

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
MAT02035 - Modelos para dados correlacionados

Avaliação pontual 01

Exercício 1. (1.5 ponto) Apresente as principais características de um estudo longitudinal.

Em um estudo longitudinal as medidas dos mesmos indivíduos são tomadas repetidamente através do tempo, permitindo, assim, o estudo direto da mudança da variável resposta ao longo do tempo. Os dados longitudinais podem ser considerados um caso especial de dados agrupados. Estes fornecem observações dentro de um grupo: um indivíduo representa um agrupamento em um estudo longitudinal; uma família/domicílio representa um agrupamento. Tipicamente, dados agrupados irão exibir correlação positiva, e esta correlação deve ser levada em conta na análise.

Outra característica presente em estudos longitudinais é a perda de dados. Em estudos com muitos participantes e/ou com um longo período de acompanhamento é comum que algumas medidas não sejam realizadas. A perda de uma visita (ao centro de investigação), ou o acompanhamento (indivíduo não deseja mais participar do estudo) são alguns dos motivos para a perda de informação.

Exercício 2. (1.5 ponto) Para um estudo longitudinal, descreva as diferenças entre os delineamentos balanceado e desbalanceado.

Quando o número e o momento das medidas repetidas são os mesmos para todos os indivíduos do estudo, dizemos que o estudo segue um delineamento balanceado. Em estudos com delineamento balanceado, as ocasiões de medição podem ser igualmente ou desigualmente distribuídas ao longo da duração do estudo. Quando a sequência dos tempos de observação não é mais comum a todos os indivíduos no estudo, dizemos que o estudo possui um delinamento desbalanceado. Isto geralmente ocorre em alguns estudos longitudinais, especialmente aqueles em que as medidas repetidas se estendem por um período relativamente longo, e alguns indivíduos perdem a sua visita programada ou data de observação.

Exercício 3. (1.5 ponto) Descreva as diferenças (incluindo possíveis vantagens e desvantagens) entre estudos longitudinais e estudos transversais.

Em um estudo longitudinal, em que medidas repetidas realizadas nos indivíduos, é possível capturar a mudança dentro de indivíduo. Já em um estudo transversal, em que a resposta é medida em apenas uma ocasião, é possível apenas estimar diferenças nas respostas entre indivíduos. Ou seja, os estudos transversais permitem comparações entre subpopulações que diferem em idade, mas não fornecem qualquer informação a respeito de como os indivíduos

mudam durante o correspondente período.

Exercício 4. (1.5 ponto) Quais as consequências de se ignorar (na análise) a correlação presente entre os dados de medidas repetidas?

Ao ignorar a correlação entre as medidas repetidas irá, em geral, resultar em estimativas incorretas da variabilidade amostral, que levam a inferências bastante enganosas. Ou seja, a análise (incorreta) implicará em erros padrões muito grandes, intervalos de confiança muito largos, e valores p para testes de hipóteses para avaliar a mudança na resposta média muito grandes.

Exercício 5. (4 pontos) O estudo *Treatment of Lead-Exposed Children* (TLC) foi um estudo aleatorizado e controlado por placebo de *succimer* (um agente quelante) em crianças com níveis de chumbo no sangue de 20 a 44 microgramas/dL. Lembre-se de que os dados consistem em quatro medições repetidas dos níveis de chumbo no sangue obtidos na linha de base (ou semana 0), semana 1, semana 4 e semana 6 em 100 crianças que foram aleatoriamente designadas para tratamento de quelação com *succimer* ou placebo. Para este conjunto de problemas, focamos apenas nas 50 crianças designadas para tratamento de quelação com *succimer*.

Os dados brutos são armazenados em um arquivo externo: `lead.dta` Cada linha do conjunto de dados contém as 5 variáveis a seguir:

- ID, Y1, Y2, Y3, Y4.
- a. Leia os dados do arquivo externo e calcule as médias da amostra, os desvios padrão e as variações dos níveis de chumbo no sangue em cada ocasião.

