

## MÓDULO I

### INTRODUCCIÓN A LOS MODELOS LINEALES GENERALIZADOS: UNA APROXIMACIÓN APLICADA UTILIZANDO EL *R*

*Curso de Postgrado*

#### **Motivación y objetivos del curso**

En los últimos años se ha dado un crecimiento explosivo en la utilización de los Modelos Lineales Generalizados (MLG) aplicados a estudios ecológicos y comportamentales. Este tipo de modelos ofrece una alternativa robusta y muy versátil respecto a la “rígida” estadística paramétrica clásica. Sin embargo, la comprensión y correcta implementación de este tipo de análisis requiere de una mínima base conceptual y de un entrenamiento en programación que raramente son brindados a nivel de grado.

Por otro lado, en las disciplinas biológicas es muy frecuente la consideración de diferentes hipótesis en forma simultánea. Esta puesta a prueba de múltiples hipótesis implica un abanico de diferentes variables y modelos plausibles, siendo el proceso de selección de modelos candidatos un paso clave en la inferencia y correcta interpretación de los resultados. En este sentido, es relativamente frecuente que las conclusiones deban basarse en diferentes modelos, y no solo en un “clásico” mejor modelo obtenido en base a umbrales de significancia.

El objetivo principal de este curso de postgrado es brindar a los participantes una introducción básica aplicada a los MLG utilizando *R*, un entorno/paquete estadístico libre y gratuito. Asimismo, otro objetivo es brindar a los participantes una visión amplia de la importancia del proceso de selección de modelos en MLG. En el curso, se brindarán las herramientas para que los participantes puedan implementar MLG con variables respuesta de tipo conteo (distribución Poisson y Binomial Negativa) y dicotómicas o de proporciones (distribución binomial). Otro objetivo del curso es que los asistentes aprecien las ventajas de los MLG en relación a las clásicas estadísticas paramétrica y no paramétrica. Asimismo, se compararán los dos marcos de referencia típicos a la hora de realizar la inferencia: prueba de hipótesis y teoría de la información, discutiendo sus fortalezas y debilidades. Este curso no plantea indagar profundamente en la teoría estadística de los MLG, sino que propone brindar un amplio panorama aplicado de su flexibilidad y virtudes.

**Fecha de dictado:** 6 al 10 de abril de 2015.

**Profesor responsable:** Dr. Walter S. Svagelj.

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

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

**Forma de evaluación:** Se propone que cada alumno 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 via correo electrónico (tiempo máximo de entrega: dos semanas desde la finalización del curso).

## Programa analítico

1. ¿Qué es un modelo estadístico? Representación de la realidad, verosimilitud y discrepancia. ¿Qué es el *R*? ¿Un software gratuito de última generación para análisis estadísticos y gráficos? *R* y RStudio, ¿Por qué utilizarlos? Estadística paramétrica y no paramétrica, y su implementación en *R*.
2. Introducción a los Modelos Lineales Generalizados (MLG). Tipos de variable respuesta. Predictor lineal, distribución de errores y funciones de varianza. Función de enlace y función media. Modelo Lineal General vs. Modelos Lineales Generalizados. Ajuste del modelo. Gráficos diagnósticos. Estimación e inferencia.
3. Variables respuesta de tipo conteo (distribución Poisson y Binomial Negativa). Sobredispersión. Tasas en variables de conteo. Presencia y exceso de ceros. Variables respuesta del tipo dicotómica o de proporciones (distribución binomial).
4. Introducción a la selección de modelos. Múltiples hipótesis y complejidad óptima del modelo: sobre-ajuste y sub-ajuste. Principio de Parsimonia. Inferencia estadística utilizando prueba de hipótesis y teoría de la información. Similitudes y diferencias. Fortalezas y debilidades. Criterio de Información de Akaike y sus derivados (AIC, AICc, QAICc).
5. ¿Es correcto concluir sobre el modelo global? Conclusión basada en el mejor modelo: cuando y como. Selección de modelos e inferencia basada en múltiples modelos (MMI). Importancia relativa de los modelos y variables. Estimación de parámetros y errores en base a todos los modelos. Varianzas incondicionales. Generación de modelos candidatos *a priori*. Número y complejidad de los modelos candidatos. 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.
- Garamszegi, L. Z. 2011. Information-theoretic approaches to statistical analysis in behavioural ecology: an introduction. Behavioral Ecology and Sociobiology 65: 1-11.
- Hilbe, J. M. 2011. Negative binomial regression. Segunda edición. Cambridge University Press, New York.
- Logan, M. 2010. Biostatistical design and analysis using R: a practical guide. John Wiley and Sons, Ltd., Chichester.
- McCullagh, P., & J. A. Nelder. 1989. Generalized linear models. Segunda edición. Chapman & Hall, New York.
- R Development Core Team. 2014. 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>.