

Modelos lineales

Hortensia J. Reyes Cervantes
colab. Alonso Nahir Ramírez

Septiembre 2023

1. Regresión lineal simple

Un modelo de regresión lineal simple es aquel que tiene un único regresor x que tiene una relación con la variable de respuesta y en una línea recta. Este modelo lineal simple es

$$y = \mu_{y|x} + \varepsilon = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

donde el intercepto β_0 y la pendiente β_1 son constantes desconocidas y ε es el error aleatorio. A partir de una muestra $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ de las variables x e y , se trata de obtener la ecuación 1. $\mu_{y|x}$ es el componente determinista y ε es el componente aleatorio. Así, esta relación no es determinista, por lo cual encontraremos a la recta que mejor represente la tendencia de los datos encontrando los estimadores.

1.1. Suposiciones del modelo de regresión lineal simple

Es posible hacer una estimación de las constantes en 1 si los siguientes **supuestos** son ciertos:

1. y es una variable aleatoria con función de densidad que depende de x .

$$E(y) = \mu_{y|x}, V(y) = \sigma_{y|x}^2$$

2. Modelo de línea recta $E(y|x) = \mu_{y|x} = \beta_0 + \beta_1 x$

3. Homogeneidad de varianzas:

$$\sigma_{y|x_1}^2 = \sigma_{y|x_2}^2 = \dots = \sigma_{y|x_n}^2$$

4. Independencia entre los valores de la variable dependiente y . Los valores de y deberán ser estadísticamente independientes.

Ejemplos:

(No independencia) Se registró dos veces el peso de un individuo en un tiempo menor de una hora.

(No independencia) Grado de agresividad de dos hermanos que viven en el mismo ambiente familiar.

5. Normalidad en las observaciones de x (variable manipulable), por lo cual la variable aleatoria no observable ε es normal.

$$\varepsilon_i \sim N(0, \sigma^2) \Rightarrow V(\varepsilon_i) = \sigma^2 = E(\varepsilon_i) \Rightarrow \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ si } i \neq j \Rightarrow y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

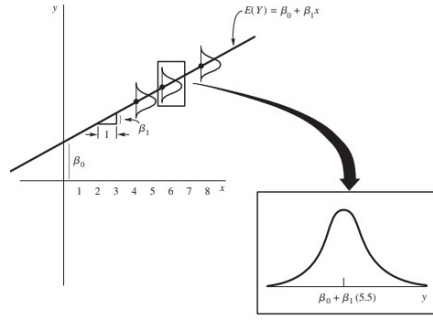


Figura 1: Gráfica del modelo probabilístico $y = \beta_0 + \beta_1 x + \varepsilon$

Si X_i es una variable observada, entonces y_i es una variable que obtiene la aleatoriedad por el error aleatorio ε_i . Es sencillo ver que

$$E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i + E(\varepsilon_i) = \beta_0 + \beta_1 x_i$$

y

$$V(y_i) = V(\beta_0 + \beta_1 x_i + \varepsilon_i) = V(\beta_0 + \beta_1 x_i) + V(\varepsilon_i) = \sigma^2$$

1.2. Estimación de la ecuación de regresión por mínimos cuadrados

Podemos dibujar un diagrama de puntos para ver la tendencia de los datos, aunque no tiene validez estadística por ser subjetivo e impreciso.

El método de mínimos cuadrados, usado por Gauss (1777-1855), produce estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimizan la suma de cuadrados de las distancias entre los valores observados y_i y los valores estimados \hat{y}_i . El modelo muestral de regresión se puede explicar por la ecuación

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

Donde nos interesa minimizar la suma positiva de los errores $\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$. Sea $f(\beta_0, \beta_1)$ una función sobre los parámetros

$$f(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^n \varepsilon_i^2 \quad (3)$$

Por simplicidad, vamos a denotar $\sum_{i=1}^n = \Sigma$. Los estimadores por mínimos cuadrados que denotados por $\hat{\beta}_0$ y $\hat{\beta}_1$ deben satisfacer

$$\begin{aligned} \frac{\partial f}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1} &= 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) (-1) = 0 \\ &\Rightarrow n y_i - n \hat{\beta}_0 - \sum \hat{\beta}_1 x_i = 0 \end{aligned} \quad (4)$$

y

$$\begin{aligned} \frac{\partial f}{\partial \beta_1} \Big|_{\hat{\beta}_0, \hat{\beta}_1} &= 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) (-x_i) = 0 \\ &\Rightarrow \sum x_i y_i - \hat{\beta}_0 \sum x_i - \sum \hat{\beta}_1 x_i^2 = 0 \end{aligned} \quad (5)$$

De aquí, se obtiene el siguiente sistema llamado **sistema de ecuaciones normales**:

$$\begin{aligned}\sum y_i &= n\hat{\beta}_0 + \sum \hat{\beta}_1 x_i \\ \sum x_i y_i &= \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2\end{aligned}$$

Dividiendo (4) entre n se obtiene

$$\frac{\sum y_i}{n} - \hat{\beta}_0 - \frac{\sum \hat{\beta}_1 x_i}{n} = 0 \Rightarrow \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0 \Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Sustituyendo esto en (5),

$$\begin{aligned}\sum x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum x_i &= \sum \hat{\beta}_1 x_i^2 \\ \sum x_i y_i - \bar{y} \sum x_i + \hat{\beta}_1 \bar{x} \sum x_i &= \sum \hat{\beta}_1 x_i^2 \\ \sum x_i y_i - \bar{y} \sum x_i &= \sum \hat{\beta}_1 x_i^2 - \hat{\beta}_1 \bar{x} \sum x_i\end{aligned}$$

Desarrollando el lado derecho

$$\begin{aligned}\sum \hat{\beta}_1 x_i^2 - \hat{\beta}_1 \bar{x} \sum x_i &= \hat{\beta}_1 (\sum x_i^2 - \bar{x} \sum x_i (\frac{n}{n})) \\ &= \hat{\beta}_1 (\sum x_i^2 - n\bar{x}^2) \\ &= \hat{\beta}_1 (\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2) \\ &= \hat{\beta}_1 (\sum x_i^2 - 2 \sum x_i \bar{x} + n\bar{x}^2) \\ &= \hat{\beta}_1 \sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \hat{\beta}_1 \sum (x_i - \bar{x})^2\end{aligned}$$

y el lado izquierdo

$$\begin{aligned}\sum x_i y_i - \bar{y} \sum x_i &= \sum (x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum (x_i y_i - \bar{y} x_i) \\ &= \sum (x_i y_i - \bar{y} x_i + 0) \\ &= \sum (x_i y_i - \bar{y} x_i) - \bar{x} \sum y_i + n\bar{x} \bar{y} \quad (*) \\ &= \sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum (x_i - \bar{x})(y_i - \bar{y})\end{aligned}$$

Veamos que la expresión en (*) es cero:

$$-\bar{x} \sum y_i + n\bar{x} \bar{y} = \bar{x} (-\sum y_i + \frac{n \sum y_i}{n}) = \bar{x} (-\sum y_i + \sum y_i) = \bar{x}(0) = 0$$

Ahora, igualando el término izquierdo y derecho

$$\begin{aligned}\sum (x_i - \bar{x})(y_i - \bar{y}) &= \hat{\beta}_1 \sum (x_i - \bar{x})^2 \\ \hat{\beta}_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}\end{aligned}$$

Por lo tanto, los estimadores para β_0 y β_1 están dados por

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \frac{\sum y_i x_i - \frac{(\sum y_i)(\sum x_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

y

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

El modelo ajustado de regresión lineal simple en (2) es, entonces,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (6)$$

Notación:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{y} = \frac{\sum y_i}{n}$$

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

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

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

Ejemplo: Los siguientes datos corresponden a la presión barométrica y la temperatura del punto de ebullición para 17 lugares en los Alpes y en Escocia.

Número de caso	Temp(°F)	Presión(In Hg)	Lpres=100*log(Presión)
1	194.5	20.79	131.79
2	194.3	20.79	131.79
3	197.9	22.40	135.02
4	198.4	22.67	135.55
5	199.4	23.15	136.46
6	199.9	23.35	136.83
7	200.9	23.89	137.82
8	201.1	23.99	138.00
9	201.4	24.02	138.06
10	201.3	24.01	138.04
11	203.6	25.14	140.04
12	204.6	26.57	142.44
13	209.5	28.49	145.47
14	208.6	27.76	144.34
15	210.7	29.04	146.30
16	211.9	29.88	147.54
17	212.2	30.06	147.80

Si nos interesa hacer un modelo de regresión lineal simple sobre las observaciones de temperatura y la transformación logarítmica de la presión podemos realizarlo con los estimadores anteriores. Sea X la temperatura

y Y como L_{pres} , entonces

$$\bar{x} = 202.95294$$

$$S_{xx} = 530.78235$$

$$\bar{y} = 139.60529$$

$$S_{xy} = 427.79402$$

De esta manera, los estimadores son

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = 0.895$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = -42.138$$

Dado que este fenómeno tiene un comportamiento físico determinista, la línea estimada (Fig. 2) tiene un gran ajuste con las observaciones. Los errores se pueden deber, por ejemplo, a los instrumentos de medición. La ecuación es

$$\hat{y} = -42.138 + 0.895Temp$$

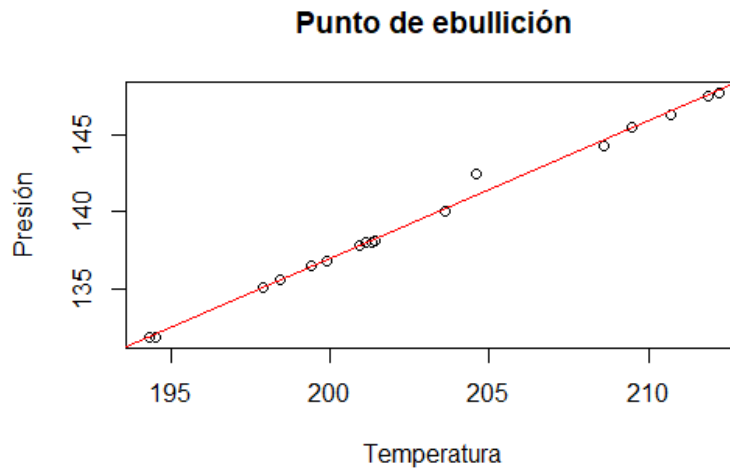


Figura 2: Línea estimada Temp L_{pres}

Teorema 1. *Los estimadores obtenidos por el método de mínimos cuadrados $\hat{\beta}_0, \hat{\beta}_1$ son los mejores estimadores lineales insesgados y de mínima varianza.*

Demostración. Sea $\hat{\beta}_1 = \sum c_i y_i$ donde los c_i son constantes

$$\begin{aligned}
 E(\hat{\beta}_1) = \beta_1 &\iff E(\hat{\beta}_1) = E(\sum c_i y_i) = E(\sum c_i (\beta_0 + \beta_1 x_i + \varepsilon_i)) \\
 &= E(\beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i \varepsilon_i) \\
 &= \beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i E(\varepsilon_i) \\
 &= \beta_0 \sum c_i + \beta_1 \sum c_i x_i \\
 &= \beta_1 \iff \sum c_i = 0 \wedge \sum c_i x_i = 1 \\
 \therefore E(\hat{\beta}_1) &= E(\sum c_i y_i) = \beta_1
 \end{aligned}$$

Por otro lado

$$\begin{aligned}
 V(\hat{\beta}_1) &= V(\sum c_i y_i) = E(\sum c_i y_i - E(\sum c_i y_i))^2 \\
 &= E(\sum c_i y_i - \beta_1)^2 \\
 &= E(\sum c_i (\beta_0 + \beta_1 x_i + \varepsilon_i) - \beta_1)^2 \\
 &= E(\beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i \varepsilon_i - \beta_1)^2 \\
 &= E(\beta_1 + \sum c_i \varepsilon_i - \beta_1)^2 \\
 &= E(\sum c_i \varepsilon_i)^2 \\
 &= E(\sum c_i \varepsilon_i + \sum_{i \neq j} \sum c_i \varepsilon_i c_j \varepsilon_j) \\
 &= \sum c_i^2 E(\varepsilon_i^2) + \sum_{i \neq j} \sum c_i c_j E(\varepsilon_i \varepsilon_j) = \sum c_i^2 \sigma^2 \\
 \therefore V(\hat{\beta}_1) &= \sum c_i^2 \sigma^2 = \sigma^2 \sum c_i^2
 \end{aligned}$$

Se desea minimizar esta varianza sujeto a $\sum c_i = 0 \wedge \sum c_i x_i = 1$. Esto es equivalente a minimizar $\sum c_i^2$, pues $\sigma^2 > 0$, con las mismas condiciones.

Utilizando multiplicadores de Lagrange. Sea $\varphi = \sum c_i^2 - 2\lambda \sum c_i - 2\delta(\sum c_i x_i - 1) = \varphi(c_i, \lambda, \delta)$. Entonces

$$\begin{aligned}
 \frac{\partial \varphi}{\partial c_i} = 2 \sum c_i - 2\lambda - 2\delta x_i &= 0 &\Rightarrow & \lambda = -\delta \bar{x} & (*) \\
 \frac{\partial \varphi}{\partial \lambda} = -2 \sum c_i &= 0 &\Rightarrow & \sum c_i = 0 \\
 \frac{\partial \varphi}{\partial \delta} = -2(\sum c_i x_i - 1) &= 0 &\Rightarrow & \sum c_i x_i = 1
 \end{aligned}$$

veamos que (*) se cumple

$$\begin{aligned}
 2c_i = 2\lambda + 2\delta x_i &\Rightarrow c_i = \lambda + \delta x_i, \quad i = \overline{1, n} && \diamond \\
 &\Rightarrow c_i = n\lambda + \delta \sum x_i \\
 \lambda &= \frac{\sum c_i - \delta \sum x_i}{n} \\
 \lambda &= \frac{\sum c_i}{n} - \delta \bar{x} \\
 \lambda &= -\delta \bar{x}
 \end{aligned}$$

Si sustituimos este último resultado en \diamond , entonces

$$\begin{aligned} c_i &= -\delta\bar{x} + \delta x_i \Rightarrow c_i = \delta(x_i - \bar{x}) \\ &\Rightarrow c_i x_i = \delta(x_i - \bar{x})x_i \end{aligned}$$

Sumando

$$\begin{aligned} 1 &= \sum c_i x_i = \delta \sum (x_i - \bar{x})x_i \\ &= \delta \sum (x_i - \bar{x})^2 \\ \therefore 1 &= \delta \sum (x_i - \bar{x})^2 \\ &\Rightarrow \delta = \frac{1}{\sum (x_i - \bar{x})^2} \end{aligned}$$

De donde el valor de c_i es

$$c_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$

Por lo tanto

$$\hat{\beta}_1 = \sum c_i y_i = \sum \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} y_i = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \hat{\beta}_1$$

Se queda como ejercicio para el lector hacer el resto de la demostración para $\hat{\beta}_0$. □

Aunque ya hemos dado un sustento estadístico a la estimación de los parámetros por mínimos cuadrados, también es posible obtener los mismos resultados **maximizando la función de verosimilitud**.

$$\begin{aligned} L(y_i, x_i, \beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i)^2\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right] \end{aligned}$$

Llamemos $\ddot{\beta}_0$, $\ddot{\beta}_1$ y $\ddot{\sigma}^2$ a los estimadores por máxima verosimilitud. La máxima verosimilitud de L es equivalente a maximizar $\ln L$, de tal modo que

$$\ln L(y_i, x_i, \beta_0, \beta_1, \sigma^2) = \frac{-n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

y los estimadores deben satisfacer

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \beta_0} &= \frac{1}{\ddot{\sigma}^2} \sum (y_i - \ddot{\beta}_0 - \ddot{\beta}_1 x_i) = 0 \\ \frac{\partial \ln(L)}{\partial \beta_1} &= \frac{1}{\ddot{\sigma}^2} \sum (y_i - \ddot{\beta}_0 - \ddot{\beta}_1 x_i) x_i = 0 \\ \frac{\partial \ln(L)}{\partial \sigma^2} &= \frac{-n}{2\ddot{\sigma}^2} + \frac{1}{2\ddot{\sigma}^2} \sum (y_i - \ddot{\beta}_0 - \ddot{\beta}_1 x_i)^2 = 0 \end{aligned}$$

de aquí llegamos a las ecuaciones normales

$$\begin{aligned} \sum y_i - n\ddot{\beta}_0 - \ddot{\beta}_1 \sum x_i &= 0 \\ \sum x_i y_i - \ddot{\beta}_0 \sum x_i - \ddot{\beta}_1 \sum x_i^2 &= 0 \\ \frac{1}{2\ddot{\sigma}^2} \sum (y_i - \ddot{\beta}_0 - \ddot{\beta}_1 x_i)^2 &= \frac{n}{2\ddot{\sigma}^2} \end{aligned}$$

de donde

$$\begin{aligned}\bar{y} &= \ddot{\beta}_0 + \ddot{\beta}_1 \bar{x} \\ \sum x_i y_i &= \ddot{\beta}_0 \sum x_i + \ddot{\beta}_1 \sum x_i^2\end{aligned}$$

Entonces

$$\ddot{\beta}_0 = \bar{y} - \ddot{\beta}_1 \bar{x}$$

y

$$\begin{aligned}\sum y_i x_i &= (\bar{y} - \ddot{\beta}_1 \bar{x}) \sum x_i + \ddot{\beta}_1 \sum x_i^2 \\ &= \bar{y} \sum x_i - \ddot{\beta}_1 \sum x_i \bar{x} + \ddot{\beta}_1 \sum x_i^2 \\ &= \bar{y} \sum x_i + \ddot{\beta}_1 (\sum x_i^2 - \sum x_i \bar{x})\end{aligned}$$

Por lo tanto

$$\ddot{\beta}_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \sum x_i \bar{x}} = \hat{\beta}_1$$

de la última ecuación podemos obtener un estimador de máxima verosimilitud sesgado para σ^2

$$\ddot{\sigma}^2 = \frac{\sum (y_i - \ddot{\beta}_0 - \ddot{\beta}_1 x_i)^2}{n} = \frac{\sum e_i^2}{n}$$

1.3. Propiedades de los estimadores

Definamos al residual como la diferencia del valor observado y_i y el valor ajustado correspondiente \hat{y}_i . Es decir

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, 2, \dots, n$$

De esta manera,

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$

El lector puede verificar que se cumplen las siguientes identidades:

$$\begin{aligned}\sum y_i &= \sum \hat{y}_i \\ \sum x_i e_i &= 0 \\ \sum \hat{y}_i e_i &= 0\end{aligned} \tag{7}$$

Teorema 2. *Los estimadores cumplen las siguientes propiedades:*

- I. $E(\hat{\beta}_1) = \beta_1$
- II. $E(\hat{\beta}_0) = \beta_0$
- III. $V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$
- IV. $V(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$

$$v. \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \bar{x}}{S_{xx}}$$

Demostración. 1. Sea $w_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$, entonces

$$\begin{aligned} \sum w_i &= \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i - n\bar{x}}{\sum (x_i - \bar{x})^2} = \frac{n\bar{x} - n\bar{x}}{\sum (x_i - \bar{x})^2} = 0 \\ \sum x_i w_i &= \frac{\sum x_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i^2 - x_i \bar{x})}{\sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2)} = \frac{\sum x_i^2 - \bar{x} \sum x_i}{\sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2} = \frac{\sum x_i^2 - n\bar{x}^2}{\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2} = \frac{\sum x_i^2 - n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} = 1 \end{aligned}$$

Ahora

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}\right) = E\left(\sum w_i y_i\right) = \sum w_i E(y_i) = \sum w_i (\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum w_i + \beta_1 \sum w_i x_i = \beta_0 (0) + \beta_1 (1) = \beta_1 \end{aligned}$$

II.

$$\begin{aligned} E(\hat{\beta}_0) &= E(\bar{y} - \bar{x} \hat{\beta}_1) = E(\bar{y}) - \bar{x} E(\hat{\beta}_1) = \frac{\sum E(y)}{n} - \bar{x} \beta_1 \\ &= \frac{\sum (\beta_0 + \beta_1 x_i)}{n} - \bar{x} \beta_1 = \frac{n\beta_0}{n} + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0 \end{aligned}$$

III.

$$\begin{aligned} V(\hat{\beta}_1) &= E[\hat{\beta}_1 - E(\hat{\beta}_1)]^2 = E[\beta_1 + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2} - \beta_1]^2 = \left[\frac{1}{\sum (x_i - \bar{x})^2}\right]^2 E[\sum (x_i - \bar{x}) \varepsilon_i]^2 \\ &= \left[\frac{1}{\sum (x_i - \bar{x})^2}\right]^2 E[\sum (x_i - \bar{x})^2 \varepsilon_i^2 + \sum_i \sum_j (x_i - \bar{x}) \varepsilon_i (x_j - \bar{x}) \varepsilon_j] \\ &= \left[\frac{1}{\sum (x_i - \bar{x})^2}\right]^2 \sum (x_i - \bar{x})^2 E[\varepsilon_i^2] + \sum_i \sum_j (x_i - \bar{x}) (x_j - \bar{x}) E[\varepsilon_i \varepsilon_j] \quad (*) \\ &= \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \end{aligned}$$

Verifiquemos (*). Como los errores son independientes, entonces

$$\begin{aligned} E(\varepsilon_i \varepsilon_j) &= E(\varepsilon_i) E(\varepsilon_j) = 0 \quad \forall i \neq j \\ V(\varepsilon_i) &= \sigma^2 = E(\varepsilon_i^2) - E(\varepsilon_i)^2 = E(\varepsilon_i^2) \end{aligned}$$

IV. Antes, algunos resultados útiles

$$\begin{aligned} V(\bar{y}) &= V(\beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}) = V(\bar{\varepsilon}) = V\left(\frac{\sum \varepsilon_i}{n}\right) = \frac{1}{n} V(\varepsilon_i) = \frac{\sigma^2}{n} \\ \text{cov}(\bar{y}, \hat{\beta}_1) &= \text{cov}\left(\frac{\sum y_i}{n}, \sum w_i y_i\right) = V\left(\sum \frac{w_i}{n} y_i\right) = \left(\frac{\sum w_i}{n}\right)^2 V(y_i) = (0)(\sigma^2) = 0 \end{aligned}$$

Ahora es más sencillo calcular la varianza

$$\begin{aligned} V(\hat{\beta}_0) &= V(\bar{y} - \bar{x} \hat{\beta}_1) = V(\bar{y}) + \bar{x}^2 V(\hat{\beta}_1) - 2\bar{x} \text{cov}(\bar{y}, \hat{\beta}_1) = \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right) = \sigma^2 \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2} \end{aligned}$$

v.

$$\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{cov}(\bar{y} - \bar{x}\hat{\beta}_1, \hat{\beta}_1) = \text{cov}(\bar{y}, \hat{\beta}_1) - \bar{x}\text{cov}(\hat{\beta}_1, \hat{\beta}_1) = 0 - \bar{x}V(\hat{\beta}_1) = -\bar{x}\frac{\sigma^2}{S_{xx}}$$

Análogo:

$$\begin{aligned} \text{cov}(\hat{\beta}_0, \hat{\beta}_1) &= E[\hat{\beta}_0\hat{\beta}_1] - E(\hat{\beta}_0)E(\hat{\beta}_1) \\ &= E[(\bar{y} - \bar{x}\hat{\beta}_1)\hat{\beta}_1] - \beta_0\beta_1 \\ &= \bar{y}\beta_1 - \bar{x}E[\hat{\beta}_1^2] - \beta_0\beta_1 \\ &= \bar{y}\beta_1 - \bar{x}\left(\frac{\sigma^2}{\sum(x_i - \bar{x})^2} + \beta_1^2\right) - \beta_0\beta_1 \\ &= \bar{y}\beta_1 - \bar{x}\frac{\sigma^2}{S_{xx}} - \bar{x}\beta_1^2 - \beta_0\beta_1 \\ &= (\bar{y} - \bar{x}\beta_1)\beta_1 - \beta_0\beta_1 - \bar{x}\frac{\sigma^2}{S_{xx}} \\ &= \beta_0\beta_1 - \beta_0\beta_1 - \frac{\bar{x}\sigma^2}{S_{xx}} \\ &= -\frac{\bar{x}\sigma^2}{S_{xx}} \end{aligned}$$

□

Estos resultados son de utilidad para crear intervalos de confianza y pruebas de hipótesis que nos servirán para medir la adecuación del modelo estimado y hacer estimaciones sobre los datos. Si no conocemos la varianza (constante por suposición) de la v.a. ε , entonces se debe estimar. Idealmente esperamos que esta estimación no dependa de la adecuación del modelo y es posible si existen distintas observaciones y para al menos un valor de x . Cuando esto no es posible, podemos obtener una estimación insesgada a partir de los residuales.

