

Estimación

Regresión: modelos y métodos

Francesc Carmona Pontaque

PID_00298351



Universitat
Oberta
de Catalunya

Francesc Carmona Pontaque

Cómo citar este recurso de aprendizaje con el estilo Harvard:

Carmona Pontaque, F. (2024) *Estimación. Regresión: modelos y métodos*. [Recurso de aprendizaje textual].

1.^a ed. Barcelona: Fundació Universitat Oberta de Catalunya (FUOC).

Primera edición: febrero 2024

© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)

Av. Tibidabo, 39-43, 08035 Barcelona

Autoría: Francesc Carmona Pontaque

Producción: FUOC

Todos los derechos reservados

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Planteamiento y objetivos

En primer lugar, concretaremos la definición general de un modelo lineal y hallaremos la estimación por mínimos cuadrados de los parámetros del modelo.

Veremos que la estimación será única si la matriz de diseño es de rango máximo. En caso contrario, resulta importante definir el concepto de función paramétrica estimable y probar, para estas funciones, la unicidad del estimador mínimo-cuadrático.

Estudiaremos las propiedades de estos estimadores, entre las que destacaremos el teorema de Gauss-Markov que demuestra que los estimadores mínimo-cuadráticos son los mejores, en el sentido de que son insesgados y de mínima varianza.

Además, con la suposición de normalidad de los errores, podremos estudiar las distribuciones de los estimadores y de otros estadísticos, así como la relación con los estimadores de máxima verosimilitud. En el siguiente módulo trabajaremos el contraste de hipótesis lineales.

1. Un ejemplo de regresión múltiple

Consideremos el ejemplo del número de especies halladas en las islas Galápagos y algunas variables que pueden ser explicativas o regresoras. Los datos de este ejemplo se consiguen así:

```
data(gala, package="faraway")
str(gala)

'data.frame': 30 obs. of 7 variables:
 $ Species : num  58 31 3 25 2 18 24 10 8 2 ...
 $ Endemics : num  23 21 3 9 1 11 0 7 4 2 ...
 $ Area      : num  25.09 1.24 0.21 0.1 0.05 ...
 $ Elevation: num  346 109 114 46 77 119 93 168 71 112 ...
 $ Nearest  : num  0.6 0.6 2.8 1.9 1.9 8 6 34.1 0.4 2.6 ...
 $ Scrutz   : num  0.6 26.3 58.7 47.4 1.9 ...
 $ Adjacent : num  1.84 572.33 0.78 0.18 903.82 ...
```

Lectura complementaria

Ver el apartado 2.6. del libro de Faraway (2014).

Otros ejemplos

Más ejemplos se pueden ver en el apartado 2.2. del libro de Carmona (2005).

En este caso, el modelo lineal de la regresión múltiple es

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad i = 1, \dots, n$$

donde las observaciones y_1, \dots, y_n con $n = 30$ son el número de especies en cada isla y las variables regresoras son todas las demás, excepto `Endemics` que no se considera.

La expresión del modelo lineal en forma matricial es

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

o en forma resumida

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Los elementos que constituyen el modelo lineal son:

1. El vector de observaciones $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$.
2. El vector de parámetros $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$.
3. La matriz del modelo o *matriz de diseño* cuyos elementos son conocidos

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}$$

4. El vector de errores o desviaciones aleatorias $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$, donde ϵ_i es la desviación aleatoria de y_i .

Matriz de diseño

La matriz \mathbf{X} contiene en la fila i los valores de las variables regresoras para los que se ha observado la respuesta y_i . Las columnas de la matriz \mathbf{X} son los valores de cada variable regresora. La primera columna de unos depende de la presencia del parámetro de intercepción β_0 .

A continuación, escribimos el modelo del ejemplo en **R** y vemos las primeras filas de la matriz \mathbf{X} .

```
lmod <- lm(Species ~ Area + Elevation + Nearest + Scrub + Adjacent,
           data = gala)
X.gala <- model.matrix(lmod); head(X.gala)
```

	(Intercept)	Area	Elevation	Nearest	Scrub	Adjacent
Baltra	1	25.09	346	0.6	0.6	1.84
Bartolome	1	1.24	109	0.6	26.3	572.33
Caldwell	1	0.21	114	2.8	58.7	0.78
Champion	1	0.10	46	1.9	47.4	0.18
Coamano	1	0.05	77	1.9	1.9	903.82
Daphne.Major	1	0.34	119	8.0	8.0	1.84

Alternativa

Otra forma de escribir el mismo modelo en **R** es `lmod <- lm(Species ~ ., data = gala[, -2])`. El punto en la fórmula significa considerar como regresoras todas las variables que no son la respuesta.

