

Econometría Avanzada

Problem Set 4: modelos de elección binaria

Estefanía Capriata, Tomás Pacheco y Abigail Riquelme

Fecha de entrega 20/05/2022

1. Consideramos que las variables que pueden ser relevantes para la deserción escolar son el ingreso per cápita, el sexo del jefe de hogar y de la persona encuestada, la cantidad de hermanos, si asiste a una institución pública o privada y el nivel educativo del jefe de hogar.

En la Tabla (1) observamos métricas importantes para las variables más relevantes para el análisis. El objetivo del trabajo es tratar de estimar la probabilidad de que una persona abandone el colegio; para esto, necesitamos tanto personas que lo hayan abandonado como personas que no. En la primera fila de esta tabla, podemos observar que de las 3783 personas de nuestra muestra, aproximadamente el 9 % desertó, lo cual nos dice que tenemos variabilidad en la muestra. En segundo lugar, podemos ver que casi el 49 % de la muestra es de sexo femenino y que en el 40 % de los hogares, la jefa del hogar es mujer. Finalmente, podemos observar que el ingreso per cápita promedio del hogar es, aproximadamente, \$10.850. El mínimo es \$166,67 y el hogar con más ingreso per cápita tiene un ingreso de \$175.000.

En el primer panel de la Figura (1) observamos la distribución del ingreso per cápita para aquellas personas que desertaron y aquellas que no lo hicieron. Si bien no es evidente que las diferencias en las distribuciones del ingreso per cápita para los dos grupos son radicalmente distintas, podemos observar que en ciertos tramos del ingreso, parecerían diferenciarse. Para la cola inferior de la distribución (ingresos más bajos), se observa que hay una densidad mayor para aquellas personas que desertaron. Al mismo tiempo, para un nivel de ingreso per cápita mayor (aproximadamente entre los 20.000 y 50.000 pesos, puede observarse que la densidad de personas que no desertaron es mayor. Este vago análisis puede darnos una primera aproximación a que podría existir una relación entre la deserción escolar y el nivel de ingreso. Sin embargo, nos vemos en la obligación de aclarar que esto es solo un análisis descriptivo. En el segundo panel de la figura en cuestión, (1) observamos el histograma de la cantidad de hermanos para las personas que desertaron y aquellas que no lo hicieron. En este caso, es posible notar que las distribuciones para ambos grupos son similares y que la mayoría de las personas tienen entre 2 y 3 hermanos.

Finalmente, en el tercer panel de la Figura (1) se presenta la proporción sobre el total de obser-

vaciones de personas que desertaron y las que no, según el nivel educativo del jefe del hogar. Podemos ver que en nuestra muestra, el nivel educativo del jefe de hogar más representado es el nivel secundario. Aproximadamente el 50 % de los jefes de hogar tienen este nivel. Luego le sigue el nivel primario. Algo importante a notar es que hay una diferencia considerable entre la proporción relativa de desertores y no desertores cuyo jefe de hogar terminó la escuela primaria. Esto junto con el hecho de que la proporción de desertores en nivel universitario y terciario es pequeña, podría darnos un indicio de que existe una correlación entre desertar y el nivel educativo del jefe del hogar. Nuevamente, nos vemos en la obligación de aclarar que este análisis es meramente descriptivo.

