Modelos de elección simple y múltiple. Regresión logit y probit. Modelos multilogit y multiprobit.

Siga

Modelos de elección discreta.

- Modelos de elección simple.
- Modelos de elección múltiple.

- Final

Modelos de elección discreta. (Alternativas binarias)

Siga

Modelos de elección simple.

- Introducción.
- Planteamiento.
- Modelos probit y logit.

- Atrás

Introducción.

- Referencias.
- Análisis aplicado.
- Relevancia.

- Atrás

Referencias.

- Goldberger, A. S. (2001), *Introducción a la Econometría*. Ariel. Capítulo 17, págs. 175-190.
- Maddala, G. S. (1996), *Introducción a la Econometría*, 2ª ed. Prentice Hall. Capítulo 8.

- Atrás

Análisis aplicado

- Muro, J. e I. Senra (2002), *Ensayo 3. ¿Influye el tamaño de una empresa sobre la probabilidad de acceder a la financiación bancaria? ¿Cambia el sentido de la influencia a partir de un tamaño determinado?*

  - Atrás

Relevancia.

- Tablas multidimensionales con condición *ceteris paribus*.
- Mecanismos de selección muestral.
- Modelos no lineales para datos de panel.
- Métodos discretos para modelos de duración.

- Atrás

Planteamiento.

- Construcción del modelo a partir de una forma reducida.
- Modelización de la probabilidad asociada a la variable dicotómica a través de las expresiones derivadas de una formulación estructural.
    - Función índice (forma estructural de v. latente).
    - Utilidad aleatoria.

- Atrás

Modelo en forma reducida.

- Analogía con el modelo de regresión (v. objetivo con 0 y 1).
- Probabilidad es una función de la matriz de características **X** y de los parámetros a estimar β.
- Hipótesis de linealidad: el argumento de la función anterior es una combinación lineal de las características y de los parámetros.
    - <u>Siga</u>

Modelo en forma reducida.

- Probabilidad de que ocurra un suceso (se tome una decisión)

$$\boldsymbol{P_i = Prob\left[y_i = 1 \mid X_i, \beta\right] = G(X_i, \beta) =}$$
$$\boldsymbol{= G(X'_i\, \beta).}$$

- Probabilidad de que no ocurra un suceso (no se tome la decisión)

$$\boldsymbol{Prob\left[y_i = 0 \mid X_i, \beta\right] = 1 - G(.).}$$

Siga

Modelo en forma reducida.

- El valor esperado de la variable a estudio, y_i, es

$$\boldsymbol{E\left(y_i\right) = 1 * P_i + 0 * \left(1 - P_i\right) = P_i = G(.)}.$$

- *¿Cuáles pueden ser las candidatas para la función G(.)?*
  - *modelo lineal de probabilidad.*
  - *modelo probit.*
  - *modelo logit.*

Atrás

Modelo lineal de probabilidad.

- La función identidad, es decir G(.) = 1.
- Expresión formal es

$$y_i = X'_i \beta + \varepsilon_i \qquad \forall i = 1,2,\ldots N.$$

$$P_i = Prob\left(y_i = 1\right) = X'_i \beta.$$

Atrás

Modelo probit.

- La función de distribución normal, es decir G(.) = F(.).
- Expresión formal es

$$P_i = Prob(y_i = 1) = F(X'_i \beta).$$

Atrás

Modelo logit.

- La función de distribución logística, es decir G(.) = Λ(.).
- Expresión formal es

$$\boldsymbol{P_i = Prob(y_i = 1) = \frac{1}{1 + exp(-X'_i\,\beta)}}.$$

Atrás

Planteamiento de función índice.

- La variable observada toma unos valores que responden al comportamiento de una variable índice (latente o inobservable).
  - Si el índice supera un determinado nivel la variable discreta toma el valor uno y, si no lo supera, toma el valor cero.
- La variable latente está relacionada con las características a través de un modelo estructural.

$$y_i^* = X'_i \beta + \varepsilon_i \qquad \forall i = 1,2,\ldots N.$$

Siga

Planteamiento de función índice.

