

# Métodos estadísticos avanzados en Ecología y Evolución.

## **Objetivo:**

Proporcionar una adecuada comprensión de los principios fundamentales de los diferentes métodos de análisis de datos comúnmente utilizados en Ecología y Evolución y adquirir la capacidad de utilizar en forma crítica y razonada estos métodos en sus actividades de investigación.

## **Docente:**

Pablo Inchausti, profesor titular de Ecología de la Universidad de la República, Uruguay.

## **Modalidad:**

El curso tendrá 6h de clase diarias (dos sesiones de 3h) durante 10 días de clase (60h en total). El curso será dictado en Español. Se empleará el programa de análisis estadístico gratuito R (y la interfase R-Studio) que se ha convertido en referencia mundial para el análisis de datos. Los estudiantes cursantes deberán traer sus laptops para emplearlas durante los prácticos de este curso. Todos materiales (clases, datos, scripts, libros y artículos) se pondrán a disposición de los estudiantes antes del comienzo del curso.

## **Requisitos:**

Este curso presupone que los estudiantes posean un conocimiento básico sobre el concepto de probabilidad, de las principales funciones de distribución de probabilidades (binomial, Poisson y normal), de nociones de estimación puntual y por intervalo (i.e. intervalos de confianza) de parámetros, del test de hipótesis estadísticas y su relación con los intervalos de confianza y del test t para diferencias de medias o proporciones. Estos temas tratados en otros cursos básicos de estadística **NO** serán cubiertos en este curso. Los estudiantes deberán además tener una familiaridad mínima (importar datos, hacer una regresión simple y hacer al menos un gráfico) con los programas R y eventualmente de R-Studio.

## **Cupos y destinatarios:**

Este curso es parte del Postgrado de Ciencias Biológicas, Facultad de Ciencias, Universidad de la República. Está principalmente destinado a estudiantes de Postgrado en Ecología, Evolución, Ciencias Ambientales y Agronomía. Se dictará con un mínimo de 5 estudiantes y habrá un cupo máximo de 30 estudiantes (debido a la logística de realización de prácticos). Los cursantes serán seleccionados en base a un CV y a una carta de motivación detallados en los que se indique claramente su formación previa en Estadística y su experiencia en análisis de datos.

**Fecha y lugar:** Centro Universitario Regional del Este, Universidad de la República, Maldonado, Uruguay. Inicio: 14 Febrero 2022, fin: 25 Febrero 2022.

## **Programa del curso:**

1. **Modelo Lineal General:** Estructura y suposiciones del Modelo Lineal General. El análisis de varianza, de regresión simple y múltiple y de covarianza como casos particulares del Modelo Lineal General. Magnitud del efecto (effect size) estimada en los Modelos Lineales Generales y su interpretación. Diseños factoriales: interacciones estadísticas entre variables explicativas, interpretación de resultados y tests a posteriori. Análisis de residuos: validación y evaluación de la calidad del ajuste de un modelo

estadístico.

**2. Modelación estadística:** El método de máxima de verosimilitud en la estimación de los parámetros y el test de hipótesis estadísticas. Propiedades fundamentales de los estimadores de máxima verosimilitud. Principios fundamentales de la Estadística Bayesiana. Los modelos estadísticos como hipótesis estadísticas y el test o selección simultánea de múltiples hipótesis estadísticas empleando la Teoría de la Información. Importancia de la parsimonia en estadística. Criterio de información de Akaike (AIC) y su uso en la selección de modelos estadísticos. Pesos de Akaike: interpretación y uso en la obtención del modelo estadístico “promedio”.

**3. Modelo Lineal Generalizado (GLM):** componentes de los GLM: predictor lineal de las variables explicativas, función de enlace y errores aleatorios. El GLM como marco teórico general para el análisis de datos univariados en escalas binaria (Binomial), de conteos (Poisson y Binomial Negativa) y continua (Normal, Beta, Gama, Lognormal, Tweedie). Sobredispersión de GLM para datos discretos: diagnóstico y soluciones. Modelos de mezcla (mixture models) para datos con exceso de ceros. Interpretación de los parámetros estimados en los GLM. Selección de modelos en los GLM. Análisis de residuos para la validación y evaluación de la calidad del ajuste de un GLM.

