

Capítulo 6

Regresión con perturbaciones normales.

6.1. Introducción.

Si a los supuestos habituales (Sección 2.3, pág. 7) añadimos¹ el de que $\vec{\epsilon} \sim N(\vec{0}, \sigma^2 I)$, todos los resultados anteriores se mantienen; obtendremos no obstante muchos adicionales, relativos a la distribución de diferentes estadísticos. Podremos también efectuar contrastes de hipótesis diversas. Buena parte de estos resultados son consecuencia casi inmediata de alguno de los siguientes lemas.

Lema 6.1 Si $\vec{u} \sim N(\vec{0}, \sigma^2 I)$ y A es una matriz simétrica idempotente de orden n y rango r , entonces: $\frac{\vec{u}' A \vec{u}}{\sigma^2} \sim \chi_r^2$.

DEMOSTRACIÓN:

Sea D la matriz diagonalizadora de A . Siendo A simétrica, D es una matriz ortogonal cuyas columnas son vectores propios de A , verificándose: $D'AD = \Lambda$, en que Λ es una matriz en cuya diagonal principal aparecen los

¹El símbolo \sim denotará en lo sucesivo que el lado izquierdo es una variable aleatoria con la distribución que especifica el lado derecho.

valores propios de A . Como A es idempotente, Λ es de la forma

$$\Lambda = \begin{pmatrix} r & (n-r) \\ I & 0 \\ 0 & 0 \end{pmatrix},$$

en que I es una matriz unidad de rango r , y los bloques de ceros que la circundan son de órdenes adecuados para completar una matriz cuadrada de orden $n \times n$.

Si hacemos el cambio de variable $\vec{v} = D'\vec{u}$ ($\Rightarrow \vec{u} = D\vec{v}$), el nuevo vector \vec{v} sigue también una distribución $N(\vec{0}, \sigma^2 I)$. Entonces,

$$\frac{\vec{u}'A\vec{u}}{\sigma^2} = \frac{\vec{v}'D'AD\vec{v}}{\sigma^2} = \frac{\vec{v}'}{\sigma} \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \frac{\vec{v}}{\sigma} = \sum_{i=1}^r \frac{v_i^2}{\sigma^2}. \quad (6.1)$$

Pero el lado derecho de (7.1) es una suma de cuadrados de r variables aleatorias $N(0, 1)$ independientes, y por tanto sigue una distribución² χ_r^2 . ■

Lema 6.2 *Sea B una matriz simétrica $n \times n$ y P una matriz simétrica idempotente del mismo orden y rango r . Sea \vec{u} un vector aleatorio n -variante, $\vec{u} \sim N(\vec{0}, \sigma^2 I)$, y supongamos que se verifica $BP = 0$. Entonces, $\vec{u}'B\vec{u}$ y $\vec{u}'P\vec{u}$ son variables aleatorias independientes.*

DEMOSTRACIÓN:

Sea D la matriz diagonalizadora de P . Al igual que antes, definamos $\vec{v} = D'\vec{u}$, (lo que implica $\vec{u} = D\vec{v}$). Tenemos que:

$$BP = 0 \Rightarrow D'BDD'PD = 0 \quad (6.2)$$

$$(6.3)$$

$$\Rightarrow D'BD \begin{pmatrix} r & (n-r) \\ I & 0 \\ 0 & 0 \end{pmatrix} = 0 \quad (6.4)$$

$$(6.5)$$

$$\Rightarrow D'BD \text{ tiene sus } r \text{ primeras columnas nulas} \quad (6.6)$$

²El recíproco es también cierto; véase en Searle (1971), Teorema 2, pag. 57 una versión más potente de este teorema.

Por tanto:

$$D'BD = \begin{matrix} & r & (n-r) \\ r & \begin{pmatrix} 0 & L_{12} \\ 0 & L_{22} \end{pmatrix} & \\ (n-r) & & \end{matrix} = 0 \quad (6.7)$$

Como, además, $D'BD$ es simétrica, L_{12} ha de ser también un bloque de ceros, y:

$$\vec{u}'B\vec{u} = \vec{v}'D'BD\vec{v} = \vec{v}' \begin{matrix} & r & (n-r) \\ \begin{pmatrix} 0 & 0 \\ 0 & L_{22} \end{pmatrix} & & \end{matrix} \vec{v} \quad (6.8)$$

Por otra parte:

$$\vec{u}'P\vec{u} = \vec{v}'D'PD\vec{v} = \vec{v}' \begin{matrix} & r & (n-r) \\ \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} & & \end{matrix} \vec{v} \quad (6.9)$$

De (7.8) y (7.9) se deduce que ambas formas cuadráticas consideradas dependen de distintas componentes del vector \vec{v} , y son por tanto independientes. ■

