



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

**CICLO II- 2024**

**SIGLAS B-606 / SP-0924**

**CREDITOS: 4**

**ESTUDIANTES:** pregrado y posgrado

**HORARIO:** 4 horas por semana (lunes 9-10:50; jueves 9-10:50)

**MODALIDAD:** 100% presencial

**REQUISITOS:** B-0304 (Ecología General) y B-0305 (Laboratorio de Ecología General)

**PROFESORES:**

- Mario Espinoza Mendiola; correo: [marioespinozamen@gmail.com](mailto:marioespinozamen@gmail.com);  
[mario.espinoza\\_m@ucr.ac.cr](mailto:mario.espinoza_m@ucr.ac.cr);  
Oficina: Biología 3B; CIMAR 130  
Tel: 2511-2208 / 2511-8683
- Marcelo Araya Salas; correo: [marcelo.araya@ucr.ac.cr](mailto:marcelo.araya@ucr.ac.cr)  
Página web: <https://marce10.github.io>  
Oficina: 2, Centro de Investigación en Neurociencias

**MEDIACION VIRTUAL:**

- Se utilizará la plataforma Mediación Virtual para que los estudiantes tengan acceso a todos los elementos del curso (clases, material complementario, tareas), así como para que puedan entregar todas las evaluaciones correspondientes.

**HORAS DE ATENCIÓN:** L: 11-12:00 / J: 11-12:00 (o a convenir con el profesor)

**DESCRIPCIÓN DEL CURSO:**

Las ciencias biológicas involucran la colección sistemática, organización, análisis y presentación de la información. En biología, la mayoría de las investigaciones generan una gran cantidad de datos cuantitativos que deben ser analizados de forma apropiada para obtener resultados confiables (Zar 2010, Zuur et al. 2010). En este sentido, la estadística es una herramienta que permite ordenar, analizar e interpretar datos biológicos. Para dicho fin, se han desarrollado una gran variedad de programas estadísticos computacionales que permiten procesar extensas bases de datos de distintos tipos a gran velocidad. Sin embargo, muchas veces estos programas o paquetes estadísticos tienen la limitante de que, son difíciles de adquirir por su alto costo, deben ser actualizados periódicamente, no proveen herramientas para la manipulación de datos, o están restringidos por un número definido de pruebas estadísticas o funciones matemáticas.

Además, debido a que existe un gran número de paquetes estadísticos, a veces es necesario aprender a manipular diferentes programas computacionales (i.e., softwares) para realizar análisis específicos. Afortunadamente, en la última década surgió el sistema computacional R,



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

el cual ofrece una plataforma de acceso libre que le permite al usuario programar sus propias funciones, realizar pruebas estadísticas, graficar, manipular bases de datos extensas y de naturaleza distinta, así como compartir nuevas herramientas desarrolladas independientemente por los usuarios (Crawley 2007, Fuchs & Barrantes 2015).

El desarrollo de la plataforma R para estadística computacional esta grandemente influenciado por la idea de “fuentes de libre acceso”: La distribución base de R y un gran número de contribuciones (extensiones) están disponibles bajo los términos de Fundación de Software Libre (R Development Core Team 2014). Esta licencia tiene dos implicaciones importantes para analistas de datos que trabajan con R. Primero, todos los códigos completos están disponibles y son de acceso libre para todos, por lo que cualquier persona puede examinar, editar y manipular el código de acuerdo a sus necesidades. Segundo, esta plataforma está constantemente creciendo, renovándose y actualizándose de acuerdo a las necesidades específicas de los usuarios. Esto ha hecho que R se haya convertido en una de las plataformas más populares y utilizadas en una amplia variedad de disciplinas (Crawley 2007, R Development Core Team 2014).

