

Analisis Ejemplo

Alicia Perdices Guerra

6 de mayo, 2021

Contents

1.GASTO SANITARIO POR FUNCIÓN.

- En primer lugar leemos el fichero:

```
library(scales)
gasto_fun<-read.csv("C:/temp/GastoSanitario_Funcion_clean.csv",sep= ",")
```

- Realicemos una breve inspección de los datos:

```
str(gasto_fun)

## 'data.frame':    2000 obs. of  6 variables:
## $ TIME      : int  2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
## $ GEO       : Factor w/ 40 levels "Austria","Belgium",...: 15 15 15 15 15 16 16 16 16 16 ...
## $ UNIT      : Factor w/ 1 level "Million euro": 1 1 1 1 1 1 1 1 1 1 ...
## $ ICHA11_HC: Factor w/ 5 levels "Curative care",...: 3 2 1 4 5 3 2 1 4 5 ...
## $ Value     : num  1213034 651215 15567 324207 8868 ...
## $ Value_imp: logi  TRUE TRUE TRUE TRUE TRUE TRUE ...

colnames(gasto_fun) #Nombre de las variables

## [1] "TIME"      "GEO"      "UNIT"      "ICHA11_HC" "Value"    "Value_imp"

nrow(gasto_fun) #Número de registros

## [1] 2000

ncol(gasto_fun) #Número de variables

## [1] 6
```

- Filtramos el dataframe para que la variable GEO aparezcan solo los paises objeto de estudio

```
gasto_fun_paises<-gasto_fun[(gasto_fun$GEO!="European Union - 27 countries (from 2020)")&
+(gasto_fun$GEO!="European Union - 27 countries (2007-2013)")&
+(gasto_fun$GEO!="Euro area - 19 countries (from 2015)")&
+(gasto_fun$GEO!="Euro area - 12 countries (2001-2006)")&
+(gasto_fun$GEO!="European Union - 28 countries +(2013-2020)")&
+(gasto_fun$GEO!="European Union - 15 countries (1995-2004)")&
+(gasto_fun$GEO!="Euro area - 18 countries (2014)"),]
gasto_fun_paises<-gasto_fun_paises[(gasto_fun_paises$ICHA11_HC=="Current health care expenditure (CHE)"]
```

- Transformamos el valor de la variable “Value” a unidades.

