PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS SOCIALES
ESPECIALIDAD DE ECONOMÍA
SEMESTRE 2013-II

PUCP

ECONOMETRÍA 1: PRÁCTICA DIRIGIDA 9

Profesor:	Gabriel Rodríguez	(gabriel.rodriguez@pucp.edu.pe)
Jefes de práctica:	Patricia Lengua Lafosse	(patricia.lengua@pucp.pe)

1 Modelos de elección discreta (Logit – Probit)

¿Podemos representar mediante un modelo un fenómeno que es discreto? Por ejemplo, la composición del mercado de trabajo, la elección de una determinada marca de producto, qué candidato votar en las próximas elecciones.

$$y_i = \begin{cases} 1 & \text{si se presenta el estado o elección} \\ 0 & \text{en otro caso} \end{cases} \tag{1}$$

La mayoría de variables analizadas bajo esta técnica son cualitativas, estas variables representan un proceso de elección o son códigos que representan un estado. El análisis de estos modelos se centra en el marco general de los modelos de probabilidad. ¿Es conveniente el uso de modelo de regresión lineal ?

$$y_i = x_i'\beta + \epsilon_i \ , \ \ i = 1, 2, ...N \tag{2}$$

Cuando $y_i = 1, \epsilon_i = 1 - x_i'\beta$ y cuando $y_i = 0, \epsilon_i = -x_i'\beta$. La interpretación del modelo de regresión es:

$$\begin{aligned} E\left[y_i|x_i\right] &= 1 \times \Pr(y_i = 1|\, x_i) + 0 \times \Pr(y_i = 0|\, x_i) \\ &= \Pr(y_i = 1|\, x_i) \end{aligned} \tag{3}$$

donde $\frac{\partial E[y_i|x_i]}{\partial x_i} = \frac{\partial \Pr(y_i=1|x_i)}{\partial x_i} = \beta$, donde β es el cambio en $\Pr(y_i = 1|\, x_i)$ cuando x_i aumenta en una unidad; es decir, el vector de parámetros β refleja el impacto que las variables X tienen sobre la probabilidad. La pregunta es cómo modelar dichos efectos. El modelo de probabilidad lineal tiene varios inconvenientes:

- El error presenta heterocedasticidad dependiente de β.
- El efecto de las variables X sobre la probabilidad es constante y no depende del valor que toma X.
- Las probabilidades predichas pueden ser menor que 0 o mayor que 1.

¿Cómo restringir $x_i'\beta$ al intervalo [0,1]? Podemos restringir el vector de información $x_i'\beta$ a una función de distribución que tenga resultados entre 0 y 1. En muchos análisis se utiliza la distribución normal estándar acumulada $\Phi(x_i'\beta)$ (modelo probit) o la distribución logística acumulada $\Lambda(x_i'\beta)$ (modelo logit).[1]

¿Cuál de las dos distribuciones deberíamos utilizar?

$$\Pr(y_i = 1) = F(x_i'\beta) \tag{4}$$

Así tenemos lo siguiente:

$$\begin{aligned} \Pr(y_i &= 1) = F(x_i'\beta) & (5)\\ \Pr(y_i &= 0) = 1 - F(x_i'\beta) & (6) \end{aligned}$$

La distribución logística tiende a dar mayores probabilidades al suceso $Y = 0$ que la normal cuando $x_i'\beta$ es pequeño (y menores probabilidades al $Y = 0$ cuando $x_i'\beta$ es grande). Pero no existen reglas a ciencia cierta para elegir entre una distribución u otra.

El modelo de probabilidad es un modelo de regresión:

$$\begin{aligned} E[y_i|x_i] &= 1 \times \Pr(y_i = 1) + 0 \times \Pr(y_i = 0) & (7)\\ E[y_i|x_i] &= 1 \times F(x_i'\beta) + 0 \times (1 - F(x_i'\beta) = F(x_i'\beta) \end{aligned}$$

Nota:
Al margen de la función de distribución utilizada, los parámetros del modelo no son los efectos marginales como se acostumbra al analizar las regresiones lineales.

¿Cuál es el modelo a estimar?

Es una función de verosimilitud que considera la realización de ambos casos (modelo de elección binaria).

$$l(\beta) = \log L(\beta) = \sum y_i \log(1 - F(-x_i'\beta)) + (1 - y_i)\log(F(-x_i'\beta)) \tag{8}$$

Efectos marginales

En general, para cualquier modelo de regresión los efectos marginales se definen de la siguiente manera:

$$\frac{\partial E(y_i|x_i)}{\partial x_i^k} = \left[\frac{\partial F(x_i'\beta)}{\partial x_i^k}\right].\beta_k \tag{9}$$

Donde $F(.)$ es la función de distribución utilizada.

