

## Capítulo 3

---

### Identificación. Colinealidad exacta

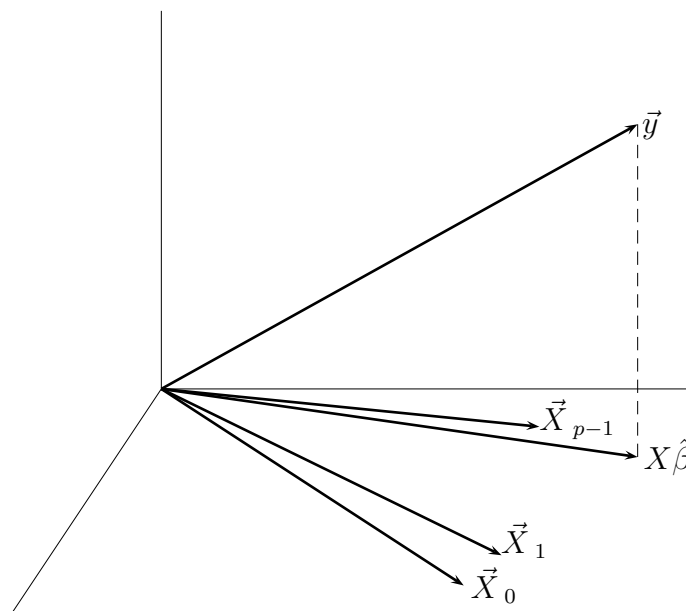
---

#### 3.1. Modelos con matriz de diseño de rango deficiente.

Uno de los que hemos llamado supuestos habituales (Sección 2.3, pág. 7, apartados 1 a 3) es que el rango de la matriz de diseño  $X$  coincide con el número de sus columnas,  $p$ . Cuando ésto no ocurre, sigue habiendo una única proyección de  $\vec{y}$  sobre  $M = R(X)$ , tal como ha quedado demostrado. (Recuérdese que  $R(X)$  designa el subespacio generado por las columnas de  $X$ .) Ocurre sin embargo (Lema 3.9) que  $\hat{\beta} = (X'X)^{-1}X'\vec{y}$  no es único.

La Figura 4.1 resulta iluminante a este respecto; el plano horizontal representa  $M$ , y en él yacen los vectores  $\vec{X}_0, \dots, \vec{X}_{p-1}$  que lo generan. La proyección  $X\hat{\beta}$  es única. Si  $\vec{X}_0, \dots, \vec{X}_{p-1}$  son linealmente independientes, forman base del espacio que generan, y los coeficientes  $\hat{\beta}_0, \dots, \hat{\beta}_{p-1}$  que permiten expresar  $P_M\vec{y}$  como combinación lineal de dichos vectores son únicos.

Si, como acontece en el caso de rango deficiente de la matriz  $X$ , los vectores  $\vec{X}_0, \dots, \vec{X}_{p-1}$  no son linealmente independientes, hay infinidad de maneras de expresar  $P_M\vec{y}$  como combinación lineal de ellos. No hay por tanto una única estimación mínimo cuadrática del vector  $\vec{\beta}$ . Se dice que hay *multicolinealidad exacta* entre las columnas de la matriz de diseño  $X$ .

Figura 3.1: Regresión en el caso de matrix  $X$  de rango deficiente.