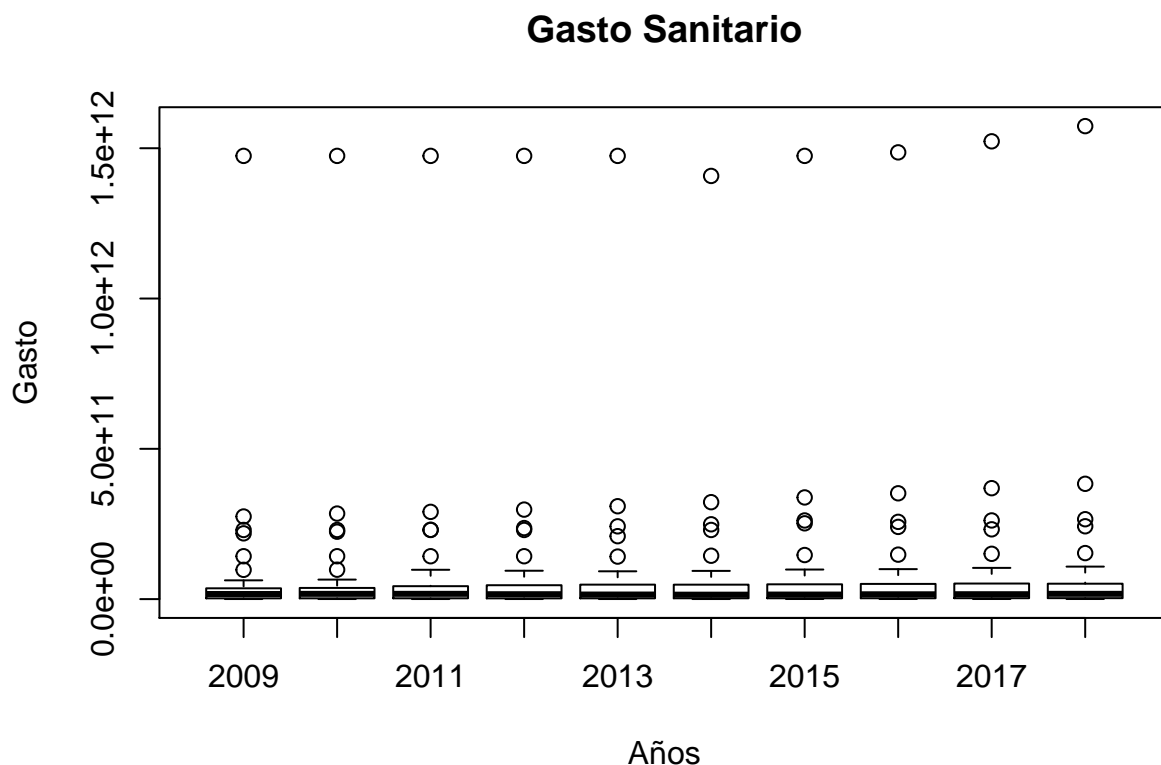
```
#Ahora reescalamos un variable "Value" del Dataframe
gasto_fun_paises$Value<-gasto_fun_paises$Value*1000000
head(gasto_fun_paises)
```

```
##      TIME                                GEO      UNIT
## 6  2009      European Union - 28 countries (2013-2020) Million euro
```

```
## 36 2009 Belgium Million euro
## 41 2009 Bulgaria Million euro
## 46 2009 Czechia Million euro
## 51 2009 Denmark Million euro
## 56 2009 Germany (until 1990 former territory of the FRG) Million euro
## ICHA11_HC Value Value_imp
## 6 Current health care expenditure (CHE) 1.474601e+12 TRUE
## 36 Current health care expenditure (CHE) 3.587939e+10 FALSE
## 41 Current health care expenditure (CHE) 3.003510e+09 TRUE
## 46 Current health care expenditure (CHE) 1.220211e+10 TRUE
## 51 Current health care expenditure (CHE) 2.516702e+10 TRUE
## 56 Current health care expenditure (CHE) 2.748410e+11 FALSE
```

- Graficamos la información:

```
boxplot(gasto_fun_paises$Value~gasto_fun_paises$TIME,xlab="Años",
        ylab="Gasto", main='Gasto Sanitario',)
```



2.INGRESOS EN EL SECTOR SANITARIO.

- En primer lugar leemos el fichero:

```
ingreso<-read.csv("C:/temp/IngresosSanitarios_Financiacion_clean.csv",sep= ",")
```

- Realicemos una breve inspección de los datos:

```
str(ingreso)
```

```
## 'data.frame': 220 obs. of 6 variables:
```

```
## $ TIME      : int  2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
## $ GEO       : Factor w/ 22 levels "Belgium","Croatia",...: 1 4 5 8 6 11 19 2 3 12 ...
## $ UNIT      : Factor w/ 1 level "Million euro": 1 1 1 1 1 1 1 1 1 1 1 ...
## $ ICHA11_FS: Factor w/ 1 level "All revenues of financing schemes": 1 1 1 1 1 1 1 1 1 1 1 ...
## $ Value     : num  35879 35879 25167 274841 926 ...
## $ Value_imp: logi  FALSE TRUE TRUE FALSE FALSE TRUE ...
```

```
colnames(ingreso) #Nombre de las variables
```

```
## [1] "TIME"      "GEO"      "UNIT"      "ICA11_FS" "Value"    "Value_imp"
```

```
nrow(ingreso) #Número de registros
```

```
## [1] 220
```

```
ncol(ingreso) #Número de variables
```

```
## [1] 6
```

- Transformamos el valor de la variable “Value” a unidades.

```
#Ahora reescalamos un variable "Value" del Dataframe
```

```
ingreso$Value<-ingreso$Value*1000000
```

