Análise de Dados Categorizados - Aula 16

Márcia D Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística
www.ime.usp.br/ mbranco - sala 295-A -

Regressão multinomial: tabelas $r \times c$ com $c > 2$

- Vamos considerar que a variável resposta tem c categorias disjuntas. Usamos a seguinte notação

$$y_{lj} = \begin{cases} 1\ , & \text{se l-ésima resposta esta na j-ésima categoria} \\ 0\ , & \text{caso contrário} \end{cases}$$

- Como as categorias devem ser disjuntas temos $\sum_{j=1}^{c} y_{lj} = 1$.
- Considere $p_j(x_l) = P(Y_{lj} = 1)$ a probabilidade associada a j-ésima categoria para a l-ésima resposta.
- A seguir apresentamos diversas maneiras de modelar essas probabilidades, sempre tomando como base a idéia de regressão linear nos logitos.

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

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

1. **Chances proporcionais:** Considera que os efeitos das covariáveis é o mesmo para todos os logitos.

$$logitoC_j = \beta_{0j} + \beta^T x \quad , \; j = 1, 2, \dots c - 1$$

Em que
$\beta^T = (\beta_1, \beta_2, \dots, \beta_q)$ e $x^T = (x_1, x_2, \dots, x_q)$.

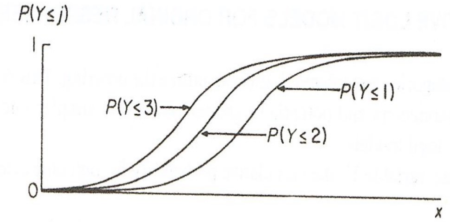

Figure 6.2 Depiction of cumulative probabilities in the cumulative logit model.

- A vantagem da suposição de chances proporcionais é que temos menos parâmetros para estimar e isso aumenta o poder dos testes.
- Como estratégia de ajuste podemos primeiro considerar um modelo mais geral, com chances não proporcionais e testar a igualdade dos coeficientes $H_0 : \beta_j = \beta$ para $j = 1, 2, \ldots c - 1$.
- O teste da razão de verossimilhança é preferido em relação ao teste de Wald.
- Se a variável explicativa também for ordinal [X e Y ordinais] podemos aumentar ainda mais o poder e reduzir a quantidade de parâmetros, substituindo x pelos escores .

Exemplo 1: Os dados na tabela abaixo (Agresti, pag. 172) referem-se a um estudo para investigar a relação entre Felicidade (Y) e Renda da família (x). Ajustou-se um MLC com chances proporcionais, considerando-se escores para x.

Renda	Felicidade		
	Pouco Feliz	Feliz	Muito Feliz
Abaixo da média (1)	37	90	45
Na média (2)	25	93	56
Acima da média (3)	06	18	13

Modelo de logitos cumulativos (MLC)

```
> Happy <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Happy.dat",
+                       header=TRUE)
> Happy # data for sampled black Americans
  income y1 y2 y3
1      1 37 90 45
2      2 25 93 56
3      3  6 18 13
> library(VGAM)
> fit <- vglm(cbind(y1,y2,y3)~ income, family=cumulative(parallel=TRUE),
+             data=Happy)
       Estimate  Std. Error  z value  Pr(>|z|)  # not showing the two
income  -0.2668      0.1510   -1.768    0.0771  # intercept estimates
---
> fit0 <- vglm(cbind(y1,y2,y3)~ 1, family=cumulative, data=Happy) # null model
> lrtest(fit, fit0)
Model 1: cbind(y1, y2, y3) ~ income # treating happiness and income as ordinal
Model 2: cbind(y1, y2, y3) ~ 1
  #Df   LogLik  Df  Chisq  Pr(>Chisq)
1   3  -14.566
2   4  -16.121   1  3.109     0.07786 .
```

- Modelo ajustado:

$$\log\left[\frac{P(Y \leq j \mid x)}{P(Y > j \mid x)}\right] = \beta_{0j} + \beta_1 x \quad , j = 1, 2.$$

- Estimativa de máxima verossimilhança para β_1 é $\hat{\beta}_1 = -0.267$.
- O teste de razão de verossimilhança para $H_0 : \beta = 0$ resulta numa estatística com valor 3.11. Considerando uma qui-quadrado com 1 grau de liberdade, temos valor-P =0.078. Rejeita-se H_0. Indicando que a covariável renda é significante.
- Cuidado na interpretação de $\hat{\beta}_1 = -0.267$!!!

