Capítulo 9

Multicolinealidad.

9.1. Introducción.

Hemos visto (Capítulo 4) que, en presencia de multicolinealidad exacta entre las columnas de la matriz de diseño X, la proyección de $\vec{y}$ sobre $M = R(X)$ sigue siendo única, pero no hay una única estimación de $\vec{\beta}$. Decíamos entonces que el vector de parámetros no estaba identificado.

Este Capítulo[1] analiza esta cuestión con mayor detalle. En particular, aborda las siguientes cuestiones:

1. ¿Es estimable una cierta combinación lineal $\vec{c}\,'\vec{\beta}$ de los parámetros?

2. Si $\vec{c}\,'\vec{\beta}$ es estimable, ¿cuál es la varianza de la estimación?. ¿De qué depende la precisión con que pueden estimarse distintas combinaciones lineales de los parámetros?

3. ¿Como escoger la matriz de diseño X —u observaciones adicionales a la misma— si el objetivo es estimar determinadas combinaciones lineales $\vec{c}\,'\vec{\beta}$ con varianza mínima?

Responder a la primera requiere que caractericemos las formas lineales estimables. Nótese que cuando $\vec{c}$ es un vector de ceros con un 1 en una única posición, la primera cuestión incluye, como caso particular, la de si un parámetro concreto es estimable.

[1] Basado en Silvey (1969).

La segunda cuestión introducirá la idea de multicolinealidad aproximada. Mientras que desde un punto de vista formal la matriz de diseño es de rango deficiente o no lo es, en la práctica interesa distinguir aquéllas situaciones en que la matriz de diseño es de rango "casi" deficiente. Cuando esto ocurra, en un sentido que se aclarará más abajo, todo es estimable, pero algunas formas lineales $\vec{c}\,'\vec{\beta}$ lo son con gran imprecisión: la varianza de su mejor estimador lineal insesgado depende de la dirección del vector $\vec{c}$ en $R(X'X)$.

La tercera cuestión hace referencia a un tema de gran interés; el de diseño óptimo. Admitido que algunas formas lineales quizá sólo pueden ser estimadas con gran varianza ¿cómo habría que escoger o ampliar X en los casos en que somos libres de ampliar la muestra?

El principal hallazgo al responder a las dos primeras cuestiones será que combinaciones lineales $\vec{c}\,'\vec{\beta}$ con $\vec{c}$ aproximadamente colineal a un vector propio de $(X\,'X)$ de valor propio asociado "pequeño", son las de estimación más imprecisa. La consecuencia será que haremos lo posible en nuestros diseños experimentales para que, si $\vec{c}\,'\vec{\beta}$ es una forma lineal de interés, no haya vectores propios de $(X\,'X)$ con valor propio pequeño aproximadamente en la misma dirección de $\vec{c}$. Recurriremos para ello a ampliar la muestra, si podemos hacerlo, o a procedimientos *ad-hoc* de manipulación de dichos valores propios pequeños para obtener estimadores diferentes del MCO. Esta cuestión se estudia en el Capítulo 11.

Realizaremos un análisis formal de la multicolinealidad en las Secciones 10.4 y siguientes. Previamente será de interés abordar la cuestión desde una perspectiva informal (en la Sección 10.2) y examinar los síntomas que evidencian problemas de multicolinealidad en una matriz de diseño (Sección 10.3).

