COMUNICACIONES DEL CIMAT

1

1

EL PROBLEMA DE LA COLINEALIDAD EN LA REGRESION LINEAL

ERNESTO BARRIOS ZAMUDIO

CENTRO DE INVESTIGACION EN MATEMATICAS

Apartado Postal 402
Guanajuato, Gto.
México
Tels. (473) 2-25-50
2-02-58

EL PROBLEMA DE LA COLINEALIDAD EN LA REGRESION LINEAL

Por Ernesto Barrios Z.

RESUMEN

Se hace una revisión amplia del tema de la Colinealidad en la Regresión Lineal. Aunque éste es un problema netamente numérico debido a la "mala calidad" de los datos, los estadísticos están involucrados, pues la calidad de los estimadores se ve afectada. Se muestran los efectos que provoca la presencia de la colinealidad y sus causas más frecuentes. Se presentan varios procedimientos utilizados para su diagnóstico, algunos de ellos apoyados en herramientas desarrolladas en el terreno del Análisis Numérico como puede ser la Descomposición en Valores Singulares. Así también se dice cómo poder remediar el problema o en su defecto se recomienda como alternativa el uso de estimadores sesgados o del tipo bayesiano. Finalmente, se propone una estrategia para atacar el problema.

I N D I C E

1 INTRODUCCION 1

2 EFECTOS DE LA COLINEALIDAD 5

3 FUENTES DE LA COLINEALIDAD 8

4 DIAGNOSTICO DE LA COLINEALIDAD 10

5 CORRECCION Y ALTERNATIVA 15

6 CONCLUSIONES 21

7 APENDICE 23

REFERENCIAS 27

1. INTRODUCCION

Considérese el modelo de regresión lineal múltiple

$$Y_i = \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + u_i \qquad i = 1,\ldots,n \tag{1}$$

o, en su forma matricial,

$$Y = X\beta + u$$

donde Y_{nx1}, X_{nxp}, β_{px1}, u_{nx1} y con los supuestos[1]

$$\begin{aligned} &\text{a) } E(u) = \theta \quad \text{(vector de ceros)} \\ &\text{b) } Cov(u) = \sigma^2 I \\ &\text{c) } X \text{ no estocástica, de rango p.} \end{aligned} \tag{2}$$

El estimador de Mínimos Cuadrados (MC) $\hat{\beta}$ del vector de parámetros β es aquel que minimiza la suma de cuadrados

$$S(\beta) = \sum_{i=1}^{n} (Y_i - \beta_1 X_{i1} - \ldots - \beta_p X_{ip})^2.$$

Una forma de obtener $\hat{\beta}$ es derivando la suma de cuadrados[2]

$$S(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)'(Y - X\beta)$$

respecto a β, igualar a θ y resolver las ecuaciones normales

$$X'X\beta = X'Y \tag{3}$$

cuya solución, si $(X'X)$ es no singular está dada por

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

La existencia y unicidad del estimador $\hat{\beta}$ está garantizada por el supuesto 2c) y en tal caso, en teoría, puede realizarse sin problema el análisis completo de la regresión.

Sin embargo, en la práctica, y a raíz del uso de la computadora en el análisis de regresión, los modelos han sufrido la inclusión indiscriminada de regresores, lo que se refleja frecuentemente en que

---

[1] Más adelante se discute el supuesto 2c).

[2] En el trabajo $\|.\|$ denotará la norma 2, o norma euclidiana ($\|x\|_2 = \|x\| = [\sum x_i^2]^{1/2}$), a menos que se indique lo contrario.

la matriz de diseño X es de rango seudodeficiente[3], por lo tanto X'X resulta casi singular y en tal caso la computadora tendrá problemas en el cálculo de la solución del sistema (3). El problema de la colinealidad en la regresión es precisamente de carácter computacional.

El que X sea de rango seudodeficiente, como se verá más adelante, puede tener consecuencias graves sobre la calidad de los estimadores $\hat{\beta}$, por lo que este hecho se identifica como el *problema de la colinealidad en la regresión*. Si X es de rango incompleto se dice que hay *colinealidad perfecta*. En Econometría el problema es común y se le conoce desde hace tiempo como *multicolinealidad*, pero a la luz de la nota 3, necesariamente se requerirá de al menos dos variables colineales para que se presente el problema, por lo que se le referirá simplemente como colinealidad. Por otro lado, los analistas numéricos identifican la colinealidad simplemente con el problema que se presenta al querer resolver las ecuaciones normales, y ésto es un reflejo del *mal condicionamiento* de la matriz X[4]. Así, la presencia de colinealidad, es el problema de multicolinealidad para los econometristas y el mal comportamiento de la matriz X para los analistas numéricos.

Para el modelo de Regresión Lineal Múltiple (1), autores como Freund y Milton (1979), afirman explícitamente que el rango de la matriz X no es un supuesto de la regresión, mientras que otros como Intriligator (1978) sí incluyen como supuesto el punto 2c, esto es, que X sea de rango completo. Considero que esto está ligado con lo que se entiende por modelo y esto dependerá del objetivo de la regresión. Si el interés del analista es observar el efecto que tienen las variables independientes $X_1, \ldots X_p$ sobre la respuesta, en base a la matriz de diseño X, ciertamente el rango no debe ser un supuesto; por otro lado, si el objetivo es hacer pronóstico de la

[3]La matriz X se dice de rango *seudodeficiente* si existen constantes $t_1, \ldots, t_p$, no todas cercanas a cero, tales que $\sum t_i X_i \approx \theta$, donde $X = (X_1, \ldots, X_p)$, $X_{i\ nx1}$, $i = 1, \ldots, p$.

[4]Seudodeficiencia, mal condicionamiento de una matriz y otros conceptos del análisis numérico, se presentan en el apéndice.

variable explicada y el rango de X es deficiente, se puede entonces trabajar con un modelo reducido, de modo que la matriz de diseño sea de rango completo. Así también, Farrar y Glauber (1967) mencionan que la colinealidad "es una faceta y un síntoma de un diseño experimental pobre", pero no siempre es un diseño lo que da lugar a la regresión (tal es el caso más común de la regresión en Econometría), mientras que por otro lado, Chatterjee y Price (1977) comentan que la colinealidad no es un problema de modelamiento, sino de deficiencia en los datos, o bien como indica Baz (1976), "es un problema de falta de información".

De manera similar, los autores no se ponen de acuerdo sobre si el problema de colinealidad es o no estadístico. Así, por ejemplo, Besley, Kuh y Welsch (1980) indican que no lo es, pues no invalida el análisis de regresión; sin embargo, considerando las fuentes más frecuentes de colinealidad que enuncian Montgomery y Peck (1982) y sus posibles remedios, bien puede considerarse un problema estadístico, pues aunque es factible llevar a cabo la estimación y el análisis bajo colinealidad, en la práctica se pueden obtener resultados inservibles. En lo que sí hay consenso es en que la colinealidad es un problema grave del Análisis de Regresión Aplicado que merece ser estudiado[5].

Como algunos de los objetivos de la regresión, Montgomery y Peck (1982) mencionan los siguientes:

- Identificar los efectos marginales de las variables explicativas X_i,
- La estimación y predicción de la variable respuesta Y,
- La selección apropiada de las variables en el modelo.