- Según establezcamos el supuesto de que la perturbación aleatoria se distribuya como una Normal o una Logística, el modelo generado será un probit o un logit, respectivamente.
- Los valores de la variable a estudio, observados, se relacionan con los de la variable latente en la forma

$$y_i = \mathbf{1}(y_i{}^* > 0) = \mathbf{1}(X'_i\,\beta + \varepsilon_i > 0).$$

Siga

Planteamiento de función índice.

- La probabilidad asociada a la realización del suceso tiene la expresión

$$\begin{aligned} P_i &= Prob\left[y_i = 1 \mid X_i, \beta\right] = Prob\left[y_i^* > 0\right] = \\ &= Prob\left[X_i\beta + \varepsilon_i > 0\right] = Prob\left[\varepsilon_i < X_i\beta\right] = \\ &= G(X'_i\,\beta). \end{aligned}$$

Siga

Planteamiento de función índice.

- El hacer la varianza del modelo igual a uno es el resultado de un mero proceso de normalización, ya que el modelo no se altera porque se multiplique por cualquier cantidad.
- El umbral considerado a superar por el índice puede ser cero o cualquier otro valor.

Atrás

Modelos probit y logit.

- Expresión de la función de verosimilitud.
- Ecuaciones de verosimilitud.
- Hessiano.
- Presentación e interpretación de resultados.
- Contrastes.
- Bondad del ajuste.

• Atrás

La función de verosimilitud de los modelos probit y logit es:

$$L(\beta \mid y_i, X_i) = \prod_{i=1}^{N} G\left(X'_i \beta\right)^{y_i} \left[1 - G\left(X'_i \beta\right)\right]^{1-y_i}.$$

$$\ln L = \sum_i \left\{ y_i \ln G\left(X'_i \beta\right) + (1 - y_i) \ln\left[1 - G\left(X'_i \beta\right)\right] \right\}.$$

Atrás

Las ecuaciones de verosimilitud de los modelos probit y logit son:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = 0.$$

$$\sum_i \left\{ y_i \frac{g(.)}{G(.)} X_i + (y_i - 1) \frac{g(.)}{1 - G(.)} X_i \right\} = 0.$$

$$\sum_i \left\{ \frac{y_i - G(.)}{G(.)\left[1 - G(.)\right]} g(.) X_i \right\} = 0.$$

Siga

Que se concretan en

$$\sum_i \frac{f(.)}{[1-F(.)]} X_i = 0. \qquad \text{Modelo probit}$$

$$\sum_i [y_i - \Lambda(.)] X_i = 0. \qquad \text{Modelo logit}$$

Atrás

El hessiano es definido negativo en ambos modelos y tiene la forma:

$$H = -\sum_i \lambda_i \left(\lambda_i + X'_i \beta \right) X_i X'_i .$$ Modelo probit

$$H = -\sum_i \Lambda(.)\left[1 - \Lambda(.)\right] X_i X'_i .$$ Modelo logit

Atrás

El hessiano es definido negativo en ambos modelos y tiene la forma:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = 0.$$

$$\sum_i \left\{ y_i \frac{g(.)}{G(.)} X_i + (y_i - 1) \frac{g(.)}{1 - G(.)} X_i \right\} = 0.$$

$$\sum_i \left\{ \frac{y_i - G(.)}{G(.)[1 - G(.)]} g(.) X_i \right\} = 0.$$

Atrás

Presentación e interpretación de resultados.

- Presentación:
  - Habitual (estimaciones parámetros).
  - Razón de probabilidades (odds ratio).
- Efectos marginales.
- Predicciones.

• Atrás

Contrastes.

- Especificación.
- Especificación errónea.

- Atrás

Contrastes de especificación errónea.

- Análisis de residuos generalizados.
- Contrastes de momentos condicionales. Pagan-Vella (1989).
- Contraste de heteroscedasticidad. Davidson y McKinnon (1993).

- Atrás

Pagan, A.R. y F. Vella (1989), “Diagnostic Tests for Models Based on Individual Data: A Survey”. *Journal of Applied Econometrics*, 4, S29-59.

Davidson, R. y J. MacKinnon (1993), *Estimation and Inference in Econometrics*. Oxford University Press.

