

Profesor: Cristhian Julian Acosta Pardo

Salón: ML - 108B.

Horario de clase: miércoles 17:00 – 18:20.

Horario de atención: jueves 11:00 – 12:00, W-826.

1. Introducción y descripción general del curso

El curso Taller de R busca que los estudiantes adquieran una comprensión general de lenguaje de R y su utilidad para el análisis de datos, la investigación económica y otras aplicaciones académicas y profesionales. El énfasis será sobre la importación, organización y manejo de datos y la presentación de resultados, a través de herramientas que permitan una comunicación efectiva.

El curso no profundiza en conceptos estadísticos o econométricos; en cambio, busca que los estudiantes entiendan la estructura general de programación en R y la complementen con los conocimientos adquiridos en otras clases. Al final se espera que los estudiantes se sientan cómodos utilizando R para resolver problemas relacionados con análisis de datos; asimismo, se espera que identifiquen las diferentes aplicaciones de R y puedan continuar utilizándolo.

2. Objetivos de la materia

- Familiarizar a los estudiantes en el lenguaje de R.
- Motivar el uso de R a partir de la presentación de casos prácticos.
- Proporcionar herramientas para el manejo y uso de bases de datos en R.
- Enseñar a los estudiantes paquetes y comandos frecuentemente usados en R.
- Enseñar a los estudiantes formas de automatizar tareas en R.
- Proporcionar herramientas para que los estudiantes puedan continuar de manera autónoma su aprendizaje de paquetes y aplicaciones de R.

3. Competencias

Al final del curso, los estudiantes podrán:

- Entender los archivos de ayuda sobre paquetes de R.
- Escribir funciones básicas.
- Escribir códigos para solucionar problemas de importación y limpieza de datos.
- Manipular datos para obtener información relevante.
- Presentar resultados estadísticos de forma adecuada.
- Elaborar documentos HTML y PDF.

4. Metodología

El curso se realizará en salas habilitadas para el uso de computadores y estarán divididas en dos partes; durante la primera parte el profesor introducirá técnicas, comandos y conceptos relacionados con la utilización de R y utilizará ejemplos que lo ilustren; durante la segunda parte los estudiantes deberán trabajar individualmente en un ejercicio, el cual deberán entregar al finalizar la clase. Se calificarán todas las tareas que realicen, pero se tendrán en cuenta solamente las mejores cuatro (4) para la nota de tareas del semestre.

Adicionalmente, en el semestre se realizarán tres (3) talleres en clase sobre temas que hayan sido cubiertos; estos talleres deben ser desarrollados por los estudiantes durante la clase, sin embargo, tendrán hasta la medianoche del día siguiente para entregarlos.

El curso tiene siete (7) módulos, los cuales buscan iniciar al estudiante en el uso de R y presentar algunas aplicaciones, que motiven su continuo aprendizaje y utilización de R durante el curso y después de haberlo terminado.

5. Contenido y cronograma

Módulo 1: Introducción a R

Semana 1:

- ¿Qué es R? ¿Por qué R?
- Instalación de R y R Studio. Instalación y carga de paquetes.
- Paneles de R Studio y *scripts*.
- Personalización de R Studio.
- Conceptos básicos de sintaxis en R.
- Tipos de datos: entero, numérico, lógico, *character*. Missings: NaN, NA.
- Estructuras de datos: vectores, matrices, listas, *arrays*, *data frames*.
- Operaciones básicas.
- Funciones relevantes para la identificación y conversión de tipos de objetos y datos (coerción).
- ¿Dónde buscar ayuda?

Módulo 2: Funciones y *loops*

Semana 2:

- Asignar datos a objetos.
- Extraer información de objetos.
- Uso de vectores lógicos para extraer información de objetos.
- Comando "*function*".
- Uso de condicionales: *if*, *else*, *ifelse*.

Semana 3:

- Depuración de errores en funciones: *debug*, *undebug*, *traceback*.
- Inclusión de controles en las funciones: *warning*, *stop*.
- Loops: *while*, *for*, *repeat*.

Semana 4:

- Controles de flujo: *next*, *break*.
- Funciones *apply*, *sapply*, *tapply*, *lapply*.
- Instalar y cargar paquetes.
- Función *ddply*

Módulo 3: Bases de datos

Semana 5:

- Manejo básico *data frames*: selección de filas y/o columnas, selección condicionada.
- Creación de muestras y subconjuntos de *data frames*: *subset*, *sample_n*.
- Importación de datos: *dta*, *sav*, *csv*, *xlsx*.
- Inspección inicial de la información.

