

Solución de problemas por medio de datos: una introducción práctica

Duración: 20h (5 sesiones de 4h)

Fechas: todos los viernes de 8 de febrero a 8 de marzo de 2019.

Horario: de 15 a 19h

Lugar: Aula de informática 2 – Planta Jardín – Facultad de Derecho

INSCRIPCIÓN

La matrícula es gratuita para todos los participantes del Programa de Iniciación Científica (PIC) del Área de Ciencia Política y de la Administración de la USAL. Los demás interesados deben inscribirse en la página de internet de Cursos Extraordinarios. (http://vaporetto.usal.es/preactform/detalleCurso?codigo_curso=4720&ano_academico=2019&convocatoria=1)

Inscripción: 150€ comunidad USAL (estudiantes y alumni)

Interesados de fuera: 250€.

DESCRIPCIÓN

Una breve inspección a los diferentes manuales de análisis de datos y de introducción a paquetes estadísticos como el SPSS, el Stata o el R revela que, en su mayor parte, estas guías se concentran en presentar conceptos estadísticos que van desde medidas descriptivas hasta modelos de regresión más o menos complejos. Este tipo de acercamiento al análisis de datos considera que el usuario o lector:

- (a) ya posee un conocimiento básico de estadística;
- (b) existe cierta familiaridad con programas estadísticos y que éstos últimos corresponden a medios de implantar conceptos y técnicas estadísticas que pueden ser utilizadas para el análisis de datos;
- (c) no hace falta orientación adicional sobre qué técnicas utilizar en qué situación y objeto de estudio; y, principalmente,
- (d) los lectores/usuarios/estudiantes saben qué son datos, qué tipos de datos existen y qué técnicas deben ser aplicadas para el ejercicio profesional en la actualidad.

Este enfoque técnico, formalista, y muy poco práctico, es quizás el principal obstáculo para un aprendizaje adecuado del análisis de datos (como la enseñanza de ecuaciones y notación teórica lo es para la enseñanza de la matemática). Su enfoque no se encuentra en el análisis de los datos en sí mismos, sino en enseñar técnicas estadísticas “de manual” y cómo implantarlas en algún paquete estadístico como el SPSS, el Stata o el R. Aún en los casos en los que se utiliza un problema como guía para la enseñanza, este problema se elige para ilustrar un concepto o técnica estadística. Raros son los ejemplos de materiales hechos con el objetivo explícito de desarrollar competencias que lleven a un individuo formular una pregunta y ser capaz de contestarla por medio de una serie de pasos lógicos a que llamamos de análisis de datos.

De ese modo, hace falta un acercamiento que ponga énfasis en el análisis de datos propiamente dicho y que utilice las técnicas como instrumentos que pueden ayudar en ciertos casos (y molestar en otros). El enfoque debe ser desplazado, por lo tanto, de la enseñanza de herramientas (sea la fórmula de la media o de la desviación estándar o hasta mismo de la noción de medida de tendencia central, dispersión o asociación) a uno

centrado en la solución de problemas (cómo entiendo el comportamiento y la estructura de un fenómeno, cómo este fenómeno reacciona frente a otros, cuáles son los posibles factores que influyen en su desarrollo).

El presente curso pretende ofrecer ese enfoque alternativo al tradicional, haciendo hincapié en los conocimientos necesarios para llevar a cabo un análisis de datos efectivo que, en términos concretos, significa:

- (a) el desarrollo de una inquietud intelectual genuina;
- (b) la formulación de cuestiones e hipótesis de trabajo;
- (c) la elección de métodos de análisis adecuados para el problema a ser analizado;
- (d) el ejercicio de creatividad para crear nuevos métodos o técnicas;
- (e) la elección de los instrumentos adecuados para el tipo de dato trabajado; y
- (f) conocer los límites y las potencialidades de cada técnica de análisis adoptada.

Se espera que los estudiantes adquieran las habilidades para utilizar datos y seleccionar las técnicas de análisis más adecuadas para cada problema a que tengan que enfrentarse. Sin embargo, resulta aún más importante que sean capaces de desarrollar insights, preguntas de investigación e hipótesis originales, que sirvan de forma práctica en su desarrollo académico y profesional posterior.

