# UNIVERSIDAD NACIONAL DE INGENIERIA

## FACULTAD DE INGENIERIA ECONOMICA Y CIENCIAS SOCIALES

# EFICIENCIA DE LOS ESTIMADORES SESGADOS EN REGRESION, EN PRESENCIA DE MULTICOLINEALIDAD

**TESIS**

**PARA OPTAR EL TITULO DE LICENCIADO EN ESTADISTICA**

**JUAN ANIBAL PARIASCA LUCIANO**

**LIMA – PERU**
**1999**

# EFICIENCIA DE LOS ESTIMADORES SESGADOS EN REGRESION, EN PRESENCIA DE MULTICOLINEALIDAD

CONTENIDO

**CAPITULO IV –APLICACIÓN VIA SIMULACION DE LA EFICIENCIA DE LOS ESTIMADORES SESGADOS, EN PRESENCIA DE MULTICOLINEALIDAD**

## SUMARIO

Este trabajo presenta la eficiencia de varios estimadores sesgados, para un modelo de regresión líneal Múltiple, cuando existe el problema de multicolinealidad. También se hace una revisión de los aspectos que involucra un problema de multicolinealidad (origen, causas, métodos para la detección y solución del problemna).

Los resultados de la simulación aplicada, revela que los procedimientos sesgados, tienen una mejor eficiencia en cuanto al error cuadratico medio y varianza, respecto al estimador tradicional de mínimos cuadrados ordinarios

# I N T R O D U C C I O N

El tema de interés de la tesis es: " EFICIENCIA DE LOS ESTIMADORES SESGADOS EN REGRESION, EN PRESENCIA DE MULTICOLINEALIDAD Lo que se intenta aqui es, cuando se presenta el fenómeno de MULTICOLINEALIDAD o dependencia lineal de las variables regresoras en un modelo de regresión lineal múltiple, los coeficientes de estas variables, se pueden estimar a través de distintos tipos de estimadores sesgados. y luego se optimiza cuál de estos estimadores es el más eficiente.

Las interrogantes que se plantean (de los tantos que existen) en esta investigación son:

- El estimador insesgado usual de Mínimos Cuadrados Ordinarios (MCO), estima bien los coeficientes del modelo de regresión lineal, en presencia de multicolinealidad?.
- Cuáles son los origenes, consecuencias y técnicas para la detección de multicolinealidad, en el modelo de regresión lineal
- Los estimadores sesgados, son más eficientes que el estimador de MCO, en presencia de multicolinealidad en el modelo de regresión lineal?. Y entre estos cuál es el más eficiente y que propiedades tienen?.
- Es posible desarrollar una estructura general para todos los estimadores presentados y además, estos estimadores sesgados, ayudan a un mejor prónostico de la variable dependiente en el modelo de regresión lineal en presencia de multicolinealidad?.

El fenómeno de multicolinealidad es un problema de tipo muestral y como tal, el presente trabajo de tesis, intenta responder las interrogantes planteadas arriba, lo que permitirá un mejor análisis y prónostico de un conjunto de datos de interés. Pues, los resultados y conclusiones que se obtengan en presencia de este fenómeno en un modelo de regresión lineal, seran equivocadas y la aplicación de ellos, genera incertidumbre. Por consiguiente lo que se intenta es:

- Presentar la naturaleza, consecuencias y técnicas de detección de multicolinealidad en el análisis de datos en un modelo de Regresión Lineal Múltiple general.
- Presentar una variedad de estimadores sesgados, con sus propiedades correspondientes, alternativos al estimador insesgado de MCO, con el objeto de ver la eficiencia de ellos respecto al estimador usual de MCO, en presencia del fenómeno de multicolinealidad.
- Conocer la existencia o no, del estimador sesgado más eficiente entre los estimadores presentados, en presencia de multicolinealidad. así como también uniformizar en una sola estructura general a todos los estimadores presentados.
- Presentar una aplicación del proyecto de tesis "Eficiencia de los estimadores sesgados en regresión, en presencia de multicolinealidad", a través de los datos obtenidos por simulación.

Por consiguiente, este trabajo de tesis es trascendente, desde el punto de vista de las estimaciones y prónosticos futuros con presencia de multicolinealidad.

En la literatura estadística no existen trabajos en forma amplia que detallen, a la vez, los origenes, consecuencias y métodos para detectar el fenómeno de multicolinealidad. El aporte de este trabajo es por lo tanto, la presentación de una variedad de estimadores sesgados alternativos al

MCO, indicando entre ellos, cuál es el más eficiente, el mismo que será de mucha utilidad en las aplicaciones del modelo de regresión lineal múltiple.

Las fuentes de información que sirvió de apoyo para desarrollar la tesis fuerón, al nivel secundario, de textos y revistas científicas de estadística y al nivel primario de un diseño muestral, el cual sirve para hacer la simulación del problema de interés. Este trabajo es de carácter científico y aplicativo, por lo tanto se intenta en ello discutir, comparar, evaluar y resumir, la eficiencia de los estimadores sesgados, en presencia de multicolinealidad en el modelo de regresión lineal múltiple, siendo además la variable estudiada de tipo cuantitativo continuo.

La organización de la presente tesis es como sigue, el capítulo I, presenta una revisión de literatura respecto a la estructura general del modelo de regresión lineal múltiple, supuestos clásicos de este, y el estimador de MCO. el Capítulo II busca explicar la naturaleza de la multicolinealidad, se examinan sus consecuencias y sugieren métodos para detectarla; el Capítulo III, presenta comparaciónes de los métodos de estimación sesgada, alternativos al MCO, de los coeficientes verdaderos del modelo (1.3), cuando existe multicolinealidad; también incluye propiedades importantes de cada uno de ellos y una formulación general de la estructura de todos los estimadores investigados. El Capítulo IV, es dedicado a la comparación, de la eficiencia de los estimadores sesgados a través de un estudio de aplicación. La aplicación es realizada a través de una simulación, la cual es diseñada en este trabajo y llevada a cabo con la ayuda de un PROGRAMA preparado en TURBO PASCAL Versión 7.0. La simulación diseñada consideró las condiciones siguientes:

i) La orientación de $\beta$ con un subconjunto de vectores latentes (autovectores) de $X'X$ (matriz de correlación para datos multicolineales).

ii) La magnitud de $\beta$ relativo a $\sigma$ ( $\sigma^2$ varianza del error "$e_i$" )

iii) La fuerza o potencia de multicolinealidad.

El Capítulo V, trata de las conclusiones y recomendaciones finales del trabajo de tesis, luego en la parte final se incluye las referencias bibliográficas y apéndice respectivamente. Se espera que esta tesis sea una contribución importante en el estudio de los modelos de regresión lineal, ya que el proposito es resumir, discutir, comparar y evaluar la eficiencia de los estimadores sesgados con respecto al estimador clásico de MCO, cuando existe multicolinealidad.

# CAPITULO I

# PRELIMINARES

En este capítulo se presentan el modelo de regresión lineal múltiple, los supuestos clásicos y la estimación de los coeficientes verdaderos por el método de mínimos cuadrados ordinarios (MCO).

## 1.1 ESTRUCTURA DEL MODELO DE REGRESION LINEAL MULTIPLE

**Definicion 1.-**

El modelo general de regresión lineal múltiple es definido por

$$Y_i = \beta_0^* + \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \beta_3^* X_{i3}^* + \ldots + \beta_p^* X_{ip}^* + e_i$$

$$i = 1, 2, \ldots, n$$

ó en notación matricial

$$Y = \beta_0^* \times \mathbf{1} + X^* \beta^* + e \qquad (1.1)$$

Donde

$Y$ Es un vector columna $n\times 1$ de la variable dependiente ó variable respuesta.

$\beta_0^*$ : Parámetro verdadero de intercepto.

$\mathbf{1}$ : Vector columna $n\times 1$ de unos.

$X^*$ :Matriz $n\times p$ de variables independientes ó constantes, conocidas también como variables predictoras, regresoras ó variables explicatorias, y es de rango completo $p \leq n$.

$\beta^*$: Vector columna p×1 de parámetros verdaderos del modelo.

e   Vector n×1 de perturbaciones ó error no observables.

n   Número total de observaciones (tamaño muestral).

p   Número total de variables regresoras.

Es bastante dificil comparar directamente los coeficientes del modelo de regresión; pues las magnitudes de los parámetros verdaderos $\beta^*_j$ reflejan las unidades de medida de las variables regresoras $X_j$. Por esta razón, es conveniente trabajar con las variables regresoras y variable respuesta en la misma escala, para que produzcan dimensiones menores de los coeficientes del modelo de regresión. Esto se hace estandarizando el modelo (1.1).

**definicion 2.-**

Sean;

$$X_{ij} = \frac{X^*_{ij} - \bar{X}^*_j}{d_j} , \quad \begin{array}{l} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, p \end{array}$$

y  (1.2)

$$Y^*_i = \frac{Y_i - \bar{Y}}{\eta} , \; i = 1, 2, \ldots ,n$$

Donde :

$$d_j = \left[ \sum_{i=1}^{n} (X^*_{ij} - \bar{X}^*_j)^2 \right]^{1/2} , \; j = 1, 2, \ldots, p$$

$$\eta^2 = \sum^{n} (Y_i - \bar{Y})^2 ,$$ llamado suma de la variación de las desviaciones al cuadrado de la variable respuesta o Suma de cuadrados totales (SCT).

$$\bar{X}^*_j = n^{-1} \sum_{i=1}^{n} X^*_{ij} , \qquad j = 1, 2, \ldots ,p$$

$$\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$$

Luego, el modelo de regresión lineal múltiple (1.1) en su forma estandarizada sólo en las variables $X^{*}$'s es:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ . \ . \ . + \beta_p X_{ip} + e_i$$

ó en notación matricial

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\beta_{p\times 1} + \mathbf{e}_{n\times 1} \qquad (1.3)$$

Este modelo, es llamado "Modelo de Regresión Lineal múltiple estandarizado", donde :

$$E(\mathbf{e}) = 0 \quad y \quad E(\mathbf{ee}') = \sigma^2\mathbf{I}_{n\times n}$$

$$X_j' 1 = 0 \quad y \quad X_j' X_j = 1 \ , \ \forall \ j = 1, 2, \ ... \ ,p$$

y además

$$\mathbf{X} = (X_1, \ ... \ ,X_p) \quad y \quad 1 = (1, \ ... \ ,1)$$

De aqui en adelante se usa el modelo (1.3), para evitar los efectos de la escala de medición.

**Definición 3.-**

En el modelo (1.3), la matriz $\mathbf{X}'\mathbf{X}$ de dimención p×p es definida como "MATRIZ DE CORRELACION SIMPLE DE LAS VARIABLES REGRESORAS", es decir

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & r_{12} & r_{13} & & r_{1p} \\ r_{12} & 1 & r_{23} & & r_{23} \\ r_{13} & r_{23} & 1 & & r_{3p} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ r_{1p} & r_{2p} & r_{3p} & & 1 \end{bmatrix} \qquad (1.4)$$

$\mathbf{X}'\mathbf{Y}$ de dimención p×1, es la relación simple entre las variables regresoras $X_j$ y la variable respuesta $\mathbf{Y}$

$$X'Y = \begin{bmatrix} r_{1y} \\ r_{2y} \\ \cdot \\ \cdot \\ \cdot \\ r_{py} \end{bmatrix}_{p\times 1}$$

## 1.2 SUPUESTOS DEL MODELO DE REGRESION LINEAL MULTIPLE

**Definición 4.-**

Para que el modelo (1.3), sea utilizado en problemas de la inferencia estadística, es necesario que cumplan con los supuestos siguientes, los cuales se detallan a continuación.

**Supuesto 1**

$$E(e_i/X_i) = 0 \quad ó \quad E(e_i) = 0 \tag{1.5}$$

La violación de este supuesto, puede llevar a pronósticos o resultados muy engañosos, afecta al intercepto de la regresión.

**Supuesto 2**

$$E(e_i' e_j) = \begin{cases} 0 & i \neq j \\ \sigma^2, & i = j \end{cases} \quad ó \quad E[ee] = \sigma^2 I \tag{1.6}$$

Donde I:matriz identidad (nxn)

Significa, que las perturbaciones $e_i$ y $e_j$ no están correlacionadas cuando $i \neq j$, es decir indica la no existencia de "**correlación serial ó autocorrelacion**". Cuando i = j significa, que la varianza de los $e_i$ para cada i, es un número positivo constante igual a $\sigma^2$, a esta característica se le llama "**Homoscedasticidad**" , "igual dispersión" ó "igual varianza".

**Supuesto 3**

$$\text{Cov}( e_i , X_i ) = 0 \quad \text{ó} \quad \mathbf{X} \qquad (1.7)$$

Significa que si, en algún momento las variables explicatorias son estocásticas, estas no están correlacionadas con las perturbaciones. Pero, en general se asume que las variables explicativas no son estocásticas y la matriz **X** se denomina "**Matriz de observación**".

**Supuesto 4**

No existe relación lineal exacta entre las variables "X´s", es decir, su significado equivalente es " NO EXISTE MULTICOLINEALIDAD " ó **X** es una matriz de rango completo e igual a **p**.

**Supuesto 5**

Los e´s (errores o perturbaciones) están normalmente distribuidos con media y varianza:

$$\left.\begin{array}{r} E( \mathbf{e} ) = \mathbf{0} \\ E( \mathbf{ee'} ) = \sigma^2 I \end{array}\right\} \qquad (1.8)$$

Es decir, $\mathbf{e} \sim N( \mathbf{0}, \sigma^2 \mathbf{I} )$,

donde $\begin{cases} \mathbf{0}: & \text{vector columna } n\times 1 \\ \mathbf{I}: & \text{matriz identidad } n\times n \end{cases}$

llamado supuesto de normalidad, este supuesto no es necesario, sí nuestro objetivo es sólo la estimación. Pero no así si se pretende realizar pruebas de hipótesis y predicción

**Definición 5.-**

El modelo de regresión lineal múltiple estandarizado (1.3), que satisface todos los supuestos señalados arriba, se le conoce como "**Modelo** de regresión lineal normal clásico". A partir de ahora lo mencionaremos en forma simplificada como "Modelo de regresión lineal".

## 1.3 ESTIMACION CLASICA DEL VECTOR DE COEFICIENTES DEL MODELO DE REGRESION LINEAL

La estimación del vector de coeficientes en el modelo (1.3), que satisface la Definición 4 en 1.2, es realizada mediante el método clásico de "Mínimos cuadrados ordinarios (MCO)". "$\beta_{MCO}$", representará el estimador de MCO de $\beta$, el cual es calculado por :

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y \qquad (1.9)$$

Este estimador, goza de las propiedades siguientes

a)

$$\left.\begin{aligned} E(\beta_{MCO}) &= \beta \\ V(\hat{\beta}_{MCO}) &= \sigma^2\sum_{j=1}^{p}\lambda_j^{-1}T_jT'_j \\ ECM(\hat{\beta}_{MCO}) &= \sigma^2\sum_{j=1}^{p}\lambda_j^{-1} \end{aligned}\right\} \qquad (1.10)$$

Donde $\lambda_j$, $j = 1, 2, \ldots, p$, son los autovalores de $X'X$

$T_j$, $j = 1, 2, \ldots, p$, son los autovectores de $X'X$

b) Es el mejor estimador lineal insesgado con varianza mínima (MELI).

c) Minimiza, la suma cuadrado de los residuales de la regresión.

**OBSERVACION:** Por convención, en este trabajo, se denota el estimador de mínimos cuadrados ordinarios (MCO) como

$$\beta_{MCO} = \hat{\beta} \qquad (1.11)$$

# C A P I T U L O II

# M U L T I C O L I N E A L I D A D

## 2.1 INTRODUCCION

El modelo de regresión lineal normal clásico (1.3), cumpliendo los supuestos mencionados en la **definición 4**, en la sección **1.2**, es utilizado para inferencias, y prónosticos, pero cuando fallan algunos de estos supuestos, la estimación de los coeficientes del modelo de regresión lineal via MCO ya no son tan confiables. Por tanto el objetivo del presente trabajo fue la investigación, cuando falla la **Definición 4** en el **Supuesto 4**, es decir cuando las variables "X´s" son multicolineales ó cuando

$$| X'X | \approx 0 \qquad (2.1)$$

En este capí tulo, se presenta los posibles origenes del fenómeno de multicolinealidad, que es un problema netamente muestral, luego las diferentes consecuencias que genera, y terminar describiendo tipos de métodos de detección del problema e indicando cuales son los más efectivos.

## 2.2 COLINEALIDAD Y MULTICOLINEALIDAD

### 2.2.1 DEFINICION

En términos estrictos, la colinealidad se refiere a una sola relación lineal entre algunas de las variables explicatorias de un modelo de regresión lineal, mientras que, multicolinealidad identifica la existencia de más de una relación lineal exacta entre algunas de las variables explicatorias de un modelo de regresión. En la práctica, no existe la distinción entre colinealidad y multicolinealidad; sólo mencionaremos el término **multicolinealidad** para referirnos a ambos casos. Si no existe relación lineal entre las regresoras, ellas son llamadas ortogonales. Desafortunadamente en problemas reales, las variables regresoras no son ortogonales, y más aún, pueden llegar a ser casi linealmente dependientes, por tanto, el estimador de "MCO" será influenciado.

### 2.2.2 ORIGEN

Sea el modelo (1.3), luego, de manera formal se define el fenómeno de multicolinealidad en términos de la dependencia lineal de las columnas de la matriz **X** de dimensión nxp como sigue. Los vectores columnas de esta matriz $X_1, \ldots, X_p$ son linealmentes dependientes si existe un conjunto de constantes $t_1, \ldots, t_p$, no todos igual a cero, tal que :

$$\sum_{j=1}^{p} t_j X_j = 0 \qquad (2.2)$$

Sí (2.2) se cumple, entonces el rango de $(X'X)$ es menor que "p", por lo tanto $(X'X)^{-1}$ no existe y se dice que hay una dependencia lineal en "$X'X$", luego, se concluye que hay mal condicionamiento en esta matriz, es decir existe multicolinealidad en la matriz de datos.

Se detectan cuatro origenes primarios de multicolinealidad

1) Método empleado para la colección de los datos.

2) Comprimiendo el modelo de la población.

3) Especificación del modelo.

4) Modelo Sobre-definido.

Es interesante distinguir estos origenes, para diagnosticar mejor el problema de multicolinealidad.

1) **Método empleado para la colección de los datos.-**

El método de colección de datos, conduce a multicolinealidad, cuando el análisis muestral solamente es realizada sólo a un subespacio de la región de las regresoras definidas en (1.3), Luego el efecto de multicolinealidad causada por las técnicas muestrales no es inherente en el modelo o en la población a ser muestreada, por lo tanto no hay forma para prevenirlo.

2) **Comprimiendo el modelo de la población.-**

Comprimiendo el modelo sólo a variables que de antemano se ve que están relacionadas, entonces la multicolinealidad esta presente, por ejemplo, si se quiere averiguar el consumo de energía (Y) según el ingreso familiar ($X_1$) y el tamaño de hogar ($X_2$) vemos que $X_1$ y $X_2$ están relacionados positivamente, a más ingreso tendrán un hogar más grande y viceversa.

3) **Especificación del modelo.-**

Al momento de especificar el modelo, es posible encontrar que dos o más variables seán linealmente dependientes, entonces el fenómeno de multicolinealidad esta presente y se sugiere trabajar con un subconjunto de regresoras desde el punto de vista multicolinealidad.

4) **Modelo sobre-definido.-**

Es cuando existen más variables regresoras que observaciones disponibles, lo usual es eliminar las variables de menor importancia. Mason, Gunst y Webster [ 1975] dan tres recomendaciones para eliminar el efecto de multicolinealidad :

a) Redefinir el modelo sobre un pequeño subconjunto de las variables regresoras.

b) Desarrollar estudios preliminares sobre la base de un subconjunto de regresores originales.

c) Usar el método de componentes principales para decidir que regresor va a ser removido desde el modelo.

## 2.2.3 CONSECUENCIA DE LA MULTICOLINEALIDAD

La multicolinealidad, trae efectos serios sobre los coeficientes estimados por "MCO" en un modelo de regresión lineal. Demostraremos algunos de estos efectos.

1) Supóngase, que hay dos variables regresoras $X_1$ y $X_2$, asumir que el modelo es estandarizado, entonces se tiene

$$Y = \beta_1 X_1 + \beta_1 X_2 + e$$

La ecuación normal de mínimos cuadrados es :

$$(X'X)\beta = X'Y$$

en forma matricial

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}_{2\times 2} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{2\times 1} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}_{2\times 1}$$

donde, $r_{12}$ es la correlación simple entre $X_1$ y $X_2$; y $r_{1y}$, $r_{2y}$ es la relación simple entre las variables regresoras con **Y** respectivamente. Luego, la matriz inversa de **X'X** es :

$$C = (X'X)^{-1} = \begin{bmatrix} \dfrac{1}{(1 - r_{12}^2)} & \dfrac{-r_{12}}{(1 - r_{12}^2)} \\ \dfrac{-r_{12}}{(1 - r_{12}^2)} & \dfrac{1}{(1 - r_{12}^2)} \end{bmatrix} \qquad (2.3)$$

por tanto la estimación del coeficiente $\beta$ del modelo de regresión (1.3), es:

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

$$= \begin{bmatrix} \dfrac{1}{(1-r_{12}^2)} & \dfrac{-r_{12}}{(1-r_{12}^2)} \\ \dfrac{-r_{12}}{(1-r_{12}^2)} & \dfrac{1}{(1-r_{12}^2)} \end{bmatrix}_{2\times 2} \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}_{2\times 1}$$

$$\longrightarrow \quad \beta = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \dfrac{r_{1y} - r_{12}r_{2y}}{(1-r_{12}^2)} \\ \dfrac{r_{2y} - r_{12}r_{1y}}{(1-r_{12}^2)} \end{bmatrix}_{2\times 1} \qquad (2.4)$$

por otro lado :

$$\text{Var-Cov}(\hat{\beta}) = E\{ [\hat{\beta} - E(\hat{\beta})] [\hat{\beta} - E(\hat{\beta})]' \}$$

y escribiendo esta expresión explícitamente para "p" variables regresoras

$$\text{Var-Cov}(\hat{\beta}) = \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1\hat{\beta}_2) & \ldots\ldots & \text{Cov}(\hat{\beta}_1\hat{\beta}_p) \\ \text{Cov}(\hat{\beta}_2\hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \ldots\ldots & \text{Cov}(\hat{\beta}_2\hat{\beta}_p) \\ \cdots & \cdots & \cdots & \cdots \\ \text{Cov}(\hat{\beta}_p\hat{\beta}_1) & \text{Cov}(\hat{\beta}_p\hat{\beta}_2) & \ldots\ldots & \text{Var}(\hat{\beta}_p) \end{bmatrix}$$

$$\text{Var-Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1} = \sigma^2 C \qquad (2.4.0)$$

$$= \sigma^2 \begin{bmatrix} c_{11} & c_{12} & \ldots\ldots & c_{1p} \\ c_{21} & c_{22} & \ldots\ldots & c_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ c_{p1} & c_{p2} & \ldots\ldots & c_{pp} \end{bmatrix}_{p\times p}$$

donde, $\sigma^2$ es la varianza homocedastica de $e_i$.

Para el caso que estamos demostrando se tendrá

$$\text{Var-Cov}(\beta) = \sigma^2 \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} \dfrac{1}{(1-r_{12})^2} & \dfrac{-r_{12}}{(1-r_{12})^2} \\ \dfrac{-r_{12}}{(1-r_{12})^2} & \dfrac{1}{(1-r_{12})^2} \end{bmatrix}$$

si existe multicolinealidad entre $X_1$ y $X_2$, entonces la correlación simple $r_{12}$ es grande, es decir

$|r_{12}| \longrightarrow 1$ entonces $V(\beta_j) = \sigma^2 c_{jj} \longrightarrow \infty$ , $j = 1,2$

y $Cov(\beta_1, \beta_2) = \sigma^2 c_{12} \longrightarrow \pm \infty$ dependiendo de si $r_{12} \longrightarrow +1$ ó $r_{12} \longrightarrow -1$

Por lo tanto, una multicolinealidad fuerte implica, que la varianza y covarianza del estimador de MCO sea grande. Efecto similar se producirá cuando se tiene más de dos variables regresoras. Otra forma de visualizar lo anterior es, demostrando que los elementos diagonales de la matriz $(X'X)^{-1}$ son

$$c_{jj} = \frac{1}{(1 - R_j^2)} \quad , \ j = 1, \ldots, p \quad [1] \qquad (2.5)$$

Donde, $R_j^2$ es el coeficiente de determinación de la regresión entre la variable $X_j$ y el resto de las (p-1) variables regresoras.

Sí el efecto de multicolinealidad es fuerte entre la variable $X_j$ y cualquier subconjunto de (p - 1) variables regresoras, entonces $R_j^2 \longrightarrow 1$, por lo tanto: $Var(\beta_j) \longrightarrow \infty$ y generalmente $Cov(\beta_i \beta_j)$ es grande si las variables explicatorias $X_i$ y $X_j$ están implicadas en multicolinealidad, para $i \neq j$.

2) Otro efecto de multicolinealidad, es que tiende a producir un estimador

[1] Se demuestra en A, apendice I, que $c_{jj} = \dfrac{\quad}{(1-R_j^2)}$, $j = 1,2,..,p$

de mínimos cuadrados $\hat{\beta}_j$ grande en valor absoluto, en efecto, considerar la distancia al cuadrado del vector $\hat{\beta}$ a $\beta$

$$L^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \qquad (2.6)$$

Aplicando la esperanza a (2.6)

$$E(L^2) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

$$= \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)^2 = \sum_{j=1}^{p} E(\hat{\beta}_j - E(\hat{\beta}_j))^2$$

$$= \sum_{j=1}^{p} V(\hat{\beta}_j)$$

$$E(L^2) = \sigma^2 Tr(X'X)^{-1} \qquad (2.7)$$

luego $$E(L^2) = \sigma^2 \sum_{j=1}^{p} (1/\lambda_j) \quad [2] \qquad (2.8)$$

Donde $\lambda_j > 0$ , son los autovalores de $X'X$ , j:1,2,...,p. Luego, si al menos uno de los $\lambda_j$ es pequeño, la distancia del estimador $\hat{\beta}$ al valor verdadero $\beta$ puede ser grande. Además si se desarrolla

$$E(L^2) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

$$= E(\hat{\beta}'\hat{\beta} - 2\hat{\beta}'\beta + \beta'\beta)$$

$$E(L^2) = E(\hat{\beta}'\hat{\beta}) - \beta'\beta$$

y por (2.7)

$$\sigma^2 Tr(X'X)^{-1} = E(\hat{\beta}'\hat{\beta}) - \beta'\beta$$

por consiguiente

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 Tr(X'X)^{-1} \qquad (2.9)$$

[2] Se demuestra en B, apendice I, que la traza de la matriz

$$Tr(X'X)^{-1} = \sum_{j=1}^{p} (1/\lambda_j)$$

grande en valor absoluto.

3) En virtud del punto 1), para casos con alta multicolinealidad, las cifras muestrales pueden ser compatibles con un conjunto de diversas hipótesis, por lo que la probabilidad de aceptar una hipótesis falsa aumenta.

4) Las estimaciones por "MCO" de los coeficientes de la regresión, son pobres cuando hay multicolinealidad, pero no necesariamente el modelo ajustado es un predictor pobre. Generalmente el modelo produce predicciones satisfactorias, esto ocurre por que las combinaciones lineales de

$$\sum_{j=1}^{p} \hat{\beta}_j X_{ij}$$

pueden ser estimada bastante bien, aún cuando los parámetros $\beta_j$ son estimados pobremente.

Como se aprecia, el fenómeno de multicolinealidad, tiene efectos negativos en la estimación correcta de los coeficientes del modelo de regresión lineal.

### 2.2.4 METODOS PARA DETECTAR LA MULTICOLINEALIDAD

Se proponen varios métodos para la detección de multicolinealidad, con el objeto de proveer información, para determinar, qué regresoras están implicadas en multicolinealidad.

**1) EXAMINACION DE LA MATRIZ DE CORRELACION: X'X**

La matriz de correlación del modelo (1.3) es **X'X**, luego, una medida para la inspección de multicolinealidad, es analizar los elementos $r_{ij}$ fuera de la diagonal principal de esta matriz:

a) Sí las regresoras $X_i$ y $X_j$ son casi linealmente dependientes entonces $|r_{ij}| \longrightarrow 1$ , donde $r_{ij}$ es la correlación simple entre $X_i$ y $X_j$, i,j =

1, 2, ... ,p, $i \neq j$

b) Desafortunadamente, cuando están involucrados más de dos variables regresoras en la dependencia lineal, no existe una medida para ver que parejas $r_{ij}$ sea grande.

Generalmente la inspección de los $r_{ij}$, no es suficiente, para detectar multicolinealidad en parejas mucho más complejas.

**2) EXAMINACION DE LOS FACTORES DE INFLACION DE LA VARIANZA**

la matriz varianza-covarianza de $\beta$ es

$$\text{Var-Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}_{p\times p} = \sigma^2 C_{p\times p}$$

además, los elementos diagonales de $C_{p\times p}$ son

$$c_{jj} = \frac{\quad}{(1 - R_j^2)}$$

luego, si la variable $X_j$ es casi linealmente dependiente con algún conjunto de variables regresoras entonces $R_j^2 \longrightarrow 1$, por tanto $c_{jj}$ es grande. Podemos ver que $c_{jj}$ es el factor por el cual la varianza de $\beta_j$ es creciente, pues $\text{Var}(\hat{\beta}_j) = \sigma^2 c_{jj}$, j = 1, 2, ...,p. Marquardt (1970) ha **llamado a $c_{jj}$ como el factor de inflación de varianza ($FIV_j$)**, es decir

$$\text{Var}(\hat{\beta}_j) = \sigma^2 FIV_j \qquad j = 1,2,\dots p. \qquad (2.10)$$

Un FIV, mide para cada término en el modelo, los efectos combinados de la dependencia entre las variables regresoras sobre la varianza del término. Por tanto la examinación de los FIV se evalúa como sigue:

a) Uno o más FIV altos indica multicolinealidad.

b) Si cualquiera de los FIV´s excede a 5 ó 10 se da la información de que el coeficiente de regresión asociado, es pobremente estimado por la existencia de multicolinealidad.

c) Sabemos que el intervalo confidencial para el coeficiente de regresión verdadero es :

$$\beta_j - t_{\frac{\alpha}{2},n-p-1} De(\beta_j) \leq \beta_j \leq \beta_j + t_{\frac{\alpha}{2},n-p-1} De(\beta_j)$$

Donde De = desviación estandar = $\hat{\sigma}_{\beta_j}$ , $j = 1, 2, \ldots, p$

= nivel de significancia

Luego, la longitud de este intervalo es

$$L_j = 2De(\beta_j)t_{\frac{\alpha}{2},n-p-1} = 2(c_{jj}\hat{\sigma}^2)^{1/2} t_{\frac{\alpha}{2},n-p-1}$$

y la longitud del intervalo correspondiente en un diseño totalmente ortogonal en el cual

$$c_{jj} = \frac{1}{(1 - R_j^2)} = 1 \ , \quad (\text{pues } R_j^2 - 0)$$

es

$$L^* = 2(\hat{\sigma}^2)^{1/2} t_{\frac{\alpha}{2},n-p-1}$$

entonces

$$L_j / L^* = (c_{jj})^{1/2} = (FIV_j)^{1/2}$$

Esto indica que el intervalo confidencial $L_j$ es más grande, a causa del efecto de multicolinealidad.

**3) EL ANALISIS DEL AUTOSISTEMA DE X'X**

Este análisis, se fundamenta en las observaciones que se hacen a los autovalores de **X'X**, los cuales son usados para medir la extensión de multicolinealidad.

Definimos EL NUMERO DE CONDICION de **X'X** así :

$$\Psi = \frac{max}{min}$$

donde $\lambda_{max}$ y $\lambda_{min}$ son los autovalores máximo y mínimo de **X'X** respectivamente. El método del análisis del autosistema sigue el siguiente algoritmo:

a) Se encuentran los autovalores de la matriz $X'X$ (matriz de correlación)

$\lambda_1, \lambda_2, \ldots, \lambda_p$

b) Se elige el máximo y mínimo de los autovalores y se evalúa $\Psi$

c) - Si $\Psi < 100$ no existe multicolinealidad.

- Si $100 \leq \Psi \leq 1000$ implica multicolinealidad moderada.

- Si $\Psi > 1000$ indica multicolinealidad muy severa.

Este método ayuda a encontrar **la naturaleza de la dependencia lineal más cercana en los datos.** Si existe una o más dependencias lineales en los datos, entonces las raices características son pequeñas; o también si uno o más autovalores son pequeños implica que existe dependencia lineal cercana entre las columnas de $X$. Esto se ve descomponiendo la matriz de correlación $X'X$ de la forma siguiente:

$$X'X = T \Lambda T'$$

donde, $\Lambda$ es una matriz diagonal $p \times p$ cuyos elementos diagonales son los autovalores $\lambda_j$, $j = 1, \ldots, p$ de $X'X$, T es la matriz ortogonal $p \times p$ cuyas columnas son los autovectores de $X'X$. Luego si $T = (T_1, T_2, \ldots, T_p)$ y $\lambda_j$ es cercano a cero los elementos del autovector $T_j$ asociado a este autovalor describe la naturaleza de esta dependencia lineal. Especificamente, los elementos del vector $T_j$ son los coeficientes $t_1$, $t_2$, $\ldots$, $t_p$ en (2.2), es decir de

$$\sum_{j=1}^{p} t_j X_j = 0.$$

**4) EL ANALISIS DE DESCOMPOSICION DE VALORES SINGULARES DE X.**

Este método, fue propuesto por Belsley, Kuh y Welsch [ 1980] , para llevarse a efecto, la matriz de datos $X_{n \times p}$ es descompuesta del siguiente modo:

$$X = U\ D\ T'$$

$U$ es de orden (n×p), cuyas columnas son los autovectores asociados a los "p" autovalores diferentes de cero de la matriz $(X'X)_{n\times n}$; $T$ de orden p×p (matriz ortogonal ya dada anteriormente), $D$ de orden p×p (matriz diagonal) con elementos no negativos $\mu_j$, j = 1,...,p. Los $\mu_j$ son llamados los "valores singulares de $X$" y $X = U\ D\ T'$ se llama "la descomposición de valores singulares de $X$", además

$$U'U = I, \qquad T'T = I, \qquad X'X = (U\ D\ T')'(U\ D\ T) = T\ D^2\ T'$$

y como se conoce que

$$X'X = T\ \Lambda\ T'$$

$$D^2 = \Lambda$$

luego:

$$\underbrace{\begin{bmatrix} \mu_1^2 & & & 0 \\ & \mu_2^2 & & \\ & & \ddots & \\ 0 & & & \mu_p^2 \end{bmatrix}}_{D^2_{p\times p}} = \underbrace{\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{bmatrix}}_{p\times p}$$

es decir

$$\mu_j^2 = \lambda_j\ ,\ j = 1,\ 2,\ \ldots,\ p$$

Este método, mide el mal condicionamiento de $X$ a través del tamaño de los valores singulares. Para cada dependencia lineal cercana hay un valor singular pequeño, luego Belsley, kuh, Welsch [ 1980] , definen "EL INDICE DE CONDICION" de la matriz $X$ como :

$$\eta_j = \frac{\mu_{max}}{\mu_j}\ ,\ j = 1,\ \ldots,\ p$$

donde, $\mu_{max}$ = Valor singular máximo de $X$.

