DETERMINACIÓN DE LA CONSTANTE K EN LA REGRESIÓN RIDGE

MI Manuel R. Piña Monarrez[1], Dr. Manuel A. Rodríguez Medina[2], Dr. Juan J. Díaz Núñez[2]

ABSTRACT

Since the Ordinary Least Square (OLS), no have inside its structure an optimisation method for determine the effect that the multicollinearity has over the estimate coefficients of the vector $\hat{\boldsymbol{\beta}}$, permit to the Ridge Regression (RR) take an important roll in solve this problem. In this paper, we present the steps for the determination of the constant of proportionality K, with we obtain a smaller variance that this estimate by OLS, over the focus of the Total mean square of the Prediction (TMSP).

RESUMEN

Desde que el método de Mínimos Cuadrados (MC), no tiene dentro de su estructura un método de optimización para determinar el efecto que la multicolinealidad tiene sobre los coeficientes estimados del vector $\hat{\boldsymbol{\beta}}$, permite a la regresión Ridge (RR), tomar importancia en resolver este problema. En este artículo, presentamos el desarrollo de la determinación de la constante de proporcionalidad K, con la que obtenemos una varianza más pequeña que la estimada por MC, bajo el enfoque del Cuadrado Medio de la Predicción Total (CMPT).

Palabras Claves: Cuadrado Medio de la Predicción, Multicolinealidad, Regresión Ridge, Mínimos Cuadrados,

1. INTRODUCCIÓN

Desde que la varianza de los coeficientes estimados $\hat{\boldsymbol{\beta}}_j$ de MC es $\mathbf{V}(\hat{\beta}_j) = \sigma_0^2 \, 1/\lambda_j$, donde σ_0^2 es la usual estimación insesgada de la varianza poblacional dada por $\sigma_0^2 = \mathbf{Y}\left[\mathbf{I} - \mathbf{X}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}\right]\mathbf{Y}/(n-p)$ y λ_j es el j-eavo eigenvalor de la matriz simétrica $(\boldsymbol{X}^t\boldsymbol{X})$ para j=1,2,...,p, la estimación de los coeficientes $\hat{\boldsymbol{\beta}}_j$ de MC, puede ser muy mala a medida que la multicolinealidad entre los regresores crece, llegando hasta cambiar de signo el coeficiente, cuando se calcula de otra muestra (Montgomery, Peck y Vining 2002). Para este problema, Hoerl y Kennard, en 1970 (a y b), propusieron el método RR, el cual, es un método para detectar la multicolinealidad dentro de un modelo de regresión. La idea del método es simple y consiste en que dado que la matriz $\mathbf{X}^t\mathbf{X}$ es altamente condicionada o cercana a singular, es posible agregar constantes positivas a los elementos de la diagonal, para asegurar que la matriz resultante, no sea altamente condicionada. Esto es, considerar el sesgo a través de la ecuación normal dada por:

$$\boldsymbol{\beta}_R = (\boldsymbol{X}^t\boldsymbol{X} + \boldsymbol{k}^{\boldsymbol{\delta}})^{-1}\boldsymbol{X}^t\boldsymbol{Y} \qquad (1)$$

El valor de la constante k en (1), está dado por

$$K = \frac{\sigma^2}{\beta^t\beta} \qquad (2)$$

donde σ^2 y $\boldsymbol{\beta}$, son la varianza poblacional y el vector de coeficientes verdadero respectivamente, los cuales son desconocidos en el estudio (ver Hoerl y Kennard 1970(a) y Piña, Rodríguez y Díaz, 2005), por lo que K, deberá de ser aproximada. Hoerl y Kennard y Baldwin 1975, y Hoerl y Kennard 1976, propusieron un método para estimar K basado en las estimaciones de $\hat{\sigma}^2$ y de

[1] Instituto Tecnológico de Ciudad Juárez. Candidato a doctor en ingeniería industrial.
[2] Instituto Tecnológico de Ciudad Juárez. Programa del Doctorado en Ingeniería Industrial.

