Estadística II - 3006915
Regresión Lineal Simple

Mateo Ochoa Medina

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín

Periodo académico 2023-2S

Contenido

Contenido

Transformaciones: Modelos intrínsecamente lineales

El término "lineal" en modelos de regresión no se refiere a que la relación de tendencia entre Y y X tiene que ser una recta. Un modelo de regresión se considera lineal cuando lo es en los parámetros, es decir, ningún parámetro de la regresión aparece como el exponente o es dividido o multiplicado por otro parámetro. La relación estadística puede tener una curvatura (no ser lineal en X y/o en Y), y sin embargo pudiera ser posible que mediante una transformación conveniente de las variables (X y/o Y) resulten aplicables las técnicas de regresión lineal sobre estas nuevas variables. Por ello, las transformaciones no implican necesariamente modelos no lineales. Denominamos modelos intrínsecamente lineales a aquellos que relacionan Y con X por medio de una transformación en Y y/o en X, originando el modelo,

$$Y^* = \beta_0 + \beta_1 X^* + \varepsilon^*, \tag{1}$$

con $Y^* = h(Y)$ y $X^* = g(X)$, las variables transformadas mediante ciertas funciones $h(\cdot)$ y $g(\cdot)$, respectivamente, que conducen a la ecuación (1).

Modelo exponencial multiplicativo

La ecuación está dada por:

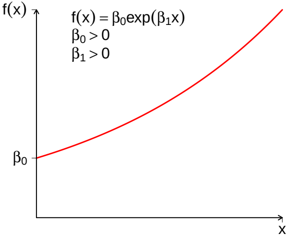
$$Y_i = \beta_0 \exp(\beta_1 X_i) \cdot \varepsilon_i,\ \beta_0 > 0,\ \text{con } \varepsilon_i \overset{iid}{\sim} \text{lognormal}(\mu = 0, \sigma). \qquad (2)$$

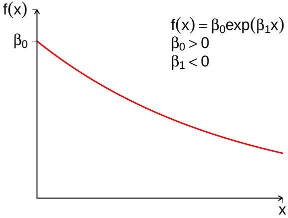
La transformación es

$$Y_i^* = \beta_0^* + \beta_1 X_i + \varepsilon_i^*,\ \varepsilon_i^* \overset{iid}{\sim} N(0, \sigma^2), \qquad (3)$$

con $Y_i^* = \log(Y_i)$, $\beta_0^* = \log(\beta_0)$ y, $\varepsilon_i^* = \log(\varepsilon_i)$.

Otra parametrización del modelo (2) es la siguiente: $Y_i = \exp(\beta_0 + \beta_1 X_i + \varepsilon_i)$, con $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, de modo que $Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ y $Y_i^* = \log(Y_i)$.





Modelo de potencia multiplicativo

La ecuación está dada por:

$$Y_i = \beta_0 X_i^{\beta_1} \cdot \varepsilon_i, \beta_0 > 0, X_i > 0, \text{con } \varepsilon_i \overset{iid}{\sim} \text{lognormal}(\mu = 0, \sigma). \tag{4}$$

La transformación es

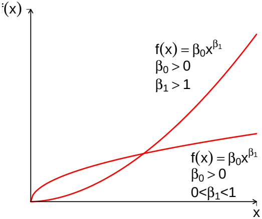
$$Y_i^* = \beta_0^* + \beta_1 X_i^* + \varepsilon_i^*, \varepsilon_i^* \overset{iid}{\sim} N(0, \sigma^2), \tag{5}$$

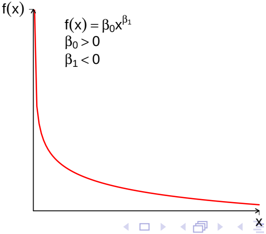
con $Y_i^* = \log(Y_i)$, $X_i^* = \log(X_i)$, $\beta_0^* = \log(\beta_0)$ y, $\varepsilon_i^* = \log(\varepsilon_i)$.

Otra parametrización del modelo (4) es la siguiente:

$Y_i = \exp(\beta_0 + \beta_1 \log(X_i) + \varepsilon_i)$, con $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, de modo que

$Y_i^* = \beta_0 + \beta_1 X_i^* + \varepsilon_i$, con $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $Y_i^* = \log(Y_i)$ y $X_i^* = \log(X_i)$.