Contraste de heteroscedasticidad. Harvey (1976). Modelo probit.

Hipótesis nula. H_0: $Var(u_i)$= cte; H_1: $Var(u_i)=\exp(2\gamma * var)$.

Regresión auxiliar. La suma de los cuadrados explicada de dicha regresión se distribuye asintóticamente bajo la hipótesis nula como una χ^2 con 1 grado de libertad (situación particular debida a que sólo hay una variable que cause la heteroscedasticidad, si hubiera más variables los grados de libertad serían iguales a su número).

La regresión auxiliar tiene la forma:

$$\frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} = \frac{f(-X'_i\theta)}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} X'_i \phi_1 + \frac{f(-X'_i\theta)*(-X'_i\theta)}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} var*\phi_2 + v_i.$$

Las p son predicciones, la X es la matriz de variables del lado derecho (que incluye la variable *var*) y, finalmente, f(.) es la función de densidad de probabilidad de una variable N(0,1).

Atrás

Bondad del ajuste.

- Contraste de la razón de verosimilitud.
- Seudo R^2.
- Tabla de predicciones.

- Atrás

Contraste de la razón de verosimilitud.

Compara de la forma habitual el modelo estimado con el modelo ingenuo (modelo que contiene sólo una constante).

Hipótesis nula. H_0: $\boldsymbol{\beta}=0$.

$$LR = -2\left(\ln L - \ln L_0\right) \underset{H_0}{\sim} \chi^2(r)$$

$$\ln L_0 = N\left[P \ln P + (1-P)\ln(1-P)\right].$$

Atrás

Seudo R^2.

Dificultades en crear un consenso sobre cuál es el modelo de comparación.

$$1 - \frac{\ln L}{\ln L_0}.$$

$$\ln L_0 = N\left[P \ln P + (1 - P)\ln(1 - P)\right].$$

Atrás

Tabla de predicciones.

Compara los resultados muestrales con los predichos para cada observación y suministra el número de aciertos y fallos.

Características de equilibrio de la muestra analizada.

Regla de construcción habitual.

$$\hat{P}_i > 0.5 \text{ implica que } \hat{y}_i = 1;$$

$$P_i \leq 0.5 \text{ implica que } \hat{y}_i = 0.$$

Atrás

Modelos de elección discreta. (Múltiples alternativas)

<u>Siga</u>

Modelos de elección discreta con múltiples alternativas.

- Introducción.
- Los temas.
  - Características comunes.
  - ¿Cuáles son las causas de la aleatoriedad?
  - Problemas que nacen del diseño muestral.
- Los modelos.

Introducción.

- Referencias.
- Thurstone, L. (1927), "A Law of Comparative Judgement". *Psychological Review*, 34, págs. 273-286.
- Luce, D. (1959), *Individual Choice Behaviour*. Wiley.

- Atrás

Referencias

- Greene, W. H. (1997), *Econometric Analysis*. 3ª ed. Macmillan. Capítulo 19, págs. 871-931.
- Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press. Capítulo 15, págs. 453-509.

- Atrás

Modelos de elección discreta con múltiples alternativas.

- Introducción.
- Los temas.
  - Características comunes.
  - ¿Cuáles son las causas de la aleatoriedad?
  - Problemas que nacen del diseño muestral.
- Los modelos.

Final

Los temas.

1. Compras de bienes duraderos (bienes indivisibles).
   - ✓ Automóviles. Cragg y Uhler (1970).
   - ✓ Fecundidad. Becker (1960).
2. Demanda de características (bien heterogéneo).
   - ✓ Transporte. McFadden (1974…….1984).
   - ✓ Vivienda; educación; ocupación……
3. Elección discreta como abstracción de demanda de un bien continuo.
   - ✓ Oferta de trabajo. Muro et al. (1986).
4. Otros temas.
   - ✓ Sentido del voto; construcción de autopistas.

Atrás

Características comunes.

- Comparación de utilidades.
- Hipótesis RUM (random utility maximization).
- IIA (Supuesto de independencia de las alternativas irrelevantes).

- Atrás

Causas de la aleatoriedad

- Heterogeneidad.
- Deficiencias de información.

