



Maestría en Analítica e Inteligencia de Negocios

Tema: Logit

PhD. Orlando Joaqui Barandica

Universidad del Valle

2022

Contenido

- 1 Modelos variable dependiente cualitativa
- 2 MLP
- 3 Logit
- 4 Probit
- 5 ¿Probit o Logit?
- 6 Predicciones correctas
- 7 HL

Contenido

1 Modelos variable dependiente cualitativa

2 MLP

3 Logit

4 Probit

5 ¿Probit o Logit?

6 Predicciones correctas

7 HL

Modelos variable dependiente cualitativa

Se consideran modelos de regresión en los que la variable dependiente puede ser una variable de tipo cualitativo: BINARIA, ORDINAL, NOMINAL o puede tratarse de una variable de CONTEO.

- **Variable binaria:** Tiene dos categorías. Normalmente indica que ha ocurrido un suceso, que alguna característica está presente o que se elige una opción.
- Ejemplos: Actividad laboral; Compra de un producto; Participación en las elecciones ...

Modelos variable dependiente cualitativa

- **Variable ordinal:** Tienen categorías que pueden ordenarse en forma ascendente o descendente.
- Ejemplo: Nivel de aceptación respecto a una afirmación, con respuestas: “completamente de acuerdo”, “de acuerdo”, “en desacuerdo”, “completamente en desacuerdo”

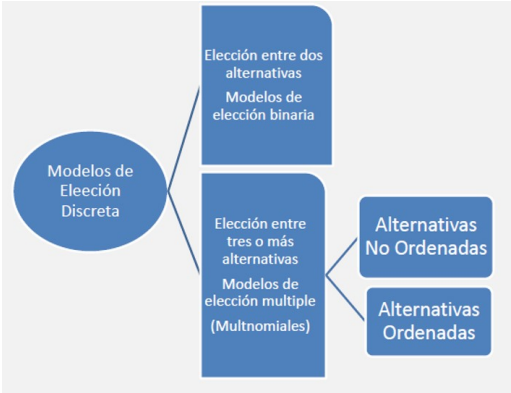
Valoración del nivel de satisfacción de un bien o servicio (“muy satisfecho”, ..., “muy insatisfecho”)

Frecuencia de una determinada acción (“nunca”, ..., “muy frecuentemente”); Nivel de educación alcanzado; ...

Modelos variable dependiente cualitativa

- **Variable nominal:** Variable con múltiples respuestas que no siguen un orden predeterminado.
- Ejemplos: Rama de actividad laboral; Estatus matrimonial; Elección política; Preferencia de marcas, ...

Modelos variable dependiente cualitativa



Modelos variable dependiente cualitativa

Los modelos de elección discreta resultan apropiados cuando el objetivo no es predecir el comportamiento medio de un agregado, sino analizar los factores determinantes de la **probabilidad** de que un agente económico individual elija un curso de acción dentro de un conjunto, generalmente finito, de opciones posibles.

- Y cuantitativa: $E(Y_i | X_{1i}, X_{2i}, \dots, X_{ki})$
- Y cualitativa: Probabilidad de que suceda el evento de interés.

Modelos de respuesta binaria

Hay cuatro métodos para crear un modelo de probabilidad para una variable de respuesta binaria:

- 1 El modelo lineal de probabilidad (MLP)
- 2 El modelo logit
- 3 El modelo probit
- 4 El modelo tobit

Contenido

1 Modelos variable dependiente cualitativa

2 MLP

3 Logit

4 Probit

5 ¿Probit o Logit?

6 Predicciones correctas

7 HL

Modelo lineal de probabilidad (MLP)

Considere el siguiente modelo simple:

$$Y = \beta_0 + \beta_1 X + u \quad (1)$$

donde X = el ingreso familiar, y $Y = 1$ si la familia tiene casa propia y 0 si la familia no tiene casa propia.

