
COLINEALIDAD: PELIGROSA PATOLOGIA DEL MODELO DE REGRESION LINEAL

Roberto Behar Gutierrez
Profesor Asociado
Depto. de Información y Sistemas
universidad del Valle
Cali - Colombia

RESUMEN

Cuando la matriz X que contiene la información sobre las variables predictoras en el modelo de regresión, presenta (cuasi) dependencia lineal entre sus vectores columna, las estimaciones de los parámetros por el método de mínimos cuadrados son muy pobres y en ocasiones carecen de sentido de acuerdo con el fenómeno que se estudia. Las predicciones son pésimas en algunas zonas de la región de predicción. Este artículo explica algunos elementos matemáticos a través de los cuales puede diagnosticarse la presencia de esta patología y la magnitud de su influencia en los procesos de estimación y predicción, igualmente esboza algunas soluciones.

INTRODUCCION

El modelo de regresión lineal es una de las herramientas estadísticas más popularizadas, más aún cuando en la actualidad proliferan toda clase de paquetes computacionales que permiten estimar los parámetros del modelo de regresión con facilidades de tipo operativo al alcance de profesionales o técnicos de todas las áreas, inclusive aquellas tradicionalmente consideradas ajenas a la matemática. Este hecho tan positivo encierra sin embargo un peligro paradójico, sobre todo cuando se tiene un respeto casi sagrado por los resultados que produce un computador y se descuida la verificación de las condiciones teóricas para las cuales el modelo es válido o eficiente.

La patología de colinealidad en el modelo de regresión es uno de esos casos en los cuales el paquete estadístico puede arrojar un conjunto de resultados que pueden carecer de sentido o de confiabilidad sin que el usuario tenga una señal explícita de estos hechos.

Uno de los objetivos de este escrito es llamar la atención sobre lo delicado de esta situación y advertir sobre la naturaleza de los peligros que pueden presentarse al ignorar la colinealidad en el modelo de regresión.

No obstante que en el desarrollo de este artículo aparecen algunos elementos matemáticos que no están al alcance de todos los usuarios del modelo de regresión, las secciones que tratan sobre los daños que puede causar la colinealidad y las causas de las mismas han sido escritas para todos los usuarios del modelo.

En cuanto a notación, se representarán los vectores con letra en negrilla.

2. ALGUNAS DEFINICIONES Y RESULTADOS UTILES

A continuación se presentan algunos resultados sobre matrices y vectores, que serán usados más adelante.

Definición 1. (Dependencia Lineal)

Se dice que un conjunto de vectores $w_1, w_2, \dots, w_p$ son linealmente dependientes, si y sólo si existen escalares $v_1, v_2, \dots, v_p$, no todos iguales a cero tal que:

$$v_1 w_1 + v_2 w_2 + \dots + v_p w_p = 0$$

Nota 1.

Si $v_1 w_1 + v_2 w_2 + \dots + v_p w_p \approx 0$, es decir es aproximadamente igual al vector cero, diremos que existe "cuasi-dependencia lineal" entre los vectores $w_1, w_2, \dots, w_p$.

Definición 2. (Espacio de columnas de una matriz).

Sea la matriz $X = (x_1, x_2, \dots, x_p)$ donde x_j es un vector columna de $\mathbb{R}^n$.

Sea $v' = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$.

Se define espacio de columnas de la matriz X como:

$$C(X) = \{z / z = Xv = \sum_{j=1}^p v_j w_j ; v \in \mathbb{R}^p\}$$

Resultado 1.

$$C(X) = C(X'X)$$

Nota 2.

Observe que el rango de la matriz X es igual al rango de $X'X$.

Resultado 2.

Sea A una matriz simétrica de orden p con valores propios $l_1, l_2, \dots, l_p$, entonces el determinante de la matriz A , puede expresarse como $|A| = l_1, l_2, \dots, l_p$.

Resultado 3.

Si los vectores columna de una matriz cuadrada $A = (a_1, a_2, \dots, a_p)$ son linealmente dependientes entonces su determinante es cero, es decir $|A| = 0$.

Resultado 4. (descomposición espectral de una matriz simétrica).

Sea A una matriz simétrica de orden p , entonces A puede descomponerse como:

$$A = V L V' \quad (R4.1)$$

donde $V = (v_1, v_2, \dots, v_p)$ es una matriz $p \times p$ ortogonal, es decir $V'V = I_p$ o sea que los vectores $v_1, v_2, \dots, v_p$ son ortonormales y constituyen vectores propios de A , asociados con los valores propios $l_1, l_2, \dots, l_p \geq 0$ además $L = \text{diag}(l_1, l_2, \dots, l_p)$.

Desarrollando el producto de matrices expresado en (R4.1) se obtiene:

$$A = \sum_{j=1}^p l_j v_j v_j' \quad (R4.2)$$

además si A^{-1} existe:

$$A^{-1} = \sum_{j=1}^p l_j^{-1} v_j v_j' \quad (R4.3)$$

Resultado 5.

En el modelo de regresión lineal:

$$y = X\beta + \varepsilon, \text{ donde } \varepsilon \sim (0, \sigma^2 I)$$

si X es de orden $n \times p$ y tiene rango $p \leq n$, el estimador de mínimos cuadrados (MC) es:

$$\hat{\beta} = (X'X)^{-1} X' Y \quad (R6.1)$$

La matriz de varianzas y covarianzas de $\hat{\beta}$ está dada por:

$$V(\hat{\beta}) = (X'X)^{-1} \sigma^2 \quad (R6.2)$$

como puede notarse de (R6.1) $\hat{\beta}$ es un estimador lineal en Y , además $\hat{\beta}$ es insesgado, es decir,

$$E(\hat{\beta}) = \beta \text{ para todo } \beta \in \mathbb{R}^p$$

además $\hat{\beta}$ es el mejor estimador lineal insesgado, desde el punto de vista de que ningún estimador lineal insesgado de $\lambda' \beta$ tiene menor varianza que:

$\hat{\lambda}'\hat{\beta}$, para cualquier $\lambda \in \mathcal{R}^p$.