9.2. Una aproximación intuitiva

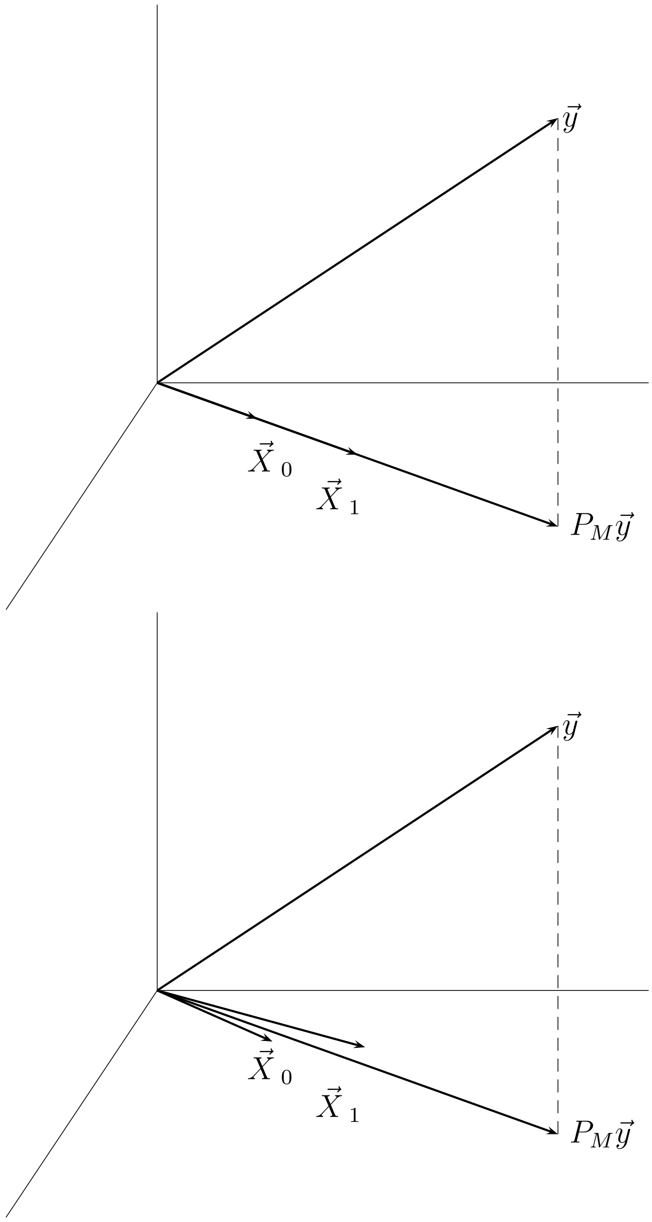
La Figura 10.1 recoge sendas situaciones de multicolinealidad exacta (en el panel superior) y multicolinealidad aproximada (en el inferior). En el panel superior,

$$P_M\vec{y} = \begin{bmatrix}5{,}3\\1{,}9\end{bmatrix} \qquad \vec{X}_0 = \begin{bmatrix}2{,}65\\0{,}95\end{bmatrix} \qquad \vec{X}_1 = \begin{bmatrix}1{,}325\\0{,}475\end{bmatrix} \tag{9.1}$$

Puede comprobarse que $\vec{X}_0 = 0{,}5 \times \vec{X}_1$, por lo que la matriz de diseño que tuviera a ambos vectores por columnas sería de rango deficiente. Consecuentemente, los estimadores MCO de los parámetros β_0 y β_1 no están unívocamente determinados. Puede comprobarse que

$$P_M\vec{y} = \hat{\beta}_0\vec{X}_0 + \hat{\beta}_1\vec{X}_1 \tag{9.2}$$

Figura 9.1: Multicolinealidad exacta (panel superior) y aproximada (panel inferior).



se verifica con $\hat{\beta}_0 = 2$ y $\hat{\beta}_1 = 0$ ó con $\hat{\beta}_0 = 0$ y $\hat{\beta}_1 = 4$, por ejemplo. De hecho, cualesquiera $\hat{\beta}_0$, $\hat{\beta}_1$ verificando $\hat{\beta}_0 + 2\hat{\beta}_1 = 2$ son una solución de (10.2).

En el panel inferior de la Figura 10.1,

$$P_M\vec{y} = \begin{bmatrix} 5{,}3 \\ 1{,}9 \end{bmatrix} \qquad \vec{X}_0 = \begin{bmatrix} 2{,}75 \\ 0{,}75 \end{bmatrix} \qquad \vec{X}_1 = \begin{bmatrix} 1{,}525 \\ 0{,}675 \end{bmatrix}; \tag{9.3}$$

puede comprobarse que ahora $P_M\vec{y} = 0{,}9544\vec{X}_0 + 1{,}7544\vec{X}_1$. Si, no obstante, $P_M\vec{y}$ fuera ligeramente diferente, con los mismos regresores,

$$P_M\vec{y} = \begin{bmatrix} 5{,}4 \\ 1{,}8 \end{bmatrix} \qquad \vec{X}_0 = \begin{bmatrix} 2{,}75 \\ 0{,}75 \end{bmatrix} \qquad \vec{X}_1 = \begin{bmatrix} 1{,}525 \\ 0{,}675 \end{bmatrix} \tag{9.4}$$

tendríamos que la solución única sería $P_M\vec{y} = 1{,}263\vec{X}_0 + 1{,}2632\vec{X}_1$. Una pequeña perturbación en $P_M\vec{y}$ ha originado un cambio drástico en los valores de los estimadores.

Si examinamos el panel inferior de la Figura 10.1, podemos entender fácilmente lo que sucede: los regresores son linealmente independientes y generan el plano horizontal, pero tienen una colinealidad acusada. Un leve cambio en la posición de $P_M\vec{y}$ hace que sea mucho más colineal con un regresor que con otro, y provoca una drástica modificación en los valores de $\hat{\beta}_0$ y $\hat{\beta}_1$.

