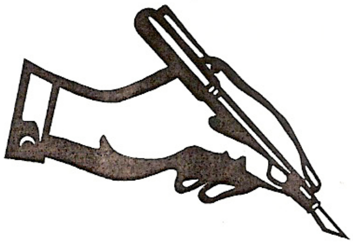
UNIVERSIDAD AUTONOMA DE MADRID

Estadística Descriptiva - EXCEL - SPSS
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández

Estadística Teórica I



EXÁMENES

EXAMEN DE ESTADÍSTICA DESCRIPTIVA - GRADO ECONOMÍA Mayo 2011

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

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

a) Represente el histograma de esta distribución.

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

c) Calcule el consumo más frecuente.

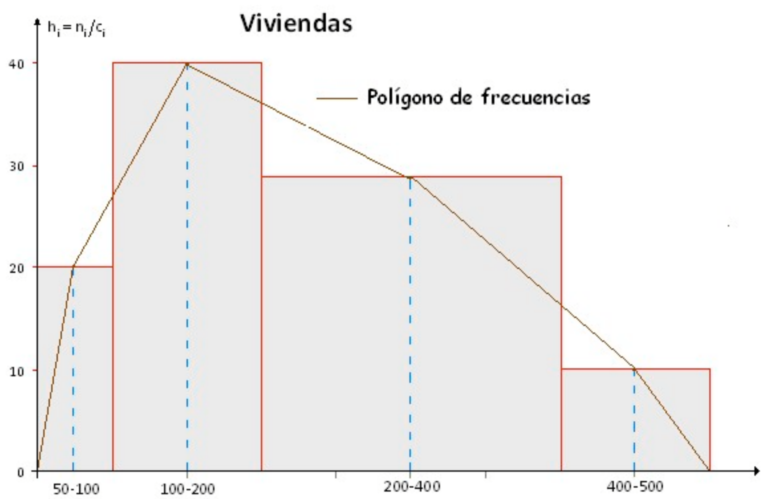
d) Averigüe el valor del tercer cuartil de la distribución del consumo de gas y explique su significado

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

Solución:

a)

Consumo gas	amplitud c_i	n_i	densidad $h_i = \frac{n_i}{c_i}$	N_i	x_i	$x_i n_i$	$x_i^2 n_i$
50 - 100	50	10	0,2	10	75	750	56250
100 - 200	100	40	0,4	50 90	150	6000	900000
200 - 400	200	60	0,3	110	300	18000	5400000
400 - 500	100	10	0,1	120	450	4500	2025000
						29250	8381250



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

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

$$\sigma_X = \sqrt{10429,6875} = 102,1258 \ m^3 \qquad C.V = \frac{\sigma_X}{\overline{x}} = \frac{102,1258}{243,75} = 0,42 \ (42\%)$$

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

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

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

La moda aproximada cuando existen distintas amplitudes:

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

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

d) El tercer cuartil: $\frac{3.120}{4} = 90$, observando en la columna N_i, $Q_3 = P_{75} = L_i + \frac{\frac{3.N}{4} - N_{i-1}}{N_i - N_{i-1}} c_i$, de donde:

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

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

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

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

$$\sigma_Y^2 = Var(20 + 0,5.X) = 0,5^2 . \sigma_X^2 \ \mapsto \ \sigma_Y = 0,5.\sigma_X = 0,5.102,1258 = 51,063 \text{ €}$$

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

La factura del gas está menos dispersa que el consumo.

NOTA.- Cambio de origen y de escala:

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

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

$$\sigma_y^2=\frac{\sum_{i=1}^{k}(y_i-\overline{y})^2.n_i}{N}=\frac{\sum_{i=1}^{k}(ax_i+b-a\overline{x}-b)^2.n_i}{N}=a^2\frac{\sum_{i=1}^{k}(x_i-\overline{x})^2.n_i}{N}=a^2\,\sigma_x^2 \;\Rightarrow\; \text{Var}(ax+b)=a\sigma_x^2$$

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

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

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

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

Solución:

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

Siendo $s_{ry}^2=\sigma_r^2=\sigma_y^2\,(1-R^2)=0 \;\rightarrow\; 1-R^2=0 \;\Rightarrow\; R^2=1$, existe una dependencia funcional, el ajuste es perfecto. Con lo cual, $\sigma_r^2=\sigma_y^2=25$.

