

Capítulo 13

Econometría II, Nelson Muriel

Andres de Alba, Cristian Cárdenas, Anderson Martínez, Gissela Rodríguez

abril 2018

ejercicio 1

```
library(plm)
library(car)
library(stats)
library(stargazer)
library(stats)
library(dplyr)
library(lmtest)
library(sandwich)

load("~/Dropbox/Econometria/datos/wooldridge6/fertil1.RData")
fertil1 <- data
rm('data', 'desc')
mod13_1 <- lm(data = fertil1, kids ~ educ + age + I(age^2) + black + east +
              northcen + west + farm + othrural + town +
              smcity + y74 + y76 + y78 + y80 + y82 + y84 )
car::linearHypothesis(mod13_1, c('farm', 'othrural', 'town', 'smcity'))

## Linear hypothesis test
##
## Hypothesis:
## farm = 0
## othrural = 0
## town = 0
## smcity = 0
##
## Model 1: restricted model
## Model 2: kids ~ educ + age + I(age^2) + black + east + northcen + west +
##          farm + othrural + town + smcity + y74 + y76 + y78 + y80 +
##          y82 + y84
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1115 2697.1
## 2    1111 2685.9   4    11.206 1.1588 0.3275

i) El estadístico F de significación conjunta para las variables de condiciones del entorno a la
edad de 16 años tiene un p-valor de 0.33 por lo cual no se puede rechazar la hipótesis de que
todas valen cero.

linearHypothesis(mod13_1, c('east', 'northcen', 'west'))

## Linear hypothesis test
##
## Hypothesis:
## east = 0
```

```
## northcen = 0
## west = 0
##
## Model 1: restricted model
## Model 2: kids ~ educ + age + I(age^2) + black + east + northcen + west +
##          farm + othrural + town + smcity + y74 + y76 + y78 + y80 +
##          y82 + y84
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1114 2707.7
## 2    1111 2685.9  3    21.843 3.0117 0.02926 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ii) Las variables de región del país a los 16 años tienen un p-valor para la prueba F de 0.029 por lo tanto si son significantes a un nivel del 97%.

```
u <- residuals(mod13_1)
test_model <- lm(I(u^2) ~ y74 + y76 + y78 + y80 + y82 + y84, data = fertil1 )
linearHypothesis(test_model, c('y74', 'y76', 'y78', 'y80', 'y82', 'y84'))
```

```
## Linear hypothesis test
##
## Hypothesis:
## y74 = 0
## y76 = 0
## y78 = 0
## y80 = 0
## y82 = 0
## y84 = 0
##
## Model 1: restricted model
## Model 2: I(u^2) ~ y74 + y76 + y78 + y80 + y82 + y84
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1128 12270
## 2    1122 12082  6    187.68 2.9048 0.008159 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

iii) Existe suficiente evidencia estadística para aceptar heterocedasticidad en los errores.

```
mod13_1 <- update(mod13_1, . ~ . + I(y74*educ) + I(y76*educ) + I(y78*educ)
+ I(y80*educ) + I(y82*educ) + I(y84*educ))
summary(mod13_1, signif.stars = true, digits = 3)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -8.477302131 3.126360276 -2.7115564 6.800684e-03
## educ        -0.022515231 0.053618261 -0.4199172 6.746276e-01
## age          0.507465834 0.138921522  3.6528957 2.714583e-04
## I(age^2)     -0.005524971 0.001570009 -3.5190710 4.506791e-04
## black        1.074054877 0.173701015  6.1833541 8.817586e-10
## east         0.206055930 0.133142996  1.5476288 1.219980e-01
## northcen     0.348286721 0.121098738  2.8760557 4.104119e-03
## west         0.177122131 0.167452303  1.0577468 2.904021e-01
## farm        -0.072162198 0.147507965 -0.4892088 6.247910e-01
```

```
## othrural      -0.191153859  0.175934037 -1.0865087  2.774910e-01
## town          0.088229487  0.124535602   0.7084680  4.788041e-01
## smcity        0.205357592  0.160210393   1.2817994  2.001819e-01
## y74           0.946914919  0.904158933   1.0472881  2.951957e-01
## y76           1.019963166  0.882033919   1.1563764  2.477772e-01
## y78           1.805985151  0.951865866   1.8973106  5.804746e-02
## y80           1.114183245  0.897601029   1.2412901  2.147620e-01
## y82           1.199807328  0.876289067   1.3691913  1.712177e-01
## y84           1.671260773  0.899050053   1.8589185  6.330439e-02
## I(y74 * educ) -0.056424832  0.072560654  -0.7776230  4.369578e-01
## I(y76 * educ) -0.092099696  0.070874767  -1.2994709  1.940534e-01
## I(y78 * educ) -0.152387255  0.075281863  -2.0242227  4.318745e-02
## I(y80 * educ) -0.097904853  0.070452286  -1.3896618  1.649115e-01
## I(y82 * educ) -0.138944737  0.068371249  -2.0322100  4.237119e-02
## I(y84 * educ) -0.176096976  0.069914918  -2.5187325  1.191811e-02

car::linearHypothesis(mod13_1, c("I(y74 * educ)", "I(y76 * educ)", "I(y78 * educ)", "I(y80 * educ)",
                                "I(y82 * educ)", "I(y84 * educ)"))

## Linear hypothesis test
##
## Hypothesis:
## I(y74 * educ) = 0
## I(y76 * educ) = 0
## I(y78 * educ) = 0
## I(y80 * educ) = 0
## I(y82 * educ) = 0
## I(y84 * educ) = 0
##
## Model 1: restricted model
## Model 2: kids ~ educ + age + I(age^2) + black + east + northcen + west +
##      farm + othrural + town + smcity + y74 + y76 + y78 + y80 +
##      y82 + y84 + I(y74 * educ) + I(y76 * educ) + I(y78 * educ) +
##      I(y80 * educ) + I(y82 * educ) + I(y84 * educ)
##
##      Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1      1111 2685.9
## 2      1105 2664.4   6    21.464  1.4836 0.1803
```

- iv) Estos terminos representan la propensión del cambio de la educación en cada año. pero no son conjuntamente significativos.

Ejercicio 2

```
load("~/Dropbox/Econometria/datos/wooldridge6/cps78_85.RData")
cps78_85 <- data
rm('data', 'desc')
mod13_2 <- lm(data = cps78_85, lwage ~ y85 + educ + y85educ + exper + expersq
              + union + female + y85fem)
summary(mod13_2)$coefficients

##              Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 0.4589328832 9.344850e-02  4.9110781 1.046281e-06
## y85         0.1178062179 1.237817e-01  0.9517254 3.414502e-01
```

```
## educ      0.0747209129 6.676431e-03 11.1917443 1.399530e-27
## y85educ   0.0184605323 9.354169e-03 1.9735085 4.869344e-02
## exper     0.0295843067 3.567312e-03 8.2931650 3.266999e-16
## expersq   -0.0003994278 7.753912e-05 -5.1513072 3.075711e-07
## union     0.2021318735 3.029449e-02 6.6722331 4.025626e-11
## female    -0.3167086481 3.662145e-02 -8.6481733 1.876348e-17
## y85fem    0.0850519706 5.130896e-02 1.6576435 9.768121e-02
```