Estos no serán fáciles de alcanzar en presencia de colinealidad pues al separarse los regresores de la ortogonalidad las variables comienzan a compartir información, lo que vuelve difícil apreciar el

[5]Un buen ejemplo es el libro de Belsley, Kuh y Welsch (1980), "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity"

efecto particular de los regresores[6]. Además, la estimación misma de los parámetros y por lo tanto de la variable respuesta será altamente sensible al mal condicionamiento de X, lo que constituye una fuente de imprecisión. En cuanto al pronóstico de la variable dependiente no habrá problema siempre y cuando el nivel de los predictores x mantenga la estructura de la matriz de diseño. Finalmente, los métodos comunes de selección de variables se verán oscurecidos en presencia de colinealidad, por lo que no será posible aplicarlos con éxito.

Una manera frecuente de trabajar en Econometría es la siguiente: Se postula el modelo y sus supuestos, se procede a la estimación de los parámetros necesarios, para después llevar a cabo la validación del modelo mediante la verificación de los supuestos, y en caso que éstos no se satisfagan, se intenta su corrección, siempre que ésta sea posible. Así por ejemplo, el supuesto 2b) considera la ausencia de correlación serial de primer orden (cov $(u_i, u_j) = 0$, $i \neq j$), este supuesto se valida mediante el estadístico d de Durbin-Watson y en caso de no cumplirse se hace la corrección de Cochrane - Orcutt. De igual forma, 2b) supone la homoscedasticidad en los errores (var $(u_i) = \sigma^2$, $i=1,\ldots,n$) y mediante la prueba de Bartlett o de White puede verificarse el supuesto; en caso de no cumplirse, pueden aplicarse, las transformaciones de variables del tipo Box y Cox, entre otras.

Con esta idea se ha estructurado el presente trabajo; en primer lugar, se muestran los efectos de la colinealidad con la intención de convencer de la necesidad de su estudio. En la siguiente sección se presentan las posibles fuentes de la colinealidad, para después, exponer algunos métodos de diagnóstico que identifiquen la presencia de colinealidad. La sección 5 trata de la corrección de la colinealidad (cuando es posible corregirla) y presenta estimadores distintos a los de Mínimos Cuadrados como alternativa. Finalmente, en la sección 6 se dan algunas conclusiones.

---

[6]Dos regresores X_j, X_k se dicen *ortogonales* si $\sum_{i=1}^{n} X_{ij} X_{ik} = 0$. El Diseño de Experimentos explota la ventaja de trabajar con este tipo de variables.

2. EFECTOS DE LA COLINEALIDAD

Considérese el modelo (1),

$$Y_i = \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + u_i \qquad i = 1, \ldots, n;$$

si se transforma el modelo a[7]

$$y_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + u_i \qquad i = 1, \ldots, n, \tag{4}$$

donde $y_i = \dfrac{Y_i - \bar{Y}}{\left[\sum_{k=1}^{n} (Y_k - \bar{Y})^2\right]^{1/2}}$, $\quad x_{ij} = \dfrac{X_{ij} - \bar{X}_j}{\left[\sum_{k=1}^{n} (X_{kj} - \bar{X}_j)^2\right]^{1/2}}$

se tiene la doble ventaja de eliminar posibles problemas de colinealidad no esencial y de poder interpretar X'X y X'Y como matrices de correlaciones entre los regresores y de los regresores con la variable dependiente respectivamente[8]. Aunque, por otro lado, Besley, Kuh y Welsch (1980) consideran que no se deben centrar las variables pues ocultarían posibles relaciones entre los regresores; para la presentación de esta sección consideraremos el modelo transformado (4).

Para el estimador de MC, $\hat{\beta} = (X'X)^{-1}X'Y = CX'Y$; donde $C = (X'X)^{-1}$, se tiene que, $E(\hat{\beta}) = \beta$ y $\mathrm{cov}(\hat{\beta}) = \sigma^2 C$. En particular, $E(\hat{\beta}_j) = \beta_j$ y $\mathrm{var}(\hat{\beta}_j) = \sigma^2 C_{jj}$.

Se sabe por el teorema de Gauss-Markov que dentro de los estimadores lineales insesgados, $\hat{\beta}$ es el mejor estimador, es decir de varianza mínima, pero esto no garantiza que ésta sea pequeña. De hecho, es posible demostrar[9] que

$$C_{jj} = \frac{1}{1 - R_j^2},$$

---

[7]Se usará la misma representación para el vector de parámetros β, pero nótese que es distinto el utilizado en el modelo (1) y el usado en el modelo (4).

[8]Véase por ejemplo, Belsley, Kuh y Welsch (1980).

[9]Véase, Johnston (1972)

donde R_j^2 es el coeficiente de determinación múltiple al hacer la regresión de X_j respecto al resto de variables independientes. En presencia de colinealidad, al menos una de las variables explicativas, por ejemplo X_j, podrá aproximarse linealmente en términos de las otras, por lo que R_j^2 será cercano a 1 , C_{jj} será muy grande y por lo tanto también $var(\hat{\beta}_j)$, lo que ocasiona una gran incertidumbre en la estimación de β_j.

Los C_{jj} son llamados *factores de inflacion de la varianza* (FIV) de los parámetros y Marquardt (1970) los interpreta como "los factores por los cuales las varianzas de los estimadores son incrementados, debido solamente a la correlación entre los predictores". Si uno o varios FIV son grandes ($\geq$ 10), esto indica que los parámetros respectivos están estimados pobremente.

Si se considera el supuesto esférico de los errores, esto es $u\ N(\Theta, \sigma^2 I)$, la longitud de los intervalos de $100(1-\alpha)\%$ de confianza para β_j es

$$L_j = 2\sqrt{C_{jj}}\ t(n-p;\alpha/2),$$

mientras que el correspondiente para el caso ortogonal es

$$L'_j = 2\ t(n-p;\alpha/2)$$

luego,

$$L_j/L'_j = \sqrt{C_{jj}}\ .$$

Así pues, otra interpretación de los FIV es que su raíz cuadrada indica qué tanto el intervalo de confianza del parámetro respectivo se ha "inflado" debido a la colinealidad.

Extendiendo la idea, Willian y Watts (1978) presentan la interpretación siguiente: si $(X'X)$ está en la forma de matriz de correlaciones, $0 \leq det(X'X) \leq 1$; si los regresores son ortogonales $det(X'X) = 1$, y en el extremo opuesto bajo colinealidad perfecta $det(X'X) = 0$. La región de confianza para una matriz de diseño ortogonal al $100(1-\alpha)\%$ está dada por la relación

$$(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \leq p\sigma^2\ F(p,\ n-p;\alpha)$$

mientras que si los regresores son no ortogonales, la región al mismo

nivel estaría dada por

$$(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta) \leq p\sigma^2 F(p, n-p;\alpha),$$

de donde el cociente de los volúmenes de las regiones es $\det(X'X)^{1/2}$. Es decir, $100(\det(X'X)^{-1/2} - 1)\%$ es el incremento del volumen de la región de confianza al separarse las variables independientes de la ortogonalidad.

