

Avaliação pontual 3

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
MAT02035 - MODELOS PARA DADOS CORRELACIONADOS

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Exercícios do Capítulo 5

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 dose alta (750 mg por dia), dose baixa (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 por causa de visitas perdidas, amostras de laboratório foram perdidas ou inadequadas, ou o acompanhamento do paciente foi encerrado.

- Os dados de colesterol sérico do NCGS estão armazenados em um arquivo externo: `colesterol.dta`.
- Cada linha do conjunto de dados contém as seguintes sete variáveis:
 - Group ID Y1 Y2 Y3 Y4 Y5 **Nota:** A variável categórica Group é codificada 1 = Alta Dose, 2 = Placebo.

Exercício 1

```
# -----  
# Carregando pacotes do R  
  
library(here)  
library(haven)  
library(tidyr)  
library(ggplot2)  
  
# -----  
# Carregando o arquivo de dados  
  
colesterol <- read_dta(file = here::here("data",  
                                         "cholesterol.dta"))  
  
colesterol$group <- factor(colesterol$group,  
                           labels = c("Alta Dose", "Placebo"))
```

```
colesterol
```

```
## # A tibble: 103 x 7
##   group      id    y1    y2    y3    y4    y5
##   <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Alta Dose      1  178  246  295  228  274
## 2 Alta Dose      2  254  260  278  245  340
## 3 Alta Dose      3  185  232  215  220  292
## 4 Alta Dose      4  219  268  241  260  320
## 5 Alta Dose      5  205  232  265  242  230
## 6 Alta Dose      6  182  213  173  200  193
## 7 Alta Dose      7  310  334  290  286  248
## 8 Alta Dose      8  191  204  227  228  196
## 9 Alta Dose      9  245  270  209  255  213
## 10 Alta Dose     10  229  200  238  259  221
## # ... with 93 more rows
```

Exercício 4

Para facilitar a resolução dos exercício 2 e 3, o exercício 4 foi resolvido previamente a estes.

- Em seguida, leia os dados do arquivo externo e coloque os dados em um formato “univariado” ou “longo”, com cinco “registros” por indivíduo.

```
# -----
# Reformatando o objeto: de "largo" para "longo"

colesterol.longo <- gather(data = colesterol,
                           key = "meses",
                           value = "col.serico", -id, -group)

# -----
# Formata variáveis

colesterol.longo$meses <- factor(colesterol.longo$meses,
                                labels = c(0, 6, 12, 20, 24))

colesterol.longo$meses.num <- as.numeric(as.character(colesterol.longo$meses))
colesterol.longo$tempo <- as.numeric(colesterol.longo$meses)

colesterol.longo
```

```
## # A tibble: 515 x 6
##   group      id meses col.serico meses.num tempo
##   <fct>    <dbl> <fct>    <dbl>    <dbl> <dbl>
## 1 Alta Dose      1  0         178         0      1
## 2 Alta Dose      2  0         254         0      1
## 3 Alta Dose      3  0         185         0      1
## 4 Alta Dose      4  0         219         0      1
## 5 Alta Dose      5  0         205         0      1
## 6 Alta Dose      6  0         182         0      1
## 7 Alta Dose      7  0         310         0      1
## 8 Alta Dose      8  0         191         0      1
## 9 Alta Dose      9  0         245         0      1
## 10 Alta Dose     10  0         229         0      1
## # ... with 505 more rows
```

Exercício 2

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

```
library(dplyr)

colesterol.resumo <- colesterol.longo %>%
  group_by(group, meses.num) %>%
  summarise(col_media = mean(col.serico, na.rm = TRUE),
            col_dp = sd(col.serico, na.rm = TRUE)) %>%
  mutate(col_var = col_dp^2)

colesterol.resumo %>%
  knitr::kable(format = "markdown",
               digits = c(0, 0, 2, 2, 2),
               col.names = c("Grupo", "Meses",
                             "Média", "Desvio Padrão",
                             "Variância"), align = "lcccc")
```

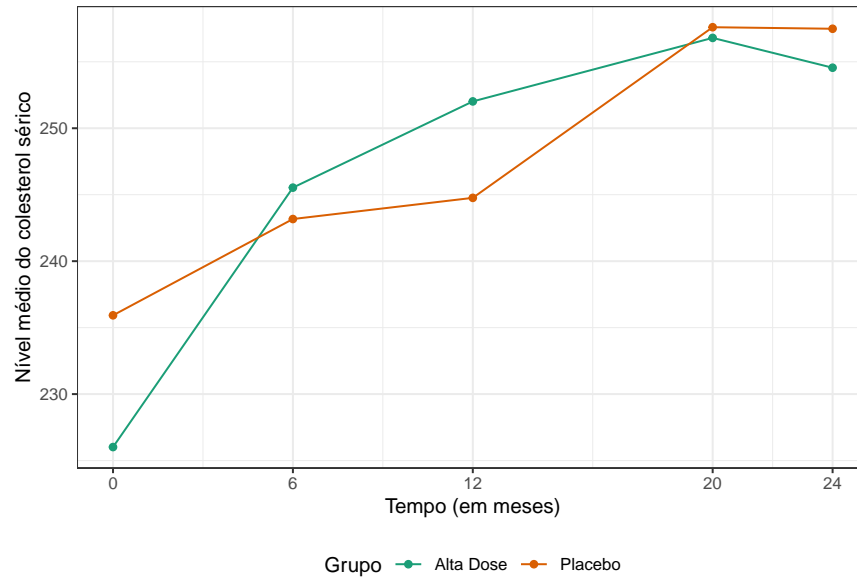
Grupo	Meses	Média	Desvio Padrão	Variância
Alta Dose	0	226.02	39.66	1573.26
Alta Dose	6	245.53	39.45	1556.48
Alta Dose	12	252.02	38.33	1469.13
Alta Dose	20	256.80	34.49	1189.52
Alta Dose	24	254.55	49.96	2496.20
Placebo	0	235.93	55.87	3121.97
Placebo	6	243.17	49.24	2424.55
Placebo	12	244.76	46.11	2126.19
Placebo	20	257.60	51.14	2615.48
Placebo	24	257.48	49.39	2439.19

- *Observação:* é importante notar que alguns indivíduos não possuem todas as cinco medidas, e é preciso remover as observações faltantes para se obter as estatísticas resumo por grupo e ocasião de medição (`na.rm = TRUE`).