$\hat{\boldsymbol{\beta}}_j$ dadas por MC. Otros intentos de estimar K, pueden ser encontrados en Hemmerle 1975, Hemmerle y Brantle 1978, Rubio y Firinguetti 2002 y Golam 2003. En el presente artículo se desarrolla el método para la obtención de la constante K a través del enfoque del cuadrado medio de la predicción total (ver Lawless y Wang 1976). El artículo está estructurado de la siguiente forma, en la sección 2, se presenta la estructura de Mínimos Cuadrados (MC), en la sección 3, se presenta la estructura del método de Regresión Ridge (RR), la sección 4, presenta el método de obtención de K, la sección 5, presenta las conclusiones, el artículo termina con las referencias.

2. Mínimos Cuadrados (MC)

Desde que OLS es una proyección ortogonal sobre el espacio **X**, de un vector de respuestas **Y** del tipo $\mathbf{Y} = \mathbf{X}^t\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$ (3) donde

$$\mathbf{x} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1P} \\ 1 & X_{21} & X_{22} & \cdots & X_{2P} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n1} & \cdots & X_{nP} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, \ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

ε, es un vector de errores aleatorio con $E(\varepsilon) = 0$ y $E(\varepsilon^t\varepsilon) = \sigma^2 \mathbf{I_n}$ y $\hat{\beta}$, es un vector de px1 de coeficientes de regresión desconocidos dado por $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$, y $\mathbf{X^tX}$ es una matriz simétrica positiva definida ya que se asume que n>p y que **X** es una matriz no estocástica de rango p. El valor esperado de los coeficientes de regresión de MC, están dados por:

$$E(\hat{\beta}) = \mathbf{E}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{E}(\mathbf{Y})$$
$$E(\hat{\beta}) = \mathbf{E}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{X}\beta + \varepsilon)$$
$$E(\hat{\beta}) = \beta + (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\varepsilon = \beta \qquad (4)$$

De igual forma, el valor esperado de $\hat{\boldsymbol{\beta}}^t\hat{\boldsymbol{\beta}}$, está dado por:

$$E(\hat{\beta}^t\hat{\beta}) = \mathbf{E}\left[(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}\right]$$
$$E(\hat{\beta}^t\hat{\beta}) = \beta^t\beta + \sigma^2 \ {}^1\!/_{\lambda_j} \qquad (5)$$

De (5), se observa que cuando la multicolinealidad aumenta (es decir λ_j tiende a cero), la longitud de $\hat{\boldsymbol{\beta}}^t\hat{\boldsymbol{\beta}}$, tiende a infinito.

La varianza de $\hat{\beta}$, está dado por:

$$V(\hat{\beta}) = V\left[\beta + (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\varepsilon\right]$$
$$V(\hat{\beta}_j) = \sigma^2 \ {}^1\!/_{\lambda_j} \qquad (6)$$

De (6), se observa que cuando λ_j, tienda a cero, $V(\hat{\beta}_j)$, tiende a infinito, por lo que los estimados de $\hat{\beta}$ serán inestables, pudiendo hasta cambiar de signo cuando se determinan de diferente muestra (ver Montgomery, Peck y Vining 2002).

3. REGRESION RIDGE (RR)

El método de RR, es un método para detectar la multicolinealidad dentro de un modelo de regresión del tipo $\mathbf{Y} = \mathbf{X}^t\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$ como se definió en (3). El método fue propuesto por Hoerl y Kennard en 1970 y es usado para trabajar con modelos que presentan sesgo. La idea del método es simple y consiste en que dado que la matriz $\mathbf{X^tX}$ es altamente condicionada o cercana a singular, es posible agregar constantes positivas a los elementos de la diagonal, para asegurar que la matriz resultante, no sea altamente condicionada. El vector de coeficientes de RR, está dado por:

