

Herramientas cuantitativas para el análisis político

Jueves, 19-22hs

(20 de septiembre al 29 de noviembre de 2018)

Juan Pablo Pilorget

jpilorget@gmail.com

Juan Pablo Ruiz Nicolini

juan.ruiznicolini@mail.utdt.edu

Descripción del taller: en su introducción a los conceptos básicos de las ciencias sociales, Jon Elster utilizó la metáfora de *Tuercas y Tornillos*¹ para fundamentar la importancia de entender los engranajes detrás de los fenómenos sociales. Partiendo de esta idea, el taller pretende introducir un conjunto de herramientas para la investigación y el análisis de fenómenos políticos que podrían ser pensados como los lubricantes de las *tuercas y tornillos*.

Utilizaremos el entorno de **RStudio** - para el manejo de datos, análisis estadístico e ilustración de resultados-, **RMarkdown** para la presentación de documentos y una variedad de recursos de fácil acceso para mejorar la comunicación del trabajo académico. Tener conocimientos básicos de estadística y comprensión lectora en inglés será de gran ayuda para los participantes.

Entre las motivaciones del taller se destacan:

- Introducir a los participantes en el descubrimiento y la utilización de herramientas de libre disposición y fácil acceso.
- Desarrollar mediante la práctica una metodología de aproximación inicial a los fenómenos políticos a través del análisis exploratorio de datos y la inferencia estadística.

¹ *Nuts and Bolts for the Social Science*, Cambridge University Press, 1989 / *Tuercas y Tornillos: una introducción a los conceptos básicos de las ciencias sociales*; Gedisa, 1990.

Lecturas

Si bien a lo largo de las sesiones se compartirá material específico sobre los temas tratados, las siguientes lecturas fueron parte importante para el armado del programa:

- * *R for Data Science*. **Autores:** Hadley Wickham, Garrett Grolemund; O' Reilly, 2016
- * *Political Analysis Using R*. **Autor:** James E. Monogan III; Springer, 2015
- * *ggplot2: Elegant Graphics for Data Analysis*. **Autor:** Hadley Wickman; Springer, 2009
- * *La introducción no tan corta a L^AT_EX*. **Autor:** Tobias Oetiker; 1995
- * *Computer Age Statistical Inference*. **Autores:** Bradley Efron, Trevor Hastie; Cambridge University Press, 2016

Objetivos del Taller

Al completar el curso se espera que los participantes:

1. Conozcan y utilicen **R** + **RStudio** como *software* para la manipulación de datos estructurados y la confección de modelos estadísticos.
2. Sean capaces de producir gráficos y tablas de calidad en **R** (utilizando **ggplot2** y **stargazer**, entre otras herramientas).
3. Usen **L^AT_EX** y **RMarkdown** para la creación y edición de documentos (artículos, presentaciones, bibliografía, etcétera).
4. Se familiaricen con herramientas adicionales para el trabajo con datos y la utilización de *softwares* estadísticos (**GitHub**, **StackOverflow**, **StackExchange**, etcétera).

Estructura del curso:

El contenido semanal podrá ser modificado en función del progreso de la clase.

Semana	Contenido
Semana 1	▪ Lubricar las <i>Tuercas y Tornillos</i> Presentación del curso y trabajo preparatorio.
Semana 2	▪ Introducción a R.
Semana 3	▪ Manejo de datos en R (“ <i>data wrangling</i> ”).
Semana 4	▪ Introducción a <i>ggplot2</i> Estructura de la <i>gramática de los gráficos</i> (aplicado a R).
Semana 5	▪ Reportes, presentaciones y documentos con RMarkdown
Semana 6	▪ Un paso más allá con <i>ggplot</i> para geolocalizar información.
Semana 7	▪ Estadística descriptiva e inferencial para el análisis político.
Semana 8	▪ Análisis de datos en R: la regresión de <i>Mínimos Cuadrados Ordinarios</i> (OLS) como puerta de entrada al mundo de las inferencias.
Semana 9	▪ Análisis de datos en R (II): <i>Modelos lineales generalizados</i> (GLM).
Semana 10	▪ Introducción a tópicos intermedios y avanzados.
Semana 11	▪ Introducción a tópicos intermedios y avanzados (II).

Semana 1

(A) Descripción de los 3 grandes módulos del taller: (1) Manejo y análisis de datos; (2) presentación de resultados y (3) producción de documentos científicos. (B) Tareas preparatorias para el taller: programas necesarios + encuesta.

Semana 2

¿Por qué R? Funcionamiento general. Primeros pasos con RStudio (consola, *scripts*, *enviroment*, ayuda, paquetes, etc.)

- **Práctica:** procesamiento de la encuesta a participantes (estadísticas descriptivas y gráficos con R `base`).

Semana 3

Presentación de `tidyverse` para el manejo y limpieza de bases de datos (*data wrangling*):

- Aplicación de `readr` para importar bases de datos, `tidyr` para limpiarlos y `dplyr` para transformarlos.

Semana 4

- ¿Qué es la *Gramática de los Gráficos*? Leland Wilkinson y una nueva manera de visualizar la información estadística.
- ¿Cómo se aplica a R mediante `ggplot2`?
- Los 5 fantásticos de `ggplot`: ejemplos prácticos para la creación de *scatterplots*, *linegraphs*, *boxplots*, *histograms* y *barplots*.
- **Ejemplo:** “*El impacto de enchufar los votos*”, (Ruiz Nicolini, 2017).

Semana 5

Uso de RMarkdown para la producción de documentos científicos (artículos, informes y presentaciones) y su relación con L^AT_EX:

- Descripción de las herramientas y versiones disponibles.
- Comandos básicos para la redacción de un documento.
- Más allá del texto: fórmulas matemáticas, manejo de figuras, tablas, citas, bibliografía, referencias e índices en la preparación de documentos científicos.
- Presentaciones usando `beamer` y RMarkdown.

Semana 6

Representación de información con mapas, grillas y polígonos. Ejemplos aplicados con:

- `ggmap`
- `geofacet`
- `sf`

Semana 7

Los métodos de estadística descriptiva e inferencial

- Medidas de centralidad y dispersión.
- Técnicas exploratorias para el análisis: la reducción de dimensiones.
- Tests de hipótesis: diferencia de medias e independencia de variables.
- Coeficientes de correlación y asociaciones entre variables.

Semana 8

Análisis estadístico en Ciencia Política.

- La regresión lineal como primera aproximación a la detección de regularidades.
- Anscombe y los riesgos de la regresión lineal.
- Los supuestos y su verificación.
- El problema de la multicolinealidad.
- Variantes robustas a los Mínimos Cuadrados Ordinarios.
- **Estadísticas descriptivas.** Ejercicio práctico: análisis de base de datos de encuesta a participantes de prueba piloto de voto electrónico en la provincia de Salta (Monogan, 2015).
- **Análisis de regresión.** Estimando la democratización de los países mediante sus fronteras.

Semana 9

Primera aproximación a *Modelos lineales generalizados* (GLM).

- ¿Qué es lo que se generaliza en un modelo *generalizado*?
- Los distintos sabores de regresiones: logística, Ridge y LASSO.
- Regularización y variaciones en los atributos a predecir.
- **Práctica:** *¿cómo predecir si una persona va a votar en las elecciones?*

Semanas 10 y 11

Introducción a tópicos intermedios y avanzados de análisis estadístico:

- Regresiones multinivel con `lme4`.
- Regresiones bayesianas con `rstanarm`.
- Análisis de series de tiempo en modelos de regresión lineal.