Exercício 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.

```
p <- ggplot(data = colesterol.resumo,
           mapping = aes(x = meses.num, y = col_media,
                         colour = group)) +
  geom_point() + geom_line() +
  scale_color_brewer(palette = "Dark2") +
  scale_x_continuous(breaks = c(0, 6, 12, 20, 24)) +
  labs(x = "Tempo (em meses)",
       y = "Nível médio do colesterol sérico",
       colour = "Grupo") +
  theme_bw() + theme(legend.position = "bottom")
p
```



Os dois grupos de tratamento apresentam trajetórias crescentes ao longo do período de acompanhamento.

Exercício 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.

```
# grupo de referência = Placebo

colesterol.longo$group <- relevel(colesterol.longo$group,
                                ref = "Placebo")

library(nlme)

# matriz de covariância não estruturada
mod.pr <- gls(col.serico ~ meses*group,
              na.action = na.omit,
              corr = corSymm(form = ~ tempo | id),
              weights = varIdent(form = ~ 1 | tempo),
              method = "REML",
              data = colesterol.longo)

summary(mod.pr)

## Generalized least squares fit by REML
## Model: col.serico ~ meses * group
## Data: colesterol.longo
## AIC BIC logLik
## 4315.08 4417.078 -2132.54
##
## Correlation Structure: General
## Formula: ~tempo | id
## Parameter estimate(s):
## Correlation:
## 1 2 3 4
```

```

## 2 0.770
## 3 0.732 0.775
## 4 0.737 0.796 0.725
## 5 0.589 0.669 0.680 0.626
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | tempo
## Parameter estimates:
##      1      2      3      4      5
## 1.0000000 0.9322277 0.8798457 0.8962063 1.0303424
##
## Coefficients:
##              Value Std.Error t-value p-value
## (Intercept)    235.92683   7.302899  32.30592  0.0000
## meses6          7.24390   4.805433   1.50744  0.1324
## meses12         8.84831   5.207205   1.69924  0.0900
## meses20        23.10280   5.297390   4.36117  0.0000
## meses24        21.12378   7.369810   2.86626  0.0044
## groupAlta Dose  -9.91070   9.412783  -1.05290  0.2930
## meses6:groupAlta Dose  12.27223   6.193773   1.98138  0.0482
## meses12:groupAlta Dose  16.41756   6.743337   2.43463  0.0153
## meses20:groupAlta Dose   4.97701   6.981935   0.71284  0.4763
## meses24:groupAlta Dose   6.90308   9.791014   0.70504  0.4812
##
## Correlation:
##              (Intr) meses6 mess12 mess20 mess24 grpAlD ms6:AD m12:AD
## meses6        -0.429
## meses12       -0.500  0.584
## meses20       -0.468  0.601  0.525
## meses24       -0.390  0.480  0.522  0.437
## groupAlta Dose -0.776  0.332  0.388  0.363  0.303
## meses6:groupAlta Dose  0.332 -0.776 -0.453 -0.466 -0.372 -0.429
## meses12:groupAlta Dose  0.386 -0.451 -0.772 -0.405 -0.403 -0.497  0.582
## meses20:groupAlta Dose  0.355 -0.456 -0.398 -0.759 -0.332 -0.458  0.587  0.512
## meses24:groupAlta Dose  0.293 -0.361 -0.393 -0.329 -0.753 -0.378  0.465  0.504
##              m20:AD
## meses6
## meses12
## meses20
## meses24
## groupAlta Dose
## meses6:groupAlta Dose
## meses12:groupAlta Dose
## meses20:groupAlta Dose
## meses24:groupAlta Dose  0.419
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.32084266 -0.68934024 -0.02685642  0.61013581  3.89366656
##
## Residual standard error: 46.76137
## Degrees of freedom: 447 total; 437 residual

```

Vamos testar a hipótese de ausência de efeito de interação *grupo* \times *tempo*:

$$H_{01} : \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0.$$

```
library(car)

Anova(mod.pr)

## Analysis of Deviance Table (Type II tests)
##
## Response: col.serico
##           Df    Chisq Pr(>Chisq)
## meses      4 65.8843 1.676e-13 ***
## group      1  0.0021   0.96380
## meses:group 4  7.8609   0.09681 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assim, ao nível de 5% de significância, não rejeitamos a hipótese de ausência de interação. Ou seja, os padrões de mudança ao longo do tempo podem não diferirem nos dois grupos de tratamento.

Exercício 6

- Exiba as matrizes 5×5 de covariância e correlação estimadas para as cinco medidas repetidas do colesterol sérico.

```
# matriz de covariância estimada
getVarCov(mod.pr)

## Marginal variance covariance matrix
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 2186.6 1570.1 1407.5 1444.5 1326.2
## [2,] 1570.1 1900.3 1390.5 1455.0 1406.0
## [3,] 1407.5 1390.5 1692.7 1250.1 1347.7
## [4,] 1444.5 1455.0 1250.1 1756.3 1264.2
## [5,] 1326.2 1406.0 1347.7 1264.2 2321.3
## Standard Deviations: 46.761 43.592 41.143 41.908 48.18