$\hat{y} = \mathbf{x}'\hat{\beta}$ es el modelo que permite estimar $E[y]$ cuando las variables predictoras toman los valores que constituyen el vector $\mathbf{x}$.

3. QUE ES COLINEALIDAD ?

Se dice que un modelo de regresión padece la patología de colinealidad cuando los vectores columna de la matriz $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ son "cuasidependientes" linealmente en el sentido de la nota 1 de la definición 1, es decir si existen $x_1, x_2, \dots, x_p$ no todos nulos, tales que:

$$v_1\mathbf{x}_1 + v_2\mathbf{x}_2 + \dots + v_p\mathbf{x}_p \approx \mathbf{0} \quad (3.1)$$

Expresado de otra forma: Una variable predictora "casi" que puede expresarse en forma lineal en términos de otras variables predictoras.

Como puede colegirse, la colinealidad es más un problema de grado (mucho o poco) que de existencia (si hay o no hay); cuanto más cerca del vector $\mathbf{0}$ esté la combinación lineal en (3.1) mayor será el grado de colinealidad, hasta el punto de que cuando haya dependencia lineal exacta como lo establece la definición 1, se dirá que la "colinealidad es perfecta".

4. QUE DAÑOS PUEDE OCASIONAR LA COLINEALIDAD EN EL MODELO DE REGRESION ?

Cuando existe fuerte colinealidad, puede suceder que:

- Las varianzas de los estimadores de mínimos cuadrados de algunos de los coeficientes de regresión β_k son exageradamente grandes, lo cual se traduce en estimaciones tan pobres que pueden llegar a ser totalmente inútiles.
- Algunas estimaciones de los coeficientes de regresión tienden a asumir valores irracionalmente grandes.
- Los signos de los estimados de los coeficientes de regresión tienden a estar más determinados por las relaciones de colinealidad que por las relaciones entre las variables predictoras y la variable de respuesta, de esta manera el investigador encontrará, por ejemplo, que de acuerdo con el modelo cuando una variable predictora particu-

lar tiende a incrementar su valor, la variable de respuesta tiende a disminuir, contradiciendo la naturaleza del fenómeno.

- Las predicciones hechas con el modelo estimado no son confiables, son inestables en gran parte del espacio de predicción.
- Las pruebas de hipótesis sobre los parámetros del modelo no son confiables.

Cuando hay colinealidad perfecta, no operan algunos algoritmos para el cálculo de las estimaciones y generalmente los paquetes computacionales sacan un aviso de precaución, que hace que el investigador solicite ayuda especializada para que finalmente se detecte la patología. Sin embargo cuando la colinealidad no es perfecta, pero es fuerte, se producen resultados con el riesgo de que el usuario no detecte el problema con las consecuencias ya mencionadas.

5. POR QUE SE PRESENTA COLINEALIDAD EN EL MODELO DE REGRESION ?

Existen razones de distinta naturaleza que dan origen a la patología de la colinealidad, las más importantes son:

1. Algunas variables predictoras están relacionadas en el contexto del fenómeno que se estudia. Por ejemplo si se desea predecir el rendimiento académico (Y) en la franja de matemática en ingeniería, con base en los exámenes de conocimiento en matemáticas (X_1), aptitud matemática (X_2) de las pruebas de estado ICFES y del examen de conocimiento en matemáticas de la Universidad del Valle (X_3), es muy posible que los resultados X_1, X_2, X_3 estén relacionados en forma lineal, lo cual generaría colinealidad en la matriz X de variables predictoras.

La colinealidad en este caso es inherente al fenómeno, a la población de individuos.

2. Otro posible origen de la colinealidad, muy relacionado con el anterior, es el descuido en la definición e inclusión de variables predictoras, así por ejemplo si se desea predecir el rendimiento de un cultivo (y) con base en el contenido de nitrógeno orgánico (X_1), nitrógeno inorgánico (X_2), nitrógeno total (X_3) en el suelo.

En este caso se genera colinealidad perfecta, puesto que nitrógeno orgánico más nitrógeno inorgánico genera la variable nitrógeno total, es decir se da que: $X_1 + X_2 = X_3$

3. La patología de colinealidad puede también tener su origen muestral, a diferencia de las situaciones anteriores. Cuando el investigador tiene control sobre los niveles que deben tomar las variables predictoras, como es el caso de los diseños experimentales, por ejemplo puede definir las concentraciones de Nitrógeno (N) y de Potasio (K) que debe agregar a las distintas parcelas con el propósito de evaluar el rendimiento de cierto cultivo. Si se escogieran los niveles de N y K que se dan:

N (kg/Ha)	20	30	50	60	100
K(kg/Ha)	30	45	75	90	150

puede observarse que: $K = 1.5 N$, es decir:
 $K - 1.5 N = 0$.

En este caso la relación existente entre las variables predictoras se da a nivel de los valores en la muestra, que definió el propio investigador.

6. COMO DIAGNOSTICAR LA PATOLOGIA DE COLINEALIDAD EN EL MODELO DE REGRESION ?

Con base en el resultado 1, de la sección 2, puede notarse que la dependencia lineal de los vectores columna de la matriz X, puede estudiarse con base en los vectores columna de la matriz simétrica $X'X$. Uniendo esto con el resultado 3, puede afirmarse que si hay colinealidad perfecta en la matriz X, debe suceder que:

$$|X'X| = 0$$

Es decir que un valor "pequeño" del determinante de la matriz $X'X$ es un síntoma de colinealidad.

