
Curso teórico–práctico

Flexibilidad y eficiencia en análisis de datos en ecología con énfasis en ornitología: una introducción a los modelos (mixtos) multinivel en **R**

Carlos Esteban Lara [†]

22 de agosto de 2019

1 Justificación

Entender, manejar y analizar datos en ecología y evolución es una de las mayores fortalezas que un investigador puede lograr. Este logro requiere un proceso de mucha dedicación. En nuestro medio (*e.g.*, muchas universidades a nivel de pregrado e incluso posgrado) nos indican el camino a seguir por medio de cursos básicos de estadística que, comúnmente, limitan sus alcances al ‘mundo normal’, es decir, datos de distribución gaussiana. Con elementos sobre esta familia de distribución se producen actualmente gran parte de los trabajos de grado en pregrados de ecología y biología. Sin embargo, el camino hacia un manejo y análisis de datos biológicos más apropiado está lejos de allí y las pautas que nos provee nuestro medio para encontrar ese camino son muy limitadas. Consecuentemente, jóvenes investigadores pierden la ruta cuando enfrentan datos no normales, y que por cierto, son más comunes en la naturaleza (*e.g.*, distribuciones de tipo binomial, poisson y gamma). Esta pérdida de ruta, no sólo genera stress en los estudiantes, sino que también los lleva a solucionar sus limitaciones bio-estadísticas por medio de análisis inapropiados. Por ejemplo, es común observar que la principal solución frente a datos que no siguen una distribución normal, es el uso de tests no paramétricos, que en casi todos los casos, presentan serias limitaciones.

El camino más apropiado a seguir, y que se ha posicionado poderosamente desde hace algunas décadas, es el uso de los modelos multinivel (también denominados modelos mixtos o GLMMs). Su fortaleza radica en la flexibilidad para analizar datos de distribución no normal y datos que no son independientes entre si (un problema casi inherente en todos los datos obtenidos en estudios de campo). Dentro de la familia GLMMs es fácil realizar tradicionales *t*-tests, anovas, anocovas, regresiones múltiples, y test no paramétricos (*e.g.*, Kruskal–Wallis, Mann–Whitney). Específicamente, para aquellos que estudian ecología del comportamiento,

[†] Ingeniero Forestal de la Universidad Nacional de Colombia. Magister en Bosques y Conservación Ambiental de la Universidad Nacional de Colombia. Doctor en Zoología del Department of Zoology, University of Otago, New Zealand. Grupo de Investigación en Ecosistemas Tropicales, Facultad de Ciencias Exactas y Naturales, Universidad de Caldas, Manizales, Colombia. E-mail: celarav@gmail.com

los modelos mixtos permiten calcular fácilmente repetibilidad (*e.g.*, variabilidad dentro de grupos y entre grupos), la cual es clave en esta área de la ecología.

Por otra parte, aunque existen muchos programas estadísticos, R (ver: r-project.org), es quizás el software más flexible, es gratuito, y aunque no parezca, es una plataforma muy amigable. Infortunadamente, el uso de R ha sido, en cierta medida, mitificado y excluido en muchos contextos (*e.g.*, muchas universidades colombianas no fomentan su uso). Un usuario novato frente a R, generalmente se desmotiva al no poder ingresar apropiadamente sus datos en la plataforma. Sin embargo, una vez los datos se presentan adecuadamente en R, es fácil usar y entender, no sólo los paquetes para ajustar GLMMs (*e.g.*, *lme4*, *nlme*, *MCMCglmm*), sino también aprovechar la versatilidad de R en otros contextos. La capacidad de R es virtualmente ilimitada. Por ejemplo, para manejar bases de datos de gran tamaño, para producir gráficos de alta calidad, para poder replicar los análisis tantas veces como sea necesario, entre otras.

Es importante enfatizar que en Colombia ha existido poca oferta sobre este tipo de cursos de análisis e interpretación de datos biológicos vía modelos multinivel, y los pocos cursos que existen no usan exclusivamente R. Como se mencionó, la combinación de ambos (modelos multinivel y R) se ha convertido en el ‘gold standard’ en análisis de datos en ecología y evolución. Consecuentemente, un curso de esta naturaleza es novedoso y muy apropiado dentro del marco del VI Congreso de Ornitología. Ya que en términos generales, este curso invitaría a todos los estudiantes de pregrado, posgrado e incluso docentes que estén interesados en familiarizarse con los modelos multinivel y su ajuste por técnicas tradicionales y bayesianas.

