

# Sesión 3 - Importación de bases de datos como panel en Rstudio

Juan Camilo Forero- Germán C. Rodriguez - Jhan Andrade

21/9/2020

## Instalación de Paquetes

Los paquetes que se van a utilizar son:

```
#install.packages("plm")           #Panel linear models
#install.packages("gplots")        #Tools for plotting data
#install.packages("stargazer")     #Tablas más estéticas
#install.packages("foreign")       #Importar datos
#install.packages("sandwich")      #Estimador de errores robustos
#install.packages("lmtest")        #Inferencia estadística para modelos lineales
#install.packages("tseries")
#install.packages("wooldridge")    #Bases de datos que se encuentran en Wooldridge
#install.packages("AER")           #AER contiene estimadores como tobit y variables instrumentales (IV)

#install.packages("dplyr")

library(plm);library(gplots);library(stargazer)
library(foreign);library(sandwich);library(lmtest)
library(tseries);library(wooldridge);library(AER);library(dplyr)
```

## Importación de base de datos

Para este ejercicio trabajaremos una base de datos que se encuentra en la memoria de Rstudio. En este caso, se empleará un ejemplo adaptado de **Introduction to Econometrics with R** de *Christoph Hanck, Martin Arnold, Alexander Gerber and Martin Schmelze*

Para quien quiera profundizar más en dicho libro, se adjunta el enlace: + <https://www.econometrics-with-r.org/index.html>

## Cargamos la base de datos

```
data("Fatalities") # Base de datos importada desde la libreria AER
help("Fatalities")
```

Para usar la función glimpse es necesario cargar el paquete dplyr.

```
# class(Fatalities) # Se importa la base de datos como un data frame
# dim(Fatalities)   # Da las dimensiones del data frame (336 obs y 34 variables)
```

Se indica la base de datos "Fatalities" en formato panel:

```
Panel= pdata.frame(Fatalities, index = c("state","year"))
```

```
attach(Panel)
class(Panel)  # el objeto ahora es un tipo especial de data frame llamado pdata.frame

## [1] "pdata.frame" "data.frame"
#help("pdata.frame")
```

Hay que tener en cuenta la estructura del código para indexar como panel la base:

- `Nombre <- pdata.frame(Base_datos, index=c("Variable del individuo", "Variable de tiempo"))`

Primero va la variable indicativa del individuo y luego la variable indicativa del tiempo.

La base de datos a emplear tiene la siguiente descripción:

- US traffic fatalities panel data for the “lower 48” US states (i.e., excluding Alaska and Hawaii), annually for 1982 through 1988 state and year are the individual and time index respectively

Se define la tasa de mortalidad como:

```
Panel$fatal_rate <- Fatalities$fatal / Fatalities$pop * 100

#Para extraer una variable del panel (e.g. pop de Panel)
#Panel$pop      #La extracción de una variable es igual que para un data frame normal
#La extracción incluye los dos índices del panel para cada observación

# Para extraer solo los índices de cada observación del panel
#index(Panel)   # Retorna un df cuyas dos variables son los dos índices
```

Para conocer las dimensiones del panel y determinar si un panel está balanceado, se emplea el comando `pdim`

```
pdim(Panel)
```

```
## Balanced Panel: n = 48, T = 7, N = 336
```

Con la siguiente función se identifica a aquellas variables que cambian a lo largo del tiempo (pero que son constantes para todos los individuos) y aquellas variables que cambian entre individuos (pero que son constantes a lo largo del tiempo)

```
pvar(Panel)
```

```
## no time variation:      state
## no individual variation: year unempus emppopus
```

## Estimación de modelos para datos panel

Para estimar los distintos modelos de datos panel se utiliza el comando **plm**. La estructura del código es bastante sencilla:

- `*plm(Var_explicada~ Var_explicativas, data=“—”, model=“—”, index=c(“Var_individuo”, “Var_tiempo”))`

Para *model* se puede elegir entre Efectos fijos (**within**), Efectos Aleatorios (**random**) y Mínimos Cuadrados Combinados (**pooling**)

Directamente en el código también se puede indexar la base de datos, en caso de que no se le haya indicado antes a RStudio que la base de datos es un panel.

