

# **Universidade de Brasília**

**Departamento de Estatística**



## **Trabalho de amostragem - Grupo 3**

### **Autores:**

Bruno Gondim Toledo 15/0167636

Giulia

Dail

Guilherme

Brasília

10 de novembro de 2023

# Conteúdo

<b>1</b>	<b>Resumo</b>	<b>2</b>
<b>2</b>	<b>Metodologia</b>	<b>2</b>
<b>3</b>	<b>Referencial teórico</b>	<b>4</b>
3.1	Amostragem Aleatória Simples . . . . .	4
3.1.1	Estimativa de parâmetros . . . . .	6
<b>4</b>	<b>Análises</b>	<b>7</b>
4.1	Análise exploratória . . . . .	7
4.1.1	Avaria . . . . .	7
4.1.2	Avaria por prateleira . . . . .	8
4.1.3	Tipos de avaria . . . . .	14
4.1.4	Tipo de avaria por prateleira . . . . .	15
4.2	Amostragem . . . . .	16
<b>5</b>	<b>Conclusão</b>	<b>16</b>
<b>6</b>	<b>Códigos Computacionais</b>	<b>17</b>

# 1 Resumo

Este projeto é um desdobramento do trabalho de amostragem “Estudo sobre a qualidade física dos livros da BCE” realizado no primeiro semestre de 2023, na disciplina Técnicas de Amostragem, pelos alunos Lucas Coelho Christo Fernandes, Luiz Gustavo Jordão Graciano e Raissa Alvim Teixeira.

## 2 Metodologia

Seguindo as orientações dos autores citados no Resumo, a avaliação do estado dos livros será feita com base em 3 critérios: o estado da capa; a oxidação das páginas e/ou costura do livro aparente; e o uso de marca textos ou canetas ou lápis. Assim, o livro será classificado com “avarias” (ou codificado como 1) se apresentar qualquer um dos critérios acima, e será classificado como “sem avarias” (ou codificado como 0), caso contrário. Como referência para essa avaliação, verificar as Figuras abaixo:

