

MAT02035 - Modelos para dados correlacionados

Avaliação Pontual 3: Solução do Problema 5.1 - Applied Longitudinal Analysis (2011)

Tábata de Bem - 207766

07/04/2021

Problema 5.1

No National Cooperative Gallstone Study (NCGS), um dos principais interesses era estudar a segurança da droga *chenodiol* para o tratamento de cálculos biliares de colesterol (Schoenfield et al., 1981; Wei e Lachin, 1984). Neste estudo, os pacientes foram aleatoriamente designados para alta dose (750 mg por dia), baixa dose (375 mg por dia) ou placebo. Nós nos concentramos em um subconjunto de dados em pacientes que tinham cálculos biliares flutuantes e que foram designados para os grupos de alta dose e placebo.

No NCGS, foi sugerido que o *chenodiol* dissolveria os cálculos biliares, mas, ao fazê-lo, poderia aumentar os níveis de colesterol sérico. Como resultado, o colesterol sérico (mg/dL) foi medido no início do estudo e aos 6, 12, 20 e 24 meses de acompanhamento. Muitas medições de colesterol estão faltando devido a visitas perdidas, amostras de laboratório foram perdidas ou inadequadas ou o acompanhamento do paciente foi encerrado.

Os dados de colesterol sérico do NCGS são armazenados em um arquivo externo: *cholesterol.dat*.

Cada linha do conjunto de dados contém as seguintes sete variáveis: Grupo, ID, Y_1 , Y_2 , Y_3 , Y_4 , Y_5 . Além disso, Grupo é uma variável categórica (1 = Alta dose, 2 = Placebo).

5.1.1 Leia os dados do arquivo externo e deixe-os em um formato “multivariável” ou “amplo”.

```
library(haven)

# Leitura dos dados
chole.df <- read_dta("data/cholesterol.dta")

# transformando a variavel grupo em fator
chole.df$group <- factor(chole.df$group, labels = c("Alta dose", "Placebo"))

# Reformando o objeto: de "Largo" para "Longo"
library(dplyr); library(tidyr)
chole.df.longo <- chole.df %>% gather("tempo", "colesterol", -id, -group)

# formatando as variaveis
chole.df.longo$tempo <- factor(chole.df.longo$tempo, labels = c("0", "6", "12", "20", "24"))
chole.df.longo$tempo.num <- as.numeric(as.character(chole.df.longo$tempo))
head(chole.df.longo)
```

##	#	A tibble: 6 x 5			
##	group	id	tempo	colesterol	tempo.num
##	<fct>	<dbl>	<fct>	<dbl>	<dbl>
##	1 Alta dose	1 0		178	0
##	2 Alta dose	2 0		254	0
##	3 Alta dose	3 0		185	0
##	4 Alta dose	4 0		219	0
##	5 Alta dose	5 0		205	0
##	6 Alta dose	6 0		182	0

5.1.2 Calcule as médias da amostra, desvios padrão e variâncias dos níveis de colesterol sérico em cada ocasião para cada grupo de tratamento

```
library(plyr)
chole.resumo <- ddply(chole.df.longo, ~ group + tempo.num, summarize, colesterol.m = round(mean
  (colesterol, na.rm = TRUE),2), sd = round(sd(colesterol, na.rm = TRUE),2), var = round(var(col
    esterol, na.rm = TRUE),2))

library(kableExtra)
# organizando as informacoes de uma forma visualmente melhor
chole.resumo %>%
  kbl() %>%
  kable_paper("hover", full_width = F)
```