Para calcular la covarianza ($s_{xy}=m_{11}$), tenemos en cuenta que

$$R=\frac{m_{11}}{\sigma_x.\sigma_y}=1 \;\mapsto\; m_{11}=\sigma_x.\sigma_y=\sqrt{36.25}=30$$

b) El coeficiente de determinación R^2 es invariante ante un cambio de origen y de escala, con lo que la bondad del ajuste será idéntico.

c) constante c = $\overline{y}$

NOTA.- Cambio de origen y de escala

Sea una distribución bidimensional (X, Y), con un cambio de origen y escala, es decir, se introducen unas nuevas variables (X', Y') relacionadas con las anteriores, de forma que $\begin{cases} X' = mX + n \\ Y' = pX + q \end{cases}$

📖 Si $a_{10}, a_{20}, m_{20}, a_{01}, a_{02}, m_{02}$ son los momentos relacionados con (x_i, y_i) y $a'_{10}, a'_{20}, m'_{20}, a'_{01}, a'_{02}, m'_{02}$ los momentos relacionados con (x'_i, y'_i), se tiene:

$$a'_{10} = \frac{\sum_{i=1}^{N} x'_i}{N} = \frac{1}{N}\sum_{i=1}^{N}(m x_i + n) = m\frac{1}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}\sum_{i=1}^{N} n = m a_{10} + n \text{ , análogamente } a'_{01} = p a_{01} + q$$

- Las medias se ven afectadas por el cambio de origen y de escala efectuado en la variable.

📖 $$m'_{20} = \sigma^2_{x'} = \frac{1}{N}\sum_{i=1}^{N}(x'_i - a'_{10})^2 = \frac{1}{N}\sum_{i=1}^{N}(m x_i + n - m a_{10} - n)^2 = m^2 \frac{1}{N}\sum_{i=1}^{N}(x_i - a_{10})^2 = m^2 \sigma^2_x = m^2 m_{20}$$

análogamente, $m'_{02} = p^2 \sigma^2_y = p^2 m_{02}$

- Las varianzas son invariantes ante un cambio de origen pero no ante un cambio de escala.

📖

$$m'_{11} = \frac{1}{N}\sum_{i=1}^{N}(x'_i - a'_{10}).(y'_i - a'_{01}) = \frac{1}{N}\sum_{i=1}^{N}(m x_i + n - m a_{10} - n).(p y_i + q - p a_{01} - q) = (mp)\frac{1}{N}\sum_{i=1}^{N}(x_i - a_{10}).(y_i - a_{01}) = (mp)\, m_{11}$$

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

📖 Sean $\beta_{Y/X}$ e $\beta_{X/Y}$, respectivamente, los coeficientes de regresión de las rectas (Y/X) e (X/Y). Análogamente, $\beta'_{Y'/X'}$ e $\beta'_{X'/Y'}$, los coeficientes de regresión de las rectas (Y'/X') e (X'/Y'). Se tiene:

$$\beta'_{Y'/X'} = \frac{m'_{11}}{m'_{20}} = \frac{(mp).m_{11}}{m^2 m_{20}} = \frac{p}{m}.\frac{m_{11}}{m_{20}} = \frac{p}{m}.\beta_{Y/X} \text{ , análogamente, } \beta'_{X'/Y'} = \frac{m}{p}.\beta_{X/Y}$$

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

📖 El coeficiente de determinación $R'^2 = \beta'_{Y/X}.\beta'_{X/Y} = \left(\frac{p}{m}.\beta_{Y/X}\right).\left(\frac{m}{p}.\beta_{X/Y}\right) = \beta_{Y/X}.\beta_{X/Y} = R^2$

- El coeficiente de determinación es invariante ante un cambio de origen y de escala. En consecuencia, también lo será el coeficiente de correlación.

3. En la siguiente tabla se observa la evolución de la inversión de una persona en dos fondos distintos, uno de renta fija, y otro de renta variable (en euros corrientes). La inversión inicial en los fondos fue de 5.000 y 8.000 euros, en 2004.

