Tarefa 3

Rafael Chaves - 00275981

02/04/2021

Foram escolhidos os exercícios do capítulo 5(páginas 140 e 141), usando o arquivo *cholesterol.dta*.

Problems

5.1 -In the National Cooperative Gallstone Study (NCGS), one of the major interests was to study the safety of the drug chenodiol for the treatment of cholesterol gallstones (Schoenfield et al., 1981; Wei and Lachin, 1984). In this study, patients were randomly assigned to high-dose (750 mg per day), low-dose (375 mg per day), or placebo. We focus on a subset of data on patients who had floating gallstones and who were assigned to the high-dose and placebo groups. In the NCGS it was suggested that chenodiol would dissolve gallstones but, in doing so, might increase levels of serum cholesterol. As a result serum cholesterol (mg/dL) was measured at baseline and at 6, 12, 20, and 24 months of follow-up. Many cholesterol measurements are missing because of missed visits, laboratory specimens were lost or inadequate, or patient follow-up was terminated. The NCGS serum cholesterol data are stored in an external file: cholesterol.dat Each row of the data set contains the following seven variables: Group ID Y1 Y2 Y3 Y4 Y5 Note: The categorical variable Group is coded 1 = High-Dose, 2 = Placebo.

5.1.1 Read the data from the external file and keep it in a "multivariate" or "wide" format.

```
colesterol.df <- read_dta(file = here::here("/home/rafaelbc/Documentos/Estudo/Facudade/202002/Dados_Cor
## Alterando pra highdose ou placebo
colesterol.df<-colesterol.df%>%
  mutate(group=case_when(group==1~"High-Dose",
                         group==2~"Placebo"))
head(colesterol.df,n = 10) # Mostrando apenas as 10 primeiras linhas do dataframe
```

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

5.1.2 Calculate the sample means, standard deviations, and variances of the serum cholesterol levels at each occasion for each treatment group.

Agrupando os valores observados por grupos, as medidas para os níveis de colesterol em cada um dos períodos medidos são:

y1-primeira medida

```
## # A tibble: 2 x 4
##   group     Média  Desvio_Padrão Variância
##   <chr>     <chr>  <chr>         <chr>
## 1 High-Dose 226.02 39.66         1573.26
## 2 Placebo   235.93 55.87         3121.97
```

y2-6 meses depois

```
## # A tibble: 2 x 4
##   group     Média  Desvio_Padrão Variância
##   <chr>     <chr>  <chr>         <chr>
## 1 High-Dose 245.53 39.45         1556.48
## 2 Placebo   243.17 49.24         2424.55
```

y3-12 meses depois

```
## # A tibble: 2 x 4
##   group     Média Desvio_Padrão Variância
##   <chr>     <chr> <chr>         <chr>
## 1 High-Dose NA    NA            NA
## 2 Placebo   NA    NA            NA
```

