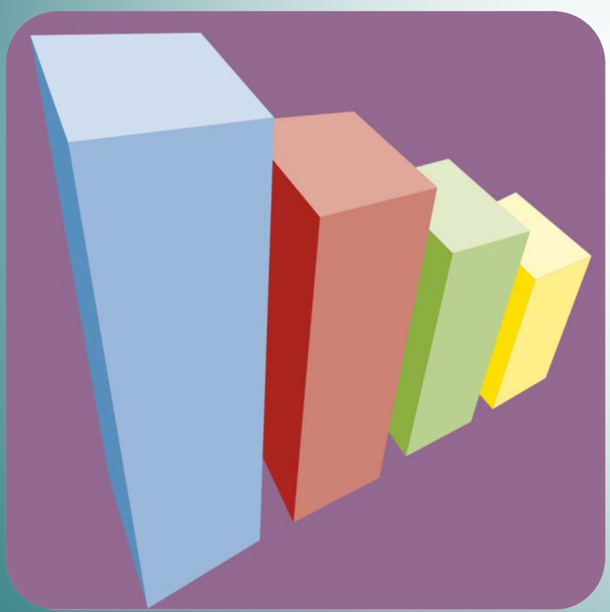
Estadística I

Tema 5. Teoría de la regresión



Carmen Trueba Salas

Lorena Remuzgo Pérez

Vanesa Jordá Gil

José María Sarabia Alegría

Dpto. de Economía

Introducción

La teoría de la regresión tiene por objeto definir la estructura de dependencia que mejor explique el comportamiento de una variable (dependiente o explicada) en función de otra (independiente o explicativa).

Dependencia funcional. Existe una función matemática que expresa los valores de Y en función de los valores de X de forma exacta.

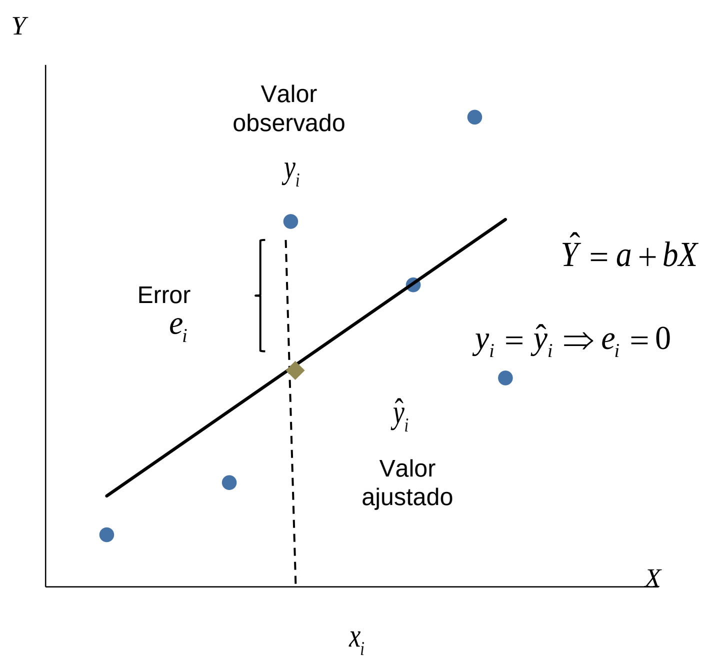
Dependencia estadística. La relación entre las variables X e Y es aproximada de modo que incluye un término aleatorio que no es posible modelizar.

$y = a + bx.$

Recta de regresión de Y sobre X

$$y = a + bx.$$

Ilustración gráfica



Recta de regresión de Y sobre X

Derivación analítica

$$a = \bar{y} - b\bar{x},$$

$$b = \frac{S_{XY}}{S_X^2}.$$

De la sustitución de estos valores en la recta $y = a + bx$., obtenemos la recta de regresión de Y sobre X:

$$y - \bar{y} = \frac{S_{XY}}{S_X^2}(x - \bar{x}),$$

Propiedades del ajuste

1. La media de los residuos es cero: $\bar{e} = 0.$
2. La media de los valores ajustados es igual a la media de los valores observados: $\bar{\hat{y}} = \bar{y}.$
3. La covarianza entre la variable dependiente y el error es cero: $S_{Ye} = 0.$
4. La covarianza entre la variable ajustada y el error es cero: $S_{\hat{Y}e} = 0.$
5. La varianza de la variable dependiente se descompone del siguiente modo

$$S_Y^2 = S_e^2 + S_{\hat{Y}}^2.$$

Recta de regresión de X sobre Y

Derivación analítica

$$a' = \bar{x} - b\bar{y},$$

$$b' = \frac{S_{XY}}{S_Y^2}.$$

Sustituyendo estos valores en la recta $x = a' + b'y$ se obtiene:

$$x - \bar{x} = \frac{S_{XY}}{S_Y^2}(y - \bar{y}),$$

Relaciones entre las rectas

- Las dos rectas de cortan en el punto $(\overline{x}, \overline{y})$.
- Las pendientes de las dos rectas (*b* y *b'*) tienen el mismo signo que a su vez es igual al signo del coeficiente de correlación lineal (r_{XY}).
- Se cumple la relación: $b = r\dfrac{S_Y}{S_X},\ b' = r\dfrac{S_X}{S_Y}.$
- El producto de los coeficientes de regresión es igual al coeficiente de determinación.
- En los casos en los que tenemos correlación perfecta ($r_{XY} = 1, r_{XY} = -1$), las rectas de regresión coinciden.
- $r_{XY} = 0,$ si $y = \overline{y}$ $x = \overline{x}.$

Medias de bondad de ajuste

$$R^2_{XY} = \frac{S^2_{XY}}{S^2_X \cdot S^2_Y}$$

- Si $R^2_{XY} = 1$, el ajuste es perfecto.
- Si R^2_{XY} está cercano a uno, la recta de regresión lineal se ajusta de forma adecuada a la nube de puntos.
- Si R^2_{XY} está cercano cero el ajuste es pobre.
- Si $R^2_{XY} = 0$, la recta de regresión no se ajusta en absoluto a la nube de puntos.

Ajustes no lineales

Modelo no lineal	*Transformación*	*Modelo lineal* $y' = b_0 + b_1 x'$
$y = ae^{bx}$	$y = \log y;\ x' = x$	$b_0 = \log a;\ b_1 = b$
$y = ax^b$	$y = \log y;\ x' = \log x$	$b_0 = \log a;\ b_1 = b$
$y = \dfrac{1}{a+bx}$	$y = \dfrac{1}{y};\ x' = x$	$b_0 = a;\ b_1 = b$

Predicción

La predicción de Y para $X = x_0$ es:

$$y^* = \overline{y} + \frac{S_{XY}}{S_X^2}(x_0 - \overline{x}).$$

La predicción de X para $Y = y_0$ viene dada por:

$$x^* = \overline{x} + \frac{S_{XY}}{S_Y^2}(y_0 - \overline{y}).$$