Dado que los efectos marginales variarán para cada valor de X (para cada individuo) es útil y práctico hacer simulaciones a partir de las medias de los regresores o en otros puntos que se quiera evaluar.

El cálculo de los efectos marginales parte del cálculo de la media muestral de los datos, como si estuviésemos computando los valores o calificaciones del individuo promedio de la muestra (individuo representativo).

---

[1]

- $\Phi(Z) = \int_{-\infty}^{Z} \phi(x)dx$ con $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}x^2)}$ $\quad -\infty < Z < \infty$
- $\Lambda(Z) = \frac{e^Z}{1+e^Z} = \frac{1}{e^{-Z}+1}$ con $-\infty < Z < \infty$

2 Laboratorio

El siguiente ejemplo fue tomado de Wooldridge (2002). Se analiza la participación en la fuerza laboral de las mujeres casadas. La variable $inlf$ toma el valor de 1 si la mujer participa en la fuerza laboral y 0 si no participa. Se estima un modelo logit que considera como variables explicativas:

1. Años de educación ($educ$).
2. Años de experiencia ($exper$).
3. Edad (age).
4. Ingresos de la familia sin considerar los ingresos de la esposa ($nwifeinc$).
5. Número de niños menores a 6 años ($kidslt6$).
6. Número de niños entre los 6 y 18 años ($kidsge6$).

El modelo a estimar es el siguiente:

$$inlf_i = \Lambda(\beta_0+\beta_1 nwifeinc_i+\beta_2 educ_i+\beta_3 \exp er_i+\beta_4 \exp er_i^2+\beta_5 age_i+\beta_6 kidslt6_i+\beta_7 kidsge6_i) + \varepsilon_i \tag{10}$$

Estadísticos clasificados por mujer en el mercado laboral

bysort inlf: summarize nwifeinc exper educ age kidslt6 kidsge6

-> inlf = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
nwifeinc	325	21.69805	12.72815	1.5	96
exper	325	7.461538	6.918567	0	45
educ	325	11.79692	2.181995	5	17
age	325	43.28308	8.467796	30	60
kidslt6	325	.3661538	.6368995	0	3
kidsge6	325	1.356923	1.327065	0	7

-> inlf = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
nwifeinc	428	18.93748	10.59135	-.0290575	91
exper	428	13.03738	8.055923	0	38
educ	428	12.65888	2.285376	5	17
age	428	41.97196	7.721084	30	60
kidslt6	428	.1401869	.3919231	0	2
kidsge6	428	1.350467	1.315935	0	8

Estimación del modelo de elección binaria

Logit

logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6, asis robust

```
Iteration 0:   log pseudo-likelihood =  -514.8732
Iteration 1:   log pseudo-likelihood = -406.94123
Iteration 2:   log pseudo-likelihood = -401.85151
Iteration 3:   log pseudo-likelihood = -401.76519
Iteration 4:   log pseudo-likelihood = -401.76515

Logit estimates                                   Number of obs   =        753
                                                  Wald chi2(7)    =     158.48
                                                  Prob > chi2     =     0.0000
Log pseudo-likelihood = -401.76515                Pseudo R2       =     0.2197

------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0213452   .0090781    -2.35   0.019     -.039138   -.0035523
        educ |   .2211704   .0444509     4.98   0.000     .1340483    .3082925
       exper |   .2058695   .0322914     6.38   0.000     .1425796    .2691594
     expersq |  -.0031541   .0010124    -3.12   0.002    -.0051384   -.0011698
         age |  -.0880244   .0144393    -6.10   0.000    -.1163248    -.059724
     kidslt6 |  -1.443354   .2031615    -7.10   0.000    -1.841543   -1.045165
     kidsge6 |   .0601122   .0798825     0.75   0.452    -.0964546     .216679
       _cons |   .4254524   .8597307     0.49   0.621    -1.259589    2.110494
------------------------------------------------------------------------------
```

Como observamos, todos los coeficientes son significativos salvo el asociado a la variable kidsge6 (niños entre los 6 y 18 años). Este resultado nos indicaría que dicha variable no permite discriminar entre mujeres casadas que participan la fuerza laboral y de las que no. Por tanto, no debería pertenecer a la especificación del modelo.