i) El coeficiente *y85* en la ecuación 13.2 refleja el control de el año 1985 en la base de datos, a lo que representa la proporción de cambio del salario a través del tiempo correspondiente a 1985 cuando son hombres o en su caso la variable *female* tiene el valor 0 al igual que las demás variables binarias, en tanto a su relevancia sería pequeña debido a que sólo explicaría una parte muy pequeña de la muestra.

```
mod13_22 <- lm(data=cps78_85, lwage~ y85 + educ + I(y85*I(educ-12)) + exper
+ expersq + union + female + y85fem)
summary(mod13_22)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.4589328832	9.344850e-02	4.911078	1.046281e-06
## y85	0.3393326055	3.400990e-02	9.977466	1.753665e-22
## educ	0.0747209129	6.676431e-03	11.191744	1.399530e-27
## I(y85 * I(educ - 12))	0.0184605323	9.354169e-03	1.973508	4.869344e-02
## exper	0.0295843067	3.567312e-03	8.293165	3.266999e-16
## expersq	-0.0003994278	7.753912e-05	-5.151307	3.075711e-07
## union	0.2021318735	3.029449e-02	6.672233	4.025626e-11
## female	-0.3167086481	3.662145e-02	-8.648173	1.876348e-17
## y85fem	0.0850519706	5.130896e-02	1.657644	9.768121e-02

ii) Queriendo saber cuál es el incremento porcentual en el salario de un hombre con 12 años de educación, entonces sería el cambio natural por el tiempo más el cambio porcentual al tener 12 años de educación, como cambia el intercepto en 1985 para un hombre con 12 años de educación, a lo el coeficiente estimado de *y85* con estas especificaciones es de .339 con una alta significatividad. $0.3393 \pm (1.96 * 0.03401)$

$IC_{95}(0.2726404, 0.4059596)$

El incremento nominal es de 33.93% y el intervalo de confianza al 95% es de

$IC_{95}(0.2726404, 0.4059596)$

que es 27.26% a 40.60%

iii) ecuación de modelo nominal *y85* 0.118 *se* \simeq .124 y la ecuación de modelo real

$y85 - 0.383 \text{ se } \simeq .124$

aquí lo que muestra la disminución de los salarios reales a través de 1978 a 1985. a diferencia de la nominal que hay una aumento en los salarios, a lo que se ve que a subido pero no proporcional al aumento de los precios.

```
library(dplyr)
cps78_85 <- cps78_85 %>% mutate(wage=exp(lwage))
cps78_85 <- cps78_85 %>% mutate(rwage= if_else(year == 78, wage, wage/1.65))
mod13_2r <- lm(data=cps78_85, log(rwage)~y85 + educ + y85educ + exper
+ expersq + union + female + y85fem)
summary(mod13_2r)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.4589328832	9.344850e-02	4.911078	1.046281e-06
## y85	-0.3829690700	1.237817e-01	-3.093906	2.026406e-03
## educ	0.0747209129	6.676431e-03	11.191744	1.399530e-27
## y85educ	0.0184605323	9.354169e-03	1.973508	4.869344e-02
## exper	0.0295843067	3.567312e-03	8.293165	3.266999e-16
## expersq	-0.0003994278	7.753912e-05	-5.151307	3.075711e-07
## union	0.2021318735	3.029449e-02	6.672233	4.025626e-11
## female	-0.3167086481	3.662145e-02	-8.648173	1.876348e-17
## y85fem	0.0850519706	5.130896e-02	1.657644	9.768121e-02

iv) modelo 13.2 Residual standard error: 0.4127 on 1075 degrees of freedom Multiple R-squared: 0.4262, Adjusted R-squared: 0.4219

modelo 13.2(iii) Residual standard error: 0.4127 on 1075 degrees of freedom Multiple R-squared: 0.3562, Adjusted R-squared: 0.3514

Comparando el R^2 del modelo con datos reales es de .3562 y para el nominal .4262, así se diferencian por a suma de los totales al cuadrado.

v) Para 1978 había 168 afiliaciones de los 550 trabajadores representando un 30.5% afiliados, y para 1985 con una muestra de 534, solo había 96 trabajadores afiliados, representando un 18%. haciendo la comparación en porcentaje cayeron las afiliaciones por 12.5% en 7 años

```
mod13_2u <- lm(data = cps78_85, lwage ~ y85 + educ + y85educ + exper + expersq
               + union + y85union + female + y85fem)
summary(mod13_2u)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.4588400102	9.458042e-02	4.851321416	1.407542e-06
## y85	0.1179666851	1.262826e-01	0.934148535	3.504372e-01
## educ	0.0747240557	6.697063e-03	11.157735441	1.978325e-27
## y85educ	0.0184566351	9.377761e-03	1.968128179	4.931022e-02
## exper	0.0295840011	3.569283e-03	8.288500628	3.392957e-16
## expersq	-0.0003994210	7.758224e-05	-5.148356567	3.123788e-07
## union	0.2022920534	3.908364e-02	5.175875546	2.706260e-07
## y85union	-0.0003962278	6.104037e-02	-0.006491241	9.948220e-01
## female	-0.3166908908	3.674048e-02	-8.619672833	2.371608e-17
## y85fem	0.0850056955	5.182549e-02	1.640229478	1.012503e-01

iv) $y85union$ al parecer tiene un disminución del alrededor de .0004 en la prima salarial, sin embargo al revisar su significatividad, notamos que su estadístico t es muy bajo, a lo que no se rechaza H_0 , considerando que el coeficiente del término de interacción es 0

vii) Se considera que no tiene algún inconveniente, como marca de manera explicativa, solo representa que hay una disminución en las personas que perciben este beneficio

ejercicio 3

```
library('stats')
library('car')
load("~/Dropbox/Econometria/datos/wooldridge6/kielmc.RData")
kielmc <- data
rm('data', 'desc')
mod3_1 <- lm(data = kielmc, lprice ~ y81 + ld1 + y81ld1)
summary(mod3_1)$coefficients
```

```
##           Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)  8.05846783 0.50843581 15.84952856 4.519980e-42
## y81         -0.01131009 0.80506217 -0.01404872 9.888000e-01
## ldist        0.31668901 0.05153226  6.14545179 2.391654e-09
## y81ldist     0.04818618 0.08179290  0.58912422 5.561975e-01
```

- i) Para δ_1 el signo es positivo es decir que si la vivienda está más lejana al incinerador, el precio de esta aumenta, y siendo β_1 positivo quiere decir que el incinerador fue construido lejos de las casas más valiosas.
- ii) $y81 * \log(dist)$ representa el cambio en el valor de las casas cuando el incinerado ya esta construido, dependiendo a la lejanía de la finca con el incinerador.

```
mod3_2 <- lm(data = kielmc, lprice ~ y81 + ldist + y81ldist + age + agesq
             + rooms + baths + lintst + lland + larea )
summary(mod3_2)$coefficients
```

```
##           Estimate  Std. Error    t value    Pr(>|t|)
## (Intercept)  7.673854e+00 5.015718e-01 15.29961130 9.349827e-40
## y81         -2.254466e-01 4.946914e-01 -0.45573182 6.489019e-01
## ldist        9.225605e-04 4.461676e-02  0.02067744 9.835163e-01
## y81ldist     6.246679e-02 5.027877e-02  1.24240888 2.150247e-01
## age         -8.007466e-03 1.417307e-03 -5.64977495 3.639212e-08
## agesq        3.569723e-05 8.707731e-06  4.09948702 5.293986e-05
## rooms        4.613892e-02 1.734420e-02  2.66019334 8.215847e-03
## baths        1.010478e-01 2.782239e-02  3.63188733 3.291194e-04
## lintst       -5.997574e-02 3.172174e-02 -1.89068262 5.959912e-02
## lland        9.534249e-02 2.472524e-02  3.85608031 1.401269e-04
## larea        3.507429e-01 5.194853e-02  6.75173936 7.204577e-11
```

- iii) Al realizar las demás características de valuación de las casas, se muestran que al parecer, las apreciaciones que están en función del incinerador, no son estadísticamente significativas, en realidad, el valor de las casas es más relevante por la cantidad de baños, el área y la antigüedad, que por la cercanía o la existencia del incinerador.
- iv) Es posible que entre las mismas variables exista una cercana tendencia que capte e interprete significatividad, sin embargo al meter las demás variables que explican más a detalle el valor de las casas, no existe estadísticamente una relación del precio de las casas.

Ejercicio 5

```
library('stats')
library('car')
load("~/Dropbox/Econometria/datos/wooldridge6/rental.RData")
rental <- data
rm('data', 'desc')
mod_rental <- lm(data=rental, lrent ~ y90 + lpop + lavginc + pctstu)
summary(mod_rental)$coefficients
```

```
##           Estimate  Std. Error    t value    Pr(>|t|)
## (Intercept) -0.568806911 0.534880801 -1.063427 2.896715e-01
## y90          0.262226733 0.034763227  7.543222 8.781472e-12
## lpop         0.040686311 0.022515387  1.807045 7.319968e-02
## lavginc      0.571446081 0.053098063 10.762089 1.937137e-19
## pctstu       0.005043561 0.001019183  4.948632 2.401385e-06
```

- i) Para el coeficiente de y_{90} , que es de .262 y es significativo, quiere decir que manteniendo todo fijo, el efecto del tiempo en las rentas creció en 10 años al rededor de un 26%, para $pctstu$ siendo el porcentaje de estudiantes de la población de la ciudad el durante el año escolar es positivo y además significativo, a lo que hace setido que si incremente en 1% el porcentaje de estudiantes en la ciudad, las rentas aumentarían cerca de medio por ciento.
- ii) Bajo el modelo

$$lrent_{it} = y_{90}t + lpop_{it} + lavginc_{it} + pctst_{it} + a_i + u_{it}$$

existe la parte de efectos fijos que pueden o no estar en los errores lo que puede ocasionar autocorrelación lo que provocaría invalidez en los errores y las t -s no serían correctas.