Por lo tanto, valores grandes de $\eta$. están asociados al mal condicionamiento de **X**. Notece, que se esta trabajando directamente con la matriz de datos **X**, con la cual es nuestro interés. Una ventaja de este método, es que, el algoritmo para generar la descomposición de los valores singulares es numéricamente más estable que el método de ANALISIS DEL AUTOSISTEMA. En la práctica, esta no es una indicación tan severa, pues, puede elegirse cualquiera de los métodos de acuerdo al problema que se afronta.

Otra indicación, que debe tener en cuenta como una medida de multicolinealidad para este método, es análizar las "Proporciones de descomposición de la varianza ($\Pi_{ij}$)" definidas por Belsley, Kuh, Welsch como:

$$\Pi_{ij} = \frac{t_{ji}^2 / \mu_i^2}{FIV}, \quad i,j = 1, \ldots, p \qquad (2.11)$$

Esta definición se ve, análizando la matriz de Var-Cov($\beta$):

$$\text{Var-Cov}(\ddot{\beta}) = \sigma^2(X'X)^{-1} = \sigma^2(T\,\Lambda\,T')^{-1} = \sigma^2[(T')^{-1}\,\Lambda^{-1}\,T^{-1}]$$

$$= \sigma^2[\,T\,\Lambda^{-1}\,T'\,]$$

$$= \sigma^2 \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{bmatrix} \underbrace{\begin{bmatrix} 1/\mu_1^2 & & & 0 \\ & 1/\mu_2^2 & & \\ & & \ddots & \\ 0 & & & 1/\mu_p^2 \end{bmatrix}}_{\text{Pues } D^2 = \Lambda} \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{bmatrix}$$

$$
= \sigma^2 \begin{bmatrix} \sum_{i=1}^{p} t_{1i}^2/\mu_i^2 & \sum_{i=1}^{p} t_{1i}t_{2i}/\mu_i^2 & \cdots\cdots & \sum_{i=1}^{p} t_{1i}t_{pi}/\mu_i^2 \\ \sum_{i=1}^{p} t_{2i}t_{1i}/\mu_i^2 & \sum_{i=1}^{p} t_{2i}^2/\mu_i^2 & \cdots\cdots & \sum_{i=1}^{p} t_{2i}t_{pi}/\mu_i^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{p} t_{pi}t_{1i}/\mu_i^2 & \sum_{i=1}^{p} t_{pi}t_{2i}/\mu_i^2 & \cdots\cdots & \sum_{i=1}^{p} t_{pi}^2/\mu_i^2 \end{bmatrix}_{p\times p}
$$

Cada elemento diagonal de esta matriz, multiplicada por $\sigma^2$ da la $Var(\beta_j)$ del coeficiente de regresión j = 1, ...,p, es decir

$$
Var(\hat{\beta}_j) = \sigma^2 \sum_{i=1}^{p} t_{ji}^2/\mu_i^2 \ , \ j = 1, \ldots, p \qquad (2.12)
$$

Luego reemplazando (2.10) en (2.11) se tiene que

$$
FIV_j = \sum_{i=1}^{p} t_{ji}^2/\mu_i^2 = \sum_{i=1}^{p} t_{ji}^2/\lambda_i \ , \ j = 1, \ldots, p \qquad (2.13)
$$

Así :

$$
\Pi_{ij} = \frac{t_{ji}^2/\mu_i^2}{FIV_j}
$$

De (2.13), se aprecia que si uno o más valores singulares (o autovalores) son pequeños, infla inmediatamente la varianza de $\hat{\beta}_j$. Haciendo el arreglo de $\Pi_{ij}$ en una matriz $\Pi$ de orden pxp

$$
\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \cdots & \Pi_{1p} \\ \Pi_{21} & \Pi_{22} & \cdots & \Pi_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \Pi_{p1} & \Pi_{p2} & \cdots & \Pi_{pp} \end{bmatrix}
$$

$$
: \begin{bmatrix}
\dfrac{t_{11}^2/\mu_1^2}{\dfrac{t_{11}^2}{\mu_1^2}+\dfrac{t_{12}^2}{\mu_2^2}+\,..+\dfrac{t_{1p}^2}{\mu_p^2}} & \dfrac{t_{21}^2/\mu_1^2}{\dfrac{t_{21}^2}{\mu_1^2}+\dfrac{t_{22}^2}{\mu_2^2}+\,..+\dfrac{t_{2p}^2}{\mu_p^2}} & \cdots & \dfrac{t_{p1}^2/\mu_1^2}{\dfrac{t_{p1}^2}{\mu_1^2}+\dfrac{t_{p2}^2}{\mu_2^2}+\,..+\dfrac{t_{pp}^2}{\mu_p^2}} \\
\dfrac{t_{12}^2/\mu_2^2}{\dfrac{t_{11}^2}{\mu_1^2}+\dfrac{t_{12}^2}{\mu_2^2}+\,..+\dfrac{t_{1p}^2}{\mu_p^2}} & \dfrac{t_{22}^2/\mu_2^2}{\dfrac{t_{21}^2}{\mu_1^2}+\dfrac{t_{22}^2}{\mu_2^2}+\,..+\dfrac{t_{2p}^2}{\mu_p^2}} & \cdots & \dfrac{t_{p2}^2/\mu_2^2}{\dfrac{t_{p1}^2}{\mu_1^2}+\dfrac{t_{p2}^2}{\mu_2^2}+\,..+\dfrac{t_{pp}^2}{\mu_p^2}} \\
\cdots & \cdots & \cdots & \cdots \\
\dfrac{t_{1p}^2/\mu_p^2}{\dfrac{t_{11}^2}{\mu_1^2}+\dfrac{t_{12}^2}{\mu_2^2}+\,..+\dfrac{t_{1p}^2}{\mu_p^2}} & \dfrac{t_{2p}^2/\mu_p^2}{\dfrac{t_{21}^2}{\mu_1^2}+\dfrac{t_{22}^2}{\mu_2^2}+\,..+\dfrac{t_{2p}^2}{\mu_p^2}} & \cdots & \dfrac{t_{pp}^2/\mu_p^2}{\dfrac{t_{p1}^2}{\mu_1^2}+\dfrac{t_{p2}^2}{\mu_2^2}+\,..+\dfrac{t_{pp}^2}{\mu_p^2}}
\end{bmatrix}
$$

$$
\Pi = \begin{bmatrix}
\dfrac{t_{11}^2/\mu_1^2}{FIV_1} & \dfrac{t_{21}^2/\mu_1^2}{FIV_2} & \cdots\cdots & \dfrac{t_{p1}^2/\mu_1^2}{FIV_p} \\
\dfrac{t_{12}^2/\mu_2^2}{FIV_1} & \dfrac{t_{22}^2/\mu_2^2}{FIV_2} & \cdots\cdots & \dfrac{t_{p2}^2/\mu_2^2}{FIV_p} \\
\cdots & \cdots & \cdots & \cdots \\
\dfrac{t_{1p}^2/\mu_p^2}{FIV_1} & \dfrac{t_{2p}^2/\mu_p^2}{FIV_2} & \cdots\cdots & \dfrac{t_{2p}^2/\mu_p^2}{FIV_p}
\end{bmatrix}_{p\times p} \qquad (2.14)
$$

Analizando esta matriz, se observa que los elementos de cada columna de la matriz $\Pi$ da la proporción de la varianza de cada $\beta_j$ contribuido por el i-ésimo valor singular (ó autovalor).

Una alta proporción de variancia para dos ó más coeficientes de regresión, es asociada con un valor singular pequeño, esto es un indicativo de que hay presencia de multicolinealidad. Es decir si $\Pi_{21}$ y $\Pi_{25}$ es grande, implica que esta asociado con el segundo valor singular, por lo tanto, infla la varianza de $\beta_1$ y $\beta_5$, por consiguiente hay signo de multicolinealidad.

Entonces, la norma general en el análisis de descomposición de valores singulares de **X**, para decir que existe multicolinealidad es :

- Indice de condición $\eta_j > 30$
- Proporcion de descomposición de la varianza $\Pi_{ij} > 0.5$

Existen otras técnicas, pero que son ocasionalmente útiles, tales como:

a) Hallar la determinante de **X'X** (matriz de correlación); como se entiende, esta determinante esta en función de las correlaciones simples de las regresiones, entonces

$$0 \leq |\, \mathbf{X'X} \,| \leq 1$$

por tanto si $|\,\mathbf{X'X}\,| \longrightarrow 0$, la multicolinealidad es fuerte, es decir, hay dependencia lineal cercana entre las regresoras, la desventaja de esta técnica es que no da la información sobre el origen del fenómeno.

b) Ver la significancia de los estadísticos **F** y **t**. Si especificamente el estadístico **F** es significativo, para todos los coeficientes de la regresión simultaneamente, y el estadístico **t** en forma individual para cada coeficiente son todos no significativos, entonces la multicolinealidad esta presente; desafortunadamente, muchos conjuntos de datos que si pueden tener multicolinealidad, no exhiben este comportamiento y su uso es cuestionable.

c) Ver la significancia del estadístico $F_j$, $j = 1, 2, \ldots, p$; esto es, hacer la regresión de cada $X_j$, $j = 1, 2, \ldots, p$, con el resto de $(p - 1)$ variables regresoras y calcular $R_j^2$ para cada regresión, entonces

$$F_j = \frac{R^2_{X_j . X_2 X_3 \cdots X_p} / p - 2}{\left[1 - R^2_{X_j . X_2 X_3 \cdots X_p}\right] / (n - p + 1)} \quad , \; j = 1, \ldots, p$$

Por consiguiente:

- Si $F_j \geq F_{Tabla\,\alpha,(p-2),(n-p+1)gl}$

  $X_j$ no es colineal con los otros X´s; se retiene $X_j$ en el modelo.

- Si $F_j < F_{Tabla\,\alpha,(p-2),(n-p+1)gl}$

  $X_j$ es colineal con los otros X´s; no sabemos si $X_j$ sale del modelo.

d) - Si adicionamos ó removemos una variable regresora en el modelo, y calculamos los coeficientes, produciendose grandes cambios, es un indicativo de presencia de multicolinealidad.

- La suspensión de algunos datos puntuales, que producen cambios grandes, también indican que el fenómeno esta presente.

- Si el signo o magnitud son contrarios al valor esperado a priori, también esta presente el fenómeno. Pero, uno debe proceder con cuidado, pues los datos no tienen un comportamiento bien definido para aplicar tal o cual método.

Concluimos que la técnica de FIV y el ANALISIS DEL AUTOSISTEMA DE **X' X**, son los mejores métodos para diagnosticar el fenómeno de multicolinealidad, esto se fundamenta en que es :

- Fácil de Calcular
- Directo para interpretar
- Util para especificar la naturaleza de multicolinealidad.

# C A P I T U L O   III

# COMPARACION DE LOS ESTIMADORES SESGADOS Y MINIMOS CUADRADOS ORDINARIOS, EN PRESENCIA DE MULTICOLINEALIDAD

## 3.1 INTRODUCCION

Este capítulo, presenta y compara, métodos alternativos al estimador clásico por mínimos cuadrados (MCO), del modelo (1.3), en presencia de multicolinealidad. Varios tipos de estimadores se han propuesto para tratar con este problema, ellos son: Componentes principales (CP), Raíz latente (RL), Contracción de STEIN (CS), Regresión RIDGE (RR), Ridge Generalizado (RG) y Selección de Variables (SV). Todos ellos reducen la varianza estimada, a costa de introducir algún sesgo. Del mismo modo, se presenta también, la comparación de acuerdo a una estructura general, para todos los estimadores sesgados.

## 3.2 ESTIMADORES DE LOS COMPONENTES PRINCIPALES [CP]

Se analizará la relación entre la variable dependiente y un conjunto de variables independientes, cuando estas últimas no son tratables por la estadística estándar. La idea es obtener las componentes principales de variables explicativas estandarizadas, calculando luego su regresión sobre una variable dependiente y proyectando los parámetros resultantes en los términos de las variables originales. Sobre este enfoque, se afirma "es

**posible ortogonalizar**" una situación de regresión para cualquier conjunto de variables X´s, para así de esta forma deshacernos de multicolinealidad.

### 3.2.1 ESTRUCTURA DE LAS COMPONENTES PRINCIPALES

El método de componentes principales es muy conocido y se ha discutido de distintos modos por diferentes autores, por tanto se presenta el desarrollo matemático para comprender los resultados del método.

El objetivo del análisis de componentes principales, es encontrar una transformación lineal de un conjunto de p-variables de **X**, en un nuevo conjunto denotado por "**P**", donde el nuevo conjunto tiene ciertas propiedades tales como:

*i*) Los elementos de **P** no están correlacionados (Se denotará **P** = $[P_1, P_2, \ldots, P_p]$, donde $\mathrm{corr}(P_i, P_j) = 0 \; i \neq j$, $i,j = 1,2,\ldots,p$) es decir son ortogonales.

*ii*) Cada elemento de **P** (de $P_1$ a $P_2$, etc) considera, tanto la varianza combinada de los X´s como sea posible, consistente con que es ortogonal a los P´s precedentes. Es decir la correlación entre ellos es cero y $P_1$ considera la cantidad de varianza de la variable $X_1$, $P_2$ considera la máxima cantidad de varianza combinada de las variables $X_1$ y $X_2$, así sucesivamente.

Las nuevas variables, corresponden a los ejes principales de la elipsoide, formada por la dispersión de puntos muestrales en el espacio p-dimensional que tienen los elementos de **X** como base. La transformación de componentes principales, es por tanto, una rotación del sistema coordenado *X* original al sistema definido por los ejes principales de esta elipsoide.

La razón para esto es que los ejes principales del espacio expandido por los elementos de la matriz original no son **INVARIANTES** a cambios en las escalas de medición de las variables, luego este problema puede hacerse a un lado, si el análisis se limita a los ejes principales de los elementos de la matriz original estandarizada; pero, se subraya, que esto no quiere decir que la multicolinealidad se ha eliminado. Además como **X** es la matriz estandarizada (vease (1.3), capít. I, pag.7), la transformación a componentes principales esta dada por

$$\mathbf{P}_{nxp} = \mathbf{X}_{nxp}\mathbf{T}_{pxp} \qquad (3.2.1)$$

En esta igualdad se encuentra la matriz $\mathbf{T}_{pxp}$ a partir del conocimiento de **X**. Veamos,

$$\mathbf{P}'\mathbf{P} = \mathbf{T}'\mathbf{X}'\mathbf{X}\ \mathbf{T} = \Lambda \qquad (3.2.2)$$

entonces $\mathbf{P}'\mathbf{P} = \Lambda$

Donde, $\mathbf{X}'\mathbf{X}$ es la matriz de correlaciones simples entre los X´s y $\mathbf{P}'\mathbf{P} = \Lambda_{pxp}$, es la matriz de varianza-covarianza para los componentes principales, luego

$$\mathbf{T}'\mathbf{X}'\mathbf{X}\ \mathbf{T} = \Lambda \qquad (3.2.3)$$

Por tanto (3.2.3), es una transformación de similaridad ortogonal, que diagonaliza la matriz simétrica $\mathbf{X}'\mathbf{X}$. Por consiguiente, por un teorema del algebra matricial, **la matriz de transformación T tiene un conjunto ortonormal de autovectores de X'X como sus columnas, y P'P tiene los autovalores de X'X como sus elementos diagonales**; por tanto $\Lambda$ es una matriz diagonal de autovalores de $\mathbf{X}'\mathbf{X}$.

Ahora, como las columnas de **T** son ortogonales se cumple $\mathbf{T}' = \mathbf{T}^{-1}$ luego, (3.2.1) se escribirá así

$$\mathbf{X} = \mathbf{P}\mathbf{T}' \qquad (3.2.4)$$

Reemplazando (3.2.4) en (1.3)

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

$$Y = PT'\beta + e$$

luego, $$Y = P\alpha + e \qquad (3.2.5)$$

donde, $$\alpha = T'\beta$$

De esta forma, las columnas de $P$ definen un nuevo conjunto ortogonal de regresores $P = [P_1, P_2, \ldots, P_p]$, llamados como los "componentes principales".

Aplicando mínimos cuadrados a (3.2.5) se obtiene

$$\hat{\alpha} = (P'P)^{-1}P'Y = (\Lambda)^{-1}P'Y \qquad (3.2.6)$$

Luego, $\beta$ se calcula ya que :

$$\hat{\alpha} = T'\hat{\beta} \qquad (3.2.6a)$$

y $$VAR(\hat{\alpha}) = \sigma^2(P'P)^{-1} = \sigma^2(\Lambda)^{-1} \qquad (3.2.7)$$

*(llamada matriz varianza-covarianza de $\alpha$)*

Para obtener el estimador de componentes principales $\alpha_{CP}$ se procede así :

- Se ordenan los autovalores de $X'X$ de acuerdo a su magnitud $\lambda_p \geq \lambda_{p-1} \geq \geq \lambda_1 > 0$
- Las columnas de $T$ se ordenan tal que, la primera columna corresponde al menor autovalor de $X'X$, la segunda al siguiente en magnitud, etc.
- Los componentes principales estarán ordenados como se especifica en *ii*) (vease pag. 32), además el último componente considera una máxima cantidad de varianza combinada de los X's, la antepenultima, la cantidad máxima tal que no sea correlacionada con la última, es decir $P$ estará ordenada así : $$P = [P_1, P_2, \ldots, P_p] \qquad (3.2.8)$$
- Los $\alpha$'s estimados seran por tanto $\alpha_1, \alpha_2, \ldots, \alpha_p$ correspondientes a los $\lambda_1, \lambda_2, \ldots, \lambda_p$ respectivamente (ordenado por magnitud).
- Si hay "r" multicolinealidades, los r primeros autovalores se igualan a cero y los componentes principales de (3.2.8) correspondientes a estos

autovalores se remueven del análisis, luego se aplica "MCO" al resto de componentes, es decir

$$\alpha_{CP} = K\alpha$$

donde, $K_1 = K_2 = \quad = K_r = 0$ y $K_{r+1} = K_{r+2} = \quad = K_p = 1$

por tanto

$$\alpha_{CP} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \hat{\alpha}_{r+1} \\ \hat{\alpha}_{r+1} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_{r+1} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\0\\\cdot\\\cdot\\\cdot\\0\end{matrix}}\right\} \text{r componentes} \\ \left.\vphantom{\begin{matrix}0\\0\\\cdot\\\cdot\\\cdot\\0\end{matrix}}\right\} \text{p - r componentes} \end{matrix} \qquad (3.2.9)$$

En términos de los regresores estándar originales

$$\beta_{CP} = T\ \alpha_{CP}$$

por consiguiente, se tiene

$$\beta_{CP} = T\ \alpha_{CP} = T\ \Lambda^{+}\ P'\, Y = T\ \Lambda^{+}\ (XT)'\, Y$$

Donde $\Lambda^{+} = \text{Diag}(\ 0,0,\ldots,0,\ \lambda_{r+1}^{-1},\ \lambda_{r+2}^{-1},\ldots,\ \lambda_{p}^{-1}\ )$ ó de otra manera

$$\beta_{CP} = \sum_{j=r+1}^{p} \lambda_j^{-1} t_j' X' Y t_j \qquad (3.2.10)$$

Donde, $t_j$ es un vector columna $p\times1$ de $T_{p\times p}$, $j = r+1, \ldots p$.

Es así como se encuentra la estructura general del estimador de componentes principales, el cual estima los coeficientes del modelo general de regresión en (1.3).

### 3.2.2 CRITERIO PARA ELIMINAR TERMINOS

Se han propuesto varios criterios para determinar "r", el número de términos de componentes principales a quitarse, ellos son:

1) r = 1 regla usada por Gunst y Mason (1977). Este criterio, se fundamenta sobre la base de los autovalores de $X'X$, que son ordenados de acuerdo a su magnitud y concluyen que debe eliminarse el autovalor más pequeño.

2) r = No de términos para los que $\lambda_j \leq 0.01$. Este criterio, también se fundamenta sobre la base de los autovalores de $X'X$, que son ordenados de acuerdo a su magnitud, el criterio funciona condicionado a que, si algunos $\lambda_j \leq 0.01$, j = 1, 2, .., p, estos se eliminan, entonces r Número de $\lambda_i$ eliminados.

3) r = No de términos para los que $F < F_{0.05,(p-r)(n-p)gl}$

donde
$$F = \frac{SCRes_q - SCRes_p}{SCRes_p/(n-p)}$$

y $SCRes_q$ y $SCRes_p$ son las sumas de cuadrados de los residuales después de incluir q y p términos respectivamente en el modelo (1.3). Este criterio es usada por Dempster, Schatzoff y Wermuth (1977) y trabaja con el procedimiento de selección de Backward realizandose de la manera siguiente:

a) Hacer la regresión con todas las p-variables de $X$ estandarizada y calcular $SCRes_p$, seguidamente, si para este valor $F < F_{0.05}$; se elimina una variable y r = 1.

b) Hacer la regresión para (p-1) variables y calcular $SCRes_{(p-1)}$, si para este valor $F < F_{.05}$; se elimina una variable y r = 2 ; caso contrario ir a d).

c) Continuar con el paso b) para p - i, i = 2,3, ...p - 1

d) Si $F \geq F_{.05}$, entonces el número de términos a quitarse igual a "r" será al del paso anterior para el cual $F < F_{.05}$

El criterio 3), no se desarrolla excepcionalmente bien (según algunos estudiosos sobre el fenómeno), la razón para esto podría ser el efecto severo de multicolinealidad en los datos para componentes principales asociadas con autovalores muy pequeños de $\mathbf{X'X}$. El criterio 2) es más usado y de más confianza, pues se ha demostrado para datos simulados y reales, que el error cuadrático medio es menor cuando se usa este que el criterio 3).

### 3.2.3 PROPIEDADES DEL ESTIMADOR PARA "CP"

Según la estructura, del estimador de componentes principales encontrado en (3.2.10), se menciona algunas propiedades importantes que ayudarán a inclinarnos a favor de este, cuando existe multicolinealidad.

**PROPIEDAD 1**

$$E(\hat{\beta}_{CP}) = \beta - \sum_{j=1}^{r} (t'_j \beta) t_j \qquad (3.2.11)$$

**Prueba**

Se sabe de (3.2.10) que

$$\hat{\beta}_{CP} = \sum_{j=r+1}^{p} \lambda_j^{-1} t'_j \mathbf{X'Y} t_j$$

esto mismo también se puede expresar como :

$$\hat{\beta}_{CP} = \sum_{j=1}^{p} \lambda_j^{-1} t'_j \mathbf{X'Y} t_j - \sum_{j=1}^{r} \lambda_j^{-1} t'_j \mathbf{X'Y} t_j \qquad (3.2.12)$$

Lugo, remplazando (1.3) en (3.2.12)

$$\hat{\beta}_{CP} = \sum_{j=1}^{p} \lambda_j^{-1} t'_j \mathbf{X'} [\mathbf{X}\beta + \mathbf{e}] t_j - \sum_{j=1}^{r} \lambda_j^{-1} t'_j \mathbf{X'} [\mathbf{X}\beta + \mathbf{e}] t_j$$

$$\hat{\beta}_{CP} = \sum_{j=1}^{p} \lambda_j^{-1} t'_j \mathbf{X'X} t_j \beta + \sum_{j=1}^{p} \lambda_j^{-1} t'_j \mathbf{X'e} t_j - \sum_{j=1}^{r} \lambda_j^{-1} t'_j \mathbf{X'X}\beta t_j - \sum_{j=1}^{r} \lambda_j^{-1} t'_j \mathbf{X'e} t_j$$

En forma matricial se tendrá

$$\beta_{CP} = \Lambda^{-1}_{pxp} T'_{pxp} (X'X)_{pxp} T_{pxp} \beta_{px1} + \Lambda^{-1}_{pxp} T'_{pxp} X'_{pxn} e_{nx1} T_{pxp}$$

$$- \Lambda^{-1}_{rxr} T'_{rxp} (X'X)_{pxp} \beta_{px1} T_{pxr} - \Lambda^{-1}_{rxr} T'_{rxp} X'_{pxn} e_{nx1} T_{pxr}$$

$$\beta_{CP} = \Lambda^{-1}\Lambda\beta + \Lambda^{-1}T'X'eT - \Lambda^{-1}_{r}\Lambda_{r} T'_{rxp}\beta T_{pxr} - \Lambda^{-1}_{r} T'_{rxp} X'eT_{pxr}$$

$$\beta_{CP} = I_{p}\beta + \Lambda^{-1}T'X'eT - I_{r} T'_{rxp}\beta_{px1} T_{pxr} - \Lambda^{-1}_{r} T'_{rxp} X'eT_{pxr}$$

**Observacion :** $T'X'XT = \Lambda \longleftrightarrow T'X'X = \Lambda T'$

Como $E(e) = 0$ según los supuestos clásicos, $\Lambda$, $X$, $T$ son matrices de números no aleatorios y aplicando la esperanza al estimador tenemos:

$$E(\hat{\beta}_{CP}) = \beta + \Lambda^{-1}T'X'E(e)T - T'_{rxp}\beta_{px1}T_{pxr} - \Lambda^{-1}_{r}T'_{rxp}X'E(e)T_{pxr}$$

$$E(\beta_{CP}) = \beta - T'_{rxp}\beta_{px1}T_{pxr} = \beta - \sum_{j=1} t'_{j}\beta t_{j} \quad \blacksquare$$

Por lo tanto se cumple (3.2.11)

**PROPIEDAD 2**

$$\text{Var - Cov}(\beta_{CP}) = \sigma^{2} \sum_{j=r+1}^{p} \lambda_{j}^{-1} t_{j} t'_{j} \qquad (3.2.13)$$

**Prueba**

La matriz Var-Cov del estimador de componentes principales se encuentra partiendo de la ecuación (3.2.6a) y (3.2.7)

$$\alpha = T'\beta \qquad \text{y} \qquad \text{Var-Cov}(\hat{\alpha}) = \sigma^{2}\Lambda^{-1}.$$

luego

$$\beta = T\alpha$$

por tanto

$$\text{Var-Cov}(\beta) = \text{Var-Cov}(T\hat{\alpha}) = T\Lambda^{-1}T'\sigma^{2}$$

finalmente

$$\text{Var-Cov}(\beta) = T\Lambda^{-1}T'\sigma^{2}$$

y cuando se quitan "r" autovalores más pequeños, esta última expresión se convierte en $$\text{Var - Cov}(\beta_{CP}) = \sigma^2 T_{p\times(p-r)} \Lambda^{-1}_{(p-r)(p-r)} T_{(p-r)\times p}$$

ahora, en forma de sumatoria se tendrá

$$\text{Var - Cov}(\beta_{CP}) = \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} t_j t'_j \blacksquare$$

De (3.2.13) se deduce que

$$\text{Var}(\hat{\beta}_{CP}) = \sigma^2 \text{tr}[\, T_{p\times(p-r)} \Lambda^{-1}_{(p-r)(p-r)} T'_{(p-r)\times p}] = \sigma^2 \sum_{j=r+1}^{p} \sum_{i=1}^{p} \lambda_j^{-1} t_{i,j}^2$$

donde $t_{i,j}$ son los elementos de la j-ésima columna para i = 1,..,p de la matriz $T_{p\times(p-r)}$, luego como

$$\sum_{i=1}^{p} t_{i,j}^2 = 1$$

por ser ortonormal cada columna, entonces

$$\text{Var}(\beta_{CP}) = \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} \qquad (3.2.14)$$

**PROPIEDAD 3**

$$\text{ECM}(\beta_{CP}) = \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} + \sum_{j=1}^{r} (t'_j \hat{\beta})^2 \qquad (3.2.15)$$

**<u>Prueba</u>**

La formula (3.2.15) es el **error cuadrático medio del estimador**, esta expresión se prueba de la siguiente forma :

Por definición de **ECM** se tiene

$$\text{ECM}(\hat{\beta}_{CP}) = E[\, \hat{\beta}_{CP} - \beta\, ]^2$$

$$= E[\, \hat{\beta}^2_{CP} + \beta^2 - 2\hat{\beta}_{CP}\beta\, ]$$

$$= E(\hat{\beta}^2_{CP}) + \beta^2 - 2\beta E(\hat{\beta}_{CP})$$

Sumando y restando el término $E^2(\hat{\beta}_{CP})$

$$= E(\hat{\beta}^2_{CP}) - E^2(\hat{\beta}_{CP}) + E^2(\hat{\beta}_{CP}) + \beta^2 - 2\beta E(\beta_{CP})$$

$$= Var(\beta_{CP}) + [\ E(\hat{\beta}_{CP}) - \beta\ ]^2$$

d (3.2.11) y (3.2.14)

$$= \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} + [\ - \sum_{j=1}^{r}(\ t_j'\beta\ )t_j\ ]^2$$

$$= \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} + [\ (\ t_1'\beta\ )^2 t_1^2 + \ldots + (\ t_r'\beta\ )^2 t_r^2$$

$$+ 2\ (\ t_1'\beta\ )t_1(\ t_2'\beta\ )t_2 \ldots + 2\ (\ t_1'\beta\ )t_1(\ t_r'\beta\ )t_r$$

$$+ 2\ (\ t_2'\beta\ )t_2(\ t_3'\beta\ )t_3 + \qquad + \qquad +$$

$$+ 2\ (t_{r-1}'\beta)t_{r-1}(t_r'\beta\ )t_r\ ]$$

Además, se sabe que T es una matriz ortogonal, por tanto

$$ECM(\ \beta_{CP}\ ) = \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} + [\ (\ t_1'\beta\ )^2 + \ldots + (\ t'\beta\ )^2\ ]$$

$$= \sigma^2 \sum_{j=r+1}^{p} \lambda_j^{-1} + \sum_{j=1}^{r}(\ t_j'\beta\ )^2 \quad \blacksquare$$

**PROPIEDAD 4**

$$(\ r_{y},p_j\ )^2 = \alpha_j t_j' X' Y = \lambda_j \hat{\alpha}_j^2\ ,\ j = r+1,\ \ldots\ ,p \qquad (3.2.16)$$

**Prueba**

$(\ r_y,p_j\ )^2$ es el coeficiente de correlación elevado al cuadrado entre la variable dependiente **Y** y el j-ésimo componente principal (j = r+1, r+2, ..., p). Además por definición de coeficiente de determinación del modelo de regresión (1.3)

$$R^2 = \frac{\beta\ X'Y\ - N\bar{Y}^2}{Y'Y\ - N\bar{Y}^2}$$

entonces

$$R^2 = \beta\ X'Y \qquad (3.2.17)$$

De (3.2.6a) y (3.2.17)

$$R^2 = \hat{\alpha}' T' X' Y = [\ \hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p\ ] \begin{bmatrix} t'_1 \\ t'_2 \\ \cdot \\ \cdot \\ \cdot \\ t'_p \end{bmatrix}_{p\times p} X' Y$$

$$R^2 = \hat{\alpha}_1 t'_1 X' Y + \hat{\alpha}_2 t'_2 X' Y + \dots + \hat{\alpha}_p t'_p X' Y \qquad (3.2.18)$$

Por otro lado de (3.2.6) :

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} \lambda_1^{-1} & & & 0 \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ 0 & & & \lambda_p^{-1} \end{bmatrix} P'Y = \begin{bmatrix} \lambda_1^{-1} & & & 0 \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ 0 & & & \lambda_p^{-1} \end{bmatrix} T X Y$$

$$= \begin{bmatrix} \lambda_1^{-1} t'_1 X' Y \\ \lambda_1^{-1} t'_1 X' Y \\ \cdot \\ \cdot \\ \cdot \\ \lambda_1^{-1} t'_1 X' Y \end{bmatrix}_{p\times p}$$

Entonces

$$\alpha_j = \lambda_j^{-1} t'_j X' Y \ , \ j = 1, 2, \dots, p$$

ó

$$\lambda_j \alpha_j = t'_j X' Y \ , \ j = 1, 2, \dots, p \qquad (3.2.19)$$

Luego, (3.2.19) en (3.2.18)

$$R^2 = \lambda_1 \hat{\alpha}_1^2 + \lambda_2 \hat{\alpha}_2^2 + \dots + \lambda_p \hat{\alpha}_p^2$$

Pero si se eliminan los primeros r-autovalores correspondientes a r-multicolinealidades se tiene

$$R^2 = \lambda_{r+1}\alpha_{r+1}^2 + \lambda_{r+2}\alpha_{r+2}^2 + \dots + \lambda_p\alpha_p^2 \qquad (3.2.20)$$

Por tanto de (3.2.18) y (3.2.20)

$$( r_y, p_j )^2 = \alpha_j t_j' X' Y = \lambda_j \hat{\alpha}_j^2 \ , \ j = r+1, \dots ,p$$

De (3.2.20) también se concluye que el coeficiente de determinación entre $Y$ y los (p - r) componentes principales es

$$R^2 = \sum_{j=r+1}^{p} \lambda_j \hat{\alpha}_j^2 \qquad (3.2.21)$$

donde $\alpha$ se calcula de (3.2.6).

**PROPIEDAD 5**

$\hat{\beta}_{CP}$ minimiza la suma de cuadrados de los residuales

$$\phi( \hat{\beta}_{CP} ) = ( Y - X\hat{\beta}_{CP} )'( Y - X\hat{\beta}_{CP} )$$

Comentario sobre las propiedades :

Como se anotó al inicio de esta sección, el $\hat{\beta}_{CP}$ es sesgado, esto se ve de la propiedad 1, donde el sesgo del estimador es $\sum_{j=1}^{r}( t_j'\beta )t_j$ , luego si este término fuera igual a cero, es decir $( t_j'\beta ) = 0$ (j = 1, ... ,r) el estimador $\hat{\beta}_{CP}$ sería insesgado lo cual matemáticamente no se puede demostrar, pero si en la práctica este sesgo tiende a ser muy pequeño se obtendrá reducciones substanciales sobre la propiedad 2 y 3, es decir sobre la varianza y ECM para $\beta_{CP}$. La propiedad 4 da la contribución de una componente principal para explicar a la variable $Y$; esta propiedad decrece a medida que componentes adicionales son usadas, en cambio el $R^2$ de (3.2.21) no es posible.

Ventajas del uso de $\hat{\beta}_{CP}$

1) la ventaja es que al quitar términos sobre la base de las magnitudes de

los autovalores se aprovecha los procedimientos de selección de variables y la teoría de la distribución exacta para el estimador.

2) Su aplicación es directa (no es tan confuso)

3) Quitando los componentes relacionados con los autovalores pequeños reduce la varianza, pero incluyendo componentes es probable reducir el sesgo en el estimador.

