Econometría para la Gestión

Variable dependiente limitada
Modelos logit y probit para respuesta binaria

Rodrigo Ortiz

Universidad Alberto Hurtado

Santiago, 2023

Planificación

Contenidos clase de hoy

1. Introducción
2. Modelos logit y probit para respuesta binaria
   - Especificación de modelos logit y probit
   - Modelo logit
   - Modelo probit
3. Estimación por máxima verosimilitud
4. Prueba de hipótesis múltiples
5. Matriz de confusión
   - Exactitud (Accuracy)
   - Sensitividad (Sensitivity)
   - Especificidad (Specificity)

Introducción

El modelo de probabilidad lineal, el cual simplemente es una aplicación del modelo de regresión múltiple a una variable dependiente binaria. Una variable dependiente binaria es un ejemplo de una variable dependiente limitada (VDL).
Una VDL se define en sentido amplio como una variable dependiente cuyo rango de valores está restringido de forma importante. Una variable binaria asume sólo dos valores, el cero y el uno.
Se han analizado otros varios ejemplos de variables dependientes limitadas:

- el porcentaje de participación en un plan de pensiones está entre 0 y 100
- el número de veces que un individuo es arrestado en un año determinado es un entero no negativo
- el promedio de calificaciones universitarias es de entre 1 y 7

Modelos logit y probit para respuesta binaria

Estimar y utilizar el modelo de probabilidad lineal es simple, pero tiene algunas desventajas.
Las dos desventajas más importantes son

- que las probabilidades ajustadas pueden ser menores que cero o mayores que uno
- el efecto parcial de cualquier variable explicativa (si aparece en la ecuación en su nivel) es constante

Estas limitaciones del MPL pueden superarse si se usan modelos de respuesta binaria más sofisticados.

Modelos logit y probit para respuesta binaria

En un modelo de respuesta binaria, el interés yace principalmente en la probabilidad de respuesta

$$P(y = 1|x) = P(y = 1|x_1, \ldots, x_k) \tag{1}$$

donde x denota el conjunto total de variables explicativas.

Por ejemplo, cuando y es un indicador del empleo, x podría contener varias características individuales como la educación, edad, estado civil y otros factores que afectan el estado del empleo, incluida una variable de indicador binario para la participación en un reciente programa de empleo.

Especificación de modelos logit y probit

En el MPL, se supone que la probabilidad de respuesta es lineal en un conjunto de parámetros, β_j.
Para evitar las limitaciones del MPL, considere una clase de modelos de respuesta binaria de la forma

$$P(y = 1|x) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots, \beta_k x_k) = G(\beta_0 + \boldsymbol{x}\boldsymbol{\beta}) \qquad (2)$$

donde G es una función que asume valores estrictamente entre cero y uno: $0 \leq G(z) \leq 1$, para todos los números reales z. Esto asegura que las probabilidades de respuesta estimada sean estrictamente entre cero y uno.
Como en los primeros capítulos, se escribe $\boldsymbol{x}\boldsymbol{\beta} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$.

Modelo logit

Se han sugerido varias funciones no lineales para la función G a fin de asegurar que las probabilidades estén entre cero y uno.

En el modelo logit, G es la función logística:

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} = \Lambda(z) \tag{3}$$

que está entre cero y uno para todos los números reales z. Esta es la función de distribución acumulada (fda) para una variable aleatoria logística estándar.

Modelo probit

En el modelo probit, G es la función de distribución acumulada normal estándar, que se expresa como una integral:

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v) dv \tag{4}$$

donde $\phi(v)$ es la densidad normal estándar

$$\phi(z) = (2\pi)^{-1/2} \exp^{-z^2/2} \tag{5}$$

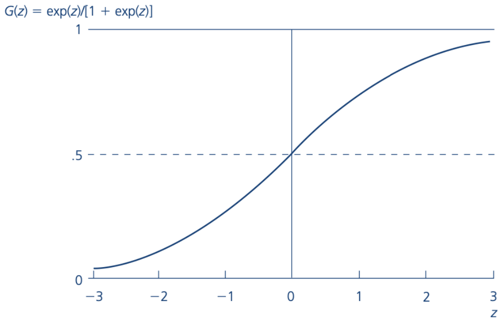
Esta elección de G nuevamente asegura que (2) esté estrictamente entre cero y uno para todos los valores de los parámetros y las x_j.

Las funciones G en (3) y (4) son funciones crecientes. Cada una aumenta con más rapidez en $z = 0$, $G(z) \longrightarrow 0$ a medida que $z \longrightarrow -\infty$, y $G(z) \longrightarrow 1$ a medida que $z \longrightarrow \infty$. La función logística está graficada en la figura 17.1. La fda normal estándar tiene una forma muy similar a la de la fda logística.

