

Capítulo 2

Estimación mínimo cuadrática.

2.1. Obtención de los estimadores de los parámetros.

Si \vec{y} es un vector $N \times 1$, consideremos $H = R^N$ y $M =$ subespacio generado por las columnas de X . Si dotamos a H del producto interno euclídeo $\langle \vec{v}, \vec{w} \rangle = \vec{v}'\vec{w}$, de las Secciones 2.4 y 2.5 inmediatamente se deduce que el vector en M más próximo a \vec{y} (en el sentido de minimizar la norma al cuadrado del vector de residuos $\hat{\epsilon}$) es la proyección de \vec{y} sobre M . Por consiguiente, ha de verificarse que $(\vec{y} - X\hat{\beta}) \perp M$. Como M es el subespacio generado por las columnas de X ,

$$\vec{X}_0 \perp (\vec{y} - X\hat{\beta}) \quad (2.1)$$

$$\vec{X}_1 \perp (\vec{y} - X\hat{\beta}) \quad (2.2)$$

$$\vdots \quad \quad \quad \vdots \quad (2.3)$$

$$\vec{X}_{p-1} \perp (\vec{y} - X\hat{\beta}) \quad (2.4)$$

que podemos reunir en la igualdad matricial

$$X'(\vec{y} - X\hat{\beta}) = \vec{0}$$

y de aquí se deduce que:

$$X'X\hat{\beta} = X'\vec{y}. \quad (2.5)$$

La igualdad matricial anterior recoge las *ecuaciones normales*. Si, como suponemos, $\text{rango}(X) = p$, entonces $(X'X)$ es de rango completo, y posee inversa. Por tanto, el vector de estimadores de los parámetros será:

$$\hat{\beta} = (X'X)^{-1}X'\vec{y}. \quad (2.6)$$

Obsérvese que el supuesto de rango total de la matriz X —y consiguientemente de $(X'X)$ — es requerido exclusivamente para pasar de (3.5) a (3.6). Las ecuaciones normales se verifican en todo caso, y la proyección de \vec{y} sobre M es también única (Teorema 2.1, pág. 10). El defecto de rango en X tiene tan solo por consecuencia que el vector $\hat{\beta}$ deja de estar unívocamente determinado. Volveremos sobre esta cuestión al hablar de multicolinealidad.

De (3.6) se deduce también que, en el caso de rango total, la proyección de \vec{y} sobre M viene dada por

$$P_M\vec{y} = X(X'X)^{-1}X'\vec{y}, \quad (2.7)$$

y el vector de residuos por

$$\hat{\epsilon} = \vec{y} - X\hat{\beta} \quad (2.8)$$

$$= \vec{y} - X(X'X)^{-1}X'\vec{y} \quad (2.9)$$

$$= (I - X(X'X)^{-1}X')\vec{y} \quad (2.10)$$

$$= (I - P_M)\vec{y}. \quad (2.11)$$

Observación 2.1 El ser $X\hat{\beta}$ proyección de \vec{y} sobre M garantiza sin más que $\|\hat{\epsilon}\|$ es mínimo. Si hubiéramos obtenido $\hat{\beta}$ derivando

$$\sum_i \left(y_i - \hat{\beta}_0 x_{i0} - \hat{\beta}_1 x_{i1} - \dots - \beta_{p-1} x_{i,p-1} \right)^2$$

e igualando las derivadas a cero (ver Observación 2.3, pág. 7), obtendríamos un $\hat{\beta}$ del que todo lo que podríamos afirmar es que corresponde a un punto estacionario de la expresión anterior (suma de cuadrados de los residuos). Para establecer que se trata de un mínimo, habríamos de tomar aún segundas derivadas y verificar el cumplimiento de las condiciones de segundo orden.

Podemos ver $X\hat{\beta}$ y $\hat{\epsilon}$ como las proyecciones de \vec{y} sobre dos espacios mutuamente ortogonales: M y M^\perp . Las matrices P_M e $(I - P_M)$ que, para aligerar la notación, denominaremos en lo sucesivo P e $(I - P)$, sobreentendiendo el subespacio M , tienen algunas propiedades que detallamos a continuación.

Teorema 2.1 Sean P e $(I - P)$ las matrices de proyección definidas en el párrafo anterior. Se verifica lo siguiente:

1. Las matrices P e $(I - P)$ son simétricas e idempotentes.
2. $\text{rango}(I - P) = N - p$.
3. Se verifica que $(I - P)X = 0$.

DEMOSTRACIÓN:

El apartado 1) es inmediato. En cuanto a 2), siendo $(I - P)$ idempotente, su rango coincide con su traza (véase Teorema A.1, pág. 229). Por tanto:

$$\text{rango}(I - P) = \text{traza}(I - P) \quad (2.12)$$

$$= \text{traza}(I) - \text{traza}(P) \quad (2.13)$$

$$= N - \text{traza}[X(X'X)^{-1}X'] \quad (2.14)$$

$$= N - \text{traza}[(X'X)^{-1}X'X] \quad (2.15)$$

$$= N - p. \quad (2.16)$$

El apartado 3), por último, se prueba sin más que efectuar el producto matricial indicado. Es además inmediato si reparamos en que la matriz $(I - P)$ proyecta sobre el subespacio M^\perp , por lo que su producto por cualquiera de los vectores columna de X (pertenecientes a M) da el vector $\vec{0}$. ■

2.2. Una obtención alternativa

La obtención del vector de estimadores $\hat{\beta}$ en la sección precedente tiene muchos méritos, y no es el menor el de proporcionar intuición geométrica acerca de la solución mínimo cuadrática ordinaria (MCO). Tendremos ocasiones abundantes de explotar esta intuición.

Podemos seguir una vía alternativa para llegar al mismo resultado: plantear el problema en forma de minimización respecto a $\vec{\beta}$ de la expresión:

$$\sum_{i=1}^N (y_i - \beta_0 x_{i0} - \beta_1 x_{i1} - \dots - \beta_{p-1} x_{i,p-1})^2, \quad (2.17)$$

tal como sugería la Observación 3.1. Con notación matricial, el problema puede reescribirse así:

$$\min_{\vec{\beta}} (\vec{y} - X\vec{\beta})'(\vec{y} - X\vec{\beta}). \quad (2.18)$$

La “suma de cuadrados” anterior es una forma cuadrática de matriz unidad. Haciendo uso de la fórmula (A.12), pág. 231, obtenemos las condiciones de primer orden

$$2X'(\vec{y} - X\vec{\beta}) = \vec{0}, \quad (2.19)$$

o equivalentemente

$$X'\vec{y} = (X'X)\vec{\beta}, \quad (2.20)$$

que son las ecuaciones normales (3.5).