Teorema 3. $\hat{\sigma}^2$ es un estimador insesgado para la varianza poblacional σ^2 , donde

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}.$$

Demostración.

$$E[\hat{\sigma}^2] = E\left[\frac{\sum (y_i - \hat{y}_i)^2}{n-2}\right] = \frac{1}{n-2}E[\sum (y_i - \hat{y}_i)^2]$$

Donde

$$\begin{aligned}
E[\sum (y_i - \hat{y}_i)^2] &= E[\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2] \\
&= E[\sum (y_i - \bar{y} + \bar{x}\hat{\beta}_1 - \hat{\beta}_1 x_i)^2] \\
&= E[\sum [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2] \\
&= E[\sum (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2] \\
&= E[\sum y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xx}] \quad (*) \\
&= \sum E(y_i^2) - nE(\bar{y}^2) - S_{xx}E(\hat{\beta}_1^2) \\
&= \sum [V(y_i) + E^2(y_i)] - n[V(\bar{y}) + E^2(\bar{y})] - S_{xx}[V(\hat{\beta}_1) + E^2(\hat{\beta}_1)] \quad (**) \\
&= n\sigma^2 + \sum (\beta_0 + \beta_1 x_i)^2 - n[\frac{\sigma^2}{n} + (\beta_0 + \beta_1 \bar{x})^2] - S_{xx}[\frac{\sigma^2}{S_{xx}} + \beta_1^2] \\
&= n\sigma^2 + \sum \beta_0^2 + 2\beta_0\beta_1 \sum x_i + \beta_1^2 \sum x_i^2 - \sigma^2 - n\beta_0^2 - 2n\beta_0\beta_1 \bar{x} - n\beta_1^2 \bar{x}^2 - \sigma^2 - S_{xx}\beta_1^2 \\
&= (n-2)\sigma^2 + n\beta_0^2 - n\beta_0^2 + 2\beta_0\beta_1 \sum x_i - 2\beta_0\beta_1 \sum x_i + \beta_1^2 \sum x_i^2 - \beta_1^2 \frac{(\sum x_i)^2}{n} - S_{xx}\beta_1^2 \\
&= (n-2)\sigma^2 + (\sum x_i^2 - \frac{1}{n}(\sum x_i)^2)\beta_1^2 - S_{xx}\beta_1^2 \\
&= (n-2)\sigma^2 + S_{xx}\beta_1^2 - S_{xx}\beta_1^2 \\
&= (n-2)\sigma^2
\end{aligned}$$

En (*), como $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x})^2 \hat{\beta}_1$, entonces $-2\hat{\beta}_1 \sum (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 = -\hat{\beta}_1 S_{xx}$.
En (**), para cualquier variable aleatoria U , se cumple $E(U^2) = V(U) + E^2(U)$.

Por lo tanto

$$\frac{1}{n-2}E[\sum (y_i - \hat{y}_i)^2] = \frac{1}{n-2}((n-2)\sigma^2) = \sigma^2$$

□

Una fórmula análoga útil para cálculos es

$$\hat{\sigma}^2 = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}$$

1.4. Pruebas de hipótesis

Hasta ahora hemos usado la esperanza y la varianza de los errores ε , pero no había sido necesario que tuvieran una distribución normal. En esta sección nos interesa crear pruebas de hipótesis sobre los parámetros del modelo y para ello es conveniente conocer la distribución de sus estimadores.

Suponiendo que deseamos probar la hipótesis de que algún beta es igual a una constante. Por ejemplo, para β_1

$$H_0 : \beta_1 = \beta_1^* \text{ vs } H_a : \beta_1 \neq \beta_1^*$$

Como los errores son normales, independientes e idénticamente distribuidos $\varepsilon_i \sim N(0, \sigma^2)$, las observaciones y_i también lo son $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$. Dado que $\hat{\beta}_1$ es una combinación lineal de estas observaciones,

entonces está normalmente distribuido. Gracias al teorema 2 conocemos la media y la varianza de los estimadores, de esta manera $\beta_1 \sim N(\beta_1, \sigma^2/S_{xx})$. Con esto, el estadístico de prueba

$$Z_0 = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\sigma^2/S_{xx}}}$$

tiene una distribución normal estándar si la hipótesis nula es cierta. Si conocemos la varianza σ^2 del error podemos utilizar Z_0 , pero es poco común saber el valor real, sin embargo, es posible estimarla con $\hat{\sigma}^2$ dada en el teorema 3. Note que $(n-2)\hat{\sigma}^2/\sigma^2$ tiene una distribución χ_{n-2}^2 y es independiente de $\hat{\beta}_1$. Por definición, el estadístico t dado por

$$t_0 = \frac{Z}{\sqrt{\chi^2/n}} = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\hat{\sigma}^2/S_{xx}}} \quad (8)$$

sigue una distribución t_{n-2} si la hipótesis nula es verdadera. Definamos al nivel de confianza α como la probabilidad de rechazar H_0 cuando es cierta. Esto nos da la posibilidad de rechazar H_0 si

$$|t_0| > t_{n-2}^{\alpha/2}$$

De manera análoga podemos encontrar un estadístico para la prueba

$$H_0 : \beta_0 = \beta_0^* \text{ vs } H_a : \beta_0 \neq \beta_0^*$$

utilizamos el estadístico

$$t_0 = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})}}$$

Note que el denominador en ambos casos es la desviación estándar del parámetro. Tomando esto en cuenta se puede hacer una pequeña generalización que toma más importancia en el caso multiparamétrico. El estadístico de prueba t_0 para

$$H_0 : \beta_i = \beta_i^* \text{ vs } H_a : \beta_i \neq \beta_i^*$$

Está dado por

$$t_0 = \frac{\hat{\beta}_i - \beta_i^*}{se(\hat{\beta}_i)} \quad (9)$$

Veremos el **enfoque de razón de verosimilitudes** para esta prueba y, posterior, concluir los mismos resultados. Sea $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ el modelo de regresión y

$$H_0 : \beta_1 = 0 \text{ vs } H_a : \beta_1 \neq 0$$

En el espacio paramétrico $\underline{\theta} = (\beta_0, \beta_1, \sigma^2)$ y función de distribución $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, tenemos la función de verosimilitud

$$L(y; \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \exp \left(\frac{-(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left(\frac{-\sum (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right)$$

Bajo H_0

$$\omega = \{(\beta_0, \beta_1, \sigma^2) | \beta_0 \in \mathbb{R}, \beta_1 = 0, \sigma^2 > 0\} \Rightarrow L(y; \beta_0, \beta_1, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left(\frac{-\sum (y_i - \beta_0)^2}{2\sigma^2} \right) \quad (10)$$

Bajo H_a

$$H = \{(\beta_0, \beta_1, \sigma^2) | \beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R} - \{0\}, \sigma^2 > 0\} \Rightarrow L(y; \beta_0, \beta_1, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(\frac{-\sum(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

Usando (10)

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum(y_i - \beta_0)^2}{2\sigma^2}$$

Encontrando los estimadores de máx-vero.

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta_0} &= \frac{-1}{2\sigma^2} 2 \sum (y_i - \beta_0)(-1) = \frac{\sum(y_i - \beta_0)}{2\sigma^2} = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{2\sum(y_i - \beta_0)^2}{4\sigma^2} = \frac{-n}{2\sigma^2} + \frac{\sum(y_i - \beta_0)^2}{2\sigma^2} = 0 \end{aligned}$$

Entonces

$$\begin{aligned} \sum y_i &= n\beta_0 \Rightarrow \frac{\sum y_i}{n} = \beta_0 \Rightarrow \bar{y} = \hat{\beta}_0 \\ \frac{n}{2\sigma^2} &= \frac{\sum(y_i - \beta_0)^2}{2\sigma^4} \Rightarrow \hat{\sigma}^2 = \frac{\sum(y_i - \beta_0)^2}{n} = \frac{\sum(y_i - \bar{y})^2}{n} \end{aligned}$$

Así, $\hat{\theta}_\omega = (\bar{y}, \hat{\sigma}^2)$. De donde

$$L(\hat{\omega}) = \left(\frac{n}{2\pi\sum(y_i - \bar{y})^2}\right)^{n/2} \exp\left(\frac{-\sum(y_i - \bar{y})^2 n}{2\sum(y_i - \bar{y})^2}\right) = \left(\frac{n}{2\pi\sum(y_i - \bar{y})^2}\right)^{n/2} \exp\left(-\frac{n}{2}\right)$$

Bajo H_a note que tenemos la función de verosimilitud completa y ya hemos el espacio paramétrico estimado por

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \bar{x}\hat{\beta}_1 \\ \hat{\beta}_1 &= \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} \\ \hat{\sigma}^2 &= \frac{\sum e_i^2}{n} \end{aligned}$$

Es decir, $\hat{\theta}_H = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$. De esta manera

$$L(\hat{H}) = \left(\frac{1}{2\pi\frac{\sum e_i^2}{n}}\right)^{n/2} \exp\left(-\frac{\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{2\frac{\sum e_i^2}{n}}\right) = \left(\frac{n}{2\pi\sum e_i^2}\right)^{n/2} \exp\left(-\frac{n}{2}\right)$$

Nuestro criterio de razón de verosimilitudes nos dice que si

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{H})} < \lambda_0$$

se rechaza H_0 . Y si

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{H})} \geq \lambda_0$$

no se rechaza H_0 . Por lo cual

$$\lambda = \frac{\left(\frac{n}{2\pi\sum(y_i - \bar{y})^2}\right)^{n/2} \exp\left(-\frac{n}{2}\right)}{\left(\frac{n}{2\pi\sum e_i^2}\right)^{n/2} \exp\left(-\frac{n}{2}\right)} = \left(\frac{\sum e_i^2}{\sum(y_i - \bar{y})^2}\right)^{n/2} < \lambda_0 \iff \lambda^* = \frac{\sum e_i^2}{\sum(y_i - \bar{y})^2} < \lambda_0^{2/n} = \lambda_0^*$$

Hagamos el desarrollo de los residuales

$$\begin{aligned} \sum e_i^2 &= \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y} + \bar{y} - \hat{y}_i)^2 = \sum [(y_i - \bar{y}) + (\bar{y} - \hat{y}_i)]^2 \\ &= \sum (y_i - \bar{y})^2 - 2 \sum (y_i - \bar{y})(\hat{y}_i - \bar{y}) + \sum (\hat{y}_i - \bar{y})^2 \end{aligned} \quad (11)$$

Donde

$$\begin{aligned} 2 \sum (y_i - \bar{y})(\hat{y}_i - \bar{y}) &= 2 \sum (y_i - \bar{y})(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}) \\ &= 2 \sum (y_i - \bar{y})(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y}) \\ &= 2 \hat{\beta}_1 \sum (y_i - \bar{y})(x_i - \bar{x}) \\ &= 2 \hat{\beta}_1 \sum [(y_i - \bar{y})x_i - (y_i - \bar{y})\bar{x}] \\ &= 2 \hat{\beta}_1 [\sum (y_i - \bar{y})x_i - \bar{x}(\sum y_i - n\bar{y})] \\ &= 2 \hat{\beta}_1 [\sum (y_i - \bar{y})x_i - \bar{x}(\sum y_i - \sum y_i)] \\ &= 2 \hat{\beta}_1 \sum (y_i - \bar{y})x_i \left(\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) \\ &= 2 \hat{\beta}_1 \sum (x_i - \bar{x})^2 \hat{\beta}_1 \\ &= 2 \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 \\ &= 2 \sum (\hat{y}_i - \bar{y})^2 \end{aligned} \quad (*)$$

En (*); sabemos que $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, entonces

$$\begin{aligned} \hat{y}_i - \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x}) \\ \therefore (\hat{y}_i - \bar{y})^2 &= \hat{\beta}_1^2 (x_i - \bar{x})^2 \end{aligned} \quad (12)$$

Regresando a (11)

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2 - 2 \sum (\hat{y}_i - \bar{y})^2 + \sum (\hat{y}_i - \bar{y})^2 = \sum (y_i - \bar{y})^2 - \sum (\hat{y}_i - \bar{y})^2$$

Por lo tanto

$$\begin{aligned} \sum (y_i - \hat{y}_i)^2 &= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 \\ \text{Realidad total} &= \text{Error} + \text{Estimación de la regresión} \end{aligned}$$

Regresando a la ecuación de λ^*

$$\lambda^* = \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2} = \frac{1}{1 + \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}} \leq \lambda_0^*$$

Entonces

$$1 + \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > \frac{1}{\lambda_0^*} \iff \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > \frac{1}{\lambda_0^*} - 1$$

De esta manera, la zona de rechazo es

$$\mathfrak{C} = \left\{ (x_i, y_i), i = \overline{1, n} \mid \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > \frac{1}{\lambda_0^*} - 1 \right\}$$

Recordemos que

$$\frac{(\hat{\beta}_1 - \beta_1)^2}{\frac{\sigma^2}{\sum(x_i - \bar{x})^2}} \sim \chi_{(1)}^2$$

Si $H_0 : \beta_1 = 0$ es verdadero, lo anterior es

$$\frac{\sum(x_i - \bar{x})\hat{\beta}_1^2}{\sigma^2} \sim \chi_{(1)}^2$$

Haciendo la estimación común S^2 para σ^2 , tenemos

$$\frac{\frac{\sum(x_i - \bar{x})^2 \hat{\beta}_1^2}{\sigma^2}}{\frac{\sum e_i^2}{\sigma^2(n-2)}} = \frac{\hat{\beta}_1^2 \sum(x_i - \bar{x})^2 (n-2)}{\sum e_i^2} \sim F_{(1, n-2)}$$

Así, para construir el criterio de decisión

$$\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > C_\alpha \Rightarrow \frac{(n-2) \sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > C_\alpha (n-2) = k_\alpha$$

Tal que

$$\alpha = P(\text{Rechazar } H_0 | H_0 \text{ es verdadera}) = P\left(\frac{(n-2) \sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} > k_\alpha \text{ con } F_{(1, n-2)}\right)$$

Así, por último, sabemos que se rechaza H_0 si

$$\frac{(n-2)}{\sum(y_i - \bar{y})^2} \sum(\hat{y}_i - \bar{y})^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\hat{\sigma}^2} > F_{(1, n-2)}^{1-\alpha}$$

1.5. Intervalos de confianza

Además de obtener la estimación de los parámetros β_0 , β_1 y σ^2 , puede resultar interesante obtener los intervalos de confianza de estos parámetros. La anchura de estos intervalos puede ser una medida de la calidad de una línea de regresión, donde un intervalo amplio implica menos certeza sobre el verdadero valor del parámetro. Con la información presentada en la sección anterior, crear un intervalo de confianza resulta casi natural. Sin embargo, se aprovecha este nuevo espacio para hacer algunas especificaciones.

Como sabemos que $\hat{\beta}_0 \sim N(\beta_0, \frac{\sigma^2 \sum x_i}{n \sum(x_i - \bar{x})^2})$ y que $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum(x_i - \bar{x})^2})$ podemos estandarizar estas variables, de tal forma que

$$z = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\sigma^2 \sum x_i}{n \sum(x_i - \bar{x})^2}}} = (\hat{\beta}_0 - \beta_0) \frac{\sqrt{n \sum(x_i - \bar{x})^2}}{\sigma \sqrt{\sum x_i^2}} \sim N(0, 1)$$

Como no conocemos a σ^2 hay que estimarla con su estimador insesgado $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \sim \chi_{n-2}^2$. Entonces, análogo a la ecuación (8)

$$\frac{Z}{\sqrt{\chi_n^2/n}} = \frac{(\hat{\beta}_0 - \beta_0) \frac{\sqrt{n \sum (x_i - \bar{x})^2}}{\sigma \sqrt{\sum x_i^2}}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2(n-2)}}} = \frac{(\hat{\beta}_0 - \beta_0) \sqrt{n \sum (x_i - \bar{x})^2}}{\hat{\sigma} \sqrt{\sum x_i^2}} \sim t_{(n-2)}$$

Observaciones:

- Este error es muy pequeño y tiende a desaparecer de la anterior expresión.
- La muestra depende de las observaciones y del valor hipotético de β_0 .
- Para saber en que región cae el valor de $\frac{Z}{\sqrt{\chi_n^2/n}}$ con la distribución $t_{(n-2)}$ obtenemos, por ejemplo, al nivel $1 - \alpha$ el intervalo de confianza para el parámetro β_0 es

$$\begin{aligned} 1 - \alpha &= P \left(-t_{n-2}^{\alpha/2} < \frac{Z}{\sqrt{\chi_n^2/n}} < t_{n-2}^{\alpha/2} \right) \\ &= P \left(-t_{n-2}^{\alpha/2} < \frac{(\hat{\beta}_0 - \beta_0) \sqrt{n \sum (x_i - \bar{x})^2}}{\hat{\sigma} \sqrt{\sum x_i^2}} < t_{n-2}^{\alpha/2} \right) \\ &= P \left(\hat{\beta}_0 - t_{n-2}^{\alpha/2} \frac{\hat{\sigma} \sqrt{\sum x_i^2}}{\sqrt{n \sum (x_i - \bar{x})^2}} < \beta_0 < \hat{\beta}_0 + t_{n-2}^{\alpha/2} \frac{\hat{\sigma} \sqrt{\sum x_i^2}}{\sqrt{n \sum (x_i - \bar{x})^2}} \right) \end{aligned}$$

Análogamente para β_1 , tenemos la estandarización

$$z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}}} = \frac{(\hat{\beta}_1 - \beta_1) \sqrt{\sum (x_i - \bar{x})^2}}{\sigma} \sim N(0, 1)$$

Nuevamente, estimamos la varianza poblacional a conciencia de la independencia que tiene con $\hat{\beta}_1$ y $\hat{\beta}_0$. Entonces

$$\frac{Z}{\sqrt{\chi_n^2/n}} = \frac{(\hat{\beta}_1 - \beta_1) \frac{\sum (x_i - \bar{x})^2}{\sigma}}{\sqrt{\frac{\hat{\sigma}^2(n-2)}{\sigma^2(n-2)}}} = \frac{(\hat{\beta}_1 - \beta_1) \sum (x_i - \bar{x})^2}{\hat{\sigma}} \sim t_{(n-2)}$$

Construyendo el intervalo de confianza

$$\begin{aligned} 1 - \alpha &= P \left(-t_{n-2}^{\alpha/2} < \frac{(\hat{\beta}_1 - \beta_1) \sum (x_i - \bar{x})^2}{\hat{\sigma}} < t_{n-2}^{\alpha/2} \right) \\ &= P \left(\hat{\beta}_1 - t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}} < \beta_1 < \hat{\beta}_1 + t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}} \right) \end{aligned}$$

Si utilizamos la notación de la ecuación (9) llegamos a la forma compacta

$$P \left(\hat{\beta}_i - t_{n-2}^{\alpha/2} \sqrt{V(\beta_i)} < \beta_i < \hat{\beta}_i + t_{n-2}^{\alpha/2} \sqrt{V(\beta_i)} \right) = 1 - \alpha \quad (13)$$

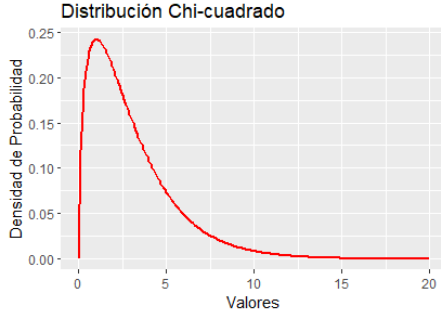


Figura 3: Distribución χ^2 .

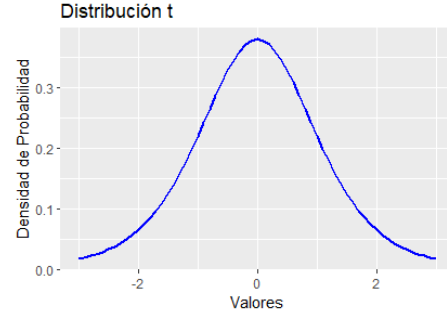


Figura 4: Distribución t .

La simetría de la distribución t fig. (4) nos da la ventaja de tener el mismo valor de $t_{(n-2)}$, sin embargo, para estimar la varianza poblacional usamos $\hat{\sigma}^2$ que sigue una distribución chi cuadrada χ^2 fig. (3) sin simetría. Es por ello que los valores críticos son distintos para cola izquierda y derecha de tal manera que $P(x \leq a) = \alpha/2$ y $P(x \geq b) = \alpha/2$. Entonces, el intervalo de confianza para σ^2 con el estimador $\hat{\sigma}^2 \sim \chi^2_{(n-2)}$ y cola izquierda para la distribución chi es

$$\begin{aligned}
 1 - \alpha &= P\left(\chi^2_{\alpha/2, n-2} < \frac{\sum e_i^2}{\sigma^2} < \chi^2_{1-\alpha/2, n-2}\right) \\
 &= P\left(\chi^2_{\alpha/2, n-2} < \frac{\hat{\sigma}^2(n-2)}{\sigma^2} < \chi^2_{1-\alpha/2, n-2}\right) \\
 &= P\left(\frac{1}{\chi^2_{\alpha/2, n-2}} > \frac{\sigma^2}{\hat{\sigma}^2(n-2)} > \frac{1}{\chi^2_{1-\alpha/2, n-2}}\right) \\
 &= P\left(\frac{\hat{\sigma}^2(n-2)}{\chi^2_{\alpha/2, n-2}} < \sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi^2_{1-\alpha/2, n-2}}\right)
 \end{aligned}$$

1.6. Predicción

Es común que la mayoría de los casos de regresión, sólo se tenga el objetivo de conocer los valores a futuro o pasado del fenómeno estudiado, por lo cual se construye una herramienta estadística que nos ayuda y nos brinda cierta confianza de los posibles resultados.

