Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Sistema de Información Científica

Manuel R. Piña Monarrez, Manuel A. Rodríguez Medina, Jesús J. Aguirre Solís

Regresión Ridge y la distribución central t

Ciencia Ergo Sum, vol. 14, núm. 2, julio-octubre, 2007, pp. 191-196,
Universidad Autónoma del Estado de México
México

Disponible en: http://www.redalyc.org/articulo.oa?id=10414210

Ciencia Ergo Sum,
ISSN (Versión impresa): 1405-0269
ciencia.ergosum@yahoo.com.mx
Universidad Autónoma del Estado de México
México

¿Cómo citar? | Fascículo completo | Más información del artículo | Página de la revista

www.redalyc.org
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Regresión Ridge y la distribución central *t*

Manuel R. Piña Monarrez*, Manuel A. Rodríguez Medina**, Jesús J. Aguirre Solís*

Recepción: 24 de mayo de 2006
Aceptación: 7 de noviembre de 2006

***División de Ciencias de la Ingenieria y Tecnología, Instituto Tecnológico Superior de Nuevo Casas Grandes, Chihuahua, México.**
Correo electrónico: mpina@itsncq.edu.mx
Correo electrónico: jesaguir@yahoo.com
****División de Ciencias de la Ingeniería y Tecnología, Instituto Tecnológico de Ciudad Juárez, México.**
Correo electrónico: mrodriguez@itcj.edu.mx

Resumen. Dado que la Regresión Ridge (RR), es una estimación sesgada que parte de la solución de la regresión de Mínimos Cuadrados (MC), es vital establecer las condiciones para las que la distribución central *t* de Student que se utiliza en la prueba de hipótesis en MC, sea también aplicable a la regresión RR. La prueba de este importante resultado se presenta en este artículo.

Palabras clave: regresión ridge, mínimos cuadrados, distribución *t*, prueba de hipótesis.

Ridge Regression and Central *t* Student Distribution

Abtract. Since Ridge Regression (RR), is a biased estimation, that begins with the Ordinary Least Square (OLS) solution, it is vital to establish the conditions for which the central Student's *t* distribution that we use in the hypothesis test in OLS, is applicable to the RR too. The proof of this important result is given in this article.

Key words: ridge regression, ordinary least square, student's *t* distribution, hypothesis test.

Introducción

Durante el proceso de optimización de las superficies de respuestas, se ajusta un modelo polinomial del tipo de la ecuación (1), a través de un método de regresión, a los datos de un diseño experimental (Box y Draper, 1987).

$$\hat{Y} = ß_0 + ß^t X + X^t BX \qquad (1)$$

donde

$$ß^t = [ß_1, ß_2, ..., ß_k], \ X = [x_1, x_2, ..., x_k] \text{ y}$$

$$B = \begin{bmatrix} ß_{11} & ß_{12}/2 \ ... \ ß_{1k}/2 \\ & ß_{22} \quad ... \ ß_{2k}/2 \\ & \ddots \quad \vdots \\ sym & \quad ß_{kk} \end{bmatrix}.$$

La funcionalidad del modelo dado en (1) para representar a la superficie de respuestas depende del grado que este polinomio representa al sistema modelado, por lo que se realizan pruebas para determinar la significancia de sus componentes.

El procedimiento para realizar estas pruebas es conocido como prueba de hipótesis. Primero se plantean dos hipótesis, la nula que establece que los coeficientes de regresión $\hat{ß}_j$ no tienen efecto sobre la variable de respuestas (H_0: $\hat{ß}_j = 0$), y la alterna que establece que al menos un coeficiente tiene efecto sobre la variable de respuesta (H_1: $\hat{ß}_j \neq 0$). La regla de decisión de rechazar o no la hipótesis establecida consiste en especificar una región crítica para un nivel de significancia preestablecido (α) de acuerdo con una distribución teórica que para nuestro caso es la distribución t de Student, cuya región crítica, por ejemplo, para un nivel de significancia de $\alpha = 5\%$ está dada por $t_{a/2,n-p}$ donde (n-p) son los grados de libertad del cuadrado medio del error. Así, el valor crítico dado por $t_{a/2,n-p}$, es comparado con el valor del estadístico de prueba dado por:

