

Apéndice. Código PS3 (STATA)

2023-08-17

```

/*****
Semana 4: Fuentes de sesgo e imprecisión

Universidad de San Andrés
Economía Aplicada
*****/

Barnes, Fasan, Legaspe y Martin
*****/
Este archivo sigue la siguiente estructura:

0) Set up environment

1) Simulaciones:

1.1) Un ejemplo para diferencias en el Tamaño Muestral
1.2) Un ejemplo para diferencias en la varianza del error
1.3) Un ejemplo para diferencias en la varianza del regresor
1.4) El valor de la suma de los residuos
1.5) ¿Los residuos son ortogonales a los regresores?
1.6) Un ejemplo para Multicolinealidad
1.7) Error no aleatorio en X
1.8) Erro no aleatorio en Y

2) Fuentes de Sesgo e Inconsistencia

*****/

* 0) Set up environment
*=====*/

global main "C:\Users\Usuario\Desktop\MAESTRIA\Economia Aplicada\TPs\PS3"
global input "$main/input"
global output "$main/output"

cd "$main"
*=====*
```

```

* 1) Simulaciones
* Para todas las simulaciones generaremos las variables que en en el
* ejemplo provisto: wage, education, intelligence, a y b.
* Para alguno casos generaremos todas, en otros sola las necesarias para
* exponer nuestra idea. 0
=====
* 1.1) Un ejemplo para diferencias en el Tamaño Muestral
* N = 100
clear
* GPD:
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100

gen education= int(invnormal(uniform()))*1+5
corr education intelligence
gen u=int(invnormal(uniform()))*1+7
gen b=int(invnormal(uniform()))*1+5
gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

* N = 200
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100
gen education= int(invnormal(uniform()))*1+5
corr education intelligence
gen u=int(invnormal(uniform()))*1+7

gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

=====
*1.2) Un ejemplo para diferencias en la varianza del error
* sigma_1 = 0.5
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100

gen education= int(invnormal(uniform()))*1+5
corr education intelligence
gen u=int(invnormal(uniform()))*0.5+7

gen wage=3*intelligence+ 2*education +6 + u

```

```

reg wage education intelligence, robust

* sigma_2 = 2
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100

gen education= int(invnormal(uniform()))*1+5)
corr education intelligence
gen u=int(invnormal(uniform()))*2+7)

gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

=====
*1.3) Un ejemplo para diferencias en la varianza del regresor
* caso 1
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100)

gen education= int(invnormal(uniform()))*1+5)
corr education intelligence
gen u=int(invnormal(uniform()))*2+7)

gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

* Caso 2
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*50+100)

gen education= int(invnormal(uniform()))*1+5)
corr education intelligence
gen u=int(invnormal(uniform()))*2+7)

gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

```

```

=====
*1.4) El valor de la suma de los residuos
* DGP
* Aca si agregamos una variable mas para que no parezca que los residuos son 0,
* por que el modelo ajusta con  $R^2 = 1$ .
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*50+100

gen education= int(invnormal(uniform()))*70+5
corr education intelligence
gen u=int(invnormal(uniform()))*2+7
gen a = int(invnormal(uniform()))*1+7
gen b = int(invnormal(uniform()))*2+15

gen wage=3*intelligence+ 2*education + 6*b + 20*a +6 + u

reg wage education intelligence, robust

predict residuos, residuals
sum residuos
* Podemos ver que dan cero, lo que ocurre por definicion
=====
*1.5) ¿Los residuos son ortogonales a los regresores?
* Una forma sencilla de ver esto es regresando los residuos sobre las variables
reg residuos education intelligence, robust
* Podemos ver que todos los coeficientes son 0

=====
*1.6) Un ejemplo para Multicolinealidad
* Sin Multicolinealidad
clear
set obs 5000
set seed 1233
gen intelligence=int(invnormal(uniform()))*20+100

gen education= int(invnormal(uniform()))*50+5
corr education intelligence
gen u=int(invnormal(uniform()))*1+7
gen wage=3*intelligence+ 2*education + u

reg wage education intelligence, robust
predict y_hat_1, xb

* Con multicolinealidad
set seed 1233
replace education=int(intelligence/10+invnormal(uniform()))*1 // multicolinealidad
corr education intelligence
replace wage=3*intelligence+ 2*education + u

```

```

reg wage education intelligence, robust
predict y_hat_2, xb

sort intelligence
twoway (lfit y_hat_1 intelligence) ///
      (lfit y_hat_2 intelligence), ///
      title(" ") ///
      legend(label(1 "Sin Multicolinealidad") label(2 "Con Multicolinealidad")) ///
      xtitle("intelligence") ytitle("Fitted Values")
graph export "$output/Multicol_intel.png", width(4000)
sort education
twoway (lfit y_hat_1 education) ///
      (lfit y_hat_2 education), ///
      title(" ") ///
      legend(label(1 "Sin Multicolinealidad") label(2 "Con Multicolinealidad")) ///
      xtitle("education") ytitle("Fitted Values")
graph export "$output/Multicol_educ.png", width(4000)

=====
* BASELINE PARA 1.7 y 1.8
clear
* GPD:
set obs 200
set seed 1234
gen intelligence=int(invnormal(uniform()))*20+100

gen education= int(invnormal(uniform()))*1+5
corr education intelligence
gen u=int(invnormal(uniform()))*2+7

gen wage=3*intelligence+ 2*education +6 + u

reg wage education intelligence, robust

=====

=====
*1.7) Error no aleatorio en X
clear
* GPD con error no aleatorio en X:
set obs 200
set seed 1234
* Generamos un error no aleatorio para intelligence
gen v = _n // error no aleatorio que aumenta con cada observacion
gen intelligence =int(invnormal(uniform()))*20*0.1*v+100

gen education= int(invnormal(uniform()))*1+5
corr education intelligence
gen u=int(invnormal(uniform()))*2+7

gen wage= 3*intelligence+ 2*education + 6 + u

reg wage education intelligence, robust

```

```

=====
*1.8) Erro no aleatorio en Y
clear
* GPD con error no aleatorio en X:
set obs 200
set seed 1234

gen intelligence=int(invnormal(uniform())*20+100)

gen education= int(invnormal(uniform())*1+5)
corr education intelligence
gen u=int(invnormal(uniform())*2+7) // error no aleatorio
* Generamos un error no aleatorio para intelligence como una secuencia
gen v = _n
gen wage= 3*intelligence+ 2*education + 6 + u + v

reg wage education intelligence, robust

=====
*2) Fuentes de sesgo (ESTO NO ERA NECESARIO PERO COMO LO HICIOS LO DEJAMOS)
=====
* GDP
clear
set obs 200
set seed 1233
gen attend = rnormal(25, 4)
=====
* X1 altamente correlacionada con X2 y X3
gen cgpa = int(attend/1.6+invnormal(uniform())*2)
corr attend cgpa

gen study = int(attend/0.7+invnormal(uniform())*3)
corr attend study

gen u=int(invnormal(uniform())*2+5)

gen score = 1.5*attend + 3*cgpa + 4*study + u

reg score attend

reg score attend cgpa study, robust
=====
* X1 no correlacionada con X2 y X3, X2 y X3 altamente correlacionadas
replace cgpa = int(invnormal(uniform())*2)
corr attend cgpa

replace study = int(cgpa/0.3+invnormal(uniform())*3)
corr attend study
corr cgpa study

```

```
replace score = 1.5*attend + 3*cgpa + 4*study + u  
reg score attend  
reg score attend cgpa study, robust
```