Puesto que: $|X'X| = l_1 \cdot l_2 \cdot \dots \cdot l_p$, donde los l_j son los valores propios de $X'X$ (Resultado 2), entonces:

$$|X'X| = 0 \Leftrightarrow l_j = 0 \text{ para algún } j \in \{1, 2, \dots, p\}$$

parece que valores propios de $X'X$ próximos a cero son indicadores de colinealidad, como se confirmará más adelante.

Nota 3.

En general, el valor del determinante de $X'X$ depende de la escala en que se expresen las variables predictoras, así por ejemplo si la variable estatura se expresara en kilómetros, se tendrá una columna de la matriz X con valores del orden de los milésimos, lo cual haría que el determinante de $X'X$ sea cercano a cero, sin que exista necesariamente relación entre las variables predictoras que definen las columnas de la matriz X, para evitar esto, se propone algún tipo de estandarización, una de las más usadas es la siguiente:

$$W_{ij} = \frac{X_{ij} - \bar{X}_i}{d_i}; \quad d_i^2 = \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

nótese que el correspondiente vector columna $W_i = (W_{i1}, W_{i2}, \dots, W_{im})'$ es tal que la media de sus componentes es cero. Además: $\|W_i\|^2 = W_i' W_i = 1$.

La matriz resultante de la estandarización, sin incluir la columna de "unos" correspondiente al término constante en el modelo, se llamará W.

Nótese que $W_i' W_j$ tiene la forma de un coeficiente de correlación entre X_i y X_j . Por tanto la matriz $W' W$ tiene la estructura de una matriz de correlación entre variables predictoras.

Nota 4.

El modelo de regresión lineal, planteado con base en la matriz de variables predictoras estandarizadas queda:

$$y = \alpha 1 + W\beta + \varepsilon; \quad V(\varepsilon) = \sigma^2 I$$

El estimador de mínimos cuadrados para α, β es:

$$\hat{\alpha} = \bar{y}; \quad \hat{\beta} = (W' W)^{-1} W' y$$

de tal forma que:

$$E(\hat{\alpha}) = \alpha; \quad E(\hat{\beta}) = \beta$$

Además:

$$V(\hat{\alpha}) = \frac{\sigma^2}{n}; \quad V(\hat{\beta}) = (W' W)^{-1} \sigma^2$$

$$V(\hat{\beta}) = Q \sigma^2 \quad (6.1)$$

donde $Q = (W' W)^{-1}$ es la inversa de la "matriz de correlaciones" entre variables predictoras.

Continuando el desarrollo de herramientas para el diagnóstico de colinealidad: Sean $l_j, V_j, j = 1, 2, \dots, p$ los valores y vectores propios de la matriz $W'W$, entonces:

$$(W'W)V_j = l_j V_j \quad \text{con} \\ \|V_j\|^2 = V_j'V_j = 1 \quad (6.2)$$

premultiplicando por el vector V_j' se obtiene:

$$V_j'(W'W)V_j = l_j V_j'V_j = l_j$$

es decir:

$$(WV_j)'(WV_j) = l_j$$

$$\|WV_j\|^2 = l_j, \text{ por lo tanto:}$$

Si $l_j \rightarrow 0$ entonces $\|WV_j\|^2 \rightarrow 0$, lo cual implica que:

$$WV_j \approx 0 \quad (6.3)$$

Si $W = (W_1, W_2, \dots, W_p)$; $V_j = (V_{j1}, V_{j2}, \dots, V_{jp})'$ donde los W_j son los vectores columna de la matriz W , la expresión (6.3) puede reescribirse como:

$$V_{j1}W_1 + V_{j2}W_2 + \dots + V_{jp}W_p \approx 0 \quad (6.4)$$

como de (6.2) se tiene que :

$$\sum_{j=1}^p V_{ji}^2 = 1$$

entonces la combinación lineal en (6.4) tiene al menos un $V_{ji} \neq 0$, lo cual implica colinealidad en la matriz W .

Además, dado que $\|V_j\| = 1$, entonces puede detectarse las variables W_j más involucradas con la patología, al observar en (6.4) aquellas W_j que posean coeficientes V_{ji} de mayor magnitud.

Algunos otros elementos para el diagnóstico se tratarán en las siguientes secciones cuando se presenten los efectos de la colinealidad.

7. LA COLINEALIDAD, COMO AFECTA A LA VARIANZA DE LOS ESTIMADORES DE MINIMOS CUADRADOS ?

Searle (1971, pp 46) y Graybill (1976, pp 19), demuestran la validez de la siguiente expresión para los elementos de la diagonal de la inversa de la matriz $W'W$

de correlaciones entre las variables predictoras, así:

Si $(W'W)^{-1} = Q = (q_{ij})$, entonces:

$$q_{jj} = [W_j'W_j - W_j'D_j(D_j'D_j)^{-1}D_j'W_j]^{-1} \quad (7.1)$$

donde D_j es la matriz W sin la columna W_j .

nótese que:

$$W_j = D_j\gamma + \epsilon^* \quad (7.2)$$

es un modelo lineal que pretende expresar la variable predictora W_j en términos de las restantes variables predictoras.

