

CAPÍTULO 4-3

MODELOS CON VARIABLE DEPENDIENTE CUALITATIVA

4-3.1 MODELOS DE ELECCIÓN BINARIA: LOGIT BINOMIAL Y PROBIT BINOMIAL

4-3.1.1 El problema

Nos proponemos, en este momento, examinar la otra cara, por así decirlo, del análisis de varianza. En el análisis de varianza la variable de reacción es continua, mientras que las variables estímulos son discretas; por lo contrario, ahora, la variable de reacción es discreta, mientras que las variables de estímulo pueden ser continuas.

Es posible deducir el modelo logit a partir del análisis de las tablas de contingencia. Sin embargo, en el análisis de las tablas de contingencia todas las variables de clasificación tienen papeles simétricos cuando, por lo contrario, en el modelo logit una de las variables toma el papel de variable dependiente y las probabilidades de pertenencia a sus diferentes categorías son condicionadas por las demás variables, las cuales, por consiguiente, toman el papel de variables independientes. Además, la posibilidad de encontrar variables continuas entre las variables independientes constituye una

generalización con respecto al logit deducido del análisis de las tablas de contingencia.

Para empezar, consideremos el caso cuando la variable dependiente es dicotómica en lugar de ser politómica; en estas condiciones, sólo existen dos posibilidades:

$$y_i = 0 \text{ o } 1$$

Ejemplo: en un estudio sobre la movilidad residencial de los hogares,²⁰¹

- $y_i = 0$ si el hogar i no se muda
- $y_i = 1$ si el hogar i se muda

Ahora bien, en el modelo estándar

$$y_i = \sum_j x_{ij} \beta_j + u_i$$

el valor de la variable dependiente y_i no puede limitarse a tomar los valores 0 y 1 cuando los términos aleatorios u_i poseen una distribución normal (esto porque una variable normal es continua y su dominio de variación se extiende de $-\infty$ a $+\infty$).

El primer paso que debemos emprender para solucionar esta dificultad es considerar que la verdadera variable dependiente no es la variable dicotómica y , más bien la probabilidad que $y = 1$:

$$\Pr[y_i = 1] = \sum_j x_{ij} \beta_j + u_i$$

Esto conlleva, sin embargo, dos dificultades:

1. Aunque la nueva variable dependiente $\Pr[y_i = 1]$ sea continua, su campo de variación se limita al intervalo $[0, 1]$, cuando el campo de variación del término aleatorio, en caso de poseer una distribución normal, se extiende de $-\infty$ a $+\infty$.

²⁰¹ Para un ejemplo, vea Mongeau (sin fecha).

2. La nueva variable dependiente $\text{Pr}[y_i = 1]$, no se puede observar, pues lo que observamos no son más que las *realizaciones* ($y_i = 1$ o $y_i = 0$); por lo tanto será necesario recurrir a un método de estimación que sea adecuado a la situación.

4-3.1.2 Modelo de comportamiento

Uno de los fundamentos teóricos posibles del modelo logit binomial es un modelo de comportamiento del tipo estímulo-reacción (stimulus-response), el cual se usa con frecuencia particularmente en biología cuando se somete a los sujetos de un experimento a condiciones (estímulos) que varían de un sujeto a otro. Entonces se observan las reacciones y se busca estimar la relación entre estímulos y reacción.

Es posible presentar este modelo en dos partes. La primera parte constituye el modelo de reacción del sujeto (modelo de comportamiento) y la segunda, el modelo de estímulo total, el cual es el resultado de las diferentes condiciones combinadas a las cuales se somete a un sujeto.

Primera parte: modelo de reacción

No es posible observar directamente el estímulo total al cual se somete al sujeto. Supongamos, no obstante, que exista una variable no observable (“latente”) w que mide este estímulo, es decir el atractivo o “lo deseable” de una elección (un economista diría la “utilidad”). En el ejemplo del estudio sobre la movilidad residencial, el hogar i muda si esta variable latente w rebasa un cierto valor crítico, un “umbral de reacción” (S):

$$y_i = 0 \text{ si } w_i < S$$

$$y_i = 1 \text{ si } w_i \geq S$$

donde w es el valor de la variable latente para el hogar i y S es el umbral crítico más allá del cual el hogar se muda. Es cómodo suponer que se define la variable latente w para que S sea igual a cero, lo que nos da:

$$y_i = 0 \text{ si } w_i < 0$$

$$y_i = 1 \text{ si } w_i \geq 0$$

Esta última hipótesis no impone ninguna restricción al modelo, puesto que w es una variable ordinal cuyo cero es arbitrario.

Segunda parte: modelo de estímulo total

Se determina el valor de la variable latente (no observada) con un cierto número de factores medidos con variables independientes apropiadas (los x). Por ejemplo, en el estudio sobre la movilidad residencial, las variables independientes podrían tomar en cuenta las características del hogar, las características de la vivienda actual, etcétera.

