

Modelos Logísticos e Hierárquicos

**Ernesto F. L. Amaral
Magna M. Inácio**

**30 de setembro de 2010
Tópicos Especiais em Teoria e Análise Política:
Problema de Desenho e Análise Empírica (DCP 859B4)**

REGRESSÃO LOGÍSTICA MULTINOMIAL

- A regressão logística multinomial pode ser vista como uma extensão do modelo logístico binário, em situações nas quais a variável dependente tem múltiplas categorias.
- É possível estimar uma regressão logística em que a variável dependente tem mais de duas categorias.
- São geradas $k - 1$ equações, sendo k o número de categorias.
- As equações geram probabilidades para predizer se uma categoria está acima/abaixo da categoria de referência.

GOVERNISTA / INDEPENDENTE / OPOSICIONISTA

- Com a intenção de estimar um modelo que explique variações na escala governista/independente/oposicionista, é possível utilizar o modelo logístico multinomial.
- Este tipo de modelagem é utilizado quando a variável dependente em questão é ordinal.
- No nível ordinal de mensuração, as categorias da variável podem ser organizados em alguma ordem.
- Sabe-se que há diferenças relativas entre os valores dos dados, mas as magnitudes das diferenças entre as categorias não são conhecidas.
- Na escala de governista/independente/oposicionista, é possível ordenar os dados, mas não sabemos se a diferença entre “governista” e “independente” é a mesma que “independente” e “oposicionista”.

VARIÁVEIS ORDINAIS

- Powers and Xie (2000) argumentam que variáveis ordinais podem ser vistas como algo entre variáveis nominais de um lado e variáveis contínuas por outro.
- As variáveis ordinais são:
 - Mais gerais que as variáveis contínuas, já que permitem a existência de diferentes distâncias entre as categorias.
 - Mais restritas que as variáveis nominais, já que suas categorias podem ser colocadas em ordem.
- Em alguns casos, a variável dependente ordinal também pode ser tratada como uma variável contínua.
- Em suma, uma variável ordinal revela a ordem de ordenamento de seus diferentes valores, mas não as magnitudes de suas diferenças em uma escala de significado substantivo.

APLICAÇÃO À ESCALA GOVERNISTA/OPOSICIONISTA

– A escala governista/independente/oposicionista possui um total de três categorias ($J=3$), o que gera três probabilidades distintas:

– Equação 1:

$$Pr(y_i = 1 | x_i) = P_{i1} = \frac{1}{1 + \exp(x'_i \beta_2) + \exp(x'_i \beta_3)}$$

– Equação 2:

$$Pr(y_i = 2 | x_i) = P_{i2} = \frac{\exp(x'_i \beta_2)}{1 + \exp(x'_i \beta_2) + \exp(x'_i \beta_3)}$$

– Equação 3:

$$Pr(y_i = 3 | x_i) = P_{i3} = \frac{\exp(x'_i \beta_3)}{1 + \exp(x'_i \beta_2) + \exp(x'_i \beta_3)}$$

PROBABILIDADES

- Nas equações anteriores, β_2 e β_3 denotam os efeitos específicos da variável independente para as segundas e terceiras categorias de respostas, tomando a primeira categoria como referência.
- A primeira equação (P_{i1}) é derivada do fato de que as três probabilidades somam uma unidade [$P_{i1}=1-(P_{i2}+P_{i3})$].
- As probabilidades de resposta da variável dependente dependem de transformações não-lineares da função linear:

$$x'_i \beta_j = \sum_{k=0}^K \beta_{jk} x_{ik}$$

- K é o número de variáveis independentes.

IIA (LONG E FREESE 2001)

- Os modelos logísticos multinomiais e os modelos logísticos condicionais se baseiam no pressuposto de independência de alternativas irrelevantes (“independence of irrelevant alternatives” - IIA).
- Com este pressuposto, as probabilidades não dependem das outras categorias que são disponíveis.
- As outras categorias alternativas são irrelevantes, o que significa que adicionar ou remover categorias da variável dependente não afetam as chances das categorias restantes.

EXEMPLO DO ÔNIBUS VERMELHO / AZUL

- Suponha que:
 - Temos a escolha de ir para o trabalho em um ônibus vermelho ou em um carro.
 - A chance de pegar um ônibus vermelho comparada ao carro seja 1:1.
- O pressuposto IIA implica que as chances continuarão 1:1 entre estas duas alternativas, mesmo se chegue à cidade uma nova companhia de ônibus azul, a qual é idêntica à companhia de ônibus vermelho (exceto pela cor).
- Esse pressuposto não é realístico, já que a chance de pegar um ônibus vermelho, comparado a um carro, reduziria a 1:2, já que metade daqueles que pegariam o ônibus vermelho agora tomarão o ônibus azul.

VOLTANDO À CIÊNCIA POLÍTICA

- Vamos retirar uma das categorias da nossa escala governista-independente-oposicionista:
 - Ao retirar a categoria “governista”, os congressistas teriam que se distribuir entre “independentes” e “oposicionistas” de tal modo que a razão entre estes dois últimos grupos não apresentasse mudança.
- Glasgow (2001: 1): “Models which assume IIA, such as multinomial logit (MNL), assume that the ratio of the probability of voting for party A over the probability of voting for party B remains unchanged when party C enters or leaves the election.”
- Esses são pressupostos razoáveis na Ciência Política?