## 3.3 ESTIMADOR DE RAICES LATENTES ( RL )

Es desarrollado por Hawkins [ 1973] y Webster and Gunst [ 1974] , sigue la misma filosofía del estimador de componentes principales, pero, considera adicionalmente la variable respuesta. Este método de estimación es dirigido a los problemas de multicolinealidad con el objeto de eliminar algunas variables explicatorias o predictoras relacionadas con el problema. Es primariamente descriptivo para indentificar el probable subconjunto a ser eliminado de la regresión original y es útil para establecer la sub-regresión. De esta forma se presenta un procedimiento utilizando los autovalores (llamados también raices latentes) y autovectores (llamados vectores latentes) de la matriz de correlación de variables dependientes (variable respuesta) é independientes (variables predictoras)

### 3.3.1 ESTRUCTURA DEL ESTIMADOR

El modelo de regresión lineal múltiple (1.1) al ser estandarizada en las variables independientes $X^{*}$'s y dependiente Y, (como fue desarrollada en la sección 1.1 , definición 2), da el modelo de regresión estándar siguiente

$$Y^{*} = X\beta + e \qquad (3.3.0)$$

tal que

$$\sum_{=1} Y_i^* = 0 \qquad\qquad Y^{*'} Y^* = 1$$

luego se define la matriz **A** de orden $n\times(p+1)$

$$\mathbf{A} = [\ Y^* : X\ ]$$

que incluye las variables independientes y dependientes estandarizadas de la regresión (3.3.0); luego la matriz de correlación de estas variables es

$$A'A = \begin{bmatrix} Y^{*'} \\ \cdot\cdot \\ X' \end{bmatrix} [\ Y^* : X\ ] = \begin{bmatrix} 1_{(1\times1)} & (Y^{*'}X)_{(1\times p)} \\ (X'Y^*)_{(p\times1)} & (X'X)_{(p\times p)} \end{bmatrix}_{(p+1)\times(p+1)}$$

Los autovalores de **A' A** se calculan de

$$|\ A'A - \xi I\ | = 0 \longleftrightarrow \begin{bmatrix} 1 - \xi & Y^*X \\ X'Y^* & [\ X'X - \xi I\ ] \end{bmatrix} = 0$$

y los autovectores de

$$(\ A'A - \xi_j I\ )\gamma_j = 0 \ , \ j = 0, 1, \ldots , p \qquad (3.3.1)$$

donde

$$\gamma_j' = [\ \gamma_{0j}, \gamma_{1j}, \gamma_{2j}, \ldots , \gamma_{pj}\ ]$$

es el j-ésimo autovector asociado al autovalor $\xi_j$, también

$$\gamma_j' = [\ \gamma_{0j} : \delta_j'\ ] , \qquad j = 0, 1, \ldots, p$$

donde,

$$\delta_j' = [\ \gamma_{1j}, \gamma_{2j}, \ldots , \gamma_{pj}\ ]$$

además

$$\gamma_j'\gamma_i = 0 \;,\quad \gamma_j'\gamma_j = 1 \;,\quad \forall\, j = 0, 1, \ldots, p$$
$$i \neq j$$

ya que son autovectores ortogonales de la matriz de correlación $A'A$.

Como se anotó al inicio del estudio de este estimador, el procedimiento es similar al estimador de componentes principales, por tanto las demostraciones ya probadas en ella, sirven también para el estimador de "RL", luego la matriz ortogonal $T_{(p+1)\times(p+1)}$, esta formada por los autovectores ó vectores latentes de $A'A$.

$$T = [\; \gamma_0, \gamma_1, \ldots, \gamma_p \;]_{(p+1)\times(p+1)}$$

y la matriz diagonal $\Lambda_{A(p+1)\times(p+1)}$ esta formada por los autovalores ó valores latentes de $A'A$.

$$\Lambda_A = \mathrm{diag}(\; \xi_0, \xi_1, \ldots, \xi_p \;)$$

Se debe recordar que los autovalores están ordenados de acuerdo a su magnitud, así

$$0 \leq \xi_0 \leq \xi_1 \leq \xi_2 \leq \ldots \leq \xi_p$$

por lo tanto si $T$ y $\Lambda_A$ están definidas según arriba, entonces

$$T'(A'A)T = \Lambda_A \qquad \text{y} \qquad A'A = T\,\Lambda_A\,T'$$

Antes de continuar buscando la estructura del estimador RL, se debe advertir la existencia de las dependencias lineales si lo hay, en las columnas de la matriz $A$

Para esto, multipliquemos la matriz $A$ por $\gamma_j$, $j = 0, 1, \ldots, p$, luego

$$\begin{bmatrix} Y_1^* & X_{11} & X_{12} & \ldots . X_{1p} \\ Y_2^* & X_{21} & X_{22} & \ldots . X_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ Y_n^* & X_{n1} & X_{n2} & \ldots . X_{np} \end{bmatrix} \begin{bmatrix} \gamma_{0j} \\ \gamma_{1j} \\ \gamma_{2j} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{pj} \end{bmatrix} = \begin{bmatrix} Y_1^* \gamma_{0j} + \sum_{r=1}^{p} X_{1r}\gamma_{rj} \\ Y_2^* \gamma_{0j} + \sum_{r=1}^{p} X_{2r}\gamma_{rj} \\ \ldots \\ Y_1^* \gamma_{0j} + \sum_{r=1}^{p} X_{nr}\gamma_{rj} \end{bmatrix}_{n\times 1}$$

$$\ldots .(3.3.2)$$

De (3.3.1) :

$$(A'A - \xi_j I)\gamma_j = \theta \longleftrightarrow A'A\,\gamma_j = \xi_j\gamma_j \longleftrightarrow \gamma'_j\, A'A\,\gamma_j = \xi_j \underbrace{\gamma'_j\,\gamma_j}$$

Entonces :

$$\xi_j = (A\gamma_j)'(A\gamma_j)$$

luego de (3.3.2)

$$\xi_j = (\, Y_1^* \gamma_{0j} + \sum_{r=1}^{p} X_{1r}\gamma_{rj}\,)^2 + (\, Y_2^* \gamma_{0j} + \sum_{r=1}^{p} X_{2r}\gamma_{rj}\,)^2 + \ldots$$

$$+ (\, Y_n^* \gamma_{0j} + \sum_{r=1}^{p} X_{nr}\gamma_{rj}\,)^2$$

$$\xi_j = \sum_{i=1}^{n} [\; Y_i^* \gamma_{0j} + \sum_{r=1}^{p} X_{ir}\gamma_{rj}\,]^2 \;;\; j = 0, 1, \ldots ,p \qquad (3.3.3)$$

entonces se concluye

a) $\xi_j = \theta$ ,$j=0,1,\ldots,p$; existe una dependencia lineal exacta entre algunas o todas las columnas de $A$.

b) $\xi_j = \theta$ y $\gamma_{0j} \neq \theta$, $j=0,1,\ldots,p$ existe un predictor perfecto, y es definido por

$$Y_i^* = -\gamma_{0j}^{-1} \sum_{r=1}^{p} X_{ir}\gamma_{rj} \;,\; i = 1, \ldots , n$$

ó también,

$$Y_i^* = (Y_i - \bar{Y})/\eta \,,\; i = 1, 2, \ldots, n \qquad \text{(según (1.2))}$$

entonces

$$Y_i = \bar{Y} \quad \eta\gamma_{0j}^{-1}\sum_{r=1}^{p} X_{ir}\gamma_{rj} \qquad i = 1, \ldots , n \qquad (3.3.3a)$$

c) $\xi_j = 0$ y $\gamma_{0j} = 0$ ,$j=0,1,\ldots,p$; existe una dependencia lineal exacta entre las columnas de $\mathbf{X}$, es decir :

$$\sum_{r=1}^{p} X_{ir}\gamma_{rj} = 0$$

Se sabe que en la práctica esto no es tan cierto, pero si las raices latentes, son muy pequeñas entonces existe sintomas de multicolinealidad.

Continuando con la búsqueda de la estructura del estimador RL, se define de (3.3.2) las (p+1) ecuaciones de predicción (si todos $\gamma_{0j} \neq 0$, $j = 0,1,\ldots,p$)

$$\mathbf{Y}_j = \bar{Y}\,\mathbf{1} - \eta\,\gamma_{0j}^{-1}\mathbf{X}\,\delta_j \qquad ,j=0,1,\ldots p \qquad (3.3.4)$$

donde, $\mathbf{Y}_j$ es de dimensión (nx1). Luego estas ecuaciones son de la forma

$$\mathbf{Y} = \beta_0\mathbf{1} + \mathbf{X}\,\beta + \mathbf{e} \qquad (3.3.4a)$$

cuyo estimador por "MCO" es

$$\hat{\mathbf{Y}} = \bar{Y}\,\mathbf{1} + \mathbf{X}\,\hat{\beta} \qquad (3.3.4b)$$

Luego (3.3.4), es de la forma de (3.3.4b)

donde, $\beta = -\eta\gamma_{0j}^{-1}\delta_j$

Si la j-ésima ecuación de predicción es usada para producir la variable dependiente, facilmente se verifica que la suma de cuadrados de los residuales (SCRes) para este predictor es

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_{ij})^2 = (\mathbf{Y} - \hat{\mathbf{Y}}_j)'(\mathbf{Y} - \hat{\mathbf{Y}}_j) = \eta^2\lambda_j/\gamma_{0j}^2 \qquad [1]$$

Las ecuaciones en (3.3.4) si se usan solas, normalmente no son buenas predictoras, pero usandolas como combinación lineal de estas ecuaciones

1 Se demuestra en C, apéndice I que

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_{ij}\right)^2 = \left(\mathbf{Y}-\hat{\mathbf{Y}}_j\right)'\left(\mathbf{Y}-\hat{\mathbf{Y}}_j\right) = \eta^2\lambda_j/\gamma_{0j}^2$$

predictoras se estiman los parámetros del modelo (3.3.0), así, considerando una combinación lineal arbitraria de estas ecuaciones predictoras

$$\hat{\mathbf{Y}} = \sum_{j=0}^{p} a_j \gamma_{0j} \hat{\mathbf{Y}}_j \tag{3.3.5}$$

con

$$\sum_{j=0}^{p} a_j \gamma_{0j} = 1$$

Luego de (3.3.4) en (3.3.5)

$$\hat{\mathbf{Y}} = \sum_{j=0}^{p} a_j \gamma_{0j} [\ \bar{Y}\mathbf{1} - \eta^{-1} \gamma_{0j} \mathbf{X}\, \delta_j\ ]$$

$$= \bar{Y}\mathbf{1} \sum_{j=0}^{p} (a_j \gamma_{0j}) - \eta\, \mathbf{X} \sum_{j=0}^{p} (a_j \gamma_{0j} \gamma_{0j}^{-1} \delta_j)$$

entonces

$$\hat{\mathbf{Y}} = \bar{Y}\mathbf{1} - \eta\, \mathbf{X} \sum_{j=0}^{p} (a_j \delta_j) \tag{3.3.6}$$

Nuevamente (3.3.6) tiene la forma de (3.3.4a), por lo tanto

$$= -\eta \sum_{j=0}^{p} (\ a_j \delta_j\ )$$

y la suma cuadrados de los residuales para este caso es

$$\begin{aligned} \text{SCRes} &= \mathbf{e}'\mathbf{e} \\ &= (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) \\ &= \eta^2 \mathbf{a}' \Lambda_{\mathbf{A}}\, \mathbf{a} \\ &= \eta^2 \sum_{j=0}^{p} a_j^2 \xi_j, \qquad \mathbf{a}' = (a_0, a_1, \ldots, a_p) \end{aligned} \tag{3.3.7}$$

como (3.3.5), es la ecuación predictora más coveniente del modelo (3.3.0), entonces se encontra los valores de $a_j$ tal que, para estos valores encontrados la suma cuadrados de los residuales de (3.3.7) sea mínima; para esto se usa los multiplicadores de Lagrange

$$f(a) = \eta^2 \sum_{j=0}^{p} a_j^2 \xi_j - 2\mu_0 \left[ \sum_{j=0}^{p} a_j \gamma_{0j} - 1 \right] \tag{3.3.8}$$

Derivando

$$\frac{\partial f(a)}{\partial a_j} = 2\eta^2 a_j \xi_j - 2\mu_o \gamma_{oj} \qquad (3.3.9)$$

Igualando (3.3.9) a cero se obtiene

$$a_j = \eta^{-2} \xi_j^{-1} \gamma_{oj} \mu_o \ , \ j = 0, 1, \ldots ,p \qquad (3.3.10)$$

Multiplicando (3.3.10) por $\gamma_{oj}$ :

$$a_j \gamma_{oj} = \eta^{-2} \xi_j^{-1} \gamma_{oj}^2 \mu_o$$

Luego sumando

$$\sum_{j=o}^{p} a_j \gamma_{oj} = \mu_o \eta^{-2} \sum_{j=o}^{p} \xi_j^{-1} \gamma_{oj}^2$$

Por la condición de (3.3.5)

$$1 = \mu_o \eta^{-2} \sum_{j=o}^{p} \xi_j^{-1} \gamma_{oj}^2$$

Luego

$$\mu_o = \eta^2 \left[ \sum_{j=o}^{p} \xi_j^{-1} \gamma_{oj}^2 \right]^{-1} \qquad (3.3.11)$$

Entonces (3.3.11) en (3.3.10) :

$$a_j = \xi_j^{-1} \gamma_{oj} \left[ \sum_{l=o}^{p} \xi_l^{-1} \gamma_{ol}^2 \right]^{-1}, \ j = 0, 1, \ldots ,p \qquad (3.3.12)$$

Por consiguiente si se reemplaza (3.3.12) en (3.3.6)

$$\hat{\mathbf{Y}} = \bar{Y}1 - \eta \mathbf{X} \sum_{j=o}^{p} \left[ \xi_j^{-1} \gamma_{oj} \left[ \sum_{l=o}^{p} \xi_l^{-1} \gamma_{ol}^2 \right]^{-1} \delta_j \right]$$

$$\hat{\mathbf{Y}} = \bar{Y}1 - \eta \mathbf{X} \left[ \sum_{l=o}^{p} \xi_l^{-1} \gamma_{ol}^2 \right]^{-1} \sum_{j=o}^{p} \left[ \xi_j^{-1} \gamma_{oj} \delta_j \right]$$

Por lo tanto el vector $\beta$ estimado es :

$$\hat{\beta} = -\eta \left[ \sum_{l=o}^{p} \xi_l^{-1} \gamma_{ol}^2 \right]^{-1} \sum_{j=o}^{p} \left[ \xi_j^{-1} \gamma_{oj} \delta_j \right] \qquad (3.3.13)$$

También reemplazando (3.3.12) en (3.3.7) se obtiene

$$SCRes = (Y - \hat{Y})'(Y - \hat{Y}) = \eta^2\left(\sum_{j=0}^{p}\left[\xi_j^{-1}\gamma_{0j}^2\right]\right)^{-1} = \mu_0 \qquad (3.3.14)$$

Supongase que al efectuar un test de multicolinealidad se encuentra que existe r-multicolinealidades, esto significa que hay "r" raices latentes con magnitud muy pequeña, luego haciendo estas r raices iguales a cero, se encuentra la estructura general para el estimador de raíz latente de (3.3.13)

$$\hat{\beta}_{RL} = -\eta\left(\sum_{l=r}^{p}\xi_l^{-1}\gamma_{0l}^2\right)^{-1}\sum_{j=r}^{p}\left[\xi_j^{-1}\gamma_{0j}\delta_j\right] \qquad (3.3.15)$$

donde $\xi_0 = \xi_1 = \xi_2 = \ldots = \xi_{r-1} = 0$

Analogamente (3.3.14) se convertirá en

$$SCRes^* = (Y - \hat{Y})'(Y - \hat{Y}) = \eta^2\left(\sum_{j=r}^{p}\left[\xi_j^{-1}\gamma_{0j}^2\right]\right)^{-1} \qquad (3.3.16)$$

donde se ha eliminado $\xi_0, \xi_1, \ldots, \xi_{r-1}$ igualando a cero.

### 3.3.2 CRITERIO PARA ELIMINAR TERMINOS

Para este estimador solo hay un criterio para eliminar términos y se traduce en que

$$\xi_j \le 0.01, \quad y \quad |\gamma_{0j}| < 0.05, \quad j = 0,1, \ldots, p.$$

Luego aplicando el procedimiento de raíz latente se debe fijar en las magnitudes de los autovalores y en su correspondiente $|\gamma_{0j}|$.

### 3.3.3 PROPIEDADES DEL ESTIMADOR

1) La suma de cuadrados de los residuales del estimador "RL" disminuye ligeramente comparado con el estimador "MCO" (propiedad muy importante)
2) Reduce la magnitud del vector de coeficientes estimados.

3) Existe indicios de que la varianza total del estimador es pequeña.

4) Identifica la multicolinealidad en **X**.

5) Determina si las multicolinealidades tienen valor predictivo.

6) Removiendo los $\xi_j$ pequeños produce un estimador con más próximidad al valor verdadero que $\beta$.

Como no se conoce las propiedades distribucionales del estimador RL, las propiedades anteriores, no serán verificadas.

### 3.3.4 DESVENTAJAS DEL ESTIMADOR "RL"

La desventaja es que no se puede calcular el valor esperado, la varianza, error cuadrático medio del estimador "RL", por que no se conocen la propiedades distribucionales y la falta de otro mecanismo para elegir $|\gamma_{oj}|$ y $\xi_j$. No se ha usado ampliamente.

## 3.4 ESTIMADOR DE REGRESION RIDGE (RR) Y RIDGE GENERALIZADO (RG)

En esta sección se presenta un nuevo procedimiento de estimación del vector de coeficientes $\beta$ del modelo de regresión lineal múltiple cuando existe multicolinealidad, esta técnica fue originalmente desarrollada por Arthur E. Hoerl y Robert W. Kennard (1970). Esta técnica, se basa en la adición de cantidades positivas pequeñas **k**, a la diagonal de la matriz **X'X** y se presenta por el **método de la traza ridge**, para la demostración en dos dimensiones de los efectos de la no-ortogonlidad. Luego, la matriz aumentada **X'X** sirve para obtener la estimación sesgada del vector $\beta$ con error cuadrático medio pequeño y suma de cuadrados de los residuales mínima.

Usando el modelo (1.3), se encontrará la estructura de este estimador, luego el valor óptimo adecuado de "k" y finalmente sus propiedades, ventajas y desventajas de este estimador.

### 3.4.1 Estructura del estimador Ridge (RR)

Sea B alguna estimación del vector de coeficientes $\beta$ en el modelo (1.3), entonces la suma de cuadrados de los residuales será

$$\begin{aligned}
\phi &= (Y - XB)'(Y - XB) \\
&= (Y - X\beta + X\beta - XB)'(Y - X\beta + X\beta - XB) \\
&= (Y - X\beta)'(Y - X\beta) + (Y - X\beta)'(X\beta - XB) + (X\beta - XB)'(Y - X\beta) \\
&\quad + (X\beta - XB)'(X\beta - XB) \\
&= (Y - X\beta)'(Y - X\beta) + Y'X\beta - Y'XB - \beta'X'X\beta + \beta'X'XB + \beta'X'Y \\
&\quad - \beta'X'X\beta - B'X'Y + B'X'X\beta + \beta'X'X\beta - \beta'X'XB - B'X'X\beta + B'X'XB \\
&= (Y - X\beta)'(Y - X\beta) + Y'X\beta - Y'XB + \beta'X'Y - B'X'Y - \beta'X'X\beta + B'X'XB \\
&= (Y - X\beta)'(Y - X\beta) + \beta'X'X\beta - \beta'X'XB + \beta'X'X\beta - B'X'X\beta - \beta'X'X\beta \\
&\quad + B'X'XB \\
&= (Y - X\beta)'(Y - X\beta) + B'X'XB - B'X'X\beta - \beta'X'XB + \beta'X'X\beta \\
&= (Y - X\beta)'(Y - X\beta) + (B - \beta)'X'X(B - \beta) \qquad (3.4.1)
\end{aligned}$$

entonces $$\Phi = \Phi_{min} + \Phi(B) \qquad (3.4.2)$$

Donde

$$\Phi_{min} = (Y - X\beta)'(Y - X\beta) \qquad y \qquad \Phi(B) = (B - \beta)'X'X(B - \beta)$$

Como se aprecia, el contorno de $\Phi$ es la superficie de la hiperelipsoide centrada en $\beta$, el estimador de cuadrados mínimos ordinario de $\beta$.

Mediante el **método de la traza Ridge**, se encuentra el estimador Ridge, esto se establece a través de la suma de cuadrados de $B'B$; tal que para un $\Phi$ fijo, un valor B de longitud mínima es elegido. Para demostrar esto, de (3.4.2), existe un valor continuo $B_o$ que satisface la relación

$$\Phi = \Phi_{min} + \Phi(B_o) = \Phi_{min} + \Phi_o \;; \qquad \Phi(B_o) = \Phi_o$$

donde $\Phi_o > 0$ es un incremento fijo, pero $\Phi$ es mínimo para este incremento, luego, minimizando $B'B$ Sujeto a

$$(B - \beta)' X' X (B - \beta) = \Phi_o \qquad (3.4.2a)$$

como un problema de Lagrange. Se tiene

$$F = B'B + (\frac{1}{k})\left[ (B - \hat{\beta})' X' X (B - \hat{\beta}) - \Phi_o \right] \qquad (3.4.3)$$

donde

$(\frac{1}{k})$ es el multiplicador [ podriá ser $(\frac{1}{k_j})$, j = 1, ,p pero se prefiere $(\frac{1}{k})$, más adelante se tiene información sobre este tema en el criterio N° 1 de Hoerl y Kennard, página 64] .

entonces $$\frac{\partial F}{\partial B} = 2B + (\frac{1}{k})\left[ 2(X'X)B - 2(X'X)\hat{\beta} \right] = 0$$

Por consiguiente

$$(\frac{1}{k})(X'X + KI)B = (\frac{1}{k})X'X\hat{\beta}$$

$$B = (X'X + KI)^{-1}X'X\hat{\beta}$$

$$B = (X'X + KI)^{-1}X'Y$$

De esta forma se encontra el valor de $B$ que hace mínimo $\Phi$ y esto es $B = B_o$ ó támbién

$$B_o = \beta_{RR} = (X'X + KI)^{-1}X'Y \qquad (3.4.4)$$

donde se incluye K para cumplir la restricción de (3.4.2a). Luego, a "$\beta_{RR}$"hallado se llama "estimador Ridge". Por lo tanto con un K elegido y luego estimando $\beta_{RR}$ se calcula $\Phi_o$. Por esta razón al proceso de estimación y análisis de la estructura de (3.4.4.) se ha llamado "Regresión Ridge". También se encuentra la relación entre el estimador Ridge y el estimador "MCO" como se aprecia a continuación

$$\beta_{RR} = (X'X + KI)^{-1}X'Y$$

$$= (X'X + KI(X'X)^{-1}(X'X))^{-1}X'Y$$

$$= (I + K(X'X)^{-1})^{-1}(X'X)^{-1}X'Y$$

$$\hat{\beta}_{RR} = (I + K(X'X)^{-1})^{-1}\hat{\beta} \qquad (3.4.4a)$$

Luego los autovalores de las matrices :

$$(X'X + KI)^{-1} \qquad y \qquad (I + K(X'X)^{-1})^{-1}$$ [2]

son : $\dfrac{1}{\lambda_j + K}$ y $\dfrac{\lambda_i}{\lambda_j + K}$ respectivamente donde $\lambda_i$ son los autovalores de $X'X$.

Otra forma de escribir (3.4.4) es

$$\hat{\beta}_{RR} = (T\Lambda^{*}T')^{-1}X'Y \qquad (3.4.4b)$$

donde $\Lambda^{*}$ es la matriz diagonal, cuyos elementos son los autovalores de la matriz $(X'X + KI)$ y T la matriz ortogonal, cuyos elementos son los autovectores correspondientes a los autovalores de la matriz $X'X$, luego

$$\hat{\beta}_{RR} = T\Lambda^{*-1}T'X'Y$$

ó en forma de sumatoria:

$$\hat{\beta}_{RR} = \sum_{j=1}^{p}(\lambda_j + K)^{-1}T_j'X'YT_j \qquad K \geq 0 \qquad (3.4.5)$$

De esta manera se encuentra la estructura del estimador Ridge para un caso particular $K \geq 0$.

### 3.4.2 ESTRUCTURA DEL ESTIMADOR REGRESION RIDGE GENERALIZADO (RG)

La estructura del estimador de regresión Ridge Generalizado se encuentra a partir del modelo general (1.3), pero expresado en la forma canónica, es decir como la matriz $(X'X)$ es simétrica y real, existe una matriz ortogonal T tal que

---

2 Se demuestra en D, apéndice I, como se encuentra los autovalores de $(X'X + KI)^{-1}$ y $(I + K(X'X)^{-1})^{-1}$

$$X'X = T \Lambda T'$$

Donde $\Lambda$ y T, son la matriz diagonal de autovalores y autovectores de $X'X$, respectivamente.

Ahora sea

$$X = X^{+}T'$$

entonces reemplazando $X$ en (1.3) se obtiene

$$Y = X^{+}\alpha + e \qquad (3.4.5a)$$

(modelo de regresion lineal en forma canonica)

Donde :

$$\alpha = T'\beta \qquad (X^{+})'(X^{+}) = \Lambda \;; \qquad \alpha'\alpha = \beta'\beta \qquad (3.4.5b)$$

Por tanto, el procedimiento de "estimación Ridge general (RG)" siguiendo la estructura del estimador $\beta_{RR}$ se define como

$$\hat{\alpha}_{RG} = [\ (X^{+})'(X^{+}) + \mathbb{K}\ ]^{-1}(X^{+})'Y \qquad (3.4.6)$$

Donde :

$\mathbb{K}$ = diag ( $K_1, K_2, ..., K_p$ ), matriz diagonal tal que $K_i \geq 0$, i= 1, 2, ..,p.

Por consiguiente de (3.4.5b) y (3.4.6)

$$\beta_{RG} = T\ \alpha_{RG} \qquad (3.4.6a)$$

Así , se encuentra la estructura del estimador "RG" en forma matricial en el cual, se busca un $K_i \geq 0$, para cada variable canónica definida por $X^{+}$.

(3.4.6a) también puede ser expresado en forma de sumatoria:

$$\beta_{RG} = \sum_{j=1}^{p} (\lambda_j + k_j)^{-1} T_j' X' Y\ T_j \qquad (3.4.6b)$$

De esta manera se encuentra la estructura del estimador Ridge Generalizado, cuya matriz diagonal $\mathbb{K}_{p\times p}$ tiene valores mayor o igual a cero, los cuales son encontrados con los algoritmos descritos abajo.

### 3.4.3 PROPIEDADES DEL ESTIMADOR $\beta_{RR}$

(Las siguientes propiedades también se cumplen para $\alpha_{RO}$ )

**PROPIEDAD 1**

$$[\ I + K(X'X)^{-1}\ ]^{-1} = I - K(\ X'X + KI\ )^{-1} \qquad (3.4.7)$$

**Prueba**

$$Z = [\ I + K(X'X)^{-1}]^{-1} \quad y \quad W = (X'X + KI)^{-1} \qquad (3.4.7a)$$

Entonces

$$Z = [\ (X'X)(X'X)^{-1} + K(X'X)^{-1}]^{-1}$$

$$Z = [\ X'X + KI\ ]^{-1}X'X \qquad (3.4.7b)$$

$$Z = WX'X$$

$$W^{-1}Z = X'X$$

$$[\ X'X + KI\ ]\ Z = X'X \qquad (3.4.7c)$$

$$(X'X)Z + K[\ I + K(X'X)^{-1}]^{-1} = X'X$$

$$(X'X)Z = -\ KI$$

$$Z = \quad K(X'X)^{-1}$$

$$Z = I - I - K(X'X)^{-1}$$

$$Z = I - K[\ X'X + KI\ ]^{-1}$$

**PROPIEDAD 2**

$\beta_{RR}$ para K > 0 es menor que $\beta$, es decir

$$\beta'_{RR}\beta_{RR} < \beta'\beta \qquad (3.4.8)$$

**Prueba**

Por (3.4.4a) y (3.4.7a) $\qquad \beta_{RR} = Z\beta$

Por la estructura de $X'X$, $Z$ es definida simetrica positiva, luego se cumple lo siguiente :

$$\beta'_{RR}\beta_{RR} = \beta' Z' Z\beta$$

pero $$\beta' Z' Z\beta \le \frac{\lambda_p^2}{(\lambda_p + K)^2} < \beta'\beta$$

Pues $$\lambda_p \ge \lambda_{p-1} \ge \quad \ge \lambda_1 > 0 \qquad \blacksquare$$

**PROPIEDAD 3**

La suma de cuadrados de los residuales es definido por

$$\phi(K) = (Y - X\beta_{RR})'(Y - X\beta_{RR})$$

entonces :

$$\phi(K) = (Y - X\beta\ )'(Y - X\beta\ ) + K^2\hat{\beta}'_{RR}(X'X)^{-1}\hat{\beta}_{RR} \qquad (3.4.9)$$

**Prueba**

Por (3.4.1)

$$\phi(K) = (Y - X\beta)'(Y - X\beta) + (\beta_{RR} - \beta\ )'(X'X)(\beta_{RR} - \beta)$$

Por (3.4.4a) :

$$\phi(k) = \phi_{min} + (\hat{\beta}_{RR} - (I + K(X'X)^{-1})\hat{\beta}_{RR})'(X'X)(\hat{\beta}_{RR} - (I + K(X'X)^{-1})\hat{\beta}_{RR})$$

$$\phi(K) = \phi_{min} + K\hat{\beta}'_{RR}(X'X)^{-1}(X'X)K(X'X)^{-1}\hat{\beta}_{RR}$$

$$\phi(K) = \phi_{min} + K^2\hat{\beta}'_{RR}(X'X)^{-1}\hat{\beta}_{RR} \qquad \blacksquare$$

**PROPIEDAD 4**

$$E(\hat{\beta}_{RR}) = \beta - K\sum_{j=1}^{p}(\lambda_j + K)^{-1}(T'_j\beta)T_j \qquad (3.4.10)$$

**Prueba**

De (3.4.4a) $$\beta_{RR} = [\ I + K(X'X)^{-1}\ ]^{-1}\hat{\beta}$$

Por (3.4.7) $$\beta_{RR} = [\ I - K(X'X + KI)^{-1}\ ]\beta$$

$$E(\ \beta_{RR}\ ) = [\ I - K(X'X + KI)^{-1}\ ]E\beta$$

$$E(\hat{\beta}_{RR}) = \beta - K(X'X + KI)^{-1}\beta$$

Por (3.4.4) y (3.4.4b)

$$E(\hat{\beta}_{RR}) = \beta - K(T\Lambda^{*}T')^{-1}\beta$$

Siguiendo la estructura de (3.4.5)

$$E(\hat{\beta}_{RR}) = \beta - K\sum^{p}(\lambda_j + K)^{-1}(T'_j\beta)T_j \qquad \blacksquare$$

**PROPIEDAD 5**

$$ECM(\hat{\beta}_{RR}) = \sigma^2\sum_{j=1}^{p}\lambda_j(\lambda_j + K)^{-2} + K^2\sum_{j=1}^{p}(\lambda_j + K)^{-2}(T'_j\beta)^2 \qquad (3.4.11)$$

<u>Prueba</u>

Por definición

$$ECM(\hat{\beta}_{RR}) = E(\hat{\beta}_{RR} - \beta)'(\hat{\beta}_{RR} - \beta)$$

También se puede denotar

$$ECM(\hat{\beta}_{RR}) = E[\ L_1^2\ ]$$

(segun Hoerl and Kennard)

De (3.4.4a) y (3.4.7a)

$$ECM(\hat{\beta}_{RR}) = E(Z\hat{\beta} - \beta)'(Z\hat{\beta} - \beta)$$

$$= E(Z\hat{\beta} - Z\beta + Z\beta - \beta)'(Z\hat{\beta} - Z\beta + Z\beta - \beta)$$

$$= E(Z(\hat{\beta} - \beta) + (Z\beta - \beta))'(Z(\hat{\beta} - \beta) + (Z\beta - \beta))$$

$$= E[\ (Z(\hat{\beta} - \beta))'(Z(\hat{\beta} - \beta)) + (Z(\hat{\beta} - \beta))'(Z\beta - \beta)$$

$$+ (Z\beta - \beta)'(z(\hat{\beta} - \beta)) + (Z\beta - \beta)'(Z\beta - \beta)]$$

$$= E[\ (\hat{\beta} - \beta)'Z'Z(\hat{\beta} - \beta)\ ] + (Z\beta - \beta)'(Z\beta - \beta) \ldots\ldots\ldots (3.4.11a)$$

$$= \sigma^2 \text{traza}\left[(X'X)^{-1}Z'Z\right] + \beta'(Z - I)'(Z - I)\beta$$

Por (3.4.7b)

$$ECM(\hat{\beta}_{RR}) = \sigma^2 \text{traza}[\ (X'X)^{-1}((X'X + KI)^{-1}X'X)'((X'X + KI)^{-1}X'X)]$$

$$+ \beta'(Z - I)'(Z - I)\beta$$

$$= \sigma^2 \text{traza}[ (X'X + KI)^{-1}(X'X + KI)^{-1}X'X] + \beta'(Z - I)'(Z - I)\beta$$

Por (3.4.7c)

$$\text{ECM}(\beta_{RR}) = \sigma^2 \text{traza}[ (X'X + KI)^{-1}(X'X + KI)^{-1}(X'X + KI)Z ]$$
$$+ \beta'(Z - I)'(Z - I)\beta$$

$$\text{ECM}(\beta_{RR}) = \sigma^2 \text{traza}[ (X'X + KI)^{-1}Z ] + \beta'(Z - I)'(Z - I)\beta$$

Finalmente por (3.4.7)

$$\text{ECM}(\beta_{RR}) = \sigma^2 \text{traza}[ (X'X + KI)^{-1}(I - K(X'X + KI)^{-1}) ]$$
$$+ \beta'((I - K(X'X + KI)^{-1}) - I)'((I - K(X'X + KI)^{-1}) - I)\beta$$

$$= \sigma^2[ \text{traza}(X'X + KI)^{-1} - K\text{traza}(X'X + KI)^{-2}] + K^2\beta'(X'X + KI)^{-2}\beta$$

$$= \sigma^2 \left[ \sum_{j=1}^{p} \frac{1}{(\lambda_j + K)} - K \sum_{j=1}^{p} \frac{1}{(\lambda_j + K)^2} \right] + K^2\beta'(X'X + KI)^{-2}\beta$$

$$= \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2} + K^2\beta'(X'X + KI)^{-2}\beta \qquad (3.4.11b)$$

Por (3.4.4b)

$$\text{ECM}(\beta_{RR}) = \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2} + K^2\beta'(T\Lambda^*T')^{-2}\beta$$

$$= \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2} + K^2 \sum_{j=1}^{p} (\lambda_j + K)^{-2}(T'_j\beta)^2$$

ó también :

$$\text{ECM}(\beta_{RR}) = \gamma_1(K) + \gamma_2(K) \qquad (3.4.11c)$$

donde $\gamma_2(K)$ es la distancia al cuadrado de $Z\beta$ a $\beta$ (esto se ve de (3.4.11a)). Esta distancia es cero si K = 0 (es decir como $Z = I - K(X'X + KI)^{-1}$, entonces $Z = I$, cuando K = 0). Luego $\gamma_2(K)$ es considerado el sesgo al cuadrado introducido al usar $\beta_{RR}$, en lugar de $\beta$. $\gamma_1(K)$ es la suma de las varianzas de los elementos del vector $\beta_{RR}$ estimado (varianza total). Otra forma de notación del error cuadrático medio es

$$\text{ECM}( \hat{\beta}_{RR} ) = E[ L_1^2(K) ] .$$

PROPIEDAD 6

$$\text{Var - Cov}(\hat{\beta}_{RR}) = \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2} T_j T_j' \qquad (3.4.12)$$

Prueba

Se sabe que $\quad \hat{\beta}_{RR} = Z\hat{\beta} = Z(X'X)^{-1}X'Y$

Entonces :

$$\text{Var}(\hat{\beta}_{RR}) = Z(X'X)^{-1}X'\,\text{Var}(Y)X(X'X)^{-1}Z'$$

$$= \sigma^2 Z(X'X)^{-1}X'X(X'X)^{-1}Z'$$

$$= \sigma^2 Z(X'X)^{-1}Z' \qquad (3.4.12a)$$

reemplazando Z por su valor :

$$\text{Var}(\hat{\beta}_{RR}) = \sigma^2 (I + K(X'X)^{-1})^{-1}(X'X)^{-1}((I + K(X'X)^{-1})^{-1})$$

$$= \sigma^2 (X'X + KI)^{-1}(I + K(X'X)^{-1})^{-1}$$

de (3.4.4) y (3.4.4a) se tiene :

$$\text{Var}(\hat{\beta}_{RR}) = \sigma^2 (X'X + KI)^{-1}(X'X + KI)^{-1}X'Y\hat{\beta}^{-1}$$

$$= \sigma^2 (X'X + KI)^{-2}(X'X)$$

$$= \sigma^2 \sum_{j=1}^{p} (\lambda_j + K)^{-2} \lambda_j T_j T_j' \qquad \blacksquare$$

La suma de las varianzas de los $\hat{\beta}_{j_{RR}}$ es la suma de los elementos diagonales de (3.4.12a).