2. Diseño de casos cruzados o *crossover* simplificado

Supongamos una experiencia clínica en la que se desean comparar dos fármacos, **a** y **b**, para combatir una determinada enfermedad. El estado de los pacientes se valora mediante una cierta variable cuantitativa Y .

En el diseño de casos cruzados, la experiencia se organiza asignando en un primer periodo a N_a pacientes el tratamiento **a** y a N_b pacientes el tratamiento **b**. En un segundo periodo, los que tomaban **a** pasan a tomar **b** y recíprocamente. En este diseño los datos son de la forma:

Grupo 1

a (primera vez) y_{11} y_{12} \dots y_{1N_a}

b (después de **a**) y_{21} y_{22} \dots y_{2N_a}

Grupo 2

b (primera vez) y_{31} y_{32} \dots y_{3N_b}

a (después de **b**) y_{41} y_{42} \dots y_{4N_b}

Si consideramos los parámetros μ = media general, α = efecto fármaco **a**, β = efecto fármaco **b** y γ = efecto recíproco entre **a** y **b**, se propone el siguiente modelo:

a (primera vez) $y_{1i} = \mu + \alpha + \epsilon_{1i}$ $i = 1, \dots, N_a$

b (después de **a**) $y_{2i} = \mu + \beta + \gamma + \epsilon_{2i}$ $i = 1, \dots, N_a$

b (primera vez) $y_{3i} = \mu + \beta + \epsilon_{3i}$ $i = 1, \dots, N_b$

a (después de **b**) $y_{4i} = \mu + \alpha + \gamma + \epsilon_{4i}$ $i = 1, \dots, N_b$

Es decir, cuando solo se ha tomado un fármaco se produce un solo efecto, pero cuando se ha tomado uno después del otro, conlleva entonces un efecto aditivo γ , el cual recoge la mejoría del enfermo que ya ha tomado el primer medicamento.

Tenemos $k = 4$ condiciones experimentales, que en el diseño de casos cruzados simplificado se consideran independientes. El vector de observaciones \mathbf{Y} consiste en poner en columna los cuatro grupos de datos, tal como se observa a continuación:

$$\mathbf{Y} = (y_{11}, \dots, y_{1N_a}, y_{21}, \dots, y_{2N_a}, y_{31}, \dots, y_{3N_b}, y_{41}, \dots, y_{4N_b})'$$

Lectura complementaria

Para profundizar en los diseños *crossover* se puede leer la lección 15 del [curso 509 de la PennState](#).

Simplificación

Una de las simplificaciones que hacemos aquí es considerar que la interacción γ es la misma en ambos casos.

En cuanto a la matriz de diseño \mathbf{X} , se trata de una matriz con cuatro columnas de ceros y unos, una por cada parámetro en el orden $\mu, \alpha, \beta, \gamma$. Sus filas deben representar la situación experimental que proporciona la observación y_{ji} , $j = 1, 2, 3, 4$. Es decir, la matriz \mathbf{X} tiene $N_{\mathbf{a}}$ filas del tipo $(1, 1, 0, 0)$ para los datos respuesta del primer grupo y primera vez, $N_{\mathbf{a}}$ filas del tipo $(1, 0, 1, 1)$ para los datos del primer grupo y segunda vez, $N_{\mathbf{b}}$ filas del tipo $(1, 0, 1, 0)$ para los datos del segundo grupo y primera vez y $N_{\mathbf{b}}$ filas del tipo $(1, 1, 0, 1)$ para los datos del segundo grupo y segunda vez.

Si $n = N_{\mathbf{a}} + N_{\mathbf{a}} + N_{\mathbf{b}} + N_{\mathbf{b}}$, para representar la matriz de diseño \mathbf{X} de orden $n \times 4$ podemos escribir la matriz reducida \mathbf{X}_R de orden 4×4 con una única fila por situación experimental.

$$\mathbf{X}_R = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

¡Atención!

En **R** hay que escribir la matriz de diseño \mathbf{X} completa.

La tabla 1 contiene los datos de dos grupos de 10 y 10 enfermos reumáticos a los que se valoró la variación del dolor respecto del estado inicial, mediante una escala convencional, con el deseo de comparar dos fármacos antirreumáticos **a** y **b**, administrados a lo largo de dos meses.

