

Capítulo 8

Inferencia simultánea.

8.1. Problemas que plantea el contrastar múltiples hipótesis simultáneas

8.1.1. Evidencia contra una hipótesis

Si examinamos la teoría sobre contrastes de hipótesis presentada en la Sección 7.2 veremos que el método ha sido el habitual en Estadística no bayesiana. Los pasos se pueden esquematizar así:

1. Fijar una hipótesis H_0 sobre los parámetros de un modelo.
2. Seleccionar un estadístico cuya distribución sea conocida cuando H_0 es cierta y que se desvía de modo predecible de dicha distribución cuando H_0 no es cierta.
3. Calcular el valor del estadístico en una determinada muestra.
4. **Si el valor de dicho estadístico es anómalo** respecto de lo que esperaríamos bajo H_0 , **rechazar H_0** .

La lógica subyacente es: “Como cuando H_0 es cierta es difícil que se de un valor del estadístico como el observado, lo más plausible es que H_0 no sea cierta.”

Cuando el estadístico que empleamos en el contraste tiene una distribución continua, todos los valores posibles tienen probabilidad cero. No obstante, podemos ordenarlos de más a menos “raros” de acuerdo con su densidad respectiva.

Ejemplo 8.1 Para una muestra X_1, \dots, X_n procedente de una distribución $N(\mu, \sigma^2)$, todos los posibles valores del estadístico \bar{X} tienen probabilidad cero. No obstante, la distribución de dicho estadístico —una $N(\mu, \sigma^2/n)$ — genera de modo frecuente observaciones en las cercanías de μ , y sólo raramente valores en las colas. Consideraremos a estos últimos “raros” y favoreciendo el rechazo de H_0 . Tienen densidad menor que los cercanos a μ .

FIN DEL EJEMPLO ■

Tendrá interés en lo que sigue la noción de *nivel de significación empírico*¹.

Definición 8.1 Llamamos nivel de significación empírico asociado al valor observado de un estadístico a la probabilidad de obtener en el muestreo (bajo H_0) valores tan o más raros que el obtenido.

Ejemplo 8.2 En el Ejemplo 9.1, supongamos que $H_0 : \mu = 0$. Supongamos conocida $\sigma^2 = 1$. Sea una muestra con $n = 100$, e imaginemos que obtenemos un valor de \bar{X} de 0.196 ($= 1,96 \times \sqrt{100^{-1}}$). El nivel de significación empírico (u *observado*) sería 0.05, porque bajo H_0 hay probabilidad 0.05 de observar valores de \bar{X} igual o más alejados de μ que el que se ha presentado.

FIN DEL EJEMPLO ■

Si en ocasiones al abordar un contraste de hipótesis pre fijamos de antemano el nivel de significación que deseamos utilizar (y la región crítica), es muy frecuente realizar el contraste sin una región crítica preespecificada y tomar el nivel de significación empírico como una medida del acuerdo (o desacuerdo) de la evidencia con la hipótesis de interés. Niveles de significación empíricos muy pequeños habrían así de entenderse como evidencia contra la hipótesis nula objeto de contraste.

¹O *p-value*, en la literatura inglesa.

8.1.2. ¿Cómo de “raro” ha de ser algo para ser realmente “raro”?

El siguiente ejemplo² ilustra que un resultado aparentemente muy raro puede no serlo tanto.

Ejemplo 8.3 Consideremos un mono frente a una máquina de escribir. Imaginemos que tras un periodo de tiempo observamos el conjunto de folios tecleados por el mono y constatamos que ¡ha escrito sin una sólo falta de ortografía *Hamlet*!

Bajo la hipótesis nula H_0 : “mono irracional”, tal resultado es absolutamente inverosímil. La probabilidad de que golpeando al azar el teclado un mono logre tal cosa es ridículamente baja. Supongamos que una obra como *Hamlet* requiriera, entre blancos y caracteres, de 635000 digitaciones. Supongamos que hay 26 letras más caracteres de puntuación, etc. totalizando 32 posibilidades de digitación. Componer *Hamlet* totalmente al azar consistiría en apretar la tecla correcta sucesivamente 635.000 veces, algo que, suponiendo las 32 posibilidades de digitación equiprobables, tendría probabilidad:

$$p = \left(\frac{1}{32}\right)^{635000} \approx 5,804527 \times 10^{-955771}. \quad (8.1)$$

La observación de un mono que teclea *Hamlet* sería prácticamente imposible bajo H_0 : habríamos de rechazar H_0 y pensar en alguna alternativa (¿quizá Shakespeare reencarnado en un mono?)