Modelos logit y probit

FIGURA 17.1

Gráfica de la función logística $G(z) = \exp(z)/[1 + \exp(z)]$.



Modelos logit y probit

Los modelos logit y probit pueden derivarse a partir de un modelo de variable latente subyacente. Sea y^* una variable inobservable, o latente, determinada por

$$y* = \beta_0 + \boldsymbol{x\beta} + \epsilon, y = 1[y* > 0] \tag{6}$$

donde se introduce la notación $1[\cdot]$ para definir un resultado binario. La función $1[\cdot]$ recibe el nombre de **función de indicador**, que asume el valor de uno si el evento dentro de los corchetes es verdadero y de cero si no lo es. Por tanto, y es uno si $y^* > 0$ y y es cero si $y^* \leq 0$.

Modelos logit y probit

Se supone que ϵ es independiente de x y que ϵ tiene la distribución logística estándar o la distribución normal estándar.

En cualquier caso, ϵ se distribuye simétricamente en torno a cero, lo cual significa que $1 - G(-z) = G(z)$ para todos los números reales z.

Los economistas tienden a favorecer el supuesto de normalidad para ϵ, lo cual es la razón por la que en econometría el modelo probit es más popular que el logit.

Además, varios problemas de especificación, que se tratarán después, se analizan fácilmente mediante probit debido a las propiedades de la distribución normal.

Modelos logit y probit

A partir de (6) y de los supuestos establecidos, se puede calcular la probabilidad de respuesta para y:

$$
\begin{aligned}
P(y = 1|\boldsymbol{x}) &= P(y^* > 0|\boldsymbol{x}) = P(\epsilon > -(\beta_0 + \boldsymbol{x\beta})|\boldsymbol{x}) \qquad (7)\\
&= 1 - G[-(\beta_0 + \boldsymbol{x\beta})] = G(\beta_0 + \boldsymbol{x\beta}) \qquad (8)
\end{aligned}
$$

lo cual es exactamente lo mismo que (2).

Modelos logit y probit

En la mayoría de las aplicaciones de los modelos de respuesta binaria, la meta principal es explicar los efectos de las x_j sobre la probabilidad de respuesta $P(y^* = 1|x)$. La formulación de la variable latente tiende a dar la impresión de que lo que principalmente interesa son los efectos de cada x_j sobre y^*.

Para logit y probit, la **dirección** del efecto de x_j sobre $E(y^*/x) = (\beta_0 + \boldsymbol{x\beta})$ y sobre $E(y/x) = P(y = 1|x) = G(\beta_0 + \boldsymbol{x\beta})$ es siempre la misma. Pero la variable latente y^* rara vez tiene una unidad de medición bien definida.

Modelos logit y probit

Por ejemplo, y^* puede ser la diferencia en niveles de utilidad de dos acciones diferentes.

Por tanto, las magnitudes de cada β_j no son, por sí mismas, especialmente útiles (en contraste con el modelo de probabilidad lineal). Para la mayoría de los propósitos, se quiere estimar el efecto de x_j sobre la probabilidad de éxito $P(y = 1|x)$, pero esto se complica por la naturaleza no lineal de $G(\cdot)$.

Estimación por máxima verosimilitud de los modelos logit y probit

Para estimar los modelos de variables dependientes limitadas, los métodos de máxima verosimilitud son indispensables.

Como la estimación de máxima verosimilitud está basada en la distribución de y dada x, la heterocedasticidad en $Var(y|x)$ automáticamente se toma en cuenta.

Estimación por máxima verosimilitud de los modelos logit y probit

Suponga que se tiene una muestra aleatoria de tamaño n. Para obtener el estimador de máxima verosimilitud, condicional sobre las variables explicativas, se necesita la densidad de y_i dada x_i. Esto se puede escribir como

$$f(y|\boldsymbol{x_i};\boldsymbol{\beta}) = [G(\boldsymbol{x_i},\boldsymbol{\beta})]^y[1-G(\boldsymbol{x_i},\boldsymbol{\beta})]^{1-y}, y = 0, 1 \qquad (9)$$

donde, por simplicidad, se absorbe el intercepto en el vector x_i.

Se puede ver con facilidad que cuando $y = 1$, se obtiene $G(\boldsymbol{x_i},\boldsymbol{\beta})$ y cuando $y = 0$, se obtiene $1 - G(\boldsymbol{x_i},\boldsymbol{\beta})$.

