Resolução da Lista 3 - Análise de Dados Longitudinais

Helen Lourenço e Vitor Kroeff

Questão 1

a)

```
dados <- tibble(foreign::read.dta("toenail.dta"))
ajuste_gee <-geeglm(y ~ trt * month - trt,family = binomial(link = "logit"),
                    data = dados,id = id,corstr = "exchangeable")
```

b)

```
exp(ajuste_gee$coefficients)
```

```
(Intercept)       month   trt:month
  0.5608934   0.8425524   0.9252438
```

O coeficiente β_2 está relacionado ao efeito do tempo no grupo de tratamento A (ou controle). O expoente e^{β_2} representa uma razão de chances. Essa razão de chances indica que, no grupo A, o coeficiente está associado a uma redução na probabilidade de ocorrência de onicólise com o passar dos meses.

c)

O coefiente β_3 está associado a interação entre o tratamento B e o tempo em meses. Assim como na alternativa anterior, vemos que $e^{\beta_3} = 0,925$ está assciado a uma reduzação das chances de ocorrência de onicólise com o passar dos meses, porém uma redução menor que a do grupo controle.

d)

Podemos observar o `summary` do modelo ajustado como sendo:

```r
summary(ajuste_gee)
```

```
Call:
geeglm(formula = y ~ trt * month - trt, family = binomial(link = "logit"),
    data = dados, id = id, corstr = "exchangeable")

 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept) -0.57822  0.13041 19.661 9.25e-06 ***
month       -0.17132  0.02957 33.574 6.86e-09 ***
trt:month   -0.07770  0.05379  2.086    0.149
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    1.088  0.5265
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.4217  0.2203
Number of clusters:   294  Maximum cluster size: 7
```

Com base no p-valor associado ao β_3, não parece ter uma diferençã significativa entre os tratamentos aplicados nos Grupos A e B. Também com base nos coeficientes do modelos, podemos observar que essa chance de desnvolver uma onicólise sereve diminui com o passar dos meses.

e)

```r
ajuste_misto <-lme4::glmer(
  formula = y ~ month + trt:month + (1 | id),
  family = binomial(link = "logit"),
  data = dados)
```

f)

Como o efeito aleatório está apenas no intercepto, vimos a seguinte relação aproximada entre entre os coeficientes do modelo marginal e aqueles do modelo de efeitos aleatórios.

$$\beta_M = \frac{\beta_{EA}}{\sqrt{1 + \frac{16\sqrt{3}}{15\pi}\sigma_b^2}}$$

```r
var_bi <- lme4::VarCorr(ajuste_misto)
var_bi
```

```
 Groups Name        Std.Dev.
 id     (Intercept) 4.54
```

```r
var_bi <- sqrt(as.numeric(var_bi)) # Conversão para númerico
fator <- sqrt(1 + (16*sqrt(3)/(15*pi)*var_bi))
fator
```

```
[1] 1.916
```

Podemos comparar as magnitudes dos efeitos da seguinte forma:

```r
b.gee <- summary(ajuste_gee)$coef[,1]
b.lme <- summary(ajuste_misto)$coef[,1]
p.gee <- summary(ajuste_gee)$coef[,4]
p.lme <- summary(ajuste_misto)$coef[,4]
round(cbind(b.gee, or.gee = exp(b.gee), p.gee, b.lme, or.lme = exp(b.lme), p.lme,
            razao.b = b.lme/b.gee), 3)
```

```
             b.gee or.gee p.gee  b.lme or.lme p.lme razao.b
(Intercept) -0.578  0.561 0.000 -2.649  0.071  0.00   4.581
month       -0.171  0.843 0.000 -0.396  0.673  0.00   2.309
month:trt   -0.078  0.925 0.149 -0.146  0.865  0.03   1.873
```

Podemos observar estimativas maiores para o modelo misto em relação ao GEE, mas com o mesmo sinal, indicando uma concordância dos efeitos das variáveis.

g)

```r
knitr::kable(
exp(summary(ajuste_misto)$coef))
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.0707	2.031	0.0238	1.00
month	0.6733	1.047	0.0002	1.00
month:trt	0.8646	1.069	0.1134	1.03

A interpretação é muito próxima a do modelo GEE, onde β_2 está associada a variação no mês no grupo A (controle), porém está sendo levado em conta a variação de cada paciente do grupo A por conta do efeito aleatório no intercepto.

h)

i)

j)

Questão 2

a)

```r
ajuste_gee_ind <- geeglm(response ~ group * time,data = dados_rats,
                         id = subject,corstr = "independence")

ajuste_gee_simetria <- geeglm(response ~ group * time,data = dados_rats,
                         id = subject,corstr = "exchangeable")

ajuste_gee_ar1 <- geeglm(response ~ group * time,data = dados_rats,
                         id = subject,corstr = "ar1")

ajuste_gee_unstructured <- geeglm(response ~ group * time,data = dados_rats,
                         id = subject,corstr = "unstructured")
```

b)

Nos ajustes realizados na alternativa anterior Y_{ij}(`response`) segue uma distribuição Normal com diferentes formas de estimar a matriz de coreelção. Dentre as ajustas estão (em ordem):

- **Independete***:* Nenhuma correlação entre as observações repetidas.
- **Simetria Composta***:* Todas as observações dentro de um indivíduo têm a mesma correlação.
- **AR(1)***:* As observações mais próximas no tempo têm maior correlação.
- **Não Estruturad***a:* A correlação entre cada par de observações é estimada de forma independente.

c)

FAZER

d)

Primeiro podemos observar a estrutura da matriz de correlação da base como um todo e depois de cada grupo. Abaixo temos a matriz de correlação geral da base:

```
dados_largo <- reshape::cast(dados_rats, subject ~ time, value = "response")
dados_largo <- na.omit(dados_largo)
round(cor(dados_largo[,2:7]),2)
```

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1.00 0.61 0.60 0.72 0.64 0.70
[2,] 0.61 1.00 0.88 0.80 0.80 0.76
[3,] 0.60 0.88 1.00 0.86 0.76 0.72
[4,] 0.72 0.80 0.86 1.00 0.67 0.79
[5,] 0.64 0.80 0.76 0.67 1.00 0.75
[6,] 0.70 0.76 0.72 0.79 0.75 1.00
```

E para os diferentes grupos:

FAZER

Questão 3

a)