



Centro de Investigación en Matemáticas, A.C.

CIMAT

Análisis de Sensibilidad en Regresión: Funciones
lineales de parámetros

T E S I S

Que para obtener el grado de:

Maestro en Ciencias

Con especialidad en:

Probabilidad y Estadística

P r e s e n t a:

Víctor Manuel Alvarado Castro

Guanajuato, Gto. México. Abril de 2004

Centro de Investigación en Matemáticas A. C.

Análisis de Sensibilidad en Regresión: Funciones
lineales de parámetros

T E S I S

Que para obtener el grado de
MAESTRO EN CIENCIAS

con especialidad en
Probabilidad y Estadística

P R E S E N T A

Víctor Manuel Alvarado Castro

DIRECTORES DE TESIS

Dr. José Antonio Díaz García
Dra. Graciela González Farías

Abril de 2004, Guanajuato, Gto., México

Agradecimientos

A mis asesores, Dr. José Antonio Díaz García y Dra. Graciela González Farías, por su disposición, paciencia y su apoyo para llevar a cabo la culminación de esta tesis.

Al Dr. Rogelio Ramos Quiroga, por sus enseñanzas, su disponibilidad y sus consejos.

A todos mis profesores del CIMAT, porque gracias a ustedes he conocido un poco más sobre la probabilidad y la estadística.

A mis amigos y compañeros de maestría, en particular a José de Jesús Ruiz Gallegos. Gracias por los momentos que me han permitido compartir con ustedes.

Al CONACYT, por apoyarme con la beca número 170565.

A mis padres, Nicasio Alvarado Alcocer y Obdulia Castro Morales, por todos sus sacrificios. Nunca terminaré de agradecerles.

A mis hermanos Hugo Alberto, Oscar Daniel y Diego Armando. Gracias por su apoyo en todo momento.

A la Señora Alicia Noemí García Beltrán y a todas sus hijas, por sus oraciones y sus palabras de ánimo.

A mi esposa Graciela. Gracias por todo el apoyo que me has dado. Gracias por la comprensión y paciencia que me has tenido.

A Graciela

Contenido

Agradecimientos	i
Resumen	vi
Símbolos y Notación	vii
1 Introducción	1
2 Análisis de Sensibilidad	5
2.1 Introducción	5
2.2 La Matriz de Predicción	6
2.2.1 Roles de P y $(I - P)$ en la Regresión Lineal	6
2.2.2 Propiedades de la Matriz de Predicción	8
2.3 Eliminación de Observaciones	12
2.3.1 Tipos de Residuales	13
2.3.2 Outliers, High-Leverages y Observaciones Influyentes	17
2.3.3 Medidas de Influencia	18
3 Una Generalización a la Distancia de Cook	22
3.1 Introducción	22
3.2 Detectando una Observación Influyente	23
3.3 Detectando un Conjunto de Observaciones Influyentes	29
3.4 Distribuciones Asociadas con las Distancias Modificadas	33

4 Ejemplos y Aplicaciones	40
4.1 Introducción	40
4.2 Ejemplo uno: mortalidad de truchas	40
4.3 Ejemplo dos: consumo de fósforo en ratas	46
4.4 Ejemplo tres: estudio dental realizado en niños y niñas	51
5 Conclusiones	57

Apéndices

A Elementos del Álgebra Lineal	59
A.1 Vectores Unitarios y Matrices Elementales	59
A.2 Traza	60
A.3 Matriz Inversa	60
A.4 Partición de Matrices	60
A.5 Inversa Generalizada	61
A.6 Producto Kronecker	62
A.7 Vectorización	63
B Estadística Multivariada	64
B.1 Distribución Normal Multivariada	64
B.2 Distribución Wishart	64
B.3 Distribución Wishart no Centrada	65
B.4 Distribución T^2 de Hotelling	66
B.5 Distribución Beta Multivariada	66
B.6 Estadísticos Traza	67
B.6.1 Estadístico Lawley-Hotelling	67
B.6.2 Estadístico Pillai	68
B.6.3 Estadístico de Roy	68
B.7 Resultados de Ellenberg	68

<i>CONTENIDO</i>	v
B.8 Distribución de Contornos Elípticos	69
C Programa	70
Bibliografía	76

Resumen

El interés por realizar estimaciones de funciones lineales del vector o de la matriz de parámetros β , surge de forma natural en modelos de curvas de crecimiento y en modelos de diseños de experimentos. Por tal razón, identificar observaciones influyentes en funciones lineales estimables de parámetros es importante para obtener mejores estimaciones.

En este trabajo se proponen algunas modificaciones a la distancia de Cook, mediante una generalización de la distancia de Mahalanobis, para detectar observaciones influyentes en funciones lineales estimables de los parámetros β , de la forma $N\beta M$, en el contexto del modelo de regresión lineal multivariado. Además, se establecen las distribuciones exactas de estadísticos pivotaes basados en estas nuevas distancias, las cuales proporcionan puntos críticos para la identificación de observaciones influyentes en funciones lineales de los parámetros; todo esto, bajo el supuesto de que los errores siguen una distribución normal y se extienden para el caso en que los errores siguen una distribución de contornos elípticos. Se ilustra el uso de las pruebas exactas bajo dos escenarios: regresión lineal multivariada múltiple y análisis de varianza multivariada (MANOVA).

Símbolos y Notación

\mathbb{R}^p	: espacio real p -dimensional.
$A \in \mathbb{R}^{p \times q}$: matriz real de dimension $p \times q$.
A, B, C, \dots, X, Y, Z	: matrices.
a_{ij}	: el ij -ésimo elemento de la matriz A .
A'	: la transpuesta de A .
I_m	: matriz identidad de orden m .
e_i^n	: vector unitario de orden n .
E_{ij}	: matriz elemental.
A_j	: j -ésima columna de A .
A'_i	: i -ésima fila de la matriz A .
$\det(A)$: determinante de A .
A^{-1}	: inversa de A .
A^-	: inversa de Moor-Penrose.
$r(A)$: rango de A .
$\text{tr}(A)$: traza de A .
$\text{vec}(A)$: vectorización de A .
$A \otimes B$: producto Kronecker de las matrices A y B .
$\text{Var}(X)$: varianza de X .
$\text{Cov}(X)$: covarianza de X .
$\text{Cor}(X_i, X_j)$: correlación entre X_i y X_j .
$\beta(\cdot, \cdot)$: distribución beta.

$t(\cdot, \cdot)$: distribución t.
$\mathcal{F}(\cdot, \cdot)$: distribución F.
gl.	: grados de libertad.
$\mathcal{N}_{n \times p}(\cdot, \cdot)$: distribución Normal Multivariada.
$\mathcal{W}_p(\cdot, \cdot)$: distribución Wishart.
$\mathcal{T}^2(\cdot, \cdot)$: distribución T^2 de Hotelling.
$\beta_m(\cdot, \cdot)$: distribución beta multivariada.
$V^{(\cdot)}$: estadístico Pillai.
$U^{(\cdot)}$: estadístico Lawley-Hotelling.
$\mathcal{P}(\cdot, \cdot, \cdot)$: distribución centrada para el estadístico Pillai.
$\mathcal{LH}(\cdot, \cdot, \cdot)$: distribución centrada para el estadístico Lawley-Hotelling.
$\mathcal{E}_{k \times p}(\cdot, \cdot, \cdot)$: distribución de contornos elípticos.

Capítulo 1

Introducción

Algunos de los objetivos de la ciencia son encontrar, describir y predecir relaciones entre variables de un conjunto de individuos. Una forma de que esto se cumpla es encontrar una fórmula o ecuación que relacione estas cantidades. Un modelo ampliamente usado, debido a su simplicidad, es el modelo de regresión lineal. Este modelo se forma para investigar la relación lineal entre dos conjuntos de variables, las variables respuesta Y y las variables independientes X . La ecuación del modelo de regresión lineal general es

$$Y = X\beta + \varepsilon, \tag{1.1}$$

donde β es el vector o la matriz de parámetros desconocidos y ε es el vector o la matriz de errores.

En términos de la cantidad de variables que tienen los dos conjuntos, el modelo de regresión lineal se clasifica en univariado y multivariado. Así, la regresión lineal univariada investiga la relación lineal entre una variable respuesta y una o más variables independientes. Si se investiga la relación lineal entre un conjunto de características, medidas en un mismo individuo, (variables respuesta) y un conjunto de variables independientes, entonces tal procedimiento es llamado regresión lineal multivariada.

Los elementos que determinan el modelo (1.1) son las observaciones, las variables y los supuestos

del modelo. Así, los resultados del ajuste del modelo pueden ser afectados por alguno de estos elementos. El estudio de las formas en las cuales los resultados del ajuste del modelo (1.1) son determinados por las observaciones, las variables y los supuestos del modelo es llamado *análisis de sensibilidad*.

Es común que, en ciertos conjuntos de datos, una observación o conjunto de observaciones determinan casi completamente los resultados del ajuste del modelo (1.1). Tales observaciones son llamadas *observaciones influyentes*. Las fuentes que originan observaciones influyentes son diversas. Algunas fuentes son: errores de captura en las bases de datos, errores de medición en las variables y que se tenga observaciones aberrantes (*outliers*) genuinas.

Con la finalidad de mejorar los resultados del ajuste del modelo de regresión, es importante contar con pruebas estadísticas que identifiquen a las observaciones influyentes. Autores como Cook (1977), Besley et al (1980), Cook y Weisberg(1982), Atkinson (1981) y Chatterjee y Hadi (1988), sólo por mencionar algunos, han estudiado el problema de identificar outliers u observaciones influyentes en la estimación del vector de parámetros β , en el caso de la regresión lineal univariada o multivariada y bajo el supuesto de que los errores tienen distribución normal. La idea original, en todos estos trabajos, es la distancia obtenida por Cook (1977) y llamada distancia de Cook. Es importante señalar que la distancia de Cook no es una prueba exacta, para detectar outliers, si no que se basa en una aproximación a la distribución \mathcal{F} centrada. Recientemente Díaz-García y González-Farías (2004) propusieron una modificación a esta distancia bajo el contexto de distribuciones elípticas, encontrando la distribución exacta de las nuevas distancias.

En algunos casos, es de interés hacer inferencia en funciones lineales del vector o de la matriz de parámetros β , de la forma $N\beta M$. Este problema surge de manera natural en modelos de curvas de crecimiento donde el modelo es de la forma

$$Y = A\beta B + E,$$

donde A , B son matrices constantes y E es una matriz aleatoria. También, en el modelo de diseños de experimentos interesa hacer inferencia sobre tales combinaciones lineales, ya que β

no es estimable para este modelo. Por tal razón, también es de interés identificar observaciones influyentes en las combinaciones lineales $N\beta M$. Este problema fue propuesto por Caroni (1987) y trabajado parcialmente por Díaz-García y González-Farías (2004).

El propósito de este trabajo es proponer algunas modificaciones a la distancia de Cook, mediante una generalización de la distancia de Mahalanobis, para detectar observaciones influyentes en funciones lineales de los parámetros β , de la forma $N\beta M$. Además, derivar las distribuciones exactas para las nuevas distancias proponiendo un punto crítico para decidir si una observación en particular (o un conjunto de observaciones) se comporta como un outlier (o como un conjunto de outliers).

El Capítulo 2 es dedicado al análisis de sensibilidad en regresión. En él se considera el modelo de regresión lineal univariado y se destaca la importancia que tiene la matriz de predicción en el análisis de regresión y en particular, en el análisis de sensibilidad. También se presenta el enfoque de eliminación de observaciones como un medio para detectar observaciones influyentes. En este capítulo se definen las principales medidas para detección de outliers, tales como, la distancia de Cook, la distancia modificada de Cook, la distancia de Welsh-Kuh, etc.

En el Capítulo 3 se deducen las modificaciones a la distancia de Cook, las cuales ayudarán a determinar si una observación o un conjunto de observaciones es influyente en funciones lineales estimables de los parámetros β , de la forma $N\beta M$. Todo esto en el contexto del modelo de regresión lineal multivariado y bajo el supuesto de errores normales con matriz de varianzas y covarianzas definida positiva. Además, se derivan las distribuciones exactas para las nuevas distancias proponiendo pruebas exactas para determinar si una observación o conjunto de observaciones es influyente. Finalmente, las distribuciones de las nuevas distancias se generalizan para el caso en que los errores siguen una distribución de contornos elípticos.

En el Capítulo 4 se presentan tres ejemplos, en ellos se ilustra el procedimiento a seguir, para poder concluir si una observación o un conjunto de observaciones es influyente en funciones lineales estimables de los parámetros β , de la forma $N\beta M$. También se intenta dar una

interpretación de las funciones lineales estimables de los parámetros, en las cuales se está interesado en medir la influencia de una observación o un conjunto de observaciones.

Las conclusiones que se obtuvieron, en este trabajo, son presentadas en el Capítulo 5. Los Apéndices A y B, muestran los resultados y aspectos sobre álgebra lineal y estadística multivariada importantes para comprender lo que se presenta y realiza en los Capítulos 2 y 3. El programa, realizado en S-plus, para hacer los análisis correspondientes a los ejemplos del Capítulo 4 es presentado en el Apéndice C.

Capítulo 2

Análisis de Sensibilidad

2.1 Introducción

Los elementos que determinan una ecuación de regresión son las observaciones, las variables, y los supuestos del modelo. El análisis de sensibilidad, en regresión lineal, es el estudio de las formas en las cuales los resultados del ajuste del modelo son determinados por estos elementos. Este capítulo se centra en el estudio del efecto que produce una observación, o un conjunto de observaciones, en la estimación de los parámetros del modelo, y particularmente en los métodos para detectar observaciones influyentes en estas estimaciones.

En este capítulo nos referiremos al modelo lineal general

$$Y = X\beta + \varepsilon, \tag{2.1}$$

donde $Y \in \mathbb{R}^n$ es el vector de variables respuesta o dependientes; $X \in \mathbb{R}^{n \times q}$, con $r(X) = q$, $\beta \in \mathbb{R}^q$, son la matriz de regresión y el vector de parámetros, respectivamente; y $\varepsilon \in \mathbb{R}^n$ es el vector de errores. Se asumirá que $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$.

2.2 La Matriz de Predicción

Considerando el modelo (2.1), con las suposiciones usuales, se puede ver que la matriz $P = X(X'X)^{-1}X'$, juega un papel importante en el análisis de regresión lineal. Esta matriz es llamada *matriz de predicción*, porque es la matriz de transformación que, aplicada a Y , produce los valores de predicción. Algunos autores la llaman *matriz hat*, porque mapea a Y en \hat{Y} .

Las interpretaciones de los elementos de P y las relaciones que tienen con los residuales son de suma importancia en la detección de observaciones influyentes. En las siguientes secciones se deducen la interpretación que tienen los elementos de P y algunas de sus propiedades. Se incluyen estos resultados y sus demostraciones, debido a la importancia de los mismos al derivar las distancias para detectar observaciones influyentes en funciones lineales de los parámetros β .

2.2.1 Roles de P y $(I - P)$ en la Regresión Lineal

Sea \mathbb{R}^n un espacio euclidiano de dimensión n y \mathbb{R}^q el subespacio de dimensión q generado por las columnas de X . La matriz de predicción P es simétrica e idempotente, i.e. $P = P'$ y $P = PP$. Así, P proyecta ortogonalmente a Y (o a cualquier $v \in \mathbb{R}^n$) en \mathbb{R}^q . Por esta razón es algunas veces llamada la matriz de proyección o el proyector ortogonal en \mathbb{R}^q (la imagen de P). La matriz de residuales $(I - P)$ también es simétrica e idempotente. Esta es la proyección ortogonal en la imagen de $(I - P)$, el subespacio de dimensión $(n - q)$ ortogonal a \mathbb{R}^q .

Sea p_{ij} el ij -ésimo elemento de P . Entonces el ii -ésimo elemento de P es

$$p_{ii} = X_i'(X'X)^{-1}X_i, \quad i = 1, 2, \dots, n. \quad (2.2)$$

Como $\hat{Y} = PY$ y $\hat{\varepsilon} = (I - P)Y$, los vectores de valores predichos y residuales, entonces $\text{Var}(\hat{y}_i) = \sigma^2 p_{ii}$ y $\text{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - p_{ii})$; Así, aparte de un factor constante σ^2 , p_{ii} determina a $\text{Var}(\hat{y}_i)$ y $\text{Var}(\hat{\varepsilon}_i)$. El ij -ésimo elemento de P es

$$p_{ij} = X_i'(X'X)^{-1}X_j, \quad i, j = 1, 2, \dots, n. \quad (2.3)$$

De $\text{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - p_{ii})$ se sigue que la covarianza entre $\hat{\varepsilon}_i$ y $\hat{\varepsilon}_j$ es

$$\text{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = -\sigma^2 p_{ij},$$

y el coeficiente de correlación entre $\hat{\varepsilon}_i$ y $\hat{\varepsilon}_j$ es

$$\text{Cor}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = \frac{-p_{ij}}{\sqrt{1 - p_{ii}}\sqrt{1 - p_{jj}}}.$$

Así, la correlación entre $\hat{\varepsilon}_i$ y $\hat{\varepsilon}_j$ está completamente determinada por los elementos de P .

Dado que P proyecta cualquier $v \in \mathbb{R}^n$ en el subespacio \mathbb{R}^q , entonces P proyecta a X en sí misma, esto es $PX = X$. De esto se sigue que $(I - P)X = 0$. Así, el vector de residuales es expresado como

$$\hat{\varepsilon} = (I - P)Y = (I - P)(X\beta + \varepsilon) = (I - P)\varepsilon, \quad (2.4)$$

de donde se observa que la relación entre $\hat{\varepsilon}$ y ε depende únicamente de P . Además, como $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}Y = PY$, entonces

$$\hat{y}_i = \sum_{j=1}^n p_{ij}y_j = p_{ii}y_i + \sum_{j \neq i} p_{ij}y_j, \quad i = 1, 2, \dots, n, \quad (2.5)$$

de donde se sigue que

$$\frac{\partial \hat{y}_i}{\partial y_i} = p_{ii}, \quad i = 1, 2, \dots, n. \quad (2.6)$$

Por tanto, p_{ii} puede ser interpretado como la cantidad de influencia que tiene cada y_i en determinar \hat{y}_i . Similarmente, p_{ij} puede ser considerado como la cantidad de influencia que tiene cada y_j en determinar \hat{y}_i .

De otra forma, p_{ii} puede ser considerado como el efecto de las observaciones que determinan \hat{y}_i . Esto es, si $p_{ii} = 1$, \hat{y}_i está determinada únicamente por y_i (una observación); así un grado de libertad se ha utilizado para ajustar y_i . Por otro lado, si $p_{ii} = 0$, y_i no tiene ninguna influencia sobre \hat{y}_i , mientras que si $p_{ii} = 0.5$, \hat{y}_i es determinado por efecto de dos observaciones.

2.2.2 Propiedades de la Matriz de Predicción

En esta sección se presenta una discusión detallada de algunas de las propiedades de la matriz de predicción junto con sus demostraciones.

Propiedad 2.1. *P es invariante bajo transformaciones lineales no singulares, de la forma $X \rightarrow XE$, con $E \in \mathbb{R}^{q \times q}$ no singular.*

Demostración. Sea $E \in \mathbb{R}^{q \times q}$ no singular, y sean P_X y P_{XE} las matrices de predicción de X y XE , respectivamente. Entonces

$$\begin{aligned} P_{XE} &= XE[(XE)'XE]^{-1}(XE)' = XEE^{-1}(X'X)^{-1}E'^{-1}E'X' \\ &= X(X'X)^{-1}X' = P_X. \end{aligned}$$

■

De la Propiedad 2.1, se tiene que $P_{XE} = P_X$, y así

$$\widehat{\varepsilon}_{Y.XE} = (I - P_{XE})Y = (I - P_X)Y = \widehat{\varepsilon}_{Y.X},$$

donde $\widehat{\varepsilon}_{Y.X}$ denota al vector de residuales de la regresión de Y y X . Así, si el modelo (2.1) contiene un término constante, los residuales y sus varianzas estimadas son invariantes bajo transformaciones de escala y localización de X , mientras que si el modelo (2.1) no tiene una constante, éstos son invariantes únicamente bajo transformaciones de escala de X . Además, la propiedad 2.1 se sigue para cualquier reparametrización no singular del modelo (2.1). Esto es, si $\alpha = E'\beta$, entonces los modelos $Y = X\beta + \varepsilon$ y $Y = XE\alpha + \varepsilon$ son equivalentes en el sentido de producir el mismo \widehat{Y} .

Propiedad 2.2. *P y $(I - P)$ son matrices simétricas e idempotentes.*

Demostración. La demostración se sigue de la definición de P . ■

Propiedad 2.3. *Sea $X \in \mathbb{R}^{n \times q}$. Entonces*

a. $\text{tr}(P) = r(P) = q,$

$$b. \operatorname{tr}(I - P) = n - q,$$

$$c. \sum_{i=1}^n \sum_{j=1}^n p_{ij}^2 = q.$$

Demostración. De la Propiedad 2.2 se tiene que $P = PP$, por lo que $\operatorname{tr}(P) = r(P)$. Además $P = X(X'X)^{-1}X' = XX^{-}$, de donde $r(P) = r(X) = q$. Esto verifica (a).

Por otro lado, $\operatorname{tr}(I - P) = \operatorname{tr}(I) - \operatorname{tr}(P)$. Así de (a) se sigue que $\operatorname{tr}(I - P) = n - q$, verificándose (b).

Finalmente se verifica (c). De la Propiedad 2.2 se sigue que $P = PP = PP'$, de donde $p_{ii} = \sum_{j=1}^n p_{ij}^2$. Así, $\sum_{i=1}^n \sum_{j=1}^n p_{ij}^2 = \sum_{i=1}^n p_{ii} = \operatorname{tr}(P) = q$. ■

Propiedad 2.4. Sea $X = (X_1 X_2)$ donde $X_1 \in \mathbb{R}^{n \times r}$ de rango r y $X_2 \in \mathbb{R}^{n \times q-r}$ de rango $q - r$. Sea $P_1 = X_1(X_1'X_1)^{-1}X_1'$ la matriz de predicción de X_1 , y $W = (I - P_1)X_2$ es la proyección de X_2 en el complemento ortogonal de X_1 . Finalmente, sea $P_2 = W(W'W)^{-1}W'$ es la matriz de predicción de W . Entonces P puede expresarse como

$$P = X(X'X)^{-1}X' = X_1(X_1'X_1)^{-1}X_1' + (I - P_1)X_2[X_2'(I - P_1)X_2]^{-1}X_2'(I - P_1),$$

o

$$P = P_1 + P_2.$$

Demostración. Note que

$$\begin{aligned} P &= X(X'X)^{-1}X' = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \left[\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \\ &= \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix}. \end{aligned} \quad (2.7)$$

De la inversa de una matriz particionada (ver Apéndice A.4) se tiene que

$$(X'X)^{-1} = \begin{pmatrix} (X'_1X_1)^{-1} + (X'_1X_1)^{-1}X'_1X_2MX'_2X_1(X'_1X_1)^{-1} & -(X'_1X_1)^{-1}X'_1X_2M \\ -MX'_2X_1(X'_1X_1)^{-1} & M \end{pmatrix} \quad (2.8)$$

donde $M = (X'_2X_2 - X'_2X_1(X'_1X_1)^{-1}X'_1X_2)^{-1} = [X'_2(I - P_1)X_2]^{-1}$.

