Análise de Dados Categóricos

Modelo de Regressão Politômica

Enrico A. Colosimo/UFMG
http://www.est.ufmg.br/~enricoc/

Departamento de Estatística
Universidade Federal de Minas Gerais

Extensões do Modelo de Regressão Logística Binário

1. Regressão Politômica: resposta com mais de duas categorias
   - Nominal;
   - Ordinal.
2. Medidas Repetidas ou Longitudinais.
   - Desenho Pareado.
   - Exemplo: Mais de um dente por paciente.
   - Exemplo: Acompanhamento longitudinal.
   - Modelos: Marginal e de Efeitos aleatórios.

Modelo de Regressão Logística para Respostas Multicategóricas

1. Resposta Nominal
   - Exemplos: escolha da marca de um produto (A, B, C e D), escolha do meio de locomoção (ônibus, avião, trem), etc.
   - Modelagem padrão: utilizar logit para pares de categorias.
2. Resposta Ordinal
   - Exemplos: qualidade de vida (excelente, boa, razoável, ruim), opção política (muito liberal, liberal, moderada, conservadora), etc.
   - Modelagem padrão: utilizar logit para probabilidades cumulativas.
3. Em ambos casos, o objetivo é modelar a dependência de π_{ij} para o i-ésimo indivíduo na j-ésima categoria,
   $$\pi_{ij} = P(Y_i = j), \quad j = 1, \ldots, r,$$
   em função de um conjunto de covariáveis x (categóricas ou contínuas).
4. Os modelos tratam as respostas y, para fixo X, como multinomiais Y_{ij} para $j = 1, \ldots, r$.

1-Regressão Logística Politômica Nominal (Modelos de logitos generalizados)

$Y : 1, 2, ..., r, \quad r > 2$

Vamos tomar a categoria r como referência

$logit\pi_j(x) = x\beta_j, \quad j = 1, 2, ..., r$

em que,

$logit\pi_1(x) = log(\frac{P[y=1|x]}{P[y=r|x]}) = log\frac{\pi_1(x)}{\pi_r(x)} = \beta_{01} + \beta_{11}X_1 + ... + \beta_{p1}X_p$

Isto significa que temos $(p+1) \times (r-1)$ parâmetros

$\pi_1(x) = p[y = 1|x]$

$\pi_r(x) = P[y = r|x] = 1 - \sum_{j=1}^{r-1} \pi_j(x)$

Modelo de logitos generalizados

$$\pi_1(x) = P[y = 1|x] = \frac{e^{X'\beta_1}}{1+\sum_{j=1}^{r-1} e^{X'\beta_j}}$$

$$\pi_{r-1}(x) = P[y = r-1|x] = \frac{e^{X'\beta_{r-1}}}{1+\sum_{j=1}^{r-1} e^{X'\beta_j}}$$

$$\pi_r(x) = 1 - \sum_{j=1}^{r-1} \pi_j(x) = 1 - \frac{\sum_{j=1}^{r-1} e^{X'\beta_j}}{1+\sum_{j=1}^{r-1} e^{X'\beta_j}} = \frac{1}{1+\sum_{j=1}^{r-1} e^{X'\beta_j}}$$

$$logit\pi_1(x) = X'\beta_1 \quad \ldots \quad logit\pi_{r-1}(x) = X'\beta_{r-1}$$

Isto é equivalente a termos $r-1$ regressões logísticas binárias.

Tudo que foi visto para o caso r=2 (binária) vale para este politômico.

Exemplo: Profa. Juliana - Traumatismo Odontológico

- Resposta: Tipo de Reabsorção
  - 0- ausente,
  - 1- substituição e
  - 2-inflamtória
- Objetivo: avaliar o efeito das variáveis na resposta:
  - 1- Idade (usei os pontos de corte de 11 e 16 anos),
  - 2- SAT - Uso de Antibiótico,
  - 3- TempoCont1 (tempo entre o reimplante e o TER),
  - 4- MeioG (1- úmido, 2- leite e 3- seco) e
  - 5-7 Genéticas (RANKL, OPG e IL10)

Exemplo: Profa. Juliana - Traumatismo Odontológico

Descritiva

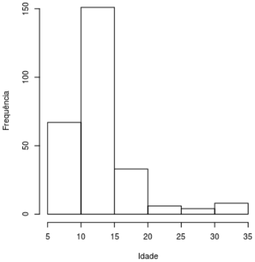
- Estudo com 269 pacientes.