Cuadro 1: Estadística descriptiva

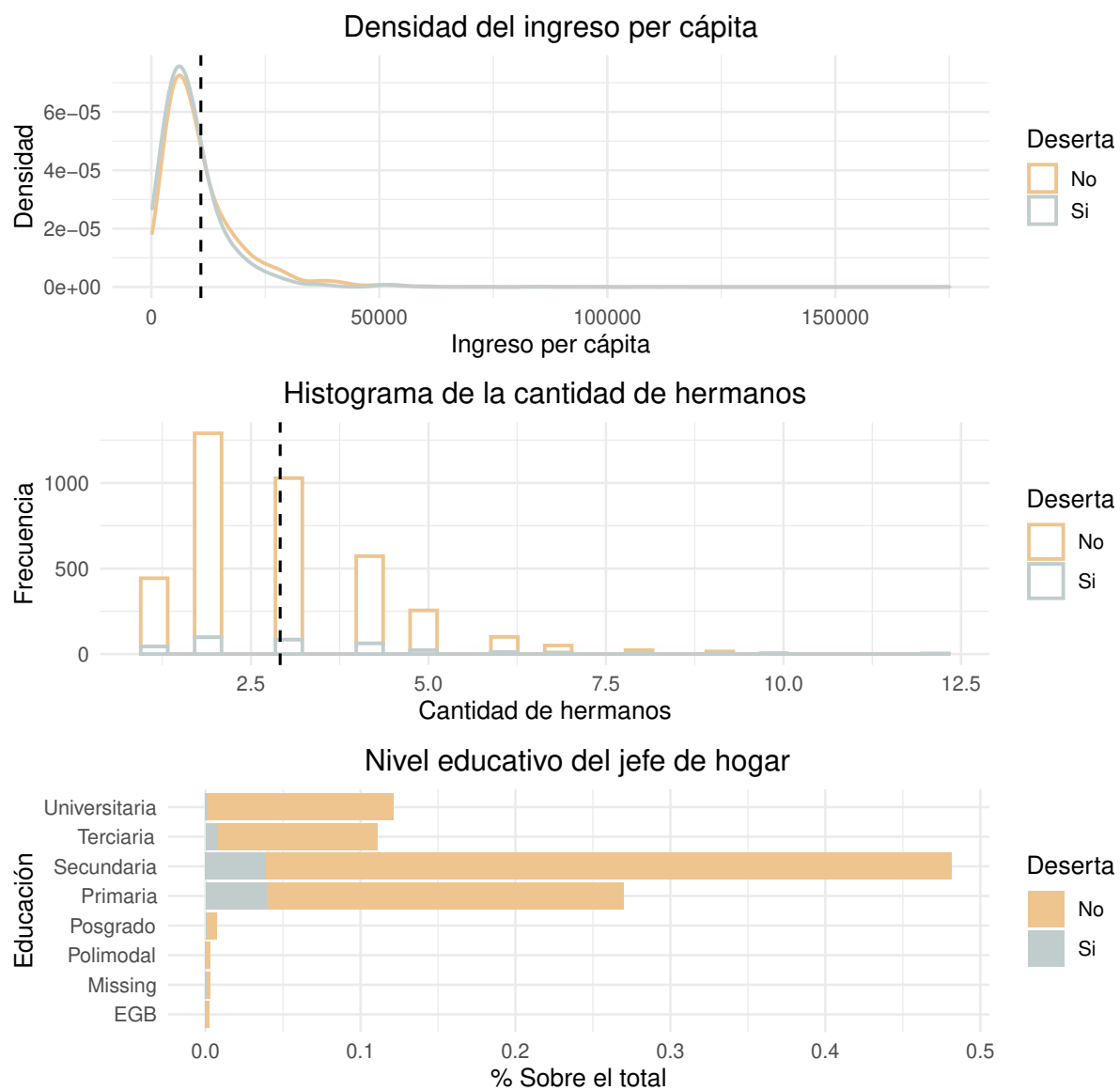
Estadístico	Obs	Media	Min	Max
Deserta	3,783	0.089	0	1
Sexo (mujer = 1)	3,783	0.489	0	1
Jefa de hogar (mujer = 1)	3,783	0.402	0	1
Ingreso per cápita	3,783	10,854.120	166.670	175,000.000

2. En la Columna 1 de la Tabla (1) obtenemos las estimaciones que surgen del modelo probit. Lo único que podemos interpretar de los coeficientes estimados es su significatividad y signo. Por un lado, podemos ver que la variable mujer, educación EGB, secundaria, terciaria y universitaria, cantidad de hermanos, ingreso per cápita y no responde a si el establecimiento en el que se educa es privado o público tienen un signo negativo. Para 5 de estas 8 variables el coeficiente es estadísticamente significativo. Por otro lado, podemos observar que las variables referidas al los datos *missing* del nivel educativo del jefe de hogar, educación polimodal, jefe de hogar mujer, establecimiento educativo público y *missings* del tipo de establecimiento educativo al que se asiste tienen signo positivo. Solo el coeficiente asociado a la variable de establecimiento educativo público es estadísticamente significativo.