Si, además, suponemos que la relación entre las variables independientes y la variable latente es lineal, obtenemos el modelo siguiente:

$$w_i = \sum_j x_{ij} \beta_j + u_i$$

donde el término aleatorio u_i representa las variaciones aleatorias entre los sujetos (los hogares). El modelo no emite la hipótesis que todos los sujetos son idénticos sino, más bien que, solamente, el modelo es lo suficiente completo para que podamos considerar las variaciones de comportamientos no explicadas por el modelo como el resultado del azar.

Integración de las dos partes del modelo

Se combina las dos partes del modelo.

$$w_i < 0 \text{ equivale a } \sum_j x_{ij} \beta_j + u_i < 0,$$

$$\text{o sea a } u_i < -\sum_j x_{ij} \beta_j$$

$$w_i \geq 0 \text{ equivale a } \sum_j x_{ij} \beta_j + u_i \geq 0,$$

$$\text{o sea a } u_i \geq -\sum_j x_{ij} \beta_j$$

Por lo tanto,

$$y_i = 0 \text{ si } u_i < -\sum_j x_{ij} \beta_j \text{ y } y_i = 1 \text{ si } u_i \geq -\sum_j x_{ij} \beta_j$$

De esta manera, la probabilidad que $y_i = 0$ es igual a la probabilidad de que $u_i < -\sum_j x_{ij} \beta_j$ y la probabilidad que

$y_i = 1$ es igual a la probabilidad que $u_i \geq -\sum_j x_{ij} \beta_j$. Estas

probabilidades dependen evidentemente de las hipótesis que hacemos en cuanto a la distribución de los u_i . Dependiendo de la hipótesis que hagamos sobre la función de densidad de probabilidad de los u_i , obtendremos un modelo logit o probit:

Hipótesis A: los u_i poseen distribuciones normales, independientes entre sí, con un promedio nulo y una varianza común σ^2 (hipótesis de Gauss-Markov); éste es el modelo probit.

Hipótesis B: los u_i se distribuyen independientemente uno del otro, y sus funciones de distribución acumulativas se definen con la función logística; éste es el modelo logit.

4-3.1.3 El modelo logit y la inducción estadística

Con un modelo que contiene una variable latente (inobservable), como el anterior descrito, es evidentemente imposible estimar los parámetros β_j por medio del método de los menores cuadrados. Es necesario recurrir al método del máximo de verosimilitud. La primera etapa en la aplicación de este método es construir la función de verosimilitud que, si recordamos bien, es la función de probabilidad de la muestra que expresamos en función de los parámetros.²⁰² La función de verosimilitud se infiere, por consiguiente, de la hipótesis que se hizo en cuanto a la distribución de los u_i , lo que constituye una especificación completa del modelo aleatorio.

Nos interesa, en este momento y de manera específica, el modelo logit, el cual se infiere de la hipótesis de una distribución logística. La función logística es la función

$$L(t) = \frac{1}{1 + e^{-t}} = \frac{e^t}{e^t + 1}$$

Si la variable aleatoria u posee una función de distribución acumulativa logística, entonces tenemos:

$$\Pr[u \leq t] = L(t) = \frac{1}{1 + e^{-t}} = \frac{e^t}{e^t + 1}$$

Se puede justificar la elección de la función logística con argumentos pragmáticos del siguiente tipo: es relativamente fácil de manejo y es muy semeja a la normal para valores alrededor del promedio (cuando se desemeja para los valores

²⁰² Al momento de presentar el principio del máximo de verosimilitud (cap. 2-2), definimos la función de verosimilitud como la función de densidad de probabilidad de la muestra. Sin embargo, en nuestro caso presente, la variable observada es discreta (dicotómica con más precisión), por lo tanto, la función de verosimilitud es, sin más, la función de probabilidad.

extremos). No obstante, la elección de la función logística puede basarse en fundamentos teóricos mucho más sólidos.²⁰³

Los tests de hipótesis que son posibles efectuar se fundamentan en unas distribuciones asintóticas; son, por lo tanto, válidos solamente cuando la muestra es lo suficientemente grande. Dicho de otra manera, más grande serán la muestra, más confiables será estos tests.

Los estimadores b_j de los parámetros β_j poseen distribuciones asintóticas normales y los estimadores de sus diferencias tipo poseen una distribución asintótica χ^2 . Habiendo cumplido con estas condiciones, es posible, ahora, aplicar los tests t de Student a los parámetros.

4-3.2 HACIA EL LOGIT MULTINOMIAL: UNA GENERALIZACIÓN HEURÍSTICA DEL BINOMIAL

Parece claro que el modelo estímulo-reacción, con umbral de reacción sólo se puede aplicar a una situación de elección binaria. En caso que haya más de dos posibilidades, es necesario un modelo más general donde el sujeto escoge la posibilidad más atrayente para él.

