Análisis de variancia para datos colineales (1)

Alipio Ordoñez Mercado (2)

Resumen

Este trabajo presenta el uso de la "Ley de los Cosenos", para realizar el análisis de variancias de modelos de regresión, en las cuales existe entre las variables explicativas el problema de "Multicolinealidad". También se presenta una aplicación numérica para detallar la implementación de la técnica.

Palabras claves: Análisis de Variancia, Datos Colineales, Regresión, Grados de Libertad Fraccional.

INTRODUCCION

El análisis de la variancia, es una de las técnicas estadísticas que más se ha extendido entre los investigadores de las diferentes disciplinas científicas. El objetivo principal que se persigue es el de realizar la prueba de hipótesis siguiente:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \ldots = \mu_k; \; k \geq 2 \\ H_1 : \mu_i \neq \mu_j \; ; \text{ para algún } \; i = 1, 2, \ldots, k \end{cases} \quad (1)$$

Para el modelo de regresión lineal múltiple, se considera el siguiente modelo:

$$Y_{nx1} = X_{nxP} \; \beta_{Px1} + \varepsilon_{nx1} \quad (2)$$

Donde las variables X e Y, están definidas en su forma correlación, como se explica en Ordoñez (1993). La técnica del análisis de variancia aplicado al modelo (2), proporciona un medio para realizar la siguiente prueba de hipótesis

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \ldots = \beta_k; \; k \geq 1 \\ H_1 : \beta_i \neq \beta_j \; ; \text{ para algún } \; i = 1, 2, \ldots, k \end{cases} \quad (3)$$

---

1 Subvencionado por el FEDU Nº 4117-93-406 del Instituto de Investigación de la Facultad de Ciencias, Univ. Nac. de Ingeniería. Lima - Perú.

2 UNI, Facultad de Ciencias. Escuela Profesional de Estadística.

Es muy conocido, que bajo las suposiciones usuales, los estimadores mínimo cuadráticos para β y σ^2, son dados por

$$\hat{\beta} = (X^TX)^{-1} X^TY \; ; \; \hat{\sigma}^2 = (Y - \hat{Y})^T (Y - \hat{Y})/(n - p - 1) \qquad (4)$$

y se constituyen como los mejores en un sentido estadístico.

Sin embargo, en casi todas las aplicaciones prácticas de la vida real se tienen con frecuencia, matrices diseños X^TX, con determinantes próximos de "0" y los estimados por mínimos cuadrados ordinarios se tornan inestables en signo y magnitud, y se utilizan procedimientos alternativos como el de **Regresión Ridge**, cuyos detalles teóricos y prácticos se describen en los trabajos de Hocking (1976), y Hoerl and Kennard (1970).

La Regresión Ridge, es una familia de estimadores indexados

$$\hat{\beta}(k) = \{X^TX + kI\}^{-1} X^TY \; ; \; k \geq 0 \qquad (5)$$

Expresiones alternativas para esta ecuación, se encuentran realizando las siguientes definiciones

$$W = (X^TX + kI)^{-1} , \quad Z = WX^TX$$

Observa que existen dos expresiones útiles para Z

$$1) \; Z = WX^TX = (X^TX + kI)^{-1} (X^TX + kI - kI) = I - k(X^TX + kI)^{-1} \qquad (6)$$

$$2) \; Z = WX^TX = \{(X^TX)^{-1} W^{-1}\}^{-1} = \{(X^TX)^{-1} (X^TX + kI)\}^{-1} = \{I - k(X^TX)^{-1}\}^{-1} \qquad (7)$$

estas expresiones, conducen a otra expresión muy útil para el estimador de regresión ridge

$$\hat{\beta}(k) = W X^TY = Z\hat{\beta} \qquad (8)$$

La matriz "Z", hace las veces de una ponderación de los estimadores clásicos $\hat{\beta}_i$, y por lo general las reduce en magnitud, por lo que es llamado de "Factor de Reducción de los β_i".