$$t = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}} \qquad (2)$$

donde $\hat{ß}_j$ es el *j-ésimo* coeficiente de regresión estimado, σ^2 es la varianza estimada y C_{ii} es el *j-ésimo* elemento diagonal de la matriz de precisión $(X^t X)^{-1}$(Montgomery Peck y Vining, 2002). Si el valor del estadístico de prueba dado en (2), es mayor que el valor crítico de la distribución teórica, se rechaza la hipótesis nula y se concluye que el coeficiente es significante para predecir el comportamiento de la variable de respuesta. Este procedimiento es ampliamente aplicado cuando el modelo polinomial se ajusta a través de MC. El polinomio dado en (1), presenta de forma inherente el problema de multicolinealidad, por lo que su ajuste deberá de realizarse a través del método de RR, ya que cuando la multicolinealidad inherente al polinomio es fuerte, algunos de los *eigenvalues* λ_j de la matriz de covarianzas $(X^t X)$ tienden a cero, haciendo que el poder de la prueba t sea pobre para detectar cuando β_j comienza a ser significante (Halawa y El Bassiouni, 2000). Dado que el estimador RR es la estimación óptima sesgada para determinar los coeficientes del polinomio cuando éste presenta multicolinealidad (ver Piña *et al.*, 2005c), y como para el sesgo de RR, se espera que tanto el coeficiente β_{Rj} estimado como su varianza $V(\beta_{Rj}) = \sigma^2 \frac{\lambda_j}{\lambda_j + k}$, sean menor que el coeficiente $\hat{ß}_j$ y su varianza $V(\hat{ß}_j) = \sigma^2 \frac{1}{\lambda_j}$, entonces, el valor absoluto esperado del estadístico de prueba de RR, se espera que sea mayor que el de MC, mejorando así el poder de detección de la prueba. El cuadrado medio del error de RR, sigue una distribución chi-cuadrada no central (ver Cheng-Ming Kuan, 2000: 62-63) con parámetro de sesgo dado por (Piña *et al.*, 2005):

$$Sesgo^2 = k^2 \sum_{1}^{p} \frac{\beta_i^2}{(\lambda_i + k)^2},$$

por lo que es necesario mostrar que debido a que RR es un múltiplo de MC, la aplicación del estadístico central t, es válida para la prueba de hipótesis de la regresión RR. En este sentido, Obenchain en 1977 derivó una prueba "exacta t" para la regresión RR dada por $t(R) = (\beta_R - b_t) / \sqrt{\mathrm{var}(\beta_R - b_i)}$ donde $\mathrm{var}(\beta_R - b_i)$ es una estimación insesgada de la varianza del numerador y

$$b_i = g_i^t \Delta G^t \left[I - (X^t X)^{-1} e_i^t \left(e_i (X^t X)^{-1} e_i^t \right)^{-1} e_i \right] \hat{\beta},$$

es el estimador insesgado de varianza mínima de β_R bajo la hipótesis nula H_0: $\beta_i = 0$, donde g es la *j-ésima* fila de G que representa los cosenos directores que orientan los ejes principales en relación con los ejes del regresor dado, Δ es una matriz de (p x p) con elementos diagonales dados por $\delta = \lambda_i / (\lambda_i + k)$ (error estándar β_R) y e_i es el *j-ésimo* elemento de la matriz identidad, el uso de esta prueba ha sido controversial ya que depende de la dirección del regresor dado. Halawa y El Bassiouni (2000), a través de la cuantificación del sesgo emplearon una prueba t no central para probar la hipótesis H_0: $b_i = 0$. El estudio consistió en generar una matriz de regresores X, con el algoritmo dado en Halawa y Assma en 1995, y haciendo uso de tres niveles pre-establecidos de desviación (σ = 0.05, 0.5 y 1) y dos niveles de correlación dados por ρ = 0.90 y ρ = 0.95, para realizar las pruebas, sólo en la dirección del máximo y mínimo eigenvalor. Encontraron que cuando el error es grande su prueba supera en poder de detección a la prueba central t de MC; cuando los errores son pequeños (como es el caso de los diseños experimentales), no hubo diferencia significativa entre los niveles de detección de las pruebas. Dado que el uso de RR mejora el poder de detección de la prueba al disminuir la varianza de la estimación, en la sección 3 se presenta la validación matemática para el uso de la distribución central t en la prueba de hipótesis de la regresión RR.

