

Trabajo completo

Estudio de la validez de los supuestos en los modelos lineales mixtos mediante un análisis de residuos

RECIBIDO: 04/10/2011

ACEPTADO: 26/07/2012

Garcia, M. del C. • Rapelli, C.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario. Bvard Oroño 1261. (2000) Rosario, Santa Fe, Argentina.

Teléfono: 0341-4802794-int.152. E-mail: mgarcia@fcecon.unr.edu.ar

RESUMEN: Los modelos mixtos se utilizan para analizar datos longitudinales. El uso apropiado de estos modelos requiere que los supuestos subyacentes en el análisis sean validados tan cuidadosamente como sea posible, incluyendo los supuestos distribucionales. Los residuos son de gran importancia para detectar violaciones de los supuestos involucrados en el modelo. En los modelos mixtos los residuos marginales, condicionales y los BLUP, que predicen los correspondientes errores y los efectos aleatorios respectivamente, se utilizan para verificar homocedasticidad de varianzas, presencia de outliers y normalidad de los errores. Los residuos con confusión mínima fueron propuestos recientemente para comprobar la normalidad del error condicional y su uso no está muy difundido entre los investigadores. En este trabajo se aborda el estudio de

los supuestos de un modelo utilizando datos obtenidos de una investigación para determinar la seguridad cardiológica de una nueva droga. Se realizan gráficos probabilísticos normales con bandas de confianza simuladas (“envelopes”) para comparar el comportamiento de los residuos con confusión mínima y los residuos condicionales clásicos en la evaluación del supuesto de normalidad del error condicional.

PALABRAS CLAVE: Estudios longitudinales, Modelos lineales mixtos, Diagnósticos, Residuos.

SUMMARY: *Study of the validity of linear mixed model assumptions through the analysis of residuals*

Mixed models are frequently used in the analysis of longitudinal data. The appropriate

use of these models requires checking the validity of the underlying assumptions, including the distributional assumptions. Residuals are used to examine the validity of model assumptions. In mixed models, the marginal and conditional residuals and BLUP are used to verify homoscedasticity, the presence of outliers and the normality of errors. The least

confounded residuals are useful to evaluate the normality of the conditional error. In this paper, we examine the assumptions of a model of evaluating the cardiology security of a drug.

KEYWORDS: longitudinal data, linear mixed models, diagnostics, residuals.

1. Introducción

Los modelos mixtos constituyen una valiosa herramienta para analizar datos compuestos por mediciones repetidas y en particular datos longitudinales. Para evaluar la validez de estos modelos se utilizan residuos, cuyo análisis se usa frecuentemente para comprobar el cumplimiento de los supuestos de los modelos estadísticos. En los modelos mixtos se distinguen los residuos marginales, condicionales y los BLUP, que predicen los correspondientes errores y los efectos aleatorios respectivamente. Cada residuo, estandarizado o no, se usa para comprobar algún supuesto. Los dos últimos residuos mencionados presentan confusión en el sentido que no sólo son función del error que predicen y pueden ser inadecuados para evaluar algún supuesto sobre los mismos (1). En los últimos años se propuso, para minimizar el efecto de la confusión, una transformación lineal de los residuos condicionados, surgiendo los residuos con confusión mínima que son resi-

duos condicionales que contienen una mínima fracción de confusión (1).

En este trabajo se presentan técnicas de diagnóstico para realizar el chequeo de los supuestos del modelo. Para la evaluación del supuesto de normalidad del error condicional se realiza un gráfico probabilístico normal con bandas de confianza simuladas (“envelopes”) de los usuales residuos condicionales y de los residuos con confusión mínima, sugeridos recientemente y cuyo uso no está muy difundido entre los investigadores. Se aplican en un estudio para comprobar la seguridad cardiológica de una nueva droga.

2. Métodos estadísticos

En los estudios longitudinales las unidades experimentales se observan repetidamente en varias ocasiones. El modelo lineal mixto, adecuado para analizar ese tipo de datos, se puede expresar como,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (2.1)$$

siendo,

$$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_1, \dots, \mathbf{Y}'_N)$$

$$\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_N)$$

$$\mathbf{b} = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_N)$$

$$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_N)$$

$$\mathbf{Z} = \oplus_{i=1}^N \mathbf{Z}$$