Dados Correlacionados - MAT02035 - Pontual 3

IME - Universidade Federal do Rio Grande do Sul
Prof. Rodrigo Citton Padilha dos Reis
Aluno: Enzo Bertoldi

05/04/2021

Applied Longitudinal Analysis - Capítulo 5 - p.140

5.1.1

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

```
df = read_dta(here::here('bancos','cholesterol.dta')) %>%
  mutate(group = ifelse(group == 1, 1, 0)) # colocando placebo como referência

knitr::kable(head(df),
             align = 'c') %>%
  kable_styling(bootstrap_options = c("striped",
                                      "bordered",
                                      'condensed'),
                position = 'center')
```

group	id	y1	y2	y3	y4	y5
1	1	178	246	295	228	274

group	id	y1	y2	y3	y4	y5
1	2	254	260	278	245	340
1	3	185	232	215	220	292
1	4	219	268	241	260	320
1	5	205	232	265	242	230
1	6	182	213	173	200	193

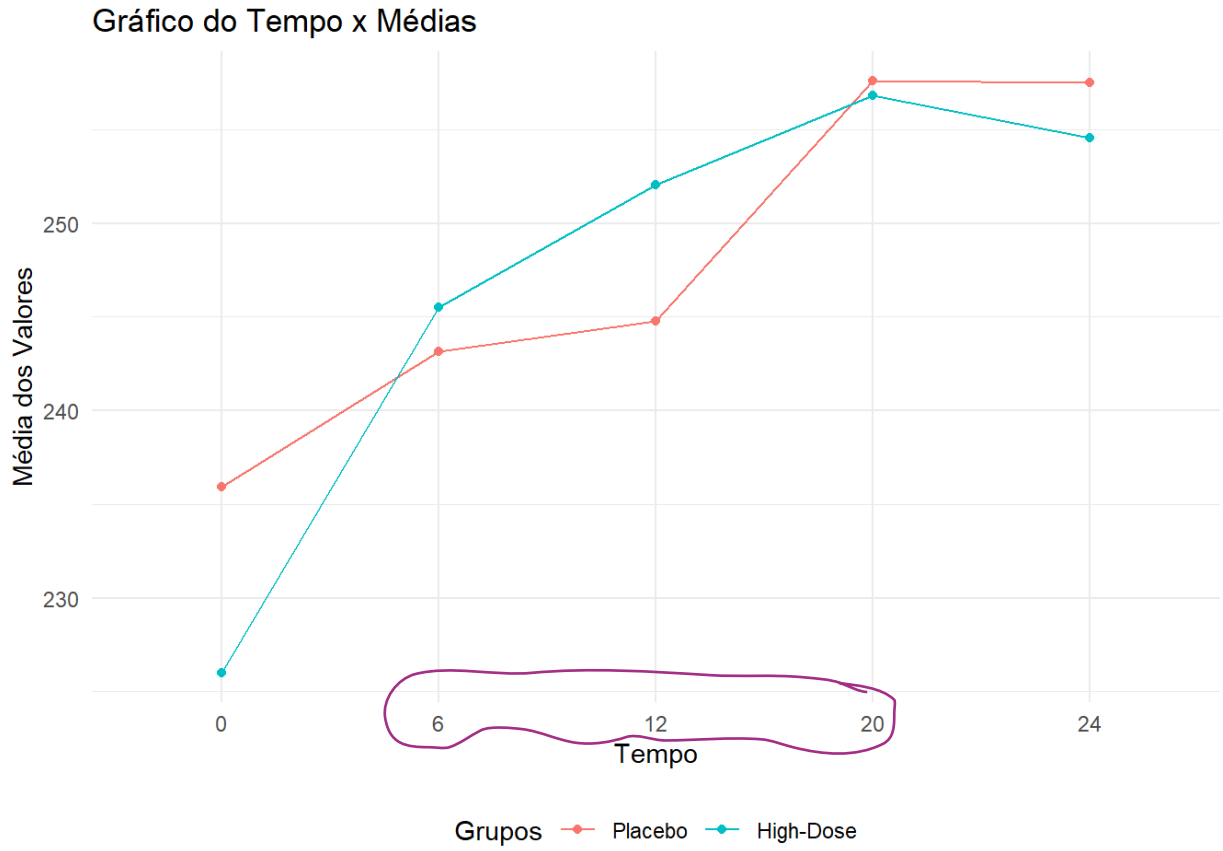
5.1.2

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

group	tempo	media	variancia	desvio
Placebo	0	235.9268	3121.970	55.87459
Placebo	6	243.1707	2424.545	49.23967
Placebo	12	244.7632	2126.186	46.11058
Placebo	20	257.6000	2615.482	51.14179
Placebo	24	257.4839	2439.191	49.38817
High-Dose	0	226.0161	1573.262	39.66437
High-Dose	6	245.5323	1556.483	39.45228
High-Dose	12	252.0182	1469.129	38.32922
High-Dose	20	256.7955	1189.515	34.48935
High-Dose	24	254.5526	2496.200	49.96198

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.

```
longo = gather(data = df,
               key = "tempo",
               value = "med",
               -id,
               -group)

longo$tempo = factor(longo$tempo,
                     labels = c('0', '6', '12', '20', '24'))

longo$group = factor(longo$group,
                     labels = c("Placebo",
                                "High-Dose"))
```

group	id	tempo	med
High-Dose	1	0	178
High-Dose	2	0	254
High-Dose	3	0	185
High-Dose	4	0	219
High-Dose	5	0	205
High-Dose	6	0	182

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.

