Pontual 3

Exercícios Fitzgerald-Laird-Ware (2017), Capítulo 5 (pg. 140-141)

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

```
# Leitura dos dados e transformação da variavel "Group" em fator.
df =  read_dta("cholesterol.dta") %>%
    mutate(group = factor(group))

colnames(df)[-c(1,2)] = c('t0', 't06', 't12', 't20', 't24')

# Placebo = Grupo de referencia.
df = df %>%
    mutate(group = ifelse(group==2, 0, 1))
```

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

```
#alta dosagem

medias1 = df %>%
    filter(group == 1) %>%
    select(t0:t24) %>%
    apply(2, mean, na.rm = T)

sd1 = df %>%
    filter(group == 1) %>%
    select(t0:t24) %>%
    apply(2, sd, na.rm = T)

var1 = df %>%
    filter(group == 1) %>%
    select(t0:t24) %>%
    apply(2, var, na.rm = T)
```

Abaixo podemos ver algumas das descritivas do grupo de alta dosagem.

Grupo de Alta Dosagem

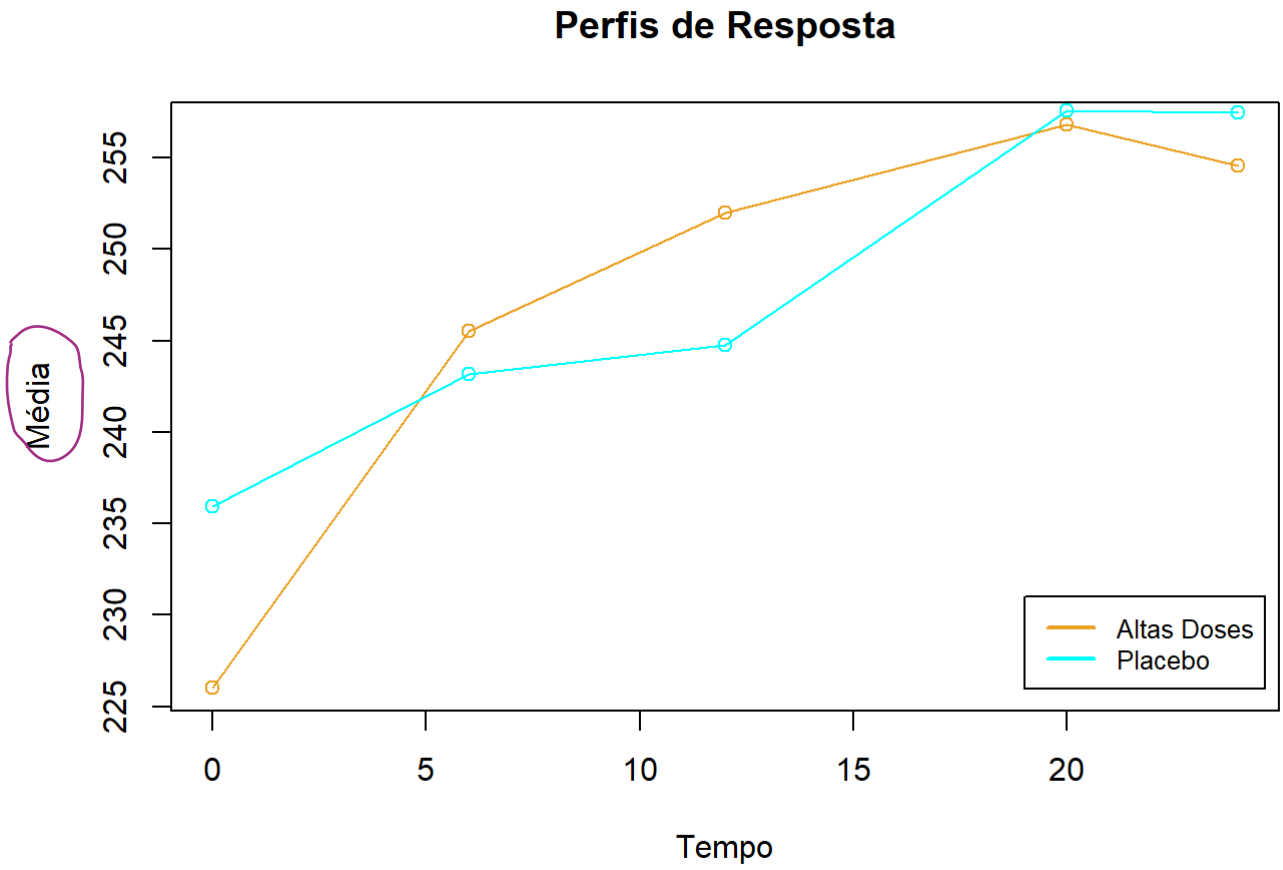
	t0	t06	t12	t20	t24
Média	226.01613	245.53226	252.01818	256.79545	254.55263
DP	39.66437	39.45228	38.32922	34.48935	49.96198
Variância	1573.26203	1556.48255	1469.12929	1189.51533	2496.19986

```
#grupo placebo

medias2 = df %>%
    filter(group == 0) %>%
    select(t0:t24) %>%
    apply(2, mean, na.rm = T)

sd2 = df %>%
    filter(group == 0) %>%
    select(t0:t24) %>%
    apply(2, sd, na.rm = T)

var2 = df %>%
    filter(group == 0) %>%
    select(t0:t24) %>%
    apply(2, var, na.rm = T)
```

