

Ejemplos Resueltos Tema 2

2012

1. Ejemplos de Rectas de Regresión y Correlación lineal

DATOS: Tasas de Paro en las Comunidades Autónomas y Ceuta y Melilla para hombres y mujeres en 2005 y 2011, distinguiendo por nacionalidad (español, extranjero).

La tabla de datos siguiente muestra las tasas de paro en las Comunidades Autónomas en el año 2005 y 2011 para las mujeres españolas y extranjeras.

ccaa	Tesp05	Tesp11	Text05	Text11
01 = Andalucía	19.88	31.17	14.41	37.83
02 = Aragón	7.97	14.01	14.26	33.92
03 = Asturias	13.46	16.34	10.79	24.88
04 = Baleares	8.13	16.56	15.85	28.31
05 = Canarias	14.32	27.81	17.37	33.58
06 = Cantabria	12.23	12.58	12.12	25.89
07 = Castilla-León	13.49	17.21	18.75	27.49
08 = Castilla-La Mancha	14.40	23.76	22.46	36.13
09 = Cataluña	7.32	15.02	16.12	29.87
10 = Comunidad Valenciana	11.77	22.07	14.31	33.09
11 = Extremadura	23.02	26.64	26.81	35.38
12 = Galicia	13.27	16.59	25.92	34.61
13 = Madrid	7.51	15.02	9.06	20.18
14 = Murcia	11.44	22.68	14.56	33.24
15 = Navarra	6.85	10.05	13.80	26.55
16 = País Vasco	9.77	11.48	16.18	21.92
17 = Rioja (La)	7.90	12.01	19.08	36.71
51 = Ceuta	27.65	37.38	44.00	45.85
52 = Melilla	23.16	22.26	26.31	55.20

La tabla de datos siguiente muestra las tasas de paro en las Comunidades Autónomas en el año 2005 y 2011 para los hombres españoles y extranjeros.

ccaa	Tesp05	Tesp11	Text05	Text11
01 = Andalucía	10.30	27.26	10.15	39.72
02 = Aragón	3.44	12.32	8.70	38.85
03 = Asturias	8.25	15.93	16.60	45.88
04 = Baleares	5.12	18.67	5.11	29.26
05 = Canarias	10.05	28.79	10.66	31.85
06 = Cantabria	5.88	14.14	12.59	36.48
07 = Castilla-León	5.43	13.57	6.07	34.53
08 = Castilla-La Mancha	5.48	18.44	5.89	29.90
09 = Cataluña	5.08	15.54	11.18	34.92
10 = Comunidad Valenciana	6.64	19.87	8.35	38.88
11 = Extremadura	11.31	21.09	12.69	36.03
12 = Galicia	7.30	15.10	9.77	32.45
13 = Madrid	5.66	13.17	10.94	24.92
14 = Murcia	4.75	20.32	10.88	37.98
15 = Navarra	3.89	9.67	5.48	24.25
16 = País Vasco	5.56	10.18	11.23	24.49
17 = Rioja (La)	3.25	12.33	5.76	33.66
51 = Ceuta	14.96	21.36	27.78	47.75
52 = Melilla	8.38	20.24	18.90	22.07

Recta de regresión de tasa de paro femenina extranjera sobre española

Construiremos la recta de regresión de tasa extranjera sobre tasa española:

CCAA	ES	EX
01 = Andalucía	19.88	14.41
02 = Aragón	7.97	14.26
03 = Asturias	13.46	10.79
04 = Baleares	8.13	15.85
05 = Canarias	14.32	17.37
06 = Cantabria	12.23	12.12
07 = Castilla-León	13.49	18.75
08 = Castilla-La Mancha	14.40	22.46
09 = Cataluña	7.32	16.12
10 = Comunidad Valenciana	11.77	14.31
11 = Extremadura	23.02	26.81
12 = Galicia	13.27	25.92
13 = Madrid	7.51	9.06
14 = Murcia	11.44	14.56
15 = Navarra	6.85	13.80
16 = País Vasco	9.77	16.18
17 = Rioja (La)	7.90	19.08
51 = Ceuta	27.65	44.00
52 = Melilla	23.16	26.31