Lema 6.3 *Sea M una matriz simétrica idempotente de rango r y dimensiones $n \times n$. Sea A una matriz que verifica $AM = 0$, y $\vec{u} \sim N(\vec{0}, \sigma^2 I)$. Entonces $A\vec{u}$ y $\vec{u}'M\vec{u}$ son variables aleatorias independientes.*

DEMOSTRACIÓN:

Sea D la matriz que diagonaliza M . Al igual que antes, definamos $\vec{v} = D'\vec{u}$ ($\Rightarrow \vec{u} = D\vec{v}$). Como $AM = 0$, y $D'MD$ es una matriz diagonal con r unos y $(n-r)$ ceros en la diagonal principal, se verifica que

$$AM = ADD'MD = 0 \Rightarrow AD = \begin{matrix} & r & (n-r) \\ \begin{pmatrix} 0 & | & L_2 \end{pmatrix}, & & \end{matrix} \quad (6.10)$$

es decir, AD tiene sus primeras r columnas de ceros. Por consiguiente,

$$A\vec{u} = AD\vec{v} = \begin{matrix} & r & (n-r) \\ \begin{pmatrix} 0 & | & L_2 \end{pmatrix} & & \end{matrix} \vec{v}. \quad (6.11)$$

Como

$$\vec{u}' M \vec{u} = \vec{v}' D' M D \vec{v} = \vec{v}' \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \vec{v}, \quad (6.12)$$

deducimos de (7.11) y (7.12) que ambas variables aleatorias consideradas dependen de distintas componentes de \vec{v} , y son consecuentemente independientes. ■

Podemos ahora, con ayuda de los Lemas precedentes, demostrar el siguiente resultado:

Teorema 6.1 Si $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$, $\vec{\epsilon} \sim N(\vec{0}, \sigma^2 I)$, y X es de orden $N \times p$ y rango p , se verifica:

1. $\hat{\beta} \sim N(\vec{\beta}, \sigma^2(X'X)^{-1})$
2. $(\hat{\beta} - \vec{\beta})'(X'X)(\hat{\beta} - \vec{\beta}) \sim \sigma^2 \chi_p^2$
3. $(N - p)\hat{\sigma}^2 = SSE \sim \sigma^2 \chi_{N-p}^2$
4. $\hat{\beta}$ y $\hat{\sigma}^2$ son variables aleatorias independientes.

DEMOSTRACIÓN:

El apartado 1) es inmediato. Si se verifican los supuestos habituales, fue ya demostrado (Teorema 3.2, pág. 21) que $\hat{\beta}$ es un estimador insesgado de $\vec{\beta}$ con la matriz de covarianzas indicada. Como, además, $\hat{\beta}$ es una combinación lineal de variables aleatorias normales e independientes, es también normal.

El apartado 2) es consecuencia inmediata del Lema 7.1, una vez que observamos que $(X'X)^{\frac{1}{2}}(\hat{\beta} - \vec{\beta}) \sim N(\vec{0}, \sigma^2 I)$.

Para demostrar el apartado 3) observemos que:

$$\frac{SSE}{\sigma^2} = \frac{(\vec{Y} - X\hat{\beta})'(\vec{Y} - X\hat{\beta})}{\sigma^2} \quad (6.13)$$

$$= \frac{(\vec{Y} - X(X'X)^{-1}X'\vec{Y})'(\vec{Y} - X(X'X)^{-1}X'\vec{Y})}{\sigma^2} \quad (6.14)$$

$$= \frac{\vec{Y}'[I - X(X'X)^{-1}X']\vec{Y}}{\sigma^2} \quad (6.15)$$

$$= \frac{(X\vec{\beta} + \vec{\epsilon})'[I - X(X'X)^{-1}X'](X\vec{\beta} + \vec{\epsilon})}{\sigma^2} \quad (6.16)$$

$$= \frac{\vec{\epsilon}'[I - X(X'X)^{-1}X']\vec{\epsilon}}{\sigma^2} \quad (6.17)$$

$$= \frac{\vec{\epsilon}'M\vec{\epsilon}}{\sigma^2} \quad (6.18)$$

$$\sim \chi_{N-p}^2, \quad (6.19)$$

donde (7.19) es consecuencia inmediata del Lema 7.1, ya que M es simétrica idempotente y de rango $N - p$.