```
rental_pan <- pdata.frame(rental, index = c("city", "year"))

mod_rental_pan <- plm(data=rental_pan, lrent ~ y90 + lpop + lavginc
+ pctstu, model = 'fd')
stargazer(mod_rental_pan, mod_rental, title="Resultados", align=TRUE, type = 'latex')
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:04:55
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
## \caption{Resultados}
## \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }
## \hline
## \hline \hline
## & \multicolumn{2}{c}{\textit{Dependent variable:}} & \\
## \cline{2-3}
## \hline & \multicolumn{2}{c}{lrent} & \\
## \hline & \multicolumn{1}{c}{\textit{panel}} & & \multicolumn{1}{c}{\textit{OLS}} \\
## & \multicolumn{1}{c}{\textit{linear}} & & \multicolumn{1}{c}{\textit{}} \\
## \hline & \multicolumn{1}{c}{(1)} & & \multicolumn{1}{c}{(2)} \\
## \hline \hline
## y90 & 0.262*** & \\
## & (0.035) & \\
## & & \\
## lpop & 0.297** & 0.041* \\
## & (0.143) & (0.023) \\
## & & \\
## lavginc & 0.940*** & 0.571*** \\
## & (0.047) & (0.053) \\
## & & \\
## pctstu & 0.019*** & 0.005*** \\
## & (0.007) & (0.001) \\
## & & \\
## Constant & -0.569 & \\
## & (0.535) & \\
## & & \\
## \hline \hline
## Observations & 64 & 128 \\
## R-squared & 0.303 & 0.861 \\
## Adjusted R-squared & 0.280 & 0.857 \\
## Residual Std. Error & 0.126 (df = 123) & \end{table}
```

```
## F Statistic & \multicolumn{1}{c}{-14.580 (df = 2; 61)} & \multicolumn{1}{c}{190.922$^{***}$ (df = 4;
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{2}{r}{\textit{\$^{*}}\$p\$<$0.1; \textit{\$^{**}}\$p\$<$0.05; \textit{\$^{***}}\$p\$<$0.01} \\\
## \end{tabular}
## \end{table}
```

iii) modelo agregado *pctstu* 0.005044 , modelo con diferencia *pctstu* 0.0187151 apesar que pierde significatividad, el coeficiente aumenta alrededor de un 1.3%, al eliminar los efectos fijos, notamos una mayor fuerza en el aumento de del porcentaje de estudiantes por ciudad.

Ejercicio 4

```
library('stats')
load("~/Dropbox/Econometria/datos/wooldridge6/injury.RData")
injury <- data
rm('data','desc')
injuryky <- filter(injury,ky==1)
mod13_12ky <- lm(data=injuryky, ldurat~ afchnge + highearn + afhigh)
mod13_12kyc <- lm(data=injuryky, ldurat~ afchnge + highearn + afhigh + male + married + head +
neck + upextr + trunk + lowback + lowextr + occdis + manuf + construc)
stargazer(mod13_12ky, mod13_12kyc, title="Resultados", align=TRUE, type = 'latex')

##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:31
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
## \caption{Resultados}
## \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }
## \[-1.8ex]\hline
## \hline \[-1.8ex]
## & \multicolumn{2}{c}{\textit{Dependent variable:}} \\\
## \cline{2-3}
## \[-1.8ex] & \multicolumn{2}{c}{ldurat} \\\
## \[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} \\\
## \hline \[-1.8ex]
## afchnge & 0.008 & 0.011 \\\
## & (0.045) & (0.045) \\\
## & & \\\
## highearn & 0.256^{***} & 0.176^{***} \\\
## & (0.047) & (0.052) \\\
## & & \\\
## afhigh & 0.191^{***} & 0.231^{***} \\\
## & (0.069) & (0.070) \\\
## & & \\\
## male & & -0.098^{**} \\\
## & & (0.045) \\\
## & & \\\
## married & & 0.122^{***} \\\
## & & (0.039) \\\
## & & \\\
```



```

## head & & -0.514^{***} \\
## & & (0.129) \\
## & & \\
## neck & & 0.270^{*} \\
## & & (0.161) \\
## & & \\
## upextr & & -0.179^{*} \\
## & & (0.101) \\
## & & \\
## trunk & & 0.126 \\
## & & (0.109) \\
## & & \\
## lowback & & -0.009 \\
## & & (0.102) \\
## & & \\
## lowextr & & -0.120 \\
## & & (0.102) \\
## & & \\
## occdis & & 0.273 \\
## & & (0.211) \\
## & & \\
## manuf & & -0.161^{***} \\
## & & (0.041) \\
## & & \\
## construc & & 0.110^{**} \\
## & & (0.052) \\
## & & \\
## Constant & & 1.126^{***} & 1.246^{***} \\
## & (0.031) & (0.106) \\
## & & \\
## \hline \\[-1.8ex]
## Observations & \multicolumn{1}{c}{5,626} & \multicolumn{1}{c}{5,349} \\
## R^2 & \multicolumn{1}{c}{0.021} & \multicolumn{1}{c}{0.041} \\
## Adjusted R^2 & \multicolumn{1}{c}{0.020} & \multicolumn{1}{c}{0.039} \\
## Residual Std. Error & \multicolumn{1}{c}{1.269 (df = 5622)} & \multicolumn{1}{c}{1.251 (df = 5334)} \\
## F Statistic & \multicolumn{1}{c}{39.540^{***} (df = 3; 5622)} & \multicolumn{1}{c}{16.372^{***} (df = 3; 5334)} \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:} & \multicolumn{2}{r}{*p < 0.1; **p < 0.05; ***p < 0.01} \\
## \end{tabular}
## \end{table}

```

- i) *afhigh* 0.191 , *afhigh* 0.231, agregando las variables de control hay un aumento de alrededor de 5%, el tiempo que reciben una compensación laboral, después del aumento en el límite de ganancias por semanas que cubría por compensación laborales para las personas con mayores ganancias, es decir que es propenso a que lo que más ganan sean cubiertos por mas semanas si sufren un accidente.
- ii) reportando una R^2 de 0.0412, $R^2 Adj$: 0.03868. representando una explicación de $\log(Durat)$ de un 4.1%, a lo que se puede considerar que para explicar $\log(Durat)$ es más complicado y puede que haya terminamos que estemos dejando fuera del modelo.

