



PROGRAMACIÓN DE ACTIVIDADES

Opt. Análisis de datos en R

Profesora Valentina Andrade / valentinaandrade@uchile.cl

Ayudante Dafne Jaime / dafne.jaime@ug.uchile.cl

No.de clase 3188

17.30 – 18.50

19.00 – 20.20

Segundo semestre 2022

I. PROPÓSITOS DE APRENDIZAJE

Se espera que al término del curso los estudiantes sean capaces de:

- ✓ Manejar R y herramientas asociadas a su utilización (*rOpensci*), utilizando prácticas que les permitan avanzar en su aprendizaje de manera autónoma (Unidad 1)
- ✓ Manipular, procesar y limpiar datos sociales utilizando R (Unidad 2)
- ✓ Aplicar herramientas para análisis estadísticos descriptivos en R (Unidad 3)
- ✓ Aplicar herramientas para análisis estadísticos inferenciales en R (Unidad 4)
- ✓ Presentar resultados, a partir de la visualización de datos y construcción de documentos (transversal a todas las unidades)

II. CONTENIDOS

Unidad 1: Elementos y herramientas básicos de R

- 1.1 R environment: interfaz de RStudio, elementos de script, workspace
- 1.2 Prácticas y herramientas de consulta: CRAN, stackoverflow, Rcommunity
- 1.3 Herramientas para la colaboración y comunicación: Rprojects, GitHub y Slack
- 1.4 Librerías y funciones para las ciencias sociales: tidyverse y sj (sjmisc y sjPlot)
- 1.5 Construcción de reportes reproducibles e integrados con código: RMarkdown

Unidad 2: Manipulación y limpieza de datos

- 2.1 Tipos de datos en R
- 2.2 Importar y exportar datos en diferentes formatos

2.3 Selección de variables y filtrar datos

2.4 Transformación de variables

Unidad 3: Análisis estadístico descriptivo en R

3.1 Análisis descriptivos univariados: medidas de tendencia central, dispersión y frecuencias

3.2 Análisis bivariado: tablas de contingencia, correlaciones y representación gráfica

3.3 Representación gráfica con *sjPlot*: Likert, proporciones agrupadas y distribuciones

Unidad 4: Análisis estadístico inferencial en R

4.1 Muestras complejas y precisión de inferencia estadística con *survey* y *srvyr*

4.2 Regresiones lineales, predictores categóricos y representación gráfica

4.3 Regresiones logísticas, exponenciación y representación gráfica

II. METODOLOGÍA

Dado el contexto de pandemia se tendrán tres espacios principales de aprendizaje:

1. **Sesiones de clases lectivas**, donde se presentarán los aspectos centrales de los contenidos correspondientes a la semana. El documento de presentación de la clase se encontrará disponible en la pestaña de Contenidos de este sitio web del curso. Estas se desarrollarán los días *lunes*.

2. **Prácticas guiadas**: cada tema de las sesiones se acompaña de una guía práctica de aplicación de contenidos. Estas guías están diseñadas para ser desarrolladas de manera autónoma por cada estudiante semana a semana. También serán desarrolladas y revisadas cada semana en grupos pequeños con supervisión de ayudantes para dar mayor oportunidad de participación y resolver las dudas respectivas. Estas se desarrollan los días

3. **Evaluaciones**: se desarrollarán **tareas** periódicas que permitirán a las/os estudiantes aplicar contenidos y replicar lo aprendido en los prácticos en base a una base de datos seleccionada por ellas/os a inicio de semestre. Esto permitirá no solo recibir retroalimentación constante, sino que aprender con datos que puedan ser útiles para otros proyectos de investigación. Al finalizar el curso, el/la estudiante deberá entregar como examen un **proyecto de investigación**.

Las instrucciones de las tareas serán publicadas con una semana de anticipación a su entrega. Además, los/as estudiantes dispondrán del bloque de clases para poder desarrollarla en conjunto con profesora y ayudante.

Recursos principales de aprendizaje y comunicación

1. Sitio web (<https://learn-r-uah.netlify.app/>)

El curso estará disponible en un sitio web programado por la docente, en tanto permite integrar texto y código de R.

2. R, RStudio y RStudio Cloud

El software que se utilizará principalmente será R y su interfaz RStudio. Ahora bien, muchos usuario/as de R presentan problemas de instalación dada la capacidad de sus computadores y sistemas operativos. Por ello, para quienes tengan estos problemas se promoverá el uso del servicio gratuito de RStudio.cloud

3. Slack

Slack es una herramienta de uso frecuente en equipos de trabajo que utilizan R pues permite integrar script de distintos lenguajes en el chat. Se tendrá un espacio de trabajo en la *app Slack* que permite que cualquier persona del curso pueda hacer preguntas y cualquiera pueda responder. Esta es una de las prácticas que se promoverán en el curso pues es probable que los estudiantes tengan dudas similares a las de sus compañeros, por lo que las respuestas de la docente, ayudante y otros compañeros/as serán de libre disposición de todo el curso. Dentro de *Slack* se tendrán canales específicos para hacer preguntas sobre las sesiones, tareas y proyectos, y el link que permite unirse a este estará disponible en el sitio del curso.

4. GitHub

GitHub es una plataforma online que permite depositar archivos y el control de versiones (VCS), por lo que se ha transformado una herramienta fácil y popular para *corregir, colaborar y compartir* códigos de distintos lenguajes (no solo R). Utilizaremos esta plataforma para subir las tareas, ayudarlos/as de manera directa con su código y darles feedback.

