

PROBABILIDAD DE APROBAR “MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA I” EN EL MARCO DE UN MODELO PROBIT

Ángel Solano García
M^a Isabel González Martínez
Universidad de Murcia

ABSTRACT

Es bien sabido por el futuro economista, que la asignatura *Métodos Cuantitativos para la Economía I*, constituye una prueba de fuego para culminar con éxito una carrera donde los contenidos matemáticos tienen un peso específico importante. Esta asignatura se imparte en el primer curso de la licenciatura, y tiene un elevado porcentaje de suspensos. En este trabajo se intentará modelizar la probabilidad de aprobar o suspender la asignatura: *Métodos Cuantitativos para la Economía I*, en función de ciertas características del alumnado: edad, sexo, nota de acceso a la Universidad, opción cursada en COU o equivalente (2º curso de bachillerato), y número de convocatoria. Para ello se aplicarán modelos de elección binaria, en concreto el modelo logit, a una muestra de alumnos matriculados en dicha asignatura, durante el curso 1999-2000. De este modo podremos analizar cuáles son los determinantes de la calificación obtenida.

Palabras clave: modelos de elección binaria, modelo probit.

1. INTRODUCCIÓN

Es bien sabido por el futuro economista, que la asignatura *Métodos Cuantitativos para la Economía I*, constituye una prueba de fuego para culminar con éxito una carrera donde los contenidos matemáticos tienen un peso específico importante. Esta asignatura se imparte en el primer curso de la licenciatura, y tiene un elevado porcentaje de suspensos. Así, la licenciatura de economía, al tratarse esta última de una ciencia social, es escogida como primera opción por estudiantes que han cursado una opción de bachillerato que no necesariamente contenían una asignatura de Matemáticas avanzadas. A priori, cabe pensar que son este tipo de alumnos los que más esfuerzo deben realizar para aprobar esta asignatura.

Por otro lado, muchos estudiantes veteranos y profesores coincidirán en que esta asignatura ha sido modificada en temario y en método de evaluación para la mayor accesibilidad del alumno a lo largo del tiempo. Sobre la base de que estas mejoras se dan tan continuamente (curso a curso) y a pesar de ellas el descontento siga, cabe preguntarse si las razones del número tan alto de suspensos no se deben a características propias del individuo (bagaje matemático).

En este trabajo se intentará modelizar la probabilidad de aprobar o suspender la asignatura: *Métodos Cuantitativos para la Economía I*, en función de ciertas características del alumnado: edad, sexo, nota de acceso a la Universidad, opción cursada en COU o equivalente (2º curso de bachillerato), y número de convocatoria. Para ello se aplicarán modelos de elección binaria, en concreto el modelo logit, a una muestra de 30 alumnos matriculados y presentados a la convocatoria de febrero de dicha asignatura en el curso 1999-2000. De este modo podremos analizar cuáles son los determinantes de la calificación obtenida

A continuación, se da unas breves nociones de los diferentes apartados en los que se divide el trabajo. En el primer epígrafe se da una idea del marco teórico de referencia del tipo de modelo que se pretende estimar. Así, al tratarse de un modelo en la que la variable dependiente es discreta (suspense, aprobado, notable, sobresaliente, matrícula) se optará por definir una variable binaria (que toma valores 0 ó 1 según las calificaciones) y se aplicará un modelo *logit* o *probit*. En el segundo epígrafe se estimarán ambos modelos (*logit*, *probit*), se seleccionará el mejor ajuste y se analizarán los coeficientes y efectos marginales más relevantes. Por último el trabajo terminará con algunas reflexiones personales sobre el estudio realizado.

2. UN MARCO TEÓRICO DE REFERENCIA: MODELOS LOGIT Y PROBIT

2.1. Modelos logit y probit

En el marco de un modelo de respuesta binaria como el que se estimará en la sección siguiente (en la que la variable dependiente solo toma como valores el 0 ó el 1), los valores ajustados de la variable dependiente son probabilidades

$$\hat{Y} = E(y_i/x_i) = \text{prob}(y=1/x)$$

Así, el método de estimación basado en un modelo de regresión lineal (ajuste lineal) parece no ser el adecuado por tres razones fundamentales:

1. La recta de regresión ajustada no está acotada por lo que pueden existir probabilidades mayores que 1 y probabilidades menores que cero.
2. En realidad, conforme aumenta cualquiera de las variables explicativas, la probabilidad de que $y=1$ o $y=0$ no aumenta o disminuye proporcionalmente.
3. Son modelos, por construcción heteroscedásticos.