1. Regresión Ridge (RR)

El método de RR permite detectar la multicolinealidad dentro de un modelo de regresión del tipo $Y = X^t \hat{\beta} + \varepsilon$. Fue propuesto por Hoerl y Kennard (1970a y b); es usado para trabajar con modelos que presentan sesgo. La idea del método es simple y consiste en que dado que la matriz $(X^t X)$ 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. El vector de coeficientes de RR está dado por:

$$\beta_R = (X^t X + K^\delta)^{-1} X^t Y \qquad (3)$$

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

$$Min: (\beta_R - \hat{\beta}) X^t X (\beta_R - \hat{\beta})$$

$$Sujeto\ a: \beta_R \beta_R \leq r^2$$

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

$$Min: (\beta_R - \hat{\beta}) X^t X (\beta_R - \hat{\beta}) + K^\delta (\beta_R \beta_R - r^2) = 0$$

derivando la función con respecto a β_R e igualándola a cero tenemos que:

$$df/d\beta_R = 2X^t X(\beta_R - \hat{\beta}) + 2K^\delta \beta_R$$

$$X^t X\beta_R - X^t X\hat{\beta} + K^\delta \beta_R = 0$$

$$\beta_R(X^t X + K^\delta) - X^t X\hat{\beta} = 0$$

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

$$\beta_R(X^t X + K^\delta) - X^t Y = 0$$

$$\beta_R = (X^t X + K^\delta)^{-1} X^t Y$$

$$\beta_R = \frac{X^t Y}{\left[I + K^\delta (X^t X)^{-1}\right] (X^t X)}$$

$$\beta_R = [I + K^\delta (X^t X)^{-1}]^{-1} (X^t X)^{-1} X^t Y$$

$$\beta_R = Z\hat{ß}, \tag{4}$$

lo cual completa la prueba.

En la estimación RR, al parámetro K se le conoce como parámetro de sesgo ya que cuando $K = 0$, $Z = I$, por lo que $\beta_R = \hat{\beta}$ y cuando $K \neq 0$, $\beta_R \neq \hat{\beta}$. Además como $\hat{\beta}$ es insesgado entonces la estimación ridge es una estimación sesgada y aunque posee esta característica, la varianza de la estimación de β_R a β es menor que la varianza de la estimación de $\hat{\beta}$ a β, por lo que los coeficientes de β_R, son más estables que los de $\hat{\beta}$ (ver Piña *et al.*, 2005a). Para ver por qué analizamos la función del cuadrado medio del error de β_R, dado por:

$$\text{CME}\beta_R = \sigma^2 \sum \frac{\lambda_j}{(\lambda_j + k_j)^2} + k_j^2 \sum \frac{ß_j^2}{(\lambda_j + k_j)^2} \tag{5}$$

La función dada en (5), debido al parámetro de sesgo, sigue una distribución chi-cuadrada no central (Chung-Ming, 2000: 62-63) que implica directamente que la prueba de significancia de sus coeficientes estimados deberá de hacerse a través de la distribución no central t. Dado que RR es un múltiplo de MC, en la sección 3 se derivan las condiciones para utilizar la prueba central t al probar la significancia de los coeficientes ajustados.

El valor de la constante de proporcionalidad K depende de la varianza poblacional y del vector de coeficientes verdaderos, los cuales son desconocidos, por lo que el valor de K deberá estimarse a partir del estimador propuesto por Lawless y Wang (1976), que está dado por:

$$K = \frac{P\hat{s}^2}{\hat{\beta}^t (X^t X)\hat{\beta}} \tag{6}$$

Una vez conocido el valor de K, la regresión RR puede ser tratada como una regresión de MC, con tan sólo incrementar la matriz de información X con una matriz diagonal cuyos elementos diagonales sean la raíz cuadrada del valor de K dado por (6) y el vector de respuestas con un vector de ceros si se está trabajando con variables escaladas, o con un vector constante equivalente a la media de Y si las variables no están escaladas. Para un procedimiento detallado de la aplicación de RR ver Piña (2006).

