,5 + \frac{7 \cdot (0,3 \cdot k - 11)}{9} \leftrightarrow \begin{aligned} 49,5 &= 2,1 \cdot k - 77 \\ k &= 126,5 / 2,1 = 60,24 \end{aligned}$$

El 39,76% $[100 - 60,24 = 39,76]$ tiene una buena comprensión lectora en el Grupo B.

En consecuencia, el Grupo A tiene una mejor comprensión lectora.

- b) La quinta parte representa el 20%. Con relación al centro (50%), cubrirán desde el 40% al 60%, se tendrá que calcular el Percentil 40 y el Percentil 60 de la distribución de comprensión lectora del Grupo A.

$$P_{40} = 13,5 + \frac{\frac{40 \cdot 40}{100} - 10}{19 - 10} \cdot 7 = 13,5 + \frac{16 - 10}{19 - 10} \cdot 7 = 18,17$$

$$P_{60} = 20,5 + \frac{\frac{60 \cdot 40}{100} - 19}{31 - 19} \cdot 7 = 20,5 + \frac{24 - 19}{31 - 19} \cdot 7 = 23,42$$

La quinta parte central del Grupo A se encuentra entre los valores $[18,17 - 23,42]$

- c) Los 12 valores representa el $(12 / 30 = 40\%)$. Con relación al centro (50%), cubrirán desde el 30% al 70%, teniendo que calcular el Percentil 30 y el Percentil 70 de la distribución de comprensión lectora del Grupo B.

$$P_{30} = 6,5 + \frac{\frac{30 \cdot 30}{100} - 4}{11 - 4} \cdot 7 = 6,5 + \frac{9 - 4}{11 - 4} \cdot 7 = 11,5$$

$$P_{70} = 20,5 + \frac{\frac{70 \cdot 30}{100} - 20}{28 - 20} \cdot 7 = 20,5 + \frac{21 - 20}{28 - 20} \cdot 7 = 21,375$$

Los 12 centrales valores centrales de comprensión del Grupo B se encuentran entre $[11,5 - 21,375]$

- d) Mayor variabilidad tendrá aquel grupo que posea mayor dispersión entre sus valores, es decir, si la media aritmética es representativa de las observaciones (no existen valores extremos exageradamente distanciados de la mayoría).

El estadístico más adecuado para medir la variabilidad relativa entre dos series es el Coeficiente de Variación de Pearson, entendiendo que un valor mayor indica menor homogeneidad, un valor menor refleja menor dispersión o variabilidad.

$$\bar{x}_A = \frac{792}{40} = 19,8 \quad s_A^2 = \frac{18798}{40} - 19,8^2 = 77,91 \quad s_A = \sqrt{77,91} = 8,83$$

$$\bar{x}_B = \frac{489}{30} = 16,3 \quad s_B^2 = \frac{9867}{30} - 16,3^2 = 63,21 \quad s_B = \sqrt{63,21} = 7,95$$

$$CV_A = \frac{8,83}{19,8} \cdot 100 = 44,59\% \quad CV_B = \frac{7,95}{16,3} \cdot 100 = 48,77\%$$

El Grupo B presenta mayor variabilidad relativa, en contra de lo obtenido comparando la desviación típica.

8. A partir de la tabla adjunta, donde $N=11$, $\bar{Y}=0$

X \ Y	- 2	0	1
0	0	1	0
1	3	n_{22}	n_{23}
2	0	1	0

- ¿Son independientes las variables estadísticamente?
- Rectas de regresión de Y/X e X/Y
- ¿Qué parte de la varianza calculada Y es explicada por la regresión? ¿Qué parte es debida a causas ajenas?

Solución:

a)

X \ Y	- 2	0	1	$n_{i\cdot}$
0	0	1	0	1
1	3	n_{22}	n_{23}	$3 + n_{22} + n_{23}$
2	0	1	0	1
$n_{\cdot j}$	3	$2 + n_{22}$	n_{23}	$5 + n_{22} + n_{23} = 11$

$$\text{De otra parte, } \bar{Y} = \frac{-2 \cdot 3 + 0 + n_{23}}{11} = 0 \quad \mapsto \quad n_{23} = 6$$

$$5 + n_{22} + 6 = 11 \quad \mapsto \quad n_{22} = 0$$

X \ Y	- 2	0	1	$n_{i\cdot}$
0	0	1	0	1
1	3	0	6	9
2	0	1	0	1
$n_{\cdot j}$	3	2	6	11

Las variables X e Y son independientes
cuando se verifica $\frac{n_{ij}}{N} = \left(\frac{n_{i\cdot}}{N}\right) \left(\frac{n_{\cdot j}}{N}\right) \quad \forall i, j$

$$\text{No son independientes porque no se verifica la relación: } \frac{1}{11} \neq \frac{1}{11} \times \frac{2}{11} \quad \left[\frac{n_{12}}{N} \neq \left(\frac{n_{1\cdot}}{N}\right) \left(\frac{n_{\cdot 2}}{N}\right) \right]$$

b)

$$a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 x_i y_j n_{ij}}{N} = \frac{1}{11} [-2 \cdot 1 \cdot 3 + 1 \cdot 1 \cdot 6] = 0$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^3 x_i n_{i\cdot}}{N} = \frac{1 \cdot 9 + 0 + 2 \cdot 1}{11} = 1 \quad a_{20} = \frac{\sum_{i=1}^3 x_i^2 n_{i\cdot}}{N} = \frac{1}{11} [1^2 \cdot 9 + 2^2 \cdot 1] = \frac{13}{11}$$

$$s_x^2 = a_{20} - a_{10}^2 = \frac{13}{11} - 1 = \frac{2}{11} \mapsto s_x = \sqrt{\frac{2}{11}} = 0,43$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^3 y_j n_{\cdot j}}{N} = 0$$

$$a_{02} = \frac{\sum_{j=1}^3 y_j^2 n_{\cdot j}}{N} = \frac{1}{11} [(-2)^2 \cdot 3 + 1^2 \cdot 6] = \frac{18}{11}$$

$$s_y^2 = a_{02} - a_{01}^2 = \frac{18}{11} - 0 = \frac{18}{11} \mapsto s_y = \sqrt{\frac{18}{11}} = 1,28$$

$$\text{covarianza: } s_{xy} = a_{11} - a_{10} \cdot a_{01} = 0 - 1 \cdot 0 = 0$$

- El coeficiente de regresión de Y sobre X (pendiente de la recta): $b = \frac{s_{xy}}{s_x^2} = \frac{0}{2/11} = 0$

$$\bar{Y} = a + b \bar{X} \mapsto 0 = a + 0 \cdot 1 \mapsto a = 0$$

$$Y/X: Y=0$$

- El coeficiente de regresión de X sobre Y (pendiente de la recta): $b' = \frac{s_{xy}}{s_y^2} = \frac{0}{18/11} = 0$

$$\bar{X} = a' + b' \bar{Y} \mapsto 1 = a' + 0 \cdot 0 \mapsto a' = 1$$

$$X/Y: X=1$$

COEFICIENTE DETERMINACIÓN: $r^2 = b \cdot b' = 0 \mapsto$ Las rectas son perpendiculares, y en consecuencia, las variables (X, Y) son INCORRELADAS

$$\text{VARIANZA RESIDUAL DE Y: } s_{ry}^2 = s_y^2 (1 - r^2) \mapsto s_{ry}^2 = \frac{18}{11} (1 - 0) = \frac{18}{11}$$

$$s_y^2 = s_{y \text{ explicada}}^2 + s_{ry}^2 \mapsto \frac{18}{11} = s_{y \text{ explicada}}^2 + \frac{18}{11} \mapsto s_{y \text{ explicada}}^2 = 0$$

9. La variable X tiene $\bar{x}=4$ y $s_x^2=1$. Determinar el coeficiente de variación de Pearson de las variables:

$$W = \frac{(X-3)}{2} \quad , \quad Z = \frac{(X-2)}{3}$$

Solución:

$$\begin{cases} E(W) = E\left[\frac{-3}{2} + \frac{1}{2}X\right] = \frac{-3}{2} + \frac{1}{2} \cdot E(X) \quad \mapsto \quad \bar{w} = \frac{-3}{2} + \frac{1}{2} \cdot \bar{x} = \frac{1}{2} \\ \text{Var}(W) = \text{Var}\left[\frac{-3}{2} + \frac{1}{2}X\right] = \left[\frac{1}{2}\right]^2 \cdot \text{Var}(X) = \frac{1}{4} \cdot s_x^2 \quad \mapsto \quad s_w = \frac{1}{2} \cdot s_x = \frac{1}{2} \end{cases}$$

$$C.V_w = \frac{s_w}{\bar{w}} = \frac{1/2}{1/2} = 1$$

$$\begin{cases} E(Z) = E\left[\frac{-2}{3} + \frac{1}{3}X\right] = \frac{-2}{3} + \frac{1}{3} \cdot E(X) \quad \mapsto \quad \bar{z} = \frac{-2}{3} + \frac{1}{3} \cdot \bar{x} = \frac{2}{3} \\ \text{Var}(Z) = \text{Var}\left[\frac{-2}{3} + \frac{1}{3}X\right] = \left[\frac{1}{3}\right]^2 \cdot \text{Var}(X) = \frac{1}{9} \cdot s_x^2 \quad \mapsto \quad s_z = \frac{1}{3} \cdot s_x = \frac{1}{3} \end{cases}$$

$$C.V_z = \frac{s_z}{\bar{z}} = \frac{1/3}{2/3} = \frac{1}{2}$$

COEFICIENTE DE VARIACIÓN DE PEARSON : CAMBIO DE ORIGEN Y DE ESCALA

$$E(Y) = E(a + b \cdot X) = a + b \cdot E(X) = a + b \cdot \bar{X}$$

$$\text{Var}(Y) = \text{Var}[a + b \cdot X] = b^2 \cdot \text{Var}(X) = b^2 \cdot s_x^2 \quad \mapsto \quad s_y = b \cdot s_x$$

$$C.V_y = \frac{b \cdot s_x}{a + b \cdot \bar{X}} = \frac{s_x}{\frac{a}{b} + \bar{X}}$$

El Coeficiente de Variación de Pearson se encuentra afectado ante un cambio de origen.

10. Si $s_y > s_x$ y $r > 0$ ¿La recta de regresión Y/X tiene mayor pendiente que la de X/Y?

Solución:

RELACIÓN ENTRE LOS COEFICIENTES DE REGRESIÓN Y CORRELACIÓN
--

$$\left. \begin{aligned} b &= \frac{s_{xy}}{s_x^2} \mapsto s_{xy} = b \cdot s_x^2 \\ b' &= \frac{s_{xy}}{s_y^2} \mapsto s_{xy} = b' \cdot s_y^2 \\ r &= \frac{s_{xy}}{s_x \cdot s_y} \mapsto s_{xy} = r \cdot s_x \cdot s_y \end{aligned} \right\} \mapsto \begin{aligned} b \cdot s_x^2 &= r \cdot s_x \cdot s_y \mapsto b = r \cdot \frac{s_y}{s_x} \\ b' \cdot s_y^2 &= r \cdot s_x \cdot s_y \mapsto b' = r \cdot \frac{s_x}{s_y} \end{aligned}$$

$$\text{Si } s_y > s_x, r > 0 \mapsto r \cdot \frac{s_y}{s_x} > r \cdot \frac{s_x}{s_y} \Rightarrow b > b'$$

11. Sean dos variables X e Y, tipificadas e incorreladas. Escribir la recta de regresión de Y sobre X

Solución:

Por ser (X, Y) variables tipificadas: $\begin{cases} \bar{x} = 0 & s_x = 1 \\ \bar{y} = 0 & s_y = 1 \end{cases}$

Por ser (X, Y) variables incorreladas: $s_{xy} = 0 \mapsto \begin{cases} b = 0 \\ b' = 0 \end{cases} \mapsto r^2 = 0$

Y/X: $y = a + b \cdot x \mapsto \begin{cases} \bar{y} = a + b \cdot \bar{x} \mapsto a = \bar{y} = 0 \\ b = 0 \end{cases}$

Y/X: $y = 0$

12. En una regresión lineal las varianza explicada por la regresión y residual son iguales. ¿Cuánto vale el coeficiente de determinación?.

Solución:

$$s_y^2 = s_{ry}^2 + s_{Ry}^2 = 2s_{ry}^2$$

$$r^2 = 1 - \frac{s_{Ry}^2}{s_y^2} \rightarrow r^2 = 1 - \frac{s_{ry}^2}{2s_{ry}^2} = 1 - \frac{1}{2} = \frac{1}{2}$$

Sea \hat{y}_i el valor teórico que correspondería a la recta de regresión de Y sobre X [$\hat{y}_i = a + b \cdot x_i$], elevando al cuadrado la descomposición $(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$:

$$\sum (y_i - \bar{y})^2 = \sum [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 + \underbrace{2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}_{=0}$$

se observa que,

$$\begin{aligned} \sum (y_i - \hat{y}_i) \cdot (\hat{y}_i - \bar{y}) &= \sum (y_i - a - bx_i) \cdot (a + bx_i - \bar{y}) = \\ &= a \underbrace{\sum (y_i - a - bx_i)}_{=0} + b \underbrace{\sum x_i (y_i - a - bx_i)}_{=0} + \bar{y} \underbrace{\sum (y_i - a - bx_i)}_{=0} \end{aligned}$$

$$\begin{array}{l} \text{suma cuadrados total} \quad \text{suma cuadrados residual} \quad \text{suma cuadrados explicada} \\ \hline \sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 \end{array}$$

$$\text{Dividiendo por N: } \underbrace{\frac{\sum (y_i - \bar{y})^2}{N}}_{s_y^2} = \underbrace{\frac{\sum (y_i - \hat{y}_i)^2}{N}}_{s_{ry}^2} + \underbrace{\frac{\sum (\hat{y}_i - \bar{y})^2}{N}}_{s_{Ry}^2} \quad [s_y^2 = s_{ry}^2 + s_{Ry}^2]$$

Dividiendo la expresión $[s_y^2 = s_{ry}^2 + s_{Ry}^2]$ por s_y^2 :

$$s_y^2 = s_{ry}^2 + s_{Ry}^2 \rightarrow 1 = \left(\frac{s_{ry}^2}{s_y^2} \right) + \left(\frac{s_{Ry}^2}{s_y^2} \right) \rightarrow \begin{cases} s_{ry}^2 = s_y^2 (1 - r^2) & s_{Ry}^2 = s_y^2 \cdot r^2 \\ r^2 = 1 - \frac{s_{ry}^2}{s_y^2} \end{cases}$$

13. Determinar si son coherentes los datos:

a) $N = 100$, $\bar{x} = 5$, $\bar{y} = 8$, $s_x^2 = 12,5$, $s_y^2 = 70$, $r^2 = 0,9$

b) La suma de residuos al cuadrado correspondientes a una de las posibles rectas de regresión vale 100

Solución:

Solo son útiles: $N = 100$, $s_x^2 = 12,5$, $s_y^2 = 70$, $r^2 = 0,9$, $\sum (y_i - \hat{y}_i)^2 = 100$

$$s_{ry}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{N} = \frac{100}{100} = 1 \quad \text{ó} \quad s_{rx}^2 = \frac{\sum (x_i - \hat{x}_i)^2}{N} = \frac{100}{100} = 1$$

