Modelo con dos regresores

Contents

1 Ecuación del modelo

Ahora se pretende estimar el siguiente modelo con dos regresores:

$$kid\_score_i = b_0 + b_1 mom\_iq_i + b_2 mom\_age_i + e_i$$

2 Modelo en notación matricial

Si escribimos la ecuación para todos los datos disponibles:

$$i = 1 \Rightarrow kid\_score_1 = b_0 + b_1 mom\_iq_1 + b_2 mom\_age_1 + e_1$$

$$i = 2 \Rightarrow kid\_score_2 = b_0 + b_1 mom\_iq_2 + b_2 mom\_age_2 + e_2$$

$$\dots$$

$$i = n \Rightarrow kid\_score_n = b_0 + b_1 mom\_iq_n + b_2 mom\_age_n + e_n$$

Agrupando:

$$\begin{bmatrix} kid\_score_1 \\ kid\_score_2 \\ \dots \\ kid\_score_n \end{bmatrix} = \begin{bmatrix} 1 & mom\_iq_1 & mom\_age_1 \\ 1 & mom\_iq_2 & mom\_age_2 \\ \dots & \dots & \dots \\ 1 & mom\_iq_n & mom\_age_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Finalmente, en notación matricial:

$$y = XB + e$$

donde B es el vector de parámetros estimados:

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

3 Estimacion del modelo

Para aplicar el método de mínimos cuadrados calculamos la suma de los residuos al cuadrado:

$$RSS(B) = \sum e_i^2 = e^T e = (y - XB)^T (y - XB)$$

Y derivando e igualando a cero se obtiene:

$$B = (X^T X)^{-1} X^T y$$

Como se observa, todo es igual que en el modelo con un regresor. Lo único que cambia es la definición de la matriz de regresores X y el vector de parámetros B.

4 Aplicación a los datos del ejemplo

- Datos

```
d = read.csv("datos/kidiq.csv")
```

- Matrices del modelo

La matriz y es la misma que en el ejemplo anterior

```
y = matrix(d$kid_score, ncol = 1)
head(y)
```

```
##      [,1]
## [1,]   65
## [2,]   98
## [3,]   85
## [4,]   83
## [5,]  115
## [6,]   98
```

```
n = nrow(d)
X = cbind(rep(1,n), d$mom_iq, d$mom_age)
head(X)
```

```
##      [,1]      [,2] [,3]
## [1,]    1 121.11753   27
## [2,]    1  89.36188   25
## [3,]    1 115.44316   27
## [4,]    1  99.44964   25
## [5,]    1  92.74571   27
## [6,]    1 107.90184   18
```

- Estimacion

```
Xt_X = t(X) %*% X
Xt_y = t(X) %*% y
( B = solve(Xt_X) %*% Xt_y )
```

```
##              [,1]
## [1,] 17.5962491
## [2,]  0.6035720
## [3,]  0.3881286
```

- valores de la recta

```r
y_e = X %*% B
```

- residuos

```r
e = y - y_e
```

- R2

```r
(RSS = sum(e^2))
```

```
## [1] 143665.4
```

```r
(TSS = sum((y-mean(y))^2))
```

```
## [1] 180386.2
```

```r
(R2 = 1 - RSS/TSS)
```

```
## [1] 0.2035673
```

Como vemos, el valor de R^2 prácticamente no ha variado, lo que nos hace pensar que los dos modelos han extraido la misma cantidad de información de la variable estudiada *kid_score* (a pesar de que el modelo con dos regresores utiliza muchos más datos, los correspondientes a la edad de las madres. A igualdad de R^2 es preferible utilizar el modelo más sencillo).