	Fondo Renta Fija	**Fondo Renta Variable**	**IPC (2004=100%)**	**IPC (2007=100%)**
2004	5.000,00	8.000,00	100,0%	
2005	5.250,00	8.640,00	103,7%	
2006	5.617,50	9.417,60	106,5%	
2007	6.123,08	10.076,83	111,0%	100,0%
2008	6.796,61	9.572,99		102,4%
2009	4.757,63	5.743,79		105,5%

a) ¿Cuál ha sido la tasa de variación media anual de su inversión entre 2004 y 2009 en cada uno de los fondos, a precios corrientes?
b) Complete la serie del IPC base 2004 y 2007 para todos los años.
c) Calcule el IPC base con base 100% en el año 2009.
d) Deflacte ambas series temporales, poniéndolas en euros constantes de 2009.
e) ¿Cuál ha sido la tasa de variación media anual de su inversión entre 2004 y 2009 en cada uno de los fondos, a precios constantes de 2009?
f) ¿Cuál ha sido la tasa de variación media anual del IPC entre 2004 y 2009?

Solución:

a) Tasa variación media anual Fondo Renta Fija entre 2004/2009, a precios corrientes:

$$t_{m0}^{t} = \sqrt[t]{1+t_0^t} - 1 = \sqrt[t]{I_0^t} - 1 \mapsto t_{m04}^{09} = \sqrt[5]{I_{04}^{09}} - 1 = \sqrt[5]{\frac{4753,63}{5000}} - 1 = -0,0099 \ \ (-0,99\%) \ \text{precios corrientes}$$

Tasa variación media anual Fondo Renta Variable entre 2004/2009, a precios corrientes:

$$t_{m0}^{t} = \sqrt[t]{I_0^t} - 1 \mapsto t_{m04}^{09} = \sqrt[5]{I_{04}^{09}} - 1 = \sqrt[5]{\frac{5743,79}{8000}} - 1 = -0,0641 \ \ (-6,41\%) \ \text{precios corrientes}$$

b - c) Para completar la tabla dada:

	Fondo Renta Fija	Fondo Renta Variable	IPC (2004=1)	IPC (2007=1)	IPC (2009=1)
2004	5000,00	8000,00	100,0%	100/1,11 90,1%	100/117,1 85,4%
2005	5250,00	8640,00	103,7%	103,7/1,11 93,4%	103,7x0,854 88,6%
2006	5617,50	9417,60	106,5%	106,5/1,11 96,0%	106,5x0,854 91,0%
2007	6123,08	10076,83	111,0%	100,0%	111x0,854 94,8%
2008	6796,61	9572,99	111x1,024 113,6%	102,4%	113,6x0,854 97%
2009	4757,63	5743,79	111x1,055 117,1%	105,5%	117,1x0,854 100,0%

d) Para deflactar ambas series temporales, en euros constantes de 2009, los valores en euros corrientes de cada serie se dividen por el IPC base 2009

	Fondo Renta Fija	Fondo Renta Variable
2004	5000/0,854 5854,80	8000/0,854 9367,68
2005	5250/0,886 5925,51	8640/0,886 9751,69
2006	6173,08/0,91 6783,60	9417,60/0,91 10349,01
2007	6123,08/0,948 6458,95	10076,83/0,948 10629,57
2008	6796,61/0,97 7006,82	9572,99/0,97 9869,06
2009	4757,63/1 4757,63	5743,79/1 5743,79

e) La tasa de variación media anual de la inversión entre 2004 y 2009 en cada uno de los fondos, a precios constantes de 2009 será:

Tasa variación media anual Fondo Renta Fija entre 2004/2009, a precios constantes de 2009:

$$t_{m0}^{t} = \sqrt[t]{1+t_0^t} - 1 = \sqrt[t]{I_0^t} - 1 \mapsto t_{m04}^{09} = \sqrt[5]{I_{04}^{09}} - 1 = \sqrt[5]{\frac{4757,63}{5854,80}} - 1 = -0,04065 \ \ (-4,06\%) \ \text{precios constantes}$$

