PRÁCTICA 4: Distribuciones unidimensionales (2)

ACCESO A CONJUNTOS DE DATOS DISPONIBLES EN R

Tanto en las librerías de la instalación básica de R como en las librerías que hemos ido instalando, contienen algún conjunto de datos. En particular, la librería **HSAUR3**, que contiene los datos de la 3 edición del libro "A Handbook of Statistical Analyses Using R" y la librería **datasets.load** que constituye un interfaz para cargar conjunto de datos existentes en R.

La sentencia

data()

nos muestra en la primera ventana de RStudio los ficheros de datos disponibles en nuestro ordenador, teóricamente los disponibles con las librerías de la instalación básica.
De la lista mostrada, si queremos cargar en memoria los datos del data.frame titulado "iris", ponemos la sentencia

data(iris)

Si queremos ver la estructura de dicho conjunto de datos escribimos el comando

str(iris)

Podemos cargar la librería HSAUR3 y ver sus ficheros de datos disponibles con las instrucciones siguientes

```
library(HSAUR3)
ls("package:HSAUR3")
```

También podemos ver **todos** los datos disponibles en R, a través de la librería **datasets.load**

```
library(datasets.load)
datasets.load()
```

cuyo efecto es abrir una ventana emergente donde nos indica el nombre del fichero, su título y su librería asociada. Pulsando sobre el fichero elegido lo cargaremos en memoria.

También existen otros comandos, véase el fichero de instrucciones, que además nos indica en que subdirectorio se encuentra dicho fichero de datos.

DESARROLLO DE LA PRÁCTICA

Una vez visto la posibilidad de acceder a los conjuntos de datos (data frame) que se encuentran disponibles con las librerías instaladas. En esta práctica vamos a en realizar la práctica anterior, y posiblemente algunos comandos de la práctica número dos, pero aplicada a otros conjuntos de datos.
Para los conjuntos de datos que se indican, se pueden importar desde txt o csv, si se buscan en internet, o cargar desde R. Posiblemente sea necesario recodificar alguna variable.
Una vez importado dicho conjunto de datos, se deben obtener tablas estadísticas para diferentes tipos de variables (cualitativa (nominal u ordinal) y cuantitativa (discreta o continua)), sus estadísticos descriptivos y representaciones gráficas. Lo más interesante, será aportar información-conclusiones sobre nuestros datos a partir de la información que obtengamos.

Primer conjunto de datos
DATA SET de la librería ISLR: **Auto**

Description
Gas mileage, horsepower, and other information for 392 vehicles.
Format
A data frame with 392 observations on the following 9 variables.
mpg: miles per gallon
cylinders: Number of cylinders between 4 and 8
displacement: Engine displacement (cu. inches)
horsepower: Engine horsepower
weight: Vehicle weight (lbs.)
acceleration: Time to accelerate from 0 to 60 mph (sec.)
year: Model year (modulo 100)
origin: Origin of car (1. American, 2. European, 3. Japanese)
name: Vehicle name

The original data contained 408 observations but 16 observations with missing values were removed.

Para este conjunto de datos, se pueden cambiar de unidades americanas a unidades del sistema métrico, las variables: mpg, displacement, weight y acceleration

(1 Milla por galón= 0.425144 km por litro, 1 libra= 0.453592 kg y 1 pulgada cúbica = 16.39 centímetros cúbicos)

Segundo conjunto de datos
DATA SET de la librería ggplot2: **diamonds**

Description
A dataset containing the prices and other attributes of almost 54,000 diamonds.
Format
A data frame with 53940 rows and 10 variables:
price: price in US dollars (\$326--\$18,823)
carat: weight of the diamond (0.2--5.01)
cut: quality of the cut (Fair, Good, Very Good, Premium, Ideal)
color: diamond colour, from J (worst) to D (best)
clarity: a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
x: length in mm (0--10.74)

.

---

y: width in mm (0--58.9)
z: depth in mm (0--31.8)
depth: total depth percentage = z / mean(x, y) = 2 * z / (x + y) (43--79)
table: width of top of diamond relative to widest point (43--95)

Tercer conjunto de datos
DATA SET de la librería HSAUR: **skulls**

Description
Measurements made on Egyptian skulls from five epochs.
Format
A data frame with 150 observations on the following 5 variables.
epoch: the epoch the skull as assigned to, a factor with levels c4000BC, c3300BC, c1850BC, c200BC, and cAD150, where the years are only given approximately, of course.
mb: maximum breaths of the skull.
bh: basibregmatic heights of the skull.
bl: basialiveolar length of the skull.
nh: nasal heights of the skull.

Details
The question is whether the measurements change over time. Non-constant measurements of the skulls over time would indicate interbreeding with immigrant populations.

.