

MAE5763 - Modelos Lineares Generalizados

2^o semestre 2020

Prof. Gilberto A. Paula

2^a Lista de Exercícios

1. Supor o modelo linear simples

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

em que $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$. Mostre a equivalência entre as estatísticas ξ_{RV} , ξ_W e ξ_{SR} para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$, supondo σ^2 conhecido. Sugestão: expressar as três estatísticas em função de $\hat{\beta}$.

2. Supor $Y_{ij}|x_j \stackrel{\text{ind}}{\sim} \text{FE}(\mu_{ij}, \phi)$ em que $g(\mu_{ij}) = \alpha_i + \beta x_j$, para $i = 1, 2$ e $j = 1, \dots, r$. Mostrar que a estatística do teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \frac{\hat{\phi}^0 [\sum_{j=1}^r x_j (y_{1j} - \bar{y}_1) + \sum_{j=1}^r x_j (y_{2j} - \bar{y}_2)]^2}{(\hat{\omega}_1^0 + \hat{\omega}_2^0) \sum_{j=1}^r (x_j - \bar{x})^2}.$$

Qual a distribuição nula assintótica da estatística do teste?

3. Supor $Y_i \stackrel{\text{iid}}{\sim} \text{NI}(\mu_i, \phi)$, para $i = 1, \dots, n$. Mostre que a estatística da razão de verossimilhanças para testar $H_0 : \phi = 1$ contra $H_1 : \phi \neq 1$ pode ser expressa na forma

$$\xi_{RV} = n(\hat{\phi}^{-1} - 1) + n \log(\hat{\phi}),$$

em que $\hat{\phi} = n/D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é a estimativa de máxima verossimilhança de ϕ . Qual a distribuição nula assintótica da estatística do teste?

4. Supor que $Y_{ij} \stackrel{\text{ind}}{\sim} P(\mu_{ij})$, para $i = 1, \dots, r$ e $j = 1, \dots, c$, com parte sistemática dada por

$$\log(\mu_{ij}) = \alpha + \beta_i + \gamma_j + \delta_{ij},$$

em que $\beta_1 = \gamma_1 = 0$, $\delta_{11} = \delta_{12} = \delta_{21} = 0$ e $\delta_{22} = \delta$. Quando há ausência de interação ($\delta = 0$), os β_i 's referem-se aos efeitos do fator A e os γ_j 's aos efeitos do fator B. Defina um modelo multinomial equivalente e mostre que a representação acima corresponde, sob a ausência de interação, à independência (no sentido probabilístico) entre os fatores A e B.

5. $Y_{ij} | \mathbf{z}_i \stackrel{\text{ind}}{\sim} \text{FE}(\mu, \phi_i)$, para $i = 1, 2$ e $j = 1, \dots, m$, em que $\log(\phi_1) = \alpha - \beta$ e $\log(\phi_2) = \alpha + \beta$. Mostre que a estatística do teste de escore para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = -\frac{m}{2d''(\hat{\phi}_0)}(\hat{t}_2 - \hat{t}_1)^2.$$

Qual a distribuição nula assintótica da estatística do teste? Particularize para o caso normal.

6. No arquivo **bioChemists** do pacote **pscl** do R são descritas as seguintes variáveis observadas numa amostra de 915 doutores formados na área de Bioquímica: (i) **art** (número de artigos publicados nos últimos 3 anos pelo doutor), (ii) **fem** (fator indicando gênero, masculino ou feminino), (iii) **mar** (fator indicando o estado civil, casado ou solteiro), (iv) **kids5** (número de crianças com até 5 anos), (v) **phd** (prestígio do departamento onde o aluno fez o doutorado) e (vi) **ment** (número de artigos publicados pelo orientador). As variáveis **fem** e **mar** são fatores, a variável **phd** é contínua e as variáveis **kids5** e **ment** são discretas.

O objetivo do estudo é explicar o número médio de artigos publicados nos últimos 3 anos pelo doutor dadas as demais variáveis como explicativas. Para disponibilizar o arquivo faça o seguinte:

```
require(gamlss)
require(pscl)
summary(bioChemists)
```

```
attach(bioChemists).
```

Fazer inicialmente uma análise descritiva dos dados, por exemplo tabelas de contingência entre o número de artigos publicados e os fatores gênero e estado civil, boxplots robustos das variáveis quantitativas e diagramas de dispersão entre `art` e as variáveis quantitativas. Ajustar modelos com resposta Poisson e resposta binomial negativa apenas com efeitos principais no `gamlss`:

```
fit1.bio = gamlss(art ~ ., family=PO, data=bioChemists)
```

```
fit2.bio = gamlss(art ~ ., family=NBI, data=bioChemists).
```

Usar o comando `stepGAIC` para selecionar um submodelo em cada caso. Deixar no modelo final apenas variáveis significativas ao nível de 10%. Comparar os 2 ajustes através de AIC e análise de resíduos. Para o modelo selecionado verificar se é possível fazer um ajuste também em zero através de uma das famílias `family=ZAP` ou `family=ZANBI`.

Selecionar os efeitos principais para o componente da média e para o componente da propabilidade de zero ao nível de 10%. Fazer análise de resíduos e interpretar os coeficientes do modelo final através de razão de médias e razão de chances.

7. No arquivo **rent** do `gamlss` são descritas 9 variáveis observadas numa amostra aleatória de 1967 unidades habitacionais da cidade de Munich em 1993. Para fins de análise iremos considerar as seguintes variáveis: (i) `R` (valor mensal líquido do alugel em DM), (ii) `F1` (área útil em m²), (iii) `A` (ano da construção), (iv) `H` (variável binária referente à existência de aquecimento central, 0: sim, 1: não) e (v) `loc` (qualidade da localização do imóvel, 1: abaixo da média, 2: na média e 3: acima da média). O arquivo pode ser disponibilizado diretamente no `gamlss` através do comando

```
attach(rent).
```

É preciso informar que a variável `loc` é categórica

```
loc=factor(loc).
```

Fazer inicialmente uma análise descritiva dos dados, tais como densidade da variável resposta, boxplots e diagramas de dispersão entre as

variáveis explicativas contínuas e a variável resposta. Procure selecionar um submodelo gama duplo no `gamlss` com ligação logarítmica para explicar o valor médio mensal do aluguel e para explicar o coeficiente de variação através do comando `stepGAIC`. Deixar no modelo final apenas variáveis significativas ao nível de 10%. Fazer análises de resíduos e interpretar os coeficientes estimados através de razão de médias. A variável explicativa **A** é considerada contínua.

Reajustar o modelo gama duplo selecionado acima comparando o ajuste com a variável explicativa **A** entrando no modelo através de um spline cúbico com 5 graus de liberdade efetivos `sc(A,df=3)`. Compare os ajustes através do AIC e gráficos de resíduos e verifique se houve mudanças importantes nas estimativas dos demais coeficientes.