

Instituciones, crecimiento económico y su relación con el Estado de derecho.

ANEXO ESTADÍSTICO

El modelo de regresión lineal simple

El modelo de regresión y correlación lineal establece la relación y el grado de dependencia entre dos variables con una ecuación basada en una línea recta:

$$Y_c = \alpha + \beta X$$

Donde Y es la variable dependiente, X es la variable independiente o explicativa, α es el intercepto al origen, y β es la pendiente de la recta.

Se debe estimar Y a partir de X mediante un análisis de *regresión* que describe la forma de la relación promedio descrita en la ecuación (en este caso lineal) entre las variables, y un análisis de *correlación* para medir el grado de dependencia o correlación entre las variables.

Hay que hacer la salvedad de que un alto grado de correlación no indica necesariamente relación *causa-efecto* entre dos variables. Se puede obtener una alta correlación entre dos variables aun cuando la relación no tenga un significado real. El alto grado de correlación indica solamente un resultado matemático.

Las conclusiones derivadas del modelo estarán basadas en razonamiento lógico e investigación inteligente sobre materias relacionadas significativamente (Shao, 1967).

El mejor modo de resolver la ecuación lineal es usar el método de mínimos cuadrados ordinarios (MCO). El MCO es una técnica para ajustar la línea recta “óptima” a la muestra de las observaciones $X Y$ (Salvatore, 1983).

Así mismo, garantiza que la sumatoria de las distancias verticales entre cada uno de los puntos de dispersión y la recta de regresión es la mínima posible, es decir, resume la relación entre dos variables cuantitativas X y Y , o minimiza la suma de las diferencias entre el valor observado y el valor estimado o error:

$$\min \sum (Y - Y_c)^2$$

La ecuación derivada de este método tendrá una línea recta de regresión con dos propiedades matemáticas:

1. La suma algebraica de las desviaciones de los valores individuales (Y) con respecto a la media (\bar{Y}) es cero: $\sum(Y - \bar{Y}) = 0$
2. La suma del cuadrado de las desviaciones de los valores individuales con respecto a la media es mínima: $\sum(Y - \bar{Y})^2 = \text{un mínimo}$

Las dos *ecuaciones normales* de la recta son:

$$\sum Y = n \alpha + \beta \sum X$$

$$\sum (XY) = \alpha \sum X + \beta \sum X^2$$

Donde n es el número de parejas de valores de X y Y (o puntos sobre la gráfica de dispersión). Resolviendo las dos ecuaciones, obtenemos las constantes α y β , también llamadas *coeficientes de regresión*:

$$\alpha = \frac{\sum Y}{n} - \beta \frac{\sum X}{n} = \bar{Y} - \beta \bar{X}$$

$$\beta = \frac{n \sum XY - \sum X \cdot \sum Y}{n \sum X^2 - (\sum X)^2}$$

Análisis de regresión

El análisis de regresión se conforma de dos operaciones principales: i) calcular la ecuación de regresión y derivar una línea (puede ser lineal o curvilínea) que represente dicha ecuación y que describa la relación funcional entre las variables, y ii) estimar una *variable dependiente* (representada por Y) a partir de otra *variable independiente* o variables (representada por X), basados en la relación descrita por la ecuación de regresión.

El término regresión fue utilizado originalmente por Francis Galton en un trabajo publicado en 1885, donde analizó la estatura media de los dos padres en una familia y la estatura media de sus hijos adultos, y construyó un diagrama para los datos bivariantes donde cada punto representa las estaturas de una familia.

Su análisis arrojó, en general, que padres altos tienden a tener hijos altos, asimismo padres de baja estatura tienden a tener hijos bajos, como era de esperarse; pero lo más interesante que encontró es que las estaturas de los hijos se desviaban menos de la estatura media del conjunto de hijos, que las estaturas de los padres de la estatura media del conjunto de padres, es decir, que en promedio, padres altos tienen hijos altos, pero no tan altos como sus padres; y por el otro lado, padres bajos tienen hijos bajos, pero no tan bajos como sus padres. Los padres altos y bajos tienen hijos más medianos que ellos mismos.

Se dice que es una regresión porque las estaturas de los hijos tienden a *regresar* hacia la estatura promedio de la población, y Galton llamó a la línea que describe esta relación promedio como *línea de regresión* (Shao, 1967).

