

Datos

La manada-02

5/13/2020

Resumen del do file

- Importar datos
- Renombrar variables
- Agregar etiquetas
- Generar nuevas variables
- Estadística descriptiva
- Regresión lineal
- Gráficos
 - Histogramas
 - Distribución de densidad
 - Scatter plot (gráficos de dispersión)

Datos

Economic Labor Market Panel Survey (ELMPS), para los países (en inglés): Egypt, Arab Rep. desde 1998 hasta el 2006.

El diccionario de los datos está disponible en https://catalog.ihsn.org/index.php/catalog/5869/data-dictionary/F9?file_name=elmps%202012%20comptbl%2088_12%20v2.1

Importar datos

Usamos la librería **haven** para importar datos de Stata en R, así también como **tidyverse** para la manipulación de datos y usar el acortador de `%>%`¹ para simplificar las funciones, además de guardar la data para guardar memoria (41mb) en la pc con el comando base **saveRds**, para luego leer la base de datos Rds (3mb) con el comando **readRds** y asignar el identificador del data frame como *panel*.

```
rm(list = ls())
library(haven)
library(tidyverse)
# qw <- read_dta("tarea/qw.dta")
read_dta("tarea/ELMPS Panel data models.dta") %>% #esta base de datos pesa 41.4mb
  saveRDS("tarea/panel.rds") #asi que se guarda como data R se reduce a 3 mb
panel <- readRDS("tarea/panel.rds")
```

Renombrar variables

Como las columnas están en inglés necesitamos actualizarla a español, usando el comando **rename** guiándonos del diccionario presentado anteriormente.

¹En términos sencillo significa que una vez realizado un proceso pasas al siguiente paso con la data generada

```
panel1 <- panel %>%
  rename( # renombrar variables
    anio = round, # anio de la encuesta
    id = indid, #identificador de persona
    id_casa = hhid, #identificador del hogar
    edad = age, # Edad de la persona
    grado_acad10 = educ2, #grado academico edad >= 10
    grado_acad6 = educ1v1, #grado academico edad >= 6
    anios_de_escuela = yrschl, # anios en la escuela
    anio_primer_trabajo = yrjob1, # yrjob1 == job1_y ano de inicio trabajo
    anios_trabajo_1_tra = job1_01, # anios trabajando en su primer empleo
    sector_01 = job1_05 # sector del primer empleo
  )
```

Agregar etiquetas

De igual manera actualizamos las etiquetas de la base de datos con el paquete `expss`,

```
library(expss) # cambiar etiquetas
panel1 <- panel1 %>%
  apply_labels( # comando de asignacion y cambio de etiquetas
    anio = "anio de la encuesta" # etiqueta para anio
  )
```

Generar nuevas variables y conservar variables importantes

Como son datos laborales nos centraremos en la experiencia total de la persona encuestada y alguna otra variable *falta contenido aqui(agregar algo)*, para la presentacion de los datos finales se seleccionara variables relevantes con el comando `select`.

```
# en el documento de referencia se compara el censo y la
# panel1 <- panel1 %>% esperar la confirmacion de la manada
# experiencia
```

