

1. Horario atención a estudiantes, correos electrónicos y nombres de los profesores complementarios

Clase magistral

Salón: ML108

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

Profesor: Eduard Fernando Martínez González

Horario de atención a estudiantes: jueves 15:00 – 16:00

Lugar de atención a estudiantes: W722

2. 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.

3. 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.

4. Organización del curso

Módulo 1: Introducción a R

Semana 1 (7 de agosto):

No hay actividades (día no laboral).

Semana 2 (14 de agosto):

- ¿Qué es R? ¿Por qué R?
- Instalación de R y R Studio. Instalación y carga de paquetes.
- Interfaz de R Studio y *scripts*.
- Conceptos básicos de sintaxis en R.
- Extraer información de objetos.
- 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).

Módulo 2: Bases de datos

Semana 3 (21 de agosto):

- Instalar y cargar paquetes.
- Importar y exportar bases de datos en formato dta, csv, xlsx, txt.
- Inspección inicial de la información.
- Descriptivas básicas de las bases de datos (*group_by*, *summary*, *summaryBy*).
- Creación de subconjuntos de *data frames* (filtrar por filas y/o columnas, selección condicionada).

Semana 4 (28 de agosto):

- Paquete *tidyverse*.
- String functions.
- Organizar y limpiar bases de datos.
- Unir bases de datos (por filas y columnas).
- Transponer bases de datos (*reshape*, *reshape2* y *data.table*).

Módulo 3: Funciones y loops

Semana 5 (4 de septiembre):

- Loops: *while*, *for*, *repeat*.

- Controles de flujo: *next*, *break*.
- Funciones *apply*, *sapply*, *tapply*, *lapply*.
- Uso de condicionales: *if*, *else*, *ifelse*.

Semana 6 (11 de septiembre):

- Función “*function*”.
- Depuración de errores en funciones: *debug*, *undebug*, *traceback*.
- Inclusión de controles en las funciones: *warning*, *stop*.

Semana 7 (18 de septiembre):

No hay actividades (Día Paiz).

Semana 8 (25 de septiembre):

- Taller en clase #1: Importar y organizar bases de datos.

Semana 9 (2 de octubre):

Semana de receso.

Módulo 4: Visualización de datos

Semana 10 (9 de octubre):

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

Semana 11 (16 de octubre):

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

Semana 12 (23 de octubre):

- Paquetes: *rgdal*, *rgeos*, *raster* y *ggmap*.
- Importar y visualizar información espacial.
- Mapas temáticos.

Módulo 5: Análisis espacial de datos

Semana 13 (30 de octubre):

- Introducción al procesamiento de imágenes satelitales

Semana 14 (5 de noviembre):

- Taller en clase #2: Analisis espacial y visualizacion de datos.

Módulo 6: Otras aplicaciones en R**Semana 15 (12 de noviembre):**

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

Semana 16 (19 de noviembre):

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

Semana 17 (26 de noviembre):

- Taller final en clase: Importar y limpiar bases de datos, análisis descriptivo, visualización de datos y presentación de conclusiones.

5. 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 calificaran todas la 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.

6. 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.
- Usar R como un SIG
- Presentar resultados estadísticos de forma adecuada.

7. Criterios de evaluación (Porcentajes de cada 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:

- I. Organización del código/*script*.
- II. Que el código genere un resultado correcto.
- III. 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.

8. Sistema de aproximación de notas definitiva

Las calificaciones definitivas de las materias serán numéricas de uno punto cinco (1.5) a cinco punto cero (5.0), en unidades y décimas.

Reclamos: Reclamos ¿Cómo y en qué tiempos debe presentar un estudiante un reclamo sobre su nota en cualquier evaluación del curso? Según los artículos 62 y 63 del Reglamento general de estudiantes de pregrado, el estudiante tendrá cuatro días hábiles después de la entrega de la evaluación calificada para presentar un reclamo. El

profesor responderá al reclamo en los cinco días hábiles siguientes. Si el estudiante considera que la respuesta no concuerda con los criterios de evaluación podrá solicitar un segundo calificador al Consejo de la Facultad en los cuatro días hábiles posteriores a la recepción de la decisión del profesor.

9. Bibliografía

- Gentleman, R., Hornik, K., & Parmigiani, G. (2009). Use R!.
- Mas, Jean-Francois. (2019). Análisis espacial con R: Usa R como un sistema de información geográfica.
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. John Wiley & Sons.
- Quintana, L., & Mendoza, M. (2016). Econometría aplicada utilizando R. *Ciudad de México: Universidad Nacional Autonoma de México*.

Fecha de entrega del 30% de las notas: 4 de octubre

Último día para solicitar retiros: 11 de octubre (6pm)

Último día para subir notas finales en banner: 18 de diciembre