A continuación, se presentan los siguientes datos de corte transversal: el PIB per cápita (Y) y el índice de Estado de Derecho (X) para 90 países en el 2020. Datos del Banco Mundial (<https://data.worldbank.org/indicador/NY.GDP.PCAP.CD>) y World Justice Project (<https://www.worldjusticeproject.org/rule-of-law-index/>)

Datos¹:

n	País	PIB per cápita (miles de dólares) Y	Índice de Estado de Derecho X	XY	X ²	Y _c = -47.8 + 108.9 X	(Y-Y _c)	(Y-Y _c) ²	(Y- \bar{Y}) ²	(Y _c - \bar{Y}) ²
1	Dinamarca	61	0.9	54.90	0.81	50.21	10.79	116.42	2110.59	1235.60
2	Noruega	67.2	0.89	59.81	0.79	49.12	18.08	326.85	2718.70	1160.23
3	Finlandia	49	0.87	42.63	0.76	46.94	2.06	4.23	1152.00	1016.60
4	Suecia	52	0.86	44.72	0.74	45.85	6.15	37.77	1364.65	948.34
5	Holanda	52.3	0.84	43.93	0.71	43.68	8.62	74.37	1386.90	818.94
6	Alemania	45.7	0.84	38.39	0.71	43.68	2.02	4.10	938.88	818.94
7	Nueva Zelanda	41.8	0.83	34.69	0.69	42.59	-0.79	0.62	715.09	757.80
8	Austria	48.1	0.82	39.44	0.67	41.50	6.60	43.59	1091.72	699.03
9	Canadá	43.2	0.81	34.99	0.66	40.41	2.79	7.79	791.92	642.63
10	Estonia	23.3	0.81	18.87	0.66	40.41	-17.11	292.72	67.92	642.63
11	Australia	51.8	0.8	41.44	0.64	39.32	12.48	155.75	1349.91	588.60
12	Singapur	59.8	0.79	47.24	0.62	38.23	21.57	465.22	2001.77	536.95
13	UK	40.3	0.79	31.84	0.62	38.23	2.07	4.28	637.11	536.95
14	Bélgica	44.6	0.79	35.23	0.62	38.23	6.37	40.56	872.68	536.95
15	Japón*	40.1	0.78	31.28	0.61	37.14	2.96	8.75	627.06	487.66
16	Sur Corea	31.5	0.73	23.00	0.53	31.70	-0.20	0.04	270.31	276.83
17	Chequia	22.8	0.73	16.64	0.53	31.70	-8.90	79.16	59.92	276.83
18	España	27	0.73	19.71	0.53	31.70	-4.70	22.06	142.59	276.83
19	Francia	38.6	0.73	28.18	0.53	31.70	6.90	47.65	554.18	276.83
20	EE. UU.	63.5	0.72	45.72	0.52	30.61	32.89	1081.88	2346.54	241.77
21	Uruguay	15.4	0.71	10.93	0.50	29.52	-14.12	199.35	0.12	209.09
22	Portugal	22.4	0.7	15.68	0.49	28.43	-6.03	36.36	53.89	178.79
23	Eslovenia	25.1	0.69	17.32	0.48	27.34	-2.24	5.02	100.82	150.85
24	Costa Rica	12.1	0.68	8.23	0.46	26.25	-14.15	200.28	8.76	125.29
25	Chile	13.2	0.67	8.84	0.45	25.16	-11.96	143.11	3.46	102.09
26	Italia	31.7	0.66	20.92	0.44	24.07	7.63	58.16	276.93	81.27
27	Polonia	15.6	0.66	10.30	0.44	24.07	-8.47	71.81	0.29	81.27
28	Rumania	12.9	0.63	8.13	0.40	20.81	-7.91	62.52	4.66	33.04
29	Namibia	4.2	0.63	2.65	0.40	20.81	-16.61	275.79	117.92	33.04
30	Ruanda	0.8	0.62	0.50	0.38	19.72	-18.92	357.89	203.32	21.71
31	Croacia	13.8	0.61	8.42	0.37	18.63	-4.83	23.32	1.58	12.75
32	Grecia	17.7	0.61	10.80	0.37	18.63	-0.93	0.86	6.98	12.75
33	Georgia	4.3	0.6	2.58	0.36	17.54	-13.24	175.30	115.75	6.16

¹ Banco Mundial (<https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>) y World Justice Project (<https://www.worldjusticeproject.org/rule-of-law-index/>)

