

Laboratório 1 - 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

Exploração e análise descritiva

Item a

Categorizariam os fatores de longevidade (Longevity), tempo em sono (Sleep) e medida do tamanho corporal (Thorax) como quantitativas contínuas, enquanto o tipo (Type), tratamento (Treatment), número de parceiros (Partners), como categóricas.

Além das variáveis citadas, também identificamos as variáveis “Partners” (número de parceiras), “Type” (tipo das parceiras), “Thorax” (comprimento do tórax em mm), “Sleep” (percentual do dia desprendido dormindo). Classificamos “Partners” como quantitativa discreta, “Type” e “Treatment” como qualitativa nominal, “Longevity”, “Thorax” e “Sleep” como quantitativas contínuas.

Item b

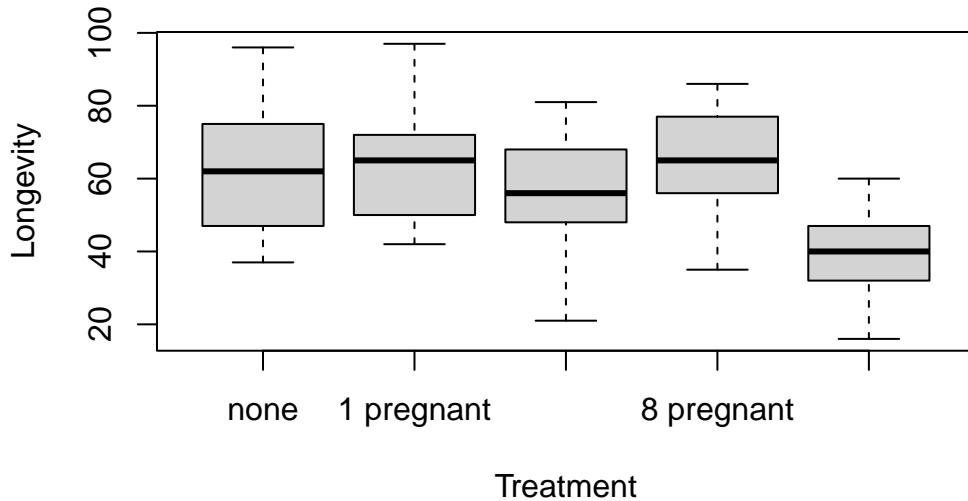
Esses fatores não se encaixam integralmente como aninhados ou cruzados, mas podemos dizer que apresentam tendência cruzada.

Item c

É importante examinar se há dados faltantes e presença de outliers.

[1] 0

Pelo resultado do R, não foram encontrados dados faltantes.

