Superioridad De La Regresión General Ridge Sobre Mínimos Cuadrados

Ing. Manuel R. Piña Monarrez[1], Dr. Manuel A. Rodríguez Medina[2] y Dr. Juan J. Diaz Núñez[3]

.

RESUMEN

El estimador de Mínimos Cuadrados (MC), es un caso especial del estimador general ridge propuesto por Hoerl y Kennard en 1970, para ajustar un modelo de regresión polinomial. En este artículo, se demuestra que la Regresión General Ridge (RGR), es mejor que el estimador de MC, para ajustar el polinomio, bajo el criterio del Cuadrado Medio del Error (CME), cuando el problema de multicolinealidad está presente. La prueba de este resultado, es dada aquí.

ABSTRACT

The ordinary least square estimator (OLS), is a special case of the general ridge estimator proposed by Hoerl and Kennard in 1970, for adjust a polynomial like a regression model. In this paper we demonstrate that the General Ridge Regression (GRR), is better than the OLS estimator, for adjust the polynomial, by the criterion of mean square error (MSE), when the multicollinearity problem is present. The proof of this result is given here.

Palabras Claves: Cuadrado Medio del Error, Regresión Ridge, Mínimos Cuadrados, Multicolinealidad.

INTRODUCCIÓN

En la literatura de regresión lineal múltiple, hay sugeridos varios estimadores para ajustar un polinomio a los datos de un experimento. Algunos de estos estimadores, por nombrar algunos, son Mínimos Cuadrados (MC), Mejor Estimador Lineal Insesgado (BLUE por sus siglas en inglés), Mínimos Cuadrados Generalizados (MCG) y Regresión General Ridge (RGR) (Creel, 2002). De hecho, en el diseño experimental, generalmente, empleamos MC, para ajustar un polinomio completo de segundo orden para representar la superficie de respuesta que queremos optimizar (Montgomery, 1991). Cuando el problema de multicolinealidad está presente en la regresión múltiple, el estimador RGR, mejora la estimación de la precisión de los coeficientes ajustados del polinomio, sobre los coeficientes estimados por MC, en el sentido de que el estimador RGR, obtiene un CME menor que el CME de MC (Box and Draper, 1987). La organización del artículo, es como sigue. En la sección 2, presentamos la prueba de la superioridad del estimador RGR sobre MC, bajo el criterio del CME, cuando el problema de multicolinealidad está presente. En la sección 3, algunas conclusiones son dadas.

SUPERIORIDAD DEL ESTIMADOR RGR SOBRE MC BAJO EL CRITERIO DEL CME

Considere el modelo de regresión lineal múltiple

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \qquad \textbf{(1)}$$

donde

[1] Estudiante del Doctorado en ingeniería Industrial. Instituto Tecnológico de Ciudad Juárez. romanpin1970@yahoo.com.mx
[2] Programa de Doctorado en Ingeniería Industrial. Instituto Tecnológico de Ciudad Juárez. Manuel_Rodriguez_itcj@yahoo.com
[3] Coordinador del Programa de Doctorado en Ingeniería Industrial. Instituto Tecnológico de Ciudad Juárez. jdiaz@uacj.mx

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} \cdots X_{1P} \\ 1 & X_{21} & X_{22} \cdots X_{2P} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n1} \cdots X_{nP} \end{bmatrix}, \quad \beta^t = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

y (X^tX) es la matriz de correlación de las variables predictoras. En orden para definir el estimador ridge generalizado, transformamos primero el modelo (1) a su forma canónica como sigue:
Sea **M** una matriz ortogonal de transformación tal que

$$M^t(X^tX)M = \Delta \qquad (2)$$

donde $\Delta = \mathbf{diag}(\lambda_1,\cdots,\lambda_p)$. Sea $\hat{X} = XM$, el i-seavo eje transformado y sea $\theta = M\beta$, el verdadero vector de regresión insesgado tal que $\theta_i = m_i\beta_i$ es el verdadero valor del i-seavo componente del vector de regresión y $\theta^t\theta = \beta^t\beta$.