Sustituyendo (2.8) en (2.7), se tiene

$$\begin{aligned} P &= P_1 + (I - P_1)X_2MX'_2(I - P_1) \\ &= P_1 + (I - P_1)X_2[X'_2(I - P_1)X_2]^{-1}X'_2(I - P_1) \\ &= P_1 + P_2. \end{aligned} \quad \blacksquare$$

Propiedad 2.5. Para $i = 1, 2, \dots$, y $j = 1, 2, \dots, n$, se tiene

- a. $0 \leq p_{ii} \leq 1$ para todo i ,
- b. $-0.5 \leq p_{ij} \leq 0.5$ para todo $j \neq i$,
- c. Si X tiene una columna constante, entonces

1. $p_{ii} \geq n^{-1}$ para todo i ,
2. $P\mathbf{1} = \mathbf{1}$, donde $\mathbf{1}$ es un vector de unos de dimensión n .

Demostración. De la Propiedad 2.2, se tiene que

$$p_{ii} = \sum_{j=1}^n p_{ij}^2 = p_{ii}^2 + \sum_{j \neq i} p_{ij}^2,$$

Así $0 \leq p_{ii}^2 \leq p_{ii}$, de donde se sigue que $0 \leq p_{ii} \leq 1$ para todo i . Esto prueba (a).

$$\text{También } p_{ii} = p_{ii}^2 + p_{ij}^2 + \sum_{k \neq i, j} p_{ik}^2,$$

por lo que $p_{ij}^2 \leq p_{ii}(1 - p_{ii})$. Como $0 \leq p_{ii} \leq 1$, entonces se sigue que $p_{ij}^2 \leq (0.5)^2$, lo cual verifica la parte (b).

Ahora, si X tiene una columna constante, defina $\tilde{X} = (\mathbf{1} \ X)$ donde $\mathbf{1}$ es un vector de unos de dimensión n . De la Propiedad 2.4 se tiene

$$\begin{aligned} P_1 &= \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = n^{-1}\mathbf{1}\mathbf{1}' \\ W &= (I - P_1)X_2 = (I - n^{-1}\mathbf{1}\mathbf{1}')X_2 = \tilde{X} \\ P_2 &= \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'. \end{aligned}$$

Así la matriz P es

$$P = P_1 + P_2 = n^{-1}\mathbf{1}\mathbf{1}' + \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'.$$

Cada elemento diagonal de P_1 es n^{-1} y dado que P_2 es una matriz de predicción, sus elementos diagonales son no negativos, esto se sigue de (a). Por tanto $p_{ii} \geq n^{-1}$ para todo i . Dado que $\tilde{X}'\mathbf{1} = 0$, entonces $P_2\mathbf{1} = 0$ y así $P\mathbf{1} = P_1\mathbf{1} = \mathbf{1}$, y la parte (c) queda demostrada. ■

Propiedad 2.6. Para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, n$,

- a. Si $p_{ii} = 1$ ó 0 , entonces $p_{ij} = 0$.
- b. $(1 - p_{ii})(1 - p_{jj}) - p_{ij}^2 \geq 0$.
- c. $p_{ii}p_{jj} - p_{ij}^2 \geq 0$.
- d. $p_{ii} + \frac{\hat{\varepsilon}_i^2}{\hat{\varepsilon}'\hat{\varepsilon}} \leq 1$.

Demostración. Recuerde que $p_{ij}^2 \leq p_{ii}(1 - p_{ii})$, de donde $|p_{ij}| \leq \sqrt{p_{ii}(1 - p_{ii})}$. Ahora, si $p_{ii} = 1$ ó 0 , entonces $p_{ij} = 0$. Esto verifica (a).

Recuerde que $|\text{Cor}(\hat{\varepsilon}_i, \hat{\varepsilon}_j)| \leq 1$, entonces $0 \leq |\text{Cor}(\hat{\varepsilon}_i, \hat{\varepsilon}_j)|^2 \leq 1$. Por lo que $0 \leq \frac{p_{ij}^2}{(1 - p_{ii})(1 - p_{jj})} \leq 1$, de donde se sigue que $(1 - p_{ii})(1 - p_{jj}) - p_{ij}^2 \geq 0$, verificándose (b).

Para verificar (c), note que $|\text{Cor}(\hat{y}_i, \hat{y}_j)| \leq 1$. Entonces $0 \leq |\text{Cor}(\hat{y}_i, \hat{y}_j)|^2 \leq 1$, por lo que $0 \leq \frac{p_{ij}^2}{p_{ii}p_{jj}} \leq 1$. De esto se sigue que $p_{ii}p_{jj} - p_{ij}^2 \geq 0$.

Finalmente se verifica (d). Defina $Z = \begin{pmatrix} X & Y \end{pmatrix}$, $P_X = X(X'X)^{-1}X'$ y $P_Z = Z(Z'Z)^{-1}Z'$. De la Propiedad 2.4 se tiene

$$P_Z = P_X + \frac{(I - P_X)YY'(I - P_X)}{Y'(I - P)Y} = P_X + \frac{\hat{\varepsilon}\hat{\varepsilon}'}{\hat{\varepsilon}'\hat{\varepsilon}}. \quad (2.9)$$

Así, $(P_Z)_{ii} = p_{ii} + \frac{\hat{\varepsilon}_i^2}{\hat{\varepsilon}'\hat{\varepsilon}}$. Como P_Z es una matriz de predicción, entonces $p_{ii} + \frac{\hat{\varepsilon}_i^2}{\hat{\varepsilon}'\hat{\varepsilon}} \leq 1$. ■

Propiedad 2.7. Los eigenvalores de P y $(I - P)$ son 0 ó 1.

Demostración. El resultado sigue del hecho de que los eigenvalores de matrices idempotentes son 0 ó 1. ■

Propiedad 2.8. Hay $n - q$ eigenvalores de P iguales a 0, y los q restantes son 1. Similarmente, q eigenvalores de $(I - P)$ son 0 y $n - q$ son iguales a 1.

Demostración. Sean $\lambda_i, i = 1, 2, \dots, n$, los eigenvalores de P . Entonces, de la Propiedad 2.3(a) se tiene

$$\sum_{i=1}^n \lambda_i = \text{tr}(P) = q$$

pero por la Propiedad 2.7, $\lambda_i = \{0, 1\}$. Por tanto q eigenvalores son iguales a 1 y $n - q$ eigenvalores son iguales a 0. Argumentos similares se utilizan para $(I - P)$. ■

2.3 Eliminación de Observaciones

Los resultados del ajuste del modelo (2.1) dado un conjunto de datos, puede ser substancialmente alterado (influenciado) al eliminar o añadir una o varias observaciones. Usualmente, no todas las observaciones tienen igual influencia en el ajuste y en las conclusiones de los resultados de tal análisis. Observaciones, que individual o colectivamente, tienen una excesiva influencia en el ajuste de la ecuación de regresión, comparadas con otras observaciones, son llamadas *observaciones influyentes*. Es importante, para el análisis de los datos, identificar las observaciones influyentes y evaluar su efecto sobre varios aspectos del análisis.

Una gran cantidad de medidas estadísticas se han propuesto para el diagnóstico de observaciones influyentes. Aquí se describirán algunas de estas medidas y se examinarán sus interrelaciones. Los siguientes son tres enfoques para detectar observaciones influyentes:

- (a) eliminación de observaciones,
- (b) diferenciación, y
- (c) geométrico.

El enfoque de eliminación de observaciones examina el cambio de los resultados del análisis de regresión (estimación de los coeficientes, valores predichos, residuales, estructura de la covarianza estimada de los coeficientes estimados, etc.) al eliminar una o varias observaciones. El enfoque de diferenciación examina las derivadas (razón de cambio) de varios resultados de regresión con respecto a ciertos parámetros del

modelo. El énfasis del tercer enfoque, geométrico, es detectar observaciones influyentes en un sentido geométrico.

Este estudio fue realizado bajo el enfoque de eliminación de observaciones. En las siguientes secciones se expondrán los conceptos y resultados importantes para abordar este enfoque. Las referencias sugeridas para los otros dos enfoques son Besley et al (1981) y Chatterjee y Hadi (1988).

2.3.1 Tipos de Residuales

Los residuales juegan un papel importante en el diagnóstico de regresión; el análisis no está completo sin una examinación de los residuales. El análisis estándar de los resultados de regresión está basado en ciertas suposiciones. Para que el análisis sea válido, es necesario asegurarse que estas suposiciones se cumplan.

El problema que se presenta es que los ε_i no pueden ser observados ni pueden ser estimados. Sin embargo, el i -ésimo residual $\hat{\varepsilon}_i$ mide, en cierta forma, el error aleatorio ε_i . Como $\hat{\varepsilon} = (I - P)\varepsilon$, entonces para que $\hat{\varepsilon}$ pueda ser una sustitución razonable para ε , los elementos fuera de la diagonal de P deben ser suficientemente pequeños. Además, si los elementos de ε son independientes y tienen la misma varianza, $\hat{\varepsilon} = (I - P)\varepsilon$ indica que los residuales no son independientes (a menos que P sea diagonal) y no tienen la misma varianza (a menos que los elementos diagonales de P sean iguales). Por lo cual, los residuales pueden ser considerados como una sustitución razonable de los ε_i si los renglones de X son homogéneos (así los elementos de la diagonal de P son aproximadamente iguales) y los elementos fuera de la diagonal de P son suficientemente pequeños.

Por estas razones es preferible usar transformaciones de los residuales ordinarios para propósitos de diagnóstico. Así, en lugar de $\hat{\varepsilon}_i$, se puede usar

$$f(\hat{\varepsilon}_i, \sigma_i) = \frac{\hat{\varepsilon}_i}{\sigma_i}, \quad (2.10)$$

donde σ_i es la desviación estándar del i -ésimo residual. Cuatro casos de (2.10) son: los residuales normalizados, los residuales estandarizados, los residuales estudentizados internos y los residuales estudentizados externos.

El i -ésimo residual normalizado se obtiene al reemplazar σ_i en (2.10) por $(\hat{\varepsilon}'\hat{\varepsilon})^{1/2}$, por lo que se obtiene

$$a_i \equiv f(\hat{\varepsilon}_i, \sqrt{\hat{\varepsilon}'\hat{\varepsilon}}) = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\varepsilon}'\hat{\varepsilon}}}, \quad i=1,2,\dots,n. \quad (2.11)$$

El i -ésimo residual estandarizado, b_i , se obtiene al sustituir $\hat{\sigma} = \sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-q}}$ por σ_i en (2.10), de donde

$$b_i \equiv f(\hat{\varepsilon}_i, \hat{\sigma}) = \frac{\hat{\varepsilon}_i}{\hat{\sigma}}, \quad i = 1, 2, \dots, n. \quad (2.12)$$

Si tomamos $\sigma_i = \hat{\sigma}\sqrt{1-p_{ii}}$, se obtiene el i -ésimo residual estudentizado interno

$$r_i \equiv f(\hat{\varepsilon}_i, \hat{\sigma}\sqrt{1-p_{ii}}) = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-p_{ii}}}, \quad i = 1, 2, \dots, n. \quad (2.13)$$

El i -ésimo residual estudentizado externo se obtiene al tomar $\sigma_i = \hat{\sigma}_{(i)}\sqrt{1-p_{ii}}$, donde

$$\hat{\sigma}_{(i)} = \frac{Y'_{(i)}(I - P_{(i)})Y_{(i)}}{n - q - 1}, \quad i = 1, 2, \dots, n,$$

es el estimador del error cuadrático medio cuando la i -ésima observación es omitida y

$$P_{(i)} = X_{(i)} \left(X'_{(i)} X_{(i)} \right)^{-1} X'_{(i)}, \quad i = 1, 2, \dots, n,$$

es la matriz de predicción para $X_{(i)}$. Así, el i -ésimo residual estudentizado externo es definido como

$$r_i^* \equiv f(\hat{\varepsilon}_i, \hat{\sigma}_{(i)}\sqrt{1-p_{ii}}) = \frac{\hat{\varepsilon}_i}{\hat{\sigma}_{(i)}\sqrt{1-p_{ii}}}, \quad i = 1, 2, \dots, n. \quad (2.14)$$

Como se mencionó anteriormente, los residuales ordinarios no son apropiados para propósitos de diagnóstico y una transformación de ellos es preferible. Pero, ¿cuál de las cuatro formas, descritas anteriormente, es preferible?. Para responder a esta pregunta, se investigarán las propiedades de las transformaciones de los residuales y las relaciones entre ellos.

Las cuatro versiones de los residuales definidos en (2.11), (2.12), (2.13) y (2.14) están muy relacionadas. Éstas son funciones de cuatro cantidades, a saber, $(n-q)$, $\hat{\varepsilon}'\hat{\varepsilon}$, $\hat{\varepsilon}_i$, y $(1-p_{ii})$. Para encontrar las relaciones entre los cuatro tipos de residuales, necesitamos escribir $\hat{\sigma}_{(i)}$ en términos de $\hat{\sigma}$. Esto se hace notando que al eliminar la i -ésima observación es equivalente a ajustar el modelo *mean-shift outlier*, esto es,

$$E(Y) = X\beta + e_i^n \theta, \quad (2.15)$$

donde θ es el coeficiente de regresión del i -ésimo vector unitario e_i^n . Así se tiene que

$$\begin{aligned} SSE_{(i)} &= Y'_{(i)}(I - P_{(i)})Y_{(i)} \\ &= Y' \left(I - P - \frac{(I - P)e_i^n e_i^{n'}(I - P)}{e_i^{n'}(I - P)e_i^n} \right) Y \\ &= SSE - \frac{\hat{\varepsilon}_i^2}{1 - p_{ii}} \end{aligned} \quad (2.16)$$

Donde SSE es la suma de cuadrados del error y $SSE_{(i)}$ es la suma de cuadrados del error cuando la i -ésima observación es eliminada. Dividiendo ambos lados de (2.16) por $(n - q - 1)$, tenemos

$$\begin{aligned}\hat{\sigma}_{(i)}^2 &= \frac{n - q}{n - q - 1} \hat{\sigma}^2 - \frac{\hat{\varepsilon}_i^2}{(n - q - 1)(1 - p_{ii})} \\ &= \hat{\sigma}^2 \left(\frac{n - q - r_i^2}{n - q - 1} \right).\end{aligned}\quad (2.17)$$

Así

$$b_i = a_i \sqrt{n - q} \quad y \quad r_i = \frac{b_i}{\sqrt{1 - p_{ii}}} = a_i \sqrt{\frac{n - q}{1 - p_{ii}}}.\quad (2.18)$$

Ahora si en (2.14) se reemplaza $\hat{\sigma}_{(i)}^2$ por $(SSE_{(i)}/(n - q - 1))^{1/2}$ y usando (2.16), r_i^* se puede escribir como

$$\begin{aligned}r_1^* &= \frac{\hat{\varepsilon}_i}{\sqrt{\frac{(1 - p_{ii})SSE_{(i)}}{n - q - 1}}} = \frac{\hat{\varepsilon}_i}{\sqrt{\frac{1 - p_{ii}}{n - q - 1} \left(SSE - \frac{\hat{\varepsilon}_i^2}{1 - p_{ii}} \right)}} \\ &= \frac{a_i \sqrt{n - q - 1}}{\sqrt{(1 - p_{ii}) - a_i^2}}.\end{aligned}\quad (2.19)$$

También, al sustituir (2.16) en (2.14), se tiene

$$r_i^* = r_i \sqrt{\frac{n - q - 1}{n - q - r_i^2}}.\quad (2.20)$$

De (2.18), (2.19) y (2.20), se observa que r_i^* es una transformación monótona de r_i , y r_i es una transformación monótona de a_i . Además, como $a_i^2 \leq (1 - p_{ii})$ (Propiedad 2.6(d)), entonces $r_i^2 \rightarrow (n - q)$ y $r_i^{*2} \rightarrow \infty$ cuando $a_i^2 \rightarrow (1 - p_{ii})$. Las propiedades distribucionales de r_i y r_i^* son dadas en el siguiente resultado.

Teorema 2.9. *Considerando el modelo (2.1), modelo lineal general, y bajo el supuesto $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$, tenemos que:*

a.

$$r_i^* \sim t(n - q - 1),\quad (2.21)$$

donde $t(n - q - 1)$ denota una distribución t central con $(n - q - 1)$ grados de libertad (gl).

b.

$$\frac{r_i}{n - q} \sim \beta(1/2, (n - q - 1)/2),\quad (2.22)$$

donde $\beta(1/2, (n - q - 1)/2)$ denota una distribución beta central con parámetros $1/2$ y $(n - q - 1)/2$.

Demstración. Para probar (a) se usará una propiedad interesante de r_i^* , normalmente, r_i^* es equivalente al estadístico F para probar la significancia del coeficiente θ en el modelo *mean-shift outlier*. Sea $H_0 : E(Y) = X\beta$ y $H_1 : X\beta + e_i^n\theta$. Bajo la suposición de normalidad, el estadístico F para probar H_0 contra H_1 es

$$F_i = \frac{[SSE(H_0) - SSE(H_1)]/1}{SSE(H_1)/(n - q - 1)} \quad (2.23)$$

Pero dado que

$$SSE(H_1) = Y' \left(I - P - \frac{(I - P)e_i^n e_i^{n'}(I - P)}{e_i^{n'}(I - P)e_i^n} \right) Y = SSE(H_0) - \frac{\widehat{\varepsilon}_i^2}{1 - p_{ii}},$$

así, (2.23) se reduce a

$$F_i = \frac{\widehat{\varepsilon}_i^2}{\widehat{\sigma}_{(i)}^2(1 - p_{ii})} = r_i^{*2}. \quad (2.24)$$

Como $F_i \sim \mathcal{F}_{(1, n-q-1)}$, $r_i^* = (F_i)^{1/2}$ es distribuido $t_{(n-q-1)}$, por lo que la parte (a) queda demostrada.

Para demostrar (b) se utiliza la igualdad $r_i^{*2} = \frac{r_i^2(n - q - 1)}{n - q - r_i^2}$, por lo que $r_i^2 = \frac{(n - q)r_i^{*2}}{r_i^{*2} + (n - q - 1)}$.

Usando la relación entre las distribuciones F y Beta, donde $G = \frac{hZ}{hZ + g}$ con $Z \sim \mathcal{F}_{(h, g)}$, por lo que $G \sim \beta(h/2, g/2)$. Así, dado que $\frac{r_i^2}{n - q} = \frac{r_i^{*2}}{r_i^{*2} + (n - q - 1)}$, entonces de (a) se sigue que

$$\frac{r_i^2}{n - q} \sim \beta(p/2, (n - q - p)/2). \quad (2.25)$$

■

Para propósitos de diagnóstico, los residuales normalizados y los residuales estandarizados son básicamente equivalentes y proporcionales a $\widehat{\varepsilon}_i$, siendo simples de calcular, pero no reflejan la varianza de $\widehat{\varepsilon}_i$. Varios autores (entre ellos Belsley et al.(1980) y Atkinson (1981)), prefieren usar r_i^* en lugar de r_i por las siguientes razones: 1) r_i^* es interpretado como el estadístico t para probar la significancia del i -ésimo vector unitario e_i^n en el modelo *mean-shift outlier* (2.15); 2) r_i^* es una transformación monótona de r_i y, dado que $r_i^{*2} \rightarrow \infty$ cuando $r_i^2 \rightarrow (n - q)$, r_i^* refleja desviaciones grandes más dramáticas que r_i ; y 3) el estimador $\widehat{\sigma}_{(i)}$ es robusto para problemas de errores grandes en la i -ésima observación.

2.3.2 Outliers, High-Leverages y Observaciones Influyentes

Violaciones en las suposiciones del modelo de regresión pueden ocurrir en una gran cantidad de formas. Frecuentemente, los datos pueden contener outliers, i.e. observaciones anómalas que no son razonablemente ajustadas por el modelo asumido. En general, una observación, o un conjunto de observaciones, puede ser influyente en el ajuste del modelo porque es una observación aberrante (outlier), o un punto palanca (al cual lo llamaremos high-leverage), o ambos. En el contexto de regresión lineal se dan las siguientes definiciones:

Definición 2.10. *[Outlier] Un outlier puede ser una observación para la cual su residual estudentizado (r_i o r_i^*) es grande en magnitud comparado con los de las otras observaciones.*

Definición 2.11. *[High-leverage] Los puntos high-leverage son puntos para los cuales su vector de entrada X_i está, en algún sentido, lejos del resto de los datos. Equivalentemente, un punto high-leverage es una observación con un p_{ii} grande en comparación con los de las otras observaciones.*

Definición 2.12. *[Observaciones influyentes] Las observaciones influyentes son aquellas observaciones que, individual o colectivamente, influyen excesivamente en los resultados del ajuste de la ecuación de regresión comparado con las otras observaciones.*

Una observación, sin embargo, no puede tener la misma influencia en todos los resultados de la regresión. La pregunta ¿Influencia sobre qué? es, por tanto, de primera importancia. Por ejemplo, una observación puede tener influencia sobre β , la varianza estimada de $\hat{\beta}$, los valores predichos, etc. Así, el primer objetivo del análisis es dar respuesta a la pregunta para saber cual influencia considerar. Por ejemplo, si la estimación de β es lo más importante, entonces lo apropiado es medir la influencia de las observaciones sobre $\hat{\beta}$, mientras que si la predicción es el objetivo principal, entonces medir la influencia sobre los valores predichos puede ser más apropiado que medir la influencia sobre $\hat{\beta}$.

Si bien es cierto que ni los outliers, ni los high-leverages, son necesariamente observaciones influyentes, potencialmente estas observaciones pueden ser influyentes. Así una forma de detectar observaciones influyentes es detectar outliers o high-leverages. Una discusión sobre algunas medidas, para detectar outliers y high-leverages, de influencia es presentada en la siguiente sección.

2.3.3 Medidas de Influencia

La influencia de una observación, o un conjunto de observaciones, es medida por el efecto que produce en los resultados del ajuste del modelo cuando es eliminada del proceso de ajuste. Por ejemplo, sea $\widehat{\beta}_{(i)}$ el estimador de β calculado de la muestra sin la i -ésima observación. Entonces la diferencia entre $\widehat{\beta}$ y $\widehat{\beta}_{(i)}$ da el punto hasta el cual la i -ésima observación afecta a la estimación de los coeficientes de regresión. Hay un gran número de medidas de influencia en la literatura, y el lector es referido a uno de los siguientes libros para mayores detalles: Belsley et al.(1980), Cook y Weisberg (1982) y Chatterjee y Hadi (1988).

Detectando una Observación Influyente

Bajo normalidad, la región de confianza conjunta $100(1 - \alpha)\%$ para β es obtenida de

$$\frac{(\beta - \widehat{\beta})'(X'X)(\beta - \widehat{\beta})}{q\widehat{\sigma}^2} \leq F_{(\alpha; q, n-q)}, \quad (2.26)$$

donde $F_{(\alpha; q, n-q)}$ es el α percentil superior de la distribución \mathcal{F} centrada con q y $n - q$ gl. Esta desigualdad define una región elipsoidal centrada en $\widehat{\beta}$. Esto sugiere una forma de medir la influencia de la i -ésima observación eliminada, que da lugar a la llama distancia de Cook.

Definición 2.13. [*Distancia de Cook, Cook(1977)*] La distancia de Cook mide la diferencia entre los coeficientes de regresión obtenidos del conjunto de datos completo y los coeficientes de regresión obtenidos eliminando la i -ésima observación. Se denota por C_i y esta definida por

$$C_i = \frac{(\widehat{\beta} - \widehat{\beta}_{(i)})'(X'X)(\widehat{\beta} - \widehat{\beta}_{(i)})}{q\widehat{\sigma}^2}, \quad (2.27)$$

donde $\widehat{\sigma}^2 = \frac{n\widetilde{\sigma}^2}{n - q}$. C_i puede ser interpretada como la distancia escalada entre $\widehat{\beta}$ y $\widehat{\beta}_{(i)}$, o alternativamente, como la distancia escalada entre \widehat{Y} y $\widehat{Y}_{(i)}$.

Notemos que $(X'X)$ no es la matriz de varianzas y covarianzas de $(\widehat{\beta} - \widehat{\beta}_{(i)})$ razón por la cual, no fue posible establecer la distribución exacta del estadístico C_i .