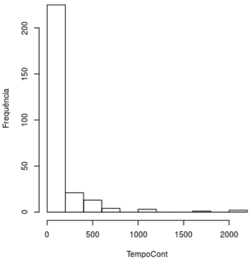
	n	%	Total
Tipo de absorção			268
Ausente	56	20,9%	
Substituição	187	69,8%	
Inflamatória	25	9,3%	
NA	1		
Meio			269
Úmido	104	38,7%	
Leite	60	22,3%	
Seco	105	39,0%	
Antibiótico			175
Sim	46	26,3%	
Não	129	73,7%	
NA	94		

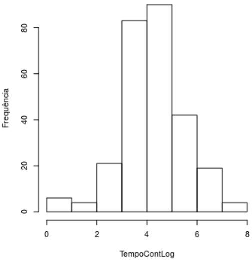
	n	%	Total
Rank			265
1	154	58,1%	
2	85	32,1%	
3	26	9,8%	
OPG			266
1	32	12,0%	
2	108	40,6%	
3	126	47,4%	
NA	3		
IL10			269
1	127	47,2%	
2	110	40,9%	
3	32	11,9%	

Exemplo: Profa. Juliana - Traumatismo Odontológico

Descritiva







Ajuste do modelo - library VGAM

- Ajuste com a variável antibiótico

```
> fit1 <- vglm(Tipo1~SAT,family="multinomial",data=dados) #
> summary(fit1)

Call:
vglm(formula = Tipo1 ~ SAT, family = "multinomial", data = dados)


Pearson residuals:
                         Min      1Q  Median      3Q    Max
log(mu[,1]/mu[,3]) -1.8073 -0.8974  0.6155  0.6671 0.6671
log(mu[,2]/mu[,3]) -0.9858 -0.1551 -0.1551 -0.1226 3.0782

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  1.38217    0.78715   1.756   0.0791 .
(Intercept):2 -1.08286    1.24795  -0.868   0.3856
SAT:1         -0.08289    0.44205  -0.188   0.8513
SAT:2          0.27193    0.68860   0.395   0.6929
```

Ajuste do modelo - library VGAM

- Ajuste com as variáveis: IdadeG11, OPGGen e TempoContLog

```
> fit11<-vglm(Tipo1~IdadeG11 + factor(OPGGen) + log(TempoCont1) -
 factor(MeioG) - factor(RanklGen), family = "multinomial", data=dados)
> summary(fit11)

Call:
vglm(formula = Tipo1 ~ IdadeG11 + factor(OPGGen) + log(TempoCont1), family = "multinomial",
    data = dados)


Pearson residuals:
                      Min      1Q  Median       3Q   Max
log(mu[,1]/mu[,3]) -5.366 -0.7593  0.4121  0.57639 1.554
log(mu[,2]/mu[,3]) -4.044 -0.2440 -0.1218 -0.05447 5.332

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept):1      -2.1840     1.1917  -1.833  0.06684 .
(Intercept):2      -3.1748     1.6597  -1.913  0.05576 .
IdadeG11:1         -0.6680     0.4195  -1.592  0.11128
IdadeG11:2         -1.7572     0.5936  -2.960  0.00308 **
factor(OPGGen)2:1  -0.1229     0.6824  -0.180  0.85704
factor(OPGGen)2:2  -1.8592     0.8648  -2.150  0.03157 *
factor(OPGGen)3:1  -0.7652     0.6548  -1.168  0.24261
factor(OPGGen)3:2  -2.1670     0.8149  -2.659  0.00783 **
log(TempoCont1):1   1.2945     0.2252   5.749 9.00e-09 ***
log(TempoCont1):2   1.7133     0.3010   5.693 1.25e-08 ***
```

Interpretação

- A interpretação é feita comparando com a categoria de referência, ou seja,
  - Logit 1: Substituição/Ausente
  - Logit 2: Inflamatória/Ausente
- Para a variável IdadeG11:
- a chance de reabsorção por substituição para os pacientes com idade maior que 11 anos é $0,2(exp(-1,757))$ (IC; 95%; 0,05; 0,55) vezes a chance daqueles com idade menor que 11 anos (comparado com ausência de reabsorção).

2 Regressão Logística Politômica Ordinal (Modelos de logitos cumulativos)

Como levar em consideração na modelagem o fato das categorias serem ordenadas?

