

Instrumentos Computacionales

Cristian Bonavida

Matías Ciaschi¹

Año 2023

El objetivo de este curso es introducir los conceptos fundamentales de los softwares estadísticos STATA y R, que se utilizan en varios cursos de la Maestría en Economía de la UNLP. El curso se desarrolla en cinco clases durante las cuales se introducen los conceptos fundamentales de cada software, se entrena en la aplicación de diferentes comandos y herramientas básicas, utilizando diversas bases de datos. Las clases se realizarán de forma presencial y, en caso de ser necesario, se incluirá una clase adicional de consulta (virtual, vía Zoom) para considerar dudas particulares y generales de los diversos temas abordados en el curso.

La evaluación del curso se realiza mediante un examen práctico domiciliario obligatorio. La misma también incluye trabajos prácticos domiciliarios. La materia consta de 5 clases dictadas los días 14/4 al 12/5, una por semana. En las primeras tres clases se enfocarán en temas similares para ambos softwares, mientras que la cuarta y la quinta se centrarán en temas específicos de STATA y R, respectivamente.

A continuación, se presenta el programa del curso, el cronograma estimativo y el contenido mínimo de las clases.

Programa y cronograma de clases de Stata

Clase 1:

- **Introducción a STATA:** interface, sintaxis general, tipos de bases de datos (encuestas de hogares, datos administrativos), creación de proyectos.
- **Comandos básicos:** apertura de la base de datos, descripción del contenido de la misma. Operadores lógicos y sentencias condicionales.

¹ Consultas: cristianbonavida@gmail.com y matiasciaschi@gmail.com.

Creación de variables y modificación de contenidos de las mismas. Variables *dummies* y variables aleatorias. Crear, ordenar, renombrar y eliminar variables. Opciones de los comandos.

Clase 2:

- **Análisis de datos:** cómputo de estadísticas descriptivas, ponderadores, correlaciones, tabulación de datos.
- **Gestión de base de datos:** importación de datos externos (formatos .xls, .csv, .txt) y exportación de una base de datos. Datos experimentales. Etiquetar variables y valores de las variables. Tipo de datos (numérico, *string*, *elapsed dates*) y modificación del tipo de dato de una variable.

Clase 3:

- **Gestión de base de datos:** Unión y combinación de bases de datos. Comandos para reestructurar la organización de los datos.
- **Análisis de regresión:** estimación por mínimos cuadrados ordinarios, interpretación de los resultados, comandos post-estimación, obtención de valores de coeficientes, y tests de hipótesis. Herramientas gráficas y opciones adicionales útiles en el análisis de regresión.

Clase 4:

- **Matrices y variables temporales:** creación de matrices y variables temporales, computación agregada de variables descriptivas, bucles simples y condicionales, exportación de matrices.
- **Bases de datos temporales:** creación de bases de datos temporales, uso y combinación entre las mismas, ventajas respecto a uso tradicional.
- **Otras herramientas y comandos:** variables globales, proyectos y sub-proyectos, organización de proyectos.

Programa y cronograma de clases de R

Clase 1:

- **Introducción y fundamentos de R:** Introducción, interface, usos y ventajas del lenguaje. Elementos fundamentales: valores, vectores, funciones y data frames. Tipo de datos, subconjuntos y manipulación. Uso de paquetes. Documentación. Organización de scripts.

- **Data frames y base de datos:** carga de base de datos, exploración y limpieza. Operadores lógicos y sentencias condicionales. Filtrado de filas y columnas.

Clase 2

- **Manipulación base de datos:** Crear, ordenar, renombrar, modificar y eliminar variables. Modificar múltiples variables a la vez. Tratamiento de strings.
- **Análisis de datos:** Agrupar base y estimación de medidas resumen, estadísticas descriptivas y tabulación de datos. Funciones propias para el análisis. Operaciones por fila.
- **Gestión de base de datos:** importación/Exportación de una base de datos. Unión y combinación. Comandos para reestructurar los datos.

Clase 3:

- **Análisis de regresión:** estimación por mínimos cuadrados ordinarios, interpretación de los resultados, comandos post-estimación, obtención de valores de coeficientes, y test de hipótesis.
- **Análisis gráfico.** Introducción a *ggplot*. Tipos de gráficos. Estructuración de datos y sintaxis gráfica. Elementos de un gráfico y opciones de personalización. Combinación de gráficos y facetado. Exportación y resolución de gráficos.
- **Mapeo:** Datos tipo shape. Mapeo de datos y escalas.

Clase 4:

- **Iteraciones y condicionales:** Loops y sentencias condicionales. Listas. Automatización de descriptivos. Gestión de data frames y eficientización de procesos.
- **Otras herramientas y usos:** Exploración de Rmarkdown. Automatización de reportes.
- **Datos no estructurados.** Introducción a web scrapping, lectura y manipulación de datos no estructurados

Bibliografía Stata

Baum, C. (2006). *An introduction to modern Econometrics using Stata*. Stata Press (capítulos 1, 2 y 4).

Hamilton, L. (2013). *Statistics with Stata (Updated for Version 12)*. Brooks/Cole-CENGAGE Learning.

Stata (2007). *User's Guide*. Stata Press.

Stata (2013). *STATA Base Reference Manual. Release 13*. Stata Press.

Bibliografía R

- Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc."
- Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2019). Introduction to Econometrics with R. University of Duisburg-Essen, 1-9.
- Bosco Mendoza Vega (2018). R para principiantes. Libro electrónico. <https://bookdown.org/jboscomendoza/r-principiantes4/>
- Tiscornia Pablo - Introducción al procesamiento de datos con R para Ciencias Sociales. Aplicación práctica en la Encuesta Permanente de Hogares (EPH-INDEC). <https://intro-r-eph.netlify.app/bio.html>
- Weksler, Kozlowski y Shokida. Curso de R para procesamiento de datos de la Encuesta Permanente de Hogares <https://diegokoz.github.io/Curso R EPH clases/>