

ESTADÍSTICA
Grado en Ingeniería Informática

Práctica 1: INTRODUCCIÓN AL USO DE R COMMANDER

Objetivos:

- Familiarizarse con el paquete *R Commander* del programa *R*.
- Crear, modificar y editar ficheros de datos.
- Recodificar un conjunto de datos.
- Crear variables.

Desarrollo de la práctica:

Ejemplo 1

Los siguientes datos se refieren al tiempo hasta el fallo (en miles de horas de trabajo) de 10 placas base Socket AM3 idénticas, junto con la causa del mismo: rotura de un condensador (R) o cortocircuito (C).

Tiempo	Causa
44	C
50	R
52	R
43	C
67	R
80	R
66	R
29	C
49	C
58	R

Introduzca los datos creando un fichero denominado *Placas.RData*.

Ejemplo 2

Los datos que aparecen en el fichero *DatosAndalucia2013.txt*, procedentes de la página web del *Instituto de Estadística y Cartografía de Andalucía*¹, corresponden a variables de carácter social y económico medidas sobre los municipios andaluces en el año 2013. Concretamente, sobre cada municipio se miden las siguientes variables:

¹<https://www.juntadeandalucia.es/institutodeestadisticaycartografia>

- *Código INE.*
- *Municipio.*
- *Población total del municipio.*
- *Nº medio de parados registrados.*
- *Rentas netas declaradas (en euros)*, definida como la suma de las rentas netas declaradas según el tipo de rendimiento: rentas netas del trabajo, rentas netas de actividades empresariales, rentas netas de actividades profesionales, otro tipo de rentas netas.
- *Nº de líneas ADSL en servicio.*
- *Ratio de hogares cubiertos por ADSL de ≥ 10 Mbps, fibra, redes inalámbricas de ≥ 30 Mbps y LTE 4G*

Se pide:

- (a) A partir de la variable *Código INE*, construya una variable tipo factor que distinga la provincia de pertenencia de cada municipio. A dicha variable tipo factor la llamaremos *Provincia*. La crearemos recodificando la variable *Código INE*, sabiendo que:

<i>Código INE</i>	<i>Provincia</i>
4000 – 4999	Almería
11000 – 11999	Cádiz
14000 – 14999	Córdoba
18000 – 18999	Granada
21000 – 21999	Huelva
23000 – 23999	Jaén
29000 – 29999	Málaga
41000 – 41999	Sevilla

- (b) La variable *Nº de líneas ADSL por municipio* está claramente afectada por la población del mismo. Para analizar la presencia de líneas *ADSL* en términos relativos a nivel municipal, construya una *tasa de líneas por cada 100 habitantes* de la siguiente forma:

$$\text{Tasa de líneas ADSL} = \frac{\text{Nº de líneas ADSL}}{\text{Población del municipio}} \times 100.$$

- (c) Construya una variable que represente las *rentas netas declaradas por habitante* de la siguiente forma:

$$\text{Renta neta por habitante} = \frac{\text{Renta netas declaradas}}{\text{Población del municipio}}.$$

Seguidamente, cree otra variable (*Renta*) tipo factor, recodificando la nueva variable, *renta neta por habitante*, según la siguiente clasificación:

<i>Rentas netas por habitante</i>	<i>Renta</i>
0 – 1199.99	Hasta 1.200
1200 – 3299.99	Entre 1.200 y 3.300
3300 – 5299.99	Entre 3.300 y 5.300
5300 – 7199.99	Entre 5.300 y 7.200
7200 – 9299.99	Entre 7.200 y 9.300
9300 – 11099.99	Entre 9.300 y 11.100
11100 – 13699.99	Entre 11.100 y 13.700
13700 – 9999999999	Más de 13.700