Es importante aclarar que para el caso de las variables categóricas la interpretación de los signos de los coeficientes depende directamente de la categoría base que usamos. Para el caso de la variable mujer utilizamos como categoría base 'ser hombre', para las categorías referidas al nivel educativo del jefe de hogar utilizamos como base 'educación primaria', por último, para las variables referidas al tipo de establecimiento educativo al que se asiste utilizamos como base a los establecimientos privados.

Por lo tanto, es posible identificar que el coeficiente positivo que observamos para la variable mujer nos dice que si la persona es mujer la probabilidad de desertar, respecto a si se es hombre, es más baja. Ahora, si el jefe de hogar tiene un nivel educativo EGB, secundario, terciario o universitario la probabilidad de desertar es más baja respecto a si el jefe de hogar tiene un nivel educativo primario. Ahora, si el dato del nivel educativo del jefe de hogar no está disponible o es educación polimodal la probabilidad de desertar es mayor respecto a cuando el nivel edu-

Figura 1: Estadística descriptiva



cativo del jefe de hogar es primario. Por último, si el tipo de establecimiento educativo al que asiste la persona es público o no está disponible esta información, la probabilidad de desertar es mayor al caso en el que el establecimiento educativo es privado. Ahora, observamos que en aquellos casos en los que la persona no responde a esta pregunta, la probabilidad de desertar es menor a la probabilidad de desertar si se asiste a un establecimiento privado.

3. Para poder interpretar los resultados numéricamente y sacar conclusiones sobre los coeficientes es necesario observar los efectos marginales en algún punto de interés. En este caso, obtendremos los efectos marginales en las medias para las siguientes variables: cantidad de hermanos e ingreso per cápita; las demás variables consideradas en el modelo serán evaluadas en su categoría base. En otras palabras, el análisis marginal que haremos es solamente válido para un individuo promedio.

En primer lugar, la diferencia esperada en la probabilidad de desertar para un mujer un hombre es de aproximadamente dos puntos porcentuales, en promedio y *ceteris paribus*. En segundo lugar, las variables de los distintos niveles de educación, con respecto a la categoría base del nivel primario, están asociados con la probabilidad de desertar de la siguiente manera: tener un jefe de hogar con un nivel EGB está asociado con una probabilidad 3.8 pp menor de desertar respecto a cuando se tiene un jefe de hogar con educación primaria. En cambio, si se tiene un jefe de hogar con un nivel secundario la probabilidad es 2.1 pp menor respecto a la categoría, si el nivel educativo es terciario dicha probabilidad es 1.8 pp menor. Por último, si el jefe de hogar tiene un nivel de estudios universitario está asociado con una probabilidad que es -10 pp respecto a la categoría base y si tiene estudios de posgrado esta probabilidad es 0.9 pp menor.

Ahora, si la jefa de hogar es mujer encontramos que esto está asociado, en promedio y *ceteris paribus*, con una probabilidad de desertar 0.8 pp mayor que si el jefe de hogar es hombre. También podemos decir, respecto a la variable referida al tipo de establecimiento educativo, que si el establecimiento es público observamos una probabilidad de desertar mayor en 3.2 pp respecto al caso en el que el tipo de establecimiento educativo es privado.

Además, respecto a las variables que son no categóricas podemos hacer un análisis de los efectos marginales en la media. En particular, observamos que si aumentamos en una unidad la cantidad de hermanos para el individuo promedio la probabilidad de deserción asociada es 1 % menor. Ahora, si aumentamos, para el mismo individuo, el ingreso per cápita en una unidad identificamos que la probabilidad de deserción tiene un cambio nulo o muy pequeño.

Para poder comparar los efectos marginales de la media con otros puntos de interés usaremos también la moda, la mediana y en cierto valor específico (cuando el individuo tiene el ingreso mínimo y la cantidad mínima de hermanos), nuevamente las variables categóricas las evaluamos en la categoría base. Observamos que los efectos marginales computados son muy similares a los obtenidos en la media del ingreso per cápita y la cantidad de hermanos.

