



UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA
SECRETARIA DE CIENCIA Y TECNOLOGIA E INSTITUTOS DE INVESTIGACIONES

Resumen Ampliado

Jornadas Anuales

“Investigaciones en la Facultad”

Ciencias Económicas y Estadística



Altamirano, Marlén
Rapelli, Cecilia
Catalano, Mara

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística.

TÉCNICAS GRÁFICAS DE DIAGNÓSTICO PARA LOS MODELOS LINEALES MIXTOS¹

Resumen

Los modelos lineales mixtos son los más utilizados en la actualidad para analizar datos longitudinales por su flexibilidad para modelar las múltiples fuentes de variabilidad presentes en este tipo de datos. Esta flexibilidad hace que el proceso de construcción del modelo no sea una tarea sencilla. Éste proceso está compuesto por reiterados ajustes y evaluaciones de modelos con el objetivo de identificar una forma funcional que describa adecuadamente la evolución de la respuesta en el tiempo, su relación con covariables y la correlación presente entre los datos. Durante este proceso, se pueden utilizar diversas herramientas diagnósticas basadas en los residuos. Éstas no solo alertan de una posible falta de ajuste del modelo a los datos, sino también, brindan información acerca de las características de los mismos, que pueden incorporarse en el modelo. Este trabajo presenta un compendio de técnicas diagnósticas gráficas basadas en los residuos y describe su utilización en el proceso de construcción del modelo y en la evaluación de los supuestos distribucionales del mismo.

Palabras clave: Datos longitudinales, Modelo lineal mixto, Análisis de residuos

Abstract

Linear mixed models are currently the most widely used to analyze longitudinal data due to their flexibility to model the multiple sources of variability present in this type of data. This flexibility makes the model building process no easy task. This process is made up of repeated adjustments and evaluations of models in order to identify a functional form that adequately describes the evolution of the response over time, its relationship with covariates and the correlation present between the data. During this process, various residual-based diagnostic tools can be used. These not only alert to a possible lack of fit of the model to the data, but also provide information about the characteristics of the data, which can be incorporated into the model. This paper presents a compendium of graphical diagnostic techniques based on residuals and describes their use in the model construction process and in the evaluation of its distributional assumptions.

Keywords: Longitudinal data, lineal mixed model, Residual analysis

Introducción

Los estudios basados en datos longitudinales son frecuentemente utilizados en una amplia variedad de disciplinas. Estos datos están conformados por mediciones repetidas de una

¹ Trabajo elaborado en el marco del Proyecto ECO214, titulado: "Métodos estadísticos avanzados en investigación en ciencias biológicas", dirigido por Cristina Cuesta



variable de interés realizadas a través del tiempo a cada unidad o individuo. Los modelos lineales mixtos son los más utilizados para analizar datos con estas características por su flexibilidad para modelar las correlaciones existentes. Éstos permiten ajustar un modelo para la variable de interés (la respuesta) mediante una parte sistemática, que está formada por una combinación de características poblacionales que son compartidas por todas las unidades; y una parte aleatoria que refleja las múltiples fuentes de heterogeneidad y correlación entre y dentro de las unidades.

La complejidad de los modelos mixtos, causada por la presencia de efectos aleatorios y las diferentes estructuras de covariancia, hace que el proceso de selección del modelo no sea una tarea sencilla. Este proceso está compuesto por reiterados ajustes y evaluaciones de modelos con el objetivo de identificar una forma funcional que describa adecuadamente la evolución de la respuesta en el tiempo, su relación con covariables y la correlación presente en los datos. El uso de herramientas diagnósticas basadas en los residuos en el proceso de construcción del modelo permite conocer mejor el problema bajo investigación y así sugerir modelos adecuados.

En este trabajo se pretende ilustrar el uso de diferentes herramientas diagnósticas para seleccionar un modelo lineal mixto y evaluar el cumplimiento de los supuestos distribucionales del mismo.