De lo anterior se puede pensar que para calcular C_i , $i = 1, 2, \dots, n$, se requieren $(n + 1)$ regresiones, una regresión usando los datos completos y n regresiones usando los datos reducidos. Sin embargo

$$\begin{aligned} C_i &= \frac{(\hat{Y} - \hat{Y}_{(i)})'(\hat{Y} - \hat{Y}_{(i)})}{q\hat{\sigma}^2} = \frac{X_i'(X'X)^{-1}X_i}{q(1 - p_{ii})} \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2(1 - p_{ii})} \\ &= \frac{1}{q} \frac{p_{ii}}{1 - p_{ii}} r_i^2, \end{aligned} \quad (2.28)$$

donde r_i es el i -ésimo residual estudentizado dado en (2.13). Así, C_i es una función de dos cantidades relacionadas con el conjunto de datos completos: p_{ii} , la cual proporciona información acerca de los high-leverages, y r_i , la cual da información sobre outliers.

Claramente, C_i será grande si p_{ii} es grande, r_i es grande, o ambos. Como respuesta a la pregunta ¿Qué tan grande es grande? Cook (1977) sugiere, por analogía con (2.26), que cada C_i sea comparado con los percentiles de la distribución \mathcal{F} centrada con q y $n - q$ gl. Estrictamente C_i no tiene una distribución \mathcal{F} , por lo cual la regla anterior no debe usarse como una prueba estricta de significancia.

La distancia C_i , al utilizar a $\hat{\sigma}^2$ como estimador de σ^2 , sólo mide la influencia de la i -ésima observación en β . Es posible que se esté interesado en medir la influencia de la i -ésima observación en σ^2 , en tal caso es preferible utilizar como estimador de σ^2 a $\hat{\sigma}_{(i)}^2$, el cual es obtenido de la misma forma que $\hat{\sigma}^2$, pero con la muestra reducida en $(n - 1)$, i.e. eliminando la i -ésima observación. Así, reemplazando $\hat{\sigma}^2$ por $\hat{\sigma}_{(i)}^2$ en (2.27), tomando la raíz cuadrada de C_i y ajustando C_i para el tamaño de muestra, se obtiene la distancia modificada de Cook.

Definición 2.14. [Distancia modificada de Cook, Atkinson (1981)] La distancia modificada de Cook es definida por

$$MC_i = |r_i^*| \sqrt{\frac{p_{ii}}{1 - p_{ii}} \frac{n - q}{q}}, \quad (2.29)$$

donde r_i^* es el residual estudentizado externo definido en (2.14).

Otras dos medidas relacionadas a la región de confianza elipsoidal para β son definidas a continuación.

Definición 2.15. [Distancia de Welsch-Kuh] La distancia de Welsch-Kuh o $DFFITs_i$ mide la influencia de la i -ésima observación, sobre \hat{Y}_i por el cambio en la predicción en X_i cuando la i -ésima

observación es eliminada, relativo al error estandar de \hat{Y}_i , como

$$\begin{aligned} DFFITS_i = WK_i &= \frac{\left| \frac{\hat{\varepsilon}_i}{1 - p_{ii}} X'_i (X'X)^{-1} X_i \right|}{\hat{\sigma}_{(i)} \sqrt{p_{ii}}} \\ &= \frac{|r_i^*|}{\sqrt{\frac{p_{ii}}{1 - p_{ii}}}} \end{aligned} \quad (2.30)$$

Definición 2.16. [Distancia de Welsch] La distancia de Welsch es definida por

$$\begin{aligned} W_i &= \sqrt{(n-1) r_i^{*2} \frac{p_{ii}}{1 - p_{ii}}} \\ &= WK_i \sqrt{\frac{p_{ii}}{1 - p_{ii}}}, \end{aligned} \quad (2.31)$$

y claramente da más énfasis a los high-leverages que WK_i .

Note que la relación entre MC_i y WK_i es

$$MC_i = WK_i \sqrt{(n-q)/q}.$$

De nuevo, para dar respuesta a la pregunta ¿Qué tan grande es grande? se sugiere, basados en la distribución de los r_i^* , que cada WK_i sea comparado con $t_{\alpha/2; n-q-1} \sqrt{q/n-q}$, donde $t_{\alpha/2; n-q-1}$ es el $\alpha/2$ percentil de la distribución t centrada con $n-q-1$ gl.

Detectando Múltiples Observaciones Influyentes

En la sección 2.3.3 se discutieron métodos para detectar observaciones que individualmente pueden ser consideradas outliers, high-leverages, u observaciones influyentes. En esta sección, se generalizan algunos de estos métodos para el caso más general de múltiples observaciones. Una discusión detallada es presentada en Chatterjee y Hadi (1988).

El problema de detectar múltiples observaciones influyentes es importante, porque pueden existir situaciones en las cuales las observaciones son conjuntamente pero no individualmente influyentes, o viceversa. Esta situación es referida como un efecto de ocultamiento, porque la influencia de una observación es ocultada por la presencia de otra observación.

Si $I = \{i_1, i_2, \dots, i_k\}$ es un conjunto de tamaño k de $\{1, 2, \dots, n\}$, tal que $(n - k) \geq q$, entonces dos medidas para observar si las correspondientes observaciones de acuerdo a los subíndices I son conjuntamente influyentes son:

Definición 2.17. [Distancia generalizada de Cook] La distancia generalizada de Cook es definida por

$$\begin{aligned} C_I &= \frac{(\hat{\beta} - \hat{\beta}_{(I)})'(X'X)(\hat{\beta} - \hat{\beta}_{(I)})}{q\hat{\sigma}^2} \\ &= \frac{\hat{\varepsilon}'_I(I - P_I)^{-1}P_I(I - P_I)^{-1}\hat{\varepsilon}_I}{q\hat{\sigma}^2}, \end{aligned} \quad (2.32)$$

donde $\hat{\beta}_{(I)}$ es el estimador de β eliminando el conjunto de observaciones correspondientes a los subíndices I , $\hat{\varepsilon}_I$ y X_I son el vector de residuales y la matriz de regresión considerando únicamente las observaciones con subíndices en I , y $P_I = X'_I(X'X)^{-1}X_I$.

Un valor grande de C_I en (2.32) indica que las observaciones indexadas por I son conjuntamente influyentes sobre $\hat{\beta}$.

Definición 2.18. [Distancia generalizada de Welsch] La distancia generalizada de Welsch se define como

$$W_I^2 = \left(\frac{n - k}{\hat{\sigma}_{(I)}^2} \right) \hat{\varepsilon}'_I(I - P_I)^{-2}P_I(I - P_I)^{-1}\hat{\varepsilon}_I, \quad (2.33)$$

donde $\hat{\sigma}_{(I)}^2$ es el estimador mínimo cuadrado de σ^2 eliminando las observaciones indexadas por I .

En el siguiente capítulo se generalizará una de las métricas, la distancia de Cook, para detectar observaciones influyentes en funciones lineales de los parámetros β , considerando el modelo lineal general multivariado.

Capítulo 3

Una Generalización a la Distancia de Cook

3.1 Introducción

Considere el modelo de regresión lineal multivariado

$$Y = X\beta + \varepsilon, \quad (3.1)$$

donde $Y \in \mathbb{R}^{n \times p}$ es la matriz respuesta; $X \in \mathbb{R}^{n \times q}$, con $r(X) = q$, $\beta \in \mathbb{R}^{q \times p}$, son la matriz de regresión y la matriz de parámetros, respectivamente; y $\varepsilon \in \mathbb{R}^{n \times p}$ es la matriz de errores, tal que $\varepsilon \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I_n)$ con Σ definida positiva.

Equivalentemente, el modelo (3.1), puede expresarse como

$$\text{vec}(Y) = (I \otimes X) \text{vec}(\beta) + \text{vec}(\varepsilon), \quad (3.2)$$

con $E(\text{vec}(\varepsilon)) = 0$ y $\text{Cov}(\text{vec}(\varepsilon)) = \Sigma \otimes I$.

Los estimadores de verosimilitud máxima de β y Σ son

$$\hat{\beta} = (X'X)^{-1}X'Y = X^-Y \quad \text{y} \quad \hat{\Sigma} = \frac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})$$

donde $P = X(X'X)^{-1}X' = XX^-$ es la matriz de predicción y X^- es la inversa de Moore-Penrose. El ij -ésimo elemento de P es $p_{ij} = X_i'(X'X)^{-1}X_j$.

En este capítulo se proponen algunas modificaciones a la distancia de Cook, mediante una generalización de la distancia de Mahalanobis, para detectar observaciones influyentes en funciones lineales de los parámetros β , de la forma $N\beta M$. Las ideas principales son retomadas de Díaz-García y González-Farías (2004). Además, se derivarán las distribuciones exactas para las nuevas distancias proponiendo un punto crítico para decidir si una observación en particular (o un conjunto de observaciones) se comporta como un outlier (o como un conjunto de outliers).

3.2 Detectando una Observación Influyente

Considere el modelo de regresión lineal multivariado

$$Y_{(i)} = X_{(i)}\beta_{(i)} + \varepsilon_{(i)}, \quad \varepsilon_{(i)} \sim \mathcal{N}_{n-1 \times p} (0, \Sigma_{(i)} \otimes I_n), \quad (3.3)$$

este modelo se obtiene de (3.1) eliminando el i -ésimo renglón de Y , X , y ε . Así, $Y_{(i)} \in \mathbb{R}^{n-1 \times p}$, $X_{(i)} \in \mathbb{R}^{n-1 \times q}$, $\beta_{(i)} \in \mathbb{R}^{q \times p}$ y $\varepsilon_{(i)} \in \mathbb{R}^{n-1 \times p}$.

Para el modelo modificado se sabe que:

$$\widehat{\beta}_{(i)} = (X'_{(i)} X_{(i)})^{-1} X'_{(i)} Y_{(i)} = X_{(i)}^- Y_{(i)} \quad \text{y} \quad \widehat{\Sigma}_{(i)} = \frac{1}{n} (Y_{(i)} - X_{(i)} \widehat{\beta}_{(i)})' (Y_{(i)} - X_{(i)} \widehat{\beta}_{(i)}),$$

donde $\widehat{\beta}_{(i)}$ y $\widehat{\Sigma}_{(i)}$ son los estimadores de verosimilitud máxima de $\beta_{(i)}$ y $\Sigma_{(i)}$, respectivamente.

Observe que

$$Y = \begin{pmatrix} Y'_1 \\ Y'_2 \\ \vdots \\ Y'_n \end{pmatrix}, Y_i \in \mathbb{R}^p \quad X = \begin{pmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_n \end{pmatrix}, X_i \in \mathbb{R}^q \quad \varepsilon = \begin{pmatrix} \varepsilon'_1 \\ \varepsilon'_2 \\ \vdots \\ \varepsilon'_n \end{pmatrix}, \varepsilon_i \in \mathbb{R}^p.$$

Por tanto

$$X'X = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix} \begin{pmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_n \end{pmatrix} = \sum_{k=1}^n X_k X'_k = X_i X'_i + \sum_{k \neq i}^n X_k X'_k = X_i X'_i + X'_{(i)} X_{(i)},$$

y

$$X'Y = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix} \begin{pmatrix} Y'_1 \\ Y'_2 \\ \vdots \\ Y'_n \end{pmatrix} = \sum_{k=1}^n X_k Y'_k = X_i Y'_i + \sum_{k \neq i} X_k Y'_k = X_i Y'_i + X'_{(i)} Y_{(i)}.$$

Note que si e_i^n es el i -ésimo vector de la base canónica en \mathbb{R}^n , esto es, el vector unitario, $e_i^n = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}'$, entonces $e_i^{n'} Y = Y'_i$, $e_i^{n'} X = X'_i$ y $e_i^{n'} \varepsilon = \varepsilon_i$.

Ahora, utilizando el Teorema de Sherman-Morrison-Woodbury (vea Apéndice A, Teorema A.1) se tiene que

$$(X'X - X_i X'_i)^{-1} = (X'_{(i)} X'_{(i)})^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1} X_i X'_i (X'X)^{-1}}{1 - p_{ii}}, \quad (3.4)$$

con $p_{ii} = X'_i (X'X)^{-1} X_i$.

De (3.4), se tiene,

$$\begin{aligned} \widehat{\beta} - \widehat{\beta}_{(i)} &= (X'X)^{-1} X'Y - (X'_{(i)} X'_{(i)})^{-1} X'_{(i)} Y_{(i)} \\ &= \left((X'_{(i)} X'_{(i)})^{-1} - \frac{(X'X)^{-1} X_i X'_i (X'X)^{-1}}{1 - p_{ii}} \right) X'Y - (X'_{(i)} X'_{(i)})^{-1} X'_{(i)} Y_{(i)} \\ &= (X'_{(i)} X'_{(i)})^{-1} (X'Y - X'_{(i)} Y_{(i)}) - \frac{(X'X)^{-1} X_i X'_i (X'X)^{-1} X'Y}{1 - p_{ii}} \\ &= (X'_{(i)} X'_{(i)})^{-1} X_i Y'_i - \frac{(X'X)^{-1} X_i X'_i (X'X)^{-1} X'Y}{1 - p_{ii}}. \end{aligned} \quad (3.5)$$

Por otro lado

$$\begin{aligned} (X'_{(i)} X'_{(i)})^{-1} X_i Y'_i &= \left((X'X)^{-1} + \frac{(X'X)^{-1} X_i X'_i (X'X)^{-1}}{1 - p_{ii}} \right) X_i Y'_i \\ &= (X'X)^{-1} X_i Y'_i + \frac{p_{ii} (X'X)^{-1} X_i Y'_i}{1 - p_{ii}} \\ &= \frac{(X'X)^{-1} X_i Y'_i}{1 - p_{ii}}. \end{aligned} \quad (3.6)$$

Sustituyendo (3.6) en (3.5) se obtiene que

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}_{(i)} &= \frac{(X'X)^{-1}X_iY'_i - (X'X)^{-1}X_iX'_i(X'X)^{-1}X'Y}{1 - p_{ii}} \\ &= \frac{(X'X)^{-1}X_i}{1 - p_{ii}}(Y'_i - X'_i(X'X)^{-1}X'Y).\end{aligned}\quad (3.7)$$

Ahora, dado que $\widehat{\varepsilon} = (Y - X\widehat{\beta}) = (I - P)Y$, entonces $\widehat{\varepsilon}'_i = e_i^{n'}(Y - X\widehat{\beta}) = Y'_i - X'_i(X'X)^{-1}X'Y$, así

$$\widehat{\beta} - \widehat{\beta}_{(i)} = \frac{(X'X)^{-1}X_i\widehat{\varepsilon}'_i}{1 - p_{ii}}.\quad (3.8)$$

Dado que estamos interesados en distinguir observaciones influyentes, en funciones lineales de los parámetros β , i.e. en funciones lineales de la forma $N\beta M$, entonces se propone la siguiente modificación a la distancia de Cook, denotada por DNM_{m_i} y definida como

$$DNM_{m_i} = \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(i)})M)\widehat{\text{Cov}}\left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M)\right)^{-} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M),\quad (3.9)$$

donde $N \in \mathbb{R}^{l \times q}$ y $M \in \mathbb{R}^{p \times s}$, con $r(N) = l$ y $r(M) = s$. Para determinar DNM_{m_i} note que $\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) = (M' \otimes N) \text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})$ y que la $\widehat{\text{Cov}}\left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M)\right) = (M' \otimes N)\widehat{\text{Cov}}\left(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})\right)(M \otimes N')$. Por tanto, sólo hay que calcular la $\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})$ y la $\widehat{\text{Cov}}\left(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})\right)$.

Dado que $\widehat{\varepsilon}'_i = e_i^{n'}(Y - X\widehat{\beta}) = e_i^{n'}(I - P)Y$, es claro que $\text{vec}(\widehat{\varepsilon}'_i) = (I_p \otimes e_i^{n'}(I - P)) \text{vec}(Y)$, por tanto

$$\begin{aligned}\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}) &= \frac{(I_p \otimes (X'X)^{-1}X_i)}{1 - p_{ii}} \text{vec}(\widehat{\varepsilon}'_i) \\ &= \frac{(I_p \otimes (X'X)^{-1}X_i)}{1 - p_{ii}} (I_p \otimes e_i^{n'}(I - P)) \text{vec}(Y) \\ &= \frac{(I_p \otimes (X'X)^{-1}X_i e_i^{n'}(I - P))}{1 - p_{ii}} \text{vec}(Y) \\ &= \frac{(I_p \otimes (X'X)^{-1}X_i P'_i)}{1 - p_{ii}} \text{vec}(Y),\end{aligned}\quad (3.10)$$

donde $P'_i = e_i^{n'}(I - P)$ es el i -ésimo renglón de la matriz $(I - P)$. Entonces

$$\begin{aligned}
 \text{Cov}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})) &= \frac{(I_p \otimes (X'X)^{-1} X_i P_i')}{1 - p_{ii}} \text{Cov}(\text{vec}(Y)) \frac{(I_p \otimes P_i X_i' (X'X)^{-1})}{1 - p_{ii}} \\
 &= \frac{(I_p \otimes (X'X)^{-1} X_i P_i')}{(1 - p_{ii})^2} (\Sigma \otimes I) (I_p \otimes P_i X_i' (X'X)^{-1}) \\
 &= \frac{\|P_i\|^2 (\Sigma \otimes (X'X)^{-1} X_i X_i' (X'X)^{-1})}{(1 - p_{ii})^2}.
 \end{aligned} \tag{3.11}$$

Note que, $\|P_i\|^2 = e_i^{n'}(I - P)(I - P)e_i^n = 1 - p_{ii}$, así sustituyendo en (3.11) se tiene

$$\text{Cov}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})) = \frac{\Sigma \otimes (X'X)^{-1} X_i X_i' (X'X)^{-1}}{(1 - p_{ii})}. \tag{3.12}$$

Sea $S = n\widehat{\Sigma}/(n - q)$ y observando que $E(S) = \Sigma$, entonces

$$\widehat{\text{Cov}}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})) = \frac{S \otimes (X'X)^{-1} X_i X_i' (X'X)^{-1}}{(1 - p_{ii})}. \tag{3.13}$$

Por lo tanto

$$\begin{aligned}
 \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) &= (M' \otimes N) \left(\frac{(I_p \otimes (X'X)^{-1} X_i P_i')}{1 - p_{ii}} \text{vec}(Y) \right) \\
 &= \frac{(M' \otimes N (X'X)^{-1} X_i P_i')}{1 - p_{ii}} \text{vec}(Y),
 \end{aligned} \tag{3.14}$$

y

$$\begin{aligned}
 \widehat{\text{Cov}}(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M)) &= (M' \otimes N) \frac{S \otimes (X'X)^{-1} X_i X_i' (X'X)^{-1}}{1 - p_{ii}} (M \otimes N') \\
 &= \frac{M' S M \otimes N (X'X)^{-1} X_i X_i' (X'X)^{-1} N'}{1 - p_{ii}}.
 \end{aligned} \tag{3.15}$$

Sea $v_i = N(X'X)^{-1} X_i$, $v_i \in \mathbb{R}^l$. Basado en los siguientes resultados:

1. Para $a \in \mathbb{R}^n$, $a^- = a' / \|a\|^2$,
2. Dada $A \in \mathbb{R}^{p \times q}$, $(AA')^- = A'^- A^-$ con $A^{-1} = A^-$ si A es no singular.

3. Dadas las matrices A y B , $(A \otimes B)^- = A^- \otimes B^-$,

se tiene que

$$\begin{aligned} \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \right) \right)^- &= \left(\frac{M' SM \otimes v_i v_i'}{1 - p_{ii}} \right)^- \\ &= \frac{1 - p_{ii}}{\|v_i\|^4} ((M' SM)^{-1} \otimes v_i v_i'). \end{aligned}$$

Por tanto la distancia modificada de Cook, DNM_{m_i} , puede ser reescrita como,

$$\begin{aligned} DNM_{m_i} &= \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \right) \right)^- \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \\ &= \left(\frac{(M' \otimes v_i P_i')}{1 - p_{ii}} \text{vec}(Y) \right)' \frac{(1 - p_{ii})((M' SM)^{-1} \otimes v_i v_i')}{\|v_i\|^4} \left(\frac{(M' \otimes v_i P_i')}{1 - p_{ii}} \text{vec}(Y) \right) \\ &= (1 - p_{ii})^{-1} \text{vec}'(Y) \frac{(M(M' SM)^{-1} M' \otimes P_i v_i' v_i v_i' P_i')}{\|v_i\|^4} \text{vec}(Y) \\ &= (1 - p_{ii})^{-1} \text{vec}'(Y) (M(M' SM)^{-1} M' \otimes P_i P_i') \text{vec}(Y). \end{aligned} \tag{3.16}$$

Alternativamente, dado que $\text{tr}(BX'CXD) = \text{vec}'(X)(B'D' \otimes C) \text{vec}(X)$, para matrices de dimensiones adecuadas, se puede escribir DNM_{m_i} como

$$DNM_{m_i} = (1 - p_{ii})^{-1} \text{tr}(M(M' SM)^{-1} M' Y' P_i P_i' Y).$$

Por otro lado, como $\widehat{\varepsilon}'_i = e_i^{n'}(Y - X\widehat{\beta}) = P_i' Y$, entonces

$$\begin{aligned} DNM_{m_i} &= (1 - p_{ii})^{-1} \text{tr}(M(M' SM)^{-1} M' \widehat{\varepsilon}_i \widehat{\varepsilon}'_i) \\ &= (1 - p_{ii})^{-1} \text{tr}(\widehat{\varepsilon}'_i M(M' SM)^{-1} M' \widehat{\varepsilon}_i) \\ &= (1 - p_{ii})^{-1} (M' \widehat{\varepsilon}_i)' (M' SM)^{-1} (M' \widehat{\varepsilon}_i) \end{aligned}$$

De esta forma se tiene que DNM_{m_i} puede expresarse de cualquiera de las formas dadas en el siguiente resultado.

Teorema 3.1. *Considere el modelo de regresión lineal multivariado dado en (3.1). Bajo el supuesto de normalidad de la matriz de errores y considerando las matrices $N \in \mathbb{R}^{l \times q}$ y $M \in \mathbb{R}^{p \times s}$, con $r(N) = l$ y $r(M) = s$. Entonces la distancia de Cook modificada, DNM_{m_i} , para detectar un dato outlier en funciones lineales de los parámetros β , $N\beta M$, puede ser escrita como:*

$$DNM_{m_i} = \begin{cases} \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \right) \right)^{-1} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M), \\ (1 - p_{ii})^{-1} \text{vec}'(Y)(M(M'SM)^{-1}M' \otimes P_i P_i') \text{vec}(Y), \\ (1 - p_{ii})^{-1} \text{tr}(M(M'SM)^{-1}M'Y'P_i P_i'Y), \\ (1 - p_{ii})^{-1} (M'\widehat{\varepsilon}_i)'(M'SM)^{-1}(M'\widehat{\varepsilon}_i). \end{cases} \quad (3.17)$$

La distancia DNM_{m_i} , al utilizar a S como estimador de Σ , sólo mide la influencia de la i -ésima observación en funciones lineales de los parámetros β , $N\beta M$. Es posible que se esté interesado en medir la influencia en $N\beta M$, tomando en cuenta la influencia al eliminar la i -ésima observación del estimador de matriz de parámetros Σ . En tal caso es preferible utilizar como estimador de Σ a $S_{(i)}$, el cual es obtenido de la misma forma que S , pero con la muestra reducida en $(n - 1)$, i.e. eliminando la i -ésima observación. Así, reemplazando S por $S_{(i)}$ en (3.17) se obtiene una nueva modificación de la distancia de Cook, la cual mide la influencia de la i -ésima observación en $N\beta M$ y Σ . Tal medida se denotará por $DNM_{m_i}^*$ y se define en el siguiente resultado.

