

ECONOMETRIA II- CURSO 2016
MODELOS DE VARIABLE DEPENDIENTE TRUNCADA
CENSURADA, MODELOS DE SELECTIVIDAD

EJERCICIO 1

Se procedió a estimar un modelo para analizar el efecto de la educación en las horas trabajadas por mujeres casadas:

Se dispone de las siguientes variables:

HS: horas trabajadas

ANIOSED: años de educación

EXP_POT: experiencia potencial (edad-aniosed-6)

EXP_CUAD: experiencia potencial al cuadrado

CANT_HIJOS: cantidad de hijos

LYLABD_HOM: ingreso laboral del marido (en logaritmos)

Se estimó el modelo a partir de una muestra de mujeres uruguayas con edades entre 25 y 60 años para el año 2009. Se estimó primero modelos MCO y Tobit utilizando todas las observaciones. Luego se estimó un modelo MCO sólo con aquellas mujeres que trabajaban y un modelo Heckman. Los resultados fueron:

```
. *Estimacion MCO utilizando todas laa observaciones
. reg horas $lista1 $lista2
```

Source	SS	df	MS	Number of obs =	19919
Model	1041072.21	4	260268.051	F(4, 19914) =	561.87
Residual	9224433.52	19914	463.213494	Prob > F =	0.0000
				R-squared =	0.1014
				Adj R-squared =	0.1012
Total	10265505.7	19918	515.388379	Root MSE =	21.522

horas	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
aniosedu	1.531808	.0437479	35.01	0.000	1.446059	1.617558
exp_pot	.4272559	.068093	6.27	0.000	.2937879	.5607239
exp_cuad	-.0093327	.0012535	-7.45	0.000	-.0117897	-.0068757
cant_hijos	-1.741745	.128056	-13.60	0.000	-1.992746	-1.490745
_cons	8.91959	1.030287	8.66	0.000	6.900142	10.93904

```
. *MODELO TOBIT
. tobit horas $lista1 $lista2, robust ll(0)
```

Tobit regression	Number of obs =	19919
	F(4, 19915) =	633.44
	Prob > F =	0.0000
Log pseudolikelihood = -69450.588	Pseudo R2 =	0.0160

horas	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Intervall]	
aniosedu	2.247643	.0617	36.43	0.000	2.126706	2.36858
exp_pot	.7852234	.0977638	8.03	0.000	.5935983	.9768485
exp_cuad	-.0171696	.0018975	-9.05	0.000	-.0208888	-.0134503
cant_hijos	-2.889338	.2050298	-14.09	0.000	-3.291214	-2.487463
_cons	-6.205808	1.447544	-4.29	0.000	-9.043116	-3.368501
/sigma	30.29703	.1940286			29.91671	30.67734

```
. *Estimacion MCO utilizando las observaciones no censuradas
. reg horas $lista1 $lista2 if horas>0
```

horas	Coef.	Std. Err.	t	P> t	[95% Conf. Intervall]	
aniosedu	.6101407	.0416848	14.64	0.000	.5284325	.691849
exp_pot	.2071339	.0656402	3.16	0.002	.0784697	.3357981
exp_cuad	-.0035754	.0012511	-2.86	0.004	-.0060278	-.001123
cant_hijos	-.8542656	.1334352	-6.40	0.000	-1.115818	-.5927133
_cons	29.95701	.9776397	30.64	0.000	28.04069	31.87332

Heckman selection model -- two-step estimates	Number of obs	=	19919
(regression model with sample selection)	Censored obs	=	6752
	Uncensored obs	=	13167
	Wald chi2(4)	=	48.97
	Prob > chi2	=	0.0000

lylabd_hom		-.0007083	.0037679	-0.19	0.851	-.0080931	.0066766
_cons		-.4373499	.0732671	-5.97	0.000	-.5809509	-.293749

mills							
lambda		20.40891	5.429278	3.76	0.000	9.767716	31.05009

rho		0.91885					
sigma		22.211331					

SE PIDE:

- 1) ¿Es la variable dependiente “HS” una variable:
 - a. Truncada
 - b. Censurada
 - c. Ninguna de las opciones anteriores?
 Justifique su respuesta.
- 2) Analice e interprete los resultados de las estimaciones Tobit y MCO utilizando todas las observaciones.
- 3) Explique en que consiste el método bietápico de Heckman. Explique la forma de la inversa del ratio de Mills (o lambda de Heckman) λ y explique su significado y la del coeficiente asociado.
- 4) Analice e interprete los resultados de la estimaciones MCO utilizando sólo las observaciones no censuradas y el modelo Heckman.
- 5) ¿En qué situaciones es más apropiado utilizar Heckman frente a: a) realizar una estimación MCO usando las observaciones no censuradas b) usar un modelo Tobit.
- 6) ¿Qué se entiende por restricciones de exclusión y por qué es importante analizar su validez en el contexto del modelo de Heckman?
- 7) Describa como obtener los efectos parciales asociados a la educación a partir de las distintas estimaciones. Considere los efectos parciales en:
 - a. El modelo latente.
 - b. La esperanza condicional en el conjunto de la población $E(y/X)$.
 - c. La esperanza condicional en las observaciones no censuradas $E(y/X, y>0)$.

Parte II:

Considere ahora que lo que quiere estimarse es el efecto de la educación en los salarios de las mujeres casadas. Revise los puntos 1 a 5 explicando cómo cambiarían las respuestas obtenidas para el caso de las horas trabajadas. Indique en cada caso si los métodos utilizados pueden ser aplicados al caso de los salarios, en caso de no ser así indique si hay otro(s) método(s) más apropiados.

EJERCICIO 2

Con datos de una muestra de 753 mujeres casadas procedentes del archivo mroz (ver Wooldridge pag. 543 a 564) se procede a estimar un modelo para el salario (en logs). En primer lugar se estiman los parámetros a través de una regresión de mínimos cuadrados ordinarios en la submuestra de mujeres trabajadoras. Después se estima un modelo bietapico a la Heckman. Los resultados son los siguientes (errores estándar entre

paréntesis). No se reportan los resultados de la primera etapa, o sea de un modelo Probit para la decisión de trabajar, esta estimación está informada en el Práctico 4.

Ecuación salarial: log(salario)		
	MCO	Heckman
Educación	0.107 (0.014)	0.109 (0.016)
Experiencia	0.042 (0.013)	0.044 (0.016)
Experiencia al cuadrado	-0.001 (0.000)	-0.001 (0.000)
Constante	-0.522 (0.199)	-0.578 (0.305)
λ		0.032 (0.134)
Núm. Observaciones	428	428

SE PIDE:

- 1) Comente las diferencias encontradas en las dos estimaciones
- 2) ¿Sugieren las estimaciones la existencia de un sesgo de selección endógeno?

EJERCICIO 3. (En base a Examen Febrero 2013)

Se poseen datos sobre **resultados obtenidos en pruebas generales de aptitud** académica para ingreso a universidades. La prueba está compuesta por dos ejercicios que tienen un máximo de 400 puntos cada uno. En cada ejercicio, los estudiantes pueden obtener un “bonus” de 50 puntos extra, por lo que en cada ejercicio podrían acceder hasta 450 puntos si logran respuestas correctas y completas.. Sin embargo, estos puntos extra sólo se dan para aquellos estudiantes que acumulando los dos ejercicios obtengan un puntaje inferior a 800, por lo que la prueba tiene un puntaje máximo de 800 puntos.

En base a ellos se busca determinar la relación entre la **aptitud** académica de una persona y su desempeño en áreas de lectura, matemática y la orientación que definen en su bachillerato.

