Universidad Técnica Federico Santa María
Departamento de Matemática.

Certamen MAT-266 Análisis de Regresión

Profesor: Felipe Osorio. 7 de diciembre de 2011.

1. (20 puntos) Muestre que si $\boldsymbol{X}$ es de rango completo
$$\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 = \|\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\boldsymbol{X}^T\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}),$$
desde ahí deduzca que $Q(\boldsymbol{\beta}) = \|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2$ es minimizado para $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}$.

2. (20 puntos) Considere el modelo
$$Y_{ij} = \theta_i x_j + \epsilon_{ij}, \qquad i=1,2; j=1,\dots,T,$$
con $\{\epsilon_{ij}\}$ variables aleatorias independientes $\mathcal{N}(0,\sigma^2)$ y $\{x_i\}$ constantes conocidas. Obtenga el estadístico F para probar $H_0 : \theta_1 = \theta_2$.

3. (20 puntos) Desde el ajuste de todas las regresiones posibles a un conjunto de 13 datos y 4 regresores, se obtuvo la siguiente información:

RSS_p	R_p^2	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$
2715.764	0.000	95.42				
1265.687	0.534	81.48	1.87			
906.336	0.666	57.42		0.79		
1939.401	0.286	110.21			-1.26	
883.867	0.675	117.57				-0.74
57.905	0.979	52.58	1.47	0.66		
1227.072	0.548	72.35	2.31		0.49	
74.762	0.972	103.10	1.44			-0.61
415.442	0.847	72.08		0.73	-1.01	
868.880	0.680	94.16		0.31		-0.46
175.738	0.935	131.28			-1.20	-0.72
48.111	0.982	48.19	1.70	0.66	0.25	
47.973	0.982	71.65	1.45	0.42		-0.24
50.836	0.981	203.64		-0.92	-1.45	-1.56
73.815	0.973	111.68	1.05		-0.41	-0.64
47.864	0.982	62.41	1.55	0.51	0.10	-0.14

   a) Encuentre el mejor subconjunto de regresores usando los critérios s_p^2, C_p y Akaike.
   b) Obtenga el mejor subconjunto de regresores mediante el procedimiento stepwise. Use $F_{IN} = F_{OUT} = 4{,}0$.

4. (20 puntos) Considere la inclusión de una nueva variable regresora $\boldsymbol{Z}$ como
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \gamma\boldsymbol{Z} + \boldsymbol{\epsilon} \tag{1}$$
con $\mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ y $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_n$. Tenemos $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$, multiplique ambos lados de (1) por $(\boldsymbol{I}-\boldsymbol{H})$ y obtenga un estimador para γ.

5. (20 puntos) Sea $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ y $z_{ij} = x_{ij} - \overline{x}_j$, para todo $i = 1, \dots, n$; $j = 1, \dots, k$ y considere $\boldsymbol{Z} = (z_{ij})$, $\boldsymbol{\beta}_{(0)} = (\beta_1, \dots, \beta_k)^T$. De este modo, tenemos el *modelo centrado* dado por:

$$\begin{aligned}\boldsymbol{Y} &= \gamma_0 \mathbf{1} + \boldsymbol{Z}\boldsymbol{\beta}_{(0)} + \boldsymbol{\epsilon} \\ &= (\mathbf{1}, \boldsymbol{Z}) \begin{pmatrix} \gamma_0 \\ \boldsymbol{\beta}_{(0)} \end{pmatrix} + \boldsymbol{\epsilon}.\end{aligned}$$

Para $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ y $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$, obtenga el estimador mínimos cuadrados de γ_0 y $\boldsymbol{\beta}_{(0)}$. Calcule también la matriz de covarianza de $\widehat{\boldsymbol{\theta}} = (\widehat{\gamma}_0, \widehat{\boldsymbol{\beta}}_{(0)}^T)^T$. ¿Son $\widehat{\gamma}_0$ y $\widehat{\boldsymbol{\beta}}_{(0)}$ independientes?