IN4402 - Aplicaciones de Probabilidades y Estadística en Gestión
Profesor: Benjamín Villena Roldán
Profesor auxiliar: Ronald Leblebici Garo

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Clase auxiliar # 6

Dif-in-dif, test Reset de Ramsey y regresión discontinua

Canción de hoy: Ilegal - Cultura profética.

¡Recordar!

- **Dif-in-dif:** método que se puede utilizar en datos de panel para estimar el efecto de un tratamiento.

$$\beta_1 = \mathbb{E}(\Delta Y_{n2}|D_{n2} = 1) - \mathbb{E}(\Delta Y_{n2}|D_{n2} = 0)$$

$$\hat{\beta}_1 = [\bar{Y}_{2,D_{n2}=1} - \bar{Y}_{1,D_{n2}=1}] - [\bar{Y}_{2,D_{n2}=0} - \bar{Y}_{1,D_{n2}=0}]$$

- **Dif-in-dif en múltiples periodos:** este método nos permite controlar heterogeneidad no observada invariante en el tiempo, como lo serían los efectos fijos α_n por individuo.

$$Y_{nt} = \alpha_n + \delta_t + \lambda D_{nt} + X_{nt}\beta + U_{nt} \Rightarrow \bar{Y}_n = \alpha_n + \bar{\delta} + \lambda \bar{D}_n + \bar{X}_n\beta + \bar{U}_n$$

$$\Rightarrow \underbrace{Y_{nt} - \bar{Y}_n}_{y_{nt}} = \delta_t - \bar{\delta} + \lambda \underbrace{(D_{nt} - \bar{D}_n)}_{d_{nt}} + \underbrace{(X_{nt} - \bar{X}_n)}_{x_{nt}} \beta + \underbrace{(U_{nt} - \bar{U}_n)}_{u_{nt}}$$

- **Test Reset de Ramsey:**

1. Especificar $Y = X\beta + U$ y estimar $\hat{\beta}_{MCO}$ e $\hat{Y}$.
2. Construir $\hat{Y}^p$ para $p \in \{1, ..., P\}$
3. Regresionar $Y = X\beta + \gamma_2\hat{Y}^2 + ... + \gamma_P\hat{Y}^P + V$
4. Testear $H_0 : \gamma_2 = ... = \gamma_P = 0$ (restricciones múltiples).

1. Dif-in-dif

Parte 1

En el curso "Emprendición e Innovamiento", el profesor auxiliar ha decidido realizar una clase "innovadora" entre los controles 1 ($t = 1$) y 2 ($t = 2$) para un subconjunto de la sección asignado aleatoriamente. El objetivo del auxiliar es estimar el efecto de la realización de esta clase sobre el promedio de notas de los alumnos, para ver si dicha metodología contribuye al aprendizaje de sus estudiantes. El docente se contacta directamente con usted y le pide que estime este efecto a partir de los datos presentados en la siguiente tabla.

	Tratamiento	Control
t=1	6	5,9
t=2	6,3	5,7

Utilizando la metodología dif-in-dif, estime dicho efecto por partes, rellenando los cuadros que se muestran a continuación:

$$\hat{\beta}_1 = \underbrace{\underbrace{[\underbrace{\bar{Y}_{2,D_{n2}=1}}_{\boxed{6,3}} - \underbrace{\bar{Y}_{1,D_{n2}=1}}_{\boxed{6,0}}]}_{\boxed{0,3}} - \underbrace{[\underbrace{\bar{Y}_{2,D_{n2}=0}}_{\boxed{5,7}} - \underbrace{\bar{Y}_{1,D_{n2}=0}}_{\boxed{5,9}}]}_{\boxed{0,2}}}_{\boxed{0,5}}$$

Parte 2

Para esta pregunta, utilizaremos la base de datos de la Universidad de Princeton contenida en la página: `http://dss.princeton.edu/training/Panel101.dta`. En dicha base se tiene información sobre una variable `y`, para años entre 1990 y 1999 y ciudades enumeradas desde el 1 hasta el 7. Nos interesa conocer el efecto sobre la variable `y` de un tratamiento que se llevó a cabo desde 1994 para las ciudades 5, 6 y 7.