Es fácil comprobar tomando las segundas derivadas que la solución (o soluciones, si hay más de una) del sistema de ecuaciones precedente corresponde a un mínimo y no a un máximo o punto de silla: la matriz de segundas derivadas ($X'X$) es por construcción (semi)definida positiva.

Importa comprobar que esta aproximación al problema, a diferencia de la que hacía uso de la noción de proyección, deja en la penumbra muchas cosas que son de interés: la ortogonalidad del vector de residuos $\hat{\epsilon} = \vec{y} - X\hat{\beta}$, la idempotencia de algunas matrices, etc.

2.3. Propiedades del estimador mínimo cuadrático $\hat{\beta}$.

Notemos que $\hat{\beta}$ es un vector aleatorio. Aunque X se mantenga fija — cosa que podemos lograr, pues los valores de los regresores se fijan por el experimentador: recuérdese los supuestos introducidos en la Sección 2.2—, en experimentos repetidos obtendremos cada vez un diferente vector \vec{y} de valores de la variable respuesta. En efecto, cada vez intervendrán en la formación de \vec{y} diferentes perturbaciones.

El vector $\hat{\beta} = (X'X)^{-1}X'\vec{y}$ por tanto es un vector aleatorio: “hereda” su condición de tal de \vec{y} , que a su vez la obtiene de $\vec{\epsilon}$. Tiene por ello sentido preguntarse por su vector de valores medios y por su matriz de covarianzas.

Recordemos que un estimador $\hat{\gamma}$ del parámetro γ se dice *insesgado* si

$$E[\hat{\gamma}] = \gamma.$$

En el caso de estimar un vector de parámetros, la condición análoga es

$$E[\hat{\beta}] = \vec{\beta}.$$

Recordemos también que la *matriz de covarianzas* de un vector aleatorio como $\hat{\beta}$ se define por:

$$\Sigma_{\hat{\beta}} = E[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]',$$

2.3. PROPIEDADES DEL ESTIMADOR MÍNIMO CUADRÁTICO $\hat{\beta}$. 21

expresión que en el caso de ser $\hat{\beta}$ insesgado como estimador de $\vec{\beta}$ se simplifica de modo obvio a

$$\Sigma_{\hat{\beta}} = E[\hat{\beta} - \vec{\beta}][\hat{\beta} - \vec{\beta}]'$$

La matriz de covarianzas $\Sigma_{\hat{\beta}}$ tiene en su diagonal principal las varianzas de los componentes del vector $\hat{\beta}$ y fuera de la diagonal principal las covarianzas.

La insesgadería de un estimador es intuitivamente atrayente: supone que no incurrimos en derivas sistemáticas al estimar el parámetro objeto de interés. Si repitiéramos el mismo experimento muchas veces y promediáramos los valores del estimador insesgado obtenidos en cada experimento, esperaríamos que este promedio se acercará progresivamente más a su objetivo (el verdadero valor del parámetro).

Acontece que el vector de estimadores $\hat{\beta}$ disfruta de esta atractiva propiedad de insesgadería. Adicionalmente, dentro de una clase particular de estimadores es el que exhibe menores varianzas en la diagonal principal de $\Sigma_{\hat{\beta}}$ —y, en este sentido, es el que estima con mayor precisión el vector $\vec{\beta}$ —. El siguiente Teorema formaliza y demuestra estas propiedades.

Teorema 2.2 *Si se verifican los supuestos habituales (Sección 2.3, pág. 7) se cumple también que:*

1. $\hat{\beta}$ es un estimador lineal insesgado de $\vec{\beta}$.
2. La matriz de covarianzas de $\hat{\beta}$ es $\Sigma_{\hat{\beta}} = \sigma^2(X'X)^{-1}$.
3. (Gauss-Markov). Si $\hat{\beta}$ es el estimador mínimo cuadrático ordinario de $\vec{\beta}$, cualquier otro estimador $\hat{\beta}_*$ de $\vec{\beta}$ que sea lineal e insesgado tiene matriz de covarianzas con elementos diagonales no menores que los de $\Sigma_{\hat{\beta}}$.

DEMOSTRACIÓN:

Tomando valor medio en (3.6):

$$\begin{aligned} E[\hat{\beta}] &= E[(X'X)^{-1}X'\vec{y}] \\ &= E[(X'X)^{-1}X'(X\vec{\beta} + \vec{\epsilon})] \\ &= \vec{\beta} + E[(X'X)^{-1}X'\vec{\epsilon}] \\ &= \vec{\beta}. \end{aligned}$$

luego $\hat{\beta}$ es insesgado. Por consiguiente, la matriz de covarianzas $\Sigma_{\hat{\beta}}$ tendrá por expresión:

$$\begin{aligned}
\Sigma_{\hat{\beta}} &= E(\hat{\beta} - \vec{\beta})(\hat{\beta} - \vec{\beta})' \\
&= E[(X'X)^{-1}X'(X\vec{\beta} + \vec{\epsilon}) - \vec{\beta}][(X'X)^{-1}X'(X\vec{\beta} + \vec{\epsilon}) - \vec{\beta}]' \\
&= E[(X'X)^{-1}X'\vec{\epsilon}][(X'X)^{-1}X'\vec{\epsilon}]' \\
&= E[(X'X)^{-1}X'\vec{\epsilon}\vec{\epsilon}'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'\sigma^2IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}$$

Para demostrar 3), consideremos cualquier estimador $\hat{\beta}_*$ alternativo a $\hat{\beta}$. Dado que restringimos nuestra atención a estimadores lineales, podemos escribir $\hat{\beta}_* = C\vec{Y}$, siendo C una matriz de orden adecuado. Siempre podremos expresar C así:

$$C = (X'X)^{-1}X' + D. \quad (2.21)$$

Puesto que nos limitamos a considerar estimadores insesgados, ha de verificarse: $E\hat{\beta}_* = EC\vec{Y} = \vec{\beta}$, y por tanto: $E[(X'X)^{-1}X' + D]\vec{Y} = \vec{\beta}$. De aquí se deduce:

$$E[(X'X)^{-1}X'(X\vec{\beta} + \vec{\epsilon}) + D(X\vec{\beta} + \vec{\epsilon})] = \vec{\beta}, \quad (2.22)$$

$$\vec{\beta} + DX\vec{\beta} = \vec{\beta}, \quad (2.23)$$

dado que $E\vec{\epsilon} = \vec{0}$. Como (3.23) se ha de verificar sea cual fuere $\vec{\beta}$, la insesgaredad de $\hat{\beta}_*$ implica $DX = 0$.