Lema 1:

$$\theta = Mc = M\Delta^{-1/2}H^tY \qquad (3)$$

Es el verdadero vector de componentes de regresión incorrelacionados

Prueba:

Denotemos la descomposición del valor singular de **X** por

$$X = H\Delta^{1/2}M^t = \sum_i^p \lambda_i^{1/2} h_i m_i^t \qquad (4)$$

Donde **H** es una matriz ortogonal tal que $1^tH = 0^t$, y $\lambda_1^{1/2},\cdots,\lambda_p^{1/2}$ son los valores singulares de Δ los cuales dan la desviación (proporcional a la desviación estándar) (Obenchain, 1975), y **M** es la matriz ortogonal definida en (2). Sea $\hat{b} = (X^tX)^{-1}XY$ el estimador de MC, entonces, este puede ser escrito a través de la descomposición del valor singular por

$$E(\hat{b}) = (X^tX)^{-1}X^tY = M\Delta^{-1/2}H^tY = Mc = \theta \qquad (5)$$

donde **c** es un vector de componentes incorrelacionados. Note que desde que la desviación de $\hat{b}$ es

$$D(\hat{b}) = \sigma^2(X^tX)^{-1} = \sigma^2 M\Delta^{-1}M^t \qquad (6)$$

implica que $D(c) = D(M^t\hat{b}) = \sigma^2\Delta^{-1}$, lo cual completa la prueba.
El estimador RGR es definido por:

$$\hat{b} = [\Delta + k]^{-1}\hat{X}^tY \qquad (7)$$

donde

$$\Delta = \begin{bmatrix} \lambda_1 & 0 \cdots & 0 \\ 0 & \lambda_2 \cdots & 0 \\ \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots \lambda_p \end{bmatrix}, \quad K = \begin{bmatrix} k_1 & 0 \cdots & 0 \\ 0 & k_2 \cdots & 0 \\ \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots k_p \end{bmatrix}$$ para $k_i \geq 0$.

Si $Z = (I + \Delta^{-1}k)^{-1}$, entonces $\hat{b} = Z\hat{b}$, donde $\hat{b}$ es el mejor estimador lineal insesgado de θ dado por

$$\hat{b} = \Delta^{-1}(X^*)^tY \qquad (8)$$

y Z es dado por:

$$Z = \left[\begin{bmatrix} 1 & 0\cdots & \mathbf{0} \\ 0 & 1\cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \mathbf{1} \end{bmatrix} + \begin{bmatrix} 1/\lambda_1 & 0\cdots & \mathbf{0} \\ 0 & 1/\lambda_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1/\lambda_p \end{bmatrix} \begin{bmatrix} \mathbf{k_1} & 0\cdots & \mathbf{0} \\ 0 & k_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \mathbf{k_p} \end{bmatrix} \right]^{-1} \quad Z = \left[\begin{bmatrix} 1 & 0\cdots & \mathbf{0} \\ 0 & 1\cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \mathbf{1} \end{bmatrix} + \begin{bmatrix} k_1/\lambda_1 & 0\cdots & \mathbf{0} \\ 0 & k_2/\lambda_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & k_p/\lambda_p \end{bmatrix} \right]^{-1}$$

$$Z = \begin{bmatrix} \mathbf{1} + k_1/\lambda_1 & 0\cdots & \mathbf{0} \\ 0 & 1 + k_2/\lambda_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & \mathbf{1} + k_p/\lambda_p \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\lambda_1 + \mathbf{k}_1}{\lambda_1} & 0\cdots & \mathbf{0} \\ 0 & \frac{\lambda_2 + \mathbf{k}_2}{\lambda_2}\cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & \frac{\lambda_p + \mathbf{k}_p}{\lambda_p} \end{bmatrix}^{-1} \quad Z = \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + \mathbf{k}_1} & 0\cdots & \mathbf{0} \\ 0 & \frac{\lambda_2}{\lambda_2 + \mathbf{k}_2}\cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & \frac{\lambda_p}{\lambda_p + \mathbf{k}_p} \end{bmatrix}$$

