Introducción al análisis de datos de panel en R

Estudio longitudinal social de Chile (ELSOC)

Enero, 2021: sesión 3

- **Cluster standard errors**: son generalizaciones a nivel de grupos (por ejemplo, g) de los errores estándares robustos a la heteroscedasiticidad.

- En el caso de OLS, si asumimos errores i.i.d tenemos

$$V_{OLS} = \sigma^2 (X'X)^{-1} \tag{1}$$

Podemos estimar σ^2 con s^2

$$s^2 = \frac{1}{N-K} \sum_{i=1}^{N} e_i^2 \tag{2}$$

N es el número de observaciones, K es el número de variables en la regresión y e_i son los residuos.

- Cluster standard errors es una versión del estimador sandwich

$$V_{Cluster} = (X'X)^{-1} \sum_{j=1}^{n_C} (u_j' * u_j)(X'X)^{-1} \quad (3)$$

donde n_C es el número total de clusters y $u_j = \sum_{j_{Cluster}} e_i * x_i$

- Algunos programas como Stata usan un ajuste de los grados de libertad para muestras pequeñas: $\frac{M}{M-1} * \frac{N-1}{N-K} * V_{Cluster}$

- M es el número de clusters, N es el tamaño muestral y K es el número de variables en el modelo.

- Problemas
  - Necesita un número grande de clusters: a lo menos 40.
  - Note que si asume cierto tipo de correlación dentro de los clusters y ocupa este tipo de corrección, en cierto modo, está diciendo que SUTVA no se cumple.
- Ver Angrist y Pischke (2009), cap. 8.

```
# Sesión 3, elsoc, enero 2021

library(foreign)
library(dplyr)

# Para tablas de calidad
library(texreg)
library(xtable)

library(plm) # Paquete para datos de panel
```

	id	time	wage	marr
1.	1	1	1000	0
2.	1	2	1050	0
3.	1	3	950	0
4.	1	4	1000	0
5.	1	5	1100	0
6.	1	6	900	0
7.	2	1	2000	0
8.	2	2	1950	0
9.	2	3	2050	0
10.	2	4	2000	0
11.	2	5	1950	0
12.	2	6	2050	0
13.	3	1	2900	0
14.	3	2	3000	0
15.	3	3	3100	0
16.	3	4	3500	1
17.	3	5	3450	1
18.	3	6	3550	1
19.	4	1	3950	0
20.	4	2	4050	0
21.	4	3	4000	0
22.	4	4	4500	1
23.	4	5	4600	1
24.	4	6	4400	1

	id	time	wage	marr	dwage	dmarr	dtime
1.	1	1	1000	0	.	.	.
2.	1	2	1050	0	50	0	1
3.	1	3	950	0	-100	0	1
4.	1	4	1000	0	50	0	1
5.	1	5	1100	0	100	0	1
6.	1	6	900	0	-200	0	1
7.	2	1	2000	0	.	.	.
8.	2	2	1950	0	-50	0	1
9.	2	3	2050	0	100	0	1
10.	2	4	2000	0	-50	0	1
11.	2	5	1950	0	-50	0	1
12.	2	6	2050	0	100	0	1
13.	3	1	2900	0	.	.	.
14.	3	2	3000	0	100	0	1
15.	3	3	3100	0	100	0	1
16.	3	4	3500	1	400	1	1
17.	3	5	3450	1	-50	0	1
18.	3	6	3550	1	100	0	1
19.	4	1	3950	0	.	.	.
20.	4	2	4050	0	100	0	1
21.	4	3	4000	0	-50	0	1
22.	4	4	4500	1	500	1	1
23.	4	5	4600	1	100	0	1
24.	4	6	4400	1	-200	0	1