Grupo Placebo

	t0	t06	t12	t20	t24
Média	235.92683	243.17073	244.76316	257.60000	257.48387
DP	55.87459	49.23967	46.11058	51.14179	49.38817
Variância	3121.96951	2424.54512	2126.18563	2615.48235	2439.19140

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.



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.

```
df_long = df %>%
    pivot_longer(
        cols = colnames(df)[-c(1,2)],
        names_to = "t" ,
        names_prefix = 't',
        values_to = 'y'
    ) %>%
    mutate(t_num = as.numeric(t), t = as.factor(t), .keep= 'unused') %>%
    na.omit
```

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.

```
mod <- gls(y ~ group * t,
              corr = corSymm(form = ~ 1 | id),
              weights = varIdent(form = ~ 1 | t),
              method = "REML",
              data = df_long)

summary(mod)
```

```
## Generalized least squares fit by REML
##   Model: y ~ group * t
##   Data: df_long
##        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 | t
##  Parameter estimates:
##         0        06        12        20        24
## 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
## group        -9.67829  9.412956 -1.02819  0.3044
## t06           7.24390  4.805425  1.50744  0.1324
## t12           8.84620  5.207262  1.69882  0.0901
## t20          23.10333  5.292171  4.36557  0.0000
## t24          21.12230  7.398137  2.85508  0.0045
## group:t06    12.21751  6.193407  1.97266  0.0492
## group:t12    16.28893  6.738391  2.41733  0.0160
## group:t20     4.75670  6.973253  0.68213  0.4955
## group:t24     6.53598  9.763271  0.66945  0.5036
##
##  Correlation:
##           (Intr) group  t06    t12    t20    t24    grp:06 grp:12 grp:20
## group     -0.776
## t06       -0.429  0.333
## t12       -0.500  0.388  0.581
## t20       -0.466  0.362  0.606  0.526
## t24       -0.392  0.304  0.476  0.522  0.438
## group:t06  0.333 -0.429 -0.776 -0.451 -0.470 -0.369
## group:t12  0.387 -0.497 -0.449 -0.773 -0.407 -0.404  0.578
## group:t20  0.354 -0.456 -0.460 -0.400 -0.759 -0.332  0.592  0.513
## group:t24  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
```

```
Anova(mod)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: y
##         Df   Chisq Pr(>Chisq)
## group    1  0.0037    0.95163
## t        4 65.4783   2.04e-13 ***
## group:t  4  7.9167    0.09468 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ou seja, a 5% de significância não rejeitamos a hipótese da ausência de efeito de interação. Portanto, não podemos dizer que os padrões de mudança mudam conforme o grupo que analisamos.

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

Matriz de Covariância:

```
covMatrix = unclass(getVarCov(mod))

covMatrix
```

```
##          [,1]     [,2]     [,3]     [,4]     [,5]
## [1,] 2188.452 1571.425 1407.913 1449.214 1320.705
## [2,] 1571.425 1901.174 1387.080 1463.678 1397.530
## [3,] 1407.913 1387.080 1691.530 1254.365 1343.293
## [4,] 1449.214 1463.678 1254.365 1762.425 1264.163
## [5,] 1320.705 1397.530 1343.293 1264.163 2322.094
```

Matriz de Correlação:

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.