Imaginemos ahora una multitud de monos a los que situamos frente a máquinas de escribir, haciéndoles teclear a su entero arbitrio 635.000 digitaciones. Específicamente, imaginemos 10^{955771} monos. Supongamos que examinando el trabajo de cada uno de ellos, nos topamos con que el mono n -ésimo ¡ha compuesto *Hamlet*! ¿Lo separaríamos de sus congéneres para homenajearlo como reencarnación de Shakespeare? Claramente no; porque, entre tantos, no es extraño que uno, por puro azar, haya tecleado *Hamlet*. De hecho, si todos los conjuntos de 635.000 digitaciones son equiprobables, del trabajo de 10^{955771} monos esperaríamos obtener en torno a 5,8045 transcripciones exactas de *Hamlet*. Lo observado no es raro en absoluto.

FIN DEL EJEMPLO ■

El ejemplo anterior, deliberadamente extremo e inverosímil, ilustra un punto importante. Algo, aparentemente lo mismo, puede ser raro o no dependiendo del contexto. Observar un mono tecleando *Hamlet* es rarísimo, pero

²Paráfrasis de un célebre comentario de Bertrand Russell.

si *seleccionamos* el mono entre una miríada de ellos *precisamente porque ha tecleado Hamlet*, ya no podemos juzgar el suceso observado del mismo modo. ¡Hemos seleccionado la observación por su rareza, no podemos extrañarnos de que sea rara!

Cuando seleccionamos la evidencia, hemos de tenerlo en cuenta al hacer inferencia. De otro modo, estaremos prejuzgando el resultado.

8.1.3. Análisis exploratorio e inferencia

Es importante entender lo que el Ejemplo 9.3 intenta transmitir. El error, frecuente en el trabajo aplicado, es *seleccionar la evidencia* e ignorar este hecho *al producir afirmaciones o resultados de tipo inferencial* como rechazar tal o cual hipótesis con nivel de significación p , construir tal o cual intervalo con confianza $(1 - p)$. Es el valor de p que reportamos el que resulta completamente irreal a menos que corriamos el efecto de la selección.

Ejemplo 8.4 Regresemos al Ejemplo 9.3. Imaginemos la segunda situación descrita en que uno entre los 10^{955771} monos examinados compone *Hamlet*. Sería incorrecto rechazar la hipótesis H_0 : “Los monos son irracionales.” atribuyendo a esta decisión un nivel de significación de $5,804525 \times 10^{-955771}$. Por el contrario, la probabilidad de que ninguno de los monos hubiera tecleado *Hamlet* sería:

$$\begin{aligned} p_0 &= (1 - p)^{10^{955771}} \\ &= \left[1 - \left(\frac{1}{32} \right)^{635000} \right]^{10^{955770}} \\ &\approx 0,0030138, \end{aligned}$$

el último valor calculado haciendo uso de una aproximación de Poisson (con media $\lambda = 5,804527$). Por tanto, la probabilidad de observar una o más transcripciones de *Hamlet* (un suceso tan raro o más raro que el observado, bajo H_0) ¡es tan grande como $1 - 0,0030138 = 0,9969862$! Difícilmente consideraríamos evidencia contra la hipótesis nula algo que, bajo H_0 , acontece con probabilidad mayor que 0.99.

FIN DEL EJEMPLO ■

Nada nos impide, sin embargo, hacer análisis exploratorio: examinar nuestros datos, y seleccionar como interesante la evidencia que nos lo parezca.

Ejemplo 8.5 De nuevo en el Ejemplo 9.3, no hay nada reproducible en examinar el trabajo de cada uno de los monos y detenernos con toda atención a examinar al animal que produce *Hamlet*. Seguramente le invitaríamos a seguir escribiendo. Sería del mayor interés que *ese mono* produjera a continuación *Macbeth*.

Lo que es reproducible es seleccionar el único mono que teclea *Hamlet* y reportar el hallazgo como si ese mono fuera el único observado.

FIN DEL EJEMPLO ■

8.1.4. Inferencia simultánea y modelo de regresión lineal ordinario

Pero ¿qué tiene esto que ver con el modelo de regresión lineal, objeto de nuestro estudio?

Bastante. En ocasiones, hemos de hacer uso de modelos con un número grande de parámetros. Cuando ello ocurre, hay muchas hipótesis que podemos plantearnos contrastar. Si lo hacemos, hemos de ser conscientes de que algunas hipótesis serán objeto de rechazo con una probabilidad mucho mayor que el nivel de significación nominal empleado para contrastar cada una de ellas. El siguiente ejemplo lo aclara.

