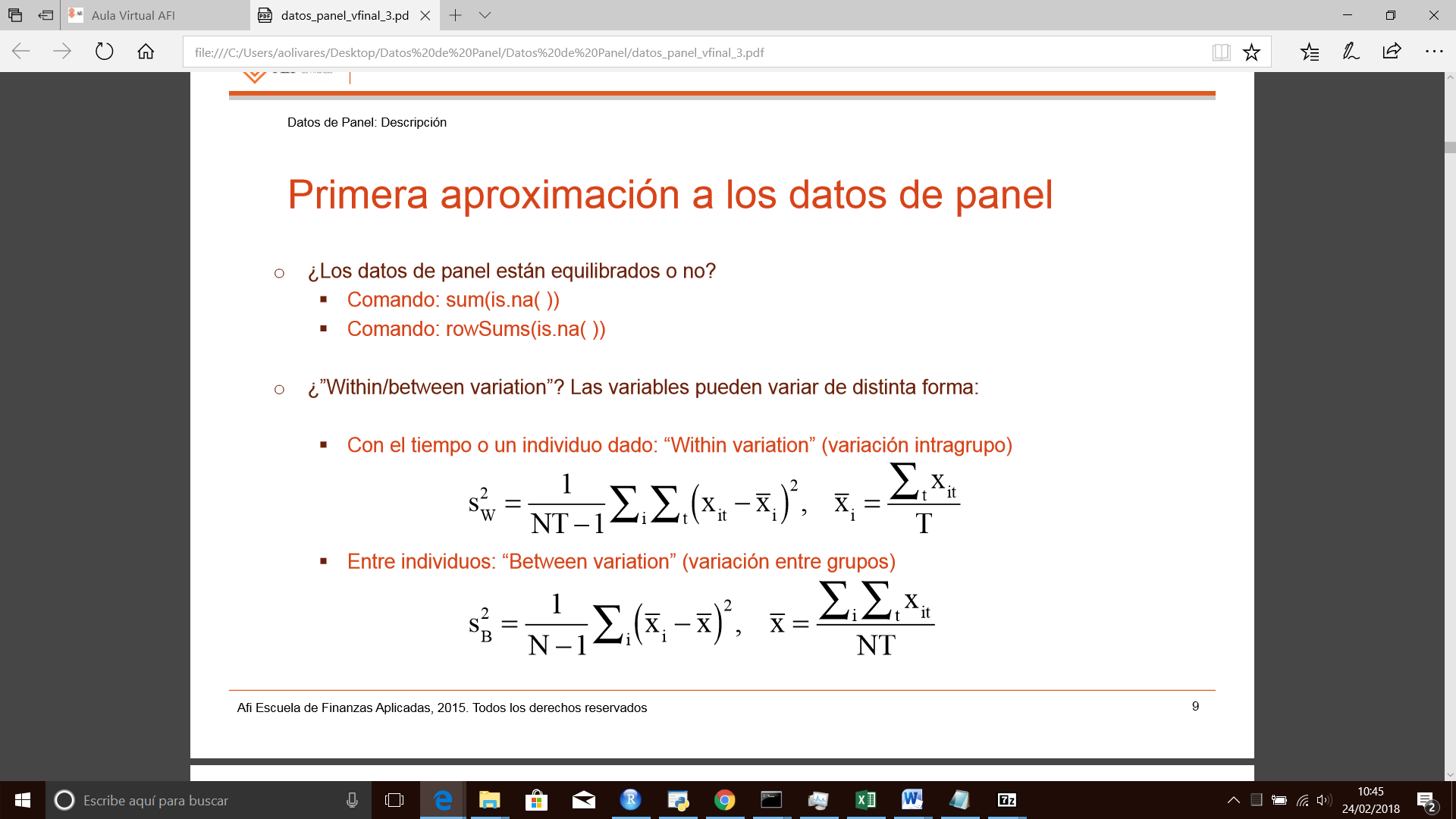
PROYECTO DE PSID – DATOS DE PANEL

Armando Olivares

PARTE DESCRIPTIVA:

1.- Determine qué variables de la base de datos son invariantes en el tiempo y argumente

La variabilidad en el tiempo es fácilmente extraíbles de la fórmula:



Las variables invariantes en el tiempo serán aquellas cuyo

Los valores para cada variable se muestran a continuación:

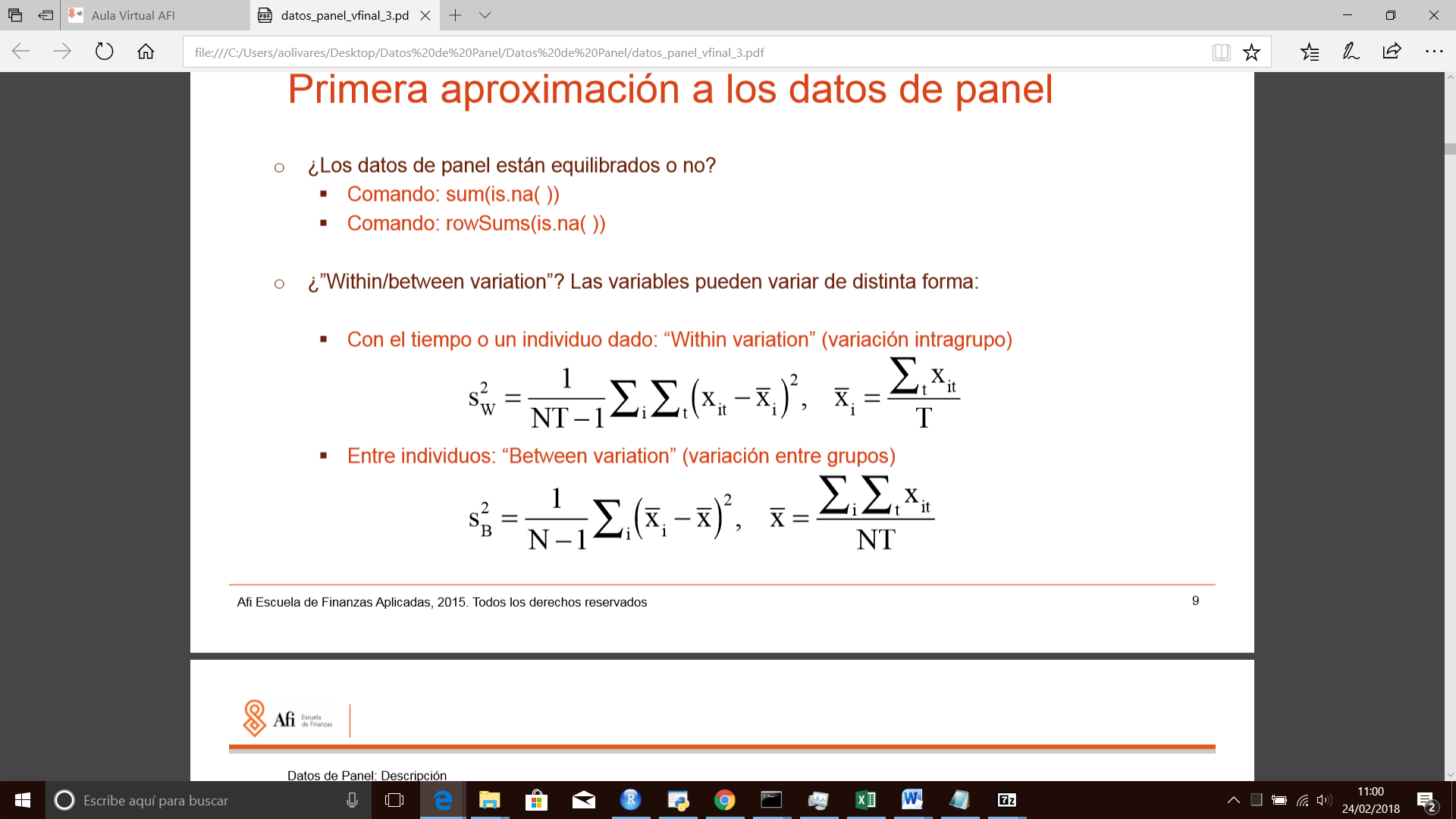
|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | Variables | |  |  |
| t | 4,67E+00 | 2,96E+04 |
| id | 0,00E+00 | 0,00E+00 |
| exp | 4,67E+00 | 1,16E+02 |
| wks | 1,14E-01 | 1,08E+01 |
| occ | 1,48E-04 | 2,20E-01 |
| ind | 4,02E-05 | 2,16E-01 |
| south | 6,32E-06 | 2,02E-01 |
| smsa | 8,47E-05 | 2,12E-01 |
| ms | 3,47E-05 | 1,36E-01 |
| fem | 0,00E+00 | 1,00E-01 |
| union | 6,58E-05 | 2,06E-01 |
| ed | 0,00E+00 | 7,78E+00 |
| blk | 0,00E+00 | 6,72E-02 |
| lwage | 4,41E-02 | 1,55E-01 |
| tdum1 | 1,43E-01 | 0,00E+00 |
| tdum2 | 1,43E-01 | 0,00E+00 |
| tdum3 | 1,43E-01 | 0,00E+00 |
| tdum4 | 1,43E-01 | 0,00E+00 |
| tdum5 | 1,43E-01 | 0,00E+00 |
| tdum6 | 1,43E-01 | 0,00E+00 |
| tdum7 | 1,43E-01 | 0,00E+00 |
| exp2 | 7,37E+09 | 2,39E+05 |

Tabla 1

De la tabla 1 se extrae que las variables invariantes en el tiempo son: **fem, blk y ed**

2.-Determine qué variables de la base de datos son invariantes en el individuo y argumente

Las variables invariantes en el individuo en el individuo son extraídas con:



Cuyo resultado arroja que las invariantes en el individuo son el tiempo mismo (**t**) y las dummies de tiempo ya que su valor , como se observa en la tabla 1.

3- Suponga que pretende estudiar la relación que existe entre las variables disponibles en la base de datos y el salario. Determine qué relación existe a través de una matriz de correlaciones.

El Grafico de correlaciones se muestra a continuación:

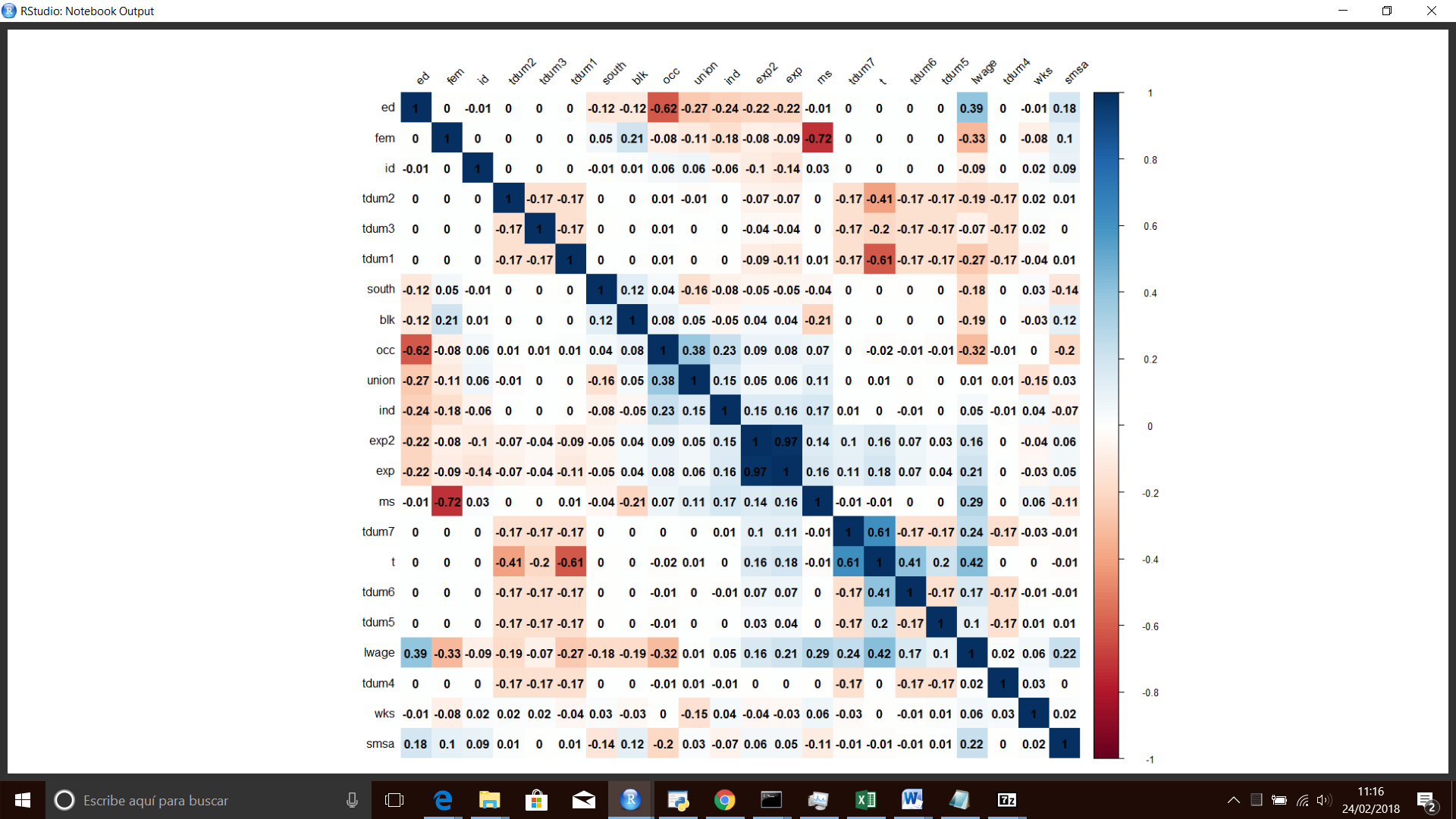


Figura 1

Del gráfico de correlaciones podemos observar que existe una correlación positiva entre el salario y la variables: *educación (ed), estado civil (ms), el vivir en una zona metropolitana(smsa), y la experiencia (exp)*, esto implica una relación lineal (directamente proporcionales es decir al aumentar una aumenta la otra), y tiene una correlación negativa con: la ocupación (occ), si vives o no en el sur de EEUU (south), si el individuo es mujer (fem) y si el individuo es afroamericano (blk)

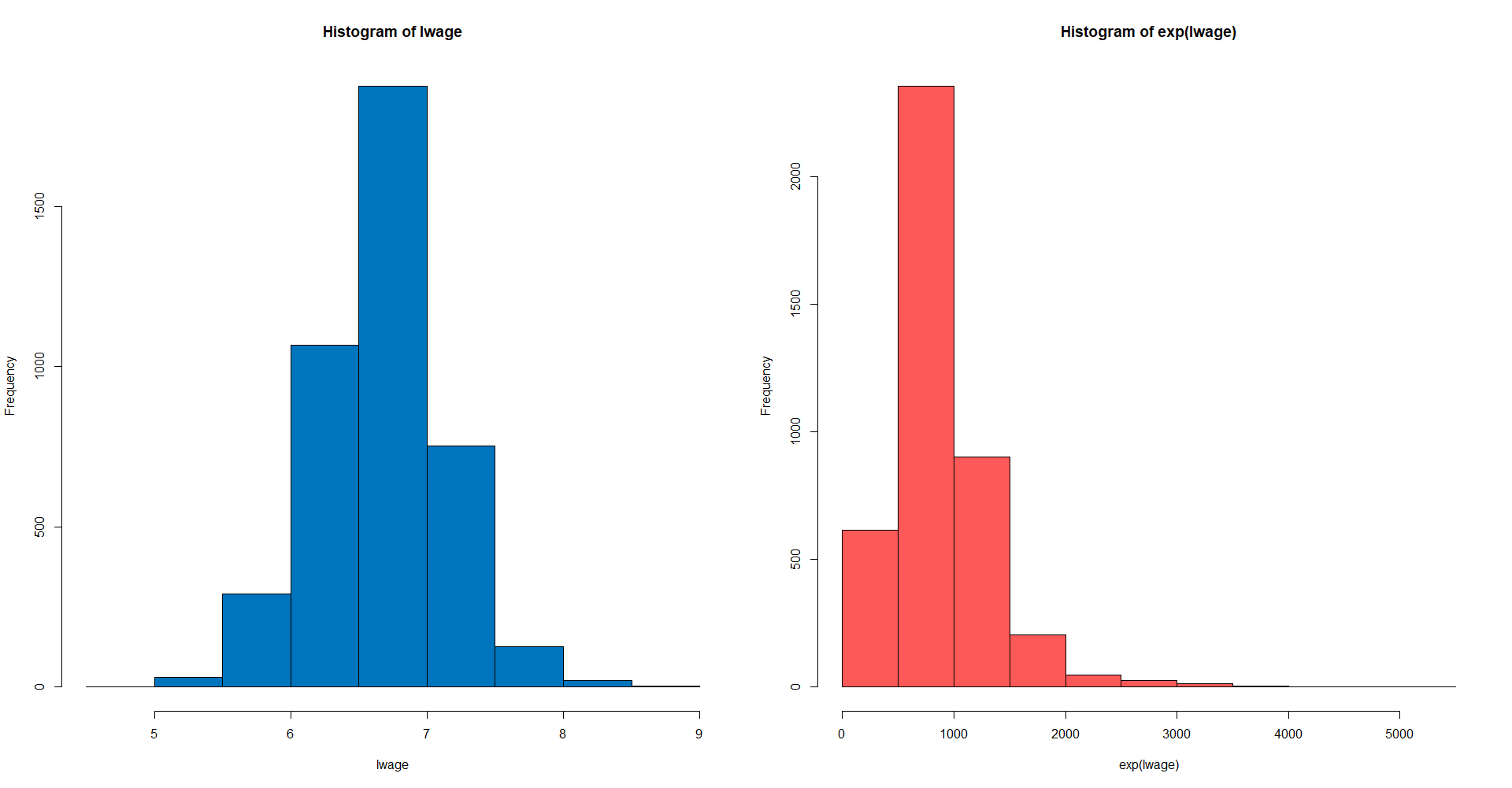
4- Compare la evolución del salario con la evolución del logaritmo neperiano del salario y ofrezca una posible explicación de por qué resulta de mayor interés utilizar el logaritmo del salario frente a la serie sin transformar

Se hace la transformación logarítmica porque esto ayuda a resolver la posible no linealidad de la variable a predecir, además que disminuye la varianza:

var(log(salario) = 0.2129935

var(salario)= 196080.4

Otra cosa es que también podemos notar que el log del salario se acerca más a una distribución normal (histograma azul),. Tal como puede ver en la siguiente figura:



5. Puntuación extra: ¿sería capaz de determinar algún valor de la variación intra-grupos y entre grupos de las siguientes variables: id, t, fem, tdum1?

Los valores de estas variables son recogidos en la tabla 1, tanto para la variación intra-grupos como para la extra-grupos.

6. Puntuación extra: Un investigador considera que la base de datos disponible resulta incompleta, porque no dispone de una variable tan importante como la edad del individuo. Ofrezca al investigador una posible solución.

La edad para este problema no es necesaria ya que sus efectos están fuertemente ligados a la experiencia (a mayor edad mayor experiencia) y por lo tanto de agregarla solo estaremos .

MODELOS LINEALES DE DATOS DE PANEL:

7. Estime el modelo de efectos individuales por los siguientes métodos que se mencionan a continuación y determine en una tabla la comparativa de coeficientes, desviaciones típicas asociadas y valores de la función de verosimilitud y del R2.