Supongamos que seguimos con el modelo lineal, donde el problema es predecir el valor medio de y que corresponde a un valor de x , sea x_0 $E(y|x_0)$, el cual puede o no puede pertenecer al rango de variación de los valores muestrales. Existen predicciones puntuales o de intervalos. Suponga que definimos a un estimador como una función lineal de los y_i , $i = \overline{1, n}$. Sea $\hat{y}_0 = \sum c_i y_i$ donde puede suceder que la c_i es una constante que permite ser a \hat{y}_0 el mejor estimador lineal e insesgado. Nos gustaría que

$$E(y_0|x_0) = E(\beta_0 + \beta_1 x_0 + \varepsilon_0) = \beta_0 + \beta_1 x_1$$

En realidad tenemos

$$\hat{y}_0 = \sum c_i (\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i \varepsilon_i \quad (14)$$

De esta manera, al obtener la esperanza condicionada a un valor particular x_0 tenemos que

$$\begin{aligned} E(\hat{y}_0|x_0) &= E(\beta_0 \sum c_i + \beta_1 \sum c_i x_i + \sum c_i \varepsilon_i) = \beta_0 E(\sum c_i) + \beta_1 E(\sum c_i x_i) + E(\sum c_i \varepsilon_i) \\ &= \beta_0 E(\sum c_i) + \beta_1 E(\sum c_i x_i) = \beta_0 + \beta_1 x_0 \iff \sum c_i = 1 \wedge \sum c_i x_i = x_0 \end{aligned} \quad (15)$$

Si lo anterior se cumple podemos afirmar que nuestro estimador es insesgado. ¿Será de varianza mínima?

$$\begin{aligned} V(\hat{y}_0) &= E[\hat{y}_0 - E(\hat{y}_0|x_0)]^2 = E[\beta_0 + \beta_1 x_0 - (\beta_0 + \beta_1 x_0 + \sum c_i \varepsilon_i)]^2 \\ &= E[-\sum c_i \varepsilon_i]^2 = E[\sum c_i \varepsilon_i]^2 = E[\sum c_i^2 \varepsilon_i^2 + \sum_{i \neq j} \sum c_i c_j \varepsilon_i \varepsilon_j] \\ &= \sum c_i^2 E(\varepsilon_i^2) + \sum_{i \neq j} \sum c_i c_j E(\varepsilon_i \varepsilon_j) = \sum c_i^2 \sigma^2 \end{aligned} \quad (16)$$

Como $\sigma^2 > 0$, utilizaremos el criterio de multiplicadores de Lagrange. De (15) vemos que $\sum c_i - 1 = 0$ y $\sum c_i x_i - x_0 = 0$. Sea $\phi = \sum c_i^2 - 2\lambda(\sum c_i - 1) - 2\mu(\sum c_i x_i - x_0) = 0$ entonces

$$\frac{\partial \phi}{\partial c_i} = 2c_i - 2\lambda - 2\mu x_i = 0 \quad \Rightarrow \quad c_i - \lambda - \mu x_i = 0 \quad (17)$$

$$\frac{\partial \phi}{\partial \lambda} = -2(\sum c_i - 1) = 0 \quad \Rightarrow \quad \sum c_i = 1 \quad (18)$$

$$\frac{\partial \phi}{\partial \mu} = -2(\sum c_i x_i - x_0) = 0 \quad \Rightarrow \quad \sum c_i x_i = x_0 \quad (19)$$

Sumando en (17)

$$\sum c_i - n\lambda - \mu \sum x_i = 0$$

Despejando λ y usando (18)

$$\lambda = \frac{\mu \sum x_i - \sum c_i}{-n} = -\mu \bar{x} + \frac{1}{n} \quad (20)$$

Sustituyendo (20) en (17)

$$c_i = \lambda + \mu x_i = -\mu \bar{x} + \frac{1}{n} + \mu x_i \quad (21)$$

Ahora, usando (21) en (19) llegamos a

$$\begin{aligned} \sum c_i x_i &= \sum (-\mu \bar{x} + \frac{1}{n} + \mu x_i) x_i = -\mu \bar{x} \sum x_i + \frac{\sum x_i}{n} + \mu \sum x_i^2 \\ &= \bar{x} + \mu \sum x_i (x_i - \bar{x}) = x_0 \end{aligned} \quad (22)$$

Si de la última igualdad despejamos a μ obtenemos

$$\mu = \frac{x_0 - \bar{x}}{\sum x_i (x_i - \bar{x})} \quad (23)$$

Con esto, ya podemos encontrar un valor explícito de c_i si sustituimos μ en (21)

$$c_i = -\mu \bar{x} + \frac{1}{n} + \mu x_i = \frac{1}{n} + \mu (x_i - \bar{x}) = \frac{1}{n} + \frac{x_0 - \bar{x}}{\sum x_i (x_i - \bar{x})} (x_i - \bar{x}) \quad (24)$$

Entonces, podemos expresar a \hat{y}_0 de (14) en términos de (24)

$$\begin{aligned}
\hat{y}_0 &= \sum c_i(\beta_0 + \beta_1 x_i + \varepsilon_i) = \sum c_i y_i \\
&= \sum \left[\frac{1}{n} + \frac{x_0 - \bar{x}}{\sum x_i(x_i - \bar{x})} (x_i - \bar{x}) \right] y_i \\
&= \frac{\sum y_i}{n} + \sum \frac{(x_0 - \bar{x})(x_i - \bar{x}) y_i}{\sum (x_i - \bar{x}) x_i} \\
&= \bar{y} + (x_0 - \bar{x}) \hat{\beta}_1 \\
&= \bar{y} + x_0 \hat{\beta}_1 - \bar{x} \hat{\beta}_1 \\
&= \bar{y} - \bar{x} \hat{\beta}_1 + x_0 \hat{\beta}_1 \\
&= \hat{\beta}_0 + x_0 \hat{\beta}_1
\end{aligned}$$

Por lo tanto, el mejor estimador de $\beta_0 + \beta_1 x_0$ es $\hat{\beta}_0 + \hat{\beta}_1 x_0$. Desde (15) conocemos la esperanza de \hat{y}_0 , ahora que conocemos c_i podemos sustituir en (16) para encontrar la varianza.

$$\begin{aligned}
V(\hat{y}_0) &= \sum c_i \sigma^2 = \sigma^2 \sum \left[\frac{1}{n} + \frac{(x_i - x_0)(x_0 - \bar{x})}{\sum (x_i - \bar{x}) x_i} \right] \\
&= \sigma^2 \sum \left[\frac{1}{n^2} + \frac{2(x_i - x_0)(x_0 - \bar{x})}{n \sum (x_i - \bar{x}) x_i} + \left(\frac{(x_i - x_0)(x_0 - \bar{x})}{\sum (x_i - \bar{x}) x_i} \right)^2 \right] \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2 \sum (x_i - \bar{x})^2}{(\sum (x_i - \bar{x}) x_i)^2} \right] \quad (*) \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \quad (**)
\end{aligned}$$

En (*), veamos que $\sum \frac{2(x_i - x_0)(x_0 - \bar{x})}{n \sum (x_i - \bar{x}) x_i} = 0$

$$\begin{aligned}
\sum \frac{2(x_i - x_0)(x_0 - \bar{x})}{n \sum (x_i - \bar{x}) x_i} &= 2(x_0 - \bar{x}) \frac{\sum (x_i - \bar{x})}{n \sum (x_i - \bar{x}) x_i} = 2(x_0 - \bar{x}) \frac{\sum x_i - \sum \bar{x}}{n \sum (x_i - \bar{x}) x_i} \\
&= 2(x_0 - \bar{x}) \left[\frac{\bar{x}}{\sum (x_i - \bar{x}) x_i} - \frac{\bar{x}}{\sum (x_i - \bar{x}) x_i} \right] = 0
\end{aligned}$$

En (**), demostremos que $\frac{\sum (x_i - \bar{x})^2}{(\sum (x_i - \bar{x}) x_i)^2} = \frac{1}{\sum (x_i - \bar{x})^2}$. Sabemos que $\sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x}) x_i$, entonces

$$1 = \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x}) x_i} \Rightarrow 1 = \frac{(\sum (x_i - \bar{x})^2)^2}{(\sum (x_i - \bar{x}) x_i)^2} \Rightarrow \frac{\sum (x_i - \bar{x})^2}{(\sum (x_i - \bar{x}) x_i)^2} = \frac{1}{\sum (x_i - \bar{x})^2}$$

Se puede obtener el mismo resultado calculando la varianza por definición, de tal forma que $V(\hat{y}_0) = E(\hat{y}_0 - E(\hat{y}_0|x_0))^2 = E(\hat{\beta}_0 + \hat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0))^2$. La demostración se deja como entretenida actividad al lector. En resumen, podemos decir que

$$\begin{aligned}
E(\hat{y}_0|x_0) &= \beta_0 + \beta_1 x_0 \\
V(\hat{y}_0) &= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \quad (25)
\end{aligned}$$

Donde \hat{y}_0 es una transformación lineal de una v.a. normal, de tal manera que

$$\hat{y}_0 \sim N\left(\beta_0 + \beta_1 x_0, \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]\right)$$

Una vez que conocemos la distribución, podemos encontrar un **intervalo de confianza** para la respuesta media de y dado un x_0 , es decir, $E(\hat{y}_0|x_0)$. Estandarizando \hat{y}_0 y dividiendo entre la raíz de una chi cuadrada llegamos a una distribución t .

$$\frac{\frac{\hat{y}_0 - E(\hat{y}_0|x_0)}{\sqrt{\sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]}}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2(n-2)}}} = \frac{\hat{y}_0 - E(\hat{y}_0|x_0)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}} \sim t_{(n-2)}$$

De esta manera, nuestro intervalo al $(100)(1 - \alpha)\%$ para $E(\hat{y}_0|x_0)$ es de la forma

$$\begin{aligned} 1 - \alpha &= P\left(-t_{(n-2)}^{\alpha/2} \leq \frac{\hat{y}_0 - E(\hat{y}_0|x_0)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}} \leq t_{(n-2)}^{\alpha/2}\right) \\ &= P\left(\hat{y}_0 - t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \leq E(\hat{y}_0|x_0) \leq \hat{y}_0 + t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}\right) \end{aligned} \quad (26)$$

Mientras los valores a predecir se alejan del centro de los datos es razonable perder predictividad. Eso lo podemos ver en el intervalo de confianza, pues mientras $|x_0 - \bar{x}|$ toma valores mayores, el intervalo se amplía.

1.6.1. Predicción de nuevas observaciones (x_0, y_0)

Hasta ahora, conocemos un intervalo de confianza para el valor medio (esperanza) de y dado un valor fijo de x , dicho x_0 . Pero, dada una pareja (x_0, y_0) observada ¿pertenece a la estructura lineal? En otras palabras, queremos saber si el modelo $y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0$ sigue siendo el apropiado. Definamos

$$\eta = y_0 - \hat{y}_0 = (\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1) + \varepsilon_0 \quad (27)$$

Como y_0 y \hat{y}_0 son independientes y normalmente distribuidos, entonces η también lo es. Veamos sus propiedades.

$$E(\eta) = E(y_0 - \hat{y}_0) = E(y_0) - E(\hat{y}_0) = \hat{y}_0 - \hat{y}_0 = 0$$

$$\begin{aligned}
V(\eta) &= E(\eta - E(\eta))^2 = E(\eta)^2 \\
&= E((\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1) + \varepsilon_0)^2 \\
&= E[\varepsilon_0^2 + ((\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1))^2 + 2\varepsilon_0((\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1))] \\
&= E[\varepsilon_0^2] + E[(\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1)]^2 + 2E[\varepsilon_0((\beta_0 + \hat{\beta}_0) + x_0(\beta_1 - \hat{\beta}_1))] \\
&= \sigma^2 + E[\beta_0 - \hat{\beta}_0]^2 + E[x_0(\beta_1 - \hat{\beta}_1)]^2 + 2E[x_0(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1)] \\
&= \sigma^2 + V(\hat{\beta}_0) + x_0^2 V(\hat{\beta}_1) + 2x_0 E[(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1)] \\
&= \sigma^2 + \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} + \frac{\sigma^2 x_0^2}{\sum (x_i - \bar{x})^2} + 2x_0 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\
&= \sigma^2 + \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} + \frac{\sigma^2 x_0^2}{\sum (x_i - \bar{x})^2} - \frac{2x_0 \bar{x} \sigma^2}{\sum (x_i - \bar{x})^2} \\
&= \sigma^2 \left[1 + \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} + \frac{x_0^2}{\sum (x_i - \bar{x})^2} - \frac{2x_0 \bar{x}}{\sum (x_i - \bar{x})^2} \right] \\
&= \sigma^2 \left[1 + \frac{\sum (x_i - \bar{x})^2 + n\bar{x}^2}{n \sum (x_i - \bar{x})^2} + \frac{x_0^2 - 2x_0 \bar{x}}{\sum (x_i - \bar{x})^2} \right] \\
&= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]
\end{aligned}$$

De esta manera, podemos dar la distribución explícita de η

$$\eta \sim N\left(0, \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]\right)$$

Ahora, con avidez, somos capaces de crear un intervalo de confianza para $y_0 - \hat{y}_0 = \eta$. Análogo al caso anterior, encontramos una distribución Student.

$$\frac{\frac{\eta - 0}{\sigma \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2(n-2)}}} = \frac{\eta}{\sqrt{\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]}} = \frac{y - \hat{y}_0}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}} \sim t_{(n-2)}$$

Así, nuestro intervalo es de la forma

$$\begin{aligned}
1 - \alpha &= P\left(-t_{(n-2)}^{\alpha/2} \leq \frac{y_0 - \hat{y}_0}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}} \leq t_{(n-2)}^{\alpha/2}\right) \\
&= P\left(\hat{y}_0 - t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \leq y_0 \leq \hat{y}_0 + t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}\right)
\end{aligned} \tag{28}$$

Igual al caso anterior, mientras los valores se alejan del centro $|x_0 - \bar{x}|$, el intervalo es más grande y deja de explicar la realidad. A este hecho se le llama *PELIGRO MATEMÁTICO*, ya que es independiente la validez del modelo, es decir, aún siendo bueno el modelo se corre el riesgo de predecir en forma poco precisa. Otro peligro es el llamado *PRÁCTICO*; este consiste en que el modelo es sólo una aproximación de la realidad y nunca un modelo puede absolutamente correcto.

Aunque los intervalos (26) y (28) son muy similares, son para diferentes elementos. El primero es un intervalo de confianza para la respuesta media (esperanza) de los valores de y dado un valor de x fijo; mientras que el segundo intervalo es para un fijo de y dado un x igualmente fijo. Note que el primer intervalo siempre será más angosto, por tener menos variabilidad en las bandas Fig. (5).

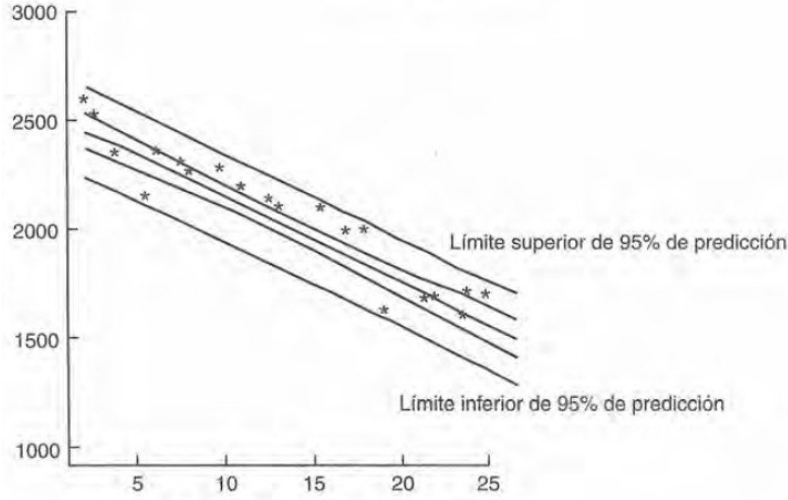


Figura 5: Intervalos de confianza para $E(y_0|x_0)$ & y_0 .

1.7. Análisis de varianza

Ya hemos definido al error e , pero se puede definir de manera más general que tanto se aleja un valor estimado de la media de los datos y del valor estimado. Suponga, para un valor x_i , un valor observado y_i y un valor estimado \hat{y}_i , además del valor medio \bar{y} para todos los valores observados de y . En la figura (6) se puede ver de manera explícita la siguiente relación

$$\begin{aligned}(y_i - \bar{y}) &= (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \\ &= (\hat{y}_i - \bar{y}) + e_i\end{aligned}$$

Si hacemos la suma para $i = 1, 2, \dots, n$ de los cuadrados, podemos llegar a

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + 2 \sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum (y_i - \hat{y}_i)^2$$

Note que

$$\begin{aligned}2 \sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= 2 \sum \hat{y}_i(y_i - \hat{y}_i) + 2\bar{y} \sum (y_i - \hat{y}_i) \\ &= 2 \sum \hat{y}_i e_i - 2\bar{y} \sum e_i = 0\end{aligned}\quad (*)$$

(*) por la propiedad [7].

Entonces, esta identidad se reduce a

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (29)$$

Note que

$$\sum (\hat{y}_i - \bar{y})^2 = \sum \hat{\beta}_1^2 S_{xx} = \hat{\beta}_1 S_{xy} = \frac{S_{xy}^2}{S_{xx}} \quad (30)$$

Estos términos son tan importantes para el estudio del modelo que reciben el nombre de la *igualdad fundamental del análisis de varianza*. $\sum (y_i - \bar{y})^2$ es la suma de cuadrados debido a la media o suma total SS_T (total sum of squares) que mide la variabilidad total de las observaciones, $\sum (\hat{y}_i - \bar{y})^2$ es la cantidad de variabilidad en las observaciones y_i explicada por la línea de regresión SS_R y $\sum (y_i - \hat{y}_i)^2$ la variación residual que queda sin explicar por la línea de regresión SS_{Res} o SS_e .

$$SS_T = SS_R + SS_e \quad (31)$$

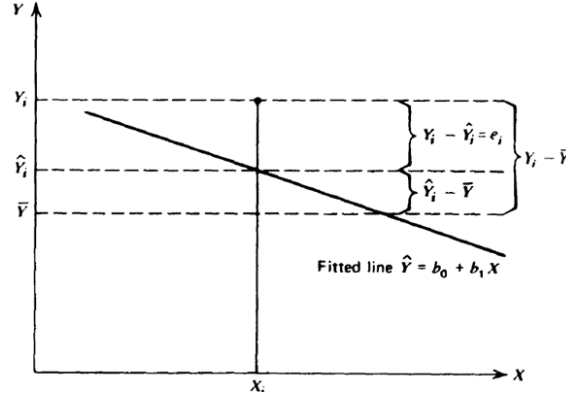


Figura 6: Suma de cuadrados.

Cada término tiene un número de grados de libertad asociado y se describe a partir de 'piezas' independientes de información involucrando los n números independientes y_1, \dots, y_n que se necesitan para completar la suma de cuadrados. SS_T tiene $(n-1)$ piezas independientes (de los números $y_1 - \bar{y}, \dots, y_n - \bar{y}$, sólo $(n-2)$ son independientes, ya que todos los n números suman cero por definición de media). Podemos calcular SS_R de una función única de y_1, \dots, y_n , llamada $\hat{\beta}_1$ (ya que $\sum (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$ como en la ec. (30)), y de esta manera sólo tiene 1 grados de libertad. Por último, SS_e tiene $(n-2)$ grados de libertad. Note como los grados de libertad son una extensión natural de la ec. (31):

$$n-1 = 1 + (n-2)$$

Esta información es útil si nos interesa conocer la significancia del modelo. Dado que β_0 casi nunca es cero, una hipótesis interesante es $H_0 : \beta_1 = 0$ contra $H_a : \beta_1 \neq 0$, donde evaluamos si la variable X realmente está explicando a Y . Para esta hipótesis conviene la prueba F del análisis de varianza Cuadro (1) que reúne la información necesaria. Sabemos que $SS_e = (n-2)MS_e$ tiene una distribución χ_{n-2}^2 ; si H_0 es cierta, entonces SS_R tiene también distribución chi-cuadrada e independiente a la anterior. Por definición de la distribución F :

$$F_0 = \frac{SS_R/1}{SS_e/(n-2)} = \frac{MS_R}{MS_e} \sim F_{1,n-2}$$

Bajo H_a , F_0 sigue una distribución F no centralizada con los mismos grados de libertad y el parámetro de no centralidad λ ;

$$\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$$

Esto nos indica que para valores grandes de F_0 se rechaza H_0 , es decir $\beta_1 \neq 0$. La regla de decisión es rechazar H_0 si $F_0 > F_{1,n-2}^\alpha$. Esta información se resume en el Cuadro (1) usualmente llamado ANOVA (analysis of variance).

Cuadro 1: Análisis de varianza de la regresión.

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	$\frac{MS_R}{MS_e}$
Residuales	$SS_e = SS_T - \hat{\beta}_1 S_{xy}$	$n - 2$	MS_e	
Total	$SS_T = \sum (y_i - \bar{y})^2$	$n - 1$		

1.7.1. Estadístico R^2

Una medida de ajuste de nuestro modelo a los datos observados es el estadístico R^2 , definido como

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{SS_R}{SS_e} \quad (32)$$

Este valor está en el rango $[0, 1]$, donde la cota inferior representa que la línea estimada no explica los datos y, por otro lado, el valor 1 representa que todas las observaciones están explicadas por la recta de regresión. Estos supuestos son teóricos, pues si $R^2 = 1$, entonces todos los puntos observados están sobre la misma recta y no existe un error aleatorio ε , de esta manera tendríamos un modelo determinista. Este cociente es la proporción de la variación total debido a la media explicado por explicado por la regresión. Una propiedad útil para evaluar este cociente es

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

1.8. Correlación entre x & y

Sea u y w dos variables aleatorias, el coeficiente de correlación se define como

$$\rho_{uw} = \frac{\text{cov}(u, w)}{[V(u)V(w)]^{1/2}}$$

Donde

$$\begin{aligned} \text{cov}(u, w) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [u - E(u)][w - E(w)]f(u, w)dudw \\ V(u) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [u - E(u)]^2 f(u, w)dudw \\ E(u) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uf(u, w)dudw \end{aligned}$$

Se puede demostrar que $-1 \leq \rho_{uw} \leq 1$. La cantidad de ρ_{uw} es una medida de la asociación lineal entre dos variables aleatorias, dichas u y w . Por ejemplo, si $\rho_{uw} = 1$, u y w están perfecta y positivamente correlacionadas, de tal manera que posibles valores de u y w ajustarían de manera perfecta sobre una línea con pendiente positiva en el plano (u, w) . Por el contrario, si $\rho_{uw} = 0$ las variables no tienen correlación lineal (esto no implica la independencia estadística). Es importante notar que el valor de la correlación entre u y w sólo es una medida de asociación lineal y no implica por si misma una relación de causalidad. Si tenemos una muestra n de la distribución conjunta $f(u, w)$, dicho $(u_1, w_1), \dots, (u_n, w_n)$, el coeficiente de correlación muestral se define como

$$r_{uw} = \frac{\sum(u_i - \bar{u})(w_i - \bar{w})}{[\sum(u_i - \bar{u})^2]^{1/2}[\sum(w_i - \bar{w})^2]^{1/2}} \quad (33)$$

Suponiendo x e y dos variables aleatorias con el modelo $y = \beta_0 + \beta_1 x + \varepsilon$ y una muestra $(x_1, y_1), \dots, (x_n, y_n)$, utilizando la ec. (33) es sencillo ver que existe una relación entre la estimación de β_1 y el factor de correlación entre x e y , de tal manera que $\hat{\beta}_1 = \frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}} r_{xy}$. Si el lector no logra discernir esta identidad, a continuación se presenta el desarrollo.

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sqrt{\sum(y_i - \bar{y})^2}}{\sqrt{\sum(y_i - \bar{y})^2}} \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\ &= \frac{\sqrt{\sum(y_i - \bar{y})^2}}{\sqrt{\sum(y_i - \bar{y})^2}} \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(x_i - \bar{x})^2}} \\ &= \frac{\sqrt{\sum(y_i - \bar{y})^2}}{\sqrt{\sum(x_i - \bar{x})^2}} \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \\ &= \frac{\sqrt{\sum(y_i - \bar{y})^2}}{\sqrt{\sum(x_i - \bar{x})^2}} \cdot r_{xy} = \frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}} \cdot r_{xy} \end{aligned}$$

De esta manera, $\hat{\beta}_1$ y r_{xy} guardan una relación, pero proveen diferentes interpretaciones. r_{xy} mide la asociación lineal entre x e y , mientras que $\hat{\beta}_1$ mide el tamaño de cambio en y que puede ser predicho para una unidad de cambio en x . Un cambio de escala en los datos afecta a $\hat{\beta}_1$, pero no a r_{xy} .