Tasa variación media anual Fondo Renta Variable entre 2004/2009, a precios constantes de 2009:

$$t_{m0}^{t} = \sqrt[t]{I_0^t} - 1 \mapsto t_{m04}^{09} = \sqrt[5]{I_{04}^{09}} - 1 = \sqrt[5]{\frac{5743,79}{9367,68}} - 1 = -0,0932 \ \ (-9,32\%) \ \text{precios constantes}$$

f) La tasa de variación media anual del IPC entre 2004 y 2009:

$$tm_{IPC0}^{t} = \sqrt[t]{I_0^t} - 1 \mapsto tm_{IPC04}^{09} = \sqrt[5]{I_{04}^{09}} - 1 = \sqrt[5]{\frac{117,1}{100}} - 1 = 0,0320 \ \ (3,20\%)$$

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

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

Suponiendo un esquema de agregación multiplicativo en la serie temporal:

a) ¿Qué características debe presentar la serie si se ha elegido dicho esquema?
b) Calcule e interprete los Índices de Variación Estacional (IVEs).
c) Desestacionalice la serie de ventas por el método de las medias móviles.

Solución:

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

Cálculo de las medias móviles:

$$\frac{1+2+4+3}{4}=2,5 \ \textit{entre segundo y tercer trimestre de 2006}$$
$$\frac{2+4+3+2}{4}=2,75 \ \textit{entre tercer y cuarto trimestre de 2006}$$
$$\frac{4+3+2+3}{4}=3 \ \textit{entre cuarto trimestre de 2006 y primer trimestre de 2007}$$
$$\frac{3+2+3+5}{4}=3,25 \ \textit{entre primer y segundo trimestre de 2007}$$
$$\frac{2+3+5+4}{4}=3,5 \ \textit{entre segundo y tercer trimestre de 2007}$$

SERIE DESCENTRADA de medias móviles

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

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

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

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

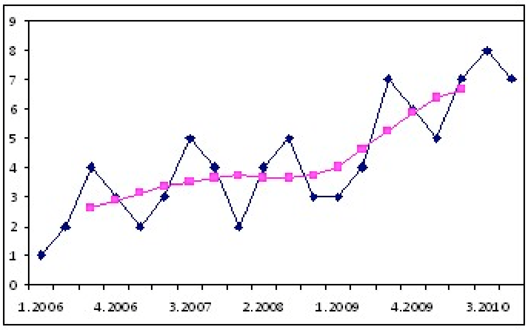
$$\frac{3,75+4,25}{2}=4 \quad \frac{4,25+5}{2}=4,625 \quad \frac{5+5,5}{2}=5,25 \quad \frac{5,5+6,25}{2}=5,875 \quad \frac{6,25+6,5}{2}=6,375$$

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

SERIE CENTRADA de las medias móviles:

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

La línea que se obtiene al representar gráficamente la serie de la tabla $(t\,,\overline{\overline{y}}_{it})$ 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}).

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	---

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

$$\frac{0{,}677+0{,}978+1{,}416+0{,}992}{4}=1{,}016$$

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
	400 %

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, método a la razón a la media móvil

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

b) MÉTODO ANALÍTICO DE LA TENDENCIA (MÍNIMOS CUADRADOS)

Se calculan las medias anuales $\bar{y}_{\bullet t}$ (medias para cada año de k = 4 subperíodos)

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

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

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

Tendencia media anual

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

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

$$T_{it}=\bar{T}_{\bullet t}+\left[i-\frac{k+1}{2}\right].\frac{b}{k} \qquad \textit{tendencia media anual para los subperíodos k-ésimos}$$

donde,

$t\ \equiv$ Año (2006, 2007, ..., 2010)
$i\ \equiv$ Subperíodo donde se calcula la tendencia (trimestral i = 1, 2, 3, 4)
$k\ \equiv$ Número total de subperíodos (datos trimestrales k = 4)
$b\ \equiv$ 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 : $T_{i2006}=2{,}25+\left[1-\frac{4+1}{2}\right].\frac{1}{4}=1{,}875$

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

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

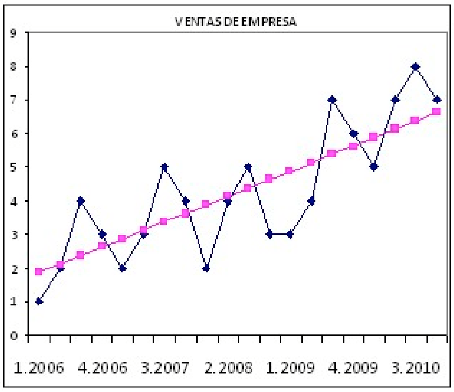
$$\text{Trimestre Primero 2007 : } T_{i2007} = 3{,}25 + \left[1 - \frac{4+1}{2} \right] . \frac{1}{4} = 2{,}875$$

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

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

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

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



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

SE ELIMINA LA TENDENCIA Y LA COMPONENTE CÍCLICA DE LA SERIE

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

Señalar que, en el esquema multiplicativo, al aplicar el método de los mínimos cuadrados, lo que se obtiene es una aproximación de , 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.