La propiedad 5, también se puede ver graficamente:

la siguiente figura muestra las relaciones entre la varianza, el sesgo al cuadrado y el parámetro K. Se aprecia cuando K crece la varianza total decrece, mientras el sesgo al cuadrado crece; la línea discontinua es la suma de $\gamma_1(K)$ y $\gamma_2(K)$ es decir ECM[$\hat{\beta}_{RR}$] o E[ $L_1^2(K)$] , por lo tanto existe la posibilidad de encontrar valores de K (valores admisibles) tal que ECM[$\hat{\beta}_{RR}$] sea menor que ECM($\hat{\beta}$)

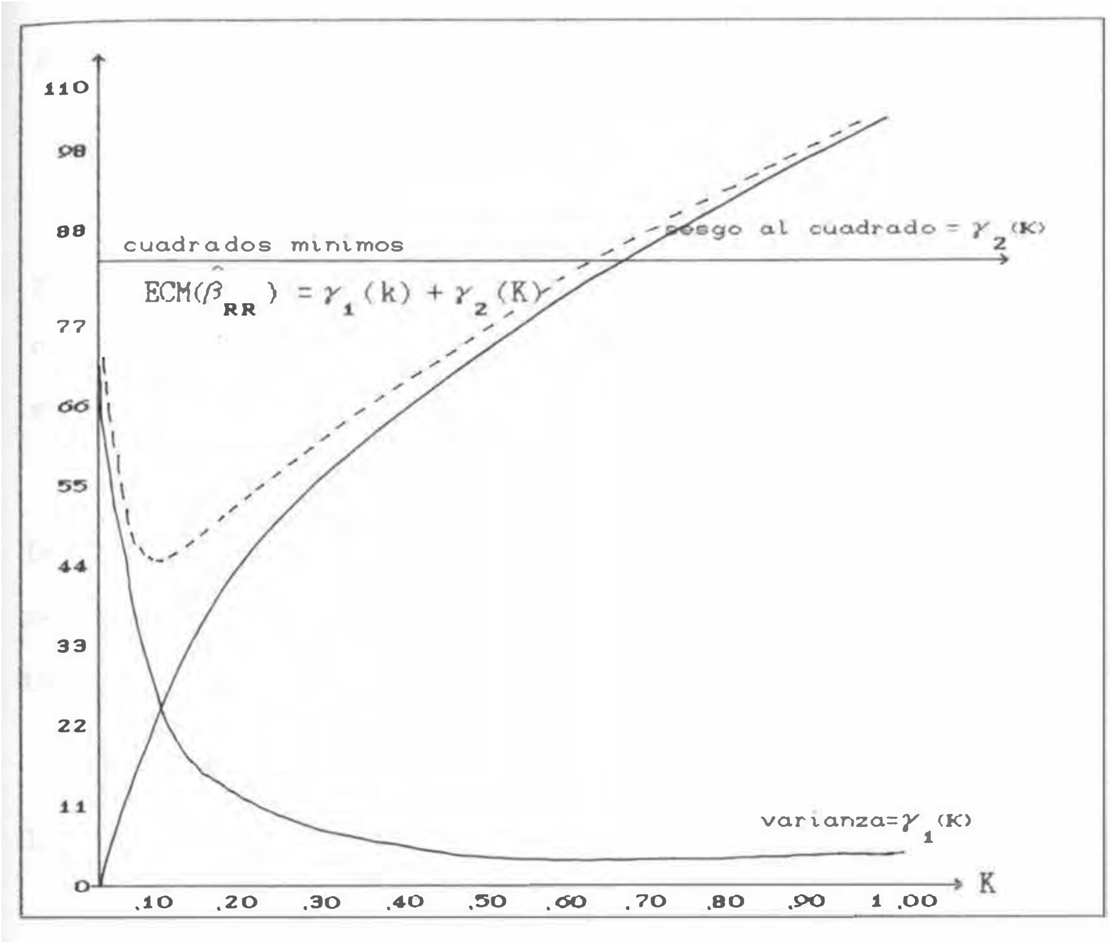


FIGURA N°1: Gráfica del comportamiento de la varianza, el sesgo al cuadrado y el ECM de $\hat{\beta}_{RR}$, cuando crece y decrece K en los intervalos anotados.

PROPIEDAD 7

Siempre existe un $K > 0$ tal que $E[L_1^2(K)] < E[L_1^2(0)]$, es decir

$$ECM(\beta_{RR}) < ECM(\beta) \qquad (3.4.13)$$

PRUEBA

de (3.4.11) se tiene

$$E[L_1^2(K)] = \sigma^2 \sum \lambda_j(\lambda_j + k)^{-2} + k^2\sum(\lambda_j + K)^{-2}(T_j'\beta)^2$$

derivando con respecto a k

$$\frac{dE[L_1^2(K)]}{dk} = \frac{d\gamma_1(K)}{dk} + \frac{d\gamma_2(K)}{dk}$$

$$= -2\sigma^2\sum\lambda_j(\lambda_j + K)^{-3} + 2k\sum\lambda_j(\lambda_j + K)^{-3}(T_j'\beta)^2 \qquad (3.4.13a)$$

Además de (3.4.11c) se tiene

$$E\left[L_1^2(\emptyset)\right] = \gamma_1(\emptyset) + \gamma_2(\emptyset)$$

$$= \sigma^2 \sum (\lambda_j)^{-1}$$

y como las funciones $\gamma_1(k)$ y $\gamma_2(K)$ son monótonamente decreciente y creciente respectivamente (según se demuestra en F, **apéndice I** ), entonces se establece

$$ECM(\beta_{RR}) < ECM(\beta\ )$$

De (3.4.13a) la primera derivada de $\gamma_1(K)$ es no positiva y de $\gamma_2(K)$ es no negativa, luego para mostrar la propiedad es necesario que exista $K > \emptyset$ tal que

$$\frac{dE\left[L_1(K)\right]}{dk} < \emptyset$$

luego de (3.4.13a) se tiene

$$-2\sigma^2 \sum \lambda_j(\lambda_j + K)^{-3} + 2k \sum \lambda_j(\lambda_j + K)^{-3}(T_j'\beta)^2 < \emptyset$$

entonces

$$k < \frac{\sigma^2}{(T_j'\beta)^2} \qquad \text{ó} \qquad K < \frac{\sigma^2}{\beta' T_j' T_j \beta}$$

finalmente

$$K < \frac{\sigma^2}{\beta'\beta}$$

por lo tanto siempre existe un " K "y además $ECM(\beta_{RR})$ es minimizado cuando

$$k = \frac{\sigma^2}{\beta'\beta}$$

pero, el problema es que en la práctica no se conoce $\sigma^2$ y $\beta$. La conclusión será entonces, que "si existen valores admisibles de K " tal que

$$ECM(\beta_{RR}) < ECM(\beta)$$

esto es sostenido por las propiedades matemáticas de $\gamma_1(K)$ y $\gamma_2(K)$, pues estas funciones son monótonas decrecientes y crecientes de K respectivamente. Luego todas estas propiedades conducen a encontrar K > 0 introduciendo un sesgo pequeño y reduciendo substancialmmente la varianza, del mismo modo mejorando el "ECM" de estimación y predicción.

Se presenta, en las líneas siguientes algunos teoremas y corolarios que prueban que $\gamma_1(k)$, $\gamma_2(k)$ son monótonas decrecientes y crecientes respectivamente así como probar la existencia de k > 0.

### 3.4.4 TEOREMAS IMPORTANTES [1]

**TEOREMA 1** La varianza total $\gamma_1(K)$ es una función continua, monótonamente decreciente de K.

**COROLARIO 1**

$$\frac{d\gamma_1(K)}{dK} \longrightarrow -\infty \text{ cuando } k \longrightarrow 0^+ \quad \text{y} \quad \lambda_1 \longrightarrow 0$$

**TEOREMA 2** El sesgo al cuadrado $\gamma_2(K)$ es una función continua, monótonamente creciente de K.

**COROLARIO 2** $\gamma_2(K)$ se aproxima a $\beta'\beta$ como un límite superior.

**COROLARIO 3**

$$\frac{d\gamma_1(K)}{dK} \longrightarrow 0 \quad \text{cuando } k \longrightarrow 0^+$$

**TEOREMA 3** (Teorema de Existencia) Siempre existe un K > 0 tal que

$$E\left[L_1^2(K)\right] < E\left[L_1^2(0)\right] = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

### 3.4.5 CRITERIOS PARA LA SELECCION OPTIMA DE K EN $\beta_{RR}$ Y $K_i$ EN $\alpha_{RG}$

En esta sección se presenta varios criterios de selección óptima de K

[1] Es demostrado en E, apéndice I, los teoremas 1,2,3 y los corolarios 2,3.

como también de $K_i$, $i = 1,2,\ldots,p$. Se aprecia primero para K.

CRITERIOS DE SELECCION OPTIMA DE K EN $\beta_{RR}$

1. CRITERIO DE HOERL Y KENNARD.

Hoerl and Kennard(1970), proponen, que la elección de la constante K debe estar en el intervalo $[0,1]$, fuera de este dominio no es aconsejable usar. Ellos usan el " METODO DE LA TRAZA RIDGE es decir se debe encontrar $\beta_{RR}(K)$ para diversos valores de $K \in [0,1]$ y luego graficar $\beta_{RR}(K)$ versus K, entonces inspeccionando la gráfica, se selecciona un K más conveniente tal que

$$ECM(\beta_{RR}(K)) < ECM(\beta)$$

y $R^2_{RR}(K)$ (coeficiente de correlación múltiple usando $\beta_{RR}(K)$ ), sea ligeramente menor que $R^2$ (coeficiente de correlación múltiple usando $\hat{\beta}$).

**Algunas criticas sobre el K elejido:**

* El problema, es cuando el efecto de multicolinealidad es bastante severa, los coeficientes estimados de $\beta_{RR}(K)$ por el método de la traza Ridge son inestable para K creciente.
* La suma de cuadrados de los residuales de $\beta_{RR}(K)$, crece y a veces es mayor al del estimador de $\beta$, cuando K crece. Esto no es de importancia cuando interesa más la estabilidad de los parámetros estimados.
* $R^2_{RR}(K)$ decrece cuando K crece.
* El método de la traza Ridge es un "procedimiento subjetivo" necesitando un juicio analítico sobre la parte que se está analizando.

2. CRITERIO DE MCDONALD Y GALARNEAU.

MacDonald.G and Galarneau.D (1975), estipulan la siguiente regla para seleccionar un valor de "K":

a) En (3.4.5) seleccionar $K = 0$, entonces $\beta_{RR}(0) = \beta$, el cual es evaluado. Enseguida se calcula

$$Q = \beta' \hat{\beta} - \hat{\sigma}^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

donde Q es el estimador de $\beta'\beta$, pues

$$ECM(\beta) = E(\beta - \beta)'(\beta - \beta) = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

$$E(\beta'\beta) - \beta'\beta = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

$$E(\beta'\beta) - \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1} = \beta'\beta$$

y

$$\hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{(n - p - 1)}$$

b) seleccionar

K > 0 tal que $\beta'_{RR}(K)\beta_{RR}(K) = Q$, si Q > 0

en otro caso seleccionar:

$$K = 0 \quad (\text{entonces } \beta_{RR} = \beta)$$

c) seleccionar

K > 0 tal que $\beta'_{RR}(K)\beta_{RR}(K) = Q$, si Q > 0

en otro caso seleccionar:

$$K = \infty \quad (\text{entonces } \beta_{RR} = 0).$$

La parte b) y c) da dos posibilidades en el sentido, que cuando el valor de K tiende a cero el estimador Ridge se transforma en el estimador de MCO; lo contrario, cuando K tiende a valores grandes el estimador Ridge es cero.

Esta regla es correcta pues la longitud al cuadrado de $\beta_{RR}$ es una función decreciente de K y se aproxima a cero cuando $K \longrightarrow \infty$.

3. CRITERIO DE HOERL, KENNARD Y BALDWIN.

Hoerl, Kennard and Baldwin (1975), usan el estimador $\alpha_{RG}$ dado en (3.4.6). Para calcular la matriz $\mathbb{K}$, proponen, el uso de la media armónica de los elementos $K_i$, i = 1,2,...,p de está matriz; para producir una sola constante K, tal que la matriz $\mathbb{K}$ se convierte ahora en $\mathbb{K} = K\mathbb{I}$, y el valor de K es

$$K = p\hat{\sigma}^2/\hat{\beta}'\hat{\beta}$$

donde, $\hat{\sigma}^2$ y $\beta$ se calculan por mínimos cuadrados.

4. CRITERIO DE LINDLEY Y SMITH (1972).

Lindley and Smith (1972), proponen, una aproximación de Bayes para la elección de la constante k, considerando que

$$Y \sim N(X\beta,\sigma^2\mathbb{I}) \qquad y \qquad \beta \sim N(\emptyset,\sigma^2\mathbb{I})$$

entonces es propuesto el siguiente algoritmo

a) Suponer que $\sigma^2$ y $\sigma^2_\beta$ son independientes y cada uno tiene una distribución inversa CHI-CUADRADO es decir

$$\nu\eta/\sigma^2 \sim \chi^2_\nu \qquad y \qquad \nu_\beta\eta_\beta/\sigma^2_\beta \sim \chi^2_{\nu_\beta}$$

b) Estimar inicialmente $\beta$ de (3.4.5) haciendo K = Ø.

c) Usando b) determinar

$$\hat{\sigma}^2 = [\,\nu\eta + (Y - X\beta\;)'(Y - X\beta\;)]\,/(\eta + \nu + 2)$$

y

$$\hat{\sigma}^2_\beta = [\,\nu_\beta\eta_\beta + \hat{\beta}'\hat{\beta}\,]\,/(p + \nu_\beta + 2)$$

d) Usando c) calcular

$$K = \hat{\sigma}^2/\hat{\sigma}^2_\beta$$

e) Usando d) hallar el estimador $\beta_{RR}$ de (3.4.5)

f) Calcular las expresiones presentadas abajo, usando e) y luego volver a d) para calcular K, pero, usando estos resultados.

$$\hat{\sigma}^2 = [\,\nu\eta + (Y - X\hat{\beta}_{RR})'(Y - X\hat{\beta}_{RR})]\,/(\eta + \nu + 2)$$

$$\hat{\sigma}^2_\beta = [\,\nu_\beta\eta_\beta + \hat{\beta}_{RR}'\hat{\beta}_{RR}]\,/(p + \nu_\beta + 2)$$

g) Proceder iterativamente hasta que la diferencia entre los valores de

$$K = \hat{\sigma}^2 / \hat{\sigma}^2_\beta$$

sea del orden de $10^{-4}$, caso contrario falla a los cuadrados mínimos ordinarios si la convergencia no es obtenida.

NOTA: En un estudio especifico la sensitividad de la solución para cambios en los valores de $\nu$ y $\nu_\beta$ pueden ser investigados. A veces se pueden fijar a cero, y en este caso el estimador K encontrado es similar al estimador iterativo de K planteado por HOERL Y KENNARD (1976).

5. 1º CRITERIO DE LAWLESS Y WANG.

Lawless and Wang (1976) adoptan la aproximación de Bayes para calcular la constante K de (3.4.5) proponiendo el siguiente algoritmo:

a) Se Hace K = 0 en (3.4.5) y evaluar

$$\hat{\beta}_{RR}(0) = \hat{\beta}$$

b) Hallar el valor de K tal que sea más próximo a satisfacer la ecuación

$$\sum_{i=1}^{p} \hat{\gamma}_i^2 \; / \; (\sigma^2_\beta + \hat{\sigma}^2 / \lambda_i) = p$$

donde $\gamma = T'\beta$

$$\hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta}{(n - p - 1)}$$

$\sigma^2_\beta = \hat{\sigma}^2/K$ , p = No de variables explicatorias

c) El valor encontrado en b) reemplazar en (3.4.5) y calcular $\beta_{RR}$.

6. CRITERIO DE HOERL Y KENNARD.

Hoerl and Kennard (1976), proponen, un criterio iterativo el cual es una modificación del <u>CRITERIO 3</u> y se desarrolla como sigue

a) Determinar

$$K_0 = p\hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$$

y evaluar $\beta_{RR}(K_0)$ en (3.4.5)

b) Determinar

$$K_1 = (p\,\hat{\sigma}^2) / \hat{\beta}'_{RR}(K_0)\hat{\beta}_{RR}(K_0)$$

y evaluar $\beta_{RR}(K_1)$ en (3.4.5).

c) Continuar el paso b) para $K_2, K_3, \ldots$; hasta que el criterio de convergencia $10^{-4}$ se cumplan para los valores sucesivos de K, en caso contrario ir a d)

d) El criterio falla al estimador de cuadrados mínimos ordinarios si la convergencia no es obtenida en 30 iteraciones

7. CRITERIO DE DEMPSTER, SCHATZOFF Y WERMUTH.

Dempster, Schatzoffy Wermuth (1977), también adoptan la APROXIMACION DE BAYES, y determinan la constante K de la manera siguiente

$$\sum_{j=1}^{p} \left[T'_j \hat{\beta}\right]^2 / (\sigma^2_\beta + \hat{\sigma}^2/\lambda_i) = p \qquad (3.4.14)$$

donde $\sigma^2_\beta = \hat{\sigma}^2 / K$ y $T = (T_1, T_2, \ldots, T_p)$

esta aproximación es implementado, evaluando (3.4.14) para una red de valores de K y determinar el valor, tal que satisfaga la igualdad (3.4.14) lo más próximo.

8. CRITERIO DE GOLUB, HEATH Y WAHBA.

Golub, Heath y Wahba (1979) proponen el criterio de la validación cruzada generalizada, para esto el parámetro K es escrito como

$$K = \tau n$$

y la solución de $\tau$ se obtiene minimizando

$$V(\tau) = \left\| [\ I - A(\tau)\ ]\, Y \right\|^2 / \left\{ \text{traza}\,[\ I - A(\tau)\ ] \right\}^2$$

donde

$$A(\tau) = X(X'X + n\tau I)^{-1}X'$$

Este procedimiento es implementado evaluando $V(\tau)$ para una red de valores de K. El valor K asociado con el mínimo observado define el estimador en (3.4.5).

Concluimos que de todos estos criterios los que mejor se han desarrollado son los criterios 3, 7 y 8, ya que varios estudios de simulación realizados, lo confirman, tales como por DIANE GALARNEAU GIBBONS (1981) a través del método de Monte Carlo.

## CRITERIOS PARA ELEGIR LA MATRIZ $\mathbb{K}$ EN UN ESTIMADOR REGRESION RIDGE GENERALIZADO ( $\alpha_{RG}$ )

Se sabe que $\alpha_{RG} = [\ \Lambda + \mathbb{K}\ ]^{-1}(\ X^{+}\ )'\ Y$

entonces según esta estructura, se presenta los siguientes criterios de elección de la matriz $\mathbb{K}$.

CRITERIO DE HOERL Y KENNARD (1970)

Llamado "Criterio Iterativo", el criterio toma como condición que los valores de la matriz diagonal $\mathbb{K}_{p\times p}$ en (3.4.6) deben ser hallados tal que minimice el error cuadratico medio

$$ECM = E[\ L_1^2\ ] = E[\ (\ \alpha_{RG} - \alpha\ )'(\ \alpha_{RG} - \alpha\ ) \qquad (3.4.15)$$

(3.4.15) también, se expresa como

$$E[\ L_1^2\ ] = \sum_{i=1}^{p} (\sigma^2\lambda_i + \alpha_i^2 K_i^2)/(\lambda_i + K_i)^2 \qquad (3.4.15a)$$

diferenciando (3.4.15a) con respecto a los $K_i$ se obtienen las ecuaciones de minimización

$$\frac{\partial E[\ L_1^2\ ]}{\partial K_i} = \frac{2\,\lambda_i(\ \lambda_i + K_i\ )(\ K_i\alpha_i^2 - \sigma^2\ )}{(\ \lambda_i + K_i\ )^4}$$

$$= 0 \qquad i = 1,\ 2,\ \ldots\ ,\ p \qquad (3.4.15b)$$

Luego de (3.4 15b) se concluye

$$K_i = \sigma^2/\alpha_i^2 \;, \quad i = 1,\, 2,\, \ldots ,p \qquad (3.4.15c)$$

Pero como $K_i$ depende de valores poblacionales desconocidos, se estima por un procedimiento iterativo

a) Hallar los valores $K_i$'s de la matriz diagonal $\mathbb{K}$, en el j-ésimo paso

$$K_{i(j)} = \frac{\hat{\sigma}^2}{\left[\hat{\alpha}_{RR\,i(j)}\right]^2} \qquad \begin{array}{l} i = 1,\, 2,\, \ldots ,p \\ j \geq 0 \end{array}$$

DONDE : $\hat{\sigma}^2$ = El estimador de $\sigma^2$ que se calcula por "MCO"

$\alpha_{RG\,i(O)} = \alpha_i$ , Estimador por mínimos cuadrados ordinarios en (3.4 .5a)

b) Luego se usa la matriz diagonal $\mathbb{K}$ hallada en a) para calcular el próximo $\alpha_{RG\,i(j+1)}$ en (3.4.6)

c) Repetir a) usando b) para j > 0.

d) El algoritmo termina cuando alcanza una estabilización $\alpha'_{RG}\alpha_{RG}$

B. CRITERIO DE HEMMERLE (1975)

El CRITERIO A, ya no es necesario, más bien lo que se hace es limitar los valores de $\alpha_{RG(j)}$, condicionado a la convergencia ó divergencia del procedimiento mencionado, a esto se le llama "Solución explicita del procedimiento iterativo". El algoritmo que se utiliza es descrito a continuación:

El procediemiento iterativo de HOERL Y KENNARD es reducido a la forma simple matricial

$$E_{j+1} = E_O(\, I + E_j\,)^2 \qquad (3.4.16)$$

donde: j = 0, 1, 2,...... es el j-ésimo paso iterativo,

$$E_j = D^{-1} A_j^{-2} \qquad (3.4.16a)$$

$$D = \frac{\Lambda}{\hat{\sigma}^2}$$

$\Lambda$, matriz diagonal de autovalores de $X'X$

$$A_{\cdot} = \begin{bmatrix} \hat{\sigma}^{1(j)}_{RG} & \emptyset & \cdots\cdots & \emptyset \\ \emptyset & \hat{\sigma}^{2(j)}_{RG} & \cdots\cdots & \emptyset \\ \vdots & \vdots & \ddots & \vdots \\ \emptyset & \emptyset & \cdots\cdots & \hat{\sigma}^{p(j)}_{RG} \end{bmatrix}$$

La formula matricial (3.4.16) también se escribir en sus p-expresiones iterativas de la forma :

$$e_{i(J+1)} = e_{i(O)}(\ 1 + e_{i(j)})^2 ,\ i = 1,\ 2, \ldots, p \qquad j \geq 0 \qquad (3.4.16b)$$

Donde

$$e_{i(j+1)} \text{ son escalares}$$

el subcrito j es usado para denotar la j-ésima iteración.

De (3.4.16b) se muestra que el proceso iterativo converge cuando

$$\emptyset < e_{i(O)} \leq 1/4$$

y diverge para

$$e_{i(O)} > 1/4$$

es decir

$$\mathcal{L}im_{j \to \infty}\ e_{i(j)} = e_{i}^{*} \qquad \emptyset < e_{i(O)} \leq 1/4 \qquad (3.4.16c)$$

$$\mathcal{L}im_{j \to \infty}\ e_{i(j)} = \infty \qquad e_{i(O)} > 1/4 \qquad (3.4.16d)$$ [2]

De (3.4.16b) y (3.4.16c) se concluye

$$e_i^{*} = e_{i(O)}(\ 1 + e_i^{*}\ )^2$$

Resolviendo esta ecuación se obtiene

Se demuestra en F, apéndice I, (3.4.16), asi como también la convergencia o divergencia de (3.4.16c) y (3.4.16d).

$$e_i^{*} = \frac{(1-2e_{i(0)}) - \sqrt{1-4e_{i(0)}}}{2e_{i(0)}} \qquad \emptyset < e_{i(0)} \leq 1/4 \qquad (3.4.16e)$$

Para $\qquad e_{i(0)} > 1/4$

diverge la expresión (3.4.16e).

De (3.4.16c) también se concluye que

$$\lim_{j \to \infty} \alpha_{RG\, i(j)} = \alpha_{RG} \; , \qquad \emptyset < e_{i(0)} \leq 1/4$$

ahora expresando (3.4.16a) en forma de sus p-expresiones:

$$e_{i(j)} = \frac{\hat{\sigma}^2}{\lambda_i (\alpha_{RG\, i(j)})^2} \qquad (3.4.16f)$$

entonces se obtiene de (3.4.16c), y las dos últimas expresiones:

$$\lim_{j \to \infty} \alpha_{RG\, i(j)} \longrightarrow \emptyset \; , \qquad e_{i(0)} > 1/4$$

Por consiguiente

$$\alpha_{RG} = \emptyset \; , \qquad e_{i(0)} > 1/4 \qquad (3.4.16g)$$

pero como $\qquad \alpha_{.} = [\, (\Lambda)^{-1} (X^{+})' Y \,]$

entonces $\qquad \alpha_{RG\, i} = \dfrac{\alpha_i}{(1 + e_i^{*})} \qquad \emptyset < e_{i(0)} \leq 1/4 \qquad (3.4.16h)$

donde: $\alpha_{.}$, es el i-esimo coeficiente estimado por MCO en (3.4.5a)

$e_i^{*}$, es evaluado en (3.4.16e) para $o < e_{i(0)} \leq 1/4$

Luego aplicando (3.4.16g) y (3.4.16h) se encuentra una solución óptima del estimador regresión Ridge Generalizado $\alpha_{RG}$ , i = 1, 2, ... ,p.

C. CRITERIO DE GUILKEY Y MURPHY (1975)

El criterio, propone, no alterar todos los elementos diagonales de la matriz de autovalores $\Lambda$, más bien sólo aquellos, cuyos autovalores son relativamente pequeños y consecuentemente la estimación es menos sesgada comparado con los otros métodos de regresión Ridge. Este criterio se conoce como "REGRESION RIDGE DIRIGIDO" y funciona como sigue:

i) Utilizarar la parte a) del "CRITERIO A" de HOERL Y KENNARD (1970), para calcular los elementos $K_i$ de la matriz diagonal $\mathbb{K}$ en (3.4.6) pera hacer $K_{i(0)} = 0$ para algún i tal que

$$\lambda_i \geq 10^{-c}\lambda_{max} \qquad i = 1, 2, 3, \dots, p$$

donde:

c = alguna constante arbitraria (puede ser c = 1,2,3, etc)

ii) Con el resultado de i) Hallar $\alpha_{RG}^{(1)}$ en (3.4.6).

iii) Reestimar $K_{i(j)} = \dfrac{\hat{\sigma}^2}{(\alpha_{RG}^{i(j)})^2}$ , para j = 1, 2, ...

iv) Utillizando iii) hallar $\alpha_{RG}^{i(j+1)}$ en (3.4.6)

v) Repetir el paso iii) y iv) hasta alcanzar la estabilización en la m-esima iteración.

vi) Determine $\beta_{RG} = T\alpha_{RG}^{(m)}$

Donde: T es la matriz ortogonal de autovectores de $\mathbf{X'X}$ y $m$ = m-esima iteración.

D. CRITERIO DE HEMMERLE Y BRAMTLE (1978)

Este se basa en la minimización de los estimadores de $E[L_1^2]$ y $E[L_2^2]$, seleccionando los valores $K_i$ , i = 1, 2, ... ,p de la matriz diagonal $\mathbb{K}$ en (3.4.6).

La función $E[L_1^2]$ es propuesta por HOERL AND KENNARD (1970) y está definida por

$$ECM(\alpha) = E[L_1^2] = E[\alpha - \alpha]'[\alpha - \alpha] \qquad (3.4.17)$$

Donde $\alpha$ es cualquier estimador del parámetro $\alpha$ en (3.4.5a). Por otro lado, la función $E[\ L_2^2\ ]$ es propuesta por THEOBALD C.M. (1974) y definida por

$$ECM(\ \tilde{\alpha}\ ) = E[\ L_2^2\ ] = E[\ \alpha - \alpha\ ]'\Lambda[\ \alpha - \alpha\ ] \qquad (3.4.17a)$$

(llamado error cuadrático medio de predicción)

Donde $\alpha$ es cualquier estimador de $\alpha$ en (3.4.5a) y $\Lambda$ la matriz diagonal de autovalores de la matriz de correlación $\mathbf{X'X}$.

Especificamente cuando el estimador $\alpha = \alpha_{RG}$ , HOCKING, SPEED y LYNN (1976) consideran, que al ser utilizado ambos criterios, el estimador RIDGE GENERALIZADO es superior a cualquier otro dentro de la clase de procedimientos de estimación Ridge, cuando los valores óptimos de $K_i$ son conocidos en términos de los parámetros poblacionales.

Fácilmente se ve que los valores óptimos que minimizan ambos criterios es

$$K_i = \sigma^2/\alpha_i^2 \qquad (3.4.17b)$$

Donde $\sigma^2$ y $\alpha$ son los parámetros poblacionales desconocidos en (3.4.5a).

Se presenta los pasos para estimar los valores de $K_i$ en función de los valores observados:

a) Simplificar las funciones $E[\ L_1^2\ ]$ y $E[\ L_2^2\ ]$ luego encontrar sus estimadores. Es decir de (3.4.6)

$$E(L_1^2) = E[\ \hat{\alpha}_{RG} - \alpha\ ]'[\ \hat{\alpha}_{RG} - \alpha\ ]$$

$$= E[\ \hat{\alpha}_{RG} - \hat{\alpha}\ ]'[\ \hat{\alpha}_{RG} - \hat{\alpha}\ ] + \sigma^2 traza[\ (\Lambda + K)\Lambda^{-1}(\Lambda + K)^{-1}\ ] \qquad (3.4.17c)$$ [3]

Analogamente

$$E[\ L_2^2\ ] = E[\ \hat{\alpha}_{RG} - \alpha\ ]'\Lambda[\ \hat{\alpha}_{RG} - \alpha\ ]$$

---

Se demuestra en G, apéndice I, como se obtiene (3.4.17c)

$$= E[\hat{\alpha}_{RG} - \hat{\alpha}]'\Lambda[\hat{\alpha}_{RG} - \hat{\alpha}] + \sigma^2 \text{traza}[(\Lambda+K)(\Lambda-K)^{-1}] \qquad (3.4.17d)$$ [4]

Por consiguiente los estimadores insesgados de $E[L_1^2]$ y $E[L_2^2]$ son

$$\hat{L}_1^2 = (\hat{\alpha}_{RG} - \hat{\alpha})'(\hat{\alpha}_{RG} - \hat{\alpha}) + \hat{\sigma}^2 \text{traza}[(\Lambda-K)\Lambda^{-1}(\Lambda+K)^{-1}]$$

$$\hat{L}_2^2 = (\hat{\alpha}_{RG} - \hat{\alpha})'\Lambda(\hat{\alpha}_{RG} - \hat{\alpha}) + \hat{\sigma}^2 \text{traza}[(\Lambda-K)(\Lambda+K)^{-1}]$$

b) De (3.4.4a) se conoce que

$$\hat{\beta}_{RR} = (I + K(X'X)^{-1})^{-1}\hat{\beta}$$

luego $\alpha_{RG}$ también se expresa según (3.4.4a) como

$$\hat{\alpha}_{RG} = (I + K\Lambda^{-1})^{-1}\hat{\alpha}$$

$$\alpha_{RG} = \begin{bmatrix} [\lambda_1/(\lambda_1 + k_1)]\hat{\alpha}_1 \\ [\lambda_2/(\lambda_2 + k_2)]\hat{\alpha}_2 \\ \vdots \\ [\lambda_p/(\lambda_p + k_p)]\hat{\alpha}_p \end{bmatrix} \qquad (3.4.17e)$$

Luego se hace

$$V_i = \lambda_i/(\lambda_i + k_i) \qquad (3.4.17f)$$

Entonces de (3.4.17e) se obtiene

$$= \alpha_i V_i \quad i = 1, 2, \ldots, p \qquad (3.4.17g)$$

c) Utilizando (3.4.17g) se obtiene una forma más simplificada $\hat{L}_1^2$ y $\hat{L}_2^2$, es decir

$$\hat{L}_1^2 = \sum_{i=1}^{p} M_i \qquad \text{y} \qquad \hat{L}_1^2 = \sum_{i=1}^{p} \lambda_i M_i$$

Donde

---

[4] Se demuestra en H, apéndice I, la formula (3.4.17d).

$$M_i = \hat{\alpha}_i^2 ( V_i - 1 )^2 + ( \hat{\sigma}^2/\lambda_i )( 2V_i - 1) \qquad (3.4.17h)$$

Del mismo modo de (3.4.16f) para j = Ø se obtiene

$$e_{i(O)} = \frac{\hat{\sigma}^2}{\lambda_i ( \hat{\alpha}_{RG\, i(O)} )^2}$$

$$= \frac{\hat{\sigma}^2}{\lambda_i ( \hat{\alpha}_i )^2} \qquad (3.4.17i)$$

Por lo tanto de (3.4.17h) y (3.4.17i) $M_i$ será

$$M_i = \hat{\alpha}_i^2 [ (V_i - 1 + e_{i(O)})^2 + \hat{\alpha}_i (e_{i(O)} - e_{i(O)}^2) ] \qquad (3.4.17j)$$

con $\quad Ø \le V_i \le 1$ , i=1,2,..,p

d) El problema es resuelto al derivar (3.4.17j) con respecto a $V_i$ e igualando a cero la expresión siguiente

$$\frac{d\, M_i}{dV_i} = 2\hat{\alpha}_i^2 ( V_i - 1 + e_{i(O)} )V_i = Ø$$

Entonces el valor que minimiza $M_i$ y por consecuencia $\hat{L}_1^2$ y $\hat{L}_2^2$ es

$$V_i = \begin{cases} 1 - e_{i(O)} , & e_{i(O)} < 1 \\ Ø , & e_{i(O)} \ge 1 \end{cases} \qquad (3.4.17k)$$

Por consiguiente de (3.4.17f), (3.4.17i) y (3.4.17k) se obtiene el valor óptimo de $K_i$

$$\begin{aligned} K_i &= \lambda_i \hat{\sigma}^2/(\lambda_i \hat{\alpha}_i^2 - \hat{\sigma}^2) , & e_{i(O)} = \hat{\sigma}^2/(\lambda_i \hat{\alpha}_i^2 ) < 1 \\ K_i &= \infty & e_{i(O)} = \hat{\sigma}^2/(\lambda_i \hat{\alpha}_i^2 ) \ge 1 \end{aligned} \qquad (3.4.17l)$$

e) A continuación se reemplaza (3.4.17l) en (3.4.17e) para calcular $\alpha_{RG}$ , y finalmente se calcula

$$\beta_{RG} = T\alpha_{RG}$$

## 3.5 ESTIMADOR DE CONTRACCION (CS)

Este estimador es propuesto inicialmente por JAMES y STEIN (1961) y su estudio está basado en, que el correspondiente ECM sea pequeño, en comparación a otros estimadores (desarrollados por otras técnicas) del coeficiente $\beta$ en el modelo (1.3).