34	Botswana	6.7	0.6	4.02	0.36	17.54	-10.84	117.51	69.87	6.16
35	Suráfrica	5	0.59	2.95	0.35	16.45	-11.45	131.13	101.18	1.94
36	Malasia	10.4	0.58	6.03	0.34	15.36	-4.96	24.62	21.71	0.09
37	Argentina	8.4	0.58	4.87	0.34	15.36	-6.96	48.47	44.34	0.09
38	Jamaica	4.7	0.57	2.68	0.32	14.27	-9.57	91.64	107.31	0.62
39	Jordania	4.3	0.57	2.45	0.32	14.27	-9.97	99.46	115.75	0.62
40	Ghana	2.3	0.57	1.31	0.32	14.27	-11.97	143.35	162.79	0.62
41	Senegal	1.4	0.55	0.77	0.30	12.10	-10.70	114.38	186.57	8.78
42	Bulgaria	10	0.55	5.50	0.30	12.10	-2.10	4.39	25.59	8.78
43	Túnez	3.3	0.54	1.78	0.29	11.01	-7.71	59.38	138.27	16.43
44	Mongolia	4	0.53	2.12	0.28	9.92	-5.92	35.01	122.30	26.44
45	Indonesia	3.9	0.53	2.07	0.28	9.92	-6.02	36.20	124.52	26.44
46	Hungría	15.9	0.53	8.43	0.28	9.92	5.98	35.80	0.71	26.44
47	Kazajistán	9	0.52	4.68	0.27	8.83	0.17	0.03	36.71	38.82
48	Panamá	12.2	0.52	6.34	0.27	8.83	3.37	11.37	8.17	38.82
49	Sri Lanka	3.7	0.52	1.92	0.27	8.83	-5.13	26.30	129.02	38.82
50	Brasil	6.8	0.52	3.54	0.27	8.83	-2.03	4.11	68.21	38.82
51	Bielorrusia	6.4	0.51	3.26	0.26	7.74	-1.34	1.79	74.98	53.58
52	India	2	0.51	1.02	0.26	7.74	-5.74	32.94	170.53	53.58
53	Tailandia	7.2	0.51	3.67	0.26	7.74	-0.54	0.29	61.76	53.58
54	Ucrania	3.7	0.51	1.89	0.26	7.74	-4.04	16.31	129.02	53.58
55	Serbia	7.7	0.5	3.85	0.25	6.65	1.05	1.10	54.15	70.71
56	Colombia	5.3	0.5	2.65	0.25	6.65	-1.35	1.82	95.24	70.71
57	Albania	5.2	0.5	2.60	0.25	6.65	-1.45	2.10	97.20	70.71
58	Marruecos	3	0.5	1.50	0.25	6.65	-3.65	13.32	145.42	70.71
59	Perú	6.1	0.5	3.05	0.25	6.65	-0.55	0.30	80.26	70.71
60	Argelia	3.3	0.49	1.62	0.24	5.56	-2.26	5.11	138.27	90.21
61	El Salvador	3.8	0.49	1.86	0.24	5.56	-1.76	3.10	126.76	90.21
62	Vietnam	2.8	0.49	1.37	0.24	5.56	-2.76	7.62	150.28	90.21
63	Ecuador	5.6	0.49	2.74	0.24	5.56	0.04	0.00	89.47	90.21
64	China	10.5	0.48	5.04	0.23	4.47	6.03	36.34	20.78	112.08
65	Dominicana	7.3	0.48	3.50	0.23	4.47	2.83	8.00	60.20	112.08
66	Filipinas	3.2	0.47	1.50	0.22	3.38	-0.18	0.03	140.63	136.33
67	Uzbekistán	1.7	0.47	0.80	0.22	3.38	-1.68	2.83	178.46	136.33
68	Tanzania	1	0.47	0.47	0.22	3.38	-2.38	5.68	197.65	136.33
69	Rusia	10.1	0.47	4.75	0.22	3.38	6.72	45.12	24.59	136.33
70	Líbano	4.9	0.45	2.21	0.20	1.21	3.69	13.65	103.20	191.93
71	Zambia	1	0.45	0.45	0.20	1.21	-0.21	0.04	197.65	191.93
72	Liberia	0.6	0.45	0.27	0.20	1.21	-0.61	0.37	209.06	191.93
73	Guatemala	4.6	0.45	2.07	0.20	1.21	3.39	11.53	109.39	191.93