El estadístico R^2 también tiene una relación directa con el coeficiente de correlación, tanto que se puede expresar como la correlación entre las observaciones de y_i y el correspondiente valor estimado \hat{y}_i .

$$R^2 = r_{y\hat{y}}^2$$

1.9. Transformación de variables

Un punto de partida conveniente en el análisis de regresión es que el modelo que describen los datos es lineal en las variables aleatorias. Para poder llevar a cabo esto el análisis se realiza muy frecuentemente sobre

variables transformadas. La necesidad de transformar los datos surge debido a que la variable original, o el modelo en términos de la variable original, violan una o más de las suposiciones estándares. Las suposiciones que se violan con más frecuencia son las que conciernen a la linealidad del modelo y a la constancia de la varianza de los errores. Un modelo de regresión es lineal cuando los parámetros presentados en el modelo ocurren linealmente. Por ejemplo, los siguientes modelos son lineales.

$$\begin{aligned}y &= \beta_0 + \beta_1 x + \mu \\y &= \beta_0 + \beta_1 x + \beta_2 x^2 + \mu \\y &= \beta_0 + \beta_1 \ln x + \mu \\y &= \beta_0 + \beta_1 \sqrt{x} + \mu\end{aligned}$$

Debido a que los parámetros β 's en el modelo entran linealmente. Pero, $y = \beta_0 + e^{\beta_1 x} + \mu$ no es un modelo lineal dado que el parámetro β_1 no es una entrada lineal en el modelo. Para satisfacer las suposiciones del modelo de regresión estándar, en lugar de trabajar con las variables originales, algunas veces trabajamos con variables transformadas. Las transformaciones pueden ser necesarias por varias razones que se pueden resumir de la siguiente manera:

- I. Las consideraciones teóricas pueden especificar que la relación entre dos variables es no lineal. Una transformación apropiada de las variables puede hacer que la relación entre las variables transformadas sea lineal. Consideremos un ejemplo en el área teórica de aprendizaje.

Un modelo de aprendizaje que es ampliamente utilizado establece que el tiempo que toma llevar a cabo una tarea por i -ésima ocasión (T_i) es

$$T_i = \alpha \beta^i \quad \alpha > 0, 0 < \beta < 1$$

La razón entre T_i e i dada de esta manera no es lineal. No se pueden aplicar las técnicas de regresión lineal directamente. Entonces, hay que transformar. Aplicando el logaritmo en ambos lados, obtenemos

$$\ln T_i = \ln \alpha + i \ln \beta$$

La transformación nos permite usar los métodos que conocemos de regresión. A pesar de que la relación entre las variables originales era no lineal, la relación entre las variables transformadas es lineal.

- II. La variable dependiente y analizada, puede tener una distribución de probabilidad cuya varianza está relacionada con la media. Si la media está relacionada con el valor de la variable independiente x , entonces la varianza de y cambiará con x y la varianza no será constante. Usualmente la distribución de las n observaciones de y será igualmente distinta de la normal bajo esta situación invalidando las pruebas de significancia.

Por ejemplo: la varianza no constante de los errores producirá estimadores que son sesgados. En estas situaciones frecuentemente se transforman los datos para asegurar la normalidad y constancia de la varianza de los errores. En la práctica, se eligen las transformaciones para asegurar la constancia de la varianza. Por fortuna, es una coincidencia que las transformaciones que estabilizan la varianza también son buenas transformaciones normalizadoras.

1.9.1. Transformaciones para conseguir linealidad

Hay varios modelos no lineales que por transformaciones adecuadas pueden ser convertidos en lineales. A continuación se dan algunas curvas linealizables y las gráficas correspondientes.

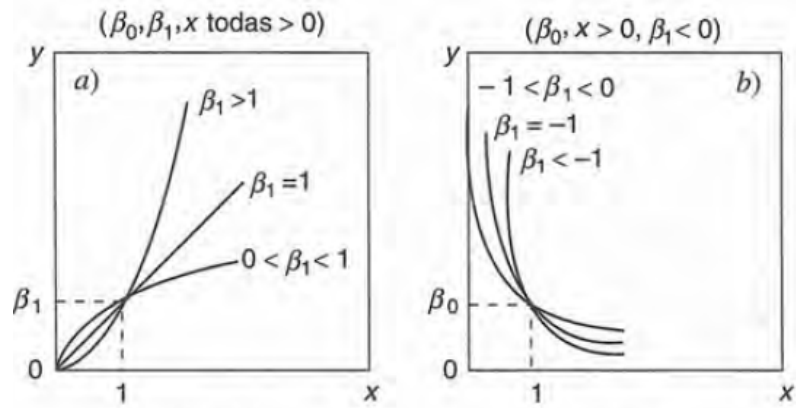


Figura 7:

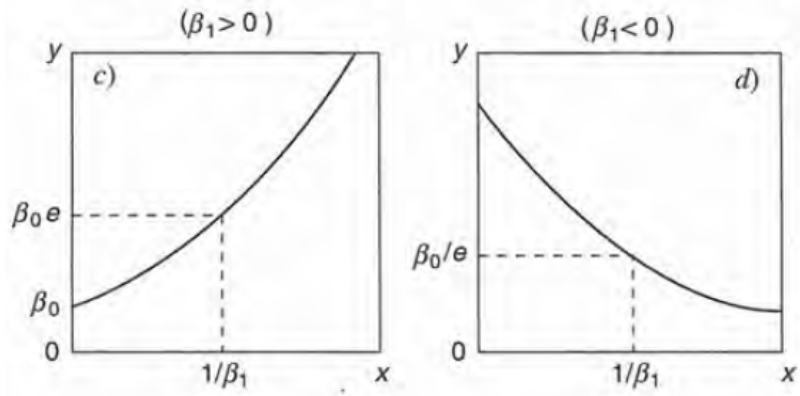


Figura 8:

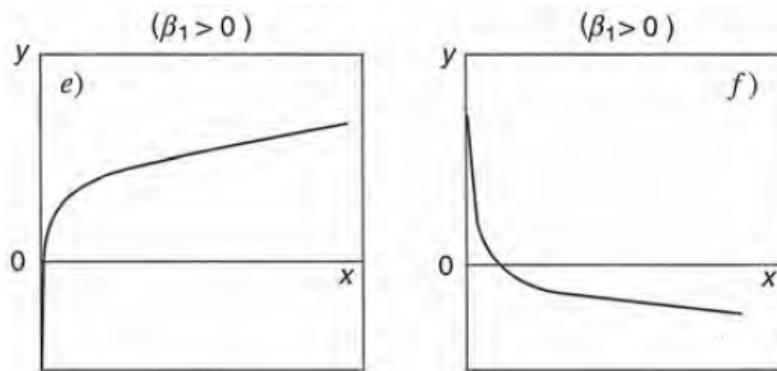


Figura 9:

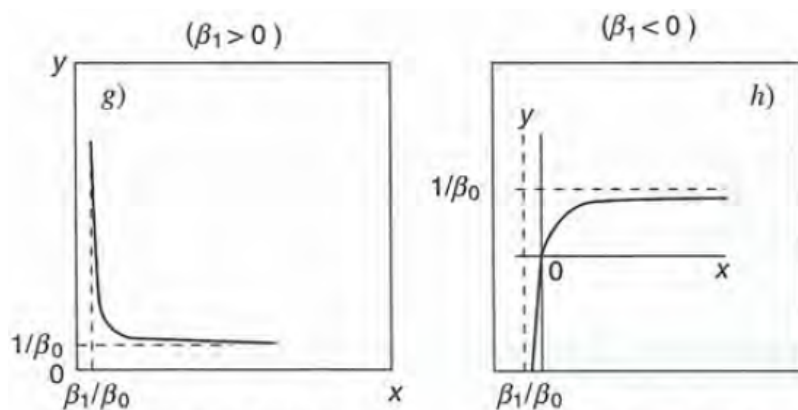


Figura 10:

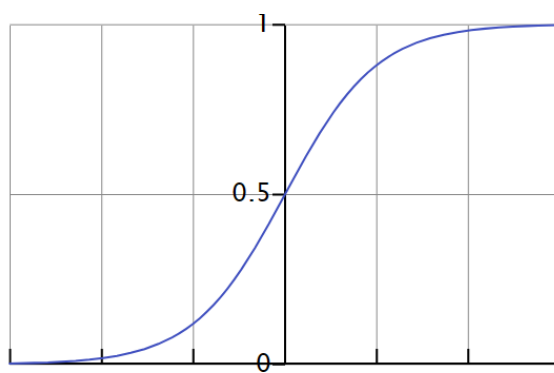


Figura 11:

Cuadro 2: Funciones linealizables.

Figura	Función inicial	Transformación	Función resultado
Fig. 7	$y = \beta_0 x^{\beta_1}$	$y^* = \log y, x^* = \log x$	$y^* = \log \beta_0 + \beta_1 x^*$
Fig. 8	$y = \beta_0 e^{\beta_1 x}$	$y^* = \ln y$	$y^* = \ln \beta_0 + \beta_1 x$
Fig. 9	$y = \beta_0 + \beta_1 \log x$	$x^* = \log x$	$y^* = \beta_0 + \beta_1 x^*$
Fig. 10	$y = \frac{x}{\beta_0 x - \beta_1}$	$y^* = \frac{1}{y}, x^* = \frac{1}{x}$	$y^* = \beta_0 - \beta_1 x^*$
Fig. 11	$y = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$	$y^* = \ln \left(\frac{y}{1-y} \right)$	$y^* = \beta_0 + \beta_1 x$

1.10. Ejercicios resueltos

Ejercicio 2.17 en [11] Para el modelo de regresión lineal simple $y = 50 + 10x + \varepsilon$, donde ε es i.i.d como $N(0, 16)$, suponer que se usan $n = 20$ pares de observaciones para ajustar este modelo. Generar 500 muestras de 20 observaciones, tomando una observación para cada valor de $x = 1, 1.5, 2, \dots, 10$ para cada muestra.

- Para cada muestra, calcular los estimados de la pendiente y la ordenada al origen por mínimos cuadrados. Trazar histogramas de los valores muestrales de $\hat{\beta}_0$ y $\hat{\beta}_1$. Comentar la forma de esos histogramas.

Se tienen que crear 500 muestras de 20 valores en x y su valor y correspondiente. Para ello, se crean 3 funciones, la primera añade la aleatoriedad a la ecuación $y = 50 + 10x + \varepsilon$, donde $\varepsilon \sim N(0, 16)$ Posterior, una función que arroje una muestra aleatoria de tamaño $2n$ con los valores (x, y) . La última función repite m cantidad de veces la muestra aleatoria de tamaño n .

Con esta última función creamos una matriz que almacene todos los valores. Posterior, calculamos los valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ para cada muestra. Por último, se presentan los histogramas. Note que presentan normalidad con centro en el valor verdadero de $\beta_0 = 50$ y $\beta_1 = 10$, tal y como se demostró en el teorema (2). Se sugiere al lector reescribir en código (en lenguaje R) y correrlo para distintos valores de n y m .

```
1 #Creamos la funcion y=50+10x+e
2 y<-function(x) {
3   valor<-50+10*x+rnorm(1,0,4)
4   return(valor)
5 }
6 #Creamos una funcion que nos arroje una muestra de tamano n*2
7 #con valores (x,y) usando la funcion anterior
8 muestra<-function(n){
9   df <- data.frame( x = rep(NA, 2*n), y = rep(NA, 2*n))
10  for (i in seq(from = 1, to = n+0.5, by = 0.5)) {
11    df[2*(i-0.5),2] <- round(y(i),4)
12    df[2*(i-0.5),1] <- i
13  }
14  return(df)
15 }
16
17 #Asi, somos capaces de crear un dataframe con m cantidad de muestras
18 #de n observaciones.
19 muestraConjunta<-function(m,n){
20   mc<-data.frame()
21   mc<-muestra(n)
22   for(i in 1:m-1){
23     mc<-cbind(mc,muestra(n))
24   }
25   return(mc)
26 }
27 #Creamos 500 muestras de tamano 10
28 muestra500<-muestraConjunta(500,10)
29
30 # a)
31 #Obtenemos los coeficientes de las 500 muestras, es decir,
32 #500 valores de beta_0 y 500 de beta_1
33 mm1<-lm(muestra500[,2]~muestra500[,1])
34 coef_estimados<-coefficients(mm1)
35 for (k in 2:500){
36   coef_estimados<-rbind(coef_estimados,
37                         coefficients(lm(muestra500[,2*k]~muestra500[,1])))
38 }
```

```

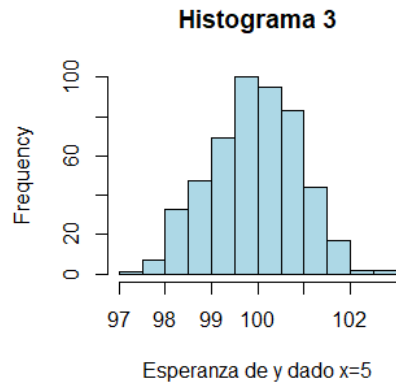
37 }
38 #Creamos el histograma de los coeficientes estimados
39 hist(coef_estimados[,1], xlab="beta_0", main = "Histograma 1",
40      col = "lightblue")
41 hist(coef_estimados[,2], xlab="beta_1", main = "Histograma 2",
42      col = "lightblue")

```



- b. Para cada muestra, calcular un estimado de $E(y|x = 5)$. Trazar un histograma de los estimados obtenidos. Comentar la forma del histograma.

El mejor estimador para $E(y|x = 5)$ es, precisamente, $\hat{\beta}_0 + \hat{\beta}_1 \cdot 5$, con valor central en $\beta_0 + \beta_1 \cdot 5$. Conocemos los valores reales de los estimadores, entonces $y = 50 + 10 \cdot 5 = 100$. El histograma 3 presenta la frecuencia de los valores estimados para las 500 muestras con aparente normalidad y centro en el valor 99.93, muy cercano al valor real de 100. Este ejercicio es muy ilustrativo al presentar de manera empírica las propiedades de normalidad y valor esperado de los parámetros estimados y de la esperanza de la variable de respuesta dado un valor en x . Al hacer la comparación de los estimadores (inciso a) y del valor esperado (inciso b) con los valores reales de la ecuación $y = 50 + 10x$ podemos ver una aproximación asertiva, como supusimos de manera teórica.



```

1 # b)
2 #Calculamos la esperanza de y dado x=5 para las 500 muestras
3 Esp_y_x5<-numeric(length = 500)
4 for (k in 1:500){
5   Esp_y_x5[k]<-coef_estimados[k,1]+coef_estimados[k,2]*5
6 }
7 #creamos el histograma
8 hist(Esp_y_x5, xlab = "Esperanza de y dado x=5",
9       main = "Histograma 3", col = "lightblue")

```

- c. Determinar un intervalo de confianza de 95 % para la pendiente en cada muestra. ¿Cuántos de los intervalos contienen el valor verdadero $\beta_1 = 10$? ¿Es lo que se esperaba?

Cuando establecemos un intervalo de confianza con un $\alpha = 0.05$ estimamos que el 95 % de las veces nuestro intervalo realmente contiene al verdadero valor del parámetro. Si aquí creamos 500 intervalos de confianza para β_1 en cada muestra, al menos 95 % de ellos deberían contener al valor de 10. Nuevamente, se puede verificar de manera empírica que el 95.2 % de los intervalos creados tiene el valor de 10.

```

1 # c)
2 #Creamos un intervalo de confianza para cada muestra
3 IC<-data.frame()
4 IC<-confint((lm(muestra500[,2]~muestra500[,1])),
5             muestra500[4,1], level = 0.95)
6 for (k in 2:500){
7   IC<-rbind(IC,confint((lm(muestra500[,2*k]~muestra500[,1])),
8                         muestra500[4,1], level = 0.95))
9 }
10 IC
11
12 #contabilizamos cuantos intervalos contienen al valor 10
13 booleano_IC10<-numeric(length=500)
14 for (k in 1:500){
15   booleano_IC10[k]<-(10 >= IC[k,1] && 10 <= IC[k,2])
16 }
17 #calculamos el cociente de los intervalos que contienen al verdadero
18 #valor 10 de beta_1, entre el total de intervalos
19 sum(booleano_IC10)/500 #0.952

```

- d. Para cada estimado de $E(y|x = 5)$ en la parte b, calcular el intervalo de confianza de 95%.
¿Cuántos de esos intervalos contienen el valor verdadero de $E(y|x = 5) = 100$? ¿Es lo que se esperaba?

A esta altura, el lector debería ser capaz de resolver este ejercicio sin mayor esfuerzo.

Ejercicio 2.12 en [11] Se cree que la cantidad de libras de vapor usadas en una planta por mes está relacionada con la temperatura ambiente promedio. A continuación se presentan los consumos y las temperaturas del último año.

Mes	Temperatura	Uso/1000
Ene.	21	185.79
Feb.	24	214.47
Mar.	32	288.03
Abr.	47	424.84
Mayo	50	454.68
Jun.	59	539.03
Jul.	68	621.55
Ago.	74	675.06
Sep.	62	562.03
Oct.	50	452.93
Nov.	41	369.95
Dic.	30	273.98

- a) Ajustar el modelo de regresión lineal simple a los datos.

Supongamos que la variable y =Uso de vapor depende de x =Temperatura ambiente promedio. Es interesante ver que matemáticamente es igual si invertimos la asignación de las variables. Sin embargo, en el contexto del problema no tiene sentido pensar que la temperatura ambiente depende de la producción de una única planta.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = 9.2084$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = -6.3320$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = 3.7854$$

- b) Probar la significancia de la regresión.

Para probar la significancia nos ayudaremos de la tabla ANOVA. Note que el valor del estadístico F_0 es tan grande es casi imposible no rechazar la hipótesis nula de no significancia. Esto se verá más claro en el gráfico adjunto al final del ejercicio.

- c) En la administración de la planta se cree que un aumento de 1 grado en la temperatura ambiente promedio hace aumentar 10 000 libras en el consumo mensual de vapor. ¿Estos datos respaldan la afirmación?

Recuerde que se hizo un escalamiento sobre mil. De esta manera, contestar a esta pregunta es análogo a la prueba de hipótesis $H_0 : \beta_1 = 10$ vs $H_a : \beta_1 \neq 10$. Supongamos un nivel de

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	$SS_R = 280589.6$	1	280589.6	74124.16
Residuales	$SS_e = 37.8547$	10	3.7854	
Total	$SS_T = 280627.4$	11		

confianza $\alpha = 0.05$ Podemos utilizar el estadístico t_0

$$t_0 = \frac{\hat{\beta}_1 - 10}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{-0.7916}{\sqrt{3.7854/3309}} = -23.4044$$

mientras que nuestro valor crítico es $t_{n-2}^{\alpha/2} = t_{10}^{0.025} = 1.8124$. Como $|t_0| > t_{10}^{0.025}$, podemos rechazar la hipótesis nula, es decir, no existe suficiente evidencia estadística para afirmar que la tasa de cambio $\beta_1 = 10$, tal que al aumentar una unidad el grado de la temperatura promedio no hace aumentar en diez mil libras el consumo mensual de vapor. Este modelo es muy estricto debido al alto nivel de ajuste. Si se calcula el valor $R^2 = 0.999$ se puede ver que existe un ajuste muy preciso. Tal parece que los datos siguen más una ley determinista que un modelo de probabilidad. Incluso si se intentara hacer la misma prueba para un valor $\beta_1 = 9$ también rechazaríamos la hipótesis nula.

- d) Determina un intervalo de predicción de 99 % para el uso de vapor en un mes con temperatura ambiente promedio de 58 %.

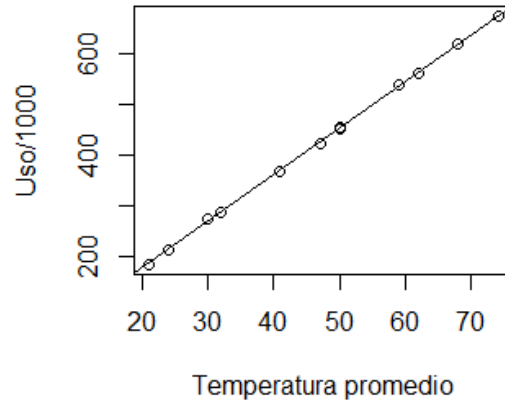
Recuerde que se puede hacer un intervalo de confianza para el valor medio de respuesta dado un valor fijo de x_0 y también un intervalo para un valor fijo de respuesta y_0 dado x_0 igual fijo. Considerando un nivel de significancia del 99 %, entonces $\alpha = 0.01$. Para el segundo caso, el intervalo de confianza es de la forma

$$\left(\hat{y}_0 - t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \leq y_0 \leq \hat{y}_0 + t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

$$\left(527.7552 - 3.1692 \sqrt{3.7854 \left(1 + \frac{1}{12} + \frac{132.25}{3309} \right)} \leq y_0 \leq 527.7552 + 3.1692 \sqrt{3.7854 \left(1 + \frac{1}{12} + \frac{132.25}{3309} \right)} \right)$$

$$(521.2200 \leq y_0 \leq 534.2903)$$

Un intervalo estrecho considerando la magnitud de los datos.



Ejemplo 11.9 en [14] En su tesis para obtener el doctorado, H. Behbahani estudió el efecto de la variación de la razón agua/cemento en la resistencia del concreto después de 28 días. Para el concreto que contiene 200 libras por yarda cúbica de cemento obtuvo los datos que se presentan en la siguiente tabla. Sea y la resistencia y x la razón de agua/cemento.

Agua/Cemento	Resistencia
1.21	1.302
1.29	1.231
1.37	1.061
1.46	1.040
1.62	0.803
1.79	0.711

a) Ajuste el modelo $E(y) = \beta_0 + \beta_1 x$.

Apliquemos las fórmulas para obtener los estimadores.

$$S_{xy} = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i = 8.709 - \frac{1}{6} (8.74)(6.148) = -0.247$$

$$S_{xx} = \sum x_i^2 - \frac{1}{n} (\sum x_i)^2 = 12.965 - \frac{1}{6} (8.74)^2 = 0.234$$

$$S_{yy} = \sum y_i^2 - \frac{1}{n} (\sum y_i)^2 = 6.569 - \frac{1}{6} (6.148)^2 = 0.269$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-0.247}{0.234} = -1.056$$

y

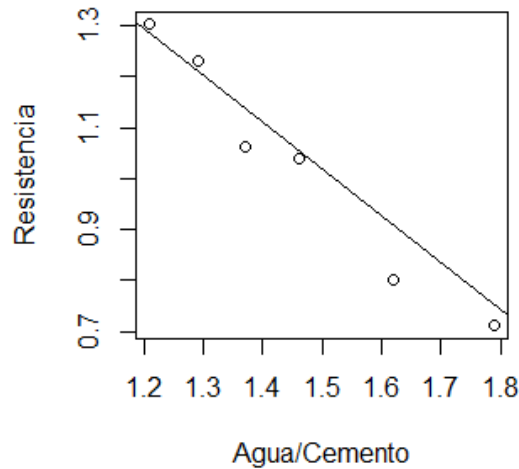
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{6.148}{6} - (-1.056) \left(\frac{8.74}{6} \right) = 2.563$$

(Para este ejemplo se llevaron a cabo los cálculos con tres decimales). Por lo tanto, el modelo de línea recta que mejor ajusta a los datos es

$$\hat{y} = 2.563 - 1.056x$$

Además, calculemos el nivel de ajuste del modelo R^2 .

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{(-0.247)^2}{(0.234)(0.269)} = 0.969$$



- b) Pruebe $H_0 : \beta_0 = 0$ frente a $H_a : \beta_1 < 0$ con $\alpha = 0.05$. (Advierta que si rechazamos H_0 , concluimos que $\beta_1 < 0$, y que la resistencia tiene a disminuir con un incremento en la razón agua/cemento). Obtenga el nivel de significancia correspondiente alcanzado.

Como deseamos probar si hay evidencia de que $\beta_1 < 0$ con $\alpha = 0.05$, el estadístico de prueba apropiado es

$$t_0 = \frac{\hat{\beta}_1 - 0}{\hat{\sigma} \sqrt{1/S_{xx}}}$$

donde

$$\hat{\sigma} = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n - 2}} = \sqrt{\frac{0.269 - (-1.056)(-0.247)}{4}} = \sqrt{\frac{0.008}{4}} = 0.045$$

Por lo tanto, el valor del estadístico de prueba apropiado para $H_0 : \beta_1 = 0$ frente a $H_a : \beta_1 < 0$ es

$$t_0 = \frac{-1.056 - 0}{0.045 \sqrt{1/0.234}} = -11.355$$

Como el estadístico anterior se basa en $n-2=4$ grados de libertad y la región de rechazo apropiada es $t < -t_{0.05} = -2.132$, rechazamos H_0 a favor de H_a a un nivel de significancia $\alpha = 0.05$. Como la prueba apropiada es de cola inferior, el valor p es $p = P(t < -11.355)$, donde t tiene una distribución con 4 grados de libertad. Por tanto, $p < 0.05$. De hecho, tablas más extensas de la distribución t indican que el valor p es considerablemente menor que 0.005, en realidad es de 0.00017. En consecuencia, para los valores de $\alpha > 0.00017$ que se usan por lo común común concluimos que hay evidencia suficiente para indicar que la resistencia disminuye con un incremento en la razón agua/cemento en la región donde se realizó el experimento. Desde un punto de vista práctico, la razón agua/cemento debe ser suficiente grande para humedecer el cemento, la arena y otros elementos que forman el concreto. Pero si la razón agua/cemento aumenta demasiado, el concreto no servirá.