Por esta razón, el manejo del ambiente de programación R resulta una herramienta fundamental para el desarrollo profesional de los estudiantes de ciencias biológicas, independientemente de su área de estudio. La carrera de biología ofrece cursos introductorios de estadística donde los estudiantes utilizan R para realizar diversas pruebas estadísticas. Además, algunos cursos introducen a sus estudiantes en aplicaciones específicas de R en sus disciplinas (ej. paquetes de análisis comparativos en el curso Análisis Comparativo Filogenético, paquetes de análisis espaciales y multivariados en el curso de ecología y conservación de tiburones y rayas). Sin embargo, muchos estudiantes de la carrera de Biología: (1) siguen teniendo una dependencia muy grande hacia programas básicos como Excel para manipular bases de datos; (2) continúan usando softwares estadísticos comerciales de interfaz gráfica de usuario (no programables) para sus análisis; (3) o circunscriben sus análisis estadísticos a las herramientas disponibles en softwares comerciales, limitando de manera importante las opciones de análisis.

Este curso pretende profundizar en los elementos de programación computacional, manipulación de bases de datos, diseño experimental, graficación personalizada, y el uso de técnicas y modelaje estadístico avanzadas utilizando R como plataforma (Touchon & McCoy 2016, Warton et al. 2016). El curso está dirigido a estudiantes avanzados de carrera, de licenciatura o postgrado. En el curso se pretende cubrir las bases de los principales análisis y técnicas de modelaje estadístico, así como análisis emergentes. Todas las semanas se realizarán prácticas de laboratorio en donde se aplicarán los conceptos desarrollados durante las clases de teoría. Además, durante el curso cada estudiante presentará un paquete o extensión (conjunto de herramientas aplicables a análisis específicos) de R, en donde profundizará sobre sus



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

aplicaciones en el campo de la biología demostrando en clase en que consiste el análisis.

**OBJETIVO GENERAL:**

- Enseñar a estudiantes elementos básicos y avanzados de programación computacional que les permitan manipular, analizar, graficar e interpretar información biológica utilizando R como plataforma de trabajo.

**OBJETIVOS ESPECIFICOS:**

- Familiarizar al estudiante con la programación en R.
- Brindar herramientas para la manipulación de bases de datos usando la plataforma R.
- Emplear métodos de visualización de datos usando la plataforma R.
- Cubrir algunos de los análisis estadísticos tradicionales usando la plataforma R.
- Realizar análisis de modelaje usando la plataforma R.
- Proveer a los estudiantes con experiencia en la aplicación de las herramientas brindadas por medio de tareas y de un proyecto de investigación.

**CONTENIDO Y CRONOGRAMA:**

- Periodo del 12 de agosto al 20 de Septiembre, 2024
  - ✓ Introducción al curso: Discusión del programa y descripción de la evaluación del curso.
  - ✓ Introducción a la programación en R: Familiarizar al estudiante con la programación en R. En este tema se abordarán las bases de programación en R:
    - ✓ Sesión de repaso
      - Objetos, vectores, operaciones con vectores y factores
      - Manipulación de datos I (matrices, data.frames y otros operadores)
      - Manipulación de datos II (funciones aggregate, tapply y ddply)
      - Importar y exportar datos
    - ✓ Métodos de visualización: graficación base
      - Histogramas, boxplots/density plots, gráficos de puntos, gráficos circulares, gráficos de barras
  - Combinar bases de datos (cambiar formatos de bases de datos: largo vs. ancho)
  - ¿Cómo trabajar con fechas y tiempo?
  - Métodos de visualización: graficación con GGLOT2
  - Reportes dinámicos



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

- Periodo del 23 de setiembre al 18 de octubre, 2024
  - ✓ Programación avanzada en R
    - Uso de operadores y funciones lógicas (“if”, “ifelse”, “for”, “while”, “Xapply” y otras)
    - Métodos para mejorar el desempeño de rutinas computacionalmente intensivas
    - Funciones personalizadas
  - ✓ Análisis estadísticos básicos en R:
    - Análisis de Frecuencias (Chi-cuadrado, Pruebas de Bondad de Ajuste, Tablas de Contingencia, “odds ratio”, etc.)
    - Pruebas t-student
    - Análisis de Varianza (“ANOVA”)
    - Análisis de Regresión Lineal y Correlación
    - Análisis de Regresión Múltiple
  - ✓ ¿Cómo crear un código reproducible?
- Periodo del 21 de octubre al 8 de noviembre, 2024
  - ✓ Modelos Lineales Generalizados (GLMs)
  - ✓ Modelos Lineales Generalizados Mixtos (GLMMs)
  - ✓ Modelos Aditivos Generalizados (“GAMs”, por sus siglas en inglés)
- Periodo del 11 de noviembre al 29 de noviembre, 2024
  - ✓ Presentaciones y trabajos finales de estudiantes.