Por lo que $\hat{\mathbf{b}} = \mathbf{Z}\hat{\mathbf{b}} = [\Delta + \mathbf{K}]^{-1}\hat{\mathbf{X}}^t\mathbf{Y}$ o equivalentemente $\hat{\mathbf{b}} = \mathbf{Z}^{-1}[\Delta + \mathbf{K}]^{-1}\hat{\mathbf{X}}^t\mathbf{Y}$ es dado por

$$\hat{\mathbf{b}} = \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + \mathbf{k}_1} & 0 & \cdots & \mathbf{0} \\ 0 & \frac{\lambda_2}{\lambda_2 + \mathbf{k}_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ 0 & 0\cdots & & \frac{\lambda_p}{\lambda_p + \mathbf{k}_p} \end{bmatrix}^{-1} \left[\begin{bmatrix} \lambda_1 & 0\cdots & \mathbf{0} \\ 0 & \lambda_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \lambda_p \end{bmatrix} + \begin{bmatrix} \mathbf{k}_1 & 0\cdots & \mathbf{0} \\ 0 & k_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \mathbf{k}_p \end{bmatrix} \right]^{-1} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{1} \\ \mathbf{X}_{11} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{p1} & \mathbf{X}_{p2} & \cdots & \mathbf{X}_{pn} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$\hat{\mathbf{b}} = \begin{bmatrix} \frac{\lambda_1 + \mathbf{k}_1}{\lambda_1} & 0 & \cdots & \mathbf{0} \\ 0 & \frac{\lambda_2 + \mathbf{k}_2}{\lambda_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ 0 & 0\cdots & & \frac{\lambda_p + \mathbf{k}_p}{\lambda_p} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{1}}{\lambda_1 + \mathbf{k}_1} & 0 & \cdots & \mathbf{0} \\ 0 & \frac{\mathbf{1}}{\lambda_2 + \mathbf{k}_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\mathbf{1}}{\lambda_p + \mathbf{k}_p} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{1} \\ \mathbf{X}_{11} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{p1} & \mathbf{X}_{p2} & \cdots & \mathbf{X}_{pn} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$\hat{\mathbf{b}} = \begin{bmatrix} 1/\lambda_1 & 0\cdots & \mathbf{0} \\ 0 & 1/\lambda_2 \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1/\lambda_p \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{1} \\ \mathbf{X}_{11} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{p1} & \mathbf{X}_{p2} & \cdots & \mathbf{X}_{pn} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

Entonces, por el teorema de existencia de la RGR (J. S Chawla, 1988), el estimador de RGR tiene superioridad sobre MC, en la estimación de los coeficientes de regresión en el sentido del CME. La prueba de el teorema, requiere de los siguientes dos lemas.

Lema 1:

$$\mathbf{CME}(\hat{\mathbf{b}}) = \sigma^2 \Sigma \left[\lambda_i / (\lambda_i + \mathbf{k}_i)^2 \right] + \Sigma \left[\mathbf{k}_i^2 \theta_i^2 / (\lambda_i + \mathbf{k}_i)^2 \right]$$

Prueba:

$$\mathbf{CME}(\hat{\mathbf{b}}) = \mathbf{E}(\hat{\mathbf{b}} - \theta)^t(\hat{\mathbf{b}} - \theta) = \mathbf{E}(\mathbf{Z}\hat{\mathbf{b}} - \theta)^t(\mathbf{Z}\hat{\mathbf{b}} - \theta) \qquad \textbf{(9)}$$

$$\mathbf{CME}(\hat{\mathbf{b}}) = \mathbf{E}(\mathbf{Z}\hat{\mathbf{b}} - \mathbf{Z}\theta)^t(\mathbf{Z}\hat{\mathbf{b}} - \mathbf{Z}\theta) + \mathbf{E}(\mathbf{Z}\theta - \theta)^t(\mathbf{Z}\theta - \theta)$$

Ahora,

$$\mathbf{E(Z\hat{b} - Z\theta)^t(Z\hat{b} - Z\theta) = Trazo\left[E(Z\hat{b} - Z\theta)^t(Z\hat{b} - Z\theta)\right]}$$
$$\mathbf{= Trazo\left[ZE(\hat{b} - \theta)^t(\hat{b} - \theta)z^t\right] = Trazo\left[Z\sigma^2\Delta^{-1}Z^t\right]} \qquad \textbf{(10)}$$