- c) Encuentre un intervalo de confianza al 90% de la resistencia esperada del concreto cuando la razón agua/cemento es de 1.5. ¿Qué pasará con el intervalo de confianza si tratamos de estimar la resistencia media para razones de agua/cemento de 0.3 o 2.7?

El intervalo de confianza se puede obtener mediante la fórmula

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}}$$

Queremos un intervalo de confianza cuando $x_0 = 1.5$; por tanto, el intervalo es de la forma

$$\left(0.979 - (2.132)(0.045) \sqrt{\frac{1}{6} + \frac{1.5 - 1.457)^2}{0.234}}, 0.979 + (2.132)(0.045) \sqrt{\frac{1}{6} + \frac{1.5 - 1.457)^2}{0.234}} \right) \\ = (0.938, 1.020)$$

Por tanto, estimaríamos la resistencia media del concreto con una razón agua/cemento de 1.5 entre 0.938 y 1.020. A partir de la expresión de la varianza, podemos ver que el intervalo de confianza se vuelve más grande conforme x_0 se aleja de $\bar{x} = 1.457$. Además, los valores $x_0 = 0.3$ y $x_0 = 2.7$ están lejos de los valores que se utilizaron en el experimento. Hay que ser muy cauteloso antes de construir un intervalo de confianza para $E(y)$ cuando los valores de x_0 se alejan de la región de experimentación. Razones de agua/cemento de 0.3 y 2.7 quizá producirían un concreto completamente inservible.

Ejemplo 11.10 en [14] En los datos de la tabla siguiente, W denota el peso (en libras) y l la longitud (en pulgadas) de la parte posterior de la cabeza a la punta de la nariz de 15 caimanes capturados en Florida. Como es más fácil observar l que W en caimanes en su hábitat natural, se desea construir un modelo que relacione el peso con la longitud. Se puede utilizar tal modelo para predecir el peso de los caimanes de longitudes específicas. Ajuste el modelo

$$\ln W = \ln \alpha_0 + \alpha_1 \ln l + \varepsilon = \beta_0 + \beta_1 x + \varepsilon$$

x=lnL	y=lnW
3.87	4.87
3.61	3.93
4.33	6.46
3.43	3.33
3.81	4.38
3.83	4.7
3.46	3.5
3.76	4.5
3.5	3.58
3.58	3.64
4.19	5.9
3.78	4.43
3.71	4.38
3.73	4.42
3.78	4.25

Comencemos calculando las cantidades que se aplican en forma rutinaria en nuestra solución.

$$S_{xy} = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i = 251.9757 + \frac{1}{15} (56.37)(66.27) = 2.933$$

$$S_{xx} = \sum x_i^2 - \frac{1}{n} (\sum x_i)^2 = 212.6933 - \frac{1}{15} (56.37)^2 = 0.8548$$

$$S_{yy} = \sum y_i^2 - \frac{1}{n} (\sum y_i)^2 = 303.0409 - \frac{1}{15} (66.27)^2 = 10.26$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{2.933}{0.8548} = 3.4312$$

y

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{66.27}{15} - (3.4312) \left(\frac{56.37}{15} \right) = -8.476$$

Ahora podemos estimar α_0 mediante

$$\hat{\alpha}_0 = e^{\hat{\beta}_0} = e^{-8.476} = 0.0002$$

y α_1 con $\hat{\alpha}_1 = \hat{\beta}_1$ para obtener el modelo estimado

$$\hat{w} = \hat{\alpha}_0 l^{\hat{\alpha}_1} = (0.0002) l^{3.4312}$$

En muchos casos α_1 estará cerca de 3, ya que el peso o el volumen generalmente son proporcionales al cubo de una medición lineal.

Para estos datos $SSE=0.1963$, $n=15$, $\hat{\sigma} = \sqrt{SSE/(n-2)} = 0.123$. Los cálculos realizados para obtener estos valores numéricos son análogos a los cálculos del ejemplo anterior (11.9 en [14]).

Para encontrar un intervalo de predicción de W donde $x = \ln l = 4$, primero debemos construir un intervalo de predicción para $y = \ln W$. Como antes, el intervalo de predicción es

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{0.05} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

donde $t_{0.05}$ tiene $n - 2 = 13$ grados de libertad. Por lo tanto, $t_{0.05} = 1.771$ y el intervalo de predicción al 90 % para $y = \ln W$ es

$$-8.476 + 3.4312(4) \pm 1.771(0.123) \sqrt{1 + \frac{1}{15} + \frac{(4 - 3.758)^2}{0.8548}}$$

o bien

$$(5.0167, 5.4809)$$

Como $\hat{y} = \ln \hat{W}$, podemos predecir W mediante $e^{\hat{y}} = e^{5.2488} = 190.3377$. El intervalo de predicción de 90 % observado para W es

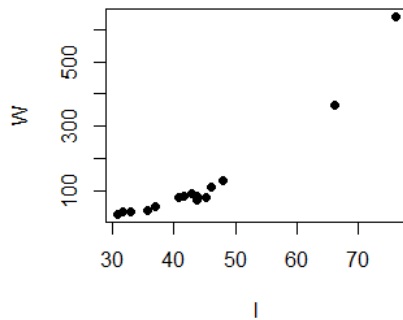
$$(e^{5.0167}, e^{5.4809})$$

o bien

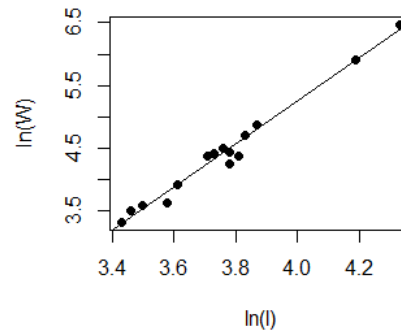
$$(150.9125, 240.0627)$$

Cuando $x = \ln x = 4$, entonces $l = e^4 = 54.598$. Por tanto, para un caimán con 54.598 pulgadas de longitud, predecimos que su peso está entre 150.91 y 240.06 libras. El intervalo hasta cierto punto estrecho en la escala logarítmica natural se vuelve muy grande cuando se transforma a la escala original.

variables sin transformación



variables transformadas



Ejemplo 6.2 en [12] Los estudiantes de la clase de estadística sugieren que hacer tarea no les ha ayudado a prepararse para el examen de medio curso. Los puntajes del examen y y los puntajes de la tarea x para los 18 estudiantes en la clase son los siguientes:

y	x
95	96
80	77
0	0
0	0
79	78
77	64
72	89
66	47
98	90
90	93
0	18
95	86
35	0
50	30
72	59
55	77
75	74
66	67

Primero, obtenemos los coeficientes estimados

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{81195 - 18(58.056)(61.389)}{80199 - 18(58.056)^2} = 0.8726$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 61.389 - (0.8726)(58.056) = 10.73$$

Así, la ecuación está dada por

$$\hat{y} = 10.73 - 0.8726x$$

En el gráfico es sencillo ver que la pendiente $\hat{\beta}_1$ es la tasa de cambio en \hat{y} a la variación en x y el intercepto $\hat{\beta}_0$ es el valor de \hat{y} para $x = 0$.

La línea aparente de tendencia no establece causalidad entre la tarea y los resultados del examen. La asunción de varianza constante $V(\varepsilon_i) = \sigma^2$ para todo $i = 1, 2, \dots, 18$ parece ser razonable.

¿Por qué calculamos los residuales al cuadrado? Como e_i es un estimador de ε_i y $E(\varepsilon_i) = 0$, se esperaría que el estimador insesgado de ε , llamado \bar{e} , también se aproxime a cero. En este ejemplo podemos ver como

$$\bar{e} = \frac{\sum e_i}{n} = \frac{1.16 \cdot 10^{-14}}{18} = 6.44 \cdot 10^{-16} \approx 0$$

Por otro lado, calculemos el estimador $\hat{\sigma}^2$ para, posteriormente, obtener pruebas de hipótesis e intervalos de confianza.

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{3071.229}{16} = 191.9518$$

Podemos probar la hipótesis $H_0 : \beta_1 = \beta_1^*$ vs $H_a : \beta_1 \neq \beta_1^*$ para cualquier valor de $\beta_1^* \in R$. Sin embargo, suele resultar interesante para $\beta_1^* = 0$. ¿Qué pasaría si $\beta_1^* = \hat{\beta}_1$? Nunca rechazaríamos la hipótesis nula

$H_0 : \beta_1 = \hat{\beta}_1$, pues nuestro estadístico quedaría de la forma

$$t_0 = \frac{\hat{\beta}_1 - \hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = 0.$$

Como la distribución t es centrada en cero, t_0 nunca caerá dentro de la región de rechazo para un $\alpha \neq 0.5$ (que es análogo a una apuesta lanzando una moneda). Entonces, hagamos la prueba de hipótesis para $\beta_1^* = 0$

$$t_0 = \frac{0.8726}{\sqrt{191.9518/19530.94}} = 8.8019$$

Al cual le corresponde un p-value de $1.57 \cdot 10^{-7} \approx 0$. Es decir, rechazamos la hipótesis nula casi para cualquier valor de α . Por lo tanto, nuestro parámetro β_1 no toma el valor cero (la variable x si tiene influencia sobre y de tal manera que es significativa). Vimos, por el consiente de verosimilitudes, que también es posible probar la significancia mediante una prueba F . Esta prueba toma más relevancia en el caso multiparamétrico, donde la hipótesis es sobre la significancia del modelo completo. Por ahora, probaremos el mismo juego de hipótesis $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. El estadístico es

$$F_0 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\hat{\sigma}^2} = 77.4832$$

Con un p-value de $1.57 \cdot 10^{-7}$. Podemos concluir lo mismo que en la prueba t antes realizada.

Ahora, creemos intervalos de confianza para los parámetros β_0 , β_1 y σ^2 con un $\alpha = 0.1$. Recordemos que el primer intervalo es de la forma

$$\left(\hat{\beta}_0 - t_{n-2}^{\alpha/2} \frac{\hat{\sigma} \sqrt{\sum x_i^2}}{\sqrt{nS_{xx}}} < \beta_0 < \hat{\beta}_0 + t_{n-2}^{\alpha/2} \frac{\hat{\sigma} \sqrt{\sum x_i^2}}{\sqrt{nS_{xx}}} \right)$$

Entonces

$$\left(10.73 - 1.7458 \frac{\sqrt{(191.9518)(80199)}}{\sqrt{18(19530.94)}} < \beta_0 < 10.73 + 1.7458 \frac{\sqrt{(191.9518)(80199)}}{\sqrt{18(19530.94)}} \right)$$

$$(-0.8261 < \beta_0 < 22.2799)$$

Para β_1

$$\left(\hat{\beta}_1 - t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} < \beta_1 < \hat{\beta}_1 + t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \right)$$

$$\left(0.8726 - 1.7458 \sqrt{\frac{191.9518}{19530.94}} < \beta_1 < 0.8726 + 1.7458 \sqrt{\frac{191.9518}{19530.94}} \right)$$

$$(0.6995 < \beta_1 < 1.0456)$$

Para σ^2 , utilizando una distribución de cola superior para chi-cuadrada

$$\begin{aligned} \left(\frac{\hat{\sigma}^2(n-2)}{\chi_{\alpha/2, n-2}^2} < \sigma^2 < \frac{\hat{\sigma}^2(n-2)}{\chi_{1-\alpha/2, n-2}^2} \right) \\ \left(\frac{191.9518(16)}{26.2962} < \sigma^2 < \frac{191.9518(16)}{7.9616} \right) \\ (116.7969 < \sigma^2 < 385.7552) \end{aligned}$$

Recordemos, de igual manera, los intervalos para la respuesta esperada de y dado un valor fijo de x y, casi igual, un valor puntual de y dado x . Supongamos que nos interesa saber el valor esperado del puntaje en el examen cuando los alumnos obtienen 80 puntos en la tarea $x = x_0 = 80$ a un $\alpha = 0.05$. Primero, vea que $\hat{y}_0 = 10.7269 + 0.8726(80) = 80.5349$. Entonces

$$\begin{aligned} \left(\hat{y}_0 - t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \leq E(\hat{y}_0|x_0) \leq \hat{y}_0 + t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \right) \\ [80.5349 - 2.1199 \sqrt{(191.9518) \left(\frac{1}{16} + \frac{(80 - 58.0555)^2}{19530.94} \right)} \leq E(\hat{y}_0|x_0) \leq \\ 80.5349 + 2.1199 \sqrt{(191.9518) \left(\frac{1}{16} + \frac{(80 - 58.0555)^2}{19530.94} \right)}] \end{aligned}$$

Esto es

$$(71.8640 \leq E(\hat{y}_0|x_0) \leq 89.5057)$$

Recuerde que con forme x_0 se aleja del centro de los datos \bar{x} , el intervalo se hace más grande y menos preciso.

De manera muy similar calculamos el intervalo para el valor puntual y_0

$$\begin{aligned} \left(\hat{y}_0 - t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \leq y_0 \leq \hat{y}_0 + t_{(n-2)}^{\alpha/2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \right) \\ (40.3769 \leq y_0 \leq 120.6928) \end{aligned}$$

Note como la imprecisión aumenta aún más si queremos estimar un valor puntual en y .

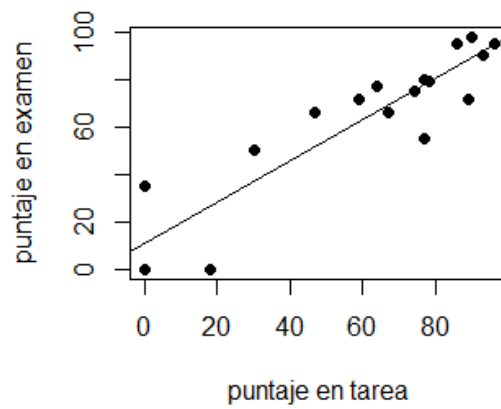
Casi por último, resulta útil medir el nivel de ajuste de nuestro modelo mediante R^2 .

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{14873.05}{17944.28} = 0.8288$$

Un nivel alto considerando que el tope es la unidad. Esto nos dice que gran parte de la variabilidad del modelo está siendo explicado por la línea estimada.

Por último, si x es una variable aleatoria, es posible calcular el coeficiente de correlación lineal entre x e y .

$$\hat{\beta}_1 = \frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}} r_{xy} \Rightarrow r_{xy} = \hat{\beta}_1 \frac{\sqrt{S_{xx}}}{\sqrt{S_{yy}}} = (0.8726) \sqrt{\frac{19530.94}{17944.28}} = 0.9110$$



Ejercicio 6.14 en [12] La siguiente tabla (Weisberg 1985, p. 231) proporciona los datos de las erupciones diurnas del géiser Old Faithful en el parque nacional de Yellowstone entre el 1 y el 4 de agosto de 1978. Las variables x =duración de la erupción e y =intervalo hasta la siguiente erupción. ¿ x puede ser usado para predecir satisfactoriamente y a través de un modelo de regresión lineal simple $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$?

x	y	x	y
4.4	78	4.5	76
3.9	74	3.9	82
4	68	4.3	84
4	76	2.3	53
3.5	80	3.8	86
4.1	84	1.9	51
2.3	50	4.6	85
4.7	93	1.8	45
1.7	55	4.7	88
4.9	76	1.8	51
1.7	58	4.6	80
4.6	74	1.9	49
3.4	75	3.5	82
4.3	80	4	75
1.7	56	3.7	73
3.9	80	3.7	67
3.7	69	4.3	68
3.1	57	3.6	86
4	90	3.8	72
1.8	42	3.8	75
4.1	91	3.8	75
1.8	51	2.5	66
3.2	79	4.5	84
1.9	53	4.1	70
4.6	82	3.7	79
2	51	3.8	60
		3.4	86

a) Encuentre $\hat{\beta}_1$ y $\hat{\beta}_0$.

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{1}{n} (\sum x_i \sum y_i)}{\sum x_i^2 - \left(\frac{1}{n} \sum x_i\right)^2} = \frac{601.1509}{52.8818} = 11.3678$$

Por consiguiente

$$\hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 = 71.1320 - 3.4641(11.3678) = 31.7523$$

b) Haga la prueba de significancia $H_0 : \beta_1 = 0$.

Estableciendo $H_a : \beta_1 \neq 0$. El estadístico de prueba es

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{11.3678}{\sqrt{55.3786 / 52.8818}} = 11.1085$$

Cuyo p-value es casi cero. Es decir, rechazamos la hipótesis nula para casi cualquier valor de α , lo que nos lleva a la afirmación de que existe suficiente evidencia estadística para suponer que el parámetro β_1 es significativo (distinto de cero). Esta muestra es lo suficiente grande para hacer una aproximación normal, de tal manera que llegaríamos a la misma conclusión.

c) Encuentre un intervalo de confianza para β_1 .

Recuerde que el intervalo es de la forma

$$\left(\hat{\beta}_1 - t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} < \beta_1 < \hat{\beta}_1 + t_{n-2}^{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \right)$$

Utilizando un $\alpha = 0.05$, entonces $t_{n-2}^{\alpha/2} = 2.0075$. Además, ya sabemos que $\hat{\sigma}^2 = 55.3786$ y $S_{xx} = 52.8818$. De esta manera, nuestro intervalo es

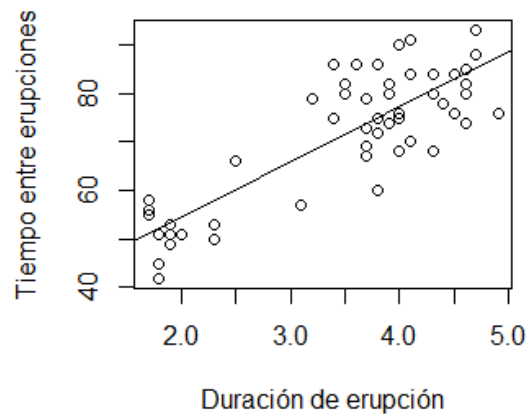
$$\left(11.3678 - (2.0075) \sqrt{\frac{55.3786}{52.8818}} < \beta_1 < 11.3678 + (2.0075) \sqrt{\frac{55.3786}{52.8818}} \right)$$

$$(9.3134 < \beta_1 < 13.4221)$$

d) Encuentre R^2 .

Por último

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{6833.766}{9658.075} = 0.7075$$



2. Regresión lineal múltiple

2.1. Modelo Lineal General

El tema de Modelo lineal General (LM) involucra los modelos de Regresión, Análisis de varianza (ANOVA) y los de Análisis de Covarianza (ANCOVA). Estos modelos tienen supuestos que deben de comprobarse una vez ajustado el modelo, los cuales son:

- i) Independencia entre las y_i por lo que también lo serán los residuales del modelo.
- ii) Linealidad entre la variable respuesta con respecto a la(s) variable(s) explicativa(s).
- iii) Normalidad en los residuos, y que cumplan que su media sea cero y σ^2 constante para todas las observaciones.
- iv) Homocedasticidad. Las varianzas tienen que ser homogéneas en los diferentes factores e iguales.

2.1.1. Modelo de Regresión Lineal

Sea Y la variable respuesta (variable dependiente) que está relacionada con las p variables explicativas (variables independientes) X_1, X_2, \dots, X_p por una función f . Tanto la variable respuesta como las explicativas son continuas y debido a que esta relación no es exacta, se escribe de la siguiente forma.

$$Y = f(X_1, X_2, \dots, X_p) + \varepsilon \quad (34)$$

donde ε es un error aleatorio.

Las Y y las X 's se observan sobre n individuos. Cuando f es lineal, la ecuación (34) observada en un individuo con $i=1, 2, \dots, n$ se denota como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (35)$$

y se llama **modelo de regresión lineal**; los parámetros $\beta_j, j = 0, 1, \dots, p$, se llaman **coeficientes de regresión**, los cuales se interesan estimar.

Escribiendo el modelo lineal (35) en forma de matricial, tenemos que con $i = 1, 2, \dots, n$. y $j = 0, 1, \dots, p$.

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{i1} & x_{i2} & x_{ij} & x_{ip} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Y una forma más compacta es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (36)$$

donde \mathbf{Y} es un vector aleatorio de dimensión $n \times 1$,
 \mathbf{X} es llamada la *matriz de diseño* de $n \times (p + 1)$ no aleatoria,
 β es un vector de parámetros desconocidos $(p + 1) \times 1$,
 ε es vector de errores aleatorio de dimensión $n \times 1$.
 Los supuestos sobre el vector de errores se expresan como

$$E(\varepsilon) = 0 \text{ y } Var(\varepsilon) = \sigma^2 I,$$

donde I es la matriz identidad de dimensión $(n \times n)$, los errores tienen varianza constante y no están correlacionadas. La esperanza de \mathbf{Y} es

$$E(\mathbf{Y}) = E(\mathbf{X}\beta + \varepsilon) = E(\mathbf{X}\beta) + E(\varepsilon) = E(\mathbf{X}\beta) = \mathbf{X}\beta$$

y la varianza de \mathbf{Y} es

$$Var(\mathbf{Y}) = Var(\mathbf{X}\beta + \varepsilon) = Var(\mathbf{X}\beta) + Var(\varepsilon) = Var(\varepsilon) = \sigma^2 I.$$

La función para realizar una regresión lineal en el paquete R es `lm` (“modelo lineal”), en [3] se detalla el uso de la función.

2.1.2. Análisis de Varianza (ANOVA)

En la modelación estadística se desea conocer el efecto de una o más variables explicativas sobre una respuesta (continua). Las variables explicativas que pueden ser controladas en un experimento reciben el nombre de *factores* y el nivel de intensidad de un factor se le denomina *nivel* del factor. Si existe un solo factor, a los niveles de ese factor se le llaman tratamientos (T) [3], [8].

Cuando la(s) variable(s) explicativa(s) son categóricas en vez de continuas y se desea hacer comparaciones entre los niveles del factor entonces se tiene un problema de análisis de varianza (ANOVA).

Cuando se modela un solo factor, la expresión:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon, \quad (37)$$

se puede escribir como:

$$Y = \mu + \beta \text{factor} + \varepsilon. \quad (38)$$

Los valores tanto de Y como del factor se observan sobre los n individuos.

En el ANOVA implica el cálculo de la variación total en la variable respuesta y la partición de ella en componentes informativos. Para el caso univariado, dividimos la variación total (SSY) en sólo dos componentes: la variación explicada (SSA) y la variación no explicada (SSE), es decir $SSY = SSA + SSE$. El Cuadro 3 expresa a detalle la información de las sumas de cuadrados.

Supongamos que hay m repeticiones de cada tratamiento y existen k niveles en el factor. Para identificar **los grados de libertad de SSE**, se tiene que hay m repeticiones de cada tratamiento por lo que se necesitan $(m - 1)$ parámetros para estimarlos y como hay k niveles de cada factor, se concluye que hay $k \times (m - 1)$ grados de libertad para el error en el experimento.

La componente de variación entre los tratamientos se representa en SSA, la suma de cuadrados de los tratamientos, explica las diferencias entre las medias de los tratamientos en promedio. Cuando se tienen dos o más variables categóricas independientes, la SSB denotará la suma de cuadrados que se atribuye a las diferencias entre las medias del segundo factor, SSC se denotará la suma de cuadrados que se atribuye al

tercer factor, así sucesivamente hasta tener todos los factores incluidos.