Serie con las COMPONENTES ESTACIONAL y ACCIDENTAL

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

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

$$\frac{0{,}533+0{,}696+0{,}516+0{,}615+0{,}851}{5}=0{,}642 \qquad \frac{0{,}941+0{,}96+0{,}97+0{,}78+1{,}143}{5}=0{,}959$$

$$\frac{1{,}684+1{,}481+1{,}143+1{,}302+1{,}255}{5}=1{,}373 \qquad \frac{1{,}143+1{,}103+0{,}649+1{,}067+1{,}057}{5}=1{,}004$$

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

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

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

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

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

1º Trimestre: (64,59 - 100 = -35,41) $\rightarrow$ un descenso de ventas del 35,41%
2º Trimestre: (96,48 - 100 = -3,52) $\rightarrow$ un descenso de ventas del 3,42%
3º Trimestre: (138,13 - 100 = 38,13) $\rightarrow$ un aumento de ventas del 38,13%
4º Trimestre: (101,01 - 100 = 1,01) $\rightarrow$ un aumento de ventas del 1,01%

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

EXAMEN DE ESTADÍSTICA DESCRIPTIVA - GRADO ECONOMÍA Junio 2011

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

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

Calcule:

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

Solución:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nos encontramos ante un cambio de escala $\begin{cases} E(k.x) = k.E(x) = k.\bar{x} \\ Var(k.x) = k^2.Var(x) = k^2\, \sigma_x^2 \end{cases}$

$$C.V_{final} = \frac{\sigma_{final}}{\bar{x}_{final}} = \frac{k.\sigma_x}{k.\bar{x}} = 0{,}4730 \;\; (47{,}30\%)$$

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

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

La concentración es medio-baja.

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

$$y = 4x + 2$$

$$y = 2x + 10$$

Sabiendo además que la covarianza entre ambas variables, S_{xy}, es 60, se pide:

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

Solución:

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

$$\text{X/y:} \quad y = 4x + 2 \quad \mapsto \quad 4x = y - 2 \quad \mapsto \quad x = \frac{1}{4}y - \frac{1}{2}$$

b) Como las dos pendientes son positivas (2 y 1/4), la recta de regresión de Y/x tiene mayor pendiente en valor absoluto que la de X/y

c) El coeficiente de determinación lineal $R^2 = \beta_{Y/x} . \beta_{X/y} = 2.\frac{1}{4} = \frac{1}{2} = 0{,}5$

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

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

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

3. Una empresa que produce tres variedades de aceite, sabe que en 2007 el valor de la producción de las variedades A, B y C fueron 100, 120 y 60 unidades monetarias, respectivamente. Por otro lado, las cantidades producidas (en miles de hectolitros) de cada variedad en el período 2007-2010 fueron:

	q_A	q_B	q_C
2007	40	50	60
2008	30	100	130
2009	60	110	150
2010	65	120	190

Calcule:

a) Los números índices cuánticos de Laspeyres con base 2007, para los años 2008, 2009 y 2010.
b) Las variaciones relativas interanuales de la producción.
c) La tasa de variación media anual de la producción para el período 2007-2010.

Solución:

a) El índice cuántico de Laspeyres $L_{Q_{i0}^{it}} = \dfrac{\sum_{i=1}^{n} q_{it}\cdot p_{i0}}{\sum_{i=1}^{n} q_{i0}\cdot p_{i0}}.100$. Por otra parte, conocemos el valor añadido bruto para el 2007: $\sum_{i=1}^{3} q_{07}\cdot p_{07} = 100+120+60=280$.

