Modelos Generales Lineales:
REGRESIÓN LINEAL SIMPLE

Métodos para la estimación de los parámetros de la recta de regresión $y = a + b x$

Supóngase que disponemos de n individuos en los que medimos los valores de las variables X e Y, obteniendo los siguientes datos experimentales:

X	**Y**
X_1	Y_1
X_2	Y_2
X_3	Y_3
$\vdots$	$\vdots$
X_n	Y_n

La ecuación de la recta de regresión puede ser obtenida a partir de los datos experimentales de la tabla anterior por dos métodos: (a) la técnica de la ecuación de la recta punto pendiente o (b) el método de los mínimos cuadrados. En un análisis de estadística descriptiva bivariante, y desde un punto de vista formal, es indiferente el uso de (a) o (b). Sin embargo, en contexto experimental inferencial lo apropiado es el uso de (b). A continuación, resumimos cada una de estas técnicas:

- Técnica de la ecuación de la recta punto pendiente: la expresión de la ecuación de la recta de regresión puede ser fácilmente obtenida a partir de la ecuación de la recta punto pendiente para el caso en el que dicha recta pasa por los puntos (x, y), $(\overline{x}, \overline{y})$:

$$y - \overline{y} = \frac{S_{xy}}{S_{xx}^2}\left(x - \overline{x}\right)$$

siendo $\overline{x}$ e $\overline{y}$ las medias aritméticas de las variables X e Y, S_{xy} la covarianza de las variables X, Y, y S_{xx}^2 la varianza de la variable X. Obsérvese que el cociente $\frac{S_{xy}}{S_{xx}^2}$ es la pendiente de la recta, y por tanto el valor del parámetro *b*. Operando, y organizando los términos de la expresión obtendremos el valor del parámetro *a* u ordenada en el origen:

$$y = a + b\,x$$

- Método de los mínimos cuadrados: el fundamento de la técnica consiste en estimar los valores de los parámetros *a* y *b*, cumpliéndose la condición de que la recta obtenida, la recta de regresión, sea tal que la distancia de todos los puntos de la nube de regresión (datos experimentales) a dicha recta sea *mínima*.

  Es decir, sea y_i el valor experimental de la variable Y medido en el individuo i, y llamemos y_i^* al valor teórico de la variable Y en el individuo i según el modelo de la recta de regresión:

$$y_i = a + b\,x$$

  Deseamos que todos los puntos experimentales estén a una distancia mínima de la recta de regresión, por tanto deberemos minimizar la siguiente expresión que se conoce como suma de cuadrados del error (SCE):

$$\text{SCE} = \sum_{i=1}^{n}\left(y_i - y_i^*\right)^2$$

  ya que $y_i - y_i^*$ es precisamente el error o distancia entre el valor experimental y el valor teórico (o del modelo) de la variable Y. Para obtener los valores de a y b que hacen mínima la SCE deberemos hallar las siguientes derivadas:

$$\begin{cases} \dfrac{\partial \,\mathrm{SCE}}{\partial a} = -2\displaystyle\sum_{i=1}^{n}\left(\mathrm{y_i} - a - b\,\mathrm{x_i}\right) \\ \dfrac{\partial \,\mathrm{SCE}}{\partial b} = -2\displaystyle\sum_{i=1}^{n}\left(\mathrm{y_i} - a - b\,\mathrm{x_i}\right)\mathrm{x_i} \end{cases}$$

A continuación, y tras igualar las derivadas a cero, reorganizar y los términos se obtienen las llamadas ecuaciones normales. Estas ecuaciones forman un sistema de ecuaciones con dos incógnitas que una vez resuelto resultarán las expresiones o estimadores que permiten estimar los coeficientes *a* y *b*:

$$b = \frac{n\displaystyle\sum_{i=1}^{n}\mathrm{x_i.y_i} - \left(\sum_{i=1}^{n}\mathrm{x_i}\right)\left(\sum_{i=1}^{n}\mathrm{y_i}\right)}{n\displaystyle\sum_{i=1}^{n}\mathrm{x_i^2} - \left(\sum_{i=1}^{n}\mathrm{x_i}\right)^2}$$

$$a = \overline{\mathrm{y}} - b\,\overline{\mathrm{x}}$$

En el ámbito experimental es habitual encontrarse con explicaciones del método en las que se calculan los sumatorios de los estimadores a partir de la siguiente tabla experimental:

X	**Y**	**X^2**	**Y^2**	**X.Y**
X_1	Y_1	X_1^2	Y_1^2	$X_1.Y_1$
X_2	Y_2	X_2^2	Y_2^2	$X_2.Y_2$
X_3	Y_3	X_3^2	Y_3^2	$X_3.Y_3$
$\vdots$	$\vdots$			
X_n	Y_n	X_n^2	Y_n^2	$X_n.Y_n$
$\sum_{i=1}^{n}\mathrm{x_i}$	$\sum_{i=1}^{n}\mathrm{y_i}$	$\sum_{i=1}^{n}\mathrm{x_i^2}$	$\sum_{i=1}^{n}\mathrm{y_i^2}$	$\sum_{i=1}^{n}\mathrm{x_i.y_i}$

Rafael Lahoz-Beltrá, Pilar López González-Nieto, Mariángeles Gómez Flechoso, María Eugenia Arribas, Mocoroa, Alfonso Muñoz Martín, María de la Luz García Lorenzo, Gloria Cabrera Gómez, Jose Antonio Alvarez Gómez, Andrea Caso Fraile, Jefferson Mark Orosco Dagan, Raul Merinero Palomar. Universidad Complutense de Madrid, 2017.

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.