

# Propuesta de curso de posgrado

## Doctorado en Biología (CRUB)

1. Título del curso: Modelos de regresión en ecología y ciencias ambientales
2. Año académico: 2023
3. Fecha de dictado: 7 al 11 de agosto
4. Carrera a la que pertenece: Doctorado en Biología (CRUB)
5. Docentes: Dres. Pablo A. E. Alarcón y Luca Schenone
6. Carga horaria: 50 horas distribuidas en cinco días (lunes a viernes)

### 7. Fundamentación:

Una parte importante de la ecología se esfuerza por hallar “leyes” generales que permitan explicar y predecir el comportamiento de los sistemas naturales, y la construcción de modelos cuantitativos es parte esencial del proceso (Begon et al. 2006, Bolker 2008, Lehman et al. 2019). Este tipo de modelos permite a los ecólogos domar las ideas con el rigor de la matemática y reducir la compleja realidad ecológica a un conjunto de expresiones comparativamente más sencillas y manipulables. En este contexto, el desarrollo de habilidades que permitan construir y evaluar modelos cuantitativos debe ser parte fundamental en la formación profesional de ecólogos e investigadores de disciplinas afines.

El enfoque empírico de la ecología se basa en promover un constante diálogo entre modelos y datos (Ford 2000). Este enfoque resulta especialmente prometedor para el desarrollo de la ecología moderna por dos factores principales. Por una parte, el desarrollo tecnológico de las últimas décadas ha generado una explosión de datos sobre una amplia diversidad de patrones y procesos ambientales. Por otra parte, el desarrollo computacional ha hecho posible implementar una variedad de técnicas analíticas en forma rápida y comparativamente mucho más sencilla que hace algunas décadas (ej. Bolker 2008). Estos factores han propiciado la reciente invasión en la literatura ecológica de términos tales como “*Linear model (LM)*”, “*Generalized Linear Model (GLM)*” y “*Linear Mixed-effect Model (LMM y GLMM)*” (Gelman y Hill 2006, Matthiopoulos 2011, Kéry y Royle 2015). Estos términos se refieren a modelos de regresión que, en su conjunto, permiten representar una amplia variedad de procesos ecológicos, incorporando aspectos de la teoría ecológica, del diseño experimental y de las propiedades de los datos de manera mucho más ajustada que las técnicas de la estadística clásica. Por estas razones, la capacitación de los ecólogos en el manejo de este tipo de modelos resulta ahora tanto conveniente como necesaria; conveniente porque permite aproximar una variedad de preguntas en ecología de manera robusta y versátil, y necesaria porque constituye parte del lenguaje cotidiano de la ecología moderna.

La ecología empírica del siglo XX (1900's) se desarrolló especialmente sobre la base de la estadística clásica (frecuentista) aunque, en las últimas décadas, la perspectiva Bayesiana ha ganado terreno en los estudios de ecología (ej. Bolker 2008, Kéry y Royle 2015). Nuevamente, el progreso tecnológico ha sido el principal responsable de esta tendencia. Si bien los métodos Bayesianos fueron conceptualmente desarrollados en el siglo XVIII (1700's), su implementación práctica se demoró hasta los años 1950, luego de que se inventaran los algoritmos computacionales genéricamente conocidos como Cadenas de Markov Monte Carlo (MCMC) (Kruschke 2014). La concepción Bayesiana de la inferencia estadística se diferencia en aspectos fundamentales de la inferencia estadística frecuentista, situación que ha derivado en un largo debate filosófico. Sin embargo, existe un consenso más generalizado respecto de las bondades prácticas de los métodos Bayesianos, entre las que se cuentan la facilidad para implementar modelos complejos (ej. modelos jerárquicos) e interpretar los modelos estimados. Finalmente, los métodos Bayesianos componen un conjunto de herramientas sumamente convenientes para la enseñanza de la estadística. El usuario de estas técnicas debe especificar cada una de las partes componentes de un modelo, cultivando el hábito de tomar decisiones debidamente reflexionadas y sobre argumentos fundados.

