Guía Estadística Descriptiva

En cualquier experimento se pueden distinguir dos fases. Una primera, que consiste en la observación y análisis de los hechos que acontecen, y otra segunda, que interpreta y obtiene conclusiones.

La Estadística Descriptiva se encarga de dar una descripción numérica, ordenar y simplificar la información recogida en la primera fase.

La Estadística Teórica, con la información obtenida en el muestreo de la primera fase, pretende generalizar o predecir los resultados para toda la población.

En la tabla se muestran las rentas (en miles de euros) y el número de personas que las perciben:

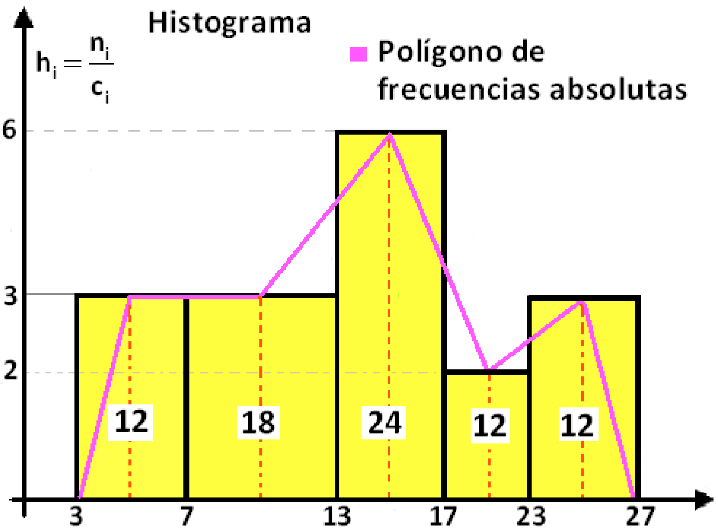
Rentas (miles euros)	3-7	7 – 13	13 – 17	17 – 23	23 – 27
número personas	12	18	24	12	12

a) Polígono de frecuencias absolutas, histograma.
b) Mediana, percentil 75 y moda.
c) Media aritmética, geométrica y armónica.
d) Desviación media (respecto a la media) y coeficiente de desviación media.
e) Coeficientes de asimetría de Pearson y de Fisher.
f) Media aritmética y desviación típica de la variable $Z = X - 15/5$
g) Coeficiente de apuntamiento o de curtosis.

Solución:

$\boxed{\text{a)}}$ Tabla de frecuencias:

Rentas (miles euros) $[L_i - L_{i+1})$	x_i	n_i	N_i	c_i amplitud	$h_i = \frac{n_i}{c_i}$ densidad
3 - 7	5	12	12	4	3
7 - 13	10	18	30	6	3
13 - 17	15	24	39 54 58,5	4	6
17 - 23	20	12	66	6	2
23 - 27	25	12	78	4	3



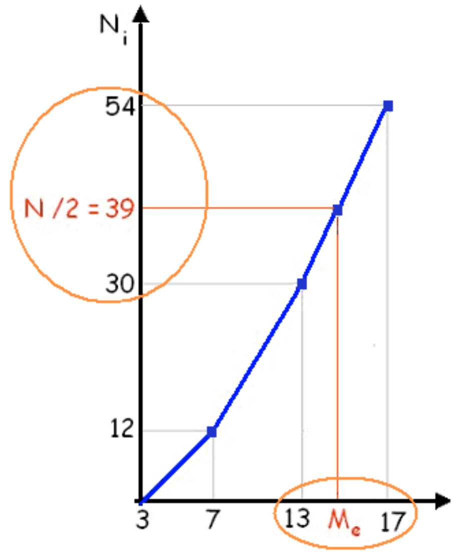
Encima de cada intervalo se levanta un rectángulo cuya área sea igual (en número) a la frecuencia absoluta de dicho intervalo.

$\boxed{\text{b)}}$ La observación $N/2 = 78/2 = 39$ se encuentra en el intervalo $[13 - 17)$

Se establece una proporcionalidad entre las bases y las alturas:

$$\frac{17-13}{54-30}=\frac{M_e-13}{N/2-30} \mapsto$$

$$\mapsto \quad M_e=\overbrace{13}^{L_i}+\overbrace{4}^{c_i}.\frac{N/2-\overbrace{30}^{N_{i-1}}}{\underbrace{54}_{N_{i+1}}-\underbrace{30}_{N_{i-1}}}$$

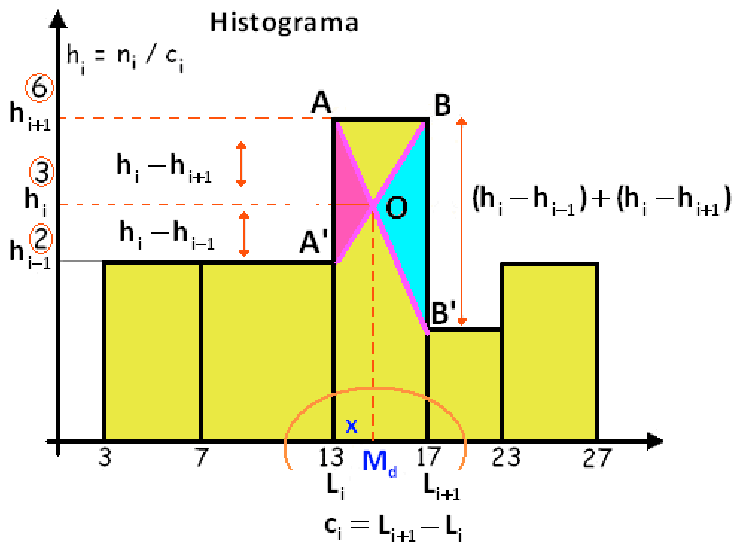


$$M_e=L_i+\frac{\frac{N}{2}-N_{i-1}}{N_i-N_{i-1}}\,c_i=13+\frac{\frac{78}{2}-30}{54-30}\,4=13+\frac{39-30}{54-30}\,4=14,5$$

- La observación para el percentil 75: $\frac{75.78}{100}=58,5$ se encuentra en el intervalo $[17-23)$

$$P_{75}=L_i+\frac{\frac{75.N}{100}-N_{i-1}}{N_i-N_{i-1}}\,c_i=17+\frac{58,5-54}{66-54}\,6=18,125$$

- La Moda es el intervalo de máxima frecuencia, por lo que el intervalo modal es $[13-17)$



Se establece una proporcionalidad de semejanza entre los triángulos $\overset{\triangle}{AOA'}$ y $\overset{\triangle}{BOB'}$

$$\frac{(h_i-h_{i-1})+(h_i-h_{i+1})}{c_i}=\frac{(h_i-h_{i-1})}{x}$$

$$x=\frac{(h_i-h_{i-1})}{(h_i-h_{i-1})+(h_i-h_{i+1})}\,c_i$$

$$M_d = L_i + \frac{(h_i - h_{i-1})}{(h_i - h_{i-1}) + (h_i - h_{i+1})}\, c_i$$

Con distintas amplitudes c_i una moda aproximada es: $M_d = L_i + \dfrac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i$

con lo cual, $M_d = 13 + \dfrac{(6-3)}{(6-3)+(6-2)}\, 4 = 13{,}428$

Moda aproximada: $M_d = L_i + \dfrac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i = 13 + \dfrac{2}{3+2}\, 4 = 14{,}6$

$\boxed{\text{c)}}$ **Media aritmética, geométrica y armónica**

$[L_i - L_{i+1})$ Rentas	x_i	n_i	$x_i.n_i$	$\frac{n_i}{x_i}$	$\log x_i$	$n_i \log x_i$
3 - 7	5	12	60	2,4	0,6989	8,3868
7 - 13	10	18	180	1,8	1	18
13 - 17	15	24	360	1,6	1,1760	28,224
17 - 23	20	12	240	0,6	1,3010	15,612
23 - 27	25	12	300	0,48	1,3979	16,7748
		78	1140	6,88		86,9976

- **Media aritmética:** $\bar{x} = \dfrac{\sum_{i=1}^{5} x_i.n_i}{N} = \dfrac{1140}{78} = 14{,}615$ *miles de euros*

- **Media geométrica:** $\bar{x}_G = \sqrt[N]{x_1^{n_1}\, x_2^{n_2}\, x_3^{n_3}\, \cdots\, x_k^{n_k}}$ para el cálculo se procede tomando logaritmos, con lo cual:

$$\log \bar{x}_G = \frac{1}{N}\log\left[x_1^{n_1}\, x_2^{n_2}\, x_3^{n_3}\, \cdots\, x_k^{n_k}\right] = \frac{1}{N}\left[\log(x_1^{n_1}) + \log(x_2^{n_2}) + \cdots + \log(x_k^{n_k})\right] =$$

$$= \frac{1}{N}\sum_{i=1}^{k} n_i \log x_i$$

finalmente, $\log \bar{x}_G = \dfrac{1}{N}\sum_{i=1}^{k} n_i \log x_i \quad \mapsto \quad \bar{x}_G = 10^{\left[\frac{1}{N}\sum_{i=1}^{k} n_i \log x_i\right]}$

En consecuencia,

$$\log \overline{x}_G = \frac{1}{N}\sum_{i=1}^{5} n_i \log x_i = \frac{1}{78}(86,9976) = 1,115 \quad \mapsto \quad \overline{x}_G = 10^{1,115} = 13,031$$

En general, se utiliza la media geométrica en los siguientes casos:

- Promediar tasas de crecimiento de una misma variable en períodos sucesivos. Por ejemplo, para calcular la tasa de crecimiento del PIB o del valor de una cartera de valores a lo largo de varios años, conocida como tasa de crecimiento acumulativa media.

- Para promediar tasas de variación o cocientes respecto a cierto nivel de referencia (aunque se trate de variaciones en un mismo período, no en períodos consecutivos). Dado que una de las propiedades de la media geométrica es que, cuando se trata de promediar cocientes, es el único promedio que cumple que la media geométrica de los cocientes es igual al cociente entre la medida geométrica de los numeradores y la de los denominadores.

Esta medida de centralización tiene la dificultad de que sólo con que un valor de la variable sea 0, ya $\overline{x}_G = 0$ y, por tanto, su representatividad sería dudosa.

- **Media armónica:** $\overline{x}_A = \dfrac{N}{\displaystyle\sum_{i=1}^{k} \frac{n_i}{x_i}}$

con lo cual, $\overline{x}_A = \dfrac{N}{\displaystyle\sum_{i=1}^{5} \frac{n_i}{x_i}} = \dfrac{78}{6,88} = 11,337$ *miles de euros*

La media armónica se suele utilizar cuando los datos a promediar están medidos en unidades expresadas en forma de cocientes.

Se emplea en promediar velocidades, tiempos, rendimientos, etc. Surge de forma natural al calcular el índice de Paasche (cambios en cantidades en cada una de las mercancías).

No es aconsejable utilizarla en distribuciones donde existan valores pequeños, no tiene sentido si algún valor de la variable es 0.

Se verifica la fórmula de Foster, para distribuciones de frecuencias con valores positivos: $\boxed{\bar{x}_A \leq \bar{x}_G \leq \bar{x}}$

$\boxed{\text{d)}}$ **Desviación Media (respecto a la media) y Coeficiente de Desviación Media:**

$$D_M(\bar{x}) = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{N} \qquad CV_{D_M}(\bar{x}) = \frac{D_{M_{\bar{x}}}}{|\bar{x}|}$$

$[L_i - L_{i+1})$ Rentas	x_i	n_i	$x_i . n_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert n_i$
3 - 7	5	12	60	9,615	115,38
7 - 13	10	18	180	4,615	83,07
13 - 17	15	24	360	0,385	9,24
17 - 23	20	12	240	5,385	64,62
23 - 27	25	12	300	10,385	124,62
		78	1140		396,93

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i . n_i}{N} = \frac{1140}{78} = 14,615$$

- **Desviación media respecto a la media aritmética:**

$$D_M(\bar{x}) = \frac{\sum_{i=1}^{5} |x_i - \bar{x}| n_i}{N} = \frac{396,93}{78} = 5,088$$

- **Coeficiente de variación media respecto a la media aritmética:**

$$CV_{D_M}(\bar{x}) = \frac{D_{M_{\bar{x}}}}{|\bar{x}|} = \frac{5,088}{14,615} = 0,3481$$

Desviación media respecto a la mediana: $D_M(M_e) = \dfrac{\sum_{i=1}^{k} |x_i - M_e| n_i}{N}$

Coeficiente de variación media respecto a la mediana: $CV_{D_M}(M_e) = \dfrac{D_M(M_e)}{|M_e|}$

e) Coeficientes de Asimetría de Pearson y de Fisher.

$[L_i - L_{i+1})$ Rentas	x_i	n_i	$x_i \cdot n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2\, n_i$	$(x_i - \bar{x})^3\, n_i$
3 - 7	5	12	60	- 9,615	1109,38	- 10666,676
7 - 13	10	18	180	- 4,615	383,368	- 1769,243
13 - 17	15	24	360	0,385	3,557	1,369
17 - 23	20	12	240	5,385	347,979	1873,865
23 - 27	25	12	300	10,385	1294,179	13440,046
		78	1140		3138,463	2879,361

- **El coeficiente de asimetría de Pearson exige el cálculo de la moda y la desviación tipica**

$$A_P = \frac{\bar{x} - M_d}{\sigma} \begin{cases} A_P > 0 & \textit{Asimetría a la derecha o positiva} \\ A_P = 0 & \textit{Simetría} \\ A_P < 0 & \textit{Asimetría a la izquierda o negativa} \end{cases}$$

Este coeficiente tiene sentido cuando la moda es única.

Se conoce: $M_d = 13,428 \quad \bar{x} = 14,615$

$$m_2 = \sigma^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2\, n_i}{N} = \frac{3183,463}{78} = 40,237 \quad \mapsto \quad \sigma = \sqrt{40,237} = 6,343$$

$$A_P = \frac{\bar{x} - M_d}{\sigma} = \frac{14,615 - 13,428}{6,343} = 0,187 > 0 \quad \mapsto \quad \begin{cases} \text{La distribucion presenta una} \\ \text{asimetría a la derecha o positiva} \end{cases}$$

- **Coeficiente de asimetría de Fisher:**

$$A_F = \frac{m_3}{\sigma^3} \begin{cases} A_F > 0 & \textit{Asimetría a la derecha o positiva} \\ A_F = 0 & \textit{Simetría} \\ A_F < 0 & \textit{Asimetría a la izquierda o negativa} \end{cases}$$

Tercer momento respecto a la media: $m_3 = \dfrac{\sum_{i=1}^{5} (x_i - \bar{x})^3\, n_i}{N} = \dfrac{2879,361}{78} = 36,915$

$$A_F = \frac{m_3}{\sigma^3} = \frac{36,915}{6,343^3} = 0,145 > 0 \quad \mapsto \begin{cases} \text{La distribucion presenta una} \\ \text{asimetría a la derecha o positiva} \end{cases}$$

$\boxed{\text{f)}}$ **Media aritmética y desviación típica de la variable $Z = X - 15 / 5$**

$[L_i - L_{i+1})$ Rentas	x_i	n_i	$z_i = \frac{x_i - 15}{5}$	$z_i . n_i$	$z_i^2 . n_i$
3 - 7	5	12	-2	-24	48
7 - 13	10	18	-1	-18	18
13 - 17	15	24	0	0	0
17 - 23	20	12	1	12	12
23 - 27	25	12	2	24	48
		78		-6	126

- $a_1 = \bar{z} = \dfrac{\sum_{i=1}^{5} z_i . n_i}{N} = \dfrac{-6}{78} = -0,077$

$$\bar{z} = E\left[\frac{x - 15}{5}\right] = \frac{1}{5}\, E\left[x - 15\right] = \frac{1}{5}(\bar{x} - 15) = \frac{1}{5}(14,615 - 15) = -0,077$$

- $\sigma_z^2 = a_2 - a_1^2 = \dfrac{\sum_{i=1}^{5} z_i^2\, n_i}{N} - (\bar{z})^2 = \dfrac{126}{78} - (-0,077)^2 = 1,61$

$$\sigma_z^2 = V\left[\frac{x - 15}{5}\right] = \frac{1}{25}.\, V(x - 15) = \frac{1}{25}.\, V(x) = \frac{1}{25}.\, 40,237 = 1,61$$

$$\sigma_z = \sqrt{1,60947} = 1,2686$$

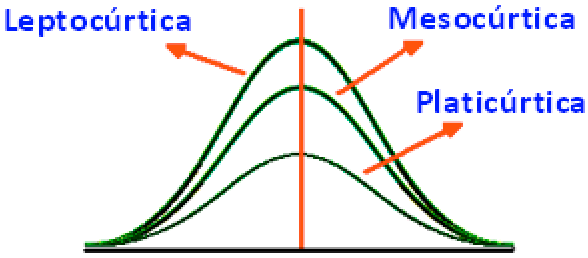
$\boxed{\text{g)}}$ **Coeficiente de apuntamiento o de curtosis.**

El coeficiente g_2 de curtosis indica cuál es el apuntamiento de la distribución comparándola con la normal (o campana de Gauss)

Bajo apuntamiento $\Leftrightarrow$ Gran aplastamiento

Coeficiente de curtosis de Fisher: $g_2 = \dfrac{m_4}{\sigma^4} - 3$

donde, $m_2 = \sigma^2 = \dfrac{\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i}{N}$ $\quad m_4 = \dfrac{\sum_{i=1}^{k}(x_i - \overline{x})^4 n_i}{N}$



$$g_2 = \frac{m_4}{\sigma^4} - 3 \begin{cases} g_2 > 0 & \text{Más apuntamiento que la normal} \mapsto \text{Leptocúrtica} \\ g_2 = 0 & \text{Igual apuntamiento que la normal} \mapsto \text{Mesocúrtica} \\ g_2 < 0 & \text{Menor apuntamiento que la normal} \mapsto \text{Platicúrtica} \end{cases}$$

En la distribución se conoce: $\overline{x} = 14{,}615 \quad \sigma_x = 6{,}343$

$[L_i - L_{i+1})$ Rentas	x_i	n_i	$x_i.n_i$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2\, n_i$	$(x_i - \overline{x})^4\, n_i$
3 - 7	5	12	60	92,4482	1109,38	102560,036
7 - 13	10	18	180	21,2982	383,368	8165,039
13 - 17	15	24	360	0,1482	3,557	0,5271
17 - 23	20	12	240	28,9982	347,979	10090,747
23 - 27	25	12	300	107,848	1294,179	139574,293
		78	1140		3138,463	260390,6421

Momento de cuarto orden respecto a la media:

$$m_4 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^4 n_i}{N} = \frac{260390{,}6421}{78} = 3338{,}3415$$

$g_2 = \dfrac{m_4}{\sigma^4} - 3 \Rightarrow \dfrac{3338{,}3415}{6{,}343^4} - 3 = -0{,}9377 < 0 \mapsto$ La distribución presenta menor apuntamiento que la distribución normal, se trata de una distribución platicúrtica.

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:

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

a) Representar el histograma de la distribución.

b) Calcular el consumo medio de gas de las viviendas. ¿El valor hallado es representativo de la distribución?

c) Calcular el consumo más frecuente.

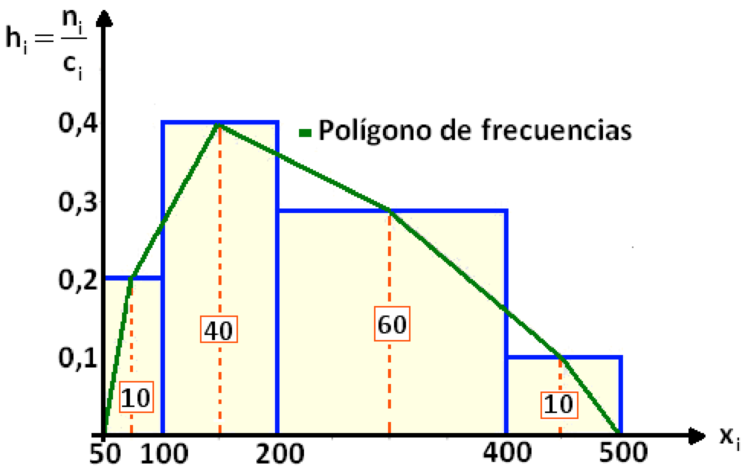
d) Averiguar el valor del tercer cuartil de la distribución del consumo de gas y explicar su significado.

e) Si la factura del gas consiste en una cantidad fija de 20 euros más 0,5 euros por m^3 consumido, calcular la factura media de las viviendas y determinar si la factura es más dispersa que el consumo.

Solución:

$\boxed{a}$ La variable estadística X = "consumo del gas" agrupada en intervalos $[L_i, L_{i+1})$

$[L_i, L_{i+1})$	c_i amplitud	n_i	$h_i = \frac{n_i}{c_i}$ densidad	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	90 110	300	18000	5400000
400 - 500	100	10	0,1	120	450	4500	2025000
						29250	8381250



$\boxed{b}$ **El consumo medio de gas de las viviendas:**

$$a_1 = \bar{x} = \frac{\sum_{i=1}^{4} x_i\, n_i}{N} = \frac{29250}{120} = 243,75\ \ m^3 \qquad a_2 = \frac{\sum_{i=1}^{4} x_i^2\, n_i}{N} = \frac{8381250}{120} = 69843,75$$

$$\sigma^2 = a_2 - a_1^2 = 69843,75 - (243,75)^2 = 10429,6875$$

$$\sigma_x = \sqrt{10429,6875} = 102,1258\ \ m^3$$

$$C.V = \frac{\sigma_x}{\bar{x}} = \frac{102,1258}{243,75} = 0,42\ \ (42\%) \quad \textit{coeficiente variación de Pearson}$$

El consumo medio de gas de las viviendas es de 243,75 m^3, con una dispersión del 42%, con lo que no es muy representativo.

$\boxed{c}$ **El consumo más frecuente se encuentra en el intervalo modal [100-200), que es donde alcanza la máxima densidad de frecuencia.**

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

• **Cuando existen distintas amplitudes, una moda aproximada viene dada por la expresión:** $M_d = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i$

con lo cual, $M_d = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i = 100 + \frac{0,3}{0,2+0,3} 100 = 160\ m^3$

Si la amplitud de los intervalos es constante: $M_d = L_i + \dfrac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} c_i$

$\boxed{d}$ El tercer cuatil $Q_3 = P_{75}$ percentil 75, siendo $\dfrac{3N}{4} = \dfrac{3.120}{4} = 90$, observando en la tabla la columna de la frecuencia absoluta acumulada N_i, se tiene:

$$Q_3 = P_{75} = L_i + \frac{\frac{3.N}{4} - N_{i-1}}{N_i - N_{i-1}} c_i \quad \mapsto \quad Q_3 = P_{75} = 200 + \frac{90 - 50}{110 - 50} 200 = 333,33 \text{ m}^3$$

El 75% de las viviendas que consumen menos, consumen como máximo 333,33 m^3 de gas.

$\boxed{e}$ Según el enunciado, la factura del gas viene expresada por la relación $Y = 20 + 0,5.X$, por lo que la variable estadística X tiene un cambio de origen y de escala.

La factura media: $\overline{Y} = 20 + 0,5.\overline{X} = 20 + 0,5.243,75 = 141,875$ euros

$$\sigma_Y^2 = \text{Var}(20 + 0,5.X) = 0,5^2.\sigma_X^2$$

$$\sigma_Y = 0,5.\sigma_X = 0,5 \,.\, 102,1258 = 51,063 \text{ euros}$$

$$C.V = \frac{\sigma_Y}{\overline{y}} = \frac{51,063}{141,875} = 0,36 \ \ (36\%)$$

La factura del gas esta menos dispersa que el consumo.

NOTA: Cambio de origen y de escala

$$\overline{y} = \frac{\sum_{i=1}^{k} y_i.n_i}{N} = \frac{1}{N}\sum_{i=1}^{k}(ax_i + b).n_i = a\frac{1}{N}\sum_{i=1}^{k} x_i.n_i + b\frac{1}{N}\sum_{i=1}^{k} n_i = a\overline{x} + b \quad \mapsto$$

$$\mapsto \quad E(Y) = E(aX + b) = a\overline{x} + b$$

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

$$\sigma_y^2 = \frac{\sum_{i=1}^{k}(y_i-\bar{y})^2.n_i}{N} = \frac{\sum_{i=1}^{k}(ax_i+b-a\bar{x}-b)^2.n_i}{N} = a^2\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2.n_i}{N} = a^2\,\sigma_x^2 \mapsto$$

$$Var(Y) = Var(ax+b) = a^2 . \sigma_x^2$$

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

Abel Grandes Pistado preguntó a sus 31 compañeros de clase qué calificación obtuvieron en el último examen de estadística. Sólo recuerda que su tocayo Escasi Lopasa tuvo un 4,6 (una de las notas más frecuentes).
Haciendo memoria ha podido completar los siguientes datos:

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

a) ¿Qué proporción de alumnos han tenido una nota superior a 5?,
¿Cómo es la distribución respecto a la moda?
b) Analizar la dispersión relativa de las notas a partir del coeficiente de variación de Pearson. Interpretar los resultados.
c) ¿Cómo afecta a la homogeneidad de la distribución que este examen sea un 60% de la calificación final?