De otra parte,

$$Y/X: s_{ry}^2 = s_y^2 (1 - r^2) \rightarrow s_{ry}^2 = 70 (1 - 0,9) = 7 \neq 1 \quad \text{No son coherentes.}$$

$$X/Y: s_{rx}^2 = s_x^2 (1 - r^2) \rightarrow s_{rx}^2 = 12,5 (1 - 0,9) = 1,25 \neq 1 \quad \text{No son coherentes.}$$

14. Dada la siguiente distribución:

x_i	5	10	15	20	25
n_i	3	7	5	3	5

- Calcular la media armónica, geométrica y aritmética
- Calcular la varianza, desviación típica y coeficiente de variación de Pearson
- Hallar la media aritmética y la desviación típica de la variable X tipificada
- Mediante la transformación $y = \frac{x-15}{5}$, hallar la media, varianza y desviación típica

Solución:

a)

x_i	5	10	15	20	25	
n_i	3	7	5	3	5	23
$x_i n_i$	15	70	75	60	125	345
$x_i^{n_i}$	125	10000000	759375	8000	9765625	$7,415771484 \cdot 10^{25}$
$\frac{n_i}{x_i}$	0,6	0,7	0,3333	0,15	0,2	1,9833
$x_i^2 n_i$	75	700	1125	1200	3125	6225

$$\bar{x}_A = \frac{N}{\sum_{i=1}^5 \frac{n_i}{x_i}} = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \frac{n_4}{x_4} + \frac{n_5}{x_5}} = \frac{23}{1,9833} = 11,597$$

$$\bar{x}_G = \sqrt[23]{\prod_{i=1}^5 x_i^{n_i}} = \sqrt[23]{x_1^{n_1} \cdot x_2^{n_2} \cdot x_3^{n_3} \cdot x_4^{n_4} \cdot x_5^{n_5}} = \sqrt[23]{7,415771484 \cdot 10^{25}} = 13,329$$

$$a_1 = \bar{x} = \frac{\sum_{i=1}^5 x_i n_i}{N} = \frac{345}{23} = 15$$

La relación entre las diferentes medias es: $\bar{x}_A \leq \bar{x}_G \leq \bar{x}$

$$b) a_2 = \frac{\sum_{i=1}^5 x_i^2 n_i}{N} = \frac{6225}{23} = 270,652 \rightarrow s_x^2 = a_2 - a_1^2 = 270,652 - 15^2 = 45,652$$

$$s_x = \sqrt{45,652} = 6,76$$

$$CV_x = \frac{s_x}{\bar{x}} = \frac{6,76}{15} = 0,45 \quad [45\% \text{ de dispersión de los datos}]$$

c) La variable X tipifica: $z_i = \frac{x_i - \bar{x}}{s_x}$

x_i	5	10	15	20	25	
n_i	3	7	5	3	5	23
$z_i = \frac{x_i - \bar{x}}{s_x}$	-1,479	-0,740	0	0,740	1,479	
$z_i n_i$	-4,438	-5,178	0	2,219	7,396	0
$z_i^2 n_i$	6,565	3,830	0	1,641	10,941	23
y_i	-2	-1	0	1	2	
$y_i n_i$	-6	-7	0	3	10	0
$y_i^2 n_i$	12	7	0	3	20	42

$$\bar{z} = \frac{\sum_{i=1}^5 z_i n_i}{N} = \frac{0}{23} = 0 \quad s_z^2 = \frac{\sum_{i=1}^5 z_i^2 n_i}{N} - (\bar{z})^2 = \frac{23}{23} - 0 = 1 \quad \left\{ \begin{array}{l} \text{Toda variable tipificada tiene} \\ \text{media} = 0 \text{ y varianza} = 1 \end{array} \right.$$

d) Con la transformación $y = \frac{x - 15}{5}$

$$\bar{y} = \frac{\sum_{i=1}^5 y_i n_i}{N} = \frac{0}{23} = 0 \quad s_y^2 = \frac{\sum_{i=1}^5 y_i^2 n_i}{N} - (\bar{y})^2 = \frac{42}{23} - 0 = \frac{42}{23} = 1,826 \quad s_y = \sqrt{\frac{42}{23}} = 1,35$$

No son necesarios los cálculos, se conoce:

$$\left\{ \begin{array}{l} \bar{y} = E \left[\frac{x - 15}{5} \right] = \frac{1}{5} E[x] - 3 = -3 + \frac{1}{5} \bar{x} = -3 + \frac{15}{5} = 0 \\ s_y^2 = \text{Var} \left[\frac{x - 15}{5} \right] = \frac{1}{5^2} \text{Var}[x] = \frac{1}{25} s_x^2 = \frac{45,652}{25} = 1,826 \end{array} \right.$$

15. Ana acude con su hijo a la consulta de un odontólogo para cuatro restauraciones dentarais, observando que el doctor aplicaba cantidades de cemento de ionómeros de vidrio con flúor y composite (Y, en gramos) conforme a los diámetros de perforación de cada pieza dental (X, en milímetros) como se refleja a continuación:

X \ Y	0 - 1	1 - 3	3 - 6	6 - 10
0 - 3		1		
3 - 5	1		1	
5 - 10				1

Se pide:

- ¿Son independientes estadísticamente ambas variables?. Razone la respuesta.
- Calcule las rectas de regresión de Y/X e X/Y. Interpretar los resultados.
- ¿Qué parte de la varianza de las perforaciones habidas (X) es explicada por la cantidad de ionómeros de vidrio consumida (Y)? ¿Qué parte no es explicada?.

Solución:

- a) Las variables X e Y son independientes cuando se verifica $\frac{n_{ij}}{N} = \left(\frac{n_{i\cdot}}{N}\right) \left(\frac{n_{\cdot j}}{N}\right) \forall i, j$

X \ Y	0,5	2	4,5	8	$n_{i\cdot}$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$
1,5		1			1	1,5	2,25
4	1		1		2	8	32
7,5				1	1	7,5	56,25
$n_{\cdot j}$	1	1	1	1	4	17	90,5
$y_j n_{\cdot j}$	0,50	2	4,50	8	15		
$y_j^2 n_{\cdot j}$	0,25	4	20,25	64	88,5		

Las variables no son independientes: $\frac{n_{12}}{N} \left(\frac{1}{4}\right) \neq \left(\frac{n_{1\cdot}}{N}\right) \left(\frac{n_{\cdot 2}}{N}\right) \equiv \left(\frac{1}{4}\right) \left(\frac{1}{4}\right)$

$$b) a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^4 x_i y_j n_{ij}}{N} = \frac{1}{4} [1,5 \cdot 2 \cdot 1 + 4 \cdot 0,5 \cdot 1 + 4 \cdot 4,5 \cdot 1 + 7,5 \cdot 8 \cdot 1] = 20,75$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^3 x_i n_{i\cdot}}{N} = \frac{17}{4} = 4,25 \quad a_{20} = \frac{\sum_{i=1}^3 x_i^2 n_{i\cdot}}{N} = \frac{90,5}{4} = 22,625$$

$$s_x^2 = a_{20} - a_{10}^2 = 22,625 - 4,25^2 = 4,5625 \quad \mapsto \quad s_x = \sqrt{4,5625} = 2,136$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^4 y_j n_{\cdot j}}{N} = \frac{15}{4} = 3,75 \quad a_{02} = \frac{\sum_{j=1}^4 y_j^2 n_{\cdot j}}{N} = \frac{88,5}{4} = 22,125$$

$$s_y^2 = a_{02} - a_{01}^2 = 22,125 - 3,75^2 = 8,0625 \mapsto s_y = \sqrt{8,0625} = 2,84$$

$$\text{covarianza: } s_{xy} = a_{11} - a_{10} \cdot a_{01} = 20,75 - 4,25 \cdot 3,75 = 4,8125$$

- El coeficiente de regresión de Y sobre X (pendiente de la recta): $b = \frac{s_{xy}}{s_x^2} = \frac{4,8125}{4,5625} = 1,055$

$$\bar{Y} = a + b \bar{X} \mapsto 3,75 = a + 1,055 \cdot 4,25 \mapsto a = -0,734$$

$$Y/X: Y = -0,734 + 1,055 X$$

- El coeficiente de regresión de X sobre Y (pendiente de la recta): $b' = \frac{s_{xy}}{s_y^2} = \frac{4,8125}{8,0625} = 0,597$

$$\bar{X} = a' + b' \bar{Y} \mapsto 4,25 = a' + 0,597 \cdot 3,75 \mapsto a' = 2,011$$

$$X/Y: X = 2,011 + 0,597 Y$$

c) COEFICIENTE DETERMINACIÓN: $r^2 = b \cdot b' = 1,055 \cdot 0,597 = 0,6298$

$$\text{VARIANZA RESIDUAL DE X: } s_{rx}^2 = s_x^2 (1 - r^2) \mapsto s_{rx}^2 = 4,5625 (1 - 0,6298) = 1,689 \text{ NO EXPLICADA}$$

$$s_x^2 = \underbrace{s_{rx}^2}_{\text{varianza residual no explicada}} + \underbrace{s_{R_x}^2}_{\text{varianza regresión explicada}} \rightarrow s_{R_x}^2 = s_x^2 - s_{rx}^2 \rightarrow s_{R_x}^2 = 8,0625 - 1,689 = 6,3735 \text{ EXPLICADA}$$

16. El salario medio mensual en cientos de euros de 160 obreros se distribuye de la siguiente forma:

Intervalos	4 - 8	8 - 12	12- 16	16 - 20	20 - 24	24 - 28	28 - 32	32 - 36
n_i	3	12	40	47	32	13	9	4

- Media aritmética, mediana, moda y percentil 75.
- Coeficiente de asimetría de Fisher.
- Realizar una redistribución en la que los intervalos tengan una amplitud de 8, y con estos nuevos intervalos calcular la media aritmética y el coeficiente de variación de Pearson. Comparar los resultados obtenidos en el apartado (a)

Solución:

a)

Intervalos	4 - 8	8 - 12	12- 16	16 - 20	20 - 24	24 - 28	28 - 32	32 - 36	
x_i	6	10	14	18	22	26	30	34	
n_i	3	12	40	47	32	13	9	4	160
N_i	3	15	55	102	134	147	156	160	
$h_i = n_i / c_i$	0,75	3	10	11,75	8	3,25	2,25	1	40
$x_i \cdot n_i$	18	120	560	846	704	338	270	136	2992
$(x_i - \bar{x})$	-12,7	-8,7	-4,7	-0,7	3,3	7,3	11,3	15,3	
$(x_i - \bar{x}) n_i$	-38,1	-104,4	-188	-32,9	105,6	94,9	101,7	61,2	0
$(x_i - \bar{x})^2 n_i$	483,87	908,28	883,6	23,03	348,48	692,77	1149,21	936,36	5425,60
$(x_i - \bar{x})^3 n_i$	-6145,149	-7902,036	-4152,92	-16,121	1149,984	5057,221	12986,073	14326,308	15303,36

$$a_1 = \bar{x} = \frac{\sum_{i=1}^8 x_i \cdot n_i}{N} = \frac{2992}{160} = 18,7$$

$$M_e = L_i + \frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} c_i \rightarrow M_e = 16 + \frac{80 - 55}{102 - 55} \cdot 4 = 18,13$$

$$M_d = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} c_i \rightarrow M_d = 16 + \frac{11,75 - 10}{(11,75 - 10) + (11,75 - 8)} \cdot 4 = 17,27$$

Se verifica la relación $\bar{x} \geq M_e \geq M_d \rightarrow$ Distribución asimétrica a la derecha o positiva

Adviértase que para calcular la moda, cuando la amplitud de los intervalos es igual, para trabajar con una escala más pequeña, se puede emplear la expresión:

$$M_d = L_i + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} c_i \rightarrow M_d = 16 + \frac{47 - 40}{(47 - 40) + (47 - 32)} \cdot 4 = 17,27$$

$$P_{75} = L_i + \frac{\frac{75}{100} N - N_{i-1}}{N_i - N_{i-1}} c_i \rightarrow Q_3 = P_{75} = 20 + \frac{120 - 102}{134 - 102} \cdot 4 = 22,25$$

b) Coeficiente de asimetría de Fisher: $g_1 = \frac{m_3}{s^3} \begin{cases} g_1 > 0 & \text{Asimetría a la derecha o positiva} \\ g_1 = 0 & \text{Simetría} \\ g_1 < 0 & \text{Asimetría a la izquierda o negativa} \end{cases}$

$$m_2 = s^2 = \frac{\sum_{i=1}^8 (x_i - \bar{x})^2 \cdot n_i}{N} = \frac{5425,60}{160} = 33,91 \text{ (varianza)} \quad s = \sqrt{33,91} = 5,82 \text{ (desviación típica)}$$

$$m_3 = \frac{\sum_{i=1}^8 (x_i - \bar{x})^3 \cdot n_i}{N} = \frac{15303,36}{160} = 95,65$$

$$g_1 = \frac{m_3}{s^3} = \frac{95,65}{5,82^3} = 0,485 > 0 \Rightarrow \text{Distribución asimétrica a la derecha o positiva.}$$

c)

Intervalos	4 - 12	12 - 20	20 - 28	28 - 36	
x_i	8	16	24	32	
n_i	15	87	45	13	160
$x_i \cdot n_i$	120	1392	1080	416	3008
$x_i^2 \cdot n_i$	960	22272	25920	13312	62464

$$a_1 = \bar{x} = \frac{\sum_{i=1}^4 x_i \cdot n_i}{N} = \frac{3008}{160} = 18,8 \quad a_2 = \frac{\sum_{i=1}^4 x_i^2 \cdot n_i}{N} = \frac{62464}{160} = 390,4 \quad s_x^2 = a_2 - a_1^2 = 390,4 - 18,8^2 = 36,96$$

$$CV = \frac{s_x}{\bar{x}} = \frac{\sqrt{36,96}}{18,8} = 0,32 \text{ (32\% de dispersión de los datos)}$$

La media aritmética cambia, se ha transformado la distribución de datos.

17. La distribución de salarios de una empresa es la siguiente:

Salario (euros)	Empleados
3000 - 5000	25
1000 - 2000	100
5000 - 9000	5
2000 - 3000	50

- Estudiar la concentración de salarios
- ¿Qué porcentaje de empleados percibe el 50% de los salarios?
- La empresa como política comercial analiza subir los salarios a todos los empleados, con un incremento del 10%, o bien con un aumento de 200 euros por empleado. ¿Cuál de las dos opciones sería más equitativa?
- ¿Cuál es la concentración de salarios si el número de empleados hubiera sido el doble?

Solución:

- La concentración de salarios se analiza mediante el Índice de Gini, que no varía mediante cambios de escala (subida porcentual del 10% a los empleados) mientras que queda modificado con cambios de origen (subida lineal de 200 euros a cada empleado).

Ordenando los salarios en forma creciente:

Salarios	x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\%p_i = \frac{N_i}{N} \cdot 100$	$\%q_i = \frac{u_i}{u_k} \cdot 100$
1000 - 2000	1500	100	100	150000	150000	55,56	36,59
2000 - 3000	2500	50	150	125000	275000	83,33	67,07
3000 - 5000	4000	25	175	100000	375000	97,22	91,46
5000 - 9000	7000	5	180	35000	410000	100	100
				410000		236,11	195,12

$$I_G = 1 - \frac{\sum_{i=1}^3 q_i}{\sum_{i=1}^3 p_i} = 1 - \frac{195,12}{236,11} = 0,174 \quad (\text{concentración de salarios del 17,4\%})$$

- En la tabla se observa que el 55,56% de los empleados percibe el 36,59% de los salarios, y el 83,33% de los empleados percibe el 67,07% de los salarios. En consecuencia, el 50% de los salarios estará distribuido entre un conjunto de empleados situado entre el 55,56 y el 83,33%.