Ejemplo 8.6 Supongamos el modelo

$$\vec{Y} = \beta_0 \vec{X}_0 + \beta_1 \vec{X}_1 + \dots + \beta_{99} \vec{X}_{99} + \vec{\epsilon}.$$

Supongamos, por simplicidad, normalidad de las perturbaciones y ortogonalidad de las columnas de la matriz de diseño. Dicho modelo tiene su origen en nuestra completa ignorancia acerca de cuál de las cien variables regresoras consideradas, si es que alguna, influye sobre la respuesta.

Si quisiéramos contrastar la hipótesis $H_0 : \beta_i = 0, i = 0, \dots, 99$, podríamos (si se verifican los supuestos necesarios) emplear el contraste presentado en la Sección 7.2.2, pág. 84. Podríamos ser más ambiciosos e intentar al mismo tiempo ver cuál o cuales β_i son distintos de cero. Sería *incorrecto* operar así:

1. Contrastar las hipótesis $H_{0i} : \beta_i = 0$ al nivel de significación α comparando cada t -ratio en valor absoluto con $t_{N-p}^{\alpha/2}$.
2. Si algún t -ratio excede $t_{N-p}^{\alpha/2}$, rechazar la hipótesis H_{0i} , y por consiguiente H_0 , reportando un nivel de significación α .

Es fácil ver por qué es incorrecto. Bajo H_0 hay probabilidad tan sólo α de que un t -ratio prefijado exceda en valor absoluto de $t_{N-p}^{\alpha/2}$. Pero la probabilidad de que *algún* t -ratio exceda de $t_{N-p}^{\alpha/2}$ es³

$$\text{Prob}(\text{Algún } \beta_i \neq 0) = 1 - (1 - \alpha)^p. \quad (8.2)$$

mayor (en ocasiones *mucho mayor*) que α . Tomemos por ejemplo el caso examinado en que $p = 100$ y supongamos $\alpha = 0,05$. La probabilidad de obtener algún t -ratio fuera de límites es $1 - 0,95^{100} = 0,9940$. Lejos de tener un nivel de significación de $\alpha = 0,05$, el que tenemos es de 0,9940. Contrastar la hipótesis H_0 de este modo tiene una probabilidad de falsa alarma de 0,9940.

Si nuestro propósito fuera puramente exploratorio, nada debe disuadirnos de estimar el modelo con los cien regresores y examinar luego las variables asociadas a t -ratios mayores, quizá estimando un modelo restringido con muestra adicional. Lo que es inadmisibles es dar un nivel de significación incorrectamente calculado.

FIN DEL EJEMPLO ■

El problema de inferencias distorsionadas es grave y muchas veces indetectable. Pensemos en el investigador que hace multitud de regresiones, quizá miles, a cuál más descabellada. Por puro azar, encuentra una pocas con R^2 muy alto, escribe un artículo y lo publica. Si el experimento es reproducible, cabe esperar que otros investigadores tratarán de replicarlo y, al no lograrlo —el R^2 alto era casualidad—, la superchería quedará al descubierto. Pero si la investigación versa sobre, por ejemplo, Ciencias Sociales, en que con frecuencia una y sólo una muestra está disponible, todo lo que sus colegas podrán hacer es reproducir sus resultados con la única muestra a mano. A menos que el primer investigador tenga la decencia de señalar que el alto R^2 obtenido era el más alto entre miles de regresiones efectuadas (lo que permitiría calcular correctamente el nivel de significación y apreciar de un modo realista su valor como evidencia), es fácil que su trabajo pase por ciencia.

De nuevo es preciso insistir: no hay nada objetable en la realización de miles de regresiones, quizá con carácter exploratorio. Tampoco es objetable el concentrar la atención en la única (o las pocas) que parecen prometedoras. Al revés, ello es muy sensato. Lo que es objetable es reportar dichas regresiones como si fueran las únicas realizadas, el resultado de estimar un modelo prefijado de antemano, dando la impresión de que la evidencia muestral sustenta

³Bajo la hipótesis de independencia entre los respectivos t -ratios, hipótesis que se verifica por la normalidad de las perturbaciones y la ortogonalidad entre las columnas de la matriz de diseño.

una hipótesis o modelo pre-establecidos, cuando lo cierto es que la hipótesis o modelo han sido escogidos a la vista de los resultados.