# matriz de correlação estimada
cov2cor(getVarCov(mod.pr))

## Marginal variance covariance matrix
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 1.00000 0.77023 0.73159 0.73713 0.58865
## [2,] 0.77023 1.00000 0.77531 0.79643 0.66945
## [3,] 0.73159 0.77531 1.00000 0.72505 0.67985
## [4,] 0.73713 0.79643 0.72505 1.00000 0.62610
## [5,] 0.58865 0.66945 0.67985 0.62610 1.00000
## Standard Deviations: 1 1 1 1 1
```

Exercício 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.

$$E(Y_i|X_i) = X_i\beta,$$

em que $\beta = (\beta_1, \dots, \beta_{10})$ é um vetor 10×1 de coeficientes de regressão, e X_i é uma matriz de delineamento.

- X_i para um indivíduo do grupo “Alta Dose” com todas as 5 medidas repetidas (caso uma medida não for obtida, basta remover a linha referente a medida repetida):

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

- X_i para um indivíduo do grupo “Placebo” (*grupo referência*) com todas as 5 medidas repetidas (caso uma medida não for obtida, basta remover a linha referente a medida repetida):

$$X_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Exercício 8

- Seja L uma matriz de pesos conhecidos e β o vetor de parâmetros de regressão linear 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 pode ser expressa como $H_0 : L\beta = 0$. Descreva uma matriz de peso L apropriada para esta hipótese nula.

$$L = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Desta forma $H_0 : L\beta = 0$ é equivalente a testar $H_0 : \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$. Note que os parâmetros $\beta_7, \beta_8, \beta_9, \beta_{10}$ são os parâmetros referentes aos termos de interação no modelo do Exercício 7 (ajustado no Exercício 5).

Exercício 9

- Mostre como os coeficientes de regressão *estimados* de uma análise de perfis de resposta podem ser usados para construir as médias específicas do tempo nos dois grupos. Compare essas médias estimadas com as médias amostrais obtidas no Problema 5.1.2.

Seja $\mu(1) = E(Y_i|X_i)$ o vetor de médias do grupo “Alta Dose”, e $\mu(2) = E(Y_i|X_i)$ o vetor de médias do grupo “Placebo”, então

$$\mu(2) = \begin{pmatrix} \mu_1(2) \\ \mu_2(2) \\ \mu_3(2) \\ \mu_4(2) \\ \mu_5(2) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ \beta_1 + \beta_4 \\ \beta_1 + \beta_5 \end{pmatrix},$$

e

$$\mu(1) = \begin{pmatrix} \mu_1(1) \\ \mu_2(1) \\ \mu_3(1) \\ \mu_4(1) \\ \mu_5(1) \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_6 \\ (\beta_1 + \beta_6) + (\beta_2 + \beta_7) \\ (\beta_1 + \beta_6) + (\beta_3 + \beta_8) \\ (\beta_1 + \beta_6) + (\beta_4 + \beta_9) \\ (\beta_1 + \beta_6) + (\beta_5 + \beta_{10}) \end{pmatrix}.$$

Substituindo os coeficientes pelas suas respectivas estimativas, obtemos estimativas para as médias de cada grupo em cada ocasião de medição.

```
# coeficientes estimados

beta.chap <- as.numeric(coef(mod.pr))
media.chap <- c(beta.chap[1],
               beta.chap[1] + beta.chap[2:5],
               (beta.chap[1] + beta.chap[6]),
               (beta.chap[1] + beta.chap[6]) + (beta.chap[2:5] + beta.chap[7:10])
               )

# Média observada
colesterol.media <- colesterol.longo %>%
  group_by(group, meses) %>%
  summarise(media_obs = mean(col.serico, na.rm = TRUE))

colesterol.media$media_ajus <- media.chap

colesterol.media %>%
  knitr::kable(format = "markdown",
               digits = c(0, 0, 2, 2),
               col.names = c("Grupo", "Meses",
                             "Média observada", "Média ajustada"),
               align = "llcc")
```

Grupo	Meses	Média observada	Média ajustada
Placebo	0	235.93	235.93
Placebo	6	243.17	243.17
Placebo	12	244.76	244.78
Placebo	20	257.60	259.03
Placebo	24	257.48	257.05
Alta Dose	0	226.02	226.02
Alta Dose	6	245.53	245.53
Alta Dose	12	252.02	251.28
Alta Dose	20	256.80	254.10
Alta Dose	24	254.55	254.04

As estimativas das médias dos níveis séricos de colesterol obtidas pelo modelo são muito próximas das médias amostrais (coluna “Média observada”), demonstrando um ajuste adequado do modelo de perfis de respostas.

Exercício 10

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


```
knitr::kable(
  summary(mod.pr)$tTable[,-4],
  digits = c(2, 2, 2),
  col.names = c("Estimativa", "EP", "Z"),
  align = "ccc")
```

	Estimativa	EP	Z
(Intercept)	235.93	7.30	32.31
meses6	7.24	4.81	1.51
meses12	8.85	5.21	1.70
meses20	23.10	5.30	4.36
meses24	21.12	7.37	2.87
groupAlta Dose	-9.91	9.41	-1.05
meses6:groupAlta Dose	12.27	6.19	1.98
meses12:groupAlta Dose	16.42	6.74	2.43
meses20:groupAlta Dose	4.98	6.98	0.71
meses24:groupAlta Dose	6.90	9.79	0.71

- $\hat{\beta}_1 = 235.93$, corresponde ao intercepto, ou seja, à média do grupo “Placebo” na linha de base;
- $\hat{\beta}_2 = 7.24$, corresponde ao quanto a média do grupo “Placebo” no tempo “6 meses” aumenta em relação à média do próprio grupo “Placebo” na linha de base;
- $\hat{\beta}_3 = 8.85$, corresponde ao quanto a média do grupo “Placebo” no tempo “12 meses” aumenta em relação à média do próprio grupo “Placebo” na linha de base;
- $\hat{\beta}_4 = 23.1$, corresponde ao quanto a média do grupo “Placebo” no tempo “20 meses” aumenta em relação à média do próprio grupo “Placebo” na linha de base;
- $\hat{\beta}_5 = 21.12$, corresponde ao quanto a média do grupo “Placebo” no tempo “24 meses” aumenta em relação à média do próprio grupo “Placebo” na linha de base;
- $\hat{\beta}_6 = -9.91$, corresponde ao quanto a média do grupo “Alto Dose” diminui em relação ao grupo “Placebo” na linha de base;
- $\hat{\beta}_7 = 12.27$, corresponde ao quanto a média do grupo “Alta Dose” no tempo “6 meses” aumenta em relação à média do próprio grupo “Alta Dose” na linha de base, descontando o aumento esperado da linha de base para o tempo “6 meses”;
- $\hat{\beta}_8 = 16.42$, corresponde ao quanto a média do grupo “Alta Dose” no tempo “12 meses” aumenta em relação à média do próprio grupo “Alta Dose” na linha de base, descontando o aumento esperado da linha de base para o tempo “12 meses”;
- $\hat{\beta}_9 = 4.98$, corresponde ao quanto a média do grupo “Alta Dose” no tempo “20 meses” aumenta em relação à média do próprio grupo “Alta Dose” na linha de base, descontando o aumento esperado da linha de base para o tempo “20 meses”;
- $\hat{\beta}_{10} = 6.9$, corresponde ao quanto a média do grupo “Alta Dose” no tempo “24 meses” aumenta em relação à média do próprio grupo “Alta Dose” na linha de base, descontando o aumento esperado da linha de base para o tempo “24 meses”.

Exercícios do Capítulo 6

Em um estudo de ganho de peso (Box, 1950), os investigadores distribuíram aleatoriamente 30 ratos em três grupos de tratamento: o tratamento 1 era um controle (sem aditivo); os tratamentos 2 e 3 consistiram de dois aditivos diferentes (tiouracil e tiroxina, respectivamente) para a ingestão de água dos ratos. O peso, em gramas, foi medido no início (semana 0) e nas semanas 1, 2, 3 e 4. Devido a um acidente no início do estudo, os dados de 3 ratos do grupo tiroxina não estão disponíveis.

- Os dados brutos são armazenados em um arquivo externo: `rat.dta`
- Cada linha do conjunto de dados contém as seguintes sete variáveis:
 - ID Group Y1 Y2 Y3 Y4 Y5

- **Nota:** A variável Group é codificada 1 = controle, 2 = tiouracil e 3 = tiroxina.

Exercício 2

Para facilitar a resolução dos demais exercício, o exercício 2 foi resolvido previamente a estes.

- Leia os dados do arquivo externo e coloque os dados no formato “univariado” ou “longo”, com cinco “registros” por indivíduo.