## Minimos Cuadrados Combinados

Para estimar un modelo de Mínimos Cuadrados Combinados el tipo de modelo es *"pooling"*

```
fatalities_pooled <- plm(fatal_rate ~ beertax + drinkage + unemp + log(income),
  index = c("state", "year"),
  model = "pooling",
  data = Panel)
```

## Efectos fijos

Para estimar un modelo de Efectos fijos el tipo de modelo es “within”

```
fatalities_FE <- plm(fatal_rate ~ beertax + drinkage + unemp + log(income),
  index = c("state", "year"),
  model = "within",
  data = Panel)
```

## Efectos aleatorios

Para estimar un modelo de Efectos aleatorios el tipo de modelo es “Random”

```
fatalities_EA <- plm(fatal_rate ~ beertax + drinkage + unemp + log(income),
  index = c("state", "year"),
  model = "random",
  data = Panel)
```

El  $\theta$  se puede calcular despejando de la formula teórica de la transformación que se realiza con Efectos Aleatorios.

$$\hat{\theta}_{it} = \frac{y_{it} - \hat{y}_{it} - \hat{u}_{it}}{\bar{y}}$$

Luego, el  $\theta$  para este ejemplo es:

```
fitted.values(fatalities_EA)[1:7] #valores ajustados de Y para el primer individuo
```

```
##          al          al          al          al          al          al
## 0.001797429 0.002174627 0.003105684 0.003809390 0.003113738 0.003841383
##          al
## 0.003956695
```

```
attach(Panel)
```

```
## The following objects are masked from Panel (pos = 3):
```

```
##
##      afatal, baptist, beertax, breath, drinkage, dry, emppop, emppopus,
##      fatal, fatal1517, fatal1820, fatal2124, gsp, income, jail, miles,
##      milestot, mormon, nfatal, nfatal1517, nfatal1820, nfatal2124, pop,
##      pop1517, pop1820, pop2124, service, sfatal, spirits, state, unemp,
##      unempus, year, youngdrivers
```

```
fatal_rate[1:7] #valores observados de Y para el primer individuo
```

```
## [1] 0.0212836 0.0234848 0.0233643 0.0219348 0.0266914 0.0271859 0.0249391
```

```
mean(fatal_rate[1:7]) #valor promedio de Y para el primer individuo
```

```
## [1] 0.02412627
```

```
residuals(fatalities_EA)[1:7] #Residuales para el primer individuo
```

```
##      al-1982      al-1983      al-1984      al-1985      al-1986
## -0.0009525292  0.0008714732 -0.0001800846 -0.0023132913  0.0031389605
##      al-1987      al-1988
##  0.0029058153  0.0005437052
```

```
theta = (fatal_rate[1:7]-fitted.values(fatalities_EA)[1:7]-residuals(fatalities_EA)[1:7])/mean(fatal_rate)
theta
```

```
##      al      al      al      al      al      al      al
## 0.8471554 0.8471554 0.8471554 0.8471554 0.8471554 0.8471554 0.8471554
```

## Primeras diferencias

Para estimar un modelo de Primeras diferencias el tipo de modelo es “*fd*”

```
fatalities_FD <- plm(fatal_rate ~ beertax + drinkage + unemp + log(income),
                     index = c("state", "year"),
                     model = "fd",
                     data = Panel)
```

Para presentar los resultados se hace uso de la función Stargazer

```
stargazer(fatalities_pooled, fatalities_FE, fatalities_EA, fatalities_FD, digits = 3,
          header = F, type = "text",
          title = "Modelo de catástrofes de tránsito por conducir en estado de ebriedad",
          column.labels = c("Pooled", "EF", "RE", "PD"))
```

```
##
## Modelo de catástrofes de tránsito por conducir en estado de ebriedad
## =====
##                               Dependent variable:
##                               -----
##                               fatal_rate
##                               EF      RE      PD
##                               (1)      (2)      (3)      (4)
## -----
## beertax      0.001*      -0.004*      0.001      0.001
##              (0.001)      (0.002)      (0.001)      (0.003)
##
## drinkage      -0.0001      -0.0004*      -0.0002      -0.0001
##              (0.0003)      (0.0002)      (0.0002)      (0.0003)
##
## unemp      -0.0003**      -0.0002      -0.0004***      -0.00000
##              (0.0001)      (0.0001)      (0.0001)      (0.0001)
##
## log(income)      -0.019***      0.006*      -0.003      0.032***
##              (0.002)      (0.004)      (0.003)      (0.006)
##
## Constant      0.209***      0.054*      -0.001***
##              (0.023)      (0.031)      (0.0002)
## -----
## Observations      336      336      336      288
## R2      0.277      0.125      0.070      0.104
## Adjusted R2      0.268      -0.032      0.059      0.091
## F Statistic  31.651*** (df = 4; 331) 10.176*** (df = 4; 284) 24.916*** 8.170*** (df = 4; 283)
## =====
```

## Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Llegados a este punto, los elementos abordados fueron: i) identificar la forma como se indexa una base de datos en formato panel; ii) estimar los modelos de datos panel de EF, EA, PD y Pooling utilizando el comando *plm*; y, iii) presentar los resultados por medio de la función stargazer.