1.1. Recta de regresión de la tasa de paro EXTRANJERA sobre tasa ESPAÑOLA (2005)

Por comodidad en la notación renombramos las variables con Y=EX y X=ES

Variable Dependiente = EX

Variable Independiente = ES

Media de EX

$$\bar{Y} = \frac{\sum y_i n_i}{N}$$

Media de ES

$$\bar{X} = \frac{\sum x_i n_i}{N}$$

Covarianza de EX y ES

$$Cov(EX, ES) = Cov(X, Y) = \frac{\sum y_i x_i n_i}{N} - \bar{X} * \bar{Y}$$

Varianza de ES

$$Var(ES) = Var(X) = \frac{\sum x_i^2 n_i}{N} - \bar{X}^2$$

Pendiente de regresión

$$b = \frac{Cov(X, Y)}{Var(X)}$$

Ordenada en el origen

$$a = \bar{Y} - b * \bar{X}$$

Recta de regresión de Y sobre X

$$Y = a + b * X$$

Nota: Observe que en este ejemplo las frecuencias n_i son todas iguales a 1.

Para determinar la recta de regresión necesitamos las siguientes columnas de cálculo:

CCAA	ES	EX	ESxEX	ES2
01 = Andalucía	19.88	14.41	286.471	395.214
02 = Aragón	7.97	14.26	113.652	63.521
03 = Asturias	13.46	10.79	145.233	181.172
04 = Baleares	8.13	15.85	128.861	66.097
05 = Canarias	14.32	17.37	248.738	205.062
06 = Cantabria	12.23	12.12	148.228	149.573
07 = Castilla-León	13.49	18.75	252.938	181.980
08 = Castilla-La Mancha	14.40	22.46	323.424	207.360
09 = Cataluña	7.32	16.12	117.998	53.582
10 = Comunidad Valenciana	11.77	14.31	168.429	138.533
11 = Extremadura	23.02	26.81	617.166	529.920
12 = Galicia	13.27	25.92	343.958	176.093
13 = Madrid	7.51	9.06	68.041	56.400
14 = Murcia	11.44	14.56	166.566	130.874
15 = Navarra	6.85	13.80	94.530	46.922
16 = País Vasco	9.77	16.18	158.079	95.453
17 = Rioja (La)	7.90	19.08	150.732	62.410
51 = Ceuta	27.65	44.00	1216.600	764.522
52 = Melilla	23.16	26.31	609.340	536.386
	253.54	352.16	5358.984	4041.074

Coefficientes de la recta de regresión:

(Intercept)	0.116
ES	1.384

Recta de regresión de tasa paro de extranjera (EX) sobre tasa de paro española (ES)

$$Y = 0,116 + 1,384X$$

Recta de regresión de la tasa de paro ESPAÑOLA sobre tasa EXTRANJERA (2005)

Variable Dependiente = ES

Variable Independiente = EX

Media de EX

$$\bar{Y} = \frac{\sum y_i n_i}{N}$$

Media de ES

$$\bar{X} = \frac{\sum x_i n_i}{N}$$

Covarianza de EX y ES

$$Cov(EX, ES) = Cov(X, Y) = \frac{\sum y_i x_i n_i}{N} - \bar{X} * \bar{Y}$$

Varianza de ES

$$Var(ES) = Var(Y) = \frac{\sum y_i^2 n_i}{N} - \bar{Y}^2$$

Pendiente de regresión

$$B = \frac{Cov(X, Y)}{Var(Y)}$$

Ordenada en el origen

$$A = \bar{X} - B * \bar{Y}$$

Recta de regresión de X sobre Y

$$X = A + B * Y$$

Para determinar la recta de regresión necesitamos las siguientes columnas de cálculo: (Observe que son las mismas usadas en la recta de regresión de EX sobre ES, salvo la que usamos para calcular la varianza, dado que se han cambiado los papeles de las variables. Ahora la variable independiente es EX):

CCAA	ES	EX	ESxEX	EX2
01 = Andalucía	19.88	14.41	286.471	207.648
02 = Aragón	7.97	14.26	113.652	203.348
03 = Asturias	13.46	10.79	145.233	116.424
04 = Baleares	8.13	15.85	128.861	251.222
05 = Canarias	14.32	17.37	248.738	301.717
06 = Cantabria	12.23	12.12	148.228	146.894
07 = Castilla-León	13.49	18.75	252.938	351.562
08 = Castilla-La Mancha	14.40	22.46	323.424	504.452
09 = Cataluña	7.32	16.12	117.998	259.854
10 = Comunidad Valenciana	11.77	14.31	168.429	204.776
11 = Extremadura	23.02	26.81	617.166	718.776
12 = Galicia	13.27	25.92	343.958	671.846
13 = Madrid	7.51	9.06	68.041	82.084
14 = Murcia	11.44	14.56	166.566	211.994
15 = Navarra	6.85	13.80	94.530	190.440
16 = País Vasco	9.77	16.18	158.079	261.792
17 = Rioja (La)	7.90	19.08	150.732	364.046
51 = Ceuta	27.65	44.00	1216.600	1936.000
52 = Melilla	23.16	26.31	609.340	692.216
	253.54	352.16	5358.984	7677.091

