Análise de Dados Categorizados - Aula 14

Márcia D Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística
mdbranco@usp.br - sala 295-A -

Modelo de logitos cumulativos (MLC)

- Para construção dos logitos cumulativos precisamos que a variável resposta Y seja do tipo ordinal.
- Os logitos serão construídos a partir das probabilidades acumuladas, denotadas por

$$\begin{aligned}
&\theta_1(x) = p_1(x) = P(Y \leq 1 \mid x) \\
&\theta_2(x) = p_1(x) + p_2(x) = P(Y \leq 2 \mid x) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\theta_{c-1}(x) = p_1(x) + p_2(x) + \cdots + p_{c-1}(x) = P(Y \leq c-1 \mid x)
\end{aligned}$$

- O j-ésimo logito cumulativo é definido como

$$logitoC_j = \log\left[\frac{\theta_j(x)}{1-\theta_j(x)}\right] \quad j = 1, \ldots, c-1$$

Modelo de logitos cumulativos (MLC)

- Para cada logito define-se uma reta de regressão

$$logitoC_j = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{qj}x_q$$

- Outras funções de ligação (g) podem ser consideradas (probito, cloglog, t-Student, etc).
- De um modo geral, considerando $g = F^{-1}$, podemos escrever

$$g(\theta_j(x)) = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{qj}x_q$$

- Lembre que o modelo de chances proporcionais é obtido quando $\beta_{ij} = \beta_i$ para $i = 1, \ldots, q$ e $j = 1, \ldots, c-1$.

Modelo multinomial com estrutura de variáveis latentes

Considere Z uma variável latente contínua tal que

$$Z = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \epsilon$$

Em que ϵ é uma v.a. contínua simétrica em torno do zero com variância constante para todos valores de x.

Sejam $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_c = \infty$ pontos de cortes.

O modelo latente é construído considerando

$$Y = j \ , \quad \text{se} \quad \alpha_{j-1} < Z < \alpha_j \quad j = 1, \ldots, c.$$

Modelo multinomial com estrutura de variáveis latentes

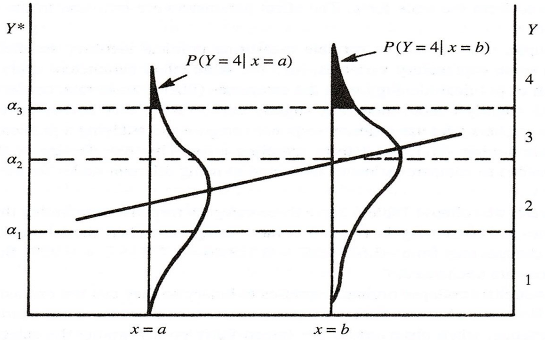

Figure 6.4 Ordinal measurement and underlying regression model for a latent variable. The left vertical axis shows values for the latent variable. The right vertical axis shows values for the observed ordinal response, its category being determined by the three cutpoints on the latent variable scale. The curves show the conditional distribution of the latent variable at two values of the explanatory variable. The line connecting their means represents the regression model for the latent variable.

Modelo multinomial com estrutura de variáveis latentes

Podemos demonstrar que

$$g(\theta_j(x)) = \alpha_j - \beta_1 x_1 - \cdots - \beta_q x_q \quad j = 1, \ldots, c-1.$$

A função de ligação g depende da suposição considerada para ϵ.

Mostrar em sala de aula!

- A ligação logito (modelo logístico) é obtida com a suposição $\epsilon \sim$ *Logistica*.
- A ligação probito é obtida com a suposição $\epsilon \sim N(0, 1)$.
- Outra suposições também podem ser consideradas, por exemplo, t-Student.
- A estrutura latente associada as variáveis ordinais Y impõe um modelo de chances proporcionais.

Regressão multinomial com chances proporcionais parciais

Considera que para algumas covariáveis as chances são proporcionais e outras tem o coeficiente variando em cada logito.

$$g(\theta_j(x)) = \beta_{0j} + \beta_j^T x \quad , \; j = 1, 2, \ldots c - 1$$

Em que $x^T = (x_1, x_2, \ldots, x_q)$ e

$$\beta_j^T = (\beta_1, \beta_2, \ldots, \beta_s, \beta_{(s+1)j}, \ldots, \beta_{qj})$$

Alguns coeficientes devem variar com o logito e outros são mantidos constantes.

Modelo logito de categorias adjacentes (MLCA)

- Para construção dos logitos de categorias adjacentes precisamos que a variável resposta Y seja do tipo ordinal.
- O j-ésimo logito adjacente é definido por

$$logitoA_j = \log\left[\frac{p_j(x)}{p_{j+1}(x)}\right] \;, \quad j = 1, \ldots, c-1$$

- O modelo de regressão é dado por

$$logitoA_j = \beta_{0j} + \beta_j^T x \;\;, \; j = 1, 2, \ldots c-1$$

- Casos especiais são obtidos com a suposição de chances proporcionais ou proporcionais parciais.