LEY DE LOS COSENOS EN LA DESCOMPOSICION DE LA VARIACION TOTAL

La no ortogonalidad de la matriz X, lleva a aplicar, la famosa ley de los cosenos, para descomponer la suma de cuadrados del total de la siguiente manera:

$$Y^TY = \|Y - XZ\hat{\beta} + XZ\hat{\beta}\|^2 = \left[(Y - XZ\hat{\beta}) + XZ\hat{\beta} \right]^T \left[(Y - XZ\hat{\beta}) + XZ\hat{\beta} \right]$$

$$= (Y\text{-}XZ\hat{\beta})^T(Y\text{-}XZ\hat{\beta}) + (XZ\hat{\beta})^T XZ\hat{\beta} + (Y\text{-}XZ\hat{\beta})^T XZ\hat{\beta} + (XZ\hat{\beta})^T (Y - XZ\hat{\beta})$$

$$= \|Y - XZ\hat{\beta}\|^2 + \|XZ\hat{\beta}\|^2 + 2\,(Y - XZ\hat{\beta})^T XZ\hat{\beta}$$

$$= \text{SC. Residual} + \text{SC. Regresión Ridge} + \text{Non} \qquad (9)$$

donde:

$$\text{Cos}\,\theta = \frac{(Y - XZ\hat{\beta})^T Z\hat{\beta}}{\left|Y - XZ\hat{\beta}\right| \left|XZ\hat{\beta}\right|} = \frac{\text{Non}/2}{\left\{\|Y - XZ\hat{\beta}\|^2 \; \|XZ\hat{\beta}\|^2\right\}^{1/2}} \qquad (10)$$

también observe que:

$$\text{Cos}\,\theta = \begin{cases} > 0 \text{ si } \theta < 90^{\circ} \\ = 0 \text{ si } \theta = 90^{\circ} \\ < 0 \text{ si } \theta > 90^{\circ} \end{cases}$$

La descomposición de la suma del total, según la ley de los cosenos, usando el estimador de regresión ridge, involucra a tres componentes no ortogonales: una suma de cuadrados debido al residual, otra debido a la regresión ridge, y finalmente otra debida a la parte no ortogonal de los dos vectores; $(Y - XZ\hat{\beta})$ y $XZ\hat{\beta}$;

$$Y^TY = (Y - XZ\hat{\beta})^T (Y - XZ\hat{\beta}) + \hat{\beta}^T Z^T X^T XZ\hat{\beta} + 2\,(Y - XZ\hat{\beta})^T XZ\hat{\beta} \qquad (11)$$

De otro lado, el uso de la Regresión Ridge origina el problema de asignar grados de libertad fraccionarios, a cada fuente de variación, estudios preliminares relacionados a este punto se encuentran en; Carmer and Hsieh (1978) Hoerl and Kennard (1990), y Ordoñez (1993).

La no ortogonalidad de la matriz X^TX, obliga a distribuir una parte de los grados de libertad correspondientes a la regresión, entre las otras fuentes restantes. Así los grados de libertad debido a la regresión, será fraccional y menor que "p".

La determinación es realizada en forma análoga a la confección del cuadro del análisis de variancia; sin embargo se requiere definir el concepto de grados de libertad en forma general.

En el cuadro Nº 1, se observa que el número de los grados de libertad ν_h, es el divisor de las sumas de cuadrados respectivamente, en la columna de los cuadrados medios; de forma que el cociente en ausencia de los β_i, son simplemente los estimadores insesgados para σ_ε^2. Esto es, sobre la hipótesis nula; $H_0 : \beta = 0$,

$$E\left\{SC_h/\nu_h\right\} = \sigma_\varepsilon^2 \text{, para cada h.} \tag{12}$$

o equivalentemente

$$\nu_h = E\{SC_h\}/\sigma_\varepsilon^2 \text{, para cada h.} \tag{13}$$

Por tanto, sobre la suposición de que la hipótesis nula $H_0 : \beta = 0$ sea verdad, entonces el multiplicador de σ_ε^2, en E {S. C_h} constituyen los grados de libertad. Este principio se aplicará en la sección siguiente, para obtener los grados de libertad para cada una de las fuentes de variabilidad, consideradas en el cuadro del análisis de la varianza.

Cuadro Nº 1: ***Tabla tradicional del análisis de varianza***

TERMINO EN EL MODELO	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	CUADRADO MEDIO
1. T_1	ν_1	$S.C_1$	$S.C_1/\nu_1$
2. T_2	ν_2	$S.C_2$	$S.C_2/\nu_2$
.	.	.	.
.	.	.	.
.	.	.	.
h. T_h	ν_h	$S.C_h$	$S.C_h/\nu_h$
Total	n-1	Y^TY	

Grados de libertad para el residual

La aplicación de la ecuación (13), sugiere el cálculo de la esperanza de las sumas de cuadrados para cada fuente y evaluarla sobre la hipótesis nula.