Por último, el curso es de carácter teórico–práctico, ya que se programarán actividades donde los estudiantes puedan aplicar, a su propio set de datos o a un set de datos asociado con el curso, ejercicios que permitan reforzar los conceptos y técnicas estudiadas durante el componente teórico del curso.

2 Objetivo general

Con el fin de fomentar un apropiado manejo y análisis de datos biológicos con un énfasis en modelos multinivel y además para fomentar el uso de R, se propone ofrecer un curso intensivo teórico–práctico. Este curso tendrá una duración de dos (2) días consecutivos. Idealmente, el grupo debería estar limitado a máximo 15 participantes. Preferiblemente estudiantes que estén cursando últimos semestres de pregrado o que estén iniciando estudios de maestría o doctorado. Cada participante debe proporcionar su propio computador portátil. Yo proporcionaré bases de datos de acceso libre (*e.g.*, Dryad o Open Science Framework), aunque es ideal si los estudiantes usan sus propios datos.

3 Programa no detallado

El programa del curso incluye una introducción básica al software R y luego a los GLMMs empezado por modelos lineales y terminando con el ajuste de modelos mixtos por técnicas

bayesianas. En las jornadas prácticas se trabajará en como ingresar datos en R, entender su estructura y el lenguaje básico de R. En otra jornada, teórica, se introducirán los tipos de algoritmos que R usa para ajustar modelos multinivel y se discutirán sus bondades y limitaciones. Se hará énfasis en el uso del paquete *MCMCglmm*. Por último se usarán algunas herramientas de R para generar gráficos apropiados para cada tipo de modelación.

Es importante resaltar que los estudiantes pueden o no tener conocimiento previo en R, de hecho, la idea del curso es eliminar las barreras hacia el uso de R y rápidamente escalar hacia el uso de GLMMs. Es indispensable que cada participante lleve su propio computador portátil.

4 Programa detallado

El curso se realizará el 20 y 21 de octubre, previo al inicio del VI Congreso de Ornitología.

Día No. 1. 20 de octubre.

Jornada No. 1. Teoría. 8:00am–10:00am. 2 horas.

- Aspectos básicos en R.
- Operaciones básicas en R.
- Funciones útiles en R.
- Vectores, dataframes y matrices en R.
- Estadísticos básicos en R.
- Manejo de datos (paquete *ddplyr*).

Receso (refrigerio) 30 minutos.

Jornada No. 2. Práctica. 10:30am–12:30pm. 2 horas.

- Ejercicios prácticos en R.
- Manipulación de datos en R.
- Graficos básicos en R.

Receso (almuerzo) 1 hora.

Jornada No. 3. Teoría. 1:30pm–3:30pm. 2 horas.

- Introducción a los modelos lineales.

- Interpretación del intercepto, coeficientes de regresión y ajuste de los modelos lineales.
- ANOVA & ANCOVA & regresión múltiple. Resolver el conundrum!

Receso (refrigerio) 30 minutos.

Jornada No. 4. Práctica. 4:00pm–6:00pm. 2 horas.

- Ajuste e interpretación de modelos lineales.
- Ajuste y comparación entre ANOVAS, ANCOVAS y regresión múltiple.
- Graficos por medio del paquete *ggplot2*.

Día No. 2. 21 de octubre

Jornada No. 5. Teoría. 8:00am–10:00am. 2 horas.

- Datos binomiales, chi-square y regresión logística.
- Conteos y dispersión.
- Proporciones.
- Construcción y selección de modelos.
- Modelos mixtos lineales.

Receso (refrigerio) 30 minutos.

Jornada No. 6. Práctica. 10:30am–12:30pm. 2 horas.

- Ajuste de modelos mixtos, selección e interpretación.
- Uso de diferentes paquetes: *nmle*, *lme4*.

Receso (almuerzo) 1 hora.

Jornada No. 7. Teoría. 1:30pm–3:30pm. 2 horas.

- Breve introducción a la estadística bayesiana.
- Inferencia estadística por métodos bayesianos.
- Introducción al ajuste de modelos mixtos por técnicas bayesianas.

Receso (refrigerio) 30 minutos.

Jornada No. 8. Práctica. 4:00pm–6:00pm. 2 horas.

- Ajuste e interpretación de modelos mixtos con el paquete *MCMCglmm*.
- Uso y selección de priors.
- Graficos y presentación de los outputs de los modelos.

5 Costo

El curso tiene un cupo máximo de 15 participantes y el costo por participante es de \$100.000 COP.

Para reservar su cupo debe contactar directamente la Asociación Colombiana de Ornitología ACO: al correo coordinacion@ornitologiacolombiana.com