Para este análisis se cuenta con las siguientes variables:

apt: Resultado en la prueba de aptitud (máximo 800)

matemática: Indicador de desempeño en el área matemática (puntaje)

lectura: Indicador de desempeño en el área de lectura y escritura (puntaje)

bac1: Toma valor 1 si la persona elige la orientación denominada “Bachillerato 1”

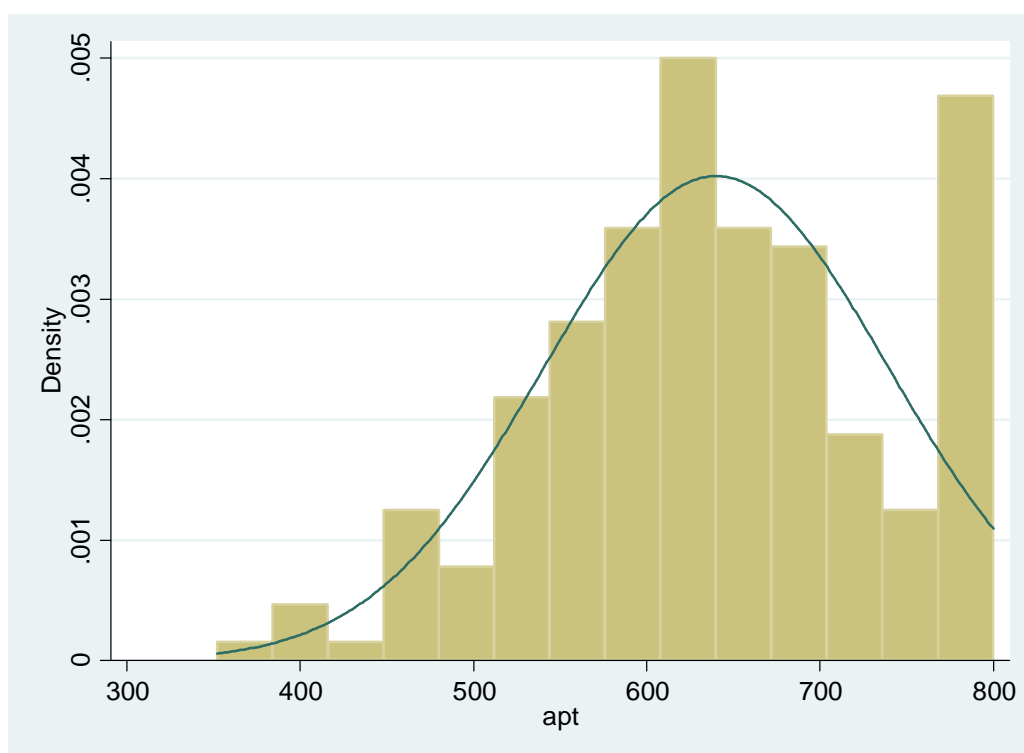
bac2: Toma valor 1 si la persona elige la orientación denominada “Bachillerato 2”

bac3: Toma valor 1 si la persona elige la orientación denominada “Bachillerato 3”

A continuación se presentan estadísticas descriptivas para estas variables

Variable	Obs	Mean	Std. Dev.	Min	Max
apt	200	640.035	99.21903	352	800
lectura	200	52.23	10.25294	28	76
matemática	200	52.645	9.368448	33	75

En el siguiente gráfico se presenta el histograma de la variable apt junto con la densidad Normal



PARTE 1

1. ¿Entiende que la variable *apt* puede presentar algún tipo de censura o truncamiento? Explique. Si entiende que se presenta un problema de censura o truncamiento, diga qué tipo de censura o truncamiento se observa y en qué punto(s). Aporte un ejemplo diferente donde se presenta el mismo tipo de cesura o truncamiento. ¿Definiría este problema como uno asociado a una solución de esquina o a recolección de datos?