Tenemos así que si en situaciones de multicolinealidad exacta los parámetros (o algunos de entre ellos) son radicalmente inestimables, cuando el rango de la matrix X es completo, pero algunas de sus columnas son acusadamente colineales, la estimación es posible, pero imprecisa. Decimos que estamos ante una situación de *multicolinealidad aproximada*.

La multicolinealidad aproximada es, en esencia, una matriz de diseño pobre, que no permite deslindar con precisión el efecto de cada regresor sobre la variable respuesta. Es una situación muy frecuente en la práctica, a medio camino entre la multicolinealidad exacta y la ortogonalidad entre los regresores. La Sección que sigue detalla algunos síntomas que permiten percibir su existencia.

9.3. Detección de la multicolinealidad aproximada

Hay algunos indicios y estadísticos que pueden ayudar en el diagnóstico de multicolinealidad.

Elevado R^2 y todos los parámetros no significativos. La multicolinealidad aproximada se pone de manifiesto en elevadas varianzas de los parámetros estimados que, como consecuencia, son de ordinario no significativos y frecuentemente toman signos contrarios a los previstos.

Una situación típica es aquélla, aparentemente paradójica, en que todos los parámetros en $\vec{\beta}$ son no significativos y sin embargo R^2 es muy elevado. ¡Parece que ningún regresor ayuda a ajustar el regresando, y sin embargo todos en conjunto lo hacen muy bien! Ello se debe a que la multicolinealidad no permite deslindar la contribución de cada regresor.

Valores propios y "número de condición" de $(X'X)$. La existencia de relaciones lineales aproximadas entre las columnas de X se traduce en relaciones lineales aproximadas entre las columnas de $(X'X)$ (ver nota al pie de la página 127). Los métodos usuales para examinar el condicionamiento de una matriz en análisis numérico son por tanto de aplicación. En particular, puede recurrirse a calcular los valores propios de la matriz $(X'X)$; uno o mas valores propios muy pequeños (cero, en caso de multicolinealidad perfecta) son indicativos de multicolinealidad aproximada.

A menudo se calcula el "número de condición" de la matriz $(X'X)$, definido como λ_1/λ_p; números de condición "grandes" evidencian gran disparidad entre el mayor y menor valor propio, y consiguientemente multicolinealidad aproximada. Hay que notar, sin embargo, que se trata de un indicador relativo, que, en particular, depende de la escala en que se miden las respectivas columnas de la matriz X —algo perfectamente arbitrario—.

Factores de incremento de varianza (VIF). Otra práctica muy usual consiste en regresar cada columna de X sobre las restantes; un R^2 muy elevado en una o más de dichas regresiones evidencia una relación lineal aproximada entre la variable tomada como regresando y las tomadas como regresores.

Llamemos $R^2(i)$ al R^2 resultante de regresar $\vec{X}_i$ sobre las restantes columnas de X. Se define el *factor de incremento de varianza* (variance inflation factor) VIF(i) así:

$$\text{VIF}(i) \stackrel{\text{def}}{=} \frac{1}{1-R^2(i)}; \tag{9.5}$$

valores de VIF(i) mayores que 10 (equivalentes a $R^2(i) > 0{,}90$) se consideran indicativos de multicolinealidad afectando a $\vec{X}_i$ junto a alguna de las restantes columnas de X.

Observación 9.1 El nombre de "factores de incremento de varianza" tiene la siguiente motivación. Supongamos que X tiene sus columnas normalizadas de modo que $(X'X)$ es una matriz de correlación (elementos diagonales unitarios). La varianza de $\hat{\beta}_i$ es $\sigma^2(X'X)^{ii}$, en que $(X'X)^{ii}$ denota el elemento en la fila y columna i de la matriz $(X'X)^{-1}$.

Si X tuviera sus columnas ortogonales, $(X'X)$ (y por tanto $(X'X)^{-1}$) serían matrices unidad y $\text{Var}(\hat{\beta}_i) = \sigma^2$; por tanto, $(X'X)^{ii}$ recoge el factor en que se modifica en general $\text{Var}(\hat{\beta}_i)$ respecto de la situación de mínima multicolinealidad (= regresores ortogonales). Se puede demostrar que $(X'X)^{ii} = (1 - R^2(i))^{-1}$, lo que muestra que se trata precisamente del VIF(i).