La matriz de covarianzas de $\hat{\beta}_*$ es:

$$\Sigma_{\hat{\beta}_*} = E[(\hat{\beta}_* - \vec{\beta})(\hat{\beta}_* - \vec{\beta})']. \quad (2.24)$$

Pero:

$$(\hat{\beta}_* - \vec{\beta}) = [(X'X)^{-1}X' + D]\vec{Y} - \vec{\beta} \quad (2.25)$$

$$= [(X'X)^{-1}X' + D](X\vec{\beta} + \vec{\epsilon}) - \vec{\beta} \quad (2.26)$$

$$= [(X'X)^{-1}X' + D]\vec{\epsilon}. \quad (2.27)$$

donde (3.27) se ha obtenido haciendo uso de $DX = 0$. Llevando (3.27) a (3.24), obtenemos:

$$\Sigma_{\hat{\beta}_*} = E\{[(X'X)^{-1}X' + D]\vec{\epsilon}\vec{\epsilon}'[(X'X)^{-1}X' + D]'\} \quad (2.28)$$

que, de nuevo haciendo uso de que $DX = 0$, se transforma en:

$$\Sigma_{\hat{\beta}_*} = (X'X)^{-1}X'\sigma^2IX(X'X)^{-1} + \sigma^2DID' \quad (2.29)$$

$$= \sigma^2(X'X)^{-1} + \sigma^2DD' \quad (2.30)$$

$$= \Sigma_{\hat{\beta}} + \sigma^2DD'. \quad (2.31)$$

La matriz DD' tiene necesariamente elementos no negativos en la diagonal principal (sumas de cuadrados), lo que concluye la demostración de 3). De forma completamente similar se puede demostrar una versión ligeramente más general: la estimación lineal insesgada con varianza mínima de cualquier forma lineal $\vec{c}'\vec{\beta}$ es $\vec{c}'\hat{\beta}$, siendo $\hat{\beta}$ el vector de estimadores mínimo cuadráticos. ■

Observación 2.2 La insesgades de un estimador es una propiedad en principio atrayente, pero de ningún modo indispensable. De hecho, un estimador insesgado de un parámetro puede incluso no existir. (Para una discusión de la condición de insesgades y de sus implicaciones puede verse Lehmann (1983), Cap. 2.)

En el Capítulo 11 comprobaremos que, en ocasiones, podemos optar con ventaja por utilizar estimadores sesgados.

2.4. Estimación de la varianza de la perturbación.

El Teorema 3.2 proporciona la matriz de covarianzas del vector de estimadores $\hat{\beta}$, $\Sigma_{\hat{\beta}} = \sigma^2(X'X)^{-1}$. Pero mientras que $(X'X)$ es conocida, σ^2 es un parámetro que necesita ser estimado. Veamos como hacerlo.

Definición 2.1 Denominamos *SSE* o suma de cuadrados de los residuos al cuadrado de la norma del vector de residuos,

$$SSE \stackrel{def}{=} \|\vec{y} - X\hat{\beta}\|^2 = \|\hat{\epsilon}\|^2$$

Teorema 2.3 Una estimación insesgada de la varianza de la perturbación viene proporcionada por

$$\hat{\sigma}^2 = \frac{SSE}{N - p}$$

DEMOSTRACIÓN:

Como

$$X\hat{\beta} = P\vec{Y} = X(X'X)^{-1}X'\vec{Y}, \quad (2.32)$$

tenemos que

$$(\vec{Y} - X\hat{\beta}) = (I - P)\vec{Y} \quad (2.33)$$

$$= (I - P)(X\vec{\beta} + \vec{\epsilon}) \quad (2.34)$$

$$= (I - P)\vec{\epsilon}, \quad (2.35)$$

y por tanto

$$SSE = \vec{Y}'(I - P)'(I - P)\vec{Y} = \vec{\epsilon}'(I - P)'(I - P)\vec{\epsilon}.$$

En virtud de la simetría e idempotencia de $(I - P)$,

$$SSE = \vec{\epsilon}'(I - P)\vec{\epsilon} \quad (2.36)$$

$$= \text{traza } \vec{\epsilon}'(I - P)\vec{\epsilon} \quad (2.37)$$

$$= \text{traza } (I - P)\vec{\epsilon}\vec{\epsilon}'. \quad (2.38)$$

Tomando valor medio en (3.38) tenemos:

$$E(SSE) = \text{traza}(I - P)(\sigma^2 I) = \sigma^2(N - p). \quad (2.39)$$

(El último paso ha hecho uso de la propiedad $\text{traza}(I - P) = N - p$, Teorema 3.1, pág. 18.) De (3.39) se deduce entonces que

$$E\left[\frac{SSE}{N - p}\right] = \sigma^2$$

y $\hat{\sigma}^2 \stackrel{\text{def}}{=} SSE/(N - p)$ es por tanto un estimador insesgado de σ^2 . ■

Observación 2.3 En lo que sigue, SSE denotará tanto la variable aleatoria definida más arriba como su valor en una experimentación concreta, contra la convención habitual con otras variables en que se emplean minúsculas para denotar sus valores en una experimentación. El contexto aclarará si nos estamos refiriendo a una variable aleatoria o a un valor experimental de la misma.

Observación 2.4 El Teorema 3.3 muestra que para obtener una estimación insesgada de la varianza de la perturbación debemos dividir la suma de cuadrados de los residuos, no entre el número de residuos N , sino entre los *grados de libertad* $N - p$. Que el número de parámetros estimado debe tomarse en consideración en el denominador del estimador es intuitivamente plausible. Después de todo, si aumentáramos el número de regresores (y parámetros estimados) p hasta que $p = N$, SSE sería idénticamente cero. (Estaríamos ante un problema *sin grados de libertad*.) Sin llegar a este extremo, es claro que aumentando el número de regresores incrementamos nuestra capacidad de aproximar \vec{y} (y de reducir SSE), y esto ha de ser contrapesado reduciendo también el denominador.