2. Distribución *t*

Sea W una variable aleatoria con distribución N(0,1), y sea V una variable aleatoria que sigue una distribución chi-cuadrada $\chi^2(r)$ con r grados de libertad, donde W y V son independientes, por lo que la distribución conjunta de W y V, representada por $\varphi(w,v)$, es el producto de las funciones densidad de probabilidad normal y chi-cuadrada, respectivamente dada por:

$$\varphi(w, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \frac{1}{\Gamma\left(\frac{1}{2}r\right)2^{\frac{r}{2}}} v^{\frac{r}{2}-1} e^{-\frac{v}{2}}$$

$-\infty < w < \infty$, $0 < v < \infty$, 0 en otro caso.

Este producto define otra variable aleatoria t, representada por:

$$T = \frac{W}{\sqrt{V/\text{r}}} \tag{7}$$

Al aplicar la técnica de cambio de variable (Hogg y Graig, 1965), obtenemos la función densidad de probabilidad $g(t)$ de T. Así, la ecuación $t = w/\sqrt{v/\text{r}}$ y $u = v$, definen la transformación uno a uno, la cual mapea al conjunto $A = \{(w, v); -\infty < w < \infty, 0 < v < \infty\}$, en $B = \{(t, u); -\infty < t < \infty, 0 < u < \infty\}$. Además, desde que $w = t\sqrt{u}/\sqrt{r}$ y $v = u$, el valor absoluto del jacobiano de la transformación es $|J| = \sqrt{u}/\sqrt{r}$. Por lo que la función densidad de probabilidad conjunta de las variables aleatorias T y $U = V$, está dada por:

$$g(t,u) = \left(\frac{t\sqrt{u}}{\sqrt{r}}, u\right)|J|$$

$$= \frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} u^{r/2-1} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]\frac{\sqrt{u}}{\sqrt{r}}$$

$-\infty < t < \infty$, $0 < u < \infty$, 0 en otro caso.

Así, la función densidad de probabilidad marginal de T es:

$$g_1(t) = \int_{-\infty}^{\infty} g(t,\mathrm{u})\mathrm{du} =$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} u^{(r+1)/2-1} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]du$$

Para esta integral, si hacemos $z = u[1+(t^2/r)]/2$, entonces:

$$g_1(t) = \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} \left(\frac{2z}{1+t^2/r}\right)^{(r+1)/2-1} \exp^{-z}\left(\frac{2}{1+t^2/r}\right)dz$$

$$g_1(t) = \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \ -\infty < t < \infty. \qquad (8)$$

Si W está N(0,1) y V sigue una distribución $\chi^2(r)$ y si W y V son independientes, entonces la variable aleatoria $T = w/\sqrt{V/r}$ tiene una función densidad de probabilidad marginal dada por (8).

A la distribución de la variable aleatoria T, se le conoce como distribución t, y está completamente determinada por el parámetro r, el cual representa los grados de libertad de la variable aleatoria que sigue una distribución $\chi^2(r)$. Finalmente, de (8) tenemos que los valores de t están dados por:

$$g_1(t) = \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/2)(1+w^2/r)^{(r+1)/2}} dw, \ -\infty < t < \infty. \qquad (9)$$

3. Prueba de hipótesis de la regresión ridge

Para simplificar los argumentos, recordemos que la hipótesis general para determinar la significancia del *j-ésimo* componente de regresión en MC, está dada por H_0: $\beta_i = 0$ generalmente contra H_1: $\beta_i \neq 0$, en donde la prueba de su significancia se realiza a través del estadístico de prueba t dado por:

$$t_i = \left[\hat{\beta}_i - 0\right] / \left[s^2 (X^tX)^{-1}_{ii}\right]^{1/2} = \hat{\beta}_i / \left[s^2(X^tX)^{-1}_{ii}\right]^{1/2}, \qquad (10)$$

donde $(X^tX)^{-1}_{ii}$, es el *i-séavo* elemento diagonal de la matriz de precisión $(X^t X)^{-1}$, s^2 es la estimación de la varianza dada por $s^2 = Y[I - X(X^tX)^{-1}X^t]Y/(n\text{-}p)$ y $[s^2(X^tX)^{-1}_{ii}]^{1/2}$ es la estimación insesgada del numerador bajo el supuesto de que los errores están $\varepsilon \approx N(0, \sigma^2 I)$. Además, como el valor esperado de $E(\beta_i) = 0$, bajo la hipótesis H_0: $\beta_i = 0$, el estadístico t, tiene una distribución central con $(n\text{-}p)$ grados de libertad. De igual forma el estadístico t para H_0: $\beta_i = 0$ utilizando el estimador RR, puede ser establecido directamente de (10) como sigue.

