Análise de Dados Categorizados - Aula 13

Márcia D Elia Branco

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

Regressão multinomial

- 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 ideia de regressão linear nos logitos.

Modelo logito de categoria de referência (MLCR)

- Primeiro escolhemos uma categoria de referência, por exemplo, podemos escolher a última como sendo a referência .
- Os logitos de referência são definidos por

$$logitoR_j = \log\left[\frac{p_j(x)}{p_c(x)}\right] \quad j = 1, 2, \ldots, c-1$$

- Para cada logito define-se uma reta de regressão

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

Modelo logito de categoria de referência (MLCR)

Exemplo 1: Os dados na tabela abaixo são referentes a um estudo realizado com crianças para investigar o seu programa de aprendizado preferido e se tal preferência estaria associada com a escola e o período escolar.

Escola	Período	Preferência			Totais
		Individual	Grupo	Sala de aula	
1	Padrão	10	17	26	53
	Integral	5	12	50	67
2	Padrão	21	17	26	64
	Integral	16	12	36	64
3	Padrão	15	15	16	46
	Integral	12	12	20	44

Modelo logito de categoria de referência (MLCR)

- Tabela $3 \times 2 \times 3$. Duas variáveis explicativas e uma variável resposta.
- $c = 3$. Vamos considerar $j = 3$ (sala de aula) como categoria de referência.
- $r = 3 \times 2 = 6$ (linhas na tabela). Modelo probabilístico considerado: produto de seis multinomias.
- As variáveis *dummy* são definidas por: $x_{11} = 1$, se escola 2 e zero, caso contrário ; $x_{12} = 1$, se escola 3 e zero, caso contrário e $x_2 = 1$, se escola padrão e zero, caso contrário.
- São definidos $c - 1 = 2$ logitos de referência e cada um deles associados à uma reta de regressão.

Modelo logito de categoria de referência (MLCR)

$$logitoR_1 = \log\left[\frac{p_1(x)}{p_3(x)}\right] = \beta_{01} + \beta_{11}x_{11} + \beta_{21}x_{12} + \beta_{31}x_2$$

$$logitoR_2 = \log\left[\frac{p_2(x)}{p_3(x)}\right] = \beta_{02} + \beta_{12}x_{11} + \beta_{22}x_{12} + \beta_{32}x_2$$

- Total de parâmetros a estimar no modelo proposto (somente com os efeitos principais, sem interação): 8
- Modelo saturado tem $2 \times 6 = 12$ parâmetros.
- Graus de liberdades associada a estatística $Q_L = \sum d_i^2$ é $g.l. = 12 - 8 = 4$.
- O modelo pode ser ajustado com o uso da função $vglm(cbind(y1, y2, y3)\ x_{11} + x_{12} + x_2, family = multinomial)$ da *library(VGAM)*.

Modelo logito de categoria de referência (MLCR)

Para avaliar a qualidade do ajuste do modelo e testar os coeficientes regressores usamos as funções desvios que podem ser obtidas por *deviance*(*ajuste*). Resultando em

$D_0 = 30.248(10 g.l.)$; $D_1 = 12.8716(6 g.l.)$ e $D_2 = 1.7776(4 g.l.)$.

Em que:

D_0 é a *deviance* do modelo somente com os interceptos;
D_1 é a *deviance* do modelo com os interceptos e a covariável escola (x_{11} e x_{12}) ;
D_2 é a *deviance* do modelo com os interceptos e as duas covariáveis (escola e período).

Modelo logito de categoria de referência (MLCR)

Modelo	TRV	g.l.	Valor-P	AIC
D_1 - Escola	17.3764	4	0.0016	73.5
D_2 - Período | Escola	11.0940	2	0.0039	66.4
Modelo com interação	1.7776	4	0.7766	72.6

- O modelo escolhido é o D_2.
- A estatística de qualidade de ajuste baseada nos desvios residuais é $Q_L = 1.7776 (valor - P = 0.776, g.l. = 4)$. Portanto, não rejeita-se o modelo ajustado .

Modelo logito de categoria de referência (MLCR)

Os resultados na tabela a seguir (Giolo, pag. 173) são referentes a estimação dos parâmetros do modelo ajustado .

Tabela 8.4 – Estimativas dos parâmetros associados ao modelo selecionado

Parâmetros	Logito 1 individual/sala de aula		Logito 2 grupo/sala de aula	
	Estimativa	Erro-padrão	Estimativa	Erro-padrão
β_{0j} intercepto	−1,9707	0,320	−1,3088	0,259
β_{1j} escola 2	1,0828	0,353	0,1801	0,317
β_{2j} escola 3	1,3147	0,384	0,6556	0,339
β_{3j} período padrão	0,7474	0,282	0,7426	0,270

Nota: $j = 1$ se logito 1 e $j = 2$ se logito 2.