$$\boldsymbol{\beta}_\mathbf{R} = \left(\mathbf{X}^t\mathbf{X} + \mathbf{K}^\delta\right)^{-1}\mathbf{X}^t\mathbf{Y} \qquad (7)$$

el cual se puede rescribir como $\beta_\mathbf{R} = \mathbf{Z}\hat{\beta}$ donde $\hat{\beta}$, es el estimador ordinario de MC dado por $\hat{\beta} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y}$ y $\mathbf{Z} = \left(\mathbf{I} + \mathbf{K}^\delta(\mathbf{X}^t\mathbf{X})^{-1}\right)^{-1}$ es la matriz que transforma a $\hat{\beta}$ en $\beta_\mathbf{R}$, es decir $\beta_\mathbf{R}$, es la solución al problema de optimización:

$$\textbf{Min:}\ \left(\beta_\mathbf{R} - \hat{\beta}\right)\mathbf{X}^t\mathbf{X}\left(\beta_\mathbf{R} - \hat{\beta}\right)$$
$$\textbf{Sujeto a:}\ \beta_\mathbf{R}\beta_\mathbf{R} \le r^2$$

Prueba: A través de los multiplicadores de Lagrange, esta función es representada por:

$$\text{Min}:\left(\beta_{\mathbf{R}}-\hat{\beta}\right)\mathbf{X}^{t}\mathbf{X}\left(\beta_{\mathbf{R}}-\hat{\beta}\right)+\mathbf{K}^{\delta}\,(\beta_{\mathbf{R}}\beta_{\mathbf{R}}-r^{2})=0$$

derivando la función con respecto a $\beta_{\mathbf{R}}$ e igualándola a cero tenemos que:

$$\text{df/d}\beta_{\mathbf{R}}=2\mathbf{X}^{t}\mathbf{X}\left(\beta_{\mathbf{R}}-\hat{\beta}\right)+2\mathbf{K}^{\delta}\beta_{\mathbf{R}}$$

$$\mathbf{X}^{t}\mathbf{X}\boldsymbol{\beta}_{\mathbf{R}}-\mathbf{X}^{t}\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{K}^{\delta}\boldsymbol{\beta}_{\mathbf{R}}=0$$

$$\boldsymbol{\beta}_{\mathbf{R}}(\mathbf{X}^{t}\mathbf{X}+\mathbf{K}^{\delta})-\mathbf{X}^{t}\mathbf{X}\hat{\boldsymbol{\beta}}=0$$

dado que $\hat{\beta}$ está dada por $\hat{\beta}=\left(\mathbf{X}^{t}\mathbf{X}\right)^{-1}\mathbf{X}^{t}\mathbf{Y}$, entonces:

$$\boldsymbol{\beta}_{\mathbf{R}}(\mathbf{X}^{t}\mathbf{X}+\mathbf{K}^{\delta})-\mathbf{X}^{t}\mathbf{Y}=0$$

$$\boldsymbol{\beta}_{\mathbf{R}}=(\mathbf{X}^{t}\mathbf{X}+\mathbf{K}^{\delta})^{-1}\mathbf{X}^{t}\mathbf{Y}$$

$$\boldsymbol{\beta}_{\mathbf{R}}=\frac{\mathbf{X}^{t}\mathbf{Y}}{\left[\,\mathbf{I}+\mathbf{K}^{\delta}(\mathbf{X}^{t}\mathbf{X})^{-1}\right](\mathbf{X}^{t}\mathbf{X})}=\left[\,\mathbf{I}+\mathbf{K}^{\delta}(\mathbf{X}^{t}\mathbf{X})^{-1}\right]^{-1}(\mathbf{X}^{t}\mathbf{X})^{-1}\mathbf{X}^{t}\mathbf{Y}$$

$$\beta_{\mathbf{R}}=\mathbf{Z}\hat{\beta} \qquad (8)$$

lo cual completa la prueba.