74	Kenia	1.9	0.45	0.86	0.20	1.21	0.69	0.48	173.16	191.93
75	México	8.3	0.44	3.65	0.19	0.12	8.18	66.98	45.68	223.29
76	Mali	0.8	0.44	0.35	0.19	0.12	0.68	0.47	203.32	223.29
77	Turquía	8.5	0.43	3.66	0.18	-0.97	9.47	89.74	43.02	257.02
78	Nigeria	2	0.43	0.86	0.18	-0.97	2.97	8.84	170.53	257.02
79	Irán	2.3	0.43	0.99	0.18	-0.97	3.27	10.71	162.79	257.02
80	Angola	1.9	0.43	0.82	0.18	-0.97	2.87	8.25	173.16	257.02
81	Etiopía	0.9	0.41	0.37	0.17	-3.15	4.05	16.41	200.47	331.60
82	Bangladesh	2	0.41	0.82	0.17	-3.15	5.15	26.53	170.53	331.60
83	Honduras	2.4	0.4	0.96	0.16	-4.24	6.64	44.09	160.25	372.45
84	Uganda	0.8	0.4	0.32	0.16	-4.24	5.04	25.40	203.32	372.45
85	Nicaragua	2	0.39	0.78	0.15	-5.33	7.33	53.71	170.53	415.67
86	Paquistán	1.1	0.39	0.43	0.15	-5.33	6.43	41.33	194.85	415.67
87	Bolivia	3.1	0.38	1.18	0.14	-6.42	9.52	90.59	143.02	461.26
88	Afganistán	0.5	0.36	0.18	0.13	-8.60	9.10	82.74	211.96	559.55
89	Camerún	1.5	0.36	0.54	0.13	-8.60	10.10	101.93	183.84	559.55
90	Egipto	3.5	0.36	1.26	0.13	-8.60	12.10	146.31	133.61	559.55
Σ		1355.3	51.95	989.57	31.89	1355.36	-0.06	6430.68	28956.9	22615.58

$$n = 90$$

$$\bar{Y} = \frac{\Sigma Y}{n} = 15.6$$

Con los datos anteriores podemos calcular los coeficientes de regresión sustituyendo en (2) y en (1):

$$\beta = \frac{90(989.57) - (51.95)(1,355.3)}{90(31.89) - (51.95)^2} = 108.9$$

$$\alpha = \frac{1,355.3}{90} - (108.9) \frac{51.95}{90} = -47.8$$

Sustituyendo los valores de los coeficientes α y β obtenemos nuestra ecuación de la regresión:

$$Y_c = -47.8 + 108.9 X$$

Puesto que el signo de β en la ecuación de regresión $Y_c = \alpha + \beta X$ es positivo, o $\beta = +108.9$, el valor de r , también la dirección de la correlación debe ser positivo.

La desviación estándar de los valores de Y de la línea de regresión (Y_c) es llamada error estándar de regresión, puesto que puede ser usada para medir el error de las estimaciones de los valores individuales de Y , basados en la línea de regresión.

Para calcular la desviación o *error estándar* de regresión de los valores de Y con respecto a Y_c se utiliza la siguiente fórmula:

$$S_{yx} = \sqrt{\frac{\sum(Y - Y_c)^2}{n}}$$

Sustituyendo los valores tenemos:

$$S_{yx} = \sqrt{\frac{6,430.68}{90}} = 8.45$$

El valor de S_{yx} nos indica el recorrido del error de las estimaciones de valores individuales de Y .

Si los valores de Y están normalmente distribuidos, 68% de los valores estarán dentro de la distancia de una vez la desviación estándar de regresión, o $1S_{yx}$, de la línea (hacia arriba o hacia abajo), aproximadamente 95% para dos desviaciones estándar, o $2S_{yx}$ con respecto a la línea, y aproximadamente 99% para tres desviaciones estándar, o $3S_{yx}$ (Shao, 1967).

Análisis de correlación

El análisis de correlación se refiere a las técnicas usadas para medir el grado de dependencia de la relación entre variables, basado en la ecuación de regresión.

Coefficiente de determinación (r^2) es una razón de la variación explicada a la variación total:

$$r^2 = \frac{\frac{\sum(Y_c - \bar{Y})^2}{n}}{\frac{\sum(Y - \bar{Y})^2}{n}} = \frac{\frac{22,615.58}{90}}{\frac{28,956.92}{90}} = \frac{251.28}{321.74} = 0.78$$

Por lo tanto, 78% de la variación de los valores de Y ha sido reducida o explicada por la línea de regresión.

Es decir, 78% de la variación en las cantidades del PIB per cápita (Y) está linealmente relacionada con la variación del índice de Estado de derecho (X).

Coefficiente de correlación (r):

$$r = \sqrt{r^2} = \sqrt{0.78} = 0.88$$

Covarianza	2.30
Coef. de determinación (r^2)	0.78
Coef. de correlación (r)	0.88