Observación 2.5 El Teorema 3.3 subsume y amplía un resultado que habitualmente aparece sin demostración en los cursos elementales de Estadística: un estimador insesgado de la varianza de una población, dada una muestra i.i.d. de la misma, viene dada por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}. \quad (2.40)$$

Este resultado puede obtenerse como caso particular del Teorema 3.3 si reparamos en lo siguiente: podemos imaginar las Y_i como generadas por

$$Y_i = \beta_0 + \epsilon_i,$$

en que β_0 es la media y ϵ_i una perturbación de media cero y misma varianza que Y_i . Si regresáramos las observaciones Y_1, \dots, Y_N sobre una columna de “unos”, $\vec{1}$, el único parámetro estimado sería:

$$\hat{\beta}_0 = (X'X)^{-1}X'\vec{Y} = (\vec{1}'\vec{1})^{-1}\vec{1}'\vec{Y} = N^{-1}\sum_{i=1}^N Y_i = \bar{Y}$$

El mejor ajuste que puede hacerse de las Y_i en términos de este único regresor es $\hat{\beta}_0\vec{1}$ y la suma de cuadrados de los residuos es por tanto $\sum_{i=1}^N (Y_i - \hat{\beta}_0\vec{1})^2 = \sum_{i=1}^N (Y_i - \bar{Y})^2$. La expresión (3.40) coincide por tanto, en este caso particular, con la dada por el Teorema 3.3.

R: Ejemplo 2.1 (*cálculo de los estimadores MCO*)

El siguiente listado crea artificialmente una matriz X y el vector respuesta \vec{y} . A continuación, realiza la regresión de dos formas. En la primera, se realizan los cálculos de modo explícito. En la segunda, se recurre a la función `lsfit` predefinida en R, que simplifica considerablemente el trabajo. Existen funciones alternativas más avanzadas que se introducen más adelante.

Al margen de la comodidad, `lsfit` realiza los cálculos de un modo mucho más eficiente en tiempo y estable numéricamente que el sugerido por la teoría: no se invierte la matriz $(X'X)$ sino que se emplea la factorización QR (ver Sección D.2, pág. 245, o Lawson and Hanson (1974)). Se trata de detalles que no necesitan preocuparnos por el momento. Generamos en primer lugar los datos y realizamos la estimación aplicando la teoría de modo más directo. Primero, la matriz de diseño,

```
> X <- matrix(c(1, 1, 1, 1,
+             1, 1, 1, 4, 12, 1, 4,
+             13, 0, 6, 7, 0, 2, 2),
+            6, 3)
> X
```

```
      [,1] [,2] [,3]
[1,]    1    1    0
[2,]    1    4    6
[3,]    1   12    7
[4,]    1    1    0
[5,]    1    4    2
[6,]    1   13    2
```

A continuación, fijamos un vector $\vec{\beta}$

```
> beta <- c(2, 3, 4)
```

Finalmente, generamos los valores de la variable respuesta del modo que prescribe el modelo lineal:

```
> y <- X %*% beta + rnorm(6)
```

(La función `rnorm(n)` genera n variables aleatorias $N(0,1)$.) A continuación, obtenemos los estimadores resolviendo las ecuaciones normales (3.5), pág. 17. Se muestran varias formas alternativas de hacerlo. Podemos por ejemplo escribir

```
> b <- solve(t(X) %*% X, t(X) %*%
+           y)
> b
```

```

      [,1]
[1,] 2.3517
[2,] 2.8129
[3,] 4.2329

```

(la función `solve(A,b)` proporciona una solución, si existe, del sistema de ecuaciones lineales $A\vec{x} = \vec{b}$). Una forma más rápida de calcular $(X'X)$ y $X'\vec{y}$ la proporciona la función `crossprod`. Podríamos sustituir lo anterior por

```

> b <- solve(crossprod(X),
+           crossprod(X, y))
> b

```

```

      [,1]
[1,] 2.3517
[2,] 2.8129
[3,] 4.2329

```

Podemos también escribir:

```

> XXinv <- solve(crossprod(X))
> b <- XXinv %*% crossprod(X,
+ y)
> b

```

```

      [,1]
[1,] 2.3517
[2,] 2.8129
[3,] 4.2329

```

Hemos obtenido separadamente $(X'X)^{-1}$ (que puede servirnos para estimar la matriz de covarianzas de los estimadores, $\hat{\sigma}^2(X'X)^{-1}$). La función `solve` con un único argumento matricial proporciona la matriz inversa. De cualquiera de las maneras que calculemos $\hat{\beta}$, la obtención de los residuos es inmediata:

```

> e <- y - X %*% b
> e

```

```

      [,1]
[1,]  0.42097
[2,] -0.29124
[3,]  0.15416
[4,] -0.61805
[5,]  0.53689
[6,] -0.20272

```

Podemos comprobar la ortogonalidad de los residuos a las columnas de la matriz X :

```

> t(e) %*% X
      [,1]      [,2] [,3]
[1,] -9.1482e-14 1.3589e-13  0
> crossprod(e, X)
      [,1]      [,2] [,3]
[1,] -9.1482e-14 1.3589e-13  0
> round(crossprod(e, X))
      [,1] [,2] [,3]
[1,]    0    0    0

```

La suma de cuadrados de los residuos y una estimación de la varianza de la perturbación pueden ahora obtenerse con facilidad:

```

> s2 <- sum(e * e)/(nrow(X) -
+          ncol(X))
> s2
[1] 0.33238

```

FIN DEL EJEMPLO ■

R: Ejemplo 2.2 Todos los cálculos anteriores pueden hacerse con mucha mayor comodidad mediante funciones de regresión especializadas. Por ejemplo,

```

> ajuste <- lsfit(X, y, intercept = FALSE)

```

hace todo lo anterior y algunas cosas más de modo mucho más eficiente. La función `lsfit` (least squares **fit**) devuelve una lista u objeto compuesto conteniendo en sus componentes los estimadores de los parámetros, los residuos y algunos resultados auxiliares asociados al método de cálculo empleado (la factorización QR aludida más arriba). Veámoslo:

```
> ajuste

$coefficients
      X1      X2      X3
2.3517 2.8129 4.2329

$residuals
[1]  0.42097 -0.29124  0.15416
[4] -0.61805  0.53689 -0.20272

$intercept
[1] FALSE

$qr
$qt
[1] -75.33003  48.78812 -23.94068
[4]  -0.66854   0.42874  -0.60529

$qr
      X1      X2
[1,] -2.44949 -14.28869
[2,]  0.40825  11.95129
[3,]  0.40825 -0.63322
[4,]  0.40825  0.28718
[5,]  0.40825  0.03616
[6,]  0.40825 -0.71690
      X3
[1,] -6.940221
[2,]  3.583992
[3,] -5.655823
[4,] -0.375532
[5,] -0.004607
[6,]  0.047314

$qraux
[1] 1.4082 1.0362 1.9256
```

```

$rank
[1] 3

$pivot
[1] 1 2 3

$tol
[1] 1e-07

attr(,"class")
[1] "qr"

> resid <- ajuste$residuals
> resid

[1] 0.42097 -0.29124 0.15416
[4] -0.61805 0.53689 -0.20272

```

El argumento `intercept=FALSE` indica a la función `lsfit` que *no* debe agregarse a la matriz de diseño X una columna de “unos” (porque ya figura entre los regresores). Ordinariamente ello no sucederá, y podremos prescindir de especificar el argumento `intercept`, con lo que tomará el valor por omisión `TRUE`.