Para probar 4), basta invocar el Lema 7.3, ya que

$$\hat{\beta} = (X'X)^{-1}X'\vec{Y}, \quad (6.20)$$

$$\hat{\sigma}^2 = \frac{SSE}{N-p} = \frac{\vec{Y}'[I - X(X'X)^{-1}X']\vec{Y}}{N-p}. \quad (6.21)$$

De la ecuación (7.20) deducimos (sustituyendo \vec{Y} por $X\vec{\beta} + \vec{\epsilon}$) que $\hat{\beta} = \vec{\beta} + (X'X)^{-1}X'\vec{\epsilon}$. La misma sustitución en (7.21) muestra que

$$\hat{\sigma}^2 = \frac{\vec{\epsilon}'[I - X(X'X)^{-1}X']\vec{\epsilon}}{N-p}.$$

Como

$$(X'X)^{-1}X'[I - X(X'X)^{-1}X'] = 0,$$

el Lema 7.3, pág. 71, demuestra la independencia de las formas lineal y cuadrática anteriores y por tanto de (7.20) y (7.21). ■

R: Ejemplo 6.1 (*ejemplo de simulación*)

El código que sigue tiene por objeto ilustrar cómo examinaríamos empíricamente la concordancia entre lo que la teoría predice y lo que podemos obtener en la práctica. Lo que se hace es generar múltiples muestras artificiales, obtener de ellas múltiples observaciones del estadístico de interés (aquí, $\hat{\beta}$) y examinar el ajuste de la distribución empírica de los mismos a la teórica.

Generemos en primer lugar la matriz de diseño X , vector de parámetros $\vec{\beta}$ y los valores medios de la respuesta $X\vec{\beta}$:

```
> X <- matrix(c(1, 1, 1, 1, 1, 1, 9, 4,
+             12, 1, 4, 13, 0, 6, 7, 0, 2, 2), 6,
+             3)
> X
      [,1] [,2] [,3]
[1,]    1    9    0
[2,]    1    4    6
[3,]    1   12    7
[4,]    1    1    0
[5,]    1    4    2
[6,]    1   13    2

> beta <- c(2, 3, 4)
> Ey <- X %*% beta
```

Definiremos ahora una matriz \mathbf{b} de dimensiones 100×3 , cada una de cuyas filas guardará los parámetros estimados $\hat{\beta}$ con una muestra artificial diferente

```
> muestras <- 100
> b <- matrix(0, muestras, 3)
```

e iteremos, generando en cada pasada del bucle `for` un nuevo vector de perturbaciones $\hat{\epsilon}$ (mediante `rnorm`), un nuevo vector de valores de la variable respuesta \vec{y} y nuevas estimaciones $\hat{\beta}$ de los parámetros $\vec{\beta}$ (`fit$coefficients`, que se almacenan en `b[i,]`):

```
> for (i in 1:muestras) {
+   y <- Ey + rnorm(6)
+   fit <- lsfit(X, y, intercept = FALSE)
+   b[i, ] <- fit$coefficients
+ }
```

La distribución teórica de los betas es Normal, con vector de medias $(2, 3, 4)'$ y matriz de covarianzas $(X'X)^{-1}$ (la varianza de las perturbaciones generadas por `rnorm` es 1 si no se especifica otra cosa).

```
> cov.betas <- solve(t(X) %*% X)
```

Por consiguiente, un modo de verificar que los resultados empíricos son congruentes con la teoría consistiría en tipificar las estimaciones de los parámetros y comparar su distribución con una $N(0, 1)$. Podemos por ejemplo comparar la media y varianza empíricas con las teóricas,

```
> beta1.tipif <- (b[, 1] - beta[1])/sqrt(cov.betas[1,
+   1])
> mean(beta1.tipif)
```

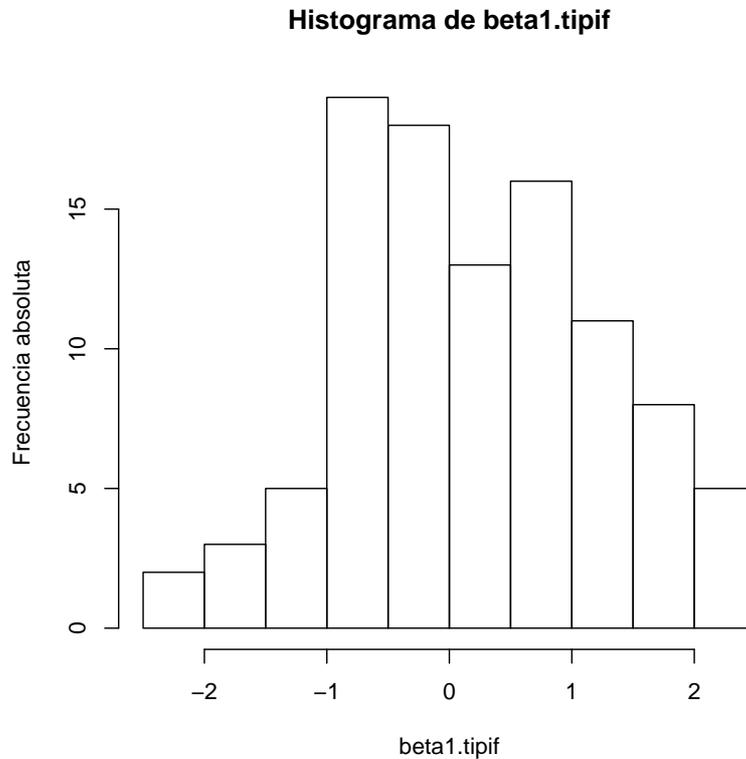
```
[1] 0.19871
```

```
> var(beta1.tipif)
```

```
[1] 1.1125
```