Figura 1: Capa com avaria

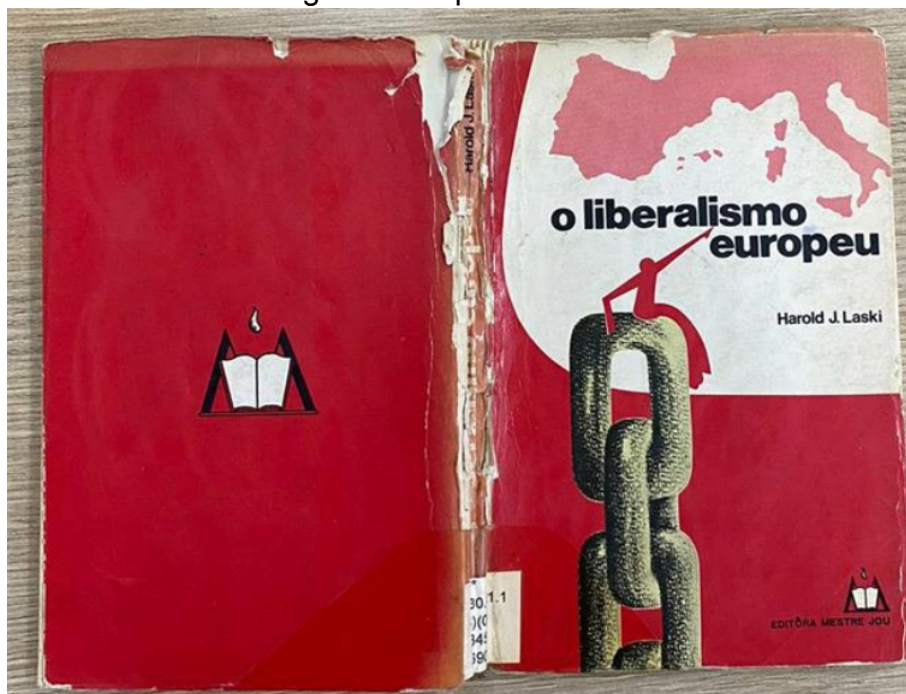


Figura 2: Oxidação/Costura com avaria

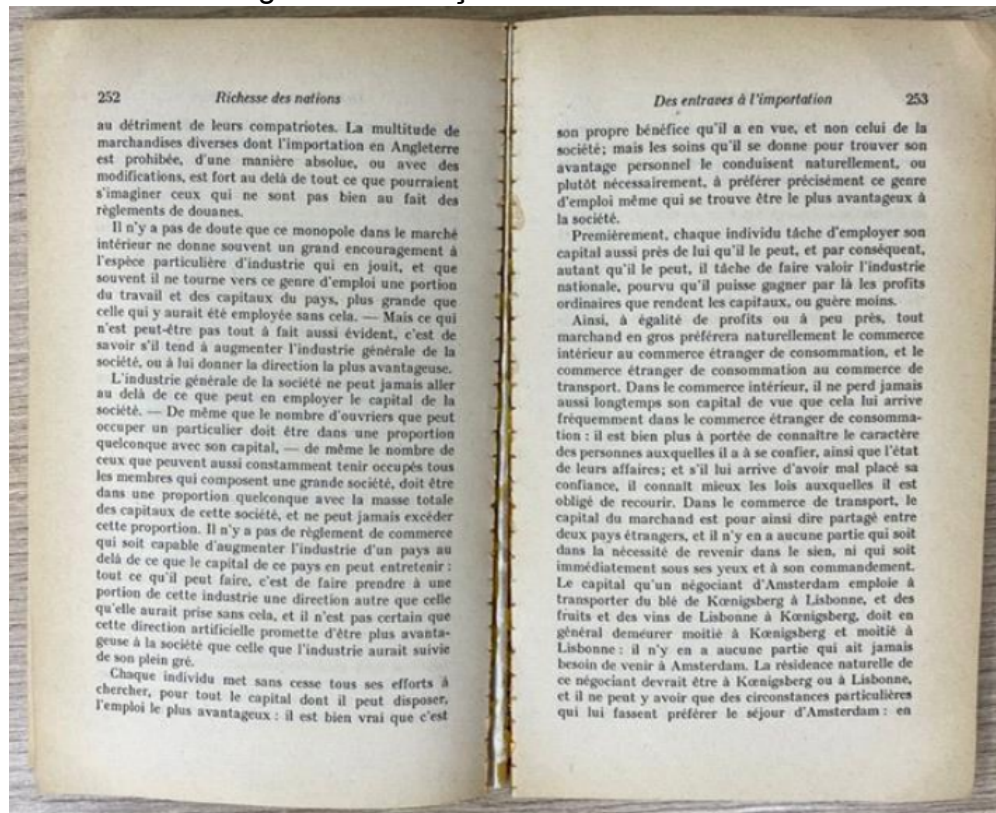
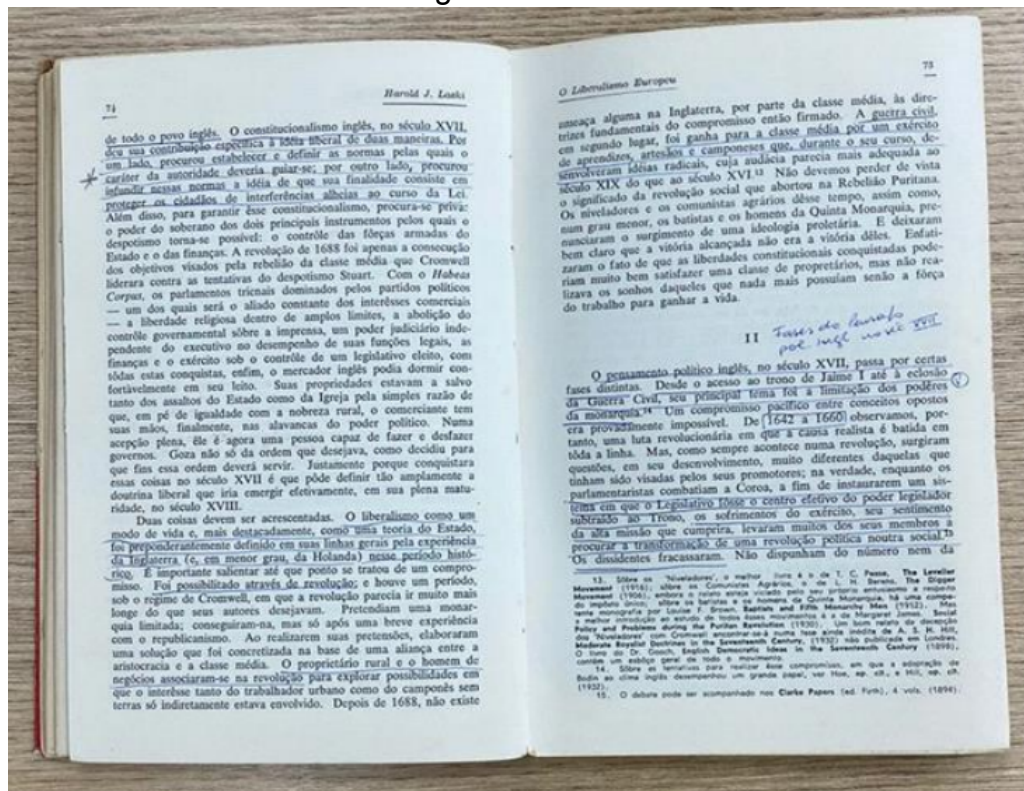