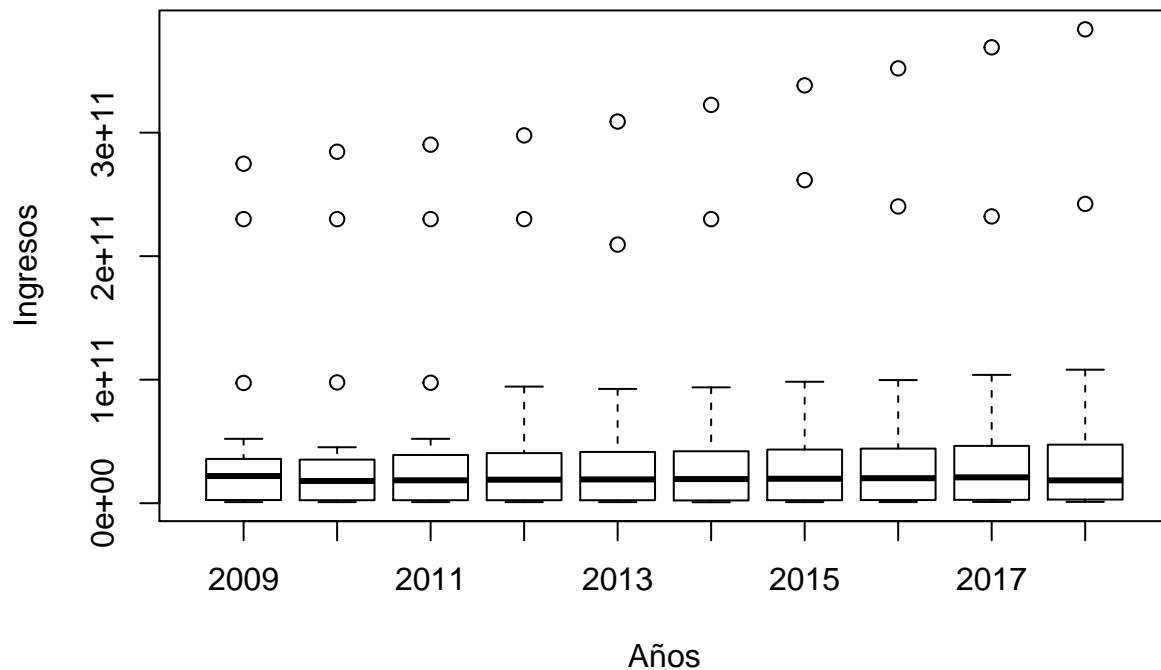
```
head(ingreso)
```

```
##      TIME                GEO      UNIT
## 1 2009                Belgium Million euro
## 2 2009                Czechia Million euro
## 3 2009                Denmark Million euro
## 4 2009 Germany (until 1990 former territory of the FRG) Million euro
## 5 2009                Estonia Million euro
## 6 2009                Ireland Million euro
##      ICA11_FS      Value Value_imp
## 1 All revenues of financing schemes 35879390000 FALSE
## 2 All revenues of financing schemes 35879390000 TRUE
## 3 All revenues of financing schemes 25167020000 TRUE
## 4 All revenues of financing schemes 274841000000 FALSE
## 5 All revenues of financing schemes 925550000 FALSE
## 6 All revenues of financing schemes 18850220000 TRUE
```

- Graficamos la información:

```
boxplot(ingreso$Value~ingreso$TIME,xlab="Años", ylab="Ingresos", main='Ingresos Sanitarios',)
```

Ingresos Sanitarios



3. ESTUDIO PARA ESPAÑA.

3.1 Creación de un DataFrame con la información para España.

- Creamos un Dataframe con toda la información de España
- Gasto Sanitario

```
spain_gasto_fun<-gasto_fun_paises[gasto_fun_paises$GEO=="Spain",]
gasto_fun_s<-spain_gasto_fun$Value
```

- Ingresos Sanitarios

```
spain_ingreso<-ingreso[ingreso$GEO=="Spain",]
ingreso_s<-spain_ingreso$Value
```

- Recursos: STAFF

```
#RECURSOS
#STAFF
medicos_e<-read.csv("C:/temp/Medicos_x_especialidad_clean.csv",sep= ",")
enfermeria<-read.csv("C:/temp/Personal_Enfermeria_Cuidados_clean.csv",sep= ",")

medicos_e<-medicos_e[(medicos_e$MED_SPEC=="Generalist medical practitioners"),]
medicos_e_s<-medicos_e[medicos_e$GEO=="Spain",]
medicos_sp<-medicos_e_s$Value
```

```

enfermeria<-enfermeria[(enfermeria$UNIT=="Number")&
  +(enfermeria$WSTATUS=="Professionally active")&
  +(enfermeria$ISCO08=="Nurses, midwives, health care assistants and home-based personal care workers")]
enfermeria_s<-enfermeria[enfermeria$GEO == "Spain",]

enfermeria_sp<-enfermeria_s$Value