Bajo la hipótesis de linealidad, se establece la relación de porcentajes:

$$\frac{67,07 - 36,59}{83,33 - 55,56} = \frac{50 - 36,59}{x - 55,56} \quad \mapsto \quad \frac{30,48}{27,77} = \frac{13,41}{x - 55,56} \quad \mapsto \quad x - 55,56 = \frac{27,77 \cdot 13,41}{30,48} \quad \Rightarrow \quad x = 67,78\%$$

c)

SUBIDA DE SALARIOS DEL 10% - *Cambio de escala en los salarios*

$x'_i = 1,1 \cdot x_i$	n_i	N_i	$x'_i n_i$	$u'_i = x'_i n_i$ acumulada	$\%p_i = \frac{N_i}{N} \cdot 100$	$\%q'_i = \frac{u'_i}{u_k} \cdot 100$
1650	100	100	165000	165000	55,56	36,59
2750	50	150	137500	302500	83,33	67,07
4400	25	175	110000	412500	97,22	91,46
7700	5	180	38500	451000	100	100
			451000		236,11	195,12

$$I_G = 1 - \frac{\sum_{i=1}^3 q_i}{\sum_{i=1}^3 p_i} = 1 - \frac{195,12}{236,11} = 0,174 \quad (\text{concentración de salarios del 17,4\%})$$

Adviértase que: $q'_i = \frac{u_i \cdot 1,1}{u_k \cdot 1,1} = \frac{u_i}{u_k} = q_i$

Con una subida del 10% a cada empleado, la equidistribución no varía.

El cambio de escala en los salarios no afecta al Índice de Gini, propiedad conocida como *Principio de la Renta relativa*.

SUBIDA LINEAL DE SALARIOS DE 200 EUROS - *Cambio de origen en los salarios*

$x'_i = 200 + x_i$	n_i	N_i	$x'_i n_i$	$u'_i = x'_i n_i$ acumulada	$\%p_i = \frac{N_i}{N} \cdot 100$	$\%q'_i = \frac{u'_i}{u_k} \cdot 100$
1700	100	100	170000	170000	55,56	38,12
2700	50	150	135000	305000	83,33	68,39
4200	25	175	105000	410000	97,22	91,93
7200	5	180	36000	446000	100	100
			446000		236,11	198,43

$$I_G = 1 - \frac{\sum_{i=1}^3 q_i}{\sum_{i=1}^3 p_i} = 1 - \frac{198,43}{236,11} = 0,16 \quad (\text{concentración de salarios del 16\%})$$

Con una subida lineal de 200 euros a cada empleado, la equidistribución de salarios es más equitativa.

Si por el contrario la empresa hubiera rebajado 50 euros a cada empleado, la equidistribución de salarios sería menos equitativa.

El cambio de origen en los salarios afecta al Índice de Gini, propiedad conocida como *Principio de Dalton*.

d) Concentración salarios si el número de empleados hubiera sido el doble:

SUBIDA LINEAL DE EMPLEADOS - Cambio de escala en la Población

x_i	$n'_i = 2 n_i$	N'_i	$x_i n'_i$	$u'_i = x_i n'_i$ acumulada	$\%p'_i = \frac{N'_i}{N} \cdot 100$	$\%q'_i = \frac{u'_i}{u_k} \cdot 100$
1500	200	200	300000	300000	55,56	36,59
2500	100	300	250000	550000	83,33	67,07
4000	50	350	200000	750000	97,22	91,46
7000	10	360	70000	820000	100,00	100,00
360			820000		236,11	195,12

$$I_G = 1 - \frac{\sum_{i=1}^3 q_i}{\sum_{i=1}^3 p_i} = 1 - \frac{195,12}{236,11} = 0,174 \quad (\text{concentración de salarios del 17,4\%})$$

El cambio de escala en la población no afecta al Índice de Gini, propiedad conocida como *Principio de la Población*. Es decir, el tamaño de la población no importa, lo que interesa son las proporciones de individuos de la población que perciben diferentes niveles de salario.

18. Dada la tabla de correlación:

X \ Y	- 1	0	1
- 1	2	1	2
0	2	4	2
1	1	0	1

Estudiar la independencia estadística, calcular las rectas de regresión y la correlación entre ambas variables.

Solución:

a)

X \ Y	- 1	0	1	$n_{i\bullet}$	$x_i \cdot n_{i\bullet}$	$x_i^2 \cdot n_{i\bullet}$
- 1	2	1	2	5	-5	5
0	2	4	2	8	0	0
1	1	0	1	2	2	2
$n_{\bullet j}$	5	5	5	N=15	-3	7
$y_j \cdot n_{\bullet j}$	-5	0	5	0		
$y_j^2 \cdot n_{\bullet j}$	5	0	5	10		

Las variables X e Y son independientes cuando se verifica $\frac{n_{ij}}{N} = \left(\frac{n_{i\bullet}}{N} \right) \left(\frac{n_{\bullet j}}{N} \right) \forall i, j$

Si alguna de las frecuencias absolutas es igual a 0 no son independientes estadísticamente:

$$\frac{n_{32}}{N} \neq \frac{n_{3\bullet}}{N} \cdot \frac{n_{\bullet 2}}{N} \equiv \frac{0}{15} \neq \frac{2}{15} \cdot \frac{5}{15}$$

$$a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 x_i \cdot y_j \cdot n_{ij}}{N} = \frac{1}{15} [-1 \cdot (-1) \cdot 2 - 1 \cdot 1 \cdot 2 + 1 \cdot (-1) \cdot 1 + 1 \cdot 1 \cdot 1] = 0$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^3 x_i \cdot n_{i\bullet}}{N} = \frac{-3}{15} = -\frac{1}{5} \quad a_{20} = \frac{\sum_{i=1}^3 x_i^2 \cdot n_{i\bullet}}{N} = \frac{7}{15} \quad s_x^2 = a_{20} - a_{10}^2 = \frac{7}{15} - \left(\frac{-1}{5}\right)^2 = \frac{6}{25}$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^3 y_j \cdot n_{\bullet j}}{N} = 0 \quad a_{02} = \frac{\sum_{j=1}^3 y_j^2 \cdot n_{\bullet j}}{N} = \frac{10}{15} \quad s_y^2 = a_{02} - a_{01}^2 = \frac{10}{15} - 0 = \frac{10}{15}$$

$$s_{xy} = a_{11} - a_{10} \cdot a_{01} = 0 - \left(\frac{-1}{5}\right) \cdot 0 = 0 \rightarrow \begin{cases} b=0 \\ b'=0 \end{cases} \rightarrow r^2 = 0 \rightarrow \begin{cases} \text{Rectas regresión perpendiculares} \\ \text{variables INCORRELADAS} \end{cases}$$

Rectas de regresión:

$$Y/X: y = a + b \cdot x \mapsto \begin{cases} \bar{y} = a + b \cdot \bar{x} \\ b=0 \end{cases} \mapsto a = \bar{y} = 0 \rightarrow y = 0$$

$$X/Y: x = a' + b' \cdot y \mapsto \begin{cases} \bar{x} = a' + b' \cdot \bar{y} \\ b'=0 \end{cases} \mapsto a' = \bar{x} = \frac{-1}{5} = -0,2 \rightarrow x = -0,2$$

19. La variable estadística X tiene $\bar{x} = 2$, $s_x = 1$. Determinar la media aritmética, la varianza y el coeficiente de variación de Pearson de $Y = \frac{X-1}{2}$

Solución:

$$\begin{cases} E(Y) = E\left[\frac{-1}{2} + \frac{1}{2}X\right] = \frac{-1}{2} + \frac{1}{2} \cdot E(X) \mapsto \bar{y} = \frac{-1}{2} + \frac{1}{2} \cdot \bar{x} = \frac{1}{2} \\ \text{Var}(Y) = \text{Var}\left[\frac{-1}{2} + \frac{1}{2}X\right] = \left[\frac{1}{2}\right]^2 \cdot \text{Var}(X) = \frac{1}{4} \cdot s_x^2 \mapsto s_y = \frac{1}{2} \cdot s_x = \frac{1}{2} \end{cases}$$

$$C.V_y = \frac{s_y}{\bar{y}} = \frac{1/2}{1/2} = 1$$

20. La varianza explicada por una regresión lineal simple es el doble de la varianza residual, ¿Cuánto vale el coeficiente de determinación?

Solución:

$$s_{RY}^2 = 2s_{ry}^2 \rightarrow s_y^2 = s_{ry}^2 + s_{RY}^2 = 3s_{ry}^2$$

$$s_{ry}^2 = s_y^2 (1 - r^2) \rightarrow r^2 = 1 - \frac{s_{ry}^2}{s_y^2} = 1 - \frac{s_{ry}^2}{3s_{ry}^2} = 1 - \frac{1}{3} = \frac{2}{3}$$

21. Dada la distribución:

x_i	2	4	8	10
n_i	3	4	1	2

a) Calcula los coeficientes de asimetría de Pearson y de Fisher, coeficiente de curtosis.

b) Siendo la variable $X = \frac{Y-1}{2}$, halla los coeficientes de asimetría de Pearson y Fisher de la variable Y

c) ¿Tienen el mismo coeficiente de Variación de Pearson las dos variables?

c) Calcula el coeficiente de curtosis de las variables X e Y

Solución:

a)

x_i	n_i	N_i	$x_i \cdot n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$	$(x_i - \bar{x})^3 \cdot n_i$	$(x_i - \bar{x})^4 \cdot n_i$
2	3	3	6	-3	9	27	-81	243
4	4	7	16	-1	1	4	-4	4
8	1	8	8	3	9	9	27	81
10	2	10	20	5	25	50	250	1250
10			50			90	192	1578

$$\bar{x} = \frac{\sum_{i=1}^4 x_i \cdot n_i}{N} = \frac{50}{10} = 5 \quad M_{ex} = 4 \quad M_{dx} = 4 \quad s_x^2 = \frac{\sum_{i=1}^4 (x_i - \bar{x})^2 \cdot n_i}{N} = \frac{90}{10} = 9 \quad s_x = \sqrt{9} = 3$$

Coeficiente asimetría de Pearson: $A_{px} = \frac{\bar{x} - M_{dx}}{s_x} = \frac{5 - 4}{3} = 0,33 > 0$ asimetría a la derecha o positiva

Coeficiente de asimetría de Fisher: $m_{3x} = \frac{\sum_{i=1}^4 (x_i - \bar{x})^3 \cdot n_i}{N} = \frac{192}{10} = 19,2 \quad s_x^3 = 3^3 = 27$

$g_{1x} = \frac{m_{3x}}{s_x^3} = \frac{19,2}{27} = 0,71 > 0$ asimetría a la derecha o positiva

Coeficiente de curtosis: $m_{4x} = \frac{\sum_{i=1}^4 (x_i - \bar{x})^4 \cdot n_i}{N} = \frac{1578}{10} = 157,8$ $s_x^4 = 3^4 = 81$

$g_{2x} = \frac{m_{4x}}{s_x^4} - 3 = \frac{157,8}{81} - 3 = -1,05 < 0$ menor apuntamiento que la normal (PLATICÚRTICA)

b) $Y = 1 + 2X$

Los coeficientes de asimetría de Pearson y de Fisher son invariantes ante un cambio de origen y de escala y, en consecuencia, la distribución Y presenta:

$A_{py} = 0,33 > 0$ asimetría a la derecha o positiva

$g_{1y} = 0,71 > 0$ asimetría a la derecha o positiva

Haciendo las operaciones:

y_i	n_i	N_i	$y_i \cdot n_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^2 \cdot n_i$	$(y_i - \bar{y})^3 \cdot n_i$	$(y_i - \bar{y})^4 \cdot n_i$
5	3	3	15	-6	36	108	-648	3888
9	4	7	36	-2	4	16	-32	64
17	1	8	17	6	36	36	216	1296
21	2	10	42	10	100	200	2000	20000
10			110			360	1536	25248

$\bar{y} = \frac{\sum_{i=1}^4 y_i \cdot n_i}{N} = \frac{110}{10} = 11$ $M_{ey} = 9$ $M_{dy} = 9$ $s_y^2 = \frac{\sum_{i=1}^4 (y_i - \bar{y})^2 \cdot n_i}{N} = \frac{360}{10} = 36$ $s_y = \sqrt{36} = 6$

Coeficiente asimetría de Pearson: $A_{py} = \frac{\bar{y} - M_{dy}}{s_y} = \frac{11 - 9}{6} = 0,33 > 0$ asimetría a la derecha o positiva

Coeficiente de asimetría de Fisher: $m_{3y} = \frac{\sum_{i=1}^4 (y_i - \bar{y})^3 \cdot n_i}{N} = \frac{1536}{10} = 153,6$ $s_y^3 = 6^3 = 216$

$g_{1y} = \frac{m_{3y}}{s_y^3} = \frac{153,6}{216} = 0,71 > 0$ asimetría a la derecha o positiva

c) El coeficiente de variación de Pearson es invariante ante un cambio de escala ($Y = 2X$) pero no ante un cambio de origen ($Y = 1 + 2X$). En este caso: $CV_y = \frac{2s_x}{1 + 2\bar{x}}$. No tienen, por tanto, el mismo coeficiente de variación.

