Juan Menduiña

Trabajo Práctico N° 1:
Modelo de Probabilidad Lineal, Logit y Probit.

Ejercicio 1: Porcentaje Correctamente Predicho.

Sea y una variable binaria y considerar algún modelo de probabilidad P (y=1|x)= F (Xβ). Mostrar que el porcentaje general predicho correctamente es un promedio ponderado del porcentaje predicho para la variable dependiente igual a 0 ($\hat{q}_0$) y del porcentaje predicho para la variable dependiente igual a 1 ($\hat{q}_1$), donde las ponderaciones son las proporciones de ceros y de unos en la muestra, respectivamente.

$$\hat{q}_0 = \frac{cantidad\ de\ observaciones\ correctamente\ predichas\ cuando\ y=0}{cantidad\ de\ observaciones\ con\ y=0} = \frac{A}{n_0}.$$

$$\hat{q}_1 = \frac{cantidad\ de\ observaciones\ correctamente\ predichas\ cuando\ y=1}{cantidad\ de\ observaciones\ con\ y=1} = \frac{B}{n_1}.$$

$$\hat{q} = \frac{cantidad\ de\ observaciones\ correctamente\ predichas}{cantidad\ de\ observaciones} = \frac{A+B}{n_0+n_1}$$

$$\hat{q} = \frac{A+B}{n_0+n_1}$$

$$\hat{q} = \frac{n_0\hat{q}_0+n_1\hat{q}_1}{n_0+n_1}$$

$$\hat{q} = \frac{n_0}{n_0+n_1}\hat{q}_0 + \frac{n_1}{n_0+n_1}\hat{q}_1.$$

Ejercicio 2: Interpretación del Modelo de Probabilidad Lineal I.

Suponer que se estima el modelo:

$y_i = \beta_0 + \beta_1 x_i + u_i$,

donde x es es una variable continua, mientras que y es una variable que sólo puede valer 0 o 1. El tamaño de la muestra es n y sea n_1 *la cantidad de elementos que verican* $y_i = 1$. *Llamar* $\bar{x}_1$ *a la media de la variable x tomada sólo para aquellos elementos que verican* $y_i = 1$ *y* $\bar{x}_0$ *a la media de la variable x tomada sobre los valores restantes. Mostrar que:*

$\hat{\beta}_1 = \frac{p(1-p)(\bar{x}_1 - \bar{x}_0)}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$,

donde $p = \frac{n_1}{n}$.

Partiendo del estimador de Mínimos Cuadrados Ordinarios (MCO) para el parámetro de pendiente (β_1) de este modelo, se tiene:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \frac{n_1}{n})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - x_i\frac{n_1}{n}}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\frac{n_1}{n})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}(\sum_{i=1}^{n_1} x_i y_i - \frac{n_1}{n}\sum_{i=1}^{n} x_i)}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}[\sum_{i=1}^{n_1} x_i - \frac{n_1}{n}(n_0\bar{x}_0 + n_1\bar{x}_1)]}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}(n_1\bar{x}_1 - \frac{1}{n}n_1 n_0\bar{x}_0 - \frac{n_1^2}{n}\bar{x}_1)}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}[n_1\bar{x}_1 - p(n-n_1)\bar{x}_0 - pn_1\bar{x}_1]}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}n_1\bar{x}_1 - \frac{1}{n}p(n-n_1)\bar{x}_0 - \frac{1}{n}pn_1\bar{x}_1}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{p\bar{x}_1 - p(1-p)\bar{x}_0 - p^2\bar{x}_1}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{p(1-p)\bar{x}_1 - p(1-p)\bar{x}_0}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \frac{p(1-p)(\bar{x}_1 - \bar{x}_0)}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

<u>Ejercicio 3:</u> Interpretación del Modelo de Probabilidad Lineal II.

Sea y una resultado binario y sean d_1, d_2, … , d_M variables binarias mutuamente excluyentes y colectivamente exhaustivas, es decir, cada persona de la población cae en una y sólo una categoría.

(a) *Mostrar que los valores ajustados de la regresión sin intercepto y_i sobre d_{1i}, d_{2i}, … , d_{Mi} están siempre en el intervalo unitario. En particular, describir qué representa cada coeficiente y el valor ajustado para cada i.*

Cada coeficiente (1 .. k) representa la proporción de observaciones que tienen un resultado binario igual a 1 (y= 1) cuando la variable binaria independiente en cuestión es igual a 1 (d_k= 1), es decir, $\bar{y}_k = \frac{\sum_{i=1}^{m_k} y_i}{m_k}$ (proporcion de "éxitos" de cada categoría), siendo m_k la cantidad de observaciones con d_k= 1, k= 1, … , M.

El valor ajustado para cada i corresponde al coeficiente asociado a la variable d_k que para esa observación sea igual a 1.

(b) *¿Qué ocurre si y_i se regresa sobre M combinaciones lineales de d_{1i}, d_{2i}, … , d_{Mi} linealmente independientes entre sí? Ayuda: Considerar 1, d_2, … , d_M.*

Lo que ocurre si y_i se regresa sobre M combinaciones lineales de d_{1i}, d_{2i}, … , d_{Mi} linealmente independientes entre sí es que se omite una de las variables independientes porque existe multicolinealidad perfecta entre el intercepto y la combinación lineal de las variables independientes (mutuamente excluyentes y colectivamente exhaustivas).

<u>Ejercicio 4:</u> Efectos Marginales.

Sea y un resultado binario y x= (x_1, … , x_k) un vector de variables explicativas. Sea G (.) la función de distribución acumulada de una variable aleatoria continua. Recordar que, si x_j es continua, su efecto marginal se obtiene como:

$\frac{\partial p(x)}{\partial x_j}$= *g* ($\beta_0$ + xβ) β_j, *donde g (z)=* $\frac{\partial G}{\partial z}$ *(z).*

(a) *Mostrar que los efectos relativos de dos variables explicativas cualesquiera no dependen de x.*

$\frac{\partial p(x)}{\partial x_1}$= g ($\beta_0$ + xβ) β_1
$\frac{\partial p(x)}{\partial x_2}$= g ($\beta_0$ + xβ) β_2

$\frac{\frac{\partial p(x)}{\partial x_1}}{\frac{\partial p(x)}{\partial x_2}} = \frac{g(\beta_0+x\beta)\beta_1}{g(\beta_0+x\beta)\beta_2}$
$\frac{\frac{\partial p(x)}{\partial x_1}}{\frac{\partial p(x)}{\partial x_2}} = \frac{\beta_1}{\beta_2}$.

Por lo tanto, los efectos relativos de dos variables explicativas cualesquiera no dependen de x.

(b) *Sea x_1 una variable binaria. ¿Cuál es el efecto parcial de cambiar x_1 de 0 a 1? ¿De qué depende? Interpretar en el caso en el que y es un indicador de empleo y x_1 es una variable binaria que indica la participación en un programa de capacitación laboral.*

El efecto parcial de cambiar x_1 de 0 a 1 es:

$\frac{\partial p(x)}{\partial x_1}$= P (y= 1 | x_1= 1) - P (y= 1 | x_1= 0)
$\frac{\partial p(x)}{\partial x_1}$= g ($\beta_0$ + xβ) β_1,

que depende de la función de densidad de la variable aleatoria continua y del coeficiente β_1.

En el caso en el que y es un indicador de empleo y x_1 es una variable binaria que indica la participación en un programa de capacitación laboral, este efecto parcial indica en cuánto varía, *céteris páribus*, la probabilidad de obtener empleo al participar en un programa de capacitación laboral respecto a no participar.

(c) *Sea x_2 una variable discreta numérica. ¿Cuál es el efecto parcial de cambiar x_2 de cierto nivel c a c + 1? ¿De qué depende? Interpretar en el caso en el que y es un indicador de si la persona i fuma y x_2 la cantidad de cigarrillos que fuma por día.*

$\frac{\partial p(x)}{\partial x_2}$= P (y= 1 | x_2= c + 1) - P (y= 1 | x_2= c)
$\frac{\partial p(x)}{\partial x_2}$= g ($\beta_0$ + xβ) β_2,

que depende de la función de densidad de la variable aleatoria continua y del coeficiente β_2.

En el caso en el que y es un indicador de si la persona i fuma y x_2 la cantidad de cigarrillos que fuma por día, este efecto parcial indica en cuánto varía, *céteris páribus*, la probabilidad de que la persona i fume cuando la cantidad de cigarrillos que fuma por día aumenta en una unidad.

Considerar, ahora, el siguiente modelo:

P (y= 1 | z)= G ($\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3$ log (z_2) + $\beta_4 z_3$).

(d) *¿Cuál es el efecto parcial de z_1 sobre P (y= 1 | z)?*

El efecto parcial de z_1 sobre P (y= 1 | z) es:

$\frac{\partial P\,(y=1 \mid z)}{\partial z_1}$= g ($\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3$ log (z_2) + $\beta_4 z_3$) β_1.

(e) *¿Cuál es el efecto parcial de z_2 sobre P (y= 1 | z)?*

El efecto parcial de z_2 sobre P (y= 1 | z) es:

$\frac{\partial P\,(y=1 \mid z)}{\partial z_2}$= g ($\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3$ log (z_2) + $\beta_4 z_3$) $\beta_3 \frac{1}{z_2}$.

(f) *¿Cuál es la elasticidad de z_3 sobre P (y= 1 | z)? ¿Siempre tiene el mismo signo que β_4?*

La elasticidad de z_3 sobre P (y= 1 | z) es:

$$\varepsilon_{z_3} = \frac{\partial P\,(y=1 \mid z)}{\partial z_3} \frac{z_3}{P\,(y=1 \mid z)} = g\,(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3)\,\beta_4 \frac{z_3}{G(\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3 log(z_2)+\beta_4 z_3)}.$$

No siempre tiene el mismo signo que β_4, ya que éste también depende del valor que tome z_3.

(g) *¿Cuál es la elasticidad de z_1 sobre P (y= 1 | z)?*

$$\varepsilon_{z_1} = \frac{\partial P\,(y=1 \mid z)}{\partial z_1}\,\frac{z_1}{P\,(y=1 \mid z)} = g\,(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3)\,\beta_1 \frac{z_1}{G(\beta_0+\beta_1 z_1+\beta_2 z_1^2+\beta_3 log(z_2)+\beta_4 z_3)}.$$

(h) *¿Cómo se obtendrían errores estándar para todos estos efectos?*

Los errroes estándar para todos estos efectos se pueden obtener utilizando la matriz de varianzas y covarianzas de los coeficientes estimados del modelo, mediante métdos analíticos, siempre que la distribución de los estimadores sea conocida, o mediante métodos de remuestreo, siempre que la distribución de los estimadores no sea conocida.

<u>Ejercicio 5:</u> MPL, Logit y Probit en Stata I.

En este ejercicio, se van a demostrar algunas propiedades de las estimaciones para modelos con variable dependiente discreta.

(a) *Estimar a ins contra retire, age, hstatusg, hhincome, educyear, married, hisp por OLS, Logit y Probit.*

<u>OLS:</u>

```
      Source |       SS           df       MS      Number of obs   =     3,206
-------------+----------------------------------   F(7, 3198)      =     41.14
       Model |  62.8403396         7  8.97719137   Prob > F        =    0.0000
    Residual |   697.78505     3,198    .2181942   R-squared       =    0.0826
-------------+----------------------------------   Adj R-squared   =    0.0806
       Total |   760.62539     3,205  .237324615   Root MSE        =    .46711

------------------------------------------------------------------------------
         ins | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.retire |   .0408508   .0182197     2.24   0.025     .0051273    .0765743
         age |  -.0028955   .0024189    -1.20   0.231    -.0076383    .0018473
  1.hstatusg |   .0655583   .0194531     3.37   0.001     .0274166    .1037001
    hhincome |   .0004921   .0001375     3.58   0.000     .0002225    .0007617
    educyear |   .0233686   .0028672     8.15   0.000      .017747    .0289903
   1.married |   .1234699   .0193618     6.38   0.000     .0855071    .1614326
      1.hisp |  -.1210059    .033666    -3.59   0.000     -.187015   -.0549969
       _cons |   .1270857   .1605628     0.79   0.429    -.1877308    .4419021
------------------------------------------------------------------------------
```

<u>Logit:</u>

```
Logistic regression                                     Number of obs =  3,206
                                                        LR chi2(7)    = 289.79
                                                        Prob > chi2   = 0.0000
Log likelihood = -1994.8784                             Pseudo R2     = 0.0677

------------------------------------------------------------------------------
         ins | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.retire |   .1969297   .0842067     2.34   0.019     .0318875    .3619718
         age |  -.0145955   .0112871    -1.29   0.196    -.0367178    .0075267
  1.hstatusg |   .3122654   .0916739     3.41   0.001     .1325878     .491943
    hhincome |   .0023036    .000762     3.02   0.003       .00081    .0037972
    educyear |   .1142626   .0142012     8.05   0.000     .0864288    .1420963
   1.married |    .578636   .0933198     6.20   0.000     .3957327    .7615394
      1.hisp |  -.8103059   .1957522    -4.14   0.000    -1.193973   -.4266387
       _cons |  -1.715578   .7486219    -2.29   0.022     -3.18285   -.2483064
------------------------------------------------------------------------------
```

Probit:

```
Probit regression                                       Number of obs =  3,206
                                                        LR chi2(7)    = 292.30
                                                        Prob > chi2   = 0.0000
Log likelihood = -1993.6237                             Pseudo R2     = 0.0683

------------------------------------------------------------------------------
         ins | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.retire |   .1183567   .0512678     2.31   0.021     .0178736    .2188397
         age |  -.0088696    .006899    -1.29   0.199    -.0223914    .0046521
  1.hstatusg |   .1977357   .0554868     3.56   0.000     .0889835    .3064878
    hhincome |    .001233   .0003866     3.19   0.001     .0004754    .0019907
    educyear |   .0707477   .0084782     8.34   0.000     .0541308    .0873647
   1.married |    .362329   .0560031     6.47   0.000      .252565    .4720931
      1.hisp |  -.4731099   .1104393    -4.28   0.000     -.689567   -.2566529
       _cons |  -1.069319   .4580794    -2.33   0.020    -1.967139   -.1715002
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS	(2) Logit	(3) Probit
main			
0.retire	0	0	0
	(.)	(.)	(.)
1.retire	0.0409**	0.197**	0.118**
	(0.0182)	(0.0842)	(0.0513)
age	-0.00290	-0.0146	-0.00887
	(0.00242)	(0.0113)	(0.00690)
0.hstatusg	0	0	0
	(.)	(.)	(.)
1.hstatusg	0.0656***	0.312***	0.198***
	(0.0195)	(0.0917)	(0.0555)
hhincome	0.000492***	0.00230***	0.00123***
	(0.000138)	(0.000762)	(0.000387)
educyear	0.0234***	0.114***	0.0707***
	(0.00287)	(0.0142)	(0.00848)
0.married	0	0	0
	(.)	(.)	(.)
1.married	0.123***	0.579***	0.362***
	(0.0194)	(0.0933)	(0.0560)
0.hisp	0	0	0
	(.)	(.)	(.)
1.hisp	-0.121***	-0.810***	-0.473***
	(0.0337)	(0.196)	(0.110)
_cons	0.127	-1.716**	-1.069**
	(0.161)	(0.749)	(0.458)
N	3206	3206	3206
R-sq	0.083		
pseudo R-sq		0.068	0.068

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

(b) *¿Cuál es el problema de estimar el modelo por OLS?*

Los problemas de estimar el modelo por OLS son que los valores estimados de la variable dependiente pueden caer fuera del rango [0, 1] y que los errores del modelo son heterocedásticos, lo cual resulta en estimadores ineficientes.

(c) *Explicar, analíticamente, cuál es la interpretación de un coeficiente β en un modelo de regresión lineal y en un modelo Probit/Logit. ¿Es constante el efecto marginal en los modelos no lineales?*

La interpretación de un coeficiente β en un modelo de regresión lineal es cuánto afecta un cambio en la variable independiente a la probabilidad de y= 1 (es decir, corresponde al efecto marginal, constante), mientras que, en un modelo Probit/Logit, es parte del efecto marginal, ya que, ahora, el efecto marginal refleja las diferentes pendientes de la curva, por lo que no es constante en los modelos no lineales.

(d) *Para evaluar la eficacia de los modelos Probit y Logit, definir el valor estimado de la variable dependiente y como:*

$$\hat{y}=\begin{cases}1, si\ P\ (\hat{y}=1)>0{,}5\\ 0, si\ P(\hat{y}=0)\leq 0{,}5\end{cases}.$$

Realizar un cuadro de doble entrada con las variables y y $\hat{y}$. Comentar.

ins	yhat_probit 0	1	Total
0	1,660	305	1,965
1	906	335	1,241
Total	2,566	640	3,206

(e) *En la literatura, se sugiere que $\beta^{logit} \approx 4\beta^{ols}$ y $\beta^{probit} \approx 2{,}5\beta^{ols}$. Comprobarlo para esta muestra.*