Por otro lado, la colinealidad en los datos provocará sobrestimación de los parámetros. Sea $L=\|\hat{\beta}-\beta\|$ la distancia entre el estimador $\hat{\beta}$ y el vector real de parámetros (no observable) β.

$$E(L^2) = E(\|\hat{\beta}-\beta\|^2) = \sum_{j=1}^{p} E(\hat{\beta}_j-\beta_j)^2 = \sum_{j=1}^{p} \mathrm{var}(\hat{\beta}_j) = \sigma^2 \sum_{j=1}^{p} \delta_j^{-2}$$

donde δ_i, $i = 1, \ldots, p$ son los valores singulares de la matriz X.[10] Bajo colinealidad (mal condicionamiento de X) uno o más de sus valores singulares serán cercanos a cero, por lo que la distancia promedio entre $\hat{\beta}$ y β será muy grande, luego

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

de donde

$$E(\|\hat{\beta}\|^2) = \|\beta\|^2 + \sigma^2\sum_{j=1}^{p} \delta_j^{-2}$$

Así pues, el estimador $\hat{\beta}$ tenderá a ser mucho más grande que el vector de parámetros β, por lo que en presencia de colinealidad se tendrán estimaciones amplificadas de los parámetros.

[10] La Descomposición en Valores Singulares (DVS) se presenta en el apéndice.

3. FUENTES DE LA COLINEALIDAD

Si frente a un problema de colinealidad se pueden identificar sus causas, se estará en posición de dar la mejor solución al problema, pero esto no siempre será posible. Habrá ocasiones en que no sea sencillo determinar las fuentes de la colinealidad, o bien que aún identificando las causas, resulte imposible eliminar el problema. Es importante de cualquier modo conocer las fuentes de colinealidad más frecuentes, para que éstas se eviten en la medida de lo posible.

Montgomery y Peck (1982) consideran como fuentes principales de la colinealidad:

1. *El método empleado en la recolección de datos.* Cuando las observaciones de las variables explicativas son tomadas solamente de un subconjunto del espacio muestra, donde su comportamiento está relacionado. Así por ejemplo, si se quiere analizar el consumo doméstico de gas como función del ingreso familiar, el número de personas por familia y el consumo de electricidad, y la muestra se restringe a un solo rumbo de la localidad, es muy probable que el consumo de electricidad sea a su vez explicado mediante el ingreso familiar y el número de personas por familia. También pueden suceder problemas de colinealidad cuando se trabaja con rangos muy pequeños de las variables explicativas.

2. *Las restricciones sobre el modelo y la población.* Las restricciones en la población son quizá la causa de colinealidad más frecuente cuando se trabaja con variables temporales en Econometría, pues en general tienden a crecer o decrecer de manera conjunta[11].

   En ocaciones, algunas variables resultan de la agregación de otras, por lo que al incluirlas en el modelo seguramente estarán relacionadas con algún otro regresor. Sucede por ejemplo, cuando se incluye el PIB (Producto Interno Bruto) en los modelos

[11]Por ejemplo, se ha mostrado que series temporales crecientes e *independientes* muestran cierta correlación; esto es lo que se ha llamado *correlación espuria.* Véase Granger (1969).

econométricos[12].

Por otro lado, en lo que se refiere a restricciones del modelo, con frecuencia se incluyen variables indicadoras exhaustivas. Por ejemplo, si I_1, I_2, I_3, I_4 son variables indicadoras que representan las estaciones del año (I_k=1 si se está en la estación k-ésima, I_k=0 de otro modo), necesariamente $\sum_{k=1}^{4} I_k = 1$, luego ,si se incluyen las cuatro variables en el modelo desde un principio se estará incluyendo relaciones de colinealidad.

3. *Elección del modelo.* Este caso puede ser el resultado de las fuentes anteriores. Por ejemplo, en un modelo polinomial de regresión, las variables X y X^2 alrededor de 1 se comportan de manera similar. Así también, es conocido el mal comportamiento de la matriz X en modelos de interpolación polinomial cuando se trabaja con grados altos.

4. *Sobreespecificación del modelo.* Esta causa también es común en modelos econométricos. Se incluyen tantos regresores que muchos de ellos resultan relacionados entre sí. También es frecuente encontrar en estudios de la salud modelos con demasiadas variables explicativas, por lo que necesariamente se dan relaciones entre ellas.

Finalmente, como se comentó al inicio de la sección, el conocer las fuentes de la colinealidad nos permite evitarla cuando es posible, y eventualmente proponer alguna solución.

[12]En pocos modelos se justifica incluir como variable explicativa el PIB; sin embargo, en la práctica diaria es muy socorrido como predictor.

4. DIAGNOSTICO DE LA COLINEALIDAD

Habiendo mostrado los efectos que tiene la colinealidad sobre el estimador $\hat{\beta}$ y sus fuentes potenciales, en esta sección se presentan algunas formas en que se ha intentado detectar el problema en el pasado; así mismo se presentan los métodos de diagnóstico utilizados actualmente aprovechando el desarrollo que han tenido los métodos numéricos en los últimos años, tanto en el terreno teórico, como en el desarrollo de bibliotecas y sistemas de cómputo de calidad. Cabe aclarar que el orden en que se presentan las formas de diagnóstico no necesariamente corresponde al de su aparición en la literatura.

Históricamente se han definido criterios con el fin de detectar la presencia de la colinealidad. Algunos de ellos surgen del estudio de los efectos del problema, otros se basan en el análisis directo de la matriz de diseño X. Ciertos criterios que se siguen para el diagnóstico son bastante simples y no requieren de gran esfuerzo, aunque quizá por lo mismo no son del todo seguros. En el extremo opuesto, hay procedimientos casi infalibles pero que requieren de trabajo extra al de un análisis de regresión estándar.

1. Entre las formas de diagnóstico elementales está el observar los elementos de la matriz (X'X) en su forma de correlación, $\mathrm{corr}(X_i, X_j) = \hat{\rho}_{ij} = r_{ij} = (X'X)_{ij}$. Aquellos elementos de la matriz cercanos a ± 1 indican que las variables correspondientes están correlacionadas. Sin embargo, es posible que aún en presencia de colinealidad perfecta los r_{ij} sean bajos y esto porque la dependencia de una de las variables independientes es de dos o más de las restantes. Es poco común que se dé el problema de colinealidad debido a la correlación entre dos variables exactamente.

2. Si después de estimar los coeficientes del modelo se tiene que la F global es significativa, pero las t-parciales no lo son, los resultados se interpretarían como si el modelo en general fuera correcto, pero ninguna de las variables regresoras tomadas independientemente aportara información importante para la explicación de la respuesta; esto es un síntoma común de

colinealidad, pero no necesariamente bajo colinealidad éste se presenta.

3. Como se comentó en la introducción, la colinealidad es el mal comportamiento de la matriz de diseño X, por lo que cálculos que involucren a $(X'X)^{-1}$, y en particular $\hat{\beta}$, serán altamente sensibles a los datos. Así pues, otra forma de intentar diagnosticar la colinealidad es observando la sensibilidad de $\hat{\beta}$ a la eliminación de observaciones y/o predictores[13]. De la misma forma, si en la estimación resultan valores o signos no esperados por los parámetros, esto puede estar indicando que hay problemas de colinealidad, aunque también podría ser resultado de una mala especificación del modelo o bien de la ignorancia que se tiene sobre la variable respuesta y su relación con los regresores.