```

injurym1 <- filter(injury,mi==1)
mod13_12mi <- lm(data=injurym1, ldurat~ afchnge + highearn + afhigh )
summary(mod13_12mi)$coefficients

```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	1.41273692	0.05671717	24.908454	3.974972e-115
## afchnge	0.09738082	0.08478791	1.148522	2.509337e-01
## highearn	0.16913878	0.10556762	1.602184	1.093227e-01
## afhigh	0.19199063	0.15416988	1.245319	2.132066e-01

- iii) *afhigh* 0.192 hay un aumento en comparación a los datos de KY sin embargo no son significativos al 10%, puede ser además porque la muestra de kentucky es mucho mayor a la de Michigan.

ejercicio 6

```
load("~/Dropbox/Econometria/datos/wooldrigde6/crime3.RData")
crime3 <- data

mod13_6 <- lm(data = crime3, lcrime ~ d78 + clrprc1 + clrprc2)
mod13_6aux <- lm(data = crime3, lcrime ~ d78 + clrprc1 + I(clrprc1 + clrprc2))
mod13_6iii <- lm(data = crime3, clcrime ~ cavgclr)
stargazer(mod13_6, mod13_6aux, mod13_6iii, title="Resultados", align=TRUE, type = 'latex')
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:32
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{Resultados}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} D{.}{.}{-3} }
##     \hline
##     & \multicolumn{3}{c}{\textit{Dependent variable:}} \\
##     \hline
##     & \multicolumn{2}{c}{lcrime} & \multicolumn{1}{c}{clcrime} \\
##     & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \multicolumn{1}{c}{(3)} \\
##     \hline
##     d78 & -0.055 & -0.055 & \\
##     & (0.094) & (0.094) & \\
##     & & & \\
##     clrprc1 & -0.018^{***} & -0.001 & \\
##     & (0.005) & (0.010) & \\
##     & & & \\
##     clrprc2 & -0.017^{***} & & \\
##     & (0.005) & & \\
##     & & & \\
##     I(clrprc1 + clrprc2) & & -0.017^{***} & \\
##     & & (0.005) & \\
##     & & & \\
##     cavgclr & & & -0.017^{***} \\
##     & & & (0.005) \\
##     & & & \\
##     Constant & 4.181^{***} & 4.181^{***} & 0.099 \\
##     & (0.188) & (0.188) & (0.063) \\
##     & & & \end{tabular}
```

```
## \hline \[-1.8ex]
## Observations & \multicolumn{1}{c}{106} & \multicolumn{1}{c}{106} & \multicolumn{1}{c}{53} \\\
## R2 & \multicolumn{1}{c}{0.471} & \multicolumn{1}{c}{0.471} & \multicolumn{1}{c}{0.175} \\\
## Adjusted R2 & \multicolumn{1}{c}{0.455} & \multicolumn{1}{c}{0.455} & \multicolumn{1}{c}{0.159} \\\
## Residual Std. Error & \multicolumn{1}{c}{0.455 (df = 102)} & \multicolumn{1}{c}{0.455 (df = 102)} & \\\
## F Statistic & \multicolumn{1}{c}{30.269*** (df = 3; 102)} & \multicolumn{1}{c}{30.269*** (df = 3; 102)} & \\\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{3}{c}{*p<$0.1; **p<$0.05; ***p<$0.01} \\\
## \end{tabular}
## \end{table}
```

En el modelo auxiliar vimos que $\theta_1 = 0$ por lo tanto comprobamos que $\beta_1 = \beta_2$

ejercicio 7

```
library(dplyr)
library(car)
load("~/Dropbox/Econometria/datos/wooldridge6/gpa3.RData")
gpa4 <- data
rm('data', 'desc')
mod1 <- lm(data=gpa4, trmgpa ~ spring + sat + hsperc + female + black + white + frstsem +
           tothrs + crsgpa + season)
summary(mod1)$coefficients
```

```
##              Estimate   Std. Error   t value   Pr(>|t|)
## (Intercept) -1.7528474371 0.3479048846 -5.0382950 5.943364e-07
## spring      -0.0580066228 0.0480368104 -1.2075453 2.276182e-01
## sat          0.0016984359 0.0001494168 11.3670984 1.174736e-27
## hsperc      -0.0086610383 0.0010362790 -8.3578245 3.280021e-16
## female       0.3504013324 0.0518524234  6.7576655 2.893544e-11
## black       -0.2541494852 0.1229215861 -2.0675741 3.903632e-02
## white       -0.0233146198 0.1173954218 -0.1985991 8.426324e-01
## frstsem     -0.0346584832 0.0760344786 -0.4558259 6.486524e-01
## tothrs      -0.0003389365 0.0007267233 -0.4663900 6.410772e-01
## crsgpa       1.0478654858 0.1041144013 10.0645585 2.195313e-22
## season      -0.0272903648 0.0490460388 -0.5564234 5.780941e-01
```

- i) *season* − 0.0273, representa que cuando es temporada deportiva, el promedio de los estudiantes cae .027 puntos, no es estadísticamente significativo,
- ii) si la habilidad no capturada está correlacionada con *season*, entonces estará sesgado y será inconsistente

```
library(plm)

gpa4.pan <- pdata.frame(gpa4, index = c("id", "term"))
mod1df <- plm(trmgpa ~ spring + sat + hsperc + female + black + white + frstsem
              + tothrs + crsgpa + season, data = gpa4.pan )
stargazer(mod1df, title="Resultados", align=TRUE, type = 'latex')
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:35
## % Requires LaTeX packages: dcolumn
```

```

## \begin{table}[!htbp] \centering
##   \caption{Resultados}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} }
## \ll[-1.8ex]\hline
## \hline \ll[-1.8ex]
## & \multicolumn{1}{c}{\textit{Dependent variable:}} \ll
## \cline{2-2}
## \ll[-1.8ex] & \multicolumn{1}{c}{trmgpa} \ll
## \hline \ll[-1.8ex]
##   spring & -0.237 \ll
##   & (0.206) \ll
##   & \ll
##   frstsem & 0.019 \ll
##   & (0.069) \ll
##   & \ll
##   tothrs & 0.012 \ll
##   & (0.014) \ll
##   & \ll
##   crsgpa & 1.136^{***} \ll
##   & (0.119) \ll
##   & \ll
##   season & -0.065 \ll
##   & (0.043) \ll
##   & \ll
## \hline \ll[-1.8ex]
## Observations & \multicolumn{1}{c}{732} \ll
## R\textsuperscript{2} & \multicolumn{1}{c}{0.208} \ll
## Adjusted R\textsuperscript{2} & \multicolumn{1}{c}{-0.603} \ll
## F Statistic & \multicolumn{1}{c}{19.016^{***}} (df = 5; 361) \ll
## \hline
## \hline \ll[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{*} p$<$0.1; **} p$<$0.05; ***} p$<$0.01} \ll
## \end{tabular}
## \end{table}

```

iii)Desaparecen las variables sat, hisperc, female, black, white. Ahora si el deporte está en temporada de ser en datos agregados season -0.0273 a diferenciados season -0.0645, sin embargo sigue siendo no significativo.

iv)Posiblemente, considerando que exista un proceso de elección de clases, el hecho de meter más o menos materias por semestres podría influir en el desempeño, tanto escolar, como deportivo y además en el promedio

ejercicio 8

```

load("~/Dropbox/Econometria/datos/wooldridge6/vote2.RData")
vote2 <- data
rm('data')
mod1c <- lm(data= vote2, cvote~ clnexp + clchexp + cincshr)
summary(mod1c)$coefficients

```

```

##           Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) -2.5559359  0.6309071 -4.0512078 8.078352e-05

```

```
## clinexp      -1.2915259  1.3821353 -0.9344425  3.515481e-01
## clchexp      -0.5985306  0.7113430 -0.8414093  4.014318e-01
## cincshr      0.1558681  0.0636806  2.4476550  1.551008e-02
```

- i) el intercepto dado por -2.56 es altamente significativo y la parte total de gastos de campaña para los titulares del cargo con un coeficiente de cincshr 0.156 a lo que es significativo al 1%

```
linearHypothesis(mod1c, c("clinexp", "clchexp"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## clinexp = 0
## clchexp = 0
##
## Model 1: restricted model
## Model 2: cvote ~ clinexp + clchexp + cincshr
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     155 9282.3
## 2     153 9102.3  2    179.98 1.5126 0.2236
```

- ii) No rechaza H0, es decir que conjuntamente no son significativas, al igual que de manera individual, con un p-value de 0.2236