```
prueba_logit[12,2]
                  Betas Logit  4 * Betas ~S
  ins:0b.retire             0             0
   ins:1.retire     .19692966     .16340327
        ins:age    -.01459553    -.01158219
ins:0b.hstatusg             0             0
 ins:1.hstatusg     .31226537     .26223337
   ins:hhincome      .0023036     .00196835
   ins:educyear     .11426256     .09347452
 ins:0b.married             0             0
  ins:1.married     .57863605     .49387952
    ins:0b.hisp             0             0
     ins:1.hisp    -.81030593    -.48402374
      ins:_cons    -1.7155784     .50834278

prueba_probit[12,2]
                 Betas Probit  2,5 * Beta~S
  ins:0b.retire             0             0
   ins:1.retire     .11835665     .10212704
        ins:age    -.00886962    -.00723887
ins:0b.hstatusg             0             0
 ins:1.hstatusg     .19773566     .16389585
   ins:hhincome     .00123304     .00123022
   ins:educyear     .07074775     .05842157
 ins:0b.married             0             0
  ins:1.married     .36232905      .3086747
    ins:0b.hisp             0             0
     ins:1.hisp    -.47310993    -.30251484
      ins:_cons    -1.0693194     .31771424
```

(f) *Computar la probabilidad esperada que ins= 1 cuando las variables están evaluadas en la media.*

La probabilidad esperada que ins= 1 cuando las variables están evaluadas en la media es:

- en el modelo OLS, 0,387;
- en el modelo Logit, 0,373; y
- en el modelo Probit, 0,374.

(g) *Definir el odds ratio como el cociente entre la probabilidad que y= 1 y y= 0. De este modo, un odds ratio de 2 implica que es dos veces más probable que y= 1 a que y= 0. Demostrar que, para el caso de un modelo Logit, se verifica que:*

$\ln\left(\frac{P(y=1|x)}{P(y=0|x)}\right) = X\beta.$

Recordar que para un modelo Logit:

$P\ (y=1 \mid x) = \frac{1}{1+e^{-X\beta}}$.

$P\ (y=1 \mid x) = \frac{e^{X\beta}}{1+e^{X\beta}}$

$P\ (y=1 \mid x) = \frac{e^{X\beta}}{e^{X\beta}(\frac{1}{e^{X\beta}}+1)}$

$P\ (y=1 \mid x) = \frac{1}{1+\frac{1}{e^{X\beta}}}$

$P\ (y=1 \mid x) = \frac{1}{1+e^{-X\beta}}$.

$P\ (y=0 \mid x) = 1 - P\ (y=1 \mid x)$

$P\ (y=0 \mid x) = 1 - \frac{1}{1+e^{-X\beta}}$

$P\ (y=0 \mid x) = \frac{1+e^{-X\beta}-1}{1+e^{-X\beta}}$

$P\ (y=0 \mid x) = \frac{e^{-X\beta}}{1+e^{-X\beta}}$.

$\frac{P(y=1|x)}{P(y=0|x)} = \frac{\frac{1}{1+e^{-X\beta}}}{\frac{e^{-X\beta}}{1+e^{-X\beta}}}$

$\frac{P(y=1|x)}{P(y=0|x)} = \frac{1}{e^{-X\beta}}$

$\frac{P(y=1|x)}{P(y=0|x)} = e^{X\beta}$

$\ln\left(\frac{P(y=1|x)}{P(y=0|x)}\right) = \ln e^{X\beta}$

$\ln\left(\frac{P(y=1|x)}{P(y=0|x)}\right) = X\beta \ln e$

$\ln\left(\frac{P(y=1|x)}{P(y=0|x)}\right) = X\beta * 1$

$\ln\left(\frac{P(y=1|x)}{P(y=0|x)}\right) = X\beta$.

Ejercicio 6: MPL, Logit y Probit en Stata II.

Utilizar la base de datos de Mroz, T. A. (1987): "The Sensitiviy of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", Econometrica, 55, 765-799. La misma posee datos sobre el desempleo de las mujeres en Estados Unidos en 1975.

(a) *Para comenzar, realiza un análisis exploratorio simple de los datos. Para esto, se puede ayudar de los comandos describe, summarize, browse, tab.*

Variable	Obs	Mean	Std. dev.	Min	Max
inlf	753	.5683931	.4956295	0	1
hours	753	740.5764	871.3142	0	4950
kidslt6	753	.2377158	.523959	0	3
kidsge6	753	1.353254	1.319874	0	8
age	753	42.53785	8.072574	30	60
educ	753	12.28685	2.280246	5	17
wage	753	2.374565	3.241829	0	25
repwage	753	1.849734	2.419887	0	9.98
hushrs	753	2267.271	595.5666	175	5010
husage	753	45.12085	8.058793	30	60
huseduc	753	12.49137	3.020804	3	17
huswage	753	7.482179	4.230559	.4121	40.509
faminc	753	23080.59	12190.2	1500	96000
mtr	753	.6788632	.0834955	.4415	.9415
motheduc	753	9.250996	3.367468	0	17
fatheduc	753	8.808765	3.57229	0	17
unem	753	8.623506	3.114934	3	14
city	753	.6427623	.4795042	0	1
exper	753	10.63081	8.06913	0	45
nwifeinc	753	20.12896	11.6348	-.0290575	96
lwage	428	1.190173	.7231978	-2.054164	3.218876
expersq	753	178.0385	249.6308	0	2025

(b) *Crear una variable de educación centrada. Recordar que se le llama variable centrada a una variable transformada como* $\tilde{x}_i = x_i - \bar{x}$.

Variable	Obs	Mean	Std. dev.	Min	Max
educ	753	12.28685	2.280246	5	17
educ_cent	753	-165.7517	2.280246	-173.0385	-161.0385

(c) *Estudiar, gráficamente, la relación entre el salario y la educación. Se puede también desagregar por las variables inlf, kidslt6. Para esto, se puede ayudar de los comandos graph, twoway, scatter, lfit y sus opciones.*

Stata.

(d) *¿Hay valores faltantes o duplicados en la muestra? Intentar resolver esto sin el comando browse ni edit.*

Variable	Missing	Total	Percent Missing
inlf	0	753	0.00
hours	0	753	0.00
kidslt6	0	753	0.00
kidsge6	0	753	0.00
age	0	753	0.00
educ	0	753	0.00
wage	0	753	0.00
repwage	0	753	0.00
hushrs	0	753	0.00
husage	0	753	0.00
huseduc	0	753	0.00
huswage	0	753	0.00
faminc	0	753	0.00
mtr	0	753	0.00
motheduc	0	753	0.00
fatheduc	0	753	0.00
unem	0	753	0.00
city	0	753	0.00
exper	0	753	0.00
nwifeinc	0	753	0.00
lwage	325	753	43.16
expersq	0	753	0.00
educ_cent	0	753	0.00

Sí, en la variable *lwage*, hay 325 valores faltantes en la muestra de 753 observaciones. No, no hay valores duplicados en la muestra.

(e) *Estimar un modelo de probabilidad lineal de inlf sobre educ, city, exper, kidslt6, expersq. Además, generar la predicción del modelo.*

OLS:

```
      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(5, 747)       =     37.62
       Model |  37.1605056         5  7.43210111   Prob > F        =    0.0000
    Residual |   147.56725       747   .19754652   R-squared       =    0.2012
-------------+----------------------------------   Adj R-squared   =    0.1958
       Total |  184.727756       752  .245648611   Root MSE        =    .44446

------------------------------------------------------------------------------
        inlf | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0388373   .0073171     5.31   0.000     .0244729    .0532018
        city |  -.0574649   .0343425    -1.67   0.095    -.1248842    .0099544
       exper |   .0444919   .0058467     7.61   0.000      .033014    .0559698
     kidslt6 |  -.1691606    .031841    -5.31   0.000    -.2316691   -.1066522
     expersq |  -.0009058   .0001881    -4.82   0.000    -.0012751   -.0005366
       _cons |  -.1433578   .0917196    -1.56   0.118    -.3234167     .036701
------------------------------------------------------------------------------
```

(f) *¿Se puede realizar inferencia con este modelo? Estimar el modelo con errores estándares robustos. ¿Cómo cambian los resultados?*

OLS (con errores estándar robustos):

```
Linear regression                               Number of obs     =        753
                                                F(5, 747)         =      52.82
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2012
                                                Root MSE          =     .44446

------------------------------------------------------------------------------
             |               Robust
        inlf | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0388373   .0069696     5.57   0.000     .0251549    .0525197
        city |  -.0574649   .0342117    -1.68   0.093    -.1246275    .0096976
       exper |   .0444919   .0055926     7.96   0.000     .0335128     .055471
     kidslt6 |  -.1691606   .0300823    -5.62   0.000    -.2282165   -.1101047
     expersq |  -.0009058   .0001738    -5.21   0.000     -.001247   -.0005647
       _cons |  -.1433578   .0852798    -1.68   0.093    -.3107744    .0240588
------------------------------------------------------------------------------
```

Sí, se puede realizar inferencia con este modelo. Si se estima el modelo con errores estándares robustos, mejora la significatividad estadística de las variables.

(g) *¿Qué ocurre si se elimina la constante del modelo?*

OLS (con errores estándar robustos y sin constate):

```
Linear regression                               Number of obs     =        753
                                                F(5, 748)         =     310.35
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6541
                                                Root MSE          =     .44489

------------------------------------------------------------------------------
             |               Robust
        inlf | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |    .028748   .0035986     7.99   0.000     .0216835    .0358125
        city |  -.0617278   .0340414    -1.81   0.070    -.1285558    .0051002
       exper |   .0425785    .005629     7.56   0.000     .0315281     .053629
     kidslt6 |  -.1700338   .0300221    -5.66   0.000    -.2289713   -.1110963
     expersq |  -.0008588   .0001749    -4.91   0.000    -.0012023   -.0005154
------------------------------------------------------------------------------
```

Lo que ocurre si se elimina la constante del modelo es que aumenta la significatividad estadística de la variable *city*.

(h) *¿Qué ocurre si estima el modelo sólo para una ciudad?*

OLS (con errores estándar robustos y sólo para una ciudad):

```
Linear regression                               Number of obs     =        484
                                                F(4, 479)         =      46.75
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2065
                                                Root MSE          =     .44379

------------------------------------------------------------------------------
             |               Robust
        inlf | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0413565   .0090158     4.59   0.000     .0236411    .0590718
        city |          0  (omitted)
       exper |   .0497399   .0068528     7.26   0.000     .0362745    .0632052
     kidslt6 |  -.1426504   .0416024    -3.43   0.001    -.2243963   -.0609046
     expersq |  -.0009985   .0002023    -4.94   0.000     -.001396    -.000601
       _cons |  -.2781658   .1143471    -2.43   0.015    -.5028497    -.053482
------------------------------------------------------------------------------
```

Lo que ocurre si se estima el modelo sólo para una ciudad es que se omite la variable *city* porque existe multicolinealidad perfecta entre el intercepto del modelo y esta variable.

(i) *Estimar un modelo Logit de inlf sobre educ, city, exper, kidslt6, expersq.*

Logit:

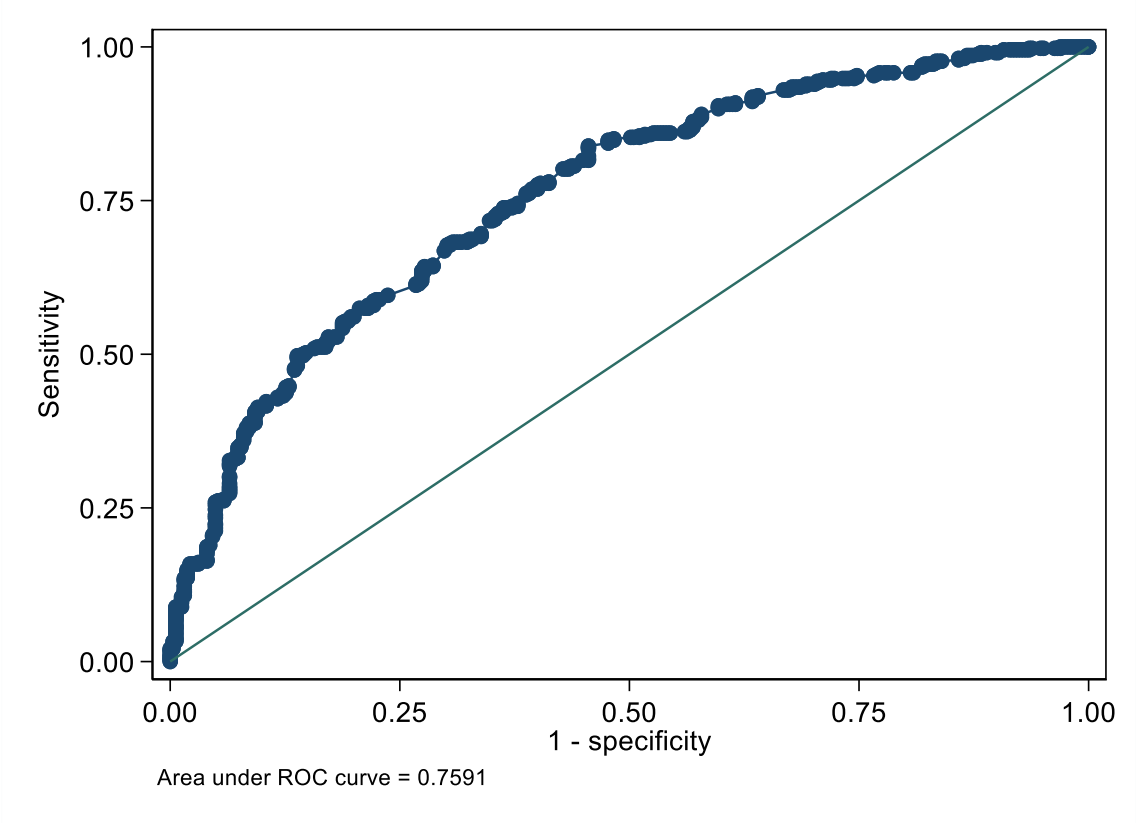
```
Logistic regression                                     Number of obs =    753
                                                        LR chi2(5)    = 163.38
                                                        Prob > chi2   = 0.0000
Log likelihood = -433.18195                             Pseudo R2     = 0.1587

------------------------------------------------------------------------------
        inlf | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .1991157    .039264     5.07   0.000     .1221596    .2760717
        city |  -.2786654    .176285    -1.58   0.114    -.6241777    .0668469
       exper |   .2041167   .0302627     6.74   0.000      .144803    .2634304
     kidslt6 |  -.8274419   .1684161    -4.91   0.000    -1.157531   -.4973525
     expersq |  -.0040423   .0009801    -4.12   0.000    -.0059633   -.0021213
       _cons |  -3.199722   .5019472    -6.37   0.000     -4.18352   -2.215924
------------------------------------------------------------------------------
```

(j) *Calcular la predicción del modelo.*

Stata.

(k) *Generar la curva ROC.*



(l) *Calcular los efectos marginales en las medias.*

Efectos marginales (condicionales en las medias) en Logit:

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: Pr(inlf), predict()
dy/dx wrt:  educ city exper kidslt6 expersq
At: educ    = 12.28685 (mean)
    city    = .6427623 (mean)
    exper   = 10.63081 (mean)
    kidslt6 = .2377158 (mean)
    expersq = 178.0385 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0485166   .0095555     5.08   0.000     .0297881    .0672452
        city |  -.0678998   .0429316    -1.58   0.114    -.1520443    .0162447
       exper |   .0497352    .007403     6.72   0.000     .0352256    .0642448
     kidslt6 |   -.201615   .0411714    -4.90   0.000    -.2823095   -.1209206
     expersq |  -.0009849   .0002397    -4.11   0.000    -.0014547   -.0005152
------------------------------------------------------------------------------
```

(m) *Calcular los efectos marginales en valores particulares de la variable que le resulten de interés.*

Efectos marginales (condicionales en valores particulares) en Logit:

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: Pr(inlf), predict()
dy/dx wrt:  educ city exper kidslt6 expersq
At: educ    =  10
    city    =   1
    exper   =  20
    kidslt6 =   3
    expersq = 400

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0296194   .0096332     3.07   0.002     .0107386    .0485001
        city |  -.0414528   .0272418    -1.52   0.128    -.0948456    .0119401
       exper |   .0303633   .0117144     2.59   0.010     .0074035    .0533231
     kidslt6 |  -.1230858   .0197055    -6.25   0.000    -.1617079   -.0844637
     expersq |  -.0006013   .0002532    -2.37   0.018    -.0010976    -.000105
------------------------------------------------------------------------------
```

(n) *Estimar un modelo Probit con las mismas variables que en el inciso (i) y crear una tabla con las estimaciones de todos los modelos.*

Probit:

```
Probit regression                                       Number of obs =    753
                                                        LR chi2(5)    = 163.97
                                                        Prob > chi2   = 0.0000
Log likelihood = -432.88971                             Pseudo R2     = 0.1592

------------------------------------------------------------------------------
        inlf | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .1209674   .0231872     5.22   0.000     .0755213    .1664136
        city |   -.169242   .1051678    -1.61   0.108    -.3753671    .0368831
       exper |   .1251388   .0181038     6.91   0.000      .089656    .1606216
     kidslt6 |  -.5046704   .1003243    -5.03   0.000    -.7013024   -.3080385
     expersq |  -.0025089   .0005879    -4.27   0.000    -.0036611   -.0013567
       _cons |  -1.945429    .294419    -6.61   0.000    -2.522479   -1.368378
------------------------------------------------------------------------------
```

Tabla comparativa:

```
------------------------------------------------------------
                      (1)             (2)             (3)
                      OLS           Logit          Probit
------------------------------------------------------------
main
educ                0.0388***       0.199***        0.121***
                   (0.00697)       (0.0393)        (0.0232)

city                -0.0575*        -0.279          -0.169
                    (0.0342)        (0.176)         (0.105)

exper               0.0445***       0.204***        0.125***
                   (0.00559)       (0.0303)        (0.0181)

kidslt6             -0.169***       -0.827***       -0.505***
                    (0.0301)        (0.168)         (0.100)

expersq          -0.000906***    -0.00404***     -0.00251***
                  (0.000174)      (0.000980)      (0.000588)

