

Capítulo 13 Ejercicios En Computadora

Econometría II, Nelson Muriel

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

April 16, 2018

1 capítulo 13

1.1 Ejercicio 1

```
In [177]: options(warn=-1)
library(plm)
library(car)
library(stats)
library(stargazer)
library(stats)
library(dplyr)
library(lmtest)
library(sandwich)
```

```
In [178]: load("data/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'))
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1115	2697.104	NA	NA	NA	NA
1111	2685.898	4	11.20549	1.158765	0.327458

- 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.

```
In [179]: linearHypothesis(mod13_1, c('east', 'northcen', 'west'))
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1114	2707.741	NA	NA	NA	NA
1111	2685.898	3	21.84249	3.011656	0.02925802

- 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%.

```
In [180]: 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'))
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1128	12269.90	NA	NA	NA	NA
1122	12082.22	6	187.6803	2.904782	0.008158624

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

```
In [181]: 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
car::linearHypothesis(mod13_1, c("I(y74 * educ)", "I(y76 * educ)", "I(y78 * educ)", "I(y80 * educ)", "I(y82 * educ)", "I(y84 * educ)"))
```

	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	
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1111	2685.898	NA	NA	NA	NA
1105	2664.435	6	21.46357	1.483569	0.1803366

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

1.2 Ejercicio 2

```
In [182]: load("data/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.

```
In [183]: 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 $y_{85} 0.118 \text{ se } \simeq .124$ y la ecuación de modelo real

$$y_{85} - 0.383 \text{ se } \simeq .124$$

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.

```
In [184]: 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 habia 96 trabajadores afiliados, representando un 18%. haciendo la comparación en porcentaje cayeron las afiliaciones por 12.5% en 7 años

```
In [185]: 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 una 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

1.3 Ejercicio 3

```
In [186]: load("data/kielmc.RData")
          kielmc <- data
          rm('data', 'desc')
          mod3_1 <- lm(data = kielmc, lprice ~ y81 + ldist + y81ldist)
          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.

```
In [187]: mod3_2 <- lm(data = kielmc, lprice ~ y81 + ldist + y81ldist + age + agesq + rooms + ba
          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.

1.4 Ejercicio 5

```
In [188]: load("data/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 y90, 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

$$lrent_{it} = y90t + lpop_{it} + lavginc_{it} + pctst_{it} + a_i + u_{it}$$

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.

```
In [189]: 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 = 'text')
```

Resultados

Dependent variable:		
	lrent	
	panel linear (1)	OLS (2)
y90		0.262*** (0.035)
lpop	0.297** (0.143)	0.041* (0.023)
lavginc	0.940*** (0.047)	0.571*** (0.053)
pctstu	0.019*** (0.007)	0.005*** (0.001)
Constant		-0.569 (0.535)
Observations	64	128
R2	0.303	0.861
Adjusted R2	0.280	0.857
Residual Std. Error		0.126 (df = 123)
F Statistic	-14.580 (df = 2; 61)	190.922*** (df = 4; 123)
Note: *p<0.1; **p<0.05; ***p<0.01		

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.

1.5 Ejercicio 4

```
In [190]: load("data/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
          stargazer(mod13_12ky, mod13_12kyc, title="Resultados", align=TRUE, type = 'text')
```

Resultados

Dependent variable:		
	ldurat	
	(1)	(2)
afchnge	0.008 (0.045)	0.011 (0.045)
highearn	0.256*** (0.047)	0.176*** (0.052)
afhigh	0.191*** (0.069)	0.231*** (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*** (0.031)	1.246*** (0.106)

Observations	5,626	5,349
R2	0.021	0.041
Adjusted R2	0.020	0.039
Residual Std. Error	1.269 (df = 5622)	1.251 (df = 5334)
F Statistic	39.540*** (df = 3; 5622)	16.372*** (df = 14; 5334)

=====

Note: *p<0.1; **p<0.05; ***p<0.01

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.

```
In [191]: injurymi <- filter(injury,mi==1)
          mod13_12mi <- lm(data=injurymi, 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.

1.6 Ejercicio 6

```
In [192]: load("data/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", header=FALSE, type = 't
```

Resultados

Dependent variable:			
	lcrime		clcrime
	(1)	(2)	(3)
d78	-0.055 (0.094)	-0.055 (0.094)	
clrprc1	-0.018*** (0.005)	-0.001 (0.010)	
clrprc2	-0.017*** (0.005)		
I(clrprc1 + clrprc2)		-0.017*** (0.005)	
cavgclr			-0.017*** (0.005)
Constant	4.181*** (0.188)	4.181*** (0.188)	0.099 (0.063)
Observations	106	106	53
R2	0.471	0.471	0.175
Adjusted R2	0.455	0.455	0.159
Residual Std. Error	0.455 (df = 102)	0.455 (df = 102)	0.345 (df = 51)
F Statistic	30.269*** (df = 3; 102)	30.269*** (df = 3; 102)	10.798*** (df = 1; 51)
Note: *p<0.1; **p<0.05; ***p<0.01			

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

1.7 ejercicio 7

