Statistics/Data analysis

```
*ECONOMIA APLICADA PROBLEM SET 3
*QUISPE, CASIANO, SAMBRANA, RIGIROZZI


************PUNTO 1***************
****AUMENTAMOS EL TAMAÑO MUESTRAL PARA VER QUE SUCEDE CON LOS ERRORES STANDARD
*Primero seteamos el modelo de base mostrado en clases.
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

*notar la correlacion entre educacion e inteligencia.

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligence a b
eststo OLS1


**Definimos un nuevo tamanio muestral mayor, de 1000
clear
set obs 1000
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)


gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u

reg wage education intelligence a b
eststo OLS2
*Lo que se observa con este mayor tamanio muestral es que se da una reduccion en el va
> lor de los correspondientes errores estandar. Adicionalmente, vemos que se estima me
> jor los coeficientes, puntualmente el de educacion se aproxima aun mas a su verdader
> o valor, 0.

esttab OLS1 OLS2 using "ejercicio1.tex",  replace se label


*******************************************************************************
************PUNTO 2***************
***AUMENTAMOS LA VARIANZA DEL TERMINO DE ERROR

*Seteamos nuevamente el modelo original:
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

*notar la correlacion entre educacion e inteligencia.

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u
```

```
reg wage education intelligence a b
eststo OLS3
*Luego proponemos la modificacion en la varianza del termino de error
clear

set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

gen education=int(intelligence/10+invnormal(uniform())*1)


gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*30+7)
gen wage=3*intelligence+a+2*b+u

reg wage education intelligence a b
eststo OLS4
*Con este aumento en la varianza del termino de error, vemos claramente que los valore
> s de los errores estandar aumentan de forma significativa. Otra consecuencia es que
> se estima los coeficientes de manera muy sesgada con respecto a su verdadero valor.

esttab OLS3 OLS4 using "ejercicio2.tex",  replace se label

**************************************************************************
***********PUNTO 3******************

*Ahora tomamos un modelo que solo considere a la variable inteligencia como regreso
* Con varianza de X=20

clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

*notar la correlacion entre educacion e inteligencia.

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligence a b
eststo OLS5
**********************************
*CON VARIANZA X=50


clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*50+100)

*notar la correlacion entre educacion e inteligencia.

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u
```

```
reg wage education intelligence a b
eststo OLS6
*Vemos que los errores estandar del regresor intelligence se reducen significativament
> e al contar con mayor variabilidad. Esto se explica, en parte porque al tener mayor
> variabilidad en el regresor,se puede medir su efecto sobre y de una manera mas preci
> sa.

esttab OLS5 OLS6 using "ejercicio3.tex",  replace se label

*******************************************************************************
******PUNTO 4**********
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

*notar la correlacion entre educacion e inteligencia.

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligence a b

predict residuals, res
tabstat residuals, s (sum)

****veamos la suma de residuos con mayor variabilidad del regresor intelligence

clear

set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*50+100)

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligence a b


predict residuals, res
tabstat residuals, s (sum)
*Lo que observamos es que, al aumentar la variabilidad en el regresor, al mismo tiempo
>  se reduce la suma de los residuos.
*******************************************************************************
******PUNTO 5**********

clear all

set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence
```

```
gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*30+7)
gen wage=3*intelligence+a+2*b+u

reg wage education intelligence a b


predict residuals, res
tabstat residuals, s (sum)

// instalar corrtex con el siguiente comando: ssc install corrtex
corrtex residuals intelligence a b education, file(ejercicio5)
*Vemos la ortogonalidad via la correlacion de los errores con cada uno de los regresor
> es. Efectivamente, al ser 0 para cada uno, se da ortogonalidad.

*************************************************************************
******PUNTO 6**********

*probemos predecir sin la variable que genera multicolinealidad y con ella.
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)


gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage intelligence a b
predict y_hat_1

reg wage education intelligence a b
predict y_hat_2
corrtex y_hat_1 y_hat_2, file(ejercicio6)
*Como la correlacion entre los y estimados es perfecta, la introduccion de un regresor
>  altamente correlacionado con otro, no produce problemas al momento de estimar y.


*************************************************************************
******PUNTO 7**********
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u
```

```
* Include education that is not in the Data Generating Process and it is highly correl
> ated with intelligence. Note that coefficients and SE for a and b do not change, but
>  the SE for the coefficient of intelligence changes, and a lot.
reg wage education intelligence a b
eststo OLS7
*planteamos el modelo en el que tenemos un regresor, inteligencia, con un error de med
> icion no aleatorio.

clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)
gen intelligencemed=int(invnormal(uniform())*20+100) + 4
/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligencemed a b
eststo OLS8
*Vemos un sesgo extremadamente grande para la variable inteligencia, que se contagia a
>  la variable educacion, que tambien esta exageradamente sesgada, al estar altamente
> correlacionada.



*planteamos el modelo en el que tenemos un regresor, inteligencia, con un error de med
> icion aleatorio.

clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)
gen intelligencemed=int(invnormal(uniform())*20+100) + int(invnormal(uniform())*1+2)
/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligencemed a b
eststo OLS9
*Vemos un sesgo extremadamente grande para la variable inteligencia(aunque mayor que c
> on un error no aleatorio), que se contagia a la variable educacion, que tambien esta
>  exageradamente sesgada, al estar altamente correlacionada.

esttab OLS7 OLS8 OLS9 using "ejercicio7.tex",  replace se label

*******************************************************************************
******PUNTO 8**********
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/
```

```
gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u


reg wage education intelligence a b
eststo OLS10

*suest ols11 ols12, robust
predict residuals
tabstat residuals,s(v mean)



*******************************
*errores en Y (no aleatorio)

clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)

/* We set the standard error of this variable so the correlation between education and
>  intelligence is high (0.90 approximate).*/

gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u
gen wagemed=3*intelligence+a+2*b+u + 8


reg wagemed education intelligence a b
eststo OLS11

predict residuals
tabstat residuals,s(v mean)

*********************************

*errores en Y (aleatorio)
clear
set obs 100
set seed 1234
gen intelligence=int(invnormal(uniform())*20+100)


gen education=int(intelligence/10+invnormal(uniform())*1)
corr education intelligence

gen a=int(invnormal(uniform())*2+10)
gen b=int(invnormal(uniform())*1+5)
gen u=int(invnormal(uniform())*1+7)
gen wage=3*intelligence+a+2*b+u
gen wagemed=3*intelligence+a+2*b+u + int(uniform()*4+100)


reg wagemed education intelligence a b
eststo OLS12
predict residuals
tabstat residuals,s(v mean)

*Notar el aumento en la media de los residuos.
```

```
esttab OLS10 OLS11 OLS12 using "ejercicio8.tex",  replace se label
```