_cons               -0.143*         -3.200***       -1.945***
                    (0.0853)        (0.502)         (0.294)
------------------------------------------------------------
N                     753             753             753
R-sq                0.201
pseudo R-sq                         0.159           0.159
------------------------------------------------------------
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
```

Ejercicio 7: Estimar el Efecto de la Educación sobre la Probabilidad de estar Desempleado.

Utilizar la EPH con datos de individuos del segundo trimestre de 2015, disponible en http://www.indec.gob.ar/bases-de-datos.asp. Usar la muestra de jefes de hogar, hombres, 25-65 años, para todos los conglomerados disponibles. Estudiar cómo se define el desempleo de acuerdo al INDEC. Rentrinjir la muestra a personas empleadas o desempleadas, es decir, excluir aquellos que están fuera de la fuerza laboral (no buscan trabajo, estudian, retirados, etc.). Usar las ponderaciones pondera.

(a) *Utilizar un modelo de probabilidad lineal para estimar el efecto de la educación sobre la probabilidad de estar desempleado, controlando por ubicación geográfica, edad y estado civil. Construir las probabilidades para cada individuo. ¿Qué proporción de la muestra tiene probabilidades predecidas mayores a 1 o menores a 0?*

Stata.

La proporción de la muestra que tiene probabilidades predecidas mayores a 1 y menores a 0 es 0 y 0,101, respectivamente.

(b) *Estimar el modelo del inciso (a) usando los modelos Probit y Logit. ¿Cómo cambian los resultados?*

Stata.

(c) *Estimar la probabilidad de estar desempleado para un hombre casado, para cada área metropolitana de la EPH, para todos los años posibles de edad 25-65. Graficar los efectos marginales de la edad sobre la probabilidad de estar desempleado, junto con los errores estándar de la estimación.*

Stata.

Juan Menduiña

Trabajo Práctico N° 2:
Extensiones de Modelos Logit y Probit.

Ejercicio 1.

Considerar la siguiente armación: "La estimación de un modelo de probabilidad lineal es más robusta que Probit o Logit porque el modelo de probabilidad lineal no asume homocedasticidad ni tiene supuestos acerca de la distribución de los errores."

En esta afirmación, se propone una comparación que no es adecuada.

<u>Ejercicio 2:</u> Probit con una Variable no Observable.

Considerar el modelo Probit:

P (y= 1 | z, q)= Φ ($z_1\delta_1 + \gamma_1 z_2 q$),

donde q es independiente de z y distribuido normal (0, 1); el vector z es observado, pero el escalar q no lo es.

(a) *Encontrar el efecto parcial de z_2 sobre la probabilidad de respuesta, a saber,* $\frac{\partial P(y=1|z,q)}{\partial z_2}$.

$\frac{\partial P(y=1|z,q)}{\partial z_2} = \phi\,(z_1\delta_1 + \gamma_1 z_2 q)\,\gamma_1 q$.

(b) *Mostrar que P (y= 1 | z)= Φ (*$\frac{z_1\delta_1}{(1+\gamma_1^2 z_2^2)^{\frac{1}{2}}}$*).*

Se escribe:

$y^* = z_1\delta_1 + r$,

con $r = \gamma_1 z_2 q + e$, donde $e \sim \mathcal{N}(0, 1)$ y es independiente de (z, q).

Como se asume que q es independiente de z, se tiene:

E (r | z)= E ($\gamma_1 z_2 q$ + e | z)
E (r | z)= E ($\gamma_1 z_2 q$ | z) + E (e | z)
E (r | z)= $\gamma_1 z_2$ E (q | z) + E (e)
E (r | z)= $\gamma_1 z_2$ E (q) + 0
E (r | z)= $\gamma_1 z_2$ * 0 + 0
E (r | z)= 0 + 0
E (r | z)= 0.

Var (r | z)= Var ($\gamma_1 z_2 q$ + e | z)
Var (r | z)= Var ($\gamma_1 z_2 q$ | z) + Var (e | z) + $2\gamma_1 z_2$ Cov (q, e | z)
Var (r | z)= $\gamma_1^2 z_2^2$ Var (q | z) + Var (e) + $2\gamma_1 z_2$ * 0
Var (r | z)= $\gamma_1^2 z_2^2$ Var (q) + 1 + 0
Var (r | z)= $\gamma_1^2 z_2^2$ * 1 + 1 + 0
Var (r | z)= 1 + $\gamma_1^2 z_2^2$.

Entonces, se puede armar la distribución de $\frac{r}{(1+\gamma_1^2 z_2^2)^{\frac{1}{2}}}$ y ver que:

P (y= 1 | z)= Φ (*$\frac{z_1\delta_1}{(1+\gamma_1^2 z_2^2)^{\frac{1}{2}}}$*).

(c) *Definir* $\rho_1 \equiv \gamma_1^2$. *¿Cómo se testearía la hipótesis* H_0*:* $\rho_1 = 0$*?*

Definiendo $\rho_1 \equiv \gamma_1^2$, la hipótesis H_0: $\rho_1 = 0$ se podría testear usando un Score Test o un LM Test.

(d) *Si se tuvieran motivos para creer que* $\rho_1 > 0$*, ¿cómo se estimaría* δ_1 *junto con* ρ_1*?*

Si se tuvieran motivos para creer que $\rho_1 > 0$, δ_1 se estimaría junto con ρ_1 mediante el método de máxima verosimilitud.

Ejercicio 3.

Considerar una gran muestra aleatoria de trabajadores en un momento dado. Sea $sick_i$ *una variable que vale 1 si la persona i se reportó enferma durante los últimos 90 días y vale 0 en caso contrario. Sea* z_i *un vector de características del individuo y del empleador. Sea* $cigs_i$ *el número de cigarrillos que fuma el individuo i por día (en promedio).*

(a) *Explicar el experimento subyacente de interés cuando se quieren examinar los efectos del tabaquismo en los días de trabajo perdidos.*

El experimento subyacente de interés cuando se quieren examinar los efectos del tabaquismo en los días de trabajo perdidos es qué analizar qué efecto tendrá sobre la probabilidad de que una persona se reporte enferma durante los últimos 90 días el cambio exógeno del número de cigarrillos que fuma por día esa persona. En otras palabras, se quiere inferir causalidad, no sólo encontrar una correlación entre el ausentismo en el trabajo y el tabaquismo.

(b) *¿Por qué* $cigs_i$ *podría estar correlacionada con variables no observables que afectan a* $sick_i$*?*

Dado que las personas eligen si fumar y cuánto, ciertamente, no se puede tratar a los datos como si provinieran del experimento que se tiene en mente en el inciso (a). Es decir, no se puede asignar a las personas, aleatoriamente, un consumo de cigarrillos diario.

El consumo de cigarrilos diario puede estar correlacionado con variables no observables que afectan la falta en el trabajo. Por ejemplo, los fumadores pueden ser menos saludables o tener otros atributos que les hagan faltar al trabajo con más frecuencia; o, por el contrario, el consumo de cigarrillos puede estar relacionado con rasgos de la personalidad que hacen que las personas trabajen más. En cualquier caso, el consumo de cigarrillos diarios podría estar correlacionado con elementos no observables de la ecuación.

(c) *Una forma de escribir el modelo de interés es:*

$P(sick = 1 \mid z, cigs, q_1) = \Phi(z_1\delta_1 + \gamma_1 cigs + q_1)$,

donde z_1 *es un subconjunto de z y* q_1 *es una variable no observable que, posiblemente, esté correlacionada con cigs. ¿Qué sucede si se ignora* q_1 *y se estima el Probit de sick sobre* z_1 *y cigs?*

Lo que sucede si se ignora q_1 y se estima el Probit de *sick* sobre z_1 y *cigs* es que los estimadores serán incosistentes.

(d) *¿Puede cigs tener una distribución normal condicional en la población? Explicar.*

Dado que, en la población, hay muchas personas que no fuman, la distribución (condicional o incondicional) de consumo de cigarrillos diarios se "apila" en cero. Además, la variable *cigs* toma valores enteros positivos, por lo que no puede tener una distribución normal condicional en la población.

(e) *Explicar cómo probar si cigs es exógeno. ¿Esta prueba se basa en cigs que tienen una distribución normal condicional?*

Para probar si *cigs* es exógeno, se puede utilizar el procedimiento de dos etapas de Rivers y Vuong (1988).

(f) *Suponer que algunos de los trabajadores viven en estados que, recientemente, implementaron leyes de no fumar en el lugar de trabajo. ¿La presencia de las nuevas leyes sugiere un buen candidato IV para cigs?*

Suponiendo que las personas no se mudarán, inmediatamente, de su estado de residencia cuando el estado implemente leyes de no de fumar en el lugar de trabajo y que ese estado de residencia es, aproximadamente, independiente de la salud general de la población, un indicador *dummy* que diga si la persona trabaja en un estado con una nueva ley puede funcionar como una variable exógena. Estas situaciones, a menudo, se denominan "experimentos naturales". Además, es probable que la variable *cigs* esté correlacionada con el indicador de la ley estatal porque las personas no podrán fumar tanto como lo harían de no existir la ley. Por tanto, la presencia de las nuevas leyes sugiere un buen candidato IV para *cigs*.

Ejercicio 4.

Utilizar el conjunto de datos "BWGHT.dta" para este problema.

(a) *Definir una variable binaria, smokes, si la mujer fuma durante el embarazo. Estimar un modelo Probit que relacione smokes con motheduc, white y log(faminc). En white= 0 y faminc evaluado en el promedio de la muestra, ¿cuál es la diferencia estimada en la probabilidad de fumar para una mujer con 16 años de educación y una con 12 años de educación?*

Probit:

```
Probit regression                                       Number of obs =  1,387
                                                        LR chi2(3)    =  92.67
                                                        Prob > chi2   = 0.0000
Log likelihood = -546.76991                             Pseudo R2     = 0.0781

------------------------------------------------------------------------------
      smokes | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    motheduc |  -.1450599   .0207899    -6.98   0.000    -.1858074   -.1043124
       white |   .1896765   .1098805     1.73   0.084    -.0256853    .4050383
     lfaminc |  -.1669109   .0498894    -3.35   0.001    -.2646923   -.0691296
       _cons |   1.126276   .2504611     4.50   0.000     .6353817    1.617171
------------------------------------------------------------------------------
```

La diferencia estimada en la probabilidad de fumar para una mujer con 16 años de educación y una con 12 años de educación es -0,086.

(b) *¿faminc es exógena en la ecuación de smokes? ¿Qué pasa con motheduc?*

faminc puede llegar a ser endógena en la ecuación de *smokes*.

(c) *Suponer que motheduc y white son exógenos en el Probit del inciso (a). Suponer, también, que fatheduc es exógeno a esta ecuación. Estimar la forma reducida de log(faminc) para ver si fatheduc está parcialmente correlacionada con log(faminc).*

Probit:

```
      Source |       SS           df       MS      Number of obs   =     1,191
-------------+----------------------------------   F(3, 1187)      =    119.23
       Model |  140.936735         3  46.9789115   Prob > F        =    0.0000
    Residual |  467.690904     1,187  .394010871   R-squared       =    0.2316
-------------+----------------------------------   Adj R-squared   =    0.2296
       Total |  608.627639     1,190  .511451797   Root MSE        =     .6277

------------------------------------------------------------------------------
     lfaminc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    motheduc |   .0709044   .0098338     7.21   0.000     .0516109     .090198
       white |   .3452115    .050418     6.85   0.000     .2462931    .4441298
    fatheduc |   .0616625    .008708     7.08   0.000     .0445777    .0787473
       _cons |   1.241413   .1103648    11.25   0.000     1.024881    1.457945
------------------------------------------------------------------------------
```

(d) *Contrastar la hipótesis nula de que log(faminc) es exógena en el Probit del inciso (a).*

```
Probit regression                                 Number of obs =  1,191
                                                  LR chi2(4)    =  79.43
                                                  Prob > chi2   = 0.0000
Log likelihood = -432.06242                       Pseudo R2     = 0.0842

------------------------------------------------------------------------------
      smokes | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    motheduc |  -.0826247   .0465204    -1.78   0.076     -.173803    .0085536
       white |   .4611075   .1965245     2.35   0.019     .0759265    .8462886
     lfaminc |  -.7622559   .3652949    -2.09   0.037    -1.478221    -.046291
       v2hat |   .6107298   .3708071     1.65   0.100    -.1160387    1.337498
       _cons |    1.98796   .5996374     3.32   0.001     .8126927    3.163228
------------------------------------------------------------------------------
```

```
(1)  [smokes]v2hat = 0

           chi2(  1) =    2.71
         Prob > chi2 =    0.0996
```

Por lo tanto, con un nivel de significancia del 10%, estos datos aportan evidencia suficiente para indicar que *log(faminc)* es endógena.

Ejercicio 5.

Una preocupación común cuando se utilizan precios autoinformados en la estimación de la prevalencia del tabaquismo con una base de datos de corte transversal (por ejemplo, Global Adult Tobacco Survey o GATS) es la potencial endogeneidad de esta variable. Para abordar este problema potencial, se construyen dos variables de precios diferentes. La primera variable de precio asigna a los fumadores el precio autoinformado pagado por la última compra y utiliza una imputación de regresión aleatoria (random regression imputation, a veces denominada imputación de regresión estocástica) para asignar un precio a los no fumadores de la muestra. La segunda variable de precio asigna a fumadores y no fumadores el promedio del precio autoinformado por unidad primaria de muestreo (UPM, o PSU por Primary Sampling Unit). Siguiendo las recomendaciones en "Economics of Tobacco Toolkit: Economic Analysis of Demand Using Data from the Global Adult Tobacco Survey (GATS)" (John et al., 2019), se puede verificar la endogeneidad del precio autoinformado utilizando el test de Rivers-Vuong (1988).

(a) *¿Por qué podrían ser endógenos los precios autoinformados?*

Los precios autoinformados podrían ser endógenos porque pueden estar correlacionados con variables omitidas en el modelo, que, a su vez, correlacionen con la variable dependiente.

(b) *Realizar el test de Rivers-Vuong para los datos provistos en "pricedata.dta" utilizando las variables X en la primera etapa y Z en la segunda etapa.*

```
Adjusted Wald test

 (1)  [SmokeCigs]resid1 = 0

       F(  1,  5976) =   18.77
            Prob > F =    0.0000
```

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los precios autoinformados son endógenos.

(c) *En función de los resultados, estimar la elasticidad de la prevalencia del tabaquismo con respecto a los precios.*

Stata.

Ejercicio 6.

Se busca simular el siguiente modelo:

Pr (y= 1)= F $\{\frac{\beta_0+\beta_1 x}{e^{\gamma_1 x_{het}}}\}$.

Generar un dataset vacío con 1000 obsevaciones. Generar las siguientes variables:

$x \sim U(-1, 1)$,
$x_{het} \sim U(0, 1)$,
$\sigma \sim e^{1,5 x_{het}}$,
$p \sim \mathcal{N}\ (\frac{\beta_0+\beta_1 x}{\sigma})$,

con β_0 *= 0,3 y* β_1 *= 2 y definir la variable dependiente y como una variable binaria que vale 1 si p es mayor o igual a una variable aleatoria uniforme en el intervalo (0, 1) y 0 en caso contrario. Estimar el modelo Probit heterocedástico y comparar con las estimaciones del Probit usual.*

Probit heterocedástico:

```
Heteroskedastic probit model                    Number of obs     =      1,000
                                                Zero outcomes     =        468
                                                Nonzero outcomes  =        532

                                                Wald chi2(1)      =      78.21
Log likelihood = -563.0256                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
y            |
           x |   2.484689   .2809507     8.84   0.000     1.934036    3.035342
       _cons |   .2876884   .0939283     3.06   0.002     .1035924    .4717845
-------------+----------------------------------------------------------------
lnsigma      |
        xhet |   1.734142   .2630328     6.59   0.000     1.218608    2.249677
------------------------------------------------------------------------------
LR test of lnsigma=0: chi2(1) = 51.24                   Prob > chi2 = 0.0000
```

Probit:

```
Probit regression                                     Number of obs =  1,000
                                                      LR chi2(1)    = 204.90
                                                      Prob > chi2   = 0.0000
Log likelihood = -588.64728                           Pseudo R2     = 0.1482

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           x |   1.054521   .0772801    13.65   0.000      .903055    1.205987
       _cons |   .1085126   .0424302     2.56   0.011      .025351    .1916742
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) Probit Het~o	(2) Probit
y		
x	2.485***	1.055***
	(0.281)	(0.0773)
_cons	0.288***	0.109**
	(0.0939)	(0.0424)
lnsigma		
xhet	1.734***	
	(0.263)	
N	1000	1000
pseudo R-sq		0.148

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Trabajo Práctico N° 3:
Modelos para Variables Categóricas No Ordenadas.

Ejercicio 1: Alternativas de Pesca.

La variable dependiente y toma el valor 1, 2, 3 o 4, dependiendo de cuál de los cuatro modos alternativos de pesca, respectivamente, playa, muelle, barco privado y barco chárter, se elija. En la base de datos, estos son beach, pier, private o charter. Los datos provienen de Herriges, J. A. y Kling, C. L. (1999): "Nonlinear Income Effects in Random Utility Models", Review of Economics and Statistics, 81, 62-72.