```
In [193]: load("data/gpa3.RData")
          gpa4 <- data
          rm('data', 'desc')
          mod1 <- lm(data=gpa4, trmgpa ~ spring + sat + hsperc + female + black + white + frstsem
                    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

```
In [194]: 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 = 'text')
```

Resultados

```
=====
                Dependent variable:
                -----
                        trmgpa
                -----
spring                -0.237
                      (0.206)

frstsem                0.019
                      (0.069)

tothrs                 0.012
                      (0.014)

crsgpa                1.136***
```

```

(0.119)
season          -0.065
                (0.043)

-----
Observations      732
R2                0.208
Adjusted R2       -0.603
F Statistic      19.016*** (df = 5; 361)
=====
Note:            *p<0.1; **p<0.05; ***p<0.01

```

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

1.8 Ejercicio 8

```

In [195]: load("data/vote2.RData")
          vote2 <- data
          rm('data')
          mod1c <- lm(data= vote2, cvote~ clineexp + clchexp + cincshr)
          summary(mod1c)$coefficients

```

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

```

In [196]: linearHypothesis(mod1c, c("clineexp", "clchexp"))

```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
155	9282.273	NA	NA	NA	NA
153	9102.292	2	179.9807	1.512643	0.2236031

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

```
In [197]: 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

```
In [198]: 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.

1.9 Ejercicio 9

```
In [199]: load("data/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", type = 'text')
```

Resultados

```
=====
Dependent variable:
-----
lcrmte
```

	(1)	(2)
d82	0.008 (0.017)	0.020 (0.021)
d83	-0.084*** (0.023)	-0.071** (0.033)
d84	-0.125*** (0.029)	-0.089** (0.043)
d85	-0.122*** (0.033)	-0.069 (0.058)
d86	-0.086** (0.037)	-0.018 (0.070)
d87	-0.038 (0.040)	0.041 (0.081)
lprbarr	-0.327*** (0.030)	-0.323*** (0.030)
lprbconv	-0.238*** (0.018)	-0.240*** (0.018)
lprbpris	-0.165*** (0.026)	-0.169*** (0.026)
lavgsen	-0.022 (0.022)	-0.016 (0.022)
lpolpc	0.398*** (0.027)	0.398*** (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)

-----
Observations      540                540
R2                0.433                0.445
Adjusted R2       0.422                0.424
F Statistic  40.318*** (df = 10; 529) 21.903*** (df = 19; 520)
=====
Note:                *p<0.1; **p<0.05; ***p<0.01

```

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 signifiatividad hasta .5% d85 y d86 perdieron su significatividad.

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

```

In [200]: library(car)
          linearHypothesis(mod2c, c("lwcon", "lwtuc", "lwtrd", "lwfir", "lwser", "lwmfg",
                                     "lwfed", "lwsta", "lwloc"))

```

Res.Df	Df	Chisq	Pr(>Chisq)
529	NA	NA	NA
520	9	11.26062	0.2582668

ii) Son conjuntamente no significativo, al no rechazar H0 siendo las variables no son distintas a 0

1.10 Ejercicio 10

```

In [201]: load("data/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.
-62.400  -5.130   0.704   0.059   4.640  137.000

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 el subsidio lo que provoca es que estaría empeorando el funcionamiento de la operación y desperdiciarías 32.36 más que si no lo tuvieras.

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

1.11 Ejercicio 11

```

In [202]: load("data/mathpn1.RData")
math_pane <- pdata.frame(data, index = c('distid','year'))
model11iii <- plm(math4 ~ year + lrexp + lenrol + lunch, data = math_pane, model = 'fd')

```



```

model11.3 <- plm(math4 ~ year + lrexpp + lag(lrexpp) + lenrol + lunch , data = math_pa
stargazer(model11iii,model11.3, title="Resultados", align=TRUE, type = 'text')

