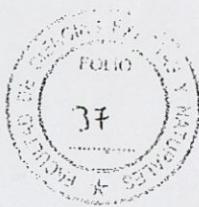


CD-0803-17



Curso de postgrado

FUNDAMENTOS DE R EN BIOLOGÍA

Dept. de Ecología, Genética y Evolución

Descripción general

Dirigido a: licenciados en Ciencias Biológicas y carreras afines

Profesores a cargo: Juan Hurtado y Gerardo de la Vega

Coordinadora: Adriana Pérez

Duración: 42 horas

Puntaje para el doctorado: 1,5 puntos

Fechas de dictado: Lunes 15 de Mayo de 2017 al sábado 20 de Mayo de 2017 en el horario de 0900 a 1600h.

Modalidad: Curso teórico/práctico a desarrollarse en laboratorio de computación. Con evaluación final.

Arancel: \$500

Exentos: alumnos inscriptos en doctorado de la UBA, docentes de la FCEyN

Cupo: 30 alumnos

Fundamentación

El paquete R es un programa que proporciona un entorno útil para análisis y gráficos estadísticos a partir de un sinfín de tipos de datos. Además de ser un programa es un lenguaje de programación que lo hace mucho más flexible y versátil que otros programas de análisis estadísticos tales como Infostat, SAS, SPSS o Statistica. Esta mayor versatilidad permite mejorar el ajuste de los análisis a intereses particulares dando más cobertura que cualquier otro programa de uso corriente. Además, R es gratuito y forma parte de un proyecto colaborativo y abierto que permite que sus usuarios extiendan su configuración continuamente. Por ello, R suele disponer de aplicaciones nuevas que no pueden encontrarse en otros programas.

Por estos motivos R es uno de los paquetes estadísticos de referencia y el programa elegido por una proporción cada vez mayor de analistas y científicos de múltiples disciplinas, en particular de ecología, genética y evolución.

No obstante las bondades de este paquete, para los egresados de esta casa de estudios que recién se inician en su uso, puede resultar poco amigable y complicado. Frente a estas dificultades, muchos optan por utilizar otros programas de análisis estadísticos aunque sean subóptimos. Para superar estas dificultades y poder aprovecharlo es necesario primero aprender algunas nociones básicas de su lenguaje y comprender cómo funciona. En este sentido, para los licenciados en Ciencias Biológicas, contar con un curso introductorio especialmente orientado a las ciencias biológicas y que no requiera conocimientos previos en cálculo o informática es condicionante para iniciarse como usuario.

Objetivo

El objetivo general es brindar las herramientas necesarias para que los licenciados en Ciencias Biológicas y carreras afines "pierdan el miedo a R": se introduzcan en su lenguaje, se inicien como usuarios y se desenvuelvan de manera autónoma. Ofrecemos para ello una guía útil para superar las dificultades iniciales más importantes del lenguaje R y para ilustrar la amplia cobertura y versatilidad del programa en ciencias naturales. El curso está dirigido a estudiantes de postgrado o investigadores de ciencias biológicas que pretendan realizar análisis estadísticos, bases de datos, análisis bioinformáticos, programación o gráficos mediante R.

Metodología

El curso es de carácter teórico/práctico y consta de ocho unidades. Para cerrar cada unidad (con excepción de la primera y la última) desafiamos a los alumnos a resolver con R un problema concreto de biología, adecuado para asimilar los contenidos.

Al finalizar el curso, y para aprobarlo, cada alumno deberá resolver un examen consistente en el tratamiento y análisis de datos con R.

Programa
UNIDAD 1

CABA, 10/03/17

Juan Hurtado

Qué es R. Por qué y para qué usar R en ciencias biológicas. Diferencias entre R y otros programas de análisis estadístico. Versatilidad y cobertura. Historia y evolución del programa. Iniciación: cómo bajar el programa, sitios web de interés.

UNIDAD 2

Cómo funciona R: entorno global, marcos y clausuras. Introducción al lenguaje R: comandos y tipos de objeto. Creación y edición de objetos con datos. Creación de funciones propias. Paquetes. El desafío para cerrar esta unidad será el de crear y nombrar objetos que les permitan cargar datos biológicos propios de cada alumno en la memoria de la computadora (o *workspace*).

UNIDAD 3

Interfaces amigables: Rstudio (y otros programas para comandar R). El desafío para cerrar esta unidad será el enviar los objetos propios creados en la unidad anterior a un compañero/a y cargar en su *workspace* los recibidos de otro compañero/a. Desde Rstudio, los alumnos deberán visualizar gráficamente y salvar en el disco de la computadora los objetos recibidos y extraerles información a partir de la creación de una función propia (e.g. una función que calcule el desvío estándar).

UNIDAD 4

Introducción a la programación con bucles. Construcción y consistencia de bases de datos: confección de tablas. Dificultades frecuentes y problemas comunes de biólogos. Búsqueda de ayuda. Cuándo no usar R. El desafío para cerrar esta unidad será el de crear una base de datos a partir de datos propios de cada alumno para luego aplicarle algunos filtros. Se espera que durante este desafío los alumnos adquieran buenas prácticas para confeccionar tablas de datos y tablas resumen típicas de las ciencias naturales, que programen alguna operación básica para filtrar las tablas y que aprendan a utilizar las herramientas de ayuda que ofrece el programa.

UNIDAD 5

Gráficos: funciones para crear, editar y personalizar figuras. Integración de diversas funciones para crear gráficos complejos. Introducción al paquete ggplot2. Gráficos para publicaciones científicas. Utilizando su propio set de datos creado en la unidad anterior, el desafío para cerrar esta unidad será el de comparar los distintos tipos de gráficos vistos en esta unidad y determinar cuál de ellos es el más adecuado para mostrar sus datos.

UNIDAD 6

Estadística descriptiva: abordaje gráfico y mediante estadísticos. Métodos de detección de datos atípicos. Datos faltantes. El desafío para cerrar esta unidad será el de curar un set de datos biológicos y decidir si la variable respuesta puede ser modelada con pocos parámetros.

UNIDAD 7

Simulaciones en biología: mi primer script. De la hipótesis biológica a la hipótesis en R. El desafío para cerrar esta unidad será el de responder mediante una simulación una pregunta de índole biológico. Los alumnos podrán optar por una pregunta de ecología de poblaciones, una de secuencias nucleotídicas o una de estadística descriptiva sobre un set de datos de una variable biológica.

UNIDAD 8

Integración y ejemplos de aplicaciones completas. Acceso desde R a distintas páginas web para descargar y utilizar bases de datos. Utilización de otros programas y análisis llamados desde R (Dropbox, Maxent, etc). Visualización de resultados en distintos tipos de mapas de GoogleEarth (terreno, satélite).

Bibliografía

- Andrew P. Beckerman AP & Petchey OL, 2012. *Getting Started with R: An introduction for biologists*.
 Crawley M, 2012. *The R book*. 2nd ed. Wiley.
 Dalgaard P, 2008. *Introductory statistics with R*. 2nd ed. Springer.
 Everitt B, 2007. *An R and S-PLUS companion to multivariate analysis*. 2nd ed. Springer.
 Everitt B & Hothorn T, 2006. *A handbook of statistical analyses using R*. Chapman & Hall.

CABA, 10/03/17

Juan Martínez