Grupo 1		Grupo 2	
a (mes 1)	b (mes 2)	b (mes 1)	a (mes 2)
17	17	21	10
34	41	20	24
26	26	11	32
10	3	26	26
19	-6	42	52
17	-4	28	28
8	11	3	27
16	16	3	28
13	16	16	21
11	4	-10	42

Tabla 1: Datos de los enfermos reumáticos

Con los datos de la tabla 1 (ejemplo 5.3.2 del libro de Carmona (2005)) tenemos

```
y <- c(17,34,26,10,19,17,8,16,13,11,      # a (grupo 1, mes 1)
      17,41,26,3,-6,-4,11,16,16,4,      # b (grupo 1, mes 2)
      21,20,11,26,42,28,3,3,16,-10,      # b (grupo 2, mes 1)
      10,24,32,26,52,28,27,28,21,42)     # a (grupo 2, mes 2)
x1 <- c(rep(1,10),rep(0,10),rep(0,10),rep(1,10)) # alpha
x2 <- c(rep(0,10),rep(1,10),rep(1,10),rep(0,10)) # beta
x3 <- c(rep(0,10),rep(1,10),rep(0,10),rep(1,10)) # gamma
cmod <- lm(y ~ x1 + x2 + x3)
X.co <- model.matrix(cmod)
head(X.co)
```

```
(Intercept) x1 x2 x3
1           1  1  0  0
2           1  1  0  0
3           1  1  0  0
4           1  1  0  0
5           1  1  0  0
6           1  1  0  0
```

El modelo `lm()`

En la definición del modelo, el parámetro μ es equivalente al valor de intercepción de la regresión y, por defecto, está representado con una columna de unos.

3. Suposiciones básicas del modelo lineal

En los modelos lineales definidos para los dos ejemplos anteriores, se supone que los errores ϵ_i son desviaciones que se comportan como variables aleatorias que verifican las **condiciones de Gauss-Markov**:

1. Los errores tienen media teórica (esperanza) cero.
2. La varianza σ^2 de todos los errores es la misma.
3. Los errores son incorrelacionados entre sí.

Como sabemos, la condición (2) es la llamada condición de *homocedasticidad* del modelo y el parámetro desconocido σ^2 es la llamada varianza del modelo. La condición (3) se puede endurecer si exigimos que las n desviaciones sean estocásticamente independientes.

Estas condiciones pueden expresarse en forma matricial como

$$E(\epsilon) = \mathbf{0} \quad \text{var}(\epsilon) = \sigma^2 \mathbf{I}_n$$

donde $E(\epsilon)$ es el vector de esperanzas matemáticas y $\text{var}(\epsilon)$ es la matriz de covarianzas de $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$.

Si además suponemos que cada ϵ_i tiene distribución normal $N(0, \sigma)$ y que $\epsilon_1, \dots, \epsilon_n$ son estocásticamente independientes, entonces diremos que el modelo definido es un *modelo lineal normal*. En este caso, podemos asumir que la distribución de $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ es también normal, en la forma:

$$\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$$

Es decir, \mathbf{Y} sigue la distribución normal multivariante de vector de medias $\mathbf{X}\beta$ y matriz de covarianzas $\sigma^2 \mathbf{I}_n$.

Se llama rango del diseño al rango de la matriz \mathbf{X} , $r = \text{rango } \mathbf{X}$, y es un elemento muy importante en la discusión de los modelos. Si m es el número de columnas de \mathbf{X} , evidentemente $r \leq m$. El valor de r es el número efectivo de parámetros del diseño, en el sentido de que si $r < m$ es posible reparametrizar el modelo para que r sea igual al número de parámetros. En muchos casos, el diseño verifica directamente que $r = m$ y entonces se dice que es de *rango máximo*.

Rango de una matriz

Si suponemos que $n > m$, el rango es el número de columnas linealmente independientes. Se dice que una columna es linealmente independiente cuando no es combinación lineal de las otras.

4. Estimación de los parámetros

La estimación de los parámetros $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ se hace con el criterio de los mínimos cuadrados. Se trata de hallar el conjunto de valores de los parámetros $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)'$ que minimicen la siguiente suma de cuadrados

$$\begin{aligned}\epsilon'\epsilon &= \sum_{i=1}^n (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip}\beta_p)^2 \\ &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\end{aligned}$$

La estimación $\hat{\beta}$ de β la llamaremos estimación MC, abreviación de mínimo-cuadrática, o LS del inglés *least squares*.

Toda estimación MC de β es solución del sistema de ecuaciones

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y}$$

que reciben el nombre de *ecuaciones normales*.