III. EVALUACIÓN DE APRENDIZAJES

Las evaluaciones del curso se componen de **tareas** (70% de la nota final) y la entrega de una **investigación** (30% de la nota final), en dónde en ambos casos la/el estudiante deberá seleccionar datos y temas de interés de modo de acercar la aplicación del software a contextos de investigación propios de la/el estudiante. En concreto, cada evaluación consiste en:

1. Tareas (70% de la nota final): consisten en evaluaciones parciales temáticas que buscan poner en práctica los aprendizajes expuestos en la sesión de *clases* y herramientas reforzadas en los *prácticos*. Durante el semestre se realizarán 4 tareas (25% c/u).

2. Investigación final (30% de la nota final): consiste en una evaluación final individual que aplica los conocimientos y herramientas entregadas a lo largo de curso, a un proyecto de investigación de elección por el/la estudiante.

Información general

Los justificativos por inasistencia a clases y/o evaluaciones por enfermedad y/o otras razones deben hacerse a la coordinación de la carrera en el plazo establecido por el reglamento. Quienes no lo hagan serán evaluados con nota 1,0 o con inasistencia a clases, según corresponda. Para presentarse a examen se requiere nota promedio 3,5 y haber rendido todas las evaluaciones (incluyendo la tarea 0) . En caso de no cumplir tales requisitos no podrá rendir examen.

Reglamento Académico del Estudiante de Pregrado. Art. 23.- Cualquier conducta de un estudiante que tienda a viciar la evaluación de actividades académicas o que constituya fraude académico, figura que contempla irregularidades tales como copia, suplantación o alteración de evaluaciones, plagio, faltas a la ética profesional, sin que esta enumeración sea taxativa, dará origen a las siguientes sanciones, según la gravedad de la falta cometida: (i) nota mínima 1,0 (uno) en la respectiva evaluación; (ii) reprobación del curso respectivo; (iii) amonestación; (iv) permanencia condicional; (v) suspensión de actividades académicas por un período académico; (vi) expulsión de la Universidad.

Asimismo, toda actividad de un estudiante que entorpezca gravemente y/o dificulte el normal desarrollo académico, podrá ser sancionada de conformidad a las disposiciones establecidas en el Reglamento de Conducta y Convivencia de la Universidad Alberto Hurtado.

Art. 24.- Las dos primeras sanciones previstas en el artículo anterior, a saber (i) Nota mínima 1,0; y (ii) Reprobación del Curso respectivo, son prerrogativa del docente a cargo de la asignatura, quien deberá informarlas a la Dirección de la Carrera.

***Para evitar el plagio todo trabajo, composición o material documental que los estudiantes realicen debe citar adecuadamente las fuentes utilizadas, ya sea a través del sistema APA (American Psychological Association) <http://www.apastyle.org> o MLA (Modern Language Association) <http://www.mla.org/>.

IV. CALENDARIZACIÓN DE ACTIVIDADES

Sesión	Semana	Unidad	Contenido de cada sesión	Práctico	Evaluaciones
1	8-agosto	Unidad 1. Elementos y herramientas básicos de R	1.1 R environment: interfaz de RStudio, elementos de script, workspace	Práctico 1: Creando un script, R project y Github	
	1.2 Prácticas y herramientas de consulta: CRAN, stackoverflow, Rcommunity				
	1.3 Herramientas para la colaboración y comunicación: Rprojects, GitHub y Slack				
2	22-agosto		1.4 Librerías y funciones para las ciencias sociales: tidyverse y sj (sjmisc y sjPlot)	Práctico 2: Reporte en Rmarkdown	Tarea 0: Reporte en Rmarkdown
		1.5 Construcción de reportes reproducibles e integrados con código: Rmarkdown			
3	29-agosto	Unidad 2. Manipulación y limpieza de datos	2.1 Tipos de datos en R	Práctico 3: Tipos de datos en R	Tarea 1: Procesamiento de datos
4	12-sep		2.2 Importar/exportar datos en diferentes formatos	Práctico 4: Importación de datos	
5	26-sep		2.3 Selección de variables y filtrar datos	Práctico 5: Limpieza de datos	
6	3-oct		2.4 Transformación de variables	Práctico 6: Transformación de variables Práctico 7: Análisis descriptivo univariado	
7	17-oct	Unidad 3. Análisis estadístico descriptivo en R	3.1 Análisis descriptivos univariados: medidas de tendencia central, dispersión y frecuencias	Práctico 8: Análisis descriptivo bivariado	Tarea 2: Análisis descriptivo en R
8	24-oct		3.2 Análisis bivariado: tablas de contingencia, correlaciones y ANOVA		
9	7-nov	Unidad 4. Análisis estadístico inferencial en R	4.1 Muestras complejas y precisión de inferencia estadística con survey y srvyr	Práctico 9: Muestras en R	Tarea 3: Muestras complejas
10	14-nov		4.2 Regresiones lineales y logísticas, predictores categóricos y representación gráfica	Práctica 10: Regresiones en R	Tarea 4: Regresiones en R
	21-nov				Entrega investigación final (30%)

V. RECURSOS PEDAGÓGICOS

1. Bibliografía Básica

- Wickham, H., & Golemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data (First edition). Sebastopol: O'Reilly. Libro con enfoque en el aprendizaje de R. Disponible en español como "R para ciencia de datos"
- Daniel Lüdecke (2021) Data Visualization for Statistics in Social Science R package versión 2.8.7. <https://CRAN.R-project.org/package=sjPlot>
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Yihui Xie, J. J. Allaire, Garrett Golemund (2021) R Markdown: The Definitive Guide
- Bryan, Jenny (2019) Happy Git in R

2. Sitios de consulta

- *rOpensci* (R Open Science Tools)
- Laboratorio de Ciencia Social Abierta, Centro de Estudios de Conflicto y Cohesión Social (LISA-COES)
- Stackoverflow
- RStudio Community
- RMarkdown
- sjPlot
- tidyverse