En la estimación RR, al parámetro k_j, se le conoce como parámetro de sesgo, ya que cuando $\mathbf{K}^{\delta}=0$, $\mathbf{Z}=\mathbf{I}$, es decir $\beta_{\mathbf{R}}=\hat{\beta}$ y cuando $\mathbf{K}^{\delta}\neq 0$, $\beta_{\mathbf{R}}\neq\hat{\beta}$. Además como $\hat{\beta}$ es insesgado, entonces la estimación ridge, es una estimación sesgada. Afortunadamente aunque la estimación es sesgada, la varianza de la estimación de $\beta_{\mathbf{R}}$ a β es menor que la varianza de la estimación de $\hat{\beta}$ a β, por lo que los coeficientes de $\beta_{\mathbf{R}}$, son más estables que los de $\hat{\beta}$. Para ver por que, analizamos la función del cuadrado medio del error de $\beta_{\mathbf{R}}$, la cual está dada por:

$$\text{CME}\beta_{\mathbf{R}}=\sigma^{2}\sum\frac{\lambda_j}{(\lambda_j+k_j)^{2}}+k_{\mathbf{j}}^{2}\sum\frac{\beta_{\mathbf{j}}^{2}}{(\lambda_j+k_j)^{2}} \qquad (9)$$

(Ver Piña, Rodríguez y Díaz 2005). De esta función del CME tenemos:

Teorema 3.1: La varianza de $\beta_{\mathbf{R}}$ dada por $V(\beta_{\mathbf{R}})=\sigma^{2}\sum\frac{\lambda_j}{(\lambda_j+k_j)^{2}}$, es una función continua de k y monoticamente decreciente:

Prueba: Tomando la derivada con respecto a k cuando $k\to 0^{+}$ tenemos que $\lim\limits_{k\to 0^{+}}(\text{dv/dk})=-2\sigma^{2}\sum_{1}^{p}(1/\lambda_j^{2})$ y cuando $k\to\infty$ el $\lim\limits_{k\to\infty}(\text{dv/dk})=-\infty$, lo cual completa la prueba.

Colorario 3.1.1: La varianza de $\beta_{\mathbf{R}}$ es la varianza de $\hat{\beta}$, cuando K=0**.**

Prueba: La varianza de $\hat{\beta}$, está dada por $V(\hat{\beta})=\sigma^{2}\sum 1/\lambda_j$, por lo que sustituyendo K=0 en la función de varianza de $\beta_{\mathbf{R}}$, tenemos que $V(\beta_{\mathbf{R}})=\sigma^{2}\sum\frac{\lambda_j}{(\lambda_j+0)^{2}}=\sigma^{2}\sum 1/\lambda_j$. Lo cual completa la prueba.

Teorema 3.2: El $\textbf{Sesgo}^{2}=k_j^{2}\sum_{1}^{p}\beta_{\mathbf{j}}^{2}\,/(\lambda_j+k_j)^{2}$ de $\beta_{\mathbf{R}}$ es una función continua de K y monoticamente creciente:

Prueba: Para la función $\textbf{Sesgo}^{2}=k_j^{2}\sum_{1}^{p}\beta_{\mathbf{j}}^{2}\,/(\lambda_j+k_j)^{2}$, tenemos directamente que si k=0, entonces $\textbf{Sesgo}^{2}=0^{2}\sum_{1}^{p}\beta_{\mathbf{j}}^{2}\,/(\lambda_j+0)^{2}=0$. Además, desde que la función de sesgo puede ser re-escrita como

$\mathbf{Sesgo}^2 = \sum_{1}^{p} \beta_j^2 / \left[1 + (\lambda_j / k_j)\right]^2$ y dado que $\lambda_j > 0$ para toda j=1,2,…,p, y como la función λ_j / k_j, es una función monótica decreciente para valores positivos de (K>0), implica directamente que $\mathbf{Sesgo}^2$, es una función creciente de k, lo cual completa la prueba.

Colorario 3.2.1: El $\mathbf{Sesgo}^2$ de $\beta_{\mathbf{R}}$, tiene como límite superior a $\beta^t\beta$.