4. Otro procedimiento de diagnóstico es por medio de los factores de inflación de la varianza $FIV_j = C_{jj}$. Basándose en su experiencia práctica, Montgomery y Peck (1980) consideran que FIV mayores que 10 son indicadores de una pobre estimación debido a la colinealidad. Este criterio tiene la ventaja de ser simple y poder llevarse a cabo utilizando una salida estándar de regresión lineal. Sin embargo, no aporta información acerca de qué variables están involucradas en la o las interdependencias.

5. Nuevamente, si los datos están centrados y escalados, la matriz (X'X) toma forma de matriz de correlaciones y en tal caso $0 \leq \det(X'X) \leq 1$. De hecho, en los casos extremos, si se tiene colinealidad perfecta $\det(X'X) = 0$; en el opuesto, si los regresores son ortogonales $\det(X'X) = 1$. De esta situación se sigue que det(X'X) pequeños es síntoma del problema, pero ¿qué tan pequeño ha de ser det(X'X) para poder considerar que se tiene colinealidad?. Esto es equivalente a tratar de determinar el rango numérico de una matriz,

[13]Se pueden tener estimadores de ciertas combinaciones lineales de los parámetros con varianza moderada y sin problemas de sensibilidad en los datos. Estas se conocen como *funciones estimables*. Véase Theil (1971) o Scheffé (1959).

problema aún en estudio[14].

6. Precisamente en esta linea es el ya famoso artículo de Farrar y Glauber (1967), que durante algún tiempo se pensó resolvería el problema de diagnóstico. Bajo supuestos de normalidad y ortogonalidad en los regresores (contrario al supuesto 2c, X no estocástica) y bajo cierta transformación llegan a que det(X'X) se distribuye aproximadamente como χ^2, lo que permitiría comparar estadísticamente det(X'X) con cero. Sin embargo, estos supuestos son demasiado fuertes y pocas veces se cumplen en la práctica, además su trabajo tiene omisiones importantes como lo muestran Kumar (1975) y Wichers (1975) en sus artículos y Belsley, Kuh y Welsch (1980) en su libro.

7. Por otro lado, Silvey (1969) aborda el problema de la colinealidad mediante la Descomposición Espectral de la matriz (X'X). Uno o más valores característicos pequeños son síntoma de colinealidad. Asociado a cada valor característico pequeño hay una relación lineal entre los regresores. En su análisis por Componentes Principales, Silvey propone además, formas de corregir la colinealidad añadiendo información apropiada que pueda romper la estructura de los datos. La manera de diagnóstico que él propone resulta incluída en la que se presentará más adelante utilizando DVS, además de que el uso de la descomposición singular arroja mayor información y los algoritmos desarrollados para este último resultan más estables numéricamente que los utilizados usualmente para los sistemas característicos[15]. Cabe sin embargo, destacar su trabajo.

8. Considérese entonces la Descomposición en Valor Singular (DVS) de la matriz X, $X = UDV'$, donde U_{nxn}, V_{pxp} matrices ortogonales y $D = \begin{bmatrix}\Delta \\ \Theta\end{bmatrix}$, $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_p)$, $\delta_1 \geq \ldots \geq \delta_p > 0$ y Θ matriz de ceros. Si X es de rango incompleto, entonces al menos uno de los valores singulares

[14]El *rango numérico* de una matriz es el rango que tiene su representación interna en la computadora. Una forma de determinar el rango de una matriz es por el número de valores singulares distintos de cero, pero nuevamente, ¿ cuáles podrían considerarse cero?. Véase Golub, Klema y Stewart (1976).

[15]Paquetes como EISPACK (1976) y LINPACK (1979) para computadoras grandes, LOBOLIB (1985) para Computadores Personales.

de X será exactamente cero[16]. Si X de rango seudodeficiente, entonces uno o más de los valores singulares serán cercanos a cero.

El número de condición de la matriz X, $K(X):=\delta_1/\delta_p$, es un indicador de lo mal comportada (numéricamente) que puede resultar la matriz X. O bien, extendiendo la idea de Boley y Bittanti (1986), $K^{-1}(X)$ es la "distancia relativa de la matriz X a la colinealidad perfecta". En la práctica, si $10 \leq K(X) \leq 1000$, se considera que la colinealidad es moderada, mientras que si $K(X) > 1000$, entonces el problema es grave.

9. Besley, Kuh y Welsch (1980), definen lo que ellos llaman *índices de condición*,

$$\eta_j = \frac{\delta max}{\delta j} \qquad j = 1, \ldots, p$$

y consideran que uno o más valores de η_j mayores que 30 son un síntoma de colinealidad (nótese que $K(X) = \max\{\eta_j\}$).

En la sección 2, se vió que $var(\hat{\beta}_j) = \sigma^2 C_{jj} = \sigma^2 FIV_j$, luego, a partir de DVS de X se puede expresar,

$$FIV_j = \sum_{i=1}^{p} \frac{v_{ij}^2}{\delta_i^2} \qquad \text{donde } v_{ij} = [V]_{ij}.$$

De aquí, es claro que valores singulares pequeños tienen como consecuencia una estimación pobre del parámetro . Besley, Kuh y Welsch (1980), utilizan esta información para descomponer porcentualmente los FIV_j y definir las *proporciones de descomposición de la varianza*

$$\pi_{ij} = \frac{v_{ij}^2 / \delta_j^2}{FIV_j} \qquad i, j = 1, \ldots, p,$$

de tal modo que si se construye la matriz $\Pi = (\pi_{ij})$, la suma de los elementos de una columna es 1, y los elementos de una misma fila están asociados a un mismo valor singular y sólo a uno. Valores de π_{ij} que resulten por encima de 0.50 significa que más de un 50% de la varianza de un estimador se concentra en una componente, situación

[16]En este caso, por el supuesto 2c) $\rho(X) = p$, por lo que $\delta_p > 0$. Véase el apéndice para una mayor explicación y referencias bibliográficas.

que sería indicadora de posible colinealidad.

Analizando la matriz Π de proporciones, es posible detectar qué valores singulares son los que más influyen en la varianza del estimador $\hat{\beta}_j$, observar todas las relaciones de colinealidad que se presentan y la magnitud de éstas, además de identificar las variables que intervienen en cada una de las relaciones. Así por ejemplo, si se tiene que π_{22}, π_{23} y π_{25} tienen valores altos y el índice de condición η_2 también lo es, entonces las variables X_2, X_3, X_5 establecen (aproximadamente) una relación lineal. Si de igual forma se identificara una relación entre las variables X_1 y X_3 se pensaría entonces en la posibilidad de eliminar la variable X_3 del modelo por aparecer en más de una relación colineal.

Así pues, en esta sección se han presentado distintas formas de diagnóstico de la colinealidad. Algunas de ellas son bastante simples y no requieren más allá de los resultados usuales de un análisis de regresión. Otras, más complejas, ofrecen mayor información y seguridad en el diagnóstico a costa de un trabajo extra (no simple) de cálculo.