Este curso capacitará a los alumnos en la construcción, la implementación y el diagnóstico de un conjunto de modelos de regresión haciendo uso de un marco de inferencia estadística Bayesiana. En las secciones que siguen se describen los contenidos temáticos, el objetivo y los aspectos metodológicos y organizacionales del curso.

#### 8. Contenido (programa analítico):

Unidad 1. Conceptos generales sobre la construcción de modelos en ciencia. El concepto de modelo. Tipos de modelos en ciencia: verbales vs. cuantitativos; determinísticos vs. estocásticos; fenomenológicos vs. mecanísticos; descriptivos vs. predictivos; generales vs. particulares; inductivos vs. deductivos. Escuelas filosóficas en las ciencias y su relación con la construcción de modelos.

Unidad 2. Conceptos generales sobre los modelos de regresión. Los modelos de regresión en el universo de modelos. Estructura general de los modelos de regresión: componente determinístico y componente estocástico. Funciones matemáticas y distribuciones de probabilidad. Tipos de preguntas que pueden ser exploradas a partir de los modelos de regresión.

Unidad 3. Estructura del Modelo Lineal General (LM). Función lineal de primer grado como componente determinístico. Distribución Normal como componente estocástico. Modelos de la media y de la varianza. Predictores continuos y categóricos. Modelos aditivos y con interacción entre variables. Supuestos de un LM. Diagnóstico de un LM. Ejemplos aplicados a la ecología.

Unidad 4. Estructura del Modelo Lineal Generalizado (GLM). Función lineal de primer grado como componente determinístico. Distribuciones binomial, Poisson y gamma como componentes estocásticos. Función de enlace. Predictores continuos y categóricos. Modelos aditivos y con interacción entre variables. Supuestos de un GLM. Diagnóstico de un GLM. Ejemplos aplicados a la ecología.

Unidad 5. Estructura del Modelo Lineal y Mixto (LMM). Estructuras de organización jerárquica en la naturaleza. Conceptos de “no pooling”, “complete pooling” y “partial pooling”. Modelos de intercepto y pendientes variables. Modelos para diferentes niveles de jerarquía. Supuestos de un LMM. Diagnóstico de un LMM. Ejemplos aplicados a la ecología.

Unidad 6. Estructura del Modelo Lineal Generalizado y Mixto (GLMM). Función lineal de primer grado como componente determinístico. Distribuciones binomial, Poisson y gamma como componentes estocásticos. Función de enlace. Supuestos de un LMM. Diagnóstico de un Modelo Lineal Generalizado y Mixto. Ejemplos aplicados a la ecología.

Unidad 7. Marco de inferencia clásica. Concepto de inferencia estadística. Probabilidad clásica (frecuentista). Métodos de estimación por Mínimo Cuadrados Ordinarios y Máxima Verosimilitud. Estimador puntual por Máxima Verosimilitud (MLE), intervalos de confianza, pruebas de hipótesis y valor p.

Unidad 8. Marco de inferencia Bayesiana. Probabilidad condicional y probabilidad subjetiva (Bayesiana). Teorema de Bayes. Distribución posterior, distribución previa y función de *likelihood*. Aproximación analítica y numérica de la distribución posterior de un modelo. Cadenas de Markov Monte Carlo (MCMC): algoritmo de Metropolis y Gibbs sampler. Intervalos de credibilidad. Posterior predictive check.

Unidad 9. Implementación de modelos regresión en R y JAGS. Vectores, matrices, data frames y listas como objetos para almacenar datos. Comandos para estimar modelos de regresión bajo un marco de inferencia clásica: *lm()*, *gls()*, *glm()*, *glmm()*. Comandos para vincular R y JAGS para estimar modelos de regresión bajo un marco de inferencia Bayesiana: *jags()*. Comandos para explorar datos y graficar los modelos estimados: *plot()*, *barplot()*, *boxplot()*, *curve()*.