```
# -----
# Carregando pacotes do R

library(here)
library(haven)
library(tidyr)
library(ggplot2)

# -----
# Carregando o arquivo de dados

rat <- read_dta(file = here::here("data",
                                "rat.dta"))

rat$group <- factor(rat$group,
                   labels = c("Controle", "Tiouracil", "Tiroxina"))

rat

## # A tibble: 27 x 7
##       id group      y1    y2    y3    y4    y5
##   <dbl> <fct>   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 1 Controle    57    86   114   139   172
## 2     2 2 Controle    60    93   123   146   177
## 3     3 3 Controle    52    77   111   144   185
## 4     4 4 Controle    49    67   100   129   164
## 5     5 5 Controle    56    81   104   121   151
## 6     6 6 Controle    46    70   102   131   153
## 7     7 7 Controle    51    71    94   110   141
## 8     8 8 Controle    63    91   112   130   154
## 9     9 9 Controle    49    67    90   112   140
## 10    10 10 Controle    57    82   110   139   169
## # ... with 17 more rows

# -----
# Reformatando o objeto: de "largo" para "longo"

rat.longo <- gather(data = rat,
                   key = "semanas",
                   value = "peso", -id, -group)

# -----
# Formata variáveis

rat.longo$semanas <- factor(rat.longo$semanas,
                          labels = 0:4)
```

```
rat.longo$semanas.num <- as.numeric(as.character(rat.longo$semanas))
rat.longo$tempo <- as.numeric(rat.longo$semanas)
```

```
rat.longo
```

```
## # A tibble: 135 x 6
##       id group   semanas  peso semanas.num tempo
##   <dbl> <fct>   <fct>   <dbl>      <dbl> <dbl>
## 1     1     1 Controle 0       57         0     1
## 2     2     2 Controle 0       60         0     1
## 3     3     3 Controle 0       52         0     1
## 4     4     4 Controle 0       49         0     1
## 5     5     5 Controle 0       56         0     1
## 6     6     6 Controle 0       46         0     1
## 7     7     7 Controle 0       51         0     1
## 8     8     8 Controle 0       63         0     1
## 9     9     9 Controle 0       49         0     1
## 10    10    10 Controle 0       57         0     1
## # ... with 125 more rows
```

Exercício 1

- Em um único gráfico, construa um gráfico de tempo que exiba o peso médio em relação ao tempo (em semanas) para os três grupos. Descreva as características gerais das tendências temporais para os três grupos.

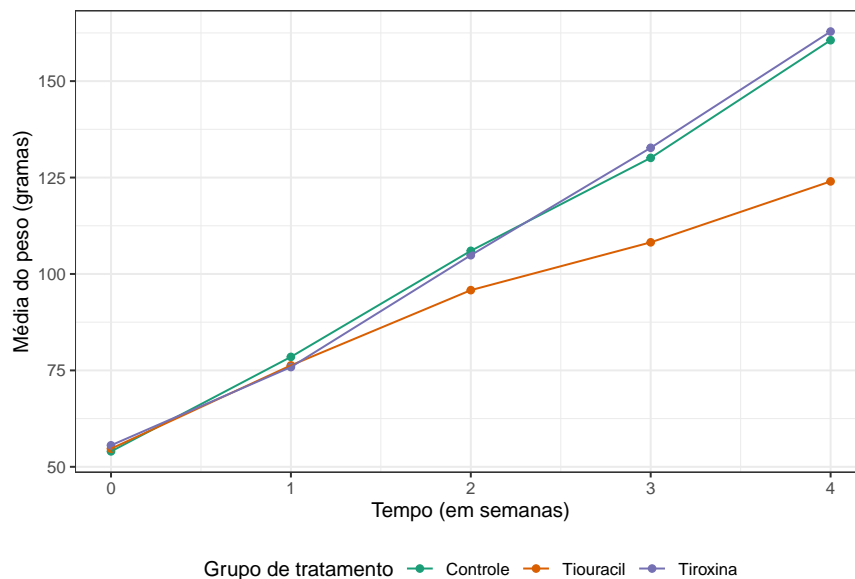
```
library(dplyr)
```

```
rat.resumo <- rat.longo %>%
  group_by(group, semanas.num) %>%
  summarise(peso_media = mean(peso, na.rm = TRUE))
```

```
rat.resumo
```

```
## # A tibble: 15 x 3
## # Groups:   group [3]
##       group   semanas.num peso_media
##   <fct>         <dbl>      <dbl>
## 1 Controle         0         54
## 2 Controle         1        78.5
## 3 Controle         2        106
## 4 Controle         3        130.
## 5 Controle         4        161.
## 6 Tiouracil        0         54.7
## 7 Tiouracil        1         76.3
## 8 Tiouracil        2         95.8
## 9 Tiouracil        3        108.
## 10 Tiouracil       4        124
## 11 Tiroxina         0         55.6
## 12 Tiroxina         1         75.9
## 13 Tiroxina         2        105.
## 14 Tiroxina         3        133.
## 15 Tiroxina         4        163.
```

```
p <- ggplot(data = rat.resumo,
            mapping = aes(x = semanas.num, y = peso_media,
                          colour = group)) +
  geom_point() + geom_line() +
  scale_color_brewer(palette = "Dark2") +
  scale_x_continuous(breaks = 0:4) +
  labs(x = "Tempo (em semanas)",
       y = "Média do peso (gramas)",
       colour = "Grupo de tratamento") +
  theme_bw() + theme(legend.position = "bottom")
p
```



Os três grupos de tratamento apresentam trajetórias de peso crescente. Os grupos Controle e Tiroxina apresentam comportamentos semelhantes, enquanto que o grupo Tiroxina parece ter uma taxa de crescimento distinto dos dois outros grupos, a partir segunda semana.

Exercício 3

- Assume-se que a taxa de crescimento em cada grupo é, aproximadamente, constante durante toda a duração do estudo. Assumindo uma matriz de covariância não estruturada, construa um teste para saber se a taxa de aumento difere nos grupos.