5. CORRECCION Y ALTERNATIVA

El objetivo de la sección es presentar distintas opciones a seguir una vez que se haya identificado la presencia de colinealidad. En ocasiones será posible disminuir la gravedad del problema (quizá hasta eliminarlo); en otras, habrá que "vivir con el problema" y en tal caso se sugiere utilizar estimadores distintos a los de Mínimos Cuadrados.

A diferencia del supuesto del rango, la mayoría de los autores coinciden en que el mejor remedio para corregir colinealidad es añadir información de "mejor calidad". Esto es, incorporar información extra de modo que rompa la estructura colineal[17]. Así, por ejemplo, Silvey (1969) describe cómo debe ser la información adicional de modo que ésta sea "óptima", pero esto no siempre es posible por razones económicas o físicas (repetir o extender un censo puede ser muy costoso, así como resulta imposible tomar cierta información una vez que concluye el experimento o proceso en estudio).

Sin embargo, debe notarse que hay situaciones en que la colinealidad no puede resolverse aumentando datos de "calidad". Tal es el caso cuando la fuente del problema son las restricciones sobre el modelo o su especificación. Aquí, la mejor solución es precisamente la reespecificación del modelo, tratando de evitar el uso de *todas* las variables explicativas que intervienen en la relación colineal, o bien utilizar (cuando sea posible) métodos o técnicas del Diseño de Experimentos.

En algunos casos es posible definir combinaciones de predictores que conservan la información de las variables pero evitan usar todas ellas, lo que disminuye la posibilidad de relaciones colineales. Sin embargo, muchas veces esto es imposible por la naturaleza misma del análisis de regresión, como cuando se desea estudiar la relación entre un subconjunto específico de regresores y la variable respuesta, y en tal situación la alternativa posible es el uso de

[17] Véase, Farrar y Glauber (1967).

otro tipo de estimadores diferentes a los MC, no necesariamente lineales o insesgados. A continuación presentamos algunos de estos estimadores.

a) *Estimadores Bayesianos*. Incorporan información previa al modelo a manera de distribución a priori del parámetro β. Así por ejemplo, si se considera a priori que β $N(\beta_0, \Sigma_0)$, el estimador bayesiano resulta dado por[18]

$$\hat{\beta}_B = \left[\frac{1}{\sigma^2} X'X + \Sigma_0^{-1}\right]^{-1} \cdot \left[\frac{1}{\sigma^2} X'y + \Sigma_0^{-1}\beta_0\right]$$

b) *Estimadores Mixtos*. Estos estimadores fueron desarrollados por Theil y Goldberger (1961). Al igual que los bayesianos, los estimadores mixtos utilizan información previa, pero traducen ésta aumentando los datos del modelo en lugar de suponer una distribución a priori del parámetro. Se considera que, en base a esa información el analista puede construir r (con r<p) restricciones sobre el parámetro β,

$$\alpha = R\beta + v$$

con α y v vectores aleatorios rx1, tales que $E(v) = \theta$, $var(v) = V$, R matriz rxp conocida, de rango completo. Las restricciones se incorporan al modelo original para formar el modelo ampliado

$$\begin{bmatrix} y \\ \alpha \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix}\beta + \begin{bmatrix} u \\ v \end{bmatrix}$$

y por MC se obtine el estimador mixto[19]

$$\hat{\beta}_M = \left[\frac{1}{\sigma^2} X'X + R'V^{-1}R\right]^{-1} \cdot \left[\frac{1}{\sigma^2} X'y + R'V^{-1}\alpha\right]$$

---

[18] Leamer (1973) estudia el problema de la colinealidad desde el punto de vista bayesiano. Baz (1976) presenta en particular estimadores bayesianos sesgados y la justificación bayesiana de otro tipo de estimadores sesgados.
En general, para un tratamiento bayesiano del modelo lineal, véase Box y Tiao (1973) y Zellner (1971).

[19] Puede verse por las restricciones y el hecho de estimar por MC en el modelo ampliado que los estimadores mixtos resultan insesgados. Nótese también su semejanza con los estimadores bayesianos; de hecho los EM tienen justificación bayesiana. Véase Besley, Kuh y Welsch (1980) y Theil (1971) para ejemplos.

c) *Estimadores Cordillera*[20]. Hoerl y Kennard (1970 a.b.) presentaron un procedimiento al que llamaron *Ridge Regression*, como extensión a la regresión de trabajos anteriores de Hoerl a los que llamó "Ridge Analysis". A diferencia de los estimadores bayesianos y mixtos los estimadores cordillera y de componentes principales que más adelante se presentan son estimadores sesgados del vector de parámetros β.

Como se vió en la sección 2, uno de los efectos de la colinealidad es producir estimadores de MC con gran varianza. Se sabe por el teorema de Gauss-Markov que dentro de la familia de estimadores lineales insesgados los de MC son los de varianza mínima, pero esto no garantiza que la varianza sea pequeña; por ello el uso de estimadores sesgados se justifica al reducir la variación del estimador o mejorar su estabilidad, aún a costa de introducir un sesgo. Existen, por supuesto, criterios para definir cuánto se está dispuesto a aceptar de sesgo con tal de reducir la variabilidad del estimador; uno de esos criterios es el Error Cuadrático Medio (ECM)[21].

Hoerl y Kennard (1970a) presentan los estimadores cordillera

$$\hat{\beta}_\gamma = (X'X + \gamma I)^{-1}X'Y \qquad \gamma \geq 0$$

para "el control de la inflación e inestabilidad general asociada con el uso de estimaciones MC". De manera simplista se puede argumentar que en el caso extremo de que la matriz $X'X$ sea singular, la matriz $(X'X + \gamma I)$ resulta de rango completo para $\gamma > 0$, (se sigue al notar que $(\delta^2_j + \gamma)$, $j = 1, \ldots, p$ son los valores singulares de $(X'X + \gamma I)$), o bien, para γ suficientemente grande, $(X'X + \gamma I)^{-1} \simeq \gamma^{-1}I$, sin importar lo mal comportada que pueda ser X[22].

[20]Traducción del inglés "Ridge estimators".
[21]$ECM(\tilde{\beta}) = E(L^2(\tilde{\beta})) = E(\|\tilde{\beta}-\beta\|^2)$.
[22]Existen varias justificaciones de los estimadores cordillera tales como el bayesiano (Baz (1976)), como resultado de un proceso regularizante a la Tijonof (Barrios y López (1986)), etc. Así mismo se recomienda la lectura del artículo de Marquardt (1970) para una exposición de los estimadores sesgados y su uso con datos no ortogonales.

Ahora bien, en la sección 2 se vió que la distancia "promedio" del estimador $\hat{\beta}$ a β es

$$E(L^2(\hat{\beta})) = E(\|\hat{\beta}-\beta\|^2) = \sum \mathrm{var}(\hat{\beta}_j) = \sigma^2 \sum \delta_j^{-2},$$

en forma análoga, la distancia de $\hat{\beta}_\gamma$ a β es

$$E(L^2(\hat{\beta}_\gamma)) = \sigma^2 \sum \frac{\delta_j^2}{(\delta_j^2+\gamma)^2} + \gamma^2 \sum \frac{\alpha_j^2}{(\delta_j^2+\gamma)^2}$$

$$=: g_1(\gamma) + g_2(\gamma)$$

donde $\alpha = V\beta = (\alpha_1, \ldots, \alpha_p)'$. Aquí, g_1 puede interpretarse como la variación total (suma de varianzas) y g_2 como el cuadrado del sesgo introducido al usar $\hat{\beta}_\gamma$ en lugar de $\hat{\beta}$.