Ejemplo: elección del idioma de uso:

- $y_i = 0$ si el idioma de uso es el francés.
- $y_i = 1$ si el idioma de uso es el inglés.
- $y_i = 2$ si el idioma de uso es el italiano.
- $y_i = 3$ si el idioma de uso es otro idioma.

El modelo teórico de comportamiento subyacente al modelo logit multinomial es el modelo de utilidad aleatoria (random utility). Según este modelo, existe una variable latente que mide la “utilidad” o el atractivo de cada opción en función de los atributos del sujeto y de las características de la opción. El sujeto escoge racionalmente la posibilidad que

²⁰³ Veá, en particular, Ben Akiva y Lerman (1985).

posee el más grande atractivo.²⁰⁴ No obstante, la utilidad de cada opción no es una función determinista de las variables independientes, más bien, la suma de dos términos la constituye, a saber la utilidad sistemática y un término aleatorio.

En una situación de elección politómica, al lugar de tener una sola variable no observada (“latente”) w , tantas variables latentes son necesarias como hay posibilidades para que cada una de estas variables mida el atractivo de la posibilidad correspondiente. Escribamos:

w_{ij} : medición del atractivo de la posibilidad j para el individuo i .

Como primer paso, reformulemos el modelo binario según este esquema. Tenemos como resultado:

$$y_i = 0 \text{ si } w_{i0} > w_{i1}$$

$$y_i = 1 \text{ si } w_{i1} > w_{i0}$$

lo que equivale a

$$y_i = 0 \text{ si } w_{i0} - w_{i1} > 0, \text{ o sea } w_{i1} - w_{i0} < 0$$

$$y_i = 1 \text{ si } w_{i1} - w_{i0} > 0$$

Por lo tanto, es posible interpretar la medición del estímulo w del modelo binario como una medición de la diferencia entre el atractivo de la primera y de la segunda posibilidad. En estas condiciones, el modelo del estímulo se convierte en

$$w_i = w_{i1} - w_{i0} = z_{i1} - z_{i0} + u_i$$

donde $z_{ij} = \sum_h x_{ih} \beta_{hj}$ es la parte determinista del modelo del atractivo de la posibilidad k .

²⁰⁴ Es justamente en este punto que el modelo de utilidad aleatoria se distingue de los modelos con utilidad “constante” donde la probabilidad que una opción sea seleccionada aumenta con su utilidad, pero donde no es seguro que un sujeto escoja la posibilidad más ventajosa para él. En los modelos con utilidad “constante”, la elección de la posibilidad más ventajosa es solamente más probable que las otras. Es en este sentido que nos alejamos de un comportamiento racional.

Teniendo:

$p_{ij} = \Pr[y_i = j]$: la probabilidad que el individuo i escoja la posibilidad j .

Tenemos:

$$p_{i1} = \frac{e^{z_{i1}-z_{i0}}}{e^{z_{i1}-z_{i0}} + 1}$$

$$p_{i0} = 1 - p_{i1} = \frac{1}{e^{z_{i1}-z_{i0}} + 1}$$

lo que implica para los puntos de ventaja (como en apuestas; *odds* en inglés) de $[y_i = 1]$ contra $[y_i = 0]$

$$\frac{p_{i1}}{1 - p_{i1}} = \frac{p_{i1}}{p_{i0}} = e^{z_{i1}-z_{i0}} = \frac{e^{z_{i1}}}{e^{z_{i0}}}$$

Supongamos, sin demostrarlo, que se pueda generalizar este resultado con un número cualquiera de posibilidades y que, para todo par de posibilidad j, k , tengamos

$$\frac{p_{ij}}{p_{ik}} = e^{z_{ij}-z_{ik}} = \frac{e^{z_{ij}}}{e^{z_{ik}}}$$

Al efectuar la suma de todas las posibilidades j , obtenemos

$$\sum_j \frac{p_{ij}}{p_{ik}} = \frac{\sum_j e^{z_{ij}}}{e^{z_{ik}}}$$

donde $\sum_j p_{ij} = 1$, de tal manera que

$$\sum_j \frac{p_{ij}}{p_{ik}} = \frac{\sum_j p_{ij}}{p_{ik}} = \frac{1}{p_{ik}} = \frac{\sum_j e^{z_{ij}}}{e^{z_{ik}}}$$

de lo que se deduce

$$p_{ik} = \frac{e^{z_{ik}}}{\sum_j e^{z_{ij}}}$$

¡Y aquí tiene usted el modelo logit multinomial!

Los parámetros de este modelo se estiman por medio del método del máximo de verosimilitud, como en el caso del logit binomial. Los procedimientos de inducción estadística son análogos.