Estimación de los efectos marginales del logit

- mfx compute

```
Marginal effects after logit
      y  = Pr(inlf) (predict)
         =  .58277201
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
nwifeinc |  -.0051901      .00221   -2.35   0.019  -.009523 -.000857   20.129
    educ |   .0537773      .01086    4.95   0.000   .032498  .075057  12.2869
   exper |   .0500569      .00788    6.35   0.000   .034604   .06551  10.6308
 expersq |  -.0007669      .00025   -3.11   0.002  -.001251 -.000283  178.039
     age |   -.021403      .00353   -6.07   0.000  -.028317 -.014489  42.5378
 kidslt6 |  -.3509498      .04988   -7.04   0.000  -.448718 -.253182  .237716
 kidsge6 |   .0146162      .01941    0.75   0.451  -.023428   .05266  1.35325
------------------------------------------------------------------------------
```

El efecto marginal asociado a la variable kidslt6 , significa que si el número de niños menores de 6 años aumenta en una unidad, la probabilidad de pertenecer al mercado laboral se reducirá en 0.35. Por lo tanto, los efectos marginales están expresados en unidades probabilísticas.

- mfx compute, eyex

```
Elasticities after logit
      y  = Pr(inlf) (predict)
         =  .58277201
------------------------------------------------------------------------------
variable |      ey/ex    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
nwifeinc |  -.1792646      .07704   -2.33   0.020  -.33026  -.028269   20.129
    educ |   1.133812      .23647    4.79   0.000   .670343  1.59728  12.2869
   exper |   .9131284      .15044    6.07   0.000   .618268  1.20799  10.6308
 expersq |  -.2342952       .0765   -3.06   0.002  -.384235 -.084355  178.039
     age |  -1.562255      .26893   -5.81   0.000  -2.08935 -1.03516  42.5378
 kidslt6 |  -.1431543      .02171   -6.59   0.000  -.185705 -.100603  .237716
 kidsge6 |   .0339403      .04502    0.75   0.451    -.0543   .12218  1.35325
------------------------------------------------------------------------------
```

Si el número de niños menores de 6 años se incrementa en 1%, la probabilidad de pertenecer al mercado laboral se reducirá en -0.14%.

- mfx compute, dyex

```
Elasticities after logit
      y  = Pr(inlf) (predict)
         =  .58277201
```

variable	dy/ex	Std. Err.	z	P>|z|	[95% C.I.	]	X
nwifeinc	-.1044704	.0445	-2.35	0.019	-.191691	-.01725	20.129
educ	.6607539	.1334	4.95	0.000	.399295	.922213	12.2869
exper	.5321457	.08382	6.35	0.000	.367864	.696427	10.6308
expersq	-.1365407	.04393	-3.11	0.002	-.222651	-.05043	178.039
age	-.9104385	.15006	-6.07	0.000	-1.20456	-.616322	42.5378
kidslt6	-.0834263	.01186	-7.04	0.000	-.106667	-.060185	.237716
kidsge6	.0197794	.02627	0.75	0.451	-.031704	.071262	1.35325

Si el número de niños se incrementa en 1%, la probabilidad de pertenecer al mercado laboral se reducirá en -0.08 unidades probabilísticas.

Calculando los efectos marginales para el individuo promedio (simulaciones)

El comando para realizar estas simulaciones es nuevamente mfx, con la diferencia que ahora se hace explícito que se trata de un individuo en particular.

mfx compute, dydx at (20.12 12.28 10.63 178.03 42.54 0 1.35 1)

```
Marginal effects after logit
      y  = Pr(inlf) (predict)
         =  .66271621
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
nwifeinc	-.0047711	.00203	-2.35	0.019	-.008756	-.000786	20.12
educ	.0494368	.00999	4.95	0.000	.029851	.069023	12.28
exper	.0460167	.00736	6.25	0.000	.031586	.060447	10.63
expersq	-.000705	.00023	-3.09	0.002	-.001152	-.000258	178.03
age	-.0196755	.00314	-6.26	0.000	-.025839	-.013512	42.54
kidslt6	-.3226235	.04175	-7.73	0.000	-.40446	-.240787	0
kidsge6	.0134365	.01788	0.75	0.452	-.021598	.048471	1.35

mfx compute, dydx at (20.12 12.28 10.63 178.03 42.54 1 1.35 1)