(a) *Abrir la base y describir las categorías.*

```
-------------------------------------------------
Fishing   |
mode      |    N(income)  mean(income)    sd(income)
----------+--------------------------------------
    beach |          134      4.051617       2.50542
     pier |          178      3.387172      2.340324
  private |          418      4.654107      2.777898
  charter |          452      3.880899      2.050029
-------------------------------------------------
```

```
-----------------------------------------------------------------------
Fishing   |
mode      |   mean(pbeach)     mean(ppier)  mean(pprivate)  mean(pcharter)
----------+------------------------------------------------------------
    beach |       35.69949        35.69949        97.80914        125.0032
     pier |       30.57133        30.57133        82.42908        109.7634
  private |       137.5271        137.5271        41.60681        70.58408
  charter |       120.6483        120.6483        44.56376        75.09694
-----------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
Fishing   |
mode      |   mean(qbeach)     mean(qpier)  mean(qprivate)  mean(qcharter)
----------+------------------------------------------------------------
    beach |       .2791948        .2190015        .1593985        .5176089
     pier |       .2614444        .2025348        .1501489        .4980798
  private |       .2082868        .1297646        .1775412        .6539167
  charter |       .2519077        .1595341        .1771628        .6914998
-----------------------------------------------------------------------
```

(b) *Estimar un modelo logit multinomial.*

Logit multinomial (betas):

```
Multinomial logistic regression                         Number of obs =  1,182
                                                        LR chi2(3)    =  41.14
                                                        Prob > chi2   = 0.0000
Log likelihood = -1477.1506                             Pseudo R2     = 0.0137

------------------------------------------------------------------------------
        mode | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
beach        |  (base outcome)
-------------+----------------------------------------------------------------
pier         |
      income |  -.1434029   .0532884    -2.69   0.007    -.2478463   -.0389595
       _cons |   .8141503    .228632     3.56   0.000     .3660399    1.262261
-------------+----------------------------------------------------------------
private      |
      income |   .0919064   .0406637     2.26   0.024     .0122069    .1716058
       _cons |   .7389208   .1967309     3.76   0.000     .3533352    1.124506
-------------+----------------------------------------------------------------
charter      |
      income |  -.0316399   .0418463    -0.76   0.450    -.1136571    .0503774
       _cons |   1.341291   .1945167     6.90   0.000     .9600457    1.722537
------------------------------------------------------------------------------
```

Logit multinomial (relative-risk ratios):

```
Multinomial logistic regression                         Number of obs =  1,182
                                                        LR chi2(3)    =  41.14
                                                        Prob > chi2   = 0.0000
Log likelihood = -1477.1506                             Pseudo R2     = 0.0137

------------------------------------------------------------------------------
        mode |        RRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
beach        |  (base outcome)
-------------+----------------------------------------------------------------
pier         |
      income |   .8664049   .0461693    -2.69   0.007     .7804799    .9617896
       _cons |   2.257257    .516081     3.56   0.000     1.442013      3.5334
-------------+----------------------------------------------------------------
private      |
      income |   1.096262   .0445781     2.26   0.024     1.012282     1.18721
       _cons |   2.093675   .4118906     3.76   0.000     1.423808    3.078697
-------------+----------------------------------------------------------------
charter      |
      income |   .9688554    .040543    -0.76   0.450     .8925639    1.051668
       _cons |   3.823979   .7438278     6.90   0.000     2.611816    5.598715
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.
```

(c) *Estimar un modelo logit condicional.*

Logit condicional:

```
Alternative-specific conditional logit          Number of obs      =      4,728
Case ID variable: id                            Number of cases    =       1182

Alternatives variable: fishmode                 Alts per case: min =          4
                                                               avg =        4.0
                                                               max =          4

                                                   Wald chi2(5)    =     252.98
Log likelihood = -1215.1376                        Prob > chi2     =     0.0000

------------------------------------------------------------------------------
           d | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
fishmode     |
           p |  -.0251166   .0017317   -14.50   0.000    -.0285106   -.0217225
           q |    .357782   .1097733     3.26   0.001     .1426302    .5729337
-------------+----------------------------------------------------------------
beach        |  (base alternative)
-------------+----------------------------------------------------------------
charter      |
      income |  -.0332917   .0503409    -0.66   0.508     -.131958    .0653745
       _cons |   1.694366   .2240506     7.56   0.000     1.255235    2.133497
-------------+----------------------------------------------------------------
pier         |
      income |  -.1275771   .0506395    -2.52   0.012    -.2268288   -.0283255
       _cons |   .7779593   .2204939     3.53   0.000     .3457992    1.210119
-------------+----------------------------------------------------------------
private      |
      income |   .0894398   .0500671     1.79   0.074    -.0086898    .1875694
       _cons |   .5272788   .2227927     2.37   0.018     .0906132    .9639444
------------------------------------------------------------------------------
```

<u>Ejercicio 2:</u> Predicción de Calificaciones de Clientes.

Net Promoter Score®, o NPS®, mide la experiencia del cliente y predice el crecimiento del negocio. Es utilizada por empresas que brindan servicios al consumidor final (bancos, telefónicas, etc). EL NPS se calcula usando la respuesta a una pregunta usando una escala de 0 a 10: ¿Qué tan probable es que recomiende a un amigo o colega? Los encuestados se agrupan de la siguiente manera:

- *Los promotores (puntuación 9-10) son entusiastas leales que seguirán comprando y recomendarán a otros, lo que impulsará el crecimiento.*
- *Los neutrales (puntuación 7-8) son clientes satisfechos pero poco entusiastas que son vulnerables a las ofertas de la competencia.*
- *Los detractores (puntuación 1-6) son clientes insatisfechos que pueden dañar su marca e impedir el crecimiento a través del boca a boca negativo.*

Al restar el porcentaje de detractores del porcentaje de promotores, se obtiene el puntaje neto del promotor, que puede oscilar entre un mínimo de -100 (si todos los clientes son detractores) y un máximo de 100 (si todos los clientes son promotores). Estas encuestas se utilizan para generar estrategias de originacion (nuevos clientes) y de reducción de churn (fuga de clientes). La base con la que se va a hacer la primera parte de la práctica consiste en la encuesta de NPS que se le hace a los clientes de un Banco luego de efectuar una transacción en caja. En base a esto, utilizando la base "NPS.dta", responder las siguientes preguntas.

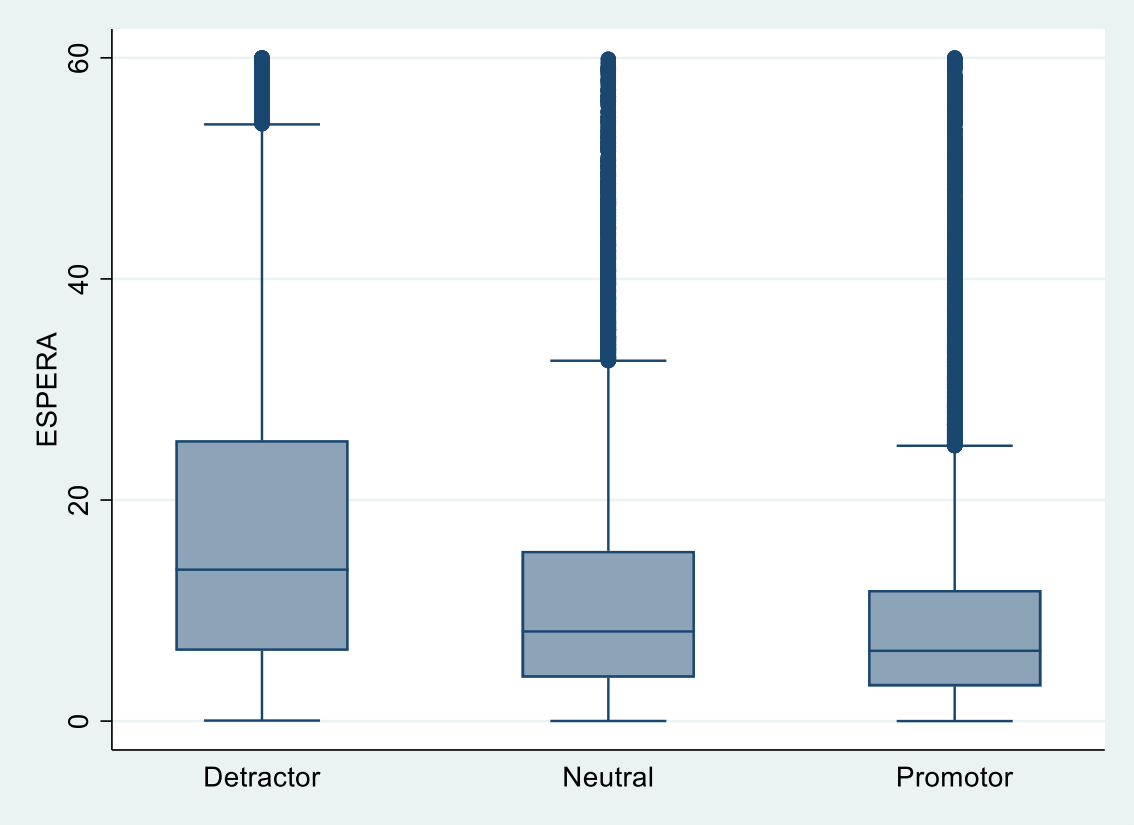
(a) *Abrir y describir la base.*

Variable	Obs	Mean	Std. dev.	Min	Max
nps	42,019	8.369975	2.263878	1	10
marital_st~e	0				
gender_code	0				
edad	42,020	52.16497	12.56996	19	101
branch_desc	0				
segmento	0				
operaciones	42,020	1.728439	1.476585	1	31
mes	42,020	6.736292	3.241668	1	12
nps_anterior	0				
hora	42,020	11.7812	1.743031	7	18
dia	42,020	14.91792	8.634796	1	31
dia_semana	0				
espera	42,020	10.89938	10.70589	0	60
cliente	42,020	21372.36	12335.51	1	42760

(b) *Generar una variable que clasifique a los clientes en función de si son promotores, detractores o neutrales.*

clasificacion	Freq.	Percent	Cum.
Detractor	6,265	14.91	14.91
Neutral	9,579	22.80	37.71
Promotor	26,175	62.29	100.00
Total	42,019	100.00	

(c) *Analizar cómo cambia la variable de espera en función de la clasificación de los clientes.*



(d) *Tomar una muestra del 10% de los datos. Estimar un logit multinomial para predecir cómo cambian las clasificaciones en función de la espera, condicionando en explicativas que se considere relevantes.*

Logit (betas):

```
Multinomial logistic regression                         Number of obs =  4,202
                                                        LR chi2(14)   = 418.26
                                                        Prob > chi2   = 0.0000
Log likelihood = -3647.5859                             Pseudo R2     = 0.0542

------------------------------------------------------------------------------
clasificacion | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
Detractor     |  (base outcome)
--------------+---------------------------------------------------------------
Neutral       |
_Igender_co_2 |  -.0022797   .1117659    -0.02   0.984    -.2213368    .2167774
        edad  |   .0106823   .0042832     2.49   0.013     .0022873    .0190772
 _Isegmento_2 |   12.80348   730.9035     0.02   0.986    -1419.741    1445.348
 _Isegmento_3 |   .0192837   .1868698     0.10   0.918    -.3469745    .3855418
 _Isegmento_4 |  -.7049277   .1983862    -3.55   0.000    -1.093758   -.3160979
 _Isegmento_5 |  -.5423917   .2023154    -2.68   0.007    -.9389226   -.1458607
       espera |  -.0234117   .0044156    -5.30   0.000     -.032066   -.0147573
        _cons |   .4567806   .2819115     1.62   0.105    -.0957557    1.009317
--------------+---------------------------------------------------------------
Promotor      |
_Igender_co_2 |  -.0740685   .0991182    -0.75   0.455    -.2683366    .1201995
        edad  |   .0222569   .0038062     5.85   0.000     .0147969    .0297169
 _Isegmento_2 |   13.38895    730.903     0.02   0.985    -1419.155    1445.933
 _Isegmento_3 |    .254493   .1689136     1.51   0.132    -.0765715    .5855575
 _Isegmento_4 |  -.6899248   .1774649    -3.89   0.000     -1.03775      -.3421
 _Isegmento_5 |  -.7035198   .1827513    -3.85   0.000    -1.061706   -.3453338
       espera |  -.0479308   .0040826   -11.74   0.000    -.0559326    -.039929
        _cons |   1.070479   .2520943     4.25   0.000     .5763835    1.564575
------------------------------------------------------------------------------
```

Logit multinomial (relative-risk ratios):

```
Multinomial logistic regression                         Number of obs =  4,202
                                                        LR chi2(14)   = 418.26
                                                        Prob > chi2   = 0.0000
Log likelihood = -3647.5859                             Pseudo R2     = 0.0542

------------------------------------------------------------------------------
clasificacion |        RRR   Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
Detractor     |  (base outcome)
--------------+---------------------------------------------------------------
Neutral       |
_Igender_co_2 |   .9977229   .1115114    -0.02   0.984     .8014467    1.242068
        edad  |    1.01074   .0043292     2.49   0.013      1.00229     1.01926
 _Isegmento_2 |   363481.5   2.66e+08     0.02   0.986            0           .
 _Isegmento_3 |   1.019471   .1905084     0.10   0.918     .7068233    1.470411
 _Isegmento_4 |   .4941443   .0980314    -3.55   0.000     .3349555    .7289881
 _Isegmento_5 |   .5813562   .1176173    -2.68   0.007     .3910489    .8642781
       espera |   .9768603   .0043134    -5.30   0.000     .9684427     .985351
        _cons |   1.578982   .4451333     1.62   0.105     .9086859    2.743726
--------------+---------------------------------------------------------------
Promotor      |
_Igender_co_2 |   .9286081   .0920419    -0.75   0.455     .7646504    1.127722
        edad  |   1.022506   .0038919     5.85   0.000     1.014907    1.030163
 _Isegmento_2 |   652751.9   4.77e+08     0.02   0.985            0           .
 _Isegmento_3 |   1.289808    .217866     1.51   0.132     .9262867    1.795992
 _Isegmento_4 |   .5016138   .0890188    -3.89   0.000      .354251    .7102772
 _Isegmento_5 |   .4948405   .0904327    -3.85   0.000     .3458654     .707984
       espera |   .9531997   .0038915   -11.74   0.000     .9456029    .9608576
        _cons |   2.916777   .7353029     4.25   0.000     1.779591    4.780643
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.
```

(e) *Calcular los efectos marginales.*

Efectos marginales en Logit multinomial (detractor):

```
Marginal effects after mlogit
      y  = Pr(clasificacion==Detractor) (predict, pr outcome(1))
         =  .13172136
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|   .0062526      .01127    0.56   0.579  -.015827  .028332   .678486
    edad |  -.0021919       .0012   -1.83   0.067  -.004541  .000157   52.2109
_Isegm~2*|  -.1331684      .00569  -23.41   0.000  -.144317  -.12202   .000952
_Isegm~3*|  -.0220274      .02219   -0.99   0.321  -.065524  .021469   .567587
_Isegm~4*|   .0931274      .05089    1.83   0.067  -.006608  .192863   .183246
_Isegm~5*|   .0890482      .04974    1.79   0.073  -.008432  .186529   .148263
  espera |   .0047328      .00246    1.92   0.055  -.000097  .009562   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Logit multinomial (neutral):

```
Marginal effects after mlogit
      y  = Pr(clasificacion==Neutral) (predict, pr outcome(2))
         =  .23194672
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|     .01049      .01435    0.73   0.465  -.017644  .038624   .678486
    edad |   -.001382      .00072   -1.91   0.056  -.002801  .000037   52.2109
_Isegm~2*|  -.0628502       .1635   -0.38   0.701  -.383304  .257604   .000952
_Isegm~3*|   -.034214      .02341   -1.46   0.144   -.08009  .011662   .567587
_Isegm~4*|    -.02724      .02669   -1.02   0.307  -.079548  .025068   .183246
_Isegm~5*|   .0021924      .02992    0.07   0.942  -.056453  .060838   .148263
  espera |   .0029036      .00123    2.37   0.018   .000501  .005306   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Logit multinomial (promotor):

```
Marginal effects after mlogit
      y  = Pr(clasificacion==Promotor) (predict, pr outcome(3))
         =  .63633192
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|  -.0167426      .01648   -1.02   0.310  -.049037  .015551   .678486
    edad |   .0035739       .0009    3.99   0.000   .001817  .005331   52.2109
_Isegm~2*|   .1960187      .16356    1.20   0.231  -.124551  .516589   .000952
_Isegm~3*|   .0562415      .02657    2.12   0.034   .004172  .108311   .567587
_Isegm~4*|  -.0658873      .04487   -1.47   0.142  -.153828  .022054   .183246
_Isegm~5*|  -.0912406      .04211   -2.17   0.030  -.173769 -.008712   .148263
  espera |  -.0076364       .0016   -4.77   0.000  -.010776 -.004497   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

(f) *Repetir el análisis con un Probit multinomial y comparar.*

Probit multinomial:

```
Multinomial probit regression                           Number of obs =  4,202
                                                        Wald chi2(14) = 416.83
Log likelihood = -3635.6144                             Prob > chi2   = 0.0000
```

clasificacion	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
Detractor	(base outcome)					
Neutral						
_Igender_co_2	-.0536078	.0798336	-0.67	0.502	-.2100787	.1028631
edad	.0092218	.003062	3.01	0.003	.0032205	.0152231
_Isegmento_2	-.352632	.5757431	-0.61	0.540	-1.481068	.7758037
_Isegmento_3	.0867023	.1308623	0.66	0.508	-.1697831	.3431876
_Isegmento_4	-.7015738	.1429056	-4.91	0.000	-.9816635	-.421484
_Isegmento_5	-.3109711	.1472973	-2.11	0.035	-.5996685	-.0222737
espera	-.0138713	.0033331	-4.16	0.000	-.020404	-.0073386
_cons	.2119086	.2034099	1.04	0.298	-.1867675	.6105848
Promotor						
_Igender_co_2	-.097611	.0738029	-1.32	0.186	-.242262	.0470399
edad	.012833	.002822	4.55	0.000	.007302	.018364
_Isegmento_2	-1.411541	.6475008	-2.18	0.029	-2.680619	-.1424626
_Isegmento_3	.2629534	.1220016	2.16	0.031	.0238348	.5020721
_Isegmento_4	-.6144694	.1313595	-4.68	0.000	-.8719294	-.3570095
_Isegmento_5	-.4984651	.1378136	-3.62	0.000	-.7685749	-.2283554
espera	-.0350071	.0031476	-11.12	0.000	-.0411763	-.0288379
_cons	1.035228	.1878494	5.51	0.000	.6670502	1.403406

Efectos marginales en Probit multinomial (detractor):