Coeficiente de variación de Pearson de X: $CV_x = \frac{s_x}{\bar{x}} = \frac{3}{5} = 0,6$ (60% de dispersión de los datos)

Coeficiente de variación de Pearson de Y: $CV_y = \frac{2 \cdot s_x}{1 + \bar{x}} = \frac{6}{11} = 0,54$ (54% de dispersión de los datos)

d) El coeficiente de curtosis o apuntamiento es invariante ante un cambio de origen y de escala ($Y = 1 + 2X$) y, en consecuencia:

$g_{2y} = -1,05 < 0$ menor apuntamiento que la normal (PLATICÚRTICA)

Haciendo operaciones: $m_{4y} = \frac{\sum_{i=1}^4 (y_i - \bar{y})^4 \cdot n_i}{N} = \frac{25248}{10} = 2524,8$ $s_y^4 = 6^4 = 1296$

$g_{2y} = \frac{m_{4y}}{s_y^4} - 3 = \frac{2524,8}{1296} - 3 = -1,05 < 0$ menor apuntamiento que la normal (PLATICÚRTICA)

PARCIALILLO 22 DE FEBRERO 2013

1. Se ha realizado un estudio entre 100 mujeres mayores de 25 años, observándose el número de hijos de las mismas. El resultado ha sido:

Número de hijos (x_i)	Número de mujeres (n_i)
0	13
1	20
2	25
3	20
4	11
5	7
6	4

- Calcular el número medio de hijos, la mediana, la moda y el tercer cuartil
- ¿Cuál es el número máximo de hijos que tiene el 70% de las mujeres que menos hijos tienen?
- Calcular el coeficiente de variación de Pearson
- Calcular el coeficiente de asimetría de Fisher y el coeficiente de curtosis

Solución:

a)

x_i	n_i	N_i	$f_i = \frac{n_i}{N}$	$F_i = \frac{N_i}{N}$	$x_i n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
0	13	13	0,13	0,13	0	-2,33	5,43	70,58	-164,44	383,15
1	20	33	0,20	0,33	20	-1,33	1,77	35,38	-47,05	62,58
2	25	58	0,25	0,58	50	-0,33	0,11	2,72	-0,90	0,30
3	20	78	0,20	0,78	60	0,67	0,45	8,98	6,02	4,03
4	11	89	0,11	0,89	44	1,67	2,79	30,68	51,23	85,56
5	7	96	0,07	0,96	35	2,67	7,13	49,90	133,24	355,75
6	4	100	0,04	1	24	3,67	13,47	53,88	197,72	725,65
	100		1,0		233			252,11	175,82	1617,01

$$\text{Media aritmética: } \bar{x} = \frac{\sum_{i=1}^7 x_i n_i}{N} = \frac{233}{100} = 2,33$$

Mediana: $M_e = 2$ (pasa de la mitad 50%)

$M_d = 2$ ($n_3 = 25$, el más grande)

3º Cuartil $\left[\frac{100 \cdot 3}{4} = 75 \right]$: $Q_3 = 3$ hijos (F_4 pasa del 75%)

- b) El número máximo de hijos que tiene el 70% de las mujeres que menos hijos tienen es el Decil 7 (Percentil 70)

Decil 7 ó Percentil 70: 3 hijos (F_4 pasa de 0,7)

c) Varianza: $m_2 = s^2 = \frac{\sum_{i=1}^7 [x_i - \bar{x}]^2 n_i}{N} = \frac{252,11}{100} = 2,5211 \text{ hijos}^2$

Desviación típica: $s = \sqrt{2,5211} = 1,59 \text{ hijos}$

Coeficiente de Variación de Pearson: $C.V = \frac{s}{\bar{x}} = \frac{1,59}{2,33} = 0,6824$ una dispersión del 68,24%

- d) Coeficiente de asimetría de Fisher:

$$g_1 = \frac{m_3}{s^3} = \frac{\frac{1}{N} \sum_{i=1}^7 (x_i - \bar{x})^3 n_i}{s^3} = \frac{1,76}{1,59^3} = 0,4378 > 0 \rightarrow \text{Asimetría a la derecha o positiva}$$

- Coeficiente de curtosis:

$$g_2 = \frac{m_4}{s^4} - 3 = \frac{\frac{1}{N} \sum_{i=1}^7 (x_i - \bar{x})^4 n_i}{s^4} = \frac{16,17}{1,59^4} - 3 = -0,47 < 0 \rightarrow \text{PLATICÚRTICA}$$

2. Los salarios de los empleados de la cadena de producción de una empresa se distribuyen según la tabla adjunta:

Salarios	10 - 20	20 - 40	40 - 50	50 - 100	100 - 200
Nº empleados	12000	6000	1000	800	200

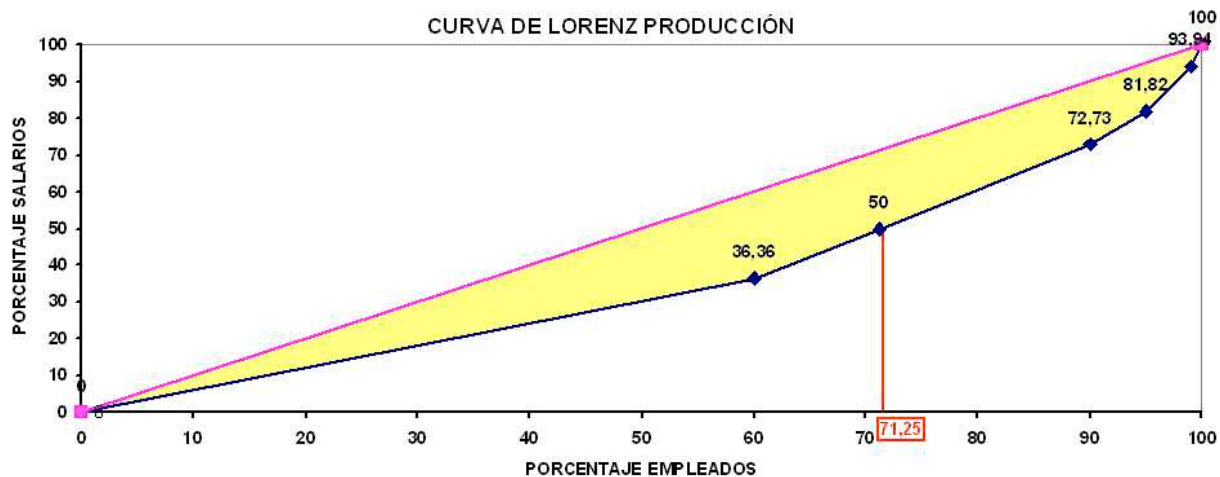
¿Qué porcentaje de empleados que percibe el 50% de los salarios? ¿Es equilibrada la distribución de salarios?

Solución:

Salarios [$L_i - L_{i+1}$)	x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\%p_i = \frac{N_i}{N} \cdot 100$	$\%q_i = \frac{u_i}{u_k} \cdot 100$
10 - 20	15	12000	12000	180000	180000	60	36,36
						x	50
20 - 40	30	6000	18000	180000	360000	90	72,73
40 - 50	45	1000	19000	45000	405000	95	81,82
50 - 100	75	800	19800	60000	465000	99	93,94
100 - 200	150	200	20000	30000	495000	100	100
		$\sum_{i=1}^5 n_i = 20000$	$\sum_{i=1}^5 N_i = 19800$	$\sum_{i=1}^5 x_i n_i = 495000$		$\sum_{i=1}^4 p_i = 99$	$\sum_{i=1}^4 q_i = 93,94$

En la tabla se observa que el 60% de los empleados percibe el 36,36% de los salarios y que el 90% de los empleados percibe el 72,73% de los salarios. Para estimar el porcentaje (x) de empleados que percibe el 50% de los salarios se necesita realizar una interpolación lineal:

$$\frac{x - 60}{50 - 36,36} = \frac{90 - 60}{72,73 - 36,36} \rightarrow \frac{x - 60}{13,64} = \frac{90 - 60}{36,37} \Rightarrow x = 71,25\%$$



$$I_G = 1 - \frac{\sum_{i=1}^4 q_i}{\sum_{i=1}^4 p_i} = 1 - \frac{284,85}{344} = 0,17$$

La concentración es pequeña, pudiendo concluir que la distribución de salarios es equilibrada.

3. Sea la distribución bidimensional, donde las variables X e Y son estadísticamente independientes.

X \ Y	3	4
1	3	c
2	2	6

Se pide:

- Calcular las medias y varianzas marginales.
- Hallar la covarianza y las rectas de regresión.

Solución:

X \ Y	3	4	$n_{i\bullet}$
1	3	c	3 + c
2	2	6	8
$n_{\bullet j}$	5	6 + c	11 + c

Por ser independientes: $\frac{n_{ij}}{N} = \frac{n_{i\bullet}}{N} \cdot \frac{n_{\bullet j}}{N} \quad \forall i, j$

$$\frac{c}{11+c} = \frac{c}{3+c} \times \frac{c}{6+c} \rightarrow (3+c) \cdot (6+c) = c \cdot (11+c) \rightarrow c = \frac{18}{2} = 9$$

MEDIAS Y VARIANZAS MARGINALES:

X \ Y	3	4	n _{i.}
1	3	9	12
2	2	6	8
n _{.j}	5	15	20

MARGINAL DE LA VARIABLE X:

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^2 x_i n_{i.}}{N} = \frac{1}{20} [1 \cdot 12 + 2 \cdot 8] = 1,4$$

$$a_{20} = \frac{\sum_{i=1}^2 x_i^2 n_{i.}}{N} = \frac{1}{20} [1^2 \cdot 12 + 2^2 \cdot 8] = 2,2$$

$$s_x^2 = a_{20} - a_{10}^2 = 2,2 - 1,4^2 = 0,24$$

MARGINAL DE LA VARIABLE Y:

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^2 y_j n_{.j}}{N} = \frac{1}{20} [3 \cdot 5 + 4 \cdot 15] = 3,75$$

$$a_{02} = \frac{\sum_{j=1}^2 y_j^2 n_{.j}}{N} = \frac{1}{20} [3^2 \cdot 5 + 4^2 \cdot 15] = 14,25$$

$$s_y^2 = a_{02} - a_{01}^2 = 14,25 - 3,75^2 = 0,1875$$

b) covarianza: $s_{xy} = a_{11} - a_{10} a_{01}$

X \ Y	X \ Y	3	4	n _{i.}
1	1	3	9	12
2	2	2	6	8
n _{.j}	n _{.j}	5	15	20

$$a_{11} = \frac{\sum_{i=1}^2 \sum_{j=1}^2 x_i y_j n_{ij}}{N} = \frac{1 \cdot 3 \cdot 3 + 1 \cdot 4 \cdot 9 + 2 \cdot 3 \cdot 2 + 2 \cdot 4 \cdot 6}{20} = \frac{105}{20} = 5,25$$

$$s_{xy} = a_{11} - a_{10} a_{01} = 5,25 - 1,4 \cdot 3,75 = 0$$

Sin calcular la covarianza, se conocía que la covarianza $s_{xy} = 0$ por ser (X, Y) variables independientes.

Si (X, Y) independientes $\mapsto s_{xy} = 0$

Si $s_{xy} = 0 \mapsto (X, Y)$ No son independientes

Por otra parte, se conoce que en las rectas de regresión:

$$Y/X: Y = a + bX$$

$$X/Y: X = a' + b'Y$$

Los coeficientes de regresión respectivos (b, b') dependen de la covarianza s_{xy} , dado que vienen

expresados: $b = \frac{s_{xy}}{s_x^2}$, $b' = \frac{s_{xy}}{s_y^2}$.

Si $s_{xy} = 0 \mapsto b = 0, b' = 0$

Con lo cual, las rectas de regresión solicitadas son: $\begin{cases} Y/X: Y = a \\ X/Y: X = a' \end{cases}$

Los coeficientes respectivos (a, a') se calculan teniendo en cuenta:

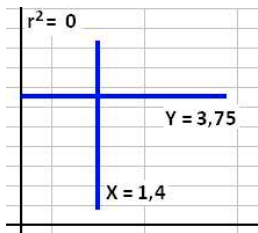
$$Y/X: Y = a + bX \mapsto \bar{Y} = a + b\bar{X} \mapsto 3,75 = a + 0 \times 1,4 \rightarrow a = 3,75 \xrightarrow{\text{recta regresión } Y/X} Y = 3,75$$

$$X/Y: X = a' + b'Y \mapsto \bar{X} = a' + b'\bar{Y} \mapsto 1,4 = a' + 0 \times 3,75 \rightarrow a' = 1,4 \xrightarrow{\text{recta regresión } X/Y} X = 1,4$$

Adviértase que cuando las variables (X, Y) son independientes, la covarianza $s_{xy} = 0$

En consecuencia:

- Las coeficientes de regresión $b = 0, b' = 0$
- La recta de regresión de $Y/X: Y = \bar{Y} = 3,75$
- La recta de regresión de $X/Y: X = \bar{X} = 1,4$
- El coeficiente de determinación $r^2 = b \cdot b' = 0$, es decir, las dos rectas son perpendiculares y las variables son INCORRELADAS.



$$\text{Si } (X, Y) \text{ independientes} \mapsto s_{xy} = 0 \mapsto \begin{cases} b = 0 \\ b' = 0 \end{cases} \mapsto r^2 = 0$$

4. En una distribución bidimensional se conoce:

$$r = 0,7 \quad s_x = 1,2 \quad \bar{y} = 4 \quad X/Y: X = 0,6 + 0,44Y$$

Obtener: a) Recta de regresión de Y/X
b) Varianza de Y

Solución:

a) Recta de regresión de X sobre Y :

$$X = 0,6 + 0,44Y \mapsto \begin{cases} a' = 0,6 & b' = 0,44 \\ \bar{X} = 0,6 + 0,44\bar{Y} \end{cases} \mapsto \bar{X} = 0,6 + 0,44 \cdot 4 = 2,36$$

De otra parte, el coeficiente de determinación r^2 :

$$r^2 = b \cdot b' \mapsto 0,7^2 = b \cdot 0,44 \mapsto b = \frac{0,7^2}{0,44} = 1,114$$

La recta de regresión de Y sobre X: $Y = a + bX \mapsto \bar{Y} = a + b\bar{X} \mapsto 4 = a + 1,114 \cdot 2,36 \mapsto a = 1,37$

Y/X: $Y = 1,37 + 1,114 X$

b) Varianza de la Y: Sabemos que, $s_x = 1,2$ $b' = 0,44$ $b = 1,114$

$$b = \frac{s_{xy}}{s_x^2} \mapsto 1,114 = \frac{s_{xy}}{1,2^2} \mapsto s_{xy} = 1,114 \cdot 1,2^2 = 1,604$$

GRADO EN ECONOMÍA

14 de Mayo 2013

1. Una institución pública decidió estudiar el gasto mensual en alimentación en una ciudad, para lo cual se seleccionó un distrito y se tomó muestras cuyo resultado fue el que sigue:

Distrito 1	
Gasto (\$)	Nº Familias
100 – 200	24
200 – 300	36
300 – 400	20
400 – 500	20
500 – 1000	50

- Halle el gasto medio y el mediano en alimentación del distrito
- Si existe un segundo distrito de 120 familias con un gasto medio de 419,4 \$ y una desviación típica de 242,701 \$, ¿cuál de los dos tiene un gasto medio más representativo?
- Halle el gasto medio y la desviación típica del conjunto de los dos distritos.
- ¿Cuál es el nivel de gasto realizado por un mayor número de familias en el distrito 1?
- ¿Cuál es el máximo gasto realizado entre las 50 familias con menor gasto del distrito 1?
- Un índice de Gini de 0,10 en esta distribución ¿qué nos indicaría?