Teorema 3.2. *Bajo las mismas condiciones del Teorema 3.1 y utilizando como estimador de Σ a $S_{(i)}$, la distancia de Cook modificada, $DNM_{m_i}^*$, para medir la influencia de la i -ésima observación en funciones lineales de los parámetros β , $N\beta M$, puede ser escrita como:*

$$DNM_{m_i}^* = \begin{cases} \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M) \right) \right)^{-1} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(i)})M), \\ (1 - p_{ii})^{-1} \text{vec}'(Y)(M(M'S_{(i)}M)^{-1}M' \otimes P_i P_i') \text{vec}(Y), \\ (1 - p_{ii})^{-1} \text{tr}(M(M'S_{(i)}M)^{-1}M'Y'P_i P_i'Y), \\ (1 - p_{ii})^{-1} (M'\widehat{\varepsilon}_i)'(M'S_{(i)}M)^{-1}(M'\widehat{\varepsilon}_i). \end{cases} \quad (3.18)$$

En la sección 3.4 se obtendrán las distribuciones exactas de DNM_{m_i} y $DNM_{m_i}^*$. Se verá, como para obtener la distribución exacta de DNM_{m_i} se requiere de un procedimiento más complejo que para obtener la distribución exacta de $DNM_{m_i}^*$. Esto se debe a que DNM_{m_i} utiliza a $\widehat{\varepsilon}_i$ y a S , los cuales no son estocasticamente independientes. En cambio, $DNM_{m_i}^*$ utiliza a $\widehat{\varepsilon}_i$ y a $S_{(i)}$, los cuales obviamente son estocasticamente independientes, esto facilita la deducción de su distribución exacta.

3.3 Detectando un Conjunto de Observaciones Influyentes

Sea $I = \{i_1, i_2, \dots, i_k\}$ un conjunto de tamaño k de $\{1, 2, \dots, n\}$, tal que $(n - k) \geq q$. Ahora, para el modelo (3.1) denote $X_{(I)}$, $Y_{(I)}$ y $\epsilon_{(I)}$ las matrices de regresión, de datos y de errores, respectivamente, obtenidas después de borrar las correspondientes observaciones de acuerdo a los subíndices I . Sean $\widehat{\beta}_{(I)}$ y $\widehat{\Sigma}_{(I)}$ los correspondientes estimadores de verosimilitud máxima en el modelo

$$Y_{(I)} = X_{(I)}\beta_{(I)} + \epsilon_{(I)}, \quad \epsilon_{(I)} \sim \mathcal{N}_{(n-k) \times p}(\mathbf{0}, \Sigma_{(I)} \otimes I_n) .$$

Analogamente al caso de una observación, observe que:

$$X'X = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix} \begin{pmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_n \end{pmatrix} = \sum_{j=1}^n X_j X'_j = \sum_{j \in I^c} X_j X'_j + \sum_{j \in I} X_j X'_j = X'_{(I)} X_{(I)} + X_I X'_I,$$

donde $X_{(I)} \in \mathbb{R}^{n-k \times q}$, $X_I \in \mathbb{R}^{q \times k}$ con $r(X_I) = k$. X_I es la matriz formada por las filas de X en I .

Ahora, usando Teorema de Sherman-Morrison-Woodbury (vea Apéndice A, Teorema A.1) se tiene

$$(X'X - X_I X'_I)^{-1} = (X'_{(I)} X_{(I)})^{-1} = (X'X)^{-1} + (X'X)^{-1} X_I (I_k - P_I)^{-1} X'_I (X'X)^{-1}, \quad (3.19)$$

donde $P_I = X'_I (X'X)^{-1} X_I$.

Como $\widehat{\beta}_{(I)} = (X'_{(I)} X_{(I)})^{-1} X'_{(I)} Y_{(I)}$ y $X'Y = X'_{(I)} Y_{(I)} + X_I Y_I$, entonces aplicando (3.19) se obtiene que

$$\begin{aligned} \widehat{\beta}_{(I)} &= ((X'X)^{-1} + (X'X)^{-1} X_I (I_k - P_I)^{-1} X'_I (X'X)^{-1}) (X'Y - X_I Y_I) \\ &= \widehat{\beta} + (X'X)^{-1} X_I (I_k - P_I)^{-1} X'_I \widehat{\beta} - (X'X)^{-1} X_I Y_I \\ &\quad - (X'X)^{-1} X_I (I_k - P_I)^{-1} P_I Y_I. \end{aligned}$$

Ahora, sumando y restando $(X'X)^{-1} (I_k - P_I)^{-1} X_I Y_I$ se tiene

$$\begin{aligned} \widehat{\beta}_{(I)} &= \widehat{\beta} + (X'X)^{-1} X_I (I_k - P_I)^{-1} X'_I \widehat{\beta} - (X'X)^{-1} (I_k - P_I)^{-1} X_I Y_I \\ &\quad - (X'X)^{-1} X_I (I - (I_k - P_I)^{-1} + (I_k - P_I)^{-1} P_I) Y_I \\ &= \widehat{\beta} - (X'X)^{-1} X_I (I_k - P_I)^{-1} (Y_I - X'_I \widehat{\beta}) \end{aligned}$$

Por tanto

$$\widehat{\beta} - \widehat{\beta}_{(I)} = (X'X)^{-1}X_I(I_k - P_I)^{-1}\widehat{\varepsilon}_I. \quad (3.20)$$

Dado que estamos interesados en distinguir múltiples observaciones influyentes, en funciones lineales de los parámetros β , $N\beta M$, entonces se propone la siguiente modificación a la distancia de Cook, denotada por DNM_{m_I} y definida como

$$DNM_{m_I} = \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(I)})M)\widehat{\text{Cov}}\left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M)\right)^{-1}\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M), \quad (3.21)$$

donde $N \in \mathbb{R}^{l \times q}$ y $M \in \mathbb{R}^{p \times s}$, con $r(N) = l$ y $r(M) = s$. Para determinar DNM_{m_I} sólo hay que calcular $\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})$ y $\widehat{\text{Cov}}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}))$.

Dado que $\widehat{\varepsilon}_I = U_I'(Y - X\widehat{\beta}) = U_I'(I - P)Y$, donde $U_I = (e_{i_1}^n, e_{i_2}^n, \dots, e_{i_2}^n)$, es claro que $\text{vec}(\widehat{\varepsilon}_I) = (I_p \otimes U_I'(I - P))\text{vec}(Y)$. Así,

$$\begin{aligned} \text{vec}(\widehat{\beta} - \widehat{\beta}_{(I)}) &= (I_p \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}U_I'(I - P))\text{vec}(Y) \\ &= (I_p \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}H_I)\text{vec}(Y), \end{aligned} \quad (3.22)$$

donde $H_I = U_I'(I - P)$. Por lo que

$$\begin{aligned} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M) &= (M' \otimes N)(I_p \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}H_I)\text{vec}(Y) \\ &= (M' \otimes N(X'X)^{-1}X_I(I_k - P_I)^{-1}H_I)\text{vec}(Y). \end{aligned} \quad (3.23)$$

Por otro lado,

$$\begin{aligned} \text{Cov}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(I)})) &= \Sigma \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}H_I H_I'(I_k - P_I)^{-1}X_I'(X'X)^{-1} \\ &= \Sigma \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}X_I'(X'X)^{-1}. \end{aligned} \quad (3.24)$$

Sustituyendo Σ por S , se tiene

$$\widehat{\text{Cov}}(\text{vec}(\widehat{\beta} - \widehat{\beta}_{(I)})) = S \otimes (X'X)^{-1}X_I(I_k - P_I)^{-1}X_I'(X'X)^{-1}, \quad (3.25)$$

De donde

$$\widehat{\text{Cov}}(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M)) = M'SM \otimes N(X'X)^{-1}X_I(I_k - P_I)^{-1}X_I'(X'X)^{-1}N', \quad (3.26)$$

luego

$$\left(\widehat{\text{Cov}}(N(\text{vec}(\hat{\beta} - \hat{\beta}_{(I)}))M) \right)^{-} = (M'SM)^{-1} \otimes (NRN')^{-}, \quad (3.27)$$

donde $R = (X'X)^{-1}X_I(I_k - P_I)^{-1}X_I'(X'X)^{-1}$, por tanto $R^{-} = (X'X)X_I^{-}(I_k - P_I)X_I^{-}(X'X)$. Note que si $r(N) < r(R)$, entonces NRN' es no singular.

Así, la distancia modificada de Cook propuesta para detectar múltiples observaciones influyentes, DNM_{m_I} , puede ser reescrita como:

$$\begin{aligned} DNM_{m_I} &= \text{vec}'(N(\hat{\beta} - \hat{\beta}_{(I)})M) \widehat{\text{Cov}} \left(\text{vec}(N(\hat{\beta} - \hat{\beta}_{(I)})M) \right)^{-} \text{vec}(N(\hat{\beta} - \hat{\beta}_{(I)})M) \\ &= \text{vec}'(\hat{\beta} - \hat{\beta}_{(I)})(M \otimes N') \left((M'SM)^{-1} \otimes (NRN')^{-} \right) (M' \otimes N) \text{vec}(\hat{\beta} - \hat{\beta}_{(I)}) \\ &= \text{vec}'(\hat{\beta} - \hat{\beta}_{(I)}) \left(M(M'SM)^{-1}M' \otimes N'(NRN')^{-}N \right) \text{vec}(\hat{\beta} - \hat{\beta}_{(I)}). \end{aligned}$$

Alternativamente, aplicando $\text{tr}(BX'CXD) = \text{vec}'(X)(B'D' \otimes C) \text{vec}(X)$, para matrices de dimensiones adecuadas, se puede escribir DNM_{m_I} como

$$\begin{aligned} DNM_{m_I} &= \text{tr}(\hat{\beta} - \hat{\beta}_{(I)})' N'(NRN')^{-} N(\hat{\beta} - \hat{\beta}_{(I)})M(M'SM)^{-1}M' \\ &= \text{tr}(M'SM)^{-1} \left(N(\hat{\beta} - \hat{\beta}_{(I)})M \right)' (NRN')^{-} \left(N(\hat{\beta} - \hat{\beta}_{(I)})M \right). \end{aligned} \quad (3.28)$$

Pero $N(\hat{\beta} - \hat{\beta}_{(I)})M = N(X'X)^{-1}X_I(I_k - P_I)^{-1}\hat{\varepsilon}_I M$, por lo que

$$\begin{aligned} &\left(N(\hat{\beta} - \hat{\beta}_{(I)})M \right)' (NRN')^{-} \left(N(\hat{\beta} - \hat{\beta}_{(I)})M \right) \\ &= M'\hat{\varepsilon}_I'(I_k - P_I)^{-1}X_I'(X'X)^{-1}N'(NRN')^{-}N(X'X)^{-1}X_I(I_k - P_I)^{-1}\hat{\varepsilon}_I M \\ &= M'\hat{\varepsilon}_I'(I_k - P_I)^{-1}X_I'(X'X)^{-1}N'N'^{-}(X'X)X_I^{-}(I_k - P_I) \\ &\quad X_I^{-}(X'X)N^{-}N(X'X)^{-1}X_I(I_k - P_I)^{-1}\hat{\varepsilon}_I M \\ &= M'\hat{\varepsilon}_I'(I_k - P_I)^{-1}X_I'X_I^{-}(I_k - P_I)X_I^{-}X_I(I_k - P_I)^{-1}\hat{\varepsilon}_I M, \end{aligned}$$

note que para $N \in \mathbb{R}^{l \times q}$, con $r(N) = l$, se tiene

$$(X'X)^{-1}N'N'^{-}(X'X) = \left(N'^{-}(X'X) \right)^{-} \left(N'^{-}(X'X) \right) = I_l.$$

Por tanto,

$$\begin{aligned}
& \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right)' (NRN')^{-1} \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right) \\
&= M' \widehat{\varepsilon}'_I (I_k - P_I)^{-1} (X_I^- X_I)' (I_k - P_I) X_I^- X_I (I_k - P_I)^{-1} \widehat{\varepsilon}_I M \\
&= M' \widehat{\varepsilon}'_I (I_k - P_I)^{-1} \widehat{\varepsilon}_I M.
\end{aligned}$$

Así

$$DNM_{m_I} = \text{tr}(M' SM)^{-1} M' \widehat{\varepsilon}'_I (I_k - P_I)^{-1} \widehat{\varepsilon}_I M. \quad (3.29)$$

De esta forma se tiene que DNM_{m_I} puede expresarse de cualquiera de las formas dadas en el siguiente resultado.

Teorema 3.3. *Considere el modelo de regresión lineal multivariado dado en (3.1). Bajo el supuesto de normalidad de la matriz de errores y considerando las matrices $N \in \mathbb{R}^{l \times q}$ y $M \in \mathbb{R}^{p \times s}$, con $r(N) = l$ y $r(M) = s$. Entonces la distancia de Cook modificada, DNM_{m_I} , para detectar k outliers en funciones lineales de los parámetros β , $N\beta M$, puede ser escrita como:*

$$DNM_{m_I} = \begin{cases} \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(I)})M) \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M) \right) \right)^{-1} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M), \\ \text{vec}'(\widehat{\beta} - \widehat{\beta}_{(I)}) \left(M(M' SM)^{-1} M' \otimes N'(NRN')^{-1} N \right) \text{vec}(\widehat{\beta} - \widehat{\beta}_{(I)}), \\ \text{tr}(M' SM)^{-1} \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right)' (NRN')^{-1} \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right), \\ \text{tr}(M' SM)^{-1} M' \widehat{\varepsilon}'_I (I_k - P_I)^{-1} \widehat{\varepsilon}_I M. \end{cases} \quad (3.30)$$

Análogamente al caso de una observación, en distancia DNM_{m_I} , se puede estar interesado en medir la influencia de un conjunto de k observaciones, I , en funciones lineales de los parámetros β , $N\beta M$, considerando el estimador de Σ , $S_{(I)}$, el cual es obtenido de la misma forma que S , pero eliminando las correspondientes k observaciones de acuerdo a los subíndices I . Así, reemplazando S por $S_{(I)}$ en (3.30) se obtiene una nueva modificación de la distancia de Cook, la cual mide la influencia de un conjunto de k observaciones en $N\beta M$ y Σ . Tal medida se denota por $DNM_{m_I}^*$ y se define en el siguiente resultado.

Teorema 3.4. *Bajo las mismas condiciones del Teorema 3.3 y utilizando como estimador de Σ a $S_{(I)}$, la distancia de Cook modificada, $DNM_{m_I}^*$, para medir la influencia de un conjunto de k observaciones, I , en funciones lineales de los parámetros β , $N\beta M$, y en la matriz de parámetros Σ , puede ser escrita como:*

$$DNM_{m_I}^* = \begin{cases} \text{vec}'(N(\widehat{\beta} - \widehat{\beta}_{(I)})M) \left(\widehat{\text{Cov}} \left(\text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M) \right) \right)^{-1} \text{vec}(N(\widehat{\beta} - \widehat{\beta}_{(I)})M), \\ \text{vec}'(\widehat{\beta} - \widehat{\beta}_{(I)}) \left(M(M' S_{(I)} M)^{-1} M' \otimes N'(NRN')^{-1} N \right) \text{vec}(\widehat{\beta} - \widehat{\beta}_{(I)}), \\ \text{tr}(M' S_{(I)} M)^{-1} \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right)' (NRN')^{-1} \left(N(\widehat{\beta} - \widehat{\beta}_{(I)})M \right), \\ \text{tr}(M' S_{(I)} M)^{-1} M' \widehat{\varepsilon}'_I (I_k - P_I)^{-1} \widehat{\varepsilon}_I M. \end{cases} \quad (3.31)$$

3.4 Distribuciones Asociadas con las Distancias Modificadas

La principal razón para explorar las modificaciones, dadas en las secciones 3.2 y 3.3, de la distancia de Cook es la de establecer una regla de decisión, para concluir si una observación, o un conjunto de observaciones, es influyente en funciones lineales de los parámetros β , $N\beta M$. En lugar de usar una aproximación a una distribución, se busca cuantificar la incertidumbre, de tal conclusión, derivando la distribución exacta de las distintas modificaciones de la distancia de Cook.

Las distribuciones exactas de DNM_{m_i} y $DNM_{m_i}^*$, utilizando estadísticos pivotaes, se muestran en el siguiente resultado.

Teorema 3.5. *Considerando el modelo de regresión lineal multivariado (3.1) y las definiciones de DNM_{m_i} y $DNM_{m_i}^*$, dadas en los Teoremas 3.1 y 3.2, respectivamente. Suponiendo que $\varepsilon \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I_n)$, se tiene,*

1.

$$\frac{DNM_{m_i}}{n - q} \sim \beta(s/2, (n - q - s)/2), \quad (3.32)$$

donde $\beta(s/2, (n - q - s)/2)$ denota una distribución beta centrada con parámetros $s/2$ y $(n - q - s)/2$.

2.

$$\frac{(n - q - s)DNM_{m_i}^*}{s(n - q - 1)} \sim \mathcal{F}(s, n - q - s), \quad (3.33)$$

donde $\mathcal{F}(s, n - q - s)$ denota una distribución \mathcal{F} centrada con s y $(n - q - s)$ grados de libertad (gl).

Demostración.

1. Dado que

$$\begin{aligned} \widehat{\Sigma} &= \frac{1}{n} Y'(I - P)Y \\ &= \frac{1}{n} \varepsilon'(I - P)\varepsilon, \end{aligned}$$

entonces, si $\varepsilon \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I_n)$ se tiene

$$n\widehat{\Sigma} = \varepsilon'(I - P)\varepsilon \sim \mathcal{W}_p(\Sigma, n - q).$$

Así,

$$(n - q)S \sim \mathcal{W}_p(\Sigma, n - q),$$

donde $S = \frac{n}{n - q} \widehat{\Sigma}$.

Por tanto

$$(n - q)M'SM \sim \mathcal{W}_s(M'\Sigma M, n - q), \quad (3.34)$$

donde $M \in \mathbb{R}^{p \times S}$, con $r(M) = s$.

Por otro lado, $\widehat{\varepsilon}_i \sim \mathcal{N}_p(0, (1 - p_{ii})\Sigma)$, entonces $(1 - p_{ii})^{-1/2}\widehat{\varepsilon}_i \sim \mathcal{N}_p(0, \Sigma)$, por lo que $(1 - p_{ii})^{-1/2}M'\widehat{\varepsilon}_i \sim \mathcal{N}_s(0, M'\Sigma M)$. Así,

$$M'\widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i'M \sim \mathcal{W}_s(M'\Sigma M, 1). \quad (3.35)$$

Pero S y $\widehat{\varepsilon}_i$ no son independientes. Sin embargo, basados en el Lema 1 de Ellenberg(1973) (vea Apéndice B.7),

$$(n - q)S = (n - q - 1)S_{(i)} + \widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i',$$

por lo que

$$(n - q)M'SM = (n - q - 1)M'S_{(i)}M + M'\widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i'M, \quad (3.36)$$

mostrando que $S_{(i)}$ y $\widehat{\varepsilon}_i$ son independientes. Además, como $(n - q - 1)S_{(i)} \sim \mathcal{W}_p(\Sigma, n - q - 1)$, entonces $(n - q - 1)M'S_{(i)}M \sim \mathcal{W}_s(M'\Sigma M, n - q - 1)$.

Ahora, note que

$$\begin{aligned} \frac{DNM_{m_i}}{n - q} &= (1 - p_{ii})^{-1}(M'\widehat{\varepsilon}_i)'((n - q)M'SM)^{-1}(M'\widehat{\varepsilon}_i) \\ &= \text{tr } M'\widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i'M((n - q)M'SM)^{-1} \\ &= \text{tr } M'\widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i'M((n - q - 1)M'S_{(i)}M + M'\widehat{\varepsilon}_i(1 - p_{ii})^{-1}\widehat{\varepsilon}_i'M)^{-1}. \end{aligned}$$

Así, se tiene que $\frac{DNM_{m_i}}{n - q} = V^{(1)}$, donde $V^{(1)}$ es el estadístico de Pillai el cual, según Seber (p. 38, 1984), tiene distribución $\beta((|1 - s| + 1)/2, (n - q - 1 - s + 1)/2)$. Por tanto

$$\frac{DNM_{m_i}}{n - q} \sim \beta(s/2, (n - q - s)/2).$$

2. Para demostrar 2, note que

$$DNM_{m_i}^* = (n - q - 1)(1 - p_{ii})^{-1/2}(M'\widehat{\varepsilon}_i)'((n - q - 1)M'S_{(i)}M)^{-1}(M'\widehat{\varepsilon}_i)(1 - p_{ii})^{-1/2},$$

entonces

$$DNM_{m_i}^* \sim \mathcal{T}_{(s, n-q-1)}^2,$$

donde $\mathcal{T}_{(s, n-q-1)}^2$ representa la distribución \mathcal{T}^2 de Hotelling con parámetros s , y $n - q - 1$. Usando relaciones entre la distribución \mathcal{T}^2 de Hotelling y la distribución \mathcal{F} , se tiene que $\mathcal{T}_{(s, n-q-1)}^2 = \frac{s(n-q-1)}{n-q-s} \mathcal{F}_{(s, n-q-s)}$.

Así

$$\frac{n-q-s}{s(n-q-1)} DNM_{m_i}^* \sim \mathcal{F}_{(s, n-q-s)}.$$

■

Del Teorema 3.5, dado un nivel de significancia α , podemos escribir la siguiente regla de decisión:

Y_i , $i = 1, 2, \dots, n$, es una observación influyente en $N\beta M$ si

$$\frac{DNM_{m_i}}{n-q} \geq \beta_{\alpha: s/2, (n-q-s)/2}, \quad (3.37)$$

donde $\beta_{\alpha: s/2, (n-q-s)/2}$ es el correspondiente α -percentil superior de una distribución β con parámetros $s/2$ y $(n-q-s)/2$. O alternativamente

Y_i , $i = 1, 2, \dots, n$, es una observación influyente en $N\beta M$ si

$$\frac{(n-q-s)DNM_{m_i}^*}{s(n-q-1)} \geq F_{\alpha: s, (n-q-s)}, \quad (3.38)$$

donde $F_{\alpha: s, (n-q-s)}$ es el correspondiente α -percentil superior de una distribución \mathcal{F} con s y $(n-q-s)$ gl. Debido al problema que se tiene al realizar pruebas múltiples, punto por punto, Draper y Smith (1981) proponen utilizar un nivel de significancia corregido. Así, el estadístico $\frac{(n-q-s)DNM_{m_i}^*}{s(n-q-1)}$ se llevaría a la forma del estadístico F de Draper y Smith [p. 174, 1981] y posteriormente se compararía con el α/n -percentil superior de una distribución $\mathcal{F}_{(s, n-q-s)}$.

Para el caso de múltiples observaciones influyentes, las distribuciones exactas de DNM_{m_I} y $DNM_{m_I}^*$, utilizando estadísticos pivotaes, se muestran en el siguiente resultado.

Teorema 3.6. *Considerando el modelo de regresión lineal multivariado (3.1) y las definiciones de DNM_{m_I} y $DNM_{m_I}^*$, dadas en los Teoremas 3.3 y 3.4, respectivamente. Suponiendo que $\varepsilon \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I_n)$, se tiene,*

1.

$$\frac{DNM_{m_I}}{n - q} \sim \mathcal{P}(w, m, h), \quad (3.39)$$

donde $\mathcal{P}(w, m, h)$ denota la distribución centrada para el estadístico Pillai con parámetros w , m , y h .

2.

$$\frac{DNM_{m_I}^*}{n - q - k} \sim \mathcal{LH}(w, m, h), \quad (3.40)$$

donde $\mathcal{LH}(w, m, h)$ denota la distribución centrada para el estadístico Lawley-Hotelling con parámetros w , m , y h .

En ambos casos los parámetros son definidos como $w = \min(s, k)$, $m = (|s - k| - 1)/2$ y $h = (n - q - s - 1)/2$.

Demostración.