```
Marginal effects after mprobit
      y  = Pr(clasificacion==Detractor) (predict, pr outcome(1))
         =  .13404784
```

variable	dy/dx	Std. err.	z	P>|z|	[	95% C.I.	]	X
_Igend~2*	.0136968	.01125	1.22	0.223	-.008346		.03574	.677297
edad	-.0019418	.00044	-4.41	0.000	-.002806		-.001078	52.1844
_Isegm~2*	.2216672	.14863	1.49	0.136	-.06965		.512984	.002618
_Isegm~3*	-.0345801	.01966	-1.76	0.079	-.07312		.00396	.578058
_Isegm~4*	.1251906	.02726	4.59	0.000	.071753		.178628	.183484
_Isegm~5*	.0823211	.02707	3.04	0.002	.02926		.135382	.140171
espera	.0046788	.00048	9.74	0.000	.003737		.005621	11.1349

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Probit multinomial (neutral):

```
Marginal effects after mprobit
      y  = Pr(clasificacion==Neutral) (predict, pr outcome(2))
         =  .23112599
```

variable	dy/dx	Std. err.	z	P>|z|	[	95% C.I.	]	X
_Igend~2*	.0048601	.01424	0.34	0.733	-.023054		.032774	.677297
edad	-.00013	.00055	-0.24	0.812	-.0012		.00094	52.1844
_Isegm~2*	.1369606	.15124	0.91	0.365	-.159457		.433378	.002618
_Isegm~3*	-.0261368	.0229	-1.14	0.254	-.071021		.018747	.578058
_Isegm~4*	-.0571084	.02392	-2.39	0.017	-.103993		-.010224	.183484
_Isegm~5*	.0120419	.02754	0.44	0.662	-.041943		.066027	.140171
espera	.0029577	.00066	4.52	0.000	.001674		.004242	11.1349

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Efectos marginales en Probit multinomial (promotor):

```
Marginal effects after mprobit
      y  = Pr(clasificacion==Promotor) (predict, pr outcome(3))
         =  .63482617
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|  -.0185569      .01634   -1.14   0.256   -.05058  .013466   .677297
    edad |   .0020718      .00063    3.31   0.001   .000845  .003298   52.1844
_Isegm~2*|  -.3586278      .15141   -2.37   0.018  -.655388 -.061868   .002618
_Isegm~3*|   .0607169       .0265    2.29   0.022   .008771  .112663   .578058
_Isegm~4*|  -.0680822       .0313   -2.18   0.030   -.12943 -.006734   .183484
_Isegm~5*|  -.0943629      .03271   -2.88   0.004  -.158476  -.03025   .140171
  espera |  -.0076366      .00076  -10.01   0.000  -.009132 -.006141   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

(g) *Realizar un test de la significatividad de las variables.*

Stata.

Ejercicio 3.

Utilizando la EPH del cuarto trimestre de 2016, estimar un modelo multinomial que permita predecir la condición de actividad de una persona, entre inactivo, ocupado o desocupado.

Stata.

Trabajo Práctico N° 4:
Modelos para Variables Categóricas Ordenadas.

Ejercicio 1: Predicción de Calificaciones de Clientes.

Considerar el ejercicio del Problem Set anterior con el mismo título que éste. Repetir el análisis utilizando un modelo ordenado.

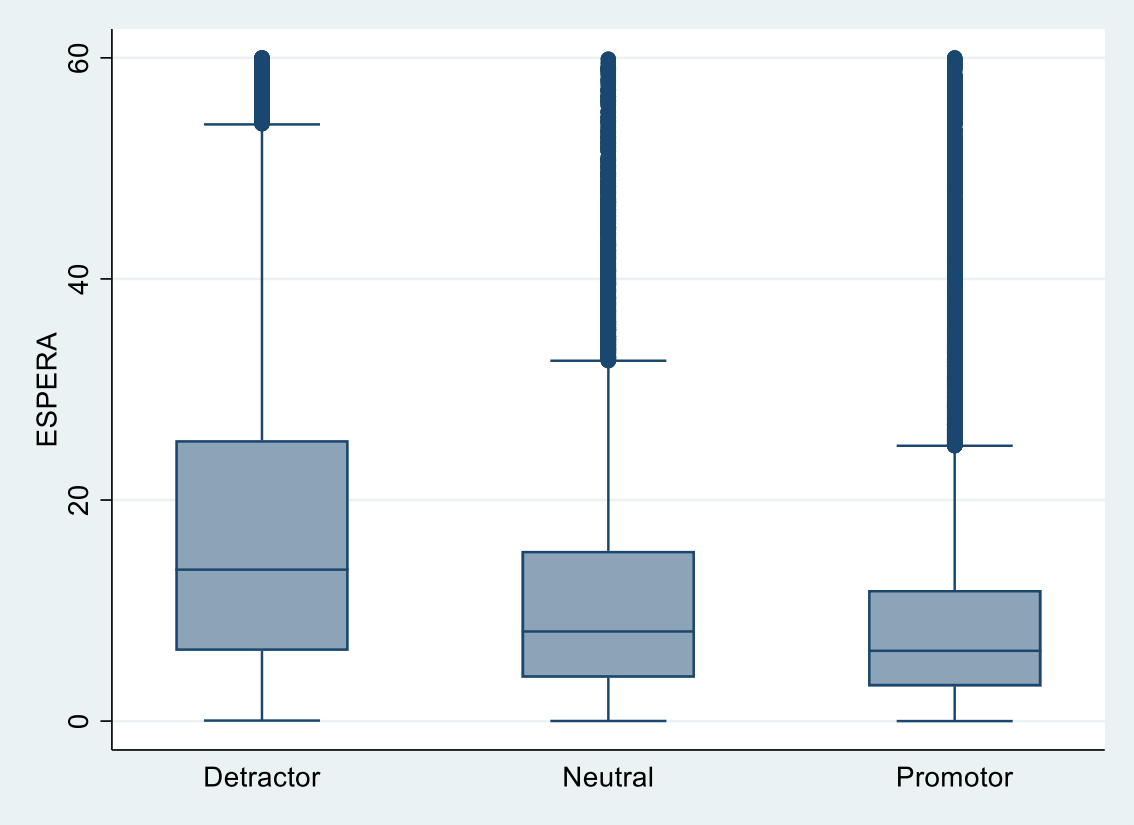
(a) *Abrir y describir la base.*

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         nps |     42,019    8.369975    2.263878          1         10
marital_st~e |          0
 gender_code |          0
        edad |     42,020    52.16497    12.56996         19        101
 branch_desc |          0
-------------+---------------------------------------------------------
    segmento |          0
 operaciones |     42,020    1.728439    1.476585          1         31
         mes |     42,020    6.736292    3.241668          1         12
nps_anterior |          0
        hora |     42,020     11.7812    1.743031          7         18
-------------+---------------------------------------------------------
         dia |     42,020    14.91792    8.634796          1         31
  dia_semana |          0
      espera |     42,020    10.89938    10.70589          0         60
     cliente |     42,020    21372.36    12335.51          1      42760
```

(b) *Generar una variable que clasifique a los clientes en función de si son promotores, detractores o neutrales.*

clasificacion	Freq.	Percent	Cum.
Detractor	6,265	14.91	14.91
Neutral	9,579	22.80	37.71
Promotor	26,175	62.29	100.00
Total	42,019	100.00	

(c) *Analizar cómo cambia la variable de espera en función de la clasificación de los clientes.*



(d) *Tomar una muestra del 10% de los datos. Estimar un logit multinomial ordenado para predecir cómo cambian las clasificaciones en función de la espera, condicionando en explicativas que se considere relevantes.*

Logit multinomial ordenado (betas):

```
Ordered logistic regression                             Number of obs =  4,202
                                                        LR chi2(7)    = 394.03
                                                        Prob > chi2   = 0.0000
Log likelihood = -3659.6981                             Pseudo R2     = 0.0511

------------------------------------------------------------------------------
clasificacion | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
_Igender_co_2 |  -.0667762   .0687925    -0.97   0.332     -.201607    .0680546
        edad  |   .0163998   .0026606     6.16   0.000     .0111852    .0216144
 _Isegmento_2 |   .7313334   .8363563     0.87   0.382    -.9078948    2.370562
 _Isegmento_3 |   .2147268   .1144579     1.88   0.061    -.0096065    .4390601
 _Isegmento_4 |   -.478007   .1271562    -3.76   0.000    -.7272286   -.2287855
 _Isegmento_5 |  -.3697909   .1299945    -2.84   0.004    -.6245754   -.1150063
      espera  |  -.0359647   .0030773   -11.69   0.000     -.041996   -.0299333
--------------+---------------------------------------------------------------
       /cut1  |  -1.429497    .178061                      -1.77849   -1.080504
       /cut2  |  -.1286401   .1758049                     -.4732113    .2159311
------------------------------------------------------------------------------
```

Logit multinomial ordenado (odds ratios):

```
Ordered logistic regression                             Number of obs =  4,202
                                                        LR chi2(7)    = 394.03
                                                        Prob > chi2   = 0.0000
Log likelihood = -3659.6981                             Pseudo R2     = 0.0511

------------------------------------------------------------------------------
clasificacion | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
_Igender_co_2 |   .9354045   .0643488    -0.97   0.332     .8174161    1.070424
        edad  |   1.016535   .0027045     6.16   0.000     1.011248     1.02185
 _Isegmento_2 |   2.077849   1.737822     0.87   0.382     .4033725     10.7034
 _Isegmento_3 |   1.239523   .1418732     1.88   0.061     .9904395    1.551249
 _Isegmento_4 |   .6200178   .0788391    -3.76   0.000     .4832464    .7954992
 _Isegmento_5 |   .6908788   .0898104    -2.84   0.004     .5354887    .8913605
      espera  |   .9646744   .0029686   -11.69   0.000     .9588736    .9705103
--------------+---------------------------------------------------------------
       /cut1  |  -1.429497    .178061                      -1.77849   -1.080504
       /cut2  |  -.1286401   .1758049                     -.4732113    .2159311
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation to odds ratios.
```

(e) *Calcular los efectos marginales.*

Efectos marginales en Logit multinomial ordenado (clasificación 1):

```
Marginal effects after ologit
      y  = Pr(clasificacion==1) (predict, pr outcome(1))
         =  .13947074
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|   .0079445      .00812    0.98   0.328  -.007962  .023851   .682532
    edad |  -.0019683      .00032   -6.13   0.000  -.002597 -.001339   51.9412
_Isegm~2*|  -.0671776      .05629   -1.19   0.233  -.177512  .043157   .001666
_Isegm~3*|  -.0261068       .0141   -1.85   0.064  -.053748  .001534   .580438
_Isegm~4*|    .063994      .01889    3.39   0.001    .02698  .101008   .179914
_Isegm~5*|   .0486811      .01869    2.60   0.009   .012049  .085313   .149929
  espera |   .0043164      .00038   11.41   0.000   .003575  .005058   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Logit multinomial ordenado (clasificación 2):

```
Marginal effects after ologit
      y  = Pr(clasificacion==2) (predict, pr outcome(2))
         =  .23365369
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|   .0076247      .00788     0.97   0.333  -.007815  .023064   .682532
    edad |  -.0018677      .00031    -6.00   0.000  -.002478 -.001258   51.9412
_Isegm~2*|  -.0833501      .08871    -0.94   0.347   -.25722   .09052   .001666
_Isegm~3*|  -.0242958      .01287    -1.89   0.059  -.049518  .000926   .580438
_Isegm~4*|   .0511031      .01252     4.08   0.000   .026558  .075648   .179914
_Isegm~5*|   .0401242      .01326     3.03   0.002   .014131  .066118   .149929
  espera |   .0040958      .00038    10.66   0.000   .003342  .004849   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Logit multinomial ordenado (clasificación 3):

```
Marginal effects after ologit
      y  = Pr(clasificacion==3) (predict, pr outcome(3))
         =  .62687557
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.      z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|  -.0155692      .01599    -0.97   0.330  -.046903  .015764   .682532
    edad |    .003836      .00062     6.17   0.000   .002618  .005054   51.9412
_Isegm~2*|   .1505277      .14494     1.04   0.299  -.133553  .434608   .001666
_Isegm~3*|   .0504026      .02693     1.87   0.061  -.002378  .103183   .580438
_Isegm~4*|  -.1150971       .0312    -3.69   0.000  -.176246 -.053948   .179914
_Isegm~5*|  -.0888053      .03183    -2.79   0.005  -.151197 -.026414   .149929
  espera |  -.0084122      .00072   -11.67   0.000  -.009825 -.006999   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

(f) *Repetir el análisis con un Probit multinomial ordenado y comparar.*

Probit multinomial ordenado:

```
Ordered probit regression                         Number of obs =  4,202
                                                  LR chi2(7)    = 450.86
                                                  Prob > chi2   = 0.0000
Log likelihood = -3631.2859                       Pseudo R2     = 0.0585

------------------------------------------------------------------------------
clasificacion | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
_Igender_co_2 |   .0079382    .040891     0.19   0.846    -.0722067    .0880831
        edad |   .0088488   .0015378     5.75   0.000     .0058348    .0118628
 _Isegmento_2 |   .3043314   .5146773     0.59   0.554    -.7044176     1.31308
 _Isegmento_3 |   .1132939   .0664574     1.70   0.088    -.0169602    .2435479
 _Isegmento_4 |  -.3667121   .0740819    -4.95   0.000    -.5119099   -.2215144
 _Isegmento_5 |  -.3928859   .0764502    -5.14   0.000    -.5427255   -.2430463
      espera |  -.0202896   .0018409   -11.02   0.000    -.0238977   -.0166814
-------------+----------------------------------------------------------------
       /cut1 |  -.9159207   .1025571                     -1.116929   -.7149126
       /cut2 |  -.1473535   .1017522                     -.3467842    .0520771
------------------------------------------------------------------------------
```

Efectos marginales en Progit multinomial ordenado (clasificación 1):

```
Marginal effects after oprobit
      y  = Pr(clasificacion==1) (predict, pr outcome(1))
         =  .13680723
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|   -.001742      .00899   -0.19   0.846  -.019358  .015873   .680866
    edad |  -.0019388      .00034   -5.73   0.000  -.002602 -.001276   52.1171
_Isegm~2*|   -.055947      .07726   -0.72   0.469  -.207383  .095489   .001428
_Isegm~3*|  -.0250324      .01481   -1.69   0.091  -.054056  .003991   .567111
_Isegm~4*|   .0905003      .02035    4.45   0.000   .050614  .130386   .186578
_Isegm~5*|    .099153      .02183    4.54   0.000    .05637  .141936   .147787
  espera |   .0044455      .00041   10.79   0.000   .003638  .005253   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Probit multinomial ordenado (clasificación 2):

```
Marginal effects after oprobit
      y  = Pr(clasificacion==2) (predict, pr outcome(2))
         =  .23532568
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|  -.0012622      .00649   -0.19   0.846  -.013992  .011467   .680866
    edad |  -.0014084      .00025   -5.61   0.000  -.001901 -.000916   52.1171
_Isegm~2*|  -.0520378      .09116   -0.57   0.568  -.230701  .126626   .001428
_Isegm~3*|  -.0179079      .01044   -1.71   0.086  -.038376   .00256   .567111
_Isegm~4*|   .0518055       .0091    5.69   0.000   .033965  .069646   .186578
_Isegm~5*|    .053895      .00872    6.18   0.000   .036803  .070987   .147787
  espera |   .0032294      .00032   10.13   0.000   .002605  .003854   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Efectos marginales en Probit multinomial ordenado (clasificación 3):

```
Marginal effects after oprobit
      y  = Pr(clasificacion==3) (predict, pr outcome(3))
         =   .6278671
------------------------------------------------------------------------------
variable |      dy/dx    Std. err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
_Igend~2*|   .0030042      .01548    0.19   0.846   -.02734  .033349   .680866
    edad |   .0033472      .00058    5.75   0.000   .002207  .004487   52.1171
_Isegm~2*|   .1079848      .16839    0.64   0.521   -.22206   .43803   .001428
_Isegm~3*|   .0429403      .02522    1.70   0.089  -.006492  .092373   .567111
_Isegm~4*|  -.1423059      .02913   -4.88   0.000  -.199406 -.085205   .186578
_Isegm~5*|  -.1530479      .03018   -5.07   0.000  -.212194 -.093902   .147787
  espera |   -.007675       .0007  -11.01   0.000  -.009042 -.006308   11.1349
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

(g) *Realizar un test de la significatividad de las variables.*

Stata.

<u>Ejercicio 2:</u> Modelo Secuencial.

(a) *Considerar la base de datos "nlsw88.dta". En la misma, hay datos de un grupo de mujeres de entre 30 y 40 años para estudiar los patrones de la fuerza laboral. Estimar un logit secuencial con la decisión de educación utilizando el comando seqlogit y mostrar que se pueden obtener los mismos resultados estimando varios modelos logit por separado.*

<u>Logit secuencial:</u>

```
                                                  Number of obs =  2,244
                                                  LR chi2(9)    = 108.50
Log likelihood = -2882.1386                       Prob > chi2   = 0.0000