Teorema 3.1. El estadístico de prueba t para la Regresión Ridge, es el estadístico de prueba t de Mínimos Cuadrados.

Demostración: permita a β_i^* denotar el *j-ésimo* elemento de β_R y sea b_i el estimador lineal insesgado de varianza mínima de β_i^* con valor esperado $E(b_i) = 0$ cuando H_0: $\beta_i = 0$ es verdadera, por lo que el estadístico de prueba para RR, está dado por:

$$t_i = \left[\beta_i^* - b_i\right] / \left[s^2 W_i^2\right]^{1/2} \qquad (11)$$

donde $W_i^2 = Z_i^2(X^tX)^{-1}_{ii}$ es una constante conocida y $s^2W_i^2$ es el estimador insesgado de la varianza del numerador. Note que $s^2 = Y[I - X(X^tX)^{-1}X^t]Y/(n\text{-}p)$. Dado que $\beta_R = Z\hat{\beta}$, entonces $\beta_i^* = Z_i\beta_i$ y $b_i = Z_i b_i$, por lo que (11), puede ser escrita como:

$$t_i = \left[Z_i\hat{\beta}_i - Z_i b_i\right] / \left[s^2 Z_i^2 (X^tX)^{-1}\right]^{1/2}$$

$$= Z_i\left[\hat{\beta}_i - 0\right] / Z_i\left[s^2(X^tX)^{-1}\right]^{1/2}$$

$$= \hat{\beta}_i / \left[s^2(X^tX)^{-1}\right]^{1/2}$$

lo cual completa la demostración.

4. Una aplicación

El experimento fue originalmente realizado por Box (1954), con el objetivo de maximizar la salida de una reacción química sujeta a dos estados y cinco factores, cuyas variables codificadas son:

$$X_1 = \frac{(T_1 - 122.5)}{7.5}$$ Temperatura en el estado 1,

$$X_2 = \left\{\frac{2(\log t_1 - \log 5)}{\log 2}\right\} + 1$$ tiempo de reacción en el estado 1,

$X_3 = \dfrac{(C-70)}{10}$ concentración del reactante,

$X_4 = \dfrac{(T_2 - 32.5)}{7.5}$ Temperatura en el estado 2, y

$X_5 = \left\{ \dfrac{2(\log t_2 - \log 1.25)}{\log 5} \right\} + 1$ tiempo de reacción en el estado 2.

El diseño inicial fue factorial, fraccionado estándar 2^{5-1}, al cual se le aplicó el método de ascenso acelerado a un polinomio de primer grado ajustado por MC a este diseño (ver Box y Draper, 1987), hasta que el experimentador determinó que la región actual modelada contenía el punto estacionario óptimo, por lo que el experimentador agregó corridas experimentales para convertir el diseño en un diseño central compuesto (con 11 puntos centrales) y ajustar un polinomio de segundo grado completo para realizar la caracterización y exploración de esa región. La matriz de diseño en variables codificadas y el vector de respuestas, se presenta en la tabla 1.

El polinomio cuadrático completo ajustado por Minitab a los datos escalados a través del método de escalonamiento unitario (Montgomery *et al.*, 2002) es:

Tabla 1. Matriz de variables codificadas

	X_1	X_2	X_3	X_4	X_5		Y
	–1	–1	–1	–1	1		49.80
	1	–1	–1	–1	–1		51.20
	–1	1	–1	–1	–1		50.40
	1	1	–1	–1	1		52.40
	–1	–1	1	–1	–1		49.20
	1	–1	1	–1	1		67.10
	–1	1	1	–1	1		59.60
	1	1	1	–1	–1		67.90
	–1	–1	–1	1	–1		59.30
	1	–1	–1	1	1		70.40
	–1	1	–1	1	1		69.60
	1	1	–1	1	–1		64.00
	–1	–1	1	1	1		53.10
	1	–1	1	1	–1		63.20
X =	–1	1	1	1	–1	Y =	58.40
	1	1	1	1	1		64.30
	3	–1	–1	1	1		63.00
	1	–3	–1	1	1		63.80
	1	–1	–3	1	1		53.50
	1	–1	–1	3	1		66.80
	1	–1	–1	1	3		67.40
	1.23	–0.56	–0.03	0.69	0.70		72.30
	0.77	–0.82	1.48	1.88	0.77		57.10
	1.69	–0.30	–1.55	–0.50	0.62		53.40
	2.53	0.64	–0.10	1.51	1.12		62.30
	–0.08	–1.75	0.04	–0.13	0.27		61.30
	0.78	–0.06	0.47	–0.12	2.32		64.80
	1.68	–1.06	–0.54	1.50	–0.93		63.40
	2.08	–2.05	–0.32	1.00	1.63		72.50
	0.38	0.93	0.25	0.38	–0.24		72.00
	0.15	–0.38	–1.20	1.76	1.24		70.40
	2.30	–0.74	1.13	–0.38	–0.15		71.80

