T0. Taller encuestas. Anotate*

VIII Jornadas R Albacete 2016**

*Álvaro Hernández Vicente, Elvira Ferre Jaén, Antonio José Perán Orcajada, Ana Belén Marín Valverde, Antonio Maurandi López****

17 de noviembre de 2016

Índice

1. Lectura de datos

Leemos los datos que están contenidos en fichero `csv` que hemos preparado, añadiendo algunas variables, partiendo de la base de datos `raq.data` del libro *Discovering Statistics Using R* (A. Field, Miles, and Field 2012).

```
df <- read.table( "saeraq.csv", sep = "\t", header = TRUE )
summary( df )
```

```
##        id            sexo        actividadS       ingresos     origen
##  id1    :   1   Min.   :1.00   Min.   :1.000   Min.   :15850   A:1178
##  id10   :   1   1st Qu.:1.00   1st Qu.:2.000   1st Qu.:20968   H: 281
##  id100  :   1   Median :1.00   Median :3.000   Median :27771   M:1112
##  id1000 :   1   Mean   :1.25   Mean   :2.513   Mean   :29672
##  id1001 :   1   3rd Qu.:2.00   3rd Qu.:3.000   3rd Qu.:37932
##  id1002 :   1   Max.   :2.00   Max.   :4.000   Max.   :44710
##  (Other):2565
##   nivelIngles          Q01             Q02             Q03
##  Min.   : 1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.: 3.000   1st Qu.:3.000   1st Qu.:4.000   1st Qu.:3.000
##  Median : 4.000   Median :4.000   Median :5.000   Median :3.000
##  Mean   : 4.033   Mean   :3.626   Mean   :4.377   Mean   :3.415
##  3rd Qu.: 5.000   3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000
##  Max.   :10.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
##       Q04             Q05             Q06             Q07
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
```

*doc:T0_anotate.Rmd

**http://r-es.org/8jornadasR/

***Servicio de Apoyo Estadístico; alvarohv@um.es, elvira@um.es, antoniojose.peran@um.es, anabelen.marin4@um.es, amaurandi@um.es

```
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.000
##  Median :3.000   Median :3.000   Median :4.000   Median :3.000
##  Mean   :3.214   Mean   :3.278   Mean   :3.773   Mean   :3.076
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
##       Q08             Q09             Q10             Q11
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:3.000
##  Median :4.000   Median :3.000   Median :4.000   Median :4.000
##  Mean   :3.763   Mean   :3.154   Mean   :3.719   Mean   :3.745
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
##       Q12             Q13             Q14             Q15
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000
##  Median :3.000   Median :4.000   Median :3.000   Median :3.000
##  Mean   :2.841   Mean   :3.551   Mean   :3.124   Mean   :3.234
##  3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
##       Q16             Q17             Q18             Q19
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000
##  Median :3.000   Median :4.000   Median :4.000   Median :4.000
##  Mean   :3.121   Mean   :3.533   Mean   :3.431   Mean   :3.708
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
##       Q20             Q21             Q22             Q23
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :3.000   Median :3.000   Median :2.000
##  Mean   :2.376   Mean   :2.829   Mean   :3.112   Mean   :2.566
##  3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##
```

2. Limpieza y recodificación de variables

Fundamentalmente empleamos las funciones:

- `factor()`
- `lapply()`

Ambas del paquete base de R (R Core Team 2016).

2.1. Variables `sexo`, `actividadS`, `origen`, `nivelIngles`

```
# la variable sexo: 1 = Hombre, 2 = Mujer
df$sexo <- factor( df$sexo, labels = c( "Mujer", "Hombre" ) )
```

```
# actividadS: 1 = Nada, 2 = Poco, 3 = Mucho, 4 = Muchísimo
df$actividadS <- factor( df$actividadS, labels = c( "Nada", "Poco", "Mucho", "Muchísimo" ) )
```