El estimador de mínimos cuadrados para el parámetro γ está dado por:

$$\hat{\gamma} = (D_j'D_j)^{-1}D_j'W_j$$

Las sumas de cuadrados asociados al modelo estimado son:

$$SCT = W_j'W_j = 1 \quad (\text{Ver nota 3, sección 6})$$

$$SCR = W_j'D_j(D_j'D_j)^{-1}D_j'W_j$$

El coeficiente de determinación R_j^2 para el modelo (7.2) es:

$$R_j^2 = \frac{SCR}{SCT} = W_j'D_j(D_j'D_j)^{-1}D_j'W_j \quad (7.3)$$

Reemplazando en (7.1) se obtiene:

$$q_{jj} = (1 - R_j^2)^{-1} \quad (7.4)$$

Recordando la expresión (6.1) para el modelo:

$$y = \alpha_1 + W\beta + \epsilon \quad \text{se tiene que:}$$

$$V(\hat{\beta}) = (W'W)^{-1}\sigma^2$$

es decir que:

$$V(\hat{\beta}_j) = q_{jj}\sigma^2 \quad (7.5)$$

Al coeficiente q_{jj} se le denomina factor de inflación de varianza para β_j .

Nótese que cuando la variable W_j está relacionada en forma lineal con las restantes variables predictoras se da

que $R_j^2 \rightarrow 1$, por lo tanto $q_{jj} \rightarrow \infty$, es decir :

$$V(\hat{\beta}_j) \rightarrow \infty$$

- Fuerte colinealidad genera varianzas muy grandes en los estimadores de mínimos cuadrados.

Nota 5.

Cuando la variable W_j no está involucrada en colinealidad, entonces $R_j^2 \rightarrow 0$, por lo tanto $q_{jj} \rightarrow 1$, es decir :

$V(\hat{\beta}_j) \rightarrow \sigma^2$, así el valor mínimo que puede tomar un factor de inflación de varianza es la unidad.

Note que si W_j está relacionada con las restantes variables de tal forma que $R_j^2 = 0.75$, entonces $q_{jj} = 4$ y :

$$V(\hat{\beta}_j) = 4 \sigma^2 \text{ si } R_j^2 = 0.90, \text{ entonces } q_{jj} = 10 \text{ y}$$

$$V(\hat{\beta}_j) = 10 \sigma^2.$$

De esta manera los factores de inflación de varianza q_{jj} pueden usarse para el diagnóstico de colinealidad.

Nota 6.

Usando el resultado 4 de la sección 2, puede escribirse:

$$Q = (W'W)^{-1} = \sum_{k=1}^p l_k^{-1} V_k V_k' \quad (7.6)$$

de donde:

$$q_{jj} = \sum_{k=1}^p l_k^{-1} V_{kj}^2 \quad (7.7)$$

Nótese que si la variable W_j no está involucrada en colinealidad, entonces $V_{kj} \rightarrow 0$ para todo k , así que aunque $l_k \rightarrow 0$, hay tendencia a neutralizarse los efectos en (7.7) haciendo que el factor de inflación de varianza q_{jj} no sea muy grande.

Por el contrario si la variable W_j estuviese involucrada en fuerte colinealidad, debe suceder que algún l_k sea muy pequeña, cuando su correspondiente V_{kj} es grande, lo cual se traduce en un factor de inflación de varianza q_{jj} grande.

En resumen: Variables involucradas en colinealidad tenderán a poseer sus coeficientes estimados con altos factores de inflación de varianza.

Por otro lado de (7.6) puede escribirse:

$$q_{jr} = \sum_{k=1}^p l_k^{-1} V_{jk} V_{rk} \quad (7.8)$$

Si las variables W_j y W_r están involucradas en una colinealidad, sus coeficientes V_{jk} y V_{rk} estarán lejos de cero y si la colinealidad es fuerte l_k debe estar cercano a cero, produciéndose un valor q_{jr} grande, por lo tanto:

$$\text{COV}(\hat{\beta}_j, \hat{\beta}_r) = q_{jr} \sigma^2 \text{ será grande.}$$

Resumiendo: Estimadores de coeficientes de regresión asociados con variables relacionadas en colinealidad pueden tener grande su covarianza.

8. LA COLINEALIDAD, COMO PUEDE AFECTAR LAS ESTIMACIONES DE MINIMOS CUADRADOS ?

De la nota 4 en la sección 6, se sabe que:

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

$$\hat{\beta} = \left[\sum_{j=1}^p l_j^{-1} V_j V_j' \right] W'Y \text{ (usando el resultado 4)}$$

$$\hat{\beta} = \sum_{j=1}^p l_j^{-1} C_j V_j, \text{ con } C_j = V_j' W'Y$$

así:

$$\hat{\beta}_k = \sum_{j=1}^p l_j^{-1} C_j V_{jk} \quad (8.1)$$

Si la variable W_k está involucrada en por lo menos una colinealidad, entonces V_{jk} estará lejos de cero, lo cual podría traducirse en un valor grande para $\hat{\beta}_k$, siempre y cuando C_j no esté cercano a cero.

Resumiendo: Variables involucradas en colinealidad pueden presentar estimaciones grandes de sus coeficientes de regresión respectivos.

Nota 7.

Si un valor propio, digamos l_1 , es mucho más pequeño que todos los demás, es decir:

si $l_1 \ll l_2 < l_3 < \dots < l_p$, entonces con base en (8.1) puede decirse en forma aproximada:

$$\hat{\beta}_k \approx l_1^{-1} C_1 V_{1k} \quad (8.2)$$

Si la variable W_k está involucrada en la colinealidad asociada con l_1 , entonces V_{jk} no está cerca de cero, por lo tanto $\hat{\beta}_k$ tiende a ser grande y además su signo estará determinado por el signo de V_{jk} (que no contiene ninguna información sobre la variable de respuesta).