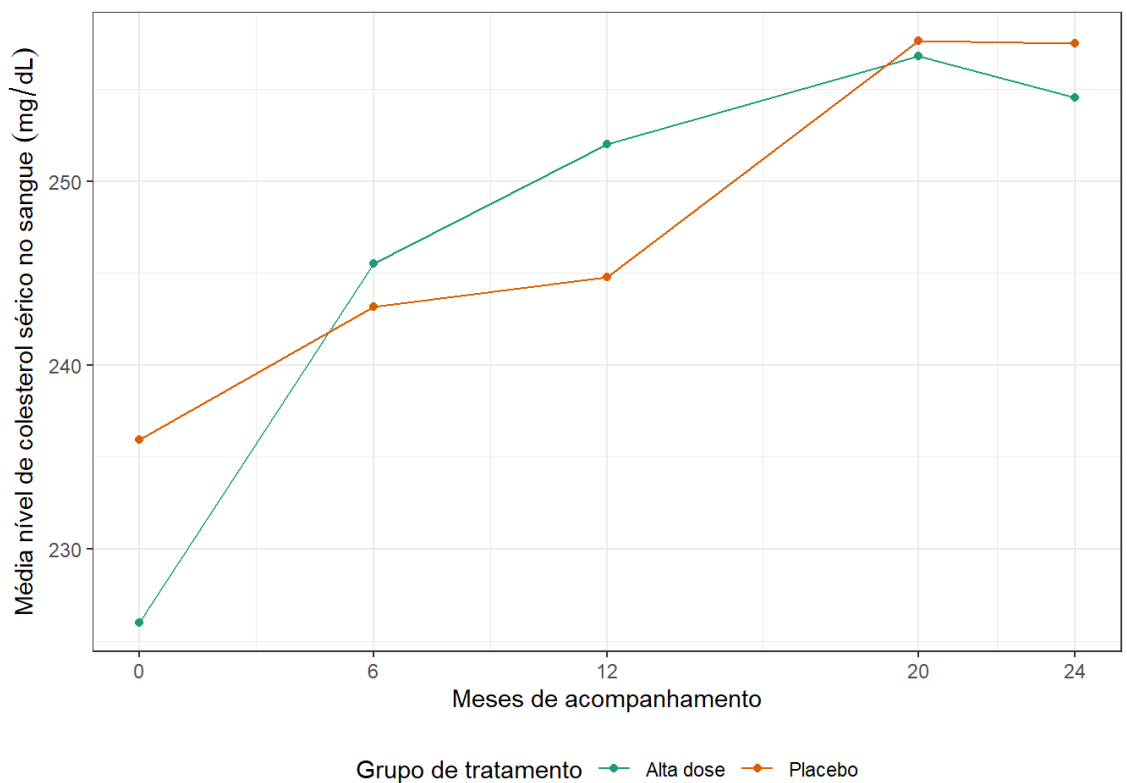
group	tempo.num	colesterol.m	sd	var
Alta dose	0	226.02	39.66	1573.26
Alta dose	6	245.53	39.45	1556.48
Alta dose	12	252.02	38.33	1469.13
Alta dose	20	256.80	34.49	1189.52
Alta dose	24	254.55	49.96	2496.20
Placebo	0	235.93	55.87	3121.97
Placebo	6	243.17	49.24	2424.55
Placebo	12	244.76	46.11	2126.19
Placebo	20	257.60	51.14	2615.48
Placebo	24	257.48	49.39	2439.19

Analisar o perfil do grupo que recebeu o tratamento de alta dose, é possível observar que os níveis médio de colesterol sérico aumentaram em aproximadamente 13% ao longo do tempo de estudo. Ao passo que o grupo que recebeu o tratamento placebo, teve um aumento de aproximadamente 9% nos níveis médio de colesterol para o mesmo período.

Além disso, é possível observar que os níveis médio de colesterol apresentaram uma baixa variabilidade entre os grupos para cada nível de tempo de coleta, ou seja, quando os grupos de alta dose e placebo são comparados em cada tempo, eles não apresentam diferenças expressivas nos níveis de colesterol médio.