Cuadro 2: Estimaciones

	<i>Dependent variable:</i>		
	Deserta		
	(1)	(2)	(3)
Mujer	−0.221*** (0.065)	−0.027*** (0.008)	−0.193*** (0.162)
Educación JH (missing)	0.493 (0.468)	0.105 (0.122)	0.305 (0.467)
Educación JH (EGB)	−0.408 (0.527)	−0.056 (0.075)	−0.296 (0.526)
Educación JH (Secundario)	−0.229*** (0.071)	−0.041*** (0.011)	−0.353*** (0.070)
Educación JH (Polimodal)	0.045 (0.526)	−0.001 (0.086)	−0.372 (0.525)
Educación JH (Terciario)	−0.188 (0.129)	−0.039*** (0.016)	−0.480*** (0.128)
Educación JH (Universitario)	−1.115*** (0.210)	−0.088*** (0.013)	−1.004*** (0.210)
Educación JH (Posgrado)	−0.096 (0.431)	−0.027 (0.048)	−0.479 (0.430)
Cantidad de hermanos	−0.015 (0.023)	−0.001 (0.003)	0.008 (0.022)
Ingreso per cápita	−0.00000 (0.00000)	−0.00000 (0.00000)	
Log ingreso per capita			0.005 (0.047)
Jefe de hogar mujer	0.084 (0.066)	0.013 (0.009)	0.145** (0.065)
Establecimiento educativo (missing)	1.693*** (0.146)	0.357*** (0.024)	
Establecimiento educativo (público)	0.336** (0.141)	0.013 (0.008)	
Establecimiento educativo (no responde)	−2.926*** (0.253)	−0.082*** (0.013)	
Constant	−1.649*** (0.163)	0.085*** (0.017)	−0.976*** (0.162)
Modelo	Probit	MLP	Probit
Observations	3,783	3,783	4,518

Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Cuadro 3: Efectos marginales

Efecto marginal	Mujer	Missing	EGB	Secundario	Polimodal	Terciario	Universitario
Media	-0.021	0.046	-0.038	-0.022	0.004	-0.018	-0.105
Mediana	-0.021	0.046	-0.038	-0.022	0.004	-0.018	-0.105
Moda	-0.021	0.048	-0.04	-0.022	0.004	-0.018	-0.108
Caso particular	-0.022	0.049	-0.041	-0.023	0.004	-0.019	-0.111

Efecto marginal	Posgrado	Hermanos	Ingreso PC	Jefa de hogar	Missing	Publico	No responde
Media	-0.009	-0.001	0	0.008	0.159	0.032	-0.275
Mediana	-0.009	-0.001	0	0.008	0.159	0.032	-0.275
Moda	-0.009	-0.001	0	0.008	0.164	0.033	-0.283
Caso particular	-0.01	-0.002	0	0.008	0.169	0.034	-0.292

4. Para estimar el modelo lineal de probabilidad utilizaremos las variables de mujer, educación del jefe de hogar, cantidad de hermanos, ingreso per cápita, jefa de hogar mujer y tipo de establecimiento educativo. Las variables que eran estadísticamente significativas en el modelo anterior continúan igual en este modelo lineal e incluso mantienen su significatividad al mismo nivel, con la excepción de la variable que indica si el jefe de hogar tiene educación terciaria y la que indica si el establecimiento educativo es público. En el primer caso el coeficiente estimado pasa de no ser estadísticamente significativa al 1 % a serlo, en el segundo caso, el coeficiente estimado deja de ser significativo a los niveles usuales de significatividad.

Teniendo esto en cuenta, interpretaremos aquellos coeficientes que son estadísticamente distintos de cero como mínimo al 10 %. En particular, el coeficiente de la variable mujer nos dice que la diferencia de la probabilidad de desertar entre mujeres y hombres es de -2.7 puntos porcentuales, en promedio y ceteris paribus. Además, podemos decir que la diferencia en términos de probabilidad para personas cuyo jefe de hogar tiene educación secundaria, en comparación a escuela primaria, es de 4.1 puntos porcentuales, en promedio y dejando el resto de las variables constantes. Como mencionamos anteriormente, la variable de la educación a nivel terciario es significativa y muestra que, nuevamente con respecto a la categoría base, estaría asociada a diferencia 3.9 puntos porcentuales en la probabilidad de desertar. Lo mismo sucede con el nivel universitario: la diferencia de probabilidades es de 8.8 puntos porcentuales, en promedio y con todo el resto constante.