### 3.5.1 ESTRUCTURA DEL ESTIMADOR DE CONTRACCION

Se denota por $\beta_{CS}$ al estimador de contracción a partir del modelo (1.3). El estimador de MCO de $\beta$ es (1.9), que también se expresa como

$$\ddot{\beta} = \mathbf{T}\,\Lambda^{-1}\mathbf{T}'\mathbf{X}'\mathbf{Y} \qquad (3.5.1)$$

Luego, si algunos autovalores $\lambda_j$, j = 1, 2,..,p, encontrados arriba, son bastante pequeños entonces $\lambda_j^{-1}$ será demasiado grande. Por tanto para disminuir el peso de $\lambda_j^{-1}$ es necesario multiplicar por una cantidad $d_j$ con $0 \leq d_j \leq 1$. Luego para solucionar esta situación a (3.5.1) se multiplica únicamente al lado derecho, por $\mathbf{D}_{p\times p}$ que es una matriz diagonal cuyos valores son $0 \leq \mathbf{d}_j \leq 1$, $\forall$ j = 1, 2, ..,p. Entonces (3.5.1) se convertirá en un estimador transformado que se nombra como

$$\hat{\beta}_{CS} = \mathbf{D}\mathbf{T}\,\Lambda^{-1}\mathbf{T}'\mathbf{X}'\mathbf{Y} \qquad (3.5.2)$$

en forma de sumatorias

$$\beta_{CS} = \sum_{j=1}^{p} \mathbf{d}_j\lambda_j^{-1}\mathbf{T}'_j\mathbf{X}'\mathbf{Y}\ \mathbf{T}_j \qquad (3.5.2a)$$

También de (3.5.2) se deduce que

$$\beta_{CS} = \mathbf{D}_{p\times p}\beta \qquad (3.5.2b)$$

Se debe anotar, que la matriz $\mathbf{D}_{p\times p}$ puede ser

$$\mathbf{d}_j = \mathbf{d}\,,\ \forall\ j = 1, 2, \ldots ,p$$

En este trabajo de tesis, nos ocuparemos de este caso, donde $\mathbf{d}_j = \mathbf{d}$ es una constante tal que $\emptyset \le \mathbf{d} \le 1$. Entonces (3.5.2b) se convierte en

$$\hat{\beta}_{cs} = d(X'X)^{-1}X'Y \qquad (3.5.2c)$$

De aqui en adelante cuando se hace referencia de $\beta_{cs}$, será el obtenido a través de (3.5.2c).

## 3.5.2 PROPIEDADES DEL ESTIMADOR $\beta_{cs}$

**PROPIEDAD 1**

$$E(\beta_{cs}) = D\beta \qquad (3.5.3)$$

**Prueba :**

De (3.5.2b) se deduce la esperanza.

**PROPIEDAD 2**

$$\text{Var - Cov}(\beta_{cs}) = \sigma^2 \sum_{j=1}^{p} d_j^2 \lambda_j^{-1} T_j T'_j \qquad (3.5.4)$$

**Prueba :**

De (3.5.2b) se tiene :

$$\beta_{cs} = D\beta$$

$$\beta_{cs} = D(X'X)^{-1}X'Y$$

entonces

$$\text{Var}(\beta_{cs}) = D(X'X)^{-1}X'\,\text{Var}(Y)X(X'X)^{-1}D'$$

$$\text{Var}(\beta_{cs}) = \sigma^2 D(X'X)^{-1}X'X(X'X)^{-1}D'$$

$$\text{Var}(\beta_{cs}) = \sigma^2 D(X'X)^{-1}D'$$

la última expresión en forma de sumatorias

$$\text{Var}(\hat{\beta}_{cs}) = \sigma^2 \sum_{j=1}^{p} d_j^2 \lambda_j^{-1} T_j T'_j \quad \blacksquare$$

**PROPIEDAD 3**

$$\text{ECM}(\hat{\beta}_{cs}) = \sigma^2 \sum_{j=1}^{p} d_j^2 \lambda_j^{-1} + \sum_{j=1}^{p} (d_j - 1)^2(\beta'\beta) \qquad (3.5.5)$$

**Prueba :**

Por definición el error cuadrático medio de un estimador es

$ECM(\beta_{cs}) = E[\beta_{cs} - \beta]'[\beta_{cs} - \beta]$

Por (3.5.2b)

$$= E(D\hat{\beta} - \beta)'(D\hat{\beta} - \beta)$$

$$= E(D\beta - D\beta + D\beta - \beta)'(D\beta - D\beta + D\beta - \beta)$$

$$= E(D(\beta - \beta) + D\beta - \beta)'(D(\beta - \beta) + D\beta - \beta)$$

$$= E[\left(D(\hat{\beta} - \beta)\right)' D(\hat{\beta} - \beta) + \left(D(\hat{\beta} - \beta)\right)'(D\beta - \beta)$$

$$+ (D\beta - \beta)' D(\beta - \beta) + (D\beta - \beta)'(D\beta - \beta)]$$

$$= E[(\beta - \beta)' D' D(\beta - \beta)] + (D\beta - \beta)'(D\beta - \beta)$$

$$= \sigma^2 \operatorname{traza}(X'X)^{-1} D'D + \beta'(D - I)'(D - I)\beta$$

$$= \sigma^2 \sum_{j=1}^{p} d_j^2 \lambda_j^{-1} + \sum_{j=1}^{p} (d_j - 1)^2 (\beta'\beta) \quad \blacksquare$$

### 3.5.3 CRITERIOS IMPORTANTES PARA ESTIMAR LA MATRIZ D

Se conocen pocos criterios para estimar la matriz **D** en (3.5.2b). Para el caso matricial no se ha desarrollado ninguna teoría y para el caso $\mathbf{d}_j = \mathbf{d}$, $\forall\ j = 1, 2, \ldots, p$, con $0 \le \mathbf{d} \le 1$, se presenta los siguientes criterios:

1) CRITERIO DE JAMES y STEIN (1961)

Para estimar la constante **d**, se considera un modelo de regresión lineal, donde las variables explicatorias son independientes ortogonales, es decir el modelo se define como

$$Y = \mathbb{M}\alpha + e \qquad (3.5.6)$$

Donde

i) Nuevamente $e \sim N(0, \sigma^2 I)$,

ii) $\mathbb{M}'\mathbb{M} = I$,

iii) $p \ge 3$ (p = número de variables explicatorias)

Luego el algoritmo para estimar **d** funciona como sigue

a) Estimar $\alpha$ en (3.5.6) por mínimos cuadrados ordinarios

$$\alpha = ( \mathbb{M}'\mathbb{M} )^{-1} \mathbb{M}'\mathbf{Y}$$

$$\alpha = \mathbb{M}'\mathbf{Y} \qquad (3.5.6a)$$

b) Calcular la suma de cuadrados de los residuales

$$SCRes = \upsilon = \mathbf{Y}'\mathbf{Y} - \alpha'\alpha$$

además $\alpha$ y $\upsilon$ son independientes y

$$\alpha \sim N(\alpha, \sigma^2 \mathbf{I}), \qquad \upsilon \sim \sigma^2 \chi^2_{n-p}$$

c) seleccionar un número "c" tal que

$$0 < c < 2(p - 2)/(n - p + 2)$$

Finalmente

d) seleccionar

$$\mathbf{d} = \max\left[ 0, (1 - \frac{c\upsilon}{\hat{\alpha}'\hat{\alpha}}) \right] \qquad (3.5.6b)$$

tal que $0 \leq \mathbf{d} \leq 1$

e) Luego según (3.5.2b)

$$\beta_{cs} = \mathbf{d}\alpha \qquad (3.5.6c)$$

donde $\mathbf{d}$ está dado en (3.5.6b)

Continuando con este criterio se tiene el siguiente teorema.

**TEOREMA 1**

Para $p \geq 3$ el estimador

$$\hat{\beta}_{cs} = \left[ 1 - \frac{c\upsilon}{\hat{\alpha}'\hat{\alpha}} \right]\hat{\alpha}$$

tiene la propiedad siguiente

$$ECM( \hat{\beta}_{cs} ) < ECM( \hat{\alpha} ), \quad \forall\, \alpha \qquad (3.5.7)$$

Además para cualquier $\alpha$, el ECM( $\beta_{cs}$ ) es pequeño cuando $c = \frac{(p - 2)}{(n - p + 2)}$.

**Prueba :**

Por definición de ECM

$$ECM( \hat{\alpha} ) = E( \hat{\alpha} - \alpha )'(\alpha - \alpha )$$

$$= E(DM'Y - DM'Y + DM'e)'(DM'Y - DM'Y + DM'e)$$

$$= E(e'_{1xn} DM_{nxp} DM'_{pxn} e_{nx1})$$

$$= \begin{bmatrix} E(e_1^2 Z_{11}^2) + \dots + E(e_n^2 Z_{n1}^2) \\ E(e_1^2 Z_{12}^2) + \dots + E(e_n^2 Z_{n2}^2) \\ \vdots \\ E(e_1^2 Z_{1p}^2) + \dots + E(e_n^2 Z_{np}^2) \end{bmatrix}_{p\times 1} = p\sigma^2$$

Es decir

$$ECM(\hat{\alpha}) = p\sigma^2 \qquad (3.5.7a)$$

Por otro lado

$$ECM(\hat{\beta}_{cs}) = E(\hat{\beta}_{cs} - \alpha)'(\hat{\beta}_{cs} - \alpha)$$

Utilizando **d**, en lugar de **D** en (3.5.2b) se obtiene

$$ECM(\hat{\beta}_{cs}) = E(\mathbf{d}\hat{\alpha} - \alpha)'(\mathbf{d}\hat{\alpha} - \alpha)$$

$$= \mathbf{d}^2 E(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)$$

$$ECM(\hat{\beta}_{cs}) = \mathbf{d}^2 p\sigma^2 \qquad (3.5.7b)$$

Luego cuando

$$\mathbf{d} = \max\left[ 0,\ 1 - \frac{c\upsilon}{\hat{\alpha}'\hat{\alpha}} \right]$$

(3.5.7b) se convierte en

$$ECM(\hat{\beta}_{cs}) = \mathbf{d}\, p\sigma^2 \qquad (3.5.7c)$$

de (3.5.7a), (3.5.7c) y (3.5.6b)

$$ECM(\beta_{cs}) < ECM(\alpha)$$

Además

$$\text{Cuando } \alpha'\alpha \longrightarrow \infty \qquad ECM(\beta_{cs}) = ECM(\alpha)$$

Por consiguiente el mejoramiento respecto al estimador usual, puede ser algo substancial. JAMES y STEIN (1961) concluyen que para p = 1, p = 2 (número de variables explicatorias ortogonales), el estimador $\beta_{cs}$ no desarrolla tan bien en la práctica, es decir

$$\text{ECM}(\beta_{cs}) \text{ no es menor a } \text{ECM}(\beta), \text{ para } p = 1, 2.$$

2) CRITERIO DE STANLEY SCLOVE (1968) (Criterio de Generalización)

Este criterio usa el mismo "**d**", definido en (3.5.6b), pero se inicia con el modelo (1.3). Entonces, para usar (3.5.6b) se debe primero transformar (1.3) de la manera siguiente

$$\mathbf{L\ X'X\ L'} = \mathbf{I} \qquad (3.5.8)$$

Luego $\mathbb{M} = \mathbf{X\ L'}$ y $\beta = \mathbf{L'}\alpha$ (3.5.8a)

<u>**Observación**</u> **:** **L** no necesariamente es Ortogonal.

Por consiguiente el modelo (1.3) se convierte en

$$\mathbf{Y} = \mathbb{M}\alpha + \mathbf{e} \qquad (3.5.8b)$$

De (3.5.8b) se concluye que igual a (3.5.6)

$$\mathbb{M}'\mathbb{M} = \mathbf{L\ X'X\ L'} = \mathbf{I}$$

Entonces, "**d**" es igual al valor encontrado en (3.5.6b).

Finalmente el estimador de los coeficientes del modelo (1.3) en función de (3.5.2b) es $\beta_{cs} = \mathbf{L'd}\alpha$

es decir

$$\beta_{cs} = \mathbf{L'}\max\left[\,0,\ 1 - \frac{cv}{\hat{\alpha}'\hat{\alpha}}\,\right]\hat{\alpha}$$

3) CRITERIO DE DEMPSTER, SCHATZOFF, WERMUTH (1977)

Sea $\beta$ el estimador por MCO del modelo (1.3). luego, generalizando este, se tendrá

$$\hat{\beta} = (\mathbf{X'X} + \mathbf{KQ})^{-1}\mathbf{X'Y} \qquad (3.5.9)$$

Donde **Q** es una matriz simétrica definida positiva, que depende sólo del diseño de la matriz **X**, y K es un escalar no negativo, que depende generalmente de **Y**. luego, si se selecciona **Q** = **(X'X)**, (3.5.9) se convierte en un estimador, llamado estimador de STEIN ($\beta_{cs}$)

$$\beta_{cs} = ( X'X + KX'X )^{-1}X'Y$$
$$= ( 1 + K )^{-1}(X'X)^{-1}X'Y$$

y según (3.5.2b)

$$\beta_{cs} = d\beta \qquad (3.5.9a)$$

entonces
$$d = \frac{1}{1 + K}$$

De esta forma se encuentra el valor de "d" que es diferente al criterio de JAMES y STEIN (1961), donde el valor de K es

$$K = \frac{\hat{\sigma}^2}{\beta'\beta}$$

y

$$\hat{\sigma}^2 = \frac{( Y - X\beta )'( Y - X\beta )}{( n - p )}$$

Otra forma posible del criterio, es transformando el modelo (1.3) en su forma canónica como fue establecido en (3.4.5a), cuyo estimador por MCO es

$$\hat{\alpha} = \Lambda^{-1}(X^{+\prime})Y$$

Luego
$$\alpha_{cs} = d\alpha \qquad (3.5.9b)$$

entonces

$$d = \frac{1}{1 + K}$$

donde
$$K = \frac{\hat{\sigma}^2}{\alpha'\alpha}$$

y

$$\hat{\sigma}^2 = \frac{( Y - X^+\hat{\alpha} )'( Y - X^+\hat{\alpha} )}{n - p}$$

Por otra parte se deriva del criterio, una regla, correspondiente a encontrar la región elipsoide confidencial del parámetrado $\beta$ centrado al estimador de mínimos cudrados $\beta$, esto es,

$$\frac{(\beta - \beta)' X' X(\beta - \beta)}{ps^2} \leq F_{p,n-p,1-\alpha} \qquad (3.5.9c)$$

entonces, se debe limitar la desviación de $\beta$ a $\beta$, la cual implica que $\beta_{CS}$ esté dentro del elipsoide de (3.5.9c), es decir

$$\frac{(\beta_{CS} - \beta)' X' X(\beta_{CS} - \beta)}{ps^2} \leq F_{p,n-p,1-\alpha} \qquad (3.5.9d)$$

Ahora si para

$$d = \frac{1}{1 + K} \qquad y \qquad K = \frac{\hat{\sigma}^2}{\alpha' \alpha}$$

el estimador $\beta_{CS}$ no cumple con (3.5.9d) se ajusta el K, el cual es calculado de acuerdo a los criterios planteados en el estimador de Regresion Ridge $\beta_{RR}$.

## 3.6 ESTIMADOR DE LA SELECCION DE VARIABLES ( SV )

En esta sección, se presenta el último estimador sesgado, llamado SELECCION DE VARIABLES (SV), alternativo al estimador de mínimos cuadrados ordinarios, su aplicación es cuando el modelo (1.3) es inestable, es decir cuando hay presencia de multicolinealidad entre las variables explicatorias. Por lo tanto esta técnica se fundamenta en que debe seleccionarse un subconjunto de variables explicatorias independientes para que el efecto de multicolinealidad sea menos seria. Se incluye abajo, su estructura, propiedades, formas y criterios para encontrar un subconjunto óptimo de variables del modelo (1.3), de manera que, en este nuevo conjunto ya no exista multicolinealidad. Finalmente se termina con comentarios sobre la ventaja de usar el estimador de SV.

### 3.6.1. ESTRUCTURA DEL ESTIMADOR DE SELECCION DE VARIABLES (SV)

El procedimiento de selección de variables (SV),es un estimador sesgado de los coeficientes de las variables independientes seleccionadas del modelo (1.3), pero aún así puede proveer predicciones más estables. Supóngase, que para eliminar la multicolinealidad del modelo (1.3), se ha seleccionado "r" variables explicatorias, las cuales deben ser eliminadas del modelo general, luego (1.3) se expresará como

$$\mathbf{Y} = \mathbf{X}_q\beta_q + \mathbf{X}_r\beta_r + \mathbf{e} \qquad (3.6.1)$$

Donde $\mathbf{X}$ matriz estándar de (1.3), es particionada en $\mathbf{X}_q$ de dimensión n×q y $\mathbf{X}_r$ de dimensión n×r. De forma similar el vector $\beta$ es particionado, y además p = q + r, siendo "q" las variables retenidas en el modelo. Sea

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{(q)} \\ \hat{\beta}_{(r)} \end{bmatrix} \qquad (3.6.1a)$$

el estimador particionado de mínimos cuadrados de $\beta$, y $\beta_{SV}$ la estimación por mínimos cuadrados ordinarios del vector de coeficientes $\beta_q$ si las r-variables del modelo (3.6.1) son eliminadas, esto es

$$\hat{\beta}_{SV} = (\mathbf{X}'_q\mathbf{X}_q)^{-1}\mathbf{X}'_q\mathbf{Y} \qquad (3.6.2)$$

Por consiguiente (3.6.2) es conocido como estimador de SELECCION DE VARIABLES, pues sólo es calculado en función de las "q" variables regresoras retenidas en el modelo (3.6.1). De esta forma se ha encontrado la estructura del estimador $\beta_{SV}$. También (3.6.2) se expresa como

$$\hat{\beta}_{SV} = \sum_{j=1}^{q} S_j^{-1} H'_j \mathbf{X}'_q \mathbf{Y}\ H_j \qquad (3.6.2a)$$

Donde $H_j$ de dimensión q×1 son autovectores ortogonales correspondientes a los autovalores $S_j$ , j = 1, 2,...,q , de la matriz $(\mathbf{X}'_q\mathbf{X}_q)$ de dimensión q×q. Se debe anotar que los autovalores en (3.6.2) o (3.6.2a) no están ordenados por magnitud como en otros estimadores presentados.

### 3.6.2. PROPIEDADES DEL ESTIMADOR $\beta_{sv}$

**PROPIEDAD 1**

$\hat{\sigma}^2_{sv}$ representa el cuadrado medio residual para la selección de variables, es decir

$$\hat{\sigma}^2_{sv} = Y'[\ I - X_q(X'_qX_q)^{-1}X'_q\ ]Y/(n-q) \qquad (3.6.3)$$

**Prueba :**

Si en (3.6.1) las "r" variables son eliminados, entonces por definición de $\sigma^2$ se tiene

$$\sigma^2_{sv} = \frac{e'e}{n-q}$$

$$= \frac{(Y - X_q\beta_q)'(Y - X_q\beta_q)}{(n-q)}$$

luego, reemplazando por su estimación, así como también $\beta_q$ (de acuerdo a 3.6.2), se obtiene

$$= \frac{(Y - X_q(X'_qX_q)^{-1}X'_qY)'(Y - X_q(X'_qX_q)^{-1}X'_qY)}{(n-q)}$$

$$= \frac{Y'Y - Y'X_q(X'_qX_q)^{-1}X'_qY}{(n-q)}$$

$$= Y'[\ I - X_q(X'_qX_q)^{-1}X'_q\ ]Y/(n-q) \quad \blacksquare$$

**PROPIEDAD 2**

$$E(\beta_{sv}) = \beta_q + (X'_qX_q)^{-1}X'_qX_r\beta_r \qquad (3.6.4)$$

**Prueba :**

De (3.6.2) $\qquad \beta_{sv} = (X'_qX_q)^{-1}X'_qY$

reemplazando Y por sus valores de acuerdo a (3.6.1)

$$\beta_{sv} = (X'_q X_q)^{-1} X'_q [\; X_q \beta_q + X_r \beta_r + e \;]$$

y aplicando esperanza

$$E(\; \beta_{sv} \;) = \beta_q + (X'_q X_q)^{-1} X'_q X_r \beta_r \,. \quad \blacksquare$$

**PROPIEDAD 3**

$$\text{VAR} - \text{COV}(\; \beta_{sv} \;) = (\; X'_q X_q \;)^{-1} \sigma^2 \qquad (3.6.5)$$

**Prueba :**

De (3.6.2) $\qquad \beta_{sv} = (X'_q X_q)^{-1} X'_q Y$

y de (3.6.1) $\qquad \beta_{sv} = (X'_q X_q)^{-1} X'_q (\; X_q \beta_q + X_r \beta_r + e \;)$

$$\beta_{sv} = \beta_q + (X'_q X_q)^{-1} X'_q X_r \beta_r + (X'_q X_q)^{-1} X'_q e$$

Luego se obtiene

$$\beta_{sv} - [\; \beta_q + (X'_q X_q)^{-1} X'_q X_r \beta_r \;] = (X'_q X_q)^{-1} X'_q e \qquad (3.6.5a)$$

Según (3.6.4), (3.6.5a) se convierte en

$$\beta_{sv} - E(\beta_{sv}) = (X'_q\; X_q\;)^{-1} X'_q e \qquad (3.6.5b)$$

Por otro lado, por definición de VAR - COV

$$\text{VAR} - \text{COV}(\; \beta_{sv} \;) = E\Big[(\; \hat{\beta}_{sv} - E(\hat{\beta}_{sv}\;)\Big]\Big[(\; \hat{\beta}_{sv} - E(\hat{\beta}_{sv}\;)\Big]'$$

de (3.6.5b)

$$= E\Big[\; (X'_q\; X_q\;)^{-1} X'_q\; e \;\Big]\Big[\; (X'_q\; X_q\;)^{-1} X'_q\; e \;\Big]'$$

$$= E\Big[\; (X'_q\; X_q\;)^{-1} X'_q\; ee'\; X_q (X'_q\; X_q\;)^{-1}\Big]$$

$$= (X'_q\; X_q\;)^{-1} X'_q\, E(ee'\,) X_q (X'_q\; X_q\;)^{-1}$$

$$= (X'_q\; X_q\;)^{-1} \sigma^2 \,. \quad \blacksquare$$

también es posible expresar como

$$\text{VAR} - \text{COV}(\; \beta_{sv}\;) = \sigma^2 \sum_{j=1}^{q} S_j^{-1}\; H_j\; H'_j$$

donde $H_j$ es definida en (3.6.2a).

**PROPIEDAD 4**

$$ECM(\beta_{sv}) = \sigma^2 traza(X'_q X_q)^{-1} + (X'_q X_q)^{-1} X'_q X_r \beta_r \beta'_r [(X'_q X_q)^{-1} X'_q X_r]' \qquad (3.6.6)$$

**Prueba :**

Por definición

$$ECM(\beta_{sv}) = E(\beta_{sv} - \beta_q)(\beta_{sv} - \beta_q)'$$
$$= E(\beta_{sv}\beta'_{sv} - \beta_{sv}\beta'_q - \beta_q\beta'_{sv} + \beta_q\beta'_q)$$
$$= E(\beta_{sv}\beta'_{sv}) - E(\beta_{sv})\beta'_q - \beta_q E(\beta'_{sv}) + \beta_q\beta'_q$$
$$= E(\hat{\beta}_{sv}\hat{\beta}'_{sv}) - E^2(\hat{\beta}_{sv}) + E^2(\hat{\beta}_{sv}) - 2E(\hat{\beta}_{sv})\beta'_q + \beta_q\beta'_q$$
$$= VAR(\beta_{sv}) + [E(\beta_{sv}) - \beta_q][E(\beta_{sv}) - \beta_q]'$$
$$= \sigma^2 traza(X'_q X_q)^{-1} + [\beta_q + (X'_q X_q)^{-1} X'_q X_r \beta_r - \beta_q][\beta_q + (X'_q X_q)^{-1} X'_q X_r \beta_r - \beta_q]'$$
$$= \sigma^2 traza(X'_q X_q)^{-1} + (X'_q X_q)^{-1} X'_q X_r \beta_r \beta'_r [(X'_q X_q)^{-1} X'_q X_r]' \quad \blacksquare$$

también se expresará como

$$ECM(\hat{\beta}_{sv}) = \sigma^2 \sum_{j=1}^{q} S_j^{-1} + (X'_q X_q)^{-1} X'_q X_r \beta_r \beta'_r [(X'_q X_q)^{-1} X'_q X_r]' \qquad (3.6.6a)$$

Observación:

- $\beta_{sv}$ está distribuida normalmente con esperanza y varianza especificada anteriormente.
- $(n - q)\hat{\sigma}_{sv} / \sigma^2$ está distribuida como una CHI-CUADRADA no central, con esperanza dada por

$$E(\sigma_{sv}) = \frac{\sigma^2 + (\beta'_r X'_r [I - X_q (X'_q X_q)^{-1} X'_q] X_r \beta_r)}{(n - q)} \qquad (3.6.6b)$$

**PROPIEDAD 5**

$$\beta_{sv} \text{ es sesgado excepto cuando: } \begin{array}{l} a) \ \beta = \emptyset \\ b) \ X'_q X_r = \emptyset \end{array} \qquad (3.6.7)$$

**Prueba :**

Los resultados se ven rapidamente de (3.6.4)

**PROPIEDAD 6**

La matriz $VAR(\beta_{(q)}) - VAR(\beta_{SV})$ es definida semipositiva (3.6.8)

**Prueba :**

Se sabe que
$$VAR(\beta) = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

luego de (3.6.1a) se tiene
$$VAR(\hat{\beta}_{(q)}) = \sigma^2 \sum_{j=1}^{q} \lambda_j^{-1} \qquad (3.6.8a)$$

Aqui $\lambda_j^{-1}$ son los autovalores de la matriz $(X'X)$ donde $X = [\ X_q, X_r\ ]$. Se observa que el cálculo de $\lambda_j$, $j = 1, 2, \ldots, p$ está determinada por todas las combinaciones de las correlaciones entre todas las variables explicatorias del modelo (1.3), por tanto si algunas de las variables son proximamente correlacionadas, habrán algunos $\lambda_j$ demasiado pequeños y por consiguiente la inversa $\lambda_j^{-1}$ será demasiado grande.

Ahora si son eliminados las variables que causan multicolinealidad en el modelo (1.3), se tendrá del primer factor del lado derecho de (3.6.6a)
$$V(\hat{\beta}_{SV}) = \sigma^2 \sum_{j=1}^{q} S_j^{-1} \qquad (3.6.8b)$$

De (3.6.8b), cada $S_j$ no será demasiado pequeño, por tanto $S_j^{-1}$ no será demasiado grande; luego de (3.6.8a) y (3.6.8b) se concluye
$$V(\hat{\beta}_{(q)}) - V(\hat{\beta}_{SV}) = \sigma^2 \sum_{j=1}^{q} (\lambda_j^{-1} - S_j^{-1}) \geq 0$$

debido a que
$$(\lambda_j^{-1} - S_j^{-1}) \geq 0$$

Las estimaciones de las componentes de $\beta_q$ por $\beta_{(q)}$ son generalmente mas variables que los encontrados por $\beta_{SV}$.

**PROPIEDAD 7**

$\hat{\sigma}^2_{SV}$ es más sesgado que $\hat{\sigma}^2$ (hallado por MCO).

**PROPIEDAD 8**

Sea $x' = (x'_q\ ,\ x'_r)$, una entrada particular en el modelo (1.3) ($x'$ es un vector fila 1xp). Se usa un subconjunto del modelo con $x'_r$ eliminados, el valor predicho de la variable respuesta es

$$Y_q = x'_q\beta_{sv} \quad (Y_q \text{ es un valor real})$$

entonces

$Y_q$ es sesgado, a menos que $X'_q X_r \beta_r = \emptyset$ (3.6.9)

**Prueba :**

como $$Y_q = x'_q\beta_{sv} \qquad (3.6.9a)$$

entonces $$E(Y_q) = x'_q E(\beta_{sv})$$

De (3.6.4) $$E(Y_q) = x'_q[\beta_q + (X'_qX_q)^{-1}X'_qX_r\beta_r]$$

$$E(Y_q) = x'_q\beta_q + x'_q(X'_qX_q)^{-1}X'_qX_r\beta_r] \qquad (3.6.9b)$$

Por consiguiente de (3.6.9b) y (3.6.9a) se observa que $Y_q$ es sesgado. Si en (3.6.9b) $X'_qX_r\beta_r = \emptyset$, entonces $Y_q$ será insesgado.

**PROPIEDAD 9**

Sea VARPre( $Y$ ) la varianza de predicción de las observaciones predichas $Y$ usando la estimación de mínimos cuadrados y VARPre( $Y_q$) de $Y_q$ cuando se usa la selección de variables, entonces

$$\text{VARPre}(Y) \geq \text{VARPre}(Y_q) \qquad (3.6.10)$$

$Y$, es el valor predicho de la variable respuesta, para una entrada particular $x' = (x'_q, x'_r)$ en el modelo (1.3), $Y_q$ es el valor predicho de la variable respuesta, para una entrada particular como arriba pero después de eliminar las $x'$

**Prueba :**

Del modelo (1.3), cuando se tiene una entrada particular de datos de las variables explicatorias, por ejemplo $x' = (x'_q, x'_r)$, entonces

$$\mathrm{VARPre}(\hat{Y}) = \sigma^2[\ 1 + x'(X'X)^{-1}x] \qquad (3.6.10a)$$

Si se eliminan las "r" variables en (1.3) se tendrá

$$\mathrm{VARPre}(\ Y_q\ ) = \sigma^2[\ 1 + x'_q(X'X)^{-1}x_q] \qquad (3.6.10b)$$

Luego, es evidente de (3.6.10a) y (3.6.10b) que

$$\mathrm{VARPre}(\ Y) \geq \mathrm{VARPre}(\ Y_q\ )$$

De la propiedad (6) y (9), se concluiye que $\beta_{sv}$ puede ser estimada ó la respuesta futura puede ser predecida con varianza pequeña usando un subconjunto de variables explicatorias del modelo (1.3), el problema, es el sesgo.

### 3.6.3. SELECCION OPTIMA DE q VARIABLES, ( $q \leq p$ ) SEGUN PROCEDIMIENTOS Y CRITERIOS

Una variedad de métodos se han propuesto, para la elección de las q variables ($q \leq p$) a ser incluidas en el modelo (1.3), cuando se tiene presencia de multicolinealidad. Estos métodos se distinguen según el procedimiento para encontrar el "mejor subconjunto q de variables independientes, tal que, éste subconjunto explique mejor a la variable respuesta " **Y** ".

En esta subsección, primero se presenta los criterios estadísticos, para seleccionar el "mejor subconjunto" q, de variables independientes al modelo (1.3), el cual tiene inicialmente "p" variables, y luego, se desarrolla los procedimientos.

DEFINICION y NOTACION .- Sean las siguientes funciones

$$SCRes_p = (Y - X\beta)'(Y - X\beta)$$

$$SCRes_q = (Y - X_q\beta_q)'(Y - X_q\beta_q)$$

"$SCRes_p$", es la suma de cuadrados de los residuales para los "p" términos del modelo (1.3) y "$SCRes_q$" es la suma de cuadrados de los residuales para el subconjunto de "q" términos del modelo (1.3), después de ser eliminadas las r = p - q variables que influencian en multicolinealidad.

1. CRITERIO DEL CUADRADO MEDIO RESIDUAL MINIMO

El cuadrado medio residual para las q-variables, luego que las r-variables comprometidas con multicolinealidad son eliminadas del modelo (1.3), es

$$CMRes_q = \frac{SCRes_q}{n - q} \qquad (3.6.11)$$

"n", es el número de observaciones muestrales. $CMRes_q$ es decreciente cuando "q" es creciente, es decir, inicialmente decrece, luego se estabiliza y posteriormente puede crecer, si la reducción de la SCRes (al adicionar un regresor al modelo de q-términos) no es suficiente para compensar la pérdida de un grado de libertad en el denominador de (3.6.11), más claramente

$$CMRes_{(q+1)} > CMRes_q$$

por consiguiente el criterio se establece como sigue:

- seleccionar el **mínimo** $CMRes_q$, luego,
- seleccionar el valor q ,tal que, $CMRes_q$ es aproximadamente igual a $CMRes_p$ , ó
- seleccionar el valor "q" próximo al punto donde el valor $CMRes_q$ es mínimo.

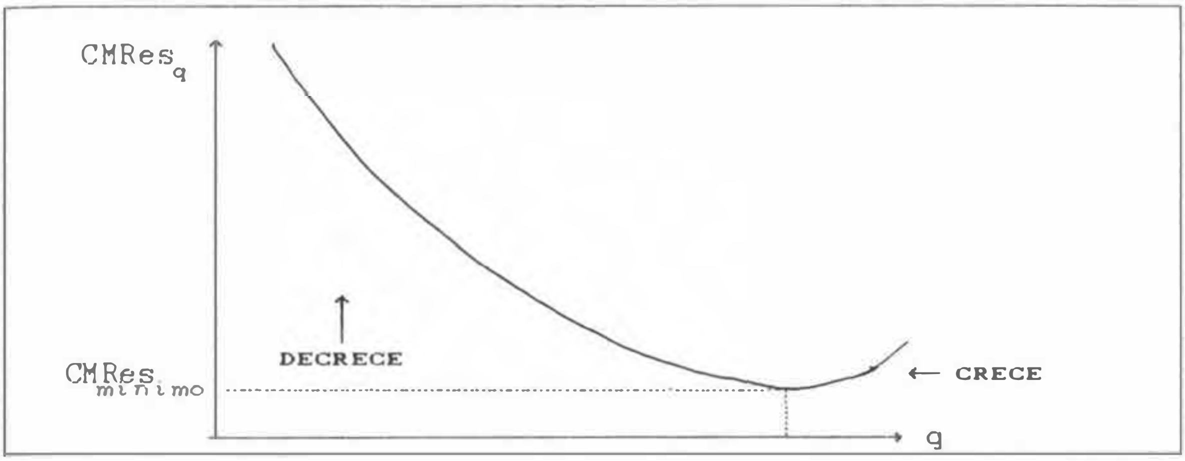


FIGURA No 2, GRAFICA DEL COMPORTAMIENTO DE CMRes vs q

2. CRITERIO DEL COEFICIENTE DE DETERMINACION MULTIPLE

El coeficiente de Determinación múltiple se define como,

$$R_q^2 = 1 - \frac{SCRes_q}{SCT_p} \qquad (3.6.12)$$

"q", son las variables incluidas en el modelo después de eliminar las r-variables en (1.3), y $SCT_p$, es la suma de cuadrados totales con las p-variables originales. $R_q^2$ crece cuando q crece y toma un máximo cuando q=p y existen $\binom{p}{q}$ valores posibles de $R_q^2$ para cada subconjunto de modelos posibles de tamaño "q".

El funcionamiento del criterio es:

Se adicionan regresores al modelo, hasta el punto que la adición de una variable más ya no es útil, esto es, cuando el crecimiento $R_q^2$ es pequeño. Entonces se selecciona este valor "q", como el número de regresoras a incluir en el modelo (1.3). Para esto se necesita un juicio de parte del analista y $R_q^2$ alto, implica que los modelos son preferibles.

