

Contexto e 5.1.1

5.1.2

5.1.3

5.1.4

5.1.5

5.1.6

5.1.7

5.1.8

5.1.9

5.1.10

Avaliação 3

Modelos para dados correlacionados - 2020/2

Rafaela Galetto - 00288943

01 de abril de 2021

Contexto e 5.1.1

O objetivo do problema do capítulo 5 era estudar a segurança do medicamento chenodiol para o tratamento de cálculos biliares de colesterol. Para tanto, foi atribuído aos pacientes, de forma aleatória, 3 tratamentos:

- Dose alta (750 mg por dia);
- Dose baixa (375 mg por dia);
- Placebo.

Nos nossos dados, vamos trabalhar somente com os pacientes da dose alta e do placebo, sendo “1” e “2” na variável “grupo” respectivamente.

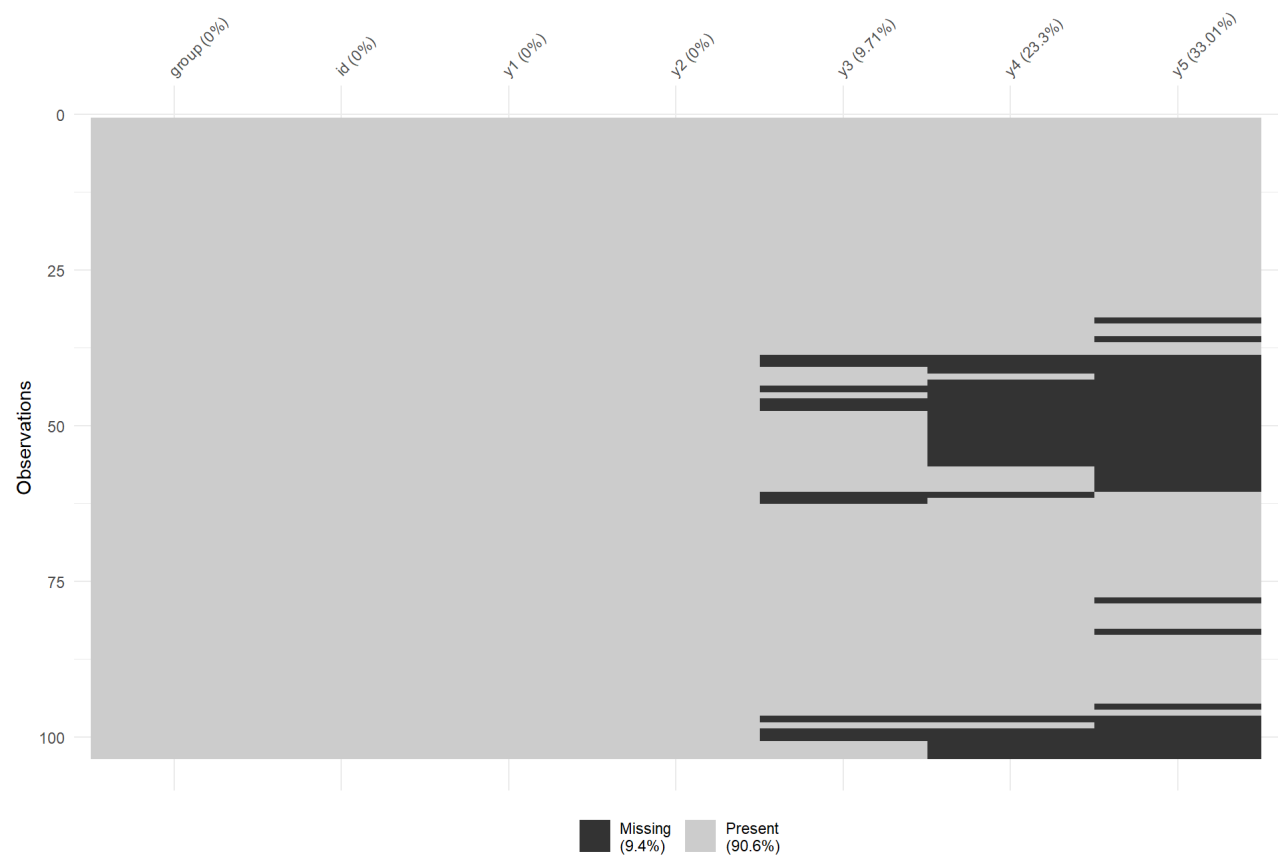
A nossa variável de interesse é o nível de colesterol sérico dos pacientes, pois, se tem a hipótese que o chenodiol causa poderia causar o aumento. Portanto, o colesterol sérico foi medido nos pacientes em 5 tempos diferentes:

- Início do estudo (tempo base);
- Após 6 meses de acompanhamento;
- Após 12 meses de acompanhamento;
- Após 20 meses de acompanhamento;
- Após 24 meses de acompanhamento.

Com o gráfico abaixo, pode-se perceber a ausência de alguns dados a partir do terceiro mês de acompanhamento, caracterizando o nosso estudo como incompleto.

Foi um estudo com **delineamento completo** e balanceado.

```
dados <- read_dta(file = "cholesterol.dta")  
  
missing <- vis_miss(dados)  
  
plot(missing)
```



O primeiro item do exercício pede que o banco de dados seja lido no formato “largo”.

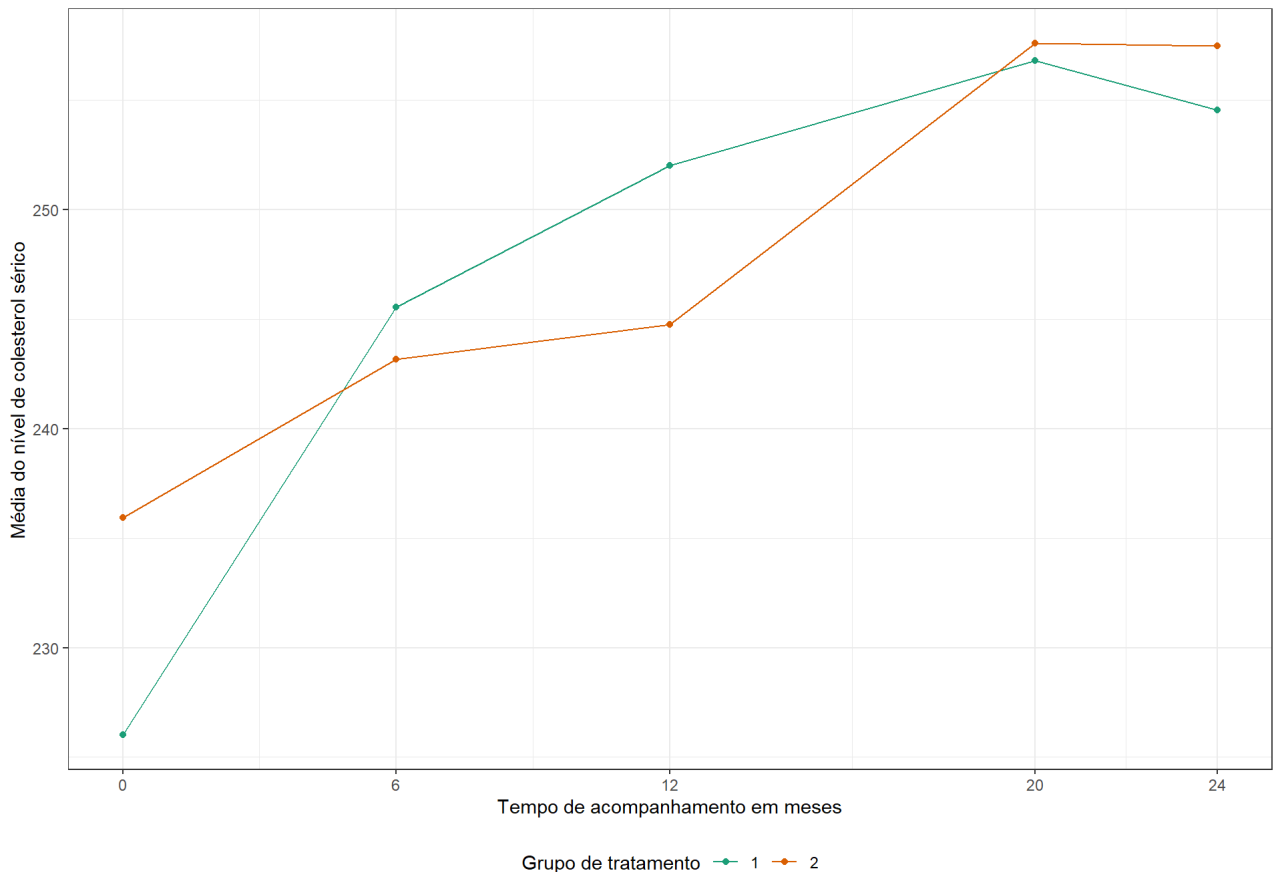
5.1.2

Enunciado: Calcular as médias da amostra, desvios padrão e variâncias do soro níveis de colesterol em cada ocasião para cada grupo de tratamento.

GRUPO	Média	Média	Média	Média	Média	Desvio	Desvio	Desvio	Desvio	Desvio	Variância	Variância	Variância
	tempo 0	após 6m	após 12m	após 20m	após 24m	padrão	padrão	padrão	padrão	padrão			
1	226.0161	245.5323	252.0182	256.7955	254.5526	39.66437	39.45228	38.32922	34.48935	49.96198	1573.262	1556.483	1469.129
2	235.9268	243.1707	244.7632	257.6000	257.4839	55.87459	49.23967	46.11058	51.14179	49.38817	3121.970	2424.545	2126.186

5.1.3

Enunciado: Em um único gráfico, construa um gráfico de tempo que exiba o colesterol sérico médio versus tempo (em meses) para os dois grupos de tratamento. Descreva características das tendências temporais para os dois grupos.



Conforme o gráfico apresentado, percebe-se que o nível médio de colesterol sérico no tempo base é menor no grupo que recebeu a dose alta do remédio em relação ao grupo placebo. Ambos os grupos apresentam, de fato, um aumento no nível de colesterol sérico ao longo dos primeiros 20 meses de acompanhamento. Porém, o grupo que recebeu a medicação parece ter esse aumento de uma forma muito mais rápida, tendo um crescimento intenso que só apresenta um decaimento no tempo 5 (24 meses). O