```

- Prevención

```

cancer<-read.csv("C:/temp/Deteccion_Cancer_Mama_Cervix_clean.csv",sep= ",")
cancer_e<-cancer[cancer$SOURCE=="Survey data",]
cancer_e<-cancer_e[cancer_e$GEO=="Spain",]
cancer_e<-cancer_e[cancer_e$ICD10 == "Malignant neoplasm of breast",]
cancer_ep<-cancer_e$Value
cancer_ep

```

```
## [1] 79.80 77.10 79.80 79.80 79.80 79.80 79.80 81.48 79.80 79.80
```

- Causas de Muerte

```

#CAUSAS DE MUERTE
mortalidad<-read.csv("C:/temp/Mortalidad_Tratable_Prevenible_clean.csv",sep= ",")
mortalidad_e<-mortalidad[mortalidad$MORTALIT=="Total",]
mortalidad_e<-mortalidad_e[mortalidad_e$SEX=="Total",]
mortalidad_e<-mortalidad_e[mortalidad_e$UNIT=="Number",]
mortalidad_e<-mortalidad_e[mortalidad_e$GEO=="Spain",]
mortalidad_ep<-mortalidad_e$Value
mortalidad_ep<-c(mortalidad_ep,NA,NA,NA)
mortalidad_ep

```

```
## [1] 75150 74282 73920 73685 74653 74599 74651 NA NA NA
```

- Creamos un DataFrame con la información para España.

```

#Creamos el dataframe para España
year<-enfermeria_s$TIME
spain<-data.frame("year"=year,"gastos"=gasto_fun_s,
  "ingresos"=ingreso_s,"medicos"=medicos_sp,
  "enfermeros"=enfermeria_sp,
  "diagnostico_cancer_mama"=cancer_ep,
  "mortalidad"=mortalidad_ep)

spain

```

```

##   year      gastos      ingresos medicos enfermeros diagnostico_cancer_mama
## 1  2010  97384010000  97384010000   34620     206145                79.80
## 2  2011  97815780000  97815780000   34722     218540                77.10
## 3  2012  97532090000  97532090000   35061     232771                79.80
## 4  2013  94417660000  94417660000   35017     245866                79.80
## 5  2014  92518800000  92518800000   34880     250633                79.80
## 6  2015  93824250000  93824250000   34735     256033                79.80
## 7  2016  98350220000  98350220000   34797     266667                79.80
## 8  2017  99715250000  99715250000   35378     277533                81.48
## 9  2018 103899870000 103899870000   35798     284733                79.80
## 10 2019 108109700000 108109700000   35378     168590                79.80
##   mortalidad
## 1      75150
## 2      74282

```

```
## 3      73920
## 4      73685
## 5      74653
## 6      74599
## 7      74651
## 8      NA
## 9      NA
## 10     NA
```

- Imputamos los valores de la variable **mortalidad** para los años 2017-2019

```
#Imputamos valores de la variable mortalidad en los años 2017-2019
idx<-which(is.na(spain$mortalidad))
length(idx)
```

```
## [1] 3
```

```
library(VIM)
```

```
## Loading required package: colorspace
```

```
## Loading required package: grid
```

```
## VIM is ready to use.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
```

```
##
```

```
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
##      sleep
```

```
output<-kNN(spain, variable=c("mortalidad"),k=3)
spain<-output
```