Análisis de residuos

Debido a las diferentes fuentes de variabilidad presentes en los modelos lineales mixtos, se pueden definir distintos tipos de residuos: marginales, condicionales y de efectos aleatorios. Cada uno de ellos es de utilidad para evaluar algún supuesto del modelo. Los marginales se usan para estudiar la linealidad de la respuesta con respecto a las variables explicativas y para evaluar si la estructura de covariancias intra unidad es adecuada; los condicionales para comprobar la homocedasticidad y normalidad del error condicional y los residuos de efectos aleatorios para verificar la normalidad de los efectos aleatorios.

El análisis de residuos involucra técnicas gráficas que se utilizan para evaluar el cumplimiento de los supuestos del modelo. Su uso en las diferentes etapas del proceso de construcción es fundamental debido a que aportan información para la correcta especificación del mismo.

Verificar si la estructura de covariancia intra unidad es adecuada es un paso no trivial en el análisis. Puesto la estimación de todos los parámetros depende de su correcta especificación, el proceso de diagnóstico debe comenzar por la evaluación de este supuesto.

La especificación del modelo para la covariancia del error condicional se compone de un modelo para la correlación y otro para la variancia. El modelo para la variancia se evalúa mediante los residuos condicionales. Dado que estos residuos pueden tener diferentes variancias, algunos autores sugieren la utilización de los residuos condicionales estandarizados, que se obtienen dividiendo el residuo ordinario por una estimación de su desvío estándar.

El modelo para la correlación se puede evaluar utilizando los residuos de covariancia propuestos por Singer et al. (2016), que son una transformación de los residuos marginales. Valores grandes de estos residuos identifican las unidades para las cuales la estructura de correlación no es adecuada. Por otra parte, Fitzmaurice et al. (2004) y Pinheiro y Bates (2000) propusieron el uso del variograma de los residuos marginales como técnica gráfica para evaluar la estructura de la matriz de covariancias. Ésta herramienta no sólo permite identificar si la estructura seleccionada no es adecuada, sino que también brinda información acerca del motivo de la falta de ajuste permitiendo proponer modificaciones en el modelo.



UNR

El supuesto de linealidad en los modelos lineales mixto se evalúa a través de los residuos marginales, de la misma forma que en los modelos lineales clásicos, mediante el gráfico de dispersión de los residuos marginales estandarizados versus las covariables o los valores ajustados.

Habitualmente la normalidad del error condicional se comprueba utilizando un gráfico probabilístico normal de los residuos condicionales. Hilden-Minton (1995) y Singer et al. (2016) argumentan que los residuos condicionales están confundidos por otras fuentes de error ajenas y esto debilita su habilidad para detectar alejamientos de la normalidad. Hilden-Minton (1995) define los residuos con mínima confusión para comprobar este supuesto. De esta forma se elimina la influencia de la distribución de los efectos aleatorios sobre la distribución del error condicional.

Para evaluar la normalidad de los efectos aleatorios, Pinheiro y Bates (2000) proponen utilizar como herramienta de diagnóstico un gráfico probabilístico normal para cada uno de los efectos aleatorios predichos. Este gráfico presenta una desventaja, ya que evalúa la normalidad de los efectos aleatorios por separado (normalidad marginal), siendo necesario evaluar la normalidad multivariada. Singer et al. (2016) proponen evaluar este supuesto utilizando un gráfico probabilístico chi-cuadrado de la distancia de Mahalanobis entre los efectos aleatorios predichos y su esperanza.

Resultados

Las herramientas presentadas se ilustran utilizando un conjunto de datos provenientes de un estudio clínico realizado en un efector de salud de la ciudad de Rosario. El objetivo era comparar el efecto de dos drogas sobre la profundidad de la hipnosis lograda durante la cirugía en pacientes que tienen una alta probabilidad de padecer el síndrome de apnea obstructiva del sueño (SAOS). El estudio cuenta con 48 pacientes, a 25 de los cuales se les suministró "dexmedetomidina", y a los 23 restantes se les suministró "midazolam". La profundidad de la hipnosis se midió mediante el índice Biespectral (BIS). Este índice proporciona una medida del nivel de conciencia del paciente y toma valores entre 0 y 100, donde 0 representa la ausencia de función cerebral y 100 un nivel de alerta completa. A todos los pacientes se les midió el índice BIS en las mismas 5 ocasiones, al inicio del estudio (minuto 0) y luego cada 15 minutos hasta completar una hora de seguimiento.