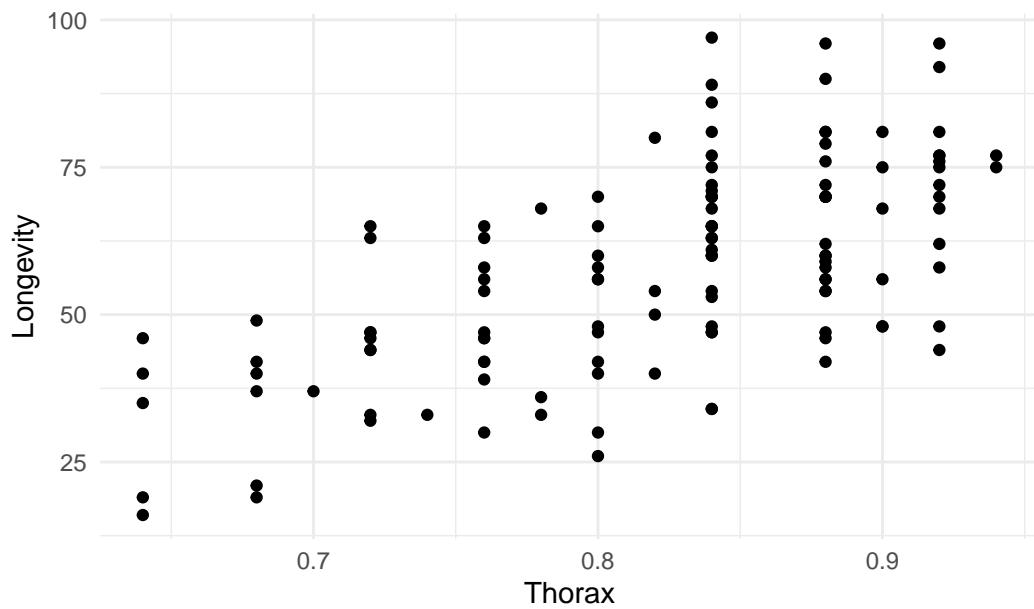


O boxplot aponta ausência de outliers. Também verificamos que todos os dados de cada coluna da tabela são do tipo esperado, isso é, as categorias são as apresentadas no contexto e os valores respeitam os tipos de variáveis explicitados em a).

Item d

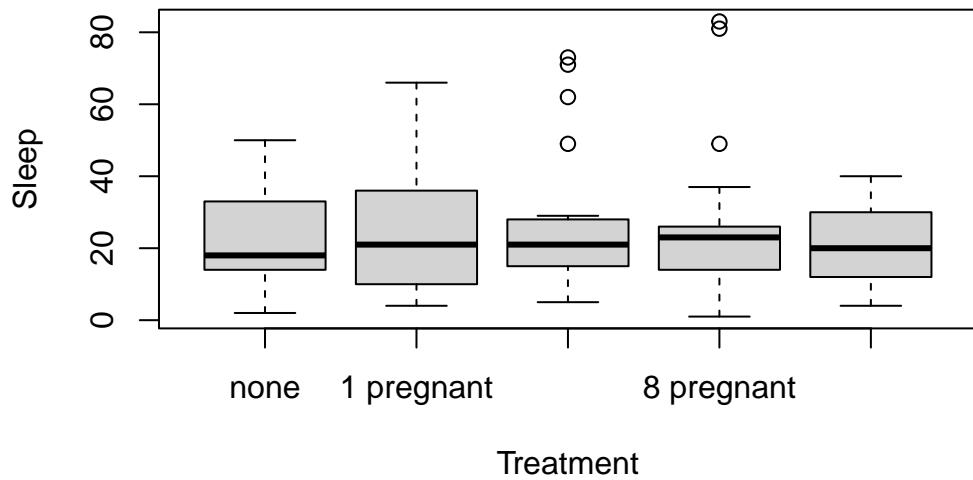
Além do boxplot criado para verificar o tratamento, podemos criar gráficos de dispersão para comparar a longevidade e o tórax (intrínseco):