9.4. Caracterización de formas lineales estimables.

Teorema 9.1 *La forma lineal $\vec{c}\,'\vec{\beta}$ es estimable si, y solo si, $\vec{c}$ es una combinación lineal de los vectores propios de $X'X$ asociados a valores propios no nulos.*

DEMOSTRACIÓN:

Observemos que el enunciado no es sino una paráfrasis del Teorema 4.1, pág. 47. La siguiente cadena de implicaciones, que puede recorrerse en ambas direcciones, establece la demostración.

$$\vec{c}\,'\vec{\beta} \text{ estimable} \iff \exists \vec{d}\colon \vec{c}\,'\vec{\beta} = E[\vec{d}\,'\vec{Y}\,] \tag{9.6}$$
$$\iff \vec{c}\,'\vec{\beta} = \vec{d}\,'X\vec{\beta} \tag{9.7}$$
$$\iff \vec{c}\,' = \vec{d}\,'X \tag{9.8}$$
$$\iff \vec{c} = X'\vec{d} \tag{9.9}$$
$$\iff \vec{c} \in R(X') \tag{9.10}$$
$$\iff \vec{c} \in R(X'X) \tag{9.11}$$
$$\iff \vec{c} = \alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j} \tag{9.12}$$

siendo $\vec{v}_1, \ldots, \vec{v}_{p-j}$ los vectores propios de $(X'X)$ asociados a valores propios no nulos. El paso de (10.10) a (10.11) hace uso del hecho de que tanto las columnas de X' como las de $X'X$ generan el mismo subespacio[2] de R^p. La

[2] Es inmediato ver que $R(X'X) \subseteq R(X')$, pues si $\vec{v} \in R(X'X) \Rightarrow \exists \vec{a}\colon\ \vec{v} = X'X\vec{a} = X'\vec{d}$, siendo $\vec{d} = X\vec{a}$. Por otra parte, $R(X'X)$ no es subespacio propio de $R(X')$, pues ambos tienen la misma dimensión. Para verlo, basta comprobar que toda dependencia lineal entre las columnas de $X'X$ es una dependencia lineal entre las columnas de X. En efecto, $X'X\vec{b} = \vec{0} \Rightarrow \vec{b}\,'X'X\vec{b} = \vec{d}\,'\vec{d} = \vec{0} \Rightarrow \vec{d} = \vec{0} \Rightarrow X\vec{b} = \vec{0}$.

equivalencia entre (10.11) y (10.12) hace uso del hecho de que los vectores propios de $R(X'X)$ asociados a valores propios no nulos generan $R(X'X)$.

∎

Hay una forma alternativa de llegar al resultado anterior, que resulta interesante en sí misma y útil para lo que sigue. Sea V la matriz diagonalizadora de $X'X$, y definamos:

$$Z = XV \tag{9.13}$$

$$\vec{\gamma} = V'\vec{\beta} \tag{9.14}$$

Entonces, como $VV' = I$ tenemos que:

$$X\vec{\beta} = XVV'\vec{\beta} = Z\vec{\gamma} \tag{9.15}$$

y por consiguiente el modelo $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ se transforma en: $\vec{Y} = Z\vec{\gamma} + \vec{\epsilon}$.

El cambio de variables y parámetros ha convertido la matriz de diseño en una matriz de columnas ortogonales:

$$Z'Z = (XV)'(XV) = V'X'XV = \Lambda \tag{9.16}$$

siendo Λ una matriz cuya diagonal principal contiene los valores propios de $X'X$. Sin pérdida de generalidad los supondremos ordenados de forma que los $p-j$ primeros $\lambda's$ son no nulos, y los restantes j son cero: $\lambda_p = \lambda_{p-1} = \cdots = \lambda_{p-j+1} = 0$.

Observemos que de (10.14) se deduce, dado que V es ortogonal, que $\vec{\beta} = V\vec{\gamma}$. Por consiguiente, es equivalente el problema de estimar $\vec{\beta}$ al de estimar $\vec{\gamma}$, pues el conocimiento de un vector permite con facilidad recuperar el otro. Las ecuaciones normales al estimar $\vec{\gamma}$ son:

$$(Z'Z)\hat{\gamma} = \Lambda\hat{\gamma} = Z'\vec{y} \tag{9.17}$$

o en forma desarrollada:

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \lambda_{p-j} & \ldots & 0 \\ 0 & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \ldots & 0 \end{pmatrix} \hat{\gamma} = Z'\vec{y} \tag{9.18}$$

El sistema (10.18) es indeterminado; solo los $(p-j)$ primeros $\hat{\gamma}'s$ pueden obtenerse de él. Obsérvese además que de (10.18) se deduce que var$(\hat{\gamma}_i)$ $\propto 1/\lambda_i$, $(i = 1, \ldots, p-j)$.