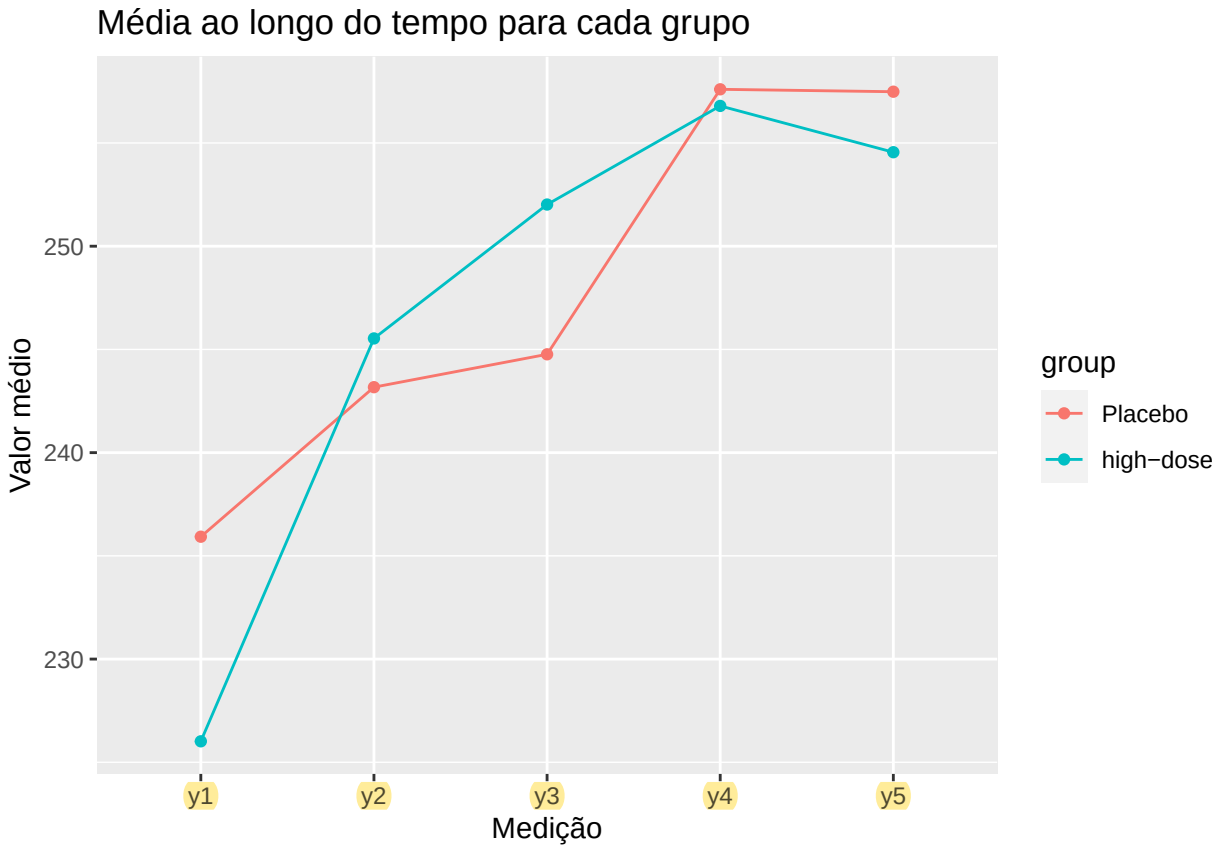
y4-20 meses depois

```
## # A tibble: 2 x 4
##   group     Média Desvio_Padrão Variância
##   <chr>     <chr> <chr>         <chr>
## 1 High-Dose NA    NA            NA
## 2 Placebo   NA    NA            NA
```

y5-24 meses depois

```
## # A tibble: 2 x 4
##   group     Média Desvio_Padrão Variância
##   <chr>     <chr> <chr>         <chr>
## 1 High-Dose NA    NA            NA
## 2 Placebo   NA    NA            NA
```

5.1.3 On a single graph, construct a time plot that displays the mean serum cholesterol versus time (in months) for the two treatment group. Describe the general characteristics of the time trends for the two groups.



Podemos ver que a média de soro colesterol em relação ao tempo foi maior para o grupo *Placebo*, enquanto a média para o grupo que recebeu a dose alta, $High - dose$ teve uma queda a partir dos 20 meses.

5.1.4 Next read the data from the external file and put the data in a "univariate" or "long" format, with five "records" per subject.

```
## # A tibble: 10 x 4
##    group        id variable value
##    <fct>     <dbl> <chr>    <dbl>
##  1 high-dose     1 y1         178
##  2 high-dose     2 y1         254
##  3 high-dose     3 y1         185
##  4 high-dose     4 y1         219
##  5 high-dose     5 y1         205
##  6 high-dose     6 y1         182
##  7 high-dose     7 y1         310
##  8 high-dose     8 y1         191
##  9 high-dose     9 y1         245
## 10 high-dose    10 y1         229
```

Carregados as primeiras 10 linhas do datafame.