Modelo logito de categorias adjacentes (MLCA)

Para o MLCA as probabilidades de sucesso em cada casela são obtidas por

$$p_j(x) = \frac{\exp\left\{\sum_{l=j}^{c-1} \beta_{0l} + \beta_l^T x\right\}}{\left[1 + \sum_{k=1}^{c-1} \exp\left\{\sum_{l=k}^{c-1} \beta_{0l} + \beta_l^T x\right\}\right]} \quad j = 1, \ldots, c-1$$

$$p_c(x) = 1 - \sum_{j=1}^{c-1} p_j(x)$$

Observação: Tem erro na expressão da pag.188 do livro da Suely Giolo.

Vamos ver a prova para o caso especial $c = 4$.
Usamos a notação $\eta_j = \beta_{0j} + \beta_j^T x$.

Da expressão dos logitos adjacentes, temos que:

$$\begin{aligned} p_1(x) &= p_2(x)\exp\{\eta_1\} \\ p_2(x) &= p_3(x)\exp\{\eta_2\} \\ p_3(x) &= p_4(x)\exp\{\eta_3\} \end{aligned}$$

Reescrevendo:

$$\begin{aligned} p_1(x) &= p_4(x)\exp\{\eta_1 + \eta_2 + \eta_3\} \\ p_2(x) &= p_4(x)\exp\{\eta_2 + \eta_3\} \\ p_3(x) &= p_4(x)\exp\{\eta_3\} \end{aligned}$$

Somando as equações anteriores, obtemos

$$p_1(x)+p_2(x)+p_3(x) = p_4(x)[\exp\{\eta_1+\eta_2+\eta_3\}+\exp\{\eta_2+\eta_3\}+\exp\{\eta_3\}]$$

Equivalente a

$$1 - p_4(x) = p_4(x)[\exp\{\eta_1 + \eta_2 + \eta_3\} + \exp\{\eta_2 + \eta_3\} + \exp\{\eta_3\}]$$

Portanto

$$p_4(x) = [\exp\{\eta_1 + \eta_2 + \eta_3\} + \exp\{\eta_2 + \eta_3\} + \exp\{\eta_3\} + 1]^{-1}$$

Substituindo esse valor no sistema de equações da página anterior, obtemos o resultado.

Modelo logito de categorias adjacentes (MLCA)

Exemplo 1: O objetivo é comparar dois tratamentos para artrite raumatoide em pacientes de ambos os sexos. A resposta da doença aos tratamentos foi registrada como: remissão total (RT), remissão parcial (RP), nenhuma mudança (NM) ou progressão da doença (PD).

Tratamento	Sexo	Resposta				Totais
		RT	RP	NM	PD	
A	M	25	30	46	27	128
	F	07	15	35	13	70
B	M	20	19	44	42	125
	F	08	10	21	36	75

Modelo logito de categorias adjacentes (MLCA)

- Considere $x_1 = 1$ (tratamento B) e $x_2 = 1$ (sexo masculino).
- Temos 4 categorias e 3 logitos.
- Realizado o TRV para $H_0 : \beta_j = \beta \ , \ j = 1, 2, 3$, resultou numa estatística igual a 8.99 com valor-P $= 0.06$.
- Decidiu-se pela rejeição da hipótese H_0 e por testar separadamente os coeficientes.
- Considerando-se somente x_1 no modelo, o teste de igualdade dos coeficientes resultou em TRV=8.56 com valor-P=0.013.
- Considerando-se somente x_2 no modelo, o teste de igualdade dos coeficientes resultou em TRV=0.41 com valor-P=0.81.
- Os graus de liberdades do modelo saturado é 12 .

Modelo logito de categorias adjacentes (MLCA)

Tabela 8.13 – Resultados dos modelos ajustados aos dados de artrite

Modelo	*g.l.*	*Deviance*	TRV	$\neq$ de *g.l.*	Valor p	AIC
Nulo	$12-3=9$	27,357				86,30
X_1	$12-6=6$	9,398	17,958	3	$< 0{,}001$	74,34
$X_2 \mid X_1$	$12-7=5$	4,652	4,746	1	0,029	71,59
$X_1 * X_2 \mid X_1, X_2$	$12-8=4$	4,578	0,074	1	0,786	73,52

Nota: X_1 = tratamento, X_2 = sexo, $g.l.$ = graus de liberdade e $\neq$ denota diferença.