```
corMatrix = matrix(nrow = 5, ncol = 5)


for(i in 1:nrow(corMatrix)){
    for( j in 1:ncol(corMatrix)){
        corMatrix[i,j] = covMatrix[i,j]/(sqrt(covMatrix[i,i] * covMatrix[j,j]))
    }
}

corMatrix
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 1.0000000 0.7703962 0.7317579 0.7379188 0.5858649
## [2,] 0.7703962 1.0000000 0.7734828 0.7996121 0.6651360
## [3,] 0.7317579 0.7734828 1.0000000 0.7264887 0.6777836
## [4,] 0.7379188 0.7996121 0.7264887 1.0000000 0.6248963
## [5,] 0.5858649 0.6651360 0.6777836 0.6248963 1.0000000
```

```
x2 = matrix(0, nrow = 5, ncol = 2*5)
rownames(x2) = colnames(df)[3:7]
x2[,1] = 1
diag(x2) = 1
x2
```

```
##     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## t0     1    0    0    0    0    0    0    0    0     0
## t06    1    1    0    0    0    0    0    0    0     0
## t12    1    0    1    0    0    0    0    0    0     0
## t20    1    0    0    1    0    0    0    0    0     0
## t24    1    0    0    0    1    0    0    0    0     0
```

```
x1 = matrix(0, nrow = 5, ncol = 5)
rownames(x1) = colnames(df)[3:7]
x1[,1] = 1
diag(x1) = 1
x1 = cbind(x1,x1)
x1
```

```
##     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## t0     1    0    0    0    0    1    0    0    0     0
## t06    1    1    0    0    0    1    1    0    0     0
## t12    1    0    1    0    0    1    0    1    0     0
## t20    1    0    0    1    0    1    0    0    1     0
## t24    1    0    0    0    1    1    0    0    0     1
```

De forma que, ao escrevermos o modelo de regressão temos:

Para o grupo Placebo

$$\mathbb{E}(Y_2|\mathbf{X}_2) = \mathbf{X}_2\beta = \begin{pmatrix} \beta_1 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ \beta_1 + \beta_4 \\ \beta_1 + \beta_5 \end{pmatrix}$$

Para o grupo de alta dosagem

$$\mathbb{E}(Y_1|\mathbf{X}_1) = \mathbf{X}_1\beta = \begin{pmatrix} (\beta_1 + \beta_6) \\ (\beta_1 + \beta_6) + (\beta_2 + \beta_7) \\ (\beta_1 + \beta_6) + (\beta_2 + \beta_8) \\ (\beta_1 + \beta_6) + (\beta_2 + \beta_9) \\ (\beta_1 + \beta_6) + (\beta_2 + \beta_{10}) \end{pmatrix}$$

5.1.8. The null hypothesis that the patterns of change over time do not differ in the two treatment groups can be expressed as H0 : $L\beta = 0$. Describe an appropriate weight matrix L for this null hypothesis.

A matriz L é dada por

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

De forma que

$$L\beta = \begin{pmatrix} \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{pmatrix}$$

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.

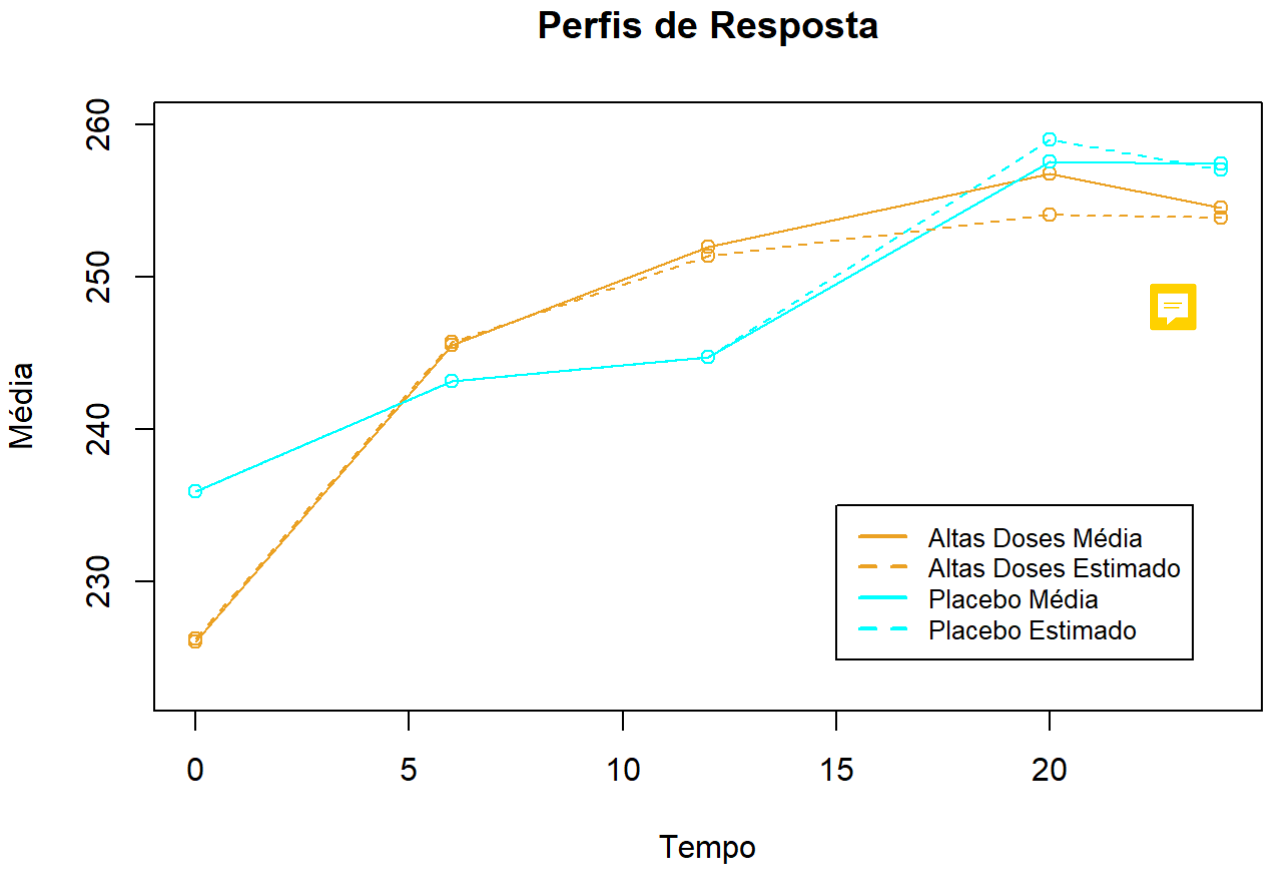
Podemos utilizar os coeficientes estimados para estimar a resposta média através de

$$\hat{\mu}_i = \mathbf{X}_i\hat{\beta}$$

Os obtemos através do atributo “coefficients” do nosso modelo. Foi necessária uma alteração na ordem que os coeficientes estavam sendo apresentados para que fosse coerente com a forma que apresentamos anteriormente. Podemos ver que as estimativas são bastante parecidas com as médias.

