Lista 01 - Planejamento e Análise de Experimentos (MAE0316)

Caio M. de Almeida - 15444560   Eduardo Yukio G. Ishihara - 15449012
Gustavo S. Garone - 15458155   Ian B. Loures - 15459667
João Victor G. de Sousa - 15463912

19 de setembro de 2025

Nesta lista, usaremos "." como separador decimal.

Exercício 1

Item a

Analisamos um estudo de coorte sobre os efeitos da "COVID longa" ou Síndrome pós-COVID-19 no trabalho de ROCHA et al. (2024). O estudo utilizou o método de coorte ambidirecional com indivíduos de três hospitais de Cuiabá. Observações foram colhidas do prontuário desses pacientes e posteriormente 6 e 12 meses após alta hospitalar por telefone. Foram perguntados fatores socioeconômicos dos indivíduos, além de sintomas comuns da "COVID longa", como fadiga e problemas de memória.

Na análise de dados, foram consideradas pelos pesquisadores comorbidades como hipertensão, diabetes, obesidade e doenças cardíacas. Para os sintomas, classificaram como musculares, neuropsiquiátricos, dermatológicos, cardiovasculares e pulmonares.

Acreditamos que o estudo tenha uma base metodológica, no geral, sólida, mas que a dependência na descrição por telefone dos entrevistados pode ter comprometido a integridade das conclusões. Isso foi parcialmente reconhecido pelos autores, que perceberam que a obesidade era subrepresentada quando comparada com o IMC calculado a partir da altura e peso dos entrevistados. Isto é, dos entrevistados, quando perguntavam se estavam com sobrepeso, responderam positivamente apenas 11 dos 46 identificados com obesidade pelo IMC. Os autores consideraram isso na análise dos dados, mas não há discussão sobre outras imprecisões.

Item b

Pode ser que haja diferença entre as conclusões dos estudos. No primeiro estudo, fatores como viés de seleção (por exemplo, selecionar pacientes que optaram pelo estudo, pois utilizam a medicina convencional, e paciente que optaram por não usar o remédio, pois utilizam exclusivamente a medicina alternativa) e menor aleatorização. No segundo estudo, ao selecionar previamente o grupo, e então efetuar um ensaio randomizado, espera-se que o processo de amostragem aleatória faça com que os dois grupos, na média, tenham indivíduos semelhantes. Logo, a amostragem aleatória forneceria resultados mais robustos e conclusões mais acuradas ao atenuar fatores intrínsecos das unidades amostrais dentre os grupos, como a gravidade da doença, estilos de vida ou fatores biológicos.

Item c

Um resultado possível, que não considera vieses dos investigadores ou resultados, é que, no primeiro estudo, o remédio já estava sendo aplicado em pacientes nos estágios avançados da doença como medida emergencial, enquanto no segundo estudo o remédio foi aplicado tanto em pacientes com doença avançada quanto leve ou moderada. Logo, uma taxa de cura menor no primeiro estudo poderia ser explicada pela ineficiência do medicamento de conter a doença no estágio avançado, enquanto funciona bem no geral, como aponta o segundo estudo.

Exercício 2

É apresentado um estudo prospectivo, aleatorizado, experimental (analítico) e controlado por grupo controle, sobre a aplicação de determinado tratamento - oxigenação específica - em ratos com diabete (população objetivo). Como hipótese, buscaram descobrir se existe efeito do oxigênio hiperbárico na cicatrização de feridas cirúrgicas nesta população. Como fator extrínseco explicativo e premeditado consta a exposição ou não - caracterizando dois níveis para esse fator - das unidades experimentais (ratos com diabete induzida) ao tratamento com oxigênio hiperbárico, uma exposição com intervenção. Uma possível fonte de variação intrínseca (externa) é a própria variação biológica dos ratos, que pode acelerar ou retardar a cicatrização. A repetição foi realizada com diversos ratos por grupo (que receberam ou não o tratamento), cujo número pode ser aumentado a fim de gerar mais repetições. Variações externas como níveis glicêmicos inadequados no sangue foram tratadas ao desconsiderar 6 desses ratos. Variações acidentais tentaram ser minimzadas por preocaucões dos pesquisadores. Por se tratar de um estudo experimental, não podemos descrever unidades observacionais, nem se encaixa como longitudinal ou transversal. A unidade observacional são as amostras de tecido coletadas. Fontes de mascaramento, como o viés dos pesquisadores, foram relatadas como minimizadas.