8.2. Desigualdad de Bonferroni.

Consideremos k sucesos, E_i , ($i = 1, \dots, k$), cada uno de ellos con probabilidad $(1 - \alpha)$. Designamos por \overline{E}_i el complementario del suceso E_i . La probabilidad de que todos los sucesos E_i , ($i = 1, \dots, k$) acaezcan simultáneamente es:

$$\text{Prob}\{\cap_{i=1}^k E_i\} = 1 - \text{Prob}\{\overline{\cap_{i=1}^k E_i}\} = 1 - \text{Prob}\{\cup_{i=1}^k \overline{E}_i\} \geq 1 - k\alpha \quad (8.3)$$

Se conoce (9.3) como *desigualdad de Bonferroni de primer orden*. Es una igualdad si los \overline{E}_i son disjuntos. Muestra que la probabilidad conjunta de varios sucesos puede, en general, ser muy inferior a la de uno cualquiera de ellos. Por ejemplo, si $k = 10$ y $\text{Prob}\{E_i\} = 0,95 = 1 - 0,05$, la desigualdad anterior solo permite garantizar que $\text{Prob}\{\cap_{i=1}^k E_i\} \geq 1 - 10 \times 0,05 = 0,50$.

Consideremos ahora el modelo $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ y los siguientes sucesos:

$$E_1 : [(\hat{\beta}_1 \pm \hat{\sigma}_{\hat{\beta}_1} t_{N-p}^{\alpha/2}) \quad \text{cubre } \beta_1] \quad (8.4)$$

$$\vdots \quad (8.5)$$

$$E_k : [(\hat{\beta}_k \pm \hat{\sigma}_{\hat{\beta}_k} t_{N-p}^{\alpha/2}) \quad \text{cubre } \beta_k] \quad (8.6)$$

Cada E_i por separado es un suceso cuya probabilidad es $1 - \alpha$. De acuerdo con (9.3), sin embargo, todo cuanto podemos asegurar acerca de $\text{Prob}\{\cap_{i=1}^k E_i\}$ es que su probabilidad es superior a $1 - k\alpha$.

Las implicaciones son importantes. Si regresáramos \vec{Y} sobre $\vec{X}_0, \dots, \vec{X}_{p-1}$ y quisiéramos obtener intervalos de confianza *simultáneos* α para los parámetros $\beta_0, \dots, \beta_{p-1}$, sería claramente incorrecto emplear los que aparecen en (9.4)–(9.6). Si actuásemos de este modo, el nivel de confianza conjunto no sería el deseado de $1 - \alpha$, sino que tan sólo podríamos afirmar que es mayor que $1 - k\alpha$.

Si queremos intervalos de confianza *simultáneos* al nivel $1 - \alpha$, podríamos construir intervalos para cada uno de los parámetros con un nivel de confianza $\psi = \frac{\alpha}{k}$. Haciendo ésto, tendríamos que la probabilidad de que *todos* los β_i fueran cubiertos por sus respectivos intervalos, sería mayor, de acuerdo con (9.3), que $1 - k\psi = 1 - k(\frac{\alpha}{k}) = 1 - \alpha$. Ello se logra, sin embargo, al coste de ensanchar el intervalo de confianza correspondiente a cada β_i quizá más de lo necesario. En lo que sigue veremos procedimientos para lograr el mismo resultado con intervalos en general más estrechos.

8.3. Intervalos de confianza basados en la máxima t .

Supongamos que tenemos k variables aleatorias independientes, t_1, \dots, t_k con distribución t -Student, y número común n de grados de libertad. La variable aleatoria $\max\{|t_1|, \dots, |t_k|\}$ sigue una distribución que se halla tabulada⁴.

Sea $u_{k,n}^\alpha$ el cuantil $1 - \alpha$ de dicha distribución, es decir, un valor que resulta superado con probabilidad α por $\max\{|t_1|, \dots, |t_k|\}$. Entonces,

$$\text{Prob}\{\cap_{i=1}^k [|t_i| \leq u_{k,n}^\alpha]\} = 1 - \alpha,$$

dado que si $u_{k,n}^\alpha$ acota con probabilidad $1 - \alpha$ al máximo, acota simultáneamente con la misma probabilidad la totalidad de las variables aleatorias.

Si $\vec{a}_i' \hat{\beta} / \hat{\sigma}_{\vec{a}_i' \hat{\beta}}$ ($i = 1, \dots, k$) fueran independientes, y la hipótesis nula $h : \vec{a}_i' \vec{\beta} = 0$ ($i = 1, \dots, k$) fuera cierta, tendríamos que:

$$\text{Prob} \left\{ \bigcap_{i=1}^k \left[\left| \frac{\vec{a}_i' \hat{\beta}}{\hat{\sigma}_{\vec{a}_i' \hat{\beta}}} \right| \leq u_{k,n}^\alpha \right] \right\} = 1 - \alpha \quad (8.7)$$

Es claro que $\vec{a}_i' \hat{\beta} / \hat{\sigma}_{\vec{a}_i' \hat{\beta}}$ ($i = 1, \dots, k$) **no** son independientes. Sin embargo, la distribución aludida del máximo valor absoluto de k variables t de Student está también tabulada cuando dichas variables tienen correlación ρ por pares. (Esto sucede en algunos casos particulares, como el de ciertos diseños de Análisis de Varianza equilibrados: la correlación ρ entre parejas de t -ratios es la misma, y fácil de calcular.)