Si el rango es máximo y $r = m$, entonces la matriz cuadrada $\mathbf{X}'\mathbf{X}$ tiene inversa y la única solución de las ecuaciones normales es

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

donde $(\mathbf{X}'\mathbf{X})^{-1}$ es la matriz inversa de $\mathbf{X}'\mathbf{X}$.

Si $r < m$ el sistema de ecuaciones es indeterminado y su solución no es única. De hecho, hay infinitas soluciones.

Con los datos de las islas Galápagos y el modelo lineal que hemos planteado, calculamos en primer lugar el rango de la matriz de diseño

```
qr(X.gala)$rank == dim(X.gala)[2]
```

```
[1] TRUE
```

La solución es

Rango de la matriz $\mathbf{X}'\mathbf{X}$

El rango de \mathbf{X} y de $\mathbf{X}'\mathbf{X}$ coinciden. La demostración puede verse en A.2.4. del apéndice A de Seber y Lee (2003).

Cálculo del rango de una matriz

El cálculo numérico del rango de una matriz es un tema delicado. Hay diversos algoritmos que no siempre funcionan correctamente.

```
X <- X.gala
y.gala <- gala$Species
XtX <- t(X) %*% X
coef.gala <- solve(XtX) %*% t(X) %*% y.gala
```

y comprobamos que los resultados coinciden con los de la función `lm()`.

```
cbind(coef(lmod), coef.gala)
```

	[,1]	[,2]
(Intercept)	7.068220709	7.068220709
Area	-0.023938338	-0.023938338
Elevation	0.319464761	0.319464761
Nearest	0.009143961	0.009143961
Scruz	-0.240524230	-0.240524230
Adjacent	-0.074804832	-0.074804832

Veamos ahora qué ocurre en el caso del diseño de datos cruzados simplificado. El rango de la matriz de diseño es

```
qr(X.co)$rank
```

```
[1] 3
```

luego es inferior a $m = 4$, que es el número de parámetros con el que hemos construido el modelo. Así pues, en este caso, las ecuaciones normales tienen infinitas soluciones.

Para salir del mal paso, una posibilidad (ver Apéndice A) es considerar

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{Y}$$

donde $\mathbf{A}^- = (\mathbf{X}'\mathbf{X})^-$ es una g-inversa de $\mathbf{A} = \mathbf{X}'\mathbf{X}$. Sin embargo, hay que ser conscientes de que simplemente hemos hallado una de las infinitas soluciones posibles.

En **R** las instrucciones son

```
library(MASS)
X <- X.co
XtX <- t(X) %*% X
coef.co <- ginv(XtX) %*% t(X) %*% y
```

La solución que da **R** con la función `lm()` es distinta

```
cbind(coef(cmod), coef.co)
```

	[,1]	[,2]
(Intercept)	12.125	11.033333
x1	8.850	9.941667
x2	NA	1.091667
x3	4.150	4.150000

Esta indeseable situación se puede sobrellevar si sabemos, como veremos más adelante, que otros elementos del modelo serán únicos para cualquiera de las soluciones que hallemos en las ecuaciones normales.

5. Propiedades de la estimación de los parámetros

El modelo lineal $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, bajo las hipótesis de Gauss-Markov, verifica

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

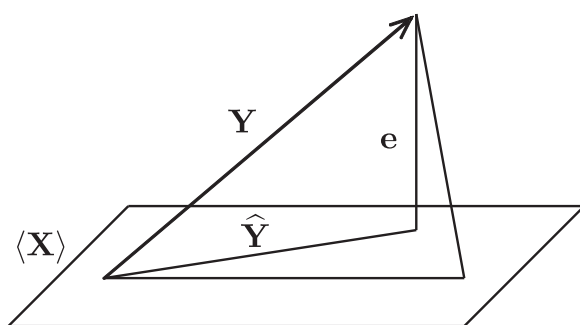
Sea $\Omega = \langle \mathbf{X} \rangle \subset \mathbb{R}^n$ el subespacio vectorial generado por las columnas de \mathbf{X} de dimensión $\dim \langle \mathbf{X} \rangle = r = \text{rango } \mathbf{X}$. Entonces, para cualquier β

$$E(\mathbf{Y}) = \mathbf{X}\beta \in \Omega = \langle \mathbf{X} \rangle$$

ya que es combinación lineal de las columnas de \mathbf{X} . En particular, si $\hat{\beta}$ es una estimación MC,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} \in \Omega = \langle \mathbf{X} \rangle$$

Además, el vector de residuos $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ es ortogonal a $\Omega = \langle \mathbf{X} \rangle$. Este último resultado se puede ver en el gráfico de la figura 1. Se observa que $\hat{\mathbf{Y}}$ es la proyección ortogonal del vector de resultados \mathbf{Y} sobre el subespacio $\langle \mathbf{X} \rangle$.



