

Curso libre:

*Econometría básica con
Python*

Monitor encargado:

*Juan Felipe Acevedo
Pérez*

Correo: uniic_bog@unal.edu.co

Teléfono: 3165000 ext 12301

Variable dependiente discreta

Juan Felipe Acevedo Pérez
Monitor (a) Unidad de Informática

Variable dependiente discreta

Correo: uniic_bog@unal.edu.co

Teléfono: 3165000 ext 12301

¿Variables discretas?

Clasificación de variables:

- Continuas: Toman valores de un conjunto **no finito**.
- Discretas: Toman valores de un conjunto **finito**.

Las variables categóricas corresponden únicamente a un número finito de valores (alternativas o categorías).

¿Caso binario?

Una variable categórica contiene k categorías (alternativas).

Si $k = 2$ entonces se trata de una variable **binaria** (solo puede tomar dos valores).

Ejemplos:

- Sexo de una persona (masculino o femenino).
- Hombres casados y no casados.
- Estudiantes que han perdido asignaturas y estudiantes que no han perdido asignaturas.

Codificación de variables

Para codificar una variable binaria se emplean los valores 0 y 1.

Por **convención**, 1 indica que una condición específica se cumple y 0 que dicha condición no se cumple.

Ejemplos:

- Sexo de una persona. *femenino* $\begin{cases} 0 \text{ si hombre} \\ 1 \text{ si mujer} \end{cases}$
- Hombres casados. *casado* $\begin{cases} 0 \text{ si no casado} \\ 1 \text{ si casado} \end{cases}$
- Estudiantes y pérdida de asignaturas. $\begin{cases} 0 \text{ si no perdió} \\ 1 \text{ si perdió} \end{cases}$

Modelos con variable dependiente discreta

Lo que se estima no es el valor que toma la variable dependiente sino la probabilidad de que esta tome un valor específico.

$$Prob[y_i = 1|x_i] = P_i = F(x_i^T \beta)$$

en donde x_i^T es el vector transpuesto de variables explicativas para el individuo i (tamaño $1 \times k$) y β es el vector de parámetros (tamaño $k \times 1$).

Dependiendo de la forma que adopte $F(x_i^T \beta)$ se tiene un modelo distinto.

Modelo Lineal de Probabilidad

No es más que un modelo de regresión lineal en el que la variable dependiente es discreta binaria.

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

Desventajas:

- Término de error problemático.
- No hay garantía de obtener probabilidades “lógicas”

Modelo Logit

Una alternativa común al MLP y que sí garantiza $0 \leq P_i \leq 1 \forall i$ es el **Modelo Logit**:

$$P_i = F(z_i) = \frac{1}{1 + e^{-z_i}}$$

en donde $\mathbf{x}_i^T \boldsymbol{\beta} = z_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$

Nota: Los β miden el impacto de una variable sobre el **logit**. El logit es el *logaritmo de la razón de probabilidades*.

Modelo Probit

Otra alternativa común al MLP y que garantiza $0 \leq P_i \leq 1 \forall i$ es el **Modelo Probit**:

$$P_i = F(\mathbf{x}_i^T \boldsymbol{\beta}) = \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} \phi(z) dz$$

en donde $\Phi(\mathbf{x}^T \boldsymbol{\beta})$ corresponde a la Función de **Distribución Acumulada** de la Distribución Normal Estándar.

Interpretación

- El MLP no es más que un modelo de regresión **lineal** con estimación por **MCO**. El coeficiente asociado a cada variable mide su impacto sobre la probabilidad de que la variable dependiente tome el valor de 1.
- Los modelos Logit y Probit son modelos de regresión **no lineal** con estimación por **Máxima Verosimilitud**. El coeficiente asociado a cada variable **no** mide su impacto sobre la probabilidad de que la variable dependiente tome el valor de 1.
- Para cuantificar el impacto de una variable se utilizan los **efectos marginales**.

Efectos marginales

- Los efectos marginales cuantifican la relación entre la variación de una variable explicativa y la variación en la probabilidad de que la variable dependiente tome el valor de 1.
- El efecto marginal que tiene la variable x_j sobre la probabilidad para la observación i es:

$$\frac{\partial P_i}{\partial x_{ji}} = \frac{\partial F(x_i^T \beta)}{\partial x_{ji}}$$

Efectos marginales: MLP

En el MLP el coeficiente asociado a una variable mide el impacto de dicha variable sobre la probabilidad de que la variable dependiente tome el valor de 1. Por lo tanto, el efecto marginal de una variable corresponda a su coeficiente. Efectivamente:

$$F(\mathbf{x}_i^T \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}$$

$$\frac{\partial P_i}{\partial x_{ji}} = \beta_j$$

Nota: Sin importar i , el efecto marginal de x_j siempre será β_j .

Efectos marginales: Logit

En el modelo Logit, el coeficiente asociado a una variable no corresponde a su efecto marginal, como se demuestra a continuación:

$$F(z_i) = \frac{1}{1 + e^{-z_i}}$$

$$\frac{\partial P_i}{\partial x_{ji}} = \beta_j \frac{e^{z_i}}{(1 + e^{z_i})^2}$$

Nótese que el efecto marginal de una variable no depende únicamente de ella misma sino también de las demás y que tendrá un valor distinto para cada i .

Efectos marginales: Probit

En el modelo Probit, el coeficiente asociado a una variable no corresponde a su efecto marginal, como se demuestra a continuación:

$$F(\mathbf{x}_i^T \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} \phi(z) dz$$
$$\frac{\partial P_i}{\partial x_{ji}} = \phi(\mathbf{x}_i^T \boldsymbol{\beta}) \beta_j$$

Nótese que el efecto marginal de una variable no depende únicamente de ella misma sino también de las demás y que tendrá un valor distinto para cada i .

Efectos marginales: una medida general

- En el MLP el efecto marginal de una variable tiene un valor único, pues corresponde a su coeficiente y éste es igual para todas las observaciones.
- En los modelos Logit y Probit, existen n valores distintos para el efecto marginal de la misma variable. Esta situación no resulta práctica; se requiere de una medida “general” del efecto marginal de una variable en específico.
- ¿Cómo se obtiene una medida general del efecto marginal de una variable?
 - Efectos marginales en la media.
 - Efectos marginales promedio.

Efectos marginales en la media

Efecto marginal de la variable calculado con los valores correspondientes a la “observación promedio”.

Para **Logit**:

Para **Probit**:

Nota: * hace referencia al promedio de * para todo *.

(0)

Efectos marginales en la media

Efecto marginal de la variable calculado con los valores correspondientes a la “observación promedio”.

Para **Logit**:

$$\frac{\partial P}{\partial x_j} = \beta_j \frac{e^{\bar{z}}}{(1 + e^{\bar{z}})^2}$$

Para **Probit**:

$$\frac{\partial P}{\partial x_j} = \phi(\bar{\mathbf{x}}^T \boldsymbol{\beta}) \beta_j$$

Efectos marginales promedio

Efecto marginal de la variable corresponde al promedio de los efectos marginales individuales de dicha variable.

Para **Logit**:

$$\frac{\partial P}{\partial x_j} = \beta_j \frac{\sum_{i=1}^n \frac{e^{z_i}}{(1 + e^{z_i})^2}}{n}$$

Para **Probit**:

$$\frac{\partial P}{\partial x_j} = \beta_j \frac{\sum_{i=1}^n \phi(x_i^T \boldsymbol{\beta})}{n}$$

Referencias

- Wooldridge, J.M. (2010). Introducción a la econometría. Un enfoque moderno. México: Cengage Learning Editores, S.A. de C.V.