Bajo las siguientes suposiciones:

$$Y \sim \mathcal{N}_n \{X\beta;\ \sigma_\varepsilon^2\ I\}$$

$$\hat{\beta} \sim \mathcal{N}_p \{\beta; (X^TX)^{-1}\ \sigma_\varepsilon^2\}$$

Y usando la expresión de "Z" en la ecuación (6), se tiene:

$$Y - XZ\hat{\beta} = Y - X\,[I - k\,(X^TX + kI)^{-1}]\,\hat{\beta}$$

$$= (Y - X\hat{\beta}\,) + kX\,(X^TX + kI)^{-1}\,\hat{\beta} = (Y - X\hat{\beta}) + kXW\hat{\beta}$$

Luego:

$$(Y-XZ\hat{\beta})^T\,(Y-XZ\hat{\beta}) = (Y-X\hat{\beta})^T\,(Y-X\hat{\beta}) + kY^TXW\hat{\beta} - k\hat{\beta}^TX^TXW\hat{\beta} + k\hat{\beta}^TWX^TY$$

$$-k\hat{\beta}^TX^TXW\hat{\beta} + k\hat{\beta}^TWX^TX\hat{\beta} + k^2\hat{\beta}^TWX^TXW\hat{\beta}$$

y por uso de la ecuación (7), se obtiene:

$$= (Y - X\hat{\beta}\,)^T\,(Y-X\hat{\beta}) + k^2\hat{\beta}^T\,Z^T\,(X^TX)^{-1}Z\hat{\beta}$$

$$= Y^T\,\{I-X\,(X^TX)^{-1}X^T\}\,Y + k^2\hat{\beta}^T\,Z^T\,(X^TX)^{-1}Z\hat{\beta}$$

Por tanto aplicando el operador de esperanza para esta última expresión:

$$E\,\{\,(Y - XZ\hat{\beta}\,)^T\,(Y-XZ\hat{\beta})\,\} = (n\text{-}p\text{-}1)\,\sigma_\varepsilon^2 + k^2\,\{Tr[Z^T\,(X^TX)^{-1}Z\,(X^TX)^{-1}]\,\sigma_\varepsilon^2$$

$$+\ \beta^T\,Z^T\,(X^TX)^{-1}\,Z\beta\,\}$$

$$= (n\text{-}p\text{-}1)\,\sigma_\varepsilon^2 + k^2\,Tr\,[W^TW]\,\sigma_\varepsilon^2 + k^2\,\beta^T\,Z^T\,(X^TX)^{-1}Z\beta$$

$$= (n\text{-}p\text{-}1)\,\sigma_\varepsilon^2 + k^2\sum_{i=1}^{p}\frac{1}{(\lambda_i + k)^2}\,\sigma_\varepsilon^2 + k2\,\beta^T\,Z^T\,(X^TX)^{-1}Z\beta \qquad (14)$$

Y sobre la hipótesis nula $H_0 : \beta = 0$;

$$E\,\{SC.\ Residual\} / \sigma_\varepsilon^2 = n-m-1 \tag{15}$$

donde:

$$m = p - k^2 \sum_{i=1}^{p} \frac{1}{(\lambda_i + k)^2}$$

Grados de libertad para la Regresión Ridge

En forma análoga el procedimiento de la sección anterior, se necesita hallar la esperanza de la suma de cuadrados debido a la Regresión Ridge

$$E\{\,||XZ\hat{\beta}||^2\} = E\,\{[XZ\hat{\beta}]^T\ [XZ\hat{\beta}]\} = E\{\ \hat{\beta}^T Z^T X^T XZ\hat{\beta}\}$$

$$= \text{Traza}\ \{\ Z^T X^T XZ\ (X^T X)^{-1}\}\ \sigma_\varepsilon^2\ +\ \beta^T Z^T X^T XZ\beta$$

$$= \text{Traza}\ \{\ \Lambda\ (\Lambda + k\,I)^{-1}\ \Lambda\ (\Lambda + k\,I)^{-1}\ \}\ \sigma_\varepsilon^2 + \beta^T Z^T X^T XZ\beta$$

y sobre la suposición de que la hipótesis $H_0 : \beta = 0$, es cierta se obtiene que:

$$E\left\{\frac{S.C.Regr.Ridge}{\sigma_\varepsilon^2}\right\} = \sum_{i=1}^{p} \frac{\lambda_i^2}{(\lambda_i + k)^2} = q \tag{16}$$