FIN DEL EJEMPLO ■

2.5. El coeficiente R^2

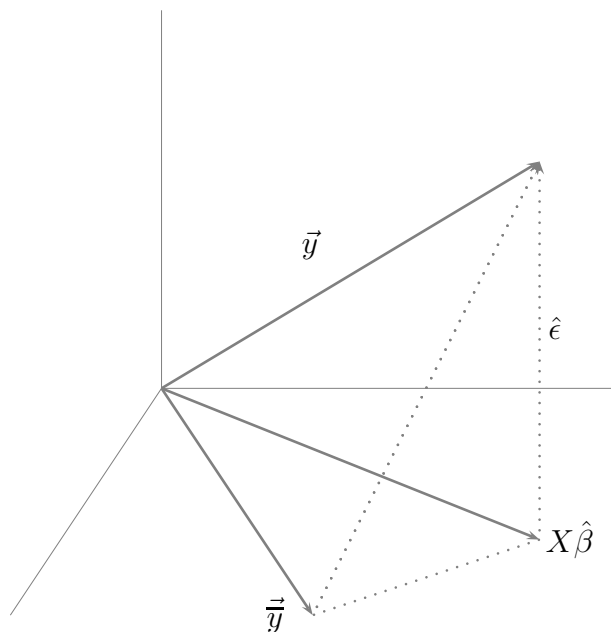
Hay una relación interesante entre SSE y otras dos sumas de cuadrados que definimos a continuación. Sea $\vec{\bar{y}}$ el vector $N \times 1$ siguiente:

$$\vec{\bar{y}} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}$$

en que \bar{y} denota la media aritmética de las observaciones en \vec{y} . Definamos:

$$\begin{aligned} SST &= \|\vec{y} - \vec{\bar{y}}\|^2 \\ SSR &= \|X\hat{\beta} - \vec{\bar{y}}\|^2 \end{aligned}$$

Se verifica entonces el Teorema a continuación.

Figura 2.1: $X\hat{\beta}$ es la proyección de \vec{y} sobre M . $R^2 = \cos^2 \alpha$ 

Teorema 2.4 Si \vec{y} pertenece al subespacio M generado por las columnas de la matriz X —lo que acontece, por ejemplo, siempre que dicha matriz tiene una columna de “unos”—, se verifica:

$$SST = SSR + SSE \quad (2.41)$$

DEMOSTRACIÓN:

$$SST = \|\vec{y} - \vec{\bar{y}}\|^2 \quad (2.42)$$

$$= \|\vec{y} - X\hat{\beta} + X\hat{\beta} - \vec{\bar{y}}\|^2 \quad (2.43)$$

$$= \langle (\vec{y} - X\hat{\beta}) + (X\hat{\beta} - \vec{\bar{y}}), (\vec{y} - X\hat{\beta}) + (X\hat{\beta} - \vec{\bar{y}}) \rangle \quad (2.44)$$

$$= \|\vec{y} - X\hat{\beta}\|^2 + \|X\hat{\beta} - \vec{\bar{y}}\|^2 + 2 \langle \vec{y} - X\hat{\beta}, X\hat{\beta} - \vec{\bar{y}} \rangle \quad (2.45)$$

Pero si $\vec{\bar{y}} \in M$, $(X\hat{\beta} - \vec{\bar{y}}) \in M$, y como quiera que $\hat{\epsilon} = (\vec{y} - X\hat{\beta}) \perp M$, el último producto interno es nulo. Por consiguiente (3.45) se reduce a (3.41).

Definimos $R^2 = SSR/SST$; se denomina a R *coeficiente de correlación múltiple*. Claramente, $0 \leq R^2 \leq 1$, siempre que X contenga una columna constante, ya que de (3.41) se obtiene:

$$\frac{SSR}{SST} = \frac{SSR}{SST} + \frac{SSE}{SST},$$

luego $1 = R^2 + \frac{SSE}{SST}$, y como ambos sumandos son no negativos (son cocientes de sumas de cuadrados), R^2 necesariamente ha de tomar valores entre 0 y 1.

La igualdad (3.41) es fácil de visualizar con ayuda de la ilustración esquemática en la Fig. 3.1; es una generalización N -dimensional del teorema de Pitágoras. Obsérvese que si \vec{y} no perteneciera a M , que hemos representado como el plano horizontal, ya no podría asegurarse que $\hat{\epsilon}$ y $(X\hat{\beta} - \vec{y})$ son ortogonales.

Observación 2.6 En la Figura 3.1 puede visualizarse R^2 como el coseno al cuadrado del ángulo que forman los vectores $(\vec{y} - \bar{y})$ y $(X\hat{\beta} - \vec{y})$. Un valor “pequeño” de R^2 significa que este coseno es “pequeño”, y el ángulo correspondiente “grande”; es decir, que \vec{y} está muy elevado sobre el plano M . Por el contrario, R^2 grande implica que el ángulo referido es pequeño, y que \vec{y} está próximo a su proyección en M .

Observación 2.7 Si regresamos \vec{y} solamente sobre una columna de “unos”, obtenemos un único coeficiente de regresión estimado, $\hat{\beta}_0$ que resulta ser igual a \bar{y} (se comprobó en la Observación 3.5, pág. 25). SST puede interpretarse como la suma de cuadrados de los residuos de este modelo mínimo.

Si regresamos \vec{y} sobre varios regresores *incluyendo la columna de “unos”* obtenemos una suma de cuadrados de los residuos igual a SSE que nunca puede ser superior a SST . En efecto: al añadir regresores el ajuste no puede empeorar (*¿por qué?*). El coeficiente R^2 puede verse como una medida de la mejora en el ajuste atribuible a los regresores distintos de la columna de “unos”. En efecto, el numerador de R^2 es $SST - SSE$, diferencia de suma de cuadrados entre el modelo ampliado y el mínimo. El denominador SST meramente normaliza el numerador anterior para que tome valores entre 0 y 1.