```
betas = mod$coefficients
betas2 = betas[c(1,3,4,5,6,2,7,8,9,10)]
```

```
plot(c(0, 6, 12, 20, 24),medias1,
     type = 'o', col = '#eba123',
     xlab = "Tempo", ylab = "Média",
     ylim = c(223, 260),
     main = "Perfis de Resposta")
lines(c(0, 6, 12, 20, 24), medias2, type = 'o', col = 'cyan')
lines(c(0, 6, 12, 20, 24), x2%*%betas2, type = 'o', col = 'cyan', lty = 2)
lines(c(0, 6, 12, 20, 24), x1%*%betas2, type = 'o', col = '#eba123', lty = 2)
legend(15, 235, legend = c("Altas Doses Média", "Altas Doses Estimado", "Placebo Média", "Pla
cebo Estimado"), col = c('#eba123', '#eba123', 'cyan', 'cyan'), lty = c(1,2,1,2), cex=0.8, lw
d = 2)
```



tempo	grupo	betas	medias	estimativas
t0	placebo	235.926829	235.9268	235.9268
t06	placebo	7.243902	243.1707	243.1707
t12	placebo	8.846205	244.7632	244.7730
t20	placebo	23.103334	257.6000	259.0302
t24	placebo	21.122304	257.4839	257.0491
t0	alta dosagem	-9.678291	226.0161	226.2485
t06	alta dosagem	12.217511	245.5323	245.7100
t12	alta dosagem	16.288928	252.0182	251.3837
t20	alta dosagem	4.756696	256.7955	254.1086
t24	alta dosagem	6.535983	254.5526	253.9068

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.

β_1: Baseline, ou seja, valor inicial de colesterol sérico de dois grupos.

$\beta_2, \beta_3, \beta_4, \beta_5$: Efeito do tempo no nível de colesterol medido.

β_6: Efeito do tratamento, seja placebo ou alta dosagem.

$\beta_7, \beta_8, \beta_9, \beta_{10}$: Interação entre grupo e tempo, ou seja, o quanto o tempo tem influencia no grupo em que o tratamento está sendo aplicado.