Generalmente esta información se presenta en la tabla de Análisis de varianza llamada Anova Tabla [38], ésta tiene seis columnas, de izquierda a derecha se describen sus características, la primera columna es la fuente de variación, le sigue la suma de cuadrados que depende de la fuente, le sigue los grados de libertad para la fuente, cuadrados medios de la fuente (varianza) y la proporción F donde se esta probando una hipótesis, la cual es

H_0 : la fuente de variación no es significativamente distinta de cero vs $H_1: \neg H_0$, el valor de p asociado a F (si $p < 0.05$ se rechaza H_0). Los cuadrados medios se obtienen haciendo el cociente de la suma de cuadrados con sus respectivos grados de libertad. La varianza del error, s^2 también es llamada cuadrado medio residual o cuadrado medio de la variación no explicada o varianza del error no agrupado debido a se calcula por medio de todos los tratamientos.

Cuadro 3: Sumas de Cuadrados ANOVA unifactorial.

La suma total de cuadrados, SSY , es la suma de los cuadrados de las diferencias entre los puntos de datos, y_{ij} , y la media general, \bar{y} .

$$SSY = \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y})^2$$

donde $\sum_{j=1}^m y_{ij}$ significa la suma sobre las m repeticiones dentro de cada uno de los k niveles de factor.

La suma de cuadrados del error, SSE , es la suma de los cuadrados de las diferencias entre los puntos de datos, y_{ij} , y sus medias de los tratamiento individuales, \bar{y}_i

$$SSE = \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2.$$

La suma de los cuadrados de tratamiento, SSR , es la suma de los cuadrados de las diferencias entre el tratamiento individual, \bar{y}_i y la media general, \bar{y}

$$SSR = \sum_{i=1}^k \sum_{j=1}^m (\bar{y}_i - \bar{y})^2 = m \sum_{i=1}^k (\bar{y}_i - \bar{y})^2.$$

Elevando el término al cuadrado en paréntesis y aplicando la suma nos da

$$m \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 = \sum \bar{y}_i^2 - 2\bar{y} \sum \bar{y}_i + k\bar{y}^2.$$

Denotando el total de todos los valores de la variable respuesta $\sum_{i=1}^k \sum_{j=1}^m y_{ij} = \sum y$;

Ahora reemplazamos \bar{y}_i por T_i/m (donde T es el nombre convencional para los k tratamientos totales individuales) y reemplazando \bar{y} por $\sum y/km$ se obtiene

$$\frac{\sum_{i=1}^k T_i^2}{m^2} - 2 \frac{\sum y \sum_{i=1}^k T_i}{km} + k \frac{\sum y \sum y}{km^2}.$$

Note que $\sum_{i=1}^k T_i = \sum_{j=1}^m y_{ij}$, Por lo que los términos de la derecha positivos y negativos ambos tienen la forma $(\sum y)^2/km^2$. Finalmente, multiplicando por m se tiene

$$SSR = \frac{\sum_{i=1}^k T_i^2}{m} - \frac{(\sum y)^2}{km}.$$

Se puede probar que $SSY = SSR + SSE$.

Información obtenida de [11].

El cálculo de las sumas de cuadrados se presenta de manera tradicional como el Cuadro 4. Hay seis columnas que indican, de izquierda a derecha, la fuente de variación, la suma de cuadrados atribuibles a esa fuente, los grados de libertad para esa fuente, la varianza para esa fuente (tradicionalmente llamados el cuadrado medio en lugar de la varianza), el estadístico F y el valor p asociado con el valor F .

Cuadro 4: ANOVA.

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F	p valor $Pr(> F)$
Tratamiento	SSR	k	$\frac{SSR}{k}$	$\frac{SSR/k}{SSE/m-k-1}$	p
Error	SSE	$m - k - 1$	$\frac{SSE}{m-k-1}$		
Total	SSY	$m - 1$			

Los cuadrados medios se obtienen simplemente dividiendo cada suma de los cuadrados por sus respectivos grados de libertad (en la misma fila). La varianza del error, s^2 , es el cuadrado medio residual (el cuadrado medio de la variación no explicada).

2.1.3. Análisis de Covarianza (ANCOVA)

El análisis de covarianza (ANCOVA) combina elementos de la regresión y el análisis de varianza. La variable de respuesta es continua, y hay al menos una variable explicativa continua (covariables) y con al menos una variable explicativa categórica. Según [3] el ANCOVA se resumen con las siguientes relaciones:

- Colocar dos o más regresiones lineales de Y contra X (uno para cada nivel del factor).
- Estimar diferentes pendientes e interceptos para cada nivel.
- Usar la simplificación del modelo (las pruebas de eliminación) para los parámetros innecesarios.

Los modelos antes descritos, se pueden resumir en el Cuadro 5.

Cuadro 5: Variables explicatorias de los modelos de Regresión, ANOVA y ANCOVA.

MODELO	VARIABLE RESPUESTA	VARIABLES EXPLICATIVAS
Regresión	Continua	Continua
ANOVA	Continua	Categórica
ANCOVA	Continua	Continuas y Categóricas

2.2. Métodos de estimación

A continuación se explican distintos métodos de estimación que ayudan a estimar los parámetros β y σ^2 de los modelos (36) y (37). En el contexto de los LM clásicos, el método de estimación más común es el método de Mínimos Cuadrados Ordinarios.

2.2.1. Mínimos Cuadrados Ordinarios (OLS)

Los valores de β son desconocidos, pero se pueden estimar, utilizando los datos de la muestra. Para estimarlos se usa el método “**Mínimos Cuadrados Ordinarios**” [11], que en inglés es Ordinary Least Squares (*OLS*).

Cuando la matriz \mathbf{X} tiene rango completo p , el estimador *OLS* se obtiene minimizando la suma de cuadrados de los residuos, donde el i -ésimo residuo es la diferencia entre el valor observado y_i y el ajustado \hat{y}_i .

Los residuos o residuales se pueden escribir en forma matricial como sigue:

$$e_{(n \times 1)} = \mathbf{Y} - \hat{\mathbf{Y}}. \quad (39)$$

La suma de cuadrados de los residuos es:

$$\begin{aligned} S(\hat{\beta}) &= \sum_{i=1}^n e_i^2 \\ &= e'e \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}. \end{aligned} \quad (40)$$

Observemos que $\hat{\beta}'\mathbf{X}'\mathbf{Y}$ es una matriz de dimensión 1×1 , es decir, un escalar, y que su transpuesta $(\hat{\beta}'\mathbf{X}'\mathbf{Y})' = \mathbf{Y}'\mathbf{X}\hat{\beta}$, es el mismo escalar. Los estimadores de mínimos cuadrados deben satisfacer

$$\frac{\partial S}{\partial \hat{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0, \quad (41)$$

que se simplifica en

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y}. \quad (42)$$

El sistema lineal de (42) se denominan **ecuaciones normales de mínimos cuadrados** [6]. Para resolver las ecuaciones normales se multiplican ambos lados de (42) por la inversa de $\mathbf{X}'\mathbf{X}$. El estimador $\hat{\beta}$ por mínimos cuadrados es un vector de dimensión $p \times 1$ cuya expresión es

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (43)$$

Los estimadores mínimos cuadrados son insesgados y tienen matriz de varianza y covarianza $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, como se muestra a continuación.

$$\begin{aligned} E(\hat{\beta}_{OLS}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= \beta. \end{aligned} \quad (44)$$

$$\begin{aligned}
Cov(\hat{\beta}_{OLS}) &= Cov[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Cov[\mathbf{Y}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2 I(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned} \tag{45}$$

Se puede demostrar que el estimador *OLS* es el mejor estimador lineal insesgado, que en inglés se escribe Best Linear Unbiased Estimator (*BLUE*) de los parámetros del modelo (36). Esto significa que, de entre los estimadores que son insesgados y lineales con respecto a las observaciones, el estimador *OLS* tiene la menor varianza (teorema Gauss-Markov, [6]). Sin embargo, esto es sólo cuando los supuestos (varianza constante y no correlación) en los residuos se mantengan.

2.2.2. Máxima Verosimilitud (ML)

El método de Máxima Verosimilitud, que en inglés se escribe Maximum Likelihood (*ML*) es un método alternativo para estimar los parámetros en (36), suponiendo que los errores son independientes e idénticamente distribuidos según la normal con varianza constante igual a σ^2 , esto es $N(0, \sigma^2 I)$ [11] y [5]. El método *ML* inicia con la función de densidad de los errores

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2}, \quad i = 1, \dots, n.$$

La función de verosimilitud es:

$$L(\varepsilon, \beta, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \varepsilon' \varepsilon}, \tag{46}$$

de (36) tenemos que $\varepsilon = \mathbf{Y} - \mathbf{X}\beta$. Así (46) se transforma en

$$L(\varepsilon, \beta, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)}. \tag{47}$$

Ahora la función log-verosimilitud es:

$$\begin{aligned}
l &= \log L(\varepsilon, \beta, \sigma^2). \\
&= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta).
\end{aligned} \tag{48}$$

Para un valor fijo de σ , la función log-verosimilitud se maximiza cuando se minimiza el término $(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$.

Por lo tanto, el estimador ML de β bajo los errores normales equivale al estimador de mínimos cuadrados $\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ [6]. El estimador de máxima verosimilitud de σ^2 es

$$\hat{\sigma}_{ML}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta}_{OLS})'(\mathbf{Y} - \mathbf{X}\hat{\beta}_{OLS})}{n}. \tag{49}$$

donde $\hat{\sigma}_{ML}^2$ es un estimador sesgado, pero asintóticamente insesgado (consistencia, ver más en Apéndice 5.1).

2.2.3. Máxima Verosimilitud Restringida (REML)

El método de Máxima Verosimilitud Restringida, que en inglés se escribe Restricted maximum Likelihood (REML) soluciona el problema del sesgo en la estimación de σ^2 .

El REML, consiste en aplicar el método de máxima verosimilitud a un vector $K'Y$ en vez de aplicarlo al vector de observaciones originales Y . La matriz K se define de manera que se elimina del vector Y toda la variación que se explica por la matriz X del modelo. Una diferencia importante entre los vectores Y y $K'Y$ es que la longitud de $K'Y$ es $n - p$ [10]. Por lo tanto, un ajuste ML de un modelo lineal de n observaciones ofrece un estimador de la varianza residual con n en el denominador, mientras que el estimador correspondiente para el vector $K'Y$ ofrece un estimador con $(n - p)$ en el denominador.

Surge la cuestión de cómo encontrar una matriz K tal que elimine toda esa variación de Y que puede ser explicada por X . La condición clave para la eliminación de toda la variación explicada por X es definir cada columna de la matriz K , denotada por k_1, \dots, k_{n-p} , tal que $k_i'X = 0$ para $i = 1, 2, 3, \dots, n - p$. En [17] hay un resultado matricial que establece que el número máximo de columnas linealmente independientes que cumplan la condición anterior es $n - p$, esto es la diferencia entre el número de filas y columnas de la matriz de modelo (la matriz X tiene rango completo). Por lo tanto, solo necesitamos encontrar un número máximo de tales vectores linealmente independientes y apilarlos en la matriz K . Una forma de encontrar al vector k es utilizar la propuesta en [9].

$$k' = c'[I - XX^-], \quad (50)$$

donde c' es un vector arbitrario de longitud n y X^- es una matriz inversa generalizada de X .

Hay que tener en cuenta que puede haber varios valores posibles de X^- . Sin embargo, las estimaciones finales REML no se ven afectados por la elección de X^- .

Una vez que la matriz K se encuentra, vemos que multiplicando el modelo.

$$Y = X\beta + \varepsilon,$$

por K' , por la izquierda, se obtiene

$$K'Y = K'X\beta + K'\varepsilon. \quad (51)$$

Por construcción de la matriz K , es decir, $K'X = 0$, obtenemos

$$K'Y = K'\varepsilon.$$

Si $Var(\varepsilon) = V$, entonces $Var(K'\varepsilon) = K'VK$.

En forma más general, si $Y \sim N(X\beta, V)$, entonces

$$K'Y \sim N(0, K'VK). \quad (52)$$

Así

$$\hat{\beta} = (X'\hat{V}X)^{-1}X'\hat{V}Y. \quad (53)$$

Para (52), la función de verosimilitud se transforma en

$$L(\sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)}. \quad (54)$$

Ahora la función log-verosimilitud es:

$$l = \log L(\sigma^2) \quad (55)$$

$$= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log K'VK - \frac{1}{2K'VK} (K'Y - 0)'(K'Y - 0). \quad (56)$$

tomando $V = \sigma^2 I$, la función log-verosimilitud se transforma en :

$$l = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \sigma^{2(n-p)} (K'K) - \quad (57)$$

$$\frac{1}{2(n-p)} (K'K)Y'K(K'K)^{-1}K'Y. \quad (58)$$

Para estimar σ^2 , diferenciamos el log- verosimilitud con respecto a σ^2

$$\frac{\partial l}{\partial \sigma^2} = \frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4} Y'K(K'K)^{-1}K'Y. \quad (59)$$

igualando a cero, y despejando σ^2 . Así tenemos que el estimador *REML* de σ^2 es:

$$\hat{\sigma}_{REML}^2 = \frac{1}{n-p} Y'K(K'K)^{-1}K'Y. \quad (60)$$

Por lo tanto, el modelo se puede ajustar utilizando una verosimilitud restringida basado en la normalidad de $K'Y$. Una diferencia esencial entre estas dos verosimilitudes, además de las longitudes de Y y $K'Y$, es que la verosimilitud *REML* no involucra a $X\beta$. Por lo tanto, *REML* se puede utilizar solo para la estimación de parámetros relacionados con V . Mientras que el vector de parámetros de β puede ser estimado usando el estimador *GLS* que se explica a continuación.

2.2.4. Mínimos Cuadrados Generalizados (*GLS*)

Si las suposiciones sobre la varianza constante y correlación cero entre los residuales no se cumplen, el estimador *OLS* del LM sigue siendo insesgado. Sin embargo, ya no es un estimador de mínima varianza. Ahora se puede utilizar el estimador por mínimos cuadrados generalizados, que en inglés se escribe Generalized Least Squares (*GLS*).

Como V es la matriz de varianza y covarianza del error, es decir, en forma más general $Var(\varepsilon) = \sigma^2 V$, el estimador *GLS* de β minimiza la suma de cuadrados $(Y - X\beta)'V^{-1}(Y - X\beta)$. Así el estimador *GLS* es

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y. \quad (61)$$

Para el modelo lineal con $Var(\varepsilon) = \sigma^2 V$, este estimador es el *BLUE*, es decir, tiene la variación más pequeña de entre todos las posibles estimadores insesgados (generalización del teorema de Gauss-Markov).

El estimador *GLS* es insesgado y tiene matriz de varianza covarianza $X'V^{-1}X^{-1}$

$$\begin{aligned} E(\hat{\beta}) &= E[(X'V^{-1}X)^{-1}X'V^{-1}Y] \\ &= (X'V^{-1}X)^{-1}X'V^{-1}E[Y] \\ &= (X'V^{-1}X)^{-1}X'V^{-1}X\beta \\ &= \beta, \end{aligned}$$

$$\begin{aligned}
Cov(\hat{\beta}) &= Cov((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Cov(\mathbf{Y})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})' \\
&= \sigma^2((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}) \\
&= \sigma^2((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}) \\
&= \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.
\end{aligned}$$

El estimador *OLS* se puede derivar de los métodos de estimación anteriores, sustituyendo $\sigma^2\mathbf{V}$ por $\sigma^2\mathbf{I}$ [10].

2.3. Bondad de ajuste del modelo

En esta sección se presentan distintas maneras de evaluar la bondad de ajuste del LM, lo que se refiere al grado en que éste es conveniente como modelo que representa a las variables implicadas en él. Con el uso del software R, anteriormente se comentó que el ajuste de los Modelos Lineales se realizan usando la función `lm()`, mientras que para ver la bondad del ajuste se usa la función `summary()` [3]. Esta función da el resumen del modelo ajustado, permitiendo observar las distintas pruebas de la bondad de ajuste del modelo, estas pruebas se describen a continuación.

2.3.1. Prueba F para el ajuste del modelo

La prueba *F* del modelo nos permite determinar estadísticamente si las variables explicativas (en conjunto) tienen efecto o no sobre la variable respuesta. Este procedimiento suele considerarse como una prueba general o global del ajuste del modelo. La prueba de hipótesis es:

$$\begin{aligned}
H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0 \\
\text{vs} \\
H_a : \beta_j \neq 0 \text{ al menos para un } j \text{ (} j = 1, \dots, p \text{)}.
\end{aligned}$$

El rechazo de la hipótesis nula implica que al menos uno de los regresores contribuye al modelo en forma significativa.

Aunque en el Cuadro 4 desarrollamos como elaborar una ANOVA que muestra el cálculo del estadístico *F* para esta prueba, creemos conveniente repetirlo con la notación comúnmente utilizada en los modelos de regresión.

Este método consiste en una partición de la variabilidad total de la variable *y* de respuesta. Para obtener esta partición se comienza con la identidad

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i). \quad (62)$$

Ahora procedemos a elevar al cuadrado en ambos lados de la ecuación (62), y se suman para todas las *n* observaciones. Así se obtiene la siguiente ecuación:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (63)$$

Por otro lado,

$$\begin{aligned}
2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= 2 \sum_{i=1}^n \hat{y}_i(y_i - \hat{y}_i) - 2 \sum_{i=1}^n \bar{y}(y_i - \hat{y}_i) \\
&= 2 \sum_{i=1}^n \hat{y}_i e_i - 2\bar{y} \sum_{i=1}^n e_i = 0.
\end{aligned} \tag{64}$$

Ya que la suma de los residuales siempre es igual a cero y la suma de los residuales ponderados por el valor ajustado $(\bar{y} - y_i)$ correspondiente también es igual a cero, la ecuación (63), se reduce a

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \tag{65}$$

Esta igualdad nos dice que **suma total de cuadrados** que se denota por SS_T se puede escribir como una **suma de cuadrados debido a la regresión**, SS_R , y una **suma de cuadrados de residuales**, SS_{Res} .

$$SS_T = SS_R + SS_{Res}. \tag{66}$$

En [11] se demuestra que $\frac{SS_R}{\sigma^2}$ tiene una distribución χ_p^2 , con el mismo número de grados de libertad que la cantidad de variables regresoras, $\frac{SS_{Res}}{\sigma^2} \sim \chi_{n-p-1}^2$ y que SS_R y SS_{Res} son independientes. Tomando a

$$F_0 = \frac{SS_R/p}{SS_{Res}/(n-p-1)} = \frac{MS_R}{MS_{Res}}$$

este tiene una distribución $F_{p,n-p-1}$ y rechazamos H_0 si

$$F_0 > F_{\alpha,p,n-p-1}.$$

El procedimiento se resume en el Cuadro 6 de análisis de varianza de la regresión.

Cuadro 6: Análisis de varianza de la regresión.

Fuente de Variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	P valor $0 < P < 1$
Regresión	SS_R	p	MS_R	$\frac{MS_R}{MS_{Res}}$	p
Residuales	SS_{Res}	$n - p - 1$	MS_{Res}		
Total	SS_T	$n - 1$			

citado en [11].

2.3.2. Coeficiente de determinación (R^2) y ajustado (R_{Adj})

El coeficiente de determinación nos permite expresar la cantidad de la variabilidad presente en las observaciones de Y , que se explica mediante el modelo LM. La cantidad

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}, \tag{67}$$

se llama **coeficiente de determinación**, R^2 indica la proporción de variabilidad explicada por la regresión. Ya que $0 \leq SS_{Res} \leq SS_T$, entonces $0 \leq R^2 \leq 1$. Los valores de R^2 cercanos a 1 implican que la mayor parte

de la variabilidad de Y está explicada por el modelo de regresión. A medida que el coeficiente se aproxime a cero el modelo deja de ser adecuado, ya que la cantidad de la variabilidad explicada mediante el modelo es pobre [11].

En general, R^2 aumenta siempre que se agrega un regresor al modelo, independientemente del valor de la contribución de esa variable. En consecuencia, es difícil juzgar si un aumento de R^2 dice en realidad algo importante.

Es posible usar el estadístico R_{Adj}^2 , que se define como sigue:

$$R_{Adj}^2 = 1 - \frac{SS_{Res}/(n - (p + 1))}{SS_T/(n - 1)}. \quad (68)$$

En vista de que $\frac{SS_{Res}}{(n - (p + 1))}$ es el cuadrado medio de los residuales, $\frac{SS_T}{(n - 1)}$ es constante e independiente de cuántas variables hay en el modelo, R_{Adj}^2 sólo aumentará al agregar una variable al modelo si esa variable reduce el cuadrado medio residual. La R_{Adj}^2 penaliza el aumento de términos que no son útiles. Tanto R^2 como R_{Adj}^2 , suelen utilizarse como procedimientos para evaluar y comparar los posibles modelos de regresión.

2.3.3. Prueba t de Student sobre coeficientes individuales

Una vez determinado que al menos una de las variables independientes es importante, la siguiente pregunta es: ¿Cuál(es) variable(s) (son) es importante(s)? Si agregamos una variable al modelo de regresión, la suma de cuadrados de la regresión aumenta y la suma de cuadrados residuales disminuye. Se debe decidir si el aumento de la suma de cuadrados de la regresión es suficiente para garantizar el uso del regresor adicional en el modelo.

En [11] se sugiere tener cuidado al agregar una variable explicativa, ya que también aumenta la varianza del valor ajustado \hat{Y} , por lo que se sugiere incluir sólo variables explicativas que tengan valor para explicar la respuesta. Además, si agregamos una variable explicativa que no es importante se puede aumentar el cuadrado medio de residuales y con eso disminuye la utilidad del modelo.

Las hipótesis para probar la significancia de cualquier coeficiente $\beta_i, i = 1, 2, \dots, p$, está dado por:

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_a : \beta_i \neq 0 \quad \text{para} \quad i = 1, 2, \dots, p.$$

El estadístico de prueba para esta hipótesis es,

$$t_0 = \frac{\hat{\beta}_i}{\sqrt{\sigma^2 C_{ii}}}, \quad (69)$$

donde C_{ii} es el elemento diagonal de $(X'X)^{-1}$ que corresponde a $\hat{\beta}_i$. Se rechaza la hipótesis nula $H_0 : \beta_i = 0$ si

$$|t_0| > t_{\alpha/2, n-p-1}.$$

Se observa que ésta es una prueba parcial, porque el coeficiente de regresión $\hat{\beta}_i$ depende de todas las demás variables explicativas x_j , que hay en el modelo. Esto es una prueba de la contribución de x_i dadas las demás variables del modelo.

En general, el cuadrado de una variable aleatoria t con f grados de libertad es una variable aleatoria F con 1 y f grados de libertad, respectivamente [11]. Aunque la prueba t para $H_0 : \beta_1 = 0$ equivale a la prueba F en la regresión lineal simple, la prueba t es algo más adaptable, porque se podría usar para probar hipótesis alternativas unilaterales (Sea $H_1 : \beta_1 < 0$ o $H_1 : \beta_1 > 0$), mientras que la prueba F sólo considera la alternativa bilateral [11].

2.3.4. Intervalos de confianza

Para construir intervalos de confianza par los coeficientes en β es importante reiterar que $\varepsilon \sim N(0, \sigma^2 I)$ y $y_i \sim N(0, \beta X)$ son independientes para cualesquiera $i \neq j$. Note que $\hat{\beta}$ es una combinación lineal de las observaciones, entonces se distribuye normal con media cero y matriz de covarianza $\sigma^2(X'X)^{-1}$. De esta manera, cada estadístico

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \quad j = \overline{0, k} \quad (70)$$

se distribuye $t_{(n-p)}$. Con esta información es sencillo construir un intervalo de confianza para el j-ésimo coeficiente

$$\left(\hat{\beta}_j - t_{(n-p)}^{\alpha/2} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{(n-p)}^{\alpha/2} \sqrt{\hat{\sigma}^2 C_{jj}} \right) \quad (71)$$

Donde C_{jj} es el j-ésimo elemento en la diagonal de $(X'X)^{-1}$.