* Estimación por MCO
* Estimación por MCO con desviaciones típicas robustas a heterogeneidad
* Estimación según el modelo de agrupación (PA)
* Estimación según Primeras Diferencias (FD)
* Estimación según el Modelo de Efectos Aleatorios (RE)
* Estimación según el Modelo de Efectos Fijos (FE)

La elección de las variables de nuestro modelo será estimado en base a lo observado en la matriz de correlaciones lineales por lo que nuestro modelo será

Con las variables

La tabla comparativa se muestra a continuación:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ==================================================================== | | | | | |
|  | Fixed1 | Random1 | Pool1 | Fdif1 | ols |
| ------------------------------------------------------------------------------------ | | | | |  |
| Exp | 0.113\*\*\* | 0.082\*\*\* | 0.040\*\*\* |  | 0.040\*\*\* |
|  | -0.002 | -0.003 | -0.002 |  | -0.002 |
| exp2 | -0.0004\*\*\* | -0.001\*\*\* | -0.001\*\*\* | 0.002\*\*\* | -0.001\*\*\* |
| South | -0.002 | -0.017 | -0.056\*\*\* | -0.018 | -0.056\*\*\* |
|  | -0.034 | -0.027 | -0.013 | -0.048 | -0.013 |
| Occ | -0.021 | -0.050\*\* | -0.140\*\*\* | -0.028 | -0.140\*\*\* |
|  | -0.014 | -0.017 | -0.015 | -0.014 | -0.015 |
| Smsa | -0.042\* | -0.014 | 0.152\*\*\* | -0.060\* | 0.152\*\*\* |
|  | -0.019 | -0.02 | -0.012 | -0.025 | -0.012 |
| Ms | -0.03 | -0.075\*\* | 0.048\* | -0.058\* | 0.048\* |
|  | -0.019 | -0.023 | -0.021 | -0.024 | -0.021 |
| Wks | 0.001 | 0.001 | 0.004\*\*\* | 0 | 0.004\*\*\* |
|  | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 |
| ind | 0.019 | 0.004 | 0.047\*\*\* | 0.022 | 0.047\*\*\* |
|  | -0.015 | -0.017 | -0.012 | -0.017 | -0.012 |
| union | 0.033\* | 0.063\*\*\* | 0.093\*\*\* | 0.019 | 0.093\*\*\* |
|  | -0.015 | -0.017 | -0.013 | -0.016 | -0.013 |
| (Intercept) |  | 4.264\*\*\* | 5.251\*\*\* |  | 5.251\*\*\* |
|  |  | -0.098 | -0.071 |  | -0.071 |
| ed |  | 0.100\*\*\* | 0.057\*\*\* |  | 0.057\*\*\* |
|  |  | -0.006 | -0.003 |  | -0.003 |
| blk |  | -0.210\*\*\* | -0.167\*\*\* |  | -0.167\*\*\* |
|  |  | -0.058 | -0.022 |  | -0.022 |
| fem |  | -0.339\*\*\* | -0.368\*\*\* |  | -0.368\*\*\* |
|  |  | -0.051 | -0.025 |  | -0.025 |
| ------------------------------------------------------------------------------------------------------------ | | | | |  |
| R-squared | 0.658 | 0.39 | 0.429 | 0.001 | 0.429 |
| adj. R-squared | 0.6 | 0.388 | 0.427 | -0.001 | 0.427 |
| F (omnibus) | 761.751 | 221.229 | 259.544 | -40.272 | 259.544 |
| p-val (omnibus) | 0 | 0 | 0 | 1 | 0 |
| N | 4165 | 4165 | 4165 | 3570 | 4165 |
| sigma |  |  |  |  | 0.349 |
| ================================================================================ | | | | | |

Tabla 2

Como se puede observar existen marcadas diferencias entre los modelos en la elección de parámetros y en los resultados de los estadísticos, si tomamos como referencia el r-squared el cual nos indica qué tan bien se ajusta nuestro modelo a los datos. Siendo el mejor modelo el de efectos fijos (fixed effect) con diferencias claras sobre el resto de los demás.

8. Argumente con los resultados obtenidos hasta el momento qué proceso de estimación se adecúa mejor a los datos y la estructura del modelo. Pista: Para argumentarlo utilice la teoría explicada, confirme de alguna forma si los supuestos que subyacen al modelo se cumplen, y utilice instrumentos como el contraste de Hausman.

De la tabla 2, observando el valor del r-squared se podrá seleccionar el modelo de efectos fijos, sin embargo se aplicaron test estadísticos para garantizar la consistencia de nuestra elección:

Para Comprobar si se debe elegir entre un modelo de efecto aleatorios y un modelo de regresión lineal (agrupada) utilizamos el Breusch-Pagan Lagrange multiplier (LM): El cual no arroja que un modelo de efectos aleatorios es apropiado para nuestros datos.

Luego de comprobar que podemos elegir un modelo RE, aplicamos el Test de Haussman, el cual es básicamente un test que indica si lo errores están correlados con los regresores, en el test realizado se observa un p-valor menor al 0.05, lo que nos indica que debemos rechazar la hipótesis nula (los errores no están correlados con los regresores) y decidimos entonces tomar como modelo al FE.

Los resultados de los test aplicados:

Contraste de Breusch-Pagan –

* Hipótesis nula: se trata de un modelo agrupado, frente a la alternativa de efectos aleatorios.´

Normal = 59.136, p-value < 2.2e-16, se rechaza la hipótesis nula

Contraste de Hausman –

* Hipótesis nula: Los estimadores de los errores no estan correlados

chisq = 5075.3, df = 9, p-value < 2.2e-16, se rechaza la hipótesis nula´

Cabe destacar que las variables ed, blk y fem desaparecen al emplear la estimación´ de efectos fijos al no presentar variabilidad en el tiempo.

9. En el modelo elegido, interprete el resultado de las siguientes variables respecto al logaritmo neperiano del salario y respecto al salario:

* Un año más de edad

Un año más de edad implicaría una año más de experiencia por lo que observando los coeficientes de la tabla (3) para el modelo de efecto fijos tenemos que con cada año más de experiencia el salario se incrementaría un factor 11.3 más con respecto al logaritmo neperiano del salario del año anterior y un 87.94 con respecto al salario, adicionalmente se reduciría en un -0.004 y un 0.02 del neperiano y del salario respectivamente, esto por la relación cuadrática de la experiencia con el salario.

* Una semana más trabajada

Una semana más de trabajo implicaría un incremento con respecto al logaritmo neperiano de 0.008 y con respecto al salario de -0.64, lo que quiere decirnos es que a mayor semanas trabajadas no se traduce en un incremento del salario.

* Un año más de educación

Esta variable al ser invariante en el tiempo su influencia se elimina del modelo y sus características son absorbidas por el efecto fijo o individual, sin embargo si observamos otros modelos podemos notar que el coeficiente de educación es positivo lo que indicaría que incrementos en la educación podría significar u incrementos en el salario.