Solución:

a)

$[L_i - L_{i+1})$	x_i	n_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$	N_i	$f_i = \frac{n_i}{N}$	$F_i = \frac{N_i}{N}$	c_i	$d_i = \frac{n_i}{c_i}$
100 – 200	150	24	3600	540000	24	0,16	0,16	100	0,24
200 – 300	250	36	9000	2250000	60	0,24	0,40	100	0,36
300 – 400	350	20	7000	2450000	80	0,13	0,53	100	0,2
400 – 500	450	20	9000	4050000	100	0,13	0,66	100	0,2
500 – 1000	750	50	37500	28125000	150	0,33	1	500	0,1
		150	66100	37415000					

$$\text{Gasto medio: } \bar{x} = \frac{\sum_{i=1}^5 x_i n_i}{N} = \frac{66100}{150} = 440,67 \$$$

El Gasto mediano se encuentra en el intervalo [300 – 400)

$$\text{Mediana } \left[\frac{150}{2} = 75 \right]: M_e = L_i + \underbrace{\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}}}_{n_i} c_i = 300 + \frac{\frac{150}{2} - 60}{80 - 60} 100 = 300 + \frac{75 - 60}{80 - 60} 100 = 375 \$$$

b) El coeficiente de variación de Pearson mide el grado de homogeneidad de una distribución

$$\text{Distrito 1} \left\{ \begin{array}{l} s_x^2 = a_2 - \bar{x}^2 = \frac{\sum_{i=1}^5 x_i^2 n_i}{N} - \bar{x}^2 = \frac{37415000}{150} - 440,67^2 = 55243,28 \\ s_x = \sqrt{55243,28} = 235,04 \\ CV_x = \frac{s_x}{\bar{x}} = \frac{235,04}{440,67} = 0,5334 \quad (55,34\%) \end{array} \right.$$

$$\text{Distrito 2} \left\{ \begin{array}{l} \bar{y} = 419,4 \quad s_y = 242,701 \\ CV_y = \frac{s_y}{\bar{y}} = \frac{242,701}{419,4} = 0,5787 \quad (57,87\%) \end{array} \right.$$

Al tener el Distrito 1 un Coeficiente de Variación de Pearson más pequeño (menor dispersión del gasto medio) indica que tiene una media más representativa que el Distrito 2.

c) El gasto medio y desviación típica conjunta de los dos distritos:

Distrito 1:

(X ; $n_1 = 150$, $\bar{x} = 440,67$, $s_x = 235,04$)

Distrito 2:

(Y ; $n_2 = 120$, $\bar{y} = 419,4$, $s_y = 242,701$)

$$N = n_1 + n_2 = 150 + 120 = 270$$

$$\bar{x+y} = \frac{n_1 \bar{x} + n_2 \bar{y}}{n_1 + n_2} = \frac{150 \cdot 440,67 + 120 \cdot 419,4}{150 + 120} = 431,22 \quad \text{media ponderada}$$

$$\underbrace{s_{x_1+x_2}^2}_{\text{varianza total}} = \underbrace{\frac{\sum_{i=1}^2 s_i^2 n_i}{N}}_{\text{intra-grupos}} + \underbrace{\frac{\sum_{i=1}^2 (\bar{x}_i - \bar{x})^2 n_i}{N}}_{\text{entre-grupos}}$$

media ponderada de las varianzas parciales varianza ponderada de las medias parciales

media ponderada de las varianzas parciales

$$\frac{\sum_{i=1}^2 s_i^2 n_i}{N} = \frac{n_1 s_x^2 + n_2 s_y^2}{n_1 + n_2} = \frac{150 \cdot 235,04^2 + 120 \cdot 242,701^2}{150 + 120} = 56870,45$$

varianza ponderada de las medias parciales

$$\frac{\sum_{i=1}^2 (\bar{x}_i - \bar{x})^2 n_i}{N} = \frac{(440,67 - 431,22)^2 \cdot 150 + (419,4 - 431,22)^2 \cdot 120}{270} = 111,71$$

varianza total

$$\overbrace{s_{x+y}^2} = 56870,45 + 111,71 = 56982,16 \quad \mapsto \quad s_{x+y} = \sqrt{56982,16} = 238,71$$

d) El intervalo modal es [200 – 300) por tener mayor densidad de frecuencia $d_2 = 0,36$

$$M_d = L_i + \frac{(d_i - d_{i-1})}{(d_i - d_{i-1}) + (d_i - d_{i+1})} c_i \quad \text{Moda aproximada: } M_d = L_i + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i$$

$$\text{Intervalo [200 – 300): } M_d = 200 + \frac{(0,36 - 0,24)}{(0,36 - 0,24) + (0,36 - 0,2)} 100 = 242,86 \text{ \$}$$

$$\text{Moda aproximada: } M_d = L_i + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i = 200 + \frac{0,2}{0,24 + 0,2} 100 = 245,45 \text{ \$}$$

e) Máximo gasto realizado entre las 50 familias con menor gasto del distrito 1

$$\left[\frac{50}{150} = 0,33 \text{ (33,33\%)} \right] \mapsto P_{33,33} = L_i + \frac{\frac{33,33 \cdot N}{100} - N_{i-1}}{\underbrace{N_i - N_{i-1}}_{n_i}} c_i$$

$$P_{33,33} = L_i + \frac{\frac{33,33 \cdot 150}{100} - N_{i-1}}{\underbrace{N_i - N_{i-1}}_{n_i}} c_i = 200 + \frac{50 - 24}{60 - 24} 100 = 272,22 \text{ \$}$$

f) Un índice de Gini de 0,10, al ser próximo a cero, indica que el gasto se encuentra bastante bien repartido entre las familias.

2. Se ha realizado un estudio para determinar la recta de regresión que explique el gasto diario de los clientes del hotel (Y, medida en €) en función de la edad de los mismos (X, medida en años). Tras analizar los datos se ha obtenido la siguiente recta de regresión Y/X:

$$Y = 25 + 2,9 X$$

- Interprete los resultados de la recta de regresión.
- Si se sabe que $s_x = 10$ y que $s_y = 30$, determine la bondad del ajuste de esta recta de regresión a partir del coeficiente de correlación lineal e interprétela.
- Calcule los parámetros de la regresión de X sobre Y sabiendo que la media de edad de los clientes es de 30 años.
- ¿Cuál sería la edad esperada para un huésped que ha gastado diariamente 100 euros? ¿La predicción será fiable?. Razone la respuesta.

Solución:

- 2'9 es el coeficiente de regresión lineal. Al ser positivo cuando X crece, Y crece e indica el aumento de gasto de un cliente cuando su edad aumenta en una unidad.

25 euros es el valor de Y para X=0 años. En este caso no tiene sentido.

b) La bondad del ajuste viene dado por el coeficiente de determinación:

$$Y/X: Y = 25 + 2,9 X \mapsto \begin{cases} a = 25 \\ b = 2,9 \end{cases} \mapsto b = \frac{s_{xy}}{s_x^2} \Rightarrow s_{xy} = b \cdot s_x^2 \Rightarrow s_{xy} = 2,9 \cdot 10^2 = 290$$

$$\text{Coeficiente determinación: } r^2 = b \cdot b' = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = \frac{290^2}{10^2 \cdot 30^2} = 0,934$$

La relación lineal es bastante buena ya que el 93,4% de la variabilidad de Y se explica a partir de su dependencia con la variable X.

$$c) \bar{x} = 30 \rightarrow \bar{y} = 25 + 2,9 \cdot \bar{x} \Rightarrow \bar{y} = 25 + 2,9 \cdot 30 = 112 \quad Y = 25 + 2,9 X$$

$$X = a' + b' Y \begin{cases} b' = \frac{r^2}{b} \mapsto b' = \frac{0,934}{2,9} = 0,322 \\ \bar{x} = a' + b' \bar{y} \mapsto 30 = a' + 0,322 \cdot 112 \Rightarrow a' = -6,064 \end{cases}$$

$$\text{Recta de regresión de X/Y: } X = a' + b' Y \mapsto X = -6,064 + 0,322 \cdot Y$$

d) Edad esperada para un huésped con un gasto diario de 100 euros

$$X/Y: \text{ para } Y = 100 \Rightarrow X = -6,064 + 0,322 \cdot 100 = 26,136 \text{ euros}$$

La predicción es con una fiabilidad del 93,4% ($r^2 = 0,934$)

3. Un sector de la economía nacional dispone del valor de producción a precios corrientes de cada año (miles de euros) y los índices de precios de Laspeyres y Fisher.

Año	Producción (precios corrientes)	L_p (%)	F_p (%)
2007	78.147	100	100
2008	91.357	104,22	105,34
2009	88.854	107,25	108,94
2010	92.892	109,05	111,36
2011	101.336	114,87	117,67
2012	102.578	126,35	130,18

Utilizando el deflactor más idóneo, calcular la producción anual en precios constantes de 2007.

Solución:

Para calcular el valor real (precios constantes) de una magnitud se requiere deflactar el valor nominal (precios corrientes), *eliminando la influencia que han experimentado los precios*. Para ello, se deflacta la serie dividiendo el valor nominal entre un índice de precios.

$$\underbrace{\text{Valor Real}}_{\text{(precios constantes)}} = \frac{\overbrace{\text{Valor Nominal}}^{\text{(precios corrientes)}}}{\text{Índice Precios}}$$

$$V_t^R = \frac{V_t^N}{I_{p,0}^t} \cdot 100$$

El deflactor más adecuado es el de Paasche, ya que con éste índice de precios se obtiene una relación entre valores monetarios corrientes y valores monetarios constantes.

$$\text{Índice de Paasche: } P_p = \frac{\sum_{i=1}^n p_{it} \cdot q_{it}}{\sum_{i=1}^n p_{i0} \cdot q_{it}} \quad V_t^R = \frac{V_t^N}{P_p} = \frac{\sum_{i=1}^n p_{it} \cdot q_{it}}{\sum_{i=1}^n p_{i0} \cdot q_{it}} = \sum_{i=1}^n p_{i0} \cdot q_{it}$$

$$\text{El índice de precios de Fisher } F_p = \sqrt{L_p \cdot P_p} \mapsto P_p = \frac{(F_p)^2}{L_p}$$

Año	Producción (precios corrientes) V_t^N	% L_p	% F_p	% $P_p = \frac{(F_p)^2}{L_p}$	Producción (precios constantes 2007) $V_t^R = \frac{V_t^N}{P_p}$
2007	78.147	100	100	100	78147
2008	91.357	104,22	105,34	106,47	85803,75
2009	88.854	107,25	108,94	110,66	80297,04
2010	92.892	109,05	111,36	113,72	81685,61
2011	101.336	114,87	117,67	120,54	84069,58
2012	102.578	126,35	130,18	134,13	76478,78

4. En la tabla adjunta se reflejan las ventas trimestrales de una empresa en millones de euros. Halle la serie desestacionalizada por el método de las medias móviles.

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	2	3	2	4	5
Segundo	2	4	4	5	6
Tercero	3	5	5	7	8
Cuarto	3	4	4	3	5

Solución:

Se obtienen las medias móviles de tamaño 4 (período de las variaciones estacionales), que al ser un número par, serán descentradas y corresponderán a los períodos intermedios entre cada dos trimestres consecutivos:

$$\begin{aligned}\bar{Y}_{2,5} &= \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{2 + 2 + 3 + 3}{4} = 2,5 & \bar{Y}_{3,5} &= \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{2 + 3 + 3 + 3}{4} = 2,75 \\ \bar{Y}_{4,5} &= \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} = \frac{3 + 3 + 3 + 4}{4} = 3,25 & \bar{Y}_{5,5} &= \frac{Y_4 + Y_5 + Y_6 + Y_7}{4} = \frac{3 + 3 + 4 + 5}{4} = 3,75 \\ \dots\dots\dots & & \dots\dots\dots & \\ \bar{Y}_{17,5} &= \frac{Y_{16} + Y_{17} + Y_{18} + Y_{19}}{4} = \frac{3 + 5 + 6 + 8}{4} = 5,5 & \bar{Y}_{18,5} &= \frac{Y_{17} + Y_{18} + Y_{19} + Y_{20}}{4} = \frac{5 + 6 + 8 + 5}{4} = 6\end{aligned}$$

SERIE DESCENTRADA

Trimestres \ Años	2008	2009	2010	2011	2012
Primero - Segundo	-----	3,75	3,75	5	5,5
Segundo - Tercero	2,5	4	3,75	4,75	6
Tercero - Cuarto	2,75	3,75	4,25	5	-----
Cuarto - Primero	3,25	3,75	4,5	5,25	-----

Para corregir la nueva serie de móviles descentrada, a partir de ella se calcula la media aritmética de cada dos valores sucesivos, asignando este nuevo valor al instante central de los dos periodos considerados, es decir:

$$\begin{aligned}\bar{\bar{Y}}_3 &= \frac{\bar{Y}_{2,5} + \bar{Y}_{3,5}}{2} = \frac{2,5 + 2,75}{2} = 2,625 & \bar{\bar{Y}}_4 &= \frac{\bar{Y}_{3,5} + \bar{Y}_{4,5}}{2} = \frac{2,75 + 3,25}{2} = 3 \\ \dots\dots\dots & & \dots\dots\dots & \\ \bar{\bar{Y}}_{17} &= \frac{\bar{Y}_{16,5} + \bar{Y}_{17,5}}{2} = \frac{5,25 + 5,5}{2} = 5,375 & \bar{\bar{Y}}_{18} &= \frac{\bar{Y}_{17,5} + \bar{Y}_{18,5}}{2} = \frac{5,5 + 6}{2} = 5,75\end{aligned}$$

SERIE CENTRADA: COMPONENTES TENDENCIA Y CÍCLICA

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	-----	3,5	3,750	4,750	5,375
Segundo	-----	3,875	3,750	4,875	5,750
Tercero	2,625	3,875	4	4,875	-----
Cuarto	3	3,750	4,375	5,125	-----

La línea que une los puntos $\left[\bar{\bar{Y}}_3, \bar{\bar{Y}}_4, \dots, \bar{\bar{Y}}_{18} \right]$ se toma como **línea de tendencia**.

El inconveniente que presenta el **método de las medias móviles** es que no permite efectuar predicciones, puesto que con él no se obtiene la expresión de una fórmula matemática que facilite obtener el valor de la tendencia para un instante futuro. Este motivo hace que el método se utilice poco para determinar la tendencia, aunque sí se utiliza en el cálculo de los índices de variación estacional (**IVE**).

Al aplicar el **método de las medias móviles**, en el esquema multiplicativo $Y_{it} = T_{it} \cdot E_{it} \cdot C_{it} \cdot A_{it}$, lo que realmente se obtiene es una aproximación de $T_{it} \cdot C_{it}$ (componentes tendencia y cíclica), quedando sin analizar las componentes estacional (E_{it}) y accidental (A_{it}).

La tendencia T_{it} y la componente cíclica C_{it} se eliminarán dividiendo cada dato de la serie original Y_{it} por la correspondiente media móvil:

$$\frac{Y_{it}}{T_{it} \cdot C_{it}} = \frac{T_{it} \cdot E_{it} \cdot C_{it} \cdot A_{it}}{T_{it} \cdot C_{it}} = E_{it} \cdot A_{it} \quad \text{quedando la componente estacional y accidental}$$

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	---	3/3,5	2/3,75	4/4,75	5/5,375
Segundo	---	4/3,875	4/3,75	5/4,875	6/5,75
Tercero	3/2,625	5/3,875	5/4	7/4,875	---
Cuarto	3/3	4/3,75	4/4,375	3/5,125	---

COMPONENTES ESTACIONAL Y ACCIDENTAL

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	-----	0,857	0,533	0,842	0,930
Segundo	-----	1,032	1,067	1,026	1,043
Tercero	1,143	1,290	1,250	1,436	-----
Cuarto	1	1,067	0,914	0,585	-----

El Índice Bruto de Variación Estacional (**IBVE**) se calcula eliminando la componente accidental A_{it} .