Consideremos una forma lineal cualquiera $\vec{c}\,'\vec{\beta}$. Tenemos que:

$$\vec{c}\,'\vec{\beta} = \vec{c}\,'VV'\vec{\beta} = (\vec{c}\,'V)\vec{\gamma} = (V'\vec{c})'\vec{\gamma} \tag{9.19}$$

y consiguientemente una estimación de $\vec{c}\,'\hat{\beta}$ vendrá dada por $(V'\vec{c})'\hat{\gamma}$. Por tanto, $\vec{c}\,'\vec{\beta}$ será estimable si $\hat{\gamma}$ es estimable, *o si $\vec{c}\,'\hat{\beta}$ depende sólo de aquellos $\hat{\gamma}'s$ que pueden ser estimados*. Es decir, en el caso de rango $(p-j)$ correspondiente a las ecuaciones normales (10.18), $\vec{c}\,'\vec{\beta}$ podrá estimarse si $(V'\vec{c})'$ tiene nulas sus últimas j coordenadas, lo que a su vez implica:

$$\vec{c} \perp \vec{v}_p \tag{9.20}$$
$$\vec{c} \perp \vec{v}_{p-1} \tag{9.21}$$
$$\vdots \tag{9.22}$$
$$\vec{c} \perp \vec{v}_{p-j+1} \tag{9.23}$$

Para que $\vec{c}\,'\vec{\beta}$ sea estimable, $\vec{c}$ debe poder escribirse como combinación lineal de los vectores propios de $(X'X)$ que no figuran en (10.20)–(10.23): $\vec{c} = \alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j}$. Toda forma estimable debe por tanto ser expresable así:

$$\vec{c}\,'\vec{\beta} = (\alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j})'\vec{\beta}, \tag{9.24}$$

resultado al que habíamos llegado.

Recapitulemos: una forma lineal $\vec{c}\,'\vec{\beta}$ es estimable si $\vec{c} = \alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j}$, es decir, no depende de vectores propios de $(X'X)$ asociados a valores propios nulos. Tal como sugería la Sección 10.2, podemos sin embargo esperar que formas lineales que son estrictamente estimables lo sean muy imprecisamente, en situaciones de multicolinealidad aproximada. La Sección que sigue formaliza esta intuición, mostrando que si $\vec{c}$ depende de vectores propios de valor propio cercano a cero, la forma lineal $\vec{c}\,'\vec{\beta}$ será estimable sólo con gran varianza.

9.5. Varianza en la estimación de una forma lineal.

Si premultiplicamos ambos lados de las ecuaciones normales $(X'X)\hat{\beta} = X'\vec{Y}$ por $\vec{v}_i$, $(i = 1, \ldots, p-j)$, tenemos:

$$\begin{aligned}
\vec{v}_i{}'(X'X)\hat{\beta} &= \vec{v}_i{}'X'\vec{Y} \\
\lambda_i\vec{v}_i{}'\hat{\beta} &= \vec{v}_i{}'X'\vec{Y}
\end{aligned}$$

y tomando varianzas a ambos lados:

$$\begin{aligned}
\lambda_i^2\text{var}(\vec{v}_i{}'\hat{\beta}) &= \text{var}(\vec{v}_i{}'X'\vec{Y}\,) \\
&= \vec{v}_i{}'X'\sigma^2 IX\vec{v}_i \\
&= \vec{v}_i{}'X'X\vec{v}_i\sigma^2 \\
&= \lambda_i\sigma^2 \qquad (9.25)
\end{aligned}$$

De la igualdad (10.25) se deduce que:

$$\text{var}(\vec{v}_i{}'\hat{\beta}) = \frac{\sigma^2}{\lambda_i} \qquad (9.26)$$

Además, para cualquier $i \neq j$ se tiene:

$$\begin{aligned}
\text{cov}(\vec{v}_i{}'\hat{\beta}, \vec{v}_j{}'\hat{\beta}) &= \vec{v}_i{}'\Sigma_{\hat{\beta}}\vec{v}_j \\
&= \vec{v}_i{}'(X\,'X)^{-1}\vec{v}_j\sigma^2 \\
&= \vec{v}_i{}'\lambda_j{}^{-1}\vec{v}_j\sigma^2 \\
&= \sigma^2\lambda_j{}^{-1}\vec{v}_i{}'\vec{v}_j \\
&= 0 \qquad (9.27)
\end{aligned}$$