Comentario

Para seguir el razonamiento de este apartado hay que saber un poco de espacios vectoriales y de álgebra matricial. En caso contrario, simplemente hay que quedarse con los resultados del final.

Lectura complementaria

Las demostraciones rigurosas de estas propiedades se pueden ver en el apartado 2.4. del libro de Carmona (2005).

Figura 1

Representación de la proyección ortogonal del vector de resultados \mathbf{Y} sobre el subespacio $\Omega = \langle \mathbf{X} \rangle$.

La longitud del vector \mathbf{e} es justamente la raíz cuadrada de la suma de cuadrados de los residuos

$$\|\mathbf{e}\| = \sqrt{\sum_{i=1}^n e_i^2}$$

que es mínima para la solución MC. Así, cualquier otra proyección proporcionaría un vector de residuos de mayor longitud. Por otra parte, la proyección ortogonal es única de forma que, sea cual sea $\hat{\beta}$, el vector de predicciones $\hat{\mathbf{Y}}$ es único y los residuos \mathbf{e} también. El vector de predicciones es

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{P}\mathbf{Y}$$

donde la matriz $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es la matriz proyección.

Ahora podemos definir la suma de cuadrados residual (*residual sum of squares* o RSS) como

$$\begin{aligned} \text{SCR} &= \mathbf{e}'\mathbf{e} = \|\mathbf{e}\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 \\ &= (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} \end{aligned}$$

Como veremos, SCR, entendido como una función estadística de la muestra \mathbf{Y} , desempeña un papel fundamental en el análisis de la varianza y el contraste de hipótesis.

Veamos el cálculo de SCR con \mathbf{R} en el diseño *crossover* y su unicidad.

```
SCR1 <- sum(residuals(cmod)^2)
e <- y - X %*% coef.co
SCR2 <- sum(e^2)
c(SCR1, SCR2)

[1] 6147.925 6147.925
```

También disponemos de una función `deviance()` que aplicada a un objeto `lm` nos da directamente SCR.

```
deviance(cmod)

[1] 6147.925
```

La longitud de un vector

En este caso la longitud del vector es su norma euclídea.

La matriz *hat*

En la literatura estadística anglosajona la matriz de proyección se denomina *hat matrix* y se escribe \mathbf{H} . Ver el apartado 2.4. de Faraway (2014).

Otro parámetro desconocido e importante del modelo lineal es la varianza común de los errores. Esta varianza σ^2 se puede estimar con la suma de cuadrados residual con el estadístico

$$\hat{\sigma}^2 = \text{ECM} = \text{SCR}/(n - r)$$

donde n es el tamaño de la muestra y r el rango de \mathbf{X} .

Se trata de un estimador insesgado ya que se demuestra que

$$E(\text{SCR}) = E(\mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y}) = \sigma^2 \text{rg}(\mathbf{I} - \mathbf{P}) = \sigma^2(n - r)$$

gracias a las propiedades de la matriz $\mathbf{I} - \mathbf{P}$.

Con los datos de las islas Galápagos, la estimación de σ (desviación estándar) de los errores se halla en el `summary` del `lm`.

```
summary(lmod)$sigma
```

```
[1] 60.97519
```

Del mismo modo, en el diseño *crossover* tenemos

```
summary(cmod)$sigma
```

```
[1] 12.89031
```

y esta estimación es única.

Mean square error

En los libros escritos en inglés, el estimador $\hat{\sigma}^2$ se denomina *mean square error* o MSE.

Demostración

La demostración se puede ver en el apartado 2.5. de Carmona (2005).

Unidades

La estimación de σ tiene las mismas unidades que la variable respuesta.

En resumen, para cualquier solución $\hat{\beta}$ de las ecuaciones normales, las predicciones, los residuos, la suma de cuadrados residual SCR y la estimación ECM de σ^2 son únicos.

6. Distribuciones de los estimadores

Vamos ahora a establecer algunas propiedades de los estimadores MC para un modelo de rango máximo. El caso de rango no máximo lo trataremos más adelante.