Hoerl y Kennard (1970a) mostraron que existe $\gamma > 0$, para el cual $ECM(\hat{\beta}_\gamma) < ECM(\hat{\beta})$, por lo que podrían pensarse los estimadores cordillera como alternativa a los estimadores de mínimos cuadrados. Sin embargo no es posible asignar un valor específico a γ que satisfaga la desigualdad anterior (y en tal caso se dice que el estimador es *admisible*) pues las condiciones sobre ella quedan en términos de σ^2 y β, en general desconocidos. Por lo que es necesario definir criterios que conduzcan a una buena elección de γ. Hoerl y Kennard (1970b) proponen el uso de la *traza cordillera*, que consiste en graficar las $\hat{\beta}_j(\gamma)$ $(j = 1, \ldots, p)$ para distintos valores de γ y elegir aquel valor γ_0 para el cual las trazas de los estimadores parecen estabilizarse[23].

d) *Regresión en Componentes Principales*. El procedimiento de componentes principales en la regresión, como alternativa cuando hay colinealidad, ha mostrado ser muy eficiente[24]. Como se presentan aquí estos estimadores resultan también ser sesgados, pero poseen la ventaja de poder controlar la varianza del estimador.

[23] Véase Barrios y López (1986) y Montgomery y Peck (1982), donde se presentan otros métodos para seleccionar γ. Dempster, Schatzoff y Wermuth (1977), presentan un estudio de 56 simulaciones, comparando entre distintos procedimientos de selección de γ. Sin embargo hasta ahora, no existe una forma de escoger γ que domine en general al resto de los procedimientos.

[24] Véase Gunst y Mason (1977), para un estudio de simulación.

De manera muy simple, se puede decir que el procedimiento de componentes principales consiste en la transformación del modelo original de modo que los nuevos regresores resulten ortogonales entre sí, con las ventajas que esto implica. El modelo transformado en su *forma canónica* es

$$Y = Z\alpha + u$$

donde, $Z = XV$ y $\alpha = V'\beta$, para cualquier matriz ortogonal V_{pxp}. Se tiene entonces que $V'X'XV = Z'Z = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$, donde si V se elige adecuadamente, $\lambda_1 \geq \ldots \geq \lambda_p > 0$, son los valores característicos de la matriz $X'X$. Los nuevos regresores ortogonales $Z_1, \ldots, Z_p$ se conocen como las *componentes principales*. El estimador de MC en el modelo transformado queda

$$\hat{\alpha} = (Z'Z)^{-1}Z'y = \Lambda^{-1}Z'y$$

con $$\text{var}(\hat{\alpha}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$$

y $$\text{var tot}(\hat{\alpha}) = \sum \text{var}(\hat{\alpha}_j) = \sigma^2\sum 1/\lambda_j$$

Como ya se ha comentado, asociado a la colinealidad se tienen valores singulares pequeños δ_j, lo que repercute en la inflación de la varianza total del estimador. Una manera de reducir esta varianza es eliminar aquellas componentes $\hat{\alpha}_k$ con valor característico asociado λ_k pequeño[25]. En tal forma, si las últimas r componentes son eliminadas, el modelo reducido

$$Y = Z^*\alpha^* + u = [Z_1, \ldots, Z_{p-r}]\begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_{p-r}\end{bmatrix} + u$$

resulta mal especificado pues al vector de parámetros le faltarán componentes, con el correspondiente estimador de MC $\hat{\alpha}^* = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{p-r})'$ sesgado pero de menor variación que $\hat{\alpha}$.

Al igual que en los estimadores cordillera, hay ciertos criterios que indican el número de regresores que podrían ser eliminados[26]. En

[25] Nótese que si δ_j son los valores singulares de X y λ_j son los valores propios de $(X'X)$, entonces $\delta_j^2 = \lambda_j$, $j = 1, \ldots, p$, pues $\Delta^2 = V'(X'X)V = \Lambda$.

[26] Véase, Silvey (1969).

esta línea, Marquardt (1970) introduce la idea de *rango fraccionario*, que a su vez resulta relacionado con el problema de determinar el rango de una matriz.

Además de la gran ventaja de contar con un respaldo teórico importante, el uso de componentes principales tiene la de que, aún con datos mal condicionados, el procedimiento resulta ser de los más eficientes para efectos de pronóstico. Sin embargo, tiene el inconveniente de transformar el modelo original, lo que dificulta en ocasiones su interpretación.

Finalmente, vale mencionar nuevamente que aunque la mayoría de los autores coinciden en que la mejor manera de remediar el problema de la colinealidad es mediante información extra de mejor calidad, esto no siempre es posible, y en tal caso se presenta como alternativa el uso de estimadores sesgados o del tipo bayesiano [27].

[27] En el libro de Montgomery y Peck (1982), sec. 8.5.7, se comenta una buena selección de artículos respecto a la comparación y evolución de los estimadores sesgados.

6. CONCLUSIONES

La colinealidad es uno de los problemas que aparecen con mayor frecuencia en la regresión y, visto fríamente, es un problema numérico de mal condicionamiento de la matriz de diseño X, debido a la "mala calidad" de los datos. Esto se hace evidente al intentar resolver las ecuaciones normales de los Mínimos Cuadrados. Como consecuencia, los estimadores obtenidos resultan pobremente calculados, de gran variabilidad o altamente sensibles, llegando inclusive a ser inservibles en la práctica. Sin embargo, el diagnóstico y su remedio son materia de estudio de la Estadística.

Una vez especificado el modelo puede resultar que la matriz de datos tenga rango deficiente, lo que obliga a su reespecificación de modo que la matriz correspondiente sea de rango completo (para este momento ya se ha comenzado con el análisis de la regresión). Así pues, definidos los regresores que intervendrán en el modelo y estimados los parámetros, se deberá evaluar la calidad de las estimaciones. Para esto se sugiere usar desde un principio métodos numéricos de calidad en la regresión - como la descomposición QR - evitando en lo posible trabajar con las matrices X y X'X. En esta linea véase por ejemplo, Golub (1979), Stewart (1973), Montgomery y Peck (1982) (versiones recientes de paquetería estadística para cómputo han incorporado este tipo de algoritmos).

Para diagnosticar la presencia de colinealidad en la regresión, sobre los resultados de una salida estándar de computadora se verifica si se detectan algunos de los síntomas comentados en la sección 4, como son valores o signos no esperados, F global significativa y t-parciales pequeñas; alternativamente, utilizando los errores estándar de los coeficientes se calculan los FIV correspondientes, y si es factible, se averigua la sensibilidad de los estimadores ejecutando varias corridas ordenadas eliminando observaciones y/o regresores.

