

Trabajo de Estadística con R

Objetivo

El objetivo de este trabajo es poner en práctica los conocimientos y destrezas con R adquiridas en las prácticas de la asignatura de Probabilidad y Estadística, así como comprobar si se dominan los principales conceptos de Estadística vistos en el curso.

Descripción

Cada alumno tendrá que realizar un proyecto de análisis de datos a partir de una base de datos elegida por él mismo. La base de datos debe contener suficiente información y tener la complejidad suficiente para poder poner en práctica la mayoría de lo aprendido durante el curso, por lo que debe cumplir los siguientes requisitos:

- Un mínimo de 6 variables (columnas), que incluyesen tanto variables cuantitativas como cualitativas (al menos 2).
- Un mínimo de 500 registros (filas). Antes de empezar a trabajar con las bases de datos, estas deben ser validadas tanto por los profesores de

A continuación se proponen algunas fuentes de datos para facilitar la búsqueda de una base de datos con las características pedidas:

- (<https://data.world/datasets/marketing>)
- <https://www.kaggle.com/datasets>
- <https://datasetsearch.research.google.com/>
- <https://www.cmswire.com/digital-marketing/6-datasets-for-marketers-should-know-about/>
- <https://www.census.gov/data.html>
- <https://www.census.gov/acs/www/data/data-tables-and-tools/>
- https://catalog.data.gov/dataset?res__format=CSV

Antes de empezar el trabajo, el alumno debe comunicarle al profesor la base de datos que quiere usar y un breve resumen con los objetivos de su análisis de datos para que el profesor los valide.

Requisitos mínimos

Para conseguir el aprobado en el trabajo, el análisis de datos debe incorporar al menos lo siguiente:

- Crear un DataFrame a partir de un fichero csv o excel con el índice, separador de campos y separador de decimales adecuado.
- Preprocesar el fichero para seleccionar las columnas necesarias y eliminar las filas con valores desconocidos.
- Generar nuevas columnas a partir de la aplicación de fórmulas a las columnas existentes (tanto numéricas, como cadena y booleanas).
- Generar nuevas columnas a partir de la recodificación de otras.
- Aplicar filtros.
- Calcular tablas de frecuencias simples y agrupadas según una o varias variables cualitativas.
- Calcular los principales estadísticos de las variables cualitativas y cuantitativas, simples y agrupando según una o varias variables cualitativas, e interpretarlos.
- Realizar diagramas de barras o líneas simples y agrupando según una variable cualitativa e interpretarlos.
- Realizar histogramas e interpretarlos.
- Realizar diagramas de cajas simples y agrupando según una variable cualitativa e interpretarlos.
- Poner título a los gráficos, a los ejes y añadir una leyenda cuando sea necesaria.
- Construir modelos de regresión lineales simples y agrupando según una o varias variables cualitativas, e interpretarlos.
- Construir modelos de regresión no lineales simples y agrupando según una o varias variables cualitativas, e interpretarlos.
- Comparar distintos modelos de regresión.
- Hacer predicciones con un modelo de regresión.
- Realizar un informe en RMarkdown con todo lo anterior.

Requisitos extra

Los siguientes requisitos no son necesarios para aprobar el trabajo pero subirán su nota.

- Pivotar el Data Frame de formato ancho a largo o viceversa.
- Realizar diagramas (barras, histogramas, diagramas de cajas o sectores) agrupando con facetas.
- Personalizar el aspecto de los gráficos con colores, tipos de marcadores y tipos de líneas.
- Comparar la distribución observada de una variable cuantitativa discreta con su distribución teórica a partir de sus percentiles.

- Comparar la distribución observada de una variable cuantitativa continua con su distribución teórica a partir de sus percentiles.
- Programar funciones para automatizar los análisis de datos.
- Simular un experimento aleatorio mediante la generación de números aleatorios a partir de una distribución discreta o continua y describir la variable resultante.

Plazo de entrega

22 de mayo a las 23:59.