La varianza de cualquier forma estimable $\vec{c}\,'\vec{\beta}$, teniendo en cuenta que puede escribirse como en (10.24), y haciendo uso de (10.26) y (10.27), será:

$$\begin{aligned}
\text{var}(\vec{c}\,'\hat{\beta}) &= \text{var}[(\alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j})'\hat{\beta}] \\
&= \alpha_1^2\text{var}(\vec{v}_1{}'\hat{\beta}) + \cdots + \alpha_{p-j}^2\text{var}(\vec{v}_{p-j}{}'\hat{\beta}) \\
&= \alpha_1^2\left[\frac{\sigma^2}{\lambda_1}\right] + \cdots + \alpha_{p-j}^2\left[\frac{\sigma^2}{\lambda_{p-j}}\right] \\
&= \sigma^2\left[\frac{\alpha_1^2}{\lambda_1} + \cdots + \frac{\alpha_{p-j}^2}{\lambda_{p-j}}\right] \qquad (9.28)
\end{aligned}$$

La expresión (10.28) es reveladora; la varianza en la estimación de $\vec{c}\,'\vec{\beta}$ dependerá de la varianza de la perturbación σ^2 y de la dirección de $\vec{c}$. Si $\vec{c}$ no puede expresarse como combinación lineal de los vectores propios con valor propio no nulo, $\vec{c}\,'\vec{\beta}$ no es estimable. Si $\vec{c} = \alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j}$ y los $\alpha's$ multiplicando a vectores propios con reducido valor propio son sustanciales, los correspondientes sumandos tenderán a dominar la expresión (10.28).

En definitiva, la varianza en la estimación de una forma lineal $\vec{c}\,'\vec{\beta}$ depende, fundamentalmente, de cuán colineal es $\vec{c}$ con vectores propios de reducido valor propio.

Hemos razonado en esta Sección y la precedente en el caso de que j valores propios de $X'X$ son exactamente cero. Es claro que si todos los valores propios son mayores que cero, todas las formas lineales serán estimables, con varianza:

$$\begin{aligned}
\mathrm{var}(\vec{c}\,'\hat{\beta}) &= \mathrm{var}[(\alpha_1\vec{v}_1 + \cdots + \alpha_{p-j}\vec{v}_{p-j})'\hat{\beta}] && (9.29)\\
&= \alpha_1^2\mathrm{var}(\vec{v}_1{}'\hat{\beta}) + \cdots + \alpha_p^2\mathrm{var}(\vec{v}_p{}'\hat{\beta})\\
&= \alpha_1^2\left[\frac{\sigma^2}{\lambda_1}\right] + \cdots + \alpha_p^2\left[\frac{\sigma^2}{\lambda_p}\right]\\
&= \sigma^2\left[\frac{\alpha_1^2}{\lambda_1} + \cdots + \frac{\alpha_p^2}{\lambda_p}\right] && (9.30)
\end{aligned}$$

9.6. Elección óptima de observaciones.

La expresión (10.28) y comentario posterior muestran que, para guarecernos de varianzas muy grandes en la estimación de algunas formas lineales, debemos actuar sobre los valores propios más pequeños de $(X'X)$, incrementándolos[3]. En lo que sigue, examinamos esta cuestión con más detalle.

Supongamos que tenemos un conjunto de N observaciones $(\vec{y} \mid X)$, y nos planteamos ampliar X con una fila adicional $\vec{x}_{N+1}\,'$ (e $\vec{y}$ con el correspondiente valor observado de Y) de modo que se reduzca al máximo la varianza en la estimación de una determinada forma lineal $\vec{c}\,'\vec{\beta}$ en que estamos interesados.

Supondremos también en lo que sigue $(X'X)$ de rango completo, aunque quizá con acusada multicolinealidad[4]. Emplearemos los subíndices $N+1$ y N para designar estimaciones respectivamente con y sin esta observación

[3]O suprimiéndolos. Los métodos de regresión sesgada del Capítulo 11 hacen explícita esta idea.

[4]Los resultados se pueden generalizar al caso en que $(X'X)$ es de rango deficiente, y sólo mediante la nueva fila $\vec{x}_{N+1}\,'$ se hace $\vec{c}\,'\vec{\beta}$ estimable.

adicional. Tenemos entonces que:

$$\begin{aligned}
\Sigma_{\hat{\beta}_N} &= \sigma^2 (X'X)^{-1} && (9.31)\\
\Sigma_{\hat{\beta}_{N+1}} &= \sigma^2 (X'X + \vec{x}_{N+1}\vec{x}_{N+1}')^{-1} && (9.32)\\
\sigma^2_{\vec{c}\,'\hat{\beta}_N} &= \sigma^2 \vec{c}\,'(X'X)^{-1}\vec{c} && (9.33)\\
\sigma^2_{\vec{c}\,'\hat{\beta}_{N+1}} &= \sigma^2 \vec{c}\,'(X'X + \vec{x}_{N+1}\vec{x}_{N+1}')^{-1}\vec{c} && (9.34)
\end{aligned}$$