dibujar el histograma

```
> hist(beta1.tipif, ylab = "Frecuencia absoluta",
+   main = "Histograma de beta1.tipif")
```



o llevar a cabo algún contraste de normalidad especializado:

```
> ks.test(beta1.tipif, "pnorm")
      One-sample Kolmogorov-Smirnov test

data:  beta1.tipif
D = 0.1036, p-value = 0.2334
alternative hypothesis: two-sided

> shapiro.test(beta1.tipif)
      Shapiro-Wilk normality test

data:  beta1.tipif
W = 0.9874, p-value = 0.4679
```

Lo que antecede ilustra, reducido a sus rasgos esenciales, el llamado método de Monte-Carlo. Puede parecer un ejercicio ocioso en el caso que nos ocupa (ya “sabíamos” cómo se distribuye $\hat{\beta}$ ¿a que viene comprobarlo mediante una simulación?). Sin embargo, tiene una enorme aplicación práctica por varias razones:

1. En ocasiones no conocemos la distribución teórica de los estadísticos de interés para muestras finitas. Todo lo que podemos obtener teóricamente es la distribución asintótica (la distribución cuando el tamaño muestral tiende a infinito). En este caso, la simulación permite ver si la aproximación asintótica es aceptable para un cierto tamaño muestral.
2. En otras ocasiones, ni siquiera la distribución asintótica es obtenible analíticamente. Este es el caso más frecuente en la práctica. De nuevo el método de Monte-Carlo proporciona un método para obtener aproximaciones a la distribución de cualquier estadístico.

El uso del método de Monte-Carlo reposa en la posibilidad de generar mediante un ordenador números aleatorios con la distribución que deseemos. En este ejemplo, se ha empleado `rnorm` para generar variables aleatorias normales. (R ofrece generadores de números aleatorios de las distribuciones más usuales, como casi cualquier otro paquete estadístico.)

FIN DEL EJEMPLO ■

6.2. Contraste de hipótesis lineales.

El problema que nos planteamos es el siguiente: dado el modelo lineal $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ con los supuestos habituales más normalidad, queremos, con ayuda de una muestra, contrastar la siguiente hipótesis lineal

$$h: A\vec{\beta} = \vec{c} \quad (\text{rango de } A = q < p), \quad (6.22)$$

siendo A de dimensiones $q \times p$. Cualquier hipótesis lineal sobre los parámetros se puede expresar en la forma (7.22). En particular, mediante adecuada elección de A se pueden hacer contrastes de nulidad de uno o varios parámetros, de igualdad de dos o más de ellos, etc.

Observación 6.1 Llamamos hipótesis lineales a las que pueden expresarse del modo (7.22); multitud de hipótesis de interés admiten tal expresión, como se verá en lo que sigue. Hay hipótesis, sin embargo, que no pueden escribirse de tal forma. Por ejemplo, restricciones de no negatividad sobre los parámetros ($\beta_i > 0$) o sobre el módulo de $\vec{\beta}$ (cosas como $\beta_1^2 + \beta_2^2 = 1$).

La forma de efectuar el contraste es la habitual. Se busca un estadístico que bajo la hipótesis nula h siga una distribución conocida; si el valor obtenido en el muestreo de dicho estadístico es “raro” de acuerdo con lo esperable cuando h es cierta, rechazaremos la hipótesis nula. El estadístico de contraste y su distribución se deducen del siguiente teorema:

Teorema 6.2 *Sea $h: A\vec{\beta} = \vec{c}$ una hipótesis lineal, $\hat{\beta}_h$ el vector de estimadores mínimo cuadráticos condicionados por h , y $SSE_h = \|\vec{Y} - X\hat{\beta}_h\|^2$. Bajo los supuestos habituales más el de normalidad en las perturbaciones, se verifica:*

1. $SSE_h - SSE = (A\hat{\beta} - \vec{c})'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - \vec{c})$
2. Si $h: A\vec{\beta} = \vec{c}$ es cierta,

$$Q_h = \frac{(SSE_h - SSE)/q}{SSE/(N-p)} \sim \mathcal{F}_{q, N-p}$$

en que $q \leq p$ es el rango de A .