Solución:

$\boxed{a}$ Se sabe que $M_d = 4,6$

$[L_i - L_{i+1})$	c_i amplitud	x_i	n_i	$h_i = \frac{n_i}{c_i}$ densidad
0 - 4	4	2	8	2
4 - 5	1	4,5	n_2	h_2
5 - 7	2	6	n_3	h_3
7 - 9	2	8	6	3
9 - 10	1	9,5	6	6
			32	

Para hallar n_2 y n_3 se recurre a la moda:

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

$$4,60 = L_2 + \frac{h_{2+1}}{h_{2-1} + h_{2+1}}.1 = 4 + \frac{h_3}{h_1 + h_3} = 4 + \frac{h_3}{2 + h_3}$$

$$0,60 = \frac{h_3}{2 + h_3} \mapsto h_3 = \frac{1,2}{0,4} = 3 \mapsto n_3 = h_3 . c_3 = 3.2 = 6$$

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

La proporción de alumnos que obtiene una nota superior:

$[L_i - L_{i+1})$	x_i	n_i	$x_i.n_i$	$x_i^2.n_i$
0 - 4	2	8	16	32
4 - 5	4,5	6	27	121,5
5 - 7	6	6	36	216
7 - 9	8	6	48	384
9 - 10	9,5	6	57	541,5
		32	184	1295

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

$\boxed{b}$ La dispersión relativa de las notas se calcula a partir del coeficiente de variación de Pearson:

$$a_1 = \bar{x} = \frac{\sum_{i=1}^{5} x_i\, n_i}{N} = \frac{184}{32} = 5,75 \qquad a_2 = \frac{\sum_{i=1}^{5} x_i^2\, n_i}{N} = \frac{1295}{32} = 40,46875$$

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

$$C.V = \frac{\sigma_x}{\bar{x}} = \frac{2,72}{5,75} = 0,4730 \ \ (47,30\,\%)$$

La dispersión es del 47,3% , es decir, una dispersión media.

c **Para analizar la homogeneidad de la distribución cuando el examen es un 60% de la calificación final hay que calcular el coeficiente de variación de Pearson.**

$$\text{Cambio de escala: } \begin{cases} E(k.x) = k.E(x) = k.\overline{x} \\ Var(k.x) = k^2.Var(x) = k^2\,\sigma_x^2 \end{cases} \mapsto \quad C.V_{final} = \frac{\sigma_{final}}{\overline{x}_{final}} = \frac{k.\sigma_x}{k.\overline{x}} = \frac{\sigma_x}{\overline{x}}$$

El cambio de escala no afecta al coeficiente de variación de Pearson.

En consecuencia, no afecta a la homogeneidad de la distribución que el examen sea un 60% de la calificación final.

La media se ve afectada por el cambio de origen y de escala.

La varianza no se ve afectada por el cambio de origen pero sí por el cambio de escala.

El coeficiente de variación de Pearson no se ve afectado por el cambio de escala pero sí por el cambio de origen.

Intervalos

$$\text{Moda: } M_d = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})}\, c_i$$

$$\text{Mediana: } M_e = L_i + \frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}}\, c_i$$

El índice de Gini verifica:

Principio del Anonimato

Principio de la Renta Relativa

Principio de Dalton

Principio de la Población

La curva de Lorenz es la representación de la concentración de recursos entre individuos.

El índice de Gini expresa el grado de concentración de los recursos.

$$I_G = 1 - \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i}$$

El cambio de escala no afecta al Índice

El cambio de origen afecta al Índice

En la tabla se muestran las rentas (en miles de euros) y el número de personas que las perciben:

Rentas (miles euros) $[L_i, L_{i+1})$	$3-7$	$7-13$	$13-17$	$17-23$	$23-27$
número personas n_i	12	18	24	12	12

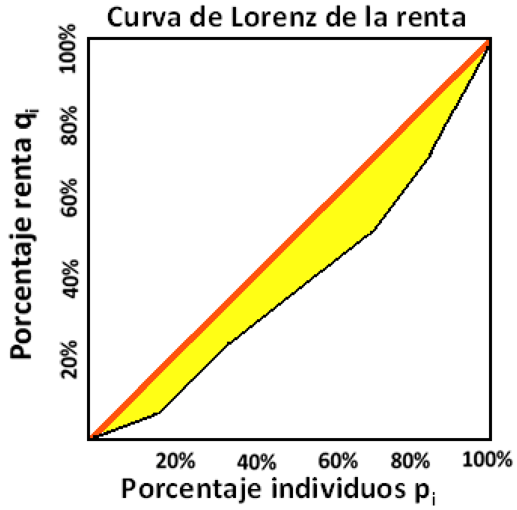
Hallar la concentración de la renta (curva de Lorenz, índice de Gini).

Solución:

Se elabora la tabla:

Rentas $[L_i - L_{i+1})$	x_i	n_i	N_i	$\%p_i = \frac{N_i}{N} 100$	$x_i . n_i$	$u_i = x_i\, n_i$ acumulada	$\%q_i = \frac{u_i}{u_k} 100$	$\%(p_i - q_i)$
3 - 7	5	12	12	15,38	60	60	5,26	10,12
7 - 13	10	18	30	38,46	180	240	21,05	17,41
13 - 17	15	24	54	69,23	360	600	52,63	16,60
17 - 23	20	12	66	84,62	240	840	73,68	10,93
23 - 27	25	12	78	~~100~~	300	1140	~~100~~	~~0~~
		78		207,69	1140		152,63	55,06

• Curva de concentración o curva de Lorenz:



En el caso de *equidistribución* la curva de Lorenz está sobre la diagonal del cuadrado de lado unidad, entonces $p_i = q_i \quad \mapsto \quad I_G = 0$

En el caso de *concentración máxima* (un individuo se lleva el total de los recursos) la curva de Lorenz está sobre los lados del cuadrado.

Entonces $q_1 = q_2 = \cdots = q_{k-1} = 0 \quad \mapsto \quad I_G = 1$

La idea de medir el área da como resultado el llamado *índice de concentración de Gini*, que se define como el área comprendida entre la diagonal y la curva de Lorenz.

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = 1 - \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i} \qquad 0 \le I_G \le 1$$

en este caso,

$$I_G = \frac{\sum_{i=1}^{4}(p_i - q_i)}{\sum_{i=1}^{4} p_i} = \frac{55,06}{207,69} = 0,2651 \quad \text{o bien,} \quad I_G = 1 - \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i} = 1 - \frac{152,63}{207,69} = 0,2651$$

La renta, aunque no equidistribuida, no esta muy concentrada.

En la tabla adjunta se expresa la distribución de rentas de determinada región expresada en 10.000 euros). ¿Qué porcentaje de individuos percibe el 50% de la renta?

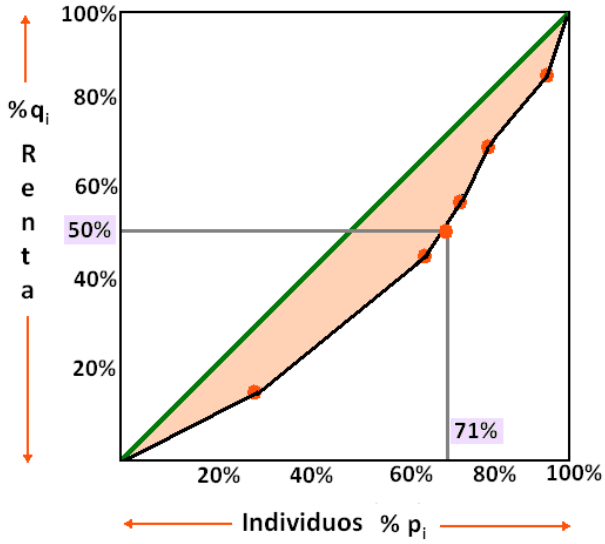
Niveles de renta	$0,5-1,5$	$1,5-2,5$	$2,5-3,5$	$3,5-4,5$	$4,5-5,5$
Número individuos	583	435	194	221	67

Solución:

Nivel Renta	x_i	Individuos n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
0,5 - 1,5	1	583	583	583	583	38,87	17,92
1,5 - 2,5	2	435	1018	870	1453	67,87	44,65
						x	50
2,5 - 3,5	3	194	1212	582	2035	80,80	62,54
3,5 - 4,5	4	221	1433	884	2919	95,53	89,70
4,5 - 5,5	5	67	1500	335	3254	~~100~~	~~100~~
		1500		3254			

Se observa que el 67,86% de los individuos percibe el 44,65% de la renta, y el 80,8% de los individuos percibe el 62,54 % de la renta. En consecuencia, el 50% de la renta se encuentra distribuida entre un conjunto de individuos situado entre el 67,86 y el 80,8%.

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

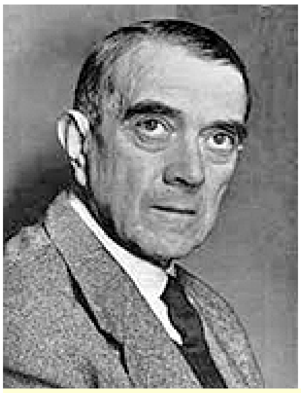


$$\frac{80{,}80 - 67{,}87}{62{,}54 - 44{,}65} = \frac{x - 67{,}87}{50 - 44{,}65} \Rightarrow$$

$$x = 67{,}87 + \frac{5{,}35\,.\,12{,}93}{17{,}89} = 71{,}74\,\%$$

El 50% de la renta se reparte entre el 71,74 % de los individuos.

Max Otto Lorenz (1876-1959) fue un economista estadounidense que desarrolló en 1905 el concepto de *curva de Lorenz* para describir las desigualdades en las rentas. Asesor y empleado en diversas ocasiones por la Oficina del Censo de los Estados Unidos y otras instituciones de información estadística. El término *curva de Lorenz* aparece por primera vez en 1912 en el libro de texto The Elements of Statistical Method.



Corrado Gini (1884-1965) fue un estadístico, demógrafo y sociólogo italiano, ideólogo del fascismo, que desarrolló el *coeficiente de Gini*, una medida de la desigualdad en los ingresos en una sociedad. Generalmente se utiliza para medir la desigualdad en los ingresos dentro de un país, aunque puede utilizarse para medir cualquier forma de distribución desigual.

En la práctica, la renta de muchos países se aproxima a una distribucion Gamma (con parámetro $k < 5$), lo que lleva a índices de Gini entre 0,50 y 0,25.

Los países con índice de Gini superior a 0,50 tienen una distribución aún más desigual que la distribución exponencial.

Operadores de una cadena del sector turístico por sus ventas en plazas hoteleras obtienen los siguientes incentivos mensuales en euros:

Incentivos x_i	100	200	500	1000	1500
Nº operadores n_i	5	6	12	4	3

a) Estudiar la concentración de incentivos.

b) La cadena turística como politica comercial estudia subir a todos los operadores los incentivos: con un incremento porcentual del 10%, o bien con un aumento de 100 euros por operador. ¿Cuál de los dos sería más equitativo?

c) ¿Cuál es la concentración de incentivos si el número de operadores hubiera sido el doble?

Solución:

En todo análisis de concentración debe imperar el Principio del Anonimato: **Si se produce una modificación en una distribución de renta consistente en que dos individuos intercambien sus rentas, el valor del Índice no debe variar.**

a) La concentración de incentivos se analiza mediante el Índice de Gini, que no varía mediante cambios de escala (subida porcentual del 10% a los operadores), mientras queda modificado en cambios de origen (subida lineal de 100 euros a cada operador)

x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
100	5	5	500	500	16,67	3,09
200	6	11	1200	1700	36,67	10,49
500	12	23	6000	7700	76,67	47,53
1000	4	27	4000	11700	90	72,22
1500	3	30	4500	16200	~~100~~	~~100~~
			16200		220	133,33

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i}{\sum_{i=1}^{4} p_i} = 1 - \frac{133,33}{220} = 0,394$$ *concentración de incentivos del 39,4%*

b) SUBIDA DE INCENTIVOS DEL 10% - Cambio de escala en la renta

$x_i^* = 1,1.x_i$	n_i	N_i	$x_i^* n_i$	$u_i^* = x_i^* n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i^* = \frac{u_i^*}{u_k^*}.100$
110	5	5	550	550	16,67	3,09
220	6	11	1320	1870	36,67	10,49
550	12	23	6600	8470	76,67	47,53
1100	4	27	4400	12870	90	72,22
1650	3	30	4950	17820	~~100~~	~~100~~
			17820		220	133,33

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i^*}{\sum_{i=1}^{4} p_i} = 1 - \frac{133,33}{220} = 0,394$$ *concentración de incentivos del 39,4%*

Adviértase que, $q_i = \frac{u_i}{u_k} = \frac{u_i \,.\, 1,1}{u_k \,.\, 1,1} = q_i^*$

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

El cambio de escala en la renta no afecta al Índice de Gini, propiedad conocida como Principio de la Renta relativa: El Índice debe mantenerse invariante frente a las variaciones proporcionales en todas las rentas.

SUBIDA LINEAL DE INCENTIVOS DE 100 EUROS - Cambio de origen en la renta

$x_i^* = 100 + x_i$	n_i	N_i	$x_i^* n_i$	$u_i^* = x_i^* n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i^* = \frac{u_i^*}{u_k^*}.100$
200	5	5	1000	1000	16,67	5,21
300	6	11	1800	2800	36,67	14,58
600	12	23	7200	10000	76,67	52,08
1100	4	27	4400	14400	90	75
1600	3	30	4800	19200	~~100~~	~~100~~
			19200		220	146,88

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i^*}{\sum_{i=1}^{4} p_i} = 1 - \frac{146,88}{220} = 0,332$$ *concentración de incentivos del 33,2%*

Con una subida lineal de 100 euros a cada operador, la equidistribución es más equitativa.

El cambio de origen en la renta afecta al Índice de Gini, propiedad conocida como Principio de Dalton: Toda transferencia de renta de un individuo a otro más rico ha de aumentar el valor de la desigualdad, y recíprocamente toda transferencia de un individuo a otro más pobre ha de reducir el índice, siempre que la ordenación relativa de los individuos se mantenga.

- **En este sentido, si la subida lineal a cada operador hubiera sido de 200 euros se obtendría una equidistribución más equitativa.**

SUBIDA LINEAL DE INCENTIVOS DE 200 EUROS - Cambio de origen en la renta

$x_i^{\bullet} = 200 + x_i$	n_i	N_i	$x_i^{\bullet} n_i$	$u_i^{\bullet} = x_i^{\bullet} n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i^{\bullet} = \frac{u_i^{\bullet}}{u_k^{\bullet}}.100$
300	5	5	1500	1500	16,67	6,76
400	6	11	2400	3900	36,67	17,57
700	12	23	8400	12300	76,67	55,41
1200	4	27	4800	17100	90	77,03
1700	3	30	5100	22200	~~100~~	~~100~~
			22200		220	156,76

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i^{\bullet}}{\sum_{i=1}^{4} p_i} = 1 - \frac{156,76}{220} = 0,287 \quad \textit{concentración de incentivos del 28,7\%}$$

Con una subida lineal de 200 euros a cada operador, la equidistribución resulta más equitativa.

- **Por el contrario, si la cadena del sector turístico hubiera decidido incentivar menos a sus operadores, con una rebaja de 50 euros a cada operador, se tendría una equidistribución menos equitativa.**

REBAJA LINEAL DE INCENTIVOS DE 50 EUROS - Cambio de origen en la renta

$x_i^{\cdot} = x_i - 50$	n_i	N_i	$x_i^{\cdot} n_i$	$u_i^{\cdot} = x_i^{\cdot} n_i$	$\% p_i = \frac{N_i}{N}.100$	$\% q_i^{\cdot} = \frac{u_i^{\cdot}}{u_k^{\cdot}}.100$
50	5	5	250	250	16,67	1,70
150	6	11	900	1150	36,67	7,82
450	12	23	5400	6550	76,67	44,56
950	4	27	3800	10350	90,00	70,41
1450	3	30	4350	14700	~~100~~	~~100~~
			14700		220	124,49

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i^{\cdot}}{\sum_{i=1}^{4} p_i} = 1 - \frac{124,49}{220} = 0,434 \quad \textit{concentración de incentivos del 43,4\%}$$

Con una rebaja lineal de 50 euros a cada operador, la equidistribución resulta menos equitativa.

c) La concentración de incentivos si el número de operadores hubiera sido el doble:

SUBIDA LINEAL DE LA POBLACIÓN - Cambio de escala en la población

x_i	$n_i^{\cdot} = 2 n_i$	$N_i^{\cdot}$	$x_i n_i^{\cdot}$	$u_i^{\cdot} = x_i n_i^{\cdot}$	$\% p_i^{\cdot} = \frac{N_i^{\cdot}}{N}.100$	$\% q_i^{\cdot} = \frac{u_i^{\cdot}}{u_k^{\cdot}}.100$
100	10	10	1000	1000	16,67	3,09
200	12	22	2400	3400	36,67	10,49
500	24	46	12000	15400	76,67	47,53
1000	8	54	8000	23400	90	72,22
1500	6	60	9000	32400	~~100~~	~~100~~
			32400		220	133,33

$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i^{\cdot}}{\sum_{i=1}^{4} p_i^{\cdot}} = 1 - \frac{133,33}{220} = 0,394 \quad \textit{concentración de incentivos del 39,4\%}$$

El cambio de escala en la población no afecta al Índice de Gini, propiedad conocida como Principio de la Población: **Si se multiplica por un mismo escalar el tamaño de todos los conjuntos de individuos con la misma renta, el valor del Índice no debe variar. Es decir, el tamano de la población no importa, lo que interesa son las proporciones de individuos de la población que perciben diferentes niveles de renta.**

A finales del pasado año, una empresa tenía mil seiscientos cincuenta accionistas distribuidos de la siguiente forma:

Número de acciones	0 – 20	20 – 60	60 – 100	100 – 500	500 – 1000
Número de accionistas	1030	380	180	50	10

a) Número medio de acciones por accionista y su desviación típica

b) ¿Cuál es el número de acciones que como máximo posee la mitad del accionariado?

c) Con base estadística, comentar el grado de concentración de las acciones

d) ¿Qué porcentaje del total de las acciones poseen los accionistas mayoritarios?, sabiendo que los accionistas mayoritarios son aquellos que poseen más de 500 acciones.

e) ¿Qué porcentaje de los accionistas minoritarios posee el 20% del total de acciones?

Solución:

a) Sea la variable aleatoria X = "número de acciones"

$[L_i - L_{i+1})$	x_i	n_i	N_i	$x_i n_i$	$x_i^2 n_i$
0 - 20	10	1030	1030 825	10300	103000
20 - 60	40	380	1410	15200	608000
60 - 100	80	180	1590	14400	1152000
100 - 500	300	50	1640	15000	4500000
500 - 1000	750	10	1650	7500	5625000
				62400	11988000

$$a_1 = \bar{x} = \frac{\sum_{i=1}^{5} x_i n_i}{N} = \frac{62400}{1650} = 37,82 \quad \text{número medio de acciones por accionista}$$

$$a_2 = \frac{\sum_{i=1}^{5} x_i^2 n_i}{N} = \frac{11988000}{1650} = 7265,45$$

$$\sigma^2 = a_2 - a_1^2 = 7265,45 - 37,82^2 = 5835,09 \quad \mapsto \quad \sigma = \sqrt{10570,86} = 76,39$$

En muchas ocasiones, la palabra *concentración* se ve como la opuesta a la *dispersión*. Con la curva de Lorenz y el índice de Gini no se contempla solamente a los valores observados, sino que el objeto de análisis será más bien cómo *se reparten* el total de los recursos entre todos los individuos que intervienen en la distribución.

b) El número de acciones que posee como máximo la mitad del accionariado $(1650 / 2) = 825$ es la Mediana: $M_e = 10$

c) El grado de concentración viene expresado por el índice de Gini:

$[L_i - L_{i+1})$	x_i	n_i	N_i	$\%p_i = \frac{N_i}{N}.100$	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\%q_i = \frac{u_i}{u_k}.100$
0 - 20	10	1030	1030	62,42	10300	10300	16,51
20 - 60	40	380	1410	85,45	15200	25500	40,87
60 - 100	80	180	1590	96,36	14400	39900	63,94
100 - 500	300	50	1640	99,39	15000	54900	87,98
500 - 1000	750	10	1650	~~100~~	7500	62400	~~100~~
				343,64	62400		209,3

12%

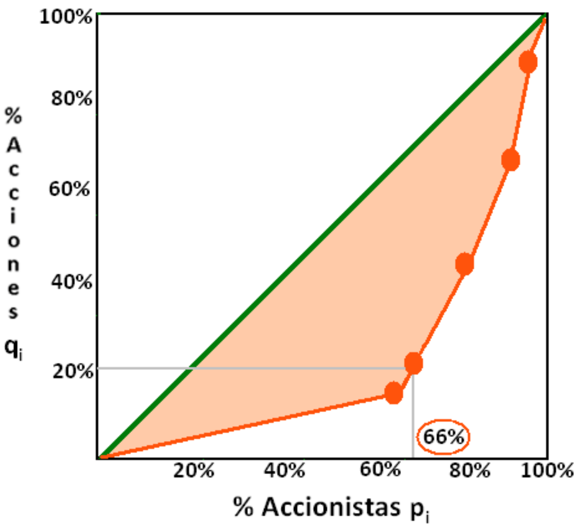
$$I_G = 1 - \frac{\sum_{i=1}^{4} q_i}{\sum_{i=1}^{4} p_i} = 1 - \frac{209,3}{343,64} = 0,3909 \quad \textit{grado de concentración 39,09\%}$$

Cuanto más próximo a cero se encuentre el Índice de Gini más equitativo será el reparto del número de acciones.

d) Los accionistas mayoritarios (más de 500 acciones) poseen el 12% del total de las acciones: $100\% - 87,98\% = 12\%$

e) El porcentaje de accionistas minoritarios que posee el 20% del total de las acciones basta realizar una interpolación:

$[L_i - L_{i+1})$	x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
0 - 20	10	1030	1030	10300	10300	62,42	16,51
						x	20
20 - 60	40	380	1410	15200	25500	85,45	40,87
60 - 100	80	180	1590	14400	39900	96,36	63,94
100 - 500	300	50	1640	15000	54900	99,39	87,98
500 - 1000	750	10	1650	7500	62400	~~100~~	~~100~~



$$\frac{x - 62,42}{20 - 16,51} = \frac{85,45 - 62,42}{40,87 - 16,51}$$

$$x - 62,42 = \frac{3,49.23,33}{24,36} \quad \Rightarrow \quad x = 65,762\,\%$$

La curva de Lorenz presenta coherencia con el índice de Gini calculado, cuanto más próxima se encuentre la curva a la diagonal menor será la concentración y, en consecuencia, más equitativo será el reparto del número de acciones.

El testamento de un hombre de negocios lega 2500 euros a su familia repartiéndose de la forma siguiente: a su cónyuge le asigna el doble que a su hijo primogénito y a éste el doble que a cada uno de sus otros dos hermanos.

a) Considerando que cada heredero ha de aplicar un impuesto de sucesiones proporcional del 20%. ¿Cuáles serán los índices de Gini en los dos casos: antes de pagar impuestos y después de haberlo hecho?. ¿Cuál de las distribuciones es más equitativa?.

b) Si a cada heredero se le aplicase un impuesto fijo de 125 euros, ¿cómo se vería afectado el índice de Gini original?

Solución:

a) Cónyuge: $4x$ Primogénito: $2x$ Cada uno de los dos hijos: x

$x = 2500 / 8 = 312,5$

ANTES DE PAGAR IMPUESTOS

x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
312,5	2	2	625	625	50	25
625	1	3	625	1250	75	50
1250	1	4	1250	2500	~~100~~	~~100~~
					125	75

$$I_G = 1 - \frac{\sum_{i=1}^{2} q_i}{\sum_{i=1}^{2} p_i} = 1 - \frac{75}{125} = 0,40$$

DESPUES DEL PAGO DE LOS IMPUESTOS (20% CADA UNO):

x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
250	2	2	500	500	50	25
500	1	3	500	1000	75	50
1000	1	4	1000	2000	~~100~~	~~100~~
					125	75

$$I_G = 1 - \frac{\sum_{i=1}^{2} q_i}{\sum_{i=1}^{2} p_i} = 1 - \frac{75}{125} = 0,40$$

Las dos distribuciones son igualmente equitativas, lo que las diferencia únicamente es un cambio de escala que no afecta al nivel de concentración, con lo que tienen el mismo índice de Gini.

b) DESPUES DE PAGAR DE LOS IMPUESTOS (125 EUROS CADA UNO):

x_i	n_i	N_i	$x_i n_i$	$u_i = x_i n_i$ acumulada	$\% p_i = \frac{N_i}{N}.100$	$\% q_i = \frac{u_i}{u_k}.100$
187,5	2	2	375	375	50	18,75
500	1	3	500	875	75	43,75
1125	1	4	1125	2000	~~100~~	~~100~~
					125	62,5

$$I_G = 1 - \frac{\sum_{i=1}^{2} q_i}{\sum_{i=1}^{2} p_i} = 1 - \frac{62,5}{125} = 0,50$$

La distribución es más concentrada, menos equitativa, el cambio de origen afecta al índice de Gini. Quitando la misma cantidad a todos, representa una proporción mayor para los que menos reciben.

La tabla refleja los ingresos (millones de euros) por quintiles del turismo en España:

Quintiles	Primero	Segundo	Tercero	Cuarto	Quinto
Turismo 2010	2,1	4,9	8,9	16,8	67,3
Turismo 2011	2	6,3	11,6	20,3	59.8

Calcular los índices de Gini y la curva de Lorenz.