- Atrás

Diseño muestral

- Exogeneidad.
- Endogeneidad.
  - Referencia útil: Pudney (1989).

- Atrás

Los modelos.

1. La información disponible.
2. Planteamiento general.
3. Modelos para datos no ordenados.
4. Modelos para datos ordenados.
5. Líneas de investigación.

Atrás

La información disponible.

- Individuos y alternativas.
  - n individuos; $i = 1,2,3,\ldots\ldots,n$.
  - J alternativas; $j = 0,1,2,3,\ldots\ldots,J-1$.
- Características de los individuos.
  - W_i es un vector (K x 1).
- Atributos de las alternativas.
  - X_{ij} es un vector (L x 1).
- En suma: $Z_{ij}= (X_{ij}\ W_i)$.

- Atrás

- Utilidad estocástica $U_{ij} = S_{ij} + \varepsilon_{ij}$.
  - Donde $S_{ij} = Z'_{ij}\gamma$. $\gamma = (\beta' \ \alpha')'$.

- El individuo i escoge la alternativa que maximiza su utilidad.
  - Prob $(y_i = j) = P_{ij} = \text{prob}\,(U_{ij} > U_{ik})$ $\surd\, k \neq j$.

- Si la distribución del término de error es
  - Normal $\rightarrow$ modelo probit;
  - Weibull $\rightarrow$ modelo logit.

Atrás

Modelos para un conjunto no ordenado de alternativas.

1. Modelo logit multinomial.
2. Modelo logit condicional.
3. Modelos anidados.
4. ¿Por qué especificaciones logit?

- Atrás

Modelo logit multinomial.

- McFadden (1973).
- Datos de características de los individuos (pero no de atributos de las alternativas).
- Planteamiento del modelo.
- Identificación.
- Razón de probabilidades (odds ratio).
- Estimación.
- Contrastes.
- Ejemplo 1; ejemplo 2.

Atrás

$$\text{Prob}\,(y_i = j) = P_{ij} = \frac{e^{W'_i \alpha_j}}{\sum_k e^{W'_i \alpha_k}}.$$

Donde i representa individuos y k corre para las alternativas (J).

Atrás

- Si $\alpha^*_j = \alpha_j + \alpha_0$.

$$\text{Prob}(y_i = j) = P_{ij} = \frac{e^{W'_i(\alpha^*_j - \alpha_0)}}{\sum_k e^{W'_i(\alpha^*_k - \alpha_0)}} = \frac{e^{W'_i \alpha^*_j}}{\sum_k e^{W'_i \alpha^*_k}}.$$

- Una mera traslación no altera el valor de las probabilidades (indeterminación).
- Manifestación de la propiedad de suma unitaria del conjunto de probabilidades.

Siga

La expresión normalizada queda (se hace la medida relativa a la alternativa cero, por lo que los coeficientes α_0 de la alternativa cero se hacen nulos):

$$\text{Prob}(y_i = j) = P_{ij} = \frac{e^{W'_i \alpha_j}}{1 + \sum_{k \neq 0} e^{W'_i \alpha_k}}.$$

Atrás

$$\frac{P_{ij}}{P_{i0}}=e^{W'_i\alpha_j};\quad \frac{P_{ij}}{P_{ik}}=\frac{e^{W'_i\alpha_j}}{e^{W'_i\alpha_k}}.$$

Si calculamos la relación en logaritmos (log odds ratio).

$$\ln\left(\frac{P_{ij}}{P_{i0}}\right)=W'_i\alpha_j;\quad \ln\left(\frac{P_{ij}}{P_{ik}}\right)=W'_i(\alpha_j-\alpha_k).$$

Atrás

$$\begin{aligned}&Hacemos \quad : \\ &d_{ij} = 1 \quad si \quad y_i = j; \\ &d_{ij} = 0 \quad si \quad y_i \neq j.\end{aligned}$$

La función de verosimilitud del modelo será:

$$L(\alpha_j, j = 1,2,....J \mid W_i ...) = \prod_{i=1}^{n} \prod_{j=0}^{J} \left[prob\,(d_{ij} = 1)\right]^{d_{ij}}.$$

$$\ln L = \sum_i \sum_j d_{ij} \ln \left\{ \frac{e^{W'_i \alpha_j}}{1 + \sum_{k \neq 0} e^{W'_i \alpha_k}} \right\}.$$

Atrás

Las ecuaciones de verosimilitud del modelo son:

$$\frac{\partial \ln L}{\partial \alpha_j} = \sum_i \left[d_{ij} - P_{ij} \right] W_i = 0.$$

El término entre corchetes se puede utilizar para realizar contrastes oportunos por el método de los contrastes de momentos condicionales.

