



PROGRAMACIÓN DE ACTIVIDADES

Opt. Análisis de datos estadísticos en R

Profesora Valentina Andrade [/valentinaandrade@uchile.cl](mailto:valentinaandrade@uchile.cl)

No.de clase 2255

Segundo semestre 2021

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 de 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 Importar y exportar datos en diferentes formatos
- 2.2 Validación y limpieza de variables (missing values)
- 2.3 Transformación y selección de variables
- 2.4 Transformación de datos en tidydata con tidy

Unidad 3: Análisis estadístico descriptivo en R

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

3.2 Análisis descriptivo bivariado: tablas de contingencia, correlaciones y ANOVA

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 Pruebas de hipótesis y representación gráfica

4.3 Regresiones lineales, predictores categóricos y representación gráfica

4.4 Regresiones logísticas, exponenciación y representación gráfica

4.5 Ajuste de modelos (performance) y otras representaciones gráficas (predicción, efectos marginales e interacciones)

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 vía Zoom. Tanto el documento de presentación como el video de la clase se encontrará disponible en la pestaña de Contenidos de este sitio web del curso.

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.

Existe un **reporte de progreso** asociado a estas guías que deberá ser completado semanalmente con fines de monitoreo y retroalimentación.

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**.

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. Esto permitirá que los participantes del curso puedan interactuar comentando sus dudas asociadas a cada ejercicio práctico a través del foro Disqus. Sumado a ello, permite registrar y dar disponibilidad el material del curso de manera libre.

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 se promoverá el uso del servicio gratuito de RStudio.cloud , que permite ejecutar la interfaz de RStudio en el navegador web y compartir el código de manera sincrónica con la docente. Dado el enfoque *rOpenSci*, las plantillas para ejemplos, ejercicios y mini proyectos podrán ser implementados en esta plataforma. Si

bien no es ideal pues tiene limitantes de memoria, mientras las/os estudiantes asimilan el programa será una buena herramienta

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 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.

5. Zotero

Zotero es un gestor bibliográfico que permite sistematizar las referencias y archivos utilizados en una investigación o informe. Enseñaremos este de manera complementaria pues este software se puede integrar en los documentos escritos hechos en R.

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*. El **promedio de notas las de tareas** será calculado solo con las **cuatro mejores entregas** a partir de la **Tarea 1** (25% c/u)¹. La evaluación es de carácter **individual o en parejas**.

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

¹ Es decir, a lo largo del semestre deberá entregar 6 tareas: 5 calificadas (Tarea 1 a Tarea 5) y 1 no (Tarea 0). De las 5 tareas restantes, solo las 4 mejores serán consideradas en su promedio de las tareas.

IV. CALENDARIZACIÓN DE ACTIVIDADES

Sesión	Fecha	Unidad	Contenido de cada sesión	Práctico	Evaluaciones	
1	09-ago	Unidad 1. Elementos y herramientas básicos de R	1.1 R enviroment: interfaz de RStudio, elementos de script, workspace	Práctico 1: Creando un script, R project y Github	Tarea 0: Colaborando en Github	
			1.2 Prácticas y herramientas de consulta: CRAN, stackoverflow, Rcommunity			
			1.3 Herramientas para la colaboración y comunicación: Rprojects, GitHub y Slack			
2	16-ago		1.4 Librerías y funciones para las ciencias sociales: tidyverse y sj (sjmisc y sjPlot)	Práctico 2: Reporte en Rmarkdown	Tarea 1: Reporte en Rmarkdown	
			1.5 Construcción de reportes reproducibles e integrados con código: Rmarkdown			
3	23-ago	Unidad 2. Manipulación y limpieza de datos		Taller N°1: Datos y proyecto de investigación	Tarea 2: Procesamiento de datos	
			2.1 Importar/exportar datos en diferentes formatos			
			2.2 Validación y limpieza de variables			
4	30-ago			2.3 Transformación y selección de variables		Práctico 3: Limpieza de datos
5	06-sept		2.4 Transformación de datos en tidydata	Práctico 4: Transformación y selección de variables		
6	13-sept	Unidad 3. Análisis estadístico descriptivo en R	3.1 Análisis descriptivos univariados: medidas de tendencia central, dispersión y frecuencias	Práctico 5: Transformación de estructura de datos	Tarea 3: Análisis descriptivo	
			3.2 Análisis bivariado: tablas de contingencia, correlaciones y ANOVA			
	20-sept			TallerN°2: Proyecto de investigación + Zotero		
			Receso			
7	27-sept	Unidad 4. Análisis estadístico inferencial en R		Práctico 6: Análisis descriptivos y representación gráfica	Tarea 4: Muestras complejas	
			4.1 Muestras complejas y precisión de inferencia estadística con survey y srvyr			
8	04-oct			4.2 Pruebas de hipótesis y representación gráfica	Práctico 7: Muestras en R	
	11-oct			Práctico 8: Pruebas de hipótesis		
				Feriado		
9	18-oct			4.3 Regresiones lineales, predictores categóricos y representación gráfica	Práctica 9: Regresiones lineales en R	Tarea 5: Regresiones en R
10	25-oct			4.4 Regresiones logísticas, exponenciación y representación gráfica	Práctico 10: Regresiones logísticas en R	
	01-nov			Feriado		
11	08-nov				Práctico 11: Representación gráfica de modelamiento	
			4.5 Calidad de modelos y otras técnicas de estimación			
15	15-nov		Receso			
16	22-nov				Entrega investigación final (30%)	
17	29-nov					

1. Bibliografía Básica

- Wickham, H., & Grolemund, 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 Grolemund (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