Para ello, se hace el cálculo de las medias aritméticas trimestrales, es decir, la media aritmética de cada fila de la tabla anterior (donde solo aparecía el producto de $E_{it} \cdot A_{it}$):

$$\frac{0,857 + 0,533 + 0,842 + 0,930}{4} = 0,791$$

$$\frac{1,143 + 1,290 + 1,250 + 1,436}{4} = 1,280$$

$$\frac{1,032 + 1,067 + 1,026 + 1,043}{4} = 1,042$$

$$\frac{1 + 1,067 + 0,914 + 0,585}{4} = 0,892$$

COMP. ESTACIONAL Y ACCIDENTAL

COMP. ESTACIONAL Y ACCIDENTAL						COMPONENTE ESTACIONAL	
Trim \ Años	2008	2009	2010	2011	2012	IBVE	% IVE
Primero	-----	0,857	0,533	0,842	0,930	0,791	$(0,791 / 1,001) \cdot 100 = 78,990$
Segundo	-----	1,032	1,067	1,026	1,043	1,042	$(1,042 / 1,001) \cdot 100 = 104,095$
Tercero	1,143	1,290	1,250	1,436	-----	1,280	$(1,280 / 1,001) \cdot 100 = 127,847$
Cuarto	1	1,067	0,914	0,585	-----	0,892	$(0,892 / 1,001) \cdot 100 = 89,067$
						1,001	400

$$IBVE = \frac{4,004}{4} = 1,001$$

Adviértase que los índices de variación estacional (**IVE**) tienen que sumar 4 (400%)

Sobre un nivel medio de ventas, la influencia de la variación estacional (% IVE - 100) produce:

- 1º Trimestre: $(78,990 - 100) = -21,01\%$ descenso de ventas del 21,01%
 2º Trimestre: $(104,095 - 100) = 4,095\%$ aumento de ventas del 4,095%
 3º Trimestre: $(127,847 - 100) = 27,847\%$ aumento de ventas del 27,847%
 4º Trimestre: $(89,067 - 100) = -10,933\%$ descenso de ventas del 10,933%

La **DESESTACIONALIZACIÓN** (aplicando el método a la razón a la media móvil) consiste en dividir cada valor de la serie original por cada Índice de Variación Estacional correspondiente, en porcentaje

$$\frac{Y_{it}}{\% IVE_t} \cdot 100$$

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	(2/78,99).100	(3/78,99).100	(2/78,99).100	(4/78,99).100	(5/78,99).100
Segundo	(2/104,095).100	(4/104,095).100	(4/104,095).100	(5/104,095).100	(6/104,095).100
Tercero	(3/127,847).100	(5/127,847).100	(5/127,847).100	(7/127,847).100	(8/127,847).100
Cuarto	(3/89,067).100	(4/89,067).100	(4/89,067).100	(3/89,067).100	(5/89,067).100

SERIE DESESTACIONALIZADA

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	2,532	3,798	2,532	5,064	6,330
Segundo	1,921	3,843	3,843	4,803	5,764
Tercero	2,347	3,911	3,911	5,475	6,257
Cuarto	3,368	4,491	4,491	3,368	5,614

GRADO EN ECONOMÍA

21 de Junio 2013

1. En una fábrica trabajan 20.000 personas en la cadena de producción, cuyos salarios, en miles de euros, se distribuyen según la tabla adjunta:

Salarios	10 - 20	40 - 50	20 - 40	50 - 100	100 - 200
Nº trabajadores	12.000	1.000	6.000	800	200

- Determine el grado de concentración de los salarios
- ¿Qué parte de la nómina percibe el 5% del personal mejor pagado?
- ¿Qué porcentaje de los trabajadores percibe el 50% de los salarios?
- Si la empresa hace una reestructuración del 60% de plantilla en cada uno de los tramos de los salarios, ¿cuál sería el índice de Gini?

Solución:

- Ordenando los datos de forma creciente:

Salarios	x_i	n_i	N_i	$x_i n_i$	$\%p_i = \frac{N_i}{N} \cdot 100$	$U_i = x_i n_i$ acumulada	$\%q_i = \frac{U_i}{U_k} \cdot 100$
10 - 20	15	12000	12000	180000	60	180000	36,36
					x		50
20 - 40	30	6000	18000	180000	90	360000	72,73
40 - 50	45	1000	19000	45000	95	405000	81,82
50 - 100	75	800	19800	60000	99	465000	93,94
100 - 200	150	200	20000	30000	-----	495000	-----
					344		284,85

$$\text{Índice de Gini: } I_G = 1 - \frac{\sum_{i=1}^5 q_i}{\sum_{i=1}^5 p_i} = 1 - \frac{284,85}{344} = 0,1719 \quad (17,19\%)$$

- Comenzando por los salarios más bajos, se observa que el 81,82% de los salarios, es percibido por el 95% de la plantilla. En consecuencia, el 5% del personal mejor pagado percibe el 18,18%
- Se observa que el 60% de los trabajadores percibe el 36,36% de los salarios, mientras que el 90% de los trabajadores percibe el 72,73% de los salarios. Para estimar el porcentaje x de trabajadores que percibe el 50% de los salarios, se realiza una interpolación lineal:

$$\frac{90 - 60}{72,73 - 36,36} = \frac{x - 60}{50 - 36,36} \quad \mapsto \quad x = 71,25 \%$$

- d) El índice de Gini tiene que ser coherente con el Principio de la Población, es decir, el índice de Gini no varía cuando el conjunto de individuos con la misma renta se multiplican por un escalar.

En consecuencia, si la empresa hace una modificación de la plantilla del 60% en todos los tramos de salarios el índice de Gini tiene que ser el mismo: $I_g = 0,1719$

2. Dada la tabla de correlación:

X \ Y	0	3	6
1	1	5	2
2	4	4	1

- a) Hallar las rectas de regresión mínimo cuadráticas asociadas.
b) Hallar la varianza explicada por la regresión y la varianza residual de la recta Y/X, explicando los resultados.

Solución:

- a) Se efectúan los cálculos necesarios para obtener los momentos respecto al origen:

X \ Y	0	3	6	$n_{i\cdot}$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$x_i y_j n_{ij}$			
1	1	5	2	8	8	8	0	15	12	63
2	4	4	1	9	18	36	0	24	12	
$n_{\cdot j}$	5	9	3	17	26	44				
$y_j n_{\cdot j}$	0	27	18	45						
$y_j^2 n_{\cdot j}$	0	81	108	189						

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^2 x_i n_{i\cdot}}{N} = \frac{26}{17} = 1,53$$

$$a_{20} = \frac{\sum_{i=1}^2 x_i^2 n_{i\cdot}}{N} = \frac{44}{17} = 2,59$$

$$s_x^2 = a_{20} - a_{10}^2 = 2,59 - 1,53^2 = 0,25$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^3 y_j n_{\cdot j}}{N} = \frac{45}{17} = 2,65$$

$$a_{02} = \frac{\sum_{j=1}^3 y_j^2 n_{\cdot j}}{N} = \frac{189}{17} = 11,12$$

$$s_y^2 = a_{02} - a_{01}^2 = 11,12 - 2,65^2 = 4,1$$

$$a_{11} = \frac{\sum_{i=1}^2 \sum_{j=1}^3 x_i y_j n_{ij}}{N} = \frac{63}{17} = 3,71$$

$$s_{xy} = a_{11} - a_{10} \cdot a_{01} = 3,71 - 1,53 \cdot 2,65 = -0,34$$

$$\text{Recta regresión Y/X: } Y = a + b X \begin{cases} b = \frac{s_{xy}}{s_x^2} = \frac{-0,34}{0,25} = -1,36 \\ \bar{y} = a + b \bar{x} \mapsto a = \bar{y} - b \bar{x} = 2,65 + 1,36 \cdot 1,53 = 4,73 \end{cases}$$

$$Y/X: Y = 4,73 - 1,36 X$$

$$\text{Recta regresión X/Y: } X = a' + b' Y \begin{cases} b' = \frac{s_{xy}}{s_y^2} = \frac{-0,34}{4,1} = -0,083 \\ \bar{x} = a' + b' \bar{y} \mapsto a' = \bar{x} - b' \bar{y} = 1,53 + 0,083 \cdot 2,65 = 1,75 \end{cases}$$

$$X/Y: X = 1,75 - 0,083 Y$$

$$b) \text{ Coeficiente de determinación: } r^2 = b \cdot b' = (-1,36) \cdot (-0,083) = 0,1129$$

$$\text{Varianza residual de Y: } s_{ry}^2 = s_y^2 (1 - r^2) = 4,1 (1 - 0,1129) = 3,637$$

$$\text{Varianza explicada por la regresión: } \begin{cases} s_y^2 = s_{ry}^2 + s_{Ry}^2 \mapsto s_{Ry}^2 = s_y^2 - s_{ry}^2 = 4,1 - 3,637 = 0,463 \\ s_{Ry}^2 = s_y^2 \cdot r^2 = 4,1 \cdot 0,1129 = 0,463 \end{cases}$$

$$\text{La mayor parte de la variable dependiente Y resulta ser residual, un } \frac{3,637}{4,1} \cdot 100 = 88,7\% .$$

En consecuencia, una pequeña parte queda explicada por la regresión:

$$\begin{cases} r^2 \cdot 100 = 0,1129 \cdot 100 = 11,29\% \\ (0,463 / 4,1) \cdot 100 = 11,29\% \end{cases}$$

Al ser la varianza explicada muy pequeña, el ajuste no es bueno y las rectas de regresión no pueden utilizarse de manera fiable para hacer predicciones.

3. Un trabajador ha recibido los siguientes salarios en los años 2005 y 2006:

Salario 2005 = 18.565 euros

Salario 2006 = 19.005 euros

Esta persona quiere saber si su poder adquisitivo ha aumentado en el año 2006 respecto al 2005. Para ello dispone de la siguiente información relativa al Índice de Precios de Consumo con base el año 2002

$$IPC_{2002}^{2005} = 109,93\% \quad \text{e} \quad IPC_{2002}^{2006} = 113,63\%$$

- Interprete el valor de los números índice proporcionados
- Determine e interprete la tasa de variación que ha sufrido el poder adquisitivo de este asalariado entre los años 2005 y 2006, en términos nominales y en términos reales (constantes del 2002)
- Si el salario del trabajador en el año 2002 fue de 16.000 euros, ¿cuál fue la tasa media anual acumulativa en términos nominales y reales (constantes del 2002) en el periodo 2002-2006?

Solución:

a)

$IPC_{2002}^{2005} = 109,93\% \rightarrow$ En el año 2005 los precios se han incrementado un 9,93% respecto al año 2002

$IPC_{2002}^{2006} = 113,63\% \rightarrow$ En el año 2006 los precios se han incrementado un 13,63% respecto al año 2002

b) Para calcular el salario real (precios constantes) se requiere deflactar el salario nominal (precios corrientes), *eliminando la influencia que han experimentado los precios*. Para ello, se deflacta la serie dividiendo el valor nominal entre el IPC

$$\underbrace{\text{Salario real}}_{\text{precios constantes}} = \frac{\underbrace{\text{Salario nominal}}_{\text{precios corrientes}}}{IPC_{2002}^t} \rightarrow \begin{cases} SR_{2005}^{\text{constante}} = \frac{SN_{2005}^{\text{corriente}}}{IPC_{2002}^{2005}} = \frac{18565}{1,0993} = 16888,02 \text{ euros} \\ SR_{2006}^{\text{constante}} = \frac{SN_{2006}^{\text{corriente}}}{IPC_{2002}^{2006}} = \frac{19005}{1,1363} = 16725,34 \text{ euros} \end{cases}$$

$$\text{Tasas de variación} \begin{cases} \text{Nominal: } TV_{2005}^{2006} = \left[\frac{19005}{18665} - 1 \right] \cdot 100 = 2,37\% \\ \text{Real: } TV_{2005}^{2006} = \left[\frac{16725,34}{16888,02} - 1 \right] \cdot 100 = -0,963\% \end{cases}$$

En términos nominales el salario ha crecido un 2,37%, aunque en términos reales (eliminado el efecto de la inflación), el salario ha disminuido un 0,963%.

c) La tasa media anual acumulativa en términos nominales y reales (constantes del 2002) en el periodo 2002-2006

$$\begin{cases} I_{\text{salario nominal}} = \frac{SN_{2006}^{\text{corriente}}}{SN_{2002}} = \frac{19005}{16000} = 1,1878 \\ I_{\text{salario real}} = \frac{SR_{2006}^{\text{constante}}}{SR_{2002}} = \frac{16725,34}{16000} = 1,0453 \end{cases}$$

- Tasa de variación media anual en términos nominales:

$$TM_{\text{nominal}} = \sqrt[4]{I_{\text{salario nominal}}} - 1 = \sqrt[4]{1,1878} - 1 = 1,04396 - 1 = 0,4396 \quad (4,396\%)$$

- Tasa de variación media anual en términos reales:

$$TM_{\text{real}} = \sqrt[4]{I_{\text{salario real}}} - 1 = \sqrt[4]{1,0453} - 1 = 1,0111 - 1 = 0,111 \quad (1,11\%)$$

4. Tras analizar los datos referentes a un año y medio (desde 2004.1 hasta 2005.2) de una determinada serie temporal (Y), de periodicidad trimestral, se han obtenido los siguientes resultados con $t = 0, 1, \dots, 5$:

$$\sum t = 15 \quad \sum t^2 = 55 \quad \sum ty_t = 71.950 \quad \sum y_t = 19.073 \quad \sum y_t^2 = 97.199.705$$

Los índices de variación estacionales han sido:

$$IVE_1 = 1,033 \quad IVE_2 = 0,87 \quad IVE_3 = 0,97 \quad IVE_4 = 1,127$$

- a) Realice un ajuste lineal de la tendencia de la serie. Determine a partir del coeficiente de determinación lineal si el ajuste es bueno o malo, y prediga el valor de la serie para el tercer y cuarto trimestre del año 2005.
- b) Interprete estadísticamente los IVEs

Solución:

a) Recta de regresión de Y sobre t: $Y = a + b \cdot t$
$$\begin{cases} b = \frac{s_{ty}}{s_t^2} \\ a = \bar{y} - b \bar{t} \end{cases}$$

$$\bar{t} = \frac{\sum_{i=1}^6 t}{N} = \frac{15}{6} = 2,5 \quad \bar{y} = \frac{\sum_{t=1}^6 y_t}{N} = \frac{19073}{6} = 3178,83 \quad a_{11} = \frac{\sum_{t=1}^6 ty_t}{N} = \frac{71950}{6} = 11991,67$$

$$s_{ty} = a_{11} - \bar{t} \cdot \bar{y} = 11991,67 - 2,5 \cdot 3178,83 = 4044,59 \quad s_t^2 = \frac{\sum_{i=1}^6 t^2}{6} - \bar{t}^2 = \frac{55}{6} - 2,5^2 = 2,92$$

con lo que, $Y = -283,99 + 1385,13 \cdot t$
$$\begin{cases} b = \frac{4044,59}{2,92} = 1385,13 \\ a = 3178,83 - 1385,13 \cdot 2,5 = -283,99 \end{cases}$$

El Coeficiente de determinación lineal: $R^2 = b \cdot b'$

$$b' = \frac{s_{ty}}{s_y^2} \quad s_y^2 = \frac{\sum_{t=1}^6 y_t^2}{N} - \bar{y}^2 = \frac{97199705}{6} - 3178,83^2 = 6094990,66 \quad b' = \frac{4044,59}{6094990,66} = 0,00066$$

$$R^2 = b \cdot b' = 1385,13 \cdot 0,00066 = 0,914$$

El modelo es bueno porque explica el 91,4% ($R^2 = 0,914$) de la variabilidad de Y_t en función de t .