Exercício 3

Item a

O contraste $p = 1$ pode ser usado para testar se o efeito da dose zero difere do efeito da dose um, enquanto o contraste $p = 2$ pode ser usado para comparar o efeito da dose dois com a média das

Tabela 1: Doses

Dose (mg/kg/dia)	Coeficiente (q) SSQ	GL		MQE
Controle	9.4864	11	0	0.8624
10	5.3824	11	0	0.4893
20	7.6729	11	3	0.6975
30	5.3824	11	-1	0.4893
30	3.4969	11	-1	0.3179

doses zero e dose um. Alternativamente, o contrate $p = 2$ é usado para comparar o dobro do efeito da dose dois com a soma dos efeitos da dose zero e da dose um. Ambos os contrastes seriam nesse exemplo usados para encontrar a dose limiar de efeito.

Item b

Assumindo tamanho amostral igual para as doses, os contrastes são ortogonais:

$$(-1) \cdot (-1) + 1 \cdot (-1) + 0 \cdot 2 = 0$$
$$(-1) \cdot (-1) + 1 \cdot (-1) + 0 \cdot (-1) + 0 \cdot 3 = 0$$
$$(-1) \cdot (-1) + (-1) \cdot (-1) + 2 \cdot (-1) + 0 \cdot 3 = 0$$

Item c

A hipótese $\mathcal{H}_2$ é a mais apropriada para testar a dose limiar uma vez que busca uma dose j tal que essa possui uma média significativamente maior que as anteriores $(1, 2, ..., j-1)$, que são iguais, assumindo doses crescentes. Caso não exista, caímos na hipótese de não efeito $\mathcal{H}_0$. Não sugeriríamos outra hipótese para este problema.

Item d

Construímos a tabela a partir dos dados do enunciado. Usamos os desvios padrões ao quadrado para determinar o MQE dentro de cada dose.

Note, portanto, que $SSE = 104.3504 + 59.2064 + 84.4019 + 59.2064 + 38.4659 = 345.631$ e $GL_{Total} = 5 \times 11 = 55$. Portanto, $MQE = \frac{SSE}{GL_{Toal}} = 6.2842$, que também é o estimador não viesado para σ^2.

Construímos a tabela de contrastes de Helmert:

Para uma hipótese $\mathcal{H} : L_p$ temos o estimador:

$$\hat{L}_j = \sum_i^a c_{j,i} \bar{Y}_i$$

Tabela 2: Contrastes usados

	Coeficiente (q)				
Contraste(p)	0	1	2	3	4
	-1	1	0	0	0
	-1	-1	2	0	0
	-1	-1	-1	3	0
	-1	-1	-1	-1	4

em que c_i são os elementos do contraste e

$$\text{Var}(\hat{L}_j) = \text{MSE}\sum_{i=1}^{a}\frac{c_{j,i}^2}{n_i}$$

Disso, construímos a estatística do teste

$$T_p = \frac{\hat{L}_j - 0}{\sqrt{\hat{\text{Var}}(\hat{L}_j)}} \overset{\mathcal{H}_0}{\sim} t_{12}$$

Estamos testando, em todas as hipóteses, $L = 0$. Logo, para cada contraste, temos as seguintes estatísticas $\hat{L}_j$:

$$\begin{cases} \hat{L}_1 = -0.060 \\ \hat{L}_2 = 0.740 \\ \hat{L}_3 = 4.130 \\ \hat{L}_4 = 10.930 \end{cases}$$

Com as seguintes variâncias:

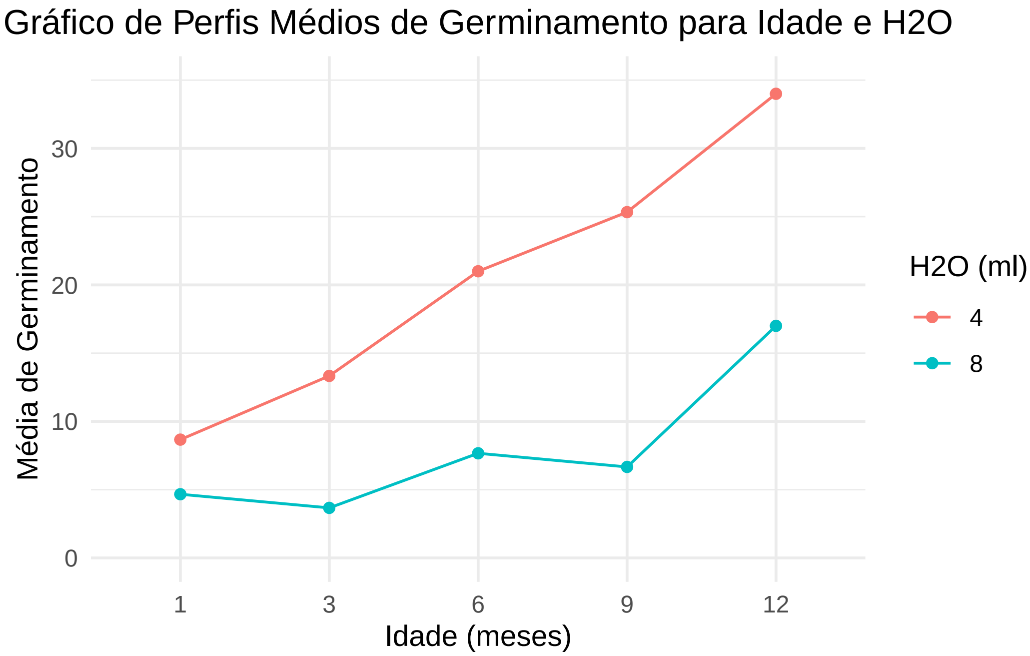
$$\begin{cases} \hat{\text{Var}}(\hat{L}_1) = 1.0474 \\ \hat{\text{Var}}(\hat{L}_2) = 3.142 \\ \hat{\text{Var}}(\hat{L}_3) = 6.284 \\ \hat{\text{Var}}(\hat{L}_4) = 10.474 \end{cases}$$

Então, são as estatísticas dos testes T_p:

$$\begin{cases} T_1 = -0.059 \\ T_2 = 0.417 \\ T_3 = 1.647 \\ T_4 = 3.377 \end{cases}$$

Temos que, para o grau de significância de 5%, o quantil de corte é 1.673. Portanto, a 5% de significância, há evidências para rejeitar a hipótese nula apenas para o contraste 4.