De esta forma, nos planteamos un **modelo no lineal** cuyos valores ajustados de la variable dependiente (probabilidad) se encuentren en el intervalo (0,1).

$$Y_i = F(x_i' \cdot \beta) + \varepsilon_i$$

Por tanto, según si la función de distribución se asemeja más a una logística o a una normal tendremos un modelo logit o probit. Siendo las funciones de distribución de cada uno de los modelos:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp(-1/2 \cdot t^2) dt \quad \text{f.d. modelo probit.}$$

$$\Lambda(x) = \exp(x) / (1 + \exp(x)) \quad \text{f.d. modelo logit.}$$

Estas funciones de distribución, como antes se ha dicho, distinguen a estos dos tipos de modelos. El mejor ajuste será aquel cuya función de distribución se asemeje más a una de estas funciones de distribución cuya representación podría ser la que se representa en la figura 1:

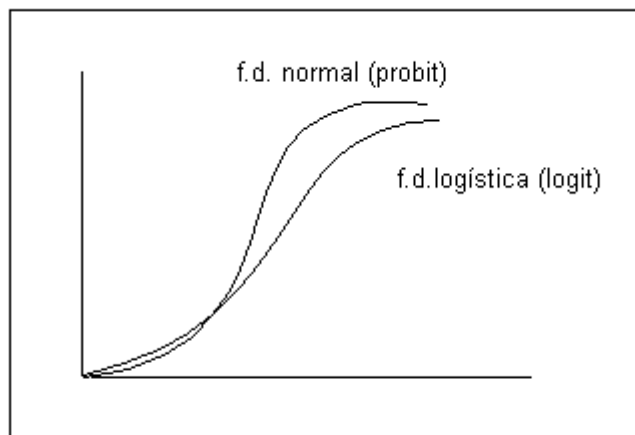


Figura 1

Estos modelos no se pueden estimar por MCO, por lo que se utiliza como método de estimación el de Máxima Verosimilitud (MV). Este método, como indica su nombre, se basa en la maximización de la función de verosimilitud de la muestra (L). Un problema que surge en la estimación es la inexistencia de una expresión que nos permita extraer el vector de coeficientes (β) directamente. Habrá que utilizar, entonces, métodos de optimización numérica.

A pesar de esta inconveniencia, que es subsanada gracias al desarrollo de las técnicas informatizadas adecuadas para la resolución de estos problemas, la estimación Máximo Verosímil cumple muy buenas propiedades asintóticas (consistencia, eficiencia asintótica, normalidad asintótica).

2.2. Efectos marginales

Se entiende por efecto marginal de una variable explicativa de un modelo cualquiera la variación experimentada por la variable explicada cuando la variable explicativa se incrementa en una unidad, es decir: la derivada de y_i con respecto de x_j .

En un modelo lineal cualquiera el efecto marginal de la variable coincide con el coeficiente asociado a la variable en cuestión. Esto no sucede en los modelos no lineales (logit y probit) ya que dicha derivada, la cual representa el incremento en la probabilidad que se da cuando se incrementa en una unidad la variable explicativa a la que vaya referida tal efecto marginal, no solo depende del coeficiente asociado a la variable explicativa sino que también depende de la función de densidad del valor ajustado de la explicada.

$$\partial y_i / \partial x_j = f(x_i; \beta) \beta_j \quad \text{Efecto marginal de } x_j \text{ en un modelo no lineal}$$

2.3. Contrastes y bondad de ajuste

En este tipo de modelos que se estiman por máxima verosimilitud, el estadístico idóneo para realizar contrastes de hipótesis es el r atio de verosimilitud, esto es:

$$\text{Estadístico: } -2 (\ln L(\beta_R) - \ln L(\beta_{MV})) \sim \chi_q^2$$

En cuanto a la bondad de ajuste, se medir a por el **R² de McFaden** el cual se calcula de la siguiente forma:

$$\text{R}^2 \text{ de McFaden} = 1 - \ln L(\beta_{MV}) / \ln L(\beta_R)$$

3. RESULTADOS EMP IRICOS

3.1. Estimaci n del modelo

Dada una muestra de 30 alumnos de la licenciatura de Econom a los cuales se presentaron a la asignatura *M todos Cuantitativos para la econom a I* en la convocatoria de febrero de 2000 se ha llevado a cabo la estimaci n de un modelo *Probit*. La variable dependiente es una variable binaria que toma valor 1 si el alumno aprob  la asignatura, y toma valor 0 si el alumno no aprob  dicha asignatura. Las variables explicativas incluidas en el modelo de partida han sido: *edad*, *sexo*, *opci n* que curs  en COU o equivalente y *nota de entrada a la universidad (nota "apto")*. Dicha estimaci n nos informa sobre las variables que influyen en la probabilidad de aprobar dicha asignatura. Los resultados de la estimaci n se muestran en el cuadro 1.