Modelo logarítmico

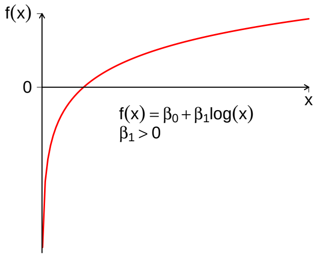
La ecuación está dada por:

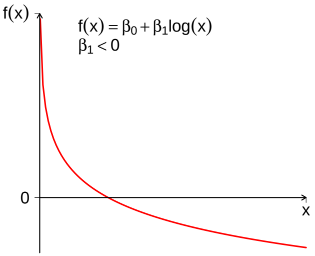
$$Y_i = \beta_0 + \beta_1 \log(X_i) + \varepsilon_i, \; X_i > 0, \text{ con } \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right). \tag{6}$$

La transformación es

$$Y_i = \beta_0 + \beta_1 X_i^* + \varepsilon_i, \; \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right), \tag{7}$$

con $X_i^* = \log(X_i)$.





Modelo recíproco

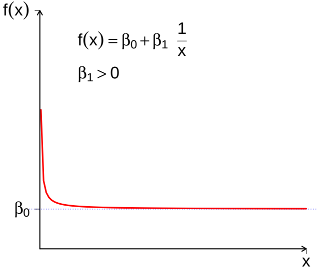
La ecuación está dada por:

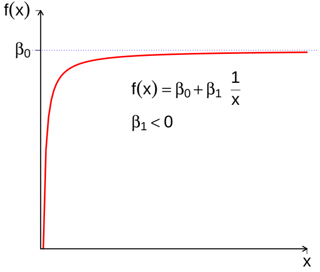
$$Y_i = \beta_0 + \beta_1 (1/X_i) + \varepsilon_i, \text{ con } \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2). \tag{8}$$

La transformación es

$$Y_i = \beta_0 + \beta_1 X_i^* + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2), \tag{9}$$

con $X_i^* = (1/X_i)$.





Consideraciones

Notas:

- No son intrínsecamente lineales los modelos exponenciales y de potencia aditivos:

$$Y_i = \beta_0 \exp\left(\beta_1 X_i\right) + \varepsilon_i,\, \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right),$$
$$Y_i = \beta_0 X_i^{\beta_1} + \varepsilon_i,\, \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right).$$

- El supuesto necesario cuando el término de error ε_i es transformado, es que esta variable transformada deberá distribuirse $N\left(0,\, \sigma^2\right)$, por ello deben examinarse los residuales del ajuste del modelo en la escala transformada.
- Los parámetros del modelo original no lineal se pueden estimar al destransformar (cuando resulte necesario) los estimadores hallados para los parámetros del modelo transformado y si el modelo lineal transformado satisface todas las suposiciones para la regresión lineal simple, las estimaciones de los parámetros originales a través de transformaciones inversas resultan razonables aunque no insesgadas.

Consideraciones

Notas:

- En los casos con modelos exponenciales y de potencia multiplicativos, si σ es pequeño, se puede obtener un intervalo de confianza aproximado para $\mu_{Y\,|\,x}$ tomando antilogaritmos sobre los límites del intervalo hallado para la media de la respuesta transformada, $\mu_{Y^*\,|\,x}$. Sin embargo cuando hacemos esto, en términos generales, estamos hallando un intervalo de confianza para la mediana de Y, pues se tiene que $Y\,|x \sim \text{lognormal}\left(\mu_{Y^*\,|\,x},\, \sigma\right)$ si $Y^*\,|x \sim N\left(\mu_{Y^*\,|\,x},\, \sigma^2\right)$, con $Y^* = \log(Y)$, y en consecuencia, $\exp\left(\mu_{Y^*\,|\,x}\right)$ es la mediana de $Y\,|\,x$. Podemos mejorar la aproximación de los valores ajustados de la respuesta en la escala original, teniendo en cuenta que en los modelos exponencial y de potencia multiplicativo bajo el supuesto de $\varepsilon_i \overset{iid}{\sim} \text{lognormal}\,(\mu = 0,\, \sigma)$, se cumple que,

$$Y_i = \exp\left[E\left(Y_i^*\right)\right] \times \varepsilon_i \quad \Rightarrow \quad \mu_{Y\,|\,x_i} = \exp\left(\mu_{Y^*\,|\,x_i}\right) \times \exp\left(\sigma^2/2\right),$$

  por tanto, podemos aproximar la respuesta ajustada en la escala original como,

$$\hat{Y}_i = \hat{\mu}_{Y\,|\,x_i} \approx \exp\left(\hat{Y}_i^*\right) \times \exp\left(MSE^*/2\right),$$

  donde MSE^* es la suma de cuadrados medios de los residuos del ajuste por mínimos cuadrados en el modelo linealizado. $\exp\left(MSE^*/2\right)$ es conocido como factor de corrección por transformación lognormal.

Contenido

Referencias

Álvarez, N. G. (2022). Notas de Clase Análisis de Regresión - 3006918, Capítulo 2: Regresión Lineal Simple. Notas no publicadas.

Álvarez, N. G. y Gómez, C. M. L. (2018). Notas de Clase - Estadística II (3006918): Análisis de Regresión Lineal e Introducción al Muestreo.