

Curso de Doctorado: “Fundamentos básicos del lenguaje R”

Unidad Académica organizadora:

- Doctorado en Ciencias Biológicas

Responsable Académico:

- Dr. Pablo Yair Huais (Profesor Asistente, Dpto. de Matemática, FCEFyN, UNC y Becario Postdoctoral en Instituto de Diversidad y Ecología Animal, IDEA (CONICET-UNC))

Temario a Desarrollar

1) Introducción al lenguaje R

1.1 R y paradigma POO. RStudio como entorno de desarrollo: consola, área de edición, área de gráficos, gestor de paquetes, memoria de trabajo. Uso de la ayuda nativa de R *online* y *offline*. Foros y bitácoras de la comunidad (i.e. Stack Overflow, Medium, R-bloggers, etc). Consideraciones sobre la asistencia con IA.

1.2 Instalación de paquetes. CRAN y otros repositorios (i.e. GitHub). Problemas comunes y cómo resolverlos. Librerías y dependencias en diferentes sistemas operativos. Compatibilidad: versiones de R y de paquetes.

2) Flujo de trabajo

2.1 Flujo de trabajo: consola y rutinas. Directorio de trabajo. Rutas relativas y absolutas. Ejecución de rutinas. Buenas prácticas de desarrollo de código. Uso de # para comentarios e índice en RStudio. Exportación de environment. Exportación/importación de objetos con saveRDS y readRDS.

3) Objetos, variables y funciones básicas

3.1 Operadores básicos. Creación de objetos. Tipos de objetos: escalares, vectores, *arrays*, matrices, listas, *dataframes*. Variables. Tipo de variables: *character*, *integer*, *numeric*, *factor*, *logical*, *NA*, *NULL*.

3.2 Trabajando con vectores: funciones básicas *round*, *floor*, *ceiling*, *mean*, *median*, *quantile*, *sd*, etc. Formas de generar vectores: *c*, *seq*, *rep*. Llamado a funciones mediante la forma *paquete::funcion*.

4) Tablas y estructuras complejas

4.1 Carga y manipulación básica de datos. Acceso a datos por filas y columnas. Filtrado de datos por condiciones. Formatos de tablas largo y ancho: función

reshape. Creación de nuevas columnas. Uso de las funciones *ifelse* y *cut*. Agregación de tablas: funciones *aggregate*, *rbind*, *cbind*, *merge*. Exportación de tablas. Uso de *pipes*, tablas *tibble* y universo *Tidyverse*.

4.2 Estructuras complejas: listas. Objetos de clase S3 y S4.

5) Control de flujo y funciones

5.1 Estructuras condicionales y bucles. Operadores booleanos y lógicos. Declaraciones *if...else*. Bucles *for* y *while*. Expresiones *next* y *break*. Funciones vectorizadas: *sapply*, *lapply*, *mapply*.

5.2 Declaración de funciones. Parámetros y argumentos. Valor de retorno. Recursividad. Variables locales y globales. Carga de scripts con *source*.

6) Gráficos

4.1 Funciones básicas de gráficos: *plot*, *points*, *lines*, *barplot*, *boxplot*, *hist*.

4.2 Universo *ggplot2* y su lógica programática. Funciones básicas: *geom_points*, *geom_bar*, *geom_boxplot*, *geom_histogram*, *geom_density*, *geom_violin*, *piecharts* con *coord_polar*. Uso de funciones accesorias: *theme*, *scale_x_continuous*, *scale_x_discrete*, *scale_fill_manual*, etc. Funciones *facet_grid* y *facet_wrap*. Exportación de gráficos.

Justificación y propósitos del curso

En los últimos 20 años, el desarrollo de nuevos lenguajes de programación y programas específicos ha facilitado el análisis de información científica. Entre las numerosas opciones, el lenguaje R se ha establecido como uno de los principales lenguajes para el análisis estadístico, así como para el análisis de gran cantidad de datos y el desarrollo de programas orientados a la resolución de problemas específicos de investigación. Particularmente en las Ciencias Biológicas, el lenguaje R ha ganado gran popularidad, siendo uno de los lenguajes más utilizados actualmente. La versatilidad de sus herramientas, impulsadas por el desarrollo de paquetes específicos para diversas tareas, abre un gran abanico de posibilidades de uso para la investigación científica. No obstante, el lenguaje presenta una curva de aprendizaje pronunciada y puede resultar un gran desafío para aquellos/as estudiantes que se inician en la actividad científica. Debido a la falta de formación básica durante los trayectos académicos de grado, es común que los/las estudiantes de doctorado no posean los conocimientos básicos necesarios para iniciarse en el uso habitual del

lenguaje. Esto suele derivar en frustraciones recurrentes que llevan al abandono del uso de una herramienta muy útil para la investigación científica. A pesar de que existe una amplia disponibilidad de cursos de posgrado formales, así como una oferta inabarcable de recursos didácticos para el autoaprendizaje (ya sean estos libres y gratuitos o de pago), son pocos los entornos de aprendizaje con formadores que se centren en consolidar los fundamentos básicos del lenguaje, en particular, y de la lógica programática, en general. De esta forma, espacios de aprendizaje pensados para enseñar aspectos aplicados (e.g. análisis ecológicos, modelos estadísticos, análisis genómicos, etc.) utilizando el lenguaje R suelen verse afectados por la falta de capacidades previas de las y los estudiantes en su uso, no pudiendo ser aprovechados de forma óptima para el desarrollo y aprendizaje de los conceptos aplicados. Debido a ello, se propone la oferta de un curso básico con el propósito de servir como fundamento de conceptos y destrezas de programación en el lenguaje R.