Cuadro1

Variables	Coeficientes	Desviaci�n t�pica
<i>Constante</i>	-45.52(*)	22.06
<i>Edad</i>	1.65(*)	0.83
<i>Sexo</i>	-0.46	0.91
<i>Opci�n</i>	0.46	0.29
<i>Nota</i>	1.69(*)	0.77

Nota: (*) significa que las variables son significativas al 5%.

Como se puede ver en el cuadro 1, las variables *sexo* y *opci n* que cursaron en COU no son relevantes por lo que se pueden eliminar del modelo. Puede tener especial inter s la no

relevancia de la *opción* cursada en COU ya que esto significa que el bagaje matemático del alumno no es tan importante de cara a las posibilidades de éste de aprobar la asignatura. Una vez eliminadas estas variables nos quedaremos con lo que va a ser nuestro modelo ajustado (donde realizaremos el análisis e interpretación de los coeficientes).

Cuadro 2

VARIABLES	COEFICIENTES	DESVIACIÓN TÍPICA
Constante	-34.03(*)	13.88
Edad	1.26(*)	0.55
Nota	1.32(**)	0.49

Nota: (*) significa que las variables son significativas al 5%.
 (**) significa que las variables son significativas al 1%

La variable más relevante es la *nota “apto”* de entrada a la universidad cuyo coeficiente tiene un valor positivo como cabría esperar. En la medida que dicha nota se forma mediante una media ponderada de las notas obtenidas en BUP, COU y Selectividad, este resultado nos puede indicar que el hábito de estudio, así como su rendimiento y en definitiva la capacidad de sacar una buena calificación independientemente del contenido de la asignatura afecta de un modo muy positivo a la probabilidad de aprobar *Métodos Cuantitativos para la Economía I*.

En cuanto a la *edad*, también aparece como una variable muy relevante. El signo positivo del coeficiente asociado a dicha variable refleja como a mayor edad, en general, existe más probabilidad de aprobar esta asignatura. Esto puede resultar lógico debido a la posibilidad de que alumnos de mayor edad pudieran agotar un número de convocatorias superior al de alumnos de menor edad.

3.2. Análisis de los efectos marginales

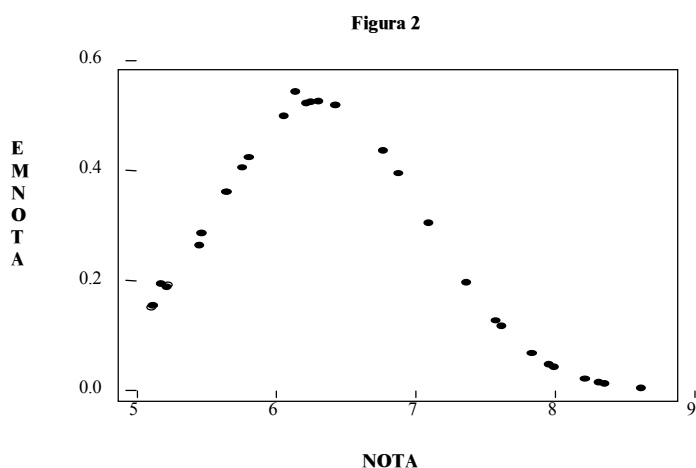
El análisis de los efectos marginales del modelo lo centraremos en analizar los efectos marginales de las variables *edad* y *nota “apto”* del individuo medio. Debido a la gran variabilidad de los valores referentes a los efectos marginales de cada individuo se ha optado por calcular el efecto marginal del individuo medio a través del cálculo de la media de los efectos marginales (*medem_edad* y *medem_nota*). Así se obtiene:

$$Medem_edad = 0.239336$$

$$Medem_nota = 0.249597$$

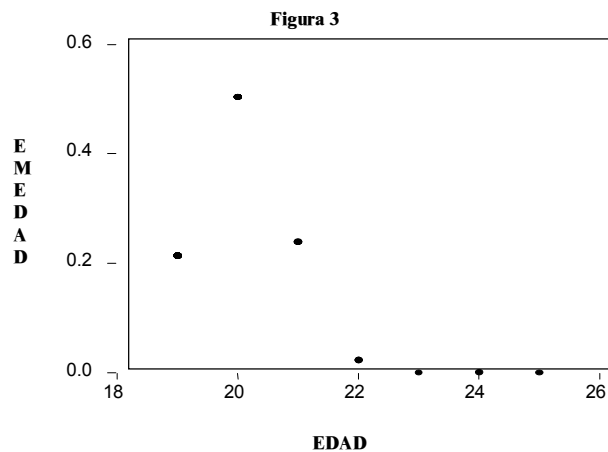
Estos valores nos indican que, para el individuo medio, la probabilidad de aprobar *Métodos Cuantitativos para la Economía I* se incrementa en 0.239336 y en 0.249597, cuando la *edad* y la *nota "apto"* se incrementan en una unidad respectivamente.

Veamos ahora la variación de dichos efectos marginales cuando varía cada una de las variables.



En la figura 2 se puede ver como varía el efecto marginal del individuo medio sobre la variable *nota "apto"* cuando varía dicha variable. Este gráfico nos aporta información sobre el nivel de la variable donde existe una mayor probabilidad de aprobar ante incrementos de dicha variable. Así, se observa como en el intervalo de nota que va aproximadamente de 6 a 6.5 es donde este efecto marginal alcanza su máximo para el individuo medio. Este valor que maximiza el efecto marginal puede resultar lógico ya que a partir de valores altos (7), conforme aumenta la nota, las posibilidades de aprobar ante nuevos incrementos de nota irán disminuyendo.

La variación del efecto marginal de la edad para el individuo medio se representa en la figura 3.



De igual modo que en el caso anterior, el efecto marginal de la variable edad obtiene su máximo en un valor intermedio de dicha variable (20 años). Este resultado nos puede llevar a plantearnos en qué medida es importante o relevante la experiencia de al menos un año a la hora de aprobar esta asignatura.

4. CONCLUSIONES

El objetivo que nos planteamos al iniciar este trabajo fue intentar determinar cuáles son los factores realmente influyentes en la probabilidad de aprobar la asignatura *Métodos Cuantitativos para la Economía I*. Esta asignatura constituye una dura prueba para muchos estudiantes de la Licenciatura en Economía.

Las conclusiones que podemos sacar de este análisis son las siguientes. Primera, las variables relevantes para explicar la probabilidad de aprobar esta asignatura son la *edad* y la *nota "apto"*. Ambas variables afectan de modo positivo, siendo la segunda la que tiene el efecto más importante. Segunda, la hipótesis inicial sobre la relevancia de la opción que cursó el alumno en COU, y por tanto sobre la importancia de la preparación matemática anterior a su matriculación ha sido rechazada. Estos resultados conducen a pensar que son la eficacia y capacidad en el estudio –reflejo de la nota de acceso –, y la constancia y la experiencia –reflejo de la edad –, los que condicionan los resultados obtenidos en esta asignatura.

Para finalizar diremos que dicho estudio puede servir como mensaje de ánimo para aquellos estudiantes que en COU cursaron una opción en la que no se impartieron muchas matemáticas.

5. REFERENCIAS BIBLIOGRÁFICAS

- AMEMIYA, T. (1981): "Qualitative Response Models: A Survey". *Journal of Economic Literature*, 19, 4, pp. 481-536.
- AMEMIYA, T. (1985): "Advanced Econometrics". Cambridge: Harvard University Press.
- CECCHETTI, S. (1986): "The Frequency of Price Adjustment: A Study of the Newsstand Price of Magazines". *Journal of Econometrics*, 31, pp. 255-274.
- CRAMER, J.S. (1991): "The Logit model: an introduction for economist". Edward Arnold.
- DAVIDSON, R. y MACKINNON, J.G. (1993): "Estimation and Inference in Econometrics". Oxford University Press.
- GREENE, W.H. (1999): "Análisis Económico", tercera edición. Prentice Hall Iberia, Madrid.
- HEIN, D.(1987): "Predicted and Actual Frecuencies in Binominal Response Models". *Economics Letters*, 23, pp.104-107.
- MCFADDEN, D.(1984): "Econometric Analysis of Qualitative Response Models", Z. Griliches y M. Intriligator, eds., *Handbook of Econometrics*, 2, Amsterdam: North Holland.
- MADDALA, G.S. (1996): "Introducción a la Econometría", segunda edición, Prentice Hall Hispanoamericana.
- NOVALES, A. (1993): "Econometría", segunda edición, McGraw-Hill.
- SPECTOR, L. y MAZZEO, M.(1980): "Probit Analysis and Economic Education". *Journal of Economic Education*, 11, pp. 312-320.