

Microeconometría
Problem Set 1
Modelo de Probabilidad Lineal, Logit y Probit

• **Ejercicio 1.** *Porcentaje correctamente predicho.*

Sea y una variable binaria y considere algún modelo de probabilidad $P(y = 1|x) = F(X\beta)$. Muestre 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.

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

Suponga que se estima el modelo

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

donde x 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 verifican $y_i = 1$. Llame \bar{x}_1 a la media de la variable x tomada solo para aquellos elementos que verifican $y_i = 1$ y \bar{x}_0 a la media de la variable x tomada sobre los valores restantes. Muestre 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 $p = \frac{n_1}{n}$.

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

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

(a) Muestre que los valores ajustados de la regresión sin intercepto

$$y_i \text{ sobre } d_{1i}, d_{2i}, \dots, d_{Mi}$$

están siempre en el intervalo unitario. En particular, describa qué representa cada coeficiente y el valor ajustado para cada i .

(b) ¿Qué ocurre si y_i se regresa sobre M combinaciones lineales de d_1, d_2, \dots, d_M linealmente independientes entre sí? *Ayuda: considere $1, d_2, \dots, d_M$.*

• **Ejercicio 4.** *Efectos marginales I.*

Sea y un resultado binario y $\mathbf{x} = (x_1, \dots, x_k)$ un vector de variables explicativas. Sea $G(\cdot)$ la función de distribución acumulada de una variable aleatoria continua. Recuerde que si x_j es continua, su efecto marginal se obtiene como

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = g(\beta_0 + \mathbf{x}\beta) \beta_j, \quad \text{donde } g(z) \equiv \frac{dG}{dz}(z)$$

(a) Muestre que los efectos relativos de dos variables explicativas cualesquiera no dependen de \mathbf{x} .

(b) Sea x_1 una variable binaria. ¿Cuál es el efecto parcial de cambiar x_1 de 0 a 1? ¿De qué depende? Interprete 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.

(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? Interprete 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.

Considere ahora el siguiente modelo

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

(d) ¿Cuál es el efecto parcial de z_1 sobre $P(y = 1 | \mathbf{z})$?

(e) ¿Cuál es el efecto parcial de z_2 sobre $P(y = 1 | \mathbf{z})$?

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

(g) ¿Cuál es la elasticidad de z_1 sobre $P(y = 1 | \mathbf{z})$?

(h) ¿Cómo obtendría errores estándar para todos estos efectos?

Microeconometría
Ejercicios Introdutorios a Stata 1
Modelo de Probabilidad Lineal, Logit y Probit

• **Ejercicio 1.** *MPL, Logit y Probit en Stata I.*

En este ejercicio usted va a demostrar algunas propiedades de las estimaciones para modelos con variable dependiente discreta.

- (a) Estime a `ins` contra `retire`, `age`, `hstatusg`, `hhincome`, `educyear`, `married`, `hisp` por OLS, Logit y Probit.
- (b) ¿Cuál es el problema de estimar el modelo por OLS?
- (c) Explique 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?
- (d) Para evaluar la eficacia de los modelos Probit y Logit defina como el valor estimado de la variable dependiente \hat{y} como

$$\hat{y} = \begin{cases} 1 & \text{si } P(\widehat{y=1|x}) > 0.5 \\ 0 & \text{si } P(\widehat{y=0|x}) \leq 0.5 \end{cases} \quad (1)$$

Realice un cuadro de doble entrada con las variables y y \hat{y} . Comente.

- (e) En la literatura se sugiere que $\beta^{\text{logit}} \approx 4\beta^{\text{ols}}$ y $\beta^{\text{probit}} \approx 2.5\beta^{\text{ols}}$. Compruébelo para esta muestra.
- (f) Compute la probabilidad esperada que `ins`=1 cuando las variables están evaluadas en la media.
- (g) Defina 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 $y = 0$. Demuestre que para el caso de un modelo Logit se verifica que

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

Recuerde que para un modelo Logit

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

• **Ejercicio 2.** *MPL, Logit y Probit en Stata II.*

Utilice la base de datos de Mroz, T.A. (1987): "*The Sensitivity 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, realice un análisis exploratorio simple de los datos. Para esto, puede ayudarse de los comandos `describe`, `summarize`, `browse`, `tab`.
- (b) Cree una variable de educación centrada. Recuerde que se le llama variable centrada a una variable transformada como $\tilde{x}_i = x_i - \bar{x}$.
- (c) Estudie gráficamente la relación entre el salario y la educación. Puede también desagregar por las variables `inlf`, `kidslt6`. Para esto, puede ayudarse de los comandos `graph`, `twoway`, `scatter`, `lfit` y sus opciones.
- (d) ¿Hay valores faltantes o duplicados en la muestra? Intente resolver esto sin el comando `browse` ni `edit`.
- (e) Estime un modelo de probabilidad lineal de `inlf` sobre `educ`, `city`, `exper`, `kidslt6`, `expersq`. Además, genere la predicción del modelo.
- (f) ¿Puede realizar inferencia con este modelo? Estime el modelo con errores estándares robustos. ¿Cómo cambian los resultados?
- (g) ¿Qué ocurre si elimina la constante del modelo?
- (h) ¿Qué ocurre si estima el modelo solo para una ciudad?
- (i) Estime un modelo logit de `inlf` sobre `educ`, `city`, `exper`, `kidslt6`, `expersq`.
- (j) Calcule la predicción del modelo.
- (k) Genere la curva ROC.

- (l) Calcule los efectos marginales en las medias.
- (m) Calcule los efectos marginales en valores particulares de la variable que le resulten de interés.
- (n) Estime un modelo probit con las mismas variables que en (i) y cree una tabla con las estimaciones de todos los modelos.

• **Ejercicio 3.** *Estimar el efecto de la educación sobre la probabilidad de estar desempleado.*

Utilice la EPH con datos de individuos del segundo trimestre de 2015 , disponible en [http://www. indec.gov.ar/bases-de-datos.asp](http://www.indec.gov.ar/bases-de-datos.asp). Use la muestra de jefes de hogar, hombres, 25-65 años, para todos los conglomerados disponibles. Estudie como se define el desempleo de acuerdo al INDEC. Rentrinja la muestra a personas empleadas o desempleadas, es decir excluya aquellos que están fuera de la fuerza laboral (no buscan trabajo, estudian, retirados, etc.). Use las ponderaciones **pondera**.

- (a) Utilice 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. Construya las probabilidades para cada individuo. ¿Qué proporción de la muestra tiene probabilidades predecidas mayores a 1 o menores a 0 ?
- (b) Estimar el modelo del inciso (a) usando los modelos Probit y logit. ¿Cómo cambian los resultados?
- (c) Estimar la probabilidad de estar desempleado para un hombre universitario (de grado) completo, casado, para cada área metropolitana de la EPH, para todos los aos 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.