**4. Modelo Lineal y Generalizado Mixto (GLMM):** definición de efectos fijos y efectos aleatorios y su interpretación biológica. Uso de los GLMM para modelar restricciones de aleatorización de unidades muestrales y el diseño experimental. Principios fundamentales del diseño experimental. El ajuste de los GLMM por Máximo de Verosimilitud Restringido (REML). Test de efectos y selección de modelos en los GLMM. Uso del bootstrap paramétrico en la selección de modelos y la evaluación de la significancia estadística en los GLMM. Ejemplos de análisis con diseños experimentales en bloques aleatorizados, anidados, parcelas divididas (split-plot) y medidas repetidas. Interpretación de resultados (efectos fijos y aleatorios) y análisis de residuos de un GLMM para validar el ajuste de un modelo a los datos.

**5. Modelo Aditivo Generalizado (GAM):** los GAM como generalizaciones no paramétricas y no lineales de los GLM. Componentes de un GAM: función lineal, estructura de los errores, función de enlace y función de suavizado. Introducción a las funciones de suavizado comúnmente empleadas en Ecología. Problemas de sobre-ajuste (overfitting) de los GAM dependiendo de la función de suavizado. Utilidad y limitaciones de los GAM ilustrada con ejemplos. Validación y evaluación de la calidad del ajuste de un GAM. Introducción a las regresiones multinomiales y a los árboles de regresión (CART: Classification And Regression Trees) y su interpretación.

### **Evaluación:**

Los alumnos serán evaluados a través de un proyecto final de análisis de datos, sobre el cual entregarán un informe individual con la interpretación detallada de los resultados. Se aprobará el curso obteniendo al menos 65% de la nota máxima como exigido en el Postgrado de Ciencias Biológicas, Facultad de Ciencias, Universidad de la República. Este informe será enviado por e-mail dos semanas después de la fecha final del curso. Además de la nota del curso, los estudiantes recibirán por e-mail la corrección comentada de su informe final. Cada estudiante admitido en el curso se compromete formalmente a realizar la evaluación final del mismo.

### **Bibliografía Básica:**

Agresti A. 2015. Foundations of Linear and Generalized Linear Models. John Wiley, New York.

- Bolker B. 2008. *Ecological Models and Data in R*. Princeton University Press. Princeton.
- Chang W. 2012. *R Graphics Cookbook*. O'Reilly. New York.
- Crawley M. 2007. *The R Book*. John Wiley. London.
- Burnham K. & Anderson D. 2002. *Model selection and multimodel inference*. John Wiley, New York.
- Dunn P. & Smyth G. 2018. *Generalized Linear Models With Examples in R*. CRC Chapman & Hall. New York.
- Dobson A. & Barnett A. 2018. *An introduction to generalized linear models*. CRC Chapman & Hall. New York.
- Faraway J. 2014. *Extending the linear model with R*. Chapman & Hall. New York.
- Fox G. et al. 2015. *Ecological Statistics: Contemporary Theory and Application*. Oxford University Press. Oxford.
- McCarthy M. 2007. *Bayesian Methods for Ecology*. Oxford University Press. Oxford.
- Pinheiro M. & Bates D. 2004. *Mixed Effects Models in S and S+*. Springer-Verlag, New York.
- Wickman H. 2009. *ggplot2 Elegant Graphics for Data Analysis*. John Wiley, New York.
- Wood S. 2006. *Generalized Additive Models*. Chapman & Hall. New York.
- Zuur A. et al. 2008. *Mixed models and extensions in Ecology with R*. Springer-Verlag, New York.

### **Bibliografia complementaria (lista parcial):**

- Bolker B. et al 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* 24:127-135.
- Garamszegi L. et al 2009. Changing philosophies and tools for statistical inferences in behavioral ecology- *Behavioral Ecology* 17:1364-1375.
- Hoobs N. & Hilborn R. 2006. Alternatives to statistical hypothesis testing in ecology: A guide to self teaching. *Ecological Applications* 16: 5–19.
- Johnson P. et al 2014. Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution* 6: 133–142.
- Martin T. et al. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8: 1235–1246
- Nakagawa S. & Schielzeth H. 2012. A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 1:547-553.
- O'Hara R. 2009. How to make models add up — a primer on GLMMs. *Ann Zool Fenn.* 46:124-137.
- O'Hara R. & Lotze H. 2010. Do not log-transform count data. *Methods in Ecology and Evolution* 1: 118–122.
- Schielzeth H. 2010. Simple means to improve the interpretability of regression coefficients *Methods in Ecology & Evolution* 1: 103–113.
- Warton D & Hui F. 2011. The arcsine is asinine: the analysis of proportions in Ecology. *Ecology* 92: 3-10.