1. Sea $\hat{\varepsilon}_I = U_I'(I - P)Y = U_I'(I - P)\varepsilon$. Si $\varepsilon \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I_n)$, entonces $\hat{\varepsilon}_I \sim \mathcal{N}_{k \times p}(0, \Sigma \otimes (I - P_I))$, por lo que $(I - P_I)^{1/2}\hat{\varepsilon}_I \sim \mathcal{N}_{k \times p}(0, \Sigma \otimes I_k)$, de donde $(I - P_I)^{1/2}\hat{\varepsilon}_I M \sim \mathcal{N}_{k \times p}(0, M'\Sigma M \otimes I_k)$ y $\hat{\varepsilon}_I'(I - P_I)^{-1}\hat{\varepsilon}_I \sim \mathcal{W}_p(\Sigma, k)$.

Note que hay dos casos, $k < p$ y $k \geq p$. Además, $(n - q)S \sim \mathcal{W}_p(\Sigma, n - q)$, pero S y $\hat{\varepsilon}_I$ no son independientes. Sin embargo, basados en el Lema 1 de Ellenberg(1973),

$$(n - q)S = (n - q - k)S_{(I)} + \hat{\varepsilon}_I'(I - P_I)^{-1}\hat{\varepsilon}_I,$$

por lo que

$$(n - q)M'SM = (n - q - k)M'S_{(I)}M + M'\hat{\varepsilon}_I'(I - P_I)^{-1}\hat{\varepsilon}_I M, \quad (3.41)$$

mostrando que $S_{(I)}$ y $\hat{\varepsilon}_I$ son independientes. Además, como $(n - q - k)S_{(I)} \sim \mathcal{W}_p(\Sigma, n - q - k)$, entonces $(n - q - k)M'S_{(I)}M \sim \mathcal{W}_s(M'\Sigma M, n - q - k)$.

Ahora, la matriz $\mathbf{B} \in \mathbb{R}^{k \times k}$, definida como

$$\begin{aligned}\mathbf{B} &= (I - P_I)^{-1/2} \widehat{\varepsilon}_I M ((n - q) M' S M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1/2} \\ &= (I - P_I)^{-1/2} \widehat{\varepsilon}_I M ((n - q - k) M' S_{(I)} M + M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1/2},\end{aligned}$$

tiene una distribución multivariada tipo beta y su densidad es dada en Srivastava (1968) y Srivastava y Khatri (1979), para el caso $k < p$. Por otro lado cuando $k \geq p$, es posible definir una matriz $\mathbf{B}_1 \in \mathbb{R}^{s \times s}$ como sigue

$$\begin{aligned}& (M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{1/2} ((n - q - k) M' S_{(I)} M + M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{-1} \\ & \times (M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{1/2},\end{aligned}$$

donde su densidad puede ser encontrada en Díaz-García y Gutiérrez-Jáimes (2001). $G^{1/2}$ es la raíz cuadrada definida no-negativa de G tal que $G = G^{1/2} G^{1/2}$. Una propiedad muy atractiva es que podemos obtener la densidad de la matriz \mathbf{B} a partir de la densidad de \mathbf{B}_1 , considerando las sustituciones dadas en Muirhead(1982). En este caso las sustituciones son:

$$s \rightarrow k, \quad k \rightarrow s, \quad n - q - k \rightarrow n - q - s. \quad (3.42)$$

Ahora, note que $\text{tr } \mathbf{B} = \text{tr } \mathbf{B}_1$, por lo que

$$\begin{aligned}& \text{tr}(I - P_I)^{-1/2} \widehat{\varepsilon}_I M ((n - q - k) M' S_{(I)} M + M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1/2} = \\ & \text{tr}((n - q - k) M' S_{(I)} M + M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M = \frac{DNM_{m_I}}{n - q} \\ & = V^{(w)} \sim \mathcal{P}(w, m, h),\end{aligned}$$

el cual es exactamente el estadístico $V^{(w)}$ de Pillai (notación de Rencher(1995)) bajo la hipótesis nula, y su densidad es dada en Khatri y Pillai (1968) para $k \geq p$. Si $k < p$ la correspondiente densidad es obtenida usando la sustitución dada en (3.42). Observe que si $k = 1$ entonces $w = 1$, y por tanto $DNM_{m_I}/(n - q) = V^{(1)} = DNM_{m_i}/(n - q)$.

2. De forma similar, $M' \widehat{\varepsilon}'_I (I - P_I)^{-1} \widehat{\varepsilon}_I M \sim \mathcal{W}_s(M' \Sigma M, k)$ independientemente de $(n - q - k) M' S_{(I)} M \sim \mathcal{W}_s(M' \Sigma M, n - q - k)$. Y de nuevo, se tienen dos casos; $k < p$ y $k \geq p$. Entonces la matriz $\mathbf{F} \in \mathbb{R}^{k \times k}$, definida como

$$\begin{aligned}\mathbf{F} &= (I - P_I)^{-1/2} \widehat{\varepsilon}_I M ((n - q - k) M' S_{(I)} M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1/2} \\ &= \frac{(I - P_I)^{-1/2} \widehat{\varepsilon}_I M (M' S_{(I)} M)^{-1} M' \widehat{\varepsilon}'_I (I - P_I)^{-1/2}}{n - q - k},\end{aligned}$$

tiene una distribución Wishart Studentised, y su densidad puede ser encontrada en el Teorema 10.4.4 de Muirhead (1982), para $k < p$. Si $k \geq p$, podríamos definir $\mathbf{F}_1 \in \mathbb{R}^{s \times s}$ como,

$$\mathbf{F}_1 = \frac{(M' \hat{\varepsilon}'_I (I - P_I)^{-1} \hat{\varepsilon}_I M)^{1/2} (M' S_{(I)} M)^{-1} (M' \hat{\varepsilon}'_I (I - P_I)^{-1} \hat{\varepsilon}_I M)^{1/2}}{n - q - k},$$

donde su función de densidad es dada en el Teorema 10.4.1 de Muirhead (1982). Para encontrar la densidad de \mathbf{F} a partir de la densidad de \mathbf{F}_1 se usarán las sustituciones de la ecuación (3.42). Note que $\text{tr } \mathbf{F} = \text{tr } \mathbf{F}_1$, así

$$\begin{aligned} \text{tr} \frac{(I - P_I)^{-1/2} \hat{\varepsilon}_I M (M' S_{(I)} M)^{-1} M' \hat{\varepsilon}'_I (I - P_I)^{-1/2}}{n - q - k} &= \\ \frac{\text{tr} (M' S_{(I)} M)^{-1} M' \hat{\varepsilon}'_I (I - P_I)^{-1} \hat{\varepsilon}_I M}{(n - q - k)} &= \frac{DNM_{m_I}^*}{n - q - k} = U^{(w)} \sim \mathcal{LH}(w, m, h), \end{aligned}$$

coincidiendo con el estadístico $U^{(w)}$ de Lawley-Hotelling (notación de Rencher(1995)) bajo la hipótesis nula. Su función de densidad es dada en el Corolario 10.6.3 de Muirhead (1982) para $k \geq p$. Si $k < p$ la correspondiente densidad es encontrada usando (3.42). Note que si $k = 1$ entonces $w = 1$, y por tanto $DNM_{m_I}^*/(n - q - 1) = U^{(w)} = DNM_{m_i}^*/(n - q - 1)$. ■

Las distribuciones $LH(w, m, h)$ y $P(w, m, h)$ son llamadas distribución de Lawley-Hotelling y distribución de Pillai, respectivamente, y han sido tabuladas por diferentes autores como una función de sus parámetros w , m , y h . Tomando las definiciones en Rencher[p. 183, 1995], las equivalencias de las notaciones aquí son, $w = \min(s, k)$ (Rencher utiliza s en lugar de w), $m = (|s - k| - 1)/2$ y $h = (n - q - s - 1)/2$. Note que esas tablas fueron preparadas para $k \geq p$, así que para el caso $k < p$ se tiene que usar las sustituciones dadas en (3.42). Si los valores que se necesitan no están disponibles en tablas, Rencher[p. 185-186, 1995] sugiere excelentes aproximaciones a través de una distribución \mathcal{F} , ver también Seber[p. 38-39, 1984].

Del Teorema 3.6, dado un nivel de significancia α , podemos escribir la siguiente regla de decisión:

$Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$, es un conjunto de observaciones influyentes en $N\beta M$ si

$$\frac{DNM_{m_I}}{n - q} \geq P_{\alpha:w,m,h}, \quad (3.43)$$

donde $P_{\alpha:w,m,h}$ es el correspondiente α -percentil superior de una distribución Pillai con parámetros $w = \min(s, k)$, $m = (|s - k| - 1)/2$ y $h = (n - q - s - 1)/2$. O alternativamente

$Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}$, es un conjunto de observaciones influyentes en $N\beta M$ si

$$\frac{DNM_{m_I}^*}{n - q - k} \geq LH_{\alpha:w,m,h}, \quad (3.44)$$

donde $LH_{\alpha:w,m,h}$ es el correspondiente α -percentil superior de una distribución Lawley-Hotelling con parámetros $w = \min(s, k)$, $m = (|s - k| - 1)/2$ y $h = (n - q - s - 1)/2$.

Observe que como consecuencia del Teorema B.4 (vea Apéndice B), los resultados en el Teorema 3.6 se pueden extender al caso en que los errores siguen una distribución de contornos elípticos, con lo cual $\widehat{\varepsilon}_I \sim \mathcal{E}_{k \times p}(0, \Sigma \otimes (I - P_I))$, esto es:

Teorema 3.7. *Considerando el modelo de regresión lineal multivariado (3.1) y las definiciones de DNM_{m_I} y $DNM_{m_I}^*$, dadas en los Teoremas 3.3 y 3.4, respectivamente. Suponiendo ahora que $\varepsilon_I \sim \mathcal{E}_{k \times p}(0, \Sigma \otimes I)$, se tiene,*

1.

$$\frac{DNM_{m_I}}{n - q} = \frac{\text{tr}(M'SM)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_IM}{n - q} \sim \mathcal{P}(w, m, h), \quad (3.45)$$

donde $\mathcal{P}(w, m, h)$ denota la distribución centrada para el estadístico Pillai con parámetros w , m , y h .

2.

$$\frac{DNM_{m_I}^*}{n - q - k} = \frac{\text{tr}(M'S_{(I)}M)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_IM}{n - q - k} \sim \mathcal{LH}(w, m, h), \quad (3.46)$$

donde $\mathcal{LH}(w, m, h)$ denota la distribución centrada para el estadístico Lawley-Hotelling con parámetros w , m , y h .

Análogamente, los resultados del Teorema 3.5 se pueden extender al caso en que los errores siguen una distribución de contornos elípticos.

Observaciones:

1. Este trabajo fué motivado en el estudio de los modelos de diseños y de medidas repetidas, en general para estos modelos, la matriz X es de rango incompleto, esto es, $r(X) = k \leq p$. Sin embargo, hasta el momento todos los resultados han sido propuestos para el caso de un modelo de rango completo, $r(X) = p$. A pesar de ello observe que cuando $r(X) = k \leq p$, las conclusiones en todos los resultados obtenidos en el presente capítulo son válidos, simplemente observando lo siguiente:

- (a) Sustituya la frase *funciones lineales de los parámetros* β , por **funciones lineales estimables de los parámetros** β en los enunciados de los teoremas.
- (b) Observe que, ahora el sistema de ecuaciones normales $X'X\tilde{\beta} = X'Y$, sigue siendo consistente, pero no tiene solución única, pudiéndose escribir todas sus soluciones como:

$$\tilde{\beta} = \begin{cases} (X'X)^c X'Y & \text{sobre todo c-inverso de } (X'X) \text{ o} \\ (X'X)^c X'Y + (I - (X'X)^c(X'X))h & \text{para } h : p \times 1 \text{ arbitrario} \end{cases}$$

donde todo c-inverso de A se denota por A^c y es tal que $AA^cA = A$.

- (c) Sin embargo, la matriz de predicciones es invariante bajo todo inverso generalizado de $X'X$, más aún, $P = X(X'X)^c X' = X^-X$, como en el caso de rango completo.
- (d) Análogamente, observe que el estimador de las funciones lineales estimables de los parámetros β , $N\beta M$, también es invariante bajo toda solución $\tilde{\beta}$ del sistema de ecuaciones normales, pues, $\widehat{N\beta M} = N\tilde{\beta}M = NX^-YM$. Con lo cual los resultados son inmediatos para el caso de rango incompleto.
2. En la definición de la distancia DNM_{m_i} se está empleando la matriz de covarianzas de $\text{vec}(\hat{\beta} - \hat{\beta}_{(i)})$. Bajo un modelo elíptico, hay que tomar en consideración que algunas distribuciones no tienen momentos. Sin embargo DNM_{m_i} , puede seguir siendo definida de la misma manera, pues en el caso elíptico la matriz definida por la ecuación (3.13), estima a la matriz de escala Σ y no a la matriz de covarianzas $c_0\Sigma$. Observando además que en el caso de una distribución normal $c_0 = 1$, vea Díaz-García y González-Farías (2004).

Capítulo 4

Ejemplos y Aplicaciones

4.1 Introducción

En este capítulo se ilustra el uso de las pruebas exactas, para detectar observaciones influyentes, obtenidas en el Capítulo 3. Se presentan tres ejemplos, en ellos se ilustra el procedimiento a seguir, para poder concluir si una observación o un conjunto de observaciones es influyente en funciones lineales estimables de los parámetros β , de la forma $N\beta M$. También se intenta dar una interpretación de las funciones lineales de los parámetros, en las cuales se está interesado en medir la influencia de una observación o un conjunto de observaciones. El programa, desarrollado en S-plus, para realizar los análisis correspondientes es presentado en el Apéndice C.

4.2 Ejemplo uno: mortalidad de truchas

Los datos de la Tabla 4.1 fueron analizados por Srivastava y Carter(1983). Veinticinco depósitos de truchas se sometieron a diferentes dosis de cobre en miligramos por litro. El peso promedio de los peces fue registrado. Las proporciones (p_{ij} , $i = 1, \dots, 25$, $j = 1, 2, 3, 4, 5$) de peces muertos después de 8, 14, 24, 36 y 48 horas fueron registradas.

El modelo a considerar es

$$Y = X\beta + \varepsilon.$$

donde $Y \in \mathbb{R}^{25 \times 5}$,

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{15} \\ y_{21} & \cdots & y_{25} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{n5} \end{pmatrix},$$

es la matriz respuesta, con $n = 25$ y $y_{ij} = \arcseno \sqrt{p_{ij}}$ (las variables respuesta p_{ij} fueron transformadas con arcseno para estabilizar la varianza), donde p_{ij} = proporción de peces muertos en el depósito i al tiempo t_j , con tiempos $t = (t_1, t_2, t_3, t_4, t_5) = (8, 14, 24, 36, 48)$; $X \in \mathbb{R}^{n \times 3}$ es la matriz de covariables, su primera columna está formada por unos correspondientes al intercepto, la segunda columna tiene logaritmos de dosis de cada depósito y la tercer columna es el peso promedio de los peces en los depósitos; finalmente, $\beta \in \mathbb{R}^{3 \times 5}$,

$$\beta = \begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} & \beta_{04} & \beta_{05} \\ \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} & \beta_{25} \end{pmatrix},$$

es la matriz de parámetros, siendo sus renglones los vectores de parámetros correspondientes al intercepto, a $\log(\text{Dosis})$ y a peso promedio, respectivamente.

El estimador de verosimilitud máxima de β es

$$\hat{\beta} = (X'X)^{-1}X'Y = X^{-1}Y = \begin{pmatrix} -94.60 & -228.67 & -258.80 & -246.05 & -232.96 \\ 19.27 & 43.65 & 48.99 & 47.33 & 45.10 \\ 24.68 & -20.59 & 3.32 & 5.22 & 8.24 \end{pmatrix}.$$

Supongamos, por ejemplo, que estamos interesados en comparar las mortalidades de peces a 14 y 36 horas después de inducir la contaminación con cobre; esta comparación se puede efectuar estimando las funciones paramétricas

$$\beta_{02} - \beta_{04}$$

$$\beta_{12} - \beta_{14},$$

$$\beta_{22} - \beta_{24}$$

o equivalentemente βM , donde $M = \begin{pmatrix} 0 & 1 & 0 & -1 & 0 \end{pmatrix}'$.

Table 4.1: Proporción de peces muertos después de 8, 14, 24, 36 y 48 horas de que se les suministro diferentes dosis de cobre.

Y_1	Y_2	Y_3	Y_4	Y_5	$\log(\text{Dosis})$ X_1	Peso Promedio X_2
0.00	0.00	30.00	30.00	30.00	5.60	0.67
0.00	18.44	33.21	33.21	33.21	6.02	0.64
0.00	45.00	60.00	71.57	71.57	6.41	0.73
22.79	53.73	90.00	90.00	90.00	6.85	0.77
42.13	90.00	90.00	90.00	90.00	7.28	0.57
0.00	12.92	26.57	26.57	26.57	5.60	0.78
12.92	18.44	33.21	33.21	33.21	6.02	0.81
12.92	42.13	77.08	90.00	90.00	6.41	0.82
18.44	56.79	90.00	90.00	90.00	6.85	0.87
26.57	67.21	90.00	90.00	90.00	7.28	0.84
0.00	0.00	0.00	0.00	12.92	5.60	0.86
0.00	12.92	22.79	30.00	33.21	6.02	0.91
0.00	22.79	77.08	77.08	77.08	6.41	1.03
0.00	46.79	77.08	90.00	90.00	6.85	1.05
18.44	67.21	90.00	90.00	90.00	7.28	1.05
0.00	0.00	0.00	12.92	18.44	5.60	0.62
0.00	12.92	22.79	26.57	30.00	6.02	0.53
18.44	42.13	77.08	77.08	77.08	6.41	0.60
18.44	56.79	90.00	90.00	90.00	6.85	0.64
36.27	77.08	90.00	90.00	90.00	7.28	0.67
0.00	12.92	26.57	26.57	26.57	5.60	0.57
0.00	0.00	22.79	30.00	30.00	6.02	0.60
0.00	39.23	71.57	90.00	90.00	6.41	0.63
12.92	53.73	90.00	90.00	90.00	6.85	0.69
33.21	67.21	90.00	90.00	90.00	7.28	0.72

En este caso también es de interés ver si existen observaciones influyentes en la combinación lineal

$$\begin{aligned} & \widehat{\beta}_{02} - \widehat{\beta}_{04} \\ & \widehat{\beta}_{12} - \widehat{\beta}_{14} \cdot \\ & \widehat{\beta}_{22} - \widehat{\beta}_{24} \end{aligned}$$

Cantidades ó matrices importantes para el cálculo de las métricas, DNM_{m_i} y $DNM_{m_i}^*$, presentadas en los Teoremas 3.1 y 3.2 son: los elementos diagonales de la matriz de predicción P , la matriz de residuales $\widehat{\varepsilon}$ y la estimación de la matriz de varianzas y covarianzas (S ó $S_{(i)}$). Para este ejemplo en particular, los resultados obtenidos son:

$$S = \frac{1}{n - q} (Y - X\widehat{\beta})'(Y - X\widehat{\beta}) = \begin{pmatrix} 52.5 & 20.3 & -12.2 & -35.6 & -31.6 \\ 20.3 & 59.6 & 37.1 & 34.8 & 33.3 \\ -12.2 & 37.1 & 183.1 & 182.3 & 162.5 \\ -35.6 & 34.8 & 182.3 & 216.4 & 196.1 \\ -31.6 & 33.3 & 162.5 & 196.1 & 182.7 \end{pmatrix}.$$

Para obtener $S_{(i)}$, se elimina la i -ésima observación del conjunto de observaciones. Así, al eliminar la observación 1 se tiene que

$$S_{(1)} = \frac{(Y_{(i)} - X_{(i)}\widehat{\beta}_{(i)})'(Y_{(i)} - X_{(i)}\widehat{\beta}_{(i)})}{(n - 1) - q}$$

$$= \begin{pmatrix} 54.5 & 21.6 & -14.9 & -38.6 & -33.9 \\ 21.6 & 62.3 & 40.2 & 37.2 & 35.4 \\ -14.9 & 40.2 & 183.6 & 186.0 & 166.9 \\ -38.6 & 37.2 & 186.0 & 223.6 & 203.4 \\ -33.9 & 35.4 & 166.9 & 203.4 & 190.0 \end{pmatrix}.$$

Además,

$$diag(P) = \begin{pmatrix} 0.12 \\ 0.07 \\ 0.04 \\ 0.06 \\ 0.22 \\ 0.13 \\ 0.08 \\ 0.05 \\ 0.08 \\ 0.13 \\ 0.17 \\ 0.12 \\ 0.19 \\ 0.20 \\ 0.24 \\ 0.13 \\ 0.13 \\ 0.08 \\ 0.09 \\ 0.15 \\ 0.15 \\ 0.09 \\ 0.06 \\ 0.07 \\ 0.13 \end{pmatrix}, y \hat{\varepsilon} = \begin{pmatrix} 3.22 & -1.90 & 12.31 & 7.60 & 4.97 \\ -5.55 & -2.30 & -4.84 & -8.81 & -10.42 \\ -11.02 & 8.75 & 2.19 & 10.28 & 9.29 \\ 4.44 & -0.55 & 10.86 & 8.03 & 7.88 \\ 10.38 & 12.59 & -9.69 & -11.41 & -9.98 \\ 6.00 & 13.34 & 8.51 & 3.57 & 0.61 \\ 11.61 & 1.24 & -5.41 & -9.71 & -11.83 \\ 4.18 & 7.78 & 18.96 & 28.23 & 26.96 \\ 2.53 & 4.55 & 10.54 & 7.51 & 7.06 \\ 1.58 & -4.55 & -10.60 & -12.84 & -12.24 \\ 7.96 & 2.05 & -18.32 & -23.41 & -13.69 \\ 0.97 & -2.37 & -16.14 & -13.40 & -12.60 \\ -3.64 & -7.31 & 18.27 & 14.23 & 12.33 \\ -11.57 & -1.84 & -2.97 & 6.60 & 5.62 \\ -1.46 & -0.31 & -11.28 & -13.92 & -13.94 \\ 1.99 & -2.93 & -17.52 & -9.22 & -6.19 \\ -8.26 & -10.07 & -14.90 & -14.88 & -12.73 \\ 4.15 & 3.16 & 19.70 & 16.48 & 15.88 \\ -3.16 & -0.20 & 11.30 & 8.72 & 8.96 \\ 6.97 & 1.71 & -10.02 & -11.93 & -10.80 \\ 0.73 & 8.94 & 9.21 & 4.69 & 2.36 \\ -6.45 & -21.49 & -15.15 & -11.84 & -13.34 \\ -13.41 & 0.99 & 14.07 & 29.21 & 28.51 \\ -7.53 & -2.31 & 11.15 & 8.48 & 8.58 \\ 5.35 & -6.95 & -10.21 & -12.24 & -11.28 \end{pmatrix},$$

donde el vector $diag(P)$ representa los elementos diagonales de la matriz de predicción $P = X(X'X)^{-1}X'$ y $\hat{\varepsilon} = (I - P)Y$, es la matriz de residuales.

Las distancias de Cook modificadas, DNM_{m_i} y $DNM_{m_i}^*$, para medir la influencia de la i -ésima observación en la combinación lineal $\hat{\beta}M$ son calculadas y presentadas en la Figura 4.1.

