# Estimadores y su distribución

## Contents



# 1 Modelo generador de datos

```
d = read.csv("datos/kidiq.csv")
d$mom_hs = factor(d$mom_hs, labels = c("no", "si"))
```

Para el modelo

```
m = lm(kid_score ~ mom_hs, data = d)
summary(m)
```

```
##
## Call:
## lm(formula = kid_score ~ mom_hs, data = d)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -57.55 -13.32   2.68  14.68  58.45
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   77.548      2.059  37.670  < 2e-16 ***
## mom_hssi      11.771      2.322   5.069 5.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.85 on 432 degrees of freedom
## Multiple R-squared:  0.05613,    Adjusted R-squared:  0.05394
## F-statistic: 25.69 on 1 and 432 DF,  p-value: 5.957e-07
```

se tiene que:

- la puntuación media de los chicos cuya madre NO ha terminado secundaria es 77.55
- la puntuación media de los chicos cuya madre SI ha terminado secundaria es $77.55 + 11.77 = 89.32$

Por tanto, la puntuación de los segundos es superior a la de los primeros. Pero este resultado se ha obtenido con los datos de 434 madres y sus hijos. ¿Es generalizable? ¿Si repetimos los cálculos con otras madres e hijos obtendremos las mismas conclusiones? ¿Se puede afirmar que si una madre termina secundaria influye positivamente en la capacidad cognitiva de sus hijos?

Para resolver este problema se hace lo siguiente:

- se considera que el modelo que representa la relación entre puntuación del test y haber acabado secundaria es:

$$kid\_score_i = \beta_0 + \beta_1 mom\_hs_i + U_i, \quad U_i \sim N(0, \sigma)$$

- Este modelo es válido para **todas** las madres y sus hijos. Por tanto, $\beta_0$ representa la puntuación media de **todos** los chicos cuyas madres no terminaron secundaria. Igualmente, $\beta_0 + \beta_1$ representa la puntuación media de **todos** los chicos cuyas madres terminaron secundaria, también de manera general. Estos parámetros no son conocidos.

- se considera que los datos que tenemos, los 434 datos en este caso, han sido generados por el modelo anterior. Es decir, es como si hubiésemos simulado el modelo anterior y hubiésemos obtenido dichos datos particulares.

- estamos interesados en los parámetros del **modelo generador** de los datos, ya que es el que indica conclusiones generales. Pero dichos parámetros son desconocidos y **no se pueden conocer**. Sin embargo, podemos calcular un valor aproximado a partir de los datos observados.

# 2 Estimadores puntuales de los parámetros del modelo

El modelo de la población tiene tres parámetros: $\beta_0$, $\beta_1$ y $\sigma$. Lo primero que haremos es estimar un valor para dichos parámetros.

Se denomina **estimador de un parámetro** a cualquier expresión que asigna un valor a dicho parámetro. Y se denomina **estimación** al valor asignado.

Por ejemplo, podríamos utilizar los siguientes estimadores (se indica poniendo el símbolo $\hat{\square}$):

$$\hat{\beta}_1 = suma(kid\_score_i),\ \hat{\beta}_0 = media(kid\_score_i),\ \hat{\sigma} = Var(kid\_score_i)$$

lo que daría lugar a las siguientes estimaciones:

```
# para beta_0
sum(d$kid_score)
```

```
## [1] 37670
```

```
# para beta_1
mean(d$kid_score)
```

```
## [1] 86.79724
```

```
# para sigma
var(d$kid_score)
```

```
## [1] 416.5962
```

Sin embargo, unos estimadores tienen mejores propiedades que otros. Nosotros vamos a utilizar como estimadores los valores calculados con mínimos cuadrados:

$$kid\_score_i = b_0 + b_1 mom\_hs_i + e_i,\ i = 1, 2, \cdots, n$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X^T X)^{-1}(X^T y)$$

La estimación es:

```r
coefficients(m)
```

```
## (Intercept)    mom_hssi
##    77.54839    11.77126
```

El estimador de $\sigma$ se define como (no es un resultado de mínimos cuadrados):

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k-1}}$$

donde $k$ es el número de regresores, 1 en este caso.

