Introducción al análisis de datos con el software Stata

José-Ignacio Antón

Universidad de Salamanca

janton@usal.es

Profesorado

José-Ignacio Antón es Profesor Titular de Universidad en el Departamento de Economía Aplicada de la Universidad de Salamanca (España). Sus áreas de investigación son Economía Laboral, Economía de la Salud y Economía Pública. Es usuario del software Stata desde hace más de 15 años.

Destinatarios

Personas interesados en el análisis de bases de datos, especialmente, aquellas de grandes dimensiones. Este software, disponible en la mayor parte de universidades, se emplea en Ciencias Sociales (Economía, Ciencias Políticas, Sociología), Ciencias de la Educación, Medicina, Bioestadística, etc. Junto con el paquete estadístico R, Stata es el programa más utilizado para el análisis de datos a nivel mundial. En este sentido, el curso puede ser especialmente interesante para que investigadores en fase formativa que pretendan mejorar su capacidad de análisis de datos, aquellos que deseen cambiar su paquete estadístico de trabajo habitual (por ejemplo, que tengan experiencia con otros programas como SPSS y deseen emplear un software más eficiente) o docentes o investigadores que deseen renovar sus conocimientos.

Asimismo, el curso puede ser apropiado para aquellos docentes que deseen fomentar el empleo de Stata para ilustrar aplicaciones de Estadística o incluso ejercicios prácticos en otras materias que involucren contenidos estadísticos relevantes.

Objetivos del curso

- ▷ Adquirir nociones básicas de manejo del software Stata.
- ▷ Desarrollar la capacidad de realizar operaciones de manipulación básicas de bases de datos.
- ▷ Desarrollar la capacidad de llevar a cabo análisis estadísticos y econométricos.
- ▷ Desarrollar la capacidad de elaborar gráficos de alta calidad aptos para publicaciones científicas.
- ▷ Desarrollar la capacidad de implementar flujos de trabajo integrando el programa Stata orientados a un uso eficiente del tiempo y garantizar la replicabilidad.

Metodología, contenidos y organización del curso

El curso cuenta con una duración de 10 horas presenciales. El curso tiene carácter *hands on*, es decir, su orientación es fundamentalmente práctica, centrada en el trabajo directo con datos estadísticos reales en equipos informáticos. En particular, los ejemplos y ejercicios emplean varias olas de la Encuesta de Condiciones de Vida de España y el Latinobarómetro. No obstante, se proporcionan explicaciones teóricas básicas e intuitivas sobre las técnicas utilizadas. Asimismo, se hace hincapié en las posibilidades extensión de las capacidades de Stata mediante el empleo de módulos creados por otros usuarios y disponibles en la red. Por último, se proporcionan rudimentos de programación en Stata para ampliar las capacidades potenciales de uso de los asistentes interesados. No se pretende profundizar en conceptos estadísticos o cubrir exahustivamente cada uno de los elementos que se abordan en el curso, sino que, por la duración del curso, se persigue mostrar las potencialidades del programa y orientar a los asistentes sobre los recursos que pueden emplear para convertirse en usuarios del programa en su actividad académica o profesional.

El curso se estructura en los siguientes cinco módulos:

1. Introducción a Stata
   1.1. ¿Qué es Stata?
   1.2. Objetivos del curso
   1.3. Conceptos básicos
   1.4. Primeros pasos en Stata
   1.5. La comunidad de usuarios
2. Manipulación de datos y archivos
   2.1. Operaciones de fusión horizontal de archivos
   2.2. Operaciones de fusión vertical de archivos
   2.3. Crear una base de datos con estadísticos descriptivos
   2.4. Cambiar la disposición de los datos
   2.5. Nociones básicas de programación: preserve, restore y el uso de macros y bucles
3. Tablas de contigencia y análisis multivariante
   3.1. Tablas de contingencia
   3.2. Tablas con estadísticos descriptivos
   3.3. Análisis de varianza
   3.4. Análisis de regresión
   3.5. Otros modelos econométricos de uso común
   3.6. Otras técnicas de análisis multivariante
4. Análisis gráfico
   4.1. Consejos preliminares
   4.2. Gráficos más frecuentes y potencialidades de los gráficos de ‘Stata‘
   4.3. Elaboración de mapas
5. Reproducibilidad, flujos de trabajo e integración con otro software
   5.1. Reproducibilidad
   5.2. Stata + Microsof Excel + Microsoft Word
   5.3. Stata + LaTeX
   5.4. Documentos dinámicos y otras posibilidades

Todos los materiales necesarios para seguir el curso (bibliografía, archivos utilizados y generados y sintaxis empleada por el profesor para generar los materiales), junto con más información sobre el mismo, pueden encontrarse en la siguiente *web*:

https://jianton.github.io/curso_stata.html

Fechas y horario de impartición

El curso se impartirá de forma sincrónica en la modalidad en línea a través de la herramienta Zoom en las siguientes fechas:

- Sesión 1: Lunes, 20 de septiembre de 2021, 18-20 h. (hora uruguaya, GMT-3).
- Sesión 2: Martes, 21 de de septiembre de 2021, 18-20 h. (hora uruguaya, GMT-3).
- Sesión 3: Miércoles, 22 de septiembre de 2021, 18-20 h. (hora uruguaya, GMT-3).
- Sesión 4: Jueves, 23 de septiembre de 2021, 18-20 h. (hora uruguaya, GMT-3).
- Sesión 5: Viernes, 24 de septiembre de 2021, 18-20 h. (hora uruguaya, GMT-3).

El curso se desarolla en la plataforma Zoom en el siguiente enlace La información de acceso es la siguiente:

https://usal-es.zoom.us/j/85251220432

Evaluación

Para superar el curso, los estudiantes deberán entregar la solución a una de las alternativas de ejercicios, basados en el material cubierto en las clases, propuestas por el profesor en la plataforma EVA. Los ejercicios deben entregarse a través del enlace en EVA proporcionando al profesor, como mínimo, un archivo `.do` y un archivo log (e.g., `smcl`), aunque otras opciones son admisibles (por ejemplo, los usuarios más avanzados pueden preferir archivos `.html` o `.pdf` generados con Markdown o TEX). Para superar el curso los estudiantes deben entregar el 70 % de los ejercicios de la alternativa escogida adecuadamente resueltos.

Bibliografía recomendada

En cada uno de los módulos, se sugiere bibliografía específica para el estudiante. No obstante, con carácter general, cabe señalar los siguientes recursos:

Acock, A. C. (2018). *A gentle introduction to Stata* (6ª ed.). Stata Press.

Baum, C. F. (2006). *An introduction to modern Econometrics using Stata*. Stata Press.

Baum, C. F. (2016). *An introduction to Stata programming* (2ª ed.). Stata Press.

Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics using Stata* (2ª ed.). Stata Press.

Escobar, M., Fernández-Macías, E. & Bernardi, F. (2012). *Análisis de datos con Stata* (2ª ed.). Centro de Investigaciones Sociológicas.

Long, J. S. (2009). *The workflow of data analysis using Stata*. Stata Press.