```
mod2c <- lm(data= vote2, cvote~cincshr)
summary(mod2c)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -2.6811182  0.62527726 -4.287887 3.159612e-05
## cincshr      0.2175863  0.03209341  6.779782 2.378697e-10
```

- iii) Un aumento del 10 por ciento en el total del gasto de campaña significaría un aumento de 2.18% en los votos al candidato en cargo

```
vote2r <- filter(vote2, rptchall==1)
mod2r <- lm(data=vote2r, cvote~cincshr)
summary(mod2r)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -2.24982237  0.99860825 -2.252958 0.03148237
## cincshr      0.09239118  0.08472626  1.090467 0.28391138
```

- iv) En este caso cincshr 0.092 a lo que un aumento en el 10% en el gasto del candidato a cargo, representaría un .9% de votos a su favor, sin embargo al tener una desviación estándar de .0847, considerado alto en este caso, no muestra significatividad, además que la determinación del voto a un candidato es más complejo, y el tratar de conocer la propensión de voto por peso gastado puede que esté sesgado y otro punto es la escasez de datos, es una muestra muy pequeña al solo tratar de revisar el comportamiento contra opositores que repiten.

ejercicio 9

```
load("~/Dropbox/Econometria/datos/wooldridge6/crime4.RData")
crime4 <- data
rm('data')

crime4.pan <- pdata.frame(crime4, index = c("county", "year"))
```

```

mod1c <- plm(data=crime4.pan,lcrmte ~ d82 + d83 + d84 + d85 + d86 + d87 + lprbarr
            + lprbconv + lprbpris + lavgsen + lpolpc, model = "fd")

mod2c <- plm(data=crime4.pan,lcrmte ~ d82 + d83 + d84 + d85 + d86 + d87 + lprbarr
            + lprbconv + lprbpris + lavgsen + lpolpc + lwcon + lwtuc + lwtrd
            + lwfir + lwser + lwmfg + lwfed + lwsta + lwloc, model = "fd")

stargazer(mod1c, mod2c, title="Resultados", align=TRUE)

```

```

##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:39
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{Resultados}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }
##     \hline
##     \hline
##     & \multicolumn{2}{c}{\textit{Dependent variable:}} & \\
##     \cline{2-3}
##     \hline
##     & \multicolumn{2}{c}{lcrmte} & \\
##     \hline
##     & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \\
##     \hline
##     d82 & 0.008 & 0.020 & \\
##     & (0.017) & (0.021) & \\
##     & & & \\
##     d83 & -0.084^{***} & -0.071^{**} & \\
##     & (0.023) & (0.033) & \\
##     & & & \\
##     d84 & -0.125^{***} & -0.089^{**} & \\
##     & (0.029) & (0.043) & \\
##     & & & \\
##     d85 & -0.122^{***} & -0.069 & \\
##     & (0.033) & (0.058) & \\
##     & & & \\
##     d86 & -0.086^{**} & -0.018 & \\
##     & (0.037) & (0.070) & \\
##     & & & \\
##     d87 & -0.038 & 0.041 & \\
##     & (0.040) & (0.081) & \\
##     & & & \\
##     lprbarr & -0.327^{***} & -0.323^{***} & \\
##     & (0.030) & (0.030) & \\
##     & & & \\
##     lprbconv & -0.238^{***} & -0.240^{***} & \\
##     & (0.018) & (0.018) & \\
##     & & & \\
##     lprbpris & -0.165^{***} & -0.169^{***} & \\
##     & (0.026) & (0.026) & \\
##     & & & \\
##     lavgsen & -0.022 & -0.016 & \\
##     & (0.022) & (0.022) &

```

```

## & & \\
## lpolpc & 0.398^{***} & 0.398^{***} \\
## & (0.027) & (0.027) \\
## & & \\
## lwcon & & -0.044 \\
## & & (0.030) \\
## & & \\
## lwtuc & & 0.025^{*} \\
## & & (0.014) \\
## & & \\
## lwtrd & & -0.029 \\
## & & (0.031) \\
## & & \\
## lwfir & & 0.009 \\
## & & (0.021) \\
## & & \\
## lwser & & 0.022 \\
## & & (0.014) \\
## & & \\
## lwmfg & & -0.140 \\
## & & (0.102) \\
## & & \\
## lwfed & & 0.017 \\
## & & (0.172) \\
## & & \\
## lwsta & & -0.052 \\
## & & (0.096) \\
## & & \\
## lwloc & & -0.031 \\
## & & (0.102) \\
## & & \\
## \hline \\[-1.8ex]
## Observations & \multicolumn{1}{c}{540} & \multicolumn{1}{c}{540} \\
## R^{2}$ & \multicolumn{1}{c}{0.433} & \multicolumn{1}{c}{0.445} \\
## Adjusted R^{2}$ & \multicolumn{1}{c}{0.422} & \multicolumn{1}{c}{0.424} \\
## F Statistic & \multicolumn{1}{c}{40.318^{***}}$ (df = 10; 529) & \multicolumn{1}{c}{21.903^{***}}$ \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:} & \multicolumn{2}{r}{^{*}$p$<$0.1; ^{**}$p$<$0.05; ^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}

```

i) Han variado de manera mínima los coeficientes, sin embargo si ha habido cambio en la significatividad de los parametros como:

d83 y d84 perdieron significatividad hasta .5% d85 y d86 perdieron su significatividad.

y las variables de Criminalidad su variación es minima y no altera su significatividad

```

library(car)
linearHypothesis(mod2c, c("lwcon", "lwtuc", "lwtrd", "lwfir", "lwser", "lwmfg",
                          "lwfed", "lwsta", "lwloc"))

```

```

## Linear hypothesis test
##
## Hypothesis:

```

```
## lwcon = 0
## lwtuc = 0
## lwtrd = 0
## lwfir = 0
## lwser = 0
## lwmfg = 0
## lwfed = 0
## lwsta = 0
## lwloc = 0
##
## Model 1: restricted model
## Model 2: lcrmte ~ d82 + d83 + d84 + d85 + d86 + d87 + lprbarr + lprbconv +
##      lprbpris + lavgsen + lpolpc + lwcon + lwtuc + lwtrd + lwfir +
##      lwser + lwmfg + lwfed + lwsta + lwloc
##
##   Res.Df Df    Chisq Pr(>Chisq)
## 1      529
## 2      520  9 11.261      0.2583
```

ii) Son conjuntamente no significativo, al no rechazar H_0 siendo las variables no son distintas a 0

ejercicio 10

```
library(stats)
load("~/Dropbox/Econometria/datos/wooldridge6/jtrain.RData")
jtrain <- data
rm("data")

library(plm)

jtrain.pan <- pdata.frame(jtrain, index = c("fcode", "year"))
mod2c <- plm(data=jtrain.pan, hrsemp~d88+d89+grant+grant_1+lemploy, model = "fd" )

summary(mod2c)

## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = hrsemp ~ d88 + d89 + grant + grant_1 + lemploy,
##      data = jtrain.pan, model = "fd")
##
## Unbalanced Panel: n = 135, T = 1-3, N = 390
## Observations used in estimation: 255
##
## Residuals:
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -62.357  -5.128   0.704   0.059   4.635  137.207
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## d88          -0.75828    1.91854  -0.3952  0.6930
## d89           4.03449    3.18208   1.2679  0.2060
```



```
## grant    32.35374    2.88001 11.2339    <2e-16 ***
## grant_1  1.17407    5.17112  0.2270    0.8206
## lemploy  0.34908    4.70046  0.0743    0.9409
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    177350
## Residual Sum of Squares: 93091
## R-Squared:    0.47512
## Adj. R-Squared: 0.46672
## F-statistic: 56.5735 on 4 and 250 DF, p-value: < 2.22e-16
```

i) fueron usadas 225 observaciones...

ii) Es significativo el valor de grant y trata de decir que al recibir la el subsidio lo que provoca es que estaria empeorando el funcionamiento de la operación y desperdiciarias 32.36 más que si no lo tuvieras.

iii) puede que haya sido su poca variabilidad al momento de hacer las diferencias, lo que provocaria que si variabilidad estadística se muy alta y este sea impresiso

Ejercicio 11

