

Modelos Logísticos e Probit

Ernesto F. L. Amaral
Magna M. Inácio

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

RAZÃO DE CHANCES (ODDS RATIO)

- Probabilidade de sucesso = p
- Probabilidade de fracasso = $1 - p = q$
- Chance = (prob. de sucesso) / (prob. de fracasso)
- Por exemplo, se a probabilidade de sucesso é 0,75, a chance é igual a 3 (0,75/0,25).
- Razão de chances para variáveis dependentes binárias é a razão entre a chance de uma linha (ou coluna) de uma tabela 2x2, dividida pela chance da outra linha (ou coluna):

$$\frac{p/(1-p)}{q/(1-q)} = \frac{p(1-q)}{q(1-p)}$$

EQUAÇÃO DE REGRESSÃO

- A parte linear da equação da regressão logística é usada para encontrar a probabilidade de estar em uma categoria, baseado na combinação de variáveis independentes.

$$\bar{Y}_i = \frac{e^{A+B_1X_1+B_2X_2+B_3X_3}}{1 + e^{A+B_1X_1+B_2X_2+B_3X_3}}$$

- Os coeficientes de regressão e seus erros padrões são estimados com métodos de máxima verossimilhança.

TESTE DE WALD

- Cada coeficiente é avaliado usando o teste de Wald, que é simplesmente um teste de escore z:

$$W_j = \frac{B_j}{SE_{B_j}}$$

- Os testes dos coeficientes são aproximadamente escores z, os quais são posteriormente elevados ao quadrado, fazendo com que esta estatística tenha distribuição qui-quadrado.
- Esse teste é usado para avaliar a significância de cada coeficiente (β) no modelo.
- O teste de Wald é conhecido por ser conservador (aumenta o erro II).

ERROS TIPO I E TIPO II

- Ao testar H_0 , chegamos a uma conclusão de rejeitá-la ou de deixar de rejeitá-la.
- Tais conclusões pode estar corretas ou erradas.

		Estado verdadeiro da natureza	
		A hipótese nula é verdadeira	A hipótese nula é falsa
Decisão	Decidimos rejeitar a hipótese nula.	Erro tipo I (rejeitar uma hipótese nula verdadeira) α	Decisão Correta
	Deixamos de rejeitar a hipótese nula	Decisão Correta	Erro tipo II (deixar de rejeitar uma hipótese nula falsa) β

- α : probabilidade de erro tipo I (probabilidade de rejeitar hipótese nula quando ela é verdadeira).
- β : probabilidade de erro tipo II (probabilidade de deixar de rejeitar hipótese nula quando ela é falsa).

TESTE DE RAZÃO DE VEROSSIMILHANÇA

- Logaritmo da verossimilhança (Log-likelihood):

$$\log\text{-likelihood} = \sum_{i=1}^N [Y_i \ln(\hat{Y}_i) + (1 - Y_i) \ln(1 - \hat{Y}_i)]$$

- Modelos são comparados com uso dos logaritmos das verossimilhanças dos modelos:

$$\chi^2 = -2 [(\log\text{-likelihood do modelo restrito}) \\ - (\log\text{ likelihood do modelo irrestrito})]$$

ou

$$\chi^2 = 2 [(\log\text{-likelihood do modelo irrestrito}) \\ - (\log\text{ likelihood do modelo restrito})]$$

- Modelos precisam ser aninhados para comparação, ou seja, todas variáveis independentes do menor modelo (restrito) devem estar incluídas no maior modelo (irrestrito).

MULTINOMIAL

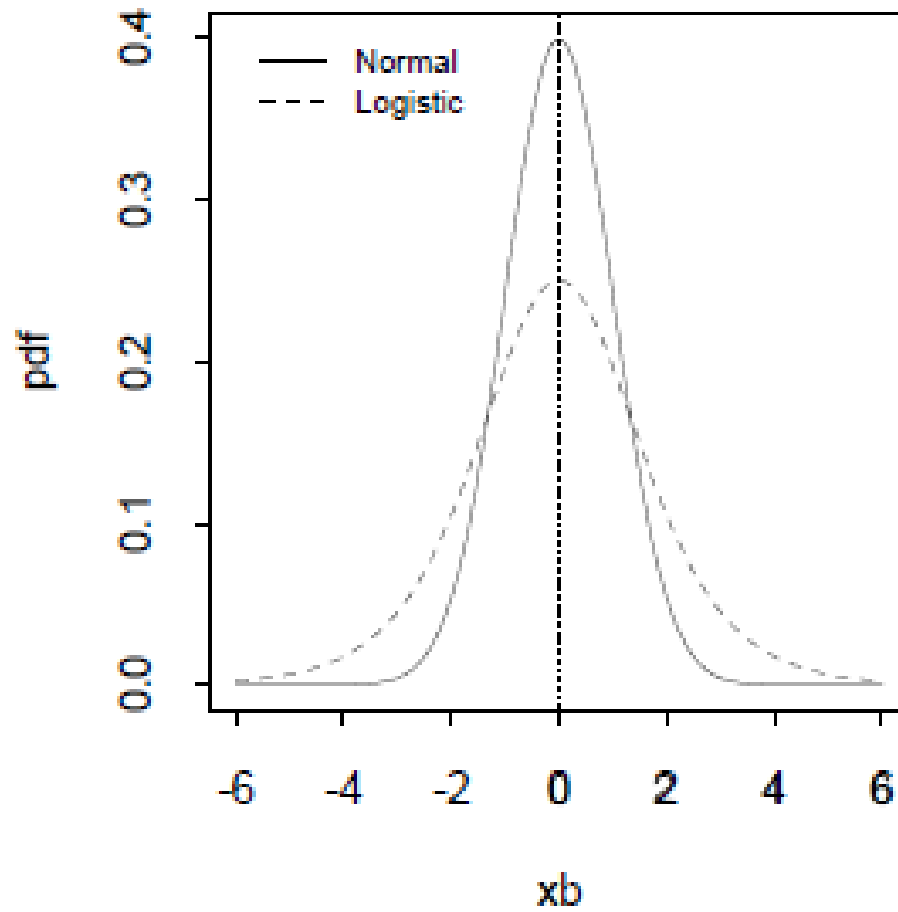
- É possível estimar uma regressão logística em que a variável dependente tem mais de duas categorias.
- São geradas $k - 1$ equações, sendo k o número de categorias.
- As equações geram probabilidades para predizer se uma categoria está acima/abaixo da categoria de referência.

PSEUDO R^2

- Há várias medidas de associação que pretendem servir como um R^2 na regressão logística.
- Porém, nenhuma destas medidas é realmente o R^2 .
- A interpretação não é a mesma, mas eles podem ser vistos como uma aproximação da variação na variável dependente, devido à variação nas variáveis independentes.

PROBIT X LOGÍSTICA

- A regressão probit assume uma distribuição normal e utiliza os escores z para estimar a proporção abaixo da curva.
- As análises são diferentes somente nos extremos das observações.
- A função de densidade de probabilidade (PDF) logística corresponde a uma distribuição t de Student com sete graus de liberdade, possuindo caudas mais grossas que a PDF normal.
- A escolha entre os dois modelos raramente gera resultados substantivamente diferentes, a não ser que os dados da variável dependente tenham distribuição muito assimétrica.
- As duas distribuições possuem variâncias distintas, o que gera parâmetros diferenciados entre os modelos.

NORMAL (PROBIT)**X****LOGÍSTICA****Probability Density
Functions (PDFs)****Cumulative Density
Functions (CDFs)**