Prueba: Desde que $\mathbf{Sesgo}^2 = \sum_{1}^{p} \beta_j^2 / \left[1 + (\lambda_j / k_j)\right]^2$, es fácil verificar que el límite de la función $\mathbf{Sesgo}^2$ cuando $k \to \infty$, es $\lim_{k \to \infty} \mathbf{Sesgo}^2 = \boldsymbol{\beta}^t\boldsymbol{\beta}$, lo cual completa la prueba.

Colorario 3.2.2: La derivada de la función $\mathbf{Sesgo}^2$, de $\beta_{\mathbf{R}}$, se aproxima a cero cuando $k \to 0^+$.

Prueba: De la función $\mathbf{Sesgo}^2 = k_j^2 \sum_{1}^{p} \beta_j^2 / (\lambda_j + k_j)^2$, es fácil verificar que $ds/dk = 2k_j \sum_{1}^{p} \lambda_j \beta_j^2 / (\lambda_j + k_j)^3$, por lo que sustituyendo k=0, tenemos que $\lim_{k \to 0^+} (ds/dk) = 0$, lo cual completa la prueba.

Teorema 3.3: Siempre existe un valor de k>0 tal que $CME\,\beta_{\mathbf{R}} < CME\,\hat{\beta}$.

Prueba: Derivando la función $CME\,\beta_{\mathbf{R}} = \sigma^2 \sum_{1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2} + k_j^2 \sum_{1}^{p} \frac{\beta_j^2}{(\lambda_j + k_j)^2}$, tenemos que

$dCME\beta_{\mathbf{R}} / dk = -2\sigma^2 \sum_{1}^{p} \lambda_j / (\lambda_j + k_j)^3 + 2k_j \sum_{1}^{p} \lambda_j \beta_j^2 / (\lambda_j + k_j)^3$, Igualando la derivada a cero y resolviendo para k, tenemos que:

$$-2\sigma^2 \sum_{1}^{p} \lambda_j / (\lambda_j + k_j)^3 + 2k_j \sum_{1}^{p} \lambda_j \beta_j^2 / (\lambda_j + k_j)^3 = 0$$

$$-2\sigma^2 \sum_{1}^{p} \lambda_j + 2k_j \sum_{1}^{p} \lambda_j \beta_j^2 = 0$$

$$k = \sigma^2 / \beta_j^2 \qquad (10)$$

por lo que los valores de k que hacen que $CME\,\beta_{\mathbf{R}} < CME\,\hat{\beta}$ son los valores que están en el intervalo $0 < k < \sigma^2 / \beta_j^2$, con lo que la prueba está completa.

Observe que σ^2 y $\boldsymbol{\beta}_j^2$, son la varianza poblacional y los coeficientes verdaderos, los cuales son desconocidos en el estudio, por lo que K, también es desconocida y se tendrá que estimar.

4. Método para la estimación de K

Dado que el objetivo, es el lograr determinar un modelo polinomial cuya varianza de predicción, sea mínima, utilizamos la función del cuadrado medio de la predicción (CMP) de MC, para determinar el valor de la constante **K.**

Teorema 4.1 La función del Cuadrado Medio de la Predicción (CMP), está dada por $CMP = \left(\hat{\mathbf{Y}} - \mathbf{Y}\right)^2 = \sigma^2 + \sum_{1}^{p} \lambda_j \left(\hat{\beta} - \beta\right)^2$.