```
load("~/Dropbox/Econometria/datos/wooldridge6/mathpnl.RData")
math_pane <- pdata.frame(data, index = c('distid', 'year'))
model11ii <- plm(math4 ~ year + lrexp + lenrol + lunch, data = math_pane, model = 'fd')
model11.3 <- plm(math4 ~ year + lrexp + lag(lrexp) + lenrol + lunch, data = math_pane, model = 'fd')
stargazer(model11ii, model11.3, title = "Resultados", align = TRUE, type = 'latex')
```

```
##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:44
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{Resultados}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }
##     \hline
##     \hline
##     & \multicolumn{2}{c}{\textit{Dependent variable:}} & \\
##     \cline{2-3}
##     \hline
##     & \multicolumn{2}{c}{math4} & \\
##     \hline
##     & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} & \\
##     \hline
##     year1993 & 5.955^{***} & 7.115^{***} & \\
##     & (0.518) & (0.674) & \\
##     & & & \\
##     year1994 & 12.431^{***} & 13.454^{***} & \\
##     & (0.759) & (0.849) & \\
##     & & & \\
##     year1995 & 25.198^{***} & 25.823^{***} & \\
##     & (1.067) & (1.091) & \\
##     & & & \\
##     year1996 & 25.918^{***} & 25.954^{***} & \end{tabular}
```

```

##      & (1.200) & (1.199) \\
##      & & \\
##      year1997 & 23.385^{***} & 23.296^{***} \\
##      & (1.333) & (1.332) \\
##      & & \\
##      year1998 & 38.308^{***} & 38.099^{***} \\
##      & (1.447) & (1.448) \\
##      & & \\
##      lrexpp & -3.447 & -2.341 \\
##      & (2.760) & (2.788) \\
##      & & \\
##      lag(lrexpp) & & 4.705^{***} \\
##      & & (1.753) \\
##      & & \\
##      lenrol & 0.635 & 1.108 \\
##      & (1.029) & (1.043) \\
##      & & \\
##      lunch & 0.025 & 0.027 \\
##      & (0.055) & (0.055) \\
##      & & \\
## \hline \\[-1.8ex]
## Observations & \multicolumn{1}{c}{3,300} & \multicolumn{1}{c}{3,299} \\
## R$^{2}$ & \multicolumn{1}{c}{0.208} & \multicolumn{1}{c}{0.210} \\
## Adjusted R$^{2}$ & \multicolumn{1}{c}{0.206} & \multicolumn{1}{c}{0.208} \\
## F Statistic & \multicolumn{1}{c}{108.028}^{***} (df = 8; 3291) & \multicolumn{1}{c}{96.976}^{***} \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:} & \multicolumn{2}{r}{*}p$<$0.1; **p$<$0.05; ***p$<$0.01 \\
## \end{tabular}
## \end{table}

```

- ii) En el primer modelo, si el gasto por alumno se incrementa en 10% generara una disminución en la tasa de quienes aprueban matemáticas de $3.5/10 \simeq 0.35\%$
- iii) El coeficiente que se agregó para $lrexpp_{t-1}$ es significativo a cualquier nivel y es de 4.7 mientras que el gasto en el año actual se vuelve insignificante. Se puede decir que un aumento del gasto real de 100% en el año anterior incrementa en $\simeq 4.7$ puntos la tasa de aprobados.

```

model11.3HAC <- plm(math4 ~ year + lrexpp + lag(lrexpp) + lenrol + lunch ,
  data = math_pane, model = 'fd', vcov = vcovNW)
stargazer(model11.3HAC,model11.3, title="Resultados", align=TRUE, type = 'latex')

```

```

##
## % Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: sáb, abr 14, 2018 - 16:05:47
## % Requires LaTeX packages: dcolumn
## \begin{table}[!htbp] \centering
##   \caption{Resultados}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }
##     \\[-1.8ex]\hline
##     \hline \\[-1.8ex]
##     & \multicolumn{2}{c}{\textit{Dependent variable:}} \\
##     \cline{2-3}
##     \\[-1.8ex] & \multicolumn{2}{c}{math4} \\
##     \\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)} \\

```

```

## \hline \[-1.8ex]
## year1993 & 7.115^{***} & 7.115^{***} \\
## & (0.674) & (0.674) \\
## & & \\
## year1994 & 13.454^{***} & 13.454^{***} \\
## & (0.849) & (0.849) \\
## & & \\
## year1995 & 25.823^{***} & 25.823^{***} \\
## & (1.091) & (1.091) \\
## & & \\
## year1996 & 25.954^{***} & 25.954^{***} \\
## & (1.199) & (1.199) \\
## & & \\
## year1997 & 23.296^{***} & 23.296^{***} \\
## & (1.332) & (1.332) \\
## & & \\
## year1998 & 38.099^{***} & 38.099^{***} \\
## & (1.448) & (1.448) \\
## & & \\
## lrexpp & -2.341 & -2.341 \\
## & (2.788) & (2.788) \\
## & & \\
## lag(lrexpp) & 4.705^{***} & 4.705^{***} \\
## & (1.753) & (1.753) \\
## & & \\
## lenrol & 1.108 & 1.108 \\
## & (1.043) & (1.043) \\
## & & \\
## lunch & 0.027 & 0.027 \\
## & (0.055) & (0.055) \\
## & & \\
## \hline \[-1.8ex]
## Observations & \multicolumn{1}{c}{3,299} & \multicolumn{1}{c}{3,299} \\
## R^2 & \multicolumn{1}{c}{0.210} & \multicolumn{1}{c}{0.210} \\
## Adjusted R^2 & \multicolumn{1}{c}{0.208} & \multicolumn{1}{c}{0.208} \\
## F Statistic (df = 9; 3289) & \multicolumn{1}{c}{96.976^{***}} & \multicolumn{1}{c}{96.976^{***}} \\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{2}{r}{*} & p < $0.1; ** & p < $0.05; *** & p < $0.01 \\
## \end{tabular}
## \end{table}

```

```

#Errores estándar robustos a heterocedasticidad
summary(model11.3, vcov = vcovHC)$coefficients

```

	Estimate	Std. Error	t-value	Pr(> t)
year1993	7.11473904	0.8866415	8.0243693	1.406362e-15
year1994	13.45400752	1.0378371	12.9635055	1.610756e-37
year1995	25.82282027	1.3384444	19.2931594	1.164999e-78
year1996	25.95449745	1.3191999	19.6744228	1.467130e-81
year1997	23.29557622	1.4118028	16.5005880	7.897643e-59
year1998	38.09868327	1.5418688	24.7094199	8.560980e-124
lrexpp	-2.34138801	4.6799973	-0.5002969	6.168995e-01
lag(lrexpp)	4.70480596	2.8144097	1.6716848	9.468161e-02
lenrol	1.10832011	1.4471623	0.7658575	4.438161e-01

```
## lunch          0.02687666  0.1482078  0.1813444  8.561084e-01
#según la documentación de plm::vcovNW asi deberia funcionar para
#errores estándar a heterocedasticidad y autocorrelación
#pero da Error in if (names(u)[i] == names(v)[j]) {: missing value where TRUE/FALSE needed
#coeftest(model11.3, vcov=vcovNW)
```

Ejercicio 12

```
load('~/.Dropbox/Econometria/datos/wooldridge6/murder.RData')
murder_pane <- pdata.frame(data, index('id', 'year'))

## Warning in pdata.frame(data, index("id", "year")): column 'id' overwritten
## by id index

model12i <- plm(mrdрте ~ year + exec + unem, model="pooling", data = murder_pane)
summary(model12i)$coefficients
```

```
##              Estimate Std. Error   t-value   Pr(>|t|)
## (Intercept) -23.8103958 26.7903317 -0.8887682 0.375560227
## year         0.2684415  0.2961822  0.9063389 0.366219834
## exec         0.1579371  0.1941408  0.8135185 0.417218950
## unem         1.2601587  0.4376239  2.8795471 0.004569543
```

- i) Según esta regresión, el efecto de a pena de muerte no disminuye el número de homicidios, por el contrario los incrementa en 0.16 pero no es significativo estadísticamente.

