

Modelo con dos regresores

Contents

1	Ecuación del modelo	1
2	Modelo en notación matricial	1
3	Estimación del modelo	2
4	Aplicación a los datos del ejemplo	2

1 Ecuación del modelo

Ahora se pretende estimar el siguiente modelo con dos regresores:

$$kid_score_i = \beta_0 + \beta_1 mom_iq_i + \beta_2 mom_age_i + e_i$$

2 Modelo en notación matricial

Si escribimos la ecuación para todos los datos disponibles:

$$i = 1 \Rightarrow kid_score_1 = \beta_0 + \beta_1 mom_iq_1 + \beta_2 mom_age_1 + e_1$$

$$i = 2 \Rightarrow kid_score_2 = \beta_0 + \beta_1 mom_iq_2 + \beta_2 mom_age_2 + e_2$$

...

$$i = n \Rightarrow kid_score_n = \beta_0 + \beta_1 mom_iq_n + \beta_2 mom_age_n + e_n$$

Agrupando:

$$\begin{bmatrix} kid_score_1 \\ kid_score_2 \\ \dots \\ kid_score_n \end{bmatrix} = \begin{bmatrix} 1 & mom_iq_1 & mom_age_1 \\ 1 & mom_iq_2 & mom_age_2 \\ \dots & \dots & \dots \\ 1 & mom_iq_n & mom_age_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Finalmente, en notación matricial:

$$y = X\beta + e$$

donde β es el vector de parámetros estimados:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

3 Estimacion del modelo

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

$$RSS(\beta) = \sum e_i^2 = e^T e = (y - X\beta)^T (y - X\beta)$$

Y derivando e igualando a cero se obtiene:

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

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 β .

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
( beta = solve(Xt_X) %*% Xt_y )
```

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

- valores de la recta

```
y_e = X %*% beta
```

- residuos

```
e = y - y_e
```

- R^2

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

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

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

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

```
(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 extraído 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).