Si asumimos que los errores son insesgados $E(\epsilon) = \mathbf{0}$, que es la primera condición de Gauss-Markov, entonces $\hat{\beta}$ es un estimador insesgado de β

$$E(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta$$

Si asumimos, además, que los errores ϵ_i son incorrelacionados y con la misma varianza, es decir $\text{var}(\epsilon) = \sigma^2\mathbf{I}$, resulta que

$$\text{var}(\mathbf{Y}) = \text{var}(\mathbf{Y} - \mathbf{X}\beta) = \text{var}(\epsilon) = \sigma^2\mathbf{I}$$

ya que $\mathbf{X}\beta$ no es aleatorio y en consecuencia

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

La inversa $\mathbf{X}'\mathbf{X}$ se puede hallar en el summary del objeto lm.

```
XtXinv <- summary(lmod)$cov.unscaled
```

Si multiplicamos esta inversa por la estimación de σ^2 o ECM, obtenemos la matriz de varianzas (en la diagonal) y covarianzas de las estimaciones de los parámetros.

Veamos a continuación algunos resultados acerca de la distribución de $\hat{\beta}$ y SCR bajo las hipótesis del modelo lineal normal en el caso de rango máximo.

Si $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ con rango $\mathbf{X} = m$, entonces

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

Es decir, los estimadores tienen también distribución normal.

Además, se demuestra que la suma de cuadrados residual verifica

$$\text{SCR}/\sigma^2 \sim \chi_{n-m}^2$$

Demostraciones

Las demostraciones de estos resultados se pueden ver en el apartado 2.6. de Carmona (2005).

Eficiencia

La varianza de las estimaciones depende de σ^2 y de la inversa de $\mathbf{X}'\mathbf{X}$.

Consecuencias

Estos resultados nos permiten construir intervalos de confianza para los parámetros. En el caso de los $\hat{\beta}_i$ se basan en la distribución t de Student. En cambio, observemos que el intervalo de confianza para σ^2 se basará en la distribución χ_{n-m}^2 .

Por último, si asumimos la normalidad de las observaciones, nos podemos plantear la utilización de estimaciones de máxima verosimilitud en lugar de las estimaciones mínimo cuadráticas.

La estimación MC de β coincide con la estimación de la máxima verosimilitud. Además es insesgada y de mínima varianza.

La estimación MC de σ^2 es insesgada, mientras que la estimación de máxima verosimilitud $\hat{\sigma}_{MV}^2 = SCR/n$ no lo es.

7. Matrices de diseño con rango no máximo

Cuando el modelo lineal corresponde al análisis de los datos de un diseño experimental, la matriz \mathbf{X} tiene todos sus elementos con valores 0 o 1 y sus columnas acostumbran a ser linealmente dependientes. Es lo que nos ha pasado en el caso del diseño *crossover*. Ya sabemos que en este caso no es posible hallar un estimador MC de β único, ya que, por desgracia, hay infinitas soluciones de las ecuaciones normales. En todo caso y como veremos en el próximo apartado, estamos interesados en concretar un valor para los parámetros β aunque no sea única.

Ya hemos visto que una solución es utilizar una g-inversa de $\mathbf{X}'\mathbf{X}$ en lugar de la inversa que no existe. Sin embargo, esta es una solución elegante pero más bien teórica. En la práctica hay soluciones más sencillas.

Una primera propuesta consiste en eliminar de la matriz de diseño las columnas linealmente dependientes para que la matriz resultante sea de rango máximo. La elección de esas columnas es bastante sencilla en el caso de variables dicotómicas que representan los niveles de un factor. Es lo que hace **R** y la mayoría de programas estadísticos. Por ejemplo, en el diseño *crossover* tenemos

```
coef(cmod)
```

(Intercept)	x1	x2	x3
12.125	8.850	NA	4.150

lo que significa que **R** ha decidido eliminar la columna x2 o, lo que es lo mismo, el parámetro $\beta = 0$. Un detalle importante es que no siempre los diferentes programas estadísticos hacen la misma elección. También hay que tener en cuenta que la interpretación de los parámetros que quedan es distinta.

Otra posibilidad es reparametrizar el modelo para reducir el número de parámetros y que ese número coincida con el rango de la matriz de diseño. Esta solución no se practica mucho, ya que modifica la interpretación de los parámetros. Por ejemplo, en el caso del diseño *crossover* podemos considerar un parámetro $\theta_{\mathbf{a}} = \mu + \alpha$ y otro $\theta_{\mathbf{b}} = \mu + \beta$, de forma que los tres parámetros $\theta_{\mathbf{a}}, \theta_{\mathbf{b}}, \gamma$ forman una matriz de rango 3.

Por último, una solución que se utiliza en diseño de experimentos es añadir una o varias restricciones a los parámetros de forma que el rango de la matriz de diseño sea máximo. Esto se consigue al añadir al modelo lineal filas adicionales que hagan que el rango coincida con el número de columnas de la matriz \mathbf{X} .