Tabela 8.5 – Expressões e estimativas da chance de ocorrência da categoria de resposta j em relação à categoria r, $p_j(\mathbf{x})/p_r(\mathbf{x})$, $j = 1, 2$ e $r = 3$

Escola	Período	Individual/Sala de aula		Grupo/Sala de aula	
		$p_1(\mathbf{x})/p_3(\mathbf{x})$	Estimativa	$p_2(\mathbf{x})/p_3(\mathbf{x})$	Estimativa
1	Padrão	$\exp(\beta_{01} + \beta_{31})$	0,294	$\exp(\beta_{02} + \beta_{32})$	0,567
1	Integral	$\exp(\beta_{01})$	0,139	$\exp(\beta_{02})$	0,270
2	Padrão	$\exp(\beta_{01} + \beta_{11} + \beta_{31})$	0,869	$\exp(\beta_{02} + \beta_{12} + \beta_{32})$	0,679
2	Integral	$\exp(\beta_{01} + \beta_{11})$	0,411	$\exp(\beta_{02} + \beta_{12})$	0,323
3	Padrão	$\exp(\beta_{01} + \beta_{21} + \beta_{31})$	1,095	$\exp(\beta_{02} + \beta_{22} + \beta_{32})$	1,093
3	Integral	$\exp(\beta_{01} + \beta_{21})$	0,519	$\exp(\beta_{02} + \beta_{22})$	0,520

Nota: $\mathbf{x} = (x_1, x_2)$ corresponde ao vetor de valores de X_1 (escola) e X_2 (período).

Modelo logito de categoria de referência (MLCR)

- Note que as chances de preferência de aprendizado individual relativa à sala de aula são quase todas menores que 1 (exceto uma que é aproximadamente 1). Isso indica que em geral há uma preferência pelo aprendizado em sala de aula.
- Análise similar pode ser feita quando observamos as chances de preferência de aprendizado em grupo relativa à sala de aula.
- A partir dos valores da tabela podemos também obter as chances de aprendizado individual relativa ao grupo, basta dividir os valores da primeira coluna numérica pelos da última coluna resultando estimativas para $p_1(x)/p_2(x)$.

Modelo logito de categoria de referência (MLCR)

- Para o logito 1, a estimativa da razão de chances entre períodos é

$$\hat{OR}_{P/I} = \exp\{\hat{\beta}_{31}\} = 2.11$$

- Logo, a chance de preferência pelo aprendizado individual entre as crianças do período padrão foi aproximadamente 2 vezes a das crianças do período integral.
- Para o logito 2, a estimativa da razão de chances entre períodos é

$$\hat{OR}_{P/I} = \exp\{\hat{\beta}_{32}\} = 2.10$$

- Logo, a chance de preferência pelo aprendizado em grupo entre as crianças do período padrão foi aproximadamente 2 vezes a das crianças do período integral.

Modelo logito de categoria de referência (MLCR)

Comparando agora as escolas.

- Para o logito 1 temos as seguintes razões de chances:

$$\hat{OR}_{2/1} = \exp\{\hat{\beta}_{11}\} = 2.95$$
$$\hat{OR}_{3/1} = \exp\{\hat{\beta}_{21}\} = 3.72$$
$$\hat{OR}_{3/2} = \exp\{\hat{\beta}_{21} - \hat{\beta}_{11}\} = 1.26.$$

- A chance de preferência pelo aprendizado individual entre as crianças da escola 2 foi aproximadamente 3 vezes a das crianças da escola 1. Analogamente para interpretação dos outros valores.
- O mesmo tipo de interpretação pode ser feito para o logito 2 (Exercício!!)

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

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

- 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)

Exemplo 2: Na tabela abaixo apresentamos dados de um ensaio clínico sobre tratamentos para dores de artrite (Giolo, pag. 183).

Sexo	Tratamento	Grau de melhora			Totais
		Acentuada	Alguma	Nenhuma	
F	A	16	5	6	27
	Placebo	6	7	19	32
M	A	5	2	7	14
	Placebo	1	1	9	11

Modelo de logitos cumulativos (MLC)