KIEWIET E ZENG (1993)

- O modelo empregado por Kiewiet e Zeng (1993) foi o “**mother logit model**”, elaborado para relaxar o pressuposto de IIA e mantendo a forma de logito para as probabilidades.
- Estes autores estimaram o modelo com a introdução de fatores que influenciariam a utilidade de “concorrer a cargos superiores” nas funções de utilidade de “decidir pela aposentadoria” e “concorrer à reeleição”.
- Portanto, as variáveis que influenciariam a escolha entre “concorrer a cargos superiores” ou “concorrer à reeleição” foram usadas na estimação da escolha entre “decidir pela aposentadoria” ou “concorrer à reeleição”.
- Eles também testaram o “nested logit” que não toma o IIA como pressuposto, mas as probabilidades estimadas não foram adequadas, em comparação ao “mother logit”.

GLASGOW (2001): CORRELAÇÃO COM O TEMPO

- Glasgow reestimou o modelo descrito por Kiewiet e Zeng, usando o modelo “mixed logit” (MXL), o qual apresenta avanços em relação ao “mother logit” por **dois aspectos**.
- **Primeiro**, Kiewiet e Zeng tomaram como pressuposto que as escolhas individuais não são correlacionadas com tempo.
- MXL não utiliza esse pressuposto, ao permitir que as escolhas feitas por indivíduos esteja correlacionada com o tempo.
- Portanto, o MXL permite correlação entre alternativas (relaxando a propriedade de IIA) e permite correlação no decorrer do tempo.

GLASGOW (2001): COMPORTAMENTO INDIVIDUAL

- **Segundo**, modelos “mother logit” podem aproximar as probabilidades de escolha de qualquer modelo de escolha discreta com a forma logística, mas não são consistentes com a maximização da utilidade aleatória (“random utility maximization”).
- No “mother logit”, os atributos de cada alternativa podem entrar nas funções de utilidade das outras alternativas.
 - Por exemplo, componentes da função de utilidade de “concorrer a cargo superior” entram na escolha entre “aposentadoria” ou “concorrer à reeleição”, embora esta escolha deveria depender somente da utilidade da “aposentadoria” ou da utilidade de “concorrer à reeleição”.
- No “mixed logit”, além de não haver pressuposto de IIA, há consistência com teorias do comportamento individual (“random utility maximization”).

GLASGOW (2001): PRESSUPOSTOS

- Modelos “mixed logit” são também conhecidos como “random parameters logit”, “mixed multinomial logit”, “random coefficients logit” e “error components logit”.
- Estes modelos se baseiam no pressuposto de que as preferências do eleitor frente às alternativas é uma escolha que pode ser descrita como uma **função de utilidade**.
- Essa função depende de **atributos das alternativas** e de **características do indivíduo**. Ao escolher uma alternativa, indivíduos selecionam aquela que gera maior utilidade.
- A função de utilidade representa a soma: (1) do **componente sistemático** que depende de atributos observados das alternativas e de características individuais; e (2) do **componente estocástico** que representa a influência de fatores não-observados na escolha individual.

GLASGOW (2001): ALGUMAS OBSERVAÇÕES

- O termo genérico de “multinomial logit” foi utilizado por Glasgow (2001) se referindo ao:
 - Logito multinomial (“**multinomial logit**”): que possui variáveis específicas para os indivíduos; e
 - Logito condicional (“**conditional logit**”): que possui variáveis específicas para as alternativas e, possivelmente, para os indivíduos.
 - A forma geral dos dois modelos é idêntica, com a diferença de escolha das variáveis incluídas.
- Ao se referir a esta classe de modelos como “multinomial logit”, Glasgow (2001) utiliza a mesma terminologia do modelo “multinomial probit”, que é chamado desta forma independente das variáveis incluídas:
 - Assim como o “mixed logit”, o “multinomial probit” (MNP) também não utiliza o pressuposto de IIA.

MODELO HIERÁRQUICO

- O modelo de logit hierárquico (valor extremo generalizado - GEV) permite violações do pressuposto IIA na decisão entre os grupos de alternativas, mas exige que o pressuposto seja válido dentro dos grupos.
- Esse modelo exige a especificação de quais variáveis afetam cada nível de decisão.
- Em Leoni et al. (2003), a variável dependente é a escolha do legislador no final do seu mandato.
- É possível então estimar modelo em que:
 - Algumas variáveis influenciam na decisão entre aposentar-se, concorrer a cargo mais alto ou reeleição.
 - Algumas variáveis influenciam apenas sobre a decisão entre concorrer a cargo mais alto e reeleição.

EXEMPLIFICAÇÃO DE MODELO HIERÁRQUICO

– Primeiro nível:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + r$$

– Segundo nível:

$$\begin{aligned}\beta_1 &= \gamma_{10} + \gamma_{11} z_1 + u_1 \\ \beta_2 &= \gamma_{20} + \gamma_{21} z_1 + \gamma_{22} z_2 + u_1\end{aligned}$$

REFERÊNCIAS

- Glasgow, G. 2001. “Mixed logit models in political science.” Presented at the Eighteenth Annual Political Methodology Conference. Emory University, July 19.
- Kiewiet, D.R., e L. Zeng. 1993. “An analysis of congressional career decisions, 1947-1986.” *American Political Science Review*, 87(4): 928-941.
- Leoni, E., C. Pereira, e L. Rennó. 2003. “Estratégias para sobreviver politicamente: escolhas de carreiras na Câmara de Deputados do Brasil.” *Opinião Pública* 9(1): 44-67.
- Long, J.S., e J. Freese. 2001. “Regression models for categorical dependent variables using Stata.” College Station, Texas: Stata Press Publication.
- Powers, D.A., e Y. Xie. 2000. “Statistical methods for categorical data analysis.” San Diego: Academic Press.
- Triola, M.F. 2008. “Introdução à estatística.” Rio de Janeiro: LTC.