El estudio se centró en la utilización de técnicas gráficas basadas en los diferentes tipos de residuos tanto en la construcción del modelo como en su diagnóstico, resultando las siguientes consideraciones:

- El uso de los gráficos de residuos de covariancia y el variograma estimados permitieron seleccionar de manera sencilla una estructura adecuada para modelar la correlación entre las mediciones repetidas. Ambas herramientas se complementan. La primera detecta los individuos para los cuales el modelo propuesto no es adecuado, y la segunda aporta información sobre las características de la dependencia entre las mediciones repetidas.
- El gráfico de los residuos condicionales estandarizados versus los valores ajustados del modelo permitió comprobar la homocedasticidad de los errores condicionales, tanto en la etapa de construcción del modelo para la covariancia como en la etapa diagnóstico del modelo final.
- El gráficos probabilístico normal de los residuos condicionales para evaluar el supuesto de normalidad de los errores condicionales, mostró un leve alejamiento de la normalidad en las colas. Debido a que estos residuos pueden presentar confusión, es decir que



UNR

puede estar influenciado por otro residuo diferente al que se evalúa, este gráfico presenta limitaciones. En este estudio, la utilización de residuos condicionales con mínima confusión sostiene el supuesto de normalidad.

- El uso de cualquiera de los gráficos propuestos para comprobar la normalidad de los efectos aleatorios resultó indistinto en esta aplicación debido a que se incluyó en el modelo un único efecto aleatorio. Cuando se incluyen más efectos aleatorios, los gráficos probabilísticos normales de los efectos aleatorios predichos no evalúan la normalidad multivariada como supone el modelo, sino la normalidad marginal. Se prefiere en ese caso el uso del gráfico probabilístico Chi-cuadrado de las distancias de Mahalanobis entre el vector de efectos aleatorios predichos y su esperanza.

REFERENCIAS BIBLIOGRÁFICAS

Diggle, P. J.; Heagerty, P.; Liang, K.; Zeger, S. L. (2001). *Analysis of Longitudinal Data*. Oxford Statistical Science Series, second edition.

Fitzmaurice, G. M.; Laird, N. M.; Ware, J. H. (2004). *Applied Longitudinal Analysis*. John Wiley & Sons.

García, M. C.; Rapelli, C. (2012). Estudio de la validez de los supuestos en los modelos lineales mixtos mediante un análisis de residuos. *Revista FABICIB*, vol. 16, pág. 11 a 20.

García, M. C.; Rapelli, C. (2018). Una comparación de métodos de diagnósticos para los modelos lineales mixtos. Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística.

Hilden-Minton, J. A. (1995). *Multilevel Diagnostics for Mixed and Hierarchical Linear Models*. University of California.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *The Annals of Statistics* 17, 624–642.

Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, 54, 570–582.

Nobre, J. S. (2004). *Métodos de Diagnóstico para Modelos Lineares Mistos*. Instituto de Matemática e Estatística da Universidade de Sao Paulo.

Nobre, J. S.; Singer, J. M. (2007). Residual Analysis for Lineal Mixed Model. *Biometrical Journal*, vol. 49, 6, 863-875.

Pinheiro, J. C.; Bates, D. M. (2000). *Mixed-Effects Model in S and S-plus*. Springer-Verlag.

Singer, J. M.; Rocha, F. M. M.; Nobre, J. S. (2016). Graphical Tools for Detecting Departures from Linear Mixed Model Assumptions and Some Remedial Measures. *International Statistical Review*, 0, 0, 1-35.

Verbeke, G.; Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.