2.3.5. Predicción de nuevas observaciones

Análogo al caso de regresión lineal simple, el modelo puede ser utilizado para hacer predicciones de observaciones futuras y, para valores particulares de X . Si $X'_0 = [1, x_{01}, \dots, x_{0k}]$, entonces el punto estimado de una observación futura y_0 es

$$\hat{y}_0 = X'_0 \hat{\beta}$$

y el intervalo de confianza para y_0 es

$$\left(\hat{y}_0 - t_{(n-p)}^{\alpha/2} \sqrt{\hat{\sigma}^2 (1 + X'_0 (X'X)^{-1} X_0)} \leq y_0 \leq \hat{y}_0 + t_{(n-p)}^{\alpha/2} \sqrt{\hat{\sigma}^2 (1 + X'_0 (X'X)^{-1} X_0)} \right) \quad (72)$$

Que es una generalización de la Ec. (28).

3. Revisión de la adecuación del modelo

Hasta ahora se ha discutido sobre el ajuste de la recta (o plano en R^n) de regresión y sus propiedades asumiendo que los supuestos se cumplen. Sin embargo, es igual de importante verificar la validez de las estas premisas para que el modelo sea insesgado y de mínima varianza, además de la posibilidad de las pruebas estadísticas. Para recordar, los supuestos son:

1. La relación entre la respuesta y y los regresores es lineal, al menos en forma aproximada.
2. El error ε tiene media cero.
3. El error ε tiene varianza σ^2 constante.
4. Los errores no están correlacionados.
5. Los errores tienen distribución normal.

En este capítulo se introducen análisis del cumplimiento de estos puntos.

3.1. Residuales, escalamiento y error puro

Note que 4 de los 5 supuestos están relacionados con el error ε de nuestro modelo $Y = X\beta + \varepsilon$, de tal manera que la violación de alguno de estos puntos se ve reflejada en los errores. Aunque no es del todo preciso, se puede entender a los residuales como los errores observados ($e = Y - \hat{Y}$), siendo estos la desviación entre los datos y el ajuste. Sin embargo, los residuales no son independientes, ya que los n residuales sólo tienen $n - p$ grados de libertad asociados a ellos. Si $n \gg p$, la dependencia tiene poco efecto al utilizar los residuales para la comprobación del modelo. Como se ha visto, los residuales tienen media cero y matriz de covarianza es

$$\text{cov}(e) = \sigma^2[I - X(X'X)^{-1}X'] = \sigma^2[I - H]$$

donde $H = X(X'X)^{-1}X'$ es la matriz gorro. Algunas características interesantes de la matriz gorro son, por ejemplo, que transforma a Y en \hat{Y} , $\hat{Y} = HY$; además, $HX = X(X'X)^{-1}X'X = X$; es simétrica e idempotente; los residuales se pueden expresar como

$$e = (I - H)Y = (I - H)(X\beta + \varepsilon) = X\beta - HX\beta + (I - H)\varepsilon = X\beta - X(X'X)^{-1}X'X\beta + (I - H)\varepsilon = (I - H)\varepsilon$$

es decir, si h_{ij} es el elemento de la i -ésima fila y la j -ésima columna, entonces $e_i = \varepsilon - \sum_{j=1}^n h_{ij}\varepsilon_j$ $i = \overline{1, n}$.

La varianza se estima por medio de

$$\hat{\sigma}^2 = \frac{1}{n - k - 1} (Y - X\hat{\beta})'(Y - X\hat{\beta}) = \frac{Y'Y - \hat{\beta}'X'Y}{n - k - 1} = \frac{SSE}{n - k - 1}$$

La matriz de covarianza de $\hat{\beta}$ es

$$\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

y se puede estimar sustituyendo σ^2 por $\hat{\sigma}^2$.

Como los residuales tienen distribución normal, es posible obtener **residuales estandarizados**

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \quad (73)$$

Dado que la varianza del i -ésimo residual se encuentra en la matriz $(I - H)$, entonces una aproximación más precisa se logra sabiendo que $V(e_i) = \sigma^2(1 - h_{ii})$ y $cov(e_i, e_j) = -\sigma^2 h_{ij}$. Ahora bien, ya que $0 \leq h_{ii} \leq 1$, si se usa la estimación $\hat{\sigma}^2$, en realidad se sobre estima $V(e_i)$. Además, como h_{ii} es una medida de ubicación del i -ésimo punto en el espacio de X , la varianza de e_i depende de dónde esta el punto x_i . Bajo los supuestos, $V(r_i) = 1$, que es independiente de la posición de x_i , donde r_i es el i -ésimo **residual estudentizado** dado por

$$r_i = \frac{e_i}{\sqrt{\sigma^2(1 - h_{ii})}} \quad (74)$$

Un tercer método de escalar los residuales usa una estimación de σ^2 que ha excluido la i -ésima observación

$$t_i = \frac{e_i}{\sqrt{\hat{\sigma}_{(i)}^2(1 - h_{ii})}} \quad (75)$$

donde $\hat{\sigma}_{(i)}$ es calculado con $n - 1$ observaciones restantes después de omitir $(y_i, x'_i) = (y_i, 1, x_{i1}, \dots, x_{ik})$. Si la observación omitida es atípica, será más probable que aparezca con esta nueva estandarización llamada estudentización externa de los residuales o sólo R de Student.

Otra opción es examinar los residuales omitidos **PRESS** (prediction error sum of equares). El i -ésimo residual omitido $e_{(i)}$ es calculado con $\hat{\beta}_{(i)}$, que se basa en $n - 1$ observaciones restantes después de omitir (y_i, x'_i) , es decir

$$e_{(i)} = y_i - \hat{y}_{(i)} = y_i - x'_i \hat{\beta}_{(i)}$$

Por definición,

$$\hat{\beta}_{(i)} = (X'_{(i)} X_{(i)})^{-1} X'_{(i)} Y_{(i)} \quad (76)$$

Donde $X_{(i)}$ e $Y_{(i)}$ siguen el mismo sentido de omisión del i -ésimo elemento. Parecería que es necesario hacer un modelo de regresión para cada subconjunto de valores después de ir iterando el valor omitido, pero no es así, ya que

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{e_i}{1 - h_{ii}} (X'X)^{-1} x_i \quad (77)$$

Si siguiendo esto, se puede calcular el valor PRESS

$$e_{(i)} = \frac{e_i}{1 - h_{ii}} \quad (78)$$

Así, los n residuales PRESS pueden ser calculados sin hacer n regresiones. Los residuales escalados t_i pueden expresarse en términos de PRESS como

$$t_i = \frac{e_{(i)}}{\sqrt{\hat{V}(e_{(i)})}} \quad (79)$$

Una manera de detectar los puntos atípicos es graficar los residuales ordinarios contra los residuales PRESS. Si no existe un cambio sustancial al calcular $\hat{\beta}$ sin alguna observación, el anterior gráfico debería seguir, aproximadamente, una línea recta de pendiente 1. Cualquier punto lejano a esta recta es potencialmente un valor atípico.

Si un punto atípico viene de una distribución con una media diferente, el modelo puede ser expresado como $E(y_i) = x'_i \beta + \theta$. La distribución de t_i en (75) o (79) es una t de Student con $(n - k - 1)$ grados de libertad, de tal manera que t_i sirve para la prueba de hipótesis $H_0 : \theta = 0$. Como se harán n pruebas, podemos

sólo enfocarnos en el tamaño de los valores de t_i . Con los residuales PRESS es posible definir el **estadístico PRESS**

$$PRESS = \sum e_{(i)}^2 = \sum [y_i - \hat{y}_{(i)}]^2 = \sum \left(\frac{e_i^2}{1 - h_{ii}} \right) \quad (80)$$

Los residuales con valores grandes en h_{ii} contribuyen más al estadístico PRESS. Para un conjunto dado de datos, PRESS puede ser un mejor evaluador de la calidad predictiva del modelo que SSE. Se prefiere valores pequeños de este estadístico para un modelo con mejor predictibilidad. Si un valor alto de algún residual PRESS se presenta, este dato es lejano al resto, tanto que podría ser un dato atípico dado por la naturaleza del experimento y no precisamente por un error de medición. En esta situación el experimento tiene comportamientos diferentes para valores cercanos a la observación correspondiente a dicho residual. Por ejemplo, si se estima la proliferación de una bacteria dada una temperatura, es posible que haya un punto de inflexión donde en vez de seguir aumentando el número de bacterias, la mayoría muera.

Si se tienen observaciones repetidas Y para un mismo punto en el espacio X , es posible hacer un nuevo análisis sobre el **error puro**. Estas repeticiones deben ser genuinas, es decir, replicar el experimento para el mismo punto en X y obtener las observaciones. Una repetición no genuina está dada por observar varias veces un experimento que no se ha repetido. Por ejemplo, si queremos estimar la relación entre la inteligencia IQ y la altura, una repetición genuina estaría dada si observamos el IQ de dos personas con la misma altura. Por el contrario, no sería genuina si observamos dos veces el IQ de la misma persona, en ese caso se recomienda mejor solo registrar una observación.

Spongamos que tenemos m diferentes valores de X y para el j -ésimo valor existen n_j observaciones. Es decir, existen $y_{j1}, y_{j2}, \dots, y_{jn_j}$ observaciones para x_j . Todas las observaciones juntas son

$$n = \sum_{j=1}^m \sum_{u=1}^{n_j} 1 = \sum_{j=1}^m n_j$$

observaciones. La contribución de la suma de cuadrados del error puro de n_1 observaciones en x_1 es la suma interna de cuadrados de Y_{1u} sobre su promedio \bar{y}_1 ; que es

$$\sum_{u=1}^{n_1} (y_{1u} - \bar{y}_1)^2 = \sum_{u=1}^{n_1} y_{1u}^2 - n_1 \bar{y}_1^2 = \sum_{u=1}^{n_1} y_{1u}^2 - \frac{1}{n_1} \left(\sum_{u=1}^{n_1} y_{1u} \right)^2 \quad (81)$$

Siempre que estemos seguros que la variación del error puro es de la misma magnitud a través de los datos, es conveniente hacer la agrupación de las sumas internas de cuadrados de todos los lugares con observaciones repetidas para obtener el error puro general SS como

$$\sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2$$

con grados de libertad

$$n_e = \sum_{j=1}^m (n_j - 1) = \sum_{j=1}^m (n_j - m)$$

Por tanto, el cuadrado medio del error puro es

$$S_e^2 = \frac{\sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2}{\sum_{j=1}^m (n_j - m)}$$

Se puede entender esta cantidad como la suma total de cuadrados dentro de repeticiones entre sus grados de libertad y es una estimación de σ^2 .

La **prueba de Barlett** muestra si existe homogeneidad en el error puro. Esta prueba requiere de normalidad y es sensible a la violación de este supuesto, es decir, bajo no normalidad la validez de la prueba es afectada.

Sea $S_1^2, S_2^2, \dots, S_m^2$ estimaciones de σ^2 provenientes de m de repeticiones con v_1, \dots, v_m grados de libertad, respectivamente, donde $v_j = n_j - 1$,

$$S_j^2 = \frac{\sum_{u=1}^{n_j} (Y_{uj} - \bar{Y}_j)^2}{n_j - 1},$$

$$S_e^2 = \frac{v_1 S_1^2 + \dots + v_m S_m^2}{v_1 + \dots + v_m}$$

Denotemos $v = \sum_{u=1}^m v_u$ y la constante C como

$$C = 1 + \frac{v_1^{-1} + \dots + v_m^{-1} - v^{-1}}{3(m-1)}$$

Entonces, el estadístico de prueba es

$$B = \frac{v \ln S_e^2 - \sum_{j=1}^m v_j \ln S_j^2}{C} \quad (82)$$

Cuando las varianzas de los m grupos son iguales, B se distribuye aproximadamente como una χ_{m-1}^2 . Valores grandes de B podrían indicar heterogeneidad de las varianzas, así como falta de normalidad.

3.2. Gráficos de los residuales

Gráfica de probabilidad normal

Las pequeñas desviaciones respecto a la suposición de normalidad no afectan mucho al modelo, pero tener desviaciones grandes de no normalidad es potencialmente peligroso, porque los estadísticos t o F dependen de la suposición de normalidad. En [11] se comentó que si los errores provienen de una distribución con colas más gruesas (cuando la frecuencia de ocurrencia de eventos que están situados en los extremos de la distribución no es muy baja) que la normal, el ajuste por mínimos cuadrados será sensible a un subconjunto menor de datos.

Un método muy sencillo de comprobar la suposición de **normalidad** es trazar una gráfica de **probabilidad normal** de los residuales. Esta es una gráfica diseñada para que se dibuje una línea recta, que representa a una normal acumulada.

Sea $e_1 < e_2 < \dots < e_n$ los residuales ordenados en orden creciente. Si se grafica e_i en función de la probabilidad acumulada $P_i = (i - \frac{1}{2})/n, i = 1, 2, \dots, n$, los puntos que resulten deberían estar aproximadamente sobre una línea recta.

La recta se suele determinar en forma visual, con énfasis en los valores centrales (por ejemplo, los puntos de probabilidad acumulada 0.33 y 0.67) y no en los extremos. Las diferencias apreciables en distancia respecto a la recta indican que la distribución no es normal.

A veces, las gráficas de probabilidad normal se trazan graficando el residual clasificando e_i en función del valor normal esperado, $\phi^{-1}[(i - \frac{1}{2})/n]$, donde ϕ representa a la función de distribución acumulada de la distribución normal estándar. Esto es consecuencia de $E(e_i) \simeq \phi^{-1}[(i - \frac{1}{2})/n]$ [11].

El estudio de las gráficas ayuda a adquirir un grado de percepción de cuánta desviación de la recta es aceptable. Con frecuencia, los tamaños pequeños de muestra ($n \leq 16$) producen gráficas de probabilidad normal que se desvían bastante de línea recta que representa la normal acumulada. Para muestras mayores ($n \geq 32$), las gráficas se comportan mucho mejor. Por lo general, se requieren alrededor de 20 puntos para producir gráficas de probabilidad suficientemente estables como para poder interpretarse con facilidad.

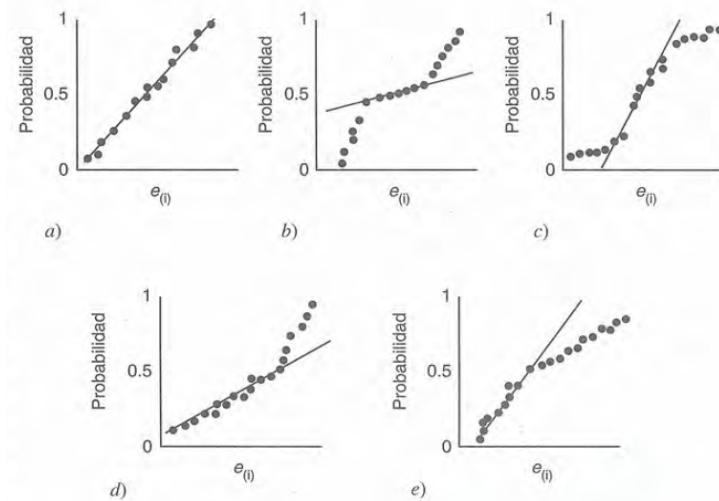


Figura 12: Gráficas de probabilidad normal: a) ideal; b) distribución con colas gruesas; c) distribución con colas delgadas; d) asimetría positiva; e) asimetría negativa. Fuente: [11].

3.2.1. Gráfica de residuales en función de los valores ajustados \hat{y}_i

Para poder detectar algunas inadecuaciones del modelo, es útil tener una gráfica de los residuales en función de los valores ajustados correspondientes \hat{y}_i . Esta gráfica permite detectar diferentes problemas, tales como:

- **Heterocedasticidad**, la varianza no es constante y se deben de transformar los datos (la variable Y) o aplicar otros métodos de estimación.
- **Error en el análisis**, se ha realizado mal el ajuste y se verifica que los residuos negativos se corresponden con los valores pequeños \hat{y}_i y los errores positivos se corresponden con los valores grandes de y_i , o al revés.
- El modelo es inadecuado por **falta de linealidad** (no lineal) y se deben transformar los datos o introducir nuevas variables que pueden ser cuadrados de las existentes o productos de las mismas, o bien se deben introducir nuevas variables explicativas.
- Existencia de **observaciones atípicas** o puntos extremos.
- **Falta de independencia**, los residuales se presentan formando grupos (clusters).

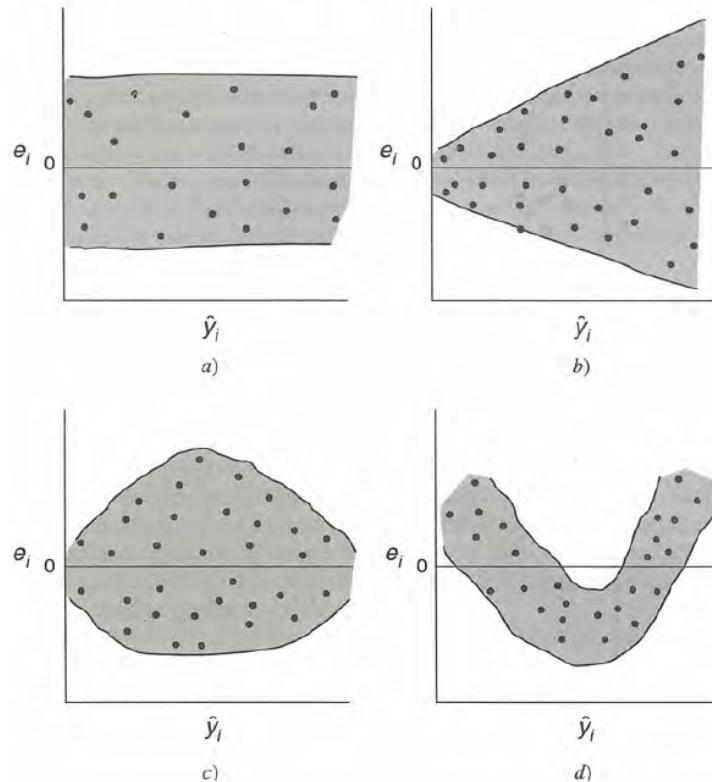


Figura 13: Patrones en las gráficas de residuales: a) satisfactorio; b) en embudo; c) en doble arco; d) no lineal
Fuente: [11].

En la Fig. 13, las distribuciones en las partes b y c indican que la varianza de los errores no es constante. La figura de embudo abierto hacia afuera en la parte b implica que la varianza es función creciente de y . También es posible un embudo abierto hacia dentro, que indica que $V(e)$ aumenta a medida que y disminuye. La distribución en doble arco en la parte c se presenta con frecuencia cuando y es una proporción entre 0 y 1. Se grafican los residuales contra \hat{y}_i y no y_i porque usualmente existe una correlación entre e_i y y_i , formando patrones incluso si no existe ningún problema en el modelo. También es plausible tener a los residuales e_i en el eje de las ordenadas y algún regresor x_{ji} en las abscisas. Se espera, de manera ideal, tener aleatoriedad como en la Figura 13(a). De otra manera, como en (b) y (c), puede existir varianza no constante o una relación de orden superior entre la variable Y y X_j , como se muestra en 13(d).

Un supuesto del modelo de regresión lineal es la independencia entre las observaciones o en análogo la independencia entre los errores. Cuando se toma una muestra que tiene un orden temporal marcado es posible tener **autocorrelación temporal**. Si conocemos el orden en el tiempo de los datos, se puede graficar los residuales con la esperanza de que sigan mostrando aleatoriedad, similar a la Figura 13(a), de lo contrario es posible que se necesite una transformación, o incluso cambiar el enfoque de regresión lineal a series de tiempo. Incluso, en el caso multivariado, si deseamos conocer la correlación entre variables explicativas X se puede hacer una gráfica de X_j contra X_i para $i \neq j$. Como se espera que la correlación sea cercana a

cero, no debe existir ningún patrón en este gráfico. La correlación es una medida de asociación lineal entre dos variables, de tal manera que si la relación es no lineal, puede que no se vea reflejada en el factor de correlación, por lo cual un gráfico no da una visión más general de la influencia de una variable regresora con otra.

Si conocemos información extra sobre el experimento, como la fuente de las observaciones, algún orden temporal o espacial, además de alguna relación a priori de las variables, es mejor hacer cualquier gráfica que resulte de la imaginación del investigador para no perder generalidad de la información y establecer una correcta adecuación al modelo de regresión lineal.

3.3. homogeneidad en las varianzas

En la ecuación (82) se expuso una prueba estadística para la homogeneidad de las varianzas. Aquí se hace una continuación de estas pruebas. Existe una ventaja sustancial de una prueba estadística sobre un gráfico al no ser necesaria una interpretación que puede ser hasta arbitraria sujeta a la experiencia del investigador. Por otro lado, la prueba estadística resume tanto la información que para un conjunto de datos obtenemos un único resultado lógico de verdadero o falso y no es posible discernir el tipo de comportamiento que posiblemente tenga la varianza, como en un gráfico.

En la **prueba de Barlett modificada para curtosis** el estadístico B de la Ec. (82) es multiplicado por $d = 2/(\hat{\beta} - 1)$, donde

$$\hat{\vartheta} = \frac{N \sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \bar{Y}_j)}{\left(\sum_{j=1}^m \sum_{u=1}^{n_j} (Y_{ju} - \bar{Y}_j)^2 \right)^2}$$

estima la curtosis de los conjuntos de repeticiones. Para una distribución normal de los datos se espera que el verdadero valor de ϑ sea 3 y d tenga un valor cercano a 1. La misma prueba χ^2 es utilizada. N es el número total de observaciones (usualmente reducidas) del conjunto de datos usados para la prueba, que es, el número total de observaciones en todos los conjuntos de repeticiones, ignorando las observaciones sin repeticiones.

La **prueba de Levene con medias** considera, en el j -ésimo grupo de repeticiones, la desviación absoluta

$$z_{ju} = |Y_{ju} - \bar{Y}_j|, \quad u = 1, 2, \dots, n_j$$

de los Y 's de las medias de sus grupos repetidos. Entiéndase esto como una manera de clasificación y comparación de los cuadrados medios "entre grupos" con los cuadrados medios "dentro de grupos." a través del estadístico F. El estadístico F apropiado es

$$F_0 = \frac{\sum_{j=1}^m n_j (\bar{z}_j - \bar{z})^2 / (m-1)}{\sum_{j=1}^m \sum_{u=1}^{n_j} (z_{ju} - \bar{z}_j)^2 / \sum_{j=1}^m (n_j - 1)} \quad (83)$$

donde

$$\bar{z}_j = \sum_{u=1}^{n_j} \frac{z_{ju}}{n_j} \quad \& \quad \bar{z} = \sum_{j=1}^m \sum_{u=1}^{n_j} \frac{z_{ju}}{\sum_{j=1}^m n_j}$$

El valor F se refiere a $F_{(m-1), \sum_{j=1}^m (n_j-1)}$ de cola superior.

3.4. normalidad en los residuales

Usualmente asumimos que los residuales se distribuyen normal $e_i \sim N(0, \sigma^2)$, supuesto fuerte para los intervalos de confianza y pruebas de hipótesis, y todos los errores son independientes unos de otros. Pero sus estimados, los residuales, no pueden ser independientes unos de otros. La estimación de los parámetros nos dice que n residuales sólo tienen $(n - p)$ grados de libertad. Las p ecuaciones normales son restricciones sobre los e_i . A menos que p sea grande en comparación con n , esto tiene poco efecto sobre la revisión de la normalidad.