Número de restricciones

El número de restricciones sobre los parámetros debe ser $m - r$.

Por ejemplo, en el caso *crossover* hay que añadir $m - r = 4 - 3 = 1$ restricción. Si añadimos la restricción $\alpha + \beta = 0$ tenemos

¡Atención!

En la definición del objeto `lm` debemos poner `0 + Xnew` ya que la columna de unos ya está en la matriz `Xnew`.

```
Xnew <- rbind(X.co, c(0, 1, 1, 0))  
y.new <- c(y, 0)  
qr(Xnew)$rank
```

```
[1] 4
```

```
cmod.new <- lm(y.new ~ 0 + Xnew)  
coef(cmod.new)
```

Xnew(Intercept)	Xnewx1	Xnewx2	Xnewx3
16.550	4.425	-4.425	4.150

Observemos que la solución verifica la restricción propuesta. Además, por lo que hemos visto, algunos elementos del modelo lineal son únicos como los residuos, las predicciones o la estimación de σ .

```
summary(cmod.new)$sigma
```

```
[1] 12.89031
```

8. Funciones paramétricas estimables

8.1. Definición y caracterizaciones

En los modelos lineales, además de la estimación de los parámetros β_i y de σ^2 , interesa también la estimación de ciertas funciones lineales de los parámetros β_i . Como vamos a ver, esto es especialmente necesario cuando los parámetros carecen de una estimación única.

Llamaremos *función paramétrica* a toda función lineal ψ de los parámetros

$$\psi = a_0\beta_0 + a_1\beta_1 + \cdots + a_m\beta_p = \mathbf{a}'\boldsymbol{\beta}$$

y diremos que una función paramétrica ψ es *estimable* si existe un estadístico $\hat{\psi}$, combinación lineal de las observaciones y_1, \dots, y_n

$$\hat{\psi} = b_1y_1 + \cdots + b_ny_n = \mathbf{b}'\mathbf{Y}$$

tal que

$$E(\hat{\psi}) = \psi$$

es decir, $\hat{\psi}$ es estimador lineal insesgado de ψ .

Es evidente que nos interesan las funciones paramétricas estimables (FPE), ya que se pueden estimar con un estimador lineal insesgado, pero no sabemos cuáles son. La siguiente caracterización lo clarifica:

Una función paramétrica $\psi = \mathbf{a}'\boldsymbol{\beta}$ es estimable si y solo si el vector fila \mathbf{a}' es combinación lineal de las filas de \mathbf{X} .

Cuando el rango de la matriz \mathbf{X} es máximo, todas las funciones paramétricas ψ son estimables. Sin embargo, cuando el rango no es máximo habrá que ir con cuidado y trabajar únicamente con las estimables.

Demostración

Ver la demostración del teorema 3.1.1. de Carmona (2005).

En la práctica

Para saber si ψ es estimable, basta con añadir la fila \mathbf{a}' a la matriz \mathbf{X} y comprobar que el rango se mantiene.

Por ejemplo, en el diseño *crossover* de apartados anteriores sabemos que el rango es 3. Consideremos la función paramétrica α , es decir

$$\mathbf{a}'\boldsymbol{\beta} = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{pmatrix}$$

entonces

```
X.a <- rbind(X.co, c(0,1,0,0))
qr(X.a)$rank

[1] 4
```

y el rango no es 3, luego no es estimable.

Ahora vamos a probar con $\psi = \alpha - \beta$, es decir

$$\mathbf{a}'\boldsymbol{\beta} = \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{pmatrix}$$

entonces

```
X.a <- rbind(X.co, c(0,1,-1,0))
qr(X.a)$rank

[1] 3
```

y el rango se mantiene, luego $\alpha - \beta$ es estimable.

Ahora nos queda calcular el estimador y estudiar sus propiedades.

8.2. Propiedades

En primer lugar, cuando el rango de la matriz de diseño no es máximo y, por tanto, la estimación MC de los parámetros no es única, la estimación de

Otra caracterización

Se puede comprobar que $\psi = \mathbf{a}'\boldsymbol{\beta}$ es estimable si y solo si $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{a}'$.

cualquier función paramétrica estimable, utilizando cualquiera de los estimadores MC, sí es única.

Si $\psi = \mathbf{a}'\beta$ es una función paramétrica estimable y $\hat{\beta}$ es un estimador MC de β , entonces el estimador $\hat{\psi} = \mathbf{a}'\hat{\beta}$ de ψ es único.

Demostración

Ver la demostración del teorema 3.2.1. de Carmona (2005).