5.1.3 Em um único gráfico, construa um gráfico de tempo que exibe o colesterol sérico médio em relação ao tempo (em meses) para os dois grupos de tratamento. Descreva as características gerais das tendências temporais para os dois grupos

```
library(ggplot2)
p <- ggplot(data = chole.resumo,
  mapping = aes(x = tempo.num, y = colesterol.m, group = group, colour = group)) +
  geom_point() +
  geom_line() +
  scale_x_continuous(breaks = c(0, 6, 12, 20, 24)) +
  scale_color_brewer(palette = "Dark2") +
  labs(x = "Meses de acompanhamento",
    y = expression("Média nível de colesterol sérico no sangue"~(mg/dL)),
    colour = "Grupo de tratamento") +
  theme_bw() +
  theme(legend.position = "bottom")
p
```



É possível observar que o grupo de tratamento que recebeu alta dose da droga, apresentou um aumento nos níveis de colesterol sérico no sangue ao longo dos vinte primeiros meses de acompanhamento do estudo, onde ao final do período, apresentou uma ligeira baixa no mesmo. É importante lembrar que, para ambos os grupos, há dados faltantes ao longo do acompanhamento, que acaba impactando nos resultados. Já o grupo de tratamento que recebeu placebo ao invés de alguma dose de chenodiol, registrou um ligeiro aumento nos níveis médios de colesterol no sangue nos primeiros 6 meses de tratamento e outro aumento mais expressivo no decorrer entre os 12 e 20 meses de acompanhamento, no restante do tempo, os níveis de colesterol não apresentaram grandes mudanças.

Ao analisar as mudanças nos níveis de colesterol sérico ao longo do tempo, observa-se que nos primeiros seis meses de acompanhamento ambos os grupos apresentam um aumento em suas médias, ao passo que entre o segundo semestre de estudo, o grupo placebo mantém os níveis médio de colesterol e o grupo que recebeu a droga continuou apresentando um aumento nos mesmos. Nos oito meses seguintes, os dois grupos registraram um aumento nas médias de colesterol no sangue e, finalmente, o grupo placebo praticamente manteve o mesmo nível médio de colesterol em comparação ao tempo anterior, por outro lado, o grupo que recebeu alta dose de tratamento registrou uma queda nos níveis de colesterol sérico.

5.1.4 Leia os dados do arquivo externo e coloque os dados em um formato “univariado” ou “longo”, com cinco “registros” por sujeito.

```
library(dplyr); library(tidyr)
chole.df.longo2 <- chole.df %>% gather("tempo", "grupo", -group, -id)
head(chole.df.longo2)
```

```
## # A tibble: 6 x 4
##   group      id tempo grupo
##   <fct>    <dbl> <chr> <dbl>
## 1 Alta dose     1 y1     178
## 2 Alta dose     2 y1     254
## 3 Alta dose     3 y1     185
## 4 Alta dose     4 y1     219
## 5 Alta dose     5 y1     205
## 6 Alta dose     6 y1     182
```

