Adaptacion R 01

La manada-01

5/13/2020

# comparaciones de codigo Stata y R

* clear al == rm(list = ls())
  + elimina toda la base de datos y variables anteriormente utilizadas
* use “pathdata” == haven::read\_dta(“data.dta”)
  + leen la base de datos, en R se automatiza en proyectos
* rename datos\_viejos datos\_nuevos == data\_nombre <- haven::read\_dta(“data.dta”)
* label var salario " = expss::apply\_labels(data, col = “label”, col2 = “label2”)
  + Anades etiquetas a las variables o columnas de los datos
* generate lvar0 = ln(var0) == mutate(data, lvar0 =log(10))
  + genera nuevas variables, de ecuaciones
* egen var = mean(var0) este codigo es usado para extraer el dato del resultado como un escalar, para presentar como datos
* gen dummy = 1 if educ > 14 == mutate(data, dummy = ifelse(educ > 14, 1, 0))
  + Genera variables con condicionantes, en el ejemplo con stata creas una nueva columna llamada ‘dummy’ si la variable educ es mayor o igual a 14, pero deja vacios si no lo son, en R se resume asi: dummy = ifelse(salario >= 14, 1, 0) si la variable educ es mayor o igual a 14 imprime 1, en otros casos 0
* reg y x1 x2 == lm(y ~ x1 + x2, data = nombre de la base de datos)
  + Regresion lineal
  + En R puedes guardar los datos de la regresion asignando un nombre `nombre\_reg <- lm(y ~ x1 + x2, data = nombre de la base de datos)
* histogram y twoway == ggplot
  + de datos a graficos

# En trabajo

## Paquetes necesarios

# install.packages('tidyverse')  
# install.packages('remotes')  
library(tidyverse)  
# install.packages('expss')  
library(expss)  
# remotes::iinstall\_github("jacob-long/jtools")  
library(jtools)  
# remotes::install\_github("datalorax/equatiomatic")  
library(equatiomatic)

## Base de datos

En el dofile, de las lineas 5 - 29, sin incluir la media y la desviacion standar de las variables

rm(list = ls())  
salario <- haven::read\_dta("Stata/salarios.dta")  
# Generar variables   
salario1 <- salario %>% # creamos una nueva base de datos para comparar  
 # con la anterior base de datos  
 mutate(lsalario = log(Salario),  
 exper2 = exper^2,  
 educ2 = educ^2,  
 dumm = ifelse(educ >= 14, 1, 0)  
 ) %>%   
 apply\_labels(Salario = "salarios en miles de soles para Junín",   
 educ = "años de educación",   
 exper = "años de experiencia",  
 lsalario = "logaritmo neperiano del salario",  
 educ2 = "años al cuadrado de educación",   
 exper2 = "años al cuadrado de experiencia"  
 )  
  
tibble(  
 variable = c('Salarios', 'educacion'),   
 media = c(mean(salario1$Salario), mean(salario1$educ)),  
 sd = c(sd(salario1$Salario, sd(salario1$educ))))

## # A tibble: 2 x 3  
## variable media sd  
## <chr> <dbl> <dbl>  
## 1 Salarios 958. 404.  
## 2 educacion 13.5 404.

## Despues de la linea 31

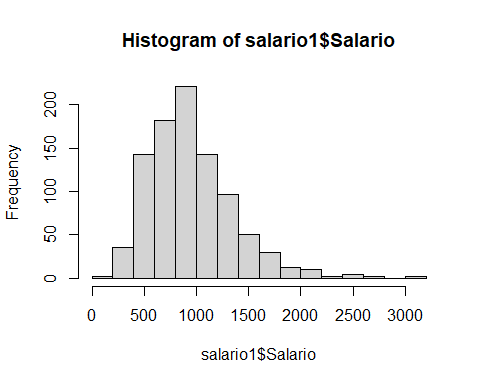
r1 <- lm(lsalario ~ educ + exper, data = salario1)  
#regresion  
summ(r1)

## MODEL INFO:  
## Observations: 935  
## Dependent Variable: lsalario  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,932) = 70.16, p = 0.00  
## R² = 0.13  
## Adj. R² = 0.13   
##   
## Standard errors: OLS  
## -----------------------------------------------  
## Est. S.E. t val. p  
## ----------------- ------ ------ -------- ------  
## (Intercept) 5.50 0.11 49.12 0.00  
## educ 0.08 0.01 11.83 0.00  
## exper 0.02 0.00 5.99 0.00  
## -----------------------------------------------

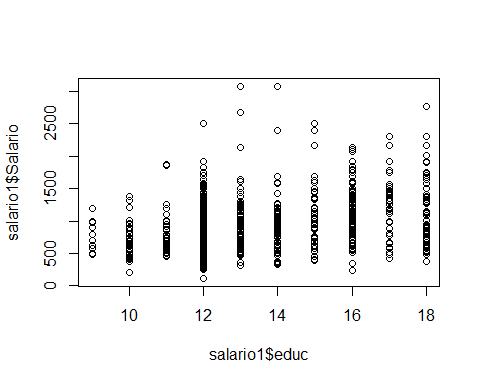
## Graficos RBASE

### Solo R

hist(salario1$Salario)

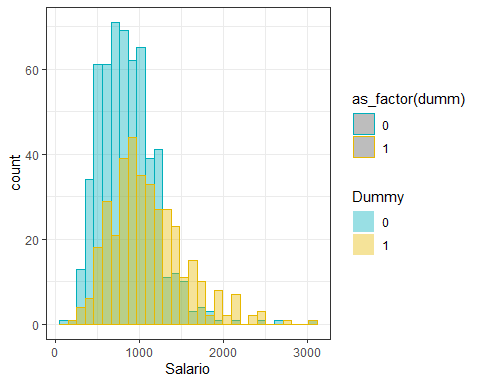


plot(salario1$educ, salario1$Salario)



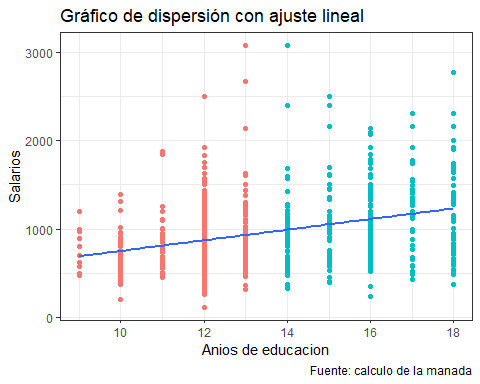
## Graficos ggplot2

salario1 %>%   
 ggplot(aes(Salario)) +   
 geom\_histogram(  
 aes(color = as\_factor(dumm),  
 fill = as\_factor(dumm)  
 ),   
 alpha=0.4, position = 'identity', bin = 50  
 )+  
 theme\_bw() +  
 labs(fill="Dummy") +   
 scale\_fill\_manual(values = c("#00AFBB", "#E7B800")) +  
 scale\_color\_manual(values = c("#00AFBB", "#E7B800"))



### scatter plot

salario1 %>%   
 ggplot(aes(educ, Salario)) + geom\_point(aes(color = as\_factor(dumm))) +   
 theme\_bw() + geom\_smooth(method = lm, se = F) +   
 labs(title ="Gráfico de dispersión con ajuste lineal",   
 caption = "Fuente: calculo de la manada",   
 x = "Anios de educacion", y = "Salarios"   
 ) +   
 theme(legend.position = "none")



## extraccion automatica de los modelos predictivos

extract\_eq(r1)

extract\_eq(r1, use\_coefs = TRUE)