Entonces,

$$\sigma^2_{\vec{c}\,'\hat{\beta}_N} - \sigma^2_{\vec{c}\,'\hat{\beta}_{N+1}} = \sigma^2 \vec{c}\,'[(X'X)^{-1} - (X'X + \vec{x}_{N+1}\vec{x}_{N+1}')^{-1}]\vec{c} \qquad (9.35)$$

y el problema es encontrar $\vec{x}_{N+1}$ maximizando esta expresión. Sea V la matriz que diagonaliza a $(X'X)$. Denominemos:

$$\begin{aligned}
\vec{a} &= V'\vec{c} && (9.36)\\
\vec{z} &= V'\vec{x}_{N+1} && (9.37)\\
D &= V'(X'X)V && (9.38)
\end{aligned}$$

Entonces, (10.35) puede transformarse así:

$$\begin{aligned}
\sigma^2_{\vec{c}\,'\hat{\beta}_N} - \sigma^2_{\vec{c}\,'\hat{\beta}_{N+1}} &= \sigma^2 \vec{c}\,'VV'[(X'X)^{-1} - (X'X + \vec{x}_{N+1}\vec{x}_{N+1}')^{-1}]VV'\vec{c}\\
&= \sigma^2 \vec{a}\,'[D^{-1} - V'(X'X + \vec{x}_{N+1}\vec{x}_{N+1}')^{-1}V]\vec{a}\\
&= \sigma^2 \vec{a}\,'[D^{-1} - (V'(X'X + \vec{x}_{N+1}\vec{x}_{N+1}')V)^{-1}]\vec{a}\\
&= \sigma^2 \vec{a}\,'[D^{-1} - (D + \vec{z}\,\vec{z}\,')^{-1}]\vec{a} && (9.39)
\end{aligned}$$

Pero (véase Teorema A.2, pág. 230):

$$(D + \vec{z}\,\vec{z}\,')^{-1} = D^{-1} - \frac{D^{-1}\vec{z}\,\vec{z}\,'D^{-1}}{1 + \vec{z}\,'D^{-1}\vec{z}} \qquad (9.40)$$

Sustituyendo (10.40) en (10.39):

$$\begin{aligned}
\sigma^2_{\vec{c}\,'\hat{\beta}_N} - \sigma^2_{\vec{c}\,'\hat{\beta}_{N+1}} &= \sigma^2 \vec{a}\,'\left[\frac{D^{-1}\vec{z}\,\vec{z}\,'D^{-1}}{1 + \vec{z}\,'D^{-1}\vec{z}}\right]\vec{a} && (9.41)\\
&= \sigma^2 \frac{\left(\sum_i \frac{a_i z_i}{\lambda_i}\right)^2}{\left(1 + \sum_i \frac{z_i^2}{\lambda_i}\right)} && (9.42)
\end{aligned}$$

Obsérvese que el problema de maximizar (10.35) carece de sentido si no imponemos restricciones, pues la expresión equivalente (10.42) es monótona creciente al multiplicar $\vec{z}$ por una constante k mayor que la unidad[5]. Necesitamos una restricción del tipo $\vec{z}\,'\vec{z} = \sum_i z_i^2 = K^2$ para obtener una solución única. Formando entonces el lagrangiano,

$$\Phi(\vec{z}\,) = \sigma^2 \frac{\left(\sum_i \frac{a_i z_i}{\lambda_i}\right)^2}{\left(1 + \sum_i \frac{z_i^2}{\lambda_i}\right)} - \mu\left(\sum_i z_i^2 - K^2\right) \tag{9.43}$$

y derivando respecto a z_i, $(i = 1, \ldots, p)$, obtenemos p igualdades de la forma:

$$\sigma^2 \frac{\left(\sum_i \frac{a_i z_i}{\lambda_i}\right)\frac{a_i}{\lambda_i}\left(1 + \sum_i \frac{z_i^2}{\lambda_i}\right) - \left(\sum_i \frac{a_i z_i}{\lambda_i}\right)^2 \frac{z_i}{\lambda_i}}{\left(1 + \sum_i \frac{z_i^2}{\lambda_i}\right)^2} - \mu z_i = 0 \tag{9.44}$$

