

## **Bases de datos:**

### **numberofwords.csv**

La mayoría de los bebés dicen sus primeras palabras a los 15 meses de nacer.

Los siguientes datos proceden de un estudio, en el que se pidió a los padres que completaran una lista de verificación con varios cientos de palabras. Para cada palabra, los padres tenían que marcar si su hijo/a ya había comenzado a decirla. Se contó el número total de palabras pronunciadas por cada niño/a, variable “words”, y de estas cuántas son sustantivos “nouns” y cuántos verbos “verbs”. También se preguntó a los padres a qué edad el niño/a pronunció la primera palabra “aget1w”. Además, la base de datos dispone de la edad (meses) de los niños “age”, género “gender” y tipo de hábitat (rural/urbano) “región”, en el momento del estudio.

### **Tobacco.csv**

Datos de la encuesta de presupuestos de belga de 1995–1996. La muestra contiene 2,724 hogares para los que se observan los gastos en una amplia gama de bienes, así como una serie de variables de fondo relacionadas con la composición familiar y el estatus ocupacional.

occupation: activo (bluecol/whitecol) inactivo o autónomo (inactself)

region: flanders, wallon, brussels

nkids: número de niños mayores de 2 años

nkids2: número de niños de 2 o menos años

nadults: número de adultos

lnx: logaritmo natural de los gastos totales del hogar

stobacco: porcentaje del presupuesto dedicado a comprar tabaco

salcohol: porcentaje del presupuesto dedicado a comprar alcohol

age: edad del entrevistado en intervalos: menor de 30 años, 30-40, 40-50, 50-60, 60 y más años

### **SmokingD.csv**

Datos simulados N=1,000 observaciones sobre consumo de tabaco y estado de salud.

gender: M (male) F (female)

age. edad continua

age.gr: edad agrupada en 4 grupos

BMI: índice de masa corporal

smoker: fumador Yes, no fumador No

cigs.per.day: cigarrillos al día

diseased: padece alguna enfermedad Yes/No

disease: tipo de enfermedad

samp.wgts: ponderador o factor de elevación individual

#### Medidas.xlsx

El fichero contiene un conjunto de medidas corporales de un grupo de estudiantes de la ULPGC:

edad: edad del estudiante en años

peso: peso del estudiante en kilogramos

altura: estatura del estudiante en centímetros

sexo: género del estudiante

muñeca: diámetro de la muñeca derecha en centímetros

biceps: diámetro del bíceps derecho en centímetros

#### heartrate.dta

Fichero de datos de Stata. Contiene dos medidas de los latidos de corazón (nº de latidos por minuto) en una muestra de 100 pacientes. La variable sex está codificada como 1 Hombre, 2 Mujer.

#### healthdisparities.dta

Fichero de datos de Stata. Variables:

gender: 1 male; 2 female

race: 1 latino; 2 pacific islander; 3 american Indian; 4 asian; 5 african American; 6 white; 7 otros

age: edad en años

poverty: pobreza 1 sí; 0 No

doctor: visitó al médico los últimos 12 meses, 1 sí; 0 No

racecat: raza recategorizada, 1 White, non-Hispanic; 2 Hispanic; 3 All other races

urban: habitat urbano 1, rural 2

#### calcium.csv

Peso de ratones/cobayas. Fichero de datos separados por comas.

Variables:

dose: dosis de fármaco administrada (mg)

sex: 0 male, 1 female

weight: peso en gramos

#### hiaa.dta

Prueba AR-HIA en orina. Fichero de datos de Stata.

Variables:

hiaa: excreción urinaria de 5-HIAA ( $\mu$  mol/l)

sero: nivel plasmático de serotonina ( $\mu$  mol/l)

sex: 1 male, 0 female

age: edad en años

#### physact.dta

phyact: índice de actividad física

sex: 1 male, 0 female

region: rural 0; urban 1

age: edad en años

#### allergy1.dta

childnr: edad del niño en días

allergyc: 1 alergia en el niño; 0 sin alergia

allergym: 1 alergia en la madre; 0 sin alergia

smokem: 1 madre fumadora; 0 madre no fumadora

#### allergy2.dta

Añade a la base anterior

allergyf: 1 alergia en la padre; 0 sin alergia

smokef: 1 padre fumador; 0 padre no fumador

backpain.dta

sex: 1 male; 2 female

age: edad en años

bp5: 1 refiere dolor de espalda a los 5 años de comenzar el estudio; 0 caso contrario

bp0: 1 refiere dolor de espalda al comenzar el estudio; 0 caso contrario

phyocc: desgaste físico de su ocupación

social: clase social en tres niveles

#### veteran.csv

Datos de un ensayo aleatorizado de los resultados de dos tratamientos para el cáncer de pulmón.

trt: 1= tratamiento estándar 1; 2= tratamiento de prueba

celltype: squamous; smallcell; adeno; large

time: tiempo de supervivencia en días

status: 0 censurado; 1 no censurado

karno: Puntuación de rendimiento de Karnofsky (100 = bueno)

diagtime: meses desde el diagnóstico hasta la aleatorización

age: edad en años

prior: tratamiento previo 0=no; 1=sí

#### lowbwt.csv

Se trata de un estudio que tiene como objetivo identificar los factores de riesgo asociados con el parto de un bebé con bajo peso al nacer (que pesa menos de 2500 gramos).

Las variables que se consideraron importantes fueron:

Low: variable categórica con dos posibilidades ( $< 2500$  g ;  $\geq 2500$  g)

Age: edad de la madre en años

Lwt: peso de la madre (libras) en su último periodo menstrual

race: raza de la madre (black, White, other)

smoke: fuma durante el embarazo (Yes; No)

ptl: semanas de gestación 0= (36-37 semanas) 1 (34-36) 2 (32-34) 3 ( $<32$  semanas)

ht: hipertensión arterial de la madre (Yes, No)

ui: infección de orina de la madre (Yes, No)

ftv: número de visitas al ginecólogo

bwt: peso del niño al nacer en gramos

#### breast.dta

Fichero de Stata, variables:

age: edad en años

grad: estadio 1,2 o 3

nodestat: ganglios linfáticos afectados, =0 ninguno, =1 entre 1 y 3, =3 4 o más

tumorsize: tamaño del tumor

survtime: supervivencia en días

died: =1 muerto, =0 vivo (censurado)

#### cigsales.dta

state: estado (3 California) (1 Alabama, 2 Arkansas y 36 West Virginia)

cigsale: ventas per-cápita en paquetes

income: renta media disponible

beer: consumo de cerveza

age15to24: proporción de la población entre 15-24 años

retprice: precio medio por cajetilla