Exercício 4




```
             Df Sum Sq Mean Sq F value   Pr(>F)    
H2O           1 1178.1  1178.1  19.723 0.000251 ***
Idade         4 1321.1   330.3   5.529 0.003645 ** 
H2O:Idade     4  208.9    52.2   0.874 0.496726    
Residuals    20 1194.7    59.7                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Item a

Vemos que pode-se supor que não há interação entre idade e H2O, que pode ser inicialmente verificado pelo paralelismo do gráfico de perfis médios e confirmado pela não significância do efeito no modelo.

Item b

O modelo de ANOVA a uma via parece ser o mais indicado para esse caso, pois há apenas um fator (Idade), e conseguiríamos obter os efeitos de cada tratamento pelos τ_i.

Item c

Iremos considerar que o índice i modela as idades (1-(1), 2-(3), 3-(6), 4-(9), 5-(12)) e o índice j modela a quantidade de H2O (1-(4), 2-(8))

Modelo de Médias

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

Em que μ_{ij} será a média do tratamento: i-ésimo nível do fator 1 e j-ésimo nível do 2.

Modelo a Duas Vias com Interação

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

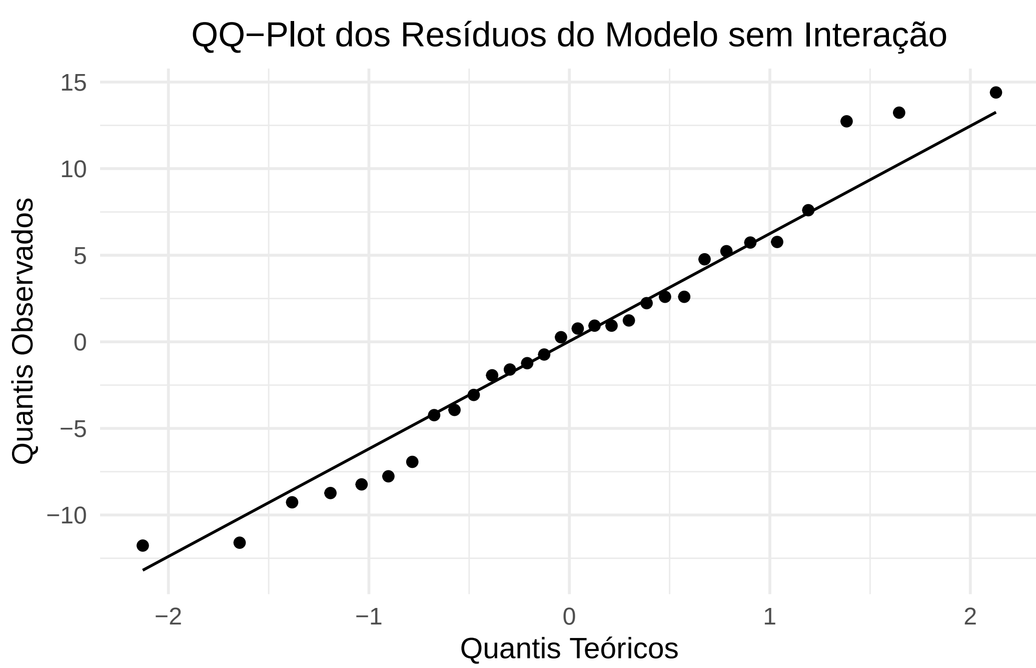
Em que μ é uma constante que representa a média geral, τ_i é o efeito do i-ésimo nível do fator 1, β_i é o efeito do j-ésimo nível do fator 2 e $(\tau\beta)_{ij}$ é o efeito da interação entre τ_i e β_j.

Item d

Excluiremos as interações já que não são significantes. Verificação das suposições do modelo:

```
Analysis of Variance Table

Response: Germinou
          Df Sum Sq Mean Sq F value    Pr(>F)
Idade      4 1321.1  330.28  5.6477 0.0023801 **
H2O        1 1178.1 1178.13 20.1457 0.0001525 ***
Residuals 24 1403.5   58.48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

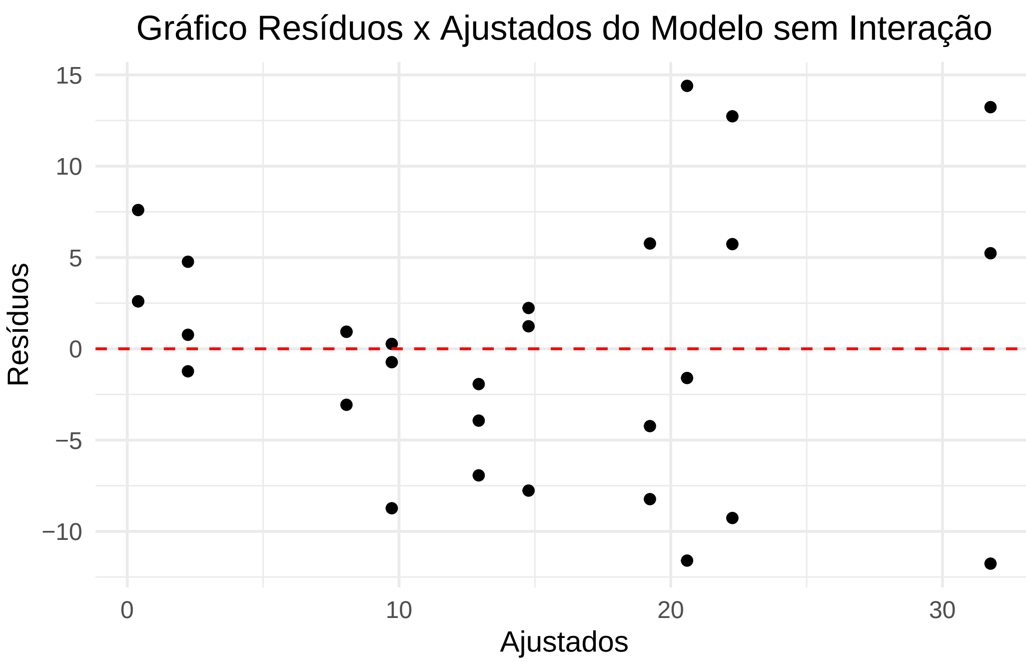


```
    Shapiro-Wilk normality test

data:  residuals(modelo1)
W = 0.96806, p-value = 0.4875


    Bartlett test of homogeneity of variances

data:  Germinou by interaction(Idade, H2O)
Bartlett's K-squared = 12.49, df = 9, p-value = 0.1871
```