```

Resultados

Dependent variable:		
	math4	
	(1)	(2)
year1993	5.955*** (0.518)	7.115*** (0.674)
year1994	12.431*** (0.759)	13.454*** (0.849)
year1995	25.198*** (1.067)	25.823*** (1.091)
year1996	25.918*** (1.200)	25.954*** (1.199)
year1997	23.385*** (1.333)	23.296*** (1.332)
year1998	38.308*** (1.447)	38.099*** (1.448)
lrexpp	-3.447 (2.760)	-2.341 (2.788)
lag(lrexpp)		4.705*** (1.753)
lenrol	0.635 (1.029)	1.108 (1.043)
lunch	0.025 (0.055)	0.027 (0.055)
Observations	3,300	3,299
R2	0.208	0.210
Adjusted R2	0.206	0.208
F Statistic	108.028*** (df = 8; 3291)	96.976*** (df = 9; 3289)
Note:	*p<0.1; **p<0.05; ***p<0.01	

- ii) En el primer modelo, si el gasto por alumno se incrementa en 10% generara una disminuci3n en la tasa de quienes aprueban matemáticas de $3.5/10 \simeq 0.35\%$
- iii) El coeficiente que se agreg3 para $lrexpp_{t-1}$ es sginificativo a cualquier nivel y es de 4.7 mientras que el gasto en el a3o actual se vuelve insignificante. Se puede decir que un aumento del gasto real de 100% en el a3o anterior incrementa en $\simeq 4.7$ puntos la tasa de aprobados.

```
In [203]: 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 = 'text')
```

Resultados

Dependent variable:		
	math4	
	(1)	(2)
year1993	7.115*** (0.674)	7.115*** (0.674)
year1994	13.454*** (0.849)	13.454*** (0.849)
year1995	25.823*** (1.091)	25.823*** (1.091)
year1996	25.954*** (1.199)	25.954*** (1.199)
year1997	23.296*** (1.332)	23.296*** (1.332)
year1998	38.099*** (1.448)	38.099*** (1.448)
lrexpp	-2.341 (2.788)	-2.341 (2.788)
lag(lrexpp)	4.705*** (1.753)	4.705*** (1.753)
lenrol	1.108 (1.043)	1.108 (1.043)

lunch	0.027 (0.055)	0.027 (0.055)
-------	------------------	------------------

```
-----
Observations      3,299      3,299
R2                0.210      0.210
Adjusted R2       0.208      0.208
F Statistic (df = 9; 3289) 96.976*** 96.976***
=====
```

Note: *p<0.1; **p<0.05; ***p<0.01

```
In [204]: #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

```
In [205]: #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 ne
#coeftest(model11.3, vcov=vcovNW)
```

1.12 Ejercicio 12

```
In [206]: load('data/murder.RData')
murder_pane <- pdata.frame(data, index('id', 'year'))
model12i <- plm(mrdrt ~ 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.

```
In [207]: mur90 <- filter(murder_pane, year==90)
mur93 <- filter(murder_pane, year==93)
mur_diff <- mur90 - mur93
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.

```
In [208]: e <- residuals(model12Fd)
BP <-lm(I(e^2) ~ exec + unem, data = mur_diff)
linearHypothesis(BP, c('exec','unem'))
fit <- model12Fd$fitted.values
white.test<- lm(I(e^2) ~ fit + I(fit^2))
linearHypothesis(white.test, c('fit','I(fit^2)'))
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
50	117.9781	NA	NA	NA	NA
48	115.1171	2	2.861061	0.5964837	0.5547754
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
50	117.9781	NA	NA	NA	NA
48	115.1786	2	2.799555	0.583349	0.5619332

- 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.

```
In [209]: 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.

1.13 Ejercicio 13

In [210]: `load("data/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.

In [211]: `wage.panpan <- pdata.frame(wage.pan,index = c("nr","year"))`

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

Res.Df	Df	Chisq	Pr(>Chisq)
3807	NA	NA	NA
3800	7	2.13734	0.9518492

- 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.

In [212]: `mod1c <- plm(data=wage.panpan,lwage ~ (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)

Res.Df	Df	Chisq	Pr(>Chisq)
3807	NA	NA	NA
3800	7	7.02579	0.4261996

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.

