Pub. Mat. UAB
N° 22 Nov. 1980
Actes VII JMHL

COMPARACION DE ESTIMADORES DE LOS PARAMETROS EN UN MODELO LINEAL

Alfredo García Olaverri

Dpto. de Cálculo de Probabilidades y Estadistica Matemática
Universidad de Zaragoza

S U M M A R Y

This paper investigates the behaviour of residuals from linear model $Y = X\beta + u$, when we use estimates β_1^c, found for the least weigthed sum of residuals by c or c-1 , $c \in [0,1]$, depending of its sign. Anagously, we work out estimates wich are a linear combination from β_1^c .

The emphasis here is comparing probability plots of residuals --- from this estimation procedure with least-squares residuals .

. . . .

En el modelo lineal

$$Y = X\beta + u \qquad (1)$$

dadas n observaciones de las variables Y , x^j , J=2,...,K definimos para cada $c \in [0,1]$ como estimador β_1^c o cuantil c aquel que minimiza la función

$$S^c(\beta) = \sum_{i\in S} u_i\, c - \sum_{i\in I} u_i (1-c)$$

siendo

$$u_i = y_i - \beta x_i' \quad , \quad i=1,\ldots,n$$

$$S = \{i \mid u_i > 0\} \ , \quad I = \{i \mid u_i < 0\}$$

En [4] se encuentran métodos de obtención sistemáticos de estos estimadores β_1^c, demostrándose que son un número finito m y que podemos descomponer el intervalo $[0,1]$ en m subintervalos $H_i = [c_i, c_{i+1}]$, $i=0,m-1$, tal que todo $c \in H_i$ tiene el mismo estimador β_i , siendo en el punto c_i estimadores tanto β_i como β_{i-1} .

Así mismo si es $\delta_i = c_{i+1} - c_i$ se definen entre otros los siguientes estimadores :

c-mínimo : El cuantil en el intervalo $[c_0, c_1]$ es decir β_0 .

c-máximo : El cuantil en $[c_{m-1}, c_m]$, es decir β_{m-1}

c-medio : $\beta = \sum_{i=0}^{m-1} \beta_i \, \delta_i$

Gastwirth: $\beta = 0.3\,\beta_i + 0.4\beta_j + 0.3\beta_k$, donde i,j,k se eligen de forma que $c_i \leq 1/3 < c_{i+1}$, $c_j \leq 1/2 \leq c_{j+1}$, $c_k \leq 2/3 < c_{k+1}$

1/2(c-mínimo+c-máximo) : $\beta = 1/2\,(\beta_0 + \beta_{m-1})$

c-mediano: El β_i tal que $0.5 \in [c_i, c_{i+1}]$

Como señala Andrews [1] , [2] , el estimador mínimo cuadrático da pocas indicaciones de cuando el modelo no es adecuado, bien por existir errores groseros en las observaciones, o bien no seguir los residuales una ley normal, o no ser lineal el modelo etc., así si simulamos datos con alguna de estas características, el estimador mínimo cuadrático -- tiende a difuminarlas .

Si suponemos que los residuales u_i son una muestra de tamaño n de una distribución F , para comprobarlo hacemos gráficos de probabilidad, ver por ej. [3] , estos gráficos se obtienen trazando los puntos ---- (u_i, E_i), donde consideraciones computacionales aconsejan tomar

$$E_i = F^{-1}((i-\delta)/(n+1-2\delta))$$

siendo usualmente $0 \leq \delta \leq 1/2$.

Si los u_i siguen la ley F (usualmente normal), el gráfico tiende a ser una línea recta, si la configuración es distinta se toma como evidencia de lo inadecuado del modelo y este ha de ser modificado .

Hemos estudiado estos estimadores para los datos que Andrews trata en [1] (21 datos, 4 variables), el estimador mínimo cuadrático es

$$y = -3929+0.72x_1 + 1.30x_2 - 0.15x_3$$

Andrews propone

$$y = -37.2 + 0.82x_1 + 0.52x_2 - 0.07\, x_3$$

Nosotros obtenemos para los estimadores (2)

c-mínimo : $y = -29.014+0.315x_1 + 1.2224x_2 - 0.028x_3$

mediano : $y = -39.690 + 0.832x_1 + 0.574x_2 - 0.061\, x_3$

c-máximo : $y = -58.462 + 0.525x_1 + 1.858x_2 + 0.106\, x_3$

c-medio : $y = -44.094 + 0.688x_1 + 1.028x_2 - 0.019\, x_3$

Gastwirth: $y = -43.655 + 0.847x_1 + 0.743x_2 - 0.064\, x_3$

1/2 (c-mínimo+c-máximo) : $y = -43.738+0.420x_1 +1.541x_2 + 0.039\, x_3$

Los gráficos de probabilidad muestran muy claramente 4 puntos fuera de la línea general, eliminados estos obtenemos los nuevos estimadores :

mediano : $y = -35.939 + 0.823x_1 + 0.438x_2 - 0.070x_3$

Gastwirth: $y = -36.938 + 0.821x_1 + 0.449x_2 - 0.060\, x_3$

Mínimo cuadrático : $y = -37.6 + 0.80x_1 + 0.58x_2 - 0.07\, x_3$

Los otros estimadores varían bastante al eliminar esos cuatro puntos .

Es de destacar su proximidad al estimador robusto propuesto por --Andrews .

Hemos estudiado igualmente estos estimadores para datos procedentes de un modelo lineal con residuales no normales y para datos procedentes de un modelo no lineal con residuales normales, comparando los estimadores definidos anteriormente con el de mínimos cuadrados .

Los estimadores mínimo y máximo puede diferir mucho de los que consideramos razonables. El estimador c-medio es siempre muy próximo al mínimo cuadrático, siendo pues los más interesantes los estimadores mediano y Gastwirth, estos señalan más claramente la no adecuación del modelo en el caso que este tenga errores groseros, pero no se detecta ventaja ---- apreciable cuando el modelo es no lineal o con residuales no normales .

BIBLIOGRAFIA

1 ANDREWS D.F.(1974) : "A robust method for multiple linear regression" Technometrics, Vol. 16, no. 4 .

2 ANDREWS D.F. (1979) : "The robustness of residual displays" Ed. Academic Press .

3 DANIEL AND WOOD (1971) : "Fitting equations to data". Ed. J. Wiley .

4 GARCIA A. (1980) : "Algunos problemas en modelos de regresión". Tesis Doctoral .

5 RAO C.R. (1965) : "Linear statistical inference and its applications" Ed. J. Wiley .