9. LA COLINEALIDAD, COMO PUEDE AFECTAR LAS PREDICCIONES ?

En esta sección se usará el resultado sobre descomposición singular de una matriz, el cual se enuncia a continuación:

Sea X una matriz de orden $n \times p$ y de rango $r \leq p$, existen matrices U , D , V tales que:

$$X = UDV^T;$$

$$D = \text{diag}(\theta_1, \theta_2, \dots, \theta_r) \text{ con } \theta_k > 0 \text{ para todo } k$$

además:

$$U^T U = I_r; \quad V^T V = I_r$$

$\theta_k = l_k^{1/2}$, $k = 1, 2, \dots, r$; donde los l_k son los valores propios no nulos de $X^T X$.

Las columnas de la matriz V son r vectores propios unitarios de $X^T X$ asociados con los valores propios no nulos $l_1, l_2, \dots, l_r$.

Las columnas de la matriz U son r vectores propios unitarios de XX^T asociados con los valores propios no nulos $l_1, l_2, \dots, l_r$.

De acuerdo con el resultado anterior, el modelo:

$$Y = X\beta + \varepsilon; E[\varepsilon] = 0; V(\varepsilon) = \sigma^2 I \quad (9.1)$$

puede escribirse como:

$$Y = UDV^T \beta + \varepsilon \quad (9.2)$$

Haciendo la reparametrización:

$$\alpha = DV^T \beta \quad (9.3)$$

El modelo en (9.1) se plantea en forma equivalente como:

$$Y = U\alpha + \varepsilon; E[\varepsilon] = 0; V(\varepsilon) = \sigma^2 I \quad (9.4)$$

Nota 8.

- Nótese que si $r(X) = r < p$, entonces las columnas de X son linealmente dependientes, es decir existe colinealidad perfecta. Como $(p - r)$ valores propios son cero, habrá $(p - r)$ combinaciones lineales, no equivalentes, de las columnas de X , que generan el vector 0 .
- Nótese que el vector β tiene p parámetros $\beta_1, \beta_2, \dots, \beta_p$, mientras que el vector α contiene sólo r parámetros $\alpha_1, \alpha_2, \dots, \alpha_r$.
- Note que el espacio $C(X)$, generado por los vectores columna de la matriz X es de dimensión r . Una base ortonormal para $C(X)$ la constituyen los vectores columna de la matriz U ; por lo tanto todo vector $x \in C(X)$ tiene su expresión en términos de la base U .
- El modelo (9.4) expresa a Y en términos de r predictores ortogonales $u_1, u_2, \dots, u_r$.
- Nótese que estimando α puede realizarse la predicción correspondiente a un punto x , hallando su correspondiente expresión u en términos de la base en U .

El estimador de mínimos cuadrados del parámetro α en (9.1) es:

$$\hat{\alpha} = (U^T U)^{-1} U^T Y = U^T Y \quad (9.5)$$

$$E(\hat{\alpha}) = U^T E(Y) = U^T U \alpha = \alpha$$

$$V(\hat{\alpha}) = U^T U \sigma^2 I = \sigma^2 I \quad (9.6)$$

Es decir que en el modelo reparametrizado todos los estimadores α_k tienen igual varianza σ^2 y además son no correlacionados.

De (9.5) y la descomposición singular de X , se llega a:

$$\hat{\alpha} = DV^T \hat{\beta} \quad (9.7)$$

donde $\hat{\beta}$ es una solución de las ecuaciones normales $X'X\hat{\beta} = X'Y$.

Recordando que $\hat{\alpha}$ tiene r componentes y el vector $\hat{\beta}$ tiene p componentes ($r \leq p$), puede considerarse a (9.7) como un sistema de r ecuaciones con p incógnitas:

$$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p.$$

Si $r < p$, el sistema (9.7) tiene infinitas soluciones para $\hat{\beta}$, una de ellas podría obtenerse dando valores arbitrarios a $(p - r)$ de las incógnitas $\hat{\beta}_k$ y obteniendo la solución para las r restantes, lo cual resulta un verdadero adfeso, puesto que de esta forma los coeficientes estimados $\hat{\beta}_k$, en el modelo (9.1) dependería de la arbitrariedad como se escogieran los $(p - r)$ valores para solucionar (9.7). Veamos esto en más detalle:

Si en (9.7) se hace $Z = DV'$, entonces:

$$\hat{\alpha} = Z\hat{\beta}$$

Se hace ahora la partición de $\hat{\beta}$ y de Z así:

$$\hat{\alpha} = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}_{r \times r \quad r \times (p-r)} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad (9.8)$$

de tal manera que $\hat{\beta}_1$ contenga r componentes y $\hat{\beta}_2$ las restantes $(p - r)$, de esta manera:

$$\hat{\alpha} = Z_1 \hat{\beta}_1 + Z_2 \hat{\beta}_2 \quad (9.9)$$

De donde:

$$\hat{\beta}_1 = -Z_1^{-1} Z_2 \hat{\beta}_2 + Z_1^{-1} \hat{\alpha} \quad (9.10)$$

Supongamos ahora que se desea realizar la predicción en el punto x , es decir:

$$\hat{Y} = x' \hat{\beta} = (x'_1 \quad x'_2) \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad (9.11)$$

donde (x_1, x_2) es una partición del vector x inducida por la partición de $\hat{\beta}$ así:

$$\hat{Y} = x'_1 \hat{\beta}_1 + x'_2 \hat{\beta}_2 \quad (9.12)$$

reemplazando (9.10) en (9.12) y factorizando se obtiene:

$$\hat{Y} = (-x'_1 Z_1^{-1} Z_2 + x'_2) \hat{\beta}_2 + x'_1 Z_1^{-1} \hat{\alpha} \quad (9.13)$$

Como $\hat{\beta}_2$ es fijado en forma totalmente arbitraria, este significa que la predicción resultaría arbitraria, excepto cuando el punto $x' = (x'_1, x'_2)$ donde se desea la predicción, satisface que el coeficiente de $\hat{\beta}_2$ en (9.13) sea nulo, es decir que:

$$-x'_1 Z_1^{-1} Z_2 + x'_2 = 0 \quad (9.14)$$

Esta expresión está definiendo el lugar geométrico de los puntos x donde la predicción es confiable, a pesar de que la matriz X presenta la patología de colinealidad.