**METODOLOGÍA Y ACTIVIDADES PARA CUMPLIR CON LOS OBJETIVOS:**

- Clases prácticas guiadas con computadora y participación activa de estudiantes
- Tareas
- Presentación y discusión de librerías en R (solo estudiantes del SEP)
- Proyecto final para el curso

**EVALUACIÓN DEL CURSO ESTUDIANTES DE PREGRADO:**

- 1- Tareas (**70%**): Durante el semestre se dejarán 7 tareas cortas (**10% cada tarea**), que le permitirá al estudiante poner en práctica lo aprendido en clase. Las tareas consistirán en



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

ejercicios de: (i) manipulación de bases de datos, (ii) graficación, (iii) uso de funciones y operadores lógicos, y (iv) análisis estadísticos aplicados, en los cuales el estudiante deberá mostrar sus habilidades para resolver una situación particular. Las tareas se asignarán con una semana de antelación, y deberán ser entregadas en las siguientes fechas:

- a. Tarea 1 – viernes 30 de agosto (5:00 pm)
- b. Tarea 2 – viernes 14 de septiembre (5:00 pm)
- c. Tarea 3 – viernes 27 de septiembre (5:00 pm)
- d. Tarea 4 – viernes 12 de octubre (5:00 pm)
- e. Tarea 5 – viernes 26 de octubre (5:00 pm)
- f. Tarea 6 – viernes 8 de noviembre (5:00 pm)
- g. Tarea 7 – viernes 29 de noviembre (5:00 pm)

Cómo parte de la evaluación, cada estudiante deberá entregar el código en R que demuestre todos los pasos que realizó para resolver la tarea. Una vez que se tenga un mayor conocimiento de cómo se elaboran reportes dinámicos (a partir del 29 de septiembre) desde la plataforma de R-Studio, los estudiantes deberán utilizar los paquetes “knitr” y “Rmarkdown” para asegurar la reproducibilidad de las tareas.

- 2- Proyecto Final (**30%**): Al final del semestre, cada estudiante deberá elaborar un proyecto final en el que utilizará un set de datos real (puede ser de un curso, proyecto, artículo científico, etc.) para realizar un análisis detallado que responda a una o varias preguntas biológicas. El proyecto debe ser entregado por escrito utilizando el formato de reporte dinámico “Rmarkdown”, y además, se realizará una presentación oral. Será la responsabilidad asegurarse de que todos los datos, paquetes complementarios y demás información estén disponibles para el resto de sus compañeros al menos unos días antes de la presentación del proyecto final. El proyecto final contará con las siguientes secciones:

- a. Breve introducción al tema (justificación del problema y pregunta de investigación que intenta responder con el set de datos escogido) – **2.5%**
- b. Objetivos – **2.5%**
- c. Diseño experimental / metodología (descripción detallada de cómo fueron recolectados los datos y el diseño experimental empleado) – **5%**
- d. Resultados (todos los análisis descriptivos y estadísticos empleados para responder a la pregunta de investigación, incluyendo gráficos y cuadros que obtuvo mediante programación) – **10%**
- e. Interpretación de resultados / conclusiones – **5%**
- f. Presentación oral de proyecto final – **5%**



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

**EVALUACIÓN DEL CURSO ESTUDIANTES DE POSGRADO:**

- 1- Tareas **(60%)**: Durante el semestre se dejarán 7 tareas cortas **(8.5% cada tarea)**, que le permitirá al estudiante poner en práctica lo aprendido en clase. Las tareas consistirán en ejercicios de: (i) manipulación de bases de datos, (ii) graficación, (iii) uso de funciones y operadores lógicos, y (iv) análisis estadísticos aplicados, en los cuales el estudiante deberá mostrar sus habilidades para resolver una situación particular. Las tareas se asignarán con una semana de antelación, y deberán ser entregadas en las siguientes fechas:
  - a. Tarea 1 – viernes 30 de agosto (5:00 pm)
  - b. Tarea 2 – viernes 14 de septiembre (5:00 pm)
  - c. Tarea 3 – viernes 27 de septiembre (5:00 pm)
  - d. Tarea 4 – viernes 12 de octubre (5:00 pm)
  - e. Tarea 5 – viernes 26 de octubre (5:00 pm)
  - f. Tarea 6 – viernes 8 de noviembre (5:00 pm)
  - g. Tarea 7 – viernes 29 de noviembre (5:00 pm)