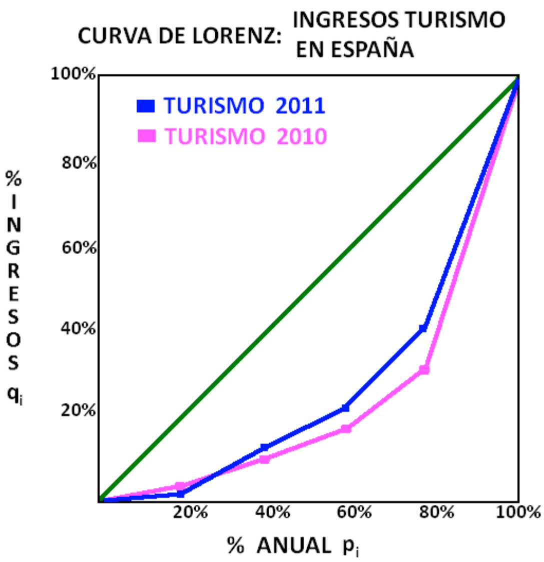
Solución:

INGRESO TURISMO 2010			
u_i	u_i acumulada	$\% p_i$	$\% q_i = \frac{u_i}{u_k}.100$
2,1	2,1	20	2,10
4,9	7	40	7
8,9	15,9	60	15,9
16,8	32,7	80	32,7
67,3	100	~~100~~	~~100~~
		200	57,7

INGRESO TURISMO 2011			
u_i	u_i acumulada	$\% p_i$	$\% q_i = \frac{u_i}{u_k}.100$
2	2	20	2
6,3	8,3	40	8,3
11,6	19,9	60	19,9
20,3	40,2	80	40,2
59,8	100	~~100~~	~~100~~
		200	70,4

$$I_G(2010) = 1 - \frac{\sum_{i=1}^{4} q_i}{\sum_{i=1}^{4} p_i} = 1 - \frac{57,7}{200} = 0,7115 \quad \textit{concentración de ingresos 2010}$$

$$I_G(2011) = 1 - \frac{\sum_{i=1}^{4} q_i}{\sum_{i=1}^{4} p_i} = 1 - \frac{70,4}{200} = 0,648 \quad \textit{concentración de ingresos 2011}$$



La obtención de ingresos por turismo en España es más equitativa en 2011, si bien presenta un alto grado de desigualdad.

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:

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

a) Representar el histograma de la distribución.

b) Calcular el consumo medio de gas de las viviendas. ¿El valor hallado es representativo de la distribución?

c) Calcular el consumo más frecuente.

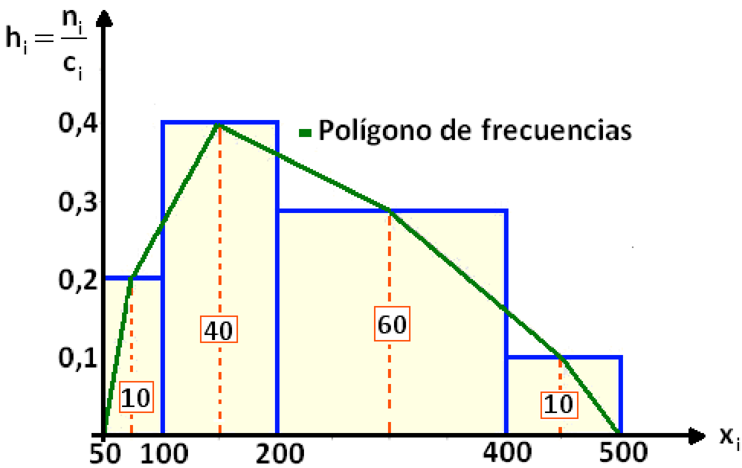
d) Averiguar el valor del tercer cuartil de la distribución del consumo de gas y explicar su significado.

e) Si la factura del gas consiste en una cantidad fija de 20 euros más 0,5 euros por m^3 consumido, calcular la factura media de las viviendas y determinar si la factura es más dispersa que el consumo.

Solución:

$\boxed{a}$ La variable estadística X = "consumo del gas" agrupada en intervalos $[L_i , L_{i+1})$

$[L_i , L_{i+1})$	c_i amplitud	n_i	$h_i = \frac{n_i}{c_i}$ densidad	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	90 110	300	18000	5400000
400 - 500	100	10	0,1	120	450	4500	2025000
						29250	8381250



$\boxed{b}$ **El consumo medio de gas de las viviendas:**

$$a_1 = \overline{x} = \frac{\sum_{i=1}^{4} x_i\, n_i}{N} = \frac{29250}{120} = 243,75 \ \ m^3 \qquad a_2 = \frac{\sum_{i=1}^{4} x_i^2\, n_i}{N} = \frac{8381250}{120} = 69843,75$$

$$\sigma^2 = a_2 - a_1^2 = 69843,75 - (243,75)^2 = 10429,6875$$

$$\sigma_x = \sqrt{10429,6875} = 102,1258 \ \ m^3$$

$$C.V = \frac{\sigma_x}{\overline{x}} = \frac{102,1258}{243,75} = 0,42 \ \ (42\%) \quad \textit{coeficiente variación de Pearson}$$

El consumo medio de gas de las viviendas es de 243,75 m^3, con una dispersión del 42%, con lo que no es muy representativo.

$\boxed{c}$ **El consumo más frecuente se encuentra en el intervalo modal [100-200), que es donde alcanza la máxima densidad de frecuencia.**

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

• **Cuando existen distintas amplitudes, una moda aproximada viene dada por la expresión:** $M_d = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i$

con lo cual, $M_d = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}}\, c_i = 100 + \frac{0,3}{0,2+0,3} 100 = 160 \ m^3$

Si la amplitud de los intervalos es constante: $M_d = L_i + \dfrac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} c_i$

$\boxed{d}$ El tercer cuatil $Q_3 = P_{75}$ percentil 75, siendo $\dfrac{3N}{4} = \dfrac{3.120}{4} = 90$, observando en la tabla la columna de la frecuencia absoluta acumulada N_i, se tiene:

$$Q_3 = P_{75} = L_i + \frac{\frac{3.N}{4} - N_{i-1}}{N_i - N_{i-1}} c_i \quad \mapsto \quad Q_3 = P_{75} = 200 + \frac{90-50}{110-50} 200 = 333,33 \text{ m}^3$$

El 75% de las viviendas que consumen menos, consumen como máximo 333,33 m^3 de gas.

$\boxed{e}$ Según el enunciado, la factura del gas viene expresada por la relación $Y = 20 + 0,5.X$, por lo que la variable estadística X tiene un cambio de origen y de escala.

La factura media: $\overline{Y} = 20 + 0,5.\overline{X} = 20 + 0,5.243,75 = 141,875$ euros

$$\sigma_Y^2 = \text{Var}(20 + 0,5.X) = 0,5^2.\sigma_X^2$$

$$\sigma_Y = 0,5.\sigma_X = 0,5 \,.\, 102,1258 = 51,063 \text{ euros}$$

$$C.V = \frac{\sigma_Y}{\overline{y}} = \frac{51,063}{141,875} = 0,36 \ \ (36\%)$$

La factura del gas esta menos dispersa que el consumo.

$\boxed{\text{NOTA: Cambio de origen y de escala}}$

$$\overline{y} = \frac{\sum_{i=1}^{k} y_i . n_i}{N} = \frac{1}{N}\sum_{i=1}^{k}(a x_i + b).n_i = a\frac{1}{N}\sum_{i=1}^{k} x_i . n_i + b\frac{1}{N}\sum_{i=1}^{k} n_i = a\overline{x} + b \quad \mapsto$$

$$\mapsto \quad E(Y) = E(aX + b) = a\overline{x} + b$$

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

$$\sigma_y^2 = \frac{\sum_{i=1}^{k} (y_i - \overline{y})^2 . n_i}{N} = \frac{\sum_{i=1}^{k} (a x_i + b - a\overline{x} - b)^2 . n_i}{N} = a^2 \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 . n_i}{N} = a^2 \sigma_x^2 \mapsto$$

$$Var(Y) = Var(a x + b) = a^2 . \sigma_x^2$$

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

A partir de la tabla adjunta, siendo $N = 11$ e $\overline{y} = 0$

X \ 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 Y/x e X/y

c) ¿Que parte de la varianza calculada de la variable Y es explicada por la regresión?, ¿Qué parte es debida a causas ajenas?

Solución:

a)

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

$$\overline{y} = \frac{-2 . 3 + 0 + n_{23}}{11} = 0 \quad \mapsto \quad n_{23} = 6$$

De otra parte, $5 + n_{22} + 6 = 11 \quad \mapsto \quad n_{22} = 0$

Las variables X e Y son independientes cuando se verifica:

$$\frac{n_{ij}}{N}=\left(\frac{n_{i\bullet}}{N}\right)\left(\frac{n_{\bullet j}}{N}\right)\ \forall i,j$$

X \ Y	-2	0	1	$n_{i\bullet}$
0	0	1 n_{12}	0	1 $n_{1\bullet}$
1	3	0	6	9
2	0	1	0	1
$n_{\bullet j}$	3	2 $n_{\bullet 2}$	6	11

$$\frac{n_{12}}{N}\neq\left(\frac{n_{1\bullet}}{N}\right)\left(\frac{n_{\bullet 2}}{N}\right)\ \mapsto$$

$$\mapsto \frac{1}{11}\neq\frac{1}{11}\ x\ \frac{2}{11}$$

b) $$a_{11}=\frac{\sum_{i=1}^{3}\sum_{j=1}^{3}x_i\,y_j\,n_{ij}}{N}=\frac{1}{11}\left[-2\,.\,1\,.\,3\ +\ 1\,.\,1\,.\,6\right]=0$$

$$a_{10}=\overline{x}=\frac{\sum_{i=1}^{3}x_i\,n_{i\bullet}}{N}=\frac{1\,.\,9+2\,.\,1}{11}=1 \qquad a_{20}=\frac{\sum_{i=1}^{3}x_i^2\,n_{i\bullet}}{N}=\frac{1}{11}\left[1^2\,.\,9\ +\ 2^2\,.\,1\right]=\frac{13}{11}$$

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

$$a_{01}=\overline{y}=\frac{\sum_{j=1}^{3}y_j\,n_{\bullet j}}{N}=0 \qquad a_{02}=\frac{\sum_{j=1}^{3}y_j^2\,n_{\bullet j}}{N}=\frac{1}{11}\left[(-2)^2\,.\,3\ +\ 1^2\,.\,6\right]=\frac{18}{11}$$

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

covarianza: $m_{11}=\sigma_{xy}=a_{11}-a_{10}\,.\,a_{01}=0-1\,.\,0=0$

- **Coeficiente de regresión de Y sobre X ($b_{Y/x}$ ≡ pendiente de la recta):**

$$b_{Y/x}=\frac{m_{11}}{\sigma_x^2}=\frac{0}{2/11}=0$$

$$\overline{y}=a+b_{Y/x}\,\overline{x}\ \mapsto\ 0=a+0\,.\,1\ \mapsto\ a=0$$

Recta de regresión de Y sobre X: $\begin{cases} y=a+b_{Y/x}\,x & \mapsto\ y=0 \\ y-\overline{y}=b_{Y/x}\,(x-\overline{x}) & \mapsto\ y=0 \end{cases}$

- **Coeficiente de regresión de X sobre Y ($b_{X/y}$ ≡ pendiente de la recta):**

$$b_{X/y} = \frac{m_{11}}{\sigma_Y^2} = \frac{0}{18/11} = 0$$

$$\overline{x} = a' + b_{X/y}\,\overline{y} \quad \mapsto \quad 1 = a' + 0\,.\,0 \quad \mapsto \quad a' = 1$$

Recta de regresión de X sobre Y: $\begin{cases} x = a' + b_{X/y}\, y & \mapsto \quad x = 1 \\ x - \overline{x} = b_{X/y}\,(y - \overline{y}) & \mapsto \quad x = 1 \end{cases}$

c) Como el coeficiente de determación $R^2 = b_{Y/x}\,.\,b_{X/y} = 0 \quad \mapsto$ Las rectas son perpendiculares. En consecuencia, las variables X e Y están incorreladas

Varianza residual de la Y: $\sigma_{rY}^2 = \sigma_y^2\,(1 - r^2) \quad \mapsto \quad \sigma_{rY}^2 = \frac{18}{11}\,(1 - 0) = \frac{18}{11}$

$$\sigma_y^2 = \sigma_{Y\,\text{explicada}}^2 + \sigma_{rY}^2 \quad \mapsto \quad \frac{18}{11} = \sigma_{Y\,\text{explicada}}^2 + \frac{18}{11} \quad \mapsto \quad \sigma_{Y\,\text{explicada}}^2 = 0$$

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 el transporte de un grupo de amigos.

$$y = 4x + 2 \qquad\qquad y = 2x + 10$$

Sabiendo además que la covarianza entre ambas variables $m_{11} = \sigma_{xy} = 60$

a) Identificar cuál es la recta de regresión de Y/x y de X/y.

b) Interpretar los coeficientes de las rectas de regresión.

c) Porcentaje de variabilidad explicada y no explicada por la recta.

d) Calcular la varianza residual en la regresión Y/x , ¿coincidirá con la varianza residual en la regresión X/y?. Justificar la respuesta.

Solución:

$\boxed{\text{a)}}$ Se analiza según los valores del coeficiente de determinación lineal

- $$y = 4x + 2 \xrightarrow[\text{regresión X/y}]{\text{recta}} x = -\frac{1}{2} + \frac{1}{4}y \mapsto \begin{array}{l} a = -\dfrac{1}{2} \\ b_{X/y} = \dfrac{1}{4} \end{array}$$

  $$y = 2x + 10 \xrightarrow[\text{regresión Y/x}]{\text{recta}} \begin{array}{l} a^{\bullet} = 10 \\ b_{Y/x} = 2 \end{array}$$

Coeficiente de determinación: $R^2 = b_{X/y}.b_{Y/x} = \frac{1}{4} \times 2 = 0,5 < 1$

- $$y = 4x + 2 \xrightarrow[\text{regresión Y/x}]{\text{recta}} \begin{array}{l} a = 2 \\ b_{Y/x} = 4 \end{array}$$

  $$y = 2x + 10 \xrightarrow[\text{regresión X/y}]{\text{recta}} x = -5 + \frac{1}{2}y \mapsto \begin{array}{l} a^{\bullet} = -5 \\ b_{X/y} = \dfrac{1}{2} \end{array}$$

Coeficiente de determinación: $R^2 = b_{X/y}.b_{Y/x} = \frac{1}{2} \times 4 = 2 > 1$

cosa que no es posible pues $0 \le R^2 \le 1$

La recta de regresión de Y/x: $y = 2x + 10$

$\boxed{\text{b)}}$ La recta de regresión de Y sobre X tiene mayor pendiente $b_{yx} = 2$

$\boxed{\text{c)}}$ Al ser el coeficiente de determinación lineal $R^2 = 0,5$, la recta de regresión de Y sobre X $(y = 2x + 10)$ explica el 50% de la variabilidad dependiente y el otro 50% es no explicado.

$\boxed{\text{d)}}$ $$\begin{cases} b_{Y/x} = \dfrac{m_{11}}{\sigma_x^2} \mapsto 2 = \dfrac{60}{\sigma_x^2} \mapsto \sigma_x^2 = 30 \\ b_{X/y} = \dfrac{m_{11}}{\sigma_y^2} \mapsto \dfrac{1}{4} = \dfrac{60}{\sigma_y^2} \mapsto \sigma_y^2 = 240 \end{cases}$$

varianzas residuales: $$\begin{cases} \sigma_{rx}^2 = \sigma_x^2.(1 - R^2) \mapsto \sigma_{rx}^2 = 30.(1 - 0,5) = 15 \\ \sigma_{ry}^2 = \sigma_y^2.(1 - R^2) \mapsto \sigma_{ry}^2 = 240.(1 - 0,5) = 120 \end{cases}$$

Las varianzas residuales no coinciden: $\sigma^2_{rx} = 15 \neq \sigma^2_{ry} = 120$

De una distribución bidimensional (X,Y) se sabe que al aumentar los valores de X aumentan los de Y. Se han obtenido la recta de regresión mínimo cuadrática de Y sobre X y se ha comprobado que la varianza residual es cero. Además se tienen los valores de los momentos respecto al origen siguientes:

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

a) Determinar la varianza debida a la regresión de Y sobre X y el valor de la covarianza.

b) Se hace el cambio de variable $X^{\bullet} = 2X$. Si se obtiene la nueva recta de regresión de Y sobre $X^{\bullet}$ ¿será bueno el ajuste?. Razonar la respuesta.

c) Se decide cambiar la función de ajuste de Y sobre X por una constante, $Y = c$. Utilizando el método de los mínimos cuadrados, determinar el valor de esta constante para el caso.

Solución:

a) Las varianzas de las variables X e Y, respectivamente, son:

$\boxed{a)}$ **Las varianzas de las variables X e Y, respectivamente, son:**

$$\begin{cases} \sigma^2_x = a_{20} - a^2_{10} = 40 - 2^2 = 36 \\ \sigma^2_y = a_{02} - a^2_{01} = 125 - 10^2 = 25 \end{cases}$$

Siendo $\sigma^2_{ry} = \sigma^2_y\,(1 - R^2) = 0 \;\rightarrow\; 1 - R^2 = 0 \;\mapsto\; R^2 = 1$ **existe una dependencia funcional, el ajuste es perfecto.**

Para calcular la covarianza $m_{11} = \sigma_{xy}$ **se tiene en cuenta que el coeficiente de correlación lineal** $r = \sqrt{R^2} = 1$**:**

$$r = \sqrt{b_{Y/x} \,.\, b_{X/y}} = \sqrt{\frac{m_{11}}{\sigma^2_x} \,.\, \frac{m_{11}}{\sigma^2_y}} = \frac{m_{11}}{\sigma_x \,.\, \sigma_y} \quad \mapsto \quad m_{11} = \sigma_x \,.\, \sigma_y = \sqrt{36 \,.\, 25} = 30$$

$\boxed{b)}$ **El coeficiente de determinación R^2 es invariante ante un cambio de origen y de escala, con lo que la bondad del ajuste ante el cambio de variable $X^{\bullet} = 2X$ es idéntico.**

$\boxed{\text{c)}}$ $E(y)=E(c) \quad \mapsto \quad c=\overline{y}$

Sea una distribución bidimensional (X,Y). Con un cambio de origen y de escala se introducen nuevas variables estadísticas $(X^{\bullet}, Y^{\bullet})$ relacionadas con las anteriores, de forma que $\begin{cases} X^{\bullet}=n+mX \\ Y^{\bullet}=q+pY \end{cases}$

Momentos variable estadística (X,Y): $a_{10}, a_{20}, a_{01}, a_{02}, a_{11}, m_{20}, m_{02}, m_{11}$

Momentos variable estadística $(X^{\bullet}, Y^{\bullet})$: $a_{10}^{\bullet}, a_{20}^{\bullet}, a_{01}^{\bullet}, a_{02}^{\bullet}, a_{11}^{\bullet}, m_{20}^{\bullet}, m_{02}^{\bullet}, m_{11}^{\bullet}$

□ Las medias se ven afectadas por el cambio de origen y de escala.

$$a_{10}^{\bullet}=\overline{x}^{\bullet}=\frac{\sum_{i=1}^{k} x_i^{\bullet}}{N}=\frac{1}{N}\sum_{i=1}^{k}(n+mx_i)=\frac{1}{N}\sum_{i=1}^{k} n+m\,\frac{1}{N}\sum_{i=1}^{k} x_i=n+m\,a_{10}$$

$$\boxed{\overline{x}^{\bullet}=n+m\,a_{10}=n+m\,\overline{x}}$$

Análogamente $a_{01}^{\bullet}=q+p\,a_{01} \quad \mapsto \quad \boxed{\overline{y}^{\bullet}=q+p\,a_{01}=q+p\,\overline{y}}$

□ Las varianzas se ven afectadas por el cambio de escala.

$$m_{20}^{\bullet}=\sigma_{x^{\bullet}}^2=\frac{1}{N}\sum_{i=1}^{k}(x_i^{\bullet}-a_{10}^{\bullet})^2=\frac{1}{N}\sum_{i=1}^{k}(\cancel{n}+mx_i-\cancel{n}-m\,a_{10})^2=m^2\,\frac{1}{N}\sum_{i=1}^{k}(x_i-a_{10})^2=$$

$$=m^2\,\sigma_x^2=m^2\,m_{20} \quad \mapsto \quad \boxed{m_{20}^{\bullet}=m^2\,m_{20} \quad \leftrightarrow \quad \sigma_{x^{\bullet}}^2=m^2\,\sigma_x^2}$$

Análogamente $m_{02}^{\bullet}=p^2\,m_{02} \quad \leftrightarrow \quad \sigma_{y^{\bullet}}^2=p^2\,\sigma_y^2$

□ La covarianza es invariante ante un cambio de origen, pero no ante un cambio de escala.

$$m_{11}^{\bullet}=\frac{1}{N}\sum_{i=1}^{k}(x_i^{\bullet}-a_{10}^{\bullet}).(y_i^{\bullet}-a_{01}^{\bullet})=\frac{1}{N}\sum_{i=1}^{k}(\cancel{n}+mx_i-\cancel{n}-m\,a_{10}).(\cancel{q}+p\,y_i-\cancel{q}-p\,a_{01})=$$

$$=mp\,\frac{1}{N}\sum_{i=1}^{k}(x_i-a_{10}).(y_i-a_{01})=mp\,m_{11} \quad \mapsto \quad \boxed{m_{11}^{\bullet}=mp\,m_{11}}$$

□ **Los coeficientes de regresión son invariantes ante un cambio de origen, pero no ante un cambio de escala.**

Sean los coeficientes de regresión de la recta Y sobre X ($b_{Y/x}$) y de la recta X sobre Y ($b_{X/y}$)

Análogamente, los coeficientes de regresión de la recta $Y^{\bullet}$ sobre $X^{\bullet}$ ($b^{\bullet}_{Y^{\bullet}/x^{\bullet}}$) y de la recta $X^{\bullet}$ sobre $Y^{\bullet}$ ($b^{\bullet}_{X^{\bullet}/y^{\bullet}}$)

$$b^{\bullet}_{Y^{\bullet}/x^{\bullet}} = \frac{m^{\bullet}_{11}}{m^{\bullet}_{20}} = \frac{\cancel{m}\, p\,.\,m_{11}}{m^{\cancel{2}}\, m_{20}} = \frac{p}{m}\,.\,\frac{m_{11}}{m_{20}} = \frac{p}{m}\,.\,b_{Y/x} \quad \mapsto \quad \boxed{b^{\bullet}_{Y^{\bullet}/x^{\bullet}} = \frac{p}{m}\,.\,b_{Y/x}}$$

Análogamente $\boxed{b^{\bullet}_{X^{\bullet}/y^{\bullet}} = \frac{m}{p}\,.\,b_{X/y}}$

□ **El coeficiente de determinación R^2 es invariantes ante un cambio de origen y ante un cambio de escala.
En consecuencia, también lo será el coeficiente de correlación r.**

$$R^{\bullet^2} = b^{\bullet}_{Y^{\bullet}/x^{\bullet}} \text{ x } b^{\bullet}_{X^{\bullet}/y^{\bullet}} = \left(\frac{p}{m}\, b_{Y/x}\right) \text{ x } \left(\frac{m}{p}\, b_{X/y}\right) = b_{Y/x} \text{ x } b_{X/y} = R^2$$

a) La variable estadística X tiene $\overline{x}=2$ y $\sigma_x=1$. Determinar la media aritmética, varianza y coeficiente de variación de Pearson de $Y=\dfrac{X-1}{2}$

b) La varianza explicada por una regresión lineal simple es el doble de la varianza residual. ¿Cuánto vale el coeficiente de deteminación?

Solución:

a) $$\overline{y}=E(Y)=E\left[\frac{1}{2}X-\frac{1}{2}\right]=\frac{1}{2}.E(X)-\frac{1}{2} \quad \mapsto \quad \overline{y}=\frac{1}{2}.\overline{x}-\frac{1}{2}=\frac{1}{2}$$

$$\sigma_Y^2=\text{Var}\left[\frac{1}{2}X-\frac{1}{2}\right]=\left(\frac{1}{2}\right)^2.\text{Var}(X)=\frac{1}{4}\sigma_x^2 \quad \mapsto \quad \sigma_Y^2=\frac{1}{4} \quad \mapsto \quad \sigma_Y=\frac{1}{2}$$

$$C.V_Y=\frac{\sigma_Y}{\overline{y}}=\frac{1/2}{1/2}=1$$

b) $$\sigma_y^2=\sigma_{explicada}^2+\sigma_{ry}^2=2\sigma_{ry}^2+\sigma_{ry}^2=3\sigma_{ry}^2 \quad \mapsto \quad \sigma_y^2=3\sigma_{ry}^2$$

$$\sigma_{ry}^2=\sigma_y^2\,(1-R) \quad \rightarrow \quad R=1-\frac{\sigma_{ry}^2}{\sigma_y^2}=1-\frac{\cancel{\sigma_{ry}^2}}{3\,\cancel{\sigma_{ry}^2}}=1-\frac{1}{3}=\frac{2}{3}$$

Se han tomado cinco muestras de glucógeno, de una cantidad fija cada una. Se les ha aplicado una cantidad X de glucogenasa (en milimoles/ litro) anotando en cada caso la velocidad de reacción Y medida en μ-moles/minuto. Obteniéndose la siguiente tabla:

X	1	2	3	0,2	0,5
Y	18	35	60	8	10

Sabiendo que la reacción no se ajusta al modelo de Michaelis-Menten, se pide:

a) ¿Se deduce de estos datos que la velocidad de reacción aumenta con la concentración de glucogenasa?. Justificar la respuesta.

b) ¿Cuál es el grado de predicción?