De esa forma la primera parte de (9) es dada por:

$$\sigma^2\mathbf{Trazo}(Z\Delta^{-1}Z^t) = \sigma^2 \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} \end{bmatrix} \begin{bmatrix} 1/\lambda_1 & 0 \cdots & 0 \\ 0 & 1/\lambda_2 \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1/\lambda_p \end{bmatrix} \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} \end{bmatrix}$$

$$\sigma^2\mathbf{Trazo}(Z\Delta^{-1}Z^t) = \sigma^2 \begin{bmatrix} \frac{1}{\lambda_1}\left(\frac{\lambda_1}{\lambda_1 + k_1}\right) & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2}\left(\frac{\lambda_2}{\lambda_2 + k_2}\right) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_p}\left(\frac{\lambda_p}{\lambda_p + k_p}\right) \end{bmatrix} \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} \end{bmatrix}$$

$$\sigma^2\mathbf{Trazo}(Z\Delta^{-1}Z^t) = \sigma^2 \begin{bmatrix} \frac{\lambda_1}{(\lambda_1 + k_1)^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{(\lambda_2 + k_2)^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{(\lambda_p + k_p)^2} \end{bmatrix}$$

$$\mathbf{Trazo\left[Z\sigma^2\Delta^{-1}Z^t\right] = \sigma^2\Sigma\left[(\lambda_i)/(\lambda_i + K_i)^2\right]} \qquad \textbf{(9a)}$$

En la misma forma el segundo término de (9) $\mathbf{E(Z\theta - \theta)^t(Z\theta - \theta)}$, puede ser escrito como $\mathbf{Trazo\theta(Z - I)^t(Z - I)\theta}$ y es dado por:

$$\mathbf{Trazo\theta(Z - I)^t(Z - I)\theta} = \theta \left[\begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} \end{bmatrix} - \begin{bmatrix} 1 & 0 \cdots & 0 \\ 0 & 1 \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1 \end{bmatrix} \right]^t \left[\begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} \end{bmatrix} - \begin{bmatrix} 1 & 0 \cdots & 0 \\ 0 & 1 \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1 \end{bmatrix} \right] \theta$$

$$\mathbf{Trazo\theta(Z - I)^t(Z - I)\theta} = \theta \left[\begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} - 1 & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} - 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} - 1 \end{bmatrix} \right]^t \begin{bmatrix} \frac{\lambda_1}{\lambda_1 + k_1} - 1 & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k_2} - 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k_p} - 1 \end{bmatrix} \theta$$

$$\mathbf{Trazo}\,\theta(Z-I)^t(Z-I)\theta=\theta\left[\begin{bmatrix} -\frac{k_1}{\lambda_1+k_1} & 0 & \cdots & 0 \\ 0 & -\frac{k_2}{\lambda_2+k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -\frac{k_p}{\lambda_p+k_p} \end{bmatrix}\right]^t\begin{bmatrix} -\frac{k_1}{\lambda_1+k_1} & 0 & \cdots & 0 \\ 0 & -\frac{k_2}{\lambda_2+k_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -\frac{k_p}{\lambda_p+k_p} \end{bmatrix}\theta$$

$$\mathbf{Trazo}\,\theta(Z-I)^t(Z-I)\theta=\theta\begin{bmatrix} \frac{k_1^2}{(\lambda_1+k_1)^2} & 0 & \cdots & 0 \\ 0 & \frac{k_2^2}{(\lambda_2+k_2)^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{k_p^2}{(\lambda_p+k_p)^2} \end{bmatrix}\theta$$

$$\mathbf{Trazo}\,\theta(Z-I)^t(Z-I)\theta=\Sigma\left[k_i^2\theta_i^2/(\lambda_i+k_i)^2\right] \qquad \textbf{(9b)}$$

Lo cual completa la prueba.

Lema 2 $\sigma^2\left[(\lambda_i)/(\lambda_i+K)^2\right]+\theta_i^2\left[K^2/(\lambda_i+K)^2\right]$ es una función monóticamente decreciente si $0\le k\le\sigma^2/\theta_i^2$.