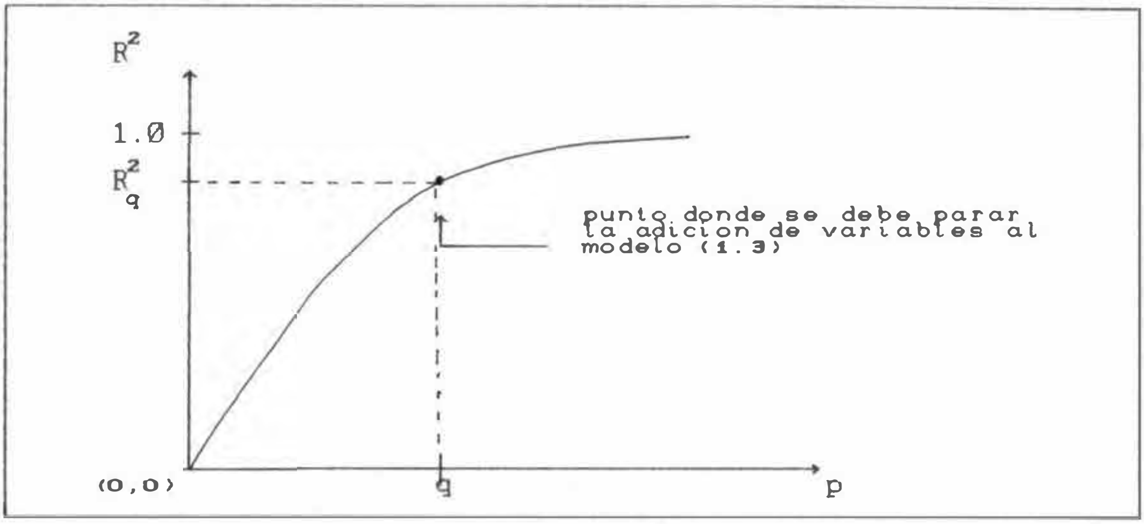


**FIGURA No3,GRAFICA DEL COMPORTAMIENTO DE $R^2_q$ PARA DIVERSOS VALORES DE q**

$(q \leq p)$

3. CRITERIO DEL $\tilde{R}^2$ AJUSTADO MAXIMO

Se define el coeficiente de determinación múltiple ajustado como

$$\tilde{R}^2_q = 1 - \frac{(n-1)(1-R^2_q)}{(n-q)} \qquad (3.6.13)$$

Transforando (3.6.13) para depender de (3.6.11), se realiza los pasos siguientes

de (3.6.12)
$$\tilde{R}^2_q = 1 - \frac{(n-1)\dfrac{(SCRes_q)}{(SCT_p)}}{(n-q)}$$

$$\tilde{R}^2 = 1 - \frac{(n-1)(SCRes\ )}{(SCT_p)(n-q)}$$

Por (3.6.11)
$$\tilde{R}^2_q = 1 - \frac{(n-1)}{(SCT\ )}\,.(CMRes_q) \qquad (3.6.14)$$

Este criterio es similar al anterior, pues se selecciona el valor "q" para el cual $\tilde{R}^2_q$ es máximo.

4. CRITERIO DE LA VARIANZA DE PREDICCION PROMEDIO MINIMO

La varianza de predicción promedio se define :

$$J_q = (n + q)\frac{(CMRes_q)}{n} \qquad (3.6.15)$$

q, es el número de variables ajustado ó corregido a partir del modelo (1.3). $J_q$, proviene del cálculo de la varianza de predicción total, sobre los datos comunes para un subconjunto dado. Por consiguiente, es posible estimar $\sigma^2$ por $CMRes_q$. Sobre la base teórica, la objeción de este estadístico, es que, ignora el sesgo en la predicción. Este criterio pocas veces es usado.

5. EL CRITERIO DEL ESTADISTICO $C_p$ DE MALLOW'S

El estadístico $C_p$ de Mallow's se define como :

$$C_p = \frac{SCRes_p}{\hat{\sigma}^2} + 2p - n \qquad (3.6.16)$$

A (3.6.16) también, se llama "Error Cuadrático total". Luego para el caso de un subconjunto de tamaño "q" del modelo (1.3), (3.6.16) será

$$C_q = \frac{SCRes_q}{\hat{\sigma}^2} + 2q - n \qquad (3.6.17)$$ [5]

El criterio $C_q$, es un procedimiento que indica, elsubconjunto más adecuado del modelo (1.3) a ser elegido, pues la interpretación de $C_q$ versus es bastante sencilla.

Puesto que $C_q$ es función de "q", es recomendado como un medio de proveer información acerca de la estructura de los datos, por consiguiente, el criterio se desarrolla como sigue:

- Se sugiere seleccionar el subconjunto con $C_q \leq q$
- Debe considerarse el $C_q$ próximo a q, tal que, el sesgo sea pequeño.

---
5 Se demuestra en J, apendice I, la formula (3.6.17).

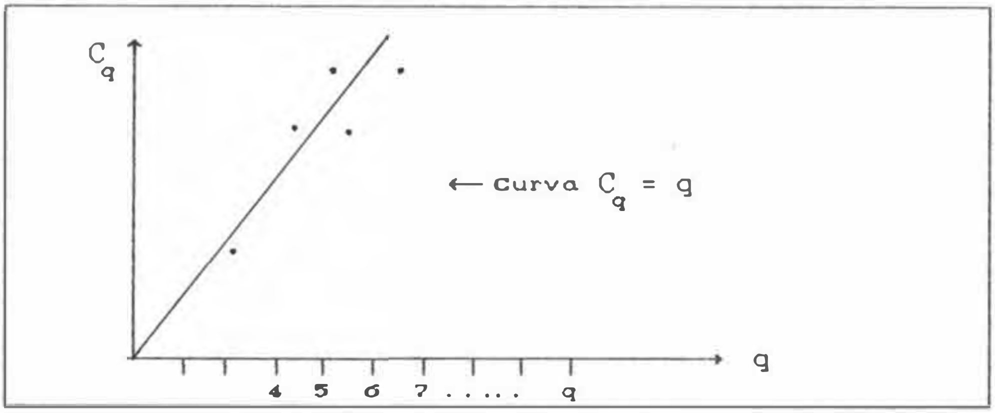


**FIGURA No 4**, Gráfica del comportamiento del estadístico $C_p$ de Mallows vs q

6. CRITERIO DEL TEST USANDO $C_q$ DE MALLOW'S

Este criterio, es usado para eligir un subconjunto de variables independientes de tamaño "q" de (1.3). Esto se hace a través del test de adecuación del mejor modelo ajustado o corregido.

El criterio considera hacer el test de Hipótesis

$$H : \beta_j = \emptyset \ , \ \begin{matrix} j \in r \\ j \neq \emptyset \end{matrix} \qquad (3.6.18)$$

Donde r es el conjunto de variables a eliminar del modelo (1.3).

Por tanto, el estadístico F para el test H es :

$$F_q = \frac{SCRes_q - \dfrac{SCRes}{(p - q)}}{\dfrac{SCRes_p}{(n - p)}} \qquad (3.6.18a)$$

$$F_q = \frac{SCRes_q - SCRes}{(p - q)\hat{\sigma}^2}$$

$$F_q = \frac{1}{(p - q)} \left[ \frac{SCRes_q - SCRes_p}{\hat{}^{2}} \right]$$

$$F_q = \frac{1}{(p - q)} \left[ \frac{SCRes_q}{\hat{\sigma}^2} \quad \frac{SCRes_p}{\dfrac{SCRes_p}{(n - p)}} \right]$$

$$F_q = \frac{1}{(p-q)} \left[ \frac{SCRes_q}{\hat{\sigma}^2} - (n-p) \right]$$

$$F_q = \frac{1}{(p-q)} \left[ C_q + p - 2q \right]$$

$$F_q = 1 + \frac{(C_q - q)}{p-q} \qquad (3.6.18b)$$

ó equivalentemente : $\quad C_q = (p - q)(F_q - 1) + q \qquad (3.6.18c)$

Luego, la indicación para aplicar este criterio se tiene a continuación:

seleccionar un "q" adecuado en (1.3) , tal que, $F_q \leq 1$

- Se debe gráficar "$F_q$ versus q", para obtener una impresión general del diagrama, y evaluar el efecto de remover ó adicionar variables.

<u>**Observación**</u> Es sencillo comparar la posición, de dos ó más valores de (3.6.18b) para diferentes subconjuntos de regresión, a través de la linea horizontal " y = 1 ", que a la línea inclinada "y = x", la cual sirve de guía para gráficar los valores de $C_q$ para combinaciones de "q".

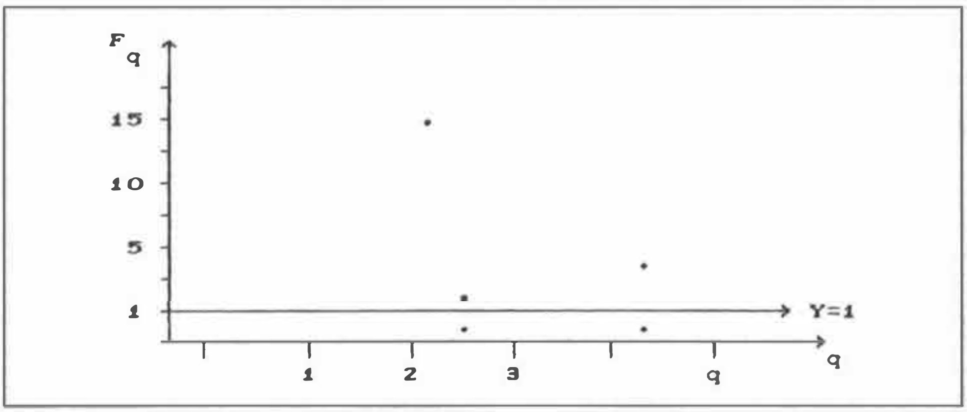


FIGURA No 5, Grafica de $F_q$ vs q

7. CRITERIO DEL ERROR CUADRATICO MEDIO DE PREDICCION INCONDICIONAL

Sea el modelo (3.6.1), que incluye las r-variables a ser eliminadas. luego, si se eliminan estás variables, el modelo se convierte en

$$Y = X_q \beta_q + e \qquad (3.6.19)$$

por consiguiente, la respuesta estimada $\mathbf{Y}_q$ es

$$\mathbf{Y}_q = \mathbf{X}_q \beta_q \qquad (3.6.19a)$$

Entonces, se define el "Error Cuadrático Medio de Predicción" de $Y_q$ para un dato particular de entrada $(y_o, x'_o)$, como:

$$\text{ECMPre}(\hat{Y}_q) = E(Y_o - \hat{Y}_q)^2 \qquad (3.6.19b)$$

$x'_o = (x'_q, x'_r)$, es nuevamente un vector de entrada particular de orden 1xp. $Y_o$, es la variable respuesta observada y es real. $Y_q$ es la predicción o estimación de $Y_o$, la cual es calculada sólo con " $x'$ "

Desarrollando (3.6.19b) se tiene

$$\begin{aligned}
\text{ECMPre}(\hat{Y}_q) &= E(Y_o^2 + \hat{Y}_q^2 - 2Y_o\hat{Y}_q) \\
&= (Y_o^2) + E(\hat{Y}_q^2) - 2Y_o E(\hat{Y}_q) \\
&= E(\hat{Y}_q^2) - E^2(\hat{Y}_q) + E^2(\hat{Y}_q) + Y_o^2 - 2Y_o E(\hat{Y}_q) \\
&= \text{VAR}(\hat{Y}_q) + [\, E(\hat{Y}_q) - Y_o \,]^2
\end{aligned}$$

en este caso particular, a $\text{VAR}(Y_q)$ también se llama $\text{VARPre}(Y_q)$. Entonces reemplazando por sus valores según (3.6.10b) y (3.6.9b) se obtiene que

$$\text{ECMPre}(\hat{Y}_q) = \sigma^2(1 + x'_q(\mathbf{X}'_q\mathbf{X}_q)^{-1}x_q) + (x'_q(\mathbf{X}'_q\mathbf{X}_q)^{-1}\mathbf{X}'_q\mathbf{X}_r\beta_r - x_r\beta_r) \qquad (3.6.19c)$$

aplicando el valor esperado a (3.6.19c) y haciendo algunas transformaciones se tiene

$$E(\text{ECMPre}(\hat{Y}_q)) = \frac{(n+1)(n \quad 2)}{n(n-q-2)}\,\sigma^2_{y.x_q} \qquad (3.6.19d)$$

$\sigma^2_{y.x_q}$, es la varianza condicional de $\mathbf{Y}$ dado las q variables en (3.6.19).

Luego, una estimación muestral de (3.6.19d) es

$$\widehat{E}(ECMPre(Y_q)) = \frac{(n^2 - n - 2)}{n(n - q - 2)} \hat{\sigma}^2_{y.x_q} \qquad (3.6.19e)$$

donde

$$\hat{\sigma}^2_{y.x} = \frac{(n - 1)S_y^2(1 - R_q^2)}{(n - q - 1)}$$

$$S_y^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n - 1}$$

$R^2$, es el coeficiente de correlación múltiple muestral entre Y y las q variables regresoras. Por consiguiente, (3.6.19e) se expresa como

$$\widehat{E}(ECMPre(Y_q)) = \frac{(n^2 - n - 2)(n - 1)S_y^2(1 - R_q^2)}{n(n - q - 1)(n - q - 2)} \qquad (3.6.19f)$$

Ahora, denotando

$$\widehat{E}(ECMPre(Y_q)) = ECMI$$

ECMI, es el Error cuadrático medio de predicción incondicional muestral. Entonces, el criterio es definido como "$V_m$" , donde

$$V_m = \text{valor muestral ECMI mínimo}$$

Por consiguiente, se selecciona un subconjunto de tamaño q, tal que, $_m$ sea el mínimo de todas las combinaciones de subconjuntos, usando el procedimiento se selección de Forward.

## 3.6.4 PROCEDIMIENTOS COMPUTACIONALES PARA LA SELECCION DE q VARIABLES

Los métodos para obtener combinaciones de subconjuntos de regresión del modelo general (1.3), son presentados abajo, los cuales, luego, son analizados por los criterios vistos arriba, considerando que no presentan multicolinealidad, para su uso en la ecuación final.

## 1. TODAS LAS REGRESIONES POSIBLES

Este procedimiento, requiere hacer todas las combinaciones posibles de regresión del modelo general (1.3), se realiza de la manera siguiente:

a) Efectuar p modelos de regresión, cada una con un solo regresor:

$$Y = \beta_j X_j + e \; , \; j = 1, 2, \ldots ,p$$

b) Efectuar $\binom{p}{2}$ modelos de regresión, cada una con dos variables regresoras

$$Y = \beta_i X_i + \beta_j X_j + e \; , \; i \neq j \; , \; i,j = 1, 2, \ldots ,p$$

c) Continuar con el paso b) considerando $\binom{p}{3}, \binom{p}{4}, \ldots, \binom{p}{p}$ modelos de regresión, cada una con 3, 4,..,p variables regresoras respectivamente

d) Instantáneamente a todos los subconjuntos de modelos hallados en a), b) y c), se les debe hacer el análisis a través de los criterios de $C_q$ de MALLOWS, $R^2$, $\tilde{R}^2$ y así seleccionar el mejor subconjunto de tamaño "q".

Para este tipo de procedimiento, se han desarrollado varios algoritmos de computación muy eficientes que desarrrollan todo el proceso presentado anteriormente; y encontrar el mejor subconjunto de regresión. Estos generalmente están basados sobre

- Reducción de GAUSS - JORDAN.
- Operadores extendidos (SWEEP OPERATOR).

De todos, se recomienda el algoritmo de FURNIVAL y WILSON (1974), llamado "BMD-P"; dicho programa, permite seleccionar el mejor subconjunto para cada tamaño $1 \leq q \leq p$, usando los criterios $C_p$, $R^2_p$, $\tilde{R}^2_p$. El algoritmo BMD-P es eficiente para un conjunto de variables regresoras menor o igual a 30, con tiempo de cómputo comparable a los algoritmos del "procedimiento STEPWISE".

Existe otro programa, llamado "SELECT" (por LA MOTTE (1972)), también se desarrolla de forma adecuada hasta 70 variables regresoras con una cantidad moderada de tiempo de computación.

## 2. PROCEDIMIENTO DE BUSQUEDA DIRECTA EN FUNCION DEL ESTADISTICO t

Se presenta los pasos del procedimiento:

a) En el modelo (1.3), estimar los coeficientes por mínimos cuadrados.

b) Encontrar la desviación estándar, para cada coeficiente estimado.

c) Hallar $|t_j|$ , $j = 1, 2, \ldots, p$ y ordenarlo en forma decreciente, donde

$$t_j = \frac{\beta_j}{De(\beta_j)} \qquad \text{De:desviación estándar}$$

d) Efectuar la regresión de la variable repuesta " Y " con la variable regresora que tiene $|t_j|$ máximo, enseguida evaluarla a través del "criterio $C_p$ de MALLOW´S".

e) Continuar con el paso d) , pero incluyendo en la regresión, la siguiente variable,en el orden de su $|t_j|$ considerando de mayor a menor según c) y así sucesivamente hasta terminar con todas las variables regresoras.

f) En el paso e) calcular el estadístico $C_p$ para cada variable incluida, de estos $C_p$ seleccionar el mejor según el criterio visto.

g) El paso f) , da el mejor subconjunto de variables regresoras de tamaño "q", tal que $C_q$ sea mínimo.

El procedimiento es efectivo, cuando el número de variables regresoras es mayor o igual a 30. Además regresoras con $|t_j|$ grande, siempre tienden a ser incluidas en el mejor subconjunto de tamaño "q".

## 3. PROCEDIMIENTO STEPWISE (Pasos Prudentes)

El método que se presenta evalúa sólamente un número pequeño de subconjuntos de regresión del modelo general (1.3), y se basa en la adición ó eliminanación de las variables regresoras, una a la véz, de acuerdo a los criterios vistos anteriormente. El procedimiento tiene 3 variaciones

- Selección hacia adelante ( Forward selection )
- Eliminación hacia atraz ( Backward elimination )
- Regresión de pasos prudentes ( Stepwise Regresión )

**A) Selección hacia adelante (forward selectión):**

Esta técnica comienza sin variables regresoras en la ecuación de regresión (1.3), y en enseguida, se va adicionando una variable a la véz al modelo, hasta que todas las variables están incluidas ó hasta que uno de los criterios presentados anteriormente es satisfecho.

Generalmente se acostumbra usar un nuevo criterio llamado "Estadístico F-Parcial", para incluir variables al modelo. El criterio se basa en el siguiente algoritmo:

i) Para incluir variables en algún paso, se selecciona el mayor F-Parcial entre los elegibles para ser incluido. Es decir, si la variable "i" es adicionada a la ecuación de "q" términos, entonces:

$$F_i = \max_i \left( \frac{SCRes_q - SCRes_{q+i}}{\hat{\sigma}^2_{q+i}} \right) > F_{in} \qquad (3.6.20)$$

donde $$\hat{\sigma}^2_{q+i} = \frac{SCRes_{q+i}}{n - (q + i)}$$

$$F_{in} = F_{(\alpha,\ 1,\ n-q-1)} \qquad y \qquad 2.0 \le F_{in} \le 4.0$$

ii) En otro caso se termina, y las variables resultantes incluidas son las elegidas.

También se puede aplicar el criterio "$C_p$ de Mallow's", "$CMRes_q$ mínimo", "$R^2$ máximo", en cada paso de esta técnica.

**B) Eliminación hacia atraz (Backward elimination):**

Al contrario de la anterior técnica, esta se inicia con todas las variables regresoras en el modelo (1.3), luego se va elimindo una a la vez en cada paso. Para la eliminación, se aplica el criterio "estadístico F-Parcial" como sigue:

**i)** Para algún paso, la variable con F-Parcial mínima, calculada desde la regresión común, es eliminada si este F-Parcial no excede al valor especificado. Es decir la variable "i" es eliminada de la ecuación de "q" términos si:

$$F_i = \min_i \left[ \frac{SCRes_{q-i} - SCRes_q}{\hat{\sigma}^2} \right] < F_{out} \qquad (3.6.21)$$

$SCRes_{q-i}$, denota la suma de cuadrados de los residuales cuando la variable "i" se elimina de la ecuación de q términos

$$F_{out} = F_{(\alpha,t,n-q)}$$

**ii)** En otro caso termina el proceso, y las variables que no sean eliminadas son las seleccionadas.

Analogamente se pueden aplicar para cada paso los criterios "$C_q$", "CMRes mínimo", "$R^2$ máximo".

**C) Regresión de pasos prudentes (Stepwise Regression):**

Esta técnica es una combinación de las dos anteriores, el algoritmo funciona como sigue:

**i)** En cada paso a realizarce, se adiciona una variable usando el criterio "selección hacia adelante", luego a todas las variables entradas en la ecuación se les calcula sus estadísticos F-Parciales, y se comparan todos con "$F_{out}$", luego ir al paso **ii)**.

**ii)** Si en el paso **i)**, uno de los F-Parcial es menor a $F_{out}$ la variable respectiva se elimina de la ecuación.

iii) Continuar con el paso i), hasta terminar con todas las variables regresoras.

**Observaciónes:**

- Algunos autores prefieren

$$F_{in} = F_{out},$$

pero generalmente es conveniente

$$F_{in} > F_{out}$$

Para el procedimiento de "pasos prudentes", se puede usar un programa de cómputo llamado "STATISTICAL ANALISIS SYSTEM (SAS)" por BARR y GOODNIGHT (1971) ó el programa BMD-P anteriormente mencionado.

El procedimiento de "pasos prudentes" es criticado, pues no asegura que el subconjunto encontrado según las técnicas A), B) ó C) sea el mejor encontrado, esto es debido a que un modelo adecuado puede ser eliminado o dejado de lado a causa de la restricción de solamente adicionar o eliminar una sola variable a la véz. Otra crítica es que emplea un orden de importancia para las variables, esta puede ser engañosa, por ejemplo no es raro hallar que la primera variable incluida en la técnica A) es bastante innecesaria dentro de la presencia de otras variables, ó que la primera variable eliminada según B) sea incluida por A). También el incumplimiento de algunos de los criterios vistos anteriormente para los subconjuntos encontrados por este procedimiento. Luego, parece razonable seleccionar el conjunto que tiene propiedades más o menos similares con el mejor subconjunto encontrado por el procedimiento 1.

- Los análistas se inclinan por la técnica C) seguida de la B).
- Para un número de regresores hasta 30, se puede usar indistintamente el procedimiento: 1., 2. ó 3. pues el costo es igual. Si el número de variables es mayor de 30, primero usar el procedimiento 3. luego a

este conjunto elegido, aplicarle el procedimiento 1 y el sentido común.

## 3.7 FORMULACION DE UNA ESTRUCTURA GENERAL PARA LOS ESTIMADORES SESGADOS.

La estructura (para estimar el vector verdadero $\beta$), de todos los estimadores sesgados en presencia de multicolinelidad, incluyendo el estimador de MCO, se pueden uniformizar en una sola estructura general, y a través de ella estimar $\beta$. De esta manera, no es necesario almacenar la estructura de cada uno de ellos.

Teniendo en consideración las estructuras de los estimadores sesgados y el de MCO, tal como se encontró en (1.9), (3.2.10), (3.3.15), (3.4.4), (3.4.6a), (3.5.2c) y (3.6.2), se observa que todos los estimadores presentan en su estructura al vector variable respuesta o dependiente **Y** (excepto RL) en su forma natural o no estandarizada, por consiguiente por un mecanismo muy simple se puede hacer que las estructuras de todos los estimadores estén en función del vector variable respuesta o dependiente, pero, en su forma estandarizada, con la condición que la estructura de todos los estimadores, no varie.

Para estandarizar **Y**, seguimos los pasos dados en en la **definición 2, sección 1.1 (capitulo I)**, luego, reemplazando en las estructuras correspondientes (excepto en RL), el vector respuesta **Y** por su forma estandarizada $\mathbf{Y}^*$, y luego, multiplicada por $\eta$, se tendrá las estructuras corregidas, de los estimadores para este caso:

$$\hat{\beta}_{\mathbf{MCO}} = \eta(\mathbf{X'X})^{-1}\mathbf{X'Y}^* \qquad (3.7.1)$$

$$\hat{\beta}_{\mathbf{CP}} = \eta(\mathbf{T\Lambda^{+}T'})\mathbf{X'Y}^* \qquad (3.7.2)$$

$$\hat{\beta}_{RL} = -\eta \left[ \sum_{k=r}^{p} l_k^{-1} \gamma_{ok}^2 \right]^{-1} \sum_{j=r}^{p} \left[ l_j^{-1} \gamma_{0j} \delta_j \right] \qquad (3.7.3)$$

$$\hat{\beta}_{RR} = \eta \ (\mathbf{X'X} + k\mathbf{I})^{-1}\mathbf{X'Y}^* \qquad (3.7.4)$$

$$\beta_{RG} = \eta \mathbf{T} \ (\Lambda + \mathbb{K})^{-1} \ \mathbf{T'X'Y}^* \qquad (3.7.5)$$

$$\beta_{CS} = \eta \mathbf{d} \ (\mathbf{X'X})^{-1}\mathbf{X'Y}^* \qquad (3.7.6)$$

$$\beta_{SV} = \eta ( \mathbf{X}_q'\mathbf{X}_q )^{-1} \ \mathbf{X}_q'\mathbf{Y}^* \qquad (3.7.7)$$

Las ecuaciones (3.7.1) hasta (3.7.7) parecen indicar varias inversiones de matrices y grandes requirimientos de memoria para calcular todos los estimadores sesgados. Realmente ellos se pueden expresar en una sola estructura general sin dificultad alguna, sólo que deben considerarse algunas anotaciones. Por consiguiente realizando algunas operaciones y simplificaciones se encuentra una estructura general de estos estimadores, el cual es expresado de la forma siguiente:

$$\beta = \eta \sum_j h_j \ m_j \qquad (3.7.8)$$

donde

$\eta$, es el mismo de la definición 2, sección 1.1 (capítulo I)

$h_j$, son variables univariadas.

$m_j$, son vectores latentes de la matriz de correlación $\mathbf{X'X}$ ó $\mathbf{A'A}$ según sea el caso.

Especificamente para los siete estimadores en estudio, $h_j$ y $m_j$ son definidos como sigue:

$$\text{MCO:}\quad m_j = T_j, \qquad h_j = \lambda_j^{-1}\, T_j' X' Y^*, \qquad j = 1, 2, \ldots, p$$

$$\text{CP:}\quad = T_j, \qquad h_j = \begin{cases} 0 & j = 1,2, \quad r \\ \lambda_j^{-1}\, T_j' X' Y^*, & j = r+1, \ldots, p \end{cases}$$

$$\text{RR:}\quad m_j = T_j, \qquad h_j = (\lambda_j + K)^{-1}\, T_j' X' Y^*, \qquad j = 1, 2, \ldots, p$$

$$\text{RG:}\quad m_j = T_j, \qquad h_j = (\lambda_j + K_j)^{-1}\, T_j' X' Y^*, \qquad j = 1, 2, \ldots, p$$

$$\text{CS:}\quad m_j = T_j, \qquad h_j = d\lambda_j^{-1}\, T_j' X' Y^*, \qquad j = 1, 2, \ldots, p$$

$$\text{RL:}\quad m_j = \delta_j, \qquad h_j = \begin{cases} 0 & = 1,2, \ldots, r \\ -\left[ \sum_{k=r}^{p} l_k^{-1} \gamma_{ok}^2 \right]^{-1} \left[ l_j^{-1} \gamma_{oj} \right] & j = r+1, \ldots, p \end{cases}$$

$$\text{SV:}\quad m_j = H_j, \qquad h_j = S_j^{-1} H_j' X_q' Y^*, \qquad j = 1, 2, \ldots, q, \quad q \leq p$$

De esta forma, (3.7.8) expresa una estructura general para todos los estimadores de (3.7.1) a (3.7.7) en presencia de multicolinealidad. Notemos que (3.7.8) es aplicable a cualquier modelo de regresión lineal múltiple cuando se detecta multicolinealidad, sólo que las matrices de entrada $X^*$ e $Y$ deben ser estandarizadas en $X$ e $Y^*$ respectivamente, y luego, desarrollar la estimación de los coeficientes verdaderos de (1.3), por cualquiera de los métodos estudiados.

# C A P I T U L O IV

# APLICACION VIA SIMULACION DE LA EFICIENCIA DE LOS ESTIMADORES SESGADOS EN PRESENCIA DE MULTICOLINEALIDAD

## 4.1 INTRODUCCION

En este capítulo, se compara la eficiencia empírica de los estimadores sesgados, por lo tanto se observa su comportamiento en la práctica, ante un problema de MULTICOLINEALIDAD. Se diseña una simulación específica del modelo de regresión lineal múltiple estandarizado (1.3), en presencia de este fenómeno, y a través de este, se observa la eficiencia de los estimadores sesgados vistos anteriormente. Los resultados de la simulación (algunos de ellos están en el **apéndice II**), son comparados y analizados, observando la eficiencia de los estimadores sesgados através de los **criterios de la VARIANZA y el ERROR CUADRATICO MEDIO**. Finalmente, se presenta, algunos resultados significativos realizados por investigadores el fenómeno de multicolinealidad.

## 4.2 DISEÑO DE LA SIMULACION DEL CONJUNTO DE DATOS.

Se presenta el diseño de simulación, que se desarrolla respecto al estudio de investigación que se considera, cuyos resultados se verá en la siguiente sección. Es preferible, comparar la eficiencia de los estimadores sesgados, através de una simulación, pues el investigador controla los valores verdaderos, que con datos reales, pero, con este último también es factible.

Para el estudio se generan dos matrices (30x6) de variables regresoras o independientes. La primera matriz, de datos casi ortogonal, consistiendo de observaciones aleatorias de una distribución Uniforme (0, 10). La segunda matriz de datos multicolineales, formada por cinco columnas de observaciones aleatorias de una distribución Uniforme (0, 10) y una columna, que es, una combinación lineal de dos de estas cinco columnas, más un vector de error aleatorio proveniente de una distribución normal con media cero y varianza dos. Por consiguiente a la última matriz se le induce una multicolinealidad de tres variables.

Seguidamente cada matriz generada, es estandarizada. El paso siguiente es obtener las matrices de correlación, obteniendo a continuación las raices latentes y vectores latentes respectivamente. Por lo tanto las **TABLAS I** y **II** nos presentan los autovalores y autovectores de la matriz de correlación $X'X$, para los datos próximamente ortogonales, como también para los datos multicolineales:

**TABLA I**

Autovalores y Autovectores de la Matriz
de Correlación $X'X$, para datos Ortogonales

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | 6 |
|---|---|---|---|---|---|
| 0.09669492 | 0.38721766 | 0.55228224 | 1.14842250 | 1.73313444 | 2.08224825 |
| T1 | T2 | T3 | T4 | T5 | T6 |
| 0.45413100 | -0.15819680 | 0.59188218 | 0.22120837 | -0.37333014 | 0.47969339 |
| 0.58911491 | -0.05648169 | 0.26436200 | -0.53564565 | 0.19212818 | -0.50600067 |
| 0.41904722 | 0.17482937 | -0.59825962 | -0.37811673 | -0.04369363 | 0.53947952 |
| 0.44647282 | 0.32736533 | -0.31582605 | 0.59066248 | -0.27496830 | -0.41140994 |
| -0.26027234 | 0.60582687 | 0.19657749 | -0.38292970 | -0.60921042 | -0.09389845 |
| 0.06346986 | 0.68339157 | 0.28893743 | 0.16019056 | 0.61241925 | 0.21152862 |

**TABLA II**

Autovalores y Autovectores de la Matriz de correlación **X'X**, para datos Multicolineales

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ |
|---|---|---|---|---|---|
| 0.00147185 | 0.19137525 | 0.64758487 | 1.15412705 | 1.88331847 | 2.12212251 |
| T1 | T2 | T3 | T4 | T5 | T6 |
| -0.52543631 | 0.21799007 | -0.48105168 | 0.13967308 | -0.59586909 | -0.26536344 |
| 0.53622374 | -0.02828457 | -0.56191489 | 0.17502332 | 0.25541735 | -0.54776323 |
| 0.66050649 | 0.19562111 | 0.08388769 | -0.02962423 | -0.68434180 | 0.22186732 |
| 0.00790755 | 0.54427480 | -0.47375853 | -0.45654569 | 0.26927909 | 0.44532019 |
| -0.00695733 | 0.61773344 | 0.46099051 | -0.27131104 | 0.05201037 | -0.57404702 |
| 0.00101291 | 0.48536897 | 0.09404462 | 0.81666213 | 0.19016307 | 0.22906883 |

De la **TABLA I**, se concluye, que la raiz latente más pequeña es $\lambda_1 = 0.097$, correspondiente a la matriz de correlación de los datos próximamente ortogonales y de la **TABLA II**, el autovalor más pequeño es $\lambda_1 = 0.0014$, con su autovector respectivo $T_1' = (-0.525,\ 0.536,\ 0.660,\ 0.007,\ -0.006,\ 0.001)$, correspondiente a la matriz de correlación de los datos múlticolineles. Esto último revela que la multicolinealidad implica los primeros tres variables independientes o regresoras.

Posteriormente conociendo las raíces latentes de las matrices de correlación arriba mencionadas, se obtiene el "**NUMERO DE CONDICION o el grado de mal condicionamiento**", el cual estadefinido en la sección **2.2.4** parte 3), como

$$\Psi = \lambda_{max} / \lambda_{min}$$

La <u>simulación es realizada teniendo en consideración el siguiente criterio:</u>

i) La orientación del vector de coeficientes $\beta$ en el modelo (1.3) con un subconjunto de vectores latentes de la matriz de correlación **X'X**.

ii) La magnitud del vector de coeficientes $\beta$ relativo a $\sigma$ (desviación

estandar verdadero del vector de error aleatorio $e$ del modelo (1.3)).
iii) la fuerza de multicolinealidad.
Por consiguiente, según el resultado de la **TABLA II**, se hacen tres orientaciones del vector de coeficientes $\beta$ con el autovector perteneciente a la raíz latente más pequeña de la matriz de correlación para datos múlticolineales, $T_1$. Por tanto para lograr lo anterior, se hace que el vector de coeficientes $\beta$ tome tres valores diferentes, donde cada uno de estos, son conbinaciones lineales de los vectores latentes de la matriz de correlación para los datos multicolineales:

$$\left.\begin{array}{l} \text{a) } \beta = T_6 \\ \text{b) } \beta = 0.5\ (T_1 + T_2 + T_5 + T_6) \\ \text{C) } \beta = T_1 \end{array}\right\} \qquad (4.1)$$

Para diferenciar los tres valores que toma $\beta$, se denota por convención, que el primer valor sea $\beta = \beta1$, el segundo $\beta = \beta2$, y el tercero $\beta = \beta3$, luego por (4.1) la relación debe ser

$$\beta1 = T_6, \qquad \beta2 = 0.5(T_1 + T_2 + T_5 + T_6), \qquad \beta3 = T_1$$

además, de lo anterior se deduce que $\beta'\beta = 1$ para cada uno de los valores de $\beta$ (por ser los $T_i'$s vectores latentes ortonormales), luego <u>los valores de las tres orientaciones que se diseñarón son</u>

$$\begin{array}{l} \text{a) } \phi = T_1'\beta1 = 0.0 \\ \text{b) } \phi = T_1'\beta2 = 0.5 \\ \text{C) } \Phi = T_1'\beta3 = 1.0 \end{array} \qquad (4.2)$$

Luego, para proveer la "**variación de la razón de la señal para el ruido (error)**", la cual esta definido por

$$\rho = \beta' \beta \,/\, \sigma^2 \qquad (4.3)$$

a "$\sigma$" (desviacion estandar del error aleatorio), se indica los valores siguientes

$$\sigma = 5.0, \quad 1.0, \quad 0.1, \quad 0.01$$

por consiguiente "$\rho$" es respectivamente

$$\rho = 0.04, \ 1.0, \ 100, \ 10000$$

De esta manera, se obtienen 24 configuraciones (dos matrices de datos estandarizados ortogonal y múlticolineal "**X**", tres orientaciones de $\beta$ con el autovalor $T_1$ para datos múlticolineales, y cuatro razones de la señal para el ruido $\rho$), y para cada uno de estas configuraciones (1.3) es usado con "$\beta_0 = 0$" para generar <u>10 muestras de tamaño n = 30 de la variable respuesta **Y**.</u> Posteriormente para cada muestra así generada, se obtiene la estimación del vector $\beta$ utilizando (3.7.8), para cada uno de los estimadores MCO, CP, RR, RL, RG, CS y SV. Así, se obtienen 10 estimaciones del vector $\beta$ para cada configuración y cada estimador sesgado. Luego, para estas estimaciones, se calculan para cada coeficiente estimado del vector $\beta$, la Varianza Estimado Individual $V(\beta_j)$ y el Error Cuadrático Estimado Individual $EC(\beta_j)$, así como el Error Cuadrático Estimado Total $EC(\beta)$, los cuales se llevan a cabo através de las formulas siguientes

$$V(\hat{\beta}_j) = \sum_{i=1}^{10} \left( \hat{\beta}_j^i - \bar{\beta}_j \right)^2 / \ 10 \ , \quad j = 1, 2, \ldots, 6 \qquad (4.4)$$

$$EC(\beta_j) = (\hat{\beta}_j - \beta_j)^2 \,/\, \sigma^2 , \qquad j = 1, 2, \ldots, 6 \qquad (4.5)$$

$$EC(\beta) = \sum_{j=1}^{p} EC(\beta_j), \qquad p = 6 \qquad (4.6)$$

Posteriormente realizado lo anterior, se calculan, para las 10 muestras de Errores Cuadráticos Estimados Individuales (obtenidos por (4.5)), de cada coeficiente, así como también para las 10 muestras de Errores Cuadráticos Estimados Totales (obtenidos por (4.6)), sus valores promedios y desviaciones Estándar para cada modelo de configuración y estimador Sesgado.