A la variable ingresos no hay que hacerle nada.

```
# la variable origen: A = Albacete, M = Murcia, H = Helsinki
df$origen <- factor( df$origen, labels = c( "Albacete", "Helsinki", "Murcia" ) )

# Cambiamos el orden para que Murcia sea la segunda
df$origen <- factor( df$origen, levels = levels( df$origen )[ c( 1, 3, 2 ) ] )
```

```
# la variable nivel de inglés: nos dan los niveles del 1 al 10
nivelesIngles  <- c( "Nulo", "CasiNulo", "A1", "A2", "B1", "B2", "C1", "C2"
                     , "IsabelII", "Shakespeare" )
df$nivelIngles <- factor( df$nivelIngles, labels = nivelesIngles )
```

2.2. Preguntas Q01 a Q23

```
colnames( df )
```

```
##  [1] "id"          "sexo"        "actividadS"  "ingresos"    "origen"
##  [6] "nivelIngles" "Q01"         "Q02"         "Q03"         "Q04"
## [11] "Q05"         "Q06"         "Q07"         "Q08"         "Q09"
## [16] "Q10"         "Q11"         "Q12"         "Q13"         "Q14"
## [21] "Q15"         "Q16"         "Q17"         "Q18"         "Q19"
## [26] "Q20"         "Q21"         "Q22"         "Q23"
```

```
# se convierten en factores
nivelesQ     <- c( "Muy en desacuerdo", "En desacuerdo", "Neutro", "De acuerdo", "Muy de acuerdo" )
df[ , 7:29 ] <- lapply( df[ , 7:29 ], factor, labels = nivelesQ )

# recomendable: evitar índices números
questions         <- c( paste0( "Q0", 1:9 ), paste0( "Q", 10:23 ) )
df[ , questions ] <- lapply( df[ , questions ], factor, labels = nivelesQ )
```

Recuperamos, mediante el diccionario de datos, los enunciados y nombres de los ítems. Creamos un vector donde guardar las etiquetas de los ítems (*nombres cortos*) y los enunciados completos (*nombres largos*). Dependiendo del uso que queramos darle usaremos unos nombres u otros (también podriamos hacer uso del diccionario directamente).

```
# diccionario con los nombres
dicc           <- read.csv( "diccionario.csv", sep = ";", header = TRUE, stringsAsFactors = FALSE )

# nombres_cortos <- colnames( df ) # guardamos los nombres originales de la bbdd
# colnames( df ) <- dicc$spanish   # cambiamos el nombre de las vbles
# nombres_largos <- colnames( df ) # guardanos los nuevos nombres (enunciados items)
```