Así pues, para estimar la FPE $\psi = \alpha - \beta$ en el diseño *crossover*, podemos utilizar cualquiera de las soluciones MC que tenemos.

```
cc <- coef(cmod)
cc[is.na(cc)] <- 0
sum(c(0,1,-1,0) * cc)

[1] 8.85

sum(c(0,1,-1,0) * coef.co)

[1] 8.85
```

¡Atención!

En los coeficientes del objeto `lm` debemos sustituir los NA por ceros.

A continuación, se destaca la principal ventaja de la utilización de los estimadores MC.

Si $\psi = \mathbf{a}'\beta$ es una función paramétrica estimable y $\hat{\beta}$ es un estimador MC de β , entonces $\hat{\psi} = \mathbf{a}'\hat{\beta}$ es el estimador de varianza mínima (BLUE: *best linear unbiased estimate*) en la clase de los estimadores lineales insesgados de ψ .

Teorema de Gaus-Markov

Esta propiedad y la anterior justifican la imposición de las condiciones que hemos propuesto al modelo lineal y la utilización de la estimación MC.

Estos resultados son válidos incluso para un modelo lineal sin la hipótesis de normalidad.

Demostración

Ver la demostración del teorema 3.2.2. de Carmona (2005).

La varianza mínima del estimador es

$$\text{var}(\hat{\psi}) = \text{var}(\mathbf{a}'\hat{\beta}) = \sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}$$

Si sustituimos σ^2 por su estimación, para la FPE $\psi = \alpha - \beta$ en el *crossover* tenemos

```
sigma2 <- summary(cmod)$sigma^2
a <- c(0,1,-1,0)
var.psi <- sigma2 * t(a) %*% ginv(XtX) %*% a
se.psi <- sqrt(var.psi)
```

El último valor corresponde al error estándar (*standard error*) del estimador de ψ .

Recordemos que el mismo resultado corresponde a los parámetros de un modelo lineal de rango máximo. Por ejemplo, en el caso del modelo de las islas Galápagos, el error estándar del parámetro de la variable Area es

```
sigma2 <- summary(lmod)$sigma^2
a <- c(0,1,0,0,0,0)
var.psi <- sigma2 * t(a) %*% XtXinv %*% a # sigma2 * XtXinv[2,2]
se.psi <- sqrt(var.psi)
```

Una situación muy problemática se da cuando \mathbf{X} tiene rango máximo pero sus columnas están cerca de ser linealmente dependientes. Entonces $\mathbf{X}'\mathbf{X}$ está cerca de ser singular (no inversible), en el sentido de que uno o varios de sus valores propios no nulos son excesivamente pequeños, casi despreciables. Y en consecuencia, las estimaciones de algunos parámetros serán muy imprecisas.

La presencia de relaciones casi lineales entre las variables regresoras se conoce con el nombre de *multicolinealidad*. Este grave problema debe ser detectado previamente a la estimación y se puede corregir de varias formas.

Errores estándar

La segunda columna del `summary` de un `lm` proporciona los errores estándar de las estimaciones de los parámetros.

Un ejemplo

En Faraway (2014, p. 27) se muestra un ejemplo de lo que pasa cuando añadimos la variable `Adiffe` que es casi una combinación lineal de `Area` y `Adjacent`.

Bibliografía

Carmona, F. (2005) *Modelos lineales*. e-UMAB, Universitat de Barcelona.

Faraway, J.J. (2014) *Linear Models with R*. 2.^a ed. Chapman and Hall/CRC.

Seber, G.A.F. and Lee, A.J. (2003) *Linear Regression Analysis*. John Wiley & Sons.

Verzani, J. (2002) *simpleR - Using R for Introductory Statistics*. College of Staten Island. Disponible en: <https://biostat.jhsph.edu/~iruczins/teaching/Rresources/simpleR.pdf>

Verzani, J. (2014) *Using R for Introductory Statistics*. 2.^a ed. Chapman and Hall/CRC.

Apéndice A: Inversa generalizada

Diremos que \mathbf{A}^- es una inversa generalizada o g-inversa de una matriz \mathbf{A} si

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

Más detalles

Ver el Apéndice A de Carmona (2005).

Un caso especial de g-inversa es la llamada inversa de Moore-Penrose \mathbf{A}^+ de \mathbf{A} que verifica

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \quad \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \quad \mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})' \quad \mathbf{A}\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^+)'$$

En R

La función `ginv()` del paquete MASS calcula la g-inversa de Moore-Penrose.

La inversa de Moore-Penrose es única.