Si el modelo ajustado es $E(Y) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, la ecuación puede ser escrita como

$$-2 \sum (y_i - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_k x_{ki}) = 0$$

Esto se reduce a

$$\sum (y_i - \hat{y}_i) = 0$$

Por lo tanto

$$\sum e_i = 0 \quad (84)$$

De esta manera, no es necesario rectificar que el residual medio $\bar{e} = \sum e_i / n$ es cero, pues es consecuencia directa de la Ec. (84). Otra manera es hacer pequeños intervalos y después crear un histograma para la frecuencia de los intervalos. Si los residuales se distribuyen aproximadamente normal deberían presentar la forma de la campana normal en este histograma. Para pruebas estadísticas de normalidad podemos encontrar la *Shapiro-Wilk* para muestras menores a 50; la prueba *Kolmogorov-Smirnov* es más general y sirve para comprobar si los datos siguen una distribución normal, uniforme, Poisson o exponencial; la prueba *Anderson-Darling* es para casi cualquier distribución y entre menor sea el estadístico, mejor es el ajuste, de tal manera que es posible comparar el ajuste de los datos a distintas distribuciones Fig. (14), incluida la normal.

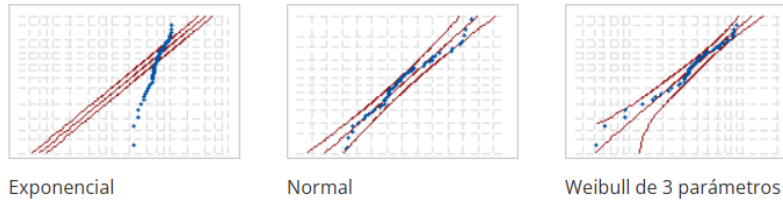


Figura 14: Prueba Anderson-Darling para los mismos datos y distintas distribuciones.

3.5. Prueba Durbin-Watson

Una prueba popular para detectar algún tipo de correlación serial es la prueba Durbin-Watson. Supongase que se quiere ajustar un modelo lineal

$$Y_u = \beta_0 + \sum_{i=1}^k \beta_i X_{iu} + \varepsilon_u$$

por mínimos cuadrados para las observaciones $(Y_u, X_{1u}, \dots, X_{ku})$, para $u = 1, \dots, n$. Usualmente asumimos que los errores son variables independientes $\varepsilon_i \sim N(0, \sigma^2)$, por lo cual todas las correlaciones seriales son

$\rho_s = 0$. Deseamos verificar este supuesto a través de los residuales. Se plantea el juego de hipótesis

$$H_0 : \rho_s = 0 \quad \forall s \quad \text{contra} \quad H_1 : \rho_s = \rho^s$$

($\rho \neq 0$, $|\rho| < 1$). La hipótesis alternativa surge de asumir que los errores ε_u son tales que

$$\varepsilon_u = \rho \varepsilon_{u-1} + z_u$$

donde $z_u \sim N(0, \sigma^2)$ y es independiente de $\varepsilon_{u-1}, \varepsilon_{u-2}, \dots$, y z_{u-1}, z_{u-2}, \dots . Se asume, también, que la media y la varianza de ε_u son constantes, independientes de u , de lo cual se sigue necesariamente que

$$\varepsilon_u \sim N(0, \sigma^2 / (1 - \rho^2))$$

Note que bajo la hipótesis nula esto se reduce a $\varepsilon_u \sim N(0, \sigma^2)$. El estadístico es

$$d = \frac{\sum_{u=2}^n (e_u - e_{u-1})^2}{\sum_{u=1}^n e_u^2} \quad (85)$$

y determina si se rechaza o no la hipótesis nula basado en el valor d . La distribución de d depende de los datos en X y no es independiente de ellos. Esta distribución cae entre 0 y 4, siendo simétrica al rededor de 2. Los puntos porcentuales también dependen de los datos de X y tendrían que calcularse para cada aplicación para realizar la prueba correctamente. Debido a la dificultad de hacer esto de manera rutinaria, la prueba generalmente se realiza utilizando límites tabulados (d_L, d_U). Por lo tanto, en lugar de buscar un único valor crítico, tenemos que buscar dos valores críticos. Además, d se utiliza sólo para prueba de cola inferior contra la alternativa $\rho < 0$, teóricamente necesitamos una prueba de cola superior; afortunadamente, esto puede manejarse simplemente como una prueba de cola inferior utilizando el estadístico $(4-d)$.

Note que los extremos 0 y 4 son alcanzables para muestras muy grandes. Los valores mínimos alcanzables dependen del tamaño de la muestra n de la siguiente manera

n	15	30	50	100	200	300	500
d mínimo	0.0437	0.0110	0.0039	0.0010	0.0002	0.0001	0.0000

El correspondiente valor d máximo es $(4 - d$ mínimo). Para tablas más detalladas con distintos n (observaciones), k (número de variables independientes) y α (nivel de significancia) revise [13], páginas 184-192, junto con las reglas de decisión que las acompañan.

3.6. Puntos atípicos

Un valor atípico es una observación extrema. Los residuales cuyo valor absoluto es bastante mayor que los demás, digamos de tres a cuatro desviaciones estándar respecto a la media, indican que hay valores atípicos potenciales en el espacio de Y . Los valores atípicos son puntos que no son representativos del resto de los datos. De acuerdo con su ubicación en el espacio de X , los valores atípicos pueden tener efectos de moderados a graves sobre el modelo de regresión. Las gráficas de residuales en función de y_i y la gráfica de probabilidad normal son útiles para identificar puntos atípicos. El examen de los residuales escalados, como por ejemplo los residuales estudentizados y los R de Student es una forma excelente de identificar puntos atípicos potenciales. Los valores atípicos se deben investigar con cuidado, para ver si se puede encontrar una razón de su comportamiento extraordinario. A veces, los valores atípicos son "malos" se deben a eventos desacostumbrados, pero explicables. Entre los ejemplos están la medición o el análisis incorrectos, el registro incorrecto de los datos y la falla de un instrumento de medición. Si éste es el caso, el valor atípico

se debería corregir (si es posible) o eliminar del conjunto de datos. Es claro que el eliminar valores malos es conveniente, porque los mínimos cuadrados jalan la ecuación ajustada hacia el valor atípico, ya que eso minimiza la suma de cuadrados de residuales, sin embargo, se hace notar que debe contarse con una fuerte evidencia no estadística de que el valor atípico es malo, para entonces descartarlo. A veces se encuentra que el valor atípico es una observación extraordinaria, pero perfectamente factible. Puede ser peligroso eliminar estos puntos para "mejorar el ajuste de la ecuación", porque puede dar al usuario una sensación falsa de precisión de la estimación o la predicción. A veces se ve que el valor atípico es más importante que el resto de los datos, porque puede controlar muchas propiedades clave del modelo. También, los valores atípicos pueden hacer resaltar inadecuaciones en el modelo, como la falla de tener un buen ajuste con los datos en cierta región del espacio de X . Si el valor atípico es un punto de respuesta especialmente deseable (por ejemplo, bajo costo o alto rendimiento), sería en extremo valioso conocer los valores de los regresores, cuando se observó esa respuesta. Los análisis de identificación y de seguimiento de los valores atípicos con frecuencia dar como resultado mejoras en el proceso, o nuevos conocimientos acerca de factores cuyo efecto sobre la respuesta se desconocía antes. Se han propuesto diversas pruebas estadísticas para detectar y rechazar los valores atípicos. Por ejemplo, basado en el residual máximo normado $|e_i|/\sqrt{\sum e_i^2}$ cuya aplicación es bastante fácil. El efecto de los valores atípicos sobre el modelo de regresión se puede comprobar con facilidad eliminándolos y volviendo a ajustar la ecuación de regresión, teniendo en cuenta que si son observaciones factibles, entonces el modelo no tiene buena predictibilidad para esa región. Se podrá encontrar que los valores de los coeficientes de regresión, o de los estadísticos de resumen como t , F o R^2 , y que el cuadrado medio de residuales pueden ser muy sensibles a los valores atípicos. Los casos en los que un porcentaje relativamente pequeño de los datos tiene un gran impacto sobre el modelo podrán no ser aceptables para el usuario de la ecuación de regresión. En general, uno se siente más cómodo suponiendo que una ecuación de regresión es válida si no es muy sensible a unas pocas observaciones. Se preferiría que la relación de regresión estuviera embebida en todas las observaciones, y no sólo fuera un artificio de unos pocos puntos.

Una forma frecuente de modelar un valor atípico es con el **modelo del valor atípico con media desplazada**. Supóngase que se ajusta el modelo $Y = X\beta + \varepsilon$, cuando el modelo verdadero es

$$Y = X\beta + \delta + \varepsilon$$

donde δ es un factor de $n \times 1$, de ceros excepto por la u -ésima observación, cuyo valor es δ_u . Así,

$$\delta = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Supóngase que tanto para el modelo que se ajusta y el modelo con el valor atípico con media desplazada, $E(\varepsilon) \sim N(0, \sigma^2 I)$. Se trata de determinar un estadístico de prueba adecuado para las hipótesis

$$H_0 : \delta_u = 0 \quad \text{contra} \quad H_1 : \delta_u \neq 0$$

En este procedimiento se supone que el interés específico es la u -ésima observación, es decir, que se cuenta con información a priori que la u -ésima observación puede ser un valor atípico. El primer paso es determinar

un estimador adecuado para δ_u . Un candidato es el u -ésimo residual. Sea $e = [I - H]Y$ el vector de los residuales de $n \times 1$. El valor esperado de e es

$$\begin{aligned} E(e) &= E([I - H]Y) \\ &= [I - H]E(Y) \\ &= [I - H][X\beta + \delta] \\ &= [I - H][X\beta] + [I - H][\delta] \\ &= [X - X]\beta + [I - H][\delta] \\ &= [I - H][\delta] \end{aligned}$$

Así,

$$E(e_u) = (1 - h_{uu})\delta_u$$

siendo h_{uu} el u -ésimo elemento de la diagonal en la matriz sombrero H . En consecuencia, un estimador insesgado de δ_u es

$$\hat{\delta}_u = \frac{e_u}{1 - h_{uu}}$$

Note que esta expresión es equivalente a la expresión en Ec. (78), de tal manera que $\hat{\delta}_u$ es un residual PRESS, cuya varianza es

$$\begin{aligned} V(e) &= V([I - H]Y) \\ &= [I - H]\sigma^2 I [I - H]' \\ &= \sigma^2 [I - H][I - H]' \\ &= \sigma [I - H] \end{aligned}$$

Así, $V(e) = (1 - h_{uu})\sigma^2$. Entonces, la varianza $\hat{\delta}_u$ es

$$\begin{aligned} V\left(\frac{e_u}{1 - h_{uu}}\right) &= \frac{1}{1 - h_{uu}} V(e_u) \\ &= \frac{(1 - h_{uu})\sigma^2}{(1 - h_{uu})^2} \\ &= \frac{\sigma^2}{1 - h_{uu}} \end{aligned}$$

A continuación, se observará que e es una combinación lineal de Y . Así, e es una combinación lineal de variables aleatorias normalmente distribuidas. Por lo anterior, e sigue una distribución normal, al igual que $\hat{\delta}_u$. Entonces, bajo la hipótesis nula, $H_0 : \delta_u = 0$,

$$\frac{e_u/(1 - h_{uu})}{\sigma/(\sqrt{1 - h_{uu}})} = \frac{e_u}{\sigma\sqrt{1 - h_{uu}}}$$

sigue una distribución normal estándar. Se ve que esta cantidad no es más que un ejemplo de un residual estudentizado, como se vio anteriormente. En general, se desconoce σ^2 . Se vio que $\hat{\sigma}^2$ es un estimador insesgado de σ^2 . Además, se vio que $\hat{\sigma}^2/\sigma^2$ tiene una distribución χ^2 , dividida entre sus grados de libertad. En consecuencia, un estadístico probable de prueba es

$$\frac{e_u}{S\sigma_{(u)}\sqrt{(1 - h_{uu})}}$$

Que es el residual estudentizado externamente como se vio en la Ec. (75). Si H_0 es verdadera, el estadístico sigue una distribución central t_{n-p-1} y bajo H_1 , sigue una distribución no central $t'_{n-p-1,\gamma}$, donde

$$\gamma = \frac{\delta_u}{\sigma/\sqrt{1-h_{uu}}} = \frac{\delta_u\sqrt{1-h_{uu}}}{\sigma}$$

Es importante notar que la potencia de esta prueba depende de h_{uu} . Recuérdese que si se ajusta una ordenada al origen al modelo, entonces $1/n \leq h_{uu} \leq 1$. La potencia máxima se tiene cuando $h_{uu} = 1/n$, que está en el centro de la nube de datos, en términos de las X . Cuando $h_{uu} \rightarrow 1$, la potencia baja a 0. En otras palabras, esta prueba tiene menos capacidad de detectar valores atípicos en los puntos de datos de alta influencia.

Apéndices

4. Matrices

Para los cálculos de la regresión lineal multiparamétrica es importante recordar algunas cuestiones importantes de las matrices.

- Sea $A = (a_{ij})$ una matriz de $n \times m$ (n renglones y m columnas), con $i \leq n$ y $j \leq m$, la matriz transpuesta se define como $A' = (a_{ji})$. Es decir, intercambiamos filas (renglones) por columnas. Note que

$$(A')' = (a_{ji})' = (a_{ij})$$

$$\begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}$$

- Si $A = A'$, entonces la matriz A se dice simétrica.
- Si $n = m$ (matriz cuadrada), la diagonal de la matriz son los elementos $a_{11}, a_{22}, \dots, a_{nn}$.
- si la matriz contiene ceros en todos sus elementos excepto en la diagonal, es llamada matriz diagonal.
- Sea A una matriz cuadrada, $B = \text{diag}(A)$ es la matriz diagonal tal que $a_{ii} = b_{ii} \forall i \leq n$.
- Si $(a_{ii} = 1)$ en una matriz diagonal, entonces es llamada matriz identidad.

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Una matriz triangular superior tiene ceros en las entradas por debajo de la diagonal.

$$T = \begin{pmatrix} 7 & 2 & 4 & 5 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

- Una matriz (o vector) unitario es tal que $(a_{ij} = 1) \forall i \leq n, j \leq m$. Análogo para una matriz nula.

$$J = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad O = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Sean $A_{n \times p}$ y $B_{n \times p}$ dos matrices, entonces la suma se define como $C_{n \times p} = A + B = (c_{ij}) = (a_{ij} + b_{ij})$. Análogo para la resta.
- Note que $A + B = B + A$ y $(A + B)' = A' + B'$.
- Sean $A_{n \times p}$ y $B_{p \times m}$ dos matrices, el producto $C = AB$ se define como $c_{ij} = \sum_k a_{ik}b_{kj}$, donde C es de tamaño $n \times m$. Note que no es conmutativa.
- sea c un escalar y A una matriz, $cA = (ca_{ij}) = Ac$.

- Sean $A_{n \times p}$ y $B_{p \times m}$ dos matrices,

$$(AB)' = B'A'.$$

- Sea $A_{n \times p}$ una matriz, entonces $A'A$ y AA' tienen las siguientes propiedades

- $A'A$ es $p \times p$ y se obtiene como producto de columnas de A .
- AA' es $n \times n$ y se obtiene como producto de filas de A .
- AA' y $A'A$ son simétricas.
- Si $A'A = O$, entonces $A = O$, donde O es una matriz nula.

- Sea I la matriz identidad, entonces

$$IA = AI = A.$$

- El rango de una matriz, llamado $\text{ran}(A)$, es el número de columnas (o filas) linealmente independientes.

- Sea $A_{n \times n}$ una matriz no singular y $\text{ran}(A) = n$. Entonces A tiene una única matriz inversa, denotada A^{-1} tal que

$$AA^{-1} = A^{-1}A = I$$

Note que $(A^{-1})^{-1} = A$. Además, se cumple que

- $(A')^{-1} = (A^{-1})'$
- $(AB)^{-1} = B^{-1}A^{-1}$

- Una inversa generalizada de la matriz $A_{n \times p}$ es cualquier matriz A^- que satisfaga

$$AA^-A = A$$

- El determinante de una matriz $A_{n \times n}$, denotado por $|A|$ o $\det(A)$, es una función escalar de A definida como la suma de todos los $n!$ posibles productos de n elementos tal que

- Cada producto contenga un elemento de todas las columnas y filas de A .
- Los factores en cada producto son escritos tal que la columna subscrita aparezca en orden de magnitud y cada producto es precedido, entonces, por un signo positivo o negativo de acuerdo al número de inversiones en la columna es impar o par.

- La traza de una matriz es la suma de todos los elementos de la diagonal. $\text{tr}(A) = \sum_{i=1}^n a_{ii}$.

5. Propiedades

5.1. Consistencia

La consistencia es una propiedad de ciertos estimadores. Se dice que un estimador es consistente cuando éste converge a su valor verdadero cuando el número de datos de la muestra tiende a infinito.

5.2. Propiedades de Varianza

Para el cálculo de la varianza de y , se usa con frecuencia la ecuación de la varianza y el valor esperado condicional, se usa la siguiente propiedad para particionar la variabilidad:

$$\text{var}(y) = \text{var}(E[y|u]) + E[\text{var}(y|u)] \quad (86)$$

5.3. Teorema de Gauss Markov

Lo que buscamos es demostrar que $\hat{\beta}$ minimiza la varianza para cualquier combinación lineal de los coeficientes estimados, $\ell' \hat{\beta}$.

Calculamos la varianza de la combinación lineal $\ell' \hat{\beta}$.

$$\text{Var}(\ell' \hat{\beta}) = \ell' \text{Var}(\hat{\beta}) \ell = \ell' \sigma^2 (X'X)^{-1} = \sigma^2 \ell' (X'X)^{-1} \ell \quad (87)$$

donde el resultado anterior es un escalar.

5.4. Función generadora de momentos

Sea X una variable aleatoria. El valor esperado

$$m_X(t) = E[\exp(tX)] \quad -c \leq t \leq c \quad (88)$$

recibe el nombre de función generadora de momentos.

Si X es una v.a. discreta

$$m_X(t) = E[\exp(tX)] = \sum_x \exp(tx) * p(x) \quad (89)$$

Si X es una v.a. continua

$$m_X(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx) * p(x) \quad (90)$$

6. Distribuciones relacionadas con la distribución normal

6.1. Distribución normal

1. Si la variable aleatoria Y tiene la distribución normal con media μ y varianza σ^2 , su función de densidad de probabilidad es que se denota por $Y \sim N(\mu, \sigma^2)$

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-[y - \mu]^2 / 2\sigma^2\}$$

2. La distribución normal con $\mu = 0$ y $\sigma^2 = 1$, $Y \sim N(0, 1)$, es llamada la **distribución normal estándar**.
3. Sea Y_1, Y_2, \dots, Y_n denota las variables aleatorias distribuidas normalmente con $Y \sim N(\mu_i, \sigma^2)$ para $i = 1, \dots, n$ y sea las covariables de Y_i y Y_j que se denota por

$$\text{Cov}(Y_i, Y_j) = \rho_{ij} \sigma_i \sigma_j,$$

donde ρ_{ij} es el coeficiente de correlación para Y_i y Y_j

4. Suponga que las variables aleatorias Y_1, Y_2, \dots, Y_n son independientes y distribuidos normalmente con las distribuciones $Y \sim N(\mu_i, \sigma_i^2)$ para $i = 1, \dots, n$. Si

$$W = a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n$$

donde los a_i 's son constantes. Entonces W es también distribuida normalmente, así que

$$W = \sum_{i=1}^n a_i Y_i \sim N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$$

6.2. Distribución Chi cuadrada

1. La distribución chi cuadrada centra con n grados de libertad está definida como la suma de cuadrados de n variables aleatorias independientes Z_1, \dots, Z_n cada una con distribución normal estándar. Se denota por

$$X^2 = \sum_{i=1}^n Z_i^2 \sim \chi^2(n).$$

2. Si X^2 tiene la distribución $\chi^2(n)$, entonces el valor esperado es $E(X^2) = n$ y su varianza es $\text{var}(X^2) = 2n$.
3. Si Y_1, Y_2, \dots, Y_n son variables aleatorias independientes distribuidas normalmente cada una con la distribución $Y_i \sim N(\mu_i, \sigma_i^2)$ entonces

$$\sum_{i=1}^n X^2 = \sum_{i=1}^n \left(\frac{Y_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)$$

porque cada variable $Z_i = \frac{Y_i - \mu_i}{\sigma_i}$ tiene una distribución normal $N(0, 1)$.

6.3. Distribución t

La distribución t con n grados de libertad está definida como el ratio de dos variables aleatorias independientes. El numerador tiene distribución estándar y el denominador es la raíz de una v. a. chi- cuadrada central dividida por sus grados de libertad: esto es,

$$T = \frac{Z}{(X^2/n)^{1/2}} \quad \text{denotado por } T \sim t(n).$$

donde $Z \sim N(0, 1)$, $X^2 \sim \chi^2(n)$, tanto Z y X^2 son independientes.

6.4. Distribución F

La distribución central con n y m grados de libertad está definida como el ratio de dos variables aleatorias independientes chi cuadradas centrales cada una dividida por sus grados de libertad

$$F = \frac{X_1^2/n}{X_2^2/m} \quad \text{denotado por } F \sim F(n, m) \quad (91)$$

donde $X_1^2 \sim \chi^2(n)$, $X_2^2 \sim \chi^2(m)$ y X_1^2 y X_2^2 son independientes.

Referencias

- [1] Calderon, JP. de los Godos, LA., *Regresión Logística Aplicada a la Epidemiología*, Rev Salud, Sexualidad y Sociedad, 1(4):78-84, 2009.
- [2] Casella, G. & Berger, R., *Statistical Inference*, Duxbury Advanced Series, 2002.
- [3] Crawley, M.J., *The R book*, Jhon Wiley & Sons, 2012.
- [4] Lindsey, J., *Applying Generalized Linear Models*, Springer-Verlag, 1997.
- [5] Gałecki A. and Burzykowski T., *Linear mixed-effects models using R: A step-by-step approach*, Springer Science & Business Media, 2013.
- [6] Heumann, C., Rao, R.C., and Shalab, Toutenburg, H., *Linear Models and Generalizations: least Squares and Alternatives*, Springer, 2008.
- [7] Hosmer, D. & Lemeshow, S., *Applied Logistic Regression*, Jhon Wiley & Sons, 2000.
- [8] Infante, G.S. y P-Zarate de Lara, G., *Métodos Estadísticos*, Trillas, 2005.
- [9] McCulloch E. C., Searle R. S. and Neuhaus M. J., *Generalized, Linear and Mixed Models*, Jhon Wiley & Sons, 2011.
- [10] Mehtätalo L., *Linear mixed-effects models with examples in R*, University of Eastern Finland, Addison-Wesley, 2013.
- [11] Montgomery, D. C., Peck E.A. and Vining G.G., *Introduction to linear regression Analysis*, Jhon Wiley & Sons, 2011.
- [12] Alvin C. Rencher and G. Bruce Schaalje., *Linear Models In Statistics*, Jhon Wiley & Sons, 2008.
- [13] Norman R. Draper, Harry Smith, *Applied Regression Analysis*, Jhon Wiley & Sons, 1998.
- [14] Dennis D. Wachterly, William Mendenhall III, Richard L. Scheaffer, *Estadística matemática con aplicaciones*, International Thomson Editores, 2002.
- [15] Navarro, E., Verbel A., Robles, D., Hurtado, K. R., *Regresión Logística Ordinal Aplicada a la identificación de factores de riesgo para cáncer de cuello uterino*, Ingeniare, 9(17):87-105, 2014.
- [16] Oroza, H. A., *Modelos mixtos en la determinación del carbono orgánico en la hojarasca en una zona de Teziutlán*, Puebla, tesis de Doctorado-BUAP, 2015.
- [17] Searle S. R. and Khuri A. I., *Matrix algebra useful for statistics*, John Wiley & Sons, 2017.
- [18] Cárdenas Julián, Investigación cuantitativa. Berlin: TrAndes, Programa de Posgrado en Desarrollo Sostenible y Desigualdades Sociales en la Región Andina, 2018. <https://networkianos.com/regresion-logistica-binaria/regresion-logistica-4/> (9 de agosto 2023)
- [19] Team, R. C., *R: A language and environment for statistical computing*, 2013. Cárdenas, Julián (2018) Investigación cuantitativa. Berlin: TrAndes, Programa de Posgrado en Desarrollo Sostenible y Desigualdades Sociales en la Región Andina