Teniendo en cuenta que $Z = DV'$ puede mostrarse como (9.14) respresenta las $(p - r)$ relaciones de dependencia lineal entre las variables predictoras, que están asociadas con los $(p - r)$ valores propios iguales a cero. Es decir que la región de puntos x en donde la predicción es confiable es el lugar geométrico definido por las dependencias lineales existentes en la matriz X .

A continuación se trata la situación en la cual no hay colinealidad perfecta, pero si se presenta cuasi-dependencia lineal. En este caso el rango de X es p .

$X = UDV'$ es decir:

$$XV = UD; \quad D = \text{diag}(\theta_1, \theta_2, \dots, \theta_p)$$

Supóngase que $\theta_1 \geq \theta_2 \geq \dots \geq \theta_p > 0$, además supóngase que los últimos $p - r$ son muy pequeños, con base en los cuales se realiza partición de las matrices D , V , así:

$$X(V_1 \quad V_2) = U \begin{bmatrix} D_1 & O \\ O & D_2 \end{bmatrix} \quad (9.15)$$

$$[XV_1 \quad XV_2] = \left[U \begin{pmatrix} D_1 \\ O \end{pmatrix} \quad U \begin{pmatrix} O \\ D_2 \end{pmatrix} \right] \quad (9.16)$$

De donde se deduce que si los valores singulares en D_2 son cercanos a cero, entonces:

$$XV_2 \approx 0 \quad (9.17)$$

Las ecuaciones en (9.17) están planteando las colinealidades existentes.

En cuanto a la incidencia sobre la varianza de las predicciones, téngase en cuenta que:

$$\hat{y} = u' \hat{\alpha} \quad (9.18)$$

donde u es el punto x expresado en términos de la base ortonormal constituida por las columnas de la matriz U que resulta en la descomposición singular de la matriz X , así que:

$$\begin{aligned} x' &= u' DV \\ u' &= x' VD^{-1} \end{aligned} \quad (9.19)$$

Calculando la varianza de $\hat{y}$ en (9.18)

$$\begin{aligned} V(\hat{y}) &= u' V(\hat{\alpha}) u = \sigma^2 u' u \\ V(\hat{y}) &= \sigma^2 \sum_{j=1}^p u_j^2 \end{aligned} \quad (9.20)$$

Es decir que la varianza de la predicción es grande para valores de x que tienen asociados vectores u con componentes grandes en magnitud. Surge la pregunta:

En qué lugar geométrico de los x se generan vectores u con componentes grandes?

Para responder esta pregunta usamos (9.19) y la partición usada en (9.16), de esta manera:

$$\begin{aligned} u' &= x' VD^{-1} = x' (V_1 \ V_2) \begin{bmatrix} D_1^{-1} & 0 \\ 0 & \bar{D}_2^{-1} \end{bmatrix} \\ &= (x' V_1 \quad x' V_2) \begin{bmatrix} D_1^{-1} & 0 \\ 0 & \bar{D}_2^{-1} \end{bmatrix} \end{aligned} \quad (9.21)$$

De esta expresión puede deducirse que si los valores singulares de la diagonal de D_2 son cercanos a cero, entonces las componentes respectivas del vector u serán grandes, excepto cuando:

$$x' V_2 \approx 0 \quad (9.22)$$

Nótese que la expresión (9.22) expresa las cuasidependencias lineales entre las variables predictoras,

puesto que las columnas de la matriz V_2 son los vectores propios de $X'X$ asociados con valores propios pequeños.

La conclusión es que la confiabilidad en las predicciones se deteriora a medida que se realiza en puntos x que se alejan del lugar geométrico determinado por la patología de colinealidad expresado en (9.22).

10. CUAL ES LA TERAPIA PARA COMBATIR LOS EFECTOS DE LA COLINEALIDAD?

El tratamiento contra esta patología está muy relacionado con su origen, de acuerdo a lo expresado en la sección 5, como se muestra a continuación:

Si la colinealidad se origina por la alta relación lineal que existe entre las variables en el contexto del fenómeno que se estudia, eliminar alguna(s) de las variables relacionadas podría no ser una buena solución puesto que se introduce sesgo en el modelo.

Supóngase que el modelo:

(10.1) $Y = X\beta + \epsilon$ es apropiado, es decir:

$E[\epsilon] = 0$ o equivalentemente $E[y] = X\beta$ al hacer una partición de la matriz X y su correspondiente vector β , se tiene que:

$$\begin{aligned} Y &= [x_1 \ x_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon \\ Y &= X_1\beta_1 + X_2\beta_2 + \epsilon \end{aligned} \quad (10.2)$$

Si las variables en X_2 están relacionadas con las variables en X_1 y se decidiera eliminar del modelo las predictoras correspondientes a X_2 , el modelo se expresaría como:

$$y = X_1 \beta_1 + \epsilon^* \quad (10.3)$$

En el modelo (10.3) resulta que: $E[\epsilon^*] \neq 0$ es decir que $E[y] \neq X\beta_1$ presentándose así falta de ajuste que será más severa entre mayor sea la importancia de X_2 en la explicación de Y .