1. Plantee un modelo de regresión que permita estimar el efecto del tratamiento.

   Para llevar a cabo el método dif-in-dif, es necesario caracterizar dos tipos de individuos en dos periodos. Es así como definiremos las siguientes variables:

   - α_n: efecto fijo correspondiente a si n es tratado o no tratado.
   - δ_t: efecto temporal que varía según si t es antes o después del tratamiento.
   - D_{nt}: variable dummy que señala si a la persona se le hizo o no el tratamiento.

   Luego se define el siguiente modelo[1]

   $$Y_{nt} = \alpha_n + \delta_t + \lambda D_{nt} + U_{nt}$$

   El efecto del tratamiento se obtiene al estimar λ por mínimos cuadrados ordinarios.

[1] Se pudo haber agregado un factor $X_{nt}\beta$ que considere otros regresores. No lo hemos hecho por simplicidad.

2. Ejecute esta regresión en *Stata* y estime el efecto del tratamiento β_{did}.

   Ver archivo `Auxiliar 06 Código.do`. Como se puede observar en la table, el efecto de tratamiento estimado es $\beta_{\text{did}} = -2,52 \cdot 10^9$ y es satisface un nivel de significancia del 10 %.

```
Linear regression                               Number of obs     =         70
                                                F(3, 66)          =       2,17
                                                Prob > F          =     0,0998
                                                R-squared         =     0,0827
                                                Root MSE          =    3,0e+09

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   2,29e+09   9,00e+08     2,54   0,013     4,92e+08    4,09e+09
     treated |   1,78e+09   1,05e+09     1,70   0,094    -3,11e+08    3,86e+09
         did |  -2,52e+09   1,45e+09    -1,73   0,088    -5,42e+09    3,81e+08
       _cons |   3,58e+08   7,61e+08     0,47   0,640    -1,16e+09    1,88e+09
------------------------------------------------------------------------------
```

3. Instale la extensión `diff` y estime nuevamente el efecto de tratamiento mediante este comando. Compare los resultados con los de la parte anterior.

   Ver archivo `Auxiliar 06 Código.do`. Como se puede observar en las tablas, los resultados son coherentes. En particular, se evidencia que el efecto diff-in-diff es equivalente al calcularlo de ambas maneras.

```
DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS
Number of observations in the DIFF-IN-DIFF: 70
            Before         After
   Control: 16             24          40
   Treated: 12             18          30
            28             42
--------------------------------------------------------
 Outcome var.   | y        | S. Err. | |t|   |  P>|t|
----------------+----------+---------+-------+---------
Before          |          |         |       |
   Control      | 3,6e+08  |         |       |
   Treated      | 2,1e+09  |         |       |
   Diff (T-C)   | 1,8e+09  | 1,1e+09 | 1,58  | 0,120
After           |          |         |       |
   Control      | 2,6e+09  |         |       |
   Treated      | 1,9e+09  |         |       |
   Diff (T-C)   | -7,4e+08 | 9,2e+08 | 0,81  | 0,422
                |          |         |       |
Diff-in-Diff    | -2,5e+09 | 1,5e+09 | 1,73  | 0,088*
--------------------------------------------------------
R-square:    0,08
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1
```

2. Test Reset de Ramsey

Utilise la base de datos `Auxiliar 06 Test Reset Ramsey BD.dta` y considere el modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + u_i$.

1. ¿Qué modelo plantearía para realizar el test Reset de Ramsey con potencias $p = 2, 3$?

   Plantearía el modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + \gamma_2 \hat{y}_i^2 + \gamma_3 \hat{y}_i^3 + v_i$, donde $\hat{y}_i$ sería la estimación por MCO del modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + u_i$.

2. Plantee un test de hipótesis (hipótesis nula, hipótesis alternativa, estadístico y región de rechazo) para dicho test.

   - H_0: $\gamma_2 = \gamma_3 = 0$.
   - H_1: $\gamma_2 \neq 0 \vee \gamma_3 \neq 0$.
   - $F = \frac{(R_L^2 - R_R^2)/r}{(1-R_L^2)/(N-K)} \sim \mathcal{F}(r, N-K)$.
   - $RR = \{F > F_{\text{crítico}}(r, N-K)\}$