* Un año más de experiencia

Mismo análisis que un año más de edad

* Ser hombre frente a ser mujer

Esta variable a ser invariante en el tiempo su influencia elimina del modelo( absorbida por el efecto fijo o individual), sin embargo parece ser (según los coeficiente de los otros modelo) que la relación es que el ser mujer influye negativamente en el salario.

10. Obtenga los valores predichos para el salario y compárelos con los verdaderos valores

El valor predicho es el resultado de evaluar la formula de regresión con efectos fijos, es decir multiplicando los coeficientes por las variables y sumando los efecto fijos individuales con los residuos

Un extracto de los valores predicho se muestra a continuación:

|  |  |
| --- | --- |
| Valor Real | Valor Predicho |
|  |  |
| 5.56068 | 4.954643 |
| 5.72031 | 3.006165 |
| 5.99645 | 5.362778 |
| 5.99645 | 3.041746 |
| 6.06146 | 5.623869 |
| 6.17379 | 4.424722 |
| 6.24417 | 4.718035 |
| 6.16331 | 3.634632 |
| 6.21461 | 6.07519 |
| 6.2634 | 5.83794 |
| 6.54391 | 4.193416 |
| 6.69703 | 4.401913 |
| 6.79122 | 4.464731 |
| 6.81564 | 5.377963 |
| 5.65249 | 4.512618 |
| 6.43615 | 3.847641 |

Tabla 3

Graficamente:

* Por Individuos

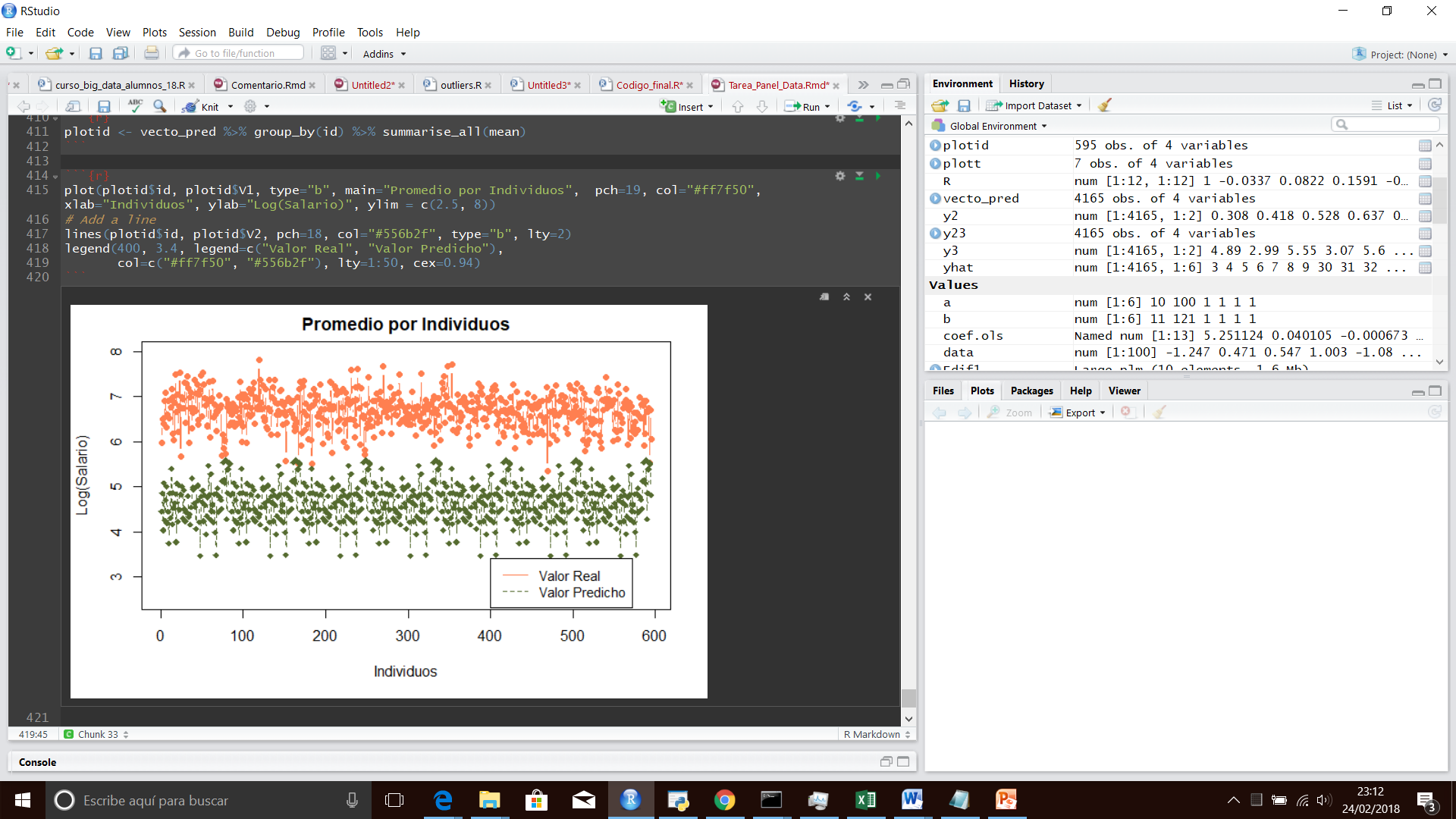


figura 2

* En el tiempo

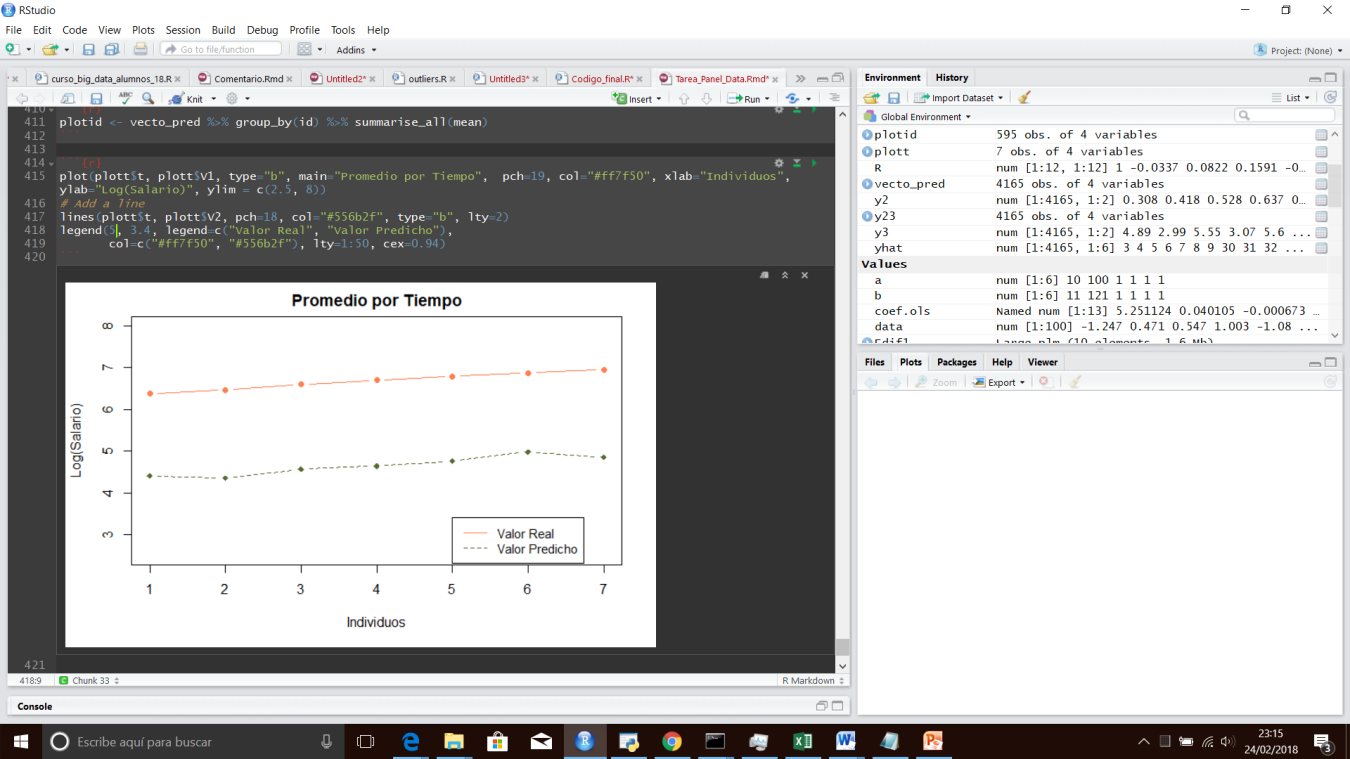


figura 3

11. Realice una predicción del salario de los siguientes individuos para el modelo elegido e indique quién cobrará más salario:

• Un varón caucásico, casado con dos hijos con una educación equivalente a 15 años, con 20 años de experiencia, 47 semanas trabajadas, informático, trabajando en el sector bancario, residiendo a las afueras de una ciudad del estado de Minnesota.

• Una mujer afroamericana, soltera, con una educación equivalente a 13 años, con 9 años de experiencia, 45 semanas trabajadas, administrativa, trabajando en una empresa que fabrica calzado, residiendo en el centro de una ciudad del estado de Luisiana.

• Un varón afroamericano, casado con un hijo, con una educación equivalente a 16 años, con 31 años de experiencia, 49 semanas trabajadas, electricista en una gran empresa que pertenece a la industria del papel, residiendo en el centro de Chicago.

• Un mujer asiática, casada, sin hijos, con una educación equivalente de 17 años, con 16 años de experiencia, 50 semanas trabajadas, ejecutiva de una empresa del sector de la construcción, que reside a las afueras de Tampa

Cabe destacar que en este caso para predecir con efectos fijos con nuevos individuos, debemos obtener los efectos individuales de estos nuevos individuos y para ello estos se deben regresar linealmente, lo que hace que para este caso el modelo de efectos fijos sea similara un modelo de OLS (mínimos cuadrados ordinarios) por lo que predeciremos usando esta última (OLS).

La tabla con los valores para los distintos individuos y la predicción se muestra a continuación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Individuo1 | Individuo2 | Individo3 | Individuo4 |
| Blk | 0 | 1 | 1 | 0 |
| Smsa | 0 | 1 | 1 | 0 |
| Occ | 0 | 0 | 1 | 0 |
| Ed | 15 | 13 | 16 | 17 |
| Exp | 20 | 9 | 31 | 16 |
| exp2 | 400 | 81 | 961 | 256 |
| Ind | 0 | 1 | 1 | 0 |
| South | 0 | 1 | 0 | 1 |
| Ms | 1 | 0 | 1 | 1 |
| Wks | 47 | 45 | 49 | 50 |
| lwage\_predicho | **6.97** | **6.18** | **6.99** | **6.61** |
| Wage | **1064.223** | **482.992** | **1085.721** | **742.483** |

Tabla 4

De la tabla 4, se nota que el individuo que cobrara mayor salario es **el individuo 3**, sin embargo la diferencia es muy poca con respecto al **individuo 1**, asumimos que es el efecto de la exp2, que reza que la experiencia llega un punto donde el salario no aumenta significativamente. En resumen los individuos con más experiencia son los que más ganan y no son necesariamente lo que tienen más años de educación.

12. Puntuación extra: Ciertas teorías económicas consideran que la educación puede tener un efecto similar en el salario que la experiencia, esto es un efecto no lineal. Genere una variable que refleje la educación al cuadrado (ed2) y confirme esta teoría con el modelo

Del modelo seleccionado (efecto fijos) tenemos que la educación al ser invariante en el tiempo es eliminada del modelo, por lo que lo mismo sucede a la educación al cuadrado, y al hacer la predicciones son resultados no difieren de nuestro anterior modelo, por lo que no notamos la diferencia. Confirmamos nuestra elección del modelo de Efecto fijos sin la variable educación al cuadrado.

EXTENSIÓN A LOS MODELOS LINEALES VARIABLES INSTRUMENTALES

Considere la opción de mejorar el modelo seleccionado en el apartado anterior a través de los métodos asociados a Variables Instrumentales. Considere también que las variables exp, exp2, wk, ed y ms son endógenas, mientras que las variables occ, south, smsa, fem y blk son exógenas.

13. Con esta información, estime para un modelo de efectos individuales en la versión de efectos fijos (FE) y otro de efectos aleatorios (RE):

• Con el método de variables instrumentales, utilizando las exógenas como instrumentos de las endógenas

• Con el método de estimación de Hausman-Taylor.