Atrás

Modelo logit condicional.

- McFadden (1974).
- Datos de atributos de las alternativas y/o características de los individuos.
- Planteamiento.
- Identificación.
- Razón de probabilidades (odds ratio). IIA.
- Estimación. ¿ Logit para datos de panel con efectos fijos?
- Contrastes de especificación (IIA).
- Ejemplo 1; ejemplo 2.

Atrás

El calificativo condicional se deriva de que elegimos una alternativa concreta condicionada al hecho de que alguna de las alternativas se elige.

$$\text{Prob}(y_i = j) = P_{ij} = \frac{e^{Z'_{ij}\gamma_j}}{\sum_k e^{Z'_{ik}\gamma_k}}.$$

Donde i representa individuos y k corre para las alternativas (J).

La mayor parte de los ejemplos de logit condicional se construyen con solo atributos de las alternativas.

Es conveniente observar que la misma expresión se obtiene de considerar nuestra muestra como el resultado de un logit para datos de panel con efectos fijos, Chamberlain (1980).

Atrás

- **Supongamos que $\gamma_j = \gamma = \beta + \alpha$. Entonces,**

$$\text{Prob}(y_i = j) = P_{ij} = \frac{e^{X'_{ij}\beta + W'_i\alpha}}{\sum_k e^{X'_{ik}\beta + W'_i\alpha}} = \frac{e^{X'_{ij}\beta}}{\sum_k e^{X'_{ik}\beta}}.$$

- **Los parámetros de las características de los individuos desaparecen.**
- **La identificación se produce por medio de términos de interacción que permiten recuperar los parámetros asociados con las características.**
- **Lo mismo ocurre con el término independiente del modelo.**

Atrás

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{X'_{ij}\beta}}{e^{X'_{ik}\beta}}.$$

Si calculamos la relación en logaritmos (log odds ratio).

$$\ln\left(\frac{P_{ij}}{P_{ik}}\right) = \left(X'_{ij} - X'_{ik}\right)\beta.$$

Se cumple el supuesto IIA (restricción implícita en el modelo).

Atrás

$$
\begin{aligned}
&Hacemos: \\
&d_{ij} = 1 \quad si \quad y_i = j; \\
&d_{ij} = 0 \quad si \quad y_i \neq j.
\end{aligned}
$$

La función de verosimilitud del modelo será:

$$L(\beta \mid X_{ij}\ldots) = \prod_{i=1}^{n}\prod_{j=1}^{J}\left[prob(d_{ij}=1)\right]^{d_{ij}}.$$

$$\ln L = \sum_i \sum_j d_{ij} \ln\left\{\frac{e^{X'_{ij}\beta}}{\sum_k e^{X'_{ik}\beta}}\right\}.$$

Atrás

El contraste del supuesto IIA, Hausman y McFadden (1984), es un contraste de tipo Hausman. El modelo sin restricciones es el que contiene todas las alternativas y el restringido el mismo modelo estimado sin una de las alternativas.
Bajo IIA ambos estimadores son consistentes y el del modelo restringido eficiente. Bajo la alternativa sólo el del modelo sin restricciones es consistente.

$$\chi^2 = \left(\hat{\beta}_r - \hat{\beta}_{sr}\right)' \left[V_r - V_{sr}\right]^{-1} \left(\hat{\beta}_r - \hat{\beta}_{sr}\right) \underset{H_0}{\sim} \chi^2_{rest.}.$$

Donde:
r significa modelo restringido; sr modelo sin restricciones; V matriz de varianzas y covarianzas de los modelos respectivos; rest. = número de restricciones implícitas en el supuesto IIA (ambos modelos proporcionan los mismos resultados; nº parámetros del modelo restringido).