------------------------------------------------------------------------------
    educ_cat | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
_2_3_4v1     |
        race |
       Black |  -.9151569   .1282466    -7.14   0.000    -1.166516   -.6637983
       Other |  -.4910998   .5511525    -0.89   0.373    -1.571339    .5891394
             |
       south |
       South |  -.4175069   .1259601    -3.31   0.001    -.6643841   -.1706298
       _cons |   2.250353   .0952967    23.61   0.000     2.063574    2.437131
-------------+----------------------------------------------------------------
_3_4v2       |
        race |
       Black |   -.173837   .1131414    -1.54   0.124    -.3955902    .0479161
       Other |   1.745005   .6241267     2.80   0.005     .5217389    2.968271
             |
       south |
       South |  -.1495226   .0968386    -1.54   0.123    -.3393228    .0402777
       _cons |   .1079773   .0617595     1.75   0.080    -.0130691    .2290237
-------------+----------------------------------------------------------------
_4v3         |
        race |
       Black |  -.3065161   .1648533    -1.86   0.063    -.6296227    .0165905
       Other |  -.3798123   .4723054    -0.80   0.421    -1.305514    .5458893
             |
       south |
       South |   .4052292    .138966     2.92   0.004     .1328609    .6775975
       _cons |   .0396236   .0855118     0.46   0.643    -.1279765    .2072237
------------------------------------------------------------------------------
```

Logit (High School):

```
Logistic regression                                     Number of obs =  2,244
                                                        LR chi2(3)    =  78.50
                                                        Prob > chi2   = 0.0000
Log likelihood = -904.78566                             Pseudo R2     = 0.0416

------------------------------------------------------------------------------
          hs | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
       Black |  -.9151569   .1282466    -7.14   0.000    -1.166516   -.6637983
       Other |  -.4910998   .5511525    -0.89   0.373    -1.571339    .5891394
             |
       south |
       South |  -.4175069   .1259601    -3.31   0.001    -.6643841   -.1706298
       _cons |   2.250353   .0952967    23.61   0.000     2.063574    2.437131
------------------------------------------------------------------------------
```

Logit (Junior College):

```
Logistic regression                                     Number of obs =  1,910
                                                        LR chi2(3)    =  18.95
                                                        Prob > chi2   = 0.0003
Log likelihood = -1314.2871                             Pseudo R2     = 0.0072

------------------------------------------------------------------------------
          sc | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
       Black |   -.173837   .1131414    -1.54   0.124    -.3955902    .0479161
       Other |   1.745005   .6241267     2.80   0.005     .5217389    2.968271
             |
       south |
       South |  -.1495226   .0968386    -1.54   0.123    -.3393228    .0402777
       _cons |   .1079773   .0617595     1.75   0.080    -.0130691    .2290237
------------------------------------------------------------------------------
```

Logit (College):

```
Logistic regression                                     Number of obs =    967
                                                        LR chi2(3)    =  11.05
                                                        Prob > chi2   = 0.0114
Log likelihood = -663.06592                             Pseudo R2     = 0.0083

------------------------------------------------------------------------------
           c | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
       Black |  -.3065161   .1648533    -1.86   0.063    -.6296227    .0165905
       Other |  -.3798123   .4723054    -0.80   0.421    -1.305514    .5458893
             |
       south |
       South |   .4052292    .138966     2.92   0.004     .1328609    .6775974
       _cons |   .0396236   .0855118     0.46   0.643    -.1279765    .2072236
------------------------------------------------------------------------------
```

(b) *Considerar la base de datos "gss.dta". La misma posee datos de la encuesta GSS (General Social Survey). Esta encuesta realiza investigaciones científicas básicas sobre la estructura y el desarrollo de la sociedad estadounidense con un programa de recopilación de datos diseñado tanto para monitorear el cambio social dentro de Estados Unidos como para comparar a Estados Unidos con otras naciones. Iniciado en 1972, el*

GSS contiene un núcleo estándar de preguntas demográficas, de comportamiento y de actitud, además de temas de especial interés. Muchas de las preguntas centrales se han mantenido sin cambios desde 1972 para facilitar los estudios de tendencias temporales, así como la replicación de hallazgos anteriores. En este ejercicio, se utilizan datos de educación similares a los del inciso anterior. Estimar un logit secuencial, interpretar los resultados y mostrar el efecto de la educación del padre en las decisiones de educación en cada transición.

Logit secuencial:

```
                                                  Number of obs =   9,842
                                                  LR chi2(18)   = 2461.15
Log likelihood = -9530.0004                       Prob > chi2   =  0.0000

------------------------------------------------------------------------------
             degree | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------------+---------------------------------------------------------
_1_2_3v0            |
              south |  -.7967635   .0736484   -10.82   0.000    -.9411116   -.6524153
                coh |   .7483053   .3414704     2.19   0.028     .0790356    1.417575
                    |
        c.coh#c.coh |  -.0482221   .0400122    -1.21   0.228    -.1266445    .0302003
                    |
             paeduc |   .1124402   .0778119     1.45   0.148    -.0400684    .2649488
                    |
      c.paeduc#c.coh |   .0469452   .0369009     1.27   0.203    -.0253792    .1192696
                    |
c.paeduc#c.coh#c.coh |  -.0050879   .0041484    -1.23   0.220    -.0132187    .0030428
                    |
              _cons |  -1.782385   .6862366    -2.60   0.009    -3.127385   -.4373864
--------------------+---------------------------------------------------------
_2_3v1              |
              south |   .0469273   .0521384     0.90   0.368     -.055262    .1491166
                coh |   .3228634   .4189998     0.77   0.441     -.498361    1.144088
                    |
        c.coh#c.coh |  -.0371565   .0445171    -0.83   0.404    -.1244084    .0500954
                    |
             paeduc |   .1222627   .0808644     1.51   0.131    -.0362286     .280754
                    |
      c.paeduc#c.coh |   .0188174   .0344105     0.55   0.584    -.0486259    .0862607
                    |
c.paeduc#c.coh#c.coh |   -.000731   .0035726    -0.20   0.838    -.0077331    .0062712
                    |
              _cons |  -3.497795    .956858    -3.66   0.000    -5.373202   -1.622388
--------------------+---------------------------------------------------------
_3v2                |
              south |   .0710731   .0976914     0.73   0.467    -.1203984    .2625446
                coh |   .9594559   .8457289     1.13   0.257    -.6981422    2.617054
                    |
        c.coh#c.coh |  -.1700969   .0872356    -1.95   0.051    -.3410755    .0008818
                    |
             paeduc |   .3357249   .1775429     1.89   0.059    -.0122528    .6837027
                    |
      c.paeduc#c.coh |  -.1217749   .0719208    -1.69   0.090     -.262737    .0191873
                    |
c.paeduc#c.coh#c.coh |   .0155494   .0071984     2.16   0.031     .0014408    .0296579
                    |
              _cons |  -.6964155   2.011413    -0.35   0.729    -4.638713    3.245882
------------------------------------------------------------------------------
```

Trabajo Práctico N° 5:
Modelos para Variables Dependientes Limitadas - Tobit.

Ejercicio 1: Variables Censuradas (Modelo Tobit I).

El modelo Tobit es relevante cuando la variable dependiente y de una regresión lineal se observa solo en algún intervalo de su soporte, porque, en este caso, los estimadores de MCC no son consistentes.

(a) *Considerar la base "auto.dta". Estimar el modelo:*

$mpg = \alpha + \beta wgt + u,$

donde $wgt = \frac{weight}{1000}$. *Luego, estimar el modelo generando una variable censurada suponiendo que no se observan autos con* $mpg \leq 17$. *Estimar por MCC y utilizando un modelo Tobit. Comparar.*

OLS:

```
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(1, 72)        =    134.62
       Model |  1591.99024         1  1591.99024   Prob > F        =    0.0000
    Residual |  851.469221        72  11.8259614   R-squared       =    0.6515
-------------+----------------------------------   Adj R-squared   =    0.6467
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.4389

------------------------------------------------------------------------------
         mpg | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         wgt |  -6.008687   .5178782   -11.60   0.000    -7.041058   -4.976316
       _cons |   39.44028   1.614003    24.44   0.000     36.22283    42.65774
------------------------------------------------------------------------------
```

OLS (ll(17)):

```
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(1, 72)        =     95.06
       Model |  1138.32073         1  1138.32073   Prob > F        =    0.0000
    Residual |  862.219806        72  11.9752751   R-squared       =    0.5690
-------------+----------------------------------   Adj R-squared   =    0.5630
       Total |  2000.54054        73  27.4046649   Root MSE        =    3.4605

------------------------------------------------------------------------------
       mpg_a | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         wgt |  -5.080912   .5211373    -9.75   0.000     -6.11978   -4.042044
       _cons |   37.12539    1.62416    22.86   0.000     33.88769     40.3631
------------------------------------------------------------------------------
```

Tobit (ll(17)):

```
Tobit regression                                  Number of obs     =         74
                                                        Uncensored  =         56
Limits: Lower =   17                              Left-censored     =         18
        Upper = +inf                              Right-censored    =          0

                                                  LR chi2(1)        =      72.85
                                                  Prob > chi2       =     0.0000
Log likelihood = -164.25438                       Pseudo R2         =     0.1815

------------------------------------------------------------------------------
       mpg_a | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         wgt |   -6.87305    .700257    -9.82   0.000    -8.268661    -5.47744
       _cons |   41.49856   2.058384    20.16   0.000      37.3962    45.60091
-------------+----------------------------------------------------------------
 var(e.mpg_a)|   14.78942   2.817609                      10.11698    21.61977
------------------------------------------------------------------------------
```

Tabla comparativa:

```
------------------------------------------------------------
                      (1)             (2)             (3)
                      OLS      OLS ll(17)    Tobit ll(17)
------------------------------------------------------------
main
wgt                -6.009***       -5.081***       -6.873***
                  (0.518)         (0.521)         (0.700)

_cons               39.44***        37.13***        41.50***
                  (1.614)         (1.624)         (2.058)
------------------------------------------------------------
/
var(e.mpg_a)                                        14.79***
                                                  (2.818)
------------------------------------------------------------
N                     74              74              74
R-sq               0.652           0.569
pseudo R-sq                                         0.182
------------------------------------------------------------
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
```

(b) *Repetir el inciso anterior suponiendo que, ahora, no se observan autos con mpg* $\geq$ *24.*

OLS (ul(24)):

```
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(1, 72)        =    186.15
       Model |  690.810491         1  690.810491   Prob > F        =    0.0000
    Residual |  267.189509        72   3.7109654   R-squared       =    0.7211
-------------+----------------------------------   Adj R-squared   =    0.7172
       Total |         958        73  13.1232877   Root MSE        =    1.9264

------------------------------------------------------------------------------
       mpg_b | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         wgt |  -3.958119   .2901034   -13.64   0.000    -4.536429   -3.379808
       _cons |   31.95138   .9041273    35.34   0.000     30.14903    33.75372
------------------------------------------------------------------------------
```

Tobit (ul(24)):

```
Tobit regression                                  Number of obs     =         74
                                                        Uncensored =         51
Limits: Lower = -inf                                Left-censored =          0
        Upper =   24                               Right-censored =         23

                                                  LR chi2(1)        =      90.72
                                                  Prob > chi2       =     0.0000
Log likelihood = -129.8279                        Pseudo R2         =     0.2589

------------------------------------------------------------------------------
       mpg_b | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         wgt |  -5.080645   .4349309   -11.68   0.000    -5.947461   -4.213829
       _cons |   36.08037   1.432059    25.19   0.000     33.22628    38.93446
-------------+----------------------------------------------------------------
 var(e.mpg_b)|   5.689927   1.166256                      3.781783    8.560846
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS	(2) OLS ul(24)	(3) Tobit ul(24)
main			
wgt	-6.009***	-3.958***	-5.081***
	(0.518)	(0.290)	(0.435)
_cons	39.44***	31.95***	36.08***
	(1.614)	(0.904)	(1.432)
/			
var(e.mpg_b)			5.690***
			(1.166)
N	74	74	74
R-sq	0.652	0.721	
pseudo R-sq			0.259

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

(c) *¿Cómo se interpretan los coeficientes del modelo? Computar los efectos marginales.*

Los coeficientes estimados miden cómo cambia la variable latente no observada con respecto a los cambios en las variables independientes, *céteris páribus*.

Efectos marginales (condicionales) con censura en Tobit (ll(17)):

```
Conditional marginal effects                                    Number of obs = 74
Model VCE: OIM

Expression: E(mpg_a*|mpg_a>17), predict(ystar(17,.))
dy/dx wrt:  wgt
1._at: wgt = 1
2._at: wgt = 2
3._at: wgt = 3
4._at: wgt = 4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wgt          |
         _at |
          1  |  -6.873035   1.389235    -4.95   0.000    -9.595886   -4.150183
          2  |  -6.855268   .7044715    -9.73   0.000    -8.236007    -5.47453
          3  |  -5.797116   .5880797    -9.86   0.000    -6.949731   -4.644501
          4  |  -1.499391   .3662326    -4.09   0.000    -2.217194   -.7815884
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con truncamiento en Tobit (ll(17)):

```
Conditional marginal effects                                    Number of obs = 74
Model VCE: OIM

Expression: E(mpg_a|mpg_a>17), predict(e(17,.))
dy/dx wrt:  wgt
1._at: wgt = 1
2._at: wgt = 2
3._at: wgt = 3
4._at: wgt = 4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wgt          |
         _at |
          1  |  -6.872705    .700472    -9.81   0.000    -8.245605   -5.499805
          2  |  -6.718373   .7348761    -9.14   0.000    -8.158703   -5.278042
          3  |  -4.345679   .4915117    -8.84   0.000    -5.309024   -3.382334
          4  |  -1.560439   .1287703   -12.12   0.000    -1.812825   -1.308054
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con censura en Tobit (ul(24)):

```
Conditional marginal effects                                    Number of obs = 74
Model VCE: OIM

Expression: E(mpg_b*|mpg_b<24), predict(ystar(.,24))
dy/dx wrt:  wgt
1._at: wgt = 1
2._at: wgt = 2
3._at: wgt = 3
4._at: wgt = 4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wgt          |
         _at |
          1  |  -.0085382   .0114991    -0.74   0.458     -.031076    .0139997
          2  |  -1.069716   .2842071    -3.76   0.000    -1.626752   -.5126807
          3  |  -4.610593   .3715716   -12.41   0.000     -5.33886   -3.882326
          4  |  -5.079249   .4349007   -11.68   0.000    -5.931638   -4.226859
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con truncamiento en Tobit (ul(24)):

```
Conditional marginal effects                                    Number of obs = 74
Model VCE: OIM

Expression: E(mpg_b|mpg_b<24), predict(e(.,24))
dy/dx wrt:  wgt
1._at: wgt = 1
2._at: wgt = 2
3._at: wgt = 3
4._at: wgt = 4

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wgt          |
         _at |
          1  |  -.3691762   .0534955    -6.90   0.000    -.4740255   -.2643269
          2  |   -1.13567   .1001953   -11.33   0.000    -1.332049    -.939291
          3  |  -3.681238   .3548315   -10.37   0.000    -4.376695   -2.985781
          4  |   -5.06274   .4362475   -11.61   0.000    -5.917769    -4.20771
------------------------------------------------------------------------------
```

<u>Ejercicio 2:</u> Variables Censuradas (Modelo Tobit II).

El siguiente ejercicio está tomado de Cameron & Trivedi. La variable dependiente para el gasto ambulatorio (ambulatory expenditure, ambexp) y los regresores (age, female, educ, blhisp, totchr, ins) se obtienen de la encuesta Medical Expenditure Panel Survey de 2001.

(a) *Abrir y describir la base "mus16datav2.dta". ¿Qué se puede decir sobre el cumplimiento de las condiciones que requiere un Tobit?*

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      ambexp |      3,328    1386.519    2530.406          0      49960
         age |      3,328    4.056881    1.121212        2.1        6.4
      female |      3,328    .5084135    .5000043          0          1
        educ |      3,328    13.40565    2.574199          0         17
      blhisp |      3,328    .3085938    .4619824          0          1
      totchr |      3,328    .4831731    .7720426          0          5
         ins |      3,328    .3650841    .4815261          0          1

                            ambexp
-------------------------------------------------------------
      Percentiles      Smallest
 1%           22              1
 5%           67              2
10%          107              2       Obs                2,802
25%          275              4       Sum of wgt.        2,802

50%          779                      Mean              1646.8
                        Largest       Std. dev.       2678.914
75%         1913          28269
90%         3967          30920       Variance         7176579
95%         6027          34964       Skewness        5.799312
99%        12467          49960       Kurtosis        65.81969
```

Lo que se puede decir sobre el cumplimiento de las condiciones que requiere Tobit es que, en principio, la asimetría y la curtosis no normal (alejadas de 0 y 3, respectivamente) de la variable dependiente *ambexp* podrían deberse a regresores que están sesgados.