Aún cuando la correlación ρ por pares de t -ratios no sea siempre la misma, (9.7) es de utilidad. Suministra intervalos simultáneos de confianza aproximada $1 - \alpha$. En caso de que conozcamos ρ , podemos emplear la expresión (9.7) con $u_{k,n}^\alpha$ reemplazado por $u_{k,n,\rho}^\alpha$, extraído éste último de la tabla correspondiente; en caso de que no conozcamos ρ , o ésta no sea constante, podemos utilizar $u_{k,n,\rho=0}^\alpha$, lo que hace en general los intervalos calculados con ayuda de (9.7) conservadores (es decir, la probabilidad conjunta en el lado izquierdo de (9.7) es *mayor* que $1 - \alpha$).

Es importante señalar que, si nuestro objetivo es contrastar una hipótesis del tipo $h : A\vec{\beta} = \vec{c}$ con $\text{rango}(A) > 1$, *tenemos* que emplear un contraste como el descrito en la Sección 7.2, pág. 77. El comparar cada una de las variables aleatorias $\left| (\vec{a}_i' \hat{\beta} - c_i) / \hat{\sigma}_{\vec{a}_i' \hat{\beta}} \right|$ ($i = 1, \dots, k$) con una $t_{N-p}^{\alpha/2}$ supone

⁴Véase, por ej., Seber (1977), Apéndice E.

emplear un nivel de significación *mayor* que α . Como caso particular, es inadecuado contrastar la hipótesis $h: \beta_1 = \dots = \beta_p = 0$ comparando cada uno de los t -ratios con $t_{N-p}^{\alpha/2}$; tal contraste tendría un nivel de significación sensiblemente superior a α , en especial si p es grande.

En el caso de que el contraste conjunto rechace $h: A\vec{\beta} = \vec{c}$ y queramos saber qué filas de A son culpables del rechazo, podríamos comparar $\left| (\vec{a}_i' \hat{\beta} - c_i) / \hat{\sigma}_{\vec{a}_i' \hat{\beta}} \right|$ ($i = 1, \dots, k$) con $u_{k,n}^\alpha$ ($k =$ número de filas de A). Nótese que es perfectamente posible rechazar la hipótesis conjunta y no poder rechazar ninguna de las hipótesis parciales correspondientes a las filas de A .

8.4. Método S de Scheffé.

Este método permite la construcción de un número arbitrario de intervalos de confianza simultáneos, de manera muy simple. Necesitaremos el siguiente lema:

Lema 8.1 *Sea L una matriz simétrica de orden $k \times k$ definida positiva, y \vec{c} , \vec{b} vectores k -dimensionales cualesquiera. Se verifica que:*

$$\sup_{\vec{c} \neq \vec{0}} \left[\frac{[\vec{c}' \vec{b}]^2}{\vec{c}' L \vec{c}} \right] = \vec{b}' L^{-1} \vec{b} \quad (8.8)$$

DEMOSTRACIÓN:

Siendo L definida positiva, existe una matriz R cuadrada no singular tal que: $L = RR'$. Si definimos:

$$\vec{v} = R' \vec{c} \quad (8.9)$$

$$\vec{u} = R^{-1} \vec{b} \quad (8.10)$$

y tenemos en cuenta que por la desigualdad de Schwarz,

$$\frac{\langle \vec{u}, \vec{v} \rangle^2}{\|\vec{u}\|^2 \|\vec{v}\|^2} \leq 1 \quad (8.11)$$

entonces sustituyendo (9.9) y (9.10) en (9.11) obtenemos (9.8). ■

Podemos ahora abordar la construcción de intervalos de confianza simultáneos por el método de Scheffé. Supongamos que tenemos k hipótesis lineales

$h_i: \vec{a}_i' \vec{\beta} = c_i$ ($i = 1, \dots, k$) cuyo contraste conjunto deseamos efectuar. Si denominamos:

$$A = \begin{pmatrix} \vec{a}_1' \\ \vec{a}_2' \\ \dots \\ \vec{a}_k' \end{pmatrix} \quad \vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_k \end{pmatrix} \quad (8.12)$$

dichas k hipótesis se pueden escribir como $h: A\vec{\beta} = \vec{c}$. Cuando h es cierta, sabemos (Sección 7.2) que:

$$\frac{(A\hat{\beta} - \vec{c})'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - \vec{c})}{q\hat{\sigma}^2} \sim \mathcal{F}_{q, N-p} \quad (8.13)$$

siendo $q = \min(d, p)$, en que $d = \text{rango } A$ y $p = \text{rango}(X'X)$. Las inversas pueden ser inversas generalizadas, si los rangos de las matrices así lo exigen.