Lendo, formatando e filtrando os dados.

```
# Carrega pacotes
library(haven)
library(tidyr)
library(dplyr)

# Carrega dados
chumbo <- read_dta(file = here::here("data", "tlc.dta"))

# Largo para longo
chumbo.longo <- gather(data = chumbo,
                      key = "tempo",
                      value = "chumbo", -id, -trt)

# Transforma dados
chumbo.longo$tempo <- as.numeric(
  as.character(factor(chumbo.longo$tempo,
                    labels = c(0, 1, 4, 6))))

chumbo.longo$trt <- factor(chumbo.longo$trt,
```

```
labels = c("Placebo", "Succimer"))

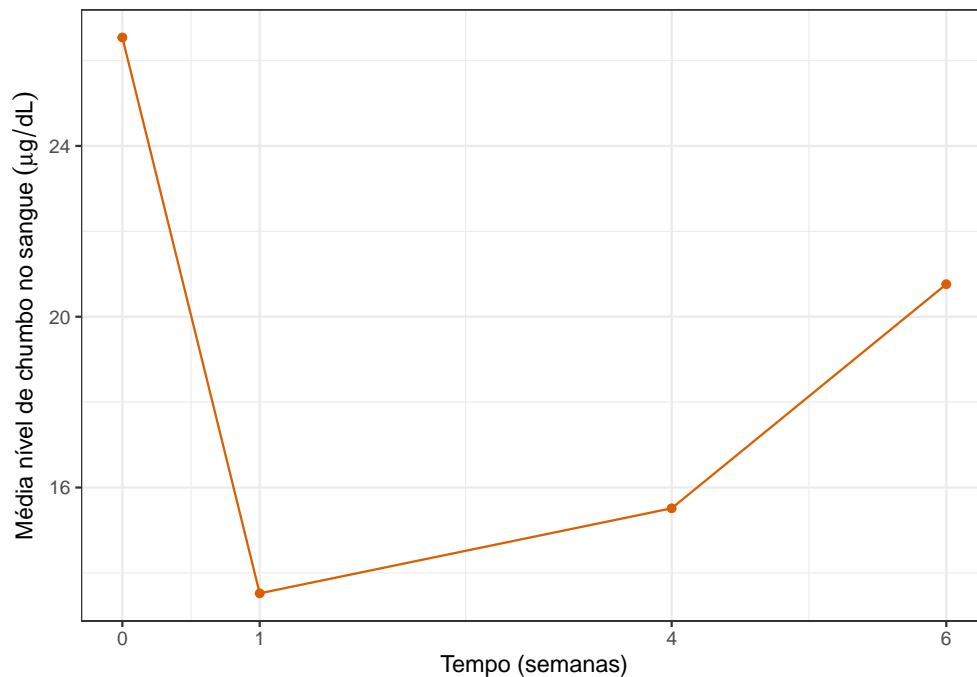
# Filtra dados do grupo succimer
chumbo.longo <- chumbo.longo %>%
  filter(trt == "Succimer")
```

A seguir são apresentados os desvios-padrões e variâncias dos níveis de chumbo no sangue em cada ocasião para os indivíduos do grupo *Succimer*.

Tempo	Média	Desvio Padrão	Variância
0	26.5	5.0	25.2
1	13.5	7.7	58.9
4	15.5	7.9	61.7
6	20.8	9.2	85.5

- b. Construa um gráfico de tempo dos níveis médios de chumbo no sangue versus tempo (em semanas). Descreva as características gerais da tendência temporal.

O gráfico de perfis médios é apresentado a seguir.



É possível perceber um decaimento acentuado no nível médio de chumbo no sangue logo após o começo do estudo. No entanto, após a semana 1, os níveis médios de chumbo no sangue volta a subir, embora não atinjam o mesmo nível apresentado na linha de base.

- c. Calcule as matrizes de covariância e correlação 4×4 para as quatro medidas repetidas dos níveis de chumbo no sangue.

A matriz de covariância das medidas repetidas dos níveis de chumbo no sangue é apresentada a seguir.

	y0	y1	y4	y6
y0	25.21	15.47	15.14	22.99
y1	15.47	58.87	44.03	35.97
y4	15.14	44.03	61.66	33.02
y6	22.99	35.97	33.02	85.49

E a seguir, é apresentada a matriz de correlação das medidas repetidas dos níveis de chumbo no sangue é apresentada a seguir.

	y0	y1	y4	y6
y0	1.00	0.40	0.38	0.50
y1	0.40	1.00	0.73	0.51
y4	0.38	0.73	1.00	0.45
y6	0.50	0.51	0.45	1.00

- d. Verifique se os elementos diagonais da matriz de covariâncias são as variâncias comparando com a estatística descritiva obtida no item (a).

Como sabemos, a $\text{Cov}(Y_j, Y_j) = \text{Var}(Y_j)$. Assim, como era de se esperar, verifica-se que os elementos diagonais da matriz de covariâncias (25.2, 58.9, 61.7, 85.5) são as variâncias das medidas em cada ocasião de tempo apresentadas no item (a).