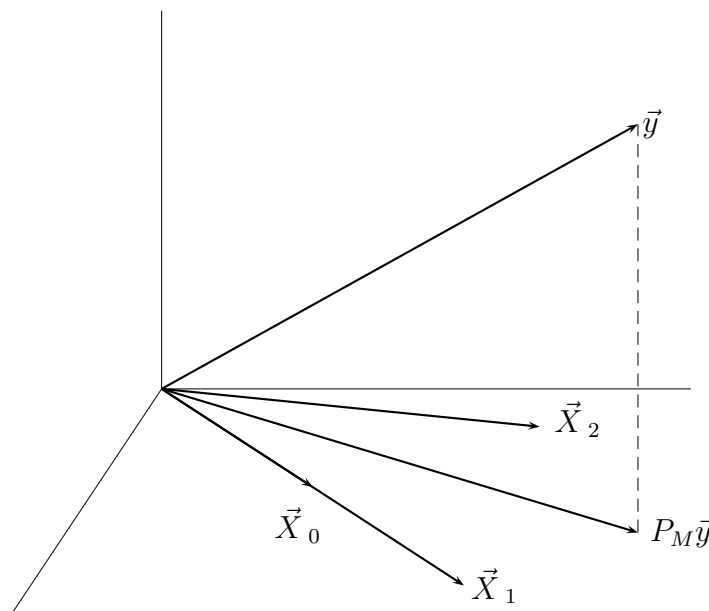
Una matriz de diseño de rango deficiente es demasiado “pobre” para deslindar todos los efectos de interés: no podemos con la información disponible deslindar la relación de cada uno de los regresores con la variable respuesta, pero puede ocurrir que si lo podamos deslindar con algunos. El Ejemplo 4.1 a continuación lo ilustra.

**Ejemplo 3.1** Imaginemos una matriz de diseño como

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 5 \\ 2 & 4 & 1 \\ 2 & 4 & 7 \\ 1 & 2 & 8 \\ 1 & 2 & 4 \end{pmatrix}.$$

Observemos que la primera columna,  $\vec{X}_0$ , es igual a la segunda,  $\vec{X}_1$ , dividida entre dos. La Figura 4.2 ilustra una situación similar. Puede verse que  $\vec{X}_0$  y  $\vec{X}_1$  yacen uno sobre otro, difiriendo sólo en el módulo.

En un caso así, la proyección,  $P_M \vec{y}$ , puede expresarse de manera única como combinación lineal de  $\vec{X}_2$  y *uno* de los vectores  $\vec{X}_0$  ó  $\vec{X}_1$ . Podemos estimar  $\beta_2$ , pero no  $\beta_0$  ó  $\beta_1$ : no es posible adscribir a uno de ellos la “parte” de  $P_M \vec{y}$  colineal con la dirección común de  $\vec{X}_0$  y  $\vec{X}_1$ .

Figura 3.2: Caso de un vector  $\vec{\beta}$  parcialmente estimable.

FIN DEL EJEMPLO ■

La noción de *función estimable* a continuación permite caracterizar situaciones como la mostrada en el ejemplo anterior.

### 3.2. Funciones estimables.

Incluso aunque el vector  $\vec{\beta}$  no sea estimable por no estar  $\hat{\beta}$  unívocamente determinado, puede haber algunos parámetros o combinaciones lineales de parámetros que sí puedan estimarse.

**Definición 3.1** Decimos que una función lineal de los parámetros  $\vec{a}'\vec{\beta}$  es estimable si existe un vector  $\vec{c}$  de constantes tal que:

$$E[\vec{c}'\vec{Y}] = \vec{a}'\vec{\beta}$$

El Teorema a continuación permite caracterizar las funciones estimables.

**Teorema 3.1** La función lineal  $\vec{a}'\vec{\beta}$  es estimable si  $\vec{a} \in R(X')$ .

DEMOSTRACIÓN:

$$\vec{a}'\vec{\beta} = E[\vec{c}'\vec{Y}] = E[\vec{c}'(X\vec{\beta} + \vec{\epsilon})] = \vec{c}'X\vec{\beta} \quad (3.1)$$

Como (4.1) ha de verificarse para cualesquiera valores de  $\vec{\beta}$ , ha de existir  $\vec{c}$  tal que:  $\vec{c}'X = \vec{a}'$ , lo que demuestra que  $\vec{a} \in R(X')$ . ■

**Observación 3.1** El teorema anterior incluye como caso particular el de parámetros aislados,  $\beta_i$ . En efecto, podemos ver  $\beta_i$  como la función lineal  $\vec{e}'_{i+1}\vec{\beta}$ , en que  $\vec{e}_i$  es un vector de ceros con un 1 en posición  $i$ -ésima. Entonces,  $\beta_i$  es estimable si  $\vec{e}_i \in R(X')$ . La totalidad de los parámetros serán estimables si  $\{\vec{e}_1, \dots, \vec{e}_p\}$  (que son linealmente independientes) están en  $R(X')$ . Esto requiere que la dimensión de  $R(X')$  sea  $p$ , es decir, que  $X$  sea de rango completo.