Solución:

a) La recta de regresión de Y sobre X: $y - \bar{y} = \dfrac{m_{11}}{\sigma_x^2}(x - \bar{x})$

$$a_{10} = \bar{x} = \frac{\sum_{i=1}^{5} x_i\, n_{i\bullet}}{N} = 1,34 \qquad a_{20} = \frac{\sum_{j=1}^{5} x_i^2\, n_{i\bullet}}{N} = 2,858 \qquad a_{11} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{5} x_i y_i\, n_{ij}}{N} = 54,92$$

$$a_{01} = \bar{y} = \frac{\sum_{j=1}^{5} y_{\bullet j}\, n_{\bullet j}}{N} = 26,2 \qquad a_{02} = \frac{\sum_{j=1}^{5} y_{\bullet j}^2\, n_{\bullet j}}{N} = 1062,6$$

$$\sigma_x^2 = a_{20} - (a_{10})^2 = 2,858 - 1,7956 = 1,0624$$

$$\sigma_y^2 = a_{02} - (a_{01})^2 = 1062,6 - 686,44 = 376,16$$

$$m_{11} = \sigma_{yx} = a_{11} - a_{10}\, a_{01} = 54,92 - 1,34\ 26,2 = 19,812$$

con lo cual, $y - 26,2 = \dfrac{19,812}{1,0624}(x - 1,34) \;\mapsto\; \hat{y} = 1,211 + 18,648\, x$

Como la pendiente $b_{Y/x} = 18,648 > 0$ se puede afirmar que la velocidad aumenta con la concentración.

b) El grado de ajuste viene dado por el coeficiente de correlación lineal:

$$r = \frac{m_{11}}{\sigma_x\,\sigma_y} = \frac{19{,}812}{\sqrt{1{,}0624}\;\sqrt{376{,}16}} = 0{,}991$$ **grado de ajuste del 99,1%**

- **SPSS: Recta de regresión de Y sobre X:** **Menú Analizar - Regresión Lineal**

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,991[a]	,982	,976	3,341

a. Variables predictoras: (Constante), GLUCOGENASA

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1847,305	1	1847,305	165,454	,001[a]
	Residual	33,495	3	11,165		
	Total	1880,800	4			

a. Variables predictoras: (Constante), GLUCOGENASA

b. Variable dependiente: MOLES

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	1,211	2,451		,494	,655
	GLUCOGENASA	18,648	1,450	,991	12,863	,001

a. Variable dependiente: MOLES

x_i	y_i	$\hat{y} = 1{,}211 + 18{,}648\,x$	$(y_i - \bar{y})^2$	$(y_i - \hat{y}_i)^2$	$(\hat{y} - \bar{y})^2$
1	18	19,859	67,24	3,456	40,23
2	35	38,507	77,44	12,299	151,46
3	60	57,155	1142,44	8,094	958,24
0,2	8	4,9406	331,24	9,360	451,98
0,5	10	10,535	262,44	0,286	245,39
			1880,8 SCT	33,495 SCR	1847,30 SCE

$SCE = \sum(\hat{y}_i - \bar{y})^2 = 1847{,}305 \qquad \sigma_e^2 = SCE/5 = 369{,}461$

$SCR = \sum(y_i - \hat{y}_i)^2 = 33{,}495 \qquad \sigma_r^2 = SCR/5 = 6{,}6991 \qquad SCT = SCR + SCE$

$SCT = \sum(y_i - \bar{y})^2 = 1880{,}8 \qquad \sigma_y^2 = SCT/5 = 376{,}16$

$y = 1{,}211 + 18{,}648\,x$

$$R^2 = \frac{SCE}{SCT} = \frac{1847,305}{1880,8} = 0,982 \qquad r = \sqrt{0,982} = 0,991$$

$$\underset{\text{SCT}}{\overset{\text{Variación TOTAL}}{\sum (y_i - \overline{y})^2}} = \underset{\text{SCR}}{\overset{\text{Variación No explicada}}{\sum (y_i - \hat{y}_i)^2}} + \underset{\text{SCE}}{\overset{\text{Variación Explicada}}{\sum (\hat{y}_i - \overline{y})^2}} \mapsto$$

$$\mapsto \; 1 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2} + \overset{\text{coeficiente determinación } R^2}{\frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2}}$$

$$R^2 = 1 - \frac{\overbrace{\sum (y_i - \hat{y}_i)^2 / N}^{\sigma^2_{ry}}}{\underbrace{\sum (y_i - \overline{y})^2 / N}_{\sigma^2_y}} = 1 - \frac{\sigma^2_r}{\sigma^2_y} \mapsto R^2 = 1 - \frac{\sigma^2_{ry}}{\sigma^2_y} \mapsto \sigma^2_r = \sigma^2_{ry}(1 - R^2)$$

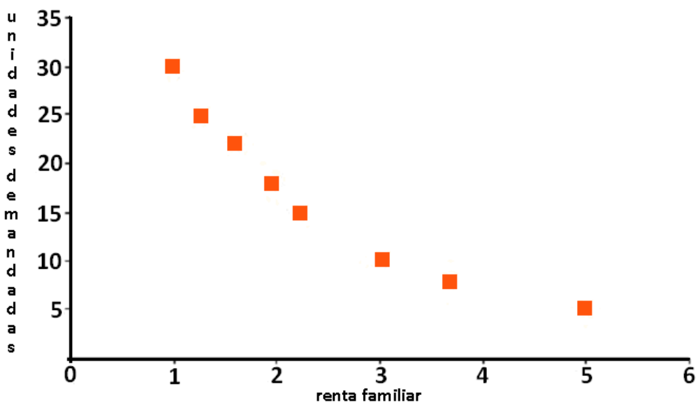
Estudiando las unidades que se demandan de cierto producto de consumo (Y, en miles) y las rentas familiares (X) en miles de euros, se tiene:

Rentas familiares	1	1,3	1,6	2	2,2	3	3,7	5
Producto consumo	30	25	22	18	15	10	8	5

Ajustar una línea de regresión al número de unidades del producto de consumo (Y) en función de las rentas familiares (X). ¿Es fiable el ajuste?.

Solución:

El diagrama de puntos sugiere un ajuste de tipo hiperbólico.



La función a ajustar será:

$$y = a + \frac{b}{x}$$

Se resuelve el sistema de ecuaciones normales:

$$\begin{cases} a\,.\,N + b\,.\displaystyle\sum_{i=1}^{n}\frac{1}{x_i} = \sum_{i=1}^{n} y_i \\ a\,.\displaystyle\sum_{i=1}^{n}\frac{1}{x_i} + b\,.\sum_{i=1}^{n}\frac{1}{x_i^2} = \sum_{i=1}^{n}\frac{1}{x_i}\,.\,y_i \end{cases}$$

rentas x_i	unidades y_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\frac{1}{x_i}\,.\,y_i$	$\hat{y}_i$	SCT $(y_i-\overline{y})^2$	SCE $(y_i-\hat{y}_i)^2$
1	30	1	1	30	32,026	178,891	4,103
1,3	25	0,769	0,592	19,231	24,636	70,141	0,132
1,6	22	0,625	0,391	13,750	20,018	28,891	3,929
2	18	0,500	0,250	9,000	16,015	1,891	3,940
2,2	15	0,455	0,207	6,818	14,560	2,641	0,194
3	10	0,333	0,111	3,333	10,678	43,891	0,460
3,7	8	0,270	0,073	2,162	8,659	74,391	0,434
5	5	0,200	0,040	1,000	6,409	135,141	1,985
19,8	133	4,152	2,663	85,294		535,875	15,178

Se tiene el sistema: $\begin{cases} 8\,.\,a + 4{,}1524\,.\,b = 133 \\ 4{,}1524\,.\,a + 2{,}6631\,.\,b = 85{,}2944 \end{cases} \mapsto \begin{array}{l} a = 0{,}00459 \\ b = 32{,}02098 \end{array}$

Regresión hiperbólica: $y = 0{,}00459 + 32{,}0209819\,.\,\dfrac{1}{x}$

con lo cual, $\hat{y}_i = 0{,}00459 + 32{,}0209819\,.\,\dfrac{1}{x_i}$

Para analizar la bondad del ajuste hay que calcular el coeficiente de determinación.
Para ello, se calcula la suma de los cuadrados de las variaciones total y residual:

$$SCT = \sum_{i=1}^{8}(y_i-\overline{y})^2 = 535{,}875 \qquad SCE = \sum_{i=1}^{8}(y_i-\hat{y}_i)^2 = 15{,}1776$$

$$SCR = SCT - SCE = 535{,}875 - 15{,}178 = 520{,}697$$

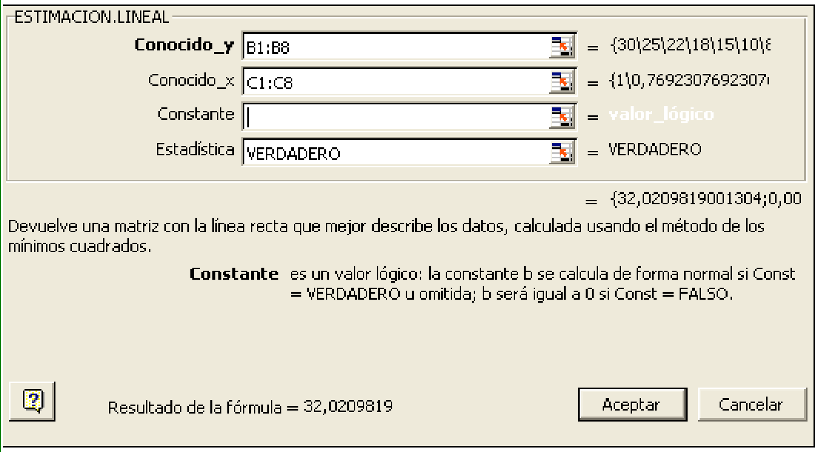
$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{15{,}1776}{535{,}875} = 0{,}9716$$

El coeficiente de determinación es del 97,16%, indicando que el ajuste mediante la hipérbola equilátera es bueno.

• EXCEL:

ESTIMACION.LINEAL es una función matricial, por lo que antes de introducir la función debe seleccionarse el rango de las celdas en las que se quiera que aparezcan los resultados (la dimensión máxima que devuelve Excel cuando se trabaja con una sola variable independiente es 5x2).
Finalmente, se pulsa simultáneamente la combinación de teclas (Control + Mayúsculas + Intro)

rentas x_i	unidades y_i	$z = \frac{1}{x_i}$
1	30	1
1,3	25	0,769
1,6	22	0,625
2	18	0,500
2,2	15	0,455
3	10	0,333
3,7	8	0,270
5	5	0,200
19,8	133	4,152



SALIDA EXCEL

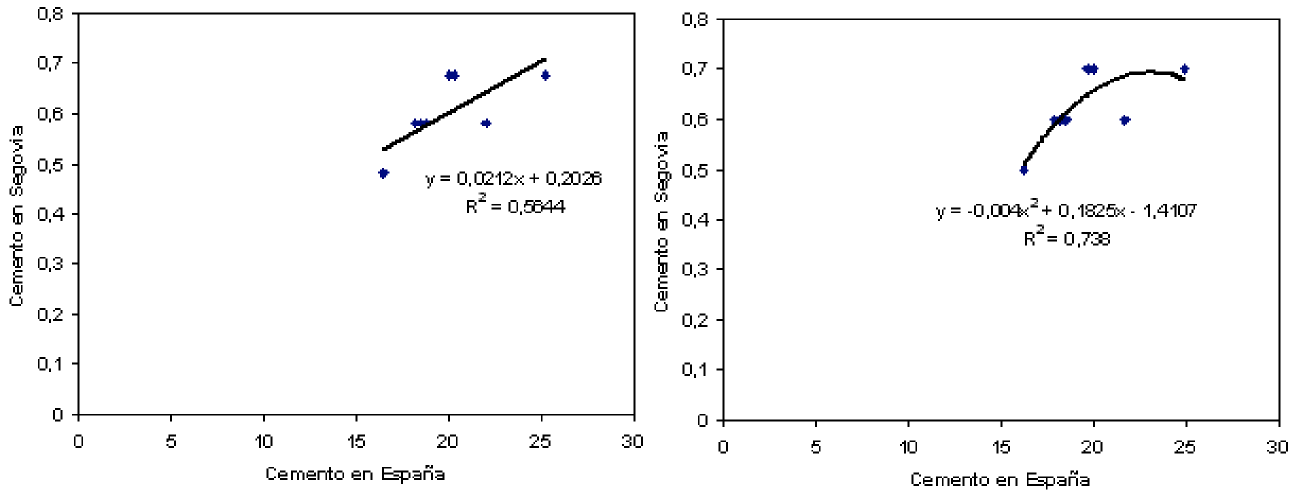
$b = 32,02098$	$a = 0,00459$
$ET_b = 2,23186$	$ET_a = 1,28771$
$R^2 = 0,97168$	$ET_{regresión} = 1,59047$
$F = 205,84194$	$g.l = 6$
$SCR = 520,69741$	$SCE = 15,17759$
$SCT = SCR + SCE = 535,875$	

Regresión hiperbólica: $y = 0,00459 + 32,0209819 \, . \, \frac{1}{x}$

En la tabla adjunta se recogen las ventas de cemento en Segovia y en todo el territorio español. ¿Qué ajuste será mejor para determinar las ventas de cemento en Segovia en función de toda España, el lineal o el parabólico?

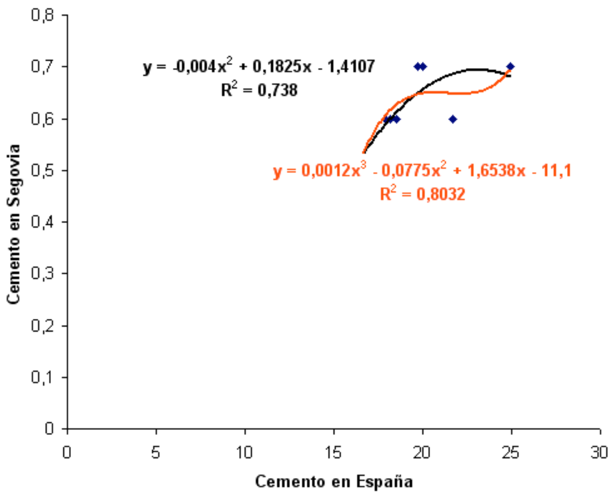
Año	España	Segovia
2000	19,7	0,7
2001	18,5	0,6
2002	18,5	0,6
2003	17,9	0,6
2004	16,2	0,5
2005	16,2	0,5
2006	18,2	0,6
2007	20	0,7
2008	21,7	0,6
2009	24,9	0,7

Solución:



Con Excel se obtiene un diagrama de dispersión y se dibuja una línea de tendencia Lineal o Polinomial presionando con el botón derecho del ratón sobre uno de los puntos de dispersión, en la pestaña Opciones se elige el grado del polinomio, se presenta ecuación en el gráfico y el coeficiente de determinación.

El coeficiente de determinación o grado de ajuste $R^2 = 0{,}738$ en el caso de un polinomio de grado dos (una parábola) es mayor que el coeficiente de determinación lineal $R^2 = 0{,}5644$, en consecuencia es aconsejable un ajuste parabólico.



Adviértase que si se hubiera ajuste un polinomio de grado tres el ajuste hubiera mejorado, con un grado de ajuste $R^2 = 0,8032$

La ecuación a ajustar por mínimos cuadrados que explica las ventas de cemento en Segovia (Y) en función de las de España (X) viene dada por la ecuación $y = a + bx + cx^2$, donde (a, b y c) son los parámetros a estimar.

Las ecuaciones normales quedan:

$$\begin{cases} aN + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2\, y_i \end{cases} \mapsto \begin{pmatrix} a \\ b \\ c \end{pmatrix} . \begin{pmatrix} N & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i\, y_i \\ \sum_{i=1}^{n} x_i^2\, y_i \end{pmatrix}$$

Años	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i\, y_i$	$x_i^2\, y_i$
2000	19,7	0,7	388,090	7645,373	150613,848	13,79	271,663
2001	18,5	0,6	342,250	6331,625	117135,063	11,1	205,35
2002	18,5	0,6	342,250	6331,625	117135,063	11,1	205,35
2003	17,9	0,6	320,410	5735,339	102662,568	10,74	192,246
2004	16,2	0,5	262,440	4251,528	68874,754	8,1	131,22
2005	16,2	0,5	262,440	4251,528	68874,754	8,1	131,22
2006	18,2	0,6	331,240	6028,568	109719,938	10,92	198,744
2007	20	0,7	400	8000	160000	14	280
2008	21,7	0,6	470,890	10218,313	221737,392	13,02	282,534
2009	24,9	0,7	620,010	15438,249	384412,400	17,43	434,007
	191,8	6,1	3740,020	74232,148	1501165,778	118,3	2332,334

$N = 10 \quad \sum_{i=1}^{10} x_i = 191,80 \quad \sum_{i=1}^{10} x_i^2 = 3740,020 \quad \sum_{i=1}^{10} x_i^3 = 74232,148$

$\sum_{i=1}^{10} x_i^4 = 1501165,778$

$$\sum_{i=1}^{10} y_i = 6,1 \quad \sum_{i=1}^{10} x_i \,.\, y_i = 118,30 \quad \sum_{i=1}^{10} x_i^2 \, y_i = 2332,334$$

Planteado el sistema de ecuaciones:

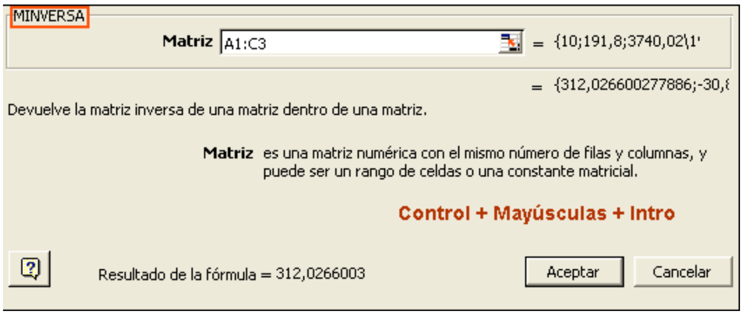
$$\begin{cases} 10\,.\,a + b\,.\,191,80 + c\,.\,3740,020 = 6,1 \\ a\,.\,191,80 + b\,.\,3740,020 + c\,.\,74232,148 = 118,3 \\ a\,.\,3740,020 + b\,.\,74232,148 + c\,.\,1501165,778 = 2332,334 \end{cases}$$

Se puede resolver por varios métodos, en forma matricial:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} . \begin{pmatrix} 10 & 191,80 & 3740,020 \\ 191,80 & 3740,020 & 74232,148 \\ 3740,020 & 74232,148 & 1501165,778 \end{pmatrix} = \begin{pmatrix} 6,1 \\ 118,3 \\ 2332,334 \end{pmatrix} \mapsto$$

$$\mapsto \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 10 & 191,80 & 3740,020 \\ 191,80 & 3740,020 & 74232,148 \\ 3740,020 & 74232,148 & 1501165,778 \end{pmatrix}^{-1} \begin{pmatrix} 6,1 \\ 118,3 \\ 2332,334 \end{pmatrix}$$

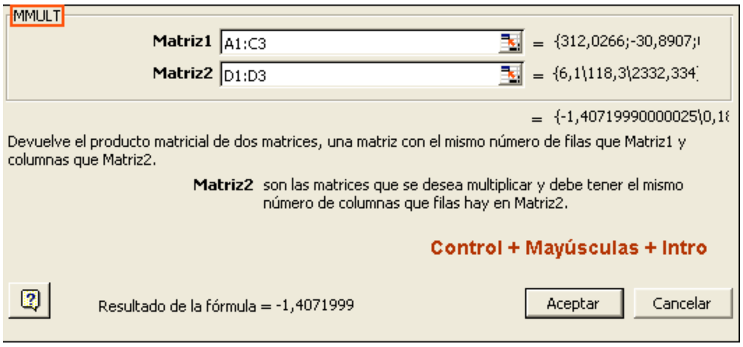
La matriz inversa se obtiene utilizando Excel con la opción MINVERSA. Como es una función matricial se tiene que seleccionar antes el rango de las celdas donde aparece el resultado.



$$\begin{pmatrix} 10 & 191,80 & 3740,020 \\ 191,80 & 3740,020 & 74232,148 \\ 3740,020 & 74232,148 & 1501165,778 \end{pmatrix}^{-1} = \begin{pmatrix} 312,027 & -30,891 & 0,750 \\ -30,891 & 3,073 & -0,075 \\ 0,750 & -0,075 & 0,002 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 312,027 & -30,891 & 0,750 \\ -30,891 & 3,073 & -0,075 \\ 0,750 & -0,075 & 0,002 \end{pmatrix} \begin{pmatrix} 6,1 \\ 118,3 \\ 2332,334 \end{pmatrix}$$

El producto de matrices se obtiene utilizando Excel con la opción MMULT.
Como es una función matricial se tiene que seleccionar antes el rango de las celdas donde aparece el resultado.



$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -1,4107 \\ 0,1825 \\ -0,0040 \end{pmatrix}$$

La ecuación de segundo grado (parábola): $y = -1,4107 + 0,1825\,x - 0,004\,x^2$

La bondad de ajuste del modelo se calcula mediante el coeficiente de determinación $R^2 = 1 - \dfrac{SCE}{SCT}$

Años	x_i	y_i	Segovia $\hat{y}_i$	SCT $(y_i - \bar{y})^2$	SCE $(\hat{y}_i - \bar{y})^2$
2000	19,7	0,7	0,6492	0,0081	0,0026
2001	18,5	0,6	0,6115	0,0001	0,0001
2002	18,5	0,6	0,6115	0,0001	0,0001
2003	17,9	0,6	0,5884	0,0001	0,0001
2004	16,2	0,5	0,5075	0,0121	0,0001
2005	16,2	0,5	0,5075	0,0121	0,0001
2006	18,2	0,6	0,6003	0,0001	0,0000
2007	20,0	0,7	0,6568	0,0081	0,0019
2008	21,7	0,6	0,6866	0,0001	0,0075
2009	24,9	0,7	0,6807	0,0081	0,0004
	191,8	6,1		0,0490	0,0128

$$SCT = \sum_{i=1}^{10} (y_i - \bar{y})^2 = 0,04900 \qquad SCE = \sum_{i=1}^{10} (\hat{y}_i - \bar{y})^2 = 0,01284$$

Coeficiente de determinación $R^2 = 1 - \dfrac{SCE}{SCT} = 1 - \dfrac{0,01284}{0,04900} = 0,7380$

El grado de ajuste es del 73,80%

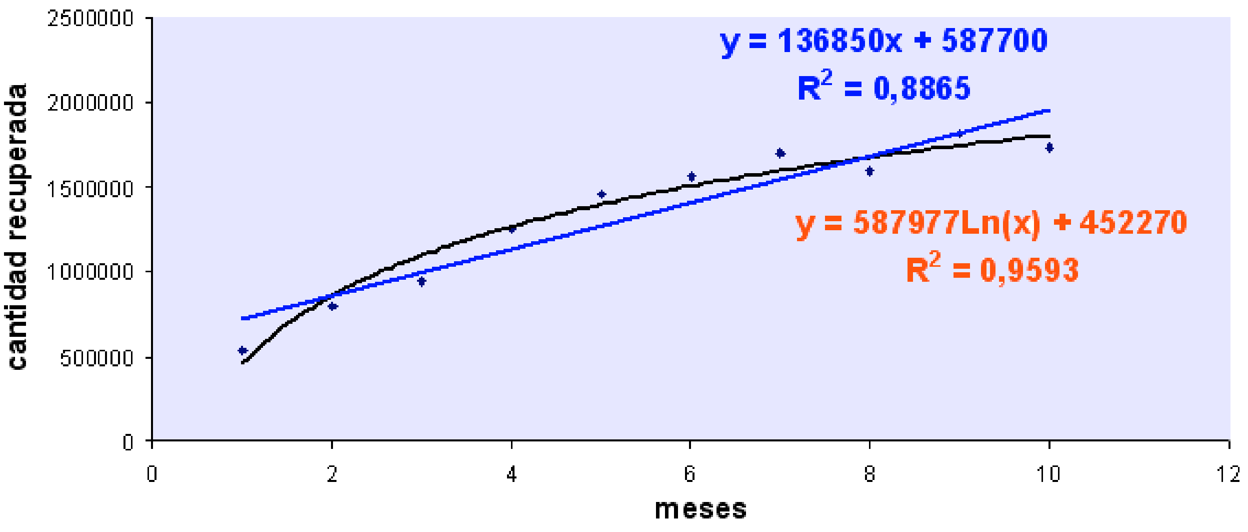
Una entidad bancaria ofrece un fondo de inversión con una duración máxima de dos años y con un riesgo alto el primer año.
Como información, ofrece la tabla adjunta, donde aparece el dinero (en euros) que podría haber recuperado una persona al haber cancelado su inversión al cabo de un número determinado de meses a partir de su inversión inicial.

Tiempo (meses)	Cantidad recuperada
1	537984,4
2	800078,5
3	942209,2
4	1251083,7
5	1461097
6	1565418
7	1703179,4
8	1595651,9
9	1814292,5
10	1732753,9

a) Determinar un modelo explicativo para los resultados en función del tiempo.
b) ¿Qué resultados podría pronosticar si una persona retira su inversión en el mes de octubre?, ¿con qué grado de predicción?

Solución:

Con el diagrama de dispersión de Excel, eligiendo Tipo y en Opciones (representar ecuación y coeficiente de determinación), se observa que la línea óptima de regresión que se ajusta a la cantidad recuperada es una función logarítmica.