	id	time	wage	marr	mwage	mmarr	wwage	wmarr
1.	1	1	1000	0	1000	0	0	0
2.	1	2	1050	0	1000	0	50	0
3.	1	3	950	0	1000	0	-50	0
4.	1	4	1000	0	1000	0	0	0
5.	1	5	1100	0	1000	0	100	0
6.	1	6	900	0	1000	0	-100	0
7.	2	1	2000	0	2000	0	0	0
8.	2	2	1950	0	2000	0	-50	0
9.	2	3	2050	0	2000	0	50	0
10.	2	4	2000	0	2000	0	0	0
11.	2	5	1950	0	2000	0	-50	0
12.	2	6	2050	0	2000	0	50	0
13.	3	1	2900	0	3250	.5	-350	-.5
14.	3	2	3000	0	3250	.5	-250	-.5
15.	3	3	3100	0	3250	.5	-150	-.5
16.	3	4	3500	1	3250	.5	250	.5
17.	3	5	3450	1	3250	.5	200	.5
18.	3	6	3550	1	3250	.5	300	.5
19.	4	1	3950	0	4250	.5	-300	-.5
20.	4	2	4050	0	4250	.5	-200	-.5
21.	4	3	4000	0	4250	.5	-250	-.5
22.	4	4	4500	1	4250	.5	250	.5
23.	4	5	4600	1	4250	.5	350	.5
24.	4	6	4400	1	4250	.5	150	.5

```
> # Cross-sectional OLS para T=4
< ####################################
> ols1 <- lm(wage~marr, data=Panel, subset=time==4)
> summary(ols1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1500.0      500.0   3.000   0.0955 .
marr          2500.0      707.1   3.536   0.0715 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Pooled OLS
> ###############
> ols2 <- lm(wage~marr, data=Panel)
> summary(ols2)

Call:
lm(formula = wage ~ marr, data = Panel)

Residuals:
    Min      1Q  Median      3Q     Max
-1266.7  -679.2  -166.7   633.3  1883.3

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2166.7      236.2   9.175 5.65e-09 ***
marr          1833.3      472.3   3.882 0.000805 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Primeras diferencias 1
> ############################
> summary(lm(dwage~dmarr,data=Panel))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.333     24.296   0.343    0.736
dmarr        441.667     76.830   5.749 1.89e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Primeras diferencias 2
> ############################
> fd2 <- plm(wage~ marr, data = Panel, model = "fd")
> summary(fd2)
Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept)   8.3333    24.2956  0.3430    0.7356
marr        441.6667    76.8295  5.7487 1.893e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> fd1 <- plm(wage~-1 + marr, data = Panel, model = "fd")
> summary(fd1)
Coefficients:
     Estimate Std. Error t-value  Pr(>|t|)
marr  450.000     71.174  6.3225 4.558e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Efectos fijos 1
> ######################
> summary(lm(wwage~wmarr,data=Panel))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.382e-15  1.269e+01    0.00        1
wmarr        5.000e+02  3.589e+01   13.93 2.15e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



> fe1 <- plm(wage~marr, data = Panel, model = "within")
> summary(fe1)
Coefficients:
     Estimate Std. Error t-value  Pr(>|t|)
marr  500.000     38.616  12.948 7.101e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Efectos fijos 2
> ####################
> fe2 <- lm(wage~marr + factor(id), data=Panel)
> summary(fe2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1000.00      27.31   36.62  < 2e-16 ***
marr          500.00      38.62   12.95 7.10e-11 ***
factor(id)2  1000.00      38.62   25.90 2.78e-16 ***
factor(id)3  2000.00      43.17   46.32  < 2e-16 ***
factor(id)4  3000.00      43.17   69.49  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> fe3 <- lm(wage~-1 + marr + factor(id), data=Panel)
> summary(fe3)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
marr          500.00      38.62   12.95  7.1e-11 ***
factor(id)1  1000.00      27.31   36.62  < 2e-16 ***
factor(id)2  2000.00      27.31   73.24  < 2e-16 ***
factor(id)3  3000.00      33.44   89.70  < 2e-16 ***
factor(id)4  4000.00      33.44  119.61  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # ------------- Cluster estandar errors ------------- #
> library(clubSandwich)

> a <- lm(wage~marr, data=Panel)
> summary(a)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2166.7      236.2   9.175 5.65e-09 ***
marr          1833.3      472.3   3.882 0.000805 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> coef_test(a,vcov="CR2",cluster=Panel$id)
        Coef. Estimate  SE t-stat d.f. p-val (Satt) Sig.
1 (Intercept)     2167 606   3.58 2.63        0.046    *
2        marr     1833 698   2.63 1.51        0.158
```

1. Cluster SE
2. Efectos fijos y cluster SE: ejemplos
3. Resumen y literatura

- Cluster SE y FE.
- Ejemplos en R.

- Literatura efectos fijos
  - Angrist, J. D. y J. Pischke (2009). *Mostly Harmless Econometrics: an Empiricist's Companion*. Cap. 5.
  - Wooldridge, Jeffrey M. (2001). *Introducción a la econometría: un enfoque moderno*. Australia: Thomson. Caps. 13 y 14.
- Literatura datos de panel
  - Andress, H. J., K. Golsch y A. W. Schmidt (2013). *Applied Panel Data Analysis for Economic and Social Surveys*. Berlin, Heildelberg: Springer.