**Observación 3.2** El enunciado del Teorema 4.1 tiene gran contenido intuitivo. Son estimables aquéllas combinaciones lineales de los parámetros cuyos coeficientes coinciden con los dados por filas de  $X$ . En efecto, si queremos estimar  $\vec{a}'\vec{\beta}$  y  $\vec{a}'$  coincide con la  $j$ -ésima fila  $\vec{x}_j'$  de la matriz  $X$ , es claro que  $Y_j$  sería un estimador insesgado de  $\vec{a}'\vec{\beta}$ , pues:

$$E[Y_j] = E[\vec{x}_j'\vec{\beta} + \epsilon_j] = E[\vec{a}'\vec{\beta} + \epsilon_j] = \vec{a}'\vec{\beta}.$$

De manera análoga se demuestra que si  $\vec{a}$  puede expresarse como combinación lineal de filas de  $X$ , la combinación lineal análoga de observaciones en el vector  $\vec{Y}$  es un estimador insesgado de  $\vec{a}'\vec{\beta}$ .

### 3.3. Restricciones de identificación.

Hemos visto que la inestimabilidad de los parámetros es consecuencia de la indeterminación del sistema de ecuaciones normales:

$$(X'X)\hat{\beta} = X'\vec{y}$$

Si contamos con información adicional sobre  $\vec{\beta}$  que podamos imponer sobre el vector de estimadores  $\hat{\beta}$ , podemos añadir al anterior sistema ecuaciones adicionales que reduzcan o resuelvan la indeterminación. Por ejemplo, si supiéramos que  $A\vec{\beta} = \vec{c}$ , podríamos formar el sistema:

$$(X'X)\hat{\beta} = X'\vec{y} \quad (3.2)$$

$$A\hat{\beta} = \vec{c} \quad (3.3)$$

y, dependiendo del rango de  $X'X$  y  $A$ , obtener estimaciones únicas de  $\vec{\beta}$ . Se dice entonces que las relaciones  $A\vec{\beta} = \vec{c}$  son *restricciones de identificación*.

**Ejemplo 3.2** Retomemos el Ejemplo 4.1. Vimos que  $\vec{\beta}$  era parcialmente estimable, y que el problema residía en que la componente de  $P_M\vec{y}$  colineal con la dirección (común) de  $\vec{X}_0$  y  $\vec{X}_1$  no puede ser “distribuida” entre ambos. Si, no obstante, supiéramos que  $\beta_0 = 1$ , el problema dejaría de existir. Por tanto,  $A\vec{\beta} = 1$  con

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

es una restricción de identificación.

FIN DEL EJEMPLO ■

Una matriz de diseño de rango incompleto se puede presentar por falta de cuidado al diseñar el experimento, pero, más frecuentemente, es intencional. El Ejemplo 4.1 ilustra este punto.

**R: Ejemplo 3.1** Supongamos que se investiga el efecto de tres diferentes tratamientos térmicos sobre la dureza de un acero. Podemos pensar en el modelo:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon; \quad (3.4)$$

Habremos de realizar mediciones de la dureza con varias probetas de acero elaborado con los distintos tratamientos, y estimar dicho modelo. La variable explicativa o regresor  $i$ -ésimo tomará el valor 1 cuando se emplee el tratamiento  $i$ -ésimo, y cero en caso contrario. Con esta especificación  $\beta_i$ , ( $i = 1, 2, 3$ ), se interpretará como la dureza estimada derivada de utilizar el tratamiento  $i$ -ésimo. Consideremos los datos siguientes:

```
> cbind(X, y)
      [,1] [,2] [,3]    [,4]
[1,]    1    0    0 4.8150
[2,]    1    0    0 4.3619
[3,]    1    0    0 4.3579
[4,]    0    1    0 4.8403
[5,]    0    1    0 5.2419
[6,]    0    1    0 6.2087
[7,]    0    0    1 3.9853
[8,]    0    0    1 4.0601
[9,]    0    0    1 3.4247
```