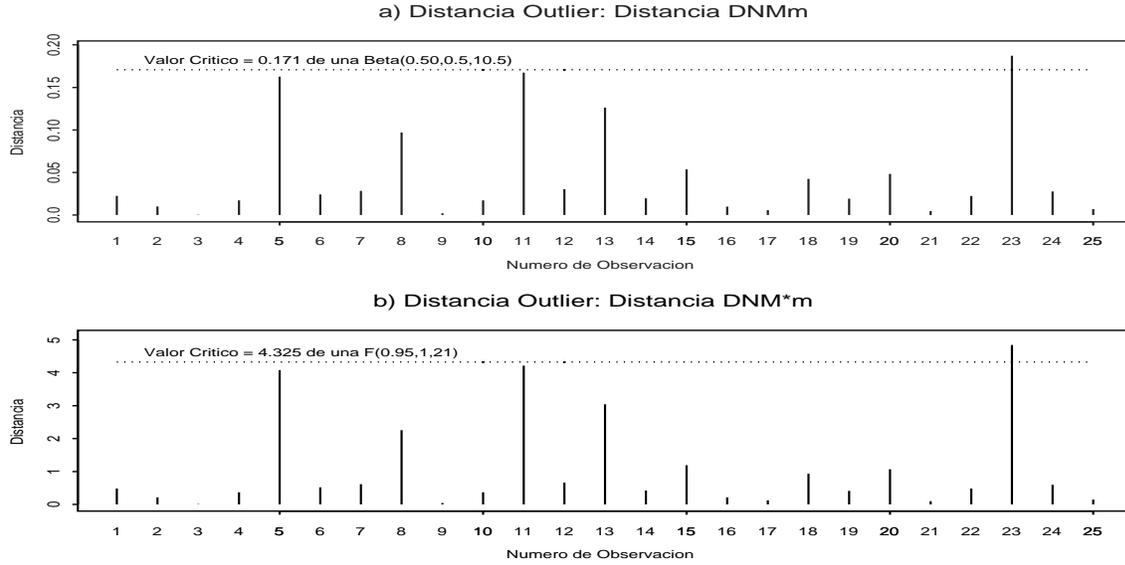


Figura 4.1: Identificación de observaciones influyentes basada en a) la distancia DNM_{m_i} y b) la distancia $DNM_{m_i}^*$.

La Figura 4.1 muestra la identificación y detección de la observación 23 como influyente en la combinación lineal $\hat{\beta}M$, dado que de las pruebas obtenidas en la sección 3.4 del capítulo 3 se tiene

$$\frac{DNM_{m_{23}}}{n - q} = \frac{4.121}{25 - 3} = 0.187 \geq \beta_{\alpha:s/2, (n-q-s)/2} = \beta_{0.95:1/2, (25-3-1)/2} = 0.171,$$

y

$$\frac{(n - q - s)DNM_{m_{23}}^*}{s(n - q - 1)} = \frac{(25 - 3 - 1)(4.840)}{1(25 - 3 - 1)} = 4.840 \geq F_{\alpha:s, (n-q-s)} = F_{0.95:1, (25-3-1)} = 4.325.$$

Así se puede concluir que la observación 23 es influyente en la en la combinación lineal $\hat{\beta}M$. Note que el valor crítico en la Figura 4.1(a) es obtenido de una distribución Beta mientras que en la Figura 4.1(b) es obtenido de una distribución \mathcal{F} . La principal diferencia entre las Figuras 4.1(a) y 4.1(b) es el estimador de la varianza que se utilizo. Para (a) se calculó el MSE (error cuadrático medio) de todas las observaciones y para (b) fue eliminada la i -ésima observación.

4.3 Ejemplo dos: consumo de fósforo en ratas

Considere la cantidad de alimento que ingirieron 32 ratas, asignadas aleatoriamente a 4 grupos, al rededor de 12 días. El alimento es de una dieta en tratamiento, la cual contiene diferentes proporciones de fósforo. La Tabla 4.2 muestra los datos; en ella se tiene que G = grupo al que se aplico una proporción de fósforo (siendo las proporciones 0.25%, 0.65%, 1.3% y 1.71%, para los grupos I, II, III y IV, respectivamente), R = rata sometida a la dieta y PI = peso inicial.

El modelo a considerar es

$$y_i = \beta_{i0} + \beta_{i1}x + \beta_{i2}x^2 + \beta_{i3}x^3 + \beta_{i4}w + \varepsilon_i \quad i = 1, \dots, 12 ,$$

donde x representa la cantidad de fósforo dada y w es la covariable peso inicial.

En notación matricial el modelo es

$$Y = X\beta + \varepsilon,$$

donde $Y \in \mathbb{R}^{32 \times 12}$,

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{15} \\ y_{21} & \cdots & y_{25} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{n5} \end{pmatrix},$$

es la matriz respuesta, con $n = 32$ y y_{ij} = cantidad de comida consumida por la rata i en el día j ; $X \in \mathbb{R}^{n \times 5}$ es la matriz de covariables, su primera columna está formada por unos correspondientes al intercepto, las columnas 2, 3 y 4 corresponden x , x^2 y x^3 , respectivamente y la última columna corresponde a w ; por último, $\beta \in \mathbb{R}^{5 \times 12}$,

$$\beta = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{41} & \beta_{42} & \cdots & \beta_{4q} \end{pmatrix},$$

es la matriz de parámetros, siendo sus renglones los vectores de parámetros correspondientes al intercepto, x , x^2 , x^3 y w , respectivamente.

Table 4.2: Dieta de fósforo para un grupo de 32 ratas.

G	R	PI	Días											
			1	2	3	4	5	6	7	8	9	10	11	12
I	1	254	12.7	10.4	5.1	8.6	7.1	9.7	3.0	4.9	3.8	5.5	6.3	5.3
I	2	262	7.2	8.4	8.5	6.8	6.3	4.5	7.9	7.0	7.7	8.7	7.1	11.4
I	3	301	14.8	13.9	8.4	7.3	8.0	10.4	6.0	7.4	7.2	10.5	12.7	8.9
I	4	311	5.6	10.2	7.8	6.1	6.4	16.5	16.7	12.8	11.6	15.9	10.6	4.5
I	5	290	13.9	12.1	8.8	8.8	8.1	7.8	4.8	6.5	4.1	7.1	7.0	7.6
I	6	300	10.4	11.2	12.5	7.0	6.9	6.9	6.0	7.3	1.2	9.0	10.8	8.1
I	7	306	16.6	17.8	14.0	6.8	5.9	5.3	3.7	2.7	1.0	7.7	7.4	11.5
I	8	286	13.9	14.3	5.9	7.7	9.2	5.7	7.3	6.1	9.2	9.7	7.3	9.0
II	1	275	11.9	7.0	5.9	6.1	0.8	5.1	4.2	9.7	10.1	7.7	15.3	10.9
II	2	282	10.7	11.3	4.4	3.9	4.7	5.3	6.6	8.1	9.9	8.8	11.4	10.1
II	3	256	10.1	6.9	7.8	6.4	9.5	7.9	5.7	9.8	4.7	5.9	4.5	7.6
II	4	276	10.8	5.2	1.3	1.3	2.1	6.3	6.6	8.9	12.9	15.0	13.0	8.6
II	5	337	14.7	14.4	11.6	7.4	7.8	14.8	9.7	9.9	8.9	15.5	11.0	5.0
II	6	296	9.7	12.1	5.2	9.1	9.7	5.2	8.8	6.5	7.3	5.7	6.0	12.2
II	7	309	5.5	7.1	7.8	3.1	1.5	8.4	12.6	13.9	4.2	11.1	16.0	8.9
II	8	296	13.1	6.5	1.3	0.9	0.8	0.5	0.9	10.7	13.4	7.9	9.8	12.9
III	1	275	8.8	17.7	11.5	6.6	5.4	12.0	2.3	0.6	1.0	7.0	9.8	4.8
III	2	292	8.3	3.2	5.2	8.9	4.3	4.4	8.1	11.8	7.3	7.0	10.3	9.5
III	3	338	16.2	11.9	10.2	15.6	15.3	13.9	5.6	2.0	0.0	0.8	3.6	8.9
III	4	248	7.7	4.9	11.7	12.7	13.2	10.7	9.0	7.5	0.7	1.4	0.4	0.3
III	5	315	14.5	14.0	16.9	8.4	13.1	9.8	6.6	6.0	9.7	9.8	5.1	6.0
III	6	295	11.6	2.5	5.5	4.5	5.8	8.6	6.0	16.4	9.9	8.8	9.4	8.4
III	7	312	5.3	6.1	1.5	4.1	6.2	2.1	11.7	14.7	13.4	7.0	4.2	13.3
III	8	286	11.2	11.0	5.7	8.1	10.0	8.1	11.2	12.1	9.2	6.6	8.9	10.4
IV	1	275	13.5	9.7	12.3	13.4	14.0	6.1	3.3	1.2	1.9	1.3	2.8	5.8
IV	2	270	11.6	2.4	9.7	14.0	10.8	10.3	5.7	15.5	3.8	1.8	6.3	1.5
IV	3	290	10.0	14.8	9.1	9.6	8.2	9.3	7.2	4.1	7.0	3.4	6.5	1.3
IV	4	260	12.3	16.2	6.6	9.2	8.3	12.6	8.1	4.8	7.9	8.2	9.9	5.2
IV	5	302	13.6	14.9	9.3	10.2	11.5	15.8	2.7	5.6	2.7	4.3	3.0	1.3
IV	6	284	12.8	13.2	11.6	11.5	11.1	10.5	6.2	6.4	2.5	4.1	4.9	4.0
IV	7	280	10.9	14.3	10.8	9.6	13.2	10.0	6.0	2.2	2.0	7.0	4.1	4.2
IV	8	329	8.3	10.5	7.5	10.6	8.5	6.2	13.1	4.1	5.1	6.2	8.3	6.0

El estimador de verosimilitud máxima $\hat{\beta}$ es

$$\begin{pmatrix} 5.23 & -2.40 & 0.43 & 7.52 & 0.09 & -1.70 & -1.54 & 8.56 & 3.31 & 2.44 & 8.77 & 5.39 \\ -1.54 & -5.38 & -2.05 & -1.17 & -0.66 & -1.26 & 1.03 & 5.16 & 3.44 & -0.66 & 0.67 & 3.00 \\ -1.73 & -3.08 & -1.26 & -1.54 & -0.16 & -0.43 & 0.87 & 2.00 & 1.60 & -1.06 & -1.04 & 0.90 \\ 1.43 & 3.44 & 1.74 & 2.18 & 1.28 & 1.13 & -0.80 & -3.03 & -2.44 & -0.25 & -0.44 & -2.35 \\ 0.02 & 0.05 & 0.03 & 0.00 & 0.02 & 0.03 & 0.03 & -0.01 & 0.01 & 0.03 & 0.00 & 0.01 \end{pmatrix}.$$

Supongamos que estamos interesados en contrastar la cantidad de comida consumida en el día 1 y la comida total consumida en los días 3 y 4. Tal contraste se puede efectuar estimando las funciones paramétricas

$$\begin{aligned} \beta_{10} - (\beta_{12} + \beta_{13}) \\ \beta_{20} - (\beta_{22} + \beta_{23}) \\ \beta_{30} - (\beta_{32} + \beta_{33}) , \\ \beta_{40} - (\beta_{42} + \beta_{43}) \\ \beta_{50} - (\beta_{52} + \beta_{53}) \end{aligned}$$

o equivalentemente βM , donde $M = \begin{pmatrix} 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}'$.

Entonces, sería de interés ver si alguna observación o conjunto de observaciones es influyente en la combinación lineal

$$\begin{aligned} \hat{\beta}_{10} - (\hat{\beta}_{12} + \hat{\beta}_{13}) \\ \hat{\beta}_{20} - (\hat{\beta}_{22} + \hat{\beta}_{23}) \\ \hat{\beta}_{30} - (\hat{\beta}_{32} + \hat{\beta}_{33}) . \\ \hat{\beta}_{40} - (\hat{\beta}_{42} + \hat{\beta}_{43}) \\ \hat{\beta}_{50} - (\hat{\beta}_{52} + \hat{\beta}_{53}) \end{aligned}$$

Análogamente al ejemplo uno, se calculan las distancias DNM_{m_i} y $DNM_{m_i}^*$ utilizando los elementos diagonales de la matriz de predicción P , la matriz de residuales $\hat{\varepsilon}$ y el estimador de la matriz de varianzas y covarianzas (S ó $S_{(i)}$).

La Figura 4.2 muestra las distancias DNM_{m_i} y $DNM_{m_i}^*$, las cuales son útiles para detectar la influencia de la i -ésima observación en la combinación lineal $\hat{\beta}M$. Las pruebas están en favor de declarar a las observaciones 12, 16 y 20 como observaciones, que individualmente, tienen una fuerte influencia en $\hat{\beta}M$.

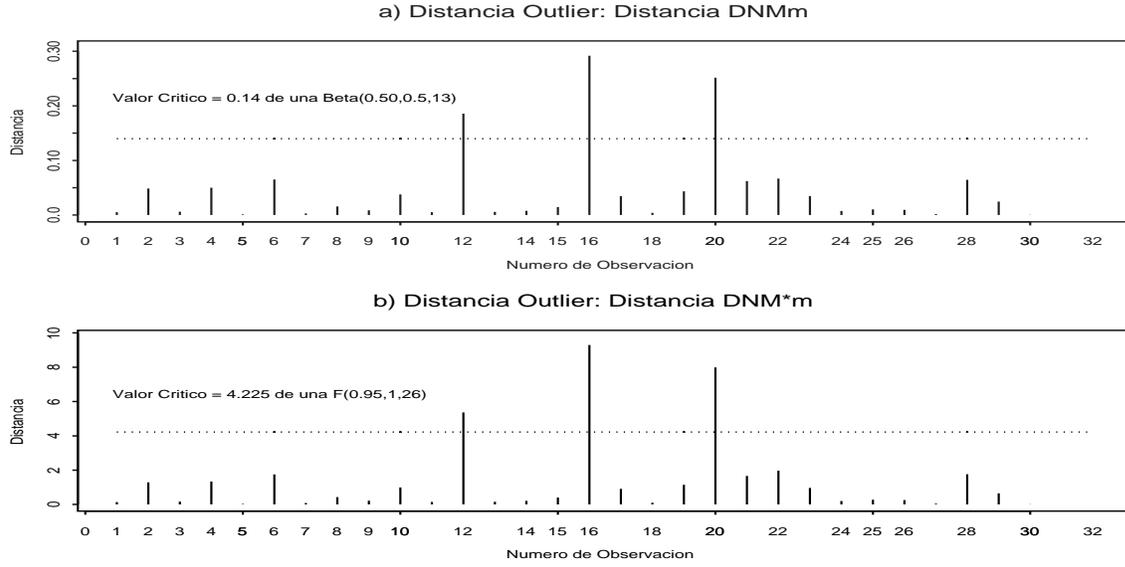


Figura 4.2: Identificación de observaciones influyentes basada en a) la distancia DNM_{m_i} y b) la distancia $DNM_{m_i}^*$.

Es posible que exista un efecto de ocultamiento, i.e. que conjuntamente las observaciones 12, 16 y 20 no tengan una fuerte influencia en la combinación lineal $\hat{\beta}M$. Por ello, las métricas obtenidas en el Teorema 6 del capítulo 4, fueron usadas para probar si las observaciones 12, 16 y 20 son conjuntamente influyentes en la combinación lineal $\hat{\beta}M$. La Tabla 4.3 resume los resultados obtenidos. Para este ejemplo se tiene

$$P_I = \begin{pmatrix} 0.065 & 0.055 & 0.065 \\ 0.055 & 0.060 & 0.037 \\ 0.065 & 0.037 & 0.144 \end{pmatrix},$$

$$\hat{\varepsilon}_I = \begin{pmatrix} 0.2 & -3.6 & -5.5 & -4.9 & -3.9 & -0.8 & -0.3 & -0.4 & 5.3 & 6.4 & 3.7 & -0.6 \\ 2.0 & -3.4 & -6.1 & -5.2 & -5.6 & -7.2 & -6.5 & 1.6 & 5.6 & -1.2 & 0.5 & 3.5 \\ -1.8 & -1.6 & 5.3 & 5.0 & 6.0 & 3.8 & 2.7 & -2.1 & -6.3 & -4.3 & -7.1 & -7.3 \end{pmatrix},$$

$$S = \begin{pmatrix} 9.5 & 4.2 & 2.9 & 2.8 & 3.3 & 1.0 & -7.8 & -4.7 & -2.3 & -2.4 & -1.7 & 0.2 \\ 4.2 & 15.4 & 4.9 & 0.7 & 2.2 & 3.8 & -3.6 & -11.6 & -3.7 & 0.7 & -0.4 & -0.9 \\ 2.9 & 4.9 & 11.9 & 4.2 & 4.9 & 4.4 & -2.6 & -5.6 & -8.4 & -1.3 & -2.2 & -4.6 \\ 2.8 & 0.7 & 4.2 & 7.3 & 6.3 & 3.1 & -0.6 & -3.5 & -6.1 & -5.1 & -4.0 & -2.2 \\ 3.3 & 2.2 & 4.9 & 6.3 & 9.7 & 3.5 & -0.9 & -4.4 & -5.1 & -4.1 & -7.5 & -2.6 \\ 1.0 & 3.8 & 4.4 & 3.1 & 3.5 & 13.6 & 1.4 & -0.7 & -3.8 & 2.8 & 0.5 & -8.3 \\ -7.8 & -3.6 & -2.6 & -0.6 & -0.9 & 1.4 & 12.5 & 6.6 & 4.5 & 4.9 & 2.6 & -0.5 \\ -4.7 & -11.6 & -5.6 & -3.5 & -4.4 & -0.7 & 6.6 & 16.9 & 7.8 & 3.9 & 3.9 & 0.5 \\ -2.3 & -3.7 & -8.4 & -6.1 & -5.1 & -3.8 & 4.5 & 7.8 & 14.9 & 6.8 & 3.8 & 3.5 \\ -2.4 & 0.7 & -1.3 & -5.1 & -4.1 & 2.8 & 4.9 & 3.9 & 6.8 & 10.0 & 6.0 & -0.9 \\ -1.7 & -0.4 & -2.2 & -4.0 & -7.5 & 0.5 & 2.6 & 3.9 & 3.8 & 6.0 & 10.8 & 1.3 \\ 0.2 & -0.9 & -4.6 & -2.2 & -2.6 & -8.3 & -0.5 & 0.5 & 3.5 & -0.9 & 1.3 & 8.4 \end{pmatrix},$$

y

$$S_I = \begin{pmatrix} 10.2 & 4.9 & 4.3 & 4.0 & 4.7 & 2.1 & -7.9 & -5.5 & -3.7 & -3.1 & -2.4 & -0.8 \\ 4.9 & 16.3 & 4.7 & 0.1 & 2.1 & 3.7 & -4.7 & -13.2 & -3.4 & 1.0 & -0.9 & -1.4 \\ 4.3 & 4.7 & 10.6 & 2.2 & 3.0 & 2.7 & -5.0 & -5.8 & -6.3 & 0.1 & -1.1 & -3.0 \\ 4.0 & 0.1 & 2.2 & 5.9 & 4.9 & 1.5 & -2.5 & -3.5 & -4.0 & -4.3 & -3.1 & -0.5 \\ 4.7 & 2.1 & 3.0 & 4.9 & 8.6 & 1.8 & -2.9 & -4.5 & -3.0 & -3.4 & -7.0 & -0.6 \\ 2.1 & 3.7 & 2.7 & 1.5 & 1.8 & 12.8 & -0.8 & -0.1 & -1.9 & 3.3 & 1.3 & -7.1 \\ -7.9 & -4.7 & -5.0 & -2.5 & -2.9 & -0.8 & 12.0 & 7.9 & 7.2 & 5.6 & 3.5 & 1.3 \\ -5.5 & -13.2 & -5.8 & -3.5 & -4.5 & -0.1 & 7.9 & 18.8 & 8.3 & 4.4 & 4.0 & -0.3 \\ -3.7 & -3.4 & -6.3 & -4.0 & -3.0 & -1.9 & 7.2 & 8.3 & 13.3 & 5.9 & 2.5 & 1.3 \\ -3.1 & 1.0 & 0.1 & -4.3 & -3.4 & 3.3 & 5.6 & 4.4 & 5.9 & 9.0 & 5.1 & -2.0 \\ -2.4 & -0.9 & -1.1 & -3.1 & -7.0 & 1.3 & 3.5 & 4.0 & 2.5 & 5.1 & 10.1 & -0.5 \\ -0.8 & -1.4 & -3.0 & -0.5 & -0.6 & -7.1 & 1.3 & -0.3 & 1.3 & -2.0 & -0.5 & 6.4 \end{pmatrix}.$$

Dado que el estadístico de prueba, en los dos casos, es mayor que el valor crítico α , entonces ambas pruebas están a favor de declarar que las observaciones 12, 16, y 20 son conjuntamente influyentes en la combinación lineal $\hat{\beta}M$.

Table 4.3: Dos métricas para detectar un conjunto de observaciones influyentes en la combinación lineal $\widehat{\beta}M$.

Métrica	Estadístico de prueba	Valor Crítico $1 - \alpha = 0.95$
$DNM_{m_I} = \text{tr}(M'SM)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_I M = 0.663964$	17.78285	2.960351
$DNM_{m_I}^* = \text{tr}(M'S_{(I)}M)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_I M = 1.337669$	12.03902	2.960351

4.4 Ejemplo tres: estudio dental realizado en niños y niñas

Los datos de la Tabla 4.4 fueron obtenidos de un estudio dental y presentados en Potthoff y Roy (1964). Los datos, consisten en cuatro mediciones de la distancia (en milímetros) del centro de la pituitaria a la fisura maxilar realizadas en edades de 8, 10, 12 y 14 años en 11 niñas y 16 niños.

El modelo a considerar es

$$Y = X\beta + \varepsilon,$$

donde $Y \in \mathbb{R}^{27 \times 4}$,

$$Y = \begin{pmatrix} y_{11t_1} & \cdots & y_{11t_4} \\ \vdots & & \vdots \\ y_{n_1 1 t_1} & \cdots & y_{n_1 1 t_4} \\ y_{12t_1} & \cdots & y_{12t_4} \\ \vdots & & \vdots \\ y_{n_2 2 t_1} & \cdots & y_{n_2 2 t_4} \end{pmatrix},$$

es la matriz respuesta, donde $n_1 = 11$, $n_2 = 16$ y y_{igt} = distancia de la pituitaria a la fisura maxilar del i individuo, del grupo g (niñas=1 o niños=2) al tiempo t , con tiempos $t = (t_1, t_2, t_3, t_4) = (8, 10, 12, 14)$.

$X \in \mathbb{R}^{27 \times 3}$ es la matriz de covariables, su primer columna está formada por unos correspondientes al intercepto, las dos columnas restantes tienen unos y ceros los cuales indican a que grupo g pertenece el individuo i . La matriz de parámetros $\beta \in \mathbb{R}^{3 \times 5}$ es

Table 4.4: distancia (en milímetros) del centro de la pituitaria a la fisura maxilar realizadas en 11 niñas y 16 niños.

Niñas					Niños				
Edad en años					Edad en años				
Obs.	8	10	12	14	Obs.	8	10	12	14
1	21.0	20.0	21.5	23.0	1	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	2	21.5	22.5	23.0	26.5
3	20.5	24.0	24.5	26.0	3	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	4	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	5	20.0	23.5	22.5	26.0
6	20.0	21.0	21.0	22.5	6	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	7	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	8	24.0	21.5	24.5	25.5
9	20.0	21.0	22.0	21.5	9	23.0	20.5	31.0	26.0
10	16.5	19.0	19.0	19.5	10	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	11	23.0	23.0	23.5	25.0
					12	21.5	23.5	24.0	28.0
					13	17.0	24.5	26.0	29.5
					14	22.5	25.5	25.5	26
					15	23.0	24.5	26.0	30
					16	22.0	21.5	23.5	25

$$\beta = \begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} & \beta_{04} \\ \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{pmatrix},$$

siendo sus renglones los vectores de parámetros correspondientes al intercepto, y a la contribución de niñas y niños en la distancia de la pituitaria a la fisura maxilar.