El modelo logarítmico que refleja la cancelación de la inversión Y respecto a un periodo de tiempo determinado t: $y = a + b\,\text{Ln}\,t$

Ecuaciones normales del modelo
$$\begin{cases} a\,.\,N + b\,.\,\sum_{i=1}^{n} \ln t_i = \sum_{i=1}^{n} y_i \\ a\,.\,\sum_{i=1}^{n} \ln t_i + b\,.\,\sum_{i=1}^{n} (\ln t_i)^2 = \sum_{i=1}^{n} y_i \ln t_i \end{cases}$$

resultando el sistema:

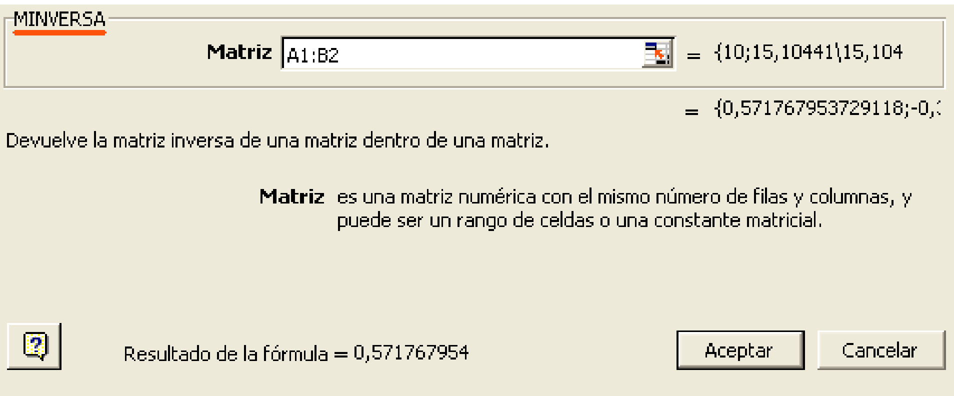
t_i	$\ln t_i$	y_i	$(\ln t_i)^2$	$y_i \ln t_i$
1	0	537984,4	0	0
2	0,69315	800078,5	0,48045	554572,1565
3	1,09861	942209,2	1,20695	1035122,6056
4	1,38629	1251083,7	1,92181	1734370,2786
5	1,60944	1461097	2,59029	2351544,9055
6	1,79176	1565418	3,21040	2804852,5248
7	1,94591	1703179,4	3,78657	3314234,0801
8	2,07944	1595651,9	4,32408	3318064,8469
9	2,19722	1814292,5	4,82780	3986408,0715
10	2,30259	1732753,9	5,30190	3989813,3000
	15,10441	13403748,5	27,65024	23088982,7695

$$\begin{cases} a\,.\,10 + b\,.\,15{,}10441 = 13403748{,}5 \\ a\,.\,15{,}10441 + b\,.\,27{,}65024 = 23088982{,}7695 \end{cases}$$

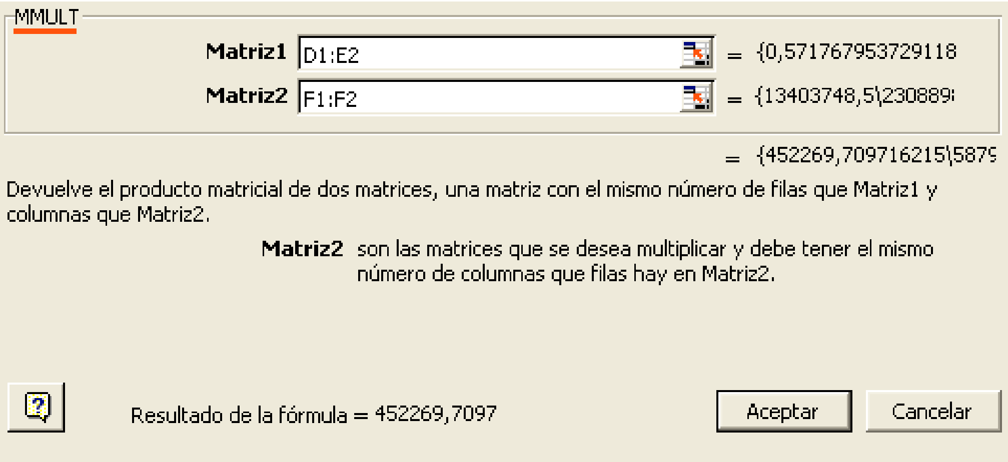
Sistema que puede resolverse por varios métodos, en forma matricial:

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} 10 & 15{,}10441 \\ 15{,}10441 & 27{,}65024 \end{pmatrix} = \begin{pmatrix} 13403748{,}5 \\ 23088982{,}7695 \end{pmatrix} \mapsto$$

$$\mapsto \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 10 & 15{,}10441 \\ 15{,}10441 & 27{,}65024 \end{pmatrix}^{-1} \begin{pmatrix} 13403748{,}5 \\ 23088982{,}7695 \end{pmatrix}$$



$$\begin{pmatrix} 10 & 15,10441 \\ 15,10441 & 27,65024 \end{pmatrix}^{-1} = \begin{pmatrix} 0,5718 & -0,3123 \\ -0,3123 & 0,2068 \end{pmatrix}$$



$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0,5718 & -0,3123 \\ -0,3123 & 0,2068 \end{pmatrix} \begin{pmatrix} 13403748,5 \\ 23088982,7695 \end{pmatrix} = \begin{pmatrix} 452270 \\ 587977 \end{pmatrix} \mapsto \begin{matrix} a = 452270 \\ b = 587977 \end{matrix}$$

Modelo logarítmico: $y = 452270 + 587977\ \text{Ln}\,t$

b) Las predicciones $\hat{y}_i$ pueden obtenerse sustituyendo los valores de la t en el modelo estimado $\hat{y}_i = 452270 + 587977\ \text{Ln}\,t_i$, con lo cual para el mes de octubre:

$$\hat{y}_8 = 452270 + 587977\ \text{Ln}\,8 = 1674933,799$$

El grado de predicción viene reflejado por el coeficiente de correlación r, siendo el coeficiente de determinación $R^2 = 1 - \dfrac{SCE}{SCT}$

donde, $SCT = \sum_{i=1}^{10} (y_i - \bar{y})^2$, $SCE = \sum_{i=1}^{10} (y_i - \hat{y}_i)^2$

Meses t_i	C. recuperada y_i	Pronóstico $\hat{y}_i = a + b\, Ln\, t_i$	SCT $(y_i - \bar{y})^2$	SCE $(y_i - \hat{y}_i)^2$
1	537984,4	452269,6380	6,43830E+11	7,34702E+09
2	800078,5	859824,5338	2,91920E+11	3,56959E+09
3	942209,2	1098228,8648	1,58536E+11	2,43421E+10
4	1251083,7	1267379,4295	7,97291E+09	2,65551E+08
5	1461097,0	1398582,8007	1,45738E+10	3,90803E+09
6	1565418,0	1505783,7605	5,06444E+10	3,55624E+09
7	1703179,4	1596420,8806	1,31627E+11	1,13974E+10
8	1595651,9	1674934,3253	6,51664E+10	6,28570E+09
9	1814292,5	1744188,0915	2,24598E+11	4,91463E+09
10	1732753,9	1806137,6964	1,53961E+11	5,38518E+09
	1340374,8500		1,74283E+12	7,09715E+10

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{7{,}09715E+10}{1{,}74283E+12} = 0{,}9593 \quad \mapsto \quad r = \sqrt{0{,}9593} = 0{,}98$$

El pronóstico se hace con una fiabilidad del 98%.

Se desea estimar los gastos en alimentación de una familia en base al número de miembros y a los ingresos mensuales. Para ello, se recoge una muestra aleatoria simple de 15 familias, cuyos resultados se facilitan en la tabla adjunta. (El gasto e ingreso se expresan en cien mil euros).

Gasto Alimentación	Ingresos	Tamaño
0,43	2,10	3
0,31	1,10	4
0,32	0,90	5
0,46	1,60	4
1,25	6,20	4
0,44	2,30	3
0,52	1,80	6
0,29	1,00	5
1,29	8,90	3
0,35	2,40	2
0,35	1,20	4
0,78	4,70	3
0,43	3,50	2
0,47	2,90	3
0,38	1,40	4

Solución:

Al realizar un ajuste lineal de la función $Y = b_0 + b_1 X_1 + b_2 X_2 + U$, con una muestra aleatoria

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + u_i \qquad \begin{cases} y_i \in N(b_0 + b_1 X_1 + b_2 X_2, \sigma^2) \text{ independientes} \\ \\ u_i \in N(0, \sigma^2) \text{ independientes} \end{cases}$$

por el método de los mínimos cuadrados ordinarios (MCO) aparece el sistema de ecuaciones normales:

$$\begin{cases} b_0 N + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i}\, x_{2i} = \sum_{i=1}^{n} x_{1i}\, y_i \\ b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{1i}\, x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i}\, y_i \end{cases}$$

y_i	x_{1i}	x_{2i}	x_{1i}^2	x_{2i}^2	$x_{1i}\, x_{2i}$	$x_{1i}\, y_i$	$x_{2i}\, y_i$
0,43	2,1	3	4,41	9	6,3	0,903	1,29
0,31	1,1	4	1,21	16	4,4	0,341	1,24
0,32	0,9	5	0,81	25	4,5	0,288	1,6
0,46	1,6	4	2,56	16	6,4	0,736	1,84
1,25	6,2	4	38,44	16	24,8	7,750	5
0,44	2,3	3	5,29	9	6,9	1,012	1,32
0,52	1,8	6	3,24	36	10,8	0,936	3,12
0,29	1	5	1	25	5	0,29	1,45
1,29	8,9	3	79,21	9	26,7	11,481	3,87
0,35	2,4	2	5,76	4	4,8	0,84	0,7
0,35	1,2	4	1,44	16	4,8	0,42	1,4
0,78	4,7	3	22,09	9	14,1	3,666	2,34
0,43	3,5	2	12,25	4	7	1,505	0,86
0,47	2,9	3	8,41	9	8,7	1,363	1,41
0,38	1,4	4	1,96	16	5,6	0,532	1,52
8,07	42	55	188,08	219	140,8	32,063	28,96

con lo cual,

$$\begin{cases} 15 b_0 + b_1 \sum_{i=1}^{15} x_{1i} + b_2 \sum_{i=1}^{15} x_{2i} = \sum_{i=1}^{15} y_i \\ b_0 \sum_{i=1}^{15} x_{1i} + b_1 \sum_{i=1}^{15} x_{1i}^2 + b_2 \sum_{i=1}^{15} x_{1i}\, x_{2i} = \sum_{i=1}^{15} x_{1i}\, y_i \\ b_0 \sum_{i=1}^{15} x_{2i} + b_1 \sum_{i=1}^{15} x_{1i}\, x_{2i} + b_2 \sum_{i=1}^{15} x_{2i}^2 = \sum_{i=1}^{15} x_{2i}\, y_i \end{cases} \mapsto \begin{cases} 15 b_0 + 42 b_1 + 55 b_2 = 8,07 \\ 42 b_0 + 188,08 b_1 + 140,08 b_2 = 32,063 \\ 55 b_0 + 140,08 b_1 + 219 b_2 = 28,96 \end{cases}$$

Sistema que se puede resolver por varios métodos, en forma matricial:

$$\overbrace{\begin{pmatrix} 15 & 42 & 55 \\ 42 & 188,08 & 140,8 \\ 55 & 140,8 & 219 \end{pmatrix}}^{X^t X} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \overbrace{\begin{pmatrix} 8,07 \\ 32,063 \\ 28,96 \end{pmatrix}}^{X^t Y} \mapsto$$

$$\mapsto \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \overbrace{\begin{pmatrix} 15 & 42 & 55 \\ 42 & 188,08 & 140,8 \\ 55 & 140,8 & 219 \end{pmatrix}^{-1}}^{(X^t X)^{-1}} \overbrace{\begin{pmatrix} 8,07 \\ 32,063 \\ 28,96 \end{pmatrix}}^{X^t Y} \mapsto$$

$$\mapsto \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1,36 & -0,093 & -0,282 \\ -0,093 & 0,017 & 0,013 \\ -0,282 & 0,013 & 0,067 \end{pmatrix} \begin{pmatrix} 8,07 \\ 32,063 \\ 28,96 \end{pmatrix}$$

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -0,16 \\ 0,149 \\ 0,077 \end{pmatrix} \mapsto \underbrace{y = -0,16 + 0,149\,x_1 + 0,077\,x_2 + u_i}_{\text{Modelo de regresión lineal}}$$

A partir de la ecuación $\hat{y}_i = \hat{b}_0 + \hat{b}_1\,x_{1i} + \hat{b}_2\,x_{2i} + u_i$ se obtienen las predicciones y residuos asociados $u_i = y_i - \hat{y}_i$ a las observaciones muestrales.

De este modo, se tiene:

y_i	$\hat{y}_i$	$u_i = y_i - \hat{y}_i$	SCT $(y_i - \bar{y})^2$	SCE $(\hat{y}_i - \bar{y})^2$	SCR $(y_i - \hat{y}_i)^2$
0,43	0,2786	0,3839	0,0117	0,0241	0,0021
0,31	0,3896	0,3119	0,0520	0,0516	0,0000
0,32	0,6094	0,3591	0,0475	0,0324	0,0015
0,46	0,4576	0,3864	0,0061	0,0234	0,0054
1,25	1,0832	1,0718	0,5069	0,2823	0,0318
0,44	0,3058	0,4137	0,0096	0,0158	0,0007
0,52	0,9788	0,5702	0,0003	0,0009	0,0025
0,29	0,623	0,3740	0,0615	0,0273	0,0071
1,29	1,2034	1,3971	0,5655	0,7327	0,0115
0,35	0,0724	0,3516	0,0353	0,0352	0,0000
0,35	0,4032	0,3268	0,0353	0,0451	0,0005
0,78	0,6322	0,7713	0,0586	0,0535	0,0001
0,43	0,222	0,5155	0,0117	0,0006	0,0073
0,47	0,3874	0,5031	0,0046	0,0013	0,0011
0,38	0,4304	0,3566	0,0250	0,0333	0,0005
8,07			1,4316	1,3595	0,0721

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\substack{\text{SCT}\\ \text{suma cuadrados total}\\ (n-1)\text{ grados libertad}}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{\substack{\text{SCE}\\ \text{suma cuadrados explicada}\\ k\text{ grados libertad}}} + \underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\substack{\text{SCR}\\ \text{suma cuadrados residual}\\ (n-k-1)\text{ grados libertad}}}$$

Variación	Suma cuadrados	Grados libertad	Media cuadrática	F Fisher-Snedecor
Explicada (regresión)	$SCE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=1,3595$	$k=2$	$\frac{SCE}{k}$	$F=\frac{SCE/k}{SCR/(n-k-1)}$
Residual	$SCR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=0,0721$	$n-k-1=12$	$\frac{SCR}{n-k-1}$	
Total	$SCT=\sum_{i=1}^{n}(y_i-\bar{y})^2=1,4316$	$n-1=14$		

- **Cálculo de los coeficientes de determinación y correlación (múltiple y simple)**

$$R^2=\frac{SCE}{SCT}=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{1,3595}{1,4316}=0,9496 \quad \text{coeficiente determinación}$$

$$r=\sqrt{0,9496}=0,9745 \quad \text{coeficiente correlación}$$

- **Coeficiente de determinación corregido por el número de grados de libertad**

$$\bar{R}^2=1-\frac{\dfrac{SCR}{n-k-1}}{\dfrac{SCT}{n-1}}=1-\frac{\dfrac{0,0721}{12}}{\dfrac{1,4316}{14}}=1-\frac{0,006}{0,1022}=0,9412 \quad \text{determinación corregido}$$

Estimación de la media condicionada

Estimar el gasto medio de una familia con unos ingresos de treinta mil euros $x_1=3$ con cuatro miembros de familia $x_2=4$, con un nivel de confianza del 90%.

Aplicando el modelo de regresión: $\hat{y}=-0,16+0,149\,x_1+0,077\,x_2$

$$\hat{y}_0=-0,16+0,149\text{ x }3+0,077\text{ x }4=0,595$$

$$IC_{E(y_0)} = \left[\hat{y}_0 \pm t_{\alpha/2,\,(n-k-1)}\, s_R \sqrt{(1 \;\; X_1 \;\; X_2)\,(X^t X)^{-1} \begin{pmatrix} 1 \\ X_1 \\ X_2 \end{pmatrix}} \right]$$

El error estándar de la estimación:

$$s_R^2 = \frac{SCR}{n-k-1} = \frac{0{,}0721}{15-2-1} = 0{,}006 \quad \mapsto \quad s_R = \sqrt{0{,}006} = 0{,}0775$$

$$t_{0{,}05,\,12} = 1{,}782$$

$$IC_{E(y_0)} = \left[0{,}595 \pm 1{,}782 \times 0{,}0775 \sqrt{(1 \;\; 3 \;\; 4) \begin{pmatrix} 15 & 42 & 55 \\ 42 & 188{,}08 & 140{,}8 \\ 55 & 140{,}8 & 219 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}} \right] =$$

$$= [0{,}557 \; ; \; 0{,}633]$$

$$(1 \;\; 3 \;\; 4) \begin{pmatrix} 15 & 42 & 55 \\ 42 & 188{,}08 & 140{,}8 \\ 55 & 140{,}8 & 219 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} = (1 \;\; 3 \;\; 4) \begin{pmatrix} 1{,}36 & -0{,}093 & -0{,}282 \\ -0{,}093 & 0{,}017 & 0{,}013 \\ -0{,}282 & 0{,}013 & 0{,}067 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} =$$

$$= (-0{,}0458 \quad 0{,}0075 \quad 0{,}025) \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} = 0{,}0765$$

El gasto medio de una familia con unos ingresos de treinta mil euros, con cuatro miembros de familia, se encuentra entre 55.700 y 63.300 euros, con una fiabilidad del 90%.

Intervalos de confianza de los parámetros del modelo de regresión

Intervalo de confianza para la varianza

$$n-k-1 = 12 \quad s_R^2 = 0{,}0721 \quad s_R = 0{,}006 \quad \chi^2_{0{,}05\,,\,12} = 21{,}026 \quad \chi^2_{0{,}95\,,\,12} = 5{,}226$$

$$IC_{1-\alpha}(\sigma^2) = \left(\frac{(n-k-1)\, s_R^2}{\chi^2_{\frac{\alpha}{2},\,(n-k-1)}} \; ; \; \frac{(n-k-1)\, s_R^2}{\chi^2_{1-\frac{\alpha}{2},\,(n-k-1)}} \right) = \left(\frac{SCR}{\chi^2_{\frac{\alpha}{2},\,(n-k-1)}} \; ; \; \frac{SCR}{\chi^2_{1-\frac{\alpha}{2},\,(n-k-1)}} \right) =$$

$$= \left[\frac{0{,}0721}{21{,}026} ; \frac{0{,}0721}{5{,}226} \right] = \left[0{,}0034 \ ; \ 0{,}0138 \right] \ \mapsto \ 0{,}0034 \leq \sigma^2 \leq 0{,}0138$$

Varianza de los estimadores del modelo $\hat{b} \in N\left(b, \sigma^2 \left[X^t X \right]^{-1} \right)$

$$Var(\hat{b}_i) = \sigma^2 \left[X^t X \right]^{-1} \simeq s_R^2 \left[X^t X \right]^{-1} = 0{,}006 \ \overbrace{\begin{pmatrix} \cancel{1{,}36} & & \\ & \cancel{0{,}017} & \\ & & \cancel{0{,}067} \end{pmatrix}}^{q_{i+1,i+1} \equiv \text{elemento de} \left[X^t X \right]^{-1}} =$$

$$= \overbrace{\begin{pmatrix} \cancel{0{,}00816} & & \\ & \cancel{0{,}000096} & \\ & & \cancel{0{,}0004} \end{pmatrix}}^{s_R^2 \, q_{i+1,i+1}}$$

de donde,

$$Var(b_0) = 0{,}00816 \ \mapsto \ \sigma_{b_0} = s_R \sqrt{q_{1,1}} = \sqrt{0{,}00816} = 0{,}0904$$

$$Var(b_1) = 0{,}000096 \ \mapsto \ \sigma_{b_1} = s_R \sqrt{q_{2,2}} = \sqrt{0{,}000096} = 0{,}0098$$

$$Var(b_2) = 0{,}0004 \ \mapsto \ \sigma_{b_2} = s_R \sqrt{q_{3,3}} = \sqrt{0{,}0004} = 0{,}02$$

Intervalo de confianza para los coeficientes

$$IC_{1-\alpha}(b_i) = \left[\hat{b}_i \ \pm \ t_{\alpha/2,(n-k-1)} \ s_R \sqrt{q_{i+1,i+1}} \ \right]$$

Datos muestrales: $\hat{b}_0 = -0{,}160 \quad \hat{b}_1 = 0{,}149 \quad \hat{b}_2 = 0{,}077 \quad s_R = 0{,}0775$

$t_{0{,}05,12} = 1{,}782$

$$IC_{0{,}90}(b_0) = \left[-0{,}160 \ \pm \ 1{,}782 \times 0{,}0904 \right] = \left[-0{,}321 \ ; \ 0{,}001 \right] \ \mapsto \ -0{,}321 \leq b_0 \leq 0{,}0006$$

$$IC_{0{,}90}(b_1) = \left[0{,}149 \ \pm \ 1{,}782 \times 0{,}0098 \right] = \left[0{,}167 \ ; \ 0{,}1665 \right] \ \mapsto \ 0{,}1315 \leq b_1 \leq 0{,}1665$$

$$IC_{0{,}90}(b_2) = \left[0{,}077 \ \pm \ 1{,}782 \times 0{,}02 \right] = \left[0{,}0414 \ ; \ 0{,}1126 \right] \ \mapsto \ 0{,}0414 \leq b_2 \leq 0{,}1126$$

Contrastes de hipótesis individuales

Se puede plantear, como ejemplo, si la variable X_2 (número de miembros familiar) influye sobre la variable de respuesta Y (Gastos de alimentación). En otras palabras, si el valor del parámetro en la población es cero o no.

Para ello, se establece la hipótesis nula $H_0: b_2 = 0$ frente a la hipótesis alternativa $H_1: b_2 \neq 0$

El estadístico observado $t = \dfrac{\hat{b}_2 - b_2}{s_R \sqrt{q_{33}}}$, bajo la hipótesis nula resulta $t = \dfrac{\hat{b}_2}{s_R \sqrt{q_{33}}}$

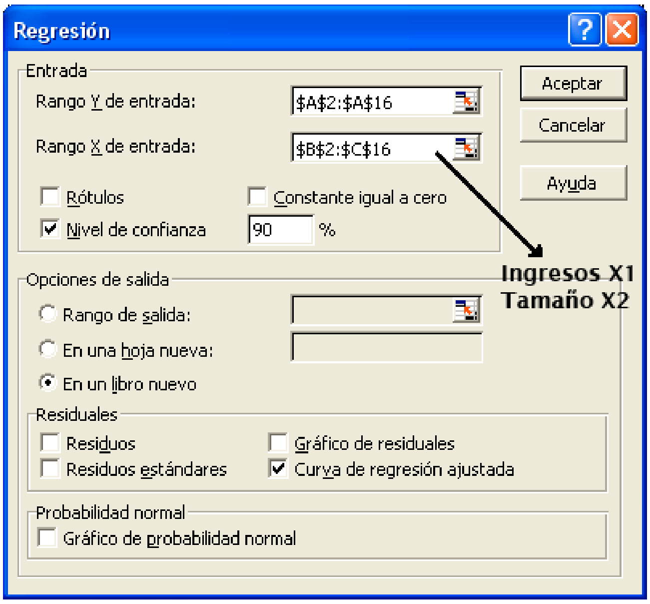
$\hat{b}_2 = 0,077 \qquad s_R q_{33} = 0,0775 \text{ x } \sqrt{0,0004} = 0,00155 \qquad t_{0,05\,,\,12} = 1,782$

Por tanto, el estadístico experimental $t = \dfrac{\hat{b}_2}{s_R \sqrt{q_{33}}} = \dfrac{0,077}{0,00155} = 49,67$

Siendo $\left|t\right| = 49,67 > t_{0,05\,,\,12} = 1,782$ se rechaza la hipótesis nula, afirmando, con un 90% de fiabilidad, que el número de miembros de la familia influye en los gastos de alimentación.