Existen distintas características que permiten comparar al modelo probit con uno de probabilidad lineal. En primer lugar, el modelo probit tendría la ventaja sobre el otro modelo ya que no es por definición heterocedástico. Sin embargo, es posible realizar una estimación robusta para corregir este problema. En segundo lugar, el modelo probit predice probabilidades que están acotadas entre cero y uno, mientras que el modelo lineal puede llegar a predecir valores inconsistentes con una probabilidad.

En tercer lugar, podemos mencionar que el modelo de probabilidad lineal puede no ser útil

para evaluar efectos marginales en los valores extremos de la distribución. Esto se debe a que el efecto marginal es constante (las derivadas parciales son constantes). En cambio, el probit o logit logran captar mejor este efecto diferencial en las puntas. Finalmente, podemos mencionar una potencial desventaja del modelo probit y esta se basa en lo que realizamos en el punto anterior. Es necesario fijar cierto punto de interés para poder interpretar los coeficientes y no únicamente los signos de las estimaciones.

6. En este inciso lo que presentaremos es la relación entre el ingreso per cápita y la probabilidad de deserción escolar. La idea de este gráfico es que sea un análisis *ceteris paribus*: para un mismo individuo, vamos a ir moviendo solo su ingreso y viendo como cambia su probabilidad de desertar. En particular, nosotros elegimos que nuestro individuo tipo sea el siguiente: tiene 3 hermanos (es el valor de la mediana), la jefa de hogar es mujer y la persona fue a escuela pública. Este análisis lo haremos por separado: para hombres y para mujeres. Los resultados se pueden observar en la imagen derecha de la Figura (2). Lo que se puede ver a simple vista es que hay una brecha considerable entre ambos grupos en favor de los hombres. Para el individuo representativo que describimos anteriormente, la diferencia de probabilidad de deserción es de aproximadamente 8 puntos porcentuales. Si el ingreso es nulo, la probabilidad de desertar es de aproximadamente 37 % para hombres y 28 % para mujeres. Para ambos sexos, la probabilidad es decreciente en el ingreso. Esto sustenta lo afirmado en el primer punto de este trabajo práctico: parecería haber una relación negativa entre el ingreso per cápita y la probabilidad de deserción. Además, observando el gráfico parecería que la pendiente de la probabilidad es igual para ambos grupos.