Un valor “grande” de R^2 podemos interpretarlo como una mejora sustancial del modelo mínimo al incluir regresores distintos de la columna de “unos”. Obsérvese que para que esta interpretación sea válida, uno de los modelos (el mínimo) ha de estar anidado en el otro, es decir, su único regresor (la columna de “unos”) ha de estar entre los regresores del otro.

Observación 2.8 Si ajustamos un modelo sin columna de “unos” podemos encontrarnos con que R^2 definido como en el Teorema 3.4 puede ser menor que cero. Es fácil de entender: puede que los regresores ensayados no den cuenta de la variabilidad de \vec{y} , y SSE sea por tanto grande. Si acontece que \vec{y} tiene poca variabilidad en torno a su media, SST será en cambio pequeño, y $SST - SSE$ puede fácilmente ser negativo.

Observación 2.9 Cuando no hay columna de “unos” algunos programas de ordenador automáticamente sustituyen SST por

$$\|\vec{y}\|^2$$

(suma de cuadrados de las desviaciones *respecto del origen* en lugar de respecto a la media). Ello da lugar a una definición alternativa de R^2 que evita que pueda ser negativa.

2.6. Algunos lemas sobre proyecciones.

Los siguientes resultados, de muy sencilla prueba en la mayoría de los casos, resultan útiles en demostraciones posteriores.

Lema 2.1 *Sea H un espacio vectorial, y M un subespacio. Todo $\vec{y} \in H$ tiene expresión única en la forma: $\vec{y} = \vec{u} + \vec{v}$, con $\vec{u} \in M$ y $\vec{v} \in M^\perp$.*

DEMOSTRACIÓN:

Es una consecuencia inmediata de la unicidad de la proyección (Teorema 2.1, pág. 10). ■

Lema 2.2 *Prefijadas las bases en H y $M \subseteq H$, la aplicación lineal que proyecta sobre M tiene por asociada una única matriz P_M .*

DEMOSTRACIÓN:

Es una especialización del resultado según el cual, prefijadas las bases en ambos espacios, la matriz que representa una aplicación lineal de uno en otro es única. La proyección es una aplicación lineal (véase solución al Ejercicio 2.14). ■

Lema 2.3 *La matriz de proyección sobre M puede ser expresada así:*

$$P_M = TT',$$

siendo T una matriz cuyas columnas forman una base ortonormal de $M \subset H$.

DEMOSTRACIÓN:

Sea N la dimensión de H y p la dimensión de M . Sea $\vec{v}_1, \dots, \vec{v}_p$ una base de M formada por vectores ortonormales, y T la matriz $N \times p$ siguiente:

$$T = (\vec{v}_1 \mid \vec{v}_2 \mid \dots \mid \vec{v}_p)$$

Siempre podemos completar $\{\vec{v}_1, \dots, \vec{v}_p\}$ con $N - p$ vectores adicionales $\{\vec{v}_{p+1}, \dots, \vec{v}_N\}$ hasta obtener una base de H (véase por ej. Grafe (1985), pág. 79). Además, los $N - p$ vectores adicionales pueden tomarse ortogonales entre sí y a los de T , y normalizados (por ejemplo, utilizando el procedimiento de ortogonalización de Gram-Schmidt; véase Grafe (1985), pág. 93). Entonces, para cualquier $\vec{y} \in H$ tendremos:

$$\vec{y} = \underbrace{\sum_{i=1}^p c_i \vec{v}_i}_{\in M} + \underbrace{\sum_{j=p+1}^N c_j \vec{v}_j}_{\in M^\perp}, \quad (2.46)$$

siendo c_i ($i = 1, \dots, N$) las coordenadas de \vec{y} en la base escogida. Premultiplicando ambos lados de (3.46) por \vec{v}_i' ($i = 1, \dots, p$), obtenemos:

$$\vec{v}_i' \vec{y} = \vec{v}_i' \sum_{j=1}^N c_j \vec{v}_j = \sum_{j=1}^N c_j (\vec{v}_i' \vec{v}_j) = c_i, \quad (2.47)$$

en virtud de la ortonormalidad de los vectores $\{\vec{v}_i\}$. Entonces, $\vec{u} = P_M \vec{y}$ puede escribirse así:

$$\begin{aligned}
\vec{u} &= P_M \vec{y} \\
&= \sum_{i=1}^p (\vec{v}_i' \vec{y}) \vec{v}_i \\
&= (\vec{v}_1 \mid \vec{v}_2 \mid \cdots \mid \vec{v}_p) \begin{pmatrix} \vec{v}_1' \vec{y} \\ \vec{v}_2' \vec{y} \\ \vdots \\ \vec{v}_p' \vec{y} \end{pmatrix} \\
&= (\vec{v}_1 \mid \vec{v}_2 \mid \cdots \mid \vec{v}_p) \begin{pmatrix} \vec{v}_1' \\ \vec{v}_2' \\ \vdots \\ \vec{v}_p' \end{pmatrix} \vec{y} \\
&= TT' \vec{y}
\end{aligned}$$

■

Lema 2.4 *La matriz P_M es simétrica idempotente.*

DEMOSTRACIÓN:

La matriz P_M es única (Lema 3.2) y puede expresarse siempre como TT' (Lema 3.3). Entonces:

$$\begin{aligned}
P_M' &= (TT')' = TT' = P_M \\
P_M P_M &= TT' TT' = T(T'T)T' = TT' = P_M.
\end{aligned}$$

■

Lema 2.5 *Denotamos por $R(C)$ el subespacio generado por las columnas de C , siendo C una matriz cualquiera. P_M denota la matriz de proyección sobre un cierto subespacio M . Entonces:*

$$R(P_M) = M.$$

DEMOSTRACIÓN:

Claramente $R(P_M) \subseteq M$. Por otra parte, para todo $\vec{x} \in M$,

$$P_M \vec{x} = \vec{x} \implies M \subseteq R(P_M).$$

■

Lema 2.6 *Si P_M es la matriz asociada al operador de proyección sobre M , $(I - P_M)$ es simétrica, idempotente, y está asociada al operador de proyección sobre M^\perp .*

DEMOSTRACIÓN:

Es consecuencia inmediata de los Lemas 3.1 y 3.4.

