

## Universidad Nacional de Colombia

### Facultad de Ciencias Económicas

### Ciencia de datos en R

### Taller de aplicación

**Docente:** Felipe Calvo Cepeda  
fcalvoc@unal.edu.co

## 1. Objetivo del taller

Poner en practica los conceptos y métodos estadísticos vistos para el análisis de datos, a partir del manejo del software R y sus paquetes estadísticos.

## 2. Contenido

- 1 Construya una base de datos en una hoja de cálculo (Excel o Google Sheets) con al menos 5 variables de su interés profesional o académico, sujeto a las siguientes necesidades:
  - a) Al menos dos variables cuantitativas.
  - b) Al menos dos variables ordinales.
  - c) Al menos una variable nominal.
  - d) Al menos 20 individuos u observaciones.
  - e) Destine una hoja exclusivamente para los datos, nombrando las columnas (nombres de las variables) bajo alguna de las convenciones de nombre vistas en clase.
  - f) Destine otra hoja exclusivamente para construir su propio *diccionario de datos*, en el cual para cada variable muestre su nombre codificado, su nombre semántico, una breve descripción, si es de tipo cuantitativo o cualitativo y en qué escala de medición está.

Si crea su base de datos en Excel adjúntela al momento de enviar su taller. Si la crea en Google Sheets, adjunte en su taller el enlace para la consulta y asegúrese de habilitar la opción de acceder a través del enlace.

- 2 Con la base de datos realice un proceso de limpieza de datos y de detección de datos atípicos.
- 3 Realice un análisis descriptivo de las variables cuantitativas de su base, incluyendo las estadísticas, tablas y gráficas que considere relevantes, y para cada una de ellas incluya un comentario o breve análisis.
- 4 Construya una tabla dinámica a partir de sus datos. Con base en la tabla:
  - a) Construya una gráfica que relacione dos de las variables que cruzó.

- b)* Plantee por escrito dos preguntas que la base de datos podría responder.
  - c)* Plantee por escrito una hipótesis para cada pregunta.
- 5 Teniendo en cuenta los temas vistos en el módulo, construya un diagrama de flujo para realizar un análisis de datos. Consulte como referencia la Guía de Lucidchart.

### 3. Entrega

Hacen parte integral del desarrollo del taller los siguientes archivos:

1. Un documento HTML con la documentación y desarrollo de los puntos 2, 3 y 4.
2. Una base de datos, tal como se especifica en el punto 1.

**Medio de envío:** correo electrónico, en un archivo comprimido.

### 4. Bibliografía complementaria

- Página web del curso
- Ott, R. L., & Longnecker, M. T. (2015). An introduction to statistical methods and data analysis. Nelson Education.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2014). Probabilidad y estadística para ingeniería y ciencias. Norma, 162, 157. Novena edición