```
library(nlme)

# matriz de covariância não estruturada
mod.cp <- gls(peso ~ semanas.num + semanas.num:group,
              na.action = na.omit,
              corr = corSymm(form = ~ tempo | id),
              weights = varIdent(form = ~ 1 | tempo),
              method = "REML",
              data = rat.longo)

summary(mod.cp)

## Generalized least squares fit by REML
```

```

## Model: peso ~ semanas.num + semanas.num:group
## Data: rat.longo
##      AIC      BIC    logLik
## 860.0732 914.7019 -411.0366
##
## Correlation Structure: General
## Formula: ~tempo | id
## Parameter estimate(s):
## Correlation:
## 1      2      3      4
## 2 0.810
## 3 0.668 0.822
## 4 0.483 0.491 0.820
## 5 0.309 0.255 0.666 0.916
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | tempo
## Parameter estimates:
##      1      2      3      4      5
## 1.000000 1.851540 2.111338 2.935646 3.761911
##
## Coefficients:
##                               Value Std.Error  t-value p-value
## (Intercept)                55.51380 0.8189181 67.78920  0.000
## semanas.num                 26.21509 1.0163220 25.79408  0.000
## semanas.num:groupTiouracil -7.09640 1.4132182 -5.02144  0.000
## semanas.num:groupTiroxina  -2.09436 1.5572906 -1.34488  0.181
##
## Correlation:
##                               (Intr) smns.n smns.nm:grpTrc
## semanas.num                  0.182
## semanas.num:groupTiouracil   0.000 -0.695
## semanas.num:groupTiroxina   0.000 -0.631  0.454
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.0788740 -0.7678052 -0.1391835  0.6333316  2.1493435
##
## Residual standard error: 4.576419
## Degrees of freedom: 135 total; 131 residual

```

Vamos testar a hipótese de ausência de efeito de interação *grupo* \times *tempo*:

$$H_{01} : \beta_3 = \beta_4 = 0.$$

```
library(car)
```

```
Anova(mod.cp)
```

```

## Analysis of Deviance Table (Type II tests)
##
## Response: peso
##      Df    Chisq Pr(>Chisq)
## semanas.num      1 1313.857 < 2.2e-16 ***

```

```
## semanas.num:group  2    26.312  1.933e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assim, ao nível de 5% de significância, rejeitamos a hipótese de ausência de interação. Ou seja, a taxa de crescimento deve diferir entre os grupos de tratamento.

Exercício 4

Em um único gráfico, construa um gráfico de tempo que exiba o peso médio estimado em relação ao tempo (em semanas) para os três grupos de tratamento a partir dos resultados gerados no Problema 6.1.3.

```
library(ggeffects)

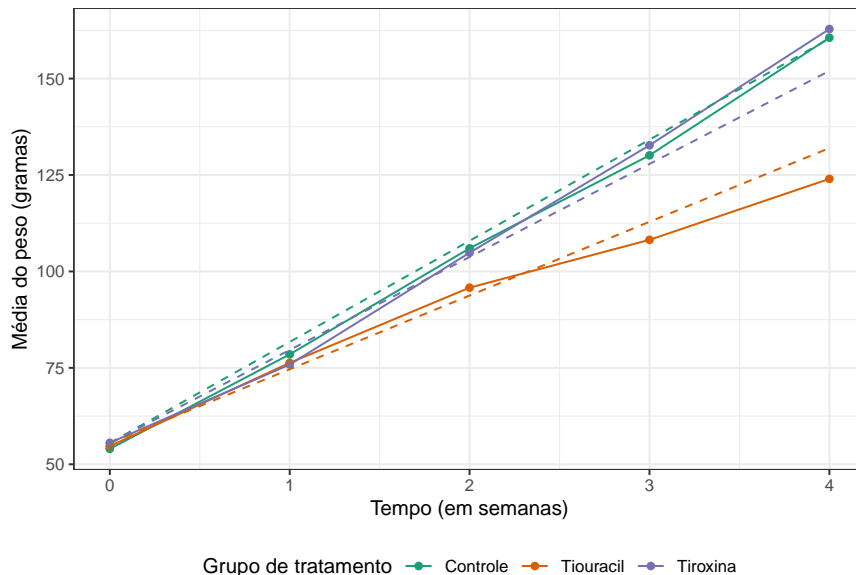
media_chap_df <- ggpredict(mod.cp, terms = c("semanas.num", "group"))
media_chap_df
```

```
##
## # Predicted values of id
## # x = semanas.num
##
## # group = Controle
##
## x | Predicted | SE | 95% CI
## -----
## 0 | 55.51 | 0.82 | [53.91, 57.12]
## 1 | 81.73 | 3.30 | [75.26, 88.19]
## 2 | 107.94 | |
## 3 | 134.16 | |
## 4 | 160.37 | |
##
## # group = Tiouracil
##
## x | Predicted | SE | 95% CI
## -----
## 0 | 55.51 | 1.42 | [52.74, 58.29]
## 1 | 74.63 | 4.29 | [66.22, 83.04]
## 2 | 93.75 | |
## 3 | 112.87 | |
## 4 | 131.99 | |
##
## # group = Tiroxina
##
## x | Predicted | SE | 95% CI
## -----
## 0 | 55.51 | 2.33 | [50.96, 60.07]
## 1 | 79.63 | |
## 2 | 103.76 | |
## 3 | 127.88 | |
## 4 | 152.00 | |
##
## Adjusted for:
## * tempo = 3.00
## * id = 14.00
```

```
df_aux <- as.data.frame(media_chap_df)[c(6, 1, 2)]
names(df_aux) <- c("group", "semanas.num", "peso_media")

p2 <- p + geom_line(data = df_aux,
                    linetype = "dashed")
```

p2



Exercício 5

- Com base nos resultados do Problema 6.1.3, qual é a taxa estimada de aumento do peso médio no grupo de controle (grupo 1)? Qual é a taxa estimada de aumento do peso médio no grupo tiouracil (grupo 2)? Qual é a taxa estimada de aumento do peso médio no grupo de tiroxina (grupo 3)?
- Taxa de crescimento estimada do grupo Controle é $\hat{\beta}_2 = 26.22$ gramas por semana.
- Taxa de crescimento estimada do grupo Tiouracil é $\hat{\beta}_2 + \hat{\beta}_3 = 19.12$ gramas por semana.
- Taxa de crescimento estimada do grupo Tiroxina é $\hat{\beta}_2 + \hat{\beta}_4 = 24.12$ gramas por semana.