Prueba:

$CMP = \left(\hat{\mathbf{Y}} - \mathbf{Y}\right)^2 = \hat{\mathbf{Y}}\hat{\mathbf{Y}} - 2\hat{\mathbf{Y}}\mathbf{Y} + \mathbf{Y}\mathbf{Y}$ donde $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$ y $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$.

$$CMP = \mathbf{X}^t\mathbf{X}\hat{\beta}\hat{\beta} - 2\mathbf{X}\hat{\beta}(\mathbf{X}\beta + \varepsilon) + (\mathbf{X}\beta + \varepsilon)(\mathbf{X}\beta + \varepsilon)$$

$$CMP = \mathbf{X}^t\mathbf{X}\hat{\beta}\hat{\beta} - 2\mathbf{X}^t\mathbf{X}\hat{\beta}\beta - 2\hat{\beta}\mathbf{X}^t\varepsilon + \mathbf{X}^t\mathbf{X}\beta\beta + 2\beta\mathbf{X}^t\varepsilon + \varepsilon^t\varepsilon$$

$$CMP = \mathbf{X^tX}\hat{\beta}\hat{\beta} - 2\mathbf{X^tX}\hat{\beta}\beta + \mathbf{X^tX}\beta\beta + \sigma^2$$
$$CMP = \mathbf{X^tX}\hat{\beta}\hat{\beta} - 2\mathbf{X^tX(X^tX)^{-1}X^tY}\beta + \mathbf{X^tX}\beta\beta + \sigma^2$$
$$CMP = \mathbf{X^tX}\hat{\beta}\hat{\beta} - 2\beta\mathbf{X^t(X}\beta + \varepsilon) + \mathbf{X^tX}\beta^t\beta + \sigma^2$$
$$CMP = \sigma^2 + \text{Trazo}\mathbf{X^tX}(\hat{\beta} - \beta)^2$$
$$CMP = \sigma^2 + \sum_{1}^{p} \lambda_j (\hat{\beta} - \beta)^2 \qquad (11)$$

Lo cual completa la prueba.

Teorema 4.2 El valor de K obtenido de la función del Cuadrado Medio de la Predicción (CMP), está dado por $K = \dfrac{P\hat{\sigma}^2}{\sum_{1}^{p} \lambda_j \hat{\beta}^t\hat{\beta}} = \dfrac{P\hat{\sigma}^2}{\hat{\beta}^t(\mathbf{X^tX})\hat{\beta}}$.

Prueba: Tomando el valor esperado de $CMP = \sigma^2 + E\left[\sum_{1}^{p} \lambda_j (\hat{\beta} - \beta)^2\right]$ y asumiendo que $\beta \approx N(0, \sigma^2\mathbf{I})$, entonces:

$$\sigma^2 + E\left[\sum_{1}^{p} \lambda_j (\hat{\beta} - \beta)^2\right] = \sigma^2 + E\left[\sum_{1}^{p} \lambda_j (\hat{\beta} - 0)^2\right] = \sigma^2 + \sum_{1}^{p} \lambda_j\, E(\hat{\beta}^t\hat{\beta})$$

$$\sigma^2 + E\left[\sum_{1}^{p} \lambda_j (\hat{\beta} - \beta)^2\right] = \sigma^2 + \sum_{1}^{p} \lambda_j\, (\beta^t\beta + \sigma^2/\lambda_j)$$

$$E\left[\sum_{1}^{p} \lambda_j \hat{\beta}^t\hat{\beta}\right] = \sum_{1}^{p} \lambda_j\, \beta^t\beta + \sum_{1}^{p} \lambda_j\, \sigma^2/\lambda_j$$

$$\frac{E\left[\sum_{1}^{p} \lambda_j \hat{\beta}^t\hat{\beta}\right]}{\sigma^2} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta + \sum_{1}^{p} \lambda_j\, \sigma^2/\lambda_j}{\sigma^2} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta}{\sigma^2} + \frac{P}{\lambda_j}$$

$$\frac{E\left[\sum_{1}^{p} \lambda_j \hat{\beta}^t\hat{\beta}\right]}{\sum_{1}^{p} \lambda_j\, \sigma^2} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta}{\sum_{1}^{p} \lambda_j\, \sigma^2} + \frac{\sum_{1}^{p} \lambda_j}{\sum_{1}^{p} \lambda_j\, \lambda_j} = \frac{\beta^t\beta}{\sigma^2} + \frac{1}{\lambda_j}$$