El estimador de verosimilitud máxima de β es

$$\hat{\beta} = X^{-1}Y = \begin{pmatrix} 14.69 & 15.35 & 16.27 & 17.19 \\ 6.50 & 6.88 & 6.82 & 6.90 \\ 8.19 & 8.47 & 9.45 & 10.28 \end{pmatrix}.$$

Un problema, que surge de manera natural, es ver si existe diferencia significativa en la distancia de la pituitaria a la fisura maxilar en los diferentes tiempos; esto se puede investigar estimando las funciones

paramétricas

$$\begin{pmatrix} \beta_{01} - \beta_{02} & \beta_{02} - \beta_{03} & \beta_{03} - \beta_{04} \\ \beta_{11} - \beta_{12} & \beta_{12} - \beta_{13} & \beta_{13} - \beta_{14} \\ \beta_{21} - \beta_{22} & \beta_{22} - \beta_{23} & \beta_{23} - \beta_{24} \end{pmatrix},$$

la primera columna contrastaría las distancia para las edades 8 y 10, la segunda contrastaría las edades 10 y 12, y la última las edades 12 y 14. Podemos expresar esta matriz como βM donde

$$M = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Con el fin de ejemplificar las herramientas desarrolladas en esta tesis, veamos si una observación o conjunto de observaciones es influyente en la combinación lineal

$$\begin{pmatrix} \hat{\beta}_{01} - \hat{\beta}_{02} & \hat{\beta}_{02} - \hat{\beta}_{03} & \hat{\beta}_{03} - \hat{\beta}_{04} \\ \hat{\beta}_{11} - \hat{\beta}_{12} & \hat{\beta}_{12} - \hat{\beta}_{13} & \hat{\beta}_{13} - \hat{\beta}_{14} \\ \hat{\beta}_{21} - \hat{\beta}_{22} & \hat{\beta}_{22} - \hat{\beta}_{23} & \hat{\beta}_{23} - \hat{\beta}_{24} \end{pmatrix},$$

esto es muy importante para la estimación de las funciones paramétricas βM . Por tanto, se procedió a detectar tales observaciones.

Los elementos diagonales de la matriz de predicción P son 0.09 y 0.06 correspondientes a las niñas y niños, respectivamente. El estimador de la matriz de varianzas y covarianzas S , el estimador de la matriz de varianzas y covarianza al eliminar la observación 2, $S_{(2)}$, y la matriz de residuales son

$$S = \begin{pmatrix} 5.6 & 2.8 & 4.1 & 2.8 \\ 2.8 & 4.4 & 3.0 & 3.5 \\ 4.1 & 3.0 & 6.7 & 4.3 \\ 2.8 & 3.5 & 4.3 & 5.2 \end{pmatrix}, \quad S_{(2)} = \begin{pmatrix} 5.9 & 2.9 & 4.3 & 3.0 \\ 2.9 & 4.5 & 3.2 & 3.7 \\ 4.3 & 3.2 & 7.0 & 4.4 \\ 3.0 & 3.7 & 4.4 & 5.3 \end{pmatrix} \quad \text{y}$$

$$\hat{\varepsilon} = \begin{pmatrix} -0.18 & -2.23 & -1.59 & -1.09 \\ -0.18 & -0.73 & 0.91 & 1.41 \\ -0.68 & 1.77 & 1.41 & 1.91 \\ 2.32 & 2.27 & 1.91 & 2.41 \\ 0.32 & 0.77 & -0.59 & -0.59 \\ -1.18 & -1.23 & -2.09 & -1.59 \\ 0.32 & 0.27 & -0.09 & 0.91 \\ 1.82 & 0.77 & 0.41 & -0.09 \\ -1.18 & -1.23 & -1.09 & -2.59 \\ -4.68 & -3.23 & -4.09 & -4.59 \\ 3.32 & 2.77 & 4.91 & 3.91 \\ 3.13 & 1.19 & 3.28 & 3.53 \\ -1.37 & -1.31 & -2.72 & -0.97 \\ 0.13 & -1.31 & -1.72 & 0.03 \\ 2.63 & 3.69 & 0.78 & -0.47 \\ -2.87 & -0.31 & -3.22 & -1.47 \\ 1.63 & 1.69 & 1.28 & 1.03 \\ -0.87 & -1.81 & -1.22 & -0.97 \\ 1.13 & -2.31 & -1.22 & -1.97 \\ 0.13 & -3.31 & 5.28 & -1.47 \\ 4.63 & 4.19 & 5.28 & 4.03 \\ 0.13 & -0.81 & -2.22 & -2.47 \\ -1.37 & -0.31 & -1.72 & 0.53 \\ -5.87 & 0.69 & 0.28 & 2.03 \\ -0.37 & 1.69 & -0.22 & -1.47 \\ 0.13 & 0.69 & 0.28 & 2.53 \\ -0.87 & -2.31 & -2.22 & -2.47 \end{pmatrix} .$$

La Figura 4.3 muestra las distancias modificadas, DNM_{m_i} y $DNM_{m_i}^*$, para detectar observaciones influyentes en la combinación lineal $\hat{\beta}M$. Se observa que las pruebas están a favor de que las observaciones 20 y 24, individualmente, tienen una fuerte influencia en la combinación lineal. La observación 15 influye en la combinación lineal βM , pero las pruebas no declaran a esta observación como influyente. Es posible que se este presentando un efecto de ocultamiento, i.e que la observación 15 individualmente

no es influyente pero si lo es en presencia de las observaciones 20 y 24.

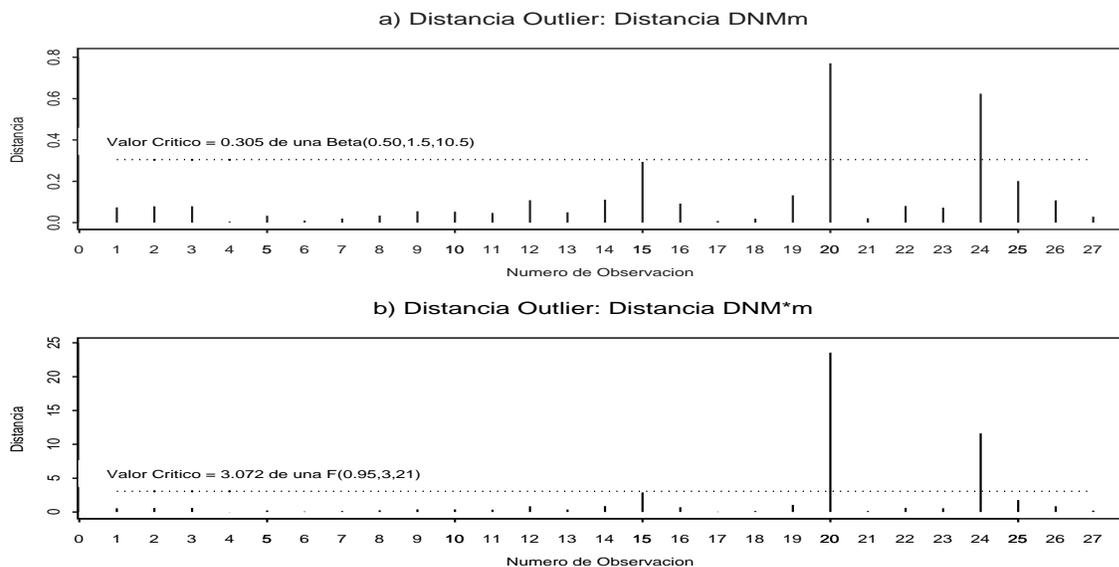


Figura 4.3: Identificación de observaciones influyentes basada en a) la distancia DNM_{m_i} y b) la distancia $DNM^*_{m_i}$.

Se utilizaron las métricas obtenidas en el Teorema 6 del capítulo 4 para probar si las observaciones 15, 20 y 24 son conjuntamente influyentes en la combinación lineal $\hat{\beta}M$. La Tabla 4.5 resume los resultados obtenidos. Dado que el estadístico de prueba, en los dos casos, es mayor que el valor crítico α , entonces ambas pruebas están a favor de declarar que las observaciones 15, 20 y 24 son conjuntamente influyentes en la combinación lineal $\hat{\beta}M$. Esto verifica que, efectivamente, existe un efecto de ocultamiento de la influencia de la observación 15 en la combinación lineal $\hat{\beta}M$.

Table 4.5: Dos métricas para detectar un conjunto de observaciones influyentes en la combinación lineal $\widehat{\beta}M$.

Métrica	Estadístico de prueba	Valor Crítico $1 - \alpha = 0.95$
$DNM_{m_I} = \text{tr}(M'SM)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_I M = 1.677703$	10.15023	2.012705
$DNM^*_{m_I} = \text{tr}(M'S_{(I)}M)^{-1}M'\widehat{\varepsilon}'_I(I_k - P_I)^{-1}\widehat{\varepsilon}_I M = 5.854709$	13.44415	2.034774

Capítulo 5

Conclusiones

De acuerdo a lo presentado en esta tesis se puede concluir que las distancias DNM_{m_i} , $DNM_{m_i}^*$, DNM_{m_I} y $DNM_{m_I}^*$ nos permiten establecer reglas estadísticas exactas, para establecer si una observación o un conjunto de observaciones es influyente en funciones lineales de los parámetros β , de la forma $N\beta M$.

El punto crítico para decidir si una observación es influyente en la función lineal $N\beta M$, esta determinado por los percentiles de una distribución β o de una distribución \mathcal{F} , por lo que en la actualidad no hay ningún problema para obtener este valor. En cambio, se tienen algunos problemas para obtener el punto crítico para determinar si un conjunto de observaciones es influyente, en la función lineal $N\beta M$. Este valor esta determinado por los percentiles de una distribución Lawley-Hotelling o de una distribución Pillai, los cuales están tabulados para algunos valores de sus parámetros. Cuando se requieren valores que no están tabulados, entonces el punto crítico estará determinado por una aproximación a una distribución \mathcal{F} .

Por otro lado, las distribuciones exactas de las distancias modificadas, deducidas en este trabajo, son las mismas bajo el supuesto que los errores sigan una distribución de contornos elípticos. Esto suugiere que las distancias definidas para medir la influencia de ciertos puntos son robustas ante una gran variedad de familias simétricas de distribución , incluyendo distribuciones de colas pesadas.

Con respecto a los ejemplos presentados podemos concluir que: frecuentemente una observación o un

conjunto de observaciones es influyente en las combinaciones lineales $\widehat{\beta}M$, y detectar estas observaciones es muy importante para la estimación de las funciones paramétricas βM .

Como continuidad de este trabajo, se podría estudiar la invarianza bajo combinaciones lineales de la forma $N\beta$ de la distancia propuesta por Díaz-García y González-Farías (2004). Esto debido a que las distancias obtenidas, en este trabajo, no dependen de la matriz N .

Apéndice A

Elementos del Álgebra Lineal

A.1 Vectores Unitarios y Matrices Elementales

El vector unitario, de orden n se denota por e_i^n y se define como un vector de ceros y un uno en la i -ésima posición. Así,

$$e_i^n = \begin{pmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \\ 1 & 2 & \cdots & i & \cdots & n \end{pmatrix}$$

La matriz elemental $E_{ij} \in \mathbb{R}^{m \times n}$, se define como la matriz que tiene un uno en la ij -ésima posición y todos sus demás elementos son ceros.

1. La matriz elemental puede ser expresada como

$$E_{ij} = e_i^m e_j^n.$$

2. La matriz identidad $I \in \mathbb{R}^{n \times n}$ está dada por

$$I = \sum_{i=1}^n E_{ii} = \sum_{i=1}^n e_i^n e_i^n.$$

3. Cualquier matriz $A \in \mathbb{R}^{m \times n}$, puede ser expresada como

$$A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} E_{ij} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} e_i^m e_j^n.$$

A.2 Traza

La traza de una matriz $A \in \mathbb{R}^{m \times m}$, se define como la suma de los elementos de la diagonal y se denota por $\text{tr}(A)$, es decir,

$$\text{tr}(A) = \sum_{i=1}^m a_{ii}.$$

Sean $A, B \in \mathbb{R}^{n \times m}$ y $\alpha, \beta \in \mathbb{R}$. Entonces,

1. $\text{tr}(\alpha A + \beta B) = \alpha \text{tr} A + \beta \text{tr} B$, para $m = n$.
2. $\text{tr} A' = \text{tr} A$, si $m = n$.
3. $\text{tr} A'B = \text{tr} B'A = \text{tr} BA'$.
4. Si A es cuadrada. Entonces,

$$\text{tr}(A) = \sum_i a_{ii} = \sum_i e_i^{n'} A e_i^n.$$

A.3 Matriz Inversa

Sea $A \in \mathbb{R}^{p \times p}$ y $\det(A) \neq 0$. Entonces, existe una única matriz $B \in \mathbb{R}^{p \times p}$ tal que $AB = I_p$ y B es llamada la inversa de A y se denota por A^{-1} .

Teorema A.1. [Teorema de Sherman-Morrison-Woodbury] Sean $A \in \mathbb{R}^{k \times k}$ y $C \in \mathbb{R}^{m \times m}$ matrices no singulares. Además, $B \in \mathbb{R}^{k \times m}$ y $D \in \mathbb{R}^{m \times k}$. Entonces, verificando que la inversa existe,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

Demostración. Se verifica multiplicando $(A+BCD)$ por el lado izquierdo de la igualdad, observándose que tal multiplicación es igual a la matriz identidad. ■

A.4 Partición de Matrices

Una matriz no singular $A \in \mathbb{R}^{p \times p}$ se dice que esta particionada en submatrices si, para $i, j = 1, 2$, A puede escribirse de la forma

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

donde $A_{11} \in \mathbb{R}^{m \times q}$, $A_{12} \in \mathbb{R}^{m \times p-q}$, $A_{21} \in \mathbb{R}^{n-m \times q}$ y $A_{22} \in \mathbb{R}^{n-m \times p-q}$.

Teorema A.2. Sea A particionada como

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

a) Si A y A_{11} son no singulares, entonces

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}MA_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}M \\ -MA_{21}A_{11}^{-1} & M \end{pmatrix},$$

$$\text{donde } M = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}.$$

b) Si A y A_{22} son no singulares, entonces

$$A^{-1} = \begin{pmatrix} N & -NA_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{12}N & A_{22}^{-1} + A_{22}^{-1}A_{21}NA_{12}A_{22}^{-1} \end{pmatrix},$$

$$\text{donde } N = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

Demostración. La prueba se obtiene mostrando que $AA^{-1} = A^{-1}A = I$. ■

A.5 Inversa Generalizada

Sea $A \in \mathbb{R}^{n \times p}$. Entonces, existe una matriz $B \in \mathbb{R}^{p \times n}$ que cumple con

$$\begin{aligned} i) AB &= (AB)', & iii) ABA &= A, \\ ii) BA &= (BA)', & iv) BAB &= B, \end{aligned}$$

tal que B es llamada la matriz inversa generalizada (o de Moore-Penrose) de A , se denota por $B = A^{-}$.

Sea $A \in \mathbb{R}^{n \times p}$. Así,

1. A^{-} es única.
2. Si $r(A) = n$, entonces $A^{-} = A'(AA')^{-1}$.
3. Si $r(A) = p$, entonces $A^{-} = (A'A)^{-1}A$.
4. Si $r(A) = n = p$, entonces $A^{-} = A^{-1}$.
5. $(A^{-})^{-} = A$.
6. $(A')^{-} = (A^{-})'$.
7. $(A'A)^{-} = A^{-}(A^{-})'$.
8. $A^{-} = (A'A)^{-}A' = A'(AA')^{-}$.
9. Si A es una proyección ortogonal, entonces $A^{-} = A$.
10. AA^{-} y $A^{-}A$ son proyecciones ortogonales.

A.6 Producto Kronecker

Considere las matrices $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ y $B \in \mathbb{R}^{p \times q}$. El producto Kronecker de dos matrices denotado por $A \otimes B$, es la matriz de orden $pm \times qn$, definida como

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix} = (a_{ij}B).$$

Sean A, B, C y D matrices de órdenes adecuados y sean $\alpha, \beta \in \mathbb{R}$. Entonces,

1. $A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C)$.
2. $(A + B) \otimes (C + D) = A \otimes C + B \otimes C + A \otimes D + B \otimes D$.
3. $(A \otimes B)(C \otimes D) = AC \otimes BD$.
4. $\alpha \otimes A = \alpha A = A\alpha = A \otimes \alpha$.
5. $(A \otimes B)' = A' \otimes B'$.
6. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, si A y B son no singulares.
7. $(A \otimes B)^- = A^- \otimes B^-$.
8. Si $A \in \mathbb{R}^{n \times n}$, con raíces características $\lambda_1, \dots, \lambda_n$, y $B \in \mathbb{R}^{p \times p}$, con raíces características $\delta_1, \dots, \delta_p$, entonces las raíces características de $A \otimes B$, están dadas por

$$\lambda_i \delta_j \quad i = 1, \dots, n, \quad j = 1, \dots, p.$$

Teorema A.3. Sean $A \in \mathbb{R}^{n \times n}$ y $B \in \mathbb{R}^{p \times p}$. Entonces,

9. $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$.
10. $\det(A \otimes B) = (\det(A))^p (\det(B))^n$.

Demostración. De A.6.8, se tiene que

$$9. \text{tr}(A \otimes B) = \sum_i \sum_j \lambda_i \delta_j = (\sum_i \lambda_i)(\sum_j \delta_j) = \text{tr}(A) \text{tr}(B).$$

$$10. \det(A \otimes B) = \prod_i \prod_j \lambda_i \delta_j = \prod_i \lambda_i^p \prod_j \delta_j = (\det(A))^p (\det(B))^n. \quad \blacksquare$$

A.7 Vectorización

Sea $A \in \mathbb{R}^{m \times n}$, tal que $A = (A_1, \dots, A_n)$. Entonces la vectorización de A es

$$\text{vec}(A) = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \in \mathbb{R}^{mn}.$$

1. Si a y b son vectores, entonces

$$\text{vec}(ab') = \text{vec}(a \otimes b') = \text{vec}(b' \otimes a) = b \otimes a.$$

Teorema A.4. Si A y B son matrices cuadradas de orden n , entonces

$$2. \text{tr}(A'B) = (\text{vec } A)' \text{vec } B.$$

Demostración. $\text{tr}(A'B) = \sum_i e_i^{n'}(A'B)e_i^n = \sum_i A_i' B_i = (\text{vec } A)' \text{vec } B.$ ■

Teorema A.5. Si A , B y C son de ordenes adecuados, entonces

$$3. \text{vec}(ABC) = (C' \otimes A) \text{vec } B.$$

Demostración.

$$\begin{aligned} \text{vec}(ABC) &= \text{vec} \left(A \left(\sum_j B_j e_j^{n'} \right) C \right) = \text{vec} \left(\sum_j AB_j e_j^{n'} C \right) \\ &= \sum_j \text{vec}(AB_j)(C' e_j^n)' = \sum_j (C' e_j^n \otimes AB_j) \\ &= \sum_j (C' \otimes A)(e_j^n \otimes B_j) = (C' \otimes A) \text{vec } B. \end{aligned}$$

■

Teorema A.6. Sean A , B , C y D matrices, tales que, el producto $ABCD$ está definido y es cuadrado.

Entonces

$$4. \text{tr } ABCD = (\text{vec } D)'(C' \otimes A) \text{vec } B = (\text{vec } D)'(A \otimes C') \text{vec } B'.$$

Demostración.

$$\text{tr } ABCD = \text{tr } DABC = (\text{vec } D)' \text{vec}(ABC) = (\text{vec } D)'(C' \otimes A) \text{vec } B.$$

Análogamente,

$$\text{tr } ABCD = \text{tr } D'C'B'A' = (\text{vec } D)' \text{vec}(C'B'A') = (\text{vec } D)'(A \otimes C') \text{vec } B'.$$

■

Apéndice B

Estadística Multivariada

B.1 Distribución Normal Multivariada

La matriz aleatoria $X \in \mathbb{R}^{n \times p}$ tiene una distribución normal multivariada, con media M y matriz de varianzas y covarianzas $C \otimes D$, si la función de densidad de X es

$$(2\pi)^{-np/2} (\det C)^{-n/2} (\det D)^{-p/2} \exp \left\{ -\frac{1}{2} \text{tr} [(X - M)' C^{-1} (X - M) D^{-1}] \right\}, \quad (\text{B.1})$$

donde $M \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{p \times p}$ y $D \in \mathbb{R}^{n \times n}$ (ambas definidas positivas), tal que $E(X) = M$ y $\text{Cov}(\text{vec}(X)) = C \otimes D$. Se denotará $X \sim \mathcal{N}_{n \times p}(M, C \otimes D)$.

1. Sean $A \in \mathbb{R}^{r \times n}$ y $B \in \mathbb{R}^{p \times s}$ matrices constantes. Si $X \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I)$ y $AA' = I_r$, entonces

$$Y = AXB \sim \mathcal{N}_{r \times s}(0, B' \Sigma B \otimes I_r).$$

B.2 Distribución Wishart

Si $Z = X'X$, donde $X \in \mathbb{R}^{n \times p}$ tal que $X \sim \mathcal{N}_{n \times p}(0, \Sigma \otimes I)$, entonces Z se dice que tiene una distribución Wishart con n grados de libertad (gl) y parámetros Σ . Se denotará $Z \sim \mathcal{W}_p(\Sigma, n)$. Cuando $n < p$, Z es singular y la distribución $\mathcal{W}_p(\Sigma, n)$ no tiene una función de densidad en el hipércono en $\mathbb{R}^{p(p+1)/2}$ definido por las matrices $Z > 0$ (matrices definidas positivas), pero tiene densidad en un subespacio de

este hipércono . Cuando $n \geq p$ la función de densidad de Z esta dada por

$$\frac{1}{2^{pn/2} \Gamma_p \left(\frac{1}{2}n \right) (\det \Sigma)^{n/2}} (\det Z)^{(n-p-1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} Z \right\}, \quad (\text{B.2})$$

donde $\Gamma_p(\cdot)$ denota la función gamma multivariada.

1. Si $Z \sim \mathcal{W}_p(\Sigma, n)$ y $B \in \mathbb{R}^{p \times q}$ es una matriz constante, entonces

$$B' Z B \sim \mathcal{W}_q(B' Z B, n).$$

2. Si $Z_1 \sim \mathcal{W}_p(\Sigma, n_1)$ y $Z_2 \sim \mathcal{W}_p(\Sigma, n_2)$, y si Z_1 y Z_2 son independientes, entonces

$$Z_1 + Z_2 \sim \mathcal{W}_p(\Sigma, n_1 + n_2).$$

B.3 Distribución Wishart no Centrada

Si $Z = X'X$, donde $X \in \mathbb{R}^{n \times p}$ tal que $X \sim \mathcal{N}_{n \times p}(\mu, \Sigma \otimes I)$, entonces Z se dice que tiene una distribución Wishart no centrada con n gl, parámetros Σ , y matriz de parámetros de no centralidad $\Omega = \Sigma^{-1} \mu' \mu$. Y se denotará como $A \sim \mathcal{W}_p(\Sigma, n, \Omega)$.