Se refiere a los Promedios de Errores Cuadráticos Estimados Totales, como el "**Error cuadrático medio estimado (ECM)**", en el resto de esta investigación. La cual es comparada, así como también la varianza estimada individual $V(\beta_j)$, para cada uno de los estimadores vistos en los capítulos anteriores.

## 4.3. ANALISIS Y COMPARACION DE LOS RESULTADOS DE LA SIMULACION.

Se presentan ordenadamente todos los resultados de la simulación (algunos se encuentran en el **Apéndice II**), y cada resultado, es analizado y algunos comparado, presentando la interpretación debida para aclarar el efecto de multicolinealidad. Ellos son presentados en dos fases, en la primera sobre

i) Generación de las matrices de datos.

ii) Configuración de las matrices de datos.

y en la segunda, sobre

i) Los coeficientes Estimados del modelo Regresión diseñado en (1.3), de acuerdo a los datos generados en la primera fase de la simulación.

ii) Los valores de la Varianza Individual de los coeficientes estimados del modelo (1.3), según la simulación en la sección **4.2**.

iii) El Error Cuadrático Medio Estimado de los coeficientes estimados del modelo (1.3), según la simulación en la sección **4.2**.

## 4.3.1. ANALISIS DE LOS RESULTADOS DE LA GENERACION DE LAS MATRICES DE DATOS.

El **CUADRO I**, **Apéndice II**, presenta la matriz aleatoria generada **A**, de dimensión 30x6. Esta es una matriz cuyos valores están en el rango entre "0 y 10", pues, consiste de observaciones aleatorias de una distribución uniforme <0,10>, según el diseño en la sección **4.2**. En el **CUADRO III**, **Apéndice II**, se tiene a la matriz de datos casi ortogonal $X^{o}1$ (la matriz **A**, luego, de ser ortogonalizada por el proceso de Grand Smith). Por otro lado el **CUADRO II, Apéndice II**, corresponde a la segunda matriz aleatoria generada **B** de dimensión 30x5, cuyos valores están en el rango de "0 a 10". Esta se genera así, con la finalidad de inducir una sola multicolinealidad al momento de ser reestructurada (según el diseño, en la sección **4.2**). Posteriormente se observa el **CUADRO IV, Apéndice II**, correspondiente a la matriz $X^{o}2$ de dimensión 30x6 (matriz **B**, luego de ser ortogonalizada y reestructurada). De estos cuadros se obtiene los **CUADROS V** y **VI, Apéndice II**, que muestra la matriz de datos casi ortogonal **X1** ($X^{o}$**1**, luego de ser estandarizada) y la matriz de datos múlticolineal **X2** ($X^{o}$**2**, luego de ser estandarizada ), respectivamente.

Como se observa, todas las matrices mencionadas arriba, cumplen con las condiciones del diseño de simulación definidas en la sección **4.2**, por ejemplo las matrices **A** y **B** para ser generadas se implementa el método "Congruencial Multiplicativo" con ciertos valores iniciales. También en la reestructuración de la matriz $X^{o}2$, se seleciona aleatoriamente las calumnas 1 y 2 de la matriz **B**. Así también, para estandarizar las matrices $X^{o}1$ y $X^{o}2$, se utiliza la fórmula (1.2), sección **1.1**.

### 4.3.2. ANALISIS SOBRE LA CONFIGURACION DE LAS MATRICES DE DATOS

En esta subsección se examina las matrices de datos estandarizados X1 y X2, mediante los métodos recomendados en la subsección 2.2.4 del Capítulo II, si son o no multicolineales.

**a) EXAMINACION DE LA MATRIZ DE CORRELACION.**

La TABLA III (vease TABLA III, página 113), presenta la matriz de correlación de las matrices de datos estandarizados X1 y X2 respectivamente. Se aprecia de la TABLA III, para la matriz de correlación X1, que la correlación entre las variables independientes $X_1$ y $X_2$; $X_2$ y $X_3$; $X_3$ y $X_4$; $X_5$ y $X_6$, están medianamente correlacionadas de manera inversa; y en el resto de las correlaciones su valor es bajo. Ahora para la matriz de correlación X2, la correlación entre las variables independientes $X_1$ y $X_3$, $X_1$ y $X_4$; $X_2$ y $X_3$, $X_2$ y $X_5$; $X_4$ y $X_5$; $X_5$ y $X_6$, están también medianamente correlacionadas, mientras que el resto es bajo. Por consiguiente se confirma lo expuesto en la subsección 2.2.4 del capítulo II, que la matriz de correlación no es un buen indicador para detectar la multicolinealidad cuando se trata más de dos variables regresoras o independientes.

**b) CALCULANDO EL DETERMINANTE DE LA MATRIZ DE CORRELACION.**

El determinante de las matrices de correlación de X1'X1 y X2'X2 respectivamente, se presenta en la TABLA IV (vease tabla IV, página 114), se obseva, que el determinante de X2, es próximo a cero. Por lo tanto se concluye que existe multicolinealidad en la matriz de datos X2, en cambio en X1 no.

**c) EXAMINACION DE LOS FACTORES DE LA VARIANZA.**

La TABLA IV, también, presenta los factores de inflación de la varianza para cada variable predictora de la matriz de datos estandarizadas X1 y X2 respectivamente. Según las candiciones de este método, para el caso de la

# TABLA III

**Comparación de las Matrices de correlación para X1 (datos casi ortogonales) y X2 (datos multicolineales), para detectar la multicolinealidad.**

**MATRIZ DE CORRELACION X1'X1 (De dimensión 6x6)**

| | | | | | |
|---|---|---|---|---|---|
| 1.00000000 | -0.65005739 | 0.28319735 | -0.18665443 | 0.21882734 | -0.08890166 |
| -0.65005739 | 1.00000000 | -0.41765781 | -0.04927327 | 0.13225882 | -0.08663078 |
| 0.28319735 | -0.41765781 | 1.00000000 | -0.55321057 | 0.07245289 | 0.07504715 |
| -0.18665443 | -0.04927327 | -0.55321057 | 1.00000000 | 0.14228077 | -0.32542825 |
| 0.21882734 | 0.13225882 | 0.07245289 | 0.14228077 | 1.00000000 | -0.56833699 |
| -0.08890166 | -0.08663078 | 0.07504715 | -0.32542825 | -0.56833699 | 1.00000000 |

**MATRIZ DE CORRELACION X2'X2 (de dimensión 6x6)**

| | | | | | |
|---|---|---|---|---|---|
| 1.00000000 | 0.22349963 | 0.61977692 | -0.45627223 | 0.10333048 | -0.21980235 |
| 0.22349963 | 1.00000000 | -0.62414096 | -0.31088437 | 0.46640021 | -0.04668289 |
| 0.61977692 | -0.62414096 | 1.00000000 | -0.12712955 | -0.27987253 | -0.14187725 |
| -0.45627223 | -0.31088437 | -0.12712955 | 1.00000000 | -0.45024267 | -0.09569074 |
| 0.10333048 | 0.46640021 | -0.27987253 | -0.45024267 | 1.00000000 | -0.43068868 |
| -0.21980235 | -0.04668289 | -0.14187725 | -0.09569074 | -0.43068868 | 1.00000000 |

## CUADRO IV

**Comparación de otros resultados, obtenidos por diferentes métodos, para detectar multicolinealidad entre las matrices de datos: X1 y X2**

| **El determinante de X1, es:** | **El determinante de X2, es:** |
|---|---|
| 0.0857010861617149 81 | 0.0008413850527306 94 |
| **FACTORES DE INFLACION DE LA VARIANZA DE LA j-ESIMA VARIABLE PREDICTORA DE X1** | **FACTORES DE INFLACION DE LA VARIANZA DE LA j-ESIMA VARIABLE PREDICTORA DE X2** |
| **j = 1** 3.5616750091 | **j = 1** 189.7515449823 |
| **j = 2** 0.8517748812 | **j = 2** 32.9191836302 |
| **j = 3** 4.6007831635 | **j = 3** 159.1962851004 |
| **j = 4** 1.9066244863 | **j = 4** 13.9498118776 |
| **j = 5** 2.0002790549 | **j = 5** 242.3786619276 |
| **j = 6** 3.7418629671 | **j = 6** 49.8582930447 |
| **ANALISIS DEL AUTOSISTEMA DE X1'X1** | **ANALISIS DEL AUTOSISTEMA DE X2'X2** |
| Número de condición: $\psi = \lambda$ máx/$\lambda$ mín | Número de condición: $\psi = \lambda$ máx/$\lambda$ mín |
| 21.534205791140202800 | 1441.803108303449040000 |

matriz X1 , todos los factores de inflación de la varianza son menores que cinco; en cambio para la matriz X2 todos los factores de inflación de la varianza son muchisimos mayores que diez, luego hay presencia de multicolinealidad en la matriz de datos estandarizados X2. Por consiguiente se concluye que los coeficientes de regresión asociados a las variables de la matriz X1 son estimados significativamente, más no, el caso de la matriz X2, por la existencia de multicolinealidad.

**d) ANALISIS DEL AUTOSISTEMA DE LA MATRIZ DE CORRELACION.**

De forma similar la **TABLA IV**, también presenta el número de condición de $X'X$, $\Psi \cong 21.53$ para la matriz de correlación $X1'X1$ y $\Psi \cong 1441.80$ para la matriz de correlación $X2'X2$. Estos resultados, de acuerdo al método 3) de la subsección 2.2.4, concluyen, que la matriz X1 no tiene multicolinealidad, en cambio en X2 si existe.

Se debe anotar también algo importante, las **TABLAS I y II** (vease **TABLAS I y II**, paginas 106 y 107), presentán los valores y vectores propios asociados de la matriz de correlación $X1'X1$ y $X2'X2$ respectivamente. Además recordar, que el capítulo II, se dice que es posible detectar qué variables están implicadas en multicolinealidad e incluso se pueden encontrar los coeficientes, tal que las variables implicadas, multiplicadas por estos coeficientes forman una combinación linealmente dependiente. Así pues, en la columna uno, de la **TABLA II**, para el valor propio $\lambda_1 = 0.00147$ se tiene su vector propio asociado $T_1$, según el cual sus primeros tres valores son altos, esto significa, que las variables regresoras implicadas en multicolinealidad son $X_1$, $X_2$, $X_3$ y los coeficientes son precisamente estos valores.

Por consiguiente, luego de realizarce el análisis de la configuración de las matrices de datos, se diagnostica que existe multicolinealidad en la

matriz de datos X2 y las variables regresoras implicadas en este fenómeno son $X_1$, $X_2$, y $X_3$. En cambio en la matriz X1 no existe multicolinealidad.
A continuación se tiene los resultados de la segunda fase de la simulación, los cuales se analizan y comparan.

### 4.3.3. ANALISIS Y COMPARACION DE LOS ESTIMADORES SESGADOS SEGUN LOS COEFICIENTES ESTIMADOS DEL MODELO DE REGRESION.

Los CUADROS VII-IX y X-XII, Apéndice II, presenta los coeficientes estimados del vector verdadero $\beta$ y sus desviaciones estándar asociadas, para las 24 configuraciones diseñadas y cada uno de los estimadores sesgados MCO, CP, RL, RR, RG, CS Y SV.
Cada uno de los coeficientes estimados, que figuran en estos cuadros, representa el promedio de un conjunto de 10 muestras de coeficientes, obtenidas para cada uno de los componentes del vector verdadero $\beta$, para cada uno de las 24 configuraciones y cada uno de los estimadores sesgados.
Se aprecia de los CUADROS VII-IX, apéndice II, para los datos X1 proximamente ortogonales, que en general, la estimación del vector de coeficientes verdadero del modelo de regresión lineal, es mejor con el estimador RR y RG, que el resto de los estimadores (incluido MCO), cuando la razón de la señal del ruido $\rho$ (según definido en (4.3)), tiene valores menores, más claramente $\rho$ = 0.04, 1.0. En cambio, cuando $\rho$ toma valores grande, los estimadores RR, MCO, y CS son bastantes mejores que RG, estas conclusiones son válidas para las tres orientaciones diseñadas según (4.1) y (4.2).
También se observa que los valores estimados de los coeficientes verdaderos, cuando se utiliza los métodos de estimación CP, RL, SV y MCO, son todos iguales. Esto significa, que el programa diseñado, realiza el

diagnostico de que la matriz **X1** , no tiene ninguna variable independiente eliminada, según las pruebas de "criterios de eliminar variables independientes", vistos en el capítulo III, para cada estimador sesgado. Por consiguiente los resultados reflejan esta interesante conclusión.

Por otro lado, los valores estimados del coeficiente verdadero $\beta$, para los estimadores de MCO, CP, SV y RL, en todo momento es siempre grande en comparación al resto de los estimadores sesgados, cuando la razón de la señal del ruido es $\rho$ = 0.04, 1.0.

Los **CUADROS X-XII, Apéndice II,** para los **datos multicolineales X2**, son analizados, apreciandose, que los valores estimados del vector de coeficientes verdadero $\beta$, por los métodos sesgados RG, CP, RL, RR, SV, son mejores, cuando se considera $\rho$ = 0.04, 1.00, en cambio el MCO lo hace pobremente. De todo este grupo, sobresale más RG.

Para $\rho$ = 100, $\phi$ = 0.5, 0.1, el que mejor estima a los coeficientes verdaderos es RR; para la razón de la señal del ruido $\rho$ = 10000, todos los estimadores presentados, incluido MCO, estiman los tres últimos coeficientes del vector $\beta$ mejor, sobresaliendo ligeramente MCO, en cambio para los tres primeros no.

Se debe indicar que las conclusiones anteriores para datos multicolineales, se cumplen para todas las orientaciones de $\phi$.

### 4.3.4. ANALISIS Y COMPARACION DE LOS ESTIMADORES SESGADOS SEGUN LA VARIANZA INDIVIDUAL DE LOS COEFICIENTES ESTIMADOS DEL MODELO DE REGRESION.

Para que se realice el análisis y comparación respecto a la varianza individual de los coeficientes estimados del vector verdadero $\beta$ del modelo de regresión (1.3), se debe en realidad análizar y comparar los valores de su desviación estándar individual, pues es más conveniente, ya que está en la misma dimensión de los coeficientes estimados. Por lo tanto los **CUADROS VII-IX** y **X-XII, Apéndice II,** presentán también estos valores para las 24 configuraciones diseñadas, los cuales se encuentran señaladas en parentesis para todos los métodos de estimación definidos.

Para los primeros **CUADROS VII-IX**, para los datos próximamente ortogonales, y considerando las tres configuraciones de $\phi$, según el diseño de la simulación, se observa, que los coeficientes estimados por el método RR, tiene menor desviación estándar, seguido de RG, cuando $\rho$=0.04, 1.00, 100; en cambio por el método de MCO es mayor que todos. Para $\rho$=10000, en general la desviación estándar de los coeficientes estimados para todos los métodos son iguales y pequeños.

Por otro lado, de los **CUADROS X-XII,** (para datos multicolineales), se aprecia, que los coeficientes estimados por el método RG, seguido de RR, CP y RL tienen su desviación estándar bastante bajo en comparación de MCO (la cual es alto), cuando la razón de la señal del ruido $\rho$ = 0.04, 1.00, 100. Mientras que para $\rho$ = 10000, todos los coeficientes estimados por todos los métodos, tienen desviación estándar bastante bajo, sobresaliendo ligeramente mejor el estimador CP. Todo los resultados anteriores se cumple para todas las orientaciones $\phi$, definidas en el diseño.

### 4.3.5 ANALISIS Y COMPARACION DE LOS ESTIMADORES SESGADOS SEGUN EL ERROR CUADRATICO MEDIO ESTIMADO DE LOS COEFICIENTES DEL MODELO (1.3)

los resultados del Error Cuadrático Medio Estimado (ECM), de los coeficientes del vector estimado $\beta$ para las 24 configuraciones y para los siete estimadores examinados (según el diseño de la sección **4.2**), se presenta en los **CUADROS XIII-XV, Apéndice II**, para los datos próximamente ortogonales y en los **CUADROS XVI-XVIII, Apéndice II**, para los datos multicolineales. Cada cuadro presenta en la parte superior, una configuración definida, luego para cada valor de la razón de la señal de ruido $\rho$, se presentan 7 columnas que corresponden respectivamente a los resultados del ECM estimado y su desviación estándar (en parentesis), para cada uno de los estimadores sesgados (incluido MCO). Los coeficientes de cada estimador, son para esta simulación seis.

Claramente se observa, de los **cuadros XIII -XV (esta incluido X1, matriz de datos próximamente ortogonales)**, que el ECM estimado para cada coeficiente del vector estimado $\beta$ es mucho mejor, para los estimadores sesgados RR, RG y CS, en cambio utilizando MCO no es bueno, todo esto ocurre cuando la razón de la señal del ruido es $\rho$ = 0.04, 1.00 y para todas las orientaciones de $\phi$; pero para $\rho$ = 100, el mejor es RR; para $\rho$ = 10000,y todas las orientaciones de $\phi$, el ECM estimado de cada coeficiente son apróximadamente iguales cuando se trabaja con todos los estimadores sesgados (incluido MCO), excepto, RG que en este caso es ligeramente alto. Por lo tanto se concluye de las comparaciones efectuadas entre los estimadores sesgados, que **en general el menor** ECM **estimado es de RR o RG.** Como observación, para los primeros tres cuadros, el ECM estimado de MCO, CP, SVy RL, son iguales (mayor explicación, vease sección **4.3.3**, página 116).

También de los cuadros **XIII -XV**, se tiene **la desviación estándar, de los errores cuadráticos estimados, para cada coeficiente del vector estimado** $\beta$, y se observa que la menor variación es con el método de RR y RG, pero en general, todos los métodos estudiados tienen pequeñas variaciones.

**Para los CUADROS XVI-XVIII, se observa:**

Ellos presentan, el ECM estimado y su desviación estándar, de cada coeficiente del vector estimado $\beta$, **cuando esta presente la matriz X2 de datos multicolineales.** Analizando los resultados, se aprecia para todas las orientaciones ($\phi$ = 0.0, 0.5, 1.0), que las tres primeras filas, correspondientes a las columnas MCO, RR, CS, para la cual $\rho$ = 0.04, 1.00, y también correspondientes a la mayoria de las columnas (para $\rho$ - 100, 10000), los valores del ECM estimado de los coeficientes del vector estimado $\beta$, son mayores, comparado con las tres filas restantes respectivamente. La explicación de lo anterior es de acuerdo a la sección **4.3.2** parte **d)**, que las variables $X_1$, $X_2$ y $X_3$ implicados en la multicolinealidad, están influyendo en los valores estimados de los coeficientes respectivos. Casos particulares ocurre con RG y RL para $\rho$ = 10000, $\phi$ = 0.00, cuyos valores del ECM estimado para todos sus coeficientes son mayores, del mismo modo con RL, para $\rho$ = 10000, $\phi$ = 0.5. Por consiguiente el ECM estimado más bajo de los coeficientes del vector estimado $\beta$, se consigue a través del método de RG, CP, RL, SV y RR, en cambio con el método usual MCO se obtiene valores demasiados altos que el resto de los métodos sesgados. Esta conclusión es cuando las configuraciones son $\rho$ = 0.04, 1.00, 100 y para todas las orientaciones de $\phi$. Entre ellos el que más sobresale es RG, CP y RR en ese orden.

Para la configuración $\rho$ = 10000 y todas las orientaciones, el ECM estimado, es alto para todos, siendo menor dentro de este grupo MCO y CS.

Sobre la desviación estándar del ECM estimado, correspondiente a cada coeficiente del vector estimado $\beta$, cuando se trabaja con datos multicolineales, se aprecia que la de menor variación (bastante baja), es de RG, seguido de CP, RR, RL y SV cuando $\rho$ = 0.04, 1.00 y 100, para todas las orientaciones de $\phi$; pero para $\rho$ = 10000, y todas las orientaciones de $\phi$, el de menor variación es CP, RG, RR.

### 4.3.6 ANALISIS Y COMPARACION DE LOS ESTIMADORES SESGADOS A TRAVES DEL $\hat{ECM}$ ( PROMEDIO $\hat{ECM}$ TOTAL )

las TABLAS V y VI (vease TABLAS V y VI , páginas 122-123), presentan el ECM (Promedio estimado del Error Cuadrático Medio Total) y su desviación estándar asociada de las 24 configuraciones diseñadas y los 7 estimadores investigados, según el diseño de la sección 4.2. Los cuadros son de doble entrada, cada uno tiene tres divisiones correspondientes a una determinada orientación $\phi$, y en cada una de estás divisiones se hace la comparación del ECM estimado para los estimadores sesgados, considerando el valor correspondiente de la razón de la señal del ruido $\rho$ diseñada.

La primera TABLA V, corresponde a los datos próximamente ortogonales y la TABLA VI , a los datos multicolineales. Para el primero, se observa que el ECM de los estimadores CP, SV, RL y MCO, son iguales para todas las configuraciones $\phi$. Este resultado se verifica, por que la matriz de datos **X1** , no es multicolinealidad, y por lo tanto los métodos mencionados, se transforman a MCO. Por ejemplo la TABLA I, para los datos próximamente ortogonales, $\lambda_1$ = 0.097 es mayor que 0.01 (condición para no eliminar ninguna variable independiente del modelo de regresión (1.3)) , luego, estos estimadores optan por el método MCO.

**Resultados para los primeros 12 configuraciones (datos ortogonales, vease TABLA V, página 122):**

Los valores del ECM (en particular para esta simulación), para las primeras 12 configuraciones, son ligeramente bajos en comparación, a los

**TABLA V**
**COMPARACION DEL PROMEDIO ESTIMADO DEL ERROR CUADRATICO TOTAL Y SUS DESVIACIONES ESTANDAR ASOCIADO (en parentesis), PARA LA MATRIZ X1 (datos casi ortogonales), Y SEGÚN LAS CONFIGURACIONES DISEÑADAS.**

| | | $\phi = 0.0$ | | |
|---|---|---|---|---|
| **ESTIMADORES** | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
| LS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| RR | 0.783 | 0.635 | 2.801 | 2.106 |
| | ( 0.219) | ( 0.226) | ( 0.558) | ( 0.629) |
| PC | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| SE | 1.638 | 1.465 | 4.101 | 2.117 |
| | ( 0.313) | ( 0.555) | ( 0.998) | ( 0.633) |
| GR | 0.660 | 0.448 | 7.449 | 3.925 |
| | ( 0.292) | ( 0.208) | ( 2.279) | ( 0.313) |
| VS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| LR | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| | | $\phi = 0.5$ | | |
| **ESTIMADORES** | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
| LS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| RR | 0.782 | 0.646 | 3.031 | 2.108 |
| | ( 0.220) | ( 0.224) | ( 0.661) | ( 0.629) |
| PC | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| SE | 1.637 | 1.466 | 4.146 | 2.110 |
| | ( 0.314) | ( 0.559) | ( 1.031) | ( 0.628) |
| GR | 0.674 | 0.470 | 4.578 | 2.386 |
| | ( 0.296) | ( 0.212) | ( 0.334) | ( 0.723) |
| VS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| LR | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| | | $\phi = 1.0$ | | |
| **ESTIMADORES** | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
| LS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| RR | 0.781 | 0.634 | 3.181 | 2.148 |
| | ( 0.219) | ( 0.229) | ( 0.532) | ( 0.658) |
| PC | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| SE | 1.634 | 1.443 | 4.221 | 2.121 |
| | ( 0.312) | ( 0.538) | ( 0.952) | ( 0.636) |
| GR | 0.680 | 0.376 | 5.211 | 33.871 |
| | ( 0.298) | ( 0.209) | ( 0.287) | ( 9.623) |
| VS | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |
| LR | 2.682 | 2.519 | 4.364 | 2.118 |
| | ( 0.413) | ( 0.690) | ( 1.056) | ( 0.636) |

**TABLA VI**
**COMPARACION DEL PROMEDIO ESTIMADO DEL ERROR CUADRATICO TOTAL Y SUS DESVIACIONES ESTANDAR ASOCIADO (en parentesis), PARA LA MATRIZ X2 (datos multicolineales), Y SEGÚN LAS CONFIGURACIONES DISEÑADAS.**

**$\phi = 0.0$**

| ESTIMADORES | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
|---|---|---|---|---|
| LS | 109.050 | 129.337 | 75.042 | 99.556 |
| | ( 32.937) | ( 125.926) | ( 21.392) | ( 4.512) |
| RR | 13.242 | 16.905 | 4.534 | 54.588 |
| | ( 3.644) | ( 25.331) | ( 1.226) | ( 3.308) |
| PC | 1.949 | 1.041 | 2.156 | 1.460 |
| | ( 1.521) | ( 0.296) | ( 0.554) | ( 0.280) |
| SE | 107.313 | 127.668 | 73.824 | 99.451 |
| | ( 32.942) | ( 125.818) | ( 21.236) | ( 4.506) |
| GR | 1.310 | 0.736 | 1.859 | 168.506 |
| | ( 0.543) | ( 0.213) | ( 0.570) | ( 499.433) |
| VS | 3.238 | 1.899 | 5.472 | 372.481 |
| | ( 1.921) | ( 0.420) | ( 0.661) | ( 11.962) |
| LR | 1.949 | 1.041 | 2.203 | 17093.617 |
| | ( 1.521) | ( 0.296) | ( 0.561) | ( 246.726) |

**$\phi = 0.5$**

| ESTIMADORES | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
|---|---|---|---|---|
| LS | 109.050 | 129.337 | 75.042 | 99.556 |
| | ( 32.937) | ( 125.926) | ( 21.392) | ( 4.512) |
| RR | 13.247 | 16.907 | 5.342 | 120.937 |
| | ( 3.639) | ( 25.454) | ( 0.480) | ( 15.579) |
| PC | 1.951 | 1.082 | 6.323 | 418.127 |
| | ( 1.521) | ( 0.296) | ( 0.554) | ( 0.280) |
| SE | 107.318 | 127.669 | 73.872 | 99.541 |
| | ( 32.936) | ( 125.812) | ( 21.275) | ( 4.470) |
| GR | 1.313 | 0.745 | 7.047 | 418.025 |
| | ( 0.540) | ( 0.205) | ( 0.846) | ( 0.265) |
| VS | 3.230 | 1.992 | 18.263 | 1783.154 |
| | ( 1.886) | ( 0.335) | ( 1.701) | ( 26.341) |
| LR | 1.951 | 1.085 | 6.801 | 15721.287 |
| | ( 1.521) | ( 0.301) | ( 0.570) | ( 4.435) |

**$\phi = 1.0$**

| ESTIMADORES | $\rho = 0.04$ | $\rho = 1.0$ | $\rho = 100$ | $\rho = 10000$ |
|---|---|---|---|---|
| LS | 109.050 | 129.337 | 75.042 | 99.556 |
| | ( 32.937) | ( 125.926) | ( 21.392) | ( 4.512) |
| RR | 13.259 | 16.943 | 11.191 | 346.659 |
| | ( 3.629) | ( 25.559) | ( 0.935) | ( 47.198) |
| PC | 1.956 | 1.207 | 18.823 | 1668.127 |
| | ( 1.521) | ( 0.296) | ( 0.554) | ( 0.280) |
| SE | 107.320 | 127.658 | 73.876 | 99.903 |
| | ( 32.937) | ( 125.809) | ( 21.290) | ( 4.440) |
| GR | 1.327 | 0.861 | 18.537 | 1668.062 |
| | ( 0.560) | ( 0.206) | ( 0.557) | ( 0.275) |
| VS | 3.231 | 1.967 | 15.108 | 1443.309 |
| | ( 1.893) | ( 0.350) | ( 1.496) | ( 23.686) |
| LR | 1.956 | 1.211 | 19.007 | 1674.048 |
| | ( 1.521) | ( 0.306) | ( 0.567) | ( 0.570) |

datos multicolineales (las ultimas 12 configuraciones, **TABLA VI**), dentro de este grupo, el ECM del estimador MCO es en general, mayor que el resto (particularmente es mayor en promedio 3 a 4 veces más que RG y RR), cuando $\rho$ = 0.04, 1.00 y para todas las orientaciones de $\phi$. Pero para $\rho$ 100, 10000 y todas las orientaciones de $\phi$, este estimador es superado por el estimador RG.

Para $\rho$ = 10000 y todas las orientaciones de $\phi$, los valores del ECM del estimador MCO, prácticamente iguala a los valores de los estimadores RR, CS, excepto a RG quien la supera. Por otro lado el ECM del estimador CS, en todo momento es inferior al valor de MCO, excepto cuando $\rho$= 10000, donde los dos estimadores tienen identicos ECM.

El desarrollo de RR es bastante bien, pues su ECM es bastante inferior a todos los estimadores, incluidos CS y MCO, para las primeras 12 configuraciones examinadas (**TABLA V**), y sólo es mayor por una unidad decimal a los valores de ECM de RG, cuando $\rho$ = 0.04, 1.00 y para todo valor de $\phi$.

El estimador RG produce un ECM bastante pequeño, conjuntamente con RR, para valores de la razón de la señal del ruido bajos, es decir para $\rho$ 0.04, 1.00 y para todas las orientaciones de $\phi$, siendo ligeramente inferior en una unidad decimal a RR; en cambio para valores mayores de $\rho$, sus valores superan a todos los estimadores en estudio. Particularmente para $\rho$ = 10000 y $\phi$ = 1.00 el ECM de RG es substancialmente mayor que el resto de los estimadores sesgados. De otro lado para $\rho$ =10000 y todas las orientaciones de $\phi$, los estimadores RR, CS y MCO son próximamente iguales.

Respecto a la desviación estándar del ECM del vector estimado $\beta$, se observa en general, sus valores bastante bajo, para todas las orientaciones de $\phi$ y $\rho$, para estas primeras 12 configuraciones (**TABLA V**), sobresaliendo mejor RR y RG cuando $\rho$ = 0.04, 1.00, 100; en cambio en las configuraciones $\rho$ = 10000, para todo valor de $\phi$, la mayoria tiene apróximadamente la misma variación, excepto para la variación de RG que es mayor al resto, cuando $\rho$ = 10000, $\phi$ = 1.00.

**Resultados para los últimos 12 configuraciones (datos multicolineales, vease TABLA VI, página 123).**

En general para estas configuraciones, el ECM del estimador MCO es sumamente grande y supera a los otros estimadores; excepto cuando la configuración tiene la razón de la señal del ruido $\rho$ = 10000 y para todas las orientaciones de $\phi$, siendo en ellas, algunas veces inferior a la mayoria de ellos. El estimador CS, tiene valores de ECM bastantes altos, pero ligeramente inferior a MCO (es menor en 1 o 2 unidades), para todas las configuraciones de datos multicolineales. Seguidamente el estimador RR se desempeña bastante bien con respecto a los estimadores CS y MCO, pues los valores correspondientes del ECM, son pequeños en comparación de los dos anteriores. Excepciones de este resultado ocurre cuando la configuración tiene $\rho$ = 10000 y $\phi$ = 0.5, 0.1, donde el ECM de RR supera a MCO y CS. Por otro lado el ECM de RR, se desempeña mejor y es el más bajo de todos los estimadores investigados, cuando $\rho$ = 100 y $\phi$ = 0.5, 1.0.

El estimador SV tiene ECM bastante pequeño en comparación a los estimadores MCO, CS y es menor que RR, para valores bajos de $\rho$ y todas las orientaciones de $\phi$. Pero para $\rho$ = 100 y todas las orientaciones de $\phi$, SV es mayor que RR, de otro modo cuando $\rho$ = 10000, para todo $\phi$, SV es mucho mayor que MCO, CS y RR.

El estimador RL se desempeña bastante bien, los valores del ECM es sumamente bajo en comparación de los estimadores MCO y CS cuando $\rho$ = 0.04, 1.00, 100 y para todo valor de $\phi$; también es menor que RR y ligeramente inferior a SV, cuando $\rho$ = 0.04, 1.00, para todas las orientaciones de $\phi$; así mismo sigue siendo ligeramente menor que RR y SV, cuando $\rho$ = 100 y $\phi$ = 0.00. Para $\rho$ = 100, $\phi$ = 0.5, el estimador RL es menor a SV, pero ligeramente superior a RR, de la misma manera en la configuración $\rho$ = 100 y $\phi$ = 1.0, nuevamente RR conjuntamente con SV tiene ECM inferior a RL. Para $\rho$ mayor, es decir, $\rho$ = 10000 y todo $\phi$, los valores del ECM de RL es enormemente mayor en comparación a todos los estimadores estudiados, incluido MCO.

Los valores del ECM del estimador CP, es enormemente pequeño en comparación a los estimadores MCO, CS, para todas las configuraciones, excepto, para $\rho$ = 10000 y $\phi$ = 0.5, 1.00, que los supera. También el ECM de CP, es menor a RR y SV, cuando $\rho$ = 0.04, 1.00 y todo valor de $\phi$, asimismo cuando $\rho$ = 100, 10000 y $\phi$ = 0.0. En esta última configuración, es decir cuando $\rho$ = 10000 y $\phi$ = 0.0, el ECM del estimador CP se desempeña mejor, pues llega a ser el más bajo con respecto a todos los estimadores investigados.

Por otro lado para estas últimas 12 configuraciones (para los datos múlticolineales, **TABLA VI**), los valores del ECM de los estimadores CP y RL son apróximadamente equivalentes (excepto para $\rho$ = 10000 y todo $\phi$). CP parece dar un resultado ligeramente inferior que RL en ECM, pero ambos resultan substancialmente pequeños en comparación a los estimadores MCO, CS; y es menor que RR, SV, especialmente cuando la razón de la señal del ruido $\rho$ es bajo. Para valores de $\rho$ altos, el ECM de CP supera a RR para la configuración $\rho$ = 100, $\phi$ = 0.5; a SV, RR para $\rho$ = 100, $\phi$ = 1.0; es bastante alto a MCO, CS, RR para $\rho$ = 10000, $\phi$ = 0.5; del mismo modo es bastante grande de MCO, CS, RR y es mayor a SV para $\rho$ = 10000, $\phi$ - 1.0, para esta última configuración, se observa que el ECM de CP, es el más alto conjuntamente con RG y RL.