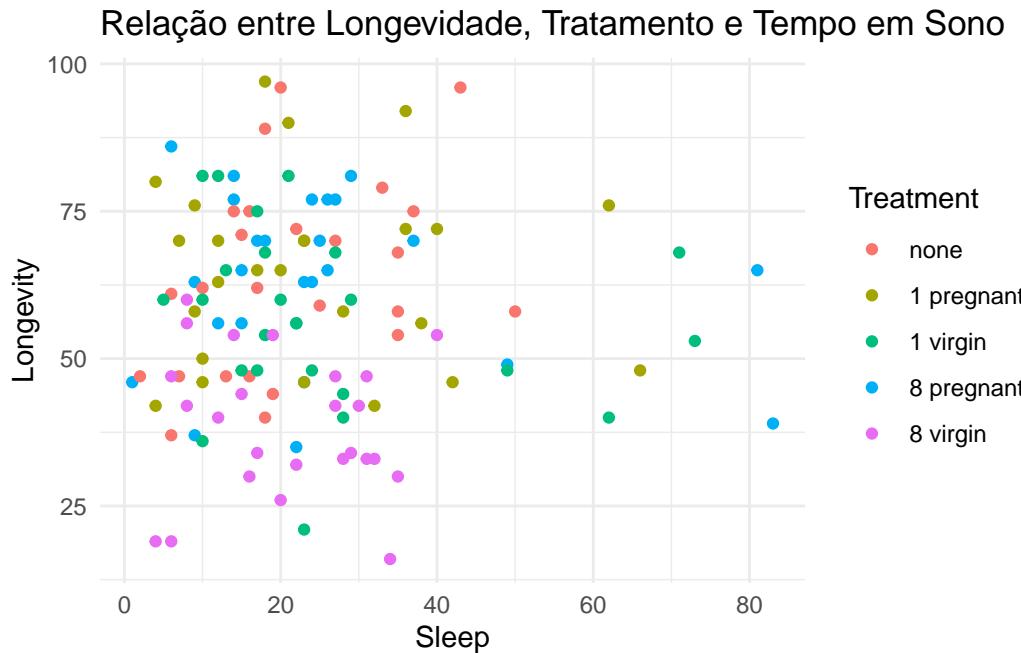
Longevidade por tamanho do tórax



E outras variáveis entre si:

% do dia despendido dormindo por tratamento





Exercício 3

Item a

Considerando os modelos com apenas um fator de tratamento, mostre a soma de quadrados total, soma de quadrados do tratamento, soma de quadrados do erro, graus de liberdade, e os quadrados médios, e monte uma tabela anova. Considerando um nível de significância de $\alpha = 0.05$, quais são as conclusões?

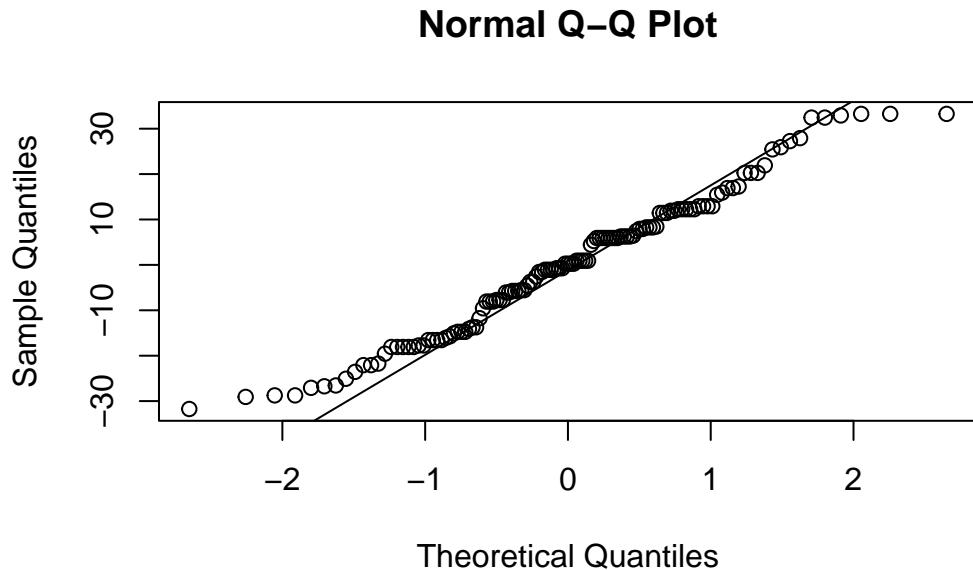
	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
Type	2	7845	3923	15.74	8.3e-07 ***						
Residuals	122	30407	249								
<hr/>											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

Table 1: Tabela de ANOVA

Causa de Variação	Soma de Quadrados (SQ)	GL	QM	F	p-valor
Tratamento	7845	2	3923	15.74	8.3e-07
Resíduo	30407	122	249		
Total	38252	124			