- As chances estimadas são

$$\left[\frac{P(PoucoFeliz \mid x)}{P(Feliz + MuitoFeliz \mid x)}\right] \quad \text{e} \quad \left[\frac{P(PoucoFeliz + Feliz \mid x)}{P(MuitoFeliz \mid x)}\right]$$

- Um valor negativo para β_1 indica que essas chances diminuem com o crescimento da renda. Portanto, a probabilidade da família ser pouco feliz é menor quando a renda aumenta.
- Alternativamente, podemos dizer que a probabilidade de muito feliz aumenta com o aumento da renda.
- Para o mesmo conjunto de dados foi também ajustado o MLCR e os resultados apresentado a seguir.

Exemplo1: MLCR

```
> fit2 <- vglm(cbind(y1,y2,y3) ~ factor(income), family=multinomial,data=Happy
> fit0 <- vglm(cbind(y1,y2,y3)~ 1, family=multinomial, data=Happy)
> # baseline cat. logit null model equivalent to cumulative logit null model
> lrtest(fit2, fit0)
Model 1: cbind(y1, y2, y3) ~ factor(income) # treats variables as nominal-scal
Model 2: cbind(y1, y2, y3) ~ 1
  #Df  LogLik  Df  Chisq  Pr(>Chisq)
1   0  -14.058                        # fit2 model is saturated
2   4  -16.121  4  4.1258    0.3892
```

Exemplo1: MLCR

- O modelo proposto considera: $x_1 = 1$ se renda baixa e zero caso contrário; $x_2 = 1$ se renda média e zero caso contrário.

$$logitoR_j = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 \quad j = 1, 2.$$

- Em que

$$logitoR_1 = \log\left[\frac{P(PoucoFeliz \mid x)}{P(MuitoFeliz \mid x)}\right]$$

$$logitoR_2 = \log\left[\frac{P(Feliz \mid x)}{P(MuitoFeliz \mid x)}\right]$$

Exemplo1: MLCR vs MLC

- Na saída do R temos o resultado do teste de razão de verossimilhnança para testar $H_0 : \beta_{j1} = \beta_{j2} = 0, \quad j = 1, 2.$
- Valor obtido é $D_0 - D_1 = 4.1258$ com 4 graus de liberdades e valor-P $= 0.3892$. Não rejeita-se H_0.
- Não podemos conlcuir pela significância do efeito de renda na felicidade da família.
- A desvantagem deste segundo ajuste é que o teste é menor poderoso. (O que é o poder de um TH?)
- A vantagem desse segundo modelo é que tem menos suposições .

MLC : 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 centrada em 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-se

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

MLC : Estrutura de variáveis latentes

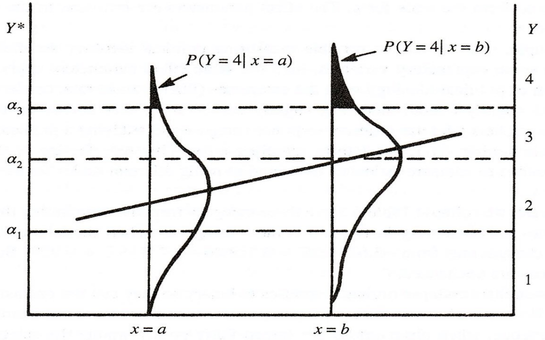

Figure 6.4 Ordinal measurement and underlying regression model for a latent variable. The left vertical axis shows values for the latent variable. The right vertical axis shows values for the observed ordinal response, its category being determined by the three cutpoints on the latent variable scale. 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.

MLC : 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.

- 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.

2. Modelo de chances proporcionais parciais: Considera que para algumas covariáveis as chances são proporcionais e outras tem o coeficiente variando em cada logito.

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

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.

Observação: O modelo de chances proporcionais ou proporcionais parciais também podem ser definidos para outros tipos de logitos.

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

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

- O modelo de regressão é dado por

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

- 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

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

Modelo logito de categorias adjacentes (MLCA)

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:

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

Reescrevendo:

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

Modelo logito de categorias adjacentes (MLCA)

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)

TAREFA: Estudar o exemplo apresentado na seção 8.4.2 do livro da S.Giolo (pag. 190-195).