Parece un modelo de regresión lineal común, pero debido a que la variable regresada es binaria, o dicótoma, se denomina modelo lineal de probabilidad (MLP).

La expectativa condicional de que Y dado X , $E(Y/X)$ puede interpretarse como la probabilidad condicional de que el suceso tenga lugar dado X , es decir: $P(Y = 1|X)$

Modelo lineal de probabilidad (MLP)

En el MLP, la variable dependiente toma únicamente dos valores:

$$Y_i = \begin{cases} 1 & \text{si ocurre un evento determinado} \\ 0 & \text{si el evento no ocurre} \end{cases} \quad (2)$$

El valor esperado de esta variable, puede interpretarse como la probabilidad de que ocurra el evento:

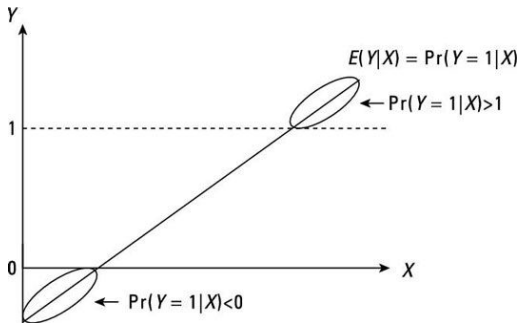
$$E(Y_i/X_i) = 1 * Pr(Y_i = 1|X_i) + 0 * Pr(Y_i = 0|X_i) \quad (3)$$

El valor esperado de Y_i dado X es la probabilidad de que $Y_i = 1$. Por tanto, el modelo de probabilidad lineal se puede escribir como: *(Asumiendo que $E(u_i) = 0$ para obtener estimadores insesgados)*

$$p_i = Pr(Y_i = 1|X_i) = \beta_0 + \beta_1 X \quad (4)$$

Problemas de MLP

Los problemas se presentan como consecuencia de que la “nube de puntos” a partir de la cual se debe ajustar la recta de regresión es en este caso un par de líneas paralelas sobre los dos únicos valores de la dependiente.



Problemas de MLP

No normalidad de las perturbaciones: Problema que no afecta la insesgadez de los estimadores puntuales, aunque el proceso de inferencia basado en una distribución normal sólo será válido si la muestra es lo suficientemente grande.

Heterocedasticidad del término de perturbación: Por tanto, los estimadores MCO son menos eficientes.



Problemas de MLP

Probabilidades predichas inconsistentes: no es posible garantizar que estén acotadas entre 0 y 1.

Interpretación de los coeficientes: El modelo supone que el efecto de las variables sobre la probabilidad es constante y lineal en todo el recorrido de las variables.

El Coeficiente de determinación: No es apropiado.

Problemas de MLP

Alternativa al modelo lineal de probabilidad

Interesa un modelo que reproduzca adecuadamente el comportamiento de una función de probabilidad.

La $Pr(Y_i = 1|X_i)$, deberá especificarse para que no supere los límites de 0 y 1, y con efectos no lineales de las variables explicativas:

$$\lim_{x_i'\beta \rightarrow -\infty} Pr[Y_i = 1|X] = 0 \quad \text{ó} \quad \lim_{x_i'\beta \rightarrow +\infty} Pr[Y_i = 1|X] = 1$$

Modelo de elección binaria

Para modelar este comportamiento, las distribuciones más empleadas han sido la normal estándar y la logística.

Cuando se emplea como función de probabilidad la distribución normal, se obtiene el denominado modelo **Probit**, mientras que la distribución logística proporciona el modelo **Logit**.