grupo placebo, como disse anteriormente, também apresenta um aumento no nível de colesterol sérico, porém, de forma mais lenta, chegando a ter períodos mais intensos (entre tempo base e até 6 meses), ficando quase constante entre 6 e 12 meses e repetindo o comportamento nos intervalos seguintes.



Ressalto um ponto de atenção para as médias do tempo 5 (24 meses), em especial, pois temos 33,01% de ausência de dados.

Parece que os perfis não são paralelos e existe efeito de interação, as médias não são constantes e os perfis não estão no mesmo nível.

5.1.4

Enunciado: Em seguida, leia os dados do arquivo externo e coloque os dados em formato "univariado" ou "longo", com cinco "registros" por assunto.

```
dados_longo <- gather(data = dados,
                      key = "tempo",
                      value = "colesterol",
                      -id, -group)
```

```
#knitr::kable(dados_longo)
```

5.1.5

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

Interpretações ao longo das demais questões.

Aqui o grupo que recebeu a dose alta é o grupo de referência.

```

dados_longo$tempo <- as.numeric(as.character(factor(dados_longo$tempo, labels = c(0, 6, 12, 20, 24))))
dados_longo$tempo <- factor(dados_longo$tempo, labels = c("0", "6", "12", "20", "24"))
dados_longo$tempo.num <- as.numeric(as.character(dados_longo$tempo))
dados_longo$colesterol <- dados_longo$colesterol

colesterol_resumo <- ddply(dados_longo, ~ group + tempo.num, summarize, MEDIA = mean(colesterol, na.rm=T))

dados_longo_sem_miss <- na.omit(dados_longo)

dados_longo_sem_miss$group <- as.factor(dados_longo_sem_miss$group)

# modelo de perfis de respostas
# com matriz de covariância não estruturada
modelo <- gls(colesterol ~ group * tempo,
              corr = corSymm(form = ~ 1 | id),
              weights = varIdent(form = ~ 1 | tempo),
              method = "REML",
              data = dados_longo_sem_miss)

summary(modelo)