$y(melhora)$: $nenhuma(1), alguma(2), muita(3)$

$\pi(x) = P[y = j|x], j = 1, 2, 3.$

- Logitos cumulativos

  $logit(\pi_{1.2-3}) = log[\frac{\pi_1(x)}{\pi_2(x)+\pi_3(x)}]$

  compara nenhuma melhora com pelo menos alguma melhora

  $logit_1(x) = log[\frac{P(y=1|x)}{P(y>1/x)}]$

  $logit_2(x) = log[\frac{P(y\leq 2|x)}{P(y=3|x)}]$

  Geral:

  $logit_k(x) = log[\frac{P(y\leq k|x)}{P(y>k|x)}]$

Modelo de Logitos Acumulados

$P(y \leq k|x) = \frac{e^{X'\beta_k}}{1+e^{X'\beta_k}}$

$P(y < k|x) = 1 - P(y \geq k|x) = \frac{1}{1+e^{X'\beta_k}}$

$logit_k(x) = X'\beta_k$

Isto significa que cada β_k está comparando as primeiras k's categorias com as restantes $r - k + 1$ categorias

- Exemplo $r = 3$

  β_1 compara $log\frac{\pi_1}{\pi_2+\pi_3}$ 1 vs 2-3

  β_2 compara $log\frac{\pi_1+\pi_2}{\pi_3}$ 1-2 vs 3

Modelo de Chances Proporcionais

- Assumimos que

$$\beta_k = \beta$$

  para $k = 1, ..., r - 1$.

- Ou seja, o mesmo efeito de X para todas as comparações categorias acumuladas.

Exemplo (Variável resposta politômica ordinal)

Uma amostra de 84 pacientes com dores de artrite foi submetida ao um tratamento experimental. O objetivo do estudo é avaliar para a resposta grau de melhora (acentuada, alguma ou nenhuma) a possível associação com o tratamento e sexo.

Variáveis:

- Sexo

  Feminino; Masculino

- Tratamento

  A; Placebo

- Graus de melhora

  Melhora acentuada; Alguma melhora; Nenhuma melhora

Exemplo (Variável resposta politômica ordinal)

A tabela a seguir mostra os dados associados ao estudo.

		Grau de Melhora			
Sexo	Tratamento	Acentuada	Alguma	Nenhuma	Total
F	A	16	5	6	27
F	Placebo	6	7	19	32
M	A	5	2	7	14
M	Placebo	1	0	10	11

Um modelo adequado para analisar esses dados é o de logitos cumulativos.

Exemplo (Variável resposta politômica ordinal)

Defina as quantidades:

π_1: probabilidade de melhora acentuada

π_2: probabilidade de alguma melhora

π_3: probabilidade de nenhuma melhora

$X_1 = 1$: sexo feminino; $X_1 = 2$: sexo masculino

$X_2 = 1$: uso do tratamento A; $X_2 = 2$: placebo

Como existem 3 categorias de respostas ordinais, defini-se dois logitos cumulativos

$$logit(P(Y \leq 1|X) = log(\frac{\pi_1}{\pi_2+\pi_3}) \qquad logit(Y \leq 2|X) = ln(\frac{\pi_1+\pi_2}{\pi_3})$$

Exemplo (Variável resposta politômica ordinal)

Análise da tabela de *deviance*

Modelos	g.l.	Deviance residual	Diferenças de deviances	Diferenças de g.l.
Nulo	6(8-2)	22.59		
X_1	5(8-3)	18.79	3.80	1
$X_2 \| X_1$	4(8-4)	2.71	16.08	1
$X_1 * X_2 \| X_1, X_2$	3(8-5)	2.40	0.31	1

Utiliza-se o teste da razão de verossimilhança para testar o modelo

Exemplo (Variável resposta politômica ordinal)

Resultados

- Iteração de sexo e tratamento não significativa

  ($p = 0.5786, g.l. = 1$)

- Sexo significativo

  ($p = 0.051$)

- Tratamento na presença de sexo significativo

  ($p < 0.0001$)

- Teste de proporcionalidade

  Hipótese nula não é rejeitada

  ($p = 0.39, g.l. = 2$)

Respostas Repetidas ou Longitudinais

1. Exemplos:
    - Medidas Repetidas: Mais de um dente em cada paciente;
    - Medidas Longitudinais: Mecanismo Evacuatório de Récem-Nascidos avaliado em cada mês, no primeiro ano de vida.
2. Dificuldade: Dados não mais independentes.

Respostas Repetidas ou Longitudinais

1. $Y_{ij}, \quad i = 1, \ldots, N;\ j = 1, \ldots, k$: binária, contagem, etc.
2. Modelos Estatísticos
   - Modelos Lineares Generalizados Mistos.
   - Modelos Marginais: GEE