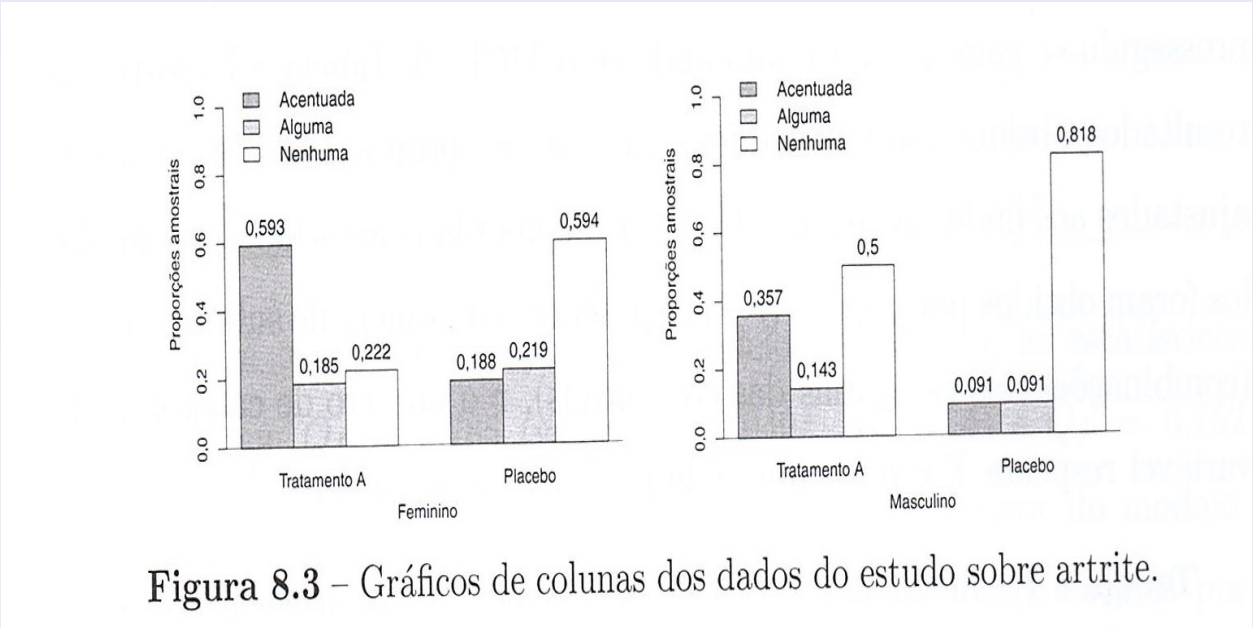

Figura 8.3 – Gráficos de colunas dos dados do estudo sobre artrite.

- Notamos que para as mulheres a maioria teve uma melhora acentuada no grupo que recebeu o tratamento; enquanto que para o grupo placebo, a maioria não obteve melhora. O que parece indicar um efeito positivo do tratamento.
- Para os homens este efeito não é tão evidente.

Modelo de logitos cumulativos (MLC)

O número de categorias é $c = 3$ portanto temos 2 logitos cumulativos:

$$logitoC_1 = \log\left[\frac{\theta_1(x)}{1-\theta_1(x)}\right] = \log\left[\frac{p_1(x)}{p_2(x)+p_3(x)}\right]$$
$$logitoC_2 = \log\left[\frac{\theta_2(x)}{1-\theta_2(x)}\right] = \log\left[\frac{p_1(x)+p_2(x)}{p_3(x)}\right]$$

Interpretação:

$\exp\{logitoC_1\}$ é a chance de melhora acentuada (relativa à alguma ou nenhuma) .

$\exp\{logitoC_2\}$ é a chance de melhora (relativa à nenhuma melhora).

Modelo de logitos cumulativos (MLC)

Modelo de Regressão somente com os efeitos fixos:

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

Em que $x_1 = 1$ para feminino e $x_2 = 1$ para tratamento A.

Para o ajuste do modelo usamos a função
$vglm(cbind(y1, y2, y3)\ x_1 + x_2, family = cumulative(parallel = FALSE))$

Após o ajuste podemos obter as funções desvios associada ao modelo e usá-las para realizar testes de hipóteses.

Modelo de logitos cumulativos (MLC)

Vamos testar a hipótese $H_0 : \beta_{11} = \beta_{12}, \beta_{21} = \beta_{22}$.

A estatística do teste da razão de verossimilhança é $TRV = 0.77$ com valor-P $= 0.68$ com $g.l. = 2$. Como o teste não rejeita H_0 vamos agora considerar um modelo mais simples denominado modelo de chances proporcionais. Neste modelo assumimos que os parâmetros regressores são os mesmos para cada um dos logitos. Assim

$$logitoC_j = \beta_{0j} + \beta_1 x_1 + \beta_2 x_2 \quad j = 1, 2$$

Modelo de logitos cumulativos (MLC)

O modelo ajustado é dado por

$$logitoC_j = \hat{\beta}_{0j} + 1.1121x_1 + 1.6738x_2 \quad j = 1, 2$$

Com $\hat{\beta}_{01} = -2.4234$ e $\hat{\beta}_{02} = -1.5332$.