Exercício 6

- Os investigadores do estudo conjecturaram que haveria um aumento no peso, mas que a taxa de aumento se estabilizaria no final do estudo. Eles também conjecturaram que esse padrão de mudança pode ser diferente nos três grupos de tratamento. Assumindo uma matriz de covariância não estruturada, construa um teste dessa hipótese.

Foi ajustado um modelo linear por partes com mudança de inclinação no ponto $t^* = 2$ semanas.

```
# modelo de curvas paramétricas
# splines lineares
# com matriz de covariância não estruturada
mod.spline <- gls(peso ~ semanas.num + I( (semanas.num - 2) * (semanas.num > 2) ) +
                  semanas.num:group + I( (semanas.num - 2) * (semanas.num > 2) ):group,
                  corr = corSymm(form = ~ tempo | id),
                  weights = varIdent(form = ~ 1 | tempo),
                  method = "REML",
```

```

data = rat.longo)

summary(mod.spline)

## Generalized least squares fit by REML
## Model: peso ~ semanas.num + I((semanas.num - 2) * (semanas.num > 2)) +      semanas.num:group + I(
## Data: rat.longo
##      AIC      BIC    logLik
## 842.3209 905.0656 -399.1604
##
## Correlation Structure: General
## Formula: ~tempo | id
## Parameter estimate(s):
## Correlation:
##  1      2      3      4
## 2 0.832
## 3 0.699 0.821
## 4 0.485 0.528 0.874
## 5 0.297 0.293 0.726 0.908
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | tempo
## Parameter estimates:
##      1      2      3      4      5
## 1.000000 1.853158 2.132319 2.934791 3.644738
##
## Coefficients:
##                                     Value Std.Error
## (Intercept)                      55.55870 0.8189965
## semanas.num                      26.57231 1.1034227
## I((semanas.num - 2) * (semanas.num > 2)) -1.03027 1.2242546
## semanas.num:groupTiouracil        -5.43092 1.5278447
## semanas.num:groupTiroxina         -2.45832 1.6836028
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil -4.94356 1.7235228
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina  1.08029 1.8992296
##                                     t-value p-value
## (Intercept)                      67.83753 0.0000
## semanas.num                      24.08172 0.0000
## I((semanas.num - 2) * (semanas.num > 2)) -0.84155 0.4016
## semanas.num:groupTiouracil        -3.55463 0.0005
## semanas.num:groupTiroxina         -1.46015 0.1467
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil -2.86829 0.0048
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina  0.56881 0.5705
##
## Correlation:
##                                     (Intr) smns.n
## semanas.num                      0.203
## I((semanas.num - 2) * (semanas.num > 2)) -0.095 -0.390
## semanas.num:groupTiouracil        0.000 -0.692
## semanas.num:groupTiroxina        0.000 -0.628
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil 0.000 0.263
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina 0.000 0.239
##                                     I((s.-2)*(.>2))
## semanas.num

```



```
## I((semanas.num - 2) * (semanas.num > 2))
## semanas.num:groupTiouracil 0.268
## semanas.num:groupTiroxina 0.243
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil -0.704
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina -0.639
## smns.nm:grpTrc
## semanas.num
## I((semanas.num - 2) * (semanas.num > 2))
## semanas.num:groupTiouracil
## semanas.num:groupTiroxina 0.454
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil -0.380
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina -0.172
## smns.nm:grpTrc
## semanas.num
## I((semanas.num - 2) * (semanas.num > 2))
## semanas.num:groupTiouracil
## semanas.num:groupTiroxina
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil -0.172
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina -0.380
## I((smns.nm-2)*(smns.nm>2)):grpTrc
## semanas.num
## I((semanas.num - 2) * (semanas.num > 2))
## semanas.num:groupTiouracil
## semanas.num:groupTiroxina
## I((semanas.num - 2) * (semanas.num > 2)):groupTiouracil
## I((semanas.num - 2) * (semanas.num > 2)):groupTiroxina 0.454
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.0847866 -0.8149819 -0.2000896  0.5465275  2.2072386
##
## Residual standard error: 4.584979
## Degrees of freedom: 135 total; 128 residual
```

```
Anova(mod.spline)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: peso
##
##              Df    Chisq Pr(>Chisq)
## semanas.num    1 1185.777 < 2.2e-16 ***
## I((semanas.num - 2) * (semanas.num > 2))    1   11.820  0.0005859 ***
## semanas.num:group    2   12.665  0.0017778 **
## I((semanas.num - 2) * (semanas.num > 2)):group    2   12.632  0.0018073 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ao nível de 5% de significância rejeitamos a hipótese de ausência de interação entre tempo e grupo de tratamento. Assim, a conjectura dos pesquisadores deve estar correta.

Exercício 7

- Compare e contraste os resultados dos Problemas 6.1.3 e 6.1.6. Um modelo com apenas uma tendência linear no tempo considera adequadamente o padrão de mudança nos três grupos de tratamento? Forneça resultados que apoiem sua conclusão.

Note que o modelo de **tendências lineares** é um caso particular (encaixado) do modelo linear por partes, quando fazemos $\beta_3 = \beta_6 = \beta_7 = 0$. Podemos utilizar o **teste da razão de verossimilhanças** para comparar os dois modelos. **Importante:** a construção de testes de razão de verossimilhanças comparando modelos encaixados para a média deve sempre ser baseada na log-verossimilhança MV, e não no REML.