```
mur90 <- filter(murder_pane, year==90)
mur93 <- filter(murder_pane, year==93)
mur_diff <- mur90 - mur93
```

```
## Warning in Ops.factor(left, right): '-' not meaningful for factors
```

```
## Warning in Ops.factor(left, right): '-' not meaningful for factors
```

```
## Warning in Ops.factor(left, right): '-' not meaningful for factors
```

```
model12Fd <- lm(mrdрте ~ d93 + exec + unem, data =mur_diff)
summary(model12Fd)$coefficients
```

```
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept) -0.41326650 0.20938479 -1.973718 0.05418497
## exec        -0.10383958 0.04341395 -2.391848 0.02072926
## unem        -0.06659144 0.15868593 -0.419643 0.67661860
```

- ii) En este modelo *exec* cambia de signo y se vuelve significativo por lo tal si se puede decir que tiene un efecto negativo en el número de homicidios.

```
e <- residuals(model12Fd)
BP <- lm(I(e^2) ~ exec + unem, data = mur_diff)
linearHypothesis(BP, c('exec', 'unem'))
```

```
## Linear hypothesis test
```

```
##
```

```
## Hypothesis:
```

```
## exec = 0
```

```
## unem = 0
```

```
##
## Model 1: restricted model
## Model 2: I(e^2) ~ exec + unem
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      50 117.98
## 2      48 115.12  2    2.8611 0.5965 0.5548

fit <- model12Fd$fitted.values
white.test<- lm(I(e^2) ~ fit + I(fit^2))
linearHypothesis(white.test, c('fit','I(fit^2)'))
```

```
## Linear hypothesis test
##
## Hypothesis:
## fit = 0
## I(fit^2) = 0
##
## Model 1: restricted model
## Model 2: I(e^2) ~ fit + I(fit^2)
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      50 117.98
## 2      48 115.18  2    2.7996 0.5833 0.5619
```

- iii) No existe suficiente evidencia estadística para decir que hay heterocedasticidad en el modelo de primera diferencia, los p-value de las pruebas de Breusch Pagan y white son de 0.55 y 0.56 respectivamente.

```
coeftest(model12Fd, vcov= vcovHC)

##
## t test of coefficients:
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.413266   0.203632 -2.0295  0.04797 *
## exec        -0.103840   0.039696 -2.6159  0.01186 *
## unem        -0.066591   0.158815 -0.4193  0.67687
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- iv) El t Robusto a HC es -2.62 como se esperaba es más grande que el anterior porque los intervalos robustos son más amplios, pero no hace falta hacer estimaciones robustas por que no se encontró heterocedasticidad.

- vi) no hace falta hacer estimaciones robustas por que no se encontró heterocedasticidad por tal razon se prefiere el estadístico usual.

Ejercicio 13

```
load("~/Dropbox/Econometria/datos/wooldridge6/wagepan.RData")
wage.pan <- data
rm(data)

mod1 <- lm(lwage ~ (educ)*factor(year) +union , data = wage.pan)
summary(mod1)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.6510101321	0.14266449	4.563224784	5.176437e-06
## educ	0.0591349499	0.01198356	4.934674039	8.329084e-07
## factor(year)1981	-0.0007600008	0.20160715	-0.003769712	9.969924e-01
## factor(year)1982	0.0052971023	0.20159798	0.026275573	9.790387e-01
## factor(year)1983	0.0219610009	0.20159386	0.108936856	9.132576e-01
## factor(year)1984	0.0852010965	0.20159173	0.422641817	6.725775e-01
## factor(year)1985	0.0612573108	0.20160074	0.303854591	7.612532e-01
## factor(year)1986	0.0874411064	0.20161691	0.433699258	6.645284e-01
## factor(year)1987	0.0968167765	0.20159111	0.480263130	6.310645e-01
## union	0.1854887430	0.01723864	10.760054177	1.149151e-26
## educ:factor(year)1981	0.0102397244	0.01694775	0.604193630	5.457465e-01
## educ:factor(year)1982	0.0146063017	0.01694707	0.861877728	3.888024e-01
## educ:factor(year)1983	0.0174085983	0.01694660	1.027261882	3.043544e-01
## educ:factor(year)1984	0.0179839800	0.01694646	1.061223244	2.886475e-01
## educ:factor(year)1985	0.0246266585	0.01694688	1.453167708	1.462494e-01
## educ:factor(year)1986	0.0277291622	0.01694777	1.636154308	1.018798e-01
## educ:factor(year)1987	0.0317960692	0.01694644	1.876267851	6.068552e-02

i) las variables de tiempo principalmente, además que las otras variables cambian poco en el tiempo.

```
wage.panpan <- pdata.frame(wage.pan, index = c("nr", "year"))

mod1c <- plm(data=wage.panpan, l wage ~ (educ) * factor(year)
             + union, model = "fd" )
coef <- names((mod1c)$coefficients)
linearHypothesis(mod1c, coef[9:15])
```

```
## Linear hypothesis test
##
## Hypothesis:
## educ:factor(year)1981 = 0
## educ:factor(year)1982 = 0
## educ:factor(year)1983 = 0
## educ:factor(year)1984 = 0
## educ:factor(year)1985 = 0
## educ:factor(year)1986 = 0
## educ:factor(year)1987 = 0
##
## Model 1: restricted model
## Model 2: l wage ~ (educ) * factor(year) + union
##
##   Res.Df Df    Chisq Pr(>Chisq)
## 1     3807
## 2     3800   7 2.1373     0.9518
```

ii) El p-valor de la prueba F es 0.95 por lo cual no hay evidencia de que el efecto de la educación en el salario cambie en el tiempo analizado. Al hacer el modelo de diferencias, se anula la variable educación.

```
mod1c <- plm(data=wage.panpan, l wage ~ (educ) * factor(year)
             + union, model = "fd" )
coef <- names((mod1c)$coefficients)
#Se realiza la prueba solo robusta a heterocedasticidad por que con vcovNW no funciona
```

```
#y con vcovSCC da valores muy raros
linearHypothesis(mod1c,coef[9:15], vcov = vcovHC)
```

```
## Linear hypothesis test
##
## Hypothesis:
## educ:factor(year)1981 = 0
## educ:factor(year)1982 = 0
## educ:factor(year)1983 = 0
## educ:factor(year)1984 = 0
## educ:factor(year)1985 = 0
## educ:factor(year)1986 = 0
## educ:factor(year)1987 = 0
##
## Model 1: restricted model
## Model 2: lwage ~ (educ) * factor(year) + union
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df    Chisq Pr(>Chisq)
## 1     3807
## 2     3800  7  7.0258      0.4262
```

- iii) Haciendo la prueba con estimadores robustos a heterocedasticidad, el p-valor de la prueba F es de 0.42 por lo cual tampoco hay evidencia de que el efecto de la educación sobre el salario cambie en los periodos.