```

```

## Generalized least squares fit by REML
## Model: colesterol ~ group * tempo
## Data: dados_longo_sem_miss
##      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
## group2        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
## group2:tempo6 -12.21751  6.193407 -1.97266  0.0492
## group2:tempo12 -16.28893  6.738390 -2.41733  0.0160
## group2:tempo20 -4.75670  6.973253 -0.68213  0.4955
## group2:tempo24 -6.53598  9.763271 -0.66945  0.5036
##
## Correlation:
## (Intr) group2 tempo6 temp12 temp20 temp24 grp2:6 gr2:12 gr2:20
## group2      -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
## group2:tempo6  0.270 -0.429 -0.631 -0.362 -0.361 -0.281
## group2:tempo12 0.313 -0.497 -0.364 -0.635 -0.313 -0.302  0.578
## group2:tempo20 0.287 -0.456 -0.373 -0.321 -0.651 -0.257  0.592  0.513
## group2:tempo24 0.234 -0.378 -0.291 -0.310 -0.258 -0.653  0.463  0.503  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

```



5.1.6

Enunciado: Exibir a matriz de covariâncias estimadas (5x5) e as matrizes de correlação para as cinco medições repetidas de colesterol sérico.

Aqui o grupo que recebeu a dose alta é o grupo de referência.

Matriz de covariâncias estimadas

```
## Marginal variance covariance matrix
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 2188.5 1513.0 1407.9 1449.2 1320.7
## [2,] 1513.0 1762.4 1335.5 1409.3 1345.6
## [3,] 1407.9 1335.5 1691.5 1254.4 1343.3
## [4,] 1449.2 1409.3 1254.4 1762.4 1264.2
## [5,] 1320.7 1345.6 1343.3 1264.2 2322.1
## Standard Deviations: 46.781 41.981 41.128 41.981 48.188
```

Matriz de correlações

```
## Correlation structure of class corSymm representing
## 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
```

5.1.7

Enunciado: 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 a análise dos perfis de resposta no Problema 5.1.5..

```
## Generalized least squares fit by REML
## Model: colesterol ~ group * tempo
## Data: dados_longo_sem_miss
##      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
## 1.000000 0.9320568 0.8791668 0.8974016 1.0300809
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)  235.92683   7.305948  32.29243  0.0000
## group1       -9.67829   9.412956  -1.02819  0.3044
##
## tempo6       7.24390   4.805425   1.50744  0.1324
## tempo12      8.84620   5.207262   1.69882  0.0901
## tempo20     23.10333   5.292171   4.36557  0.0000
## tempo24     21.12230   7.398137   2.85508  0.0045
## group1:tempo6 12.21751   6.193407   1.97266  0.0492
## group1:tempo12 16.28893   6.738391   2.41733  0.0160
## group1:tempo20  4.75670   6.973253   0.68213  0.4955
## group1:tempo24  6.53598   9.763271   0.66945  0.5036
##
## Correlation:
##      (Intr) group1 tempo6 temp12 temp20 temp24 grp1:6 gr1:12 gr1:20
## group1      -0.776
## tempo6      -0.429  0.333
## tempo12     -0.500  0.388  0.581
## tempo20     -0.466  0.362  0.606  0.526
## tempo24     -0.392  0.304  0.476  0.522  0.438
## group1:tempo6  0.333 -0.429 -0.776 -0.451 -0.470 -0.369
## group1:tempo12 0.387 -0.497 -0.449 -0.773 -0.407 -0.404  0.578
## group1:tempo20 0.354 -0.456 -0.460 -0.400 -0.759 -0.332  0.592  0.513
## group1:tempo24 0.297 -0.378 -0.361 -0.396 -0.332 -0.758  0.463  0.503  0.419
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.32029916 -0.68866948 -0.02685013  0.60855779  3.89204113
##
## Residual standard error: 46.7809
## Degrees of freedom: 447 total; 437 residual
```