Tobit:

```
Tobit regression                                  Number of obs     =  3,328
                                                        Uncensored  =  2,802
Limits: Lower =    0                                Left-censored  =    526
        Upper = +inf                                Right-censored  =      0

                                                  LR chi2(6)        = 694.07
                                                  Prob > chi2       = 0.0000
Log likelihood = -26359.424                       Pseudo R2         = 0.0130

------------------------------------------------------------------------------
       ambexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+---------------------------------------------------------------
          age |   314.1479   42.63366     7.37   0.000      230.557    397.7388
       female |   684.9918   92.85464     7.38   0.000     502.9337    867.0498
         educ |    70.8656   18.57365     3.82   0.000     34.44865    107.2825
       blhisp |   -530.311   104.2669    -5.09   0.000    -734.7448   -325.8772
       totchr |   1244.578   60.51376    20.57   0.000      1125.93    1363.226
          ins |  -167.4714   96.46088    -1.74   0.083    -356.6002    21.65734
        _cons |  -1882.591   317.4305    -5.93   0.000    -2504.971   -1260.212
--------------+---------------------------------------------------------------
 var(e.ambexp)|    6635296   179247.7                      6292994     6996217
------------------------------------------------------------------------------
```

(b) *Computar los efectos marginales.*

Efectos marginales (promedio) con censura en Tobit:

```
Average marginal effects                                  Number of obs = 3,328
Model VCE: OIM

Expression: E(ambexp*|ambexp>0), predict(ystar(0,.))
dy/dx wrt:  age female educ blhisp totchr ins

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   201.4409   27.29283     7.38   0.000     147.9479    254.9338
      female |   439.2368   59.32556     7.40   0.000     322.9608    555.5127
        educ |    45.4411   11.89795     3.82   0.000     22.12154    68.76066
      blhisp |  -340.0509   66.77218    -5.09   0.000     -470.922   -209.1799
      totchr |     798.06   38.00729    21.00   0.000     723.5671    872.5529
         ins |  -107.3876   61.86227    -1.74   0.083    -228.6354    13.86024
------------------------------------------------------------------------------
```

Efectos marginales (promedio) con truncamiento en Tobit:

```
Average marginal effects                                    Number of obs = 3,328
Model VCE: OIM

Expression: E(ambexp|ambexp>0), predict(e(0,.))
dy/dx wrt:  age female educ blhisp totchr ins

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |    147.796   20.14716     7.34   0.000     108.3083    187.2838
      female |   322.2656    43.7895     7.36   0.000     236.4397    408.0914
        educ |   33.33988   8.742173     3.81   0.000     16.20554    50.47422
      blhisp |  -249.4935   49.12834    -5.08   0.000    -345.7832   -153.2037
      totchr |   585.5322   29.01047    20.18   0.000     528.6727    642.3917
         ins |  -78.78967   45.40264    -1.74   0.083    -167.7772    10.19787
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con censura en Tobit:

```
Conditional marginal effects                                Number of obs = 3,328
Model VCE: OIM

Expression: E(ambexp*|ambexp>0), predict(ystar(0,.))
dy/dx wrt:  age female educ blhisp totchr ins
At: age    = 4.056881 (mean)
    female = .5084135 (mean)
    educ   = 13.40565 (mean)
    blhisp = .3085938 (mean)
    totchr = .4831731 (mean)
    ins    = .3650841 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |    207.526    28.2054     7.36   0.000     152.2444    262.8076
      female |   452.5052   61.30341     7.38   0.000     332.3528    572.6577
        educ |   46.81378   12.26552     3.82   0.000     22.77381    70.85375
      blhisp |  -350.3232   68.86825    -5.09   0.000    -485.3025   -215.3439
      totchr |   822.1678   40.61039    20.25   0.000     742.5729    901.7627
         ins |  -110.6315   63.74577    -1.74   0.083    -235.5709    14.30787
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con truncamiento en Tobit:

```
Conditional marginal effects                              Number of obs = 3,328
Model VCE: OIM

Expression: E(ambexp|ambexp>0), predict(e(0,.))
dy/dx wrt:  age female educ blhisp totchr ins
At: age    = 4.056881 (mean)
    female = .5084135 (mean)
    educ   = 13.40565 (mean)
    blhisp = .3085938 (mean)
    totchr = .4831731 (mean)
    ins    = .3650841 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |    145.524    19.7808     7.36   0.000     106.7543    184.2936
      female |   317.3113   42.99069     7.38   0.000     233.0511    401.5716
        educ |   32.82734   8.601086     3.82   0.000     15.96952    49.68516
      blhisp |   -245.658   48.29427    -5.09   0.000     -340.313   -151.0029
      totchr |   576.5307   28.50492    20.23   0.000     520.6621    632.3993
         ins |  -77.57842    44.7012    -1.74   0.083    -165.1912    10.03432
------------------------------------------------------------------------------
```

(c) *Computar los efectos marginales haciendo las cuentas con los comandos de escalares y matrices de Stata.*

Stata.

(d) *Considerar la variable dependiente en logaritmos. ¿Qué interpretación tiene esto sobre la variable dependiente? ¿Qué complicaciones introduce en el análisis? Estimar un Tobit para el logaritmo de ambexp.*

La variable dependiente en logaritmos introduce dos complicaciones en el análisis: un umbral distinto de cero y una variable dependiente lognormal.

OLS (con variable dependiente en logaritmos):

```
      Source |       SS           df       MS      Number of obs   =     3,328
-------------+----------------------------------   F(6, 3321)      =    169.68
       Model |  5772.79592         6  962.132653   Prob > F        =    0.0000
    Residual |  18831.0239     3,321  5.67028725   R-squared       =    0.2346
-------------+----------------------------------   Adj R-squared   =    0.2332
       Total |  24603.8199     3,327  7.39519683   Root MSE        =    2.3812

------------------------------------------------------------------------------
     lambexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .3247317    .038348     8.47   0.000     .2495436    .3999199
      female |   1.144695   .0833418    13.73   0.000     .9812886    1.308102
        educ |    .114108   .0165414     6.90   0.000     .0816757    .1465403
      blhisp |  -.7341754   .0928854    -7.90   0.000    -.9162938   -.5520571
      totchr |   1.059395   .0553699    19.13   0.000     .9508324    1.167958
         ins |   .2078343   .0869061     2.39   0.017     .0374394    .3782293
       _cons |   1.728764   .2812597     6.15   0.000     1.177304    2.280224
------------------------------------------------------------------------------
```

Tobit (con variable dependiente en logaritmos):

```
Tobit regression                                  Number of obs     =  3,328
                                                        Uncensored  =  2,802
Limits: Lower = -0.00                                   Left-censored =    526
        Upper =  +inf                                   Right-censored =      0

                                                  LR chi2(6)        = 831.03
                                                  Prob > chi2       = 0.0000
Log likelihood = -7494.29                         Pseudo R2         = 0.0525

------------------------------------------------------------------------------
       lambexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+--------------------------------------------------------------
           age |   .3630699   .0453222     8.01   0.000     .2742077    .4519321
        female |   1.341809   .0986074    13.61   0.000     1.148471    1.535146
          educ |    .138446   .0196568     7.04   0.000     .0999054    .1769866
        blhisp |  -.8731611   .1102504    -7.92   0.000    -1.089327   -.6569955
        totchr |   1.161268   .0649655    17.88   0.000     1.033891    1.288644
           ins |   .2612202    .102613     2.55   0.011     .0600292    .4624112
         _cons |   .9237178   .3350343     2.76   0.006     .2668233    1.580612
---------------+--------------------------------------------------------------
 var(e.lambexp)|   7.735265   .2181984                      7.319064    8.175133
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS (log)	(2) Tobit (log)
main		
age	0.325***	0.363***
	(0.0383)	(0.0453)
female	1.145***	1.342***
	(0.0833)	(0.0986)
educ	0.114***	0.138***
	(0.0165)	(0.0197)
blhisp	-0.734***	-0.873***
	(0.0929)	(0.110)
totchr	1.059***	1.161***
	(0.0554)	(0.0650)
ins	0.208**	0.261**
	(0.0869)	(0.103)
_cons	1.729***	0.924***
	(0.281)	(0.335)
/		
var(e.lamb~)		7.735***
		(0.218)
N	3328	3328
R-sq	0.235	
pseudo R-sq		0.053

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Ejercicio 3: Variables Censuradas (Modelo Tobit III).

Considerar la base de datos "mroz.dta", que posee datos que permiten estudiar la oferta laboral anual de mujeres casadas. Considerar las horas trabajadas, hours, y las explicativas nwifeinc, educ, exper, expersq, age, kidslt6, kidsge6. Estimar un modelo lineal y un modelo Tobit. Comparar. Computar los efectos marginales.

OLS:

```
      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(7, 745)       =     38.50
       Model |   151647606         7  21663943.7   Prob > F        =    0.0000
    Residual |   419262118       745  562767.944   R-squared       =    0.2656
-------------+----------------------------------   Adj R-squared   =    0.2587
       Total |   570909724       752  759188.463   Root MSE        =    750.18

------------------------------------------------------------------------------
       hours | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -442.0899    58.8466    -7.51   0.000    -557.6148    -326.565
     kidsge6 |  -32.77923   23.17622    -1.41   0.158     -78.2777    12.71924
         age |  -30.51163   4.363868    -6.99   0.000    -39.07858   -21.94469
        educ |   28.76112   12.95459     2.22   0.027     3.329283    54.19297
       exper |   65.67251   9.962983     6.59   0.000     46.11365    85.23138
    nwifeinc |  -3.446636      2.544    -1.35   0.176    -8.440898    1.547626
     expersq |  -.7004939   .3245501    -2.16   0.031    -1.337635   -.0633524
       _cons |   1330.482   270.7846     4.91   0.000     798.8906    1862.074
------------------------------------------------------------------------------
```

Tobit:

```
Tobit regression                                  Number of obs     =        753
                                                        Uncensored  =        428
Limits: Lower =    0                                    Left-censored =        325
        Upper = +inf                                    Right-censored =         0

                                                  LR chi2(7)        =     271.59
                                                  Prob > chi2       =     0.0000
Log likelihood = -3819.0946                       Pseudo R2         =     0.0343

------------------------------------------------------------------------------
       hours | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -894.0202   111.8777    -7.99   0.000    -1113.653   -674.3875
     kidsge6 |  -16.21805    38.6413    -0.42   0.675    -92.07668    59.64057
         age |  -54.40491   7.418483    -7.33   0.000     -68.9685   -39.84133
        educ |   80.64541   21.58318     3.74   0.000     38.27441    123.0164
       exper |    131.564   17.27935     7.61   0.000     97.64211     165.486
    nwifeinc |  -8.814226   4.459089    -1.98   0.048    -17.56808   -.0603706
     expersq |  -1.864153   .5376606    -3.47   0.001    -2.919661   -.8086455
       _cons |   965.3068   446.4351     2.16   0.031     88.88827    1841.725
-------------+----------------------------------------------------------------
 var(e.hours)|    1258927   93304.48                      1088458     1456093
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS	(2) Tobit
main		
kidslt6	-442.1***	-894.0***
	(58.85)	(111.9)
kidsge6	-32.78	-16.22
	(23.18)	(38.64)
age	-30.51***	-54.40***
	(4.364)	(7.418)
educ	28.76**	80.65***
	(12.95)	(21.58)
exper	65.67***	131.6***
	(9.963)	(17.28)
nwifeinc	-3.447	-8.814**
	(2.544)	(4.459)
expersq	-0.700**	-1.864***
	(0.325)	(0.538)
_cons	1330.5***	965.3**
	(270.8)	(446.4)
/		
var(e.hours)		1258926.8***
		(93304.5)
N	753	753
R-sq	0.266	
pseudo R-sq		0.034

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Efectos marginales (promedio) con censura en Tobit:

```
Average marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: E(hours*|hours>0), predict(ystar(0,.))
dy/dx wrt:  kidslt6 kidsge6 age educ exper nwifeinc expersq

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -526.2776   64.70619    -8.13   0.000    -653.0994   -399.4558
     kidsge6 |  -9.546986   22.75224    -0.42   0.675    -54.14056    35.04659
         age |  -32.02622    4.29211    -7.46   0.000     -40.4386   -23.61384
        educ |   47.47306    12.6214     3.76   0.000     22.73558    72.21054
       exper |   77.44703    9.99765     7.75   0.000     57.85199    97.04206
    nwifeinc |  -5.188619   2.621409    -1.98   0.048    -10.32649   -.0507514
     expersq |   -1.09736   .3155945    -3.48   0.001    -1.715914   -.4788063
------------------------------------------------------------------------------
```

Efectos marginales (promedio) con truncamiento en Tobit:

```
Average marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: E(hours|hours>0), predict(e(0,.))
dy/dx wrt:  kidslt6 kidsge6 age educ exper nwifeinc expersq

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -402.5505   50.74874    -7.93   0.000    -502.0162   -303.0848
     kidsge6 |  -7.302504   17.40426    -0.42   0.675    -41.41423    26.80922
         age |   -24.4969   3.362491    -7.29   0.000    -31.08726   -17.90654
        educ |   36.31221   9.703035     3.74   0.000     17.29461    55.32981
       exper |   59.23934    7.83368     7.56   0.000     43.88561    74.59308
    nwifeinc |  -3.968782   2.007582    -1.98   0.048    -7.903569   -.0339945
     expersq |  -.8393724   .2423183    -3.46   0.001    -1.314307   -.3644373
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con censura en Tobit:

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: E(hours*|hours>0), predict(ystar(0,.))
dy/dx wrt:  kidslt6 kidsge6 age educ exper nwifeinc expersq
At: kidslt6  = .2377158 (mean)
    kidsge6  = 1.353254 (mean)
    age      = 42.53785 (mean)
    educ     = 12.28685 (mean)
    exper    = 10.63081 (mean)
    nwifeinc = 20.12896 (mean)
    expersq  = 178.0385 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -540.2567   66.62387    -8.11   0.000    -670.8371   -409.6763
     kidsge6 |  -9.800576   23.36132    -0.42   0.675    -55.58792    35.98677
         age |  -32.87691   4.457699    -7.38   0.000    -41.61384   -24.13998
        educ |   48.73405    12.9634     3.76   0.000     23.32625    74.14185
       exper |   79.50419   10.30495     7.72   0.000     59.30685    99.70153
    nwifeinc |   -5.32644   2.690724    -1.98   0.048    -10.60016   -.0527175
     expersq |  -1.126508   .3232603    -3.48   0.000    -1.760087     -.49293
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con truncamiento en Tobit:

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: OIM

Expression: E(hours|hours>0), predict(e(0,.))
dy/dx wrt:  kidslt6 kidsge6 age educ exper nwifeinc expersq
At: kidslt6  = .2377158 (mean)
    kidsge6  = 1.353254 (mean)
    age      = 42.53785 (mean)
    educ     = 12.28685 (mean)
    exper    = 10.63081 (mean)
    nwifeinc = 20.12896 (mean)
    expersq  = 178.0385 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidslt6 |  -379.9678   46.79714    -8.12   0.000    -471.6885   -288.2471
     kidsge6 |  -6.892841   16.42951    -0.42   0.675    -39.09409     25.3084
         age |  -23.12265   3.130037    -7.39   0.000    -29.25741   -16.98789
        educ |   34.27513   9.117076     3.76   0.000     16.40599    52.14427
       exper |   55.91608   7.239109     7.72   0.000     41.72769    70.10447
    nwifeinc |  -3.746137    1.89236    -1.98   0.048    -7.455095     -.03718
     expersq |  -.7922845   .2273444    -3.48   0.000    -1.237871   -.3466976
------------------------------------------------------------------------------
```

Juan Menduiña

Trabajo Práctico N° 6:
Modelos para Variables Dependientes Limitadas - Heckman.

Ejercicio 1: Gastos Ambulatorios.

Retomar la base de datos del Ejercicio 2 del Problem Set 5. Ahora, se estimará un modelo de dos partes de Heckman. Estos modelos sirven para muestras autoseleccionadas. Se modela, explícitamente, la ecuación que determina la selección y la ecuación de interés. En este ejercicio, se pide estimar un modelo de Heckman para los gastos ambulatorios y comparar con las predicciones de un modelo Tobit.

Heckman (MLE):

```
Heckman selection model                         Number of obs     =      3,328
(regression model with sample selection)              Selected    =      2,802
                                                      Nonselected =        526

                                                Wald chi2(6)      =     288.88
Log likelihood = -5836.219                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
lambexp      |
         age |   .2119749   .0230072     9.21   0.000     .1668816    .2570682
      female |   .3481441   .0601142     5.79   0.000     .2303223    .4659658
        educ |    .018716   .0105473     1.77   0.076    -.0019563    .0393883
      blhisp |  -.2185714   .0596687    -3.66   0.000    -.3355199    -.101623
      totchr |     .53992   .0393324    13.73   0.000     .4628299      .61701
         ins |  -.0299871   .0510882    -0.59   0.557    -.1301182    .0701439
       _cons |   5.044056   .2281259    22.11   0.000     4.596938    5.491175
-------------+----------------------------------------------------------------
dambexp      |
         age |   .0879359    .027421     3.21   0.001     .0341917      .14168
      female |   .6626649   .0609384    10.87   0.000     .5432278    .7821021
        educ |   .0619485   .0120295     5.15   0.000     .0383711    .0855258
      blhisp |  -.3639377   .0618734    -5.88   0.000    -.4852073   -.2426682
      totchr |   .7969518   .0711306    11.20   0.000     .6575383    .9363653
         ins |   .1701367   .0628711     2.71   0.007     .0469117    .2933618
      income |   .0027078   .0013168     2.06   0.040      .000127    .0052886
       _cons |  -.6760546   .1940288    -3.48   0.000    -1.056344   -.2957652
-------------+----------------------------------------------------------------
     /athrho |  -.1313456   .1496292    -0.88   0.380    -.4246134    .1619222
    /lnsigma |   .2398173   .0144598    16.59   0.000     .2114767     .268158
-------------+----------------------------------------------------------------
         rho |  -.1305955   .1470772                    -.4008098    .1605217
       sigma |   1.271017   .0183786                     1.235501    1.307554
      lambda |  -.1659891   .1878698                    -.5342072    .2022291
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 0.91         Prob > chi2 = 0.3406
```