Se sugiere calcular las proporciones de descomposición de la varianza Π solamente bajo sospecha de colinealidad pues la descomposición en valores singulares puede resultar costosa. Si se detecta que el problema está presente, entonces habrá que identificar todas las relaciones colineales y los predictores que intervienen en ellas. En este momento se recomienda detenerse para investigar las posibles causas de la colinealidad o si es factible eliminar uno o más de los regresores implicados en las relaciones.

Siempre que sea posible, es conveniente aumentar datos de mejor calidad de modo que aligeren o eliminen el problema de la colinealidad. Si esto no fuera posible y el objetivo del modelo es la estimación y predicción de la variable dependiente, se recomienda el uso de componentes principales y trabajar con los niveles de variación (de la variable respuesta) que se desee. O bien, si el objetivo es el análisis del modelo propuesto o el control de la respuesta se sugiere entonces el uso de estimadores cordillera. Por otro lado, si el analista tiene información que pueda incorporar al modelo, puede utilizar estimadores del tipo bayesiano.

Por último, resulta común en la actualidad realizar análisis de regresión en computadoras personales y no siempre se dispone de paquetería especializada para poder elegir entre los distintos tipos de estimadores. Si este fuera el caso, cabe mencionar que es posible trabajar los estimadores cordillera con rutinas estándar de regresión lineal. El estimador cordillera $\hat{\beta}_\gamma$ será el estimador de MC del modelo ampliado

$$\begin{bmatrix} Y \\ \Theta \end{bmatrix} = \begin{bmatrix} X \\ \sqrt{\gamma} I \end{bmatrix} \beta + u.$$

7. APENDICE[28]

Descomposición LU. Sea $A \in \mathbb{R}^{nxn}$. Existen entonces L, $U \in \mathbb{R}^{nxn}$ triangulares inferior y superior respectivamente. L con 1's en la diagonal principal tales que A = LU.

Aplicaciones: a) El sistema Ax=b puede descomponerse en dos sistemas triangulares, Ly=b, Ux=y, donde para el primero $y_1=b_1$, $y_2=b_2-l_{21}y_1$, ...,etc. (sustitución hacia adelante) y para el segundo $x_n=y_n/u_{nn}$, $x_{n-1}=(y_{n-1}-u_{n-1,n}x_n)/u_{n-1,n-1}$, etc. (sustitución hacia atrás).

b) $\det(A) = \det(LU) = \det(L)\det(U)=1\cdot\prod_{i=1}^{n} u_{ii}$

Descomposición QR. Sea $A \in \mathbb{R}^{mxn}$ $(m\geq n)$. Existen matrices Q_{mxm} ortogonal y R_{nxn} triangular superior tal que $A = Q\begin{bmatrix}R\\ \Theta\end{bmatrix}$. $\Theta \in \mathbb{R}^{(m-n)xn}$ de elementos todos nulos.

Aplicaciones: a) Como Q es ortogonal, no modifica la norma de un vector para el problema de MC, $\min \|Ax-b\|^2 = \min \|Q(Ax-b)\|^2$ y su solución se obtiene al resolver el sistema $Rx=c_1$, donde $Q'b = \begin{bmatrix}c_1\\ c_2\end{bmatrix}$, con sustitución hacia atrás. Nótese que resolver $Rx=c_1$ es preferible que resolver las ecuaciones normales $R'Rx=R'c_1$ $(A'Ax=A'b)$ que implica más trabajo de cómputo y por lo mismo más errores.

b) Si A_{nxn}, $\det(A)=\det(QR)=\det(Q)\det(R)=1\circ\prod_{i=1}^{n} C_{jj}$

Condicionamiento numérico de un problema. Es una consideración importante siempre que se piense en el uso de computadoras digitales. Debido a la longitud finita de la palabra, los números sólo pueden

[28]Parte del material e ideas que aquí se presentan fueron extraídas de Boley y Bittanti (1986). Se recomienda el libro de Stewart (1973) para adentrarse en el Algebra Lineal Numérica.

representarse aproximadamente en la computadora. En el tratamiento de tales aproximaciones los modelos más usados son los que consideran qué sucede si los números son perturbados ligeramente.

En la solución de un problema uno debe aplicar algún algoritmo. Si el problema es *mal condicionado* o *mal planteado*, entonces pequeñas perturbaciones a los datos pueden arrojar resultados prácticamente inútiles.

Si un problema es bien condicionado y el método empleado para su solución es estable, entonces se puede creer en la respuesta. Un método se dice *estable* si introduce pocos errores en los cálculos. Si el problema es bien planteado y el método inestable, se puede pensar en usar mejores métodos. Sin embargo, no se puede pensar en mejorar la calidad de la solución si el problema es mal condicionado, independientemente de que el método utilizado sea estable o no. En este sentido, la descomposición QR es un algoritmo estable.

<u>Descomposición en Valores Singulares</u> (DVS). Sea $A \in \mathbb{R}^{m\times n}$ $(m\geq n)$. Existen matrices $U_{m\times m}$, $V_{n\times n}$ ortogonales y $\Sigma_{m\times n}$ tal que $\Sigma = \begin{bmatrix}\Delta\\ \Theta\end{bmatrix}$ $\Delta=\mathrm{diag}(\delta_1, \ldots, \delta_n)$, $\delta_1 \geq \ldots \geq \delta_n \geq 0$, llamados *valores singulares* de A, $\Theta \in \mathbb{R}^{(m-n)\times n}$ matriz de elementos nulos tal que, $A = U\Sigma V'$.

<u>Aplicaciones</u>: a) El rango de una matriz $A = U\Sigma V'$ es el número de valores singulares distintos de cero.

b) Si $A_{n\times n}$, $\det(A) = \prod_{i=1}^{n} \delta_i$

Para la norma matricial $\|A\| = \max\{\|Ax\| : \|x\|=1\}$ heredada de la norma euclidiana usual, se tiene que si Q es una matriz ortogonal, $\|Qx\| = \|x\|$ y $\|QA\| = \|A\|$, de donde se sigue que

$$\|A\| = \|\Sigma\| = \delta_{max} = \delta_1$$

Si $A_{n\times n}$ es no singular, $A^{-1} = V\Sigma^{-1}U'$, y

$$\|A^{-1}\| = \|\Sigma^{-1}\| = \delta_n^{-1}$$

Para $A_{n\times n}$ no singular, se define el *número de condición* de A,

mediante[29]

$$K(A) = \|A\| \; \|A^{-1}\| = \delta_1/\delta_n \geq 1.$$

Esta cantidad K(A) resulta útil para estimar la sensibilidad respecto a perturbaciones de la matriz A en la solución de sistemas de ecuaciones lineales , el cálculo del rango, de la inversa, etc.