Llamemos \hat{c} a $A\hat{\beta}$. Bajo h , sabemos que:

$$1 - \alpha = \text{Prob} \left\{ (\hat{c} - \vec{c})'[A(X'X)^{-1}A']^{-1}(\hat{c} - \vec{c}) \leq q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha \right\} \quad (8.14)$$

$$= \text{Prob} \left\{ (\hat{c} - \vec{c})'L^{-1}(\hat{c} - \vec{c}) \leq q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha \right\} \quad (8.15)$$

en que $L = [A(X'X)^{-1}A']$. Teniendo en cuenta el Lema 9.1, obtenemos:

$$1 - \alpha = \text{Prob} \left\{ \sup_{\vec{h} \neq \vec{0}} \left[\frac{[\vec{h}'(\hat{c} - \vec{c})]^2}{\vec{h}'L\vec{h}} \right] \leq q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha \right\} \quad (8.16)$$

$$= \text{Prob} \left\{ \bigcap_{\vec{h} \neq \vec{0}} \left[\left| \frac{\vec{h}'(\hat{c} - \vec{c})}{(\vec{h}'L\vec{h})^{\frac{1}{2}}} \right| \leq (q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha)^{\frac{1}{2}} \right] \right\} \quad (8.17)$$

La ecuación (9.17) muestra que $(q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha)^{\frac{1}{2}}$ es un valor que acota con probabilidad $1 - \alpha$ un número arbitrariamente grande de cocientes como:

$$\frac{|\vec{h}'(\hat{c} - \vec{c})|}{\sqrt{\vec{h}'L\vec{h}}} \quad (8.18)$$

Por consiguiente, cuantos intervalos para $\vec{h}'\vec{c}$ construyamos de la forma:

$$\vec{h}'\hat{c} \pm \sqrt{(\vec{h}'L\vec{h})(q\hat{\sigma}^2 \mathcal{F}_{q, N-p}^\alpha)} \quad (8.19)$$

tendrán confianza *simultánea* $1 - \alpha$.

Esto es *más* de lo que necesitamos —pues sólo queríamos intervalos de confianza simultáneos para c_1, \dots, c_k —. El método de Scheffé proporciona intervalos de confianza conservadores (más amplios, en general, de lo estrictamente necesario).

Obsérvese que, en el caso particular en que $A = I_{p \times p}$, los intervalos de confianza en (9.19) se reducen a:

$$\vec{h}'\hat{\beta} \pm \sqrt{(\vec{h}'(X'X)^{-1}\vec{h})(p\hat{\sigma}^2\mathcal{F}_{p,N-p}^\alpha)} \quad (8.20)$$

expresión que será frecuente en la práctica. Cuando el conjunto de hipótesis simultáneas que se contrastan configure una matriz A de rango $q < p$, será sin embargo conveniente tener en cuenta este hecho, ya que obtendremos intervalos menos amplios.

R: Ejemplo 8.1 (*uso del método de Scheffé*)

El siguiente código implementa el método de Scheffé para contrastar la igualdad entre todas las parejas de parámetros intervinientes en un modelo. La matriz de diseño es una matriz de ceros y unos. Si, por ejemplo, X_{kl} fuera “uno” cuando la k -ésima parcela se siembra con la variedad l -ésima de semilla y la variable respuesta recogiera las cosechas obtenidas en las diferentes parcelas, los parámetros β_i serían interpretables como la productividad de las diferentes variedades de semilla (suponemos que no hay otros factores en juego; las parcelas son todas homogéneas).

En una situación como la descrita tendría interés contrastar todas las hipótesis del tipo: $h_{ij} : \beta_i - \beta_j = 0$. Aquellas parejas para las que no se rechazase corresponderían a variedades de semilla no significativamente diferentes.

Fácilmente se ve que el contraste de todas las hipótesis de interés agrupadas ($h : A\vec{\beta} = \vec{c}$) no es de gran interés: no nos interesa saber si *hay* algunas variedades de semilla diferentes, sino *cuáles son*. Fácilmente se ve también que, incluso para un número moderado de variedades de semilla, hay bastantes parejas que podemos formar y el realizar múltiples contrastes como $h_{ij} : \beta_i - \beta_j = 0$ requerirá el uso de métodos de inferencia simultánea.

