**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 (X). Mostrar que el porcentaje general predicho correctamente es un promedio ponderado del porcentaje predicho para la variable dependiente igual a 0 () y del porcentaje predicho para la variable dependiente igual a 1 (), donde las ponderaciones son las proporciones de ceros y de unos en la muestra, respectivamente.*

= = .

= = .

= =

=

=

= + .

**Ejercicio 2: Interpretación del Modelo de Probabilidad Lineal I.**

*Suponer que se estima el modelo:*

= + + ,

*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 la cantidad de elementos que verican = 1. Llamar a la media de la variable x tomada sólo para aquellos elementos que verican = 1 y a la media de la variable x tomada sobre los valores restantes. Mostrar que:*

*= ,*

*donde p= .*

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

=

=

=

=

=

=

=

=

=

=

=

=

= .

**Ejercicio 3: Interpretación del Modelo de Probabilidad Lineal II.**

*Sea y una resultado binario y sean , , … , 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 sobre , , … , 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 (= 1), es decir, = (proporcion de “éxitos” de cada categoría), siendo la cantidad de observaciones con = 1, k= 1, … , M.

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

**(b)** *¿Qué ocurre si se regresa sobre M combinaciones lineales de , , … , linealmente independientes entre sí? Ayuda: Considerar 1, , … , .*

Lo que ocurre si se regresa sobre M combinaciones lineales de , , … , 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).

**Ejercicio 4: Efectos Marginales.**

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

*= g ( + x) , donde g (z)= (z).*

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

= g ( + x)

= g ( + x)

=

= .

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

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

El efecto parcial de cambiar de 0 a 1 es:

= P (y= 1 | = 1) - P (y= 1 | = 0)

= g ( + x) ,

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

En el caso en el que y es un indicador de empleo y 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 una variable discreta numérica. ¿Cuál es el efecto parcial de cambiar 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 la cantidad de cigarrillos que fuma por día.*

= P (y= 1 | = c + 1) - P (y= 1 | = c)

= g ( + x) ,

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

En el caso en el que y es un indicador de si la persona i fuma y 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 ( + + + log () + ).*

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

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

= g ( + + + log () + ) .

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

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

= g ( + + + log () + ) .

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

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

= = g ( + + + log () + ) .

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

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

= = g ( + + + log () + ) .

**(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.

**Ejercicio 5: 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.*

OLS:

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

------------------------------------------------------------------------------

Logit:

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) (2) (3)

OLS Logit 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 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 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:*

*= .*

*Realizar un cuadro de doble entrada con las variables y y . Comentar.*

| yhat\_probit

ins | 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 4 y 2,5. 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 ()= .

*Recordar que para un modelo Logit:*

*P (y= 1 | x)= .*

P (y= 1 | x)=

P (y= 1 | x)=

P (y= 1 | x)=

P (y= 1 | x)= .

P (y= 0 | x)= 1 - P (y= 1 | x)

P (y= 0 | x)= 1 -

P (y= 0 | x)=

P (y= 0 | x)= .

=

=

=

ln ()= ln

ln ()= ln e

ln ()= \* 1

ln ()= .

**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 = - .*

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*](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.