	Df	Chisq	Pr(>Chisq)
group	1	0.0036801	0.9516270
tempo	4	65.4783135	0.0000000
group:tempo	4	7.9167409	0.0946761

A nível de significância de 5% percebemos que apenas o tempo é significativo na quantidade de colesterol, portanto grupo e a interação grupo x tempo não são significativos, possuindo p-valor equivalente a 0.95 e 0.095 respectivamente.

5.1.6

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

2188.452	1512.997	1407.913	1449.214	1320.705
1512.997	1762.425	1335.506	1409.257	1345.568
1407.913	1335.506	1691.530	1254.365	1343.293
1449.214	1409.257	1254.365	1762.425	1264.163
1320.705	1345.568	1343.293	1264.163	2322.094

```
## 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

Nosso modelo proposto terá forma:

$$Y = \beta_1 I(High - Dose) + \beta_2 I(Tempo_6) + \beta_3 I(Tempo_{12}) + \beta_4 I(Tempo_{20}) + \beta_5 I(Tempo_{24}) + \beta_6 I(Tempo_6) \times I(Group_{High-Dose}) + \beta_7 I(Tempo_{12}) \times I(Group_{High-Dose}) + \beta_8 I(Tempo_{20}) \times I(Group_{High-Dose}) + \beta_{10} I(Tempo_{24}) \times I(Grupo_{High-Dose})$$

Onde I é a função indicadora valendo 1 caso tal característica seja satisfeita e 0 caso contrário. Podemos escrever de forma matricial a fórmula acima, primeiro para o placebo:

intercept	high.dose	t1	t2	t3	t4	hd.x.t1	hd.x.t2	hd.x.t3	hd.x.t4

intercept	high.dose	t1	t2	t3	t4	hd.x.t1	hd.x.t2	hd.x.t3	hd.x.t4
1	0	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
1	0	0	0	0	1	0	0	0	0

Agora para a *high-dose*:

intercept	high.dose	t1	t2	t3	t4	hd.x.t1	hd.x.t2	hd.x.t3	hd.x.t4
1	1	0	0	0	0	0	0	0	0
1	1	1	0	0	0	1	0	0	0
1	1	0	1	0	0	0	1	0	0
1	1	0	0	1	0	0	0	1	0
1	1	0	0	0	1	0	0	0	1

5.1.8

Let L denote a matrix of known weights and /3 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 β = 0. Describe an appropriate weight matrix L for this null hypothesis.

Sabemos que $\beta_2, \ldots, \beta_5$ são os coeficientes para a mudança média no colesterol para o grupo que recebeu o placebo, além disso $\beta_6, \ldots, \beta_9$ é a mudança média no colestorel para o grupo com *high-dose*. Logo, queremos L de forma que:

$$H_0 = \beta_6 = \beta_7 = \beta_8 = \beta_9$$

Portando L será:

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

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

Para comparar essas estimativas devemos usar a matriz criada no exercício 5.1.7 e multiplicá-la pelos coeficientes estimados. Primeiramente o placebo:

estimados	media

estimados	media
235.9268	235.9268
243.1707	243.1707
244.7730	244.7632
259.0302	257.6000
257.0491	257.4839

Agora vamos repetir o procedimento para a *high-dose*:

estimados	media
226.2485	226.0161
245.7100	245.5323
251.3837	252.0182
254.1086	256.7955
253.9068	254.5526

Podemos que em ambos os casos a estimativa se aproximou (ou foi igual) a média amostral do estudo em questão.

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.

- **Intercepto**: valor esperado no tempo 0 para indivíduos utilizando placebo;
- β_1 **(grupo)**: para indivíduos que receberam *high-dose* é esperado que possuam em média -9.67 colesterol que o grupo no placebo;
- **tempo**: alteração média no valor do colesterol de acordo com o tempo para o grupo tomando placebo;
- **interação**: diferença média no valor do colesterol entre usuários do placebo e *high-dose*;
- **interação + tempo**: alteração média no valor do colesterol de acordo com o tempo para o grupo tomando *high-dose*;

Abaixo estão os coeficientes:

	x
(Intercept)	235.926829
groupHigh-Dose	-9.678291
tempo6	7.243902
tempo12	8.846205
tempo20	23.103334
tempo24	21.122304

	x
groupHigh-Dose:tempo6	12.217511
groupHigh-Dose:tempo12	16.288928
groupHigh-Dose:tempo20	4.756696
groupHigh-Dose:tempo24	6.535983