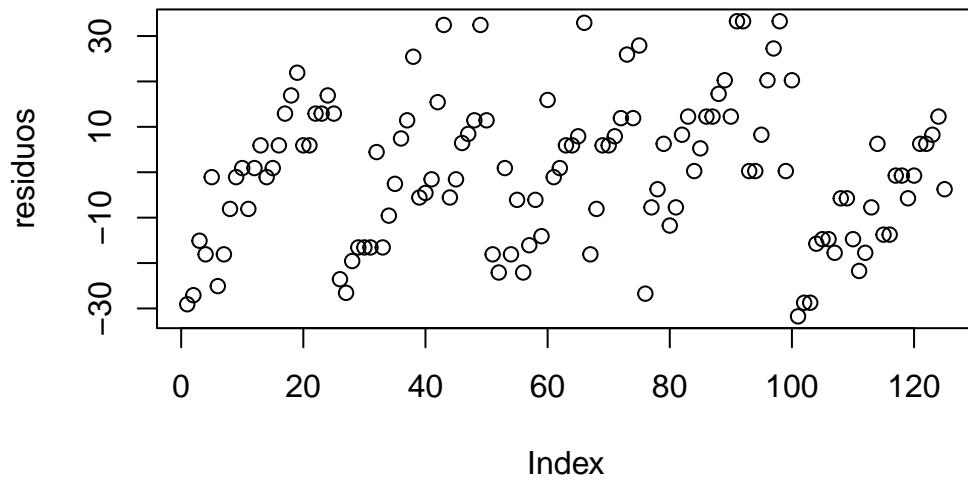
A 5% de significância estatística, conclui-se que há evidências para rejeitar a hipótese nula que os tratamentos não têm efeito sobre a observação. Ou seja, deixar os machos com fêmeas prenhas, virgens ou sem nenhuma fêmea afeta a longevidade deles. Contudo, não é possível definir qual dos tratamentos é significativamente efetivo sobre a longevidade.

Item b

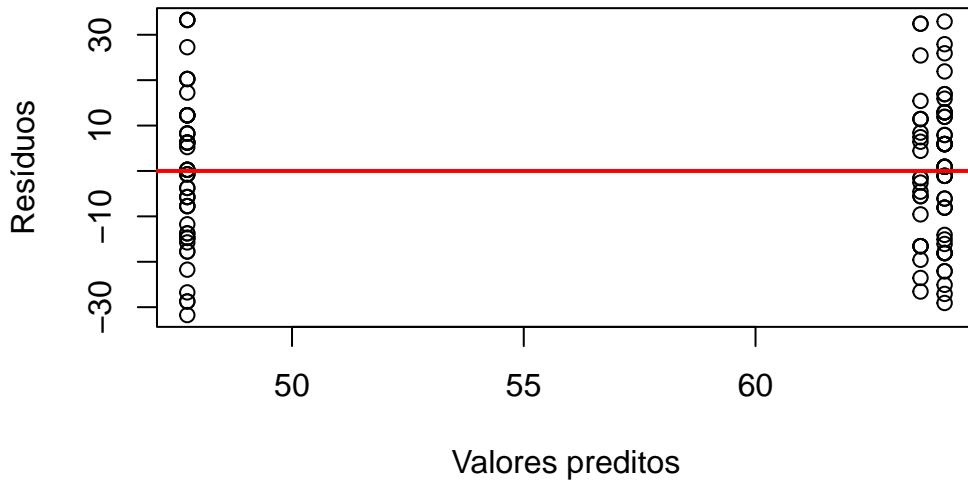


Shapiro-Wilk normality test

```
data: residuos
W = 0.97965, p-value = 0.05634
```



Resíduos vs Valores preditos



```
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group  2  0.1865 0.8301
122
```

Bartlett test of homogeneity of variances

```
data: Longevity by Type  
Bartlett's K-squared = 0.42831, df = 2, p-value = 0.8072
```

Pela análise do qq-plot, os resíduos parecem não violar a hipótese de normalidade. Pelo teste de Shapiro-Wilk, a 5% de significância estatística, não rejeitamos H_0 de que os resíduos vêm de uma distribuição normal. Portanto, podemos seguir para os próximos passos do diagnóstico do modelo.

Do plot da evolução dos resíduos com a evolução do tempo, não é possível observar nenhuma tendência de dependência entre as observações, então possivelmente não há autocorrelação.