Recordando que la colinealidad ataca a los estimadores de mínimos cuadrados, se recomienda en esta situación usar métodos como el de Componentes

Principales y el método Ridge, que hacen parte de los llamados métodos de estimación sesgada. Estos métodos producen estimadores con sesgo pero en cambio reducen grandemente su varianza sobre todo cuando se presenta fuerte colinealidad.

Si estamos ante la situación de colinealidad perfecta, como se plantea en (5.2), la solución es eliminar las variables redundantes como se muestra en seguida, retomando el modelo (10.2).

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon; E(\varepsilon) = 0$$

Se tendría en este caso que:

$$X_2 = X_1 C \quad (10.4)$$

Reemplazando (10.4) se tendría:

$$\begin{aligned} Y &= X_1\beta_1 + X_1C\beta_2 + \varepsilon \\ Y &= X_1(\beta_1 + C\beta_2) + \varepsilon \\ Y &= X_1\beta^* + \varepsilon \end{aligned} \quad (10.5)$$

Como se deduce del (10.5) expresando el modelo en términos de X_1 solamente, no se produce falta de ajuste, puesto que en el proceso descrito la variable ε no fue alterada.

Cuando la colinealidad es de origen muestral como se explicó en la sección 5.3 una buena solución es aumentar la muestra incluyendo puntos cuyos valores x se alejen de la relación de colinealidad, haciendo lo más grande posible la dispersión de dichos puntos.

Es conveniente recordar que si el objetivo del modelo de regresión es la predicción, en cualquiera de las situaciones la predicción será confiable para puntos x que satisfacen las condiciones de colinealidad. Esto es muy confortable sobre todo en la situación en que la colinealidad es inherente a la población, puesto que generalmente los individuos a los cuales se desea hacer la predicción de $E(y)$, guardan en su x correspondiente aproximadamente la misma relación que los datos usados en la construcción del modelo.

11. ALGUNAS ACOTACIONES DE TIPO PRACTICO

- La matriz $W'W$ que tiene la estructura de una matriz de correlaciones entre las variables predictoras, es un instrumento preliminar muy sencillo para detectar algunas posibles colinealidades. En la

práctica correlaciones con magnitud superior a 0.7 son indicios de colinealidad.

- Aun cuando todas las correlaciones sean bajas, sin embargo puede existir colinealidad puesto que la dependencia puede involucrar más de dos variables.
- En la sección 6, que trató sobre el diagnóstico de colinealidad se concluyó que valores propios de $X'X$ "pequeños" son indicadores de la patología, sin embargo no se acotó el significado de "pequeño" lo cual es de gran interés práctico.

Existe todo un desarrollo teórico en análisis numérico acerca de lo que se conoce como matrices "mal condicionadas" en el sentido en que pequeños cambios en sus componentes generan grandes cambios en algunas características, como por ejemplo las componentes de su matriz inversa. En lo que es pertinente a nuestro interés en este artículo, puede mencionarse un criterio para definir el grado de mal condicionamiento de una matriz, es aquel llamado número condición definido para una matriz A como:

$$N.C(A) = \sqrt{\frac{\max(l_k)}{\min(l_k)}}$$

donde los l_k son los valores propios de A .

Con base en este criterio surge un indicador que va a ser útil para decidir qué tan pequeño es un valor propio, este es el llamado **Índice de condicionamiento**:

$$I.C_j = \sqrt{\frac{\max(l_k)}{l_j}}$$

En la práctica se han definido en forma empírica algunas guías para interpretar este índice de condicionamiento así:

- $I.C_j \leq 10$ indican dependencias pobres.
- $10 < I.C_j \leq 100$ indican dependencias moderadas.
- $I.C_j > 100$ fuertes dependencias.

Nótese que al conceptuar sobre el tamaño de un valor propio, este será pequeño o grande en relación con la magnitud del mayor valor propio y no de su valor en términos absolutos.

- Algunos paquetes estadísticos tienen procedimientos para el diagnóstico de colinealidad y proporcionan indicadores como la matriz de correlaciones entre variables predictoras, factores de inflación de varianza, valores y vectores propios, índice de condicionamiento.

CONCLUSIONES

- La colinealidad en el modelo de regresión es un fenómeno que atenta contra la confiabilidad del modelo que se estima usando el método de mínimos cuadrados.
- Es muy importante ser acucioso con respecto a los valores estimados de los parámetros que arrojan los paquetes computacionales. Se debe estar atento a observar valores de estimados demasiado grandes o con signo que contradicen el conocimiento que se tiene del fenómeno que se estudia, igualmente pueden aparecer varianzas grandes para los estimadores de los parámetros del modelo. Si esto ocurre, debe examinarse la situación tratando de identificar su origen y adoptar la terapia pertinente.
- Debe tenerse presente que aún cuando exista fuerte colinealidad, las predicciones son confiables en puntos que pertenecen al lugar geométrico definido por las relaciones de colinealidad.
- Cuando sea pertinente deben considerarse métodos alternativos de estimación como los métodos de estimación sesgada, los cuales pueden producir grandes disminuciones en la varianza a cambio de algún sesgo pequeño en las estimaciones.
- No debe convertirse el problema del modelaje estadístico en un ejercicio de cómputo. Debe considerarse aspectos como la verificación de los supuestos y restricciones bajo las cuales fue deducido el modelo, además de las posibles patologías.