```
# modelo linear por partes
mod.cp <- update(mod.cp, method = "ML")

# modelo de tendências lineares
mod.spline <- update(mod.spline, method = "ML")

anova(mod.spline, mod.cp)

##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod.spline     1 22 856.3033 920.2194 -406.1517
## mod.cp         2 19 867.1784 922.3786 -414.5892 1 vs 2 16.87511 7e-04
```

Ao nível de 5% de significância, o teste indica que o modelo linear por partes melhora o ajuste global da resposta média ao longo do tempo quando comparado com o modelo linear.

Exercício 8

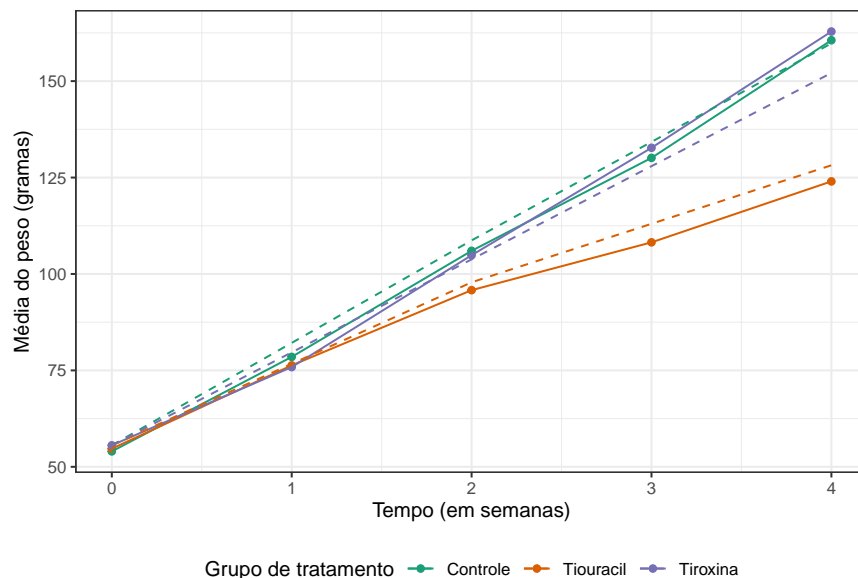
- Dado os resultados de todas as análises anteriores, que conclusões podem ser tiradas sobre o efeito dos aditivos nos padrões de mudança de peso?

```
media_chap_df <- ggpredict(mod.spline, terms = c("semanas.num", "group"))

df_aux <- as.data.frame(media_chap_df)[c(6, 1, 2)]
names(df_aux) <- c("group", "semanas.num", "peso_media")

p3 <- p + geom_line(data = df_aux,
                    linetype = "dashed")

p3
```



Os resultados anteriores nos permitem concluir que:

1. O modelo de tendências lineares por partes com um nó em $t^* = 2$ apresenta um ajuste melhor que o modelo linear de taxa de crescimento constante.
2. Os grupos Controle e Tiroxina não diferem significativamente no ganho de peso ao longo do tempo, e apresentam taxa de crescimento constante ao longo do tempo.
3. O grupo Tiouracil difere significativamente no ganho de peso dos outros grupos, apresentam diminuição da taxa de crescimento após 2 semanas de estudo.
 - Assim, o aditivo Tiouracil tem influência no ganho de peso de ratos.

Exercícios do Capítulo 7

Em um estudo de crescimento dentário, as medidas da distância (mm) do centro da glândula pituitária à fissura pteromaxilar foram obtidas em 11 meninas e 16 meninos nas idades de 8, 10, 12 e 14 (Potthoff e Roy, 1964).

- Os dados brutos são armazenados em um arquivo externo: `dental.dta`.
- Cada linha do conjunto de dados contém as seguintes seis variáveis:
 - ID Gender Y1 Y2 Y3 Y4
- **Nota:** A variável categórica `Gender` é codificada F = Feminino, M = Masculino.
- A terceira medida (aos 12 anos) no indivíduo ID = 20 é um potencial outlier.

```
# -----
# Carregando pacotes do R

library(here)
library(haven)
library(tidyr)
library(ggplot2)

# -----
# Carregando o arquivo de dados

dental <- read_dta(file = here::here("data",
                                     "dental.dta"))

dental$gender <- factor(dental$gender,
                        labels = c("Feminino", "Masculino"))

dental

## # A tibble: 27 x 6
##       id gender    y1    y2    y3    y4
##   <dbl> <fct>    <dbl> <dbl> <dbl> <dbl>
## 1     1 Feminino  21    20   21.5  23
## 2     2 Feminino  21   21.5  24   25.5
## 3     3 Feminino 20.5  24   24.5  26
## 4     4 Feminino 23.5  24.5  25   26.5
## 5     5 Feminino 21.5  23   22.5  23.5
## 6     6 Feminino  20    21    21   22.5
## 7     7 Feminino 21.5  22.5  23    25
## 8     8 Feminino  23    23   23.5  24
## 9     9 Feminino  20    21    22   21.5
## 10    10 Feminino 16.5  19    19   19.5
## # ... with 17 more rows

# -----
# Reformatando o objeto: de "largo" para "longo"
```

```
dental.longo <- gather(data = dental,
  key = "idade",
  value = "distancia", -id, -gender)

# -----
# Formata variáveis

dental.longo$idade <- factor(dental.longo$idade,
  labels = c(8, 10, 12, 14))

dental.longo$idade.num <- as.numeric(as.character(dental.longo$idade))
dental.longo$tempo <- as.numeric(dental.longo$idade)

dental.longo
```

```
## # A tibble: 108 x 6
##       id gender  idade distancia idade.num tempo
##   <dbl> <fct>   <fct>      <dbl>      <dbl> <dbl>
## 1     1  1 Feminino 8         21         8      1
## 2     2  2 Feminino 8         21         8      1
## 3     3  3 Feminino 8        20.5         8      1
## 4     4  4 Feminino 8        23.5         8      1
## 5     5  5 Feminino 8        21.5         8      1
## 6     6  6 Feminino 8         20         8      1
## 7     7  7 Feminino 8        21.5         8      1
## 8     8  8 Feminino 8         23         8      1
## 9     9  9 Feminino 8         20         8      1
## 10    10 10 Feminino 8        16.5         8      1
## # ... with 98 more rows
```

Exercício 1

Em um único gráfico, construa um gráfico de tempo que exiba a distância média (mm) versus a idade (em anos) para meninos e meninas. Descreva as tendências temporais para meninos e meninas.