■

Lema 2.7 *Toda matriz simétrica idempotente P representa una proyección ortogonal sobre el subespacio generado por las columnas de P .*

DEMOSTRACIÓN:

Consideremos la identidad $\vec{y} = P\vec{y} + (I - P)\vec{y}$. Claramente, $(I - P)\vec{y} \perp P\vec{y}$ y además $(I - P)\vec{y} = \vec{y} - P\vec{y}$ es ortogonal a $P\vec{y}$. Por tanto, $P\vec{y}$ es proyección de \vec{y} sobre un cierto subespacio, que, de acuerdo con el Lema 3.5, es el generado por las columnas de P .

■

Definición 2.2 *Sea D una matriz cualquiera, de orden $m \times n$. Decimos que D^- es una pseudo-inversa (o inversa generalizada) de D si:*

$$DD^-D = D \tag{2.48}$$

En general, D^- así definida no es única. En el caso particular de que D sea una matriz cuadrada de rango completo, $D^- = D^{-1}$.

Lema 2.8 *Sea D una matriz $m \times n$ cualquiera. Sea \vec{c} una matriz $m \times 1$ y \vec{z} un vector de variables. Si el sistema:*

$$D\vec{z} = \vec{c} \tag{2.49}$$

es compatible, una solución viene dada por $\vec{z} = D^-\vec{c}$, siendo D^- una pseudo-inversa.

DEMOSTRACIÓN:

De (3.48) deducimos:

$$DD^{-}D\vec{z} = \vec{c} \quad (2.50)$$

y sustituyendo (3.49) en (3.50):

$$DD^{-}\vec{c} = \vec{c} \quad (2.51)$$

$$D(D^{-}\vec{c}) = \vec{c} \quad (2.52)$$

lo que muestra que $D^{-}\vec{c}$ es solución de (3.49). ■

En realidad, es posible probar un resultado algo más fuerte¹; *toda* solución de (3.49) puede expresarse como $D^{-}\vec{c}$ para alguna elección de D^{-} .

Lema 2.9 *Si $M = R(X)$, entonces $P_M = X(X'X)^{-}X'$.*

DEMOSTRACIÓN:

Sea \vec{y} un vector cualquiera. Su proyección sobre $R(X)$ ha de ser de la forma $X\hat{\beta}$, y verificar las ecuaciones normales (3.5) en la pág. 17:

$$X'X\hat{\beta} = X'\vec{y} \quad (2.53)$$

Identificando $D = X'X$, $\vec{z} = \hat{\beta}$, y $\vec{c} = X'\vec{y}$, el lema anterior garantiza que $(X'X)^{-}X'\vec{y}$ será una posible solución para $\hat{\beta}$ (no necesariamente única, ya que hay múltiples $(X'X)^{-}$ en general); no obstante, $X(X'X)^{-}X'\vec{y}$ es la *única* proyección de \vec{y} sobre M , y $X(X'X)^{-}X'$ es la *única* matriz de proyección. La unicidad de la proyección se demostró en el Teorema 2.1, pág. 10. La unicidad de la matriz de proyección, fue objeto del Lema 3.2. ■

Como se ha indicado, hay en general múltiples inversas generalizadas D^{-} , cada una de las cuales da lugar a una diferente solución del sistema (3.51)–(3.52).

¹Cf. Searle (1971), Teorema 8, pág. 26.

2.7. Lectura recomendada

Sobre la teoría. Seber (1977), Cap. 3 cubre completamente la materia de este capítulo. Para las cuestiones de álgebra matricial, proyecciones, etc. Draper and Smith (1998) tiene un capítulo completo (el 20) mostrando el problema de la estimación MCO desde un punto de vista geométrico, similar al empleado aquí; Searle (1982), Searle (1971) y Abadir and Magnus (2005) son buenas referencias. Sobre matrices inversas generalizadas, en particular, pueden verse, además de Searle (1982), Ben-Israel and Greville (1974) y Rao and Mitra (1971).

Sobre R. Son de utilidad las referencias indicadas en el Capítulo precedente. Como se indicó, hay mucha documentación *on line* sobre R, como Venables et al. (1997) (hay traducción castellana, Venables et al. (2000), un poco desfasada), Maindonald (2000) o Kuhnert and Venables (2005); una relación actualizada puede obtenerse en <http://cran.r-project.org/>.

COMPLEMENTOS Y EJERCICIOS

2.1 ¿Que efecto tienen sobre los estimadores $\hat{\beta}$ cambios en la escala de los regresores en X ? Demuéstrese.

2.2 Haciendo uso del mismo argumento empleado (en (3.39), pág. 24) para mostrar que $SSE/(N-p)$ es un estimador insesgado de σ^2 , compruébese que, dada una muestra aleatoria simple Z_1, \dots, Z_n , el estimador de la varianza

$$\sigma_Z^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2$$

no es insesgado.

2.3 Extiéndase el teorema de Gauss-Markov, para probar la afirmación hecha al final de la Sección 3.4 (pág. 23): si $\vec{c}'\vec{\beta}$ es cualquier forma lineal, en el caso de rango completo el estimador insesgado de varianza mínima de $\vec{c}'\vec{\beta}$ es $\vec{c}'\hat{\beta}$.

2.4 La Definición 3.2, pág. 36, no individualiza una única inversa generalizada, salvo cuando D es cuadrada de rango completo. Las siguientes condiciones, la primera de las cuáles coincide con (3.48), proporcionan una única definición de inversa generalizada (la inversa de Moore-Penrose):

$$DD^-D = D; \quad D^-DD^- = D^-; \quad D^-D \text{ y } DD^- \text{ simétricas.}$$

A la única matriz D^- así especificada se la denomina inversa de Moore-Penrose. Sobre inversas generalizadas e inversas de Moore-Penrose puede consultarse Searle (1971) y Rao and Mitra (1971)

2.5 (\uparrow 3.4) Cuando la función `lsfit` de R encuentra una matriz de diseño de rango incompleto, proporciona no obstante *una* solución de $\hat{\beta}$, haciendo un cómputo en esencia equivalente a $\hat{\beta} = (X'X)^-X'\vec{y}$. Podemos llevar a cabo el cálculo de la inversa generalizada de Moore-Penrose mediante la función `ginv` del paquete `MASS` (asociado al libro Venables and Ripley (1999a))

```
> library(MASS)
> XX <- matrix(c(2, 0, 0, 0),
+             2, 2)
> XX
```

```

      [,1] [,2]
[1,]    2    0
[2,]    0    0
> XXig <- ginv(XX)
> XXig
      [,1] [,2]
[1,]  0.5    0
[2,]  0.0    0

```

Observemos que las condiciones que definen a la inversa de Moore-Penrose se verifican.

```

> XX %*% XXig %*% XX
      [,1] [,2]
[1,]    2    0
[2,]    0    0
> XXig %*% XX %*% XXig
      [,1] [,2]
[1,]  0.5    0
[2,]  0.0    0
> XXig %*% XX
      [,1] [,2]
[1,]    1    0
[2,]    0    0
> XX %*% XXig
      [,1] [,2]
[1,]    1    0
[2,]    0    0

```

2.6 (↑ 2.13) Resuélvase el problema 2.13, pág. 15, haciendo uso de regresión lineal. (Ayuda: basta normalizar el primer vector y regresar el segundo sobre él. El vector de residuos de esta regresión es ortogonal al primero.)