BIBLIOGRAFIA

- (1) BARNETT, V. and LEWIS T. "Outliers in Statistical Data" New York: John Wiley. 1978.
- (2) BELSLEY, D. A., KUH, E. and WELSH, R. E. "Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity". New York John Wiley. 1980.
- (3) BELSLEY, D. A., "Assessing the presence of harmful Collinearity and other form of Weak Data through a test for signal to Noise". Journal of Econometrics, 20, 211-253. 1982.
- (4) BELSLEY, D. A., "Demeaning Conditioning Diagnostics Through Centering". The American Statistician May 1984, Vol 38 No. 2, 73-93.
- (5) BERK, K. N. "Tolerance and Condition in Regression Computation". Journal of American Statistical Association. 72, 863-866. 1977.
- (6) BIBBY - TOUTENBURG. "Prediction and Improved Estimation in Linear Models". John Wiley and Akademie - Verlag Berlin GDR. 1977.
- (7) CHATTERJEE, S. and PRICE, B. "Regression Analysis by Example". New York: John Wiley. 1977.
- (8) DANIEL, C. and WOOD, F.S. "Fitting Equations to Data". New York: Wiley - Interscience. 1980.
- (9) DRAPER, N. R. and VAN NOSTRAND, R. C. "Ridge Regression and James-Stein Estimation: Review and Comments". Technometrics, 21, 451-466. 1979.
- (10) FARRAR, D. E. and GLAUBER, R. R. "Multicollinearity in Regression Analysis: The Problem Revisited". Review of Economics and Statistics, 49, 92 - 107. 1967.
- (11) GRAYBILL, F. A. "An Introduction to linear statistical Models". Vol. I. New York: Mc Graw-Hill. 1961.
- (12) GUNST, R. F. and MASON, R. L. "Advantages of Examining Multicollinearities in Regression Analysis." Biometrics, 33, 249 - 260, March 1977.
- (13) GUNST, R. F. and MASON, R. L. "Regression Analysis and its Applications: A Data Oriented Approach." New York: Marcel Dekker. 1980.
- (14) GUNST, R. F. "Regression Analysis with Multicollinear Predictor Variables: Definition, Detection and Effects." Communications in Statistics, A12 (Special Issue on Statistical Reviews). 2217-2260. 1983.
- (15) HOAGLIN, D. C. and WELSH, R. E. "The Hat Matrix in Regression and ANOVA." The American Statistician 32, 17-22. 1978.

- (16) HOCKING, R. R. and DUNN, M. R. " *Collinearity, Influential Data and Ridge Regression.*" Paper Presented at University of Delaware Symposium on Ridge Regression. 1982.
- (17) HOCKING, R. R. " *Developments in Linear Regression Methodology: 1959-1982.*" (With discussion) *Technometrics*, 25, 219- 249. 1983.
- (18) HOERL, A. E. and KENNARD, R. W. " *Ridge Regression: Biased Estimation for Non-Orthogonal Problems.*" *Technometrics* 12, 55-67. 1970.
- (19) HOERL, R. W. " *Ridge Analysis 25 Years Later.*" *The American Statistician*, Vol. 39, No. 3, 186-192. 1985.
- (20) KLEINBAUM, D. G. ; KUPPER L. L. ; MULLERKE " *Applied Regression Analysis and Other Multivariable Methods.*" PWS - KENT Publishing. Boston. 1988.
- (21) LAWSON, C. R. and HANSON, R. J. " *Solving Least - Squares Problems.*" Englewood Cliffs N. J. : Prentice - Hall. 1974.
- (22) MARQUARDT, D. W. and SNEE, R. D. " *Ridge Regression in Practice.*" *The American Statistician*, 29, 3-19. 1975.
- (23) MARQUARDT, D. W. " *You Should Standardize the Predictor Variables in Your Regression Models.*" *Journal of American Statistical Association*. 75, 74-103. 1980.
- (24) MARTINEZ, A. y CASTILLO A. " *Teoría de la Regresión con Aplicaciones Agronómicas.*" Colegio de Postgraduados, Chapingo, Mexico. 1987.
- (25) MASON, R. L.; GUNST, R. F. and WEBSTER, J. T. " *Regression Analysis and Problems of Multicollinearity.*" *Communications in Statistics*, 4, 277-292. 1975.
- (26) MASON, R. L. and GUNST, R. F. " *Outliers Induced Collinearities.*" *Technometrics*, November, Vol. 27 No. 4, 401-407. 1985.
- (27) MEYER, L. S. and WILKE, T. A. " *On biased Estimation in Linear Models.*" *Technometrics* 15, 497-508. 1973.
- (28) MULLET, G. M. " *Why Regression Coefficients Have the Wrong Sign.*" *Journal of Quality Technology*. 8, 121-126. 1976.
- (29) RENCHER, A. C. and PUN, F. C. " *Inflation on R-squared in Best Subset Regression.*" *Technometrics* 22, 49-54. 1980.
- (30) THEIL, H. " *Principles of Econometrics.*" New York: John Wiley. 1980.
- (31) WEISBERG, S. " *Applied Linear Regression.*" New York: John Wiley. 1980.
- (32) WILLAN, A. R. and WATTS, D. G. " *Meaningful Multicollinearity Measures.*" *Technometrics* 20, 407-412. 1978.
- (33) ZARATE, G. P. y ALVAREZ, O. " *Aplicaciones de las Descomposiciones Singular y Espectral de una Matriz.*" *AGROCIENCIA*. No. 61. Colegio de Posgraduados, Chapingo, Mexico. 1985.
- (34) ZARATE, G. P. y DIAZ, B. L. " *Técnicas de Diagnóstico y Manejo de Puntos de Influencia y Multicolinealidad en Regresión Lineal.*" *AGROCIENCIA*. No. 61. Colegio de Posgraduados, Chapingo, Mexico. 1985.