No plot de resíduos x valores preditos, nota-se que os resíduos apresentam apresentar uma mesma variância dentro de cada tratamento. Formalmente, realizamos os testes de Levene e de Bartlett, os quais não rejeitaram a hipótese nula de homocedasticidade a 5% de significância estatística. Note que, uma vez que não rejeitamos a hipótese de normalidade, o teste de Bertlett é mais indicado para esse conjunto de dados.

Item c

(Intercept)	Typeprenha	Typevirgem
63.56	0.52	-15.82

No contexto do experimento, a interpretação dos coeficientes estimados indica que a longevidade média dos machos mantidos sem fêmeas é de aproximadamente 63.6 dias. Para os machos mantidos com fêmeas prenhas, a longevidade é estatisticamente equivalente, apresentando um acréscimo médio muito pequeno (cerca de 0.5 dias), o que sugere que a presença de fêmeas não receptivas não altera substancialmente o tempo de vida dos machos. Já para os machos mantidos com fêmeas virgens receptivas, a longevidade apresenta uma redução expressiva, em torno de 15.8 dias a menos em média, revelando que a atividade sexual efetiva tem um efeito negativo marcante sobre a longevidade. Assim, a conclusão é que não é apenas a presença de fêmeas que importa, mas sim a demanda reprodutiva ativa, a qual implica custos fisiológicos que reduzem a sobrevivência dos machos.

Exercício 5

Item a

As comparações mais interessantes é entre exposição à virgens ou prenhas e a uma ou oito parceiras. Dessa forma, ordenamos as comparações mais em ordem decrescente de interesse:

1. Prenha1 x Virgem1
2. Virgem1 x Virgem8
3. Prenha1 x Prenha8
4. Prenha8 x Virgem8
5. Prenha8 x Virgem1
6. Prenha1 x Virgem8

Deixamos 5. e 6. nestas posições uma vez que ocorre a mudança de mais de um fator de um grupo para o outro, comprometendo a análise.

Item b

Ao preplanejar as comparações por contrastes, evitamos alguns erros no planejamento e análise do experimento. Por exemplo, ao selecionar as comparações post-hoc, o analista pode priorizar maiores diferenças, ampliando o erro Tipo I, uma vez que, das comparações selecionadas, muitas das diferenças podem ter sido causadas por erro.

Item c

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_8 \\ v_1 \\ v_8 \end{bmatrix}$$

em que p_1 : Prenha1, p_8 : Prenha8, v_1 : Virgem1 e v_8 : Virgem8

Analisando a dimensão da matriz de contrastes, tem-se que os contrastes não são ortogonais. Isso seria vantajoso uma vez que os testes são independentes com contrastes ortogonais.

Item d

Call:

```
lm(formula = Longevity ~ Partners + Type, data = dat)
```

Residuals:

Min	1Q	Median	3Q	Max
-31.61	-10.95	1.05	10.79	32.44

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	63.5600	3.0443	20.878	< 2e-16 ***
Partners	-1.3914	0.4349	-3.199	0.00176 **
Typeprenha	6.7814	4.2110	1.610	0.10991

```

Typevirgem -9.5586     4.2110 -2.270  0.02498 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.22 on 121 degrees of freedom
Multiple R-squared: 0.2671,   Adjusted R-squared: 0.2489
F-statistic: 14.7 on 3 and 121 DF, p-value: 3.193e-08

```

Linear hypothesis test:

```

Typeprenha = 0
Partners = 0
Typevirgem = 0

```

Model 1: restricted model

Model 2: Longevity ~ Partners + Type

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	124	38253				
2	121	28036	3	10217	14.699	3.193e-08 ***

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Linear hypothesis test:

```
Typeprenha - Typevirgem = 0
```

Model 1: restricted model

Model 2: Longevity ~ Partners + Type

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	122	34711				
2	121	28036	1	6674.9	28.808	3.913e-07 ***

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Oi gente tem que arrumar aqui em cima mas acho que to chegando em algo \wedge não sei como faz o partners 1 e 8, só o partners geral????