3. Realice dicho test en *Stata*. ¿Qué puede concluir?

   Ver archivo `Auxiliar 06 Código.do`. Como se puede ver en la imagen, se puede rechazar la hipótesis nula con una significancia del 5 % (e incluso menor). Lo anterior permite rechazar que γ_2 y γ_3 sean cero, lo que es señal de que al menos un regresor de segundo y/o tercer grado de x y z podría ser estadísticamente significativo al momento de explicar y.

```
. test (y_estimado_2=0) (y_estimado_3=0)

 ( 1)  y_estimado_2 = 0
 ( 2)  y_estimado_3 = 0

       F(  2, 97921) = 3564,37
            Prob > F =    0,0000
```

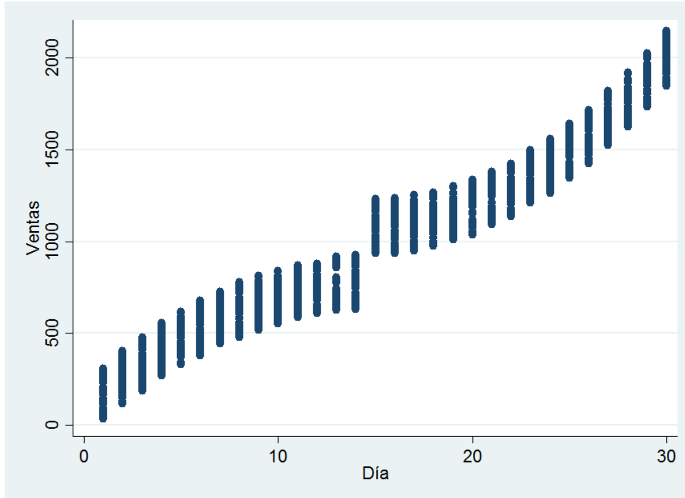
3. Regresión discontinua

El Gerente de Ventas de una famosa firma de juguetes quiere conocer el comportamiento de la demanda por *spinners* durante el mes de diciembre, de cara a la navidad. La empresa en cuestión tiene 50 multitiendas –convenientemente todas iguales– distribuidas a lo largo del país. Las ventas de cada una de estas empresas entre los días 1 y 25 de diciembre se encuentran registradas en la base de datos `Auxiliar 06 Regresión Discontinua BD.dta`.

A modo de antecedente, dicho gerente le comenta que el día 15 de diciembre se llevó a cabo una espectacular campaña de Marketing, y sospecha que esta misma pudo tener un efecto en la cantidad de compras efectuadas.

1. Obtenga un gráfico de dispersión (o *scatter*) que le permita visualizar cómo se comportó la demanda entre los días 1 y 25 de diciembre.

   Ver archivo `Auxiliar 06 Código.do`.



2. Proponga un modelo de regresión discontinua para dicha demanda.

   Sea Y_n la cantidad de ventas registradas en la observación n, X_n el día al cual corresponde la observación n (entre 1 y 25) y $D_n = \mathbb{I}_{X_n \geq 15}$.

   $$Y_n = \beta_0 + \tau D_n + \beta_1 (X_n - 15)(1 - D_n) + \beta_2 (X_n - 15)^2 (1 - D_n) + \beta_3 (X_n - 15) D_n + \beta_4 (X_n - 15)^2 D_n + U_n$$

3. Ejecute este modelo de regresión en *Stata* y señale el efecto de la propaganda sobre la demanda. ¿Es significativo?

   Ver archivo `Auxiliar 06 Código.do`. Como se puede observar en la tabla, el efecto inmediato de la propaganda sobre la demanda (τ) se estima en $\hat{\tau} = 313,659$ unidades de *spinners* vendidos y es estadísticamente significativo con un 95 % de confianza.

Source	SS	df	MS
Model	356575289	5	71315057,8
Residual	11142404,2	1.494	7458,1019
Total	367717693	1.499	245308,668

Number of obs	=	1.500
F(5, 1494)	=	9562,09
Prob > F	=	0,0000
R-squared	=	0,9697
Adj R-squared	=	0,9696
Root MSE	=	86,36

ventas	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.propaganda	313,659	13,98148	22,43	0,000	286,2336	341,0844
propaganda#c.dia						
0	-1,494995	3,490112	-0,43	0,668	-8,341035	5,351046
1	-,3296092	2,512831	-0,13	0,896	-5,25866	4,599441
propaganda#c.dia2						
0	-3,163379	,2263255	-13,98	0,000	-3,607329	-2,71943
1	4,036786	,1615977	24,98	0,000	3,719803	4,353768
_cons	778,6497	11,37947	68,43	0,000	756,3282	800,9711