Regression analysis: *Y* versus b_0; X_1;

$$\hat{Y} = 0.00 + 0.20259X_1 + 0.10896X_2 + 0.21641X_3 + 0.31819X_4$$
$$+ 0.14082X_5 - 0.31763X_1^2 - 0.24471X_2^2 - 0.48295X_3^2$$
$$- 0.40416X_4^2 - 0.26317X_5^2 - 0.28808X_1X_2 + 0.30388X_1X_3$$
$$- 0.05279X_1X_4 - 0.00597X_1X_5 + 0.07616X_2X_3 - 0.02281X_2X_4$$
$$- 0.15118X_2X_5 - 0.52389X_3X_4 - 0.09917X_3X_5 + 0.05477X_4X_5$$

Análisis de varianza

Source	DF	SS	MS	F	MS
Regression	20	0.987193	0.049359	41.00	0.000
Residual error	11	0.013244	0.001204		
Total	31	1.000427			

No replicates. Cannot do pure error test.

S = 0.03470 R-Sq = 98.7% R-Sq(adj) = 96.3%
PRESS = 0.259868 R-Sq(pred) = 74.02%

Porción de la matriz de precisión $(X^t X)^{-1}$ que contiene los (C_{jj}) utilizados en la ecuación (2) para este ajuste dado por MC, es:

$$(X^t X)^{-1} = \begin{bmatrix} \boldsymbol{1.835} & 0.510 & 0.401 & 0.496 & 0.634 \\ 0.510 & \boldsymbol{2.093} & 0.710 & 0.697 & 0.756 \\ 0.401 & 0.710 & \boldsymbol{1.564} & 0.330 & 0.584 \\ 0.496 & 0.697 & 0.330 & \boldsymbol{1.956} & 0.580 \\ 0.634 & 0.756 & 0.584 & 0.580 & \boldsymbol{1.647} \end{bmatrix} \qquad (12)$$

Del ajuste, tenemos que $p = \dfrac{(k+1)(k+2)}{2} = 21$ parámetros, $\hat{\sigma}^2 = (0.03470)^2$ y $\hat{\beta}^t (X^t X) \hat{\beta} = 0.98659$, por lo que de la ecuación (6), $K = \dfrac{(21)(0.03470)^2}{0.98659} = 0.25629$. Con este valor de K, el polinomio ajustado a través de RR utilizando Minitab, está dado por:

Regression Analysis: Y versus X1, X2, ...

$$\hat{Y} = 0.00 + 0.20969X_1 + 0.10397X_2 + 0.21210X_3 + 0.31404X_4$$
$$+ 0.14190X_5 - 0.29227X_1^2 - 0.21343X_2^2 - 0.46031X_3^2$$
$$- 0.37443X_4^2 - 0.23595X_5^2 - 0.27755X_1X_2 + 0.29870X_1X_3$$
$$- 0.06660X_1X_4 - 0.01686X_1X_5 + 0.06955X_2X_3 - 0.01381X_2X_4$$
$$- 0.14708X_2X_5 - 0.51027X_3X_4 - 0.09590X_3X_5 + 0.04442X_4X_5$$

Análisis de varianza

Source	DF	SS	MS	F	P
Regression	20	0.955046	0.047752	32.93	0.000
Residual error	31	0.044954	0.001450		
Total	51	1.000000			

No replicates. Cannot do pure error test.