Estimación por máxima verosimilitud de los modelos logit y probit

La función de log-verosimilitud para la observación i es una función de los parámetros y los datos $(\boldsymbol{x_i}, y_i)$, y se obtiene al aplicar el log a la ecuación (9):

$$l(\boldsymbol{\beta}) = y_i log[G(\boldsymbol{x_i}, \boldsymbol{\beta})] + (1 - y_i) log[1 - G(\boldsymbol{x_i}, \boldsymbol{\beta})] \tag{10}$$

Debido a que $G(\cdot)$ está estrictamente entre cero y uno para logit y probit, $l(\boldsymbol{\beta})$ está bien definida para todos los valores de $\boldsymbol{\beta}$.

La log-verosimilitud para un tamaño de muestra de n se obtiene al sumar (10) a través de todas las observaciones: $L(\boldsymbol{\beta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta})$. La EMV de $\boldsymbol{\beta}$, denotada como $\hat{\boldsymbol{\beta}}$, maximiza esta log-verosimilitud.

Estimación por máxima verosimilitud de los modelos logit y probit

Si $G(\cdot)$ es la fda logit estándar, entonces $\hat{\beta}$ es el **estimador logit**; si $G(\cdot)$ es la fda normal estándar, entonces $\hat{\beta}$ es el **estimador probit**.

Debido a la naturaleza no lineal del problema de maximización, no se pueden escribir fórmulas para las estimaciones de máxima verosimilitud logit o probit.

Además de los problemas computacionales, esto hace que la teoría estadística para logit y probit sea mucho más difícil que MCO o incluso que MC2E.

Sin embargo, la teoría general de EMV para muestras aleatorias implica que, bajo condiciones muy generales, la EMV es consistente, asintóticamente normal y asintóticamente eficiente.

Prueba de hipótesis múltiples

Existen tres formas de probar las restricciones de exclusión para modelos logit y probit.

1) El multiplicador de Lagrange o el estadístico de puntuación sólo requieren estimar el modelo bajo la hipótesis nula, tal como en el caso lineal de la sección 5.2; aquí no se abordará el estadístico de puntuación, dado que rara vez se necesita probar las restricciones de exclusión.

2) La prueba de Wald requiere la estimación sólo del modelo no restringido. En el caso del modelo lineal, el estadístico de Wald, después de una simple transformación, es esencialmente el estadístico F, así que no hay necesidad de abordar el estadístico de Wald por separado.

3) Si tanto el modelo restringido como el no restringido son fáciles de estimar, como suele ser el caso con las restricciones de exclusión, entonces la prueba de la razón de verosimilitudes (RV) se vuelve muy atractiva. La prueba RV está basada en el mismo concepto que la prueba F en un modelo lineal. La prueba F mide el incremento en la suma de los residuales cuadrados cuando las variables se desechan del modelo. La prueba RV está basada en la diferencia en las funciones de log-verosimilitud para los modelos restringido y no restringido.

Matriz de confusión

Una matriz de confusión, también conocida como matriz de error, nos permite la visualización del rendimiento de un algoritmo, generalmente uno de aprendizaje supervisado (en el aprendizaje no supervisado generalmente se denomina matriz coincidente).

Cada fila de la matriz representa las instancias en una clase predicha, mientras que cada columna representa las instancias en una clase real (o viceversa). El nombre se deriva del hecho de que hace que sea fácil ver si el sistema confunde dos clases (es decir, comúnmente etiquetar incorrectamente una como otra).

Las clasificaciones binarias, cuando el outcome puede tomar solo dos clases, nos arrojan esta siguiente matriz de confusión.

Matriz de confusión

Figura: Matriz de confusión

	Realmente es positivo	Realmente es negativo
Predicho como positivo	Verdaderos Positivos (VP)	Falsos Positivos (FP)
Predicho como negativo	Falsos Negativos (FN)	Verdaderos Negativos (VN)

Exactitud (Accuracy)

La exactitud del modelo (accuracy en inglés), la podemos calcular a partir de la matriz de confusión:

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

El accuracy del modelo es la proporción de veces que el algoritmo predijo correctamente, respecto al total de datos evaluados.

Sensitividad (Sensitivity)

La sensibilidad (también llamada tasa positiva verdadera, el recuerdo o la probabilidad de detección en algunos campos) mide la proporción de positivos reales que se identifican correctamente como tales (por ejemplo, el porcentaje de personas enfermas que se identifican correctamente como que tienen condición).

$$Sensitivity = \frac{VP}{VP + FN}$$

Especificidad (Specificity)

La especificidad (también llamada tasa negativa verdadera) mide la proporción de negativos reales que se identifican correctamente como tales (por ejemplo, el porcentaje de personas sanas que se identifican correctamente como que no tienen la afección).

$$Sensitivity = \frac{VN}{VN + FP}$$