Grados de libertad para la componente no ortogonal

De la ecuación (9), la suma de cuadrados debido a la componente no ortogonal es:

$$NON = 2\,[Y-X\hat{\beta}(k)]^T X\hat{\beta}\,(k) = 2\,[Y-X\hat{\beta}\,(k)]^T XZ\hat{\beta} = 2\,[Y^T XZ\hat{\beta} - \hat{\beta}^T Z^T X^T XZ\hat{\beta}]$$

$$= 2\,[\ \hat{\beta}^T Z^T X^T XZ\hat{\beta} - \hat{\beta}\ (I+kW)\ X^T XZ\hat{\beta}\ ] = 2k\hat{\beta}^T\ [WX^T XZ]\ \hat{\beta}$$

$$= 2k\hat{\beta}\ [I + k\ (X^T X)^{-1}]^{-1}\ Z\hat{\beta} = 2k\hat{\beta}^T Z^T Z\hat{\beta} \tag{17}$$

Por tanto:

$$E\{NON\} = 2k\, E\{\hat{\beta}^T Z^T Z \hat{\beta}\} = 2k\{Tr\{Z^T Z (X^T X)^{-1}\} \sigma_\varepsilon^2 + \beta^T Z^T Z\beta$$

$$= 2k\, \sigma_\varepsilon^2\ Tr\{\Lambda\, (\Lambda + kI)^{-2}\} + 2k\, \beta^T Z^T Z\beta$$

$$= 2k\ \ \sigma_\varepsilon^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda + k)^2} + 2k\hat{\beta}^T Z^T Z\beta$$

y sobre la suposición de que la hipótesis nula es cierta:

$$E\left\{\frac{NON}{\sigma_\varepsilon^2}\right\} = 2k \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} = 2kr \qquad (18)$$

Observe que los grados de libertad del total, también obedece a la regla aplicada para las otras fuentes de variación, y además verifica la aditividad del modelo; esto es

$$(n\text{-}1) = (n\text{-}1\text{-}m) + q + 2kr, \qquad (19)$$

Así la descomposición de la "Ley de los Cosenos", produce un cuadro análogo al del teorema de "Pitágoras":

Cuadro N°2: Anva según la "Ley de los Cosenos"

FUENTES DE VARIACION	G.L.	SUMA DE CUADRADOS	VALOR ESPERADO DE LOS CUADRADOS MEDIOS
Regresión	q	$\lvert X\hat{\beta}(k)\rvert^2$	$q\,\sigma^2 + \beta^T Z^T X^T XZ\beta$
Sesgo	2kr	$2k\,\lvert\hat{\beta}(k)\rvert^2$	$2kr\sigma^2 + 2k\,\beta^T Z^T Z\beta$
Residual	n-1-m	$\lvert Y - X\hat{\beta}(k)\rvert^2$	$(n\text{-}1\text{-}m)\sigma^2 + k\,\beta^T Z^T (X^T X)^{-1} Z\beta$
Total	n-1	$\lvert Y\rvert^2$	$(n\text{-}1)\sigma^2 + \beta^T X^T X\beta$

APLICACION: DATOS ESTATURA DE UN BEBE

Para ilustrar el comportamiento de los estimadores de Regresión Ridge en la construcción del cuadro del ANVA, vía la aplicación de la "Ley de los Cosenos"; se usará un conjunto de datos referentes a un bebé, nacido el 01/02/1993 en la clínica de San Pablo; los datos tomados por el pediatra fueron: la estatura, peso, edad y perímetro cefálico durante los meses de febrero a noviembre. El problema que se planteó, fue el de obtener un pronóstico de la estatura del bebé para el mes siguiente en función de otras variables;

$$\textit{Estatura} = f\,(\textit{edad, peso, y perímetro cefálico}) \qquad (20)$$

La matriz de correlaciones para este conjunto de datos fue:

$$X^T X = \begin{bmatrix} 1{,}0000 & 0{,}9806 & 0{,}9658 \\ & 1{,}0000 & 0{,}9916 \\ & & 1{,}0000 \end{bmatrix}$$

Cuyas raices características son:

$$\lambda_1 = 0{,}005693\,,\ \ \lambda_2 = 2{,}9587\,,\ \ \lambda_3 = 0{,}0356$$

la magnitud λ_1, está indicando que el problema de multicolinealidad está presente en la matriz de datos y en consecuencia se usará el estimador Ridge con $k = \lambda_{min}$ para combatir los efectos perturbadores de la colinealidad, los resultados se muestran en los siguientes cuadros:

Cuadro Nº 3: Estimados de los parámetros (forma correlación)
Datos: estatura de un bebé.