$$\frac{\beta^t\beta}{\sigma^2} = \frac{E\left[\sum_{1}^{p} \lambda_j \hat{\beta}^t\hat{\beta}\right]}{P\sigma^2} - \frac{1}{\lambda_j} = \frac{\sum_{1}^{p} \lambda_j\, E(\hat{\beta}^t\hat{\beta})}{P\sigma^2} - \frac{1}{\lambda_j} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta + \sum_{1}^{p} \lambda_j\, \frac{\sigma^2}{\lambda_j}}{P\sigma^2} - \frac{1}{\lambda_j}$$

$$\frac{\beta^t\beta}{\sigma^2} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta}{P\sigma^2} + \frac{P\sigma^2}{P\sigma^2\lambda_j} - \frac{1}{\lambda_j} = \frac{\sum_{1}^{p} \lambda_j\, \beta^t\beta}{P\sigma^2}$$

$$K = \frac{\sigma^2}{\beta^t\beta} = \frac{P\hat{\sigma}^2}{\sum_{1}^{p} \lambda_j\, \hat{\beta}^t\hat{\beta}} = \frac{P\hat{\sigma}^2}{\hat{\beta}^t(\mathbf{X^tX})\hat{\beta}} \qquad (12)$$

lo cual completa la prueba.

5. Conclusiones

A medida que la multicolinealidad crece entre las variables regresoras que determinan el comportamiento de una variable de respuesta, los coeficientes estimados por MC del modelo polinomial que modela ese comportamiento, se vuelven erráticos e impredecibles, debido a los efectos desastrosos que la multicolinealidad tiene sobre su varianza, afortunadamente la regresión RR, minimiza este problema al contraer los coeficientes $\hat{\beta}_j$ de MC, logrando coeficientes ajustados con menor varianza, dando estabilidad así a la predicción del modelo. La solución de la constante K desarrollada en este artículo, ofrece una función de CME al aplicar la regresión RR, menor que el CME de MC, cuando el problema de multicolinealidad está presente.

Referencias

Golam Kibria BM. 2003. *Performance of Some New Ridge Regression Estimators*. Comunication in statistics: Vol 32, No 2, pp 419 – 435.

Hemmerle WJ. 1975. *An Explicit Solution for Generalized Ridge Regression*. Technometrics, vol 17, No 3 pp 309 – 314.

Hemmerle WJ and Brantle. 1978. *Explicit and Constrained Generalized Ridge Estimation*. Technometrics, vol 20, No 2 pp 109 – 120.

Hoerl AE y RW Kennard. 1970a. *Ridge Regression: Biased estimation for nonorthogonal Problems*. Technometrics, vol 12, No, 55-67.

Hoerl AE y RW Kennard. 1970b. *Ridge Regression: Applications to nonorthogonal problems*. Technometrics, vol 12, No 1, 69-82.

Hoerl AE, RW Kennard y KF Baldwin. 1975. *Ridge Regression: Some Simulations*. Comunication in statistics, 4(2), 105-123.

Hoerl AE y RW Kennard. 1976. *Ridge Regression Iterative Estimation of the Biased Parameter*. Communication in statistics, A5(1), 77-88.

Lawless JF and PWang. 1976. *A Simulation Study of Ridge and other Regression Estimators*. Communication in Statistics – Theory and Method, A5(4), 307 – 326.

Montgomery DC, EA Peck and GG Vining. 2002. *Introducción al Análisis de Regresión Lineal*. México: Editorial Continental tercera edición.

Piña MR, MA Rodríguez y JJ Díaz. 2005. *Superioridad de la Regresión General Ridge sobre Mínimos Cuadrados*. CULCYT//Enero-Febrero, 2005, México. Año 2, N° 6.

Rubio H and L Firinguetti. 2002. *The Distribution of Stochastic Shrinkage Parameters in Ridge Regression*. Communication in Statistics – Theory and Method: Vol 39 No 9, pp 1531 – 1547.

Vías del Ferrocarril. Centro de Ciudad Juárez. Foto: Betina.