Modelos Generales Lineales:
REGRESIÓN LINEAL SIMPLE

Inferencias acerca de α y β

Los valores de *a* y *b* representan las estimaciones de los valores de los parámetros de la ecuación:

$$y = \alpha + \beta x$$

que podemos obtener aplicando el método de los mínimos cuadrados. De forma similar a como ocurre con la media, varianza y proporción poblacionales en inferencia estadística, los verdaderos valores de los parámetros α y β son desconocidos pudiéndose realizar contrastes de hipótesis acerca los valores de tales parámetros.

En general, dos son los contrastes y estadísticos que se aplican en regresión lineal:

- Contrastes sobre α (origen en ordenadas): supóngase que evaluamos el siguiente contraste con el fin de averiguar si el origen en ordenadas difiere o no significativamente de cero:

  $$\text{H0: } \alpha = 0$$
  $$\text{Ha: } \alpha \neq 0$$

  El estadístico que permite realizar contrastes de hipótesis o cualquier otra prueba inferencial acerca de α siguen una distribución *t*-student con n-2 grados de libertad:

  $$t = \frac{a - \alpha}{\sqrt{\frac{SCE}{n-2}} \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - n^2 \bar{x}^2}}}$$

- Contrates sobre β (pendiente): evaluemos el contraste que se muestra a continuación. En este caso si aceptamos la H0 la recta de regresión será paralela al eje-X y por tanto no hay relación lineal entre X e Y ya que la pendiente es cero. Es decir, el modelo de regresión lineal es válido si en el contraste rechazamos la H0, aceptando Ha (pendiente distinta de cero):

$$\text{H0: } \beta = 0$$
$$\text{Ha: } \beta \neq 0$$

siendo el estadístico:

$$t = \frac{b - \beta}{\sqrt{\frac{SCE}{n-2}}\sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}}$$

con distribución *t*-student y n-2 grados de libertad.

- ANOVA de regresión: se trata de un análisis de varianza orientado a la regresión lineal simple. El ANOVA evalúa la calidad del ajuste descomponiendo la variabilidad de la variable Y en dos componentes o fuentes de variación: por un lado el porcentaje de variabilidad de Y que es explicado por su relación de dependencia o *regresión* con X; y de otro el porcentaje de variabilidad de Y que es explicada por el error, es decir la variabilidad de la Y en torno a la recta de la regresión. Estas dos fuentes de variabilidad son reflejadas en la *media de cuadrados de regresión* (MCR) y *media de cuadrados del error* (MCE), respectivamente:

$$\text{MCR} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = b^2 - \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

que estaría dividido entre 1 grado de libertad, siendo MCE la expresión que resulta de la diferencia entra la suma de cuadrados total (SCTotal) y la SCR, es decir:

$$SCE = SCTotal - SCR$$

siendo SCTotal:

$$SCTotal = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2$$

Esto es, MCE será la expresión que resulte de dividir SCE entre n-2 grados de libertad:

$$MCE = \frac{SCE}{n-2}$$

En el ANOVA el contraste es similar al que se muestra más arriba para la pendiente, es decir:

H0: $\beta = 0$ (no linealidad)
Ha: $\beta \neq 0$ (linealidad)

evaluándose con el correspondiente F-test:

$$F = \frac{MCR}{MCE}$$

con distribución F y 1, n-2 grados de libertad. Es decir, para que el modelo de regresión sea válido p-valor deberá ser menor que el nivel de significación α, por ejemplo 0.05.

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.