Presente una tabla comparativa de los resultados en un formato idéntico al solicitado en el apartado 7 y argumente que modelo le parece más razonable.

|  |  |  |  |
| --- | --- | --- | --- |
| ================= | ============= | =========== | =============== |
|  | HT | Random2 | Fixed2 |
| ----------------- | ------------- | ----------- | --------------- |
| (Intercept) | 2.978\*\*\* | -2.202 |  |
| Occ | -0.021 |  |  |
| South | 0.029 |  |  |
| Smsa | -0.036 |  |  |
| Ind | 0.018 |  | 0.02 |
| Exp | 0.113\*\*\* | 0.101 | 0.114\*\*\* |
| exp2 | -0.000\*\*\* | 0.001 | -0.000\*\*\* |
| Wks | 0.001 | 0 | 0.001 |
| Ms | -0.03 | -0.027 | -0.032 |
| Unión | 0.032 |  |  |
| Fem | -0.129 |  |  |
| Ed | 0.132\*\*\* | 0.486 |  |
| Blk | -0.300\*\*\* |  |  |
| ----------------- | ------------- | ----------- | --------------- |
| R-squared | 0.146 | 0.606 | 0.657 |
| adj. R-squared | 0.144 | 0.605 | 0.599 |
| F (omnibus) | 1542.884 | 448.077 | 1366.164 |
| p-val (omnibus) | 0 | 0 | 0 |
| N | 4165 | 4165 | 4165 |
| ================= | ============= | =========== | =============== |

Tabla 5

En la tabla 5 se encuentran los resultados de aplicar variables instrumentales de los modelo con variables instrumentales el mejor sigue siendo el modelo de efecto fijos (columna Fixed2), al observar solo el r-squared, no obstante notamos que existen gran similitud de los coeficientes con el estimador Hausman-Taylor lo que podría indicar que nuestro modelo con VI es válido.

ANEXO

Codigo R

#DESCRIPCION DE LOS DATOS DE PANEL: INTRODUCCION

#Cargar el panel de datos:

library(foreign)

library(plm)

Panel1 <- read.csv("psid76\_82.csv",header = TRUE, sep = ";", dec=".")

#PARTE DESCRIPTIVA:

#1. Determine qué variables de la base de datos son invariantes en el tiempo y argumente

library(dplyr)

Panel1 %>% group\_by(id) %>% summarise\_all(var)

t(Panel1 %>% group\_by(id) %>% summarise\_all(funs(mean)) %>% summarise\_all(funs(var)))

t(Panel1 %>% group\_by(t) %>% summarise\_all(funs(mean)) %>% summarise\_all(funs(var)))

attach(Panel1)

Panel2 <- data.frame(Panel1, index = c(id, t), all=TRUE)

plot((union)~t, main="Variabilidad de Blk en tiempo", xlim=c(0,10),data=Panel2)

plot(blk~t, main="Variabilidad de Blk en tiempo", xlim=c(0,10),data=Panel2)

plot(exp~t, main="Variabilidad de la Experiencia en tiempo", xlim=c(0,10),data=Panel2)

library(gplots)

plotmeans(exp~t, main="Variabilidad de la Experiencia en tiempo", data=Panel2)

plotmeans(south~t, main="Comportamiento de south en media en el tiempo", data=Panel2)

plot(exp~t, main="Variabilidad de la Experiencia en tiempo", xlim=c(0,10),data=Panel2)

plot(south~t, main="Comportamiento de south en el tiempo", xlim=c(0,10),data=Panel2)

for (name in names(Panel2[, -c(1,2)])){

par(mfrow=c(1,2))

print(name)

x = Panel2$t

y = unlist(Panel2[name])

p <- plot(y~x, main= paste("Comportamiento de ", name) , xlim=c(0,10))

print(p)

p2<- plotmeans(y~x, main= paste("Comportamiento de ", name) )

print(p2)

}

#Gráficamente las variables invariantes en el tiempo son: \*\*fem, blk y ed\*\*

#Overall variation:

library(dplyr)

Panel2 %>%

group\_by(id,t) %>%

summarise\_all(funs((unique(exp))))

#2. Determine qué variables de la base de datos son invariantes en el individuo y argumente

#Gráfico para observar invariabilidad en el individuo

for (name in names(Panel2[, -c(1)])){

par(mfrow=c(1,2))

print(name)

x = Panel2$id

y = unlist(Panel2[name])

suppressWarnings(p <- plot(y~x, main= paste("Comportamiento de ", name) , xlim=c(0,10)))

print(p)

suppressWarnings((p2<- plotmeans(y~x, main= paste("Comportamiento de ", name), subset = (0:98) )))

print(p2)

}

#Overall variation:

library(dplyr)

Panel2 %>%

group\_by(id) %>%

summarise\_all(funs(sum))%>%sapply( function(x) length(unique(x)) )%>%as.data.frame()

#Candidatas \* t, t-num \*

#Suponga que pretende estudiar la relación que existe entre las variables disponibles en la base de datos y el salario. Determine qué relación existe a través de una matriz de correlaciones.

library(corrplot)

R <- cor(Panel1[,])

corrplot.mixed(R, lower = "number", upper = "color")

corrplot(R, method="shade", # visualisation method

shade.col=NA, # colour of shade line

tl.col="black", # colour of text label

tl.srt=45, # text label rotation

addCoef.col="black", # colour of coefficients

order="AOE" # ordering method

)

#\*\*Del gráfico correlaciones podemos observar que existe una correlación positiva entre el salario y la variables: \*educación (ed), estado civil (ms), el vivir en una zona metropolitana(smsa), y la experiencia (exp)\*, esto implica una relación lineal (al aumentar 1 aumenta la otra), y tiene una correlación negativa con:\* la ocupación (occ), si vives ó no en el sur de EEUU (south), si el individuo es mujer (fem) y si el individuo es afroamericano (blk) \* \*\*

#4. Compare la evolución del salario con la evolución del logaritmo neperiano del salario y ofrezca una posible explicación de por qué resulta de mayor interés utilizar el logaritmo del salario frente a la serie sin transformar

#se hacen gráficos para comparar los salarios

plotmeans(lwage~t, main= "Comportamiento de Log Salario", data = Panel2)

plotmeans(exp(lwage)~t, main= "Comportamiento de Salario", data = Panel2)

par(mfrow=c(1,2))

hist(lwage)

hist(exp(lwage))

plot(density(lwage))

lines(density(exp(lwage)), col="red")

plot(lwage, main= "Comportamiento de Log Salario", data = Panel2, type = "l")

plot(exp(lwage), main= "Comportamiento de Salario", data = Panel2, type = "l")

plot(Panel1[Panel1$id ==1, ]$lwage, x= Panel1[Panel1$id ==1, ]$t, type = "l")

plot(exp(Panel1[Panel1$id ==1, ]$lwage), x= Panel1[Panel1$id ==1, ]$t, type = "l")

#Overall variation:

lwage.aov <- aov(lwage ~ 1, data=Panel1)

summary(lwage.aov)

#Overall variation:

wage.aov <- aov(exp(lwage) ~ 1, data=Panel1)

summary(wage.aov)

#MODELOS LINEALES DE DATOS DE PANEL:

# 7. Estime el modelo de efectos individuales por los siguientes métodos que se mencionan a #continuación y determine en una tabla la comparativa de coeficientes, desviaciones típicas asociadas y #valores de la función de verosimilitud y del R2.

