

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 coeficiente 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 associado a uma redução 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:

```
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 diferença 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)

```
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 os coeficientes do modelo marginal e aqueles do modelo de efeitos aleatórios.

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

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

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

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

```
[1] 1.916
```

Podemos comparar as magnitudes dos efeitos da seguinte forma:

```
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)

```
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)

```
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 corelação. 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 Estruturada:** A correlação entre cada par de observações é estimada de forma independente.

c)

Podemos comparar cada um dos modelos ajustados na alternativa **a)** e comparar com a sua respectiva estrutura do modelo `gls()` da seguinte forma:

- Independente

Para os modelos com matriz de correlação independentes, podemos observar que

```
dados_rats <- na.omit(dados_rats)

summary(ajuste_gee_ind)$coefficients # Ajuste GEE
```

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	63.049929	1.186235	2825.059	0.0000
group	0.243826	0.678895	0.129	0.7195
time	0.203898	0.014009	211.857	0.0000
group:time	-0.008235	0.008066	1.042	0.3073

```
ajuste_gls_ind <- gls(response ~ group * time, data = dados_rats)
summary(ajuste_gls_ind) # Ajuste GLS
```

Generalized least squares fit by REML

Model: response ~ group * time

Data: dados_rats

AIC BIC logLik

1203 1221 -596.7

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	63.05	1.6633	37.91	0.0000
group	0.24	0.8163	0.30	0.7654
time	0.20	0.0217	9.39	0.0000
group:time	-0.01	0.0108	-0.76	0.4476

Correlation:

	(Intr)	group	time
group	-0.924		
time	-0.969	0.902	
group:time	0.889	-0.969	-0.924

Standardized residuals:

	Min	Q1	Med	Q3	Max
	-2.57774	-0.67881	0.04173	0.61316	2.60110

Residual standard error: 2.511

Degrees of freedom: 252 total; 248 residual

d)

Consideramos o ajuste `ajuste_gee_unstructured` (Não Estruturada), como o mais adequado aos dados. abaixo temos o resultado do `summary` do modelo.

```
summary(ajuste_gee_unstructured)$coefficients
```

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	70.55828	2.35043	901.157	0.00000
group	-2.50386	1.05934	5.587	0.01810
time	0.08166	0.04398	3.448	0.06335
group:time	0.03421	0.01848	3.428	0.06410

Com base nos resultados, podemos observar que as variáveis `time` e interação de tempo e grupo (`group::time`), não são significativas a um nível de 5%. Mas podemos ver que o efeito de grupo é significativo.

Questão 3

a)

```
ajuste_misto_intercept <- lme4::lmer(response ~ group*time + (1|subject), data = dados_rats)
ajuste_misto_tempo <- lme4::lmer(response ~ group * time + (1 + time | subject), data = dados_rats)
```

Ambos os modelos ajustados assumem que $Y_{ij}|b_i \sim N(\mu, \sigma^2)$ e que os $b_i(b_{0i}, b_{1i}) \sim N(0, \Sigma)$. Onde b_{0i} é o efeito aleatório do intercepto e b_{1i} do tempo.

b)

Nos modelos ajustados no item a) temos a seguinte relação de parâmetros:

Table 2: Modelos Mistos

Modelo	Efeitos fixos	Efeitos aleatórios	Total de parâmetros
ajuste_misto_intercept	4	2	6
ajuste_misto_tempo	4	4	8