5.1.5 Assumindo uma matriz de covariância não estruturada, conduza uma análise dos perfis de resposta. Determine se os padrões de mudança ao longo do tempo diferem nos dois grupos de tratamento.

```
library(nlme)
# modelo de perfis de respostas com matriz de covariancia nao estruturada
mod.pr <- gls(cholesterol ~ group * tempo,
              corr = corSymm(form = ~ 1 | id),
              weights = varIdent(form = ~ 1 | tempo),
              method = "REML",
              data = chole.df.longo,
              na.action = na.omit)
summary(mod.pr)
```

```
## Generalized least squares fit by REML
##   Model: colesterol ~ group * tempo
##   Data: chole.df.longo
##       AIC       BIC    logLik
##  4314.588 4416.587 -2132.294
##
## Correlation Structure: General
## Formula: ~1 | id
## Parameter estimate(s):
## Correlation:
##   1     2     3     4
## 2 0.770
## 3 0.732 0.773
## 4 0.738 0.800 0.726
## 5 0.586 0.665 0.678 0.625
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | tempo
## Parameter estimates:
##       0         6        12        20        24
## 1.0000000 0.9320569 0.8791668 0.8974016 1.0300809
##
## Coefficients:
##                               Value Std.Error  t-value p-value
## (Intercept)                226.24854   5.935222  38.11964  0.0000
## groupPlacebo                 9.67829   9.412956   1.02819  0.3044
## tempo6                      19.46141   3.907196   4.98092  0.0000
## tempo12                     25.13513   4.276720   5.87720  0.0000
## tempo20                     27.86003   4.540835   6.13544  0.0000
## tempo24                     27.65829   6.370952   4.34131  0.0000
## groupPlacebo:tempo6        -12.21751   6.193407  -1.97266  0.0492
## groupPlacebo:tempo12      -16.28893   6.738390  -2.41733  0.0160
## groupPlacebo:tempo20      -4.75670   6.973253  -0.68213  0.4955
## groupPlacebo:tempo24      -6.53598   9.763271  -0.66945  0.5036
##
## Correlation:
##              (Intr) grpPlc tempo6 temp12 temp20 temp24 grpP:6 grP:12
## groupPlacebo    -0.631
## tempo6          -0.428  0.270
## tempo12         -0.493  0.311  0.574
## tempo20         -0.441  0.278  0.573  0.494
## tempo24         -0.358  0.226  0.445  0.475  0.395
## groupPlacebo:tempo6  0.270 -0.429 -0.631 -0.362 -0.361 -0.281
## groupPlacebo:tempo12 0.313 -0.497 -0.364 -0.635 -0.313 -0.302  0.578
## groupPlacebo:tempo20 0.287 -0.456 -0.373 -0.321 -0.651 -0.257  0.592  0.513
## groupPlacebo:tempo24 0.234 -0.378 -0.291 -0.310 -0.258 -0.653  0.463  0.503
##
##              grp:20
## groupPlacebo
## tempo6
## tempo12
## tempo20
## tempo24
## groupPlacebo:tempo6
## groupPlacebo:tempo12
## groupPlacebo:tempo20
## groupPlacebo:tempo24  0.419
##
## Standardized residuals:
##       Min       Q1       Med       Q3       Max
## -2.32029892 -0.68866946 -0.02685013  0.60855777  3.89204107
##
## Residual standard error: 46.7809
## Degrees of freedom: 447 total; 437 residual
```


Nota-se que há efeito significativo (p-valor < 0.05) no efeito da interação entre o grupo placebo e os tempos 6 e 12 meses de estudo.



5.1.6 Mostre a covariância estimada 5 x 5 e as matrizes de correlação para as cinco medições repetidas de colesterol sérico

```
round(matrix(getVarCov(mod.pr),
             ncol = 5),1) %>%
  kbl() %>%
  kable_paper("hover", full_width = F)
```

2188.5	1513.0	1407.9	1449.2	1320.7
1513.0	1762.4	1335.5	1409.3	1345.6
1407.9	1335.5	1691.5	1254.4	1343.3
1449.2	1409.3	1254.4	1762.4	1264.2
1320.7	1345.6	1343.3	1264.2	2322.1

Podemos notar que a correlação até a metade do acompanhamento apresenta um comportamento decrescente e após isso, começa a aumentar. 

5.1.7 Com a linha de base (mês 0) e o grupo placebo (grupo 2) como o grupo de referência, escreva o modelo de regressão para o colesterol sérico médio que corresponde à análise dos perfis de resposta no Problema 5.1.5

```
library(car)

Anova(mod.pr)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: colesterol
##           Df    Chisq Pr(>Chisq)
## group      1  0.0037   0.95163
## tempo      4 65.4783  2.04e-13 ***
## group:tempo 4  7.9167   0.09468 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
knitr::kable(
  Anova(mod.pr),
  digits = c(0, 2, 4))
```

	Df	Chisq	Pr(>Chisq)
group	1	0.00	0.9516
tempo	4	65.48	0.0000
group:tempo	4	7.92	0.0947