Coeficientes de la recta de regresión:

(Intercept)	0.055
EX	0.718

Recta de regresión de tasa paro de española (ES) sobre tasa de paro extranjera (EX)

$$X=0.055+0.718Y$$

Coeficiente de correlación lineal: tasa de paro EXTRANJERA y tasa ESPAÑOLA (2005)

Para determinar el coeficiente de correlación lineal necesitamos las siguientes columnas de cálculo (observe que todas ellas ya se han determinado en pasos anteriores):

CCAA	ES	EX	ESxEX	ES2	EX2
01 = Andalucía	19.88	14.41	286.471	395.214	207.648
02 = Aragón	7.97	14.26	113.652	63.521	203.348
03 = Asturias	13.46	10.79	145.233	181.172	116.424
04 = Baleares	8.13	15.85	128.861	66.097	251.222
05 = Canarias	14.32	17.37	248.738	205.062	301.717
06 = Cantabria	12.23	12.12	148.228	149.573	146.894
07 = Castilla-León	13.49	18.75	252.938	181.980	351.562
08 = Castilla-La Mancha	14.40	22.46	323.424	207.360	504.452
09 = Cataluña	7.32	16.12	117.998	53.582	259.854
10 = Comunidad Valenciana	11.77	14.31	168.429	138.533	204.776
11 = Extremadura	23.02	26.81	617.166	529.920	718.776
12 = Galicia	13.27	25.92	343.958	176.093	671.846
13 = Madrid	7.51	9.06	68.041	56.400	82.084
14 = Murcia	11.44	14.56	166.566	130.874	211.994
15 = Navarra	6.85	13.80	94.530	46.922	190.440
16 = País Vasco	9.77	16.18	158.079	95.453	261.792
17 = Rioja (La)	7.90	19.08	150.732	62.410	364.046
51 = Ceuta	27.65	44.00	1216.600	764.522	1936.000
52 = Melilla	23.16	26.31	609.340	536.386	692.216
	253.54	352.16	5358.984	4041.074	7677.091

El coeficiente de correlación lineal se obtiene como cociente entre la covarianza y el producto de las desviaciones típicas de las variables

$$Cov(X, Y) = \frac{\sum x_i y_i n_i}{N} - \bar{X} * \bar{Y}$$

$$r = \frac{Cov(X, Y)}{\sigma(X) * \sigma(Y)}$$

$$r = 0,997$$

Rectas de Regresión y Correlación lineal para las tasas de paro masculinas extranjera y española en 2011

[datos](#)

Obtenga las 2 rectas de regresión y el coeficiente de correlación lineal para las tasas masculinas del año 2011.

Resultados:

Recta de EX sobre ES: $EX = 26.728 + 0.415 ES$

Recta de ES sobre EX: $ES = 9.421 + 0.231 EX$

Coeficiente de correlación lineal de Pearson

CCAA	ES	EX
01 = Andalucía	27.26	39.72
02 = Aragón	12.32	38.85
03 = Asturias	15.93	45.88
04 = Baleares	18.67	29.26
05 = Canarias	28.79	31.85
06 = Cantabria	14.14	36.48
07 = Castilla-León	13.57	34.53
08 = Castilla-La Mancha	18.44	29.90
09 = Cataluña	15.54	34.92
10 = Comunidad Valenciana	19.87	38.88
11 = Extremadura	21.09	36.03
12 = Galicia	15.10	32.45
13 = Madrid	13.17	24.92
14 = Murcia	20.32	37.98
15 = Navarra	9.67	24.25
16 = País Vasco	10.18	24.49
17 = Rioja (La)	12.33	33.66
51 = Ceuta	21.36	47.75
52 = Melilla	20.24	22.07

Cuadro 1: Tasas de Paro Hombres según Nacionalidad en CCAA

$$r=0,309$$

NOTA: Utilice los datos para obtener otras rectas de regresión. Por ejemplo, tasas de paro de 2011 sobre 2005.