2.7 (↑ 3.6) Escribase una función en R que resuelva el problema 3.6 de un modo completamente general: debe admitir como único argumento una matrix de rango completo cuyas columnas contengan los vectores a ortonormalizar, y devolver una matrix de las mismas dimensiones cuyas columnas sean los vectores ortonormalizados.

2.8 Justifíquese la afirmación hecha en la Observación 3.7, pág. 32, de acuerdo con la cual el ajuste, medido en términos de SSE , no puede empeorar al añadir regresores.

2.9 ¿Cuándo incluir y cuándo no una columna de “unos”? En general, siempre convendrá hacerlo. Las únicas situaciones en que no será conveniente son aquéllas en que la columna de unos crearía una dependencia lineal exacta entre las columnas de la matriz X .

El no incluir columna de “unos” fuerza a la recta (o hiperplano) de regresión a pasar por el origen. Salvo que haya buenos motivos para ello, no queremos forzar tal cosa en nuestra regresión, especialmente si, como sucede en multitud de ocasiones, el origen es arbitrario.

2.10 (↑ 3.1)(↑ 3.9) Pensemos en la siguiente situación: un investigador está interesado en dilucidar si la velocidad de sedimentación de un fluido (y , medida en unidades adecuadas) está influida por la temperatura (X_1 , medida en grados centígrados). Cuenta con las siguientes observaciones:

$$\vec{y} = \begin{pmatrix} 5,8 \\ 4,7 \\ 4,9 \\ 3,8 \\ 2,1 \end{pmatrix} \quad X_1 = \begin{pmatrix} -10 \\ -6,2 \\ -2,5 \\ 3,0 \\ 4,6 \end{pmatrix}$$

Imaginemos que ajusta una regresión a dichos datos. Los resultados pueden verse en el siguiente fragmento en R:

```
> y <- c(5.8, 4.7, 4.9, 3.8,
+       2.1)
> X <- c(-10, -6.2, -2.5, 3,
+       4.6)
> ajuste <- lsfit(X, y, intercept = FALSE)
> ajuste$coefficients
```

```
      X
-0.44798
```

El coeficiente que afecta a la única variable es negativo ($= -0,447984$), lo que estaríamos tentados de interpretar así: por cada grado que aumenta la temperatura, disminuye en 0.447984 la velocidad de sedimentación. (Quedaría por ver si la estimación del coeficiente de regresión es de fiar, cuestión que abordaremos más adelante.)

Supongamos ahora que otro investigador repite el mismo análisis, pero en lugar de expresar las temperaturas en grados centígrados (C)

lo hace en grados Fahrenheit (F) cuya relación con los centígrados viene dada por $C = \frac{5}{9}(F - 32)$ ($\Rightarrow F = \frac{9}{5}C + 32$). Los cálculos, siempre haciendo una regresión pasando por el origen, serían ahora:

```
> y <- c(5.8, 4.7, 4.9, 3.8,
+       2.1)
> X <- c(-10, -6.2, -2.5, 3,
+       4.6)
> X <- (9/5) * X + 32
> ajuste <- lsfit(X, y, intercept = FALSE)
> ajuste$coefficients

      X
0.12265
```

¡Ahora el coeficiente afectando a la variable temperatura es positivo, dando la impresión de una asociación *directa* entre temperatura y velocidad de sedimentación! Claramente, tenemos motivo para preocuparnos si llegamos a conclusiones diferentes dependiendo de nuestra elección de los sistemas de medida —enteramente convencionales ambos—. El problema desaparece si incluimos una columna de unos en ambos análisis, para dar cuenta de los diferentes orígenes.

```
> y <- c(5.8, 4.7, 4.9, 3.8,
+       2.1)
> X <- c(-10, -6.2, -2.5, 3,
+       4.6)
> ajuste <- lsfit(X, y)
> ajuste$coefficients

Intercept      X
   3.80119  -0.20667

> X <- (9/5) * X + 32
> ajuste <- lsfit(X, y)
> ajuste$coefficients

Intercept      X
   7.47538  -0.11482

> ajuste$coefficients[2] *
+   (9/5)

      X
-0.20667
```

Los coeficientes de X no son ahora iguales (porque los grados Fahrenheit son más “pequeños”), pero si relacionados por un factor de escala y darían lugar a la misma conclusión de asociación inversa entre ambas magnitudes. La inversión del signo del coeficiente se explica comparando en la Figura 3.2 los puntos muestrales (en escalas comparables) y las respectivas rectas de regresión. Dichas rectas de regresión y las gráficas se han generado mediante

```
> postscript(file = "demo2d.eps",
+           horizontal = FALSE, width = 5,
+           height = 10)
> par(mfcol = c(2, 1))
> y <- c(5.8, 4.7, 4.9, 3.8,
+        2.1)
> C <- c(-10, -6.2, -2.5, 3,
+        4.6)
> ajuste <- lsfit(C, y, intercept = FALSE)
> par(xlim = c(-25, 5))
> par(ylim = c(-0.5, 6))
> plot(C, y, ylim = c(-0.5,
+ 6), xlim = c(-25, 5))
> title(main = "Ajuste en grados centigrados")
> abline(a = 0, b = ajuste$coefficients)
> text(x = 0, y = 0, labels = "(0,0)")
> F <- (9/5) * C + 32
> ajuste <- lsfit(F, y, intercept = FALSE)
> plot(F, y, ylim = c(-0.5,
+ 6), xlim = c(-13, 41))
> title(main = "Ajuste en grados Fahrenheit")
> text(x = 0, y = 0, labels = "(0,0)")
> abline(a = 0, b = ajuste$coefficients)
> scratch <- dev.off()
```

Puede verse que el forzar a ambas a pasar por el origen las obliga a tener pendiente de signo opuesto para aproximar la nube de puntos.

Figura 2.2: En un ajuste sin término constante, la pendiente depende de la elección arbitraria del origen

