

MAE5763 - Modelos Lineares Generalizados

2º semestre 2020

Prof. Gilberto A. Paula

3ª Lista de Exercícios

1. Considere o banco de dados **polypharm** da biblioteca **aplore3** do R, em que uma amostra $n = 500$ pacientes é analisada com relação ao uso de medicamentos para o tratamento de doenças mentais. O objetivo principal do estudo é relacionar 11 variáveis explicativas observadas em 7 anos (2202 a 2008) com o uso de medicamentos representada pela variável resposta **polypharmacy** (=0 uso de no máximo 3 medicamentos diferentes; =1 uso de mais de 3 medicamentos diferentes). Considere para análise apenas 6 variáveis explicativas: (i) **mhv4**, número de consultas ambulatoriais relacionadas à saúde mental (0: nenhum, 1: um a cinco, 2: seis a quatorze e 3: maior do que quatorze), (ii) **inptmhv3**, número de internações hospitalares relacionadas à saúde mental (0: nenhuma, 1: uma e 2: mais do que uma), (iii) **gender**, gênero (0: Feminino, 1: Masculino), (iv) **urban**, local de residência (0: Urbana, 1: Rural)) e (v) **comorbid**, existência de comorbidade (0: Não, 1: Sim) e (vi) **age**, idade em anos. Inicialmente disponibilizar o arquivo de dados `require(aplore3)`

`attach(polypharm).`

Transformar a variável resposta em variável numérica

`resp=factor(polypharmacy)`

`resp=numeric(resp).`

Fazer inicialmente uma análise descritiva dos dados, tais como tabelas de contingência entre a variável resposta e cada variável explicativa

categórica e boxplots da idade, para cada ano. Ajustar inicialmente uma equação de estimação generalizada EEG-Bernoulli com estrutura de correlação do tipo AR(1) e considere `tage=log(age/100)` como variável explicativa contínua. Ajustar apenas os efeitos principais e utilizando a biblioteca GEE

```
fit1.poly = gee(resp ~ mhv4 + inptmhv3 + gender + urban +
  comorbid + tage, id = id, family = binomial(link = "logit"),
  corstr = "AR-M", Mv = 1).
```

Selecione as variáveis explicativas ao nível de significância de 10%. Fazer uma análise de influência e análise de resíduos. Obter estimativas intervalares de 95% para as razões de chances e interpretar.

Fazer agora um ajuste considerando efeito aleatório de paciente, ou seja ajustar um modelo linear generalizado misto Bernoulli

```
fit2.poly = gamlss(resp ~ mhv4 + inptmhv3 + gender + urban
  + comorbid + tage + random(as.factor(id)), family = BI).
```

Selecione as variáveis explicativas ao nível de significância de 10%. Fazer uma análise de resíduos. Obter estimativas intervalares de 95% para as razões de chances e interpretar.

2. No arquivo **Milk** do `gamlss` são apresentados dados referentes a um experimento longitudinal desenvolvido na Austrália com 79 vacas que foram aleatorizadas segundo três dietas e foi observado semanalmente a quantidade de proteína no leite de cada animal. O objetivo principal do estudo é verificar se há diferenças significativas entre as quantidades médias semanais de proteína sob as três dietas. Os dados estão descritos na seguinte ordem: (i) `protein` (quantidade de proteína), (ii) `Time` (semana), (iii) `Cow` (identificação do animal) e (iv) `Diet` (cevada, cevada+tremoços e tremoços). É preciso informar que a variável `Diet` é categórica através do comando

```
Diet=factor(Diet).
```

Fazer inicialmente uma análise descritiva dos dados, por exemplo, apresentando os perfis dos animais segundo a quantidade de proteína observada ao longo das semanas e para cada dieta gráficos de densidade e boxplots. Ajustar inicialmente uma equação de estimação generalizada EEG-gama com estrutura de correlação do tipo AR(1) e considere o

tempo como variável explicativa contínua. Por exemplo, o ajuste apenas com efeitos principais e utilizando a biblioteca **GEE** fica dado por

```
fit1.milk = gee(protein ~ Diet + Time + I(Time^2), id = Cow,
family = Gamma(link = "log"), corstr = "AR-M", Mv = 1).
```

Fazer agora um ajuste considerando efeito aleatório de animal, ou seja ajustar um modelo linear generalizado misto gama

```
fit2.milk = gamlss(formula = protein ~ Diet + Time + I(Time^2)
+ random(as.factor(Cow)), family = GA).
```

Verifique se é possível incluir interação entre **Diet** e **Time** em cada um dos modelos. Compare os ajustes através de análise de resíduos.

Discuta as vantagens e desvantagens de cada um dos modelos ajustados, EEG-gama e MLG misto gama.

3. Sejam $Y_{i1} \stackrel{\text{iid}}{\sim} \text{FE}(\mu_1, \phi)$ e $Y_{i2} \stackrel{\text{iid}}{\sim} \text{FE}(\mu_2, \phi)$, em que $\text{Corr}(Y_{i1}, Y_{i2}) = \rho$, para $i = 1, \dots, n$. Para testar $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_1 - \mu_2 \neq 0$ considere a estatística

$$\xi_W = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\text{Var}(\bar{Y}_1 - \bar{Y}_2)},$$

que sob H_0 segue assintoticamente distribuição χ_1^2 central. Sob a hipótese alternativa ξ_W segue assintoticamente distribuição $\chi_1^2(\lambda)$, em que $\lambda = (\mu_1 - \mu_2)^2 / 2\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ é o parâmetro de não centralidade. Seja $P_n(\lambda, \rho) = P(\xi_W > c | H_1)$ o poder do teste (probabilidade de rejeitar H_0 quando há diferença entre as médias populacionais). Calcule inicialmente $\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ e discuta o comportamento do poder do teste à medida que varia o coeficiente de correlação linear de Pearson $-1 \leq \rho \leq 1$, supondo fixados μ_1 , μ_2 , ϕ e n . Procure responder para quais situações será necessário um tamanho amostral maior para detectar a mesma diferença (entre as médias populacionais) com a mesma probabilidade.