DEMOSTRACIÓN:

$$SSE_h - SSE = \|\vec{Y} - X\hat{\beta}_h\|^2 - \|\vec{Y} - X\hat{\beta}\|^2 \quad (6.23)$$

$$= \|\vec{Y} - X\hat{\beta} + X\hat{\beta} - X\hat{\beta}_h\|^2 - \|\vec{Y} - X\hat{\beta}\|^2 \quad (6.24)$$

$$= \|\vec{Y} - X\hat{\beta}\|^2 + \|X\hat{\beta} - X\hat{\beta}_h\|^2 - 2\langle \vec{Y} - X\hat{\beta}, X\hat{\beta} - X\hat{\beta}_h \rangle \quad (6.25)$$

$$= \|X\hat{\beta} - X\hat{\beta}_h\|^2 \quad (6.26)$$

$$= (\hat{\beta} - \hat{\beta}_h)'(X'X)(\hat{\beta} - \hat{\beta}_h). \quad (6.27)$$

Se ha hecho uso en el paso de (7.25) a (7.26) de que $\hat{\epsilon}$ es ortogonal a toda combinación lineal de las columnas de X , lo que garantiza la nulidad del producto interno en (7.25).

Haciendo uso de la ecuación (5.11), pág. 58, la expresión (7.27) se convierte en:

$$SSE_h - SSE = (A\hat{\beta} - \vec{c})'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - \vec{c}). \quad (6.28)$$

Esto finaliza la demostración del primer apartado. Por otra parte, como

$$\hat{\beta} = \vec{\beta} + (X'X)^{-1}X'\vec{\epsilon},$$

tenemos que, cuando se verifica la hipótesis h ,

$$(A\hat{\beta} - \vec{c}) = (A\hat{\beta} - A\vec{\beta}) = A(X'X)^{-1}X'\vec{\epsilon},$$

resultado que llevado a (7.28) proporciona:

$$SSE_h - SSE \stackrel{h}{=} \underbrace{\vec{\epsilon}' X(X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A(X'X)^{-1}X'\vec{\epsilon}}_G \quad (6.29)$$

Esta expresión muestra que $SSE_h - SSE$ es una forma cuadrática en variables normales (las $\vec{\epsilon}$) de matriz G que fácilmente comprobamos es idempotente. Por tanto, según el Lema 7.1, pág. 69, $SSE_h - SSE$ sigue una distribución $\sigma^2\chi_q^2$, con grados de libertad q iguales al rango de G ($= \text{rango}(A)$). Tenemos además (Teorema 7.1) que:

$$SSE = \vec{Y}'(I - P_M)\vec{Y} \sim \sigma^2\chi_{N-p}^2 \quad (6.30)$$

Para demostrar que Q_h en el enunciado es una variable aleatoria con distribución \mathcal{F} de Snedecor, sólo resta comprobar que numerador y denominador son independientes: pero ésto es inmediato, ya que

$$(I - P_M) \underbrace{X(X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A(X'X)^{-1}X'}_G = 0.$$

El Lema 7.2 garantiza por tanto la independencia. ■

Observación 6.2 Hay cuestiones de interés sobre el Teorema 7.2. En primer lugar, es claro que, para un nivel de significación α , la región crítica estará formada por valores mayores que $\mathcal{F}_{q, N-p}^\alpha$. En efecto, son grandes discrepancias entre SSE_h y SSE las que cabe considerar evidencia contra h . Desde otro punto de vista, el apartado 1) del Teorema 7.2 muestra que el estadístico tiene en su numerador una forma cuadrática que crece al separarse $A\hat{\beta}$ de \vec{c} .

Observación 6.3 La presentación es puramente heurística; se ha propuesto el estadístico Q_h y encontrado su distribución, indicándose, sin otro apoyo que el sentido común, qué valores debemos considerar en la región crítica. Podríamos llegar a un resultado análogo si

construyéramos un estadístico de contraste basado en la razón generalizada de verosimilitudes:

$$\Lambda = \frac{\max_{\hat{\beta}} g(\hat{\beta}; \vec{y}, X)}{\max_{\hat{\beta}_h} g(\hat{\beta}_h; \vec{y}, X)}$$

siendo $\hat{\beta}_h$ aquellos $\hat{\beta}$ verificando $h: A\hat{\beta} = \vec{c}$. Ello proporciona una justificación al estadístico anterior.

Observación 6.4 Del enunciado del teorema anterior se sigue con facilidad que cuando h no es cierta (y en consecuencia $A\vec{\beta} - \vec{c} = \vec{d} \neq \vec{0}$, Q_h sigue una distribución \mathcal{F} de Snedecor no central, con parámetro de no centralidad $\delta^2 = \vec{t}'\vec{t}$ (véase Apéndice B.1), siendo

$$\vec{t} = [A(X'X)^{-1}A']^{-\frac{1}{2}}(A\vec{\beta} - \vec{c}).$$

Ello permite calcular fácilmente la potencia de cualquier contraste frente a alternativas prefijadas, si se dispone de tablas o ábacos de la \mathcal{F} de Snedecor no central. En R se dispone de la función `pf` que admite un parámetro de no centralidad. Alternativamente, puede estimarse la potencia por simulación.