Modelos logit y probit	
Distribución	Modelo
Normal Estándar	Probit
$\Pr[y_i = 1 X] = \int_{-\infty}^{x_i'\beta} \phi(t)dt = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x_i'\beta)$	
Logística	Logit
$\Pr[y_i = 1 X] = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}} = \frac{1}{1 + e^{-x_i'\beta}} = \Lambda(x_i'\beta)$	



Contenido

- 1 Modelos variable dependiente cualitativa
- 2 MLP
- 3 Logit
- 4 Probit
- 5 ¿Probit o Logit?
- 6 Predicciones correctas
- 7 HL

Modelo Logit

Ejemplo:

$$P_i = \beta_0 + \beta_1 X + u \quad (5)$$

X es el ingreso y $Pr(Y_i = 1|X_i)$ significa que la familia es propietaria de una casa.

Pero considere ahora la siguiente representación de la propiedad de vivienda:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad (6)$$

$$P_i = \frac{1}{1 + e^{-Z}} \quad \text{Donde: } Z = \beta_0 + \beta_1 X \quad (7)$$

$$P_i = \frac{1}{1 + e^{-Z}} = \frac{e^Z}{1 + e^Z} \Rightarrow \text{Función de distribución logística} \quad (8)$$

Modelo Logit

P_i es no lineal no sólo en X sino también en las β_i . Esto significa que no podemos estimar los parámetros con el procedimiento habitual de MCO.

Es necesario linealizar a P_i

Si P_i es la probabilidad de tener casa propia, entonces $(1 - P_i)$, es la probabilidad de no tener casa propia.

$$1 - P_i = \frac{1}{1 + e^Z} \quad (9)$$

Por consiguiente, podemos escribir

$$\frac{P_i}{1 - P_i} = \frac{1 + e^Z}{1 + e^{-Z}} = e^Z \quad (10)$$

Modelo Logit

$$\frac{P_i}{1 - P_i} \quad (11)$$

Es la razón de las probabilidades en favor de tener una casa propia: la razón de la probabilidad de que una familia posea una casa propia respecto de la probabilidad de que no la posea.

Ahora si se aplica logaritmo natural sea L

$$L = \ln \left(\frac{P_i}{1 - P_i} \right) = \ln(e^Z) = Z \quad (12)$$

$$= \beta_0 + \beta_1 X \quad (13)$$

De ahí se tiene entonces el modelo logit.

Modelo Logit

Características del modelo logit:

- P va de 0 a 1.
- Aunque L es lineal en X , las probabilidades en sí mismas no lo son. Esta propiedad contrasta con el MLP.
- Se pueden añadir tantas regresoras como indique la teoría subyacente.
- Los modelos logit se estiman usualmente empleando el **método de máxima verosimilitud (MV)**.
- Como empleamos MV, que en general es para muestras grandes, los errores estándar estimados son asintóticos.
- Como resultado, en vez del estadístico t para evaluar la importancia estadística de un coeficiente, empleamos el estadístico (normal estandarizado) Z

Modelo Logit

Características del modelo logit:

- La medida convencional de la bondad de ajuste, R^2 , no es particularmente significativa para los modelos con regresada binaria. Se pueden utilizar similares:
 - pseudo R^2
 - R^2 McFadden
 - cuenta R^2

$$\text{Cuenta } R^2 = \frac{\text{Número de predicciones correctas}}{\text{Número total de observaciones}} \quad (14)$$

Como Y en el modelo logit toma el valor de 1 o de 0, si la probabilidad pronosticada es mayor que 0.5, se clasifica como si fuese 1, pero si es menor que dicho valor, se considera 0. Así, se cuenta el número de predicciones correctas y se calcula R^2

Ejemplo

Observación	Calificación GPA	Calificación TUCE	PSI	Calificación	Calificación con letra	Observación	Calificación GPA	Calificación TUCE	PSI	Calificación	Calificación con letra
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4.00	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29	0	1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4.00	23	1	1	A
15	3.53	26	0	0	B	31	3.10	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

Notas: Calificación Y = 1 si la calificación final fue A

Y = 0 si la calificación final fue B o C

TUCE = calificación en un examen presentado al comienzo del curso para evaluar los conocimientos previos de macroeconomía

PSI = 1 con el nuevo método de enseñanza

= 0 en otro caso

GPA = promedio de puntos de calificación inicial

Fuente: L. Spector y M. Mazzero, "Probit Analysis and Economic Education", *Journal of Economic Education*, vol. 11, 1980, pp. 37-44.

