

Práctica 1 Estadística. Curso 2015-2016

Esta primera práctica tiene como objetivo repasar y consolidar los conceptos vistos en la asignatura de estadística. Para ello se propone realizar los siguientes apartados en grupos de **máximo dos personas**.

Todos los ejercicios se deben realizar utilizando el software estadístico Open Source R (<http://www.r-project.org/>). Se sugiere el uso de algún entorno de desarrollo como RStudio (<https://www.rstudio.com>).

Estadística Descriptiva y Ajuste Mínimos Cuadrados

Uno de los objetivos principales de la estadística es entender y analizar un conjunto de datos. La estadística descriptiva consiste en ordenar y representar los datos de forma que nos puedan sugerir modelos o estructuras que los expliquen.

En este apartado vamos a utilizar una pequeña variante del dataset “adult” [1].

Este dataset contiene información basada en el censo de USA del año 1996 que indica si una persona gana más de 50K USD al año en función de una serie de variables.

Asumiendo que el fichero está almacenado en el directorio RUTA_COMPLETA. El dataset “adult.prac1.csv” se puede cargar en R con el siguiente comando:

```
adult <- read.csv("/RUTA_COMPLETA/adult.prac1.csv")
```

A continuación se puede hacer referencia al data frame que contiene los datos con el nombre **adult**. A modo de ejemplo las siguientes instrucciones nos proporcionan el número de filas, número de columnas, el valor del campo [1,1] y del campo [150, 5] de este conjunto de datos.

```
nrow(adult)
ncol(adult)
adult[1,1]
adult[150,5]
```

Recuerde que **head(adult)** y **tail(adult)** le muestran los primeros 6 y los últimos 6 registros de este conjunto de datos respectivamente.

El conjunto de datos contiene las siguientes variables:

- Age: numérica
- workclass: Tipo de trabajo. Categórica.
- education: Nivel de estudios. Categórica.
- education.num: Equivale a la anterior. Numérica.
- marital.status: Categórica
- occupation: Categórica
- relationship: Categórica. *Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.*
- race
- sex
- capital.gain: Ganancias de capital. Numérica.
- capital.loss: Perdidas de Capital. Numérica.
- hours.per.week: Horas de trabajo por semana. Numérica.
- native.country: Categórica.
- Class: Si gana más de 50K al año o no.

1. Genere las gráficas necesarias para visualizar las diferencias de ingresos en función de las demás variables, tanto categóricas como numéricas. (3.5 punto)
2. Obtenga las medidas de tendencia central de cada una de las variables numéricas para cada una de las dos posibles clases. ¿Existe diferencia? (1 punto)
3. Obtenga los cuartiles y los percentiles de cada una de las variables numéricas para cada una de las dos posibles clases. ¿Existe diferencia? (1 punto)
4. Para las variables numéricas y para cada una de las clases, calcule las medidas de dispersión siguientes: (a) recorrido, (b) recorrido inter-cuantílico y (c) varianza. ¿Existe diferencia? (1 punto)
5. Con respecto a las medidas de asimetría, ¿qué podríamos indicar de la distribución de cada una de las variables numéricas de las dos clases? (1 punto)
7. Realice un ajuste de mínimos cuadrados entre la variable horas por semana y nivel de educación. Indique la precisión del ajuste. (2.5 punto)

[1] <https://archive.ics.uci.edu/ml/datasets/Adult>

Normativa y entrega de la práctica

- La práctica se realizará y entregará por grupos compuestos de 2 personas.
- La práctica se realizará utilizando el lenguaje de programación R.
- La entrega deberá realizarse mediante Blackboard antes del **17 de Abril de 2016** a las 23:59 horas (hora peninsular en España).
- La entrega se compondrá de un fichero de texto con extensión .R que permita ejecutar la práctica en cualquier interprete de lenguaje R. Las explicaciones y comentarios necesarios se pondrán en dentro del fichero con comentarios (#).
- El nombre del fichero tendrá el nombre y el primer apellido de cada miembro del grupo separados por “_” sin acentos. Ejemplo: pepe_garcia_luis_sanchez.R
- La corrección de la práctica se hará ejecutando los comandos del fichero.
- Cualquier sospecha y verificación de copia entre dos o más prácticas derivará en la calificación de 0 para todos los alumnos involucrados.