R: Ejemplo 6.2 (contraste de una hipótesis lineal)

Veamos el modo en que contrastaríamos una hipótesis lineal general sobre los parámetros de un modelo de regresión lineal. Nos serviremos de la función `lscond` para realizar estimación condicionada presentada en el Ejemplo 5.1, pág. 58.

```
> lscond <- function(X, y, A, d, beta0 = TRUE) {
+   ajuste <- lsfit(X, y, intercept = beta0)
+   betas <- ajuste$coefficients
+   xxinv <- solve(t(X) %*% X)
+   axxa <- solve(A %*% xxinv %*% t(A))
+   betas.h <- betas - xxinv %*% t(A) %*%
+     axxa %*% (A %*% betas - d)
+   betas.h <- as.vector(betas.h)
+   names(betas.h) <- names(ajuste$coefficients)
+   return(list(betas = betas, betas.h = betas.h,
+     ajuste.inc = ajuste))
+ }
```

Definiremos ahora una nueva función, `contraste.h`, que calcula SSE , SSE_h (utilizando `lscond`), el estadístico Q_h y su nivel de significación.

```

> contraste.h <- function(X, y, A, d, beta0 = TRUE) {
+   lscond.result <- lscond(X, y, A, d,
+     beta0 = beta0)
+   betas <- lscond.result$betas
+   betas.h <- lscond.result$betas.h
+   SSE <- sum((y - X %*% betas)^2)
+   SSE.h <- sum((y - X %*% betas.h)^2)
+   numer <- (SSE.h - SSE)/nrow(A)
+   denom <- SSE/(nrow(X) - ncol(X))
+   Qh <- numer/denom
+   p.value <- 1 - pf(Qh, nrow(A), nrow(X) -
+     ncol(X))
+   return(list(Qh = Qh, p.value = p.value))
+ }

```

Generemos datos artificiales:

```

> 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

> beta <- c(2, 3, 4)
> y <- X %*% beta + rnorm(6)

```

“Sabemos”, porque los datos han sido artificialmente generados, que $\beta_1 = 3$ y $\beta_2 = 4$. Probaremos a continuación a contrastar la hipótesis $\beta_1 = \beta_2$, que debiera ser rechazada. La matriz A y vector \vec{c} especificando dicha hipótesis pueden construirse así:

```

> A <- matrix(c(0, 1, -1), 1, 3, byrow = TRUE)
> d <- 0

```

El contraste puede entonces llevarse a cabo así:

82 CAPÍTULO 6. REGRESIÓN CON PERTURBACIONES NORMALES.

```
> result <- contraste.h(X, y, A = A, d = d,
+   beta0 = FALSE)
> result$Qh
[1] 161.11
> result$p.value
[1] 0.0010548
```

Rechazaríamos por consiguiente la hipótesis contrastada para cualquier nivel de significación $\alpha > 0.0010548$.

Frecuentemente podemos obtener las sumas de cuadrados requeridas para el contraste de hipótesis de interés de manera más simple. En el caso que nos ocupa, si realmente $\beta_1 = \beta_2$,

$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (6.31)$$

es equivalente a

$$Y = \beta_0 X_0 + \beta_1 (X_1 + X_2) + \epsilon \quad (6.32)$$

y las sumas de cuadrados SSE y SSE_h podrían obtenerse así:

```
> SSE <- sum(lsfite(X, y)$residuals^2)
> Xmod <- cbind(X[, 1], X[, 2] + X[, 3])
> SSE.h <- sum(lsfite(Xmod, y)$residuals^2)
> Qh <- ((SSE.h - SSE)/1)/(SSE/(nrow(X) -
+   ncol(X)))
```

Puede verse que el valor de Q_h así calculado es idéntico al obtenido más arriba:

```
> Qh
[1] 161.11
```

Esta técnica de calcular las sumas de cuadrados SSE y SSE_h en dos regresiones *ad-hoc* puede ser muy frecuentemente utilizada. En el caso frecuente de hipótesis de exclusión (alguno o varios betas iguales a cero), puede obtenerse SSE_h de una regresión en que los regresores correspondientes están ausentes. Si en nuestro ejemplo quisiéramos contrastar $h : \beta_1 = \beta_2 = 0$, podríamos obtener SSE de la regresión (7.31) y SSE_h de la regresión

$$Y = \beta_0 X_0 + \epsilon,$$

para calcular el estadístico Q_h así:

```

> SSE <- sum(lsfrit(X, y)$residuals^2)
> SSE.h <- sum(lsfrit(X[, 1], y)$residuals^2)
> Qh <- ((SSE.h - SSE)/2)/(SSE/(nrow(X) -
+      ncol(X)))
> Qh
[1] 16956

```

El valor que dicho estadístico Q_h deja en a su derecha en la distribución de referencia,

```

> 1 - pf(Qh, 2, nrow(X) - ncol(X))
[1] 8.3193e-07

```

permite rechazar contundentemente la hipótesis $h : \beta_1 = \beta_2 = 0$ contrastada.