O modelo final considerado foi de o de chances proporcionais parcias sem interação.

$$logitoA_j = \beta_{0j} + \beta_{1j}x_1 + \beta_2 x_2$$

Tabela 8.14 – Estimativas dos parâmetros e respectivos erros-padrão do modelo selecionado para a análise dos dados de artrite, $j = 1, 2, 3$

Parâmetros	Logitos categorias adjacentes		
	$\ln\left[\frac{P(Y=1\mid x)}{P(Y=2\mid x)}\right]$	$\ln\left[\frac{P(Y=2\mid x)}{P(Y=3\mid x)}\right]$	$\ln\left[\frac{P(Y=3\mid x)}{P(Y=4\mid x)}\right]$
β_{0j}	−0,502 (0,245)	−0,737 (0,199)	0,568 (0,202)
β_{1j}: tratamento B	0,307 (0,352)	−0,218 (0,291)	−0,887 (0,256)
β_{2j}: sexo M	0,228 (0,106)	0,228 (0,106)	0,228 (0,106)

Modelo logito de categorias adjacentes (MLCA)

Tabela 8.15 – Probabilidades preditas a partir do modelo selecionado

Tratamento	Sexo	$\hat{p}_1(\mathbf{x})$	$\hat{p}_2(\mathbf{x})$	$\hat{p}_3(\mathbf{x})$	$\hat{p}_4(\mathbf{x})$
A	M	0,1821	0,2395	0,3985	0,1797
A	F	0,1240	0,2049	0,4283	0,2426
B	M	0,1622	0,1569	0,3249	0,3557
B	F	0,1028	0,1250	0,3250	0,4470

Modelo logito de categorias adjacentes (MLCA)

Tabela 8.16 – Estimativas da chance de ocorrência da categoria de resposta j em relação à categoria $j+1$ para um dado vetor $\mathbf{x}$ de covariáveis

Tratamento	Sexo	$\frac{P(Y=1 \mid \mathbf{x})}{P(Y=2 \mid \mathbf{x})}$	$\frac{P(Y=2 \mid \mathbf{x})}{P(Y=3 \mid \mathbf{x})}$	$\frac{P(Y=3 \mid \mathbf{x})}{P(Y=4 \mid \mathbf{x})}$
A	M	0,76	0,60	2,21
A	F	0,60	0,48	1,76
B	M	1,03	0,48	0,91
B	F	0,82	0,38	0,73

Nota: $\mathbf{x} = (x_1, x_2)$ = vetor de valores de X_1 = tratamento e X_2 = sexo, $j = 1$ remissão total, $j = 2$ remissão parcial, $j = 3$ nenhuma mudança e $j = 4$ progressão da doença.

Modelo logito de categorias adjacentes (MLCA)

- Para quase todas as categorias (Tratamento X Sexo), a chance de Remissão Total relativa a Remissão Parcial é menor que 1. Apenas para o Tratamento B e Masculino ela é aproximadamente 1.
- Quando comparamos Remissão Parcial com Nenhuma Mudança, todas as chances são menores que 1.
- Na comparação de Nenhuma Mudança com Progressão da doença, notamos que somente para o tratamento A essas chances são maiores que 1.
- A estimativa da razão de chances para o tratamento A relativa ao tratamento B, comparando Nenhuma Mudança com Progressão da Mudança 2.4. Isso significa que o Tratamento A parece mais eficiente para conter a progressão da doença.
- Comparar Masculino e Feminino!

Resumo : Regressão Multinomial

- Variáveis nominais: Logito Categoria de Referência *logitoR*.

$$p_j(x) = \frac{\exp\{\beta_{0j} + \beta_j^T x\}}{1 + \exp\{\beta_{0j} + \beta_j^T x\}} \quad j = 1, \ldots, c-1$$

- Variáveis ordinais: Logito Cumulativo *logitoC*

$$p_j(x) = \theta_j(x) - \theta_{j-1}(x) \quad j = 1, \ldots, c;$$

$$\theta_j(x) = \frac{\exp\{\beta_{0j} + \beta_j^T x\}}{1 + \exp\{\beta_{0j} + \beta_j^T x\}} \quad j = 1, \ldots, c-1;$$

$\theta_0(x) = 0$ e $\theta_c(x) = 1$

Resumo : Regressão Multinomial

- Variáveis ordinais: Logito Categoria Adjacente *logitoA*

$$p_j(x) = \frac{\exp\left\{\sum_{l=j}^{c-1} \beta_{0l} + \beta_l^T x\right\}}{\left[1 + \sum_{k=1}^{c-1} \exp\left\{\sum_{l=k}^{c-1} \beta_{0l} + \beta_l^T x\right\}\right]} \quad j = 1, \ldots, c-1$$

$$p_c(x) = 1 - \sum_{j=1}^{c-1} p_j(x)$$

- Para todos os tipos de logitos, podemos definir modelos de chances proporcionais ou proporcionais parciais.
- Uma referência de aplicação: Paulino, Silva and Branco. *A fully Bayesian parametric approach for cytogenetic dosimetry*, BJPS, 2013.