Comencemos por construir una matriz de diseño y generar artificialmente las observaciones:

```
> X <- matrix(c(rep(1, 5), rep(0, 25)),
+           25, 5)
> X
```

```

      [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    1    0    0    0    0
[3,]    1    0    0    0    0
[4,]    1    0    0    0    0
[5,]    1    0    0    0    0
[6,]    0    1    0    0    0
[7,]    0    1    0    0    0
[8,]    0    1    0    0    0
[9,]    0    1    0    0    0
[10,]   0    1    0    0    0
[11,]   0    0    1    0    0
[12,]   0    0    1    0    0
[13,]   0    0    1    0    0
[14,]   0    0    1    0    0
[15,]   0    0    1    0    0
[16,]   0    0    0    1    0
[17,]   0    0    0    1    0
[18,]   0    0    0    1    0
[19,]   0    0    0    1    0
[20,]   0    0    0    1    0
[21,]   0    0    0    0    1
[22,]   0    0    0    0    1
[23,]   0    0    0    0    1
[24,]   0    0    0    0    1
[25,]   0    0    0    0    1

```

```

> b <- c(3, 4, 4, 5, 5)
> y <- X %*% b + rnorm(25, sd = 0.1)

```

Construyamos la matriz definiendo la hipótesis conjunta $A\vec{\beta} = \vec{c}$:

```

> p <- ncol(X)
> N <- nrow(X)
> A <- cbind(1, diag(-1, p - 1))
> A

```

```

      [,1] [,2] [,3] [,4] [,5]
[1,]    1   -1    0    0    0
[2,]    1    0   -1    0    0
[3,]    1    0    0   -1    0
[4,]    1    0    0    0   -1

```

```

> q <- nrow(A)

```

Aunque por motivos didácticos hemos construido A del modo que se ha visto, hay funciones standard que permiten hacerlo con mayor comodidad.

```
> A <- t(contrasts(as.factor(1:5)))
> A

      1 2 3 4 5
[1,] 1 0 0 0 -1
[2,] 0 1 0 0 -1
[3,] 0 0 1 0 -1
[4,] 0 0 0 1 -1
```

que es equivalente a la A precedente.

Habiendo p betas a comparar, habrá un total de $\frac{p(p-1)}{2}$ comparaciones a efectuar. Construimos una matriz cada una de cuyas filas corresponde a una comparación:

```
> H <- matrix(0, p * (p - 1)/2, p)
> j <- 0
> for (i in ((p - 1):1)) {
+   H[(j + 1):(j + i), (p - i):p] <- cbind(1,
+     diag(-1, i))
+   j <- j + i
+ }
> H

      [,1] [,2] [,3] [,4] [,5]
[1,]    1   -1    0    0    0
[2,]    1    0   -1    0    0
[3,]    1    0    0   -1    0
[4,]    1    0    0    0   -1
[5,]    0    1   -1    0    0
[6,]    0    1    0   -1    0
[7,]    0    1    0    0   -1
[8,]    0    0    1   -1    0
[9,]    0    0    1    0   -1
[10,]   0    0    0    1   -1
```

El siguiente fragmento de código construye ahora todos los intervalos de la forma dada por (9.20) y los imprime:

```

> fit <- lsfit(X, y, intercept = FALSE)
> betas <- fit$coefficients
> s2 <- sum(fit$residuals^2)/(N - p)
> qsf <- q * s2 * qf(0.05, q, N - p)
> xxi <- solve(t(X) %*% X)
> for (i in 1:nrow(H)) {
+   cat("Intervalo comp. ", H[i, ])
+   z <- sqrt(t(H[i, ]) %*% xxi %*% H[i,
+     ]) * qsf)
+   d <- t(H[i, ]) %*% betas
+   cat(" es: (", d - z, " , ", d + z,
+     ")")
+   if ((d - z < 0) && (d + z > 0))
+     cat("\n")
+   else cat(" * \n")
+ }

Intervalo comp.  1 -1 0 0 0 es: ( -1.0463 , -0.94141 ) *
Intervalo comp.  1 0 -1 0 0 es: ( -1.0631 , -0.95825 ) *
Intervalo comp.  1 0 0 -1 0 es: ( -2.0886 , -1.9837 ) *
Intervalo comp.  1 0 0 0 -1 es: ( -2.0670 , -1.9622 ) *
Intervalo comp.  0 1 -1 0 0 es: ( -0.069268 , 0.035591 )
Intervalo comp.  0 1 0 -1 0 es: ( -1.0947 , -0.98989 ) *
Intervalo comp.  0 1 0 0 -1 es: ( -1.0732 , -0.96834 ) *
Intervalo comp.  0 0 1 -1 0 es: ( -1.0779 , -0.97305 ) *
Intervalo comp.  0 0 1 0 -1 es: ( -1.0564 , -0.9515 ) *
Intervalo comp.  0 0 0 1 -1 es: ( -0.030881 , 0.073979 )

```

Vemos que la mayoría de intervalos de confianza simultáneos no cubren el cero. Los correspondientes a $\beta_2 - \beta_3$ y $\beta_4 - \beta_5$ si lo hacen, como esperábamos, ya que en ambas parejas los parámetros han sido fijados al mismo valor.