FIN DEL EJEMPLO ■

6.2.1. Contraste sobre coeficientes β_i aislados.

El Teorema 7.2 permite obtener como casos particulares multitud de contrastes frecuentemente utilizados. Por ejemplo, la hipótesis $h : \beta_{i-1} = 0$ puede contrastarse tomando $\vec{c} = \vec{0}$ y $A = (0 \ \cdots \ 1 \ \cdots \ 0)$, ocupando el único “uno” la posición i -ésima (recuérdese que los parámetros β se numeran a partir de β_0). En tal caso, Q_h puede escribirse así:

$$Q_h = \frac{(\hat{\beta}_{i-1} - 0)'[(X'X)_{ii}^{-1}]^{-1}(\hat{\beta}_{i-1} - 0)}{\hat{\sigma}^2} \quad (6.33)$$

donde $(X'X)_{ii}^{-1} = [A(X'X)^{-1}A']$ designa el elemento en la posición i -ésima de la diagonal principal de $(X'X)^{-1}$. Bajo la hipótesis h , (7.33) sigue una distribución $\mathcal{F}_{1, N-p}$, y como $\hat{\sigma}^2(X'X)_{ii}^{-1} = \hat{\sigma}_{\hat{\beta}_{i-1}}^2$ tenemos que:

$$\sqrt{Q_h} = \frac{\hat{\beta}_{i-1}}{\hat{\sigma}_{\hat{\beta}_{i-1}}} \sim \sqrt{\mathcal{F}_{1, N-p}} \sim t_{N-p} \quad (6.34)$$

La regla de decisión que se deduce de (7.34) es:

Rechazar $h : \beta_{i-1} = 0$ al nivel de significación α si

$$\left| \frac{\hat{\beta}_{i-1}}{\hat{\sigma}_{\hat{\beta}_{i-1}}} \right| > t_{N-p}^{\alpha/2}.$$

El estadístico $|\hat{\beta}_{i-1}/\hat{\sigma}_{\hat{\beta}_{i-1}}|$ recibe el nombre de *estadístico t* o *t-ratio*. De forma análoga se contrasta la hipótesis $h : \beta_{i-1} = c$.

6.2.2. Contraste de significación conjunta de la regresión.

Otra hipótesis frecuentemente de interés es: $h: \beta_1 = \dots = \beta_{p-1} = 0$ —es decir, nulidad de todos los parámetros, salvo el correspondiente a la columna de “unos”, β_0 —. En este caso,

$$SSE_h = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

y la hipótesis h puede expresarse en la forma $A\vec{\beta} = \vec{c}$ siendo:

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = (\vec{0} \mid I)$$

una matriz con $(p - 1)$ filas y p columnas, y:

$$\vec{c}' = (0 \quad 0 \quad \cdots \quad 0)$$

Pero SSE_h en este caso particular es lo que hemos definido (Teorema 3.4, pág. 30) como SST . Por tanto,

$$\begin{aligned} Q_h &= \frac{(SST - SSE)/(p - 1)}{SSE/(N - p)} \\ &= \frac{N - p}{p - 1} \times \frac{(SST - SSE)}{SSE} \\ &= \frac{N - p}{p - 1} \times \frac{R^2}{(1 - R^2)} \end{aligned}$$

siendo R el coeficiente de correlación múltiple definido en el Teorema 3.4, pág. 31. El contraste de h requiere solamente conocer R^2 . Cuando h es cierta, Q_h se distribuye como una $\mathcal{F}_{p-1, N-p}$.

6.3. Construcción de intervalos de confianza para la predicción.

Supongamos de nuevo que trabajamos sobre el modelo $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ con los supuestos habituales más el de normalidad en las perturbaciones.

Frecuentemente es de interés, además de la estimación de los parámetros, la utilización del modelo con finalidad predictiva.

Sea \vec{x}_* un vector $p \times 1$ de valores a tomar por los regresores. La correspondiente Y_* será: $Y_* = \vec{x}_*' \vec{\beta} + \epsilon_*$. Una predicción \hat{Y}_* del valor a tomar por la Y_* es: $\hat{Y}_* = \vec{x}_*' \hat{\beta}$.

