Linear Minimum Least Squares

Moreno Guerra José Alberto
División de Ciencias e Ingenierías,
Métodos Estadísticos en Cosmología de la Universidad de Guanajuato
Prof. Alma González and Gustavo Niz

August 24, 2015

1 La relación matemática determinística más simple entre dos variables x y y es una relación lineal $y = \beta_o + \beta_1 x$

Principio de los mínimos cuadrados

- La suma de las desviaciones verticales al cuadrado de los puntos $(x_1, y_1), ..., (x_n, y_n)$

  Es entonces:

$$f(b_o, b_1) = \sum_{i=1}^{n} [y_i - (b_o + b_1 x_i)]^2 \tag{1}$$

- Los valores minimizantes de b_o y b_1 se encuentran tomando las derivadas parciales de $f(b_o, b_1)$

  Donde:

$$f(b_o, b_1) = \sum_{i=1}^{n} y_i^2 - 2b_o \sum_{i=1}^{n} y_i - 2b_1 \sum_{i=1}^{n} y_i x_i + n b_o^2 + 2b_o b_1 \sum_{i=1}^{n} x_i + b_1^2 \sum_{i=1}^{n} x_i^2 \tag{2}$$

- y resolviendo las ecuaciones

$$\frac{\partial f(b_o, b_1)}{\partial b_o} = \sum_{i=1}^{n} -2(y_i - b_o - b_1 x_i) = 0 \tag{3}$$

$$\frac{\partial f(b_o, b_1)}{\partial b_1} = \sum_{i=1}^{n} -2x_i(y_i - b_o - b_1 x_i) = 0 \tag{4}$$

- La cancelación del factor -2 y reordenando se obtine el siguiente sistema de ecuaciones, llamado *ecuaciones normales*

$$n b_o + \left(\sum_{i=1}^{n} x_i\right) b_1 = \sum_{i=1}^{n} y_i \tag{5}$$

$$\left(\sum_{i=1}^{n} x_i\right) b_o + \left(\sum_{i=1}^{n} x_i^2\right) b_1 = \sum_{i=1}^{n} x_i y_i \tag{6}$$

- Siempre que por lo menos dos de las x_i sean diferentes, las estimaciones de los mínimos cuadrados son la única solución de este sistema.

- Tal que si resolvemos para b_1 y b_o:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} \tag{7}$$

$$b_o = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} \tag{8}$$

Las fórmulas de cómputo para S_{xy} y S_{xx} requieren sólo de los estadísticos resumidos $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} y_i$, $\sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} x_i y_i$ y $\sum_{i=1}^{n} y_i^2$

La *suma de los cuadrados del error* determina la cantidad de variabilidad inherente en el modelo de regresión, y la denotamos por:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} \tag{9}$$

el divisor $n - 2$ es el número de grados de libertad asociado con la estimación.