```
In [213]: mod1c <- plm(data=wage.panpan,lwage ~ (educ) * factor(year)
+union * factor(year), model ="fd" )
summary(mod1c)$coefficients
coef <- names((mod1c)$coefficients)
linearHypothesis(mod1c,coef[16:22])
linearHypothesis(mod1c,coef[16:22], vcov = vcovHC)
```

	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
Res.Df	Df	Chisq	Pr(>Chisq)	
3800	NA	NA	NA	
3793	7	7.233906	0.4049389	
Res.Df	Df	Chisq	Pr(>Chisq)	
3800	NA	NA	NA	
3793	7	8.099407	0.32391	

iv)

- v) La prueba conjunta F para los estimadores usuales tiene un P-valor de 0.40 y para errores robustos es de 0.32, por lo cual no hay evidencia de que el efecto de union está cambiando en el tiempo. individualmente, solo el año 1987 es significativo.
- w) la diferencia en 1980 de *union* es de 0.1056. En 1987, la diferencia es de -0.147 y es significativa al 97

1.14 Ejercicio 14

```
In [214]: load('data/jtrain3.RData')
jtrain <- pdata.frame(data)
mod14 <- lm(re78 ~ train, data = jtrain)
stargazer(mod14, type = 'text', header = FALSE)
```

```
=====
Dependent variable:
-----
re78
-----
train1                -15.205***
                      (1.155)
Constant              21.554***
                      (0.304)
-----
Observations          2,675
R2                    0.061
Adjusted R2           0.061
Residual Std. Error   15.152 (df = 2673)
F Statistic           173.415*** (df = 1; 2673)
=====
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

i) Si es positiva y realmente significativo el coeficiente, haciendo mención que si fue esta capacitación tuvo un efecto positivo en el sueldo pudiendo ser que a empleador mayor capacitados es mejor su sueldo.

```
In [215]: mod14_2 <- lm(I(re78 - re75) ~ train, data = jtrain)
stargazer(mod14_2, type = 'text', header = FALSE)
```

```
=====
Dependent variable:
-----
I(re78 - re75)
```

```

-----
train1                2.327***
                      (0.814)

Constant              2.491***
                      (0.214)

-----
Observations          2,675
R2                    0.003
Adjusted R2           0.003
Residual Std. Error   10.680 (df = 2673)
F Statistic            8.172*** (df = 1; 2673)
=====
Note:                  *p<0.1; **p<0.05; ***p<0.01

```

ii) quitando los efectos fijos en el tiempo se muestra que en realidad la capacitación y su relación positiva con el sueldo es en realidad menor que la del primer modelo, ya que en los efectos inobservables sobre estiman los valores, sin embargo sigue siendo significativo y positivo.

In [216]: *#pendiente hacer los intervalos del 14.3*

1.15 Ejercicio 15

```

In [217]: library(stringr)
          load('data/happiness.RData')
          happiness <- data
          for (year in unique(happiness$year)){
            print(paste(year, ' -> ', sum(select(happiness, paste('y', str_sub(year, start = -2),
            paste('Porcentaje vHappy', round(sum(happiness$vhappy)/nrow(happiness), 4)*100, '%')

[1] "1994 -> 2977"
[1] "1996 -> 2885"
[1] "1998 -> 2806"
[1] "2000 -> 2777"
[1] "2002 -> 1369"
[1] "2004 -> 1337"
[1] "2006 -> 2986"

```

'Porcentaje vHappy 30.69 %'

```

In [218]: mod14 <- lm(vhappy ~ y96 + y98 + y00 + y02 + y04 + y06, data = happiness)
          linearHypothesis(mod14, names((mod14)$coefficients[2:7]), vcov = vcovHC)

```


Res.Df	Df	F	Pr(>F)
17136	NA	NA	NA
17130	6	1.429194	0.1989899

- ii) La prueba F con errores robustos tiene un P-valor de .20 por lo cual no hay evidencia de que la variable vhappy est cambiando en los periodos de la muestra.

```
In [219]: mod14 <- update(mod14,. ~ . + occattend + regattend, data = happiness)
          summary(mod14)
```

Call:

```
lm(formula = vhappy ~ y96 + y98 + y00 + y02 + y04 + y06 + occattend +
    regattend, data = happiness)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.4148	-0.2992	-0.2881	0.6931	0.7287

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.271346	0.009030	30.049	<2e-16 ***
y96	0.016749	0.012144	1.379	0.1679
y98	0.027859	0.012193	2.285	0.0223 *
y00	0.031266	0.012266	2.549	0.0108 *
y02	0.015748	0.015083	1.044	0.2965
y04	0.025164	0.015188	1.657	0.0976 .
y06	0.022184	0.011968	1.854	0.0638 .
occattend	0.004265	0.008095	0.527	0.5983
regattend	0.112174	0.010818	10.369	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4597 on 16855 degrees of freedom

(273 observations deleted due to missingness)

Multiple R-squared: 0.007092, Adjusted R-squared: 0.006621

F-statistic: 15.05 on 8 and 16855 DF, p-value: < 2.2e-16

- iii) En referencia a esta regresin lo que muestra es que a mayor asistencia a la iglesia puede aportar a tu felicidad de manera positiva.