As probabilidades de sucesso de cada categoria são estimadas por:

$$\hat{p}_j(x) = \hat{\theta}_j(x) - \hat{\theta}_{j-1}(x) \quad j = 1, 2, 3$$

Em que $\hat{\theta}_0 = 0$ e $\hat{\theta}_3 = 1$ e

$$\hat{\theta}_j(x) = \frac{\exp\{\hat{\beta}_{0j} + 1.1121x_1 + 1.6738x_2\}}{1 + \exp\{\hat{\beta}_{0j} + 1.1121x_1 + 1.6738x_2\}} \quad j = 1, 2$$

Modelo de logitos cumulativos (MLC)

Tabela 8.10 – Probabilidades cumulativas $\theta_j(\mathbf{x})$ e não cumulativas $p_j(\mathbf{x})$ preditas a partir do modelo selecionado para análise dos dados de artrite

Sexo	Tratamento	$\hat{\theta}_1(\mathbf{x})$	$\hat{\theta}_2(\mathbf{x})$	$\hat{p}_1(\mathbf{x})$	$\hat{p}_2(\mathbf{x})$	$\hat{p}_3(\mathbf{x})$
F	A	0,5896	0,7777	0,5896	0,1881	0,2223
F	Placebo	0,2123	0,3963	0,2123	0,1840	0,6037
M	A	0,3209	0,5351	0,3209	0,2142	0,4649
M	Placebo	0,0814	0,1775	0,0814	0,0961	0,8225

Modelo de logitos cumulativos (MLC)

Tabela 8.11 – Expressões para as chances decorrentes do modelo ajustado

Sexo	Tratamento	Melhora acentuada *versus* alguma ou nenhuma	Melhora acentuada ou alguma melhora *versus* nenhuma
F	A	$\exp(\beta_{01}+\beta_1+\beta_2)$	$\exp(\beta_{02}+\beta_1+\beta_2)$
F	Placebo	$\exp(\beta_{01}+\beta_1)$	$\exp(\beta_{02}+\beta_1)$
M	A	$\exp(\beta_{01}+\beta_2)$	$\exp(\beta_{02}+\beta_2)$
M	Placebo	$\exp(\beta_{01})$	$\exp(\beta_{02})$

Modelo de logitos cumulativos (MLC)

- A chance de melhora acentuada para os indivíduos que receberam o tratamento A é $\exp\{\hat{\beta}_2\} = 5.33$ vezes a dos receberam placebo.
- A chance de melhora para indivíduos que receberam o tratamento A é 5.33 vezes a dos que receberam placebo.
- A chance de melhora para as mulheres é $\exp\{\hat{\beta}_1\} \approx 3$ vezes a dos homens.
- A chance de melhora acentuada para as mulheres é aproximadamente 3 vezes a dos homens.

Modelo de logitos cumulativos (MLC)

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

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

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

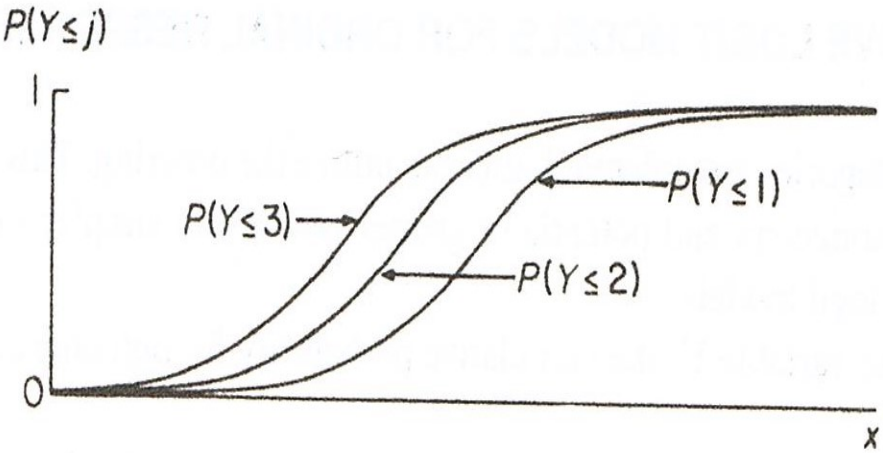

Figure 6.2 Depiction of cumulative probabilities in the cumulative logit model.

Modelo de logitos cumulativos (MLC)

- 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 3: 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

MLC com chances proporcionais

```
> 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 .
```

MLC com chances proporcionais

- 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 $H_0 : \beta_1 = 0$ 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 $\hat{\beta}_1 = -0.267$!!!

MLC com chances proporcionais

- 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.

ML Categoria de Referência

```
> 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
```

- 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

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

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

MLCR vs MLC proporcional

- Na saída do R temos o resultado do teste de razão de verossimilhança 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 concluir pela significância do efeito de renda na felicidade da família.
- O teste usando os logitos de referência é menos poderoso que o considerando os logitos acumulados.
- A vantagem do uso do MLCR é que tem menos suposições.