```
Marginal effects after logit
      y  = Pr(inlf) (predict)
         =  .31692664
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
nwifeinc	-.0046209	.00197	-2.35	0.019	-.008473	-.000769	20.12
educ	.0478799	.00971	4.93	0.000	.028839	.066921	12.28
exper	.0445675	.00758	5.88	0.000	.029703	.059432	10.63
expersq	-.0006828	.00022	-3.05	0.002	-.001121	-.000245	178.03
age	-.0190559	.00285	-6.70	0.000	-.024633	-.013479	42.54
kidslt6	-.3124633	.02763	-11.31	0.000	-.366615	-.258312	1
kidsge6	.0130133	.01747	0.74	0.456	-.021236	.047262	1.35

Además de calcular los efectos marginales también calculamos la probabilidad estimada según el modelo logit. Para $nwifeinc = 20.12$, $educ = 12.28$, $exper = 10.63$, $expersq = 178.03$, $age = 42.54$ y $kidsge6 = 1$, calculamos la probabilidad de pertenecer al mercado laboral si $kidslt6 = 1$ y $kidslt6 = 0$. Dicho valor se aproxima al efecto marginal estimado previamente.

El siguiente paso es determinar el grado de ajuste del modelo. Debemos recordar que el modelo de elección binaria arroja como resultado valores en el intervalo abierto 0 y 1, es decir, probabilidades y los valores observados son 0 y 1. La pregunta que surge es ¿cómo comparamos los resultados con los valores observados?

Evaluación del modelo logit

- Pseudo R^2 de McFadden

  Este estadístico toma valores entre 0 y 1 y puede ser "leído" como el R^2 de los modelos lineales. La forma de construir el estadístico es simple pues se trata de una relación entre el modelo actual (irrestricto) y un modelo que sea el más simple, el que provee menos información (restricto). En este modelo el modelo restricto será aquel que solo contenga como variable explicativa la constante. Por tanto, las restricciones en nuestro ejemplo serán 7 que corresponden a los 7 parámetros sin considerar la constante.

  $$R^2_{McFadden} = 1 - \frac{l(\widehat{\beta})}{l(\widetilde{\beta})} \tag{11}$$

  donde:

  $l(\widetilde{\beta})$: Valor de la función de log-verosimilitud del modelo con restricciones.

  $l(\widehat{\beta})$: Valor de la función de log-verosimilitud del modelo sin restricciones.

  Un aspecto importante es que dicha lectura del R^2 de McFadden es menos exigente que la lectura del R^2 de los modelos lineales. Por ejemplo, cuando tenemos un R^2 de McFadden mayor a 0.5 estamos ante un muy buen modelo, valores entre 0.3 y 0.5 estamos ante un buen modelo. Mientras que valores menores a 0.1 indican que el modelo analizado no es bueno para clasificar a los individuos en las elecciones establecidas. En el ejemplo, el valor obtenido es 0.22, lo que nos indica que el modelo tiene una efectividad regular al momento de la clasificación.

- LR Statistic

  H0: Todos los coeficientes son estadísticamente igual a cero (sin considerar la constante).

  $$-2(l(\widetilde{\beta}) - l(\widehat{\beta})) \sim \chi^2_L \tag{12}$$

  donde:

  $l(\widetilde{\beta})$: Valor de la función de log-verosimilitud del modelo con restricciones.

  $l(\widehat{\beta})$: Valor de la función de log-verosimilitud del modelo sin restricciones.

  L : Número de restricciones, es igual al número de coeficientes estimados sin considerar la constante.

- Tabla de predicción

Otra aproximación del nivel de predicción del modelo es calculando la tabla de predicción. Para ello es necesario establecer un punto de corte, generalmente es 0.5. Como mencionamos, el modelo arroja valores entre 0 y 1, y al establecer un punto de corte de 0.5 estamos afirmando que toda probabilidad menor a 0.5 es muy "pequeña" y el individuo con esa probabilidad estimada será clasificado entre los individuos que eligieron el valor 0 de la variable dependiente. Si la probabilidad es mayor o igual a 0.5 entonces la probabilidad será grande y clasificaremos a esos individuos como los que eligieron la opción 1 de la variable dependiente.

lstat

```
Logistic model for inlf

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       347           118  |        465
     -     |        81           207  |        288
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .5
True D defined as inlf != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   81.07%
Specificity                     Pr( -|~D)   63.69%
Positive predictive value       Pr( D| +)   74.62%
Negative predictive value       Pr(~D| -)   71.88%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   36.31%
False - rate for true D         Pr( -| D)   18.93%
False + rate for classified +   Pr(~D| +)   25.38%
False - rate for classified -   Pr( D| -)   28.13%
--------------------------------------------------
Correctly classified                        73.57%
--------------------------------------------------
```

El modelo predice correctamente el 73.57% de las observaciones.