Teorema 6.3 *Se verifica lo siguiente:*

1. $E(Y_* - \hat{Y}_*) = 0$
2. $E(Y_* - \hat{Y}_*)^2 = \sigma^2(1 + \vec{x}_*' (X'X)^{-1} \vec{x}_*)$

DEMOSTRACIÓN:

El apartado 1) se sigue inmediatamente de las ecuaciones (7.35) y (7.36) a continuación, consecuencia la primera de los supuestos habituales, y la segunda de la insesgaredad de $\hat{\beta}$ (Teorema 3.2, pág. 21).

$$E(Y_*) = E(\vec{x}_*' \vec{\beta} + \epsilon_*) = \vec{x}_*' \vec{\beta} \quad (6.35)$$

$$E(\hat{Y}_*) = E(\vec{x}_*' \hat{\beta}) = \vec{x}_*' \vec{\beta} \quad (6.36)$$

Se dice que \hat{Y}_* es una predicción *insesgada* de Y_* . Observemos que:

$$E(Y_* - \hat{Y}_*)^2 = E[\vec{x}_*' \vec{\beta} + \epsilon_* - \vec{x}_*' \hat{\beta}]^2 \quad (6.37)$$

$$= E[\vec{x}_*' (\vec{\beta} - \hat{\beta}) + \epsilon_*]^2 \quad (6.38)$$

$$= E[\vec{x}_*' (\vec{\beta} - \hat{\beta})]^2 + E[\epsilon_*]^2 \quad (6.39)$$

$$= E[\vec{x}_*' (\vec{\beta} - \hat{\beta}) (\vec{\beta} - \hat{\beta})' \vec{x}_*] + E[\epsilon_*]^2 \quad (6.40)$$

$$= \vec{x}_*' \Sigma_{\hat{\beta}} \vec{x}_* + \sigma^2 \quad (6.41)$$

$$= \vec{x}_*' \sigma^2 (X'X)^{-1} \vec{x}_* + \sigma^2 \quad (6.42)$$

$$= \sigma^2 [1 + \vec{x}_*' (X'X)^{-1} \vec{x}_*] \quad (6.43)$$

En el paso de (7.38) a (7.39) se ha hecho uso de la circunstancia de que $\hat{\beta}$ y ϵ_* son independientes ($\hat{\beta}$ depende solamente de $\vec{\epsilon}$, y ϵ_* es perturbación de una observación adicional, distinta de las que han servido para estimar $\hat{\beta}$ e independiente de ellas). ■

El examen de (7.43) muestra dos cosas. Una, que la varianza del error de predicción es *mayor o igual* que la varianza de la perturbación (ya que $\vec{x}_*' (X'X)^{-1} \vec{x}_*$ es una forma cuadrática semidefinida positiva). Esto es lógico: ϵ_* es del todo impredecible, y, *además*, la predicción \hat{Y}_* incorpora una fuente adicional de error, al emplear $\hat{\beta}$ en lugar de $\vec{\beta}$.

Por otra parte, (7.43) muestra que la varianza del error de predicción depende de \vec{x}_* . Habrá determinadas Y_* cuya predicción será más precisa que la de otras. En el Capítulo 10 volveremos sobre el particular.

6.4. Lectura recomendada.

Sobre la teoría. Pueden ser consultados los manuales repetidamente citados: Seber (1977), Cap. 4, Draper and Smith (1998) Cap. 8, Stapleton (1995) Sec. 3.8, Peña (2002) Sec. 7.7 son unos cuantos.

Sobre generadores de números aleatorios, pueden consultarse Knuth (1968), Kennedy (1980), Lange (1998), Thisted (1988) y, en general, cualquier texto sobre computación estadística.

Sobre el contraste razón generalizada de verosimilitudes, puede verse Cox and Hinkley (1974) p. 313 y para su aplicación al contraste de hipótesis lineales generales, Stapleton (1995) Sec. 3.8.

Sobre la utilización de R. En el Ejemplo 5.1, pág. 58 y siguientes, se han definido las funciones `lscond` y `contraste.h` por motivos didácticos. En R hay funciones en varios paquetes que proporcionan análoga funcionalidad. Puede consultarse por ejemplo la documentación de `linear.hypothesis` (paquete `car`) y `glh.test` (paquete `gmodels`).

Por lo que hace a intervalos de confianza, que también pueden obtenerse fácilmente de acuerdo con la teoría esbozada en la Sección 7.3, puede ser de utilidad la función `confint` (paquete `stats`).

El empleo de dichas funciones, sin embargo, presupone familiaridad con la función `lm`, que es objeto de atención en el Capítulo 8 a continuación.

COMPLEMENTOS Y EJERCICIOS

6.1 Demuéstrese que si G es la matriz definida en (7.29) con A y $(X'X)$ ambas de rango completo, entonces $\text{rango}(G) = \text{rango}(A)$.