Objetivo general del curso

Proveer a los/las estudiantes con las herramientas necesarias para iniciarse en el uso del lenguaje R, mediante el aprendizaje de su lógica programática, orientado a resolver problemas específicos de sus temas de investigación.

Objetivos específicos del curso

- Que las y los estudiantes adquieran tanto conceptos teóricos fundamentales, como habilidades prácticas básicas del lenguaje R.
- Que las y los estudiantes desarrollen un pensamiento programático en relación al uso del lenguaje R.
- Que las y los estudiantes sean capaces de trasladar las herramientas aprendidas para la resolución de problemas metodológicos específicos de sus investigaciones.

Contenidos mínimos

1. Introducción al lenguaje R
2. Flujo de trabajo
3. Objetos, variables y funciones básicas
4. Tablas y estructuras complejas

5. Control de flujo y funciones
6. Gráficos

Nombre de el/los disertante/s (se adjuntan CVs reducidos)

- Dr. Pablo Yair Huais (Profesor Asistente, Dpto. de Matemática, FCEFyN, UNC y Becario Postdoctoral en Instituto de Diversidad y Ecología Animal, IDEA (CONICET-UNC)

Disertante invitado:

- Biól. Nicolás Pastor (Profesor Asistente, Dpto. de Diversidad Biológica y Ecología, FCEFyN, UNC y Profesional Asistente en Instituto de Antropología de Córdoba, IDACOR (CONICET-UNC)

Destinatarios de la actividad

- Alumnos del Doctorado en Ciencias Biológicas y afines. Egresados de la carrera de Biología y carreras afines.

Fecha de realización

- 9 de diciembre al 13 de diciembre 2024

Duración y programa de actividad diaria

- Duración: 40 hs. Las actividades se desarrollarán de forma presencial de lunes a viernes de 9:00 a 13:00 y de 14:00 a 18:00 hs.

Metodología a utilizar en el dictado

El curso tendrá una modalidad teórico-práctica. Se realizarán disertaciones teóricas introductorias, con especial énfasis en las bases conceptuales y prácticas de los temas a desarrollar. Seguidamente, se desarrollarán sesiones de actividades prácticas por parte de las y los estudiantes, enmarcadas en el aprendizaje basado en problemas; o bien, demostraciones prácticas a cargo de los docentes, abiertas a la interacción con el alumnado, evitando las validaciones instantáneas y propiciando un espacio de preguntas abiertas y repreguntas que promuevan la interacción dialógica, tanto para con los docentes como entre estudiantes.

Bibliografía y material didáctico que se proveerá a los asistentes

Se desarrollarán materiales teórico-prácticos, incluyendo diapositivas, rutinas de R y material accesorio para el desarrollo de las actividades prácticas. Todo el material se disponibilizará en un repositorio digital para su fácil acceso por parte de los/las estudiantes.

- **Bibliografía y links útiles:**

Libros

2023. “R for Data Science (2e)”. Wickham H, Çetinkaya-Rundel M, Grolemund G. Disponible en <https://r4ds.had.co.nz/>

2024. “An Introduction to R”. Douglas A, Roos D, Mancini F, Ana Couto A, Lusseau D. Disponible en <https://intro2r.com/>

Tutoriales y recursos digitales

- <https://www.statmethods.net/>
- <https://www.geeksforgeeks.org/r-tutorial/>
- <https://r-graph-gallery.com/>
- <https://r-charts.com/>

Evaluación final, metodología y profesores propuestos para realizarla

- **Evaluación:** Sí. Se realizará una evaluación continua sobre el desempeño de las y los estudiantes; a la par se realizará un examen final escrito teórico-práctico de carácter integrador. Para la evaluación de la actividad en sí misma, esto es, si la misma incide en el alumnado significativamente en las elaboraciones conceptuales, la identificación y problematización respecto a los temas propuestos, se tomarán las evaluaciones de las y los estudiantes como un indicador del éxito de la actividad.

- **Tribunal:**

1. Dr. Pablo Yair Huais
2. Dr. Santiago Miguel Benitez-Vieyra
3. Dr. Javier Nori

- **Aranceles:** \$40.000 (Se eximirá del pago del 20% del arancel a los estudiantes de las Carreras de Doctorado en Ciencias Biológicas de la FCEFYN, UNC).
- **Cupo:** 10 alumnas/os mínimo; 25 máximo.

Presupuesto estimativo y prioridades para la asignación de recursos

- **Honorarios:** No.
- **Gastos del curso:** refrigerios durante el cursado, a definir por el Doctorado en función de la cantidad de asistentes.

Entidad que operará como unidad ejecutora de recursos

- Doctorado en Ciencias Biológicas