FIN DEL EJEMPLO ■

8.5. Empleo de métodos de inferencia simultánea.

Si el desarrollo anterior es formalmente simple, puede no ser obvio, en cambio, en que situaciones es de aplicación. Las notas siguientes esbozan algunas ideas sobre el particular⁵.

- Emplearemos inferencia simultánea cuando *a priori*, y por cualquier motivo, estemos interesados en múltiples contrastes (o intervalos de confianza) y queramos que el nivel de significación conjunto sea $1 - \alpha$. Esta situación se presenta con relativa rareza en la práctica estadística.
- Más importante, emplearemos los métodos anteriores cuando la elección de hipótesis o parámetros objeto de contraste o estimación *se haga a la vista de los resultados*. Esta situación es muy frecuente en el análisis exploratorio. Sería incorrecto, por ejemplo, estimar una ecuación con veinte regresores, seleccionar aquel $\hat{\beta}_i$ con el máximo t-ratio, y comparar dicho t-ratio con una t de Student con grados de libertad adecuados. Dado que hemos seleccionado el $\hat{\beta}_i$ de interés como el de mayor t-ratio, hemos de comparar éste con los cuantiles de la distribución del máximo de k ($k = 20$ en este caso) variables aleatorias con distribución t de Student ($u_{20, N-20}^\alpha$).
- Por último, conviene resaltar la diferencia entre el contraste de varias hipótesis simultáneas $\vec{a}_i' \vec{\beta} = c_i$ agrupadas en $A\vec{\beta} = \vec{c}$ mediante Q_h (Sección 7.2) y el que hace uso de (9.7). El primero es perfectamente utilizable; el segundo será, en general, conservador —menos rechazos de los que sugiere el nivel de significación nominal—, pero tiene la ventaja de arrojar luz sobre cuales de las “subhipótesis” $\vec{a}_i' \vec{\beta} = c_i$ son responsables del rechazo, caso de que se produzca. Esta información queda sumergida al emplear Q_h .

⁵Puede consultarse también Trocóniz (1987a) Cap. 5 y Cox and Hinkley (1974), Sec. 7.4.

COMPLEMENTOS Y EJERCICIOS

8.1 Un investigador sospecha que la concentración de una toxina en la sangre puede estar relacionada con la ingesta de algún tipo de alimento. Realiza un completo estudio en que para $N = 500$ sujetos mide la concentración de dicha toxina y las cantidades consumidas de 200 diferentes tipos de alimento. Cree razonable proponer como modelo explicativo,

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{200} X_{200} + \epsilon.$$

Tras estimar los 201 parámetros del mismo, se plantea contrastar la hipótesis como $H_0 : \beta_1 = \dots = \beta_{200}$ y considera las siguientes posibilidades:

- Comparar cada uno de los t-ratios $\hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i}$ con el cuantil $t_{N-p; \alpha/2}$.
- Idem con el cuantil correspondiente de una distribución del máximo de k variables t de Student, con grados de libertad apropiados.
- Calcular el estadístico Q_h para la hipótesis $H_0 : \hat{\beta}_1, \dots, \hat{\beta}_{200} = 0$ y comparar con $\mathcal{F}_{200, 500-201; \alpha}$.

Juzga los diferentes procedimientos, e indica con cuál (o cuáles) de ellos tendríamos garantizada una probabilidad de error de tipo I no superior al α prefijado.

8.2 Preocupado por el posible impacto de las antenas de telefonía móvil sobre la salud de los niños, un político solicita un listado completo de las 15320 escuelas del país a menos de 500 metros de una antena. Investiga la probabilidad de contraer leucemia y la probabilidad de que por puro azar se presenten los casos de leucemia que se han registrado en dichas escuelas.

Aparece un caso llamativo: en la escuela X con 650 niños hay tres que han contraído la enfermedad, lo que, de acuerdo con los cálculos realizados por nuestro político, asistido por un epidemiólogo, acontecería por azar con probabilidad 0,0003. Al día siguiente acude al Parlamento y pide la dimisión del Ministro de Sanidad: “Hay —dice— evidencia concluyente de que las antenas de telefonía móvil influyen en la prevalencia de la leucemia entre la población infantil. Un evento como el registrado en la escuela X sólo se presentaría por azar con probabilidad 0,0003”. Comenta.