Prueba:
Sea

$$f_i(k)=\sigma^2\left[(\lambda_i)/(\lambda_i+K)^2\right]+\theta_i^2\left[K^2/(\lambda_i+K)^2\right] \qquad \textbf{(11)}$$

$$f_i(k)=\frac{\sigma^2\lambda_i+\theta_i^2k^2}{(\lambda_i+K)^2}=(\sigma^2\lambda_i+\theta_i^2k^2)(\lambda_i+K)^{-2}$$

De esa forma

$$f_i'(k)=(-2\sigma^2\lambda_i+2\theta_i^2\lambda_ik)/(\lambda_i+k)^3 \qquad \textbf{(12)}$$

Desde que $f_i'(k)$ <0 si $k_i<\sigma^2/\theta_i^2$, la prueba está completa.

TEOREMA: $CME(\hat{b})<CME(\hat{b})$ si el mayor de los $k_i<\Omega$; Donde:

$$\Omega=\min\left\{\sigma^2/\theta_i^2; i=1,2,\ldots,p\right\} \qquad \textbf{(13)}$$

Prueba: del lema 2

$$\sigma^2\left[(\lambda_i)/(\lambda_i+K)^2\right]+\theta_i^2\left[K^2/(\lambda_i+K)^2\right]<\sigma^2/\lambda_i \text{ si } 0<k_i<\sigma^2/\theta_i^2 \qquad \textbf{(14)}$$

de esa forma,

$$\sigma^2\Sigma\left[(\lambda_i)/(\lambda_i+K_i)^2\right]+\Sigma\theta_i^2\left[K_i^2/(\lambda_i+K_i)^2\right]<\sigma^2\Sigma(1/\lambda_i) \qquad \textbf{(15)}$$

si el mayor de los $k_i<\min\left\{\sigma^2/\theta_i^2; i=1,2,\ldots,p\right\}$, o equivalentemente

$$CME(\hat{b})<CME(\hat{b}) \qquad \textbf{(16)}$$

Si el mayor de los $k_i<\Omega$, lo cual completa la prueba.

Colorario: Cuando k=0, la función del Cuadrado Medio del Error del estimador de Regresión General Ridge, es la función del Cuadrado Medio del Error del estimador de Mínimos Cuadrados

$$\sigma^2\Sigma\left[(\lambda_i)/(\lambda_i+K_i)^2\right]+\Sigma\theta_i^2\left[K_i^2/(\lambda_i+K_i)^2\right]=\sigma^2\Sigma(1/\lambda_i)$$ si y solo si $k_i = 0$ for i=1,...,p

Prueba:

$$\sigma^2\Sigma\left[(\lambda_i)/(\lambda_i+0)^2\right]+\Sigma\theta_i^2\left[0/(\lambda_i+0)^2\right]=\sigma^2\Sigma(1/\lambda_i)$$

$$\sigma^2\Sigma\left[\lambda_i/\lambda_i\lambda_i\right]+\Sigma\theta_i^2\left[0\right]=\sigma^2\Sigma(1/\lambda_i)$$

$$\sigma^2\Sigma\left[1/\lambda_i\right]=\sigma^2\Sigma(1/\lambda_i) \tag{17}$$

Lo cual completa la prueba.

CONCLUSIÓN

En este artículo, utilizamos el criterio del CME, para comparar los estimadores de RGR y MC como una alternativa para ajustar un modelo polinomial en regresión lineal múltiple. Aquí, también mostramos que el estimador RGR, es mejor que el estimador de MC en el sentido del CME, cuando el problema de multicolinealidad está presente. El resultado singular de que la función del CME del estimador RGR, es la función del CME del estimador de MC cuando K=0, fue establecido como colorario.

REFERENCIAS

Box G. E. P. and Draper N. R. 1987. Empirical Model-Building and Response Surfaces, New York: Jhon wiley and Sons.

Creel M. 2002. Graduate Econometrics Lecture Notes, University of Barcelona, España, version 0.4, Revised November 2002.

Montgomery D. C. 1991. Diseño y Análisis de Experimentos, México D. F. versión en español, Grupo Editorial Iberoamérica.

Obenchain R. L. 1975. "Ridge Analysis Following a Preliminary Test of the Shrunken Hypotesis", Technometrics, E.U.A. Vol. 17, No.4, 431-441.

Chawla J. S. 1988. "The Existence Theorem in General Ridge Regression" Elsevier Science Publisher, North-Holland Nova Scotia, Canada. Vol. 7, Issue 2, 135-137.