```
library(dplyr)

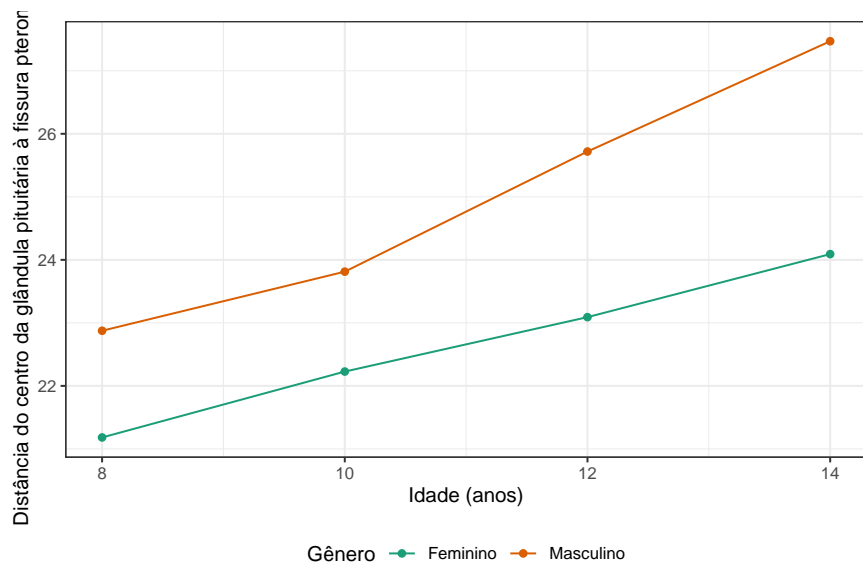
dental.resumo <- dental.longo %>%
  group_by(gender, idade.num) %>%
  summarise(distancia = mean(distancia))

dental.resumo
```

```
## # A tibble: 8 x 3
## # Groups:   gender [2]
##   gender  idade.num distancia
##   <fct>      <dbl>      <dbl>
## 1 Feminino      8      21.2
## 2 Feminino     10      22.2
## 3 Feminino     12      23.1
## 4 Feminino     14      24.1
## 5 Masculino      8      22.9
## 6 Masculino     10      23.8
## 7 Masculino     12      25.7
```

```
## 8 Masculino      14      27.5
```

```
p <- ggplot(data = dental.resumo,
            mapping = aes(x = idade.num, y = distancia,
                          colour = gender)) +
  geom_point() + geom_line() +
  scale_color_brewer(palette = "Dark2") +
  scale_x_continuous(breaks = c(8, 10, 12, 14)) +
  labs(x = "Idade (anos)",
       y = "Distância do centro da glândula pituitária à fissura pteromaxilar",
       colour = "Gênero", caption = "Médias por gênero e idade.") +
  theme_bw() + theme(legend.position = "bottom")
p
```



Médias por gênero e idade.

Meninas e meninos parecem ter tendência de crescimento linear entre 8 e 14 anos, com meninos apresentando maiores medidas de *distância do centro da glândula pituitária à fissura pteromaxilar* (em média) ao longo do tempo de estudo em relação às meninas.

Exercício 2

Realizado antes do Exercício 1.

Exercício 3

Modelo maximal para média:

$$\begin{aligned}
 E(Y_{ij}) = & \beta_1 + \beta_2 I_{\{idade=10\}} + \beta_3 I_{\{idade=12\}} + \beta_4 I_{\{idade=14\}} \\
 & + \beta_5 I_{\{genero=M\}} + \beta_6 I_{\{idade=10\} \cap \{genero=M\}} \\
 & + \beta_7 I_{\{idade=12\} \cap \{genero=M\}} + \beta_8 I_{\{idade=14\} \cap \{genero=M\}}
 \end{aligned}$$

```
library(nlme)
```

```
# matriz de covariância não estruturada
mod.ne <- gls(distancia ~ idade * gender,
```

```

na.action = na.omit,
corr = corSymm(form = ~ tempo | id),
weights = varIdent(form = ~ 1 | tempo),
method = "REML",
data = dental.longo)

knitr::kable(
  summary(mod.ne)$tTable,
  digits = c(4, 4, 2, 4),
  col.names = c("Estimativa", "EP", "Z", "Valor p"))

```

	Estimativa	EP	Z	Valor p
(Intercept)	21.1818	0.7017	30.19	0.0000
idade10	1.0455	0.6155	1.70	0.0925
idade12	1.9091	0.6068	3.15	0.0022
idade14	2.9091	0.6729	4.32	0.0000
genderMasculino	1.6932	0.9115	1.86	0.0662
idade10:genderMasculino	-0.1080	0.7995	-0.14	0.8929
idade12:genderMasculino	0.9347	0.7883	1.19	0.2386
idade14:genderMasculino	1.6847	0.8741	1.93	0.0568

Como o efeito de interação não foi estatisticamente diferente de zero, ajustamos o seguinte modelo com os efeitos principais de *idade* e *genero*:

$$\begin{aligned}
 E(Y_{ij}) = & \beta_1 + \beta_2 I_{\{idade=10\}} + \beta_3 I_{\{idade=12\}} + \beta_4 I_{\{idade=14\}} \\
 & + \beta_5 I_{\{genero=M\}}
 \end{aligned}$$

```

# matriz de covariância simetria composta
mod.sc <- gls(distancia ~ idade * gender,
  na.action = na.omit,
  corr = corCompSymm(form= ~ tempo | id),
  method = "REML",
  data = dental.longo)

# matriz de covariância simetria composta
# e var. heterogênea
mod.sch <- gls(distancia ~ idade * gender,
  na.action = na.omit,
  corr = corCompSymm(form= ~ tempo | id),
  weights = varIdent(form = ~ 1 | tempo),
  method = "REML",
  data = dental.longo)

# matriz de covariância autorregressiva
mod.ar <- gls(distancia ~ idade * gender,
  na.action = na.omit,
  corr = corAR1(form = ~ tempo | id),
  method = "REML",
  data = dental.longo)

# matriz de covariância autorregressiva

```

```
# e var. heterogênea
mod.arh <- gls(distancia ~ idade * gender,
  na.action = na.omit,
  corr = corAR1(form = ~ tempo | id),
  weights = varIdent(form = ~ 1 | tempo),
  method = "REML",
  data = dental.longo)
```

Exercício 4

(Inclua aqui o texto necessário para você responder a essa pergunta.)

```
# inclua aqui o código
```