Como competencias específicas se espera que sean capaces de:

- (a) buscar datos para sus investigaciones;
- (b) tratarlos y limpiarlos para evitar problemas e interferencias no deseadas;
- (c) realizar distintos análisis con el objetivo de generar preguntas de investigación e hipótesis;
- (d) utilizar estadísticas, gráficos y otros recursos analíticos para explorar los datos y comunicar resultados obtenidos;
- (e) definir y testar relaciones causales.

Como se puede verificar, los objetivos son los mismos. No obstante, el método difiere a los enfoques tradicionales en dos sentidos:

- 1) por centrar la enseñanza en la solución de problemas sustantivos.
- 2) se utilizan ejemplos reales, sencillos y cercanos al dominio de conocimiento de los estudiantes de las ciencias sociales.

Este curso, de 20 horas de duración, está dividido en cinco temas de cuatro horas cada uno. Sin embargo, para un aprendizaje adecuado, se recomienda que los estudiantes realicen las lecturas recomendadas en la bibliografía para cada tema y lleven a cabo los

ejercicios propuestos. Estas actividades ayudarán a que consoliden los conocimientos adquiridos en clase y facilitarán un aprendizaje más denso y completo de los contenidos y propuestas cubiertos en este curso.

CONTENIDO

Sesión 1. La práctica del análisis de datos

- a. La clave está en saber formular una buena pregunta.
- b. ¿Cómo el análisis de datos puede constituirse como una ventaja profesional?
- c. ¿Qué son datos? ¿Qué tipos existen? ¿Por qué importan?
- d. Práctica: identificación de distintos tipos de dato.
- e. La labor de analizar: identificar patrones sin perder la atención al detalle.
- f. Práctica: ¿qué podemos hacer con una lista de nombres de diputados?

Sesión 2. Encontrar y preparar los datos

- a. Dime que fuentes de información conoces y te diré qué puedes investigar.
- b. ¿Dónde encontrar datos? Principales fuentes de información para el análisis social en España, Europa y repositorios internacionales.
- c. Práctica: Paro, resultados electorales y PIB en las provincias españolas.
- d. El preparo y la limpieza de los datos: ¿por qué es necesario y cómo hacerlo?
- e. El cálculo de los indicadores sociales más utilizados
- f. Práctica: Limpieza de los datos y cálculo de indicadores

Sesión 3. Descubrir lo nuevo por el análisis de patrones y anomalías

- a. ¿Qué podemos entender por patrones y anomalías cuando analizamos datos?
- b. Utilizando preguntas sencillas para orientar el análisis de los datos y formular hipótesis.
- c. Práctica: Identificación de patrones y anomalías en un conjunto de datos
- d. El Análisis Exploratorio de Datos: recetas vs saber práctico
- e. Descripción y exploración: entender un atributo (una sola variable) o la relación entre atributos de un fenómeno (dos o más variables)
- f. El rol de los gráficos y de la visualización de datos de forma general
- g. Práctica: Análisis exploratorio de la relación entre riqueza, paro y comportamiento electoral

Sesión 4. Clasificar, asociar, comparar

- a. Tareas fundamentales del análisis de datos: clasificar, asociar, dividir
- b. La clasificación como esfuerzo de síntesis analítica
- c. Práctica: el análisis de conglomerados y la clasificación de una variable continua
- d. La asociación como proximidad, vínculo o jerarquía
- e. Práctica: medidas de asociación, árboles y redes
- f. ¿Quién ha sido mejor votado? La comparación a partir de criterios objetivos
- g. Práctica: teste de hipótesis y diferencia entre medias

Sesión 5. Explicar los resultados

- a. “Adivinando” el futuro y explicando el pasado: la predicción y la inferencia
- b. ¿Por qué asociación no es causalidad?
- c. Práctica: Modelos predictivos y correlaciones espurias
- d. Diferentes formas de llevar a cabo la inferencia causal
- e. Simplificar y generalizar: los modelos de regresión
- f. Modelos: cuándo utilizarlos y cuál el más adecuado
- g. Práctica: formulando un modelo de regresión lineal múltiple

BIBLIOGRAFÍA

La bibliografía listada abajo consiste en material de referencia y consulta. Su lectura no resulta obligatoria ni implica un requisito para el seguimiento del curso. No obstante, sirve como apoyo para todos que estén interesados en profundizar y saber más sobre el tema de cada sesión en concreto, por ello son de lectura muy recomendada.