Cómo parte de la evaluación, cada estudiante deberá entregar el código en R que demuestre todos los pasos que realizó para resolver la tarea. Una vez que se tenga un mayor conocimiento de cómo se elaboran reportes dinámicos (a partir del 29 de septiembre) desde la plataforma de R-Studio, los estudiantes deberán utilizar los paquetes “knitr” y “Rmarkdown” para asegurar la reproducibilidad de las tareas.

- 2- Presentación de Paquete de R **(15%)**: Los estudiantes de posgrado elegirán un paquete en R, el cual deberán explicar al resto de sus compañeros al final del curso. Para esto, los estudiantes trabajarán en parejas y cada pareja entregará un reporte dinámico “Rmarkdown” y realizará una presentación oral del paquete seleccionado, ilustrando mediante ejemplos, como funcionan las diferentes funciones del paquete. Será responsabilidad de cada pareja asegurarse de que todos los datos, paquetes complementarios, y demás información estén disponibles para el resto de sus compañeros al menos unos días antes de la presentación del paquete. La evaluación consiste en:
  - a. Entrega de reporte dinámico – **10%**
    - i. Introducción a las principales funciones del paquete **(2.5%)**
    - ii. Aplicaciones del paquete a diferentes áreas de la biología **(2.5%)**
    - iii. Ejemplos guiados de cómo funciona el paquete **(5%)**
  - b. Presentación oral del paquete – **5%**



**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

- 3- Proyecto Final (**25%**): Al final del semestre, cada estudiante deberá elaborar un proyecto final en el que utilizará un set de datos real (puede ser de un curso, proyecto, artículo científico, etc.) para realizar un análisis detallado que responda a una o varias preguntas biológicas. El proyecto debe ser entregado por escrito utilizando el formato de reporte dinámico “Rmarkdown”, y además, se realizará una presentación oral. Será la responsabilidad de cada estudiante asegurarse de que todos los datos, paquetes complementarios y demás información estén disponibles para el resto de sus compañeros al menos unos días antes de la presentación del proyecto final. El proyecto final contará con las siguientes secciones:
- a. Breve introducción al tema (justificación del problema y pregunta de investigación que intenta responder con el set de datos escogido) – **2.5%**
  - b. Objetivos – **2.5%**
  - c. Diseño experimental / metodología (descripción detallada de cómo fueron recolectados los datos y el diseño experimental empleado) – **2.5%**
  - d. Resultados (todos los análisis descriptivos y estadísticos empleados para responder a la pregunta de investigación, incluyendo gráficos y cuadros que obtuvo mediante programación) – **10%**
  - e. Interpretación de resultados / conclusiones – **5%**
  - f. Presentación oral de proyecto final – **2.5%**

## **BIBLIOGRAFÍA**

- Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412
- Bates D, Maechler M, Bolker BM, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. <http://cran.r-project.org/package=lme4>.
- Bolker BM (2008) *Ecological Models and Data in R*. Princeton University Press, London
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24:127–35
- Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*. Springer, London
- Braak CJF ter, Verdonschot PFM (1995) Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat Sci* 57:255–289
- Crawley MJ (2007) *The R Book*, 2nd Editio. Wiley, Southern Gate
- Fuchs EJ, Barrantes G (2015) *El lenguaje estadístico R aplicado a las ciencias biológicas*. Editorial de la Universidad de Costa Rica, San José, Costa Rica
- James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning with Applications in R*. Springer





**PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924**  
**Programación y métodos estadísticos avanzados en R**

- Legendre P, Oksanen J, Braak CJF ter (2011) Testing the significance of canonical axes in redundancy analysis. *Methods Ecol Evol* 2:269–277
- Maindonald J, Braun WJ (2010) *Data Analysis and Graphics Using R - an Example-Based Approach* (J Maindonald and W. Braun, Eds.), Third edit. Cambridge University Press, London
- R Development Core Team (2014) *R: a language and environment for statistical computing*.
- Reimann C, Filzmoser P, Garrett RG (2008) *Statistical Data Analysis Explained*.
- Touchon JC, McCoy MW (2016) The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* 7:e01394
- Warton DI, Lyonsy M, Stoklosa J, Ivesz AR (2016) Three points to consider when choosing a LM or GLM test for count data. *Methods Ecol Evol*:n/a-n/a
- Zar JH (2010) *Biostatistical Analysis*. Prentice Hall, New Jersey
- Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14
- Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, NY
- Zuur AF, Saveliev AA, Ieno EN (2014) *A beginner's Guide to Generalised Additive Mixed Models with R*. Highland Statistics Ltd., Newburgh, United Kingdom

**Recursos adicionales en internet**

- <https://cran.r-project.org>
- <https://cran.r-project.org/manuals.html>
- <https://cran.r-project.org/other-docs.html>
- <https://www.r-project.org/doc/bib/R-books.html>
- <http://research.stowers-institute.org/efg/R/Color/Chart>
- <http://research.stowers-institute.org/efg/R/Color/Chart/ColorChart.pdf>
- [http://www.stat.ubc.ca/~jenny/STAT545A/block14\\_colors.html](http://www.stat.ubc.ca/~jenny/STAT545A/block14_colors.html)
- <http://research.stowers-institute.org/efg/R/Graphics/Basics/mar-oma/index.html>
- <http://www.r-bloggers.com/setting-graph-margins-in-r-using-the-par-function-and-lots-of-cow-milk>
- <https://www.stat.auckland.ac.nz/~paul/Talks/Rgraphics.pdf>
- <http://www.statmethods.net/advgraphs/parameters.html>
- <https://www.r-bloggers.com/a-fast-intro-to-plyr-for-r/>
- <https://www.r-statistics.com/tag/visualization/>
- <http://blog.revolutionanalytics.com/2014/04/some-r-resources-for-glms.html>
- <https://www.datacamp.com/community/tutorials/r-tutorial-apply-family>





PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924  
Programación y métodos estadísticos avanzados en R



Es un acto u omisión que afecta las oportunidades de una persona o sus derechos humanos.

SON MANIFESTACIONES DE  
DISCRIMINACIÓN:

- Ataques físicos
- Burlas, bromas ofensivas
- Uso de vocabulario discriminatorio
- Trato diferencial o despectivo
- Exclusión o segregación
- Desinterés o maltrato
- Negación a brindar servicios

DENUNCIA

La denuncia puede presentarse personalmente o mediante correo electrónico ante la Comisión Institucional Contra la Discriminación (CICDI).

Ninguna de las personas involucradas en el proceso podrán sufrir prejuicios.

Si usted ha vivido una situación de discriminación puede acercarse a la Facultad de Ciencias para buscar apoyo.



2511-6345



facultad.ciencias@ucr.ac.cr





PROBLEMAS ESPECIALES EN ECOLOGÍA – B606 / SP0924  
Programación y métodos estadísticos avanzados en R



Toda conducta de naturaleza sexual indeseada por quien la recibe, que provoque efectos perjudiciales en el estado general o bienestar personal.

SON MANIFESTACIONES DE  
HOSTIGAMIENTO SEXUAL:

- Promesa o amenaza, implícita o expresa, relacionada con favores sexuales
- Propuestas o conductas de naturaleza sexual
- Humillaciones u ofensas con palabras, gestos o imágenes
- Acercamientos o formas de contacto físico no deseados
- Intentos de comunicación ajenos a la relación profesional o académica

DENUNCIA

Las denuncias se realizan en forma verbal o escrita, ante la Comisión Institucional Contra el Hostigamiento Sexual (CICHS).

CONTACTOS

Comisión Institucional contra el Hostigamiento Sexual: 2511-4898  
[comision.contrahostigamiento@ucr.ac.cr](mailto:comision.contrahostigamiento@ucr.ac.cr)  
Defensoría contra el Hostigamiento Sexual: 2511-1909  
[defensoriahs@ucr.ac.cr](mailto:defensoriahs@ucr.ac.cr)

