

MÓDULO II

TÓPICOS ESPECIALES EN MODELOS LINEALES GENERALIZADOS

Propuesta de Curso de Postgrado

Motivación y objetivos del curso

Tanto en ciencias biológicas como en ciencias sociales, frecuentemente se analizan fenómenos que, por su distribución en probabilidades y/o sus características, no pueden ser modelados utilizando estadística paramétrica clásica. Los Modelos Lineales Generalizados (MLG) ofrecen una alternativa robusta y muy versátil respecto a la estadística clásica, permitiendo diversas distribuciones de errores, funciones de varianza, etc. En este sentido, los MLG permiten incluso modelar variables respuesta con una gran cantidad de ceros, y sobre o sub-dispersión en la variación de los datos.

Más allá de su flexibilidad y ventajas, los MLG aún requieren la independencia de las observaciones a ser consideradas. Los Modelos Lineales Generalizados Mixtos (MLGM) representan una herramienta analítica poderosa ya que permiten considerar diferentes distribuciones de errores -derivadas del ajuste de los modelos a la variable a ser analizada- a la vez que permiten la inclusión y modelado de observaciones no independientes. En la última década, los MLGM han sido muy utilizados tanto en diseños experimentales complejos (e. g. diseños de medidas repetidas, parcelas divididas) como al considerar la no independencia de los datos en conjuntos de datos jerárquicos.

El objetivo principal de este curso es que los participantes conozcan diferentes extensiones y tópicos candentes en MLG, particularmente los referentes al modelado de sobre y sub-dispersión en variables de conteo, exceso de ceros, modelos no lineales y la extensión de todos ellos a MLGM. A lo largo del curso, se utilizará el programa *R* como plataforma de análisis. En el curso, se brindarán las herramientas básicas necesarias para que los participantes puedan implementar MLGM con variables respuesta del tipo continua (distribución normal), discreta o de conteo (distribución Poisson y Binomial Negativa) y dicotómicas o de proporciones (distribución Binomial). Este curso no plantea indagar en la teoría estadística de los MLGM, sino que propone brindar un amplio panorama aplicado de sus virtudes y flexibilidad.

Profesor responsable: Dr. Walter S. Svagelj.

Total de horas: 40 (cuarenta). 20hs teóricas, 10hs prácticas, 10hs teórico-prácticas.

Formato de dictado: Lunes a Viernes, 8hs diarias.

Destinatarios del curso: principalmente dirigido a estudiantes y docentes de grado y posgrado en Cs. Biológicas y afines.

Programa analítico

1. Introducción a los Modelos Lineales Generalizados (MLG). Predictor lineal, distribución de errores y función de enlace. Prueba de hipótesis. Múltiples hipótesis y complejidad óptima del modelo. Selección de modelos por teoría de la información.

Criterio de Información de Akaike y sus derivados (AIC, AICc, QAICc). Nociones básicas de diseño experimental. *R* y RStudio, ¿Por qué utilizarlos?

2. Introducción a los Modelos Lineales Generalizados Mixtos (MLGM). Efectos fijos y aleatorios. Modelo Lineal General Mixto. Distribución normal de errores. Matrices de varianza-covarianza. Funciones de varianza para modelar heteroscedácea. Estructuras de correlación para modelar dependencia. Diseños anidados. Componentes de varianza.

3. Modelo No Lineal Mixto (MNLM) y sus aplicaciones. Curvas de crecimiento. Parametrización con relevancia biológica. Modelos logístico, Gompertz, von Bertalanffy y Richards. MNLM vs. MLGM.

4. Variables respuesta de tipo conteo. Distribución Poisson. Sobredispersión y distribución Binomial Negativa. Subdispersión de la variación. Modelo Poisson de Conway-Maxwell. Presencia y exceso de ceros en la variable respuesta. Modelos inflados y alterados en ceros. Modelo de vallas. Extensión a MLGM.

5. Variables respuesta del tipo dicotómica o de proporciones. Distribución binomial. Aplicación de MLGM y MNLM a estudios ecológicos y comportamentales. Introducción a los Modelos Aditivos Generalizados (MAG) y su extensión a Modelos Aditivos Generalizados Mixtos (MAGM). Perspectivas futuras.

Bibliografía

- Burnham, K. P., & D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Segunda edición. Springer, New York.
- Crawley, M. J. 2013. The R book. Segunda edición. John Wiley and Sons, Ltd., Chichester.
- Hilbe, J. M. 2011. Negative binomial regression. Segunda edición. Cambridge University Press, New York.
- Pinheiro, J. C., & D. M. Bates. 2000. Mixed-effects models in S and S-Plus. Springer, New York.
- R Development Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Wood, S. N. 2006. Generalized additive models. An introduction with R. CRC Press, Boca Raton.
- Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, & G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer, New York.

Cantidad máxima (propuesta) de participantes: 25

Requisitos de admisión para los participantes: aunque no es excluyente, es deseable que los participantes tengan conocimientos básicos de estadística y modelos lineales generalizados, así como instalados en sus computadoras los programas *R* y RStudio.

Metodología de enseñanza-aprendizaje: modalidad teórico-práctica. El curso constará de teoría impartida por el docente y actividades prácticas relacionadas con los conocimientos brindados. Las actividades prácticas se realizarán con los programas *R* y

RStudio, los cuales son serán provistos por el docente al inicio del curso.

Forma de evaluación: a convenir. Se propone que cada participante realice un trabajo final, en el cual deberá aplicar los conocimientos adquiridos en el curso a un conjunto de datos de su interés. El trabajo deberá ser entregado por correo electrónico (tiempo máximo de entrega: dos semanas desde la finalización del curso).

Recursos necesarios para el dictado del curso: para el dictado del curso se requiere un pizarrón y un cañón para proyectar las clases (el docente llevará su propia computadora portátil). Se requiere además que los participantes tengan a disposición ordenadores durante la clase (cantidad mínima de ordenadores: 1 ordenador cada dos participantes). Se sugiere que los participantes utilicen su propio ordenador (de existir esa posibilidad).