*** Obsérvese que en el Intervalo de confianza $IC_{1-\alpha}(b_2) = \left[\, 0,0414 \;;\; 0,1126 \,\right]$ el 0 no se encuentra en el intervalo, con lo que se rechaza la hipótesis nula $H_0: b_2 = 0$, concluyendo que el número de miembros de la familia (tamaño) si influye en los gastos de alimentación (Y).**

EXCEL: **Herramientas/Análisis de datos/Regresión**



Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,9745
Coeficiente de determinación R^2	0,9496
R^2 ajustado	0,9412
Error típico	0,0775
Observaciones	15

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F
Regresión	2	1,3595	0,6798	113,1414
Residuos	12	0,0721	0,0060	
Total	14	1,4316		

$$\hat{y} = -0,160 + 0,149\, x_1 + 0,077\, x_2$$

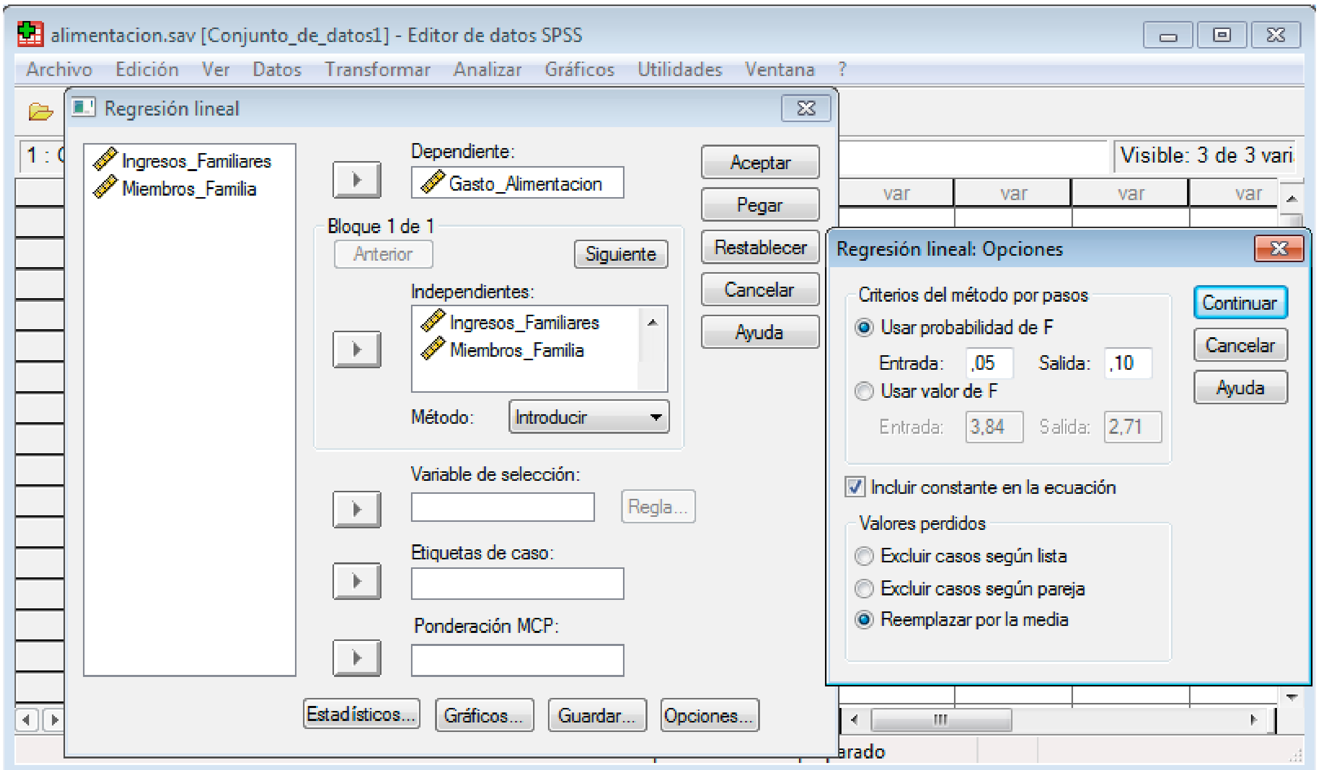
	Coeficientes	Error típico	Estadístico t	Inferior 90,0%	Superior 90,0%
Intercepción	-0,160	0,0904	-1,7752	-0,3216	0,0006
Variable X 1	0,149	0,0100	14,9155	0,1310	0,1665
Variable X 2	0,077	0,0201	3,8253	0,0411	0,1128

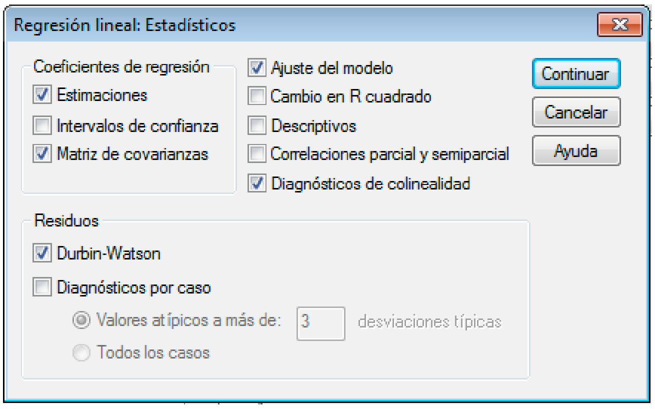
$$IC_{0,90}(b_0) = \left[-0,160 \pm 1,782 \times 0,0904\right] \mapsto -0,321 \leq b_0 \leq 0,0006$$

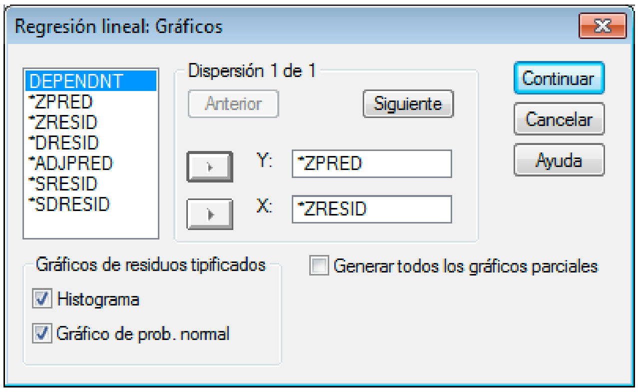
$$IC_{0,90}(b_1) = \left[0,149 \pm 1,782 \times 0,0098\right] \mapsto 0,1315 \leq b_1 \leq 0,1665$$

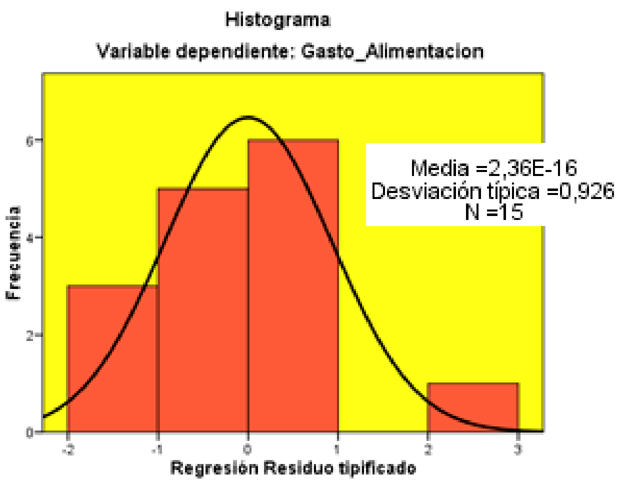
$$IC_{0,90}(b_2) = \left[0,077 \pm 1,782 \times 0,02\right] = \left[0,0414\ ;\ 0,1126\right] \mapsto 0,0414 \leq b_2 \leq 0,1126$$

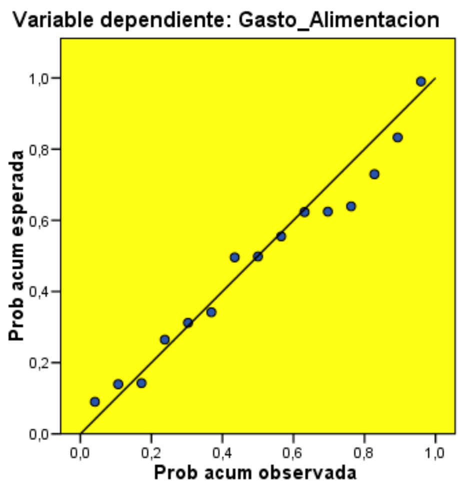
SPSS: Analizar/ Regresión / Lineal











En el histograma de los residuos se observa que el modelo de regresión se ajusta bien a una distribución normal.

El gráfico de normalidad refleja que el modelo de regresión se ajusta muy bien a la diagonal del primer cuadrante.

● **Se representan los residuos tipificados o estandarizados (ZRESID) frente a los valores pronosticados o predicciones tipificadas (ZPRED). El resultado tiene que ser una nube de puntos totalmente aleatoria, es decir, no se observan tendencias ni patrones en la representación gráfica. Si se cumple esta condición se acepta la hipótesis de linealidad y de varianza constante (homocedasticidad) de los errores. Dos supuestos del análisis de regresión.**

● **Al representar los valores observados frente a los predichos (DEPEND vs. ZPRED) se deben alinear en la diagonal del cuadrante, si hay mucha dispersión o variabilidad entonces no se cumple la hipótesis de homocedasticidad. Si la dispersión no es muy grande, existe igualdad de varianzas.**

Resumen del modelo[b]

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,9745[a]	,9496	,9412	,07751	1,177

a. Variables predictoras: (Constante), Miembros_Familia, Ingresos_Familiares

b. Variable dependiente: Gasto_Alimentacion

El estadístico de Durbin-Watson con un valor de 1,177 no deja claro la presencia o no de autocorrelación: $DW = \dfrac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2} \approx 2(1-r) \;\mapsto\; \begin{cases} DW \approx 2 & \text{si } \rho = 0 \\ DW \approx 0 & \text{si } \rho = 1 \\ DW \approx 4 & \text{si } \rho = -1 \end{cases}$

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1,3595	2	,680	113,141	,000[a]
	Residual	,0721	12	,006		
	Total	1,4316	14			

a. Variables predictoras: (Constante), Miembros_Familia, Ingresos_Familiares

b. Variable dependiente: Gasto_Alimentacion

A un nivel de confianza $(1-\alpha)$ **se rechaza la hipótesis nula** $H_0:\ b_1 = b_2 = 0$ **(el modelo no es explicativo), cuando** $F_{k,(n-k-1)} \geq F_{\alpha;\,k,(n-k-1)}$

$$F_{2,12} = \frac{SCE/2}{SCR/(15-2-1)} = 113,141 > 9,4081 = F_{0,1;2,12}$$

Así, pues, se rechaza la hipótesis nula, el contraste de la F de Fisher-Snedecor indica claramente la influencia del modelo en la variable respuesta.

De otra parte, siendo el $\text{p-valor (Sig)} = 0 < 0,05$ **se rechaza la hipótesis nula de que la variabilidad observada en la variable respuesta sea explicada por el azar. Se admite que hay algún tipo de asociación entre la variable dependiente y las independientes.**

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	-,160	,090		-1,775	,101		
	Ingresos_Familiares	,149	,010	1,044	14,915	,000	,857	1,166
	Miembros_Familia	,077	,020	,268	3,825	,002	,857	1,166

a. Variable dependiente: Gasto_Alimentacion

Modelo estimado: $y = -0,160 + 0,149\, x_1 + 0,077\, x_2$

Como la constante $b_0 = -0,160$ no resulta estadísticamente significativa $(p - valor = Sig = 0,101 > 0,05)$ se puede eliminar del modelo de regresión.

Los valores de Beta (pendiente de la recta de regresión) positivos indican una relación directa entre X_i e Y. Los valores de Beta negativos indican una relación inversa. La hipótesis nula en los contrastes de hipótesis de las pendientes señala que $Beta = 0$.

Como las dos variables independientes tienen una pendiente estadísticamente significativa $(p - valor = Sig \leq 0,05)$ no se elimina ninguna variable del modelo. Si alguna de ellas no hubiese sido estadísticamente significativa se podría eliminar del modelo de regresión.

Cuando existe colinealidad o multicolinealidad (correlación entre las variables independientes del modelo de regresión) las variables están correlacionadas entre sí y se reduce el poder predictivo de las variables independientes tomadas individualmente. En otras palabras, cuanta mayor sea la colinealidad menor será la varianza explicada por cada variable independiente.

La colinealidad se detecta examinando la matriz de correlaciones entre las variables independientes, cuando los valores son altos es probable que exista colinealidad.

Los estadísticos de multicolinealidad son el Valor de Tolerancia (TOL) y el Factor de Inflación de la Varianza (FIV).
Se tiene multicolinealidad cuando el Valor de tolerancia (TOL) es próximo a cero o El Factor de Inflación de la Varianza (FIV) es mayor que 4.

La situación ideal es tener variables independientes (X_1, X_2) altamente correlacionadas con la variable dependiente Y pero con poca correlación entre sí.

Correlaciones

		Gasto_ Alimentacion	Ingresos_ Familiares	Miembros_ Familia
Correlación de Pearson	Gasto_Alimentacion	1,000	,942	-,126
	Ingresos_Familiares	,942	1,000	-,378
	Miembros_Familia	-,126	-,378	1,000
Sig. (unilateral)	Gasto_Alimentacion	.	,000	,327
	Ingresos_Familiares	,000	.	,083
	Miembros_Familia	,327	,083	.
N	Gasto_Alimentacion	15	15	15
	Ingresos_Familiares	15	15	15
	Miembros_Familia	15	15	15

$$R_x = \begin{pmatrix} r_{YY} = 1 & r_{YX_1} = 0,942 & r_{YX_2} = -0,126 \\ r_{X_1Y} = 0,942 & r_{X_1X_1} = 1 & r_{X_1X_2} = -0,378 \\ r_{X_2Y} = -0,126 & r_{X_2X_1} = -0,378 & r_{X_2X_2} = 1 \end{pmatrix}$$

- **Coeficientes Beta de la regresión parcial:**

$$b_{yx_1.x_2} = \frac{r_{yx_1} - r_{yx_2} \times r_{x_1x_2}}{1 - r^2_{x_1x_2}} \qquad b_{yx_2.x_1} = \frac{r_{yx_2} - r_{yx_1} \times r_{x_2x_1}}{1 - r^2_{x_2x_1}}$$

$$b_{yx_1.x_2} = \frac{r_{yx_1} - r_{yx_2} \times r_{x_1x_2}}{1 - r^2_{x_1x_2}} = \frac{0,942 - (-0,126) \times (-0,378)}{1 - (-0,378)^2} = 1,044$$

$$b_{yx_2.x_1} = \frac{r_{yx_2} - r_{yx_1} \times r_{x_2x_1}}{1 - r^2_{x_2x_1}} = \frac{-0,126 - 0,942 \times (-0,378)}{1 - (-0,378)^2} = 0,268$$

- **Coeficiente de determinación múltiple:** $R^2_{YX_1X_2} = \dfrac{r^2_{YX_1} + r^2_{YX_2} - 2 \times r_{YX_1} \times r_{YX_2} \times r_{X_1X_2}}{1 - r^2_{X_1X_2}}$

$$R^2_{YX_1X_2} = \frac{0,942^2 + (-0,126)^2 - 2 \times 0,942 \times (-0,126) \times (-0,378)}{1 - (-0,378)^2} = 0,9492$$

- **Coeficientes de correlación parciales:**

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} \times r_{X_1X_2}}{\sqrt{(1 - r^2_{YX_2}) \times (1 - r^2_{X_1X_2})}}$$

Correlación parcial de las variables (Y, X_1) dejando constante la variable X_2

$$r_{YX_2.X_1} = \frac{r_{YX_2} - r_{YX_1} \times r_{X_2X_1}}{\sqrt{(1 - r^2_{YX_1}) \times (1 - r^2_{X_2X_1})}}$$

Correlación parcial de las variables (Y, X_2) dejando constante la variable X_1

$$r_{YX_1.X_2} = \frac{0,942 - (-0,126)\ x\ (-0,378)}{\sqrt{\left[(1-(-0,126)^2\right] x \left[(1-(-0,378)^2\right]}} = 0,9747$$

$$r_{YX_2.X_1} = \frac{(-0,126) - (0,942)\ x\ (-0,378)}{\sqrt{(1-0,942^2)\ x \left[1-(-0,378)^2\right]}} = 0,741$$

- **El coeficiente de determinación semiparcial** $R^2_{Y(X_1X_2)}$ **de las variables** (Y, X_1) **es el incremento del coeficiente de determinación** $R^2_{YX_1X_2}$ **que se produce al incluir la variable** X_2 **en la ecuación de regresión:** $R^2_{Y(X_1X_2)} = R^2_{YX_1X_2} - r^2_{YX_2}$

$$R^2_{Y(X_1X_2)} = R^2_{YX_1X_2} - r^2_{YX_2} = 0,9492 - (-0,126)^2 = 0,933 \ \mapsto\ r_{Y(X_1X_2)} = \sqrt{0,933} = 0,966$$

$$R^2_{Y(X_2X_1)} = R^2_{YX_2X_1} - r^2_{YX_1} = 0,9492 - 0,942^2 = 0,061 \ \mapsto\ r_{Y(X_2X_1)} = \sqrt{0,061} = 0,248$$

Coeficientes[a]

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Correlaciones		
	B	Error típ.	Beta			Orden cero	Parcial	Semiparcial
(Constante)	-,160	,090		-1,775	,101			
Ingresos_Familiares	,149	,010	1,044	14,915	,000	,942	,974	,966
Miembros_Familia	,077	,020	,268	3,825	,002	-,126	,741	,248

a. Variable dependiente: Gasto_Alimentacion

$\boxed{y = -0,160 + 0,149\ x_1 + 0,077\ x_2}$ $\boxed{y = 1,044\ x_1 + 0,268\ x_2}$

$$y = -0,160 + 0,149\ x_1 + 0,077\ x_2 \qquad y = 1,044\ x_1 + 0,268\ x_2$$

En la muestra aleatoria $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + u_i$

- $Y_i \in N(b_0 + b_1 X_1 + b_2 X_2, \sigma^2)$ independientes
- $u_i \in N(0, \sigma^2)$ independientes

En la tabla se recogen los Índices de Precios Industriales para España con base 1980 y 2010 para los meses de diciembre de cada año.

Diciembre	Base 1980	Base 2010
2007	429,70	
2008	444,49	
2009	460,67	
2010	471,12	102
2011		102,6
2012		104,2
2013		107,7
2014		113,3
2015		118,3

a) Completar la tabla de IPI.
b) Tabla donde se reflejen los Precios de Índices Industriales con base 2015

Solución:

a) Para **enlazar dos series** basta con determinar la relación existente entre los valores de dos bases del mismo período.

La relación o coeficiente de enlace con base 1980: $\dfrac{471{,}12}{102} = 4{,}6188$

Tomando 2010 como base, el coeficiente de enlace: $\dfrac{102}{471{,}12} = 0{,}2165$

Diciembre	Base 1980	Base 2010
2007	429,70	429,70 x 0,2165 = 93,03
2008	444,49	444,49 x 0,2165 = 96,23
2009	460,67	460,67 x 0,2165 = 99,73
2010	471,12	102
2011	102,6 x 4,6188 = 473,89	102,6
2012	104,2 x 4,6188 = 481,28	104,2
2013	107,7 x 4,6188 = 497,45	107,7
2014	113,3 x 4,6188 = 523,31	113,3
2015	118,3 x 4,6188 = 546,41	118,3

b) Una operación similar al **enlace de series** es el **cambio de base** para una serie concreta. Para que la serie con base 2010 tenga el valor 100 en 2015, se requiere encontrar el coeficiente que haga esta transformación:

coeficiente transformación: $\dfrac{100}{188{,}3} = 0{,}8453$

Diciembre	Base 1990	Base 2010	Base 2015 (Diciembre 2015 = 100)
2007	429,70	93,03	93,03 x 0,8453 = 78,61
2008	444,49	96,23	96,23 x 0,8453 = 81,34
2009	460,67	99,73	99,73 x 0,8453 = 84,30
2010	471,12	102	102 x 0,8453 = 86,22
2011	473,89	102,6	102,6 x 0,8453 = 86,73
2012	481,28	104,2	104,2 x 0,8453 = 88,08
2013	497,45	107,7	107,7 x 0,8453 = 91,04
2014	523,31	113,3	113,3 x 0,8453 = 95,77
2015	546,41	118,3	100

En el cuadro adjunto se reflejan las cantidades en euros que cobraba mensualmente un directivo de empresa, así como el índice de precios de consumo (IPC) elaborado por el INE con base 2005.

Años	Sueldos mensuales *euros corrientes*	IPC (Base 2005)
2012	2900	172,5
2013	3400	199,3
2014	3850	228,3
2015	4280	261,3
2016	4550	291,3

Analizar la variación del poder adquisitivo mensual del directivo.

Solución:

Como el IPC se encuentra expresado en base 2005 y la serie de sueldos se inicia en el año 2012, hay que efectuar un cambio de base del IPC con base 2012:

Coeficiente transformación: $\frac{100}{172,5} = 0,58$

Años	IPC 2005 = 100	IPC 2012 = 100	IPC unitario 2012 = 1	Sueldos mensuales *euros constantes 2012*
2012	172,5	100	1	2900
2013	199,3	$199,3 \times 0,58 = 115,6$	1,156	$3400 / 1,156 = 2941,18$
2014	228,3	$228,3 \times 0,58 = 132,4$	1,324	$3850 / 1,324 = 2907,85$
2015	261,3	$261,3 \times 0,58 = 151,5$	1,515	$4280 / 1,515 = 2825$
2016	291,3	$291,3 \times 0,58 = 169$	1,69	$4550 / 1,69 = 2692,31$

Para conocer los sueldos reales o poder adquisitivo hay que deflactar la serie para expresarlos en euros de 2012.

$$\text{Euros de 2012} = \frac{\text{Euros de cada año}}{\text{IPC unitario}}$$

Tabla de sueldos mensuales en euros corrientes y en euros constantes, expresando de bajo el porcentaje de aumento salarial mensual de cada año con respecto al anterior:

Años	Sueldos mensuales *euros corrientes 2005*	Sueldos mensuales *euros constantes 2012*
2012	2900	2900
2013	3400 $\frac{3400-2900}{2900}=17,24\%$	2941,18 $\frac{2941,18-2900}{2900}=1,42\%$
2014	3850 $\frac{3850-3400}{3400}=13,24\%$	2907,85 $\frac{2907,85-2941,18}{2941,18}=-1,13\%$
2015	4280 $\frac{4280-3850}{3850}=11,17\%$	2825 $\frac{2825-2907,85}{2907,85}=-2,85\%$
2016	4550 $\frac{4550-4280}{4280}=6,31\%$	2692,31 $\frac{2692,31-2825}{2825}=-4,7\%$

Aumento salarial entre 2012 y 2016 en euros corrientes: $\frac{4550-2900}{2900}=56,9\%$

Aumento salarial entre 2012 y 2016 en euros constantes:

$$\frac{2692,31-2900}{2900}=-7,16\%$$

El sueldo real o poder adquisitivo del directivo ha disminuido un 7,16% entre los años 2012 y 2016.

$$\boxed{\text{Sueldo real o poder adquisitivo} = \frac{\text{Sueldo corriente}}{\text{IPC unitario}}}$$

En la tabla adjunta aparecen distintos artículos, precios (en céntimos de euros) y cantidad vendida entre 2012 y 2014

Artículos	2012		2013		2014	
	precios	cantidad vendida	precios	cantidad vendida	precios	cantidad vendida
Pan	38	150	44	200	48	240
Huevos	130	400	150	580	215	560
Leche	88	700	100	780	110	925
Pollo	160	400	190	400	205	375

a) Calcular los índices compuestos sin ponderar de precios de Sauerbeck, media geométrica, media armónica, y de Bradstreet-Dûtot.
Comentar los inconvenientes de estos índices.

b) Determinar los índices de precios de Laspeyres, Paasche, Edgeworth y Fisher para 2014, con año base 2012.

Solución:

a) Utilizando la media aritmética se obtiene el índice de la media aritmética simple o índice de Sauerbeck: $I_0^t = \frac{1}{m}.\sum_{i=1}^{m}\frac{p_{it}}{p_{i0}}.100$

Artículos	Precios		
	2012	2013	2014
Pan	38	44	48
Huevos	130	150	215
Leche	88	100	110
Pollo	160	190	205

$$I_{2012}^{2013} = \frac{1}{4}.\sum_{i=1}^{4}\frac{p_{2013}}{p_{2012}}.100 = \frac{1}{4}.\left[\frac{44}{38}+\frac{150}{130}+\frac{100}{88}+\frac{190}{160}\right].100 = 115,89$$

$$I_{2012}^{2014} = \frac{1}{4}.\sum_{i=1}^{4}\frac{p_{2014}}{p_{2012}}.100 = \frac{1}{4}.\left[\frac{48}{38}+\frac{215}{130}+\frac{110}{88}+\frac{205}{160}\right].100 = 136,21$$

- **Índice media geométrica:** $I_0^t = \sqrt[m]{\prod_{i=1}^{m}\frac{p_{it}}{p_{i0}}}\;.\,100$

$$I_{2012}^{2013} = \sqrt[4]{\frac{44}{38}.\frac{150}{130}.\frac{100}{88}.\frac{190}{160}}\;.\,100 = 115,88$$

$$I_{2012}^{2014} = \sqrt[4]{\frac{48}{38}.\frac{215}{130}.\frac{110}{88}.\frac{205}{160}}\ .100 = 135,25$$

- **Índice media armónica:** $I_0^t = \dfrac{m}{\sum\limits_{i=1}^{m} \dfrac{p_{i0}}{p_{it}}}.100$

$$I_{2012}^{2013} = \frac{4}{\frac{38}{44} + \frac{130}{150} + \frac{88}{100} + \frac{160}{190}}.100 \ = 115,86$$

$$I_{2012}^{2014} = \frac{4}{\frac{38}{48} + \frac{130}{215} + \frac{88}{110} + \frac{160}{205}}.100 \ = 134,37$$

- **Índice media agregativa simple o de Bradstreet-Dûtot:** $I_o^t = \dfrac{\sum\limits_{i=1}^{m} p_{it}}{\sum\limits_{i=1}^{m} p_{i0}}.100$

$$I_{2012}^{2013} = \frac{\sum\limits_{i=1}^{4} p_{2013}}{\sum\limits_{i=1}^{4} p_{2012}}.100 = \frac{44+150+100+190}{38+130+88+160}.100 = 116,35$$

$$I_{2012}^{2014} = \frac{\sum\limits_{i=1}^{4} p_{2014}}{\sum\limits_{i=1}^{4} p_{2012}}.10 = \frac{48+215+110+205}{38+130+88+160}.100 = 138,94$$

Los diferentes precios o cualquier otro concepto que se refiera a los m artículos tienen la misma importancia o peso específico. Obviamente, esto no es correcto, cada magnitud simple tendrá distinta importancia, que variará en función de la finalidad del índice.

Otro problema que presentan los índices compuestos sin ponderar es la heterogeneidad en las unidades de medida de los diferentes artículos que intervienen en la agregación para obtener la magnitud compleja. En esta línea, estos índices cambian si se cambia la unidad de medida de un artículo.

Estos dos tipos de inconvenientes quedan solventados al introducir los índices compuestos ponderados. El primero, asignando coeficientes de ponderación a cada artículo o magnitud simple.