## 9. Objetivo:

El objetivo del curso es asistir a los alumnos en el desarrollo de habilidades que les permitan construir, implementar y diagnosticar modelos estadísticos en estudios sobre la biodiversidad.

## 10. Actividad práctica

Desde el punto de vista organizacional, este curso contempla clases teóricas, prácticas y rondas de discusión. Todas las clases se realizarán de manera virtual a través de Zoom y empleando el usuario del Doctorado en Biología (CRUB). A continuación, se describen las principales características de cada una.

**Clases teóricas.** Las clases teóricas consistirán de presentaciones expositivas por parte del docente que introducirán y conectarán los distintos contenidos del programa. En concreto, estas clases tendrán como objetivo revisar los conceptos más importantes y necesarios para transitar las tres etapas fundamentales de la construcción de un modelo: 1) la construcción conceptual del modelo, 2) la implementación y la estimación del modelo en computadora y 3) el diagnóstico del modelo estimado. Para lograr este objetivo, las clases teóricas repasarán los conceptos fundamentales de álgebra y probabilidad que permitirán construir el andamiaje de los modelos de regresión. Seguidamente, estas clases revisarán conceptos relacionados con la inferencia estadística, y particularmente con la inferencia Bayesiana, y esbozarán la mecánica de los algoritmos utilizados para estimar un modelo a partir de datos. Finalmente, las clases teóricas harán hincapié en la importancia de diagnosticar los modelos estimados y expondrán la lógica de algunas estrategias utilizadas con este propósito. El curso contempla un total de diez clases teóricas durante los primeros cuatro días de curso. Cada clase constará de dos bloques de una hora aproximadamente separados por un receso de 30 minutos.

**Clases prácticas.** Las clases prácticas consistirán de ejercicios y situaciones-problema que los alumnos deberán resolver en forma individual y grupal. Estas clases tendrán como objetivo general que los alumnos se ejerciten en la construcción conceptual de los modelos de regresión e indaguen en los aspectos técnicos que les permitirán estimarlos y diagnosticarlos. En particular, los alumnos se ejercitarán en la construcción de modelos con aplicabilidad a sistemas ecológicos y en el manejo de los programas informáticos R y JAGS, ambos de acceso libre y gratuito. Las clases prácticas se intercalarán con las clases teóricas siempre que sea posible, de manera que los alumnos puedan prontamente aplicar los conceptos revisados para resolver situaciones concretas. El curso constará de un total de ocho clases prácticas de dos horas cada una durante los primeros cuatro días de curso.

Rondas de discusión. Las rondas de discusión serán charlas-debate sobre situaciones presentadas por los alumnos y tendrán como principal objetivo la construcción conceptual de los modelos de regresión. En grupos, y en exposiciones de 15 min, los alumnos deberán: 1) esbozar el contexto ecológico de una pregunta de interés, 2) identificar una hipótesis y algunas de sus predicciones, 3) proponer un diseño experimental u observacional para la recolección de datos y 4) diseñar un modelo de regresión para medir el grado de relación entre las variables implicadas. Cada presentación será seguida de un módulo de 30 minutos en el que los alumnos y el docente discutirán sobre las debilidades y/o las fortalezas del diseño propuesto, haciendo especial énfasis en la estructura del modelo. Las rondas de discusión tendrán lugar el quinto y último día de curso durante la mañana.

El curso implementará dos estrategias de aprendizaje principales. La primera consistirá en concebir la construcción de un modelo como un proceso que transita por tres etapas principales y sucesivas: 1) la construcción conceptual del modelo, 2) la implementación y la estimación del modelo en computadora y 3) el diagnóstico del modelo estimado. Esta estrategia debería permitir a los alumnos reconocer fácilmente el valor conceptual y práctico que los contenidos del programa tienen en el contexto de los modelos de regresión y, por lo tanto, facilitar su asimilación. La segunda estrategia consistirá en explorar técnicas progresivamente más sofisticadas usando continuamente a las técnicas más sencillas como punto de referencia y comparación. Para tal fin, el curso se iniciará con la “deconstrucción” del modelo de regresión lineal simple en sus partes componentes, ejercicio que facilitará modificar y/o extender su estructura hacia versiones más sofisticadas. En forma similar, las técnicas Bayesianas se estudiarán mediante el análisis comparativo con las técnicas de la estadística clásica (frecuentista) con la que los alumnos estarán probablemente mejor familiarizados. Es importante destacar que la inferencia estadística clásica se estudiará con el único propósito de establecer un marco de referencia conceptual; sin embargo, el curso no buscará necesariamente que los alumnos aprendan a aplicar estas técnicas.