5.1.5 Assuming an unstructured covariance matrix, conduct an analysis of response profiles. Determine whether the patterns of change over time differ in the two treatment groups.

```
library(nlme)
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
colesterol.nf<-colesterol.df.longo[complete.cases(colesterol.df.longo),]
mod.pr <- gls(value ~ group * variable,
              corr = corSymm(form = ~ 1 | id),
              weights = varIdent(form = ~ 1 | variable),
              method = "REML",
              data = colesterol.nf)
summary(mod.pr)
```

```
## Generalized least squares fit by REML
##   Model: value ~ group * variable
##   Data: colesterol.nf
##        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 | variable
##  Parameter estimates:
##        y1        y2        y3        y4        y5
## 1.0000000 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
## grouphigh-dose             -9.67829  9.412956 -1.02819  0.3044
## variabley2                  7.24390  4.805425  1.50744  0.1324
## variabley3                  8.84620  5.207262  1.69882  0.0901
## variabley4                 23.10333  5.292171  4.36557  0.0000
## variabley5                 21.12230  7.398137  2.85508  0.0045
## grouphigh-dose:variabley2  12.21751  6.193407  1.97266  0.0492
## grouphigh-dose:variabley3  16.28893  6.738391  2.41733  0.0160
## grouphigh-dose:variabley4   4.75670  6.973253  0.68213  0.4955
## grouphigh-dose:variabley5   6.53598  9.763271  0.66945  0.5036
##
```

```
##  Correlation:
##                             (Intr) grphg- vrbly2 vrbly3 vrbly4 vrbly5 grp-:2
## grouphigh-dose              -0.776
## variabley2                  -0.429  0.333
## variabley3                  -0.500  0.388  0.581
## variabley4                  -0.466  0.362  0.606  0.526
## variabley5                  -0.392  0.304  0.476  0.522  0.438
## grouphigh-dose:variabley2    0.333 -0.429 -0.776 -0.451 -0.470 -0.369
## grouphigh-dose:variabley3    0.387 -0.497 -0.449 -0.773 -0.407 -0.404  0.578
## grouphigh-dose:variabley4    0.354 -0.456 -0.460 -0.400 -0.759 -0.332  0.592
## grouphigh-dose:variabley5    0.297 -0.378 -0.361 -0.396 -0.332 -0.758  0.463
##                             grp-:3 grp-:4
## grouphigh-dose
## variabley2
## variabley3
## variabley4
## variabley5
## grouphigh-dose:variabley2
## grouphigh-dose:variabley3
## grouphigh-dose:variabley4    0.513
## grouphigh-dose:variabley5    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
```

Concluímos que existe uma diferença siginificativa entre os dois grupos analisando os coeficientes. Os tempos de medição iniciais tem um aumento maior da média do grupo $High - Dose$, enquanto nos dois últimos temos um aumento maior da média do grupo placebo.

5.1.6 Display the estimated 5 x 5 covariance and correlation matrices for the five repeated measurements of serum cholesterol.

```
knitr::kable(x=matrix(getVarCov(mod.pr),ncol=4),digits=1)
```

Matriz de covariância

```
## Warning in matrix(getVarCov(mod.pr), ncol = 4): comprimento dos dados [25] não é
## um submúltiplo ou múltiplo do número de linhas [7]
```

```
## Warning in kable_pipe(x = structure(c("2188.5", "1571.4", "1407.9", "1449.2", :
## The table should have a header (column names)
```

2188.5	1387.1	1343.3	1397.5
1571.4	1463.7	1449.2	1343.3
1407.9	1397.5	1463.7	1264.2
1449.2	1407.9	1254.4	2322.1
1320.7	1387.1	1762.4	2188.5
1571.4	1691.5	1264.2	1571.4
1901.2	1254.4	1320.7	1407.9

```
mod.pr$modelStruct$corStruct
```

