Regresión lineal simple

Isaac Cortés Olmos

Universidad Arturo Prat

May 23, 2025

Esquema

Análisis de regresión simple

Análisis de regresión simple

Método de mínimos cuadrados

$$
\begin{aligned}
b &= \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \\
&= \frac{134.6 - \frac{132(9.8)}{10}}{1788 - \frac{132^2}{10}} \\
&= 0.1149 \\
a &= \overline{y} - b\overline{x} \\
&= \frac{9.8}{10} - 0.1149\left(\frac{132}{10}\right) \\
&= -0.5367 \\
\widehat{y} &= -0.5367 + 0.1149x.
\end{aligned}
$$

Análisis de regresión simple

Ajuste de la recta estimada

Familia	Ingresos (x)	Ahorros (y)	$\widehat{y}$
1	11	0.5	0.7272
2	14	1.1	1.0719
3	12	0.9	0.8421
4	9	0.6	0.4974
5	13	1.2	0.9570
6	13	0.9	0.9570
7	15	1.5	1.1868
8	17	1.3	1.4166
9	15	1.1	1.1868
10	13	0.7	0.9570

Table 1: Ingreso y ahorro mensual de 10 familias (en miles de pesos).

Análisis de regresión simple

Ejemplo:

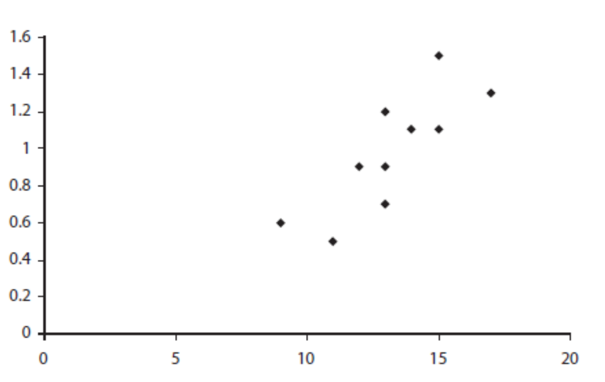

Figure 1: Diagrama de dispersión para los datos del ejemplo.

Análisis de regresión simple

Coeficiente de correlación

- Es una medida descriptiva de la intensidad de la relación lineal entre dos variables x e y.
- Los valores del coeficiente de correlación siempre estarán entre -1 a 1.
- Un valor de $+1$ indica que las dos variables x e y están perfectamente relacionadas en un sentido lineal positivo.
- Un valor de -1 indica que las dos variables x e y están perfectamente relacionadas en un sentido lineal negativo.
- Los valores del coeficiente de correlación cercanos a 0 indican que x e y no se relacionan linealmente.

Análisis de regresión simple

Coeficiente de correlación

- Se define el coeficiente correlación como:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

Ejemplo:

El gerente de un banco desea saber si puede considerarse que el ahorro de las familias (variable y) depende de sus ingresos (variable x). En la Tabla se muestran los resultados obtenidos para una muestra de 10 familias.

Ingresos (x)	Ahorro (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
11	0.5	-2.2	-0.48	1.056	4.84	0.2304
14	1.1	0.8	0.02	0.016	0.64	0.0004
12	0.9	-1.2	-0.08	0.096	1.44	0.0064
9	0.6	-4.2	-0.38	1.596	17.64	0.1444
13	1.2	-0.2	0.22	-0.044	0.04	0.0484
13	0.9	-0.2	-0.08	0.016	0.04	0.0064
15	1.5	1.8	0.52	0.936	3.24	0.2704
17	1.3	3.8	0.32	1.216	14.44	0.1024
15	1.1	1.8	0.12	0.216	3.24	0.0144
13	0.7	-0.2	-0.28	0.056	0.04	0.0784
132	9.8			5.16	45.06	0.902

Table 2: Ingreso y ahorro mensual de 10 familias (en miles de pesos).

Análisis de regresión simple

Coeficiente de correlación

$$
\begin{aligned}
r_{xy} &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \\
&= \frac{5.16}{\sqrt{45.06}\sqrt{0.902}} \\
&= 0.809
\end{aligned}
$$

Análisis de regresión simple

Coeficiente de determinación

- Toma valores entre 0 y 1.
- Se usa para evaluar la bondad de ajuste de la ecuación de regresión estimada.
- $$r^2 = r_{xy}^2 = 0.809^2 = 0.654$$
- El 65,4% de la variabilidad en los ahorros se explica por la relación lineal que existe entre éstas y los ingresos.

Referencias I

- Devore, J. L. (2009). *Probabilidad y estadística para ingeniería y ciencias*. Cengage Learning Editores.
- Díaz Mata, A. (2013). *Estadística aplicada a la Administración y Economía*. McGraw–Hill.
- Ross, S. M. (2002). *Probabilidad y estadística para ingeniería y ciencias*. McGraw–Hill.
- Spiegel, M. R., & Stephens, L. J. (2009). *Estadística*. McGraw–Hill.