Ejemplo

Sea $Y = 1$, si la calificación final de un estudiante en un curso intermedio de microeconomía fue A, y $Y = 0$ si esa calificación final fue B o C. Spector y Mazzeo utilizaron el GPA (promedio de puntos de calificación), TUCE y PSI (Sistema de Enseñanza Personalizada) de Estados Unidos como predictores de la calificación. El modelo logit en este caso se expresa como:

$$\ln \left(\frac{P_i}{1 - P_i} \right) = \beta_0 + \beta_1 GPA + \beta_2 TUCE + \beta_3 PSI + u \quad (15)$$

Ejemplo

El comando en Stata es **logit**

Ejemplo

El comando en Stata es **logit**

logit Calificacion GPA TUCE PSI

Ejemplo

El comando en Stata es **logit**

logit Calificacion GPA TUCE PSI

Variable dependiente: Calificación

Método: Logit binario MV

Convergencia lograda después de 5 iteraciones

Variable	Coeficiente	Error estándar	Estadístico Z	Probabilidad
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255
R^2 McFadden = 0.3740 Estadístico RV (3 gl) = 15.40419				

Ejemplo

Variable dependiente: Calificación

Método: Logit binario MV

Convergencia lograda después de 5 iteraciones

Variable	Coefficiente	Error estándar	Estadístico Z	Probabilidad
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

 R^2 McFadden = 0.3740 Estadístico RV (3 gl) = 15.40419

Interpretaciones:

Cada coeficiente de pendiente es un coeficiente de pendiente parcial y mide el cambio en el logit estimado correspondiente a una unidad de cambio del valor de la regresada dada (con las demás regresoras constantes).

Ejemplo

Variable dependiente: Calificación

Método: Logit binario MV

Convergencia lograda después de 5 iteraciones

Variable	Coefficiente	Error estándar	Estadístico Z	Probabilidad
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

 R^2 McFadden = 0.3740

Estadístico RV (3 gl) = 15.40419

Interpretaciones: Ejemplo

El coeficiente del GPA igual a 2.8261 significa que, mientras las demás variables se mantengan constantes, si el GPA se incrementa en una unidad, en promedio el logit estimado aumenta casi 2.83 unidades, lo cual indica una relación positiva entre ambos.

Ejemplo

Variable dependiente: Calificación

Método: Logit binario MV

Convergencia lograda después de 5 iteraciones

Variable	Coefficiente	Error estándar	Estadístico Z	Probabilidad
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255

 R^2 McFadden = 0.3740

Estadístico RV (3 gl) = 15.40419

Interpretaciones: Ejemplo “Posibilidad”

Una interpretación más significativa se da en términos de las posibilidades en favor, las cuales se obtienen al tomar el antilogaritmo de los diversos coeficientes.

Ejemplo

Variable dependiente: Calificación
 Método: Logit binario MV
 Convergencia lograda después de 5 iteraciones

Variable	Coefficiente	Error estándar	Estadístico Z	Probabilidad
C	-13.0213	4.931	-2.6405	0.0082
GPA	2.8261	1.2629	2.2377	0.0252
TUCE	0.0951	0.1415	0.67223	0.5014
PSI	2.3786	1.0645	2.2345	0.0255
R ² McFadden = 0.3740 Estadístico RV (3 gl) = 15.40419				

Interpretaciones: Ejemplo “Posibilidad”

El antilogaritmo de PSI: 10.7897 ($\approx e^{2.3786}$).

Esto indica que los estudiantes expuestos al nuevo método de enseñanza son por encima de 10 veces más propensos a obtener una A que quienes no están expuestos al nuevo método, en tanto no cambien los demás factores.