Para predecir el tercer ($t = 6$) y cuarto trimestre ($t = 7$) de 2005: $Y = -283,99 + 1385,13 \cdot t$

$$2005.3: Y = -283,99 + 1385,13 \cdot 6 = 8026,79$$

$$2005.4: Y = -283,99 + 1385,13 \cdot 7 = 9411,92$$

En el esquema multiplicativo $Y_{it} = T_{it} \cdot E_{it} \cdot C_{it} \cdot A_{it} \mapsto Y_{it} = T_{it} \cdot IVE_h \quad (h \in t)$

$$Y_{it} = T_{it} \cdot IVE_h \begin{cases} Y_{2005.3} = T_{2005.3} \cdot IVE_3 = 8026,79 \cdot 0,97 = 7785,99 \\ Y_{2005.4} = T_{2005.4} \cdot IVE_4 = 9411,92 \cdot 1,127 = 10607,23 \end{cases}$$

b) Los índices de variación estacional muestran el componente estacional en el esquema multiplicativo. El componente estacional E_{it} son las oscilaciones que sufre una serie temporal en periodos inferiores o iguales a un año.

$$IVE_1 = 1,033 \quad IVE_2 = 0,87 \quad IVE_3 = 0,97 \quad IVE_4 = 1,127$$

- $IVE_1 = 1,033$ significa que por el hecho de estar en el primer trimestre, la variable Y_{it} es un 3,3% mayor que el comportamiento habitual o tendencia de la serie.
- $IVE_2 = 0,87$ significa que por el hecho de estar en el segundo trimestre, la variable Y_{it} es un 13% menor que el comportamiento habitual o tendencia de la serie.

EXAMEN DE ESTADÍSTICA DESCRIPTIVA

28 DE JUNIO 2013

1.- Se quieren analizar los accidentes de tráfico en las provincias españolas. Se disponen de los siguientes datos:

Accidentes de Tráfico (miles)	Nº de Provincias españolas
0 - 15	25
15 - 35	15
35 - 50	10

- Obtenga el número medio de accidentes por provincia y su valor mediano.
- La media obtenida en el apartado anterior, ¿es representativa?
- ¿Se producen en España los accidentes de forma concentrada según provincias? Justifique el indicador empleado para medir la concentración de los accidentes e interprete los resultados.
- En Alemania se ha realizado un estudio similar al español. Se ha obtenido un índice de Gini del 0,70. Dibuje las curvas de Lorenz teóricas que representarían los indicadores de concentración de ambos países y explique la posición de cada una de ellas.

Solución:

a)

$[L_i - L_{i+1})$	x_i	n_i	c_i	N_i	$x_i n_i$	$x_i^2 n_i$
0 - 15	7,5	25	15	15	187,5	1406,25
15 - 35	25	15	20	40	375	9375
35 - 50	42,5	10	15	50	425	18062,5
					987,5	28843,75

Número medio accidentes: $\bar{x} = \frac{\sum_{i=1}^3 x_i n_i}{N} = \frac{987,5}{50} = 19,75$

Valor mediano: $M_e = L_i + \underbrace{\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}}}_{n_i} c_i = 15 + \frac{\frac{50}{2} - 15}{40 - 15} 20 = 15 + \frac{25 - 15}{40 - 15} 20 = 23$

- Para saber si la media obtenida es representativa se calcula el Coeficiente de Variación de Pearson:

$$a_2 = \frac{\sum_{i=1}^3 x_i^2 n_i}{N} = \frac{28843,75}{50} = 576,875$$

$$s^2 = a_2 - a_1^2 = 576,875 - 19,75^2 = 186,8125 \quad \mapsto \quad s = \sqrt{186,8125} = 13,67$$

$$CV = \frac{s}{\bar{x}} = \frac{13,67}{19,75} = 0,6911 \quad (69,11\%)$$

El Coeficiente de Variación de Pearson cuantifica el grado de dispersión (69,11%), que resulta ser alto, por lo que la media aritmética no es representativa.

c)

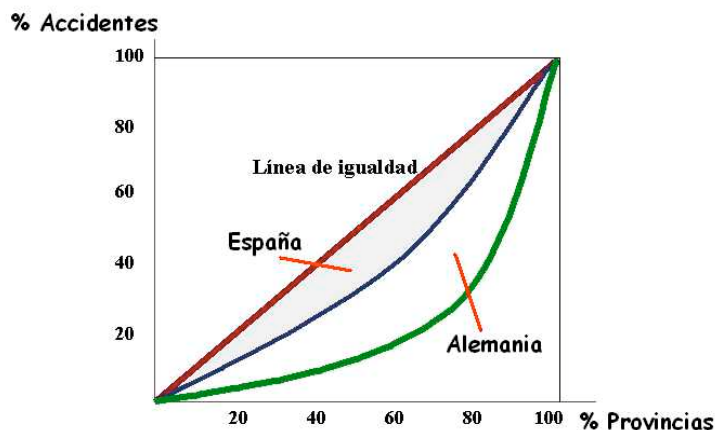
Rentas [$L_i - L_{i+1}$)	x_i	n_i	N_i	$x_i n_i$	$U_i = x_i n_i$ acumulada	$\%p_i = \frac{N_i}{N} 100$	$\%q_i = \frac{U_i}{U_k} 100$	$\%(p_i - q_i)$
0 - 15	7,5	25	15	187,5	187,5	50	18,99	31,01
15 - 35	25	15	40	375	562,5	80	56,96	23,04
35 - 50	42,5	10	50	425	987,5	100	100	0
				987,5		130	75,95	54,05

El grado de concentración de accidentes viene reflejado por el Índice de Gini:

$$I_G = 1 - \frac{\sum_{i=1}^2 q_i}{\sum_{i=1}^2 p_i} = 1 - \frac{75,95}{130} = 0,4158 \quad (41,58\%) \quad \text{o bien} \quad I_G = \frac{\sum_{i=1}^2 (p_i - q_i)}{\sum_{i=1}^2 p_i} = \frac{54,05}{130} = 0,4158$$

Cuanto más próximo a cero se encuentre el Índice de Gini será más equitativo el grado de concentración de accidentes, siendo de 41,58%, se puede concluir que existe concentración de accidentes.

d)



[$I_G(\text{Alemania}) = 0,70 > I_G(\text{España}) = 0,4158$]
concluyendo que en Alemania están más concentrados los accidentes, esto es, al dibujar las curvas teóricas, la curva de Lorenz de España se encontraría más próxima a la diagonal principal.

2.- A partir de la tabla adjunta, siendo $N=11$, $\bar{Y}=0$

$X \setminus Y$	- 2	0	1
0	0	1	0
1	3	n_{22}	n_{23}
2	0	1	0

- a) ¿Son independientes las variables estadísticamente?
- b) Rectas de regresión de Y/X e X/Y
- c) ¿Qué parte de la varianza calculada Y es explicada por la regresión? ¿Qué parte es debida a causas ajenas?

Solución:

a)

$X \setminus Y$	- 2	0	1	$n_{i\cdot}$
0	0	1	0	1
1	3	n_{22}	n_{23}	$3 + n_{22} + n_{23}$
2	0	1	0	1
$n_{\cdot j}$	3	$2 + n_{22}$	n_{23}	$5 + n_{22} + n_{23} = 11$

$$\text{De otra parte, } \bar{Y} = \frac{-2 \cdot 3 + 0 + n_{23}}{11} = 0 \quad \mapsto \quad n_{23} = 6$$

$$5 + n_{22} + 6 = 11 \quad \mapsto \quad n_{22} = 0$$

$X \setminus Y$	- 2	0	1	$n_{i\cdot}$
0	0	1	0	1
1	3	0	6	9
2	0	1	0	1
$n_{\cdot j}$	3	2	6	11

Las variables X e Y son independientes cuando se verifica $\frac{n_{ij}}{N} = \left(\frac{n_{i\cdot}}{N}\right) \left(\frac{n_{\cdot j}}{N}\right) \quad \forall i, j$

No son independientes porque no se verifica la relación: $\frac{1}{11} \neq \frac{1}{11} \times \frac{2}{11} \quad \left[\frac{n_{12}}{N} \neq \left(\frac{n_{1\cdot}}{N}\right) \left(\frac{n_{\cdot 2}}{N}\right) \right]$

b)

$$a_{11} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 x_i y_j n_{ij}}{N} = \frac{1}{11} [-2 \cdot 1 \cdot 3 + 1 \cdot 1 \cdot 6] = 0$$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^3 x_i n_{i\cdot}}{N} = \frac{1 \cdot 9 + 0 + 2 \cdot 1}{11} = 1 \quad a_{20} = \frac{\sum_{i=1}^3 x_i^2 n_{i\cdot}}{N} = \frac{1}{11} [1^2 \cdot 9 + 2^2 \cdot 1] = \frac{13}{11}$$

$$s_x^2 = a_{20} - a_{10}^2 = \frac{13}{11} - 1 = \frac{2}{11} \mapsto s_x = \sqrt{\frac{2}{11}} = 0,43$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^3 y_j \cdot n_{\bullet j}}{N} = 0$$

$$a_{02} = \frac{\sum_{j=1}^3 y_j^2 \cdot n_{\bullet j}}{N} = \frac{1}{11} [(-2)^2 \cdot 3 + 1^2 \cdot 6] = \frac{18}{11}$$

$$s_y^2 = a_{02} - a_{01}^2 = \frac{18}{11} - 0 = \frac{18}{11} \mapsto s_y = \sqrt{\frac{18}{11}} = 1,28$$

$$\text{covarianza: } s_{xy} = a_{11} - a_{10} \cdot a_{01} = 0 - 1 \cdot 0 = 0$$

- El coeficiente de regresión de Y sobre X (pendiente de la recta): $b = \frac{s_{xy}}{s_x^2} = \frac{0}{2/11} = 0$

$$\bar{Y} = a + b \bar{X} \mapsto 0 = a + 0 \cdot 1 \mapsto a = 0$$

$$Y/X: Y=0$$

- El coeficiente de regresión de X sobre Y (pendiente de la recta): $b' = \frac{s_{xy}}{s_y^2} = \frac{0}{18/11} = 0$

$$\bar{X} = a' + b' \bar{Y} \mapsto 1 = a' + 0 \cdot 0 \mapsto a' = 1$$

$$X/Y: X=1$$

COEFICIENTE DETERMINACIÓN: $r^2 = b \cdot b' = 0 \mapsto$ Las rectas son perpendiculares, y en consecuencia, las variables (X, Y) son INCORRELADAS

$$\text{VARIANZA RESIDUAL DE Y: } s_{ry}^2 = s_y^2 (1 - r^2) \mapsto s_{ry}^2 = \frac{18}{11} (1 - 0) = \frac{18}{11}$$

$$s_y^2 = s_{y \text{ explicada}}^2 + s_{ry}^2 \mapsto \frac{18}{11} = s_{y \text{ explicada}}^2 + \frac{18}{11} \mapsto s_{y \text{ explicada}}^2 = 0$$

3.- En la tabla se presenta el valor de importaciones de un país durante los años 2009 y 2010.

Importaciones	2009	2010
Alimentos	1010	1200
Otros bienes de consumo	7450	7955
Bienes de capital	2400	2210
Bienes intermedios	4755	6256
TOTAL	15615	17621

Se sabe que las importaciones tanto de alimentos como de otros bienes de consumo se pagaron un 3% más caras en 2010 que en 2009.

Las importaciones de bienes de capital subieron sus precios un 1,2% y las de bienes intermedios bajaron un 0,5%.

Se pide:

a) Calcular el índice de precios total de las importaciones en 2010 con base 2009, utilizando Laspeyres y Paasche.

b) ¿Cuánto crecieron las importaciones en cantidad en 2009 con respecto a 2010?

Solución:

a)

♦ Utilizando el índice de precios de Laspeyres:

Importaciones	Laspeyres		
	$p_{i,09} \cdot q_{i,09}$	$p_{i,10} \cdot q_{i,10}$	$p_{i,10} \cdot q_{i,09}$
Alimentos	1010	1200	$1,03 \times 1010 = 1040,3$
Otros bienes de consumo	7450	7955	$1,03 \times 7450 = 7673,5$
Bienes de capital	2400	2210	$1,012 \times 2400 = 2428,8$
Bienes intermedios	4755	6256	$0,995 \times 4755 = 4731,23$
TOTAL	15615	17621	15873,83

$$L_p = \frac{\sum_{i=1}^4 p_{i,10} \cdot q_{i,09}}{\sum_{i=1}^4 p_{i,09} \cdot q_{i,09}} \cdot 100 = \frac{15873,83}{15615} \cdot 100 = 101,66\%$$

♦ Utilizando el índice de precios de Paasche:

Importaciones	Paasche		
	$p_{i,09} \cdot q_{i,09}$	$p_{i,10} \cdot q_{i,10}$	$p_{i,09} \cdot q_{i,10}$
Alimentos	1010	1200	$1200/1,03 = 1165,05$
Otros bienes de consumo	7450	7955	$7955/1,03 = 7723,30$
Bienes de capital	2400	2210	$2210/1,012 = 2183,79$
Bienes intermedios	4755	6256	$6256/0,995 = 6287,44$
TOTAL	15615	17621	17359,58

$$P_p = \frac{\sum_{i=1}^4 p_{it} \cdot q_{it}}{\sum_{i=1}^4 p_{i0} \cdot q_{it}} \cdot 100 = \frac{17621}{17359,58} \cdot 100 = 101,51\%$$

b) Para calcular los índices cuánticos de Laspeyres y Paasche se requiere hallar previamente el índice de valor de las importaciones entre 2009 con base 2010.

$$IV_{09}^{10} = \frac{V_{10}}{V_{09}} = \frac{\sum_{i=1}^4 p_{i,10} \cdot q_{i,10}}{\sum_{i=1}^4 p_{i,09} \cdot q_{i,09}} = \frac{17621}{15615} = 1,1285 \quad (112,85\%)$$

$$\text{Siendo, } IV_0^t = L_{P0}^t \cdot P_{Q0}^t = P_{P0}^t \cdot L_{Q0}^t \quad \left\{ \begin{array}{l} P_{Q09}^{10} = \frac{IV_{09}^{10}}{L_{P09}^{10}} \cdot 100 = \frac{112,85}{101,66} \cdot 100 = 111,01\% \\ L_{Q09}^{10} = \frac{IV_{09}^{10}}{P_{P09}^{10}} \cdot 100 = \frac{112,85}{101,51} \cdot 100 = 111,17\% \end{array} \right.$$

4.- En la tabla adjunta se reflejan las ventas trimestrales de una empresa en millones de euros.

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	1	2	2	3	5
Segundo	2	3	4	4	7
Tercero	4	5	5	7	8
Cuarto	3	4	3	6	7

Suponiendo un esquema de agregación multiplicativo en la serie temporal:

- Desestacionalice la serie de ventas por el método de las medias móviles.
- Calcule los Índices de Variación Estacional (IVEs) por el método de la tendencia.

Solución:

a) Para calcular la tendencia secular de la serie por el método de las medias móviles, se obtienen primero medias móviles de tamaño 4 (período de las variaciones estacionales), que al ser un número par, se pierden 4 datos, resulta una serie descentrada y corresponderán a los períodos intermedios entre cada dos trimestres consecutivos.