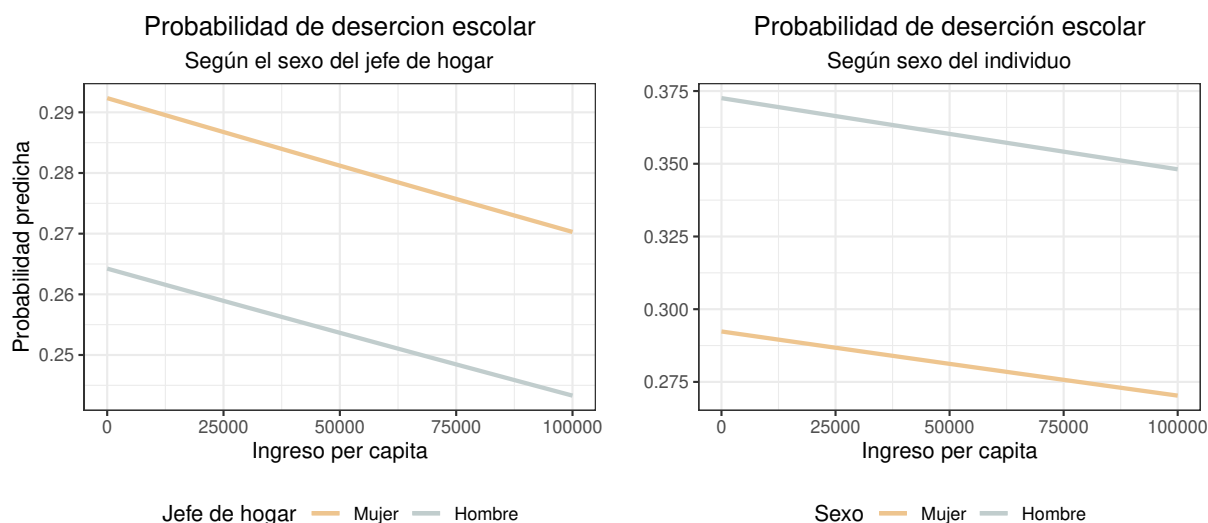
En la imagen de la izquierda de la Figura (2) podemos ver lo mismo: la probabilidad predicha de deserción escolar según el sexo del jefe de hogar. En este caso, fijamos que el individuo representativo sea mujer. Nuevamente es posible observar que la probabilidad es decreciente en el ingreso. Además, también podemos observar una brecha de acuerdo al sexo del jefe de hogar. A diferencia de lo expuesto previamente, ahora las personas cuyo jefe de hogar es mujer tienen una mayor probabilidad de desertar. Dado esto, podemos observar que la brecha es de aproximadamente 4 puntos porcentuales (mucho menor que en el caso anterior). La probabilidad de deserción predicha para una persona sin ingreso per cápita es de (aproximadamente) 29 % si el jefe de hogar es mujer y 26.5 %, si es hombre. En cambio, la probabilidad de deserción cae a 25 % para el caso de jefe de hogar mujer si el ingreso per cápita es de \$100.000, mientras que para el jefe de hogar hombre es de menos de 20 %. Esto nos da la pauta de que las pendientes no parecerían ser tan similares.

7. La consigna nos pide que estimemos el siguiente modelo:

$$p \equiv P(\text{Deserta} = 1|X) = F(\beta_0 + \beta_1 \ln \text{ing}_i + \beta_2 j_{\text{mujer}_i} + \theta Z_i + \mu_i)$$

en donde $F(\cdot)$ es la función de distribución acumulada de una normal estándar (en otras palabras, estimamos un modelo probit) y Z es un vector de variables que contiene el sexo, la

Figura 2: Probabilidad de deserción



educación del jefe de hogar y el tipo de escuela a la que asistió. Lo que tenemos que hacer es calcular cuál sería el subsidio que habría que entregarle a una jefa mujer para que su hija/o enfrente las mismas chances de deserción que uno correspondiente a una familia con jefe varón. Para esto, vamos a plantear los modelos para cada uno de los casos (jefe hombre y jefe mujer). La probabilidad de desertar dado que el jefe es hombre es $p^H = P(\text{Deserta} = 1 | X, j_{\text{mujer}} = 0) = F(\beta_1 \ln \text{ing}_i + \theta T_i)$ y si la jefa es mujer, $p^M = P(\text{Deserta} = 1 | X, j_{\text{mujer}} = 1) = F(\beta_1 \ln \text{ing}_i^* + \beta_2 + \theta T_i)$. Nótese que el ingreso para una persona cuya jefa de hogar es mujer lo notamos como $\ln \text{ing}^* = \ln[\text{ing}(1 + s)]$, es decir, incorporamos el subsidio.

Siguiendo la consigna, ahora lo que queremos es igualar las probabilidades, $p^H = p^M$. Para que esto ocurra, necesariamente tiene que suceder que: $F(\beta_1 \ln \text{ing} + \theta T) = F(\beta_1 \ln \text{ing}^* + \beta_2 + \theta T)$. Ahora, para que esto se iguale, los argumentos de la función $F(\cdot)$ debe ser iguales. Ahora lo que tenemos que hacer es despejar s de la siguiente ecuación:

$$\beta_1 \ln \text{ing} + \theta T = \beta_1 (\ln(\text{ing}(1 + s))) + \beta_2 + \theta T$$

Si despejamos, obtenemos que:

$$s = \exp\left(-\frac{\beta_2}{\beta_1}\right) - 1$$

Ahora utilizando los parámetros estimados ($\beta_1 = -0,008$ y $\beta_2 = 0,145$), podemos decir que el subsidio será 42170766.