Ejemplo

Interpretaciones: Ejemplo “Probabilidad”

los coeficientes de pendiente no dan la tasa de cambio de la probabilidad por cada unidad de cambio en la regresora. Es necesario calcularlos: **Efectos marginales: dy/dx**

Comando en Stata: **mfex**

Contenido

1 Modelos variable dependiente cualitativa

2 MLP

3 Logit

4 Probit

5 ¿Probit o Logit?

6 Predicciones correctas

7 HL

Modelo Probit

Para explicar el comportamiento de una variable dependiente dicótoma es preciso utilizar una función de distribución acumulativa (FDA) seleccionada apropiadamente. El modelo logit utiliza la función logística acumulativa, pero no es la única FDA posible.

El modelo de estimación que surge de una FDA **Normal** se conoce comúnmente como **modelo probit**.

Ejemplo

Variable dependiente: Calificación

Método: Probit binario MV

Convergencia lograda después de 5 iteraciones

Variable	Coficiente	Error estándar	Estadístico Z	Probabilidad
C	-7.4523	2.5424	-2.9311	0.0033
GPA	1.6258	0.6938	2.3430	0.0191
TUCE	0.0517	0.0838	0.6166	0.5374
PSI	1.4263	5950	2.3970	0.0165

Estadístico RV (3 gl) = 15.5458 R^2 McFadden = 0.3774

Probabilidad (estadístico RV) = 0.0014

El comando para estimar un modelo probit en Stata es **probit**

Contenido

1 Modelos variable dependiente cualitativa

2 MLP

3 Logit

4 Probit

5 ¿Probit o Logit?

6 Predicciones correctas

7 HL

¿Probit o Logit?

Aunque para el ejemplo de las calificaciones los modelos logit, probit y MLP dan cualitativamente resultados semejantes, nos centraremos en los modelos logit y probit, en vista de los problemas con el MLP ya mencionados.

De los modelos logit y probit, ¿cuál preferiría?

Para la mayoría de las aplicaciones, los modelos son muy semejantes; En la práctica, muchos investigadores eligen el modelo logit debido a su comparativa simplicidad matemática.



Contenido

- 1 Modelos variable dependiente cualitativa
- 2 MLP
- 3 Logit
- 4 Probit
- 5 ¿Probit o Logit?
- 6 Predicciones correctas
- 7 HL

Predicciones correctas

Una forma sencilla de valorar el ajuste de un modelo de elección binaria consiste en comparar las predicciones del modelo con las respuestas observadas en la muestra.

Para cada observación se predice la probabilidad y se asigna la respuesta de ese elemento a valores 0 o 1, según la probabilidad supere o no un determinado umbral. Normalmente, el criterio de asignación emplea como punto de corte una probabilidad igual a 0,5

$$P(\hat{Y}_i = 1|X) \geq 0,5 \longrightarrow \hat{Y}_i = 1 \quad (16)$$

$$P(\hat{Y}_i = 1|X) < 0,5 \longrightarrow \hat{Y}_i = 0 \quad (17)$$

Contenido

1 Modelos variable dependiente cualitativa

2 MLP

3 Logit

4 Probit

5 ¿Probit o Logit?

6 Predicciones correctas

7 HL

Hosmer-Lemeshow

La prueba de Hosmer-Lemeshow es otro método para estudiar la bondad de ajuste del modelo de regresión logística que consiste en comparar los valores previstos (esperados) por el modelo con los valores realmente observados. Ambas distribuciones, esperada y observada, se contrastan mediante una prueba de χ^2 .

La hipótesis nula del test de Hosmer-Lemeshow es que no hay diferencias entre los valores observados y los valores pronosticados (el rechazo este test indicaría que el modelo no está bien ajustado).



Preguntas?

Gracias!! ,

Jr.

orlando.joaqui@correounivalle.edu.co