
TEMARIO

Introducción a R

Descripción

El uso de software adecuado para realizar análisis es fundamental en Ciencia de datos. Sin duda, uno de los programas más usados para el análisis de datos es R. Este lenguaje tiene la ventaja de ser de acceso libre y gratuito, así como abierto a que las personas generen sus propios paquetes para realizar el análisis que ellos deseen. Esto ha hecho que la cantidad de librerías y posibles análisis que pueden hacerse usándolo sea prácticamente ilimitado. Podemos encontrar desde paquetes asociados a técnicas clásicas de análisis hasta aquellos relacionados con los últimos artículos de investigación, o desde paquetes propios de análisis en demografía, econometría o bioestadística hasta aquellos relacionados con las más modernas técnicas en Machine Learning.

El objetivo de este curso es proporcionar una introducción básica a R, desde su instalación, manejo y lectura de distintos tipos de datos hasta la elaboración de gráficas básicas. Su objetivo no es que el estudiante se vuelva un maestro en R, sino la de servir como una introducción para empezar a adquirir confianza en su uso. Así, una vez que se quieran hacer análisis específicos, se pueden entender los fundamentos con los cuales entender la estructura general del programa, para que entonces la estructura particular asociada a un paquete de interés sea menos difícil de entender.

Duración: Curso de aproximadamente 8 horas.

1. Introducción al uso de R

- a) R y el editor auxiliar RStudio.
- b) Introducción: objetos, ayuda en R, instalación de librerías y aritmética simple.
- c) Vectores, sucesiones y matrices en R: funciones asociadas básicas, operadores lógicos y operaciones con matrices y vectores
- d) Funciones y descripción breve de R como lenguaje de programación
- e) Ejemplos de algunas gráficas simples con sus argumentos
- f) Lectura de datos: ejemplos de lectura con varias extensiones (.csv, .txt, .sav)
- g) Manejo básico de datos: filtros, ponderación de acuerdo a frecuencias, generación de nuevas variables a través de operaciones básicas, etiquetar variables categóricas, bases con valores perdidos (NA) y categorización de variables continuas.

2. Práctica simple de los conceptos a través de análisis descriptivos básicos en algunas bases de datos.

Informes: lumialearning@gmail.com