Atrás

Modelo logit anidado.

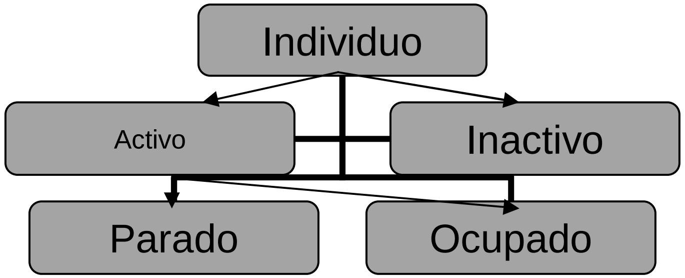
- McFadden (1984).
- Relajación del supuesto IIA.
- Modelo jerárquico. Construcción de un árbol de decisiones.
- Planteamiento del modelo.
- Estimación.
- Contrastes de especificación (logit condicional).
- Ejemplo 1; ejemplo 2.

Atrás

Supongamos que la elección se produce entre J alternativas, donde J = 3.

Ej. Mercado de trabajo: inactividad, paro y empleo.

Suponemos que la jerarquía (árbol de decisiones) se establece de la forma siguiente:



Llamamos ramas a cada una de las alternativas del primer nivel. Alternativas de cada rama a las que cuelgan de cada una de ellas.

Se admite que las ramas del árbol sean heteroscedásticas.

Atrás

Como información disponemos de los atributos de las ramas Z_l y de los atributos de las alternativas incluidas en cada rama X_{jl}.

Para el caso de un modelo para dos niveles, con L ramas en el primer nivel y con J_l alternativas en cada rama del segundo nivel, la probabilidad no condicionada de elegir la alternativa j que pertenece a la rama l es

$$P_{jl} = \frac{e^{X'_{jl}\beta + Z'_l\gamma}}{\sum_l \sum_j e^{X'_{jl}\beta + Z'_l\gamma}}.$$

Esta probabilidad puede descomponerse en el producto de la condicional por la marginal. Es decir,

$$P_{jl} = P_{j|l}P_l = \frac{e^{X'_{jl}\beta + Z'_l\gamma}}{\sum_l \sum_j e^{X'_{jl}\beta + Z'_l\gamma}} \frac{\sum_j e^{X'_{jl}\beta}}{\sum_j e^{X'_{jl}\beta}} \frac{\sum_l e^{Z'_l\gamma}}{\sum_l e^{Z'_l\gamma}}.$$

Siga

Se define el valor inclusivo, o de dentro de la rama, como

$$I_l = \ln \sum_j e^{X'_{jl}\beta}.$$

Y también la probabilidad de la alternativa j condicionada a encontrarse en la rama l como

$$P_{j|l} = \frac{e^{X'_{jl}\beta}}{\sum_j e^{X'_{jl}\beta}}.$$

Y la probabilidad marginal de la rama l como

$$P_l = \frac{e^{Z'_l\gamma+\pi_l I_l}}{\sum_l e^{Z'_l\gamma+\pi_l I_l}}.$$

Siga

En definitiva, de las definiciones anteriores se deriva que un modelo logit anidado tiene un doble comportamiento:

1. Modelo logit condicional interno para la elección de las alternativas de la rama correspondiente.

2. Modelo logit condicional para el conjunto de las ramas. En este logit condicional se consideran como atributos de las ramas no sólo las Z_l sino también los valores inclusivos I_l.

Atrás

1. **Procedimiento maximoverosímil.**
   - **Aplicación directa de la maximización de la función de verosimilitud (producto de la contribución de cada observación de la muestra).**

2. **Procedimiento en dos etapas.**
   - **Estimación del modelo logit condicional intrarrama.**
   - **Cálculo del (o de los) valor inclusivo.**
   - **Estimación del modelo logit condicional de las ramas.**

Atrás

Cabe realizar un contraste de especificación que es equivalente al contraste del supuesto IIA.