# • Estimación por MCO

# • Estimación por MCO con desviaciones típicas robustas a heterogeneidad

# • Estimación según el modelo de agrupación (PA)

# • Estimación según Primeras Diferencias (FD)

# • Estimación según el Modelo de Efectos Aleatorios (RE)

# • Estimación según el Modelo de Efectos Fijos (FE)

#Se ejecutan los distintos metodos y se observa el r^2 y los p-valores

library(sandwich)

library(lmtest)

ols <- lm(lwage~exp+exp2+ed+ blk + fem +south+ occ+ ms+ smsa+ wks+ ind+union , data=Panel1)

summary(ols)

coeftest(ols, vcov = sandwich)

#Robusta

olsr <- ols

SE\_robust <- coeftest(olsr, vcov = sandwich)

olsr2.sum <- summary(olsr)

olsr2.sum$coefficients <- unclass(SE\_robust)

olsr2.sum

# Versión Agrupada

Pool1 <- plm(lwage ~ exp+exp2+ed+ blk + fem +south+ occ+ smsa +ms+ wks+ind+union, data=Panel1, index=c("id", "t"), model="pooling")

summary(Pool1)

#Version robusta: version Sandwich aplicable -> pvcovHC

coeftest(Pool1, vcov = pvcovHC)

## Regresion con efectos fijos (FE)

Fixed1 <- plm((lwage) ~ exp+exp2+ed+ blk + fem +south+ occ+ smsa +ms +wks+ ind+ union, data=Panel1, index=c("id", "t"), model="within")

summary(Fixed1)

coeftest(Fixed1, vcov = pvcovHC)

#fixef(Fixed1)

#Time Fixed

Fixed.time <- plm(lwage ~ exp+exp2+ed+ blk + fem +south+ occ+ smsa +ms, data=Panel1, index=c("id", "t"), model="between")

summary(Fixed.time)

coeftest(Fixed1, vcov = pvcovHC)

plmtest(Pool1, type = ("bp"))

## Regresion en primeras diferencias (FD)

Fdif1 <- plm(lwage ~ exp+exp2+ed+ blk + fem +south+ occ+ smsa +ms+ wks+ind+union, data=Panel1, index=c("id", "t"), model="fd")

summary(Fdif1)

coeftest(Fdif1, vcov = pvcovHC)

# Regresion con efectos aleatorios (RE)

Random1 <- plm(lwage ~ exp+exp2+ed+ blk + fem +south+ occ+ smsa +ms+wks+ ind+union, data=Panel1, index=c("id", "t"), model="random")

summary(Random1)

coeftest(Random1, vcov = pvcovHC)

# Test de Haussman:

phtest(Fixed1,Random1)

phtest(Random1, Fixed1)

pFtest(Fixed1, ols)

plmtest(Pool1)

## el resumen de los modelos

library(tonymisc)

mtable(Fixed1, Random1, Pool1, Fdif1, ols)

#En el modelo elegido, interprete el resultado de las siguientes variables respecto al logaritmo neperiano del salario y respecto al salario:

#• Un año más de edad

#• Una semana más trabajada

#• Un año más de educación

#• Un año más de experiencia

#3• Ser hombre frente a ser mujer

#Si se elige efecto fijo debo sumar estos valores al resultado de la funcion predict

Fixed1$coefficients

lwpa <- Pool1$model[,1] - Pool1$residuals

lwfe <- Fixed1$model[,1] - Fixed1$residuals

lwfd <- Fdif1$model[,1] - Fdif1$residuals

lwre <- Random1$model[,1] - Random1$residuals

## ó sum(x[1,]\*Fixed1$coefficients) + fixef(Fixed1)[1]

#Las preguntas 8-9-10 son interpretaciones del modelo (texto)

vecto\_pred <- cbind(Panel1$lwage, predict(Fixed1, data = Panel1) + fixef(Fixed1))

vecto\_pred <- as.data.frame(vecto\_pred)

vecto\_pred["id"] <- Panel1$id

vecto\_pred["t"] <- Panel1$t

plotid <- vecto\_pred %>% group\_by(id) %>% summarise\_all(mean)

plott <- vecto\_pred %>% group\_by(t) %>% summarise\_all(mean)

plot(plott$t, plott$V1, type="b", main="Promedio por Tiempo", pch=19, col="#ff7f50", xlab="Individuos", ylab="Log(Salario)", ylim = c(2.5, 8))

# Add a line

lines(plott$t, plott$V2, pch=18, col="#556b2f", type="b", lty=2)

legend(5, 3.4, legend=c("Valor Real", "Valor Predicho"),

col=c("#ff7f50", "#556b2f"), lty=1:50, cex=0.94)

##Pregunta 12

Panel3 <- Panel1

Panel3["ed2"] <- Panel3$ed^2

Fixed4e <- plm((lwage) ~ exp+exp2+ed+ed2+ blk + fem +south+ occ+ smsa +ms +wks+ ind+ union, data=Panel3, index=c("id", "t"), model="within")

summary(Fixed4e)

coeftest(Fixed1, vcov = pvcovHC)

mtable(Fixed4e)

vecto\_pred2 <- cbind(Panel3$lwage, predict(Fixed4e, data = Panel3) + fixef(Fixed4e))

vecto\_pred2 <- as.data.frame(vecto\_pred2)

#### EXTENSIÓN A LOS MODELOS LINEALES VARIABLES INSTRUMENTALES

#Considere la opción de mejorar el modelo seleccionado en el apartado anterior a través de los métodos #asociados a Variables Instrumentales. Considere también que las variables exp, exp2, wk, ed y ms son #endógenas, mientras que las variables occ, south, smsa, fem y blk son exógenas.

#13. Con esta información, estime para un modelo de efectos individuales en la version de efectos fijos (FE) #y otro de efectos aleatorios (RE):

#Considere también que las variables exp, exp2, wk, ed y ms son endógenas(correladas con el error), mientras #que las variables occ, south, smsa, fem y blk son exógenas( no correladas con el error).

#Modelos de Variables Instrumentales:

Fixed2 <- plm(lwage ~ exp +exp2 +wks+ed+ind + ms | occ + south + smsa + fem + blk , data=Panel1, index=c("id", "t"), model="within")

summary(Fixed2)

Random2 <- plm(lwage ~ exp +exp2 +wks +ed+ ms | occ + south + smsa + fem + blk, data=Panel1, index=c("id", "t"), model="random")

summary(Random2)

Random3 <- plm(lwage ~ occ+south+smsa+ind+exp+exp2+wks+ms+union+fem+ed+blk | occ+south+smsa+blk+fem, data=Panel1, index=c("id", "t"), model="ht")

summary(Random3)

mtable(Random3, Random2, Fixed2)