PARAMETRO EN MODELO	MINIMOS CUADRADOS K=0	REGRESION RIDGE K=0,005693
$\hat{\beta}_1$	- 0,273875	- 0,130204
$\hat{\beta}_2$	0,975851	0,689624
$\hat{\beta}_3$	0,290499	0,433115

Cuadro Nº 4: Estimados de parámetros (forma original)
Datos: estatura de un bebé.

PARAMETRO EN MODELO	MINIMOS CUADRADOS K=0	REGRESION RIDGE K=0,005693
$\hat{\beta}_0$	17,7774	11,5544
$\hat{\beta}_1$	- 0,6261	- 0,2977
$\hat{\beta}_2$	3,5932	2,5393
$\hat{\beta}_3$	0,5685	0,8476

La forma original es obtenida de la forma correlación o viceversa, aplicada la siguiente relación:

$$\hat{\beta}(k) = \hat{\beta}^{*}(k)\ \{S.C_{YY} / S.C_{XX}\}^{1/2} \qquad (21)$$

Cuadro Nº 5: Anva para los datos, estatura de un bebé
Mínimos Cuadrados Ordinarios, k = 0

FUENTES DE VARIACION	G.L.	SUMA DE CUADRADOS	CUADRADO MEDIO	RAZON F
Regresión	3,0	505,2176	168,4059	595,761
Residual	7,0	1,9787	0,28267	
Total	10,0	507,1963		

Regresión Ridge k = 0,005693

FUENTES DE VARIACION	G.L.	SUMA DE CUADRADOS	CUADRADO MEDIO	RAZON F
Regresión	1,98954	500,8474	251,7398	292,017
Sesgo	0,74146	0,0826	8,9804	
Residual	7,26899	6,2664	0,8621	
Total	10,00000	507,19637	θ = 89,96	

Las ecuaciones para obtener los pronósticos se obtienen, desde el cuadro Nº 4, y son:

$$\hat{Y} = 17,7774 - 0,6261X_1 + 3,5932X_2 + 0,5685X_3,$$

$$\hat{Y}(k) = 11,5584 - 0,2977X_1 + 2,5393X_2 + 0,8476X_3,$$

Conclusiones y recomendaciones

Las siguientes conclusiones se reportan para el presente trabajo:

1. La razón "F", para la Regresión Ridge se reduce en aproximadamente el 104,02%.

2. Según el punto Nº 1, la aplicación del teorema de Pitágoras, en presencia de la multicolinealidad, puede conducir a tomar decisiones incorrectas en la prueba de hipótesis $H_0 : \beta = 0$.

3. La razón "F" para Regresión Ridge debe obtenerse mediante integración numérica de la expresión:

$$P\{(Y^T AY/Y^T BY) \leq z\} = P\{Y^T (A-zB)Y \leq 0\}$$

4. Investigaciones adicionales se requiere para determinar la distribución exacta de la razón "F" para el caso de la Regresión Ridge.

REFERENCIAS

[1] **Carmer S. and Hsieb W. (1978)**. *"A simulation study of five biased for straigbt line regression" Comm. Statit. "Simulation and Computation" B7(6), Pág. 529-548.*

[2] **Hocking R. R. (1976)**. *"Tbe analysis and selection of variables in linear regression". Biometrics Vol. 32; Pág. 1-49.*

[3] **Hoerl A. and Kennard R. (1970)**. *"Ridge regression: Biased estimation for non ortbogonal problems". Tecbonmetrics Vol 12; Pág. 55-67.*

[4] **Hoerl A. and Kennard W. (1970)**. *"Ridge regression: Degrees of fredom in tbe analysis of variance". Communications in Statistics Simulation 19 (4), Pág. 1485-1495.*

[5] **Ordoñez M. A. (1993)**. *"Regresion ridge lineal simple", XI - Coloquio Nacional de Matemáticas; 08-12 Nov. 1993; Huaraz - Perú.*