2. Proponga una estrategia de estimación adecuada, desarrollando la función de verosimilitud individual y para el conjunto de la muestra (recuerde plantear las ecuaciones de partida del modelo teórico y los supuestos que le permitan desarrollar la función de verosimilitud).
3. Comente brevemente la principal diferencia que encuentra entre la función de verosimilitud planteada anteriormente y la asociada al modelo de regresión lineal (no es necesario plantear la función de verosimilitud del modelo lineal).
4. **PARTE 2**
 1. Debajo se presentan los resultados de la estimación Tobit así como el cálculo de los efectos marginales. ¿Es la estimación Tobit apropiada para el caso bajo estudio? Justifique.
 2. Comente los resultados asociados a la salida del modelo Tobit
 3. Sabiendo que el valor de la log-verosimilitud para el modelo Tobit que sólo incluye la constante es -1135.545 pruebe la significación conjunta del modelo, especificando hipótesis nula y alternativa, estadístico de prueba, región crítica, valor del estadístico y conclusión
 4. En cuanto a los efectos marginales:
 - a. Explique cuáles son los diferentes tipos de efectos marginales que se pueden obtener en los modelos Tobit y relaciónelos con las salidas correspondientes.
 - b. Comente los efectos parciales que le parezcan más interesantes en el caso bajo estudio justificando su elección. (2 puntos)

```

. tobit apt lecturamatematica bac2 bac3, ul(800)
Tobit regression                               Number of obs   =          200
                                                LR chi2(4)        =       188.97

Prob> chi2      =          0.0000
Log likelihood = -1041.0629                    Pseudo R2        =       0.0832

-----+-----
apt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
lectura |  2.697939   .618798     4.36  0.000     1.477582    3.918296
matematica |  5.914485   .7098063    8.33  0.000     4.514647    7.314323
      bac2 | -12.71476   12.40629    -1.02  0.307    -37.18173    11.7522
      bac3 | -46.1439    13.72401    -3.36  0.001    -73.2096   -19.07821
      _cons |  209.566    32.77154     6.39  0.000    144.9359   274.1961
-----+-----
      /sigma |  65.67672   3.481272                58.81116   72.54228
-----+-----

Obs. summary:          0      left-censored observations
                    183      uncensored observations
17      right-censored observations at apt>=800

```

Efectos parciales sobre variable latente

```

. margins, dydx(lecturamatematica bac2 bac3)
Average marginal effects                               Number of obs   =          200
Model VCE      : OIM

Expression      : Linear prediction, predict()
dy/dx w.r.t.    : lecturamatematica bac2 bac3

```

```

-----+-----
              |      Delta-method
              |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
lectura |  2.697939   .618798     4.36  0.000     1.485118    3.910761
matematica |  5.914485   .7098063    8.33  0.000     4.52329    7.30568
      bac2 | -12.71476   12.40629    -1.02  0.305    -37.03065    11.60113
      bac3 | -46.1439    13.72401    -3.36  0.001    -73.04248   -19.24533
-----+-----

```

Efectos parciales sobre variable observada

```
. margins, dydx(lecturamatematica bac2 bac3) predict(e(.,800))
```

Average marginal effects

Number of obs = 200

Model VCE : OIM

Expression : E(apr|apr<800), predict(e(.,800))

dy/dx w.r.t. : lecturamatematica bac2 bac3

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
lectura	2.293154	.5220864	4.39	0.000	1.269884	3.316425
matematica	5.027106	.5842984	8.60	0.000	3.881902	6.172309
bac2	-10.8071	10.53928	-1.03	0.305	-31.46371	9.849497
bac3	-39.22071	11.6335	-3.37	0.001	-62.02194	-16.41947

Efectos parciales sobre variable condicional a que se trata de observaciones no censuradas/truncadas

```
. margins, dydx(lecturamatematica bac2 bac3) predict(ystar(.,800))
```

Average marginal effects

Number of obs = 200

Model VCE : OIM

Expression : E(apr*|apr<800), predict(ystar(.,800))

dy/dx w.r.t. : lecturamatematica bac2 bac3

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
lectura	2.500675	.5681643	4.40	0.000	1.387093	3.614256
matematica	5.482037	.6329752	8.66	0.000	4.241428	6.722646
bac2	-11.7851	11.50087	-1.02	0.305	-34.3264	10.7562
bac3	-42.77001	12.74277	-3.36	0.001	-67.74538	-17.79465