```
In [220]: happiness <- mutate(happiness, highinc = ifelse(income == '$25000 or more', 1, 0))
          mod14 <- update(mod14, . ~ . + highinc + unem10 + educ + teens)
          summary(mod14)
```

Call:

```
lm(formula = vhappy ~ y96 + y98 + y00 + y02 + y04 + y06 + occattend +
    regattend + highinc + unem10 + educ + teens, data = happiness)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.4945	-0.3413	-0.2468	0.6211	0.9087

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.196274	0.023831	8.236	< 2e-16 ***
y96	0.012344	0.015586	0.792	0.4284
y98	0.018809	0.015568	1.208	0.2270
y00	0.030608	0.015941	1.920	0.0549 .
y02	-0.016943	0.019244	-0.880	0.3786
y04	0.006819	0.019649	0.347	0.7286
y06	-0.005601	0.015434	-0.363	0.7167
occattend	-0.006507	0.010429	-0.624	0.5327
regattend	0.095997	0.014034	6.841	8.36e-12 ***
highinc	0.101143	0.010307	9.813	< 2e-16 ***
unem10	-0.088062	0.009867	-8.925	< 2e-16 ***
educ	0.003915	0.001647	2.377	0.0175 *
teens	-0.016782	0.009347	-1.795	0.0726 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4518 on 9755 degrees of freedom
(7369 observations deleted due to missingness)

Multiple R-squared: 0.03111, Adjusted R-squared: 0.02992

F-statistic: 26.1 on 12 and 9755 DF, p-value: < 2.2e-16

iv) Tuvo una ligera disminución en su magnitud, a lo que sigue siendo positivo y además significativo

v) En regattend pierden fuerza, sin embargo es significativa, a comparación de occattend que no lo es además de que tuvo un cambio de signo y una mayor fuerza. Dice entonces que si eres muy recurrente en asistir a misa aporta en tu felicidad en un 9%

```
In [221]: mod14 <- update(mod14, . ~ . + highinc + unem10 + educ + teens +
    black + female)
summary(mod14)
```

Call:

```
lm(formula = vhappy ~ y96 + y98 + y00 + y02 + y04 + y06 + occattend +
    regattend + highinc + unem10 + educ + teens + black + female,
    data = happiness)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.5005	-0.3415	-0.2483	0.6180	0.9235

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.206587	0.024568	8.409	< 2e-16 ***
y96	0.013191	0.015579	0.847	0.397156
y98	0.020499	0.015566	1.317	0.187909
y00	0.031683	0.015935	1.988	0.046807 *
y02	-0.015322	0.019239	-0.796	0.425821
y04	0.007291	0.019639	0.371	0.710453
y06	-0.004014	0.015433	-0.260	0.794782
occattend	-0.003403	0.010465	-0.325	0.745078
regattend	0.100040	0.014107	7.092	1.42e-12 ***
highinc	0.096764	0.010452	9.258	< 2e-16 ***
unem10	-0.086849	0.009871	-8.799	< 2e-16 ***
educ	0.003516	0.001650	2.131	0.033130 *
teens	-0.015069	0.009361	-1.610	0.107503
black	-0.050246	0.013940	-3.604	0.000314 ***
female	0.002185	0.009287	0.235	0.814042

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4516 on 9753 degrees of freedom

(7369 observations deleted due to missingness)

Multiple R-squared: 0.0324, Adjusted R-squared: 0.03101

F-statistic: 23.33 on 14 and 9753 DF, p-value: < 2.2e-16

Vi) Sigue teniendo un efecto considerable y significativo el hecho de la recurrencia en la asistencia a una iglesia, en cambio en el tema de raza y género, el género no es significativo, estadísticamente no es diferente a 0, pero en la raza si es significativo el ser de raza negra, existe un efecto negativo a la felicidad, además de ser altamente significativo.

1.16 Ejercicio 16

```
In [222]: county<-county <- data.frame(read.csv("data/countymurders.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 exces'      = max(county$execs),
    'nrow exces > 0' = nrow(filter(county, execs > 0))
  )
sum

```

```

Mean murdrate '0.508' Sd murdrate '0.851' \%murdrate = 0 '41.9 %' Mean execs '0.007' Sd
execs '0.112' \%execs = 0 '99.45 %' max exces '7' nrow exces > 0 '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 correlaciondo con las variables explicativas.

```

In [223]: 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\$percmales	-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.

```
In [224]: county_pane <- pdata.frame(county, index = c('countyid','year'))
          model13_fd <- plm(murdrate ~ year + execs + lag(execs) + percblack + percmales +
          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.