Cálculo de las medias móviles:

$$\frac{1+2+4+3}{4} = 2,5 \text{ entre segundo y tercer trimestre de 2006}$$

$$\frac{2+4+3+2}{4} = 2,75 \text{ entre tercer y cuarto trimestre de 2006}$$

$$\frac{4+3+2+3}{4} = 3 \text{ entre cuarto trimestre de 2006 y primer trimestre de 2007}$$

$$\frac{3+2+3+5}{4} = 3,25 \text{ entre primer y segundo trimestre de 2007}$$

$$\frac{2+3+5+4}{4} = 3,5 \text{ entre segundo y tercer trimestre de 2007}$$

SERIE DESCENTRADA de medias móviles

Trimestres \ Años	2006	2007	2008	2009	2010
Primero-Segundo	---	3,25	3,75	4,25	6,5
Segundo-Tercero	2,5	3,5	3,5	5	6,75
Tercero-Cuarto	2,75	3,5	3,75	5,5	---
Cuarto-Primero	3	3,75	3,75	6,25	---

Para centrar la serie hay que calcular la media aritmética de cada dos observaciones sucesivas, de este modo, las medias que irán apareciendo, respectivamente, serán:

$$\frac{2,5+2,75}{2} = 2,625 \quad \frac{2,75+3}{2} = 2,875 \quad \frac{3+3,25}{2} = 3,125 \quad \frac{3,25+3,5}{2} = 3,375 \quad \frac{3,5+3,5}{2} = 3,5$$

$$\frac{3,5+3,75}{2} = 3,625 \quad \frac{3,75+3,75}{2} = 3,75 \quad \frac{3,75+3,5}{2} = 3,625 \quad \frac{3,5+3,75}{2} = 3,625 \quad \frac{3,75+3,75}{2} = 3,75$$

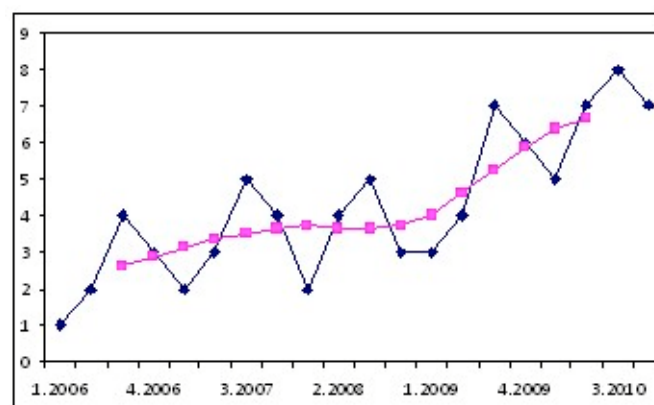
$$\frac{3,75+4,25}{2}=4 \quad \frac{4,25+5}{2}=4,625 \quad \frac{5+5,5}{2}=5,25 \quad \frac{5,5+6,25}{2}=5,875 \quad \frac{6,25+6,5}{2}=6,375$$

$$\frac{6,5+6,75}{2}=6,625$$

SERIE CENTRADA de las medias móviles:

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	---	3,125	3,75	4	6,375
Segundo	---	3,375	3,625	4,625	6,625
Tercero	2,625	3,5	3,625	5,25	---
Cuarto	2,875	3,625	3,75	5,875	---

La línea que se obtiene al representar gráficamente la serie de la tabla (t, \bar{y}_{it}) será la *línea de tendencia*, que comienza en el tercer trimestre de 2006 y finaliza en el segundo trimestre de 2010.



Al aplicar el *método de las medias móviles*, en el esquema multiplicativo $Y_{it} = T_{it} \cdot E_{it} \cdot C_{it} \cdot A_{it}$, lo que realmente se obtiene en la serie cronológica es una aproximación de $T_{it} \cdot C_{it}$, quedando sin analizar las componentes estacional (E_{it}) y accidental (A_{it}).

La tendencia y la componente cíclica se eliminarán dividiendo cada dato de la serie original por la correspondiente media móvil:

$$\frac{Y_{it}}{T_{it} \cdot C_{it}} = \frac{T_{it} \cdot E_{it} \cdot C_{it} \cdot A_{it}}{T_{it} \cdot C_{it}} = E_{it} \cdot A_{it} \quad \text{quedando la componente estacional y accidental}$$

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	---	2/3,125	2/3,75	3/4	5/6,375
Segundo	---	3/3,375	4/3,625	4/4,625	7/6,625
Tercero	4/2,625	5/3,5	5/3,625	7/5,25	---
Cuarto	3/2,875	4/3,625	3/3,75	6/5,875	---

SERIE con las componentes estacional y accidental

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	---	0,640	0,533	0,750	0,784
Segundo	---	0,889	1,103	0,865	1,057
Tercero	1,524	1,429	1,379	1,333	---
Cuarto	1,043	1,103	0,8	1,021	---

Se elimina la componente accidental A_{it} con el cálculo de las medias aritméticas trimestrales, es decir, la media aritmética de cada fila de la tabla anterior (donde solo aparecía el producto de $E_{it} \cdot A_{it}$):

$$\frac{0,640 + 0,533 + 0,750 + 0,784}{4} = 0,677$$

$$\frac{0,889 + 1,103 + 0,865 + 1,057}{4} = 0,978$$

$$\frac{1,524 + 1,429 + 1,379 + 1,333}{4} = 1,416$$

$$\frac{1,043 + 1,103 + 0,8 + 1,021}{4} = 0,992$$

Trimestres \ Años	2006	2007	2008	2009	2010	IVBE
Primero	---	0,640	0,533	0,750	0,784	0,677
Segundo	---	0,889	1,103	0,865	1,057	0,978
Tercero	1,524	1,429	1,379	1,333	---	1,416
Cuarto	1,043	1,103	0,8	1,021	---	0,992
						<u>1,016</u>

Se calcula la media aritmética de los cuatro valores obtenidos anteriormente $\frac{0,677 + 0,978 + 1,416 + 0,992}{4} = 1,016$

Se calculan los Índices de Variación Estacional, expresando para ello cada uno de los valores anteriores en forma de porcentaje sobre la media anual, obteniendo:

Trimestres \ Años	IVE (%)
Primero	$(0,677/1,016) \cdot 100 = 66,63$
Segundo	$(0,978/1,016) \cdot 100 = 96,31$
Tercero	$(1,416/1,016) \cdot 100 = 139,41$
Cuarto	$(0,992/1,016) \cdot 100 = 97,65$
	<u>400 %</u>

DESESTACIONALIZACIÓN (aplicando el método a la razón a la media móvil).- El proceso consiste en dividir cada valor de la serie original por cada Índice de Variación Estacional correspondiente:

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	1/0,6663	2/0,6663	2/0,6663	3/0,6663	5/0,6663
Segundo	2/0,9631	3/0,9631	4/0,9631	4/0,9631	7/0,9631
Tercero	4/1,3941	5/1,3941	5/1,3941	7/1,3941	8/1,3941
Cuarto	3/0,9765	4/0,9765	3/0,9765	6/0,9765	7/0,9765

Serie desestacionalizada, método a la razón a la media móvil

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	1,501	3,002	3,002	4,502	7,504
Segundo	2,077	3,115	4,153	4,153	7,268
Tercero	2,869	3,587	3,587	5,021	5,738
Cuarto	3,072	4,096	3,072	6,144	7,168

b) Los Índices de Variación Estacional (IVEs) por el método de la tendencia.
Se calculan las medias anuales $\bar{y}_{\bullet t}$ (medias para cada año de $k = 4$ subperiodos)

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	1	2	2	3	5
Segundo	2	3	4	4	7
Tercero	4	5	5	7	8
Cuarto	3	4	3	6	7
	$\bar{y}_{\bullet 2006} = 2,5$	$\bar{y}_{\bullet 2007} = 3,5$	$\bar{y}_{\bullet 2008} = 3,5$	$\bar{y}_{\bullet 2009} = 5$	$\bar{y}_{\bullet 2010} = 6,75$

$$\bar{y}_{\bullet t} = \frac{\sum_{i=1}^4 y_{it}}{4} \quad t = (2006, 2007, \dots, 2010) \text{ medias anuales}$$

La *tendencia media anual* $\bar{T}_{\bullet t}$ se obtiene ajustando una recta de regresión a los años (t_1, t_2, \dots, t_n) y a las medias anuales $\bar{y}_{\bullet t}$, donde $t \equiv (t_1, t_2, \dots, t_n)$: $\bar{T}_{\bullet t} = \hat{\bar{y}}_{\bullet t} = a + b \cdot t$

$(t_{2006}, t_{2007}, \dots, t_{2010})$	2006	2007	2008	2009	2010
$\bar{y}_{\bullet t} \equiv$ medias anuales	2,50	3,50	3,50	5,00	6,75

Por el método de los mínimos cuadrados, resulta: $a = -2003,75$ y $b = 1$

con lo que, $\bar{T}_{\bullet t} = \hat{\bar{y}}_{\bullet t} = -2003,75 + t$ $t = (t_{2006}, t_{2007}, \dots, t_{2010})$, resulta pues:

Tendencia media anual

$(t_{2006}, t_{2007}, \dots, t_{2010})$	2006	2007	2008	2009	2010
$\bar{T}_{\bullet t}$	2,25	3,25	4,25	5,25	6,25

A partir de la tendencia media anual $\bar{T}_{\bullet t}$ se obtiene el valor de la *tendencia para los distintos subperiodos*, según la expresión general:

$$T_{it} = \bar{T}_{\bullet t} + \left[i - \frac{k+1}{2} \right] \cdot \frac{b}{k} \quad \text{tendencia media anual para los subperiodos } k\text{-ésimos}$$

donde,

$t \equiv$ Año (2006, 2007, ..., 2010)

$i \equiv$ Subperíodo donde se calcula la tendencia (trimestral $i = 1, 2, 3, 4$)

$k \equiv$ Número total de subperiodos (datos trimestrales $k = 4$)

$b \equiv$ Pendiente de la recta de regresión = 1

$$\text{Trimestre Primero 2006 : } T_{i2006} = 2,25 + \left[1 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 1,875$$

$$\text{Trimestre Segundo 2006 : } T_{i2006} = 2,25 + \left[2 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 2,125$$

$$\text{Trimestre Tercero 2006 : } T_{i2006} = 2,25 + \left[3 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 2,375$$

$$\text{Trimestre Primero 2007 : } T_{i2007} = 3,25 + \left[1 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 2,875$$

$$\text{Trimestre Primero 2008 : } T_{i2008} = 4,25 + \left[1 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 3,875$$

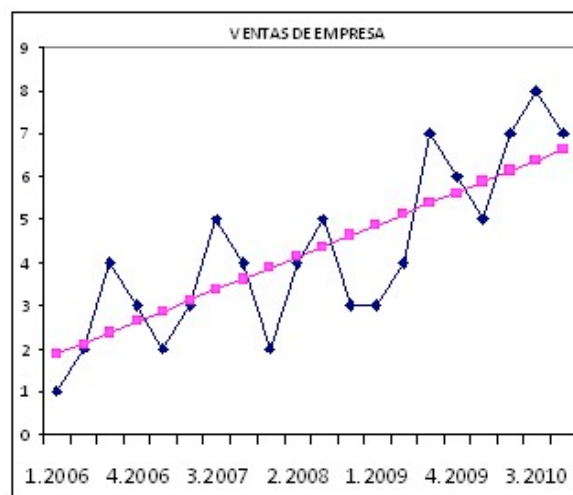
$$\text{Trimestre Primero 2009 : } T_{i2009} = 4,25 + \left[1 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 4,875$$

$$\text{Trimestre Primero 2010 : } T_{i2010} = 5,25 + \left[1 - \frac{4+1}{2} \right] \cdot \frac{1}{4} = 5,875$$

SERIE DE LA TENDENCIA

(k=4 trimestres)	i t	2006	2007	2008	2009	2010
Primero	1	1,875	2,875	3,875	4,875	5,875
Segundo	2	2,125	3,125	4,125	5,125	6,125
Tercero	3	2,375	3,375	4,375	5,375	6,375
Cuarto	4	2,625	3,625	4,625	5,625	6,625

Representación gráfica de la serie con los datos originales y la serie suavizada de tendencia



Para eliminar la tendencia y la componente cíclica se divide cada término de la serie original entre el correspondiente término de la serie teórica de tendencia.

SE ELIMINA LA TENDENCIA Y LA COMPONENTE CÍCLICA DE LA SERIE

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	1/1,875	2/2,875	2/3,875	3/4,875	5/5,875
Segundo	2/2,125	3/3,125	4/4,125	4/5,125	7/6,125
Tercero	4/2,375	5/3,375	5/4,375	7/5,375	8/6,375
Cuarto	3/2,625	4/3,625	3/4,625	6/5,625	7/6,625

Señalar que, en el esquema multiplicativo, al aplicar el método de los mínimos cuadrados, lo que se obtiene es una aproximación, ya que en el período que se considera (un año) es suficientemente pequeño, pudiendo suponer que la componente cíclica está incluida en la tendencia secular, puesto que en un período tan corto no da lugar a que se manifiesten plenamente las variaciones cíclicas.

Serie con las COMPONENTES ESTACIONAL y ACCIDENTAL

Trimestres \ Años	2006	2007	2008	2009	2010
Primero	0,533	0,696	0,516	0,615	0,851
Segundo	0,941	0,960	0,970	0,780	1,143
Tercero	1,684	1,481	1,143	1,302	1,255
Cuarto	1,143	1,103	0,649	1,067	1,057

Para eliminar la componente accidental, calculamos para cada trimestre la media aritmética de los valores obtenidos por trimestres (filas) en la serie anterior con las componentes estacional y accidental.

$$\frac{0,533+0,696+0,516+0,615+0,851}{5}=0,642$$

$$\frac{0,941+0,96+0,97+0,78+1,143}{5}=0,959$$

$$\frac{1,684+1,481+1,143+1,302+1,255}{5}=1,373$$

$$\frac{1,143+1,103+0,649+1,067+1,057}{5}=1,004$$

Trimestres \ Años	2006	2007	2008	2009	2010	IBVE
Primero	0,533	0,696	0,516	0,615	0,851	0,642
Segundo	0,941	0,960	0,970	0,780	1,143	0,959
Tercero	1,684	1,481	1,143	1,302	1,255	1,373
Cuarto	1,143	1,103	0,649	1,067	1,057	1,004
						0,994

El promedio anual de las cuatro medias aritméticas: $\frac{0,642+0,959+1,373+1,004}{4}=0,994$

Se calculan los Índices de Variación Estacional, expresando para ello cada uno de los valores obtenidos (medias aritméticas por trimestres) en forma de porcentaje sobre la media anual, obteniendo:

Trimestres \ Años	IBVE	IVE (%)
Primero	0,642	$(0,642/0,944).100 = 64,59$
Segundo	0,959	$(0,959/0,944).100 = 96,48$
Tercero	1,373	$(1,373/0,944).100 = 138,13$
Cuarto	1,004	$(1,004/0,944).100 = 101,01$

En definitiva, sobre un nivel medio de ventas, la influencia de la variación estacional produce:

1º Trimestre: $(64,59 - 100 = -35,41) \rightarrow$ un descenso de ventas del 35,41%

2º Trimestre: $(96,48 - 100 = -3,52) \rightarrow$ un descenso de ventas del 3,42%

3º Trimestre: $(138,13 - 100 = 38,13) \rightarrow$ un aumento de ventas del 38,13%

4º Trimestre: $(101,01 - 100 = 1,01) \rightarrow$ un aumento de ventas del 1,01%