Podemos ver pelos testes que os erros são normalmente distribuídos e são homocedásticos.

Usando o modelo a Duas Vias sem Interação (não significante):

```
Call:
lm(formula = Germinou ~ Idade + H2O, data = data, contrasts = list(Idade = contr.sum,
    H2O = contr.sum))

Residuals:
     Min       1Q   Median       3Q      Max
-11.7667  -4.1583   0.5167   4.2250  14.4000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.2000     1.3962  10.171 3.53e-10 ***
```

```
Idade1       -7.5333     2.7924  -2.698 0.012570 *
Idade2       -5.7000     2.7924  -2.041 0.052367 .
Idade3        0.1333     2.7924   0.048 0.962311
Idade4        1.8000     2.7924   0.645 0.525294
H2O1          6.2667     1.3962   4.488 0.000153 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.647 on 24 degrees of freedom
Multiple R-squared:  0.6404,    Adjusted R-squared:  0.5655
F-statistic: 8.547 on 5 and 24 DF,  p-value: 9.27e-05


Analysis of Variance Table

Response: Germinou
          Df Sum Sq Mean Sq F value    Pr(>F)
Idade      4 1321.1  330.28  5.6477 0.0023801 **
H2O        1 1178.1 1178.13 20.1457 0.0001525 ***
Residuals 24 1403.5   58.48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(e)

```
(Intercept)      Idade1      Idade2      Idade3      Idade4        H2O1
 14.2000000  -7.5333333  -5.7000000   0.1333333   1.8000000   6.2666667
```

- $\hat{\gamma}_0$ é a média geral do modelo ($\hat{\mu} = 14.2$)
- $\hat{\gamma}_1$ é o aumento médio do germinamento dado 1 mês de idade ($\hat{\beta}_1 = -7.53333$)
- $\hat{\gamma}_2$ é o aumento médio do germinamento dado 3 meses de idade ($\hat{\beta}_2 = -5.7$)
- $\hat{\gamma}_3$ é o aumento médio do germinamento dado 6 meses de idade ($\hat{\beta}_3 = 0.13333$)
- $\hat{\gamma}_4$ é o aumento médio do germinamento dado 9 meses de idade ($\hat{\beta}_4 = 1.8$)
- $\hat{\beta}_5 = -(-7.53333 - 5.7 + 0.13333 + 1.8) = 11.3$ é o aumento médio do germinamento dado 12 meses de idade
- $\hat{\gamma}_5$ é o aumento médio do germinamento dado 4ml de água ($\hat{\tau}_1 = 6.26666$)
- $\hat{\tau}_2 = -6.26666$ é o aumento médio do germinamento dado 8ml de água

Referências

ROCHA, R. P. S. et al. Síndrome pós-COVID-19 entre hospitalizados por COVID-19: estudo de coorte após 6 e 12 meses da alta hospitalar. **Cadernos de Saúde Pública**, v. 40, p. e00027423, 2024.