

Curso Libre:
Econometría en Python

Monitor encargado:
Juan Felipe Acevedo Pérez

Variable Dependiente Discreta

Juan Felipe Acevedo Pérez
Monitor Junior

Correo: uniic_bog@unal.edu.co

Tel: 3165000 **Ext:** 12301

1

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 no binario?

Una variable categórica contiene k categorías (alternativas).

Si $k > 2$ entonces se trata de una variable **no binaria** (puede tomar más de dos valores).

Ejemplos:

- Elección de la marca de cerveza que una persona va a consumir.
- Combo de comida que alguien elige cuando va a cine.

Codificación de variables

- Para codificar una variable no binaria se emplean m valores (un valor
- distinto para cada una de las m alternativas).

- **Ejemplo:**

Marcas de cerveza. *Marca* {
1 si Águila
2 si Póker
3 si C. Colombia
4 si Corona
5 si Heineken

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.

$$Prob[y_i = j | \mathbf{x}_i] = P_{ji} = F_j(\mathbf{x}_i, \boldsymbol{\gamma})$$

Dependiendo de la forma que adopte $F_j(\mathbf{x}_i, \boldsymbol{\gamma})$ se tiene un modelo distinto: Logit Multinomial, Probit Multinomial, Logit Condicional, etc.

Restricción: un individuo (entidad), de las m alternativas posibles, solo puede elegir una.

Datos empleados

La información empleada para trabajar con modelos con variable dependiente discreta **no binaria** puede encontrarse en dos formatos: ***wide*** y ***long***.

- En el formato ***wide***, la información para cada individuo (entidad) se presenta en una única fila.
- En el formato ***long***, la información para cada individuo (entidad) se presenta en m filas, cada una de las cuales corresponde a una de las alternativas de la variable dependiente, la cual toma el valor de 1 una única vez y de 0 $m - 1$ veces para cada individuo (¿Por qué?).

Ejemplo de formato *wide*

| Individuo | Marca elegida | y | Ingreso del individuo | \$Águila | \$Póker | \$C. Colombia | \$Corona | \$Heineken |
|-------------|---------------|---|-----------------------|-----------|-----------|---------------|-----------|------------|
| Individuo 1 | Póker | 2 | COP 2,300,000 | COP 2,600 | COP 3,100 | COP 3,300 | COP 4,200 | COP 4,100 |
| Individuo 2 | C. Colombia | 3 | COP 3,000,000 | COP 3,300 | COP 3,600 | COP 4,000 | COP 5,700 | COP 5,500 |
| Individuo 3 | Corona | 4 | COP 2,700,000 | COP 3,000 | COP 3,300 | COP 3,600 | COP 5,000 | COP 4,800 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| Individuo n | Águila | 1 | COP 2,500,000 | COP 3,000 | COP 3,400 | COP 3,900 | COP 4,300 | COP 4,500 |

Ejemplo de formato *long*

| Individuo | Marca elegida | Alternativas | y_j | Ingreso del individuo | \$ Alternativa |
|-------------|---------------|--------------|-----------|-----------------------|----------------|
| Individuo 1 | Póker | Águila | $y_1 = 0$ | COP 2,300,000 | COP 2,600 |
| Individuo 1 | Póker | Póker | $y_2 = 1$ | COP 2,300,000 | COP 3,100 |
| Individuo 1 | Póker | C. Colombia | $y_3 = 0$ | COP 2,300,000 | COP 3,300 |
| Individuo 1 | Póker | Corona | $y_4 = 0$ | COP 2,300,000 | COP 4,200 |
| Individuo 1 | Póker | Heineken | $y_5 = 0$ | COP 2,300,000 | COP 4,100 |
| Individuo 2 | C. Colombia | Águila | $y_1 = 0$ | COP 3,000,000 | COP 3,300 |
| Individuo 2 | C. Colombia | Póker | $y_2 = 0$ | COP 3,000,000 | COP 3,600 |
| Individuo 2 | C. Colombia | C. Colombia | $y_3 = 1$ | COP 3,000,000 | COP 4,000 |
| Individuo 2 | C. Colombia | Corona | $y_4 = 0$ | COP 3,000,000 | COP 5,700 |
| Individuo 2 | C. Colombia | Heineken | $y_5 = 0$ | COP 3,000,000 | COP 5,500 |
| Individuo 3 | Corona | Águila | $y_1 = 0$ | COP 2,700,000 | COP 3,000 |
| Individuo 3 | Corona | Póker | $y_2 = 0$ | COP 2,700,000 | COP 3,300 |
| Individuo 3 | Corona | C. Colombia | $y_3 = 0$ | COP 2,700,000 | COP 3,600 |
| Individuo 3 | Corona | Corona | $y_4 = 1$ | COP 2,700,000 | COP 5,000 |
| Individuo 3 | Corona | Heineken | $y_5 = 0$ | COP 2,700,000 | COP 4,800 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| Individuo n | Águila | Águila | $y_1 = 1$ | COP 2,500,000 | COP 3,000 |
| Individuo n | Águila | Póker | $y_2 = 0$ | COP 2,500,000 | COP 3,400 |
| Individuo n | Águila | C. Colombia | $y_3 = 0$ | COP 2,500,000 | COP 3,900 |
| Individuo n | Águila | Corona | $y_4 = 0$ | COP 2,500,000 | COP 4,300 |
| Individuo n | Águila | Heineken | $y_5 = 0$ | COP 2,500,000 | COP 4,500 |