Lecturas específicas

Sesión 1.

Bergin, Tiffany. 2018. “Introducing Data” y “Thinking Like a Data Analyst”. En: *An Introduction to Data Analysis: Quantitative, Qualitative, and Mix-Methods*. Londres: Sage. Pp. 1-14, 15-32.

Kabacoff, Robert I. 2011. “Introduction to R”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 3-20.

Munzner, Tamara. 2014. “What: Data Abstraction” y “Why: Task Abstraction”. En: *Visualization Analysis and Design*. Har/Psc edition. Boca Raton: A K Peters/CRC Press. pp.20-40 y 42-65.

Tufte, Edward. 1974. “Introduction to Data Analysis”. En: *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice-Hall. Pp. 1-30.

Yau, Nathan. 2013. “Understanding data”. *Data Points: Visualization That Means Something*. Edición: 1. Indianapolis, IN: Wiley John + Sons. Pp. 1-41.

Sesión 2.

Bergin, Tiffany. 2018. “Finding, Collecting, and Organizing Data”. En: *An Introduction to Data Analysis: Quantitative, Qualitative, and Mix-Methods*. Londres: Sage. Pp. 33-64.

Kabacoff, Robert I. 2011. “Creating a Dataset”, “Basic Data Management”, “Advanced Data Management”, “Advanced Methods for Missing Data”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 21-45; 73-90; 91-116; 352-371.

Sesión 3.

- Bergin, Tiffany. 2018. "Introducing Quantitative Data Analysis". En: *An Introduction to Data Analysis: Quantitative, Qualitative, and Mix-Methods*. Londres: Sage. Pp. 65-88.
- Few, Stephen. 2012. "Simple Statistics to Get You Started". En: *Show Me the Numbers. Designing Tables and Graphs to Enlighten*. El Dorado Hills, CA: Analytic Press. Pp. 15-38.
- Vaus, David de. 2002. "Which Summary Statistics to Use to Describe a Single Variable", "How to Interpret a Correlation Coefficient" y "Which Correlation?" En: *Analyzing Social Science Data: 50 Key Problems in Data Analysis*. Londres: Sage. Pp. 218-227, 267-273, 274-278.

Sesión 4.

- Bergin, Tiffany. 2018. "Applying Quantitative Data Analysis: Correlations, t-Tests, and Chi-Square Tests". En: *An Introduction to Data Analysis: Quantitative, Qualitative, and Mix-Methods*. Londres: Sage. Pp. 89-128.
- Kabacoff, Robert I. 2011. "Basic Statistics", "Analysis of Variance", "Principal Component and Factor Analysis". En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 141-170; 219-245; 331-350.

Sesión 5.

- Kabacoff, Robert I. 2011. "Regression", "Generalized Linear Models". En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 171-218; 313-330.
- Tufte, Edward. 1974. "Predictions and Projections: Some Issues of Research Design", "Two-Variable Linear Regression" y "Multiple Regression". En: *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice-Hall. Pp. 31-65, 65-134, 135-164.
- Vaus, David de. 2002. "How to Do a Multiple Regression", "How to Use Non-Interval Variables in Multiple Regression" y "What does the Multiple Regression Output Mean?" En: *Analyzing Social Science Data: 50 Key Problems in Data Analysis*. Londres: Sage. Pp. 353-367, 368-373, 374-382.

Manuales de estadística y análisis de datos

- Levine, David, and David Stephan. 2009. *Even You Can Learn Statistics: A Guide for Everyone Who Has Ever Been Afraid of Statistics*. Upper Saddle River, N.J: FT Press.
- Saskind, Neil. 2017. *Statistics for People Who (Think They) Hate Statistics*. Londres: Sage.