

UNIVERSIDAD NACIONAL DE ROSARIO  
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA  
LICENCIATURA EN ESTADÍSTICA



ANTEPROYECTO DE TESINA  
**“Técnicas gráficas para la evaluación de supuestos en modelos  
lineales mixtos para datos longitudinales”**

Alumna:  
Altamirano Marlen

Directora:  
Lic. Rapelli Cecilia

Codirectora:  
Lic. Catalano Mara

Año 2021

## **1. INTRODUCCIÓN**

Los estudios basados en datos longitudinales son frecuentemente utilizados en una amplia variedad de disciplinas. Éstos consisten en medir repetidamente la variable de interés a cada uno de los individuos o unidades a través del tiempo, bajo diferentes condiciones experimentales permitiendo caracterizar el cambio en la respuesta y los factores que influyen dicho cambio.

Una característica distintiva de los estudios longitudinales es que las mediciones realizadas sobre un mismo individuo o unidad están correlacionadas positivamente, lo cual debe ser tenido en cuenta en el análisis para obtener inferencias válidas.

Para ajustar datos longitudinales se utilizan los modelos lineales mixtos, que permiten considerar las múltiples fuentes de correlación presentes en este tipo de datos.

El proceso de modelización comienza con un análisis exploratorio, mediante la utilización de gráficos, que permite examinar el comportamiento de los datos. Posteriormente, en base a la evidencia empírica, se proponen modelos con diferentes estructuras de variancias y covariancia. Si bien el análisis exploratorio guía la especificación del modelo para describir la estructura de los datos, el análisis de residuos es un complemento muy eficaz para la selección del modelo.

Luego de seleccionar un modelo es necesario verificar el cumplimiento de los supuestos distribucionales e identificar la presencia de posibles observaciones atípicas. Para ello se realiza un análisis de residuos. Si bien para los modelos lineales clásicos este análisis es sencillo, para los modelos mixtos se complejiza debido a la presencia de efectos aleatorios y diferentes estructuras de covariancias.

## **2. OBJETIVO**

El objetivo de esta tesina es presentar un compendio de técnicas diagnósticas gráficas para modelos mixtos. Describir su uso en el proceso de selección del modelo y en la evaluación del cumplimiento de los supuestos distribucionales del mismo y la identificación de posibles valores atípicos.

### **3. METODOLOGÍA**

Los modelos mixtos son útiles para modelar datos longitudinales debido a su flexibilidad para representar las múltiples fuentes de variación y correlación. Estos modelos reconocen las tres fuentes de variabilidad presentes en los datos longitudinales: la variación aleatoria entre los individuos, la variación biológica inherente a cada individuo y los errores de medición.

En los modelos para datos longitudinales es fundamental la correcta especificación tanto de la estructura media como de la covariancia. La mala especificación del modelo puede tener un impacto considerable sobre las propiedades asintóticas de los estimadores y, por consiguiente, en las inferencias que se realizan sobre ellos.

Una etapa importante en la modelización estadística es la evaluación del modelo ajustado, para lo cual se utilizan los residuos. Este proceso, conocido como análisis de residuos, consiste en evaluar el ajuste del modelo propuesto a un conjunto de datos y verificar el cumplimiento de los supuestos postulados sobre el mismo.

El uso de esta herramienta en el proceso de construcción del modelo permite conocer mejor el problema bajo investigación y así sugerir modelos adecuados.

La presencia de múltiples fuentes de variabilidad en los datos longitudinales permite definir distintos tipos de residuos y cada uno de ellos es de utilidad para evaluar algún supuesto del modelo ajustado. Los tres tipos de residuos, marginales, condicionales y de efectos aleatorios, se estandarizan dividiéndolos por los desvíos estándares estimados de los residuos o de la variable respuesta, surgiendo los residuos estudentizados o de Pearson respectivamente. Otras opciones para estandarizar también son posibles, una de ellas es usar la transformación de Cholesky.

El gráfico de estos diferentes residuos permite examinar distintos aspectos del modelo estimado. Con los residuos marginales se puede corroborar la linealidad de la variable respuesta sobre las covariables, evaluar la validez de la estructura de covariancia intra individuo y la presencia de observaciones atípicas.

Los residuos condicionales se utilizan para chequear la normalidad de los errores intra individuo, examinar la presencia de heterocedasticidad e identificar posibles observaciones atípicas.

A través del análisis de los residuos de efectos aleatorios se evalúan los supuestos distribucionales de los mismos, como la normalidad de los efectos aleatorios, y se identifican unidades atípicas.

#### **4. MATERIALES**

Los datos que se utilizarán provienen de un estudio clínico desarrollado para comparar el efecto de dos drogas utilizadas para sedación en una población de pacientes STOP bang positivos. STOP bang es un cuestionario que se utiliza para la detección temprana de pacientes con síndrome de apneas/ hipopneas obstructivas del sueño. En el estudio participaron 56 pacientes con 3 o más criterios STOP bang que tenían cirugías programadas con indicación de bloqueo central y sedación. Los pacientes fueron asignados aleatoriamente a dos grupos, al primero (MDZ) se le suministró una dosis única de midazolam de 0,05 mg/Kg y al segundo grupo (DEX) se le suministró en infusión continua la droga dexmedetomidina en 0,4 ug/kg/h. En cada paciente se evaluó la profundidad de hipnosis en 7 ocasiones distintas dentro de la hora de suministrada la droga correspondiente a cada grupo. Se registró una medida de interés, el índice BIS, con el fin de comparar la profundidad de hipnosis producidas por ambas drogas. El índice BIS es parámetro un electroencefalográfico que se utiliza en la actualidad para vigilar el nivel de hipnosis de los pacientes bajo sedación o anestesia. Ofrece una puntuación de 0 a 100, en la que 0 representa la ausencia de función cerebral y 100 un nivel de alerta completa del paciente. Por lo tanto, valores bajos del índice BIS indican que el paciente está completamente sedado.

Se ilustrará el uso del análisis de residuos tanto en el proceso de selección del modelo como en el diagnóstico del mismo.

Para el desarrollo del análisis de los datos se utilizará el software estadístico R.

#### **5. CRONOGRAMA**

- Revisión bibliográfica exhaustiva de los residuos para modelos lineales mixtos para datos longitudinales: Febrero/ Agosto 2019.

- Procesamiento y análisis de la información: Septiembre 2019 / Marzo 2020.
- Elaboración de informe: Abril / Septiembre 2020.
- Revisión de informe: Octubre 2020.
- Fecha tentativa de presentación: Marzo 2021.

## 6. BIBLIOGRAFÍA

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