Podemos estimar los parámetros mediante

```
> ajuste1 <- lsfit(X, y, intercept = FALSE)
> ajuste1$coefficients

      X1      X2      X3
4.5116 5.4303 3.8234

> ajuste1$residuals

[1] 0.30342 -0.14972 -0.15371 -0.58995 -0.18841
[6] 0.77837  0.16193  0.23672 -0.39865

> SSE <- sum(ajuste1$residuals^2)
> SSE

[1] 1.3687
```

Podríamos pensar, sin embargo, en adoptar una diferente parametrización:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon; \quad (3.5)$$

En esta nueva parametrización,  $\beta_0$  sería una dureza “media” y  $\beta_1$  a  $\beta_3$  recogerían el efecto diferencial (respecto de dicha dureza “media”) resultado de emplear cada uno de los tres tratamientos. Para introducir en el modelo  $\beta_0$  multiplicando a una columna de “unos”, basta omitir el argumento `intercept=FALSE`, con lo que obtenemos:

```
> ajuste2 <- lsfit(X, y, intercept = TRUE)
> ajuste2$coefficients

Intercept      X1      X2      X3
 3.82339  0.68824  1.60690  0.00000

> ajuste2$residuals

[1] 0.30342 -0.14972 -0.15371 -0.58995 -0.18841
[6] 0.77837  0.16193  0.23672 -0.39865

> SSE <- sum(ajuste2$residuals^2)
> SSE

[1] 1.3687
```

Observemos que los dos ajustes son idénticos, como muestran los residuos, que son iguales, y  $SSE = 1.3687$ , igual en los dos casos; resultado lógico, dado que los subespacios que generan  $\vec{X}_1, \dots, \vec{X}_3$  y estos tres vectores más la columna de “unos” son idénticos. Las proyecciones han de serlo también.

En el segundo ajuste, `lsfit` ha proporcionado *una* estimación de los parámetros, a pesar de que el rango de la matriz  $X$  ampliada con una columna de “unos” es incompleto. `lsfit` ha tomado una restricción identificadora arbitraria —ha hecho  $\beta_3 = 0$ — y proporcionado *una* de las infinitas soluciones equivalentes.

La restricción adoptada hace  $\beta_3 = 0$ . El tratamiento 3 pasa así a convertirse en *caso de referencia* y la dureza atribuible al mismo viene medida por  $\hat{\beta}_0 = 3.8234$ . Los valores estimados  $\hat{\beta}_1$  y  $\hat{\beta}_2$  miden así las diferencias de dureza de los tratamientos 1 y 2 *respecto del caso de referencia, o tratamiento 3*.

Podríamos adoptar restricciones de identificación diferentes. Una muy habitual sería, en el caso que nos ocupa,  $\beta_1 + \beta_2 + \beta_3 = 0$ . Esto equivale a forzar que los efectos diferenciales de los tres tratamientos no puedan ser todos positivos o negativos. Con esta restricción,  $\beta_0$  tendría la interpretación de “dureza media” y  $\beta_1, \beta_2, \beta_3$  serían desviaciones respecto de esta dureza media.

FIN DEL EJEMPLO ■

### 3.4. Multicolinealidad exacta y aproximada

La existencia de dependencia lineal “exacta” entre las columnas de la matriz de diseño  $X$ , es, como se ha visto, fruto habitualmente de una decisión consciente. Escogemos un diseño de rango incompleto, pero lo suplementamos con restricciones de identificación que solventan el problema de la estimación y dotan a los parámetros de la interpretación que deseamos.

En la medida en que la matriz  $X$  sea de nuestra elección, siempre podemos eludir el problema. Si, por el contrario, no podemos diseñar nuestro experimento y nos vemos obligados a utilizar unos datos  $X, \vec{y}$  dados, puede ocurrir que la matriz  $X$ , aunque no precisamente de rango incompleto, proporcione una matriz  $(X'X)$  “casi” singular. Esto se traduce en dificultades numéricas para resolver las ecuaciones normales, dificultades para seleccionar un modelo adecuado, grandes varianzas de los estimadores y otros inconvenientes a los que nos referiremos en el Capítulo 10.

### **3.5. Lectura recomendada.**

Pueden verse Seber (1977), Sección 3.8, o Draper and Smith (1998), Sección 20.4, por ejemplo.