El segundo, transformando las diferentes magnitudes a través de una unidad de cuenta común a las mismas unidades. Para ello se introduce el concepto de valor en Economía, que es el producto del precio por la cantidad V = P.Q

b) Índices compuestos de Laspeyres, Paasche, Edgeworth y Fisher

Artículos	2012 Precios p_{2012}	2012 Cantidad q_{2012}	2013 Precios p_{2013}	2013 Cantidad q_{2013}	2014 Precios p_{2014}	2014 Cantidad q_{2014}
Pan	38	150	44	200	48	240
Huevos	130	400	150	580	215	560
Leche	88	700	100	780	110	925
Pollo	160	400	190	400	205	375

- **Índices Laspeyres:** $P_{L_0^t} = \dfrac{\sum_{i=1}^{m} p_{it}\, q_{i0}}{\sum_{i=1}^{m} p_{io}\, q_{i0}}.100 \qquad Q_{L_0^t} = \dfrac{\sum_{i=1}^{m} p_{i0}\, q_{it}}{\sum_{i=1}^{m} p_{io}\, q_{i0}}.100$

- **Índices Paasche:** $P_{P_0^t} = \dfrac{\sum_{i=1}^{m} p_{it}\, q_{it}}{\sum_{i=1}^{m} p_{io}\, q_{it}}.100 \qquad Q_{P_0^t} = \dfrac{\sum_{i=1}^{m} p_{it}\, q_{it}}{\sum_{i=1}^{m} p_{it}\, q_{i0}}.100$

Artículos	Laspeyres $p_{2014}\, q_{2012}$	Laspeyres $p_{2012}\, q_{2012}$	Paasche $p_{2014}\, q_{2014}$	Paasche $p_{2012}\, q_{2014}$
Pan	7200	5700	11520	9120
Huevos	86000	52000	120400	72800
Leche	77000	61600	101750	81400
Pollo	82000	64000	76875	60000
	252200	183300	310545	223320

Índice Precios de Laspeyres: $P_{L_{2012}^{2014}} = \dfrac{\sum_{i=1}^{4} p_{2014}\, q_{2012}}{\sum_{i=1}^{4} p_{2012}\, q_{2012}}.100 = \dfrac{252200}{183300}.100 = 137,59$

Índice Precios de Paasche: $$P_{p_{2012}}^{2014}=\frac{\sum_{i=1}^{4}p_{2014}\,q_{2014}}{\sum_{i=1}^{4}p_{2012}\,q_{2014}}.100=\frac{310545}{223320}.100=139,06$$

- **Índices Edgeworth:** $$P_{E_0^t}=\frac{\sum_{i=1}^{m}p_{it}.(q_{i0}+q_{it})}{\sum_{i=1}^{m}p_{i0}.(q_{i0}+q_{it})}.100 \qquad Q_{E_0^t}=\frac{\sum_{i=1}^{m}q_{it}.(p_{i0}+p_{it})}{\sum_{i=1}^{m}q_{i0}.(p_{i0}+p_{it})}.100$$

				Edgeworth	
Artículos	p_{2012}	p_{2014}	$(q_{2012}+q_{2014})$	$p_{2012}(q_{2012}+q_{2014})$	$p_{2014}(q_{2012}+q_{2014})$
Pan	38	48	390	14820	18720
Huevos	130	215	960	124800	206400
Leche	88	110	1625	143000	178750
Pollo	160	205	775	124000	158875
				406620	562745

Índice Precios Edgeworth:

$$E_{p2008}^{2010}=\frac{\sum_{i=1}^{4}p_{2014}\,(q_{2012}+q_{2014})}{\sum_{i=1}^{4}p_{2012}\,(q_{2012}+q_{2014})}.100=\frac{562745}{406620}.100=138,40$$

- **Fisher propuso como índices de precios y de cantidades las medias geométricas de los respectivos índices de precios y cantidades de Laspeyres y Paasche:**

$$P_{F_0^t}=\sqrt{P_{L_0^t}\,P_{P_0^t}} \qquad Q_{F_0^t}=\sqrt{Q_{L_0^t}\,Q_{P_0^t}}$$

Índice Precios Fisher: $$P_{F_{2012}^{2014}}=\sqrt{P_{L_{2012}^{2014}}\,P_{P_{2012}^{2014}}}=\sqrt{137,59\ \text{x}\ 139,06}=138,32$$

Una empresa estudia la evolución de los precios en euros de tres componentes (A, B, C) para una pieza en los últimos 5 años.

Año	A	B	C
1	3	4	1
2	4	6	1,5
3	5	6,5	2
4	4,5	7	2,5
5	7	4	3

a) Calcular un índice simple para estudiar la evolución de los precios del componente A tomando como periodo de referencia el año 1.

b) Calcular un índice conjunto de la evolución de los precios utilizando una media aritmética de índices simples y tomando como referencia el año 1.

c) Analizar cómo varían los resultados si escoge otros promedios como la media geométrica.

d) Suponiendo que en cada pieza van 5 unidades del componente A, 10 del B y 15 del C. Calcule índices de precios conjuntos para los tres componentes tomando como referencia el periodo 1 y usando una media aritmética ponderada de los índices simples. Analice cómo varían los resultados, y cuál es el incremento medio anual de precios a partir del índice compuesto media aritmética ponderada.

Solución:

a) **Índice simple de la evolución de los precios tomando como periodo de referencia el año 1:**

				Índice Simple Precios					
Año	A	B	C	A		B		C	
1	3	4	1	(3 / 3) 100	100	(4 / 4) 100	100	(1 / 1)100	100
2	4	6	1,5	(4 / 3) 100	133,33	(6 / 4)100	150	(1,5 / 1)100	150
3	5	6,5	2	(5 / 3) 100	166,67	(6,5 / 4)100	162,50	(2 / 1)100	200
4	4,5	7	2,5	(4,5 / 3) 100	150	(7 / 4)100	175	2,5. 100	250
5	7	4	3	(7 / 3) 100	233,33	(4 / 4)100	100	3.100	300

b) Índice conjunto de la evolución de los precios utilizando una media aritmética de índices simples y tomando como referencia el año 1

Año	A	B	C		Media aritmética
1	100	100	100	300/3 = 100	100
2	133,33	150	150	433,33/3= 144,44	144,44
3	166,67	162,50	200	529,17/3=176,39	176,39
4	150	175	250	575/3=191,67	191,67
5	233,33	100	300	633,33/3=211,11	211,11

c) Índice conjunto de la evolución de los precios utilizando la media geométrica:

Año	A	B	C	$\prod_{i=1}^{3} I_i$	$\sqrt[3]{\prod_{i=1}^{3} I_i}$
1	100	100	100	1000000	100
2	133,33	150	150	3000000	144,22496
3	166,67	162,50	200	5416666,67	175,62137
4	150	175	250	6562500	187,22181
5	233,33	100	300	7000000	191,29312

d) Índice conjunto de la evolución de los precios utilizando la media ponderada:

Año	A (5 unidades)	B (10 unidades)	C (15 unidades)	Media ponderada
1	100	100	100	100
2	133,33	150	150	147,22
3	166,67	162,50	200	181,94
4	150	175	250	208,33
5	233,33	100	300	222,22

1 $5.100 + 10.100 + 15.100 / (5 + 10 + 15) = 100$

2 $5.133,33 + 10.150 + 15.150 / (5 + 10 + 15) = 147,22$

3 $5.166,67 + 10.162,50 + 15.200 / (5 + 10 + 15) = 181,94$

4 $5.150 + 10.175 + 15.250 / (5 + 10 + 15) = 208,33$

5 $5.233,33 + 10.100 + 15.300 / (5 + 10 + 15) = 222,22$

El incremento (tasa) medio anual de precios a partir del índice compuesto:

Año	A 5 unidades	B 10 unidades	C 15 unidades	Media ponderada	Incremento	% Incremento (Tasa)
1	100	100	100	100		
2	133,33	150	150	147,22	0,47222	47,22
3	166,67	162,50	200	181,94	0,23584	23,58
4	150	175	250	208,33	0,14503	14,50
5	233,33	100	300	222,22	0,06667	6,67

1

2 $(147,22 / 100) - 1 = 0,47222$

3 $(181,94 / 147,22) - 1 = 0,23584$

4 $(208,33 / 181,94) - 1 = 0,14503$

5 $(222,22 / 208,33) - 1 = 0,06667$

Se conoce la información sobre la evolución de precios de los bienes y servicios consumidos por un estudiante. Completar el cuadro con las cantidades correspondientes.

Año	Índice General	Índice cafetería	Índice transporte	Índice ocio	Índice otros
2010		149%	157%	133%	142%
2011		160%	165%	143%	
Ponderación	100%	15%	35%		20%
% Variación					4,225%

Solución:

Año	Índice General	Índice cafetería	Índice transporte	Índice ocio	Índice otros
2010		149%	157%	133%	142%
2011		160%	165%	143%	148%
Ponderación	100%	15%	35%	30%	20%
% Variación		7,383%	5,096%	7,519%	4,225%

$$\% \ TV^{\text{Cafetería 2011}}_{\text{Cafetería 2010}} = \left[\frac{I_{\text{Cafetería 2011}}}{I_{\text{Cafetería 2010}}} - 1 \right] . 100 = \left[\frac{160}{149} - 1 \right] . 100 = 7,383\%$$

$$\%\ TV_{\text{Transporte 2010}}^{\text{Transporte 2011}} = \left[\frac{I_{\text{Transporte 2011}}}{I_{\text{Transporte 2010}}} - 1\right].100 = \left[\frac{165}{157} - 1\right].100 = 5{,}096\%$$

$$\%\ TV_{\text{Ocio 2010}}^{\text{Ocio 2011}} = \left[\frac{I_{\text{Ocio 2011}}}{I_{\text{Ocio 2010}}} - 1\right].100 = \left[\frac{143}{133} - 1\right].100 = 7{,}519\%$$

$$I_{\text{Otros 2011}} = 142 \text{ x } (1 + 0{,}04225) = 148\%$$

Año	Índice General	Índice cafetería	Índice transporte	Índice ocio	Índice otros
2010	145,6%	149%	157%	133 %	142%
2011	154,25%	160%	165%	143 %	148%
Ponderación	100%	15%	35%	30%	20%
% Variación	5,914%	7,383%	5,096%	7,519%	4,225%

El Índice General es como un IPC, un Índice de Laspeyres, denotando por I_i los índices de cada grupo y w_i las ponderaciones de cada bien o servicio:

$$L_p^{2010} = \frac{\sum_{i=1}^{4} I_i . w_i}{\sum_{i=1}^{4} w_i} = \frac{149.15 + 157.35 + 133.30 + 142.20}{15 + 35 + 30 + 20} = 145{,}6\%$$

$$L_p^{2011} = \frac{\sum_{i=1}^{4} I_i . w_i}{\sum_{i=1}^{4} w_i} = \frac{160.15 + 165.35 + 143.30 + 148.20}{15 + 35 + 30 + 20} = 154{,}25\%$$

$$\%\ TV_{\text{General 2010}}^{\text{General 2011}} = \left[\frac{I_{\text{General 2011}}}{I_{\text{General 2010}}} - 1\right].100 = \left[\frac{154{,}25}{145{,}6} - 1\right].100 = 5{,}914\%$$

Se adjunta la tabla de precios y cantidades vendidas de tres productos por una determinada empresa durante tres períodos:

t	P_A	P_B	P_C	Q_A	Q_B	Q_C
0	4	10	15	2	2	3
1	6	11	20	5	1	3
2	5	12	25	4	1	2

a) Obtener los índices de precios y de cantidades de Paasche, de Laspeyres y de Fisher para estos tres períodos considerando como referencia el periodo 0.

b) Obtener los índices de valor.

Solución:

Índice ponderado de precios de Laspeyres: $L_p = \dfrac{\sum_{i=1}^{n} p_{it} \cdot q_{i0}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i0}} .100$

$$L_{p0}^{1} = \frac{6 \cdot 2 + 11 \cdot 2 + 20 \cdot 3}{4 \cdot 2 + 10 \cdot 2 + 15 \cdot 3} \cdot 100 = \frac{94}{73} \cdot 100 = 128{,}77$$

$$L_{p0}^{2} = \frac{5 \cdot 2 + 12 \cdot 2 + 25 \cdot 3}{4 \cdot 2 + 10 \cdot 2 + 15 \cdot 3} \cdot 100 = \frac{109}{73} \cdot 100 = 149{,}32$$

Índice ponderado de precios de Paasche: $P_p = \dfrac{\sum_{i=1}^{n} p_{it} \cdot q_{it}}{\sum_{i=1}^{n} p_{i0} \cdot q_{it}} .100$

$$P_{p0}^{1} = \frac{6 \cdot 5 + 11 \cdot 1 + 20 \cdot 3}{4 \cdot 5 + 10 \cdot 1 + 15 \cdot 3} \cdot 100 = \frac{101}{75} \cdot 100 = 134{,}67$$

$$P_{p0}^{2} = \frac{5 \cdot 4 + 12 \cdot 1 + 25 \cdot 2}{4 \cdot 4 + 10 \cdot 1 + 15 \cdot 2} \cdot 100 = \frac{82}{56} \cdot 100 = 146{,}43$$

Índice ponderado de precios de Fisher: $F_p = \sqrt{L_p \cdot P_p}$

$$F_{p0}^{1} = \sqrt{L_{p0}^{1} \cdot P_{p0}^{1}} = \sqrt{128{,}77 \cdot 134{,}67} = 131{,}69$$

$$F_{p0}^{2} = \sqrt{L_{p0}^{2} \cdot P_{p0}^{2}} = \sqrt{149{,}32 \cdot 146{,}43} = 147{,}87$$

t	P_A	P_B	P_C	L_p	P_p	F_p
0	4	10	15	100	100	100
1	6	11	20	128,77	134,67	131,69
2	5	12	25	149,32	146,43	147,87

Índice ponderado de cantidades de Laspeyres: $L_q = \dfrac{\sum_{i=1}^{n} q_{it} \cdot p_{i0}}{\sum_{i=1}^{n} q_{i0} \cdot p_{i0}} .100$

$$L_{q0}^{1} = \frac{5\,.\,4 + 1\,.\,10 + 3\,.\,15}{2\,.\,4 + 2\,.\,10 + 3\,.\,15}\,.\,100 = \frac{75}{73}\,.\,100 = 102,74$$

$$L_{q0}^{2} = \frac{4\,.\,4 + 1\,.\,10 + 2\,.\,15}{2\,.\,4 + 2\,.\,10 + 3\,.\,15}\,.\,100 = \frac{56}{73}\,.\,100 = 76,71$$

Índice ponderado de cantidades de Paasche: $P_q = \dfrac{\sum_{i=1}^{n} q_{it} \cdot p_{it}}{\sum_{i=1}^{n} q_{i0} \cdot p_{it}} .100$

$$P_{q0}^{1} = \frac{5\,.\,6 + 1\,.\,11 + 3\,.\,20}{2\,.\,6 + 2\,.\,11 + 3\,.\,20}\,.\,100 = \frac{101}{94}\,.\,100 = 107,45$$

$$P_{q0}^{2} = \frac{4\,.\,5 + 1\,.\,12 + 2\,.\,25}{2\,.\,5 + 2\,.\,12 + 3\,.\,25}\,.\,100 = \frac{82}{109}\,.\,100 = 75,23$$

Índice ponderado de cantidades de Fisher: $F_q = \sqrt{L_q \,.\, P_q}$

$$F_{p0}^{1} = \sqrt{L_{p0}^{1}\,.\,P_{p0}^{1}} = \sqrt{102,74\,.\,107,45} = 105,07$$

$$F_{p0}^{2} = \sqrt{L_{p0}^{2}\,.\,P_{p0}^{2}} = \sqrt{76,71\,.\,75,23} = 75,97$$

t	Q_A	Q_B	Q_C	L_q	P_q	F_q
0	2	2	3	100	100	100
1	5	1	3	102,74	107,45	105,07
2	4	1	2	76,71	75,23	75,97

b) Índice de Valor: Evolución del valor de la serie a precios constantes (se deflactan los valores en precios corrientes o actuales)

$$\text{Índice Valor} = \frac{\overbrace{\text{Valor nominal}}^{\text{precios corrientes}}}{\underbrace{\text{Valor real}}_{\text{precios constantes}}} \qquad IV_0^t = \frac{V_t}{V_0} \cdot 100 = \frac{\sum_{i=1}^{n} p_{it} \cdot q_{it}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i0}} \cdot 100$$

$$IV_0^1 = \frac{6 \cdot 5 + 11 \cdot 1 + 20 \cdot 3}{4 \cdot 2 + 10 \cdot 2 + 15 \cdot 3} \cdot 100 = \frac{101}{73} \cdot 100 = 138{,}36$$

$$IV_0^2 = \frac{5 \cdot 4 + 12 \cdot 1 + 25 \cdot 2}{4 \cdot 2 + 10 \cdot 2 + 15 \cdot 3} \cdot 100 = \frac{82}{73} \cdot 100 = 112{,}33$$

	Indices Precios			Indices Cantidades			Indices Valor
Año	L_p	P_p	F_p	L_q	P_q	F_q	IV
0	100	100	100	100	100	100	100
1	128,77	134,67	131,69	102,74	107,45	105,07	138,36
2	149,32	146,43	147,87	76,71	75,23	75,97	112,33

En la tabla adjunta se adjunta el valor de importaciones de un país durante los años 2011 y 2012.

Importaciones	2011	2012
Alimentos	2010	2230
Otros bienes de consumo	8400	8650
Bienes de capital	3400	3215
Bienes intermedios	5860	7450
	19670	21545

Se sabe que las importaciones tanto de alimentos como de otros bienes de consumo se pagaron un 5% más caras en 2012 que en 2011. Las importaciones de bienes de capital subieron sus precios un 1,5% y las de bienes intermedios bajaron un 0,5%.

a) Calcular el índice de precios total de las importaciones en 2012 con base 2011, utilizando Laspeyres y Paasche.

b) ¿Cuánto crecieron las importaciones en cantidad en 2011 con respecto a 2012?

Solución:

a) Utilizando el índice de precios de Laspeyres:

	Laspeyres	
Importaciones	$p_{i2011} \cdot q_{i2011}$	$p_{i2012} \cdot q_{i2011}$
Alimentos	2010	1,05 x 2010 = 2110,5
Otros bienes de consumo	8400	1,05 x 8400 = 8820
Bienes de capital	3400	1,015 x 3400 = 3451
Bienes intermedios	5860	0,995 x 5860 = 5830,7
TOTAL	19670	20212,2

$$L_{p_{2011}}^{2012} = \frac{\sum_{i=1}^{4} p_{i2012} \cdot q_{i2011}}{\sum_{i=1}^{4} p_{i2011} \cdot q_{i2011}} . 100 = \frac{20212,2}{19670} . 100 = 102,76\%$$

Utilizando el índice de precios de Paasche:

	Paasche	
Importaciones	$p_{i2011} \cdot q_{i2012}$	$p_{i2012} \cdot q_{i2012}$
Alimentos	2010	2230/1,05 = 2123,81
Otros bienes de consumo	8400	8650/1,05 = 8238,10
Bienes de capital	3400	3215/1,015 = 3167,49
Bienes intermedios	5860	7450/0,995 = 7487,44
TOTAL	19670	21016,84

$$P_{p\,2011}^{2112} = \frac{\sum_{i=1}^{4} p_{i2012} \cdot q_{i2012}}{\sum_{i=1}^{4} p_{i2011} \cdot q_{i2012}} .100 = \frac{21545}{21016,84} .100 = 102,51\%$$

b) Para calcular los índices cuánticos de Laspeyres y Paasche se requiere hallar previamente el índice de valor de las importaciones entre 2011 con base 2012.

$$IV_{2011}^{2012} = \frac{V_{2012}}{V_{2011}} .\,100 = \frac{\sum_{i=1}^{4} p_{i2012} \cdot q_{i2012}}{\sum_{i=1}^{4} p_{i2011} \cdot q_{i2011}} .\,100 = \frac{21545}{19670} .100 = 109,53\,\%$$

siendo, $IV_{2011}^{2012} = L_{P\,2011}^{\;2012} \,.\, P_{Q\,2011}^{\;2012} = P_{P\,2011}^{\;2012} \,.\, L_{Q\,2011}^{\;2012}$

$$P_{Q\,2011}^{\;2012} = \frac{IV_{2011}^{2012}}{L_{P2011}^{\;2012}} .100 = \frac{109,53}{102,76} .100 = 106,59\%$$

$$L_{Q\,2011}^{\;2012} = \frac{IV_{2011}^{2012}}{P_{P\,2011}^{\;2012}} .100 = \frac{109,53}{102,51} .100 = 106,85\%$$

Un trabajador ha recibido los siguientes salarios en los años 2011 y 2012:

Salario 2011 = 18.565 euros   Salario 2012 = 19.005 euros

Esta persona quiere saber si su poder adquisitivo ha aumentado en el año 2012 respecto al 2011. Para ello dispone de información relativa al Índice de Precios de Consumo con base el año 2008:

$$IPC_{2008}^{2011} = 109,93\,\% \quad \text{e} \quad IPC_{2008}^{2012} = 113,63\,\%$$

a) Interprete el valor de los números índice proporcionados

b) Determine e interprete la tasa de variación que ha sufrido el poder adquisitivo de este asalariado entre los años 2011 y 2012, en términos nominales y en términos reales (constante del 2008)

c) Suponga que el índice general de precios está formado por tres grupos de artículos, con índices: I1, I2, I3. Con pesos proporcionales de: 0,3 0,4 y 0,3.
Y que las repercusiones absolutas en el incremento del índice han sido de:
1, 1 y 1,7 respectivamente.
Calcule las participaciones y repercusiones relativas de los grupos en el incremento del índice general durante el periodo 2011-2012

Solución

a)

$IPC_{2008}^{2011} = 109,93\% \mapsto$ En el año 2011 los precios se han incrementado un 9,93% respecto al año 2008

$IPC_{2008}^{2012} = 113,63\% \mapsto$ En el año 2012 los precios se han incrementado un 13,63% respecto al año 2008

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

$$\overbrace{\text{Salario real}}^{\text{precios constantes}} = \frac{\overbrace{\text{Salario nominal}}^{\text{precios corrientes}}}{IPC_{2008}^{t}} \mapsto \begin{cases} SR_{2008}^{2011} = \dfrac{SN_{2008}^{2011}}{IPC_{2008}^{2011}} = \dfrac{18565}{1,0993} = 16888,02 \text{ euros} \\[2ex] SR_{2008}^{2012} = \dfrac{SN_{2008}^{2012}}{IPC_{2008}^{2012}} = \dfrac{19005}{1,1363} = 16725,34 \text{ euros} \end{cases}$$

$$\text{Tasas de variación}\begin{cases}\text{Nominal:} \quad TV_{2011}^{2012}=\left[\dfrac{19005}{18665}-1\right].\,100=2,37\% \\ \text{Real:} \quad TV_{2011}^{2012}=\left[\dfrac{16725,34}{16888,02}-1\right].\,100=-0,963\%\end{cases}$$

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)

I_i	$\mapsto$	%	$\mapsto$	R_i
I_1	$\mapsto$	0,3	$\mapsto$	$R_1=1$
I_2	$\mapsto$	0,4	$\mapsto$	$R_2=1$
I_3	$\mapsto$	0,3	$\mapsto$	$R_3=1,7$

$$\sum_{i=1}^{3} R_i = 3,7$$

Participación: $P_i=\dfrac{R_i}{\sum_{i=1}^{3}R_i}.100$

$$\text{Grupo 1 y Grupo 2: } P_i=\frac{R_i}{\sum_{i=1}^{3}R_i}.100=\frac{1}{3,7}.100=27,07\%$$

$$\text{Grupo 3: } P_i=\frac{R_i}{\sum_{i=1}^{3}R_i}.100=\frac{1,7}{3,7}.100=45,95\%$$

Repercusión relativa: $R_i^{\bullet}=\dfrac{R_i}{I_0^{t-1}}.100$

$$\text{Grupo 1: } R_1^{\bullet}=\frac{R_1}{I_{2008}^{2011}}.100=\frac{1}{100,93}.100=0,909\%$$

$$\text{Grupo 2: } R_2^{\bullet}=\frac{R_2}{I_{2008}^{2011}}.100=\frac{1}{100,93}.100=0,909\%$$

$$\text{Grupo 3: } R_3^{\bullet}=\frac{R_3}{I_{2008}^{2011}}.100=\frac{1,7}{100,93}.100=1,684\%$$

Series Históricas

Una serie temporal o cronológica es una sucesión de observaciones de una variable tomadas en el transcurso del tiempo.

El objetivo fundamental que se persigue es analizar la evolución de una variable a lo largo del tiempo, para poder efectuar predicciones sobre el comportamiento de dicha variable en un futuro próximo, suponiendo que persista la forma estructural donde se desarrolla el fenómeno en estudio.

El comportamiento de una serie cronológica viene determinado por la acción de factores que actúan conjuntamente sobre el fenómeno en cuestión, que son: Tendencia secular (T), componente estacional (E), componente cíclica (C) y componente accidental (A).