En efecto, si en un modelo anidado la probabilidad condicional de elegir una alternativa dentro de una rama coincide con la no condicional el modelo se reduce a un modelo logit condicional de todas las alternativas contempladas (al margen de su pertenencia o no a una rama concreta)

Matemáticamente se reduce al contraste de que el parámetro, o parámetros, que afectan a los valores inclusivos son iguales a 1.

Atrás

Otras especificaciones distintas a las logísticas.

- Facilidad de cálculo.
- Imposibilidad de calcular integrales multidimensionales implícitas en modelos probit. Máximo J=3.
- Probit multinomial.
- Simulación. Nuevas investigaciones.

Atrás

Probit multinomial

- Debería llamarse probit condicional.
- Alternativa al logit condicional que evita la IIA.
- Las J perturbaciones aleatorias se distribuyen como una distribución normal J-variante, que admite correlaciones para $i \neq j$.

  - Atrás.

Modelos para un conjunto ordenado de alternativas.

- Planteamiento general.
- No se conocen los límites de los intervalos.
- Se conocen los límites de los intervalos: Variable dependiente agrupada.

Atrás

Planteamiento general.

- Los datos dan información de dos cuestiones:
  - de la alternativa escogida;
  - y de la ordenación de las alternativas posibles .
- Modelización de la información discreta y ordenada a través de una función índice inobservable.

$$y_i^* = X'_i \beta + \varepsilon_i .$$

Siga

•La conversión de los valores de la función índice en los valores discretos ordenados sigue la regla siguiente:

$$y = 0 \quad si \quad y^* \leq 0,$$
$$y = 1 \quad si \quad 0 < y^* \leq \mu_1,$$
$$y = 2 \quad si \quad \mu_1 < y^* \leq \mu_2,$$
$$y = J \quad si \quad \mu_{J-1} < y^*.$$

•La conversión puede considerarse una censura de los datos. Sólo se conoce el intervalo en el que está una observación concreta.

Siga

•En general, el modelo se representa como:

$$\boldsymbol{Prob(y=0) = Prob(y^{*} \leq 0) = G(-X\beta),}$$

$$\boldsymbol{Prob(y=1) = Prob(0 < y^{*} \leq \mu_1) =}$$

$$\boldsymbol{G(\mu_1 - X\beta) - G(-X\beta),}$$

•••••••

•Cuando G(.) es la función de distribución de una normal el modelo será un probit; cuando es la logística será un logit.

Atrás

•Si los parámetros μ_j son desconocidos, los modelos probit y logit ordenados son los oportunos.

•*Modelo probit (logit) ordenado*.

•Dado que los límites del intervalo son desconocidos podemos hacer estándar el modelo y suponer que las perturbaciones aleatorias del modelo son N(0,1).

•*Expresión formal del modelo*.

•Para la estimación empleamos el método de la máxima verosimilitud. La función de verosimilitud es

$$L(\beta, \mu_j) = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[G(\mu_j - X'_i \beta) - G(\mu_{j-1} - X'_i \beta)\right]^{d_{ij}}.$$

$$\ln L = \sum_i \sum_j d_{ij} \ln \left[G(\mu_j - X'_i \beta) - G(\mu_{j-1} - X'_i \beta)\right].$$

•Donde,

$$d_{ij} = 1 \quad si \quad \mu_{j-1} < y_i^* \leq \mu_j,$$

$$d_{ij} = 0 \quad en\ el\ resto\ de\ los\ casos.$$

Atrás

•**Si los parámetros μ_j son conocidos, el modelo apropiado es el de variables dependientes agrupadas.**

•*Modelo de variable dependiente agrupada*.

•**Al conocer los límites del intervalo, el modelo no se puede hacer estándar y los parámetros a estimar serán β y σ^2.**

•*Expresión formal del modelo*.

•**La función de verosimilitud del modelo es**

$$\ln L = \sum_i \sum_j d_{ij} \ln\left[G\left(\frac{\mu_j - X'_i \beta}{\sigma} \right) - G\left(\frac{\mu_{j-1} - X'_i \beta}{\sigma} \right) \right].$$

•**Donde las equivalencias con la expresión anterior son inmediatas.**

Atrás