Heckman (Two Step):

```
Heckman selection model -- two-step estimates   Number of obs     =      3,328
(regression model with sample selection)              Selected    =      2,802
                                                      Nonselected =        526

                                                Wald chi2(6)      =     193.43
                                                Prob > chi2       =     0.0000
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
lambexp      |
         age |   .2024668   .0242202     8.36   0.000     .1549961    .2499374
      female |   .2921341   .0725756     4.03   0.000     .1498886    .4343796
        educ |   .0123889   .0115682     1.07   0.284    -.0102844    .0350622
      blhisp |  -.1828659   .0653449    -2.80   0.005    -.3109396   -.0547922
      totchr |   .5006332   .0485548    10.31   0.000     .4054675    .5957988
         ins |  -.0465097   .0529742    -0.88   0.380    -.1503373    .0573179
       _cons |   5.288927    .288522    18.33   0.000     4.723435     5.85442
-------------+----------------------------------------------------------------
dambexp      |
         age |   .0868152   .0274556     3.16   0.002     .0330032    .1406272
      female |   .6635053   .0609648    10.88   0.000     .5440165    .7829941
        educ |    .061884    .012039     5.14   0.000      .038288    .0854801
      blhisp |  -.3657835   .0619095    -5.91   0.000    -.4871239   -.2444432
      totchr |   .7957496   .0712174    11.17   0.000      .656166    .9353332
         ins |    .169107   .0629296     2.69   0.007     .0457673    .2924467
      income |   .0026773   .0013105     2.04   0.041     .0001088    .0052458
       _cons |  -.6686471   .1941247    -3.44   0.001    -1.049125   -.2881698
-------------+----------------------------------------------------------------
/mills       |
      lambda |  -.4637133   .2825997    -1.64   0.101    -1.017598     .090172
-------------+----------------------------------------------------------------
         rho |   -0.35907
       sigma |  1.2914258
------------------------------------------------------------------------------
```

Tobit:

```
Tobit regression                                  Number of obs     =  2,802
                                                        Uncensored  =  2,801
Limits: Lower =    0                                Left-censored  =      1
        Upper = +inf                                Right-censored  =      0

                                                  LR chi2(6)        = 596.53
                                                  Prob > chi2       = 0.0000
Log likelihood = -4642.5217                       Pseudo R2         = 0.0604
------------------------------------------------------------------------------
       lambexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+--------------------------------------------------------------
           age |   .2172778   .0222024     9.79   0.000     .1737431    .2608126
        female |   .3795502   .0485335     7.82   0.000     .2843851    .4747153
          educ |   .0221958   .0097527     2.28   0.023     .0030726    .0413191
        blhisp |  -.2384675   .0551452    -4.32   0.000     -.346597   -.1303381
        totchr |   .5618619   .0304802    18.43   0.000      .502096    .6216278
           ins |  -.0210413   .0499613    -0.42   0.674     -.119006    .0769234
         _cons |   4.908076   .1679989    29.21   0.000     4.578661     5.23749
---------------+--------------------------------------------------------------
 var(e.lambexp)|   1.608909   .0429988                      1.526767     1.69547
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) Heckman (M~)	(2) Heckman (T~)	(3) Tobit
lambexp			
age	0.212***	0.202***	0.217***
	(0.0230)	(0.0242)	(0.0222)
female	0.348***	0.292***	0.380***
	(0.0601)	(0.0726)	(0.0485)
educ	0.0187*	0.0124	0.0222**
	(0.0105)	(0.0116)	(0.00975)
blhisp	-0.219***	-0.183***	-0.238***
	(0.0597)	(0.0653)	(0.0551)
totchr	0.540***	0.501***	0.562***
	(0.0393)	(0.0486)	(0.0305)
ins	-0.0300	-0.0465	-0.0210
	(0.0511)	(0.0530)	(0.0500)
_cons	5.044***	5.289***	4.908***
	(0.228)	(0.289)	(0.168)
dambexp			
age	0.0879***	0.0868***	
	(0.0274)	(0.0275)	
female	0.663***	0.664***	
	(0.0609)	(0.0610)	
educ	0.0619***	0.0619***	
	(0.0120)	(0.0120)	
blhisp	-0.364***	-0.366***	
	(0.0619)	(0.0619)	
totchr	0.797***	0.796***	
	(0.0711)	(0.0712)	
ins	0.170***	0.169***	
	(0.0629)	(0.0629)	
income	0.00271**	0.00268**	
	(0.00132)	(0.00131)	
_cons	-0.676***	-0.669***	
	(0.194)	(0.194)	
/			
athrho	-0.131		
	(0.150)		
lnsigma	0.240***		
	(0.0145)		
var(e.lamb~)			1.609***
			(0.0430)
/mills			
lambda		-0.464	
		(0.283)	
N	3328	3328	2802
pseudo R-sq			0.060

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

<u>Ejercicio 2:</u> Ecuación Salarial para las Mujeres I.

Considerar la base de datos "womenwk.dta". Describir la base. Estimar una ecuación salarial en función de la educación y la edad por Mínimos Cuadrados Clásicos. Repetir utilizando un modelo de Heckman, utilizando las variables married, children, education y age para la ecuación de selección. Utilizar el comando heckman.

<u>Descripción de la base:</u>

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      county |      2,000         4.5       2.873          0          9
         age |      2,000      36.208     8.28656         20         59
   education |      2,000      13.084    3.045912         10         20
     married |      2,000       .6705    .4701492          0          1
    children |      2,000      1.6445    1.398963          0          5
        wage |      1,343    23.69217    6.305374    5.88497   45.80979
```

```
            Hourly wage; missing, if not working
-------------------------------------------------------------
      Percentiles      Smallest
 1%     9.728734        5.88497
 5%     13.48302       6.739784
10%     15.69925        7.12612       Obs                1,343
25%     19.30873       7.328383       Sum of wgt.        1,343

50%     23.51122                      Mean            23.69217
                        Largest       Std. dev.       6.305374
75%     28.05009       43.01642
90%     31.49893       43.97919       Variance        39.75775
95%     33.98332       44.53403       Skewness        .1881963
99%     40.34642       45.80979       Kurtosis        3.048037
```

<u>OLS:</u>

```
      Source |       SS           df       MS      Number of obs   =     2,000
-------------+----------------------------------   F(3, 1996)      =    140.75
       Model |  52555.2814         3  17518.4271   Prob > F        =    0.0000
    Residual |  248439.676     1,996  124.468775   R-squared       =    0.1746
-------------+----------------------------------   Adj R-squared   =    0.1734
       Total |  300994.957     1,999  150.572765   Root MSE        =    11.157

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |    .369376   .0324995    11.37   0.000     .3056395    .4331124
   education |   1.024154   .0863307    11.86   0.000     .8548468    1.193462
     married |   1.269777   .5790207     2.19   0.028     .1342283    2.405325
       _cons |   -11.7165   1.411936    -8.30   0.000    -14.48552   -8.947476
------------------------------------------------------------------------------
```

Heckman (MLE):

```
Heckman selection model                         Number of obs     =      2,000
(regression model with sample selection)              Selected    =      1,343
                                                      Nonselected =        657

                                                Wald chi2(3)      =     508.52
Log likelihood = -5178.289                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wage         |
         age |   .2121393   .0213504     9.94   0.000     .1702933    .2539852
   education |   .9881493   .0542321    18.22   0.000     .8818563    1.094442
     married |    .066304   .3758994     0.18   0.860    -.6704452    .8030532
       _cons |   .4973339    1.07856     0.46   0.645    -1.616605    2.611273
-------------+----------------------------------------------------------------
dwage        |
         age |   .0364354   .0041745     8.73   0.000     .0282535    .0446174
   education |   .0555733   .0107731     5.16   0.000     .0344585    .0766882
     married |   .4499889    .072705     6.19   0.000     .3074898     .592488
    children |   .4385259   .0277979    15.78   0.000      .384043    .4930087
       _cons |  -2.489276   .1896044   -13.13   0.000    -2.860893   -2.117658
-------------+----------------------------------------------------------------
     /athrho |   .8753773   .1015349     8.62   0.000     .6763725    1.074382
   /lnsigma  |   1.792839   .0276367    64.87   0.000     1.738672    1.847006
-------------+----------------------------------------------------------------
         rho |   .7040959   .0511989                      .5891561    .7911065
       sigma |   6.006483   .1659993                      5.689785    6.340809
      lambda |    4.22914   .3994723                      3.446189    5.012092
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0): chi2(1) = 60.72        Prob > chi2 = 0.0000
```

Heckman (Two Step):

```
Heckman selection model -- two-step estimates   Number of obs     =      2,000
(regression model with sample selection)              Selected    =      1,343
                                                      Nonselected =        657

                                                Wald chi2(3)      =     442.08
                                                Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
wage         |
         age |   .2108123   .0225447     9.35   0.000     .1666255     .254999
   education |   .9804939   .0546614    17.94   0.000     .8733596    1.087628
     married |   .0863959   .3776478     0.23   0.819    -.6537802     .826572
       _cons |    .730021   1.249191     0.58   0.559    -1.718349    3.178391
-------------+----------------------------------------------------------------
dwage        |
         age |   .0347211   .0042293     8.21   0.000     .0264318    .0430105
   education |   .0583645   .0109742     5.32   0.000     .0368555    .0798735
     married |   .4308575    .074208     5.81   0.000     .2854125    .5763025
    children |   .4473249   .0287417    15.56   0.000     .3909922    .5036576
       _cons |  -2.467365   .1925635   -12.81   0.000    -2.844782   -2.089948
-------------+----------------------------------------------------------------
/mills       |
      lambda |   4.021226   .6126901     6.56   0.000     2.820375    5.222077
-------------+----------------------------------------------------------------
         rho |    0.67552
       sigma |  5.9528138
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS	(2) Heckman (M~)	(3) Heckman (T~)
main			
age	0.369***	0.212***	0.211***
	(0.0325)	(0.0214)	(0.0225)
education	1.024***	0.988***	0.980***
	(0.0863)	(0.0542)	(0.0547)
married	1.270**	0.0663	0.0864
	(0.579)	(0.376)	(0.378)
_cons	-11.72***	0.497	0.730
	(1.412)	(1.079)	(1.249)
dwage			
age		0.0364***	0.0347***
		(0.00417)	(0.00423)
education		0.0556***	0.0584***
		(0.0108)	(0.0110)
married		0.450***	0.431***
		(0.0727)	(0.0742)
children		0.439***	0.447***
		(0.0278)	(0.0287)
_cons		-2.489***	-2.467***
		(0.190)	(0.193)
/			
athrho		0.875***	
		(0.102)	
lnsigma		1.793***	
		(0.0276)	
/mills			
lambda			4.021***
			(0.613)
N	2000	2000	2000
R-sq	0.175		

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Ejercicio 3: Ecuación Salarial para las Mujeres II.

Conceptualmente, se va a repetir el ejercicio anterior utilizando la base de datos "mroz.dta" que ya se ha utilizado. Ahora, se pide modelar, explícitamente, la ecuación de selección con un Probit y la ecuación estructural con un modelo lineal aumentada por la inversa del ratio de Mills. Reportar el efecto marginal sobre las horas trabajadas, correctamente, estimado.

OLS:

```
      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(6, 746)       =     33.05
       Model |   119885614         6  19980935.6   Prob > F        =    0.0000
    Residual |   451024110       746   604589.96   R-squared       =    0.2100
-------------+----------------------------------   Adj R-squared   =    0.2036
       Total |   570909724       752  759188.463   Root MSE        =    777.55

------------------------------------------------------------------------------
       hours | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidsge6 |  -13.56954   23.87531    -0.57   0.570    -60.44032    33.30125
         age |  -17.10219   4.127445    -4.14   0.000    -25.20499   -8.999404
        educ |    23.9582   13.41096     1.79   0.074    -2.369512    50.28591
       exper |   74.12513   10.26049     7.22   0.000     53.98227      94.268
    nwifeinc |  -4.336964   2.633972    -1.65   0.100    -9.507843     .833916
     expersq |  -.9264192   .3349462    -2.77   0.006    -1.583968   -.2688699
       _cons |   656.2857   264.8041     2.48   0.013     136.4358    1176.136
------------------------------------------------------------------------------
```

Heckman (Two Step):

```
Heckman selection model -- two-step estimates   Number of obs     =        753
(regression model with sample selection)              Selected    =        428
                                                      Nonselected =        325

                                                Wald chi2(6)      =      26.17
                                                Prob > chi2       =     0.0002

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
hours        |
     kidsge6 |  -83.74795   33.16153    -2.53   0.012    -148.7433   -18.75256
         age |  -2.839866   6.990271    -0.41   0.685    -16.54054    10.86081
        educ |  -63.81931   21.02964    -3.03   0.002    -105.0366   -22.60196
       exper |   6.070658   21.16833     0.29   0.774     -35.4185    47.55982
    nwifeinc |   4.458736    4.03176     1.11   0.269    -3.443369    12.36084
     expersq |   .1358569   .5265464     0.26   0.796     -.896155    1.167869
       _cons |    2477.33   425.3662     5.82   0.000     1643.627    3311.032
-------------+----------------------------------------------------------------
dhours       |
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
       _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
-------------+----------------------------------------------------------------
/mills       |
      lambda |  -621.8712   199.0294    -3.12   0.002    -1011.962   -231.7808
-------------+----------------------------------------------------------------
         rho |   -0.74244
       sigma |  837.60041
------------------------------------------------------------------------------
```

Tabla comparativa:

	(1) OLS	(2) Heckman (T~)
main		
kidsge6	-13.57	-83.75**
	(23.88)	(33.16)
age	-17.10***	-2.840
	(4.127)	(6.990)
educ	23.96*	-63.82***
	(13.41)	(21.03)
exper	74.13***	6.071
	(10.26)	(21.17)
nwifeinc	-4.337	4.459
	(2.634)	(4.032)
expersq	-0.926***	0.136
	(0.335)	(0.527)
_cons	656.3**	2477.3***
	(264.8)	(425.4)
dhours		
kidsge6		0.0360
		(0.0435)
age		-0.0529***
		(0.00848)
educ		0.131***
		(0.0253)
exper		0.123***
		(0.0187)
nwifeinc		-0.0120**
		(0.00484)
expersq		-0.00189***
		(0.000600)
kidslt6		-0.868***
		(0.119)
_cons		0.270
		(0.509)
/mills		
lambda		-621.9***
		(199.0)
N	753	753
R-sq	0.210	

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Efectos marginales (promedio) con censura en Heckman (Two Step):

```
Average marginal effects                                    Number of obs = 753
Model VCE: Conventional

Expression: E(hours*|hours>0), predict(ystar(0,.))
dy/dx wrt:  kidsge6 age educ exper nwifeinc expersq kidslt6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidsge6 |  -81.38955   32.34639    -2.52   0.012    -144.7873   -17.99179
         age |  -2.759893   6.777967    -0.41   0.684    -16.04446    10.52468
        educ |  -62.02211   20.76646    -2.99   0.003    -102.7236   -21.32059
       exper |   5.899704   20.52631     0.29   0.774    -34.33112    46.13052
    nwifeinc |   4.333175   3.931451     1.10   0.270    -3.372327    12.03868
     expersq |    .132031   .5124218     0.26   0.797    -.8722971    1.136359
     kidslt6 |          0  (omitted)
------------------------------------------------------------------------------
```

Efectos marginales (promedio) con truncamiento en Heckman (Two Step):

```
Average marginal effects                                    Number of obs = 753
Model VCE: Conventional

Expression: E(hours|hours>0), predict(e(0,.))
dy/dx wrt:  kidsge6 age educ exper nwifeinc expersq kidslt6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidsge6 |  -73.14433   29.50712    -2.48   0.013    -130.9772   -15.31144
         age |    -2.4803   6.058449    -0.41   0.682    -14.35464    9.394042
        educ |  -55.73892   19.50195    -2.86   0.004    -93.96204    -17.5158
       exper |   5.302031   18.34814     0.29   0.773    -30.65967    41.26373
    nwifeinc |   3.894199   3.565572     1.09   0.275    -3.094194    10.88259
     expersq |   .1186556   .4620408     0.26   0.797    -.7869278    1.024239
     kidslt6 |          0  (omitted)
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con censura en Heckman (Two Step):

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: Conventional

Expression: E(hours*|hours>0), predict(ystar(0,.))
dy/dx wrt:  kidsge6 age educ exper nwifeinc expersq kidslt6
At: kidsge6  = 1.353254 (mean)
    age      = 42.53785 (mean)
    educ     = 12.28685 (mean)
    exper    = 10.63081 (mean)
    nwifeinc = 20.12896 (mean)
    expersq  = 178.0385 (mean)
    kidslt6  = .2377158 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidsge6 |  -81.62997   32.48895    -2.51   0.012    -145.3071    -17.9528
         age |  -2.768046    6.79893    -0.41   0.684     -16.0937    10.55761
        educ |  -62.20532   20.85318    -2.98   0.003    -103.0768   -21.33383
       exper |   5.917131   20.58963     0.29   0.774     -34.4378    46.27207
    nwifeinc |   4.345974     3.9435     1.10   0.270    -3.383144    12.07509
     expersq |   .1324211   .5138982     0.26   0.797    -.8748009    1.139643
     kidslt6 |          0  (omitted)
------------------------------------------------------------------------------
```

Efectos marginales (condicionales) con truncamiento en Heckman (Two Step):

```
Conditional marginal effects                                    Number of obs = 753
Model VCE: Conventional

Expression: E(hours|hours>0), predict(e(0,.))
dy/dx wrt:  kidsge6 age educ exper nwifeinc expersq kidslt6
At: kidsge6  = 1.353254 (mean)
    age      = 42.53785 (mean)
    educ     = 12.28685 (mean)
    exper    = 10.63081 (mean)
    nwifeinc = 20.12896 (mean)
    expersq  = 178.0385 (mean)
    kidslt6  = .2377158 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     kidsge6 |  -73.52816   29.74524    -2.47   0.013    -131.8278   -15.22856
         age |  -2.493316   6.090065    -0.41   0.682    -14.42962    9.442992
        educ |  -56.03141   19.67987    -2.85   0.004    -94.60325   -17.45957
       exper |   5.329854    18.4439     0.29   0.773    -30.81952    41.47923
    nwifeinc |   3.914635   3.586477     1.09   0.275    -3.114731      10.944
     expersq |   .1192782   .4644865     0.26   0.797    -.7910986    1.029655
     kidslt6 |          0  (omitted)
------------------------------------------------------------------------------
```