Así, por ejemplo, considérese el sistema Ax = b. Al resolverlo, se obtiene $\tilde{x}$, que es la solución *exacta* de un problema perturbado ligeramente $A\tilde{x} = \tilde{b}$. De donde $A(x-\tilde{x}) = (b-\tilde{b})$. Multiplicando ambos lados de la igualdad por A^{-1} y tomando normas

$$\|x-\tilde{x}\| \leq \|A^{-1}\| \; \|b-\tilde{b}\|.$$

Esto es, el error en la respuesta está acotado por el error en el lado derecho del sistema ampliado $\|A^{-1}\|$ veces[30]. El error relativo

$$\frac{\|x-\tilde{x}\|}{\|x\|}$$

indica el número de dígitos de precisión en el cálculo de x. En el caso del sistema Ax=b, $\|A\| \cdot \|x\| \geq \|b\|$, de donde se sigue que

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq K(A) \; \frac{\|b-\tilde{b}\|}{\|b\|}$$

Así, el número de dígitos significativos en x está acotado por los de b ampliado K(A) veces. Para perturbaciones en A, se obtiene en forma análoga que

$$\frac{\|x-\tilde{x}\|}{\|\tilde{x}\|} \leq K(A) \; \frac{\|A-\tilde{A}\|}{\|\tilde{A}\|} .$$

Por otro lado, el rango de $A = \rho(\Sigma) = \#\{i: \delta_i \neq 0\}$. Se dice que la matriz es de rango *seudodeficiente* si δ_n es muy cercano a cero. Si A_{nxn} es no singular y sólo entonces, se tiene que $\delta_1 \geq \ldots \geq \delta_n > 0$.

Sea E_{nxn} una perturbación de A tal que A+E resulte singular. Luego, la *distancia a la singularidad* de la matriz A está dada por

$$\text{dsing} = \min_{(A+E)\text{sing}} \|E\| = \delta_n$$

---

29 El número de condición depende de la norma utilizada; así, $K_\nu(A) = \|A\|_\nu \; \|A^{-1}\|_\nu$. $K(A)=\delta_1/\delta_n$ es válido sólo en norma 2.

30 Es posible alcanzar esta cota, por lo que no hay manera de mejorarla.

y la *distancia relativa*

$$\frac{\text{dsing}}{\|A\|} = \frac{\delta_n}{\delta_1} = K^{-1}(A)$$

Se sigue entonces, que K(A) también es un indicador de la sensibilidad de rango(A). Similarmente, δ_r indica el "tamaño" de la perturbación para reducir a A, a una matriz de rango r-1 y δ_r/δ_1 el "tamaño relativo".

Finalmente, nótese que si δ_n es cercano a cero, K(A) tiende a ser muy grande y las cotas a los errores tienden a explotar, por lo que uno o más valores singulares pequeños son indicadores de problemas potenciales en la estabilidad del cálculo (mal comportamiento).

REFERENCIAS

BARRIOS Z. E., J. LOPEZ E.(1986)."Estimación Gauss-Markov regularizado con estimación óptima de parámetros". Reporte de Investigación. XIX Congreso de la Soc. Mat. Mex., Guadalajara, Jal. México. Nov. 86.

BAZ T. G. (1976) "Estimación sesgada en el modelo lineal". Tesis para obtener el título de Actuario. Fac. de Ciencias, UNAM.

BELSLEY D., E. KUH, R. WELSCH (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons; New York.

BOLEY D., S. BITTANTI (1986) "Numerical Problems in Linear Systems Problems", in S. Bittanti (Ed.), Time Series and Linear Systems. Springer-Verlag, Berlin.

BOX G.E.P., G. C. TIAO (1973) Bayesian Inference Statistical Analysis. Addison-Wesley; Massachusetts.

CHATERJEE S., B. PRICE (1977). Regression Analysis by Examples. John Wiley & Sons; New York.

DEMPSTER A.P., M. SCHATZOFF, N. WERMUTH (1977). "A simulation study of alternatives to Ordinary Least Squares". Journal of the American Statistical Association, vol. 72, pp. 77-103. Con comentarios.

EISPACK, (1976). Matrix Eigensystem Routines - Eispack Guide Extension. B. S. Garbow, et al. Springer-Verlag, Berlin.

FARRAR D.E., R.R. GLAUBER (1967). "Multicollinearity in regression analysis: the problem revisited", Review of Economics and Statistics, vol. 49, pp. 92-107.

FREUND R.J., P.D. MILTON (1979). Regression Methods. A tool for data analysis. Marcel Dekker Inc., New York.

GOLUB G.H. (1969). "Matrix descomposition and statistical calculations". in R.C. Milton and J.A. Welder (Eds.). Statistical Computation, Academic Press: New York.

GOLUB G.H., V. KLEMA, G.W. STEWART (1976). "Rank Degeneracy and Least square problems", Computer Sciences Technical Report Series, #TR-456, University of Maryland.

GRANGER C.W. (1969) "Investigating causal relationships by econometric models and cross-spectral methods". Econometrica, Vol. 37, pp. 424-438.

GUNST R.F., R.L. MASON (1977). "Biased estimation in regression: An evaluation using Mean Squared Error", Journal of the American Statistical Association, vol. 72, pp. 616-628.

HOERL A.E., R.W. KENNARD (1970a). "Ridge regression: Biased estimation for nonorthogonal problems", Technometrics, Vol. 12, pp. 55-67.

HOERL A.E., R.W. KENNARD (1970b). "Ridge regression: Applications to nonorthogonal problems". Technometrics, Vol. 12, pp. 69-82.

INTRILIGATOR, M.D. (1978). Econometric Models, techniques and application. Prentice-Hall, Inc., New Jersey.

JOHNSTON, J. (1972). Econometric methods. 2nd Ed. McGraw-Hill International, Tokyo.

KUMAR, T. K. (1975). "Multicollinearity in Regression Analysis". Review of Economics and Statistics, Vol. 57, pp. 365-366.

LEAMER E.E. (1973). "Multicollinearity: A bayesian interpretation". Review of Economics and Statistics. Vol. 55, pp. 371-388.

LINPACK (1979). Linpack Users Guide - J. J. Dongarra et al., Society for Industrial and Applied Mathematics. Philadelphia, Pa.

LOBOLIB (1985), Mathematical Software Library - C. Moler. Preliminary Edition. Department of Computer Science. University of New Mexico.

MARQUARDT, D.W. (1970) "Generalized Inverses, ridge regression, biased linear estimation and nonlinear estimation", Technometrics, Vol. 12, pp. 591-612.

MONTGOMERY D.C., E. PECK (1982). Introduction to Linear Regression Analysis. John Wiley & sons, New York.,

SCHEFFE, H. (1959). The Analysis of Variance. John Wiley & Sons, New York.

SEBER G.A.F. (1977). Linear Regression Analysis. John Wiley & Sons, New York.

SILVEY S.D. (1969). "Multicollinearity and imprecise estimation". Journal of Royal Statistical Society, Ser. B., Vol. 31, pp. 539-552.

STEWART, G.W. (1973). Introduction to Matrix Computation. Academic Press. New York.

THEIL, H., and A.S. GOLDBERGER (1961). "On pure and mixed statistical estimation in economics", International Economics Review, Vol. 2, pp. 65-78.

THEIL, H. (1971). Principles of Econometrics. John Wiley & Sons: New York.

WICHERS C. R. (1975). "The detection of Multicollinearity: A comment". The Review of Economics and Statistics, Vol. 57, pp. 366-368.

WILLIAN A. R., D. G. WATTS (1978). "Meaningful multicollinearity measures". Technometrics, Vol. 20, pp. 407-412.

ZELLNER, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, New York.