

MAE0560 – Análise de dados categorizados – 2019 -

Lista 6

1) Um estudo foi realizado para investigar se a felicidade de um casal esta associada com a renda familiar. A variável felicidade foi categorizada em 3 níveis: infeliz, feliz e muito feliz. A variável explicativa renda foi categorizada em: baixa, média e alta. Ajustou-se o Modelo Logístico de Categoria de Referência (MLCR) e os resultados são apresentados a seguir.

Call:

```
vglm(formula = cbind(infeliz, feliz, mfeliz) ~ renda, family = multinomial)
```

Pearson residuals:

	$\log(\mu[,1]/\mu[,3])$	$\log(\mu[,2]/\mu[,3])$
1	0.6488	-0.6403
2	-0.9328	0.8400
3	0.6659	-0.5487

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept):1	-2.55518	0.72560	-3.521	0.000429 ***
(Intercept):2	-0.35129	0.26837	-1.309	0.190554
renda:baixa	-0.22751	0.34120	-0.667	0.504907
renda:media	-0.09615	0.12202	-0.788	0.430694

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: $\log(\mu[,1]/\mu[,3])$, $\log(\mu[,2]/\mu[,3])$

Residual deviance: 3.1909 on 2 degrees of freedom

a) Escreva o modelo ajustado, descrevendo todas as variáveis. Este modelo considera a suposição de chances proporcionais? Justifique.

b) No ajuste do modelo nulo (sem a variável renda) obteve-se um valor de *deviance* de 4.135. Qual o valor da estatística da razão de verossimilhança (TRV)? Quando graus de liberdade? Usando essa estatística, realize o teste de hipótese apropriado e conclua.

c) Com base nos resultados, determine a chance estimada do casal ser *muito feliz* em relação a ser *infeliz*.

2) Um modelo foi ajustado para prever as preferências para presidência nos Estados Unidos. Três candidatos foram considerados: Democrata, Republicano e Independente. Usando-se a renda anual do eleitor como variável explicativa (x) e o MLCR, obteve-se os seguintes ajustes:

$$\text{Logito1} = \ln (P(Y=D | x) / P(Y=I | x)) = 3.3 - 0.2 x$$

$$\text{Logito2} = \ln (P(Y=R | x) / P(Y=I | x)) = 1.0 + 0.3 x$$

a) Obtenha a equação para a chance do candidato Republicano em relação ao Democrata.

b) Para quais valores de renda (x) ocorre que o valor predito $P(Y=R | x)$ é maior que o valor predito de $P(Y=D | x)$?

c) Escreva a equação de predição de $P(Y=I | x)$.

3) O modelo de logitos cumulativos foi ajustado para relacionar a ideologia política do indivíduo (variando de muito liberal a muito conservadora) com a preferência religiosa. Usando-se variáveis indicadoras para variável explicativa preferência religiosa (R1, R2, R3, R4), obteve-se as seguintes retas ajustadas:

$$\text{Logito1} = -1.03 - 1.27 X_1 - 1.22 X_2 - 0.44 X_3$$

$$\text{Logito2} = -0.13 - 1.27 X_1 - 1.22 X_2 - 0.44 X_3$$

$$\text{Logito3} = 1.57 - 1.27 X_1 - 1.22 X_2 - 0.44 X_3$$

$$\text{Logito4} = 2.41 - 1.27 X_1 - 1.22 X_2 - 0.44 X_3$$

- Quantas categorias tem a variável resposta Y? O modelo está assumindo chances proporcionais?
- Qual dos grupos religiosos é o mais liberal? Qual o mais conservador?
- Use a razão de chances estimada para comparar a ideologia política dos grupos R1 e R2.

4) Obtenha as probabilidades $p_j(x)$, $j=1,2,\dots,r$ para os modelos:

- logitos de categoria de referência;
- logitos de razão contínua.

5) Analise a saída do R a seguir. Você não acha que algo estranho está ocorrendo? Saberá identificar? Os dados usados para o ajuste são: $x=c(10,20,30,40,60,70,80,90)$ e $y=c(0,0,0,0,1,1,1,1)$

```
Call:
glm(formula = y ~ x, family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.045e-05	-2.110e-08	0.000e+00	2.110e-08	1.045e-05

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-118.158	296046.187	0	1
x	2.363	5805.939	0	1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1.1090e+01 on 7 degrees of freedom
Residual deviance: 2.1827e-10 on 6 degrees of freedom
AIC: 4
```

Number of Fisher Scoring iterations: 25

6) Exercício 1 do capítulo 8 do livro da S.Giolo.

7) Exercício 2 do capítulo 8 do livro da S.Giolo.