La estimación tomará el valor:

```r
sqrt( sum(resid(m)^2)/(434-2) )
```

```
## [1] 19.85253
```

# 3 Distribución en el muestreo de los estimadores puntuales

## 3.1 Distribución del estimador de $\beta$

El hecho de considerar que nuestros datos han sido generados por un modelo con variables aleatorias hace que los estimadores sean variables aleatorias: si consideramos otra muestra diferente, los valores obtenidos serán diferentes, y antes de seleccionar la muestra no sabemos que valores vamos a obtener.

Como hemos visto en el apartado anterior, la distribución de los estimadores la obtenemos teniendo en cuenta que:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + U_i, \quad U_i \sim N(0, \sigma), \quad i = 1, \cdots, n$$

En forma matricial

$$Y = X\beta + U, \ U \sim N(0, \sigma \mathrm{I}) \Rightarrow$$
$$Y \sim N(X\beta, \sigma \mathrm{I})$$

### 3.1.1 Con matrices de datos

La ecuación de los estimadores es

$$\hat{\beta} = (X^T X)^{-1}(X^T y)$$

Si consideramos variables aleatorias en lugar de datos:

$$\hat{\beta} = (X^T X)^{-1}(X^T Y)$$

obtenemos la distribución de los estimadores:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

ya que

$$Y \sim Normal \Rightarrow \hat{\beta} \sim Normal$$

$$E[\hat{\beta}] = E[(X^TX)^{-1}X^TY] = (X^TX)^{-1}X^TE[Y] = (X^TX)^{-1}X^TX\beta = \beta$$

$$Var[\hat{\beta}] = Var[(X^TX)^{-1}X^TY] = (X^TX)^{-1}X^TVar[Y]((X^TX)^{-1}X^T)^T = (X^TX)^{-1}\sigma^2$$

### 3.1.2 Con matrices de covarianzas

Si en lugar de utilizar las matrices $(X^TX)^{-1}$ y $(X^Ty)$ utilizamos matrices de varianzas y convarianzas, los estimadores son:

$$\hat{\beta}_a = S_{XX}^{-1}S_{Xy} = \left(\frac{1}{n-1}X_a^TX_a\right)^{-1}\left(\frac{1}{n-1}X_a^Ty_a\right)$$

donde $\hat{\beta}_a = [\hat{\beta}_1\ \hat{\beta}_2\ \cdots\ \hat{\beta}_k]^T$. Faltaría el estimador de $\beta_0$ que se obtiene con la ecuación

$$\hat{\beta}_0 = \bar{y} - \bar{x}^T\hat{\beta}_a$$

dónde $\bar{x} = [\bar{x}_1\ \bar{x}_2\ \cdots\ \bar{x}_n]^T$. Se puede demostrar que

$$\hat{\beta}_a \sim N\left(\beta_a, \frac{\sigma^2}{n-1}S_{XX}^{-1}\right), \quad \hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{1}{n-1}\bar{x}^TS_{XX}^{-1}\bar{x}\right)\right)$$

En la primera ecuación $\beta_a = [\beta_1\ \beta_2\ \cdots\ \beta_k]^T$.

## 3.2 Distribución del estimador de $\sigma^2$

El modelo tiene un parámetro más, la varianza de los errores, $\sigma^2$. Este parámetro también hay que estimarlo. Se puede demostrar que

$$\frac{\sum e_i^2}{\sigma^2} \to \chi^2_{n-k-1}$$

donde n es el número de observaciones y k es el número de regresores. Por ello se propone el siguiente estimador

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1}$$

ya que es un estimador insesgado de $\sigma^2$. Efectivamente

$$E[\hat{\sigma}^2] = E\left[\frac{\sum e_i^2}{n-k-1}\right] = \sigma^2$$

ya que $E[\chi^2_n] = n$. Al término $\sum e_i^2/(n-k-1)$ también se lo conoce como **varianza residual** y se representa por $\hat{s}_R^2$.

$$\hat{s}_R^2 = \frac{\sum e_i^2}{n-k-1}$$

A la raiz cuadrada se le conoce como **residual standard error**. El término (n-k-1) son los *grados de libertad*. La distribución en el muestreo de la varianza residual es

$$\frac{\sum e_i^2}{\sigma^2} \rightarrow \chi^2_{n-k-1} \Rightarrow \frac{(n-k-1)\hat{s}^2_R}{\sigma^2} \rightarrow \chi^2_{n-k-1}$$