

REGRESIÓN LINEAL

Regresión lineal simple

Modelo

Se busca la ecuación de una recta de la forma:

$$y = mx + b$$

que minimice la suma de los errores elevados al cuadrado.

Medias

$$\bar{x} = \frac{\sum x_i}{N} \quad \bar{y} = \frac{\sum y_i}{N}$$

Sumas de cuadrados (fórmulas básicas)

$$S_{xx} = \sum (x_i - \bar{x})^2$$

$$S_{yy} = \sum (y_i - \bar{y})^2$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

Sumas de cuadrados (fórmulas simplificadas)

Estas fórmulas tienen la ventaja de que no requieren el cálculo previo de los promedios, y reducen los errores de redondeo al evitar restar valores muy similares.

$$S_{xx} = \sum x_i^2 - \frac{1}{N} \left(\sum x_i \right)^2$$

$$S_{yy} = \sum y_i^2 - \frac{1}{N} \left(\sum y_i \right)^2$$

$$S_{xy} = \sum x_i y_i - \frac{1}{N} \left(\sum x_i \right) \left(\sum y_i \right)$$

Pendiente e intersección

$$m = \frac{S_{xy}}{S_{xx}} \quad b = \bar{y} - m\bar{x}$$

Coefficiente de correlación

$$R = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Coefficiente de determinación

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

Error estándar

Error estándar alrededor de la línea de regresión:

$$s_r = \sqrt{\frac{S_{yy} - m^2 S_{xx}}{N - 2}}$$

Error estándar de la pendiente y la intersección:

$$s_m = \frac{s_r}{\sqrt{S_{xx}}} \quad s_b = s_r \sqrt{\frac{\sum x_i^2}{N S_{xx}}}$$

Error estándar del valor x_c correspondiente a la media \bar{y}_c de un conjunto de M réplicas de un análisis, obtenido empleando una curva de calibración con N puntos:

$$s_c = \frac{s_r}{m} \sqrt{\frac{1}{M} + \frac{1}{N} + \frac{(\bar{y}_c - \bar{y})^2}{m^2 S_{xx}}}$$