Denominando:

$$A = \left(\sum_i \frac{a_i z_i}{\lambda_i}\right) \tag{9.45}$$

$$B = \left(1 + \sum_i \frac{z_i^2}{\lambda_i}\right) \tag{9.46}$$

las p igualdades anteriores toman la forma:

$$\frac{a_i}{\lambda_i}\frac{A}{B} - \frac{z_i}{\lambda_i}\frac{A^2}{B^2} - \frac{\mu z_i}{\sigma^2} = 0 \tag{9.47}$$

Multiplicando por z_i cada una de las anteriores igualdades y sumándolas, puede despejarse:

$$\mu = \frac{A^2}{K^2 B^2}\sigma^2 \tag{9.48}$$

y por consiguiente de (10.47) se obtiene:

$$\frac{a_i}{\lambda_i}\frac{A}{B} - \frac{z_i}{\lambda_i}\frac{A^2}{B^2} - \frac{A^2}{K^2 B^2} z_i = 0 \qquad (i = 1, \ldots, p) \tag{9.49}$$

$$z_i\left(\frac{1}{\lambda_i} + \frac{1}{K^2}\right) = \frac{B}{A}\frac{a_i}{\lambda_i} \qquad (i = 1, \ldots, p) \tag{9.50}$$

[5]Observemos que al multiplicar $\vec{z}$ por k el numerador queda multiplicado por k^2, en tanto sólo una parte del denominador lo hace. Es pues claro que el numerador crece más que el denominador, y el cociente en consecuencia aumenta.

o sea:

$$z_i \propto \frac{a_i}{\lambda_i \left(\frac{1}{\lambda_i} + \frac{1}{K^2} \right)} = \frac{a_i}{1 + \frac{\lambda_i}{K^2}} \tag{9.51}$$

para $i = 1, \ldots, p$. Las anteriores p igualdades pueden expresarse en notación matricial así:

$$\vec{z} \propto (I + K^{-2}D)^{-1}\vec{a} \tag{9.52}$$

Por tanto, la fila a añadir a X para mejorar al máximo la estimación de $\vec{c}\,'\vec{\beta}$ será:

$$\begin{aligned}
\vec{x}_{N+1} &= V\vec{z} \\
\text{(por (10.52))} \quad &\propto V(I + K^{-2}D)^{-1}\vec{a} \\
&= V(I + K^{-2}D)^{-1}V'V\vec{a} \\
\text{(por (10.36))} \quad &= V(I + K^{-2}D)^{-1}V'\vec{c} \\
&= [V(I + K^{-2}D)V']^{-1}\vec{c} \\
&= [I + K^{-2}(X'X)]^{-1}\vec{c}
\end{aligned}$$

Recordemos que hemos obtenido una solución única para $\vec{z}$ (y en consecuencia $\vec{x}_{N+1}$) sólo mediante la imposición de una restricción de escala $\sum_i z_i^2 = K^2$. Es decir, podemos determinar la dirección de $\vec{z}$, pero no su norma. El examen de (10.42) hace evidente que una norma tan grande como sea posible es lo deseable.

Cabe hacer dos comentarios sobre esta última afirmación. El primero, que es lógico que así sea. Si σ^2 es fija, es claro que siempre preferiremos filas de módulo muy grande, pues si:

$$Y_i = m_i + \epsilon_i = \beta_0 + \cdots + \beta_{p-1}x_{i,p-1} + \epsilon_i \tag{9.53}$$

incrementar el módulo de $\vec{x}_{N+1}$ equivale a incrementar $|m_i|$; y haciendo $|m_i| \gg \epsilon_i$ podemos reducir en términos relativos el peso de ϵ_i en y_i.

En la práctica, sin embargo, hay un límite al valor de $|m_i|$, cuyo crecimiento desaforado podría llevarnos a regiones en las que las Y_i dejan de ser una función aproximadamente lineal de los regresores. Por ejemplo, si el modelo intenta ajustar una constante biológica como función lineal de ciertos tipos de nutrientes, hay un límite práctico a los valores que pueden tomar los regresores: el impuesto por las cantidades que los sujetos bajo estudio pueden ingerir.

En definitiva, el desarrollo anterior suministra la *dirección* en que debe tomarse una observación adicional para mejorar al máximo la varianza en la estimación de $\vec{c}\,'\vec{\beta}$. Tomaremos $\vec{x}_{N+1}$ tan grande como sea posible en dicha dirección. Si no tuviéramos una forma estimable única como objetivo, una estrategia sensata consistiría en tomar observaciones de forma que se incrementasen los menores valores propios de la matriz $(X'X)$. Podríamos también aceptar como criterio el de maximizar el determinante de $(X'X)$. Este criterio se conoce como de D-optimalidad[6].

[6]Véase Silvey (1980), una monografía que trata el tema de diseño óptimo.