3. Guardamos los datos con Rdata

Guardamos un conjunto de datos con las variables *anotadas*, a partir de ahora trabajaremos con este conjunto de datos.

```
rm( list = ls()[ -(1:2) ] )
save.image( "saeraq.RData" )
```

```
summary( df )
```

```
##        id           sexo          actividadS     ingresos
##  id1    :   1   Mujer :1927   Nada     :613   Min.   :15850
##  id10   :   1   Hombre: 644   Poco     :658   1st Qu.:20968
##  id100  :   1                 Mucho    :669   Median :27771
##  id1000 :   1                 Muchísimo:631   Mean   :29672
##  id1001 :   1                                 3rd Qu.:37932
##  id1002 :   1                                 Max.   :44710
##  (Other):2565
##       origen       nivelIngles                 Q01
##  Albacete:1178   B1      :473   Muy en desacuerdo:  41
##  Murcia  :1112   A2      :455   En desacuerdo    : 187
##  Helsinki: 281   A1      :451   Neutro           : 735
##                  B2      :377   De acuerdo       :1338
##                  Nulo    :297   Muy de acuerdo   : 270
##                  CasiNulo:289
##                  (Other) :229
##                 Q02                     Q03                     Q04
##  Muy en desacuerdo:  20   Muy en desacuerdo: 76   Muy en desacuerdo:121
##  En desacuerdo    : 101   En desacuerdo    :448   En desacuerdo    :437
##  Neutro           : 206   Neutro           :878   Neutro           :924
##  De acuerdo       : 808   De acuerdo       :672   De acuerdo       :949
##  Muy de acuerdo   :1436   Muy de acuerdo   :497   Muy de acuerdo   :140
##
##
##                 Q05                     Q06                     Q07
##  Muy en desacuerdo: 104   Muy en desacuerdo: 146   Muy en desacuerdo:219
##  En desacuerdo    : 475   En desacuerdo    : 252   En desacuerdo    :623
##  Neutro           : 746   Neutro           : 344   Neutro           :663
##  De acuerdo       :1095   De acuerdo       :1127   De acuerdo       :875
##  Muy de acuerdo   : 151   Muy de acuerdo   : 702   Muy de acuerdo   :191
##
##
##                 Q08                     Q09                     Q10
##  Muy en desacuerdo:  72   Muy en desacuerdo:212   Muy en desacuerdo:  40
##  En desacuerdo    : 147   En desacuerdo    :732   En desacuerdo    : 248
##  Neutro           : 482   Neutro           :591   Neutro           : 467
##  De acuerdo       :1487   De acuerdo       :521   De acuerdo       :1455
##  Muy de acuerdo   : 383   Muy de acuerdo   :515   Muy de acuerdo   : 361
##
##
##                 Q11                     Q12
##  Muy en desacuerdo:  58   Muy en desacuerdo: 223
##  En desacuerdo    : 165   En desacuerdo    : 589
##  Neutro           : 565   Neutro           :1193
##  De acuerdo       :1370   De acuerdo       : 507
##  Muy de acuerdo   : 413   Muy de acuerdo   :  59
##
##
##                 Q13                     Q14                     Q15
##  Muy en desacuerdo:  72   Muy en desacuerdo:177   Muy en desacuerdo: 150
##  En desacuerdo    : 302   En desacuerdo    :453   En desacuerdo    : 457
##  Neutro           : 655   Neutro           :974   Neutro           : 778
##  De acuerdo       :1222   De acuerdo       :809   De acuerdo       :1014
##  Muy de acuerdo   : 320   Muy de acuerdo   :158   Muy de acuerdo   : 172
```

```
##
##
##                Q16                      Q17                      Q18
##  Muy en desacuerdo: 143   Muy en desacuerdo:  72   Muy en desacuerdo:147
##  En desacuerdo    : 422   En desacuerdo    : 250   En desacuerdo    :299
##  Neutro           :1079   Neutro           : 702   Neutro           :793
##  De acuerdo       : 836   De acuerdo       :1329   De acuerdo       :962
##  Muy de acuerdo   :  91   Muy de acuerdo   : 218   Muy de acuerdo   :370
##
##
##                Q19                     Q20                     Q21
##  Muy en desacuerdo: 60   Muy en desacuerdo:561   Muy en desacuerdo:239
##  En desacuerdo    :376   En desacuerdo    :952   En desacuerdo    :738
##  Neutro           :556   Neutro           :636   Neutro           :865
##  De acuerdo       :842   De acuerdo       :375   De acuerdo       :681
##  Muy de acuerdo   :737   Muy de acuerdo   : 47   Muy de acuerdo   : 48
##
##
##                Q22                     Q23
##  Muy en desacuerdo:117   Muy en desacuerdo: 320
##  En desacuerdo    :656   En desacuerdo    :1091
##  Neutro           :877   Neutro           : 696
##  De acuerdo       :664   De acuerdo       : 314
##  Muy de acuerdo   :257   Muy de acuerdo   : 150
##
##
```

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: i686-pc-linux-gnu (32-bit)
## Running under: Ubuntu 16.04.1 LTS
##
## locale:
##  [1] LC_CTYPE=es_ES.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=es_ES.UTF-8        LC_COLLATE=es_ES.UTF-8
##  [5] LC_MONETARY=es_ES.UTF-8    LC_MESSAGES=es_ES.UTF-8
##  [7] LC_PAPER=es_ES.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5    assertthat_0.1  tools_3.3.2     htmltools_0.3.5
##  [5] yaml_2.1.14     tibble_1.2      Rcpp_0.12.7     stringi_1.1.2
##  [9] rmarkdown_1.1   knitr_1.15      stringr_1.1.0   digest_0.6.10
## [13] evaluate_0.10
```

Referencias

Field, Andy, Jeremy Miles, and Zoe Field. 2012. *Discovering Statistics Using R*. 1st edition. Sage Publications Ltd.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation

for Statistical Computing. https://www.R-project.org/.