Data final

```
glimpse(panel1)
```

```
## Rows: 137,951
## Columns: 63
## $ anio      <labelled> 2006, 2006, 2006, 2006, 2006, 2006, 2012, ...
## $ id        <chr> "0601000601", "0601000902", "0601000903", "0601...
## $ id_casa    <chr> "06010006", "06010009", "06010009", "06010009",...
## $ pn         <dbl> 1, 2, 3, 4, 5, 6, 5, 5, 6, 6, 3, 4, 5, 7, 4, 2,...
## $ Fhhid      <chr> "06010006", "06010009", "06010009", "06010009",...
## $ Findid     <chr> "0601000601", "0601000902", "0601000903", "0601...
## $ attr_98_06_12 <dbl+lbl> 5, 5, 5, 5, 5, 5, 4, 4, 4, 4, 5, 5, 5, 5, 5...
## $ hhid_12     <chr> "", "", "", "", "", "", "120110971", "120110971...
## $ hhid_06     <chr> "06010006", "06010009", "06010009", "06010009",...
## $ hhid_98     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",...
## $ indid_12    <chr> "", "", "", "", "", "", "12011097105", "1201109...
## $ indid_06    <chr> "0601000601", "0601000902", "0601000903", "0601..."
```

```

## $ indid_98      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",...
## $ pn_12         <dbl> NA, NA, NA, NA, NA, NA, 5, 5, 6, 6, NA, NA, NA,...
## $ pn_06         <dbl> 1, 2, 3, 4, 5, 6, 5, 5, 6, 6, 3, 4, 5, 7, 4, 2,...
## $ pn_98         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ in_2012       <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1,...
## $ in_2006       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ in_1998       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ hhid_panel_06_12 <chr> "", "", "", "", "", "", "", "06010010", "", "060100...
## $ hhid_panel_98_06 <chr> "98011283", "98011286", "98011286", "98011286",...
## $ expan         <dbl> 2443.800, 2536.223, 2536.223, 2536.223, 2536.22...
## $ sex           <dbl+lbl> 1, 1, 2, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2...
## $ edad          <dbl> 39, 38, 41, 39, 14, 73, 12, 6, 7, 1, 34, 42, 8,...
## $ qwealth       <dbl> 5, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,...
## $ grado_acad10  <dbl+lbl> 8, 8, 7, 8, 4, 8, 2, NA, NA, NA, 8, 8, 2, 6...
## $ grado_acad6   <dbl+lbl> 6, 6, 5, 6, 3, 6, 2, NA, 1, NA, 6, 6, 2, 4,...
## $ anios_de_escuela <dbl> 16, 16, 14, 17, 9, 16, 5, 0, 0, NA, 16, 17, 2, ...
## $ fthcode       <dbl> NA, NA, NA, NA, 4, NA, 1, 1, 1, NA, 1, NA, 4, N...
## $ mthcode       <dbl> 2, 1, 1, NA, 3, NA, 2, 2, 2, NA, 2, NA, 3, NA, ...
## $ anio_primer_trabajo <dbl> 1998, 2004, NA, 1996, NA, 1955, NA, NA, NA, NA,...
## $ crlfsr1       <dbl> 1, 1, 0, 1, 0, NA, 0, 0, 0, NA, 1, 1, 0, 1, NA,...
## $ cremp1        <dbl> 1, 1, 0, 1, 0, NA, 0, 0, 0, NA, 1, 1, 0, 1, NA,...
## $ crmanpwr1     <dbl> 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1,...
## $ crmanpwr2     <dbl> 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1,...
## $ crempstp      <dbl+lbl> 1, 1, NA, 1, NA, NA, NA, NA, NA, NA, 1, 1, ...
## $ crsectrp      <dbl+lbl> 3, 3, NA, 3, NA, NA, NA, NA, NA, NA, 1, 1, ...
## $ crstabl       <dbl+lbl> 1, 1, NA, 1, NA, NA, NA, NA, NA, NA, 1, 1, ...
## $ crecacid_cl   <dbl+lbl> 7, 8, NA, 5, NA, NA, NA, NA, NA, NA, 15, 14...
## $ crformal      <dbl+lbl> 1, 1, NA, 1, NA, NA, NA, NA, NA, NA, 1, 1, ...
## $ crinstsec     <dbl+lbl> 8, 8, NA, 8, NA, NA, NA, NA, NA, NA, 10, 10...
## $ RhrwgAllJobCPI12 <dbl> NA, 6.617009, NA, 6.427951, NA, NA, NA, NA, NA,...
## $ job1_y        <dbl> 1998, 2004, NA, 1996, NA, 1955, NA, NA, NA, NA,...
## $ anios_trabajo_1_tra <dbl+lbl> 1, 1, NA, 1, NA, 1, NA, NA, NA, NA, 1, 1, N...
## $ sector_01     <dbl+lbl> 3, 3, NA, 3, NA, 1, NA, NA, NA, NA, 1, 1, N...
## $ job1_07       <dbl+lbl> 1, 1, NA, 1, NA, 1, NA, NA, NA, NA, 1, 1, N...
## $ job1_08_cl    <dbl+lbl> 1, 1, NA, 1, NA, 1, NA, NA, NA, NA, 1, 1, N...
## $ job1_12       <dbl+lbl> 1, 2, NA, 2, NA, 1, NA, NA, NA, NA, 1, 1, N...
## $ job2_y        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job2_01       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job2_05       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job2_07       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job2_08_cl    <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job2_12       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_y        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_01       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_05       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_07       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_08_cl    <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ job3_12       <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ panel_wt_98_06 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ panel_wt_98_06_12 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ panel_wt_06_12 <dbl> NA, NA, NA, NA, NA, NA, 1.886362, 1.886362, 1.8...

```

Estadística descriptiva

Regresión lineal

Graficos