Matriz de correlação

```
## 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 With baseline (month 0) and the placebo group (group 2) as the reference group, write out the regression model for mean serum cholesterol that corresponds to the analysis of response profiles in Problem 5.1.5.

Considerando os coeficientes gerados nas saídas anteriores, a nossa regressão considerando o grupo placebo como o grupo de referência será representada por:

$Y = 235.92683 - 9.67829 * grouphigh - dose + 7.24390 * variabley2 + 8.84620 * variabley3 + 23.10333 * variabley4 + 21.12230 * variabley5 + 12.21751 * grouphigh - dose : variabley2 + 16.28893 * grouphigh - dose : variabley4 + 4.75670 * grouphigh - dose : variabley3 + 6.53598 * grouphigh - dose : variabley5$

5.1.8 Let L denote a matrix of known weights and β3 the vector of linear regression parameters from the model assumed in Problem 5.1.7. The null hypothesis that the patterns of change over time do not differ in the two treatment groups can be expressed as $H0 : L\beta 3 = 0$. Describe an appropriate weight matrix L for this null hypothesis.

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
Anova(mod.pr)
```

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

Podemos ver que apenas o efeito das medições ao longo do tempo foi significativo, enquanto a interação entre grupos e o efeito entre grupos não foi estatísticamente significativo.

5.1.9 Show how the estimated regression coefficients from an analysis of response profiles can be used to construct the time-specific means in the two groups. Compare these estimated means with the sample means obtained in Problem 5.1.2.

```
library(dplyr)
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:purrr':
##
##     pluck
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
coefplac <- matrix(c(235.926829,7.243902,8.846205,23.103334,21.122304),ncol=1)
covplac <- matrix(c(rep(1,5),c(0,6,12,20,24),rep(0,5)),ncol=3)
mean_plac <- getVarCov(mod.pr)%*%coefplac
mean_plac
```

```
##          [,1]
## [1,] 601530.5
## [2,] 460118.5
## [3,] 414529.3
## [4,] 431027.5
## [5,] 411850.8
```

```
coefhd <- c(-9.678291,12.217511,16.288928,4.756696,6.535983)
mean_hd <- getVarCov(mod.pr)%*%coefhd
mean_hd
```

```
##          [,1]
## [1,] 36477.39
## [2,] 46709.45
## [3,] 45620.05
## [4,] 40934.73
## [5,] 47363.38
```

5.1.10 With baseline (month 0) and the placebo group (group 2) as the reference group, provide an interpretation for each of the estimated regression coefficients in terms of the effect of the treatments on the patterns of change in mean serum cholesterol.

```
knitr::kable(summary(mod.pr)$tTable[,-4],
             digits = c(3, 3, 2),
col.names = c("Estimativa", "EP", "Z"))
```

	Estimativa	EP	Z
(Intercept)	235.927	7.306	32.29
grouphigh-dose	-9.678	9.413	-1.03
variabley2	7.244	4.805	1.51
variabley3	8.846	5.207	1.70
variabley4	23.103	5.292	4.37
variabley5	21.122	7.398	2.86
grouphigh-dose:variabley2	12.218	6.193	1.97
grouphigh-dose:variabley3	16.289	6.738	2.42
grouphigh-dose:variabley4	4.757	6.973	0.68
grouphigh-dose:variabley5	6.536	9.763	0.67

Considerando o grupo placebo como nível de referência para o grupo tratamento e as estimativas de β da tabela anterior para interpretar os coeficientes em relação ao efeito dos tratamentos sobre a mudança no colesterol sérico.

Os indivíduos que receberam altas doses da droga apresentaram um aumento maior na média e colesterol nas duas primeiras ocasiões, em relação aos indivíduos que receberam o placebo.

Tendo o grupo placebo como referência, o grupo de $High - Dose$ tem um aumento adicional de 12.218 nos níveis médios de colesterol na medição do mês 6.

No tempo até as medições seguintes, o grupo $High - Dose$ tem um aumento adicional de 16.289 nos níveis médios de colesterol do início do estudo até o mês 12; um aumento de 4.757 nos níveis médios de colesterol até o mês 20 e um aumento adicional de 6.536 nos níveis de colesterol até o mês 24.