- Normalizamos los valores del dataframe creado.

```
spain_norm<-abs(scale(spain[,2:7]))
spain_norm
```

```
##           gastos  ingresos  medicos enfermeros diagnostico_cancer_mama
## [1,] 0.20647994 0.20647994 1.11911071 0.9775872      0.09672527
## [2,] 0.11483094 0.11483094 0.84641770 0.6274411      2.46364962
## [3,] 0.17504797 0.17504797 0.05988552 0.2254297      0.09672527
## [4,] 0.83612774 0.83612774 0.05774675 0.1444907      0.09672527
## [5,] 1.23918639 1.23918639 0.42401089 0.2791537      0.09672527
## [6,] 0.96208702 0.96208702 0.81166271 0.4316982      0.09672527
## [7,] 0.00138884 0.00138884 0.64590814 0.7320979      0.09672527
## [8,] 0.28835719 0.28835719 0.90737261 1.0390514      1.68984743
## [9,] 1.17659922 1.17659922 2.03022617 1.2424441      0.09672527
## [10,] 2.07019242 2.07019242 0.90737261 2.0384779      0.09672527
##           mortalidad
## [1,] 1.7220199
## [2,] 0.1303946
## [3,] 0.9029454
## [4,] 1.4044631
## [5,] 0.6613632
## [6,] 0.5461208
```

```
## [7,] 0.6570950
## [8,] 0.9029454
## [9,] 0.9029454
## [10,] 0.6570950
## attr("scaled:center")
##          gastos          ingresos          medicos
##      9.835676e+10      9.835676e+10      3.503860e+04
##      enfermeros diagnostico_cancer_mama          mortalidad
##      2.407511e+05      7.969800e+01      7.434310e+04
## attr("scaled:scale")
##          gastos          ingresos          medicos
##      4.711126e+09      4.711126e+09      3.740470e+02
##      enfermeros diagnostico_cancer_mama          mortalidad
##      3.539950e+04      1.054533e+00      4.685776e+02

spain_norm<-cbind(spain_norm, year=c(year))
spain_norm<-data.frame(spain_norm)
```

3.2 Aplicación de un modelo lineal que explique la variable mortalidad en función de gasto y número de staff en hospital.

```
modelo<-lm(mortalidad~medicos, data=spain) #modelo que explica la variable mortalidad
#en función del número de médicos disponibles.
summary(modelo)
```

```
##
## Call:
## lm(formula = mortalidad ~ medicos, data = spain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -673.11 -257.68   74.82  184.73  543.83
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 98699.7499 12910.9770   7.645 6.05e-05 ***
## medicos      -0.6951    0.3685  -1.887  0.0959 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 413.5 on 8 degrees of freedom
## Multiple R-squared:  0.3079, Adjusted R-squared:  0.2214
## F-statistic: 3.559 on 1 and 8 DF, p-value: 0.09593
```

```
modelo2<-lm(mortalidad~gastos, data=spain) #modelo que explica la variable mortalidad
#en función del gasto aplicado al sector sanitario.
summary(modelo2)
```

```
##
## Call:
## lm(formula = mortalidad ~ gastos, data = spain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -671.88 -414.69   88.53  303.28  803.50
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.469e+04  3.460e+03   21.59 2.24e-08 ***
## gastos      -3.498e-09  3.514e-08   -0.10  0.923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 496.7 on 8 degrees of freedom
## Multiple R-squared:  0.001237, Adjusted R-squared:  -0.1236
## F-statistic: 0.00991 on 1 and 8 DF, p-value: 0.9232

modelo3<-lm(mortalidad~enfermeros, data= spain)#modelo que explica la mortalidad
                                                #en función del número de enfermeros disponibles.
summary(modelo3)
```

```
##
## Call:
## lm(formula = mortalidad ~ enfermeros, data = spain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -627.3 -199.8 -142.5  364.1  598.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.579e+04  1.013e+03  74.839 1.13e-12 ***
## enfermeros  -6.029e-03  4.166e-03  -1.447  0.186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 442.5 on 8 degrees of freedom
## Multiple R-squared:  0.2074, Adjusted R-squared:  0.1084
## F-statistic: 2.094 on 1 and 8 DF, p-value: 0.1859
```

- A continuación realizamos predicciones de los modelos sobre nuevos datos, es decir, que mortalidad cabría esperar si se disminuye el gasto y el staff hospitalario?

```
new_data=data.frame(gastos=973, medicos=50, enfermeros=23)
predict(modelo,new_data)
```

```
##          1
## 98664.99
```

```
predict(modelo2,new_data)
```

```
##          1
## 74687.2
```

```
predict(modelo3,new_data)
```

```
##          1
## 75794.4
```

3.3 Modelización del Gasto Sanitario por Países.

Buscamos agrupar los países por similitud en el Gasto aplicado al sector sanitario.


```
kmeans_res<-kmeans(gasto_fun_paises$Value, centers = 3)
kmeans_res
```

#Representamos los grupos

```
plot(gasto_fun_paises$Value, col=kmeans_res$cluster)
```