La variable Y puede adoptar las formas:

Esquema aditivo: $Y = T+E+C+A$

Esquema multiplicativo: $Y = T.E.C.A$

Ventas trimestrales de una empresa, expresadas en millones de euros, desestacionalizar la serie por el método de las medias móviles:

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	1	2	2	3	5
Segundo	2	3	4	4	7
Tercero	4	5	5	7	8
Cuarto	3	4	3	6	7

Solución:

$$\boxed{Y_{it} = T_{it} . E_{it} . C_{it} . A_{it}}$$

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.

$$\overline{Y}_{2,5} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{1+2+4+3}{4} = 2,5$$

$$\overline{Y}_{3,5} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{2+4+3+2}{4} = 2,75$$

$$\overline{Y}_{4,5} = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} = \frac{4+3+2+3}{4} = 3$$

$$\overline{Y}_{5,5} = \frac{Y_4 + Y_5 + Y_6 + Y_7}{4} = \frac{3+2+3+5}{4} = 3,25$$

$$\overline{Y}_{17,5} = \frac{Y_{16} + Y_{17} + Y_{18} + Y_{19}}{4} = \frac{6+5+7+8}{4} = 6,5$$

$$\overline{Y}_{18,5} = \frac{Y_{17} + Y_{18} + Y_{19} + Y_{20}}{4} = \frac{5+7+8+7}{4} = 6,75$$

SERIE DESCENTRADA

Trimestres \ Años	2008	2009	2010	2011	2012
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 corregir la nueva serie de móviles descentrada se calcula la media aritmética de cada dos valores sucesivos, asignando este nuevo valor al instante central de los dos periodos considerados, es decir:

SERIE CENTRADA: $T_{it}.C_{it}$

Trimestres \ Años	2008	2009	2010	2011	2012
Primero - Segundo	-----	3,125	3,750	4,000	6,375
Segundo - Tercero	-----	3,375	3,625	4,625	6,625
Tercero - Cuarto	2,625	3,500	3,625	5,250	-----
Cuarto - Primero	2,875	3,625	3,750	5,875	-----

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

La componente 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:

$$\boxed{\frac{Y_{it}}{T_{it}.C_{it}}=\frac{T_{it}.E_{it}.C_{it}.A_{it}}{T_{it}.C_{it}}=E_{it}.A_{it}}$$

quedando la componente estacional y accidental

SERIE CON COMPONENTE ESTACIONAL Y ACCIDENTAL: $E_{it}.A_{it}$

Trimestres \ Años	2008	2009	2010	2011	2012
Primero	---	2/3,125	2/3,750	3/4,000	5/6,375
Segundo	---	3/3,375	4/3,625	4/4,625	7/6,625
Tercero	4/2,625	5/3,500	5/3,625	7/5,250	---
Cuarto	3/2,875	4/3,625	3/3,750	6/5,875	---

COMPONENTE ESTACIONAL Y ACCIDENTAL

						COMPONENTE ESTACIONAL	
Trim \ Años	2008	2009	2010	2011	2012	IBVE	% IVE
Primero	-----	0,640	0,533	0,750	0,784	0,677	66,631
Segundo	-----	0,889	1,103	0,865	1,057	0,977	96,312
Tercero	1,524	1,429	1,379	1,333	-----	1,416	139,407
Cuarto	1,043	1,103	0,800	1,021	-----	0,992	97,651
						1,016	400,000

$$\text{Primero: } \frac{0,64+0,533+0,75+0,784}{4}=0,677 \qquad \text{Tercero: } \frac{1,524+1,429+1,379+1,333}{4}=1,416$$

$$\text{Segundo: } \frac{0,889+1,103+0,865+1,057}{4}=0,978 \qquad \text{Cuarto: } \frac{1,043+1,103+0,800+1,021}{4}=0,992$$

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: $\dfrac{Y_{it}}{\%\ IVE_t}.100$

Trimestres \ Años	2008	2009	2010	2011	2012	% IVE
Primero	1	2	2	3	5	66,631
Segundo	2	3	4	4	7	96,312
Tercero	4	5	5	7	8	139,407
Cuarto	3	4	3	6	7	97,651

$$\frac{1}{66,631}.100=1,501 \quad \frac{2}{66,631}.100=3,002 \quad \frac{2}{66,631}.100=3,002 \quad \frac{3}{66,631}.100=4,502 \quad \frac{5}{66,631}.100=7,504$$

$$\frac{2}{96,312}.100=2,077 \quad \frac{3}{96,312}.100=3,115 \quad \frac{4}{96,312}.100=4,153 \quad \frac{4}{96,312}.100=4,153 \quad \frac{7}{96,312}.100=7,268$$

$$\frac{4}{139,407}.100=2,869 \quad \frac{5}{139,407}.100=3,587 \quad \frac{5}{139,407}.100=3,587 \quad \frac{7}{139,407}.100=5,021 \quad \frac{8}{139,407}.100=5,739$$

$$\frac{3}{97,651}.100=3,072 \quad \frac{4}{97,651}.100=4,096 \quad \frac{3}{97,651}.100=3,072 \quad \frac{6}{97,651}.100=6,144 \quad \frac{7}{97,651}.100=7,168$$

SERIE DESESTACIONALIZADA

Trimestres \ Años	2008	2009	2010	2011	2012
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	5,739
Cuarto	3,072	4,096	3,072	6,144	7,168

Tras analizar los datos referentes a un año y medio (desde 2011.1 hasta 2012.2) de una determinada serie temporal (Y), de periodicidad trimestral, se han obtenido los siguientes resultados con $t = 0, 1, \cdots, 5$

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

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) Realizar un ajuste lineal de la tendencia de la serie. A partir del coeficiente de determinación lineal determinar sí el ajuste es bueno o malo, y predecir el valor de la serie para el tercer y cuarto trimestre del año 2012.

b) Interpretar estadísticamente los IVEs.

Solución:

a) Recta de regresión de Y sobre t: $y = a + b_{Y/t} \cdot t \quad \begin{cases} b_{Y/t} = \dfrac{\sigma_{Yt}}{\sigma_t^2} \\ a = \overline{y} - b_{Y/t}\, \overline{t} \end{cases}$

$$a_{10} = \overline{y} = \frac{\sum_{t=1}^{6} y_t}{N} = \frac{19073}{6} = 3178{,}83 \qquad a_{01} = \overline{t} = \frac{\sum_{i=1}^{6} t}{N} = \frac{15}{6} = 2{,}5$$

$$a_{11} = \frac{\sum_{t=1}^{6} t\, y_t}{N} = \frac{71950}{6} = 11991{,}67$$

$$m_{11} = \sigma_{Yt} = a_{11} - \overline{y} \cdot \overline{t} = 11991{,}67 - 3178{,}83 \cdot 2{,}5 = 4044{,}59$$

$$\sigma_t^2 = a_{02} - (a_{01})^2 = \frac{\sum_{i=1}^{6} t^2}{6} - \overline{t}^2 = \frac{55}{6} - 2{,}5^2 = 2{,}92$$

con lo que, $b_{Y/t} = \dfrac{\sigma_{Yt}}{\sigma_t^2} = \dfrac{4044{,}59}{2{,}92} = 1385{,}13$

$$a = \overline{y} - b_{Y/t}\, \overline{t} = 3178{,}83 - 1385{,}13 \cdot 2{,}5 = -283{,}99$$

Recta de regresión de Y sobre t: $y = -283{,}99 + 1385{,}13 \cdot t$

Coeficiente de determinación lineal: $R^2 = b_{Y/t} \cdot b_{T/y}$

$$\sigma_y^2 = \frac{\sum_{t=1}^{6} y_t^2}{N} - \overline{y}^2 = \frac{97199705}{6} - 3178,83^2 = 6094990,66$$

$$b_{T/y} = \frac{\sigma_{ty}}{\sigma_y^2} = \frac{4044,59}{6094990,66} = 0,00066$$

$$R^2 = b_{Y/t}\,.\,b_{T/y} = 1385,13\,.\,0,00066 = 0,914$$

El modelo es bueno porque explica el 91,4% de la variabilidad de Y_t en función de t.

Para predecir el tercer $t=6$ y cuarto trimestre $t=7$ de 2012, siendo la recta de regresión de Y sobre t: $y = -283,99 + 1385,13\,.\,t$

2012.3: $y = -283,99 + 1385,13\,.\,6 = 8026,79$

2012.4: $y = -283,99 + 1385,13\,.\,7 = 9411,92$

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

$$Y_{it} = T_{it}\,.\,IVE_h \begin{cases} Y_{2012.3} = T_{2012.3}\,.\,IVE_3 = 8026,79\,.\,0,97 = 7785,99 \\ Y_{2012.4} = T_{2012.4}\,.\,IVE_4 = 9411,92\,.\,1,127 = 10607,23 \end{cases}$$

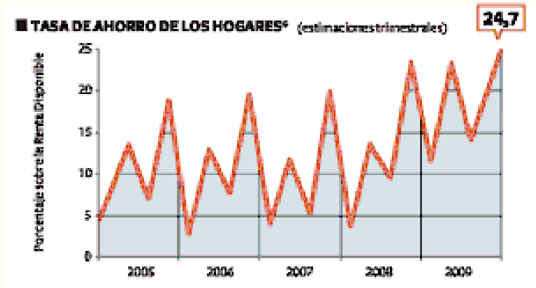
b) Los índices de variación estacional muestran el componente estacional en el esquema multiplicativo. El componente estacional E_{it} son las oscilaciones que sufre una serie temporal en periodos inferiores o iguales a un año.

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

- **$IVE_1 = 1,033$ 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.**

La siguiente gráfica, publicada en *El País*, en abril de 2010, muestra la evolución trimestral de la tasa de ahorro de los hogares entre 2005 y 2009:



a) ¿Cuál de los siguientes modelos trimestrales de estacionalidad es el más adecuado? Justificar la respuesta.

Opción 1: $IVE1 = -5 \quad IVE2 = 2,5 \quad IVE3 = -2,5 \quad IVE4 = \ ?$

Opción 2: $IVE1 = 5 \quad IVE2 = -2,5 \quad IVE3 = 2,5 \quad IVE4 = \ ?$

b) Con la información anterior, ¿se trata de un modelo aditivo o multiplicativo?

c) Suponiendo que la componente ciclo-tendencia durante los años 2005 a 2009 es $y = 10 + 0,5\,t$ siendo $t = 1$ el primer trimestre de 2005, ¿cuál es la previsión de ahorro para cada uno de los cuatro trimestres de los años 2010 y 2011?

d) Discuta, a la vista de la gráfica y de la situación económica actual, si es adecuado o no simular el componente ciclo-tendencia mediante una recta. Justificar la respuesta.

Solución:

a) La Opción 1,se observa que el primer trimestre está por debajo de la media del año.

b) Se trata de un modelo Aditivo, $Y_{it} = T_{it} + E_{it} + C_{it} + A_{it}$**, puesto que hay Índices de Variación Estacional negativos.**

c) Siendo la componente ciclo-tendencia (CT) durante los años 2005 a 2009: $y = 10 + 0,5\,t$**. La previsión de ahorro para cada uno de los cuatro trimestres de los años 2010 y 2011:**

Año	Trimestre	t	CT	IVE	CT + IVE
2010	1	21	20,5	– 5	15,5
	2	22	21	2,5	23,5
	3	23	21,5	– 2,5	19
	4	24	22	5	27
2011	1	25	22,5	– 5	17,5
	2	26	23	2,5	25,5
	3	27	23,5	– 2,5	21
	4	28	24	5	29

d) Hay un cambio de tendencia en 2008, con motivo de la crisis económica, así que el considerar el CT como una recta no es totalmente correcto.

Dada la información del I.P.C. se solicitan las repercusiones y participaciones de cada uno de los grupos.
¿Cuál es el grupo más afectado por la subida de los precios?

Grupos	Indices 2007	Ponderaciones	Indices 31/12/2008
1. Alimentos, bebidas y tabaco	100	367,2	125,9
2. Vestido y calzado	100	100,12	132,8
3. Vivienda	100	157,3	133,4
4. Menaje	100	76,1	122
5. Servicios médicos y sanitarios	100	42,65	123
6. Transportes y comunicaciones	100	92,35	126,5
7. Esparcimiento, enseñanza y cultura	100	78,15	128,4
8. Otros bienes y servicios	100	86,13	134,4
	100	1000	128,33

Solución:

El I.P.C. es un índice de Laspeyres $L_p = \dfrac{\sum_{i=1}^{n} I_i . w_i}{\sum_{i=1}^{n} w_i}$ siendo I_i los índices de cada grupo
Y w_i las ponderaciones de cada bien o servicio.

La repercusión de cada grupo i-ésimo $(i = 1, 2, \ldots, 8)$ en la variación global del I.P.C. desde 2007 a 2008:

$$R_1 = \frac{\Delta I_1 . w_1}{\sum_{i=1}^{n} w_i} = \frac{(125,9-100).367,2}{1000} = 9,51\%$$

$$R_2 = \frac{\Delta I_2 . w_2}{\sum_{i=1}^{n} w_i} = \frac{(132,8-100).100,12}{1000} = 3,284\%$$

$$R_3 = \frac{\Delta I_3 . w_3}{\sum_{i=1}^{n} w_i} = \frac{(133,4-100).157,3}{1000} = 5,254\%$$

$$R_4 = \frac{\Delta I_4 . w_4}{\sum_{i=1}^{n} w_i} = \frac{(122-100).76,1}{1000} = 1,674\%$$

$$R_5 = \frac{\Delta I_5 . w_5}{\sum_{i=1}^{n} w_i} = \frac{(123-100).42,65}{1000} = 0,981\%$$

$$R_6 = \frac{\Delta I_6 . w_6}{\sum_{i=1}^{n} w_i} = \frac{(126,5-100).92,35}{1000} = 2,447\%$$

$$R_7 = \frac{\Delta I_7 . w_7}{\sum_{i=1}^{n} w_i} = \frac{(128,4-100).78,15}{1000} = 2,219\%$$

$$R_8 = \frac{\Delta I_8 . w_8}{\sum_{i=1}^{n} w_i} = \frac{(134,4-100).86,13}{1000} = 2,963\%$$

Grupos	Índice 2007 I_i	Ponderación w_i	Índice 2008 $I_i + \Delta I_i$	Repercusión $R_i = \frac{\Delta I_i . w_i}{\sum_{i=1}^{8} w_i}$
1. Alimentos, bebidas y tabaco	100	367,2	125,9	9,510
2. Vestido y calzado	100	100,12	132,8	3,284
3. Vivienda	100	157,3	133,4	5,254
4. Menaje	100	76,1	122	1,674
5. Servicios médicos y sanitarios	100	42,65	123	0,981
6. Transportes y comunicaciones	100	92,35	126,5	2,447
7. Esparcimiento, enseñanza y cultura	100	78,15	128,4	2,219
8. Otros bienes y servicios	100	86,13	134,4	2,963
	100	1000	128,33	28,33

La suma de las Repercusiones $\sum_{i=1}^{8} R_i = 28,33\%$ es igual a la Variación Índice General $\Delta L_P = \sum_{i=1}^{8} R_i$:

$$\Delta L_P = \frac{\sum_{i=1}^{8}(I_i + \Delta I_i).w_i}{\sum_{i=1}^{8} w_i} - \frac{\sum_{i=1}^{8} I_i.w_i}{\sum_{i=1}^{8} w_i} = 128,33 - 100 = 28,33\%$$

La PARTICIPACIÓN de cada grupo en la variación del I.P.C. viene dada por la relación:

$$P_i = \frac{R_i}{\sum_{i=1}^{8} \Delta I_i.w_i}.100 = \frac{R_i}{\sum_{i=1}^{8} R_i}.100 \quad \text{así} \quad P_2 = \frac{R_2}{\sum_{i=1}^{8} R_i}.100 = \frac{3,284}{28,33}.100 = 11,59\%$$

Grupos	Repercusión $R_i = \frac{\Delta I_i.w_i}{\sum_{i=1}^{8} w_i}$	Participación $P_i = \frac{R_i}{\sum_{i=1}^{8} R_i}.100$	Repercusión en porcentaje $\%R_i = \frac{R_i}{L_P}.100$
1. Alimentos, bebidas y tabaco	9,510	33,57	9,510
2. Vestido y calzado	3,284	11,59	3,284
3. Vivienda	5,254	18,54	5,254
4. Menaje	1,674	5,91	1,674
5. Servicios médicos y sanitarios	0,981	3,46	0,981
6. Transportes y comunicaciones	2,447	8,64	2,447
7. Esparcimiento, enseñanza y cultura	2,219	7,83	2,219
8. Otros bienes y servicios	2,963	10,46	2,963
	28,33	100,00	28,33

La REPERCUSIÓN porcentual de cada uno de los grupos viene dado por la expresión:

$$\%R_i = \frac{R_i}{L_P} = \frac{\Delta I_i.w_i}{\sum_{i=1}^{n} I_i.w_i}.100 \quad \text{donde} \quad L_P = \frac{\sum_{i=1}^{8} I_i.w_i}{\sum_{i=1}^{8} w_i} = 100$$

LA VARIACIÓN (en porcentaje) DEL ÍNDICE GENERAL es la suma de las repercusiones (en porcentaje)

$$\sum_{i=1}^{8} \%R_i = 28,33 \text{ o también, } \frac{\Delta L_P}{L_P}.100 = \frac{128,33 - 100}{100}.100 = 28,33$$

El primer grupo (alimentos, bebidas y tabaco) es el que más ha influido en la subida del I.P.C., suponiendo un 33,57% de la variación total.
Es decir, en la subida del índice en un 28,33% ha tenido un peso del 9,51%.

De otra parte, el quinto grupo (servicios médicos y sanitarios) es el que menos ha influido en la subida del IPC, representando un 3,46% de la variación total.
Esto es, en la subida del índice en un 28,33% ha repercutido en 0,981%.

En la tabla adjunta se reflejan las ventas trimestrales de una empresa en millones de euros. Desestacionalizar la serie por el método de las medias móviles y el método analítico de los mínimos cuadrados.

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

Solución:

MÉTODO DE LAS MEDIAS MÓVILES

PRIMER PASO.- Para calcular la tendencia secular de la serie por el método de las medias móviles, se obtienen primero 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.

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

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

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

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

$\frac{2+3+5+4}{4}=3,5$ 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	---

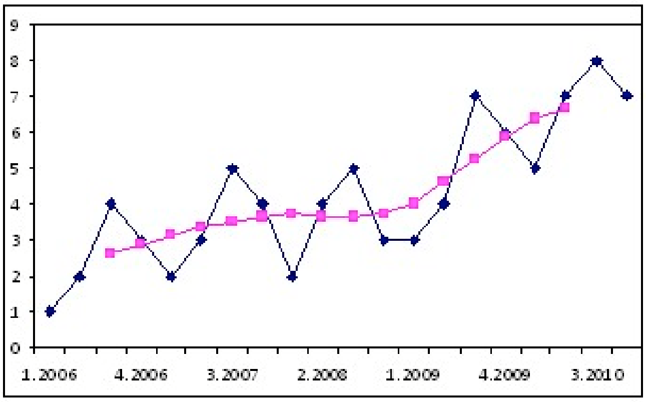
Para centrar la serie hay que calcular la media aritmética de cada dos observaciones sucesivas, de este modo, las medias que irán apareciendo, respectivamente, serán:

....	$\frac{3+3,25}{2}=3,125$	$\frac{3,75+3,75}{2}=3,75$	$\frac{3,75+4,25}{2}=4$	$\frac{6,25+6,5}{2}=6,375$
....	$\frac{3,25+3,5}{2}=3,375$	$\frac{3,75+3,5}{2}=3,625$	$\frac{4,25+5}{2}=4,625$	$\frac{6,5+6,75}{2}=6,625$
$\frac{2,5+2,75}{2}=2,625$	$\frac{3,5+3,5}{2}=3,5$	$\frac{3,5+3,75}{2}=3,625$	$\frac{5+5,5}{2}=5,25$	....
$\frac{2,75+3}{2}=2,875$	$\frac{3,5+3,75}{2}=3,625$	$\frac{3,75+3,75}{2}=3,75$	$\frac{5,5+6,25}{2}=5,875$	....

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 $(t\,,\overline{\overline{y}}_{it})$ 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 $Y_{it} = T_{it}.E_{it}.C_{it}.A_{it}$, lo que realmente *se obtiene en la serie cronológica es una aproximación de* $T_{it}.C_{it}$, quedando sin analizar las componentes estacional E_{it} y accidental A_{it}

SEGUNDO PASO.- **La tendencia y la componente cíclica se eliminarán dividiendo cada dato de la serie original por la correspondiente media móvil:**

$$\frac{Y_{it}}{T_{it}.C_{it}} = \frac{T_{it}.E_{it}.C_{it}.A_{it}}{T_{it}.C_{it}} = E_{it}.A_{it}$$ **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	---

TERCER PASO.- **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 $E_{it}.A_{it}$):**

$$\frac{0,640+0,533+0,750+0,784}{4}=0,677 \qquad \frac{0,889+1,103+0,865+1,057}{4}=0,978$$

$$\frac{1,524+1,429+1,379+1,333}{4}=1,416 \qquad \frac{1,043+1,103+0,8+1,021}{4}=0,992$$

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
						1,016

Se calcula la media aritmética de los cuatro valores obtenidos anteriormente

$$\frac{0,677+0,978+1,416+0,992}{4}=1,016$$

CUARTO PASO.- **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) 100 = 66,63
Segundo	(0,978/1,016) 100 = 96,31
Tercero	(1,416/1,016) 100 = 139,41
Cuarto	(0,992/1,016) 100= 97,65

Sobre un nivel medio de ventas, la influencia de la variación estacional produce:

1º Trimestre: (66,63 –100 = –33,37) → un descenso de ventas del 33,37%
2º Trimestre: (96,31 –100 = – 3,69) → un descenso de ventas del 3,69%
3º Trimestre: (139,41 – 100 = 39,41) → un aumento de ventas del 39,41%
4º Trimestre: (97,65 – 100 = –2,35) → un descenso de ventas del 2,35%

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, esto es:

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, aplicando el 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

MÉTODO ANALÍTICO DE LA TENDENCIA (MÍNIMOS CUADRADOS)

PRIMER PASO.- Se calculan las medias anuales $\bar{y}_{\bullet t}$ (medias para cada año de k = 4 subperíodos)

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

Trimes / 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
	$\bar{y}_{\bullet 2006} = 2,5$	$\bar{y}_{\bullet 2007} = 3,5$	$\bar{y}_{\bullet 2008} = 3,5$	$\bar{y}_{\bullet 2009} = 5$	$\bar{y}_{\bullet 2010} = 6,75$

SEGUNDO PASO.- La *tendencia media anual* $\bar{T}_{\bullet t}$ se obtiene ajustando una recta de regresión a los años $(t_1\,, t_2\,, \cdots\,, t_n)$ y a las medias anuales $\bar{y}_{\bullet t}$ donde

$t \equiv (t_1\,, t_2\,, \cdots\,, t_n):\ \bar{T}_{\bullet t} = \hat{\bar{y}}_{\bullet t} = a + b.t$

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

Tendencia media anual

$t_{2006}\,, t_{2007}\,, \cdots\,, t_{2010}$	2006	2007	2008	2009	2010
$\bar{T}_{\bullet t} = -2003,75 + t$	2,25	3,25	4,25	5,25	6,25

TERCER PASO.- **A partir de la tendencia media anual $\overline{T}_{\bullet t}$ se obtiene el valor de la *tendencia para los distintos subperíodos*, según la expresión general:**

$$T_{it} = \overline{T}_{\bullet t} + \left[i - \frac{k+1}{2} \right] \frac{b}{k}$$ **tendencia media anual para los subperíodos k-ésimos**

donde,

- **t ≡ Año (2006, 2007, ... , 2010)**
- **i ≡ subperíodos donde se calcula la tendencia (trimestral i = 1, 2, 3, 4)**
- **k ≡ Número total de subperíodos (datos trimestrales k = 4)**
- **b ≡ Pendiente de la recta de regresión = 1**

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

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

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

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

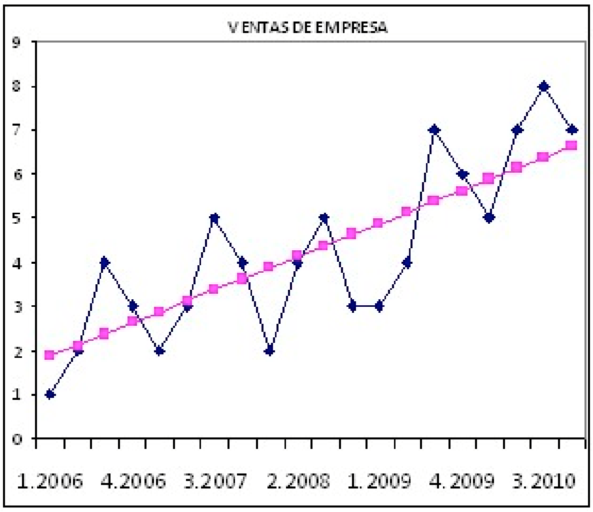
Trimestre Primero 2007: $T_{1\,2007} = 3,25 + \left[1 - \frac{4+1}{2} \right] \frac{1}{4} = 2,875$

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

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

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

Serie con los datos originales y serie suavizada de tendencia



CUARTO PASO.- **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

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 de $E_{it} \cdot A_{it}$ **, 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 manifiestes plenamente las variaciones cíclicas.**

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

QUINTO PASO.- 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.

$$\frac{0,533+0,696+0,516+0,615+0,851}{5}=0,642 \qquad \frac{0,941+0,96+0,97+0,78+1,143}{5}=0,959$$

$$\frac{1,684+1,481+1,143+1,302+1,255}{5}=1,373 \qquad \frac{1,143+1,103+0,649+1,067+1,057}{5}=1,004$$

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

Promedio anual de las cuatro medias aritméticas:

$$\frac{0,642+0,959+1,373+1,004}{4}=0,994$$

SEXTO PASO.- 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) → un descenso de ventas del 35,41%
2º Trimestre: (96,48 – 100 = –3,52) → un descenso de ventas del 3,42%
3º Trimestre: (138,13 –100 = 38,13) → un aumento de ventas del 38,13%
4º Trimestre: (101,01 –100 = 1,01) → un aumento de ventas del 1,01%

DESESTACIONALIZACIÓN (aplicando el método a la razón a la tendencia)

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	1/0,6459	2/0,6459	2/0,6459	3/0,6459	5/0,6459
Segundo	2/0,9648	3/0,9648	4/0,9648	4//0,9648	7/0,9648
Tercero	4/1,3813	5/1,3813	5/1,3813	7/1,3813	8/1,3813
Cuarto	3/1,0101	4/1,0101	3/1,0101	6/1,0101	7/1,0101

SERIE DESESTACIONALIZADA, aplicando el método a la razón a la tendencia

Trimestres / Años	2006	2007	2008	2009	2010
Primero	1,548	3,096	3,096	4,645	7,741
Segundo	2,073	3,109	4,146	4,146	7,255
Tercero	2,896	3,620	3,620	5,068	5,792
Cuarto	2,970	3,960	2,970	5,940	6,930