Como nos dan las producciones anuales para cada variedad, solo nos falta conocer los precios de cada variedad en el 2007, tarea que resulta sencilla al saber el valor añadido:

$$V_{A_{07}} = p_{A_{07}}\cdot q_{A_{07}} \mapsto 100 = p_{A_{07}}.40 \Rightarrow p_{A_{07}} = 2,5$$
$$V_{B_{07}} = p_{B_{07}}\cdot q_{B_{07}} \mapsto 120 = p_{B_{07}}.50 \Rightarrow p_{B_{07}} = 2,4$$
$$V_{C_{07}} = p_{C_{07}}\cdot q_{C_{07}} \mapsto 60 = p_{C_{07}}.60 \Rightarrow p_{C_{07}} = 1$$

Por tanto,

	2007 q_{07}	2007 p_{07}	2008 q_{08}	2009 q_{09}	2010 q_{10}
A	40	2,5	30	60	65
B	50	2,4	100	110	120
C	60	1	130	150	190

$$\sum_{i=1}^{3} q_{07}\cdot p_{07} = 100+120+60=280$$

$$L_{Q_{07}^{08}} = \frac{\sum_{i=1}^{3} q_{08}\cdot p_{07}}{\sum_{i=1}^{3} q_{07}\cdot p_{07}}.100 = \frac{30.2,5+100.2,4+130.1}{280}.100 = 158,93$$

$$L_{Q_{07}^{09}} = \frac{\sum_{i=1}^{3} q_{09}\cdot p_{07}}{\sum_{i=1}^{3} q_{07}\cdot p_{07}}.100 = \frac{60.2,5+110.2,4+150.1}{280}.100 = 201,43$$

$$L_{Q_{07}^{10}} = \frac{\sum_{i=1}^{3} q_{10} \cdot p_{07}}{\sum_{i=1}^{3} q_{07} \cdot p_{07}} . 100 = \frac{65 . 2{,}5 + 120 . 2{,}4 + 190 . 1}{280} . 100 = 228{,}75$$

b) Las variaciones relativas anuales de la producción

Años	Índice Laspeyres	Tasa de variación
2007	100	
2008	158,93	0,5893
2009	201,43	0,2674
2010	228,75	0,1356

La tasa de variación de la producción (tanto por uno) en 2007-2010, con el índice de Laspeyres como índice deflactor, viene dada por la relación:

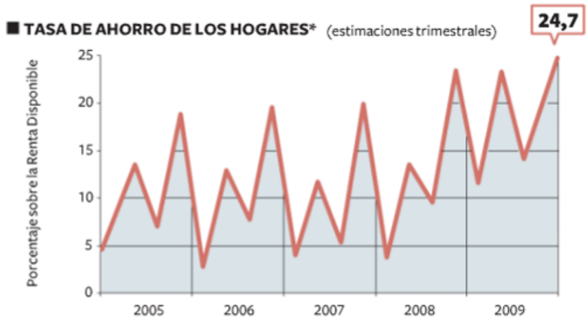
$$t_{Q07}^{08} = \frac{L_{Q_{07}^{08}}}{L_{Q_{07}^{07}}} - 1 = 1{,}5893 - 1 = 0{,}5893$$

$$t_{Q08}^{09} = \frac{L_{Q_{07}^{09}}}{L_{Q_{07}^{08}}} - 1 = \frac{2{,}0143}{1{,}5893} - 1 = 0{,}2674 \qquad t_{Q09}^{10} = \frac{L_{Q_{07}^{10}}}{L_{Q_{07}^{09}}} - 1 = \frac{2{,}2875}{2{,}0143} - 1 = 0{,}1356$$

c) La tasa media anual de variación de la producción para el período 2007-2010

$$(1 + t_{m07}^{10})^3 = L_{Q_{07}^{10}} \quad \mapsto \quad t_{m07}^{10} = \sqrt[3]{2{,}2875} - 1 = 0{,}3176$$

4. 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? Justifique su respuesta.

Opción 1: IVE_1= -5; IVE_2= 2,5; IVE_3= -2,5; IVE_4= ?

Opción 2: IVE_1= 5; IVE_2= -2,5; IVE_3= 2,5; IVE_4= ?

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:

$$y = 10 + 0{,}5 \cdot t$$

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. Justifique su 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, puesto que hay Índices de Variación Estacional negativos.

c) Siendo la componente ciclo-tendencia durante los años 2005 a 2009: $y = 10 + 0{,}5 \cdot t$

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.