Correo: uniic_bog@unal.edu.co

Teléfono: 3165000 ext 12301

Variable dependiente en formato *long* y *wide*

Para el formato *wide*:

$$y = j$$

Para el formato *long*:

$$y_j = \begin{cases} 1 & \text{si } y = j \\ 0 & \text{si } y \neq j \end{cases}$$

Variables explicativas

Las variables explicativas usadas en los modelos con variable dependiente discreta se clasifican en:

- ***Alternative-invariant regressors*** (también llamados *case-specific regressors*)
- ***Alternative-variant regressors*** (también llamados *alternative-specific regressors*)

La clasificación de las variables explicativas es muy importante ya que dependiendo de su tipo se puede o no utilizar un determinado modelo.

Alternative-invariant regressors

Las variables de tipo *alternative-invariant regressors* no tienen correspondencia directa con determinada alternativa; los valores de estas variables varían entre individuos (entidades) pero no varían entre alternativas para el mismo individuo (entidad).

Ejemplo:

- El ingreso. ¿Por qué? Porque el ingreso **no** depende de la marca de cerveza elegida. El ingreso del Individuo 1 sigue siendo COP 2,300,000 sin importar si elige Águila, Póker, Corona, etc.

Alternative-variant regressors

Las variables de tipo *alternative-variant regressors* tienen correspondencia directa con determinada alternativa, por lo que su valor depende de esta; los valores de estas variables varían entre alternativas y también pueden variar entre individuos.

Ejemplo:

- El precio. ¿Por qué? Porque el precio **sí** depende de la alternativa: el precio de Águila no es el mismo que el de Póker o el de Heineken. Dependiendo de la alternativa elegida se pagará un precio distinto.

Modelo Logit Multinomial

El Modelo Logit Multinomial acepta **únicamente** variables de tipo *alternative-invariant regressors*.

La forma funcional que adopta es:

$$P_{ji} = \text{Prob}[y_i = j] = \frac{e^{(\mathbf{w}_i^T \boldsymbol{\gamma}_j)}}{\sum_{k=1}^m e^{(\mathbf{w}_i^T \boldsymbol{\gamma}_k)}}$$

Donde P_{ji} es la probabilidad de que el individuo i elija la alternativa j ; \mathbf{w}_i es el vector de variables *alternative-invariant* para el individuo (entidad) i y $\boldsymbol{\gamma}_j$ es el vector de coeficientes correspondientes a la alternativa j .

Estimación en el Modelo Logit Multinomial

Para realizar la **estimación**, uno de los conjuntos de coeficientes se normaliza a cero, por lo que se obtienen $m - 1$ conjuntos de coeficientes.

La **categoría base** corresponde a la alternativa cuyo conjunto de coeficientes se normaliza a cero. La **interpretación** de los demás coeficientes se hace con referencia a la categoría base.

Coeficientes y Efectos Marginales

- Las magnitudes de los coeficientes en un modelo logit multinomial **no** tienen interpretación directa. Solo indican si aumenta (cuando el coeficiente es positivo) o disminuye (cuando el coeficiente es negativo) la probabilidad de que se elija la alternativa j , ***en comparación*** con la categoría base.
- Los coeficientes **no** cuantifican el impacto de una variable explicativa sobre la probabilidad; para esto se emplean los **efectos marginales**.
- **Importante:** ¡Los coeficientes son distintos dependiendo de la categoría base que se elija!

Efectos Marginales

Los efectos marginales en el Modelo Logit Multinomial están dados por:

$$\frac{\partial P_{ji}}{\partial w_f} = P_{ji} \left(\gamma_{jf} - \sum_{k=1}^m P_{ki} \gamma_{kf} \right)$$

Importante: Los efectos marginales no necesariamente tienen el mismo signo que los coeficientes asociados.

Medida “general” de efectos marginales

Al igual que en el Modelo Logit y el Modelo Probit, con n individuos (entidades) se obtienen n efectos marginales para la misma variable. Para hallar una medida “general” se puede usar efectos marginales *en la media* y efectos marginales *promedio*.

Importante: A diferencia de los coeficientes, sin importar la categoría base elegida, los efectos marginales siempre son los **mismos**.

Medida “general” de efectos marginales

Importante: Los efectos marginales (cuantificados con una medida “general”) de la misma variable deben sumar **0**. Es decir:

$$\sum_{j=1}^m \frac{\partial P_j}{\partial w_f} = 0$$

en donde $\frac{\partial P_j}{\partial w_f}$ es el efecto marginal de la variable w_f sobre la probabilidad de elegir la alternativa j y existen m alternativas.

Referencias

- Katchova, A. (2013). Econometrics – Probit and Logit Models.
- Katchova, A. (2013, 2). Econometrics – Multinomial Probit and Logit Models.