Note que cuando $\mu = 0$, se tiene que $\Omega = 0$, $Z \sim \mathcal{W}_p(\Sigma, n)$. Cuando $n < p$, Z es singular y la distribución $\mathcal{W}_p(\Sigma, n, \Omega)$ no tiene una función de densidad, más precisamente no tiene una función de densidad en el hipércono en $\mathbb{R}^{p(p+1)/2}$ definido por las matrices $Z > 0$, pero tiene densidad en un subespacio de este hipércono . Cuando $n \geq p$ la función de densidad de Z esta dada por

$$\frac{1}{2^{pn/2} \Gamma_p \left(\frac{1}{2}n \right) (\det \Sigma)^{n/2}} (\det Z)^{(n-p-1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \Sigma^{-1} Z \right\} \exp \left\{ -\frac{1}{2} \Omega \right\} {}_0F_1 \left(\frac{1}{2}n; \frac{1}{4} \Omega \Sigma^{-1} Z \right), \quad (\text{B.3})$$

donde $Z > 0$, $\Omega = \Sigma^{-1} \mu' \mu$ y en general

$${}_pF_q(a_1, \dots, a_r; b_1, \dots, b_s, X) = \sum_{k=0}^{\infty} \sum_{\lambda} \frac{(a_1)_{\lambda} \dots (a_r)_{\lambda}}{(b_1)_{\lambda} \dots (b_s)_{\lambda}} \frac{C_{\lambda}(X)}{k!}, \quad X = X'$$

define las funciones hipergeométricas con argumentos matriciales y \sum_{λ} denota la suma sobre toda partición $\lambda = (k_1, \dots, k_m)$, $k_1 \geq \dots \geq k_m \geq 0$ de k , $C_{\lambda}(X)$ es el polinomio zonal de X correspondiente a λ

y $(a)_\lambda$ son los coeficientes hipergeométricos generalizados

$$(a)_\lambda = \prod_{i=1}^m \left(a - \frac{1}{2}(i-1)\right)_k$$

$$(a)_k = a(a+1)\dots(a+k-1), \quad (a)_0 = 1.$$

B.4 Distribución T^2 de Hotelling

Si α puede ser escrita como $nX'Z^{-1}X$ donde X y Z son independientemente distribuidas $\mathcal{N}_p(0, I)$ y $\mathcal{W}_p(I, n)$, respectivamente, entonces se dice que α tiene una distribución T^2 de Hotelling con parámetros p y n . Se denotará $\alpha \sim T^2(p, n)$.

1. Si X y Z son independientemente distribuidas $\mathcal{N}_p(\mu, \Sigma)$ y $\mathcal{W}_p(\Sigma, n)$, respectivamente, entonces

$$n(X - \mu)'Z^{-1}(X - \mu) \sim T^2(p, n).$$

2.

$$T^2(s, n - q - 1) = \frac{s(n - q - 1)}{n - q - s} \mathcal{F}_{(s, n - q - s)},$$

donde $\mathcal{F}_{(s, n - q - s)}$ es la distribución \mathcal{F} centrada con $s, n - q - s$ gl.

3. Si X y Z son independientemente distribuidas $\mathcal{N}_p(0, I)$ y $\mathcal{W}_p(I, n)$, respectivamente, entonces

$$X'X \left(1 + \frac{1}{X'Z^{-1}X}\right) \sim \chi_{n+1}^2,$$

donde χ_{n+1}^2 es la distribución ji-cuadrada centrada con $n + 1$ gl., y es distribuida independientemente de $X'Z^{-1}X$.

B.5 Distribución Beta Multivariada

La distribución Beta Multivariada es una distribución que está cercanamente relacionada a la distribución Wishart. El siguiente Teorema, cuya demostración es dada en Muirhead (1982), pone en manifiesto la relación que guardan tales distribuciones.

Teorema B.1. Sean A y B matrices aleatorias independientes, donde $A \sim \mathcal{W}_m(\Sigma, n_1)$ y $B \sim \mathcal{W}_m(\Sigma, n_2)$, con $n_1 > m - 1$, $n_2 > m - 1$. Asumiendo que $A + B = T'T$ donde $T \in \mathbb{R}^{m \times m}$ triangular superior con

elementos diagonales positivos. Sea $U \in \mathbb{R}^{m \times m}$ simétrica definida por $A = T'UT$. Entonces, $A + B$ y U son independientes; $A + B$ tiene distribución $\mathcal{W}_m(\Sigma, n_1 + n_2)$ y la función de densidad de U es

$$\frac{\Gamma_m \left[\frac{1}{2}(n_1 + n_2) \right]}{\Gamma_m \left(\frac{1}{2}n_1 \right) \Gamma_m \left(\frac{1}{2}n_2 \right)} (\det U)^{(n_1 - m - 1)/2} (\det(I_m - U))^{(n_2 - m - 1)/2}, \quad (\text{B.4})$$

con $0 < U < I_m$, lo cual indica que $U > 0$ (i.e., U es definida positiva) y $I_m - U > 0$.

Así, una matriz U con función de densidad (B.4) se dice que tiene una distribución beta multivariada con parámetros $\frac{1}{2}n_1$ y $\frac{1}{2}n_2$ y se denota escribiendo $U \sim \beta_m(\frac{1}{2}n_1, \frac{1}{2}n_2)$, es obvio que si $U \sim \beta_m(\frac{1}{2}n_1, \frac{1}{2}n_2)$, entonces $(I - U) \sim \beta_m(\frac{1}{2}n_2, \frac{1}{2}n_1)$.

1. Si $U \sim \beta_m(\frac{1}{2}n_1, \frac{1}{2}n_2)$ y $U = T'T$, donde T es triangular superior entonces t_{11}, \dots, t_{mm} son todas independientes y $t_{ii}^2 \sim \text{beta}[\frac{1}{2}(n_1 - i + 1), \frac{1}{2}n_2]$; $i = 1, \dots, m$.
2. Si $U \sim \beta_m(\frac{1}{2}n_1, \frac{1}{2}n_2)$ la función de densidad conjunta de las raíces características u_1, \dots, u_m de U es

$$\frac{\pi^{m^2/2}}{\Gamma_m \left(\frac{1}{2}m \right) \Gamma_m \left(\frac{1}{2}n_1 \right) \Gamma_m \left(\frac{1}{2}n_2 \right)} \frac{\Gamma_m \left[\frac{1}{2}(n_1 + n_2) \right]}{\prod_{i=1}^m \left[u_i^{(n_1 - m - 1)/2} (1 - u_i)^{(n_2 - m - 1)/2} \right]} \cdot \prod_{i < j}^m (u_i - u_j) \quad (1 > u_1 > \dots > u_m > 0).$$

B.6 Estadísticos Traza

En términos de las raíces características $\lambda_1, \dots, \lambda_m$ de AB^{-1} , con A y B independientes, tal que $A \sim \mathcal{W}_m(\Sigma, r, \Omega)$ y $B \sim \mathcal{W}_m(\Sigma, n - p \geq m)$, se tienen tres estadísticos importantes:

B.6.1 Estadístico Lawley-Hotelling

Si $A \sim \mathcal{W}_m(\Sigma, r, \Omega)$, $B \sim \mathcal{W}_m(\Sigma, n - p \geq m)$, y A y B son independientes ($n - p \geq m, r \geq m$) entonces la función de densidad de $U^{(m)} = \text{tr} AB^{-1} = \sum_{i=1}^m \lambda_i$ puede ser expresada como

$$f_{U^{(m)}}(x) = \frac{\exp \left\{ -\frac{1}{2} \text{tr} \Omega \right\} \Gamma_m \left[\frac{1}{2}(r + n - p) \right]}{\Gamma_m \left[\frac{1}{2}(n - p) \right] \Gamma \left(\frac{1}{2}mr \right)} x^{mr/2 - 1} \cdot \sum_{k=0}^{\infty} \frac{(-x)^k}{\left(\frac{1}{2}mr \right)_k k!} \sum_{\kappa} \left[\frac{1}{2}(r + n - p) \right]_{\kappa} L_{\kappa}^{\gamma} \left(\frac{1}{2}\Omega \right), \quad (\text{B.5})$$

donde $\gamma = \frac{1}{2}(r - m - 1)$ y $L_\kappa^\beta(X)$ denota el polinomio Laguerre generalizado de la matriz simétrica X correspondiente a la partición κ de k .

B.6.2 Estadístico Pillai

El estadístico

$$V^{(m)} = \text{tr } A(A + B)^{-1} = \sum_{i=1}^m \frac{\lambda_i}{1 + \lambda_i}$$

fue sugerido por Pillai(1955, 1956). Su distribución no-nula sobre el rango $0 < V^{(m)} < 1$ ha sido encontrada por Khatri y Pillai (1968) como una complicada serie de polinomio zonal.

B.6.3 Estadístico de Roy

Si $A \sim \mathcal{W}_m(\Sigma, r, \Omega)$, $B \sim \mathcal{W}_m(\Sigma, n - p \geq m)$ con $(n - p \geq m, r \geq m)$, y A y B son independientes y si $t = \frac{1}{2}(n - p - m - 1)$ es un entero, entonces la función de densidad de λ_1 , la raíz más grande de AB^{-1} , puede ser expresada como

$$P(\lambda_1 \leq x) = \left(\frac{x}{1+x} \right)^{mr/2} \exp \left\{ -\frac{1}{2(1+x)} \text{tr } \Omega \right\} \sum_{k=0}^{mt} \sum_{\kappa}^* \frac{L_\kappa^\gamma \left(-\frac{1}{2} \frac{x}{1+x} \Omega \right)}{(1+x)^k k!}, \quad (\text{B.6})$$

donde $\gamma = \frac{1}{2}(r - m - 1)$ y \sum_{κ}^* denota la suma sobre todas las particiones $\kappa = (k_1, \dots, k_m)$ de k con la parte más grande $k_1 \leq t$.

B.7 Resultados de Ellenberg

Considere el modelo lineal general

$$Y = X\beta + u, \quad (\text{B.7})$$

donde $Y \in \mathbb{R}^{n \times 1}$ es el vector respuesta; $X \in \mathbb{R}^{n \times k}$, con $r(X) = k$, $\beta \in \mathbb{R}^k$, son la matriz de regresión y el vector de parámetros, respectivamente; y $u \in \mathbb{R}^n$ es el vector de errores, tal que $u \sim \mathcal{N}_{n \times p}(0, \sigma^2 I_n)$ donde σ^2 es desconocido (notación retomada de Ellenberg(1973)).

Así, se tiene que los estimadores mínimos cuadrados de los parámetros en el modelo (B.7) son:

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

$$\hat{u} = (Y - X\hat{\beta}) = Mu,$$

donde

$M = I_n - X(X'X)^{-1}X'$, y M es idempotente de rango $n - k$,

El estimador de la suma de cuadrados del error para el modelo esta dado por $S^2 = \hat{u}'\hat{u} = u'Mu$.

Lema B.2. [Ellenberg(1973)] *Asuma que hay n observaciones del modelo lineal general (B.7) y que los estimadores mínimos cuadrados han sido calculados. Considere un p -subconjunto de observaciones, digamos las primeras p , con sus correspondiente vector de errores estimado \hat{u}_p y la partición de M*

$$M = \begin{pmatrix} M_p & M_a \\ M'_a & M_{n-p} \end{pmatrix},$$

con $M_p \in \mathbb{R}^{p \times p}$, $M_a \in \mathbb{R}^{p \times n-p}$ y $M_{n-p} \in \mathbb{R}^{n-p \times n-p}$. Entonces si M_p^{-1} existe, la forma cuadrática

$$S_p^2 = S^2 - \hat{u}'_p M_p^{-1} \hat{u}_p \tag{B.8}$$

se distribuirá $\sigma^2 \chi^2(n - k - p)$.

Lema B.3. [Ellenberg (1973)] *Dadas las suposiciones del Lema 1, el p -subconjunto \hat{u}_p es independiente de S_p^2 .*

De forma natural estos resultados son generalizados para el modelo lineal multivariado.

B.8 Distribución de Contornos Elípticos

Sea $X \in \mathbb{R}^{p \times n}$ una matriz aleatoria. Entonces, se dice que tiene una distribución de contornos elípticos si su función característica tiene la forma $\phi_X(T) = \exp\{\text{tr } iT'M\}\psi(\text{tr}(T'\Sigma T\Phi))$, donde, $T, M \in \mathbb{R}^{p \times n}$, $\Sigma \in \mathbb{R}^{p \times p}$, $\Phi \in \mathbb{R}^{n \times n}$, $\Sigma > 0$, $\Phi > 0$ y $\psi : [0, \infty) \rightarrow \mathbb{R}$ y se denotará por

$$X \sim \mathcal{E}_{p,n}(M, \Sigma \otimes \Phi, \psi). \tag{B.9}$$

Teorema B.4. [Gupta y Varga (1993)] *Sea $X \sim E_{p,n}(0, \Sigma \otimes \Phi, \psi)$ con $P(X = 0) = 0$. Asuma que $Y \sim N_{p \times n}(0, \Sigma \otimes \Phi)$. Sea \mathcal{F} un subconjunto de matrices reales $p \times n$, tal que si $Z \in \mathbb{R}^{p \times n}$, $Z \in \mathcal{F}$, y $a > 0$ entonces $aZ \in \mathcal{F}$ y $P(X \notin \mathcal{F}) = P(Y \notin \mathcal{F}) = 0$. Sea $K(Z)$ una función definida en \mathcal{F} , si $Z \in \mathcal{F}$ y $a > 0$, entonces $K(Z) = K(aZ)$. Entonces, $K(X)$ y $K(Y)$ están definidas con probabilidad uno y $K(X)$ y $K(Y)$ son idénticamente distribuidas.*

Apéndice C

Programa

El siguiente programa fue desarrollado en S-plus. Se realizó para aplicar las pruebas exactas de detección de outliers, en funciones lineales de los parámetros β , de la forma $N\beta M$. Para llevarlo a cabo se utilizó el ejemplo de Potthoff y Roy (1964).

```
# Potthoff y Roy (1964). A generalized multivariate analysis of variance model
# useful especially for growth curve problems.
# Biometrika, núm. 51, P.p 313-326.
# Los datos, consisten en cuatro mediciones de la distancia (en milímetros)
# del centro de la pituitaria a la fisura maxilar realizadas en edades de
# 8, 10, 12 y 14 años en 11 niñas y 16 niños.

girls <- matrix( c(
21.0, 20.0, 21.5, 23.0, 21.0, 21.5, 24.0, 25.5, 20.5, 24.0, 24.5, 26.0,
23.5, 24.5, 25.0, 26.5, 21.5, 23.0, 22.5, 23.5, 20.0, 21.0, 21.0, 22.5,
21.5, 22.5, 23.0, 25.0, 23.0, 23.0, 23.5, 24.0, 20.0, 21.0, 22.0, 21.5,
16.5, 19.0, 19.0, 19.5, 24.5, 25.0, 28.0, 28.0), ncol=4, byrow=T)

boys <- matrix( c(
26.0, 25.0, 29.0, 31.0, 21.5, 22.5, 23.0, 26.5, 23.0, 22.5, 24.0, 27.5,
25.5, 27.5, 26.5, 27.0, 20.0, 23.5, 22.5, 26.0, 24.5, 25.5, 27.0, 28.5,
```

```

22.0, 22.0, 24.5, 26.5, 24.0, 21.5, 24.5, 25.5, 23.0, 20.5, 31.0, 26.0,
27.5, 28.0, 31.0, 31.5, 23.0, 23.0, 23.5, 25.0, 21.5, 23.5, 24.0, 28.0,
17.0, 24.5, 26.0, 29.5, 22.5, 25.5, 25.5, 26.0, 23.0, 24.5, 26.0, 30.0,
22.0, 21.5, 23.5, 25.0), ncol=4, byrow=T)

p<- 4
q<- 3
n<- 27

Y<- rbind(girls, boys)      # Matriz respuesta
X<- matrix(0,n,3)
X[,1]<- -1
X[1:11,2]<- -1
X[12:27,3]<- -1           # Matriz de regresión

# Matriz de combinaciones lineales por columnas
M<- matrix(c( 1, 0, 0, -1, 1, 0, 0,-1, 1, 0, 0,-1), ncol=3, byrow=TRUE)

n <- dim(Y)[1]      # Número de renglones de X y Y
k <- 1              # Número de observaciones a quitar
p <- dim(Y)[2]      # Número de columnas de Y
q <- dim(X)[2]      # Número de columnas de X
s <- dim(M)[2]      # Número de columnas de M

bg<- ginverse(t(X)%*%X)%*%t(X)%*%Y      # Estimador de beta
yg<- X%*%bg                               # Estimador de Y
s2<- (t(Y)%*(diag(n)- X%*%ginverse(X))%*%Y)/(n-q) # Estimador de  $\sigma^2$ 
itMs2M<- solve(t(M)%*%s2%*%M)           # Inversa de la matriz tMs2M
hii<- diag(X%*%ginverse(X))              # Diagonal del Proyector Ortogonal
eg<- Y - yg                               # Estimador de los errores

# Cálculos de la distancia que tiene distribución  $T^2$  cF

```

```

# Se hará una verificación de las observaciones i=1,2,...,n

DNMm <- DNMme <- rep(0,n)
ni <- n-1

for(i in 1:n)
{
yi <- Y[-i, ]
xi <- X[-i, ]
s22 <- (t(yi)%*(diag(ni)- xi%*ginverse(xi))%*yi)/(ni-q)
itMs22M <- solve(t(M)%*s22%M)

ei <- as.matrix(eg[i,])

DNMm[i] <- (1/(1-hii[i]))*(t(t(M)%*ei)%*itMs22M%(t(M)%*ei))
DNMme[i] <- (1/(1-hii[i]))*(t(t(M)%*ei)%*itMs22M%(t(M)%*ei))
}

DNMm <- DNMm/(n-q)
DNMme <- (n-q-s)*DNMme/(s*(n-q-1))
par(mfrow=c(2,1))

# Gráfica de la distancia DNMm
Dist <- DNMm
alpha <- (1-0.05)
vp <- qbeta(alpha,s/2,(n-q-s)/2)

ym <- max(c(Dist,vp))*1.05
plot(1:n,Dist, main="a) Distancia Outlier: Distancia DNMm",
xlab="Número de Observación", ylab="Distancia", type="h",ylim=c(0,ym))
axis(1,1:n,ticks=F,line=0)
leg <- paste("Valor Crítico = ",round(vp,3)," de una Beta(0.50,",s/2,",

```

```

", (n-q-s)/2, ")", sep="")
legend(0, vp, leg, bty="n")
lines(1:n, rep(vp, n), lty=2, lwd=2)

# Gráfica de la distancia DNmM modificada
Dist <- DNmMe
vp <- qf(alpha, s, n-q-s)

ym <- max(c(Dist, vp))*1.05
plot(1:n, Dist, main="b) Distancia Outlier: Distancia DNmM",
     xlab="Número de Observación", ylab="Distancia", type="h", ylim=c(0, ym))
axis(1, 1:n, ticks=F, line=0)
leg <- paste("Valor Crítico = ", round(vp, 3), " de una F(", round(alpha, 3), " ",
            s, " ", n-q-s, ")", sep="")
legend(0, vp, leg, bty="n")
lines(1:n, rep(vp, n), lty=2, lwd=2)

#####
## Quitando Varias Observaciones ##
#####

#Datos Outliers
cuales <- c(15, 20, 24)

k <- length(cuales)
XII <- as.matrix(X[cuales,])
HI <- XII%*%ginverse(t(X)%*%X)%*%t(XII)
eI <- as.matrix(eg[cuales,])
NDMmI <- sum(diag(itMs2M)%*%t(eI)%*%M)%*%solve(diag(k)-HI)%*%(eI)%*%M))
NDMmI <- NDMmI/(n-q)

NDMmI

```

```

### Calculando ahora  $s^2$  sin las  $ij$ -ésimas observaciones
cuales <- c(15, 20, 24)
ni <- n-3
XX <- X[-cuales,]
YY <- Y[-cuales,]
s2 <- (t(YY)%*(diag(ni)- XX%*ginverse(XX))%*YY)/(ni-q)
is22 <- solve(t(M)%*s2%M)

XII <- as.matrix(X[cuales,])
HI <- XII%*ginverse(t(X)%*X)%*t(XII)
eI <- as.matrix(eg[cuales,])
NDMmIe <- sum(diag(is22)%*t(eI)%*M)%*solve(diag(k)-HI)%*(eI)%*M))
NDMmIe <- NDMmIe/(n-q-k)

NDMmIe

### Regla de decisión de Pillai
me <- n-q mh <- k
n1 <- 0.5*(abs(s-mh)-1) n2 <- 0.5*(me-s-1) w <- min(s,mh)

EstF <- (2*n2+w+1)*NDMmI/((2*n1+w+1)*(w-NDMmI))
EstF
tabF <- qf(0.95,w*(2*n1+w+1),w*(2*n2+w+1))
tabF

### Regla de decisión de Lawley-Hotelling
f1 <- w*(2*n1 + w + 1)
f2 <- 2*(w*n2 + 1)

EstF <- f2*NDMmIe/(f1*w)

```

EstF

```
tabF <- qf(0.95,f1,f2)
```

tabF

Bibliografía

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

Atkinson, A. C. (1981). "Two Graphical Displays for Outlying and Influential Observations in Regression", *Biometrika*, **68**, 13-20.

Atkinson, A. C. (1985). *Plots, Transformations and Regresión: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford Science Publications, New York.

Besley, D., Kuh, E., and Welsch, R. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York.

Caroni, C. (1987). "Residuals and Influence in the Multivariate Linear Model", *The Statistician*, **36**, 365-370.

Cook, R. D. (1977). "Detection of Influential Observations in Linear Regression", *Technometrics*, **19**, 15-18.

Cook, R. D., and Weisberg, S. (1982). *Residual and Influence in Regression*, Chapman and Hall, London.

Cook, R. D., and Weisberg, S. (1994). *An Introduction to Regression Graphics*, John Wiley & Sons, New York.

Chatterjee, S., and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York.

Chatterjee, S., Hadi, A. S., and Price, B. (2000). *Regression Analysis by Example* (3rd ed.), John Wiley & Sons, New York.

Díaz-García, J. A., Galea, M., and Leiva-Sánchez, V. (2001). "Influence Diagnostics for Elliptical Regresión Linear Models", *Suministrada para publicación*.

Díaz-García, J. A. and González-Farías, G. (2004). "A Note on the Cook's Distance", *Journal of Statistical Planning and Inference*, **1/2**, 119-136.

Díaz-García, J. A. and Gutiérrez-Jáimes, R. (2001). "The Expected Value of Zonal Polynomials", *Test*, **10**, 133-145.

Draper, N., and Smith, H. (1981). *Applied Regression Analysis*, (2nd ed.), John Wiley & Sons, New York.

Ellenberg, J. H. (1973). "The Joint Distribution of the Standardized Least Squares Residual from General Linear Regression", *Journal of American Statistical Association*, **68**, 941-943.

Gupta, A. K., and Varga, T. (1993). *Elliptically Contoured Models in Statistics*, Kluwer Academic Publishers, Dordrecht.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Hossain, A., and Naik, D. N. (1989). "Detection of Influential Observations in Multivariate Regression", *Journal of Applied Statistics*, **16**, 25-37.

Khatri, C. G., and Pillai, K. C. S. (1968). "On the Noncentral Distributions of two Test Criteria in Multivariate Analysis of Variance", *Annals of Mathematical Statistics*, **39**, 215-226.

Mardia, K. V. Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press Limited, London.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York.

Olkin, I., and Rubin, H. (1964). "Multivariate Beta Distribution and Independence Properties of Wishart Distributions", *Annals of Mathematical Statistics*, **35**, 261-269.

Potthoff, R. F. and Roy, S. N. (1964). "A generalized multivariate analysis of variance model useful especially for growth curve problems", *Biometrika*, **51**, 313-326.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications* (2nd ed.), John Wiley & Sons, New York.

Rencher, A. C. (1995). *Methods of Multivariate Analysis*, John Wiley & Sons, New York.

Seber, G. A. F. (1984). *Multivariate Observations*, John Wiley & Sons, New York.

Srivastava, M. S. (1968). "On the Distributions of a Multiple Correlation Matrix: Noncentral Multivariate Beta Distributions", *Annals of Mathematical Statistics*, **39**, 227-232.

Srivastava, M. S. (2002). *Methods of Multivariate Statistics*, John Wiley & Sons, New York.

Srivastava, M. S., and Carter (1983). *An Introduction to Applied Multivariate Statistics*, North-Holland Publ., New York.

Srivastava, M. S., and Khatri, C. G. (1979). *An Introduction to Multivariate Analysis*, North-Holland Publ., New York.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.