$\hat{\gamma} = 235,9268 - 9,6783(Dose) + 7,2439(Tempo6) + 8,8462(Tempo12) + 23,1033(Tempo20) + 21,1223(Tempo24) + 12,2175(Dose * Tempo6) + 16,2889(Dose * Tempo12) + 4,7567(Dose * Tempo20) + 6,53598(Dose * Tempo24)$

5.1.8

Enunciado: Deixe L denotar uma matriz de pesos conhecidos e β o vetor de parâmetros do modelo assumido no Problema 5.1.7. A hipótese nula de que os padrões de mudança ao longo do tempo não diferem nos dois grupos de tratamento podem ser expressa como $H_0: \beta = 0$. Descreva uma matriz de peso L apropriada para essa hipótese nula.

Aqui o grupo que recebeu a dose alta é o grupo de referência.

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

Utilizando $\alpha = 0,05$, **não rejeitamos** a hipótese de ausência de efeito de interação entre grupo e tempo, pois p-valor (0,09468) > α (é possível ver a interação no gráfico das médias de colesterol sérico por grupo ao longo do tempo)

Dado que não rejeitamos a hipótese de ausência de efeito de interação entre grupo e tempo, faz sentido analisar as outras duas hipóteses:


Utilizando $\alpha = 0,05$, **rejeitamos** a hipótese de ausência de efeito de **tempo**, pois p-valor $< 0,001$. Há evidência estatística de que temos efeito de tempo no nível de colesterol sérico.

Utilizando $\alpha = 0,05$, **não rejeitamos** a hipótese de ausência de efeito de **grupo**, pois p-valor $(0,95163) > \alpha$. Não há evidência estatística de que temos efeito de grupo no nível de colesterol sérico.

A matriz de pesos L:

0	1	0	0	0
0	0	1	0	0

5.1.9

Enunciado: Mostre como os coeficientes de regressão estimados a partir de uma análise de resposta os perfis podem ser usados para construir os meios específicos de tempo nos dois grupos. Compare essas médias estimadas com as médias de amostra obtidas no Problema 5.1.2. 

5.1.10

Enunciado: Com a linha de base (mês 0) e o grupo placebo (grupo 2) como o grupo de referência, fornecer uma interpretação para cada um dos coeficientes de regressão estimados em termos do efeito dos tratamentos sobre os padrões de mudança no soro médio colesterol.

- $\beta_0 = 235,9268$. É o valor esperado de nível colesterol sérico quando o tempo de acompanhamento e o tratamento aplicados são iguais a zero.
- $\beta_1 = -9,6783$. Indicam que, mantendo-se as outras variáveis do modelo constantes, o valor esperado de nível colesterol sérico deve diminuir em 9.6783 no grupo placebo com relação ao grupo que recebeu a dose alta.
- Os coeficientes β_2 , β_3 , β_4 e β_5 indicam que, mantendo-se as outras variáveis do modelo constantes, o valor esperado de nível colesterol sérico deve aumentar em 7,2439; 8,8462; 23,1033 ou 21,1223; respectivamente no tempo de acompanhamento de 6 meses, 12 meses, 20 meses e 24 meses, com relação ao grupo que recebeu a dose alta.

Para as interações:

- Mantendo-se as outras variáveis do modelo constantes e estivermos analisando o grupo placebo, o valor esperado de nível colesterol sérico deve aumentar $\beta_{tempo6} + 12,217$. Similarmente, mantendo-se as outras variáveis do modelo constantes e estivermos analisando o tempo de acompanhamento de 6 meses, o valor esperado de nível colesterol sérico deve aumentar $\beta_{DoseAlta} + 12,217$ unidades.

As demais interações seguem o mesmo padrão.