#### 11. Evaluación total del curso

El curso incluirá cuatro instancias de evaluación parcial y una instancia de evaluación final. Las evaluaciones parciales (“parcialitos”) consistirán de un examen escrito que explorará mayormente aspectos conceptuales y se completarán durante los primeros 10 minutos de cada clase. El principal propósito de estas evaluaciones es promover en los alumnos el hábito de estudiar y revisar los conceptos trabajados, a fin de que ellos mismos puedan reconocer las temáticas en las que deben volcar mayor esfuerzo de estudio. La instancia de evaluación final consistirá de un trabajo práctico en la que los alumnos deberán implementar las técnicas estudiadas durante el curso sobre una base de datos reales. Este trabajo podrá entregarse hasta

15 días posteriores a la finalización del curso. La aprobación del curso se logrará mediante la aprobación del 75% de los exámenes parciales y un 60 % de la evaluación final.

## 12. Bibliografía

Para desarrollar conceptos generales de inferencia científica, ecología y modelado cuantitativo se utilizarán los libros de Taper y Lele (2004), Lheman y colaboradores (Lehman et al. 2019) y McElreath (McElreath 2020). Los contenidos temáticos y las estrategias de estudio especialmente combinarán aquellos desarrollados en los libros de Kruschke (Kruschke 2014), Bolker (Bolker 2008) y Gelman (Gelman y Hill 2006). Además, se utilizarán para consulta los libros de Matthiopoulos (Matthiopoulos 2011) y Kéry (Kéry y Royle 2015). Las referencias completas de estos libros se enlistan debajo.

- Bolker, B. M. *Ecological models and data in R*. Princeton University Press, 2008.
- Gelman, A., y J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
- Kéry, M., y J. A. Royle. *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. Academic Press, 2015.
- Kruschke, J. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.
- Lehman, C., S. Loberg, y A. Clark. *Quantitative Ecology: A New Unified Approach*. University of Minnesota Libraries Publishing, 2019.
- Matthiopoulos, J. *How to be a quantitative ecologist: the 'A to R' of green mathematics and statistics*. John Wiley & Sons, 2011.
- McElreath, R. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press, 2020.
- Taper y Lele. *The nature of scientific evidence: statistical, empirical and philosophical considerations*. The University of Chicago Press, 2004.

## 13. Referencias citadas en la propuesta

- Currie, D. J. 2019. Where Newton might have taken ecology. *Global Ecology and Biogeography*, 28(1), 18–27.
- Craig, M. P. A. *Ecological Political Economy and the Socio-Ecological Crisis*. Springer 2017.

- Ellison, A. M., y Dennis, B. 2010. Paths to statistical fluency for ecologists. *Frontiers in Ecology and the Environment*, 8(7), 362–370.
- Green, J. L., Hastings, A., Arzberger, P., Ayala, F. J., Cottingham, K. L., Cuddington, K., Davis, F., Dunne, J. A., Fortin, M.-J., y Gerber, L. (2005). Complexity in ecology and conservation: Mathematical, statistical, and computational challenges. *BioScience*, 55(6), 501–510.
- Maurer, B. A. *Untangling ecological complexity: The macroscopic perspective*. University of Chicago Press, 1999.
- Seidl, R. 2017. To model or not to model, that is no longer the question for ecologists. *Ecosystems*, 20(2), 222–228.
- Taper y Lele. *The nature of scientific evidence: statistical, empirical and philosophical considerations*. The University of Chicago Press, 2004.