```
mod1c <- plm(data=wage.panpan, lwage ~ (educ) * factor(year)
+union * factor(year), model = "fd" )
summary(mod1c)$coefficients
```

	Estimate	Std. Error	t-value	Pr(> t)
## factor(year)1981	-0.008967107	0.13031319	-0.06881197	0.94514292
## factor(year)1982	0.013901141	0.18391006	0.07558663	0.93975196
## factor(year)1983	0.040854230	0.22502344	0.18155544	0.85594132
## factor(year)1984	0.109818991	0.25979694	0.42271087	0.67253019
## factor(year)1985	0.062734493	0.29008649	0.21626134	0.82879567
## factor(year)1986	0.088404100	0.31761408	0.27833810	0.78076800
## factor(year)1987	0.141818811	0.34328449	0.41312327	0.67953965
## union	0.105624238	0.04807647	2.19700497	0.02808037
## educ:factor(year)1981	0.011340588	0.01089367	1.04102525	0.29793018
## educ:factor(year)1982	0.015422106	0.01539654	1.00166018	0.31657164
## educ:factor(year)1983	0.017607386	0.01885501	0.93383070	0.35045067
## educ:factor(year)1984	0.017142129	0.02177398	0.78727592	0.43116955
## educ:factor(year)1985	0.025240339	0.02433991	1.03699413	0.29980470
## educ:factor(year)1986	0.029302154	0.02666343	1.09896412	0.27185347
## educ:factor(year)1987	0.031327543	0.02879939	1.08778502	0.27675914
## factor(year)1981:union	-0.019609041	0.04660205	-0.42077637	0.67394223
## factor(year)1982:union	-0.069152558	0.05550741	-1.24582574	0.21290532
## factor(year)1983:union	-0.088143516	0.05884727	-1.49783524	0.13425927
## factor(year)1984:union	-0.058525288	0.05990589	-0.97695381	0.32865432
## factor(year)1985:union	-0.048676771	0.06136947	-0.79317563	0.42772512
## factor(year)1986:union	-0.107560424	0.06259971	-1.71822557	0.08583709
## factor(year)1987:union	-0.147148680	0.06837792	-2.15199107	0.03146098

Ejercicio 16

```
county<-county <- data.frame(read.csv("~/Dropbox/Econometria/datos/wooldridge6/county Murders.csv"))
sum = c(
  'Mean murdrate' = round(mean(county$murdrate),3),
  'Sd murdrate' = round(sd(county$murdrate),3),
  '%murdrate = 0' = paste(round(nrow(filter(county, murdrate == 0))/nrow(county),4)*100,'%'),
  'Mean execs' = round(mean(county$execs),3),
  'Sd execs' = round(sd(county$execs),3),
  '%execs = 0' = paste(round(nrow(filter(county, execs == 0))/nrow(county),4)*100,'%'),
  'max excs' = max(county$execs),
  'nrow excs > 0' = nrow(filter(county, execs > 0))
)
sum
```

##	Mean murdrate	Sd murdrate	%murdrate = 0	Mean execs	Sd execs
##	"0.508"	"0.851"	"41.9 %"	"0.007"	"0.112"
##	%execs = 0	max excs	nrow excs > 0		
##	"99.45 %"	"7"	"205"		

i & ii) La media de ejecuciones por pena de muerte es muy baja por que de los 2197 condados únicamente se aplicó esta condena en 205 y el máximo fue de 7. Si no se tuvieran en cuenta los condados que no la aplicaron, el promedio sería de 1.25.

iii) Para poder aplicar sección cruzada por MCO se debe asumir que el error idiosincrático no está correlacionando con las variables explicativas.

```
library(stringr)
i=0
teta = NULL
for(year in unique(county$year)){
  teta[i<- i+1]= paste("y",str_sub(year, start= -2), sep = "_")
  county[teta[i]] <-
    ifelse(county$year == year, 1, 0)
}
years <- select(county, c(teta[2:16]))
model16_4 <- lm(county$murdrate ~. + county$execs +
  lag(county$execs) + county$percblack + county$percmale
  + county$perc1019 + county$perc2029, data = years)
summary(model16_4)$coefficients
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	0.910960481	0.1078108204	8.4496202	3.026762e-17
##	y_81	0.019819395	0.0226134868	0.8764414	3.807958e-01
##	y_82	0.033857252	0.0226398516	1.4954715	1.347999e-01
##	y_83	-0.047600101	0.0227235420	-2.0947483	3.620009e-02
##	y_84	-0.077821516	0.0228451603	-3.4064771	6.587641e-04
##	y_85	-0.049300658	0.0229337466	-2.1496992	3.158542e-02
##	y_86	-0.078797923	0.0230052234	-3.4252188	6.149652e-04
##	y_87	-0.065725728	0.0229970780	-2.8580034	4.265516e-03
##	y_88	-0.073847859	0.0229763342	-3.2140836	1.309728e-03
##	y_89	-0.091789943	0.0229621584	-3.9974440	6.415314e-05
##	y_90	-0.043050141	0.0229652270	-1.8745794	6.085842e-02
##	y_91	-0.065563269	0.0229733817	-2.8538797	4.321264e-03
##	y_92	-0.061709037	0.0229937124	-2.6837353	7.283684e-03


```
## y_93          0.001975480 0.0224776036 0.0878866 9.299673e-01
## y_94          -0.007279477 0.0225373326 -0.3229964 7.466998e-01
## y_95          -0.061408833 0.0226073219 -2.7163250 6.604149e-03
## county$execs 0.158076707 0.0407598703 3.8782436 1.053936e-04
## lag(county$execs) 0.156309506 0.0407594589 3.8349259 1.258084e-04
## county$percblack 0.019550389 0.0003273903 59.7158536 0.000000e+00
## county$percmale -0.010719324 0.0019971801 -5.3672295 8.043131e-08
## county$perc1019 -0.006029970 0.0024246342 -2.4869605 1.288830e-02
## county$perc2029 0.002750175 0.0012658096 2.1726607 2.981214e-02
```

- iv) En esta regresión, se observa que manteniendo los demás factores constantes, por cada pena capital, se incrementan los asesinatos en 0.16. Lo que puede estar pasando aquí es que los otros factores que se encuentran en el error idiosincrático inciden más que el número de ejecuciones que se realicen.
- v) Es raro que tantos parámetros sean tan significativos, seguramente el modelo tiene problemas de autocorrelación y esto hace que se subestime los errores estándar y a su vez se sobre estime la significatividad de los parámetros.

```
county_pane <- pdata.frame(county, index = c('countyid', 'year'))
model13_fd <- plm(murdrate ~ year + execs + lag(execs) + percblack + percmale +
  perc1019 + perc2029, data = county_pane, model = 'fd')
summary(model13_fd)$coefficients
```

```
##          Estimate Std. Error    t-value  Pr(>|t|)
## year1981 -0.0462015857 0.02454723 -1.88215087 0.05982379
## year1982 -0.0218815572 0.04010384 -0.54562243 0.58532897
## year1983 -0.0891153029 0.05615470 -1.58696081 0.11253052
## year1984 -0.1020193092 0.07231059 -1.41084882 0.15829805
## year1985 -0.0533458290 0.08642973 -0.61721620 0.53709613
## year1986 -0.0596285714 0.09995251 -0.59656901 0.55079901
## year1987 -0.0211347306 0.11172498 -0.18916745 0.84996267
## year1988 -0.0046675668 0.12296830 -0.03795748 0.96972180
## year1989 0.0002027582 0.13401864 0.00151291 0.99879288
## year1990 0.0668904189 0.14402876 0.46442405 0.64234684
## year1991 0.0613162032 0.15396862 0.39823832 0.69045694
## year1992 0.0824225115 0.16372328 0.50342572 0.61466815
## year1993 0.1746024248 0.24208221 0.72125261 0.47075891
## year1994 0.1797633410 0.24802306 0.72478479 0.46858890
## year1995 0.1360879174 0.25246038 0.53904664 0.58985809
## year1996 0.1373170315 0.25572470 0.53697210 0.59129029
## execs    -0.0029340759 0.04937216 -0.05942774 0.95261176
## lag(execs) 0.0278753853 0.05269265 0.52901851 0.59679595
## percblack 0.0208243565 0.03279320 0.63502065 0.52541908
## percmale  -0.0119626404 0.01731513 -0.69087801 0.48964677
## perc1019 0.0149674562 0.03301533 0.45334866 0.65030050
## perc2029 0.0497673734 0.02413941 2.06166508 0.03924697
```

- vi) los estimadores de $execs$ y $execs_{t-1}$ son $\simeq .003$ y $\simeq .028$ es evidente a simple vista la diferencia con el modelo de sección cruzada, incluso cambia el signo.
- vii) No existe suficiente evidencia estadística para decir que la pena de muerte influye significativamente en el número de homicidios tanto $execs$ y $execs_{t-1}$ tienen valores T muy pequeños.