S = 0.03808 R-Sq = 95.5% R-Sq(adj) = 92.6%

Porción de la matriz de precisión $(X^t X)^{-1}$que contiene los (C_{jj}) utilizados en la ecuación (2) para este ajuste de RR es:

$$(X^t X)^{-1} = \begin{bmatrix} 1.753 & 0.452 & 0.362 & 0.443 & 0.573 \\ 0.452 & 1.982 & 0.646 & 0.623 & 0.678 \\ 0.362 & 0.646 & 1.515 & 0.292 & 0.533 \\ 0.443 & 0.623 & 0.292 & 1.859 & 0.519 \\ 0.573 & 0.678 & 0.533 & 0.519 & 1.579 \end{bmatrix} \qquad (13)$$

Con estos datos, y haciendo uso de la ecuación (2) y los coeficientes dados en (12) y (13), tenemos por ejemplo que el valor del estadístico de prueba para el coeficiente $\hat{ß}_{11}$ para ambos modelos es:

$$\hat{\beta}_{11(MC)} = \frac{-0.31763}{\sqrt{(0.03470)^2(1.835)}} = -6.757$$

$$\hat{\beta}_{11(RR)} = \frac{-0.29227}{\sqrt{(0.03470)^2(1.753)}} = -6.361$$

De igual forma, es posible construir un estadístico de prueba t_j para cada uno de los coeficientes $\hat{ß}_j$ estimados ya sea por MC o por RR.

Conclusiones

Aunque la regresión ridge es sesgada, es ampliamente utilizada para el ajuste del polinomio cuando existe multicolinealidad. Una vez determinada la constante de proporcionalidad K, la estimación RR puede ser tratada como una estimación de MC. Aunque el cuadrado medio del error de RR sigue una distribución chi-cuadrada no central, por el teorema 3.1 de la sección 3, la distribución central t, es perfectamente aplicable para probar la significancia de los coeficientes estimados por RR. Para al ajuste del polinomio cuadrático completo con que se modelan las superficies de respuesta, se recomienda utilizar el estimador RR para su ajuste, por presentar este polinomio de forma inherente el problema de multicolinealidad. ergo

Bibliografía

Box, G. E. P. y N. R. Draper (1987). *Empirical Model-Building and Response-surfaces*. John Wiley & Sons, Nueva York.

Bisgaard, S. y B. Ankenman (1998). ''Standard Errors for the Eigenvalues in Second Order Response Surface Models'', *Technometrics*. Vol. 36.

Chung-Ming, K. (2000). *Introduction to Econometric Theory*. Institute of Economics. Taipei, Taiwan.

Hoerl, A. E. y R. W. Kennard

_________ (1970a). "Ridge Regression: Biased Estimation for Non-orthogonal Problems", *Technometrics*. Vol. 12. Núm.1.

_________ (1970b). ''Ridge Regression: Applications to Non-orthogonal Problems'', *Technometrics*. Vol. 12. Núm 1.

Hogg, R. V. y A. T. Craig (1965). *Introduction to Mathematical Statistics*. Macmillan Company. Catalog Card Number:65-10243. New York.

Halawa, A. M. y M. Y. El Bassiouni (2000). ''Test f Regression Coefficients Under Ridge Regression Models", *Statistics Computer Simulation*. Vol. 65.

Lawless, J. F. y P. Wang (1976). "A Simulation Study of Ridge and Other Regression Estimators'', *Communication in Statistics – Theory and Method*. A5(4).

Mongomery, D. C.; E. A. Peck y G. G. Vining (2002). *Introducción al Análisis de Regresión Linneal*. 3a ed. Editorial Continental, México.

Obenchain R. L. (1977). "Classical F–Test and Confidence Regions for Ridge Regression", *Technometrics*. 19(4).

Piña, M. R.; M. A. Rodríguez y J. J.Díaz

_________ (2005a). ''Superioridad de la regresión general ridge sobre mínimos cuadrados'', *CULCYT*. Año 2. Núm. 6, enero-febrero, Universidad Autónoma de Ciudad Juárez. Chihuahua, México.

_________ (2005b). "Determinación de la constante k en la regresión ridge", *CULCYT*. Año 2. Núm. 11, enero-febrero, Universidad Autónoma de Ciudad Juárez. Chihuahua, México.

Piña, M. R. (2006). *Metodología para la caracterización y exploración de superficies de respuestas*. Tesis doctoral, Instituto Tecnológico de Ciudad Juárez Chih. México.