Semana 6:

- *Raw data vs tidy data*.
- Organización y limpieza de bases de datos con los paquetes *tidyr* y *dplyr*.
- Análisis descriptivo.
- Exportación de bases de datos: *dta*, *csv*, *xlsx*, *txt*.

Semana 7:

- Taller en clase #1: importación de base de datos *raw* y exportación de un archivo organizado (*tidy*).

Módulo 4: Visualización de datos

Semana 8:

- Gráficos del paquete base: histogramas, *scatter plot*, *bar plot*, *box plot*.
- Introducción a *ggplot2*.

Semana 9:

- Relación entre la organización de los datos y *ggplot2*: *dplyr* y *ggplot2*.
- Personalización de gráficos: ejes, colores, títulos.
- Exportación de gráficos.
- Otras herramientas para visualización de datos.

Módulo 5: Análisis espacial de datos

Semana 10:

- Análisis exploratorio de datos espaciales.
- Autocorrelación espacial (I de Moran).
- Métodos de corrección por autocorrelación espacial.

Semana 11:

- Taller en clase #2: ejercicio de visualización de datos y análisis espacial.

Módulo 6: Otras aplicaciones en R

Semana 12:

- Regresión lineal.
- Métodos de regresión para variables dicótomas.
- Aplicación en series de tiempo: modelos autorregresivos.
- Logit y Probit

Semana 13:

- Introducción a *webscrapping*: *rvest* (nociones básicas)
- Introducción a *text mining*: *tm*, *tidytext*. (nociones básicas)

Módulo 7: RMarkdown y Shiny

Semana 14:

- Introducción a RMarkdown.
- Análisis de componentes principales (PCA).
- Análisis de cluster.

Semana 15:

- Taller final en clase: importación y limpieza de datos, análisis descriptivo, visualización de datos y presentación de conclusiones utilizando RMarkdown.

Fechas importantes:

Inicio de clases: 21 de enero

Semana de trabajo individual: 15 a 19 de abril

Entrega del 30%: 15 de marzo

Último día de retiros: 22 de marzo (6 p.m.)

Calificación de cursos y profesores: 22 abril - 10 de mayo

Último día de clases: 11 de mayo

6. Criterios de evaluación

El principal criterio para la evaluación es los tres talleres que se realizarán en clase. La calificación final será determinada de la siguiente manera:

Taller en clase #1: 30%

Taller en clase #2: 25%

Taller final en clase: 30%

Ejercicios en clase: 15%

Los talleres en clase se evaluarán teniendo en cuenta: 1) Organización del código/*script*, 2) Que el código genere un resultado correcto y 3) Número de errores en el código. Los talleres deben ser presentados a través de SicuaPlus. Para aprobar el curso cada estudiante debe alcanzar una nota igual o superior a 3.00. Las calificaciones definitivas de las materias serán numéricas de uno punto cinco (1,50) a cinco punto cero (5,00), en unidades, décimas y centésimas.

7. Ajustes razonables

Si usted lo considera necesario o importante, siéntase en libertad de informar lo antes posible si usted tiene alguna condición o discapacidad visible o invisible y requiere de algún tipo de apoyo o ajuste para estar en igualdad de condiciones con los y las demás estudiantes, de manera que se puedan tomar las medidas necesarias. En caso en que decida informar, por favor, justifique su solicitud con un certificado médico o constancia de su situación.

Lo invitamos a buscar asesoría y apoyo en la Dirección de su programa, en la Decanatura de Estudiantes (<http://centrodeconsejeria.uniandes.edu.co> Bloque Ñf, ext.2330, horario de atención L-V 8:00 a. m. a 5:00 p. m.) o en el Programa de Acción por la Igualdad y la Inclusión Social (PAIIS) de la Facultad de Derecho (paiis@uniandes.edu.co).

Se entiende por ajustes razonables todas "*las modificaciones y adaptaciones necesarias y adecuadas que no impongan una carga desproporcionada o indebida, cuando se requieran en un caso particular, para garantizar a las personas con discapacidad el goce o ejercicio, en igualdad de condiciones con las demás, de todos los derechos humanos y libertades fundamentales*" Convención sobre los Derechos de las personas con discapacidad, art.2.

8. Bibliografía

- R Development Core Team. *R-Manuals*. <https://cran.r-project.org/manuals.html>.
- Wiley, M, Wiley, J. *Advanced R Data Programming and the cloud*. Apress. 2016
- Wickham, H. *Advanced R. The R Series*. CRC Press. 2015.
- S.P. Paul. *Introductory Time Series with R*.
- Chang Winston. *R graphics Cookbook*. O'Reilly.
- Matthias Kohl. *Introduction to statistical data analysis with R*.
- Shiny. <https://shiny.rstudio.com>.