El estimador RG tiene en esta simulación para los datos múlticolineales (TABLA VI), el ECM más pequeño que todos los estimadores en estudio, incluido MCO, cuando la razón de la señal del ruido $\rho$ = 0.04, 1.00, para todo valor $\phi$. Del mismo modo esta conclusión es cierta para $\rho$ = 100, $\phi$ = 0.0. Pero para $\rho$ = 100, $\phi$ = 0.5, 1.0, tiene un comportamiento más o menos parecido a CP y RL, siendo todos ellos mayor a RR. También cuando la configuración es $\rho$ = 10000, el ECM de RG es bastante alto a la mayoria, excepto a RL y SV, cuando $\phi$ = 0.0, 0.5 y cuando $\phi$ = 1.0, supera substancialmente a RR, CS, MCO, es menor a SV e igual a CP y RL. Notemos finalmente, que para todas las configuraciones de datos multicolineales (TABLA VI), excepto para $\rho$ =10000, $\phi$ = 0.0, el ECM de RG es ligeramente inferior en unas cuantas cifras decimales a CP. Por consiguiente para las configuraciones que tienen $\rho$ bajas estos estimadores en general se comportan bastante bien, y en general para todas las configuraciones, el estimador RR tiene potencial de mejora sobre CP y RG.

Respecto a la desviacion estándar del ECM, de todos los métodos (para datos multicolineales, TABLA VI), se observa, que las desviaciones de los estimadores MCO y CS, son bastantes altos para las configuraciones bajas; en cambio de RG, CP, RL, son las más pequeñas para estas mismas configuraciones, siendo mucho más, de RG. Para la configuracion alta, es decir, $\rho$= 10000, y todo valor de $\phi$, las desviaciones del ECM de MCO y CS, son algunas veces menores que sus similares, los estimadores sesgados restantes.

## 4.3.7. ALGUNOS RESULTADOS HALLADOS, SOBRE EL PROBLEMA DE MULTICOLINEALIDAD, POR lOS INVESTIGADORES.

* **Estudio de Simulación por Dempster, Schatzoff y Wermuth (1977).**

Ellos consideran, 57 variedades de estimadores sesgados, incluyendo MCO y varias versiones de cuatro tipos de estimadores sesgados como RR, CS, CP y SV. En ella se estudia el más grande conjunto de estimadores, pero curiosamente no incluyen a otros de interesante capacidad en el mejoramiento del ECM, como RL, CS. Comparan el desarrollo de los estimadores sesgados com MCO, usando 160 diferentes conjuntos de datos simulados de tamaño p = 6, n = 20.

El criterio de comparación empleado es

$$\text{Suma del error cuadrático de estimación} = \text{SECE} = \frac{(\beta-\beta)'(\beta-\beta)}{\sigma^2}$$

$$\text{Suma del error cuadrático de predicción} = \text{SECP} = \frac{(\beta-\beta)'X'X(\beta-\beta)}{\sigma^2}$$

y concluyen que el estimador RR, CP en ese orden da un mayor mejoramiento en SECE y SECP, especialmente para multicolinealidades fuertes, también cocluyen que ninguno de los estimadores sesgados domina en todos los casos, considerando que el mayor mejoramiento se da en SECE que para SECP, para los estimadores sesgados. Debemos anotar que la simulación que diseñarón, es diferente al que se presenta en este proyecto. Es decir la dependencia es variada especificando diferentes autovalores de la matriz de correlación.

* **Estudio de simulación por Lawless y Wang (1976)**

Estudian el estimador MCO, dos tipos de RR, dos versiones de CP, dos es estimadores naive de CS. Las comparaciones es hecha sobre la base del ECM y ECMP (error cuadrático medio de predicción). Con respecto a ambos criterios los autores concluyen que el estimador RR se desarrolla excelentemente mejor que los otros y que los estimadores RG y CS, no se desarrollan bien. También resumen que el estimador CP, no puede ser recomendado sobre la base de las propiedades del ECM. En este estudio el diseño de la simulación es llevado acabo variando la dependencia sobre la base de los autovalores de la matrices de correlación.

* **Estudio sobre datos reales de Hemmerle y Brantle (1978).**

Comparan tres tipos de estimadores RR, cuatro tipos de RG y MCO, sobre el criterio del ECM. Concluyen que el estimador RR, es mejor que MCO y que no es uniformemente mejor en todos los casos. El diseño del estudio se llevó a cabo variando la dependencia a través de las especificaciones de las correlaciones entre pares de variables.

* **Estudio sobre datos actuales de Gunst, Webster y Mason (1976).**

Consideran para el estudio los estimadoresd MCO y RL, sobre el criterio del ECM y Selección de Variables para datos multicolineales. Concluyen que RL es mejor que MCO, cuando la orientación del vector $\beta$ verdadero no es paralelo con el autovector correspondiente al autovalor más pequeño de la matriz de correlación $X'X$. Para el caso del paralelismo ambos métodos estiman pobremente el ECM. El diseño del estudio es llevado a cabo variando la dependencia de los datos, al hacer combinaciones lineales de las variables cuyo autovalor es próximo a cero.

# CONCLUSIONES Y RECOMENDACIONES

Despúes de llevarse a cabo esta trabajo de tesis sobre la "**EFICIENCIA DE LOS ESTIMADORES SESGADOS EN REGRESION, EN PRESENCIA DE MULTICOLINEALIDAD**", se ha llegado a conclusiones importantes y las más significativas se mencionan en las siguientes lineas.

## CONCLUSIONES

De la investigación se llegan a dos conclusiones principales.

1. El método usual de Mínimos Cuadrados Ordinarios (MCO), no es el más indicado, para estimar los parámetros verdaderos $\beta$´s en el modelo de regresión lineal, cuando en las variables independientes X´s, se declara la presencia de multicolinealidad, las razones son por que:
   * La estimación del vector verdadero $\beta$ por $\beta$, es impredecible, esto es, $\beta$ en **valor absoluto** es bastante grande (alejado de $\beta$).
   * La varianza y covarianza de $\beta$, es bastante grande e inestable.
   * El Error Cuadrático Medio Estimado (ECM) de $\beta$, es bastante grande.
2. Los estimadores sesgados **CP, RL, RR, RG, CS** y **SV**, en presencia de multicolinealidad, en general, estiman mejor al vector verdadero $\beta$ del modelo de regresión lineal que el método usual de MCO.

Otras conclusiones que contribuyen a solucionar el problema de multicolinealidad, se presentan a continuación:

- La varianza de todos los estimadores sesgados, se reducen substancialmente en comparación con el estimador MCO. Esto se observa tanto de las propiedades teóricas que se presenta en el capítulo III (para cada estimador sesgado), como en la práctica a través de los resultados de la simulación diseñada.

- El error cuadrático medio estimado (ECM), de todos los estimadores sesgados, presentan en la práctica, un mayor mejoramiento en comparación con MCO y dentro de ellos, los más eficientes son los estimadores RG, CP, RL, RR, en ese orden.
- El estimador RR en ECM, tiene potencial para mejorar, para lo cual, es necesario un mejor método de estimación del parametro K. Del mismo modo el estimador RL, tiene capacidad para mejorar, pero se necesita un mejor procedimiento de eliminación de términos y conocer sus propiedades distribucionales.
- El estimador sesgado CS, sobre la base del ECM, es útil para datos ortogonales, pero para datos multicolineales sólo muestra un ligero mejoramiento sobre **MCO** .
- Los estimadores sesgados, también muestran en la práctica, un mayor mejoramiento en ECM al de MCO para datos proximamente ortogonales, y dentro de ellos los más eficientes son los estimadores RR y RG.
- Teóricamente, los estimadores sesgados son potencialmente superiores al MCO respecto al criterio del ECM, pero en la práctica esto no es facilmente alcanzado en todos los casos o circunstancias.
- No existe un criterio optimo en general, para elejir un método de estimación en particular, como el mejor, para análizar un conjunto de datos especifico, en presencia de la multicolinealidad.
- Es posible uniformizar en una sola estructura general, a todos los estimadores estudiados, incluido MCO, con presencia de la multicolinealidad.

- En general, los estimadores sesgados desarrollan mejores pronósticos de la variable dependiente, en el modelo de regresión lineal general en presencia de multicolinealidad.

**RECOMENDACIONES:**

las siguientes sugerencias son para tener presente, cuando existe multicolinealidad:

- **Bajo ciertas circunstancias**, los estimadores sesgados son considerados mejores que el estimador usual MCO, es decir, tienen buen potencial en mejorar el ECM. Dentro de este dominio, también, son considerados los estimadores RR y RL pero se necesita un método de estimación mejor, para el parámetro K y un mecanismo de eliminación más adecuado de las variables independientes.
- Si se declara la presencia de multicolinelidad en las variables independientes, se debe usar, la estructura general hallada en el capítulo III, para los estimadores sesgados.
- Se pueden usar otros criterios de comparación para ver la eficiencia de los estimadores sesgados en un conjunto de datos con presencia de multicolinealidad, tales como Suma de Cuadrados de los Residuales (SCRes), Suma de Cuadrados de los Errores de Predicción (SCEP).
- Es posible diseñar otras formas de simulación para resolver el problema de multicolinealidad y este diseño debe estar en función de la variación de la estructura de dependencia de las variables regresoras, en el modelo de regresión lineal.

# BIBLIOGRAFIA

Se presenta la bibliografia siguiente, las cuales fuerón la fuente principal de información para realizar y culminar este proyecto de tesis.

- DIRECTED RIDGE REGRESION TECHNIQUES IN CASES OF MULTICOLLINEARITY, David K. Guilkey and James L. Murphy, Journal Statistical Association, December 1975, vol 70, No 352, pag. 769 - 775.
- A MONTE CARLO EVALUATION OF SOME RIDGE-TIPE ESTIMATORS, GaryC. McDonald and Diane I. Galarneau, Journal of the American Statistical Association, June 1975, vol 70, No 350 pag. 407 -416.
- MEAN SQUARED ERROR PROPERTIES OF GENERALIZED RIDGE ESTIMATORS, J. F. Lawless, Journal of the American Statistical Association, June 1981, vol 76, No 374 pag 462-466.
- SOME COMMENTS ON Cp, Technometrics, Nov. 1973, vol 15, No 4, pag,. 661 - 675.
- RIDGE REGRESION AND JAMES-STEIN ESTIMATION: REVIEW AND COMMENTS, Draper N. R. and Van Nostrand C., Technometrics 1979, vol 21, pag. 451 - 466.
- A SIMULATION STUDY OF SOME RIDGE ESTIMATORS, Diane Galarneau Gibbons, Journal of the American Statistical Asosociation, March 1981, vol 76, No 373, pag. 131 - 139.
- GOOD AND OPTIMAL RIDGE ESTIMATORS, R. L. Obenchain, The Annals of

Statistics, 1978, vol 6, No 5, pag. 1111 - 1121.

- A CRITIQUE OF SOME RIDGE RESGRESION METHODS, Gary Smith and Frank Campbell, Journal of the American Statistical Association, March 1980, vol 75, No 369, pag. 74 - 103.
- COMPARISON OF STOPPING RULES IN FORWARD STEPWISE REGRESION, Robert B. Bendel and A. A. Afifi, Journal of the American Statistical Association, March 1977, vol 72, No 357, pag. 46 - 53.
- THE ANALYSIS AND SELECTION OF VARIABLES IN LINEAR REGRESION, R. R. Hocking, Biometrics, March 1976, vol 32, pag. 1 - 49.
- ALTERNATIVES TO PLOTTING Cp IN MULTIPLE REGRESION, Emil SpjØtvoll, Biometrika, 1977, vol 64, No 1, pag. 1 - 8.
- BIASED REGRESSION ESTIMATORS: ANALYTIC Y SIMULATION RESULTS, J. Weber, B. Baldessari, D. Monarchi, International Journal of statistics, No 3-4, página 1-41.
- STEIN´S ESTIMATION RULE AND ITS COMPETITORS AN EMPIRICAL BAYES APPROACH, Bradley Efron and Carl Morris, Journal of the American Statistical Association, March 1973, vol 68, No 341, pag. 117 - 130.
- INTRODUCTION TO LINEAR REGRESSION ANALYSIS, Douglas C. Montgomery and Elizabeth A. Peck, Editorial John Wiley, New York, 1982.
- BIASED ESIMATION IN REGRESSION: AN EVALUATION USING MEAN SQUARED ERROR, Richard F. Gunst and Robert L. Mason, Journal of the American Statistical association, September 1977, vol 72, No 359, pag. 616 - 628.
- RIDGE REGRESSION: BIASED ESTIMATION FOR NONORTHOGONAL PROBLEMS, Arthur E. Hoerl and Robert W. Kennard, Technometrics, February 1970, vol 12, No 1, pag. 55 - 67.

MINIMUM VARIANCE PROPERTIES OF PRINCIPAL COMPONENT

REGRESSION,Edward Greenberg, Journal of the American Statistical Association, March 1975, vol 70, No 349, pag. 194 - 197.

- AN ANALYTIC VARIABLE SELECTION TECHNIQUE FOR PRINCIPAL COMPONENT REGRESSION, E. R. Mansfield, J.T.Webster and R. F. Gunst, vol 26, No 1, pag. 43 - 40.
- AN APPROACH TO THE PROGRAMMIMG OF BIASED REGRESSION ALGORITHMS, Richard F.Gunst, Commun Statist Simula Computa, vol 2, pag. 151 - 159.
- LATENT ROOT REGRESSION ANALYSIS, J. T. Webster, R. F. Gunst and R. L. Mason, Technometrics, vol 16, No 4, pag. 513 - 522.
- A SIMULATION STUDY OF ALTERNATIVES TO ORDINARY LEAST SQUARES, A. P. Dempster, Martin Schatzoff, and Nanny Wermuth, Journal of the American Statistical Association, march 1977, vol 72, No 357, pag. 77 - 106.
- EXPLICIT AND CONTRAINED GENERALIZED RIDGE ESTIMATION, W. J. Hemmerle and T.F.Brantle, Technometrics, vol. 20, No 2, May 1978.
- AN EXPLICIT SOLUTION FOR GENERALIZED RIDGE REGRESION, W.J.Hemmerle, Technometrics, vol. 17, No 3, August 1975.
- THE MINIMUM MEAN SQUARE ERROR LINEAR ESTIMATOR AND RIDGE REGRESSION, Farabrother, R.W (1975), Technometrics, Vol.17, No 1, página 127 - 128.

# A P E N D I C E   I

## DEMOSTRACIONES DE LAS NOTAS DE PIE

Los elementos diagonales de la Matriz $(X'X)^{-1}$ son:

$$C_{jj} = \frac{1}{(1 - R^2)} \quad , \; j = 1,2,\ldots,p$$

Prueba

Sabemos que

$$\text{Var} - \text{Cov}(\beta) = \sigma^2(X'X)^{-1} = \sigma^2 C_{p\times p}$$

Luego particionando la matriz de observación estandarizada $X$ es

$$X = [\; X_j \quad X_j^+ \;]$$

donde

$X_j$ es la columna de observación de la j-ésima variable independiente.

$X_j^+$ Es la matriz de observaciones del resto (p - 1) variables independientes.

entonces

$$(X'X)^{-1}_{p\times p} - \left( [\; X_j : X_j^+ \;]'[\; X_j : X_j^+ \;] \right)^{-1}$$

$$= \begin{bmatrix} X_j' X_j & X_j' X_j^+ \\ X_j^{+\prime} X_j & X_j^{+\prime} X_j^+ \end{bmatrix}^{-1}_{p\times p}$$

Por consiguiente para calcular la inversa de esta última matriz para aplicar el método de "INVERSION DE MATRICES POR PARTICION", además como se obeserva, la matriz ya está particionada en cuatro bloques.

$$(X'_jX_j)_{1\times1}\ ,\ (X^{+\prime}_jX^+_j)_{(p-1)(p-1)}\ ,\ (X'_jX^+_j)_{1x(p-1)}\ ,\ (X^{+\prime}_jX_j)_{(p-1)x1}$$

Por tanto, se tiene aplicando el método

$$(X'X)^{-1}_{p\times p}=C_{p\times p}$$

$$=\begin{bmatrix} [X'_jX_j-X'_jX^+_j(X^{+\prime}_jX^+_j)^{-1}X^{+\prime}_jX_j]^{-1}_{1x1} & (X^{+\prime}_jX_j)^{-1}-X'_jX^+(X^{+\prime}_jX^+_j)^{-1}_{1x(p-1)} \\ (X^{+\prime}_jX^+_j)^{-1}X^{+\prime}_jX_j-(X'_jX^+_j)^{-1}_{(p-1)x1} & (X^{+\prime}_jX^+_j)^{-1}(X^{+\prime}_jX_j)(X'_jX^+_j)(X^{+\prime}_jX^+_j)^{-1}_{(p-1)(p-1)} \end{bmatrix}_{p\times p}$$

entonces

$$C_{jj}=[\ X'_jX_j-X'_jX^+_j(X^{+\prime}_jX^+_j)^{-1}X^{+\prime}_jX_j\ ]^{-1} \qquad (*)$$

Además

$$X'_jX_j=1 \qquad (**)$$

(Columnas de la matriz X estandarizada).

y como el coeficiente de correlación múltiple es definida por

$$R^2=\frac{\beta'X'Y}{Y'Y}$$

Entonces similarmente $R^2_j$ será el coeficiente de correlación múltiple de la j-ésima variable independiente cuando es regresado sobre el resto (p-1) variables independientes, es decir:

$$R^2_j=\frac{\hat{\beta}_jX^{+\prime}_jX_j}{X'_jX_j}=X'_jX^+_j(X^{+\prime}_jX^+_j)^{-1}X^{+\prime}_jX_j \qquad (***)$$

Por lo tanto de (*), (**) y (***) se tiene :

$$C_{jj}=[\ 1-R^2_j\ ]^{-1},\quad \forall\ j=1,2,\ldots,p$$

l.q.q.d

$\boxed{\text{B}}$

$Tr(X'X)^{-1} = \sum_{j=1} \left[ \frac{1}{\lambda_j} \right]$ , donde $\lambda_j$, $j = 1,2,\ldots,p$ son los valores propios de $(X'X)$

<u>Prueba</u>

Por propiedad del álgebra matricial; como $(X'X)_{p\times p}$ tiene "p" valores propios distintos, entonces existe una matriz cuadrada $T$, de orden $p\times p$ no singular tal que se cumple:

$$T^{-1}(X'X)T = \Lambda \tag{B.1}$$

donde $\Lambda$ es una matriz diagonal de orden $p\times p$, cuyos elementos $\lambda_j$, $j= 1,2,\ldots,p$ son los valores propios de $(X'X)_{p\times p}$

Por consiguiente de (B.1)

$$T^{-1}(X'X)^{-1}T = \Lambda^{-1}$$

Entonces

$$Tr\left[ T^{-1}(X'X)^{-1}T \right] = Tr\left( \Lambda^{-1} \right)$$

Usando propiedades de la traza, se tiene

$$Tr\left[ TT^{-1}(X'X)^{-1} \right] = tr\left( \Lambda^{-1} \right)$$

Finalmente

$$Tr\,(X'X)^{-1} = \sum_{j=1} \left[ \frac{1}{\lambda_j} \right] \qquad \text{l.q.q.d}$$

$\boxed{\text{C}}$

$$\sum_{i=1}^{n} \left[ Y_j - \hat{Y}_{ij} \right]^2 = \left[ Y - \hat{Y}_j \right]' \left[ Y - \hat{Y}_j \right] = \eta^2 \, \frac{\lambda_j}{\gamma^2_{0j}} \tag{C.1}$$

<u>Prueba</u>

Se sabe $\qquad Y_i^* = \dfrac{(Y_i - \bar{Y})}{\qquad}$

Entonces

$$\bar{Y} = Y_i - \eta Y_i^*$$

Luego, reemplazando esta expresión en (3.3.4)

$$\hat{Y}_j = \left[Y_i - \eta Y_i^*\right] 1 - \eta \gamma_{0j}^{-1} X\delta_j, \qquad j = 0,1,2,\ldots,p$$

Por consiguiente se tiene en (C.1)

$$\sum_{i=1}^{n}\left[ Y_i - Y_i - \eta Y_i^* + \eta\gamma_{0j}^{-1} X\delta_j \right]^2 = \eta^2\gamma_{0j}^{-2} \sum_{i=1}^{n}\left[ Y_i^*\gamma_{0j} + X\delta_j \right]^2$$

$$= \eta^2\gamma_{0j}^{-2} \sum_{i=1}^{n}\left[Y_i^*\gamma_{0j} + \sum_{r=1}^{p} X_{ir}\gamma_{rj}\right]^2 \qquad j = 0,1,2,\ldots,p$$

$$= \eta^2 \frac{\lambda_j}{\gamma_{0j}^2} \qquad j = 0,1,2,..,p \qquad \text{l.q.q.d}$$

$\boxed{\text{D}}$

Sean $\xi_i(W)$ y $\xi_i(Z)$ los autovalores de las matrices $W$ y $Z$ respectivamente, entonces

$$\xi_i(W) = \frac{1}{\lambda + k} \qquad\qquad \xi_i(Z) = \frac{\lambda_i}{\lambda_i + k}$$

Donde $\qquad W = (X'X + kI)^{-1} \qquad Z = (I + k(X'X)^{-1})^{-1}$

$\lambda_i$ , $i = 1,2,\ldots,p$ son los autovalores de $X'X$

<u>Prueba</u>

De la definición de ecuación caracteristica se tiene:

$$W - \xi I \mid = 0$$

$$(X'X + kI)^{-1} - \xi I \mid = 0$$

$$\mid (X'X)^{-1} + k^{-1}I - \xi I \mid = 0 \qquad (D.1)$$

Ahora, como $T'(X'X)T = \Lambda$, donde $\Lambda$ es la matriz diagonal cuyos elementos son los autovalores de $(X'X)$ y $T$ es la matriz ortogonal cuyas columnas son los autovectores de $X'X$, entonces en (D.1).

$$T\Lambda^{-1}T' + k^{-1}TIT' - \xi TIT' = 0$$

$$T \mid\mid \Lambda^{-1} + k^{-1}I - \xi I \mid\mid T' \mid = 0$$

$$T \mid\mid T' \mid\mid \Lambda^{-1} + k^{-1}I - \xi I \mid = 0$$

$$|\,T\,T'\,|\,|\,\Lambda^{-1} + k^{-1}I - \xi I\,| = 0$$

$$|\,\Lambda^{-1} + k^{-1}I - \xi I\,| = 0 \Longrightarrow \xi_i(W) = \frac{1}{\lambda_i + k}\ ,\quad i = 1,2,\ldots,p$$

Similarmente para Z

$$|\,Z - \xi I\,| = 0$$

$$|\,(I + k(X'X)^{-1})^{-1} - \xi I\,| = 0$$

$$[(X'X)^{-1}(X'X) + k(X'X)^{-1}]^{-1} - \xi I\,| = 0$$

$$(X'X)[\,X'X + kI\,]^{-1} - \xi I\,| = 0$$

$$|\,T\Lambda T'W - \xi TIT'\,| = 0$$

$$|\,T\,|\,|\,\Lambda W - \xi I\,|\,|\,T'\,| = 0$$

$$TT'\,|\,|\,\Lambda W - \xi I\,| = 0$$

$$|\,\Lambda W - \xi I\,| = 0$$

Por lo anterior

$$|\,\Lambda(\Lambda^{-1} + k^{-1}I) - \xi I\,| = 0 \Longrightarrow \xi_i(Z) = \frac{\lambda_i}{\lambda_i + k}\ ,\quad i = 1,2,..p \qquad \text{Lqqd.}$$

E

**TEOREMA 1**

La varianza total $\gamma_1(k)$ es una función continua y monótonamente decreciente de **k** .

Prueba

Sea
$$\gamma_1(k) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} \qquad k > 0$$

Derivando con respecto a **k** , tenemos

$$\gamma_1'(k) = -2\sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^3} \qquad \text{(E.1)}$$

Luego $\gamma_1(k)$ es diferenciable para todo $k > 0$, por consiguiente es continua $\forall\ k > 0$ (Por teorema de Análisis Matemático), además:

$$\gamma_1'(k) \leq 0\ ,\quad \forall\ k > 0$$

Por lo tanto, es monótonamente decreciente. Lqqd.

**TEOREMA 2**

El sesgo al cuadrado $\gamma_2(k)$ es una función continua monótonamente creciente de **k**

Prueba

Similarmente como el **TEOREMA 1**, sea:

$$\gamma_2(k) = k^2\beta'(X'X + kI)^{-2}\beta$$

Si $\Lambda$ y T son las matrices de autovalores y autovectores de $X'X$ respectivamente, entonces:

$$X'X = T'\Lambda T$$

Luego:

$$\gamma_2(k) = k^2 \sum_{i=1}^{p} \frac{\alpha_i}{(\lambda + k)^2} \qquad \gamma_2(k) = \sum_{i=1}^{p} \frac{\alpha_i^2}{(1 + (\frac{\lambda_i}{k}))^2} \qquad (E.2)$$

Donde, $\alpha = T\beta$ , Derivando (E.2) con respecto a **k**, se tiene

$$\gamma_2'(k) = 2\sum_{i=1}^{p} \frac{\alpha_i^2\lambda_i}{(k^{2/3} + \lambda_i k^{-1/3})^3}$$

Luego $\gamma_2'(k)$ existe $\forall\ k > 0$. Por consiguiente $\gamma_2(k)$ es diferenciable $\forall\ k>0$, entonces es continua $\forall\ k>0$.

Además : $\gamma_2'(k) \geq 0 \ , \ \forall\ k > 0$

Por lo tanto es monótonamente creciente. Lqqd.

**COROLARIO 2**

$\gamma_2(k)$ se apróxima a $\beta'\beta$ como un límite superior.

Prueba

De (E.2) en el TEOREMA 2 se tiene

$$\lim_{k \to \infty} \sum_{i=1}^{p} \frac{\alpha_i}{(1 + (\frac{\lambda_i}{k}))^2} = \sum_{i=1}^{p} \alpha_i^2$$

$$= \alpha' \alpha$$

$$= \beta' T' T \beta$$

$$= \beta' \beta$$ Lqqd.

**COROLARIO 3**

$$\frac{d\, \gamma_2(k)}{dk} \longrightarrow \emptyset \text{ , cuando , } k \longrightarrow \emptyset^+$$

Prueba

Se sabe que la derivada de $\gamma_2(k)$ es:

$$\gamma_2'(k) = 2k \sum_{i=1}^{p} \frac{\alpha_i^2 \lambda_i}{(\lambda_i + k)^3} \qquad (E.3)$$

$$\lim_{k \to 0^+} \gamma_2'(k) - \lim_{k \to 0^+} 2k \sum_{i=1}^{p} \frac{\alpha_i^2 \lambda_i}{(\lambda_i + k)^3} = \emptyset$$ lqqd.

**TEOREMA 3**

Siempre existe un $k>\emptyset$ tal que:

$$E[\ L_1^2(k)\ ] < E[\ L_1^2(\emptyset)\ ] = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}$$

Prueba

Como $$E[\ L_1^2(k)\ ] = \gamma_1(k) + \gamma_2(k)$$

Luego su derivada teniendo en cuenta (E.1) y (E.3) será

$$\frac{d\, E[\ L_1^2(k)\ ]}{dk} = -2\sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum_{i=1}^{p} \frac{\alpha_i^2 \lambda_i}{(\lambda_i + k)^3} \qquad (E.4)$$

Notece que $\gamma_1(\emptyset) = \sigma^2 \sum_{i=1}^{p} \left[ \frac{1}{\lambda_i} \right]$ y $\gamma_2(\emptyset) = \emptyset$

En el TEOREMA 1 y TEOREMA 2 $\gamma_1(k)$ y $\gamma_2(k)$ son monótonamente decreciente y creciente respectivamente, sus primeras derivadas son siempre no positivas y no negativas respectivamente. Luego para probar el teorema solamente es necesario mostrar que siempre existe un $k>0$ tal que

$$\frac{d\,E[\,L_1^2(k)\,]}{dk} < 0.$$

Luego la condición para esto es mostrada por (E.4) por ser:

$$k < \frac{\sigma^2}{\alpha_{max}} \qquad \text{Lqqd.}$$

F

El "CRITERIO ITERATIVO" de HOYRL and KENNARD (1970) es reducido a la formula simple matricial:

$$E_{j+1} - E_0(I + E_j)^2$$

Prueba

Por conveniencia representemos los p-vectores $(X^*)'Y$ y $\alpha_{rg(j)}$ como matrices diagonales, en este contexto sea

$$B = \begin{bmatrix} [(X^*)'Y]_1 & 0 & \cdots & 0 \\ 0 & [(X^*)'Y]_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [(X^*)'Y]_p \end{bmatrix}$$

$$A_j = \begin{bmatrix} \hat{\alpha}_{rg}^{1(j)} & 0 & \cdots & 0 \\ 0 & \hat{\alpha}_{rg}^{2(j)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\alpha}_{rg}^{p(j)} \end{bmatrix}$$

Como consecuencia se obtiene:

$$A_o = \Lambda^{-1}B$$

Donde $\Lambda$ es conocido como la matriz diagonal de raices características de $X'X$.

Seguidamente el criterio A llamado "criterio iterativo" puede ser descrito por la formula matricial siguiente:

$$A_{j+1} = [\ \Lambda + \hat{\sigma}^2 A_j^{-2}\ ]^{-1}B \quad ó$$

$$A_{j+1} = [\ \Lambda + \hat{\sigma}^2 A_j^{-2}\ ]^{-1}\Lambda A_o$$

también, esta última expresión es escrito como:

$$A_{j+1} = [\ \Lambda(\ I + \hat{\sigma}^2\Lambda^{-1}A_j^{-2}\ )]^{-1}\Lambda A_o$$

Entonces:

$$A_{j+1} = (I + \hat{\sigma}^2\Lambda^{-1}A_j^{-2}\ )^{-1}A_o$$

Si se hace:

$$D = \frac{\Lambda}{\hat{\sigma}^2}$$

se obtiene:

$$A_{j+1} = [\ I + D^{-1}A_j^{-2}\ ]^{-1}A_o$$

Luego, una expresión para $A_{j+1}^{-2}$ es dado por:

$$A_{j+1}^{-2} = A_o^{-1}(I + D^{-1}A_j^{-2})A_o^{-1}(I + D^{-1}A_j^{-2})$$

Sin embargo, las matrices $A_o^{-1}$ y $(I + D^{-1}A_j^{-2})$ de esta expresión son diagonales y conmutativas, por consiguiente se escribirá:

$$A_{j+1}^{-2} = A_o^{-2}(I + D^{-1}A_j^{-2})^2$$

Ahora, si se multiplica a ambos lados de esta expresión por $D^{-1}$ se obtiene:

$$D^{-1}A_{j+1}^{-2} = D^{-1}A_o^{-2}(I + D^{-1}A_j^{-2})^2$$

Por tanto, si se hace:

$$E_j = D^{-1}A^{-2}$$

El "criterio iterativo" es reducido a la fórmula simple:

$$E_{j+1} = E_O(I + E_j)^2 \qquad \text{lqqd.}$$

$\boxed{G}$

$$E[\ L_1^2\ ] = E[\ \alpha_{RG} - \alpha]'[\ \alpha_{RG} - \alpha\ ] + \sigma^2 Tr[\ (\Lambda - \mathbb{K})\Lambda^{-1}(\Lambda + \mathbb{K})^{-1}]$$

Prueba

Sea

$$E[\ L_1^2\ ] = E[\ \alpha_{RG} - \alpha]'[\ \alpha_{RG} - \alpha\ ]$$

Luego, haciendo algunas simplificaciones:

$$E(\ \alpha_{RG} - \alpha)'(\ \alpha_{RG} - \alpha\ ) = \sum_{j=1}^{p} \frac{(\lambda_j \sigma^2 + K_j^2 \alpha_j^2)}{(\lambda_j + K_j)^3}$$

Similarmente:

$$E(\ \alpha_{RG} - \alpha)'(\ \alpha_{RG} - \alpha\ ) = \sum_{j=1}^{p} \frac{K_j^2\left(\dfrac{\sigma^2}{\lambda_j} + \alpha_j^2\right)}{(\lambda_j + K_j)^3}$$

Así que:

$$E(\alpha_{RG} - \alpha)'(\alpha_{RG} - \alpha) = E(\alpha_{RG} - \alpha)'(\alpha_{RG} - \hat{\alpha}) + \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j - K_j}{\lambda_j(\lambda_j + K_j)}$$

Consecuentemente:

$$E[\ L_1^2\ ] = E[\ \alpha_{RG} - \alpha]'[\ \alpha_{RG} - \alpha\ ] + \sigma^2 Tr[\ (\Lambda - \mathbb{K})\Lambda^{-1}(\Lambda + \mathbb{K})^{-1}] \qquad \text{lqqd}$$

$\boxed{H}$

$$E[\ L_2^2\ ] = E[\ \hat{\alpha}_{RG} - \alpha]'\Lambda[\ \hat{\alpha}_{RG} - \alpha\ ] + \sigma^2 Tr[\ (\Lambda - \mathbb{K})(\Lambda + \mathbb{K})^{-1}]$$

Prueba

Se demuestra en forma similar a $E[\ L_1^2\ ]$ en $\boxed{G}$

I

$$C_q = \frac{SCRes_q}{\hat{\sigma}^2} + 2q - n \qquad (I.1)$$

Donde:

$\hat{\sigma}^2 = SCRes_p /(n - p)$ es la estimación de $\sigma^2$.

Prueba

El estadístico (I.1) se deriva de la siguiente forma:

Sea $Y_q$, el valor predecido de la variable respuesta para una entrada particular tal como se ha definido en la **PROPIEDAD 8** (vease página 87), luego el error cuadrático medio del i-ésimo valor predecido $Y_{q(i)}$ para una i-ésima entrada particular es

$$ECM(Y_{q(i)}) = E\,[\,Y_{q(i)} - E(Y_i)]$$

$$= [\,E(Y_i) - E(\hat{Y}_{q(i)})]^2 + Var(\hat{Y}_{q(i)}) \qquad (I.2)$$

por tanto:

$$[\,E(Y_i) - E(Y_{q(i)})]$$

Es el sesgo del i-ésimo valor predecido .

Si (I.2) es sumado sobre un conjunto específicado de "n" entradas y dividiendo por $\sigma^2$ se tendrá :

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} E[\,\hat{Y}_{q(i)} - E(Y_i)]^2 = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n} E[\,\hat{Y}_i - E(\hat{Y}_{q(i)})]^2 + \sum_{i=1}^{n} var(Y_{q(i)})\right\}$$

(I.3)

Denotando:

$$\Gamma_q = \frac{1}{\sigma^2}\sum_{i=1}^{n} E[\,\hat{Y}_{q(i)} - E(Y_i)]^2 \qquad y \qquad SSesCT_q = \sum_{i=1}^{n} E[\,\hat{Y}_i \quad E(\hat{Y}_{q(i)})]^2$$

Donde:

$\Gamma_q$ = Error cuadrático medio total estandarizado para "q" variables

$SSesCT_q$ = Suma sesgo cuadrático total para "q" variables.

Luego de (I.3):

$$\Gamma_q = \frac{SSesCT_q}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n} var(\gamma_{q(i)}) \qquad (I.4)$$

como:

$$E(SCRes_q) = SSesCT_q + (n - q)\sigma^2 \qquad (I.5)$$

y además

$$\sum_{i=1} var(\gamma_{q(i)}) = q\sigma^2 \qquad (I.6)$$

Por (I.5) y (I.6)

$$\Gamma_q = \frac{1}{\sigma^2}\left\{E(SCRes_q) - (n - q)\sigma^2 + q\sigma^2\right\}$$

$$\Gamma = \frac{E(SCRes_q)}{\sigma^2} - n + 2q \qquad (I.7)$$

Luego una estimación de (I.7) es:

$$C_q = \frac{SCRes_q}{\sigma^2} - n + 2q \qquad (I.8)$$

Lqqd.

Ahora si no hay sesgo en (I.5) entonces:

$$E(SCRes_q) = (n - q)\sigma^2$$

Luego:

$$E\left[C_q \;/\; SSesCT_q = 0\right] = \frac{(n - q)\sigma^2}{\sigma^2} - n + 2q$$

$$E\left[C_q \;/\; SSesCT_q = 0\right] = q \qquad (I.9)$$