Figura 3: Riscos



Sendo este o grupo 3, ficamos responsáveis pela Classe 2 — Religião — da Bibli-

oteca Central da Universidade de Brasília (**BCE**).

Sem o cadastro de livros a serem pesquisados, o plano amostral mais indicado seria o conglomerado em dois estágios [1], mas o trabalho foi feito como se fosse um plano aleatório simples ou estratificado, seguindo o esquema a seguir:

Fomos à biblioteca verificar primeiramente quantas estantes existem em sua classe correspondente. No caso, haviam apenas duas estantes para esta classe, a qual dividimos em 4 (frente e verso). A seguir, usando os números aleatórios de uma página específica recebida do livro “A Million Random Digits with 100000 Normal Deviates” [2], cada componente do grupo selecionou 1 (uma) estante a ser pesquisada. Os membros do grupo garantiram que a mesma estante não foi utilizada mais de uma vez.

Como o tamanho das prateleiras era diferente para cada membro do grupo, utilizou-se critérios específicos para estimar o número total de livros para antes de realizar o sorteio. Para a prateleira 4, a menor de todas, a medida adotada foi de contar o número total de livros para obter o parâmetro, que é  $N = 196$ ; e posteriormente sortear deste  $N$  total um  $n = 25$  amostras, sorteadas segundo a tabela de números aleatórios [2] na página 265, linha 13.225.

### 3 Referencial teórico

A amostragem é uma técnica estatística que permite conseguir resultados aproximados para a população a partir de uma quantidade menor de informações, ou seja, por meio de observações de apenas um “pedaço” dessa população. Dessa forma, consegue-se, com um intervalo de confiança, reduzir os custos e otimizar o tempo de coleta de informações sem perder a credibilidade para o estudo em questão.

#### 3.1 Amostragem Aleatória Simples

Na amostragem aleatória simples, cada componente da população tem a mesma probabilidade de ser selecionado para fazer parte da amostra, ou seja, dada uma população com  $N$  indivíduos, cada um possui probabilidade igual a  $\frac{1}{N}$  de ser selecionado. Além disso, é necessário que a seleção de indivíduos seja feita de forma aleatória.

Quando a amostra é relativamente grande, o Teorema do Limite Central garante que a média amostral ( $\bar{X}$ ) aproxima-se de uma distribuição normal com média  $\mu$  e

variância  $\sigma^2/n$ , e o tamanho necessário de amostra ( $n'$ ), para um determinado erro  $\varepsilon$ , nível de confiança  $\gamma$  e população infinita, é dado pela seguinte expressão:

$$n' = \frac{z_{\frac{\alpha}{2}}^2 \times s^2}{\varepsilon^2}$$

Com:

- $z_{\frac{\alpha}{2}}$ : quantil da distribuição normal padrão e aproximadamente igual a 1,96 para  $\alpha = 5\%$  e 1,64 para  $\alpha = 10\%$
- $\alpha$ : nível de significância, equivale a  $1 - \gamma$
- $s^2$ : variância amostral da variável analisada
- $\varepsilon$ : erro sobre a estimativa do parâmetro populacional
- $\mu$ : média populacional da variável analisada
- $\sigma^2$ : variância populacional da variável analisada

O erro  $\varepsilon$  significa que, se fosse possível construir uma grande quantidade de intervalos de confiança da forma  $\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon$ , todos baseados em amostras independentes de tamanho  $n'$ ,  $100 \times \gamma\%$  (em geral, 90% ou 95%) conteriam o parâmetro populacional  $\mu$ .

Quando se conhece o tamanho da população ( $N$ ), o valor de  $n'$  pode ser corrigido para se reduzir o tamanho necessário de amostra para:

$$n = \frac{n'N}{N + n'}$$

É importante ressaltar que, como a proporção pode ser escrita como a média de variáveis indicadoras, os resultados apresentados acima também são válidos. Além disso, caso não se conheça o valor verdadeiro da variância, pode-se utilizar uma cota superior de 0,25, pois este é o valor máximo da variância de uma variável indicadora.

[1]

### 3.1.1 Estimativa de parâmetros

A média amostral é dada por:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

A média amostral é um estimador não viesado para a média populacional (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1])

A variância para uma amostra aleatória simples com reposição ( $AAS_c$ ) é dada por (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1]):

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}$$

Ainda segundo FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1], a variância para uma amostra aleatória simples sem reposição ( $AAS_s$ ) é dada por:

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N - 1}$$

Sendo assim, definem-se as variâncias da média  $\bar{x}$  como (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1]):

$$Var_{AAS_c}(\bar{x}) = \frac{S^2}{n} \frac{(N - n)}{N} = \frac{S^2}{n} (1 - f)$$

onde  $f$  é dado por  $\frac{n}{N}$ .

Essa proporção inserida na fórmula é conhecida como fator de Correção para População Finita (CPF), ou do inglês Finite Population Correction (FPC) (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1]).

É válido observar que para o cálculo dessa variância é preciso conhecer previamente alguns parâmetros populacionais tais como seu tamanho e a média de seus valores. Na prática, tais parâmetros não podem ser conhecidos, mas podem ser estimados a partir dos dados amostrais (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1]).

De acordo com Cochran (FALCÃO,2013,p.5 [3] apud COCHRAN,1977 [1])), um estimador não viciado da variância populacional estimada  $S^2$  ou  $\sigma^2$  é dado por:

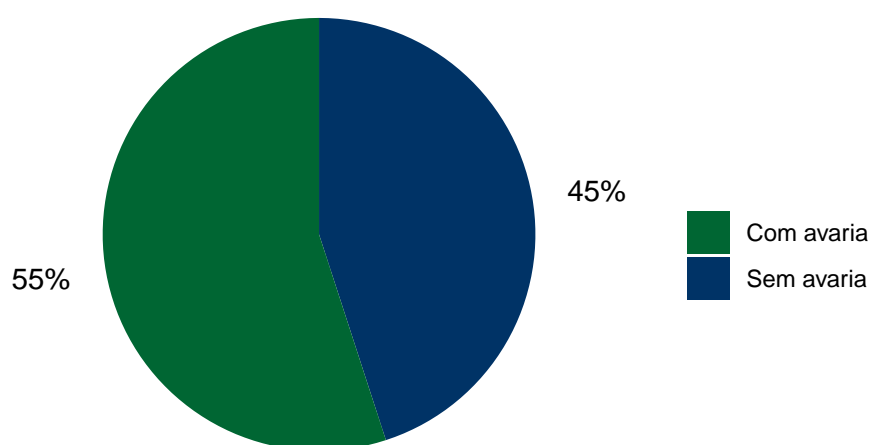
$$\widehat{Var}(\bar{x}) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = s^2$$

## 4 Análises

### 4.1 Análise exploratória

#### 4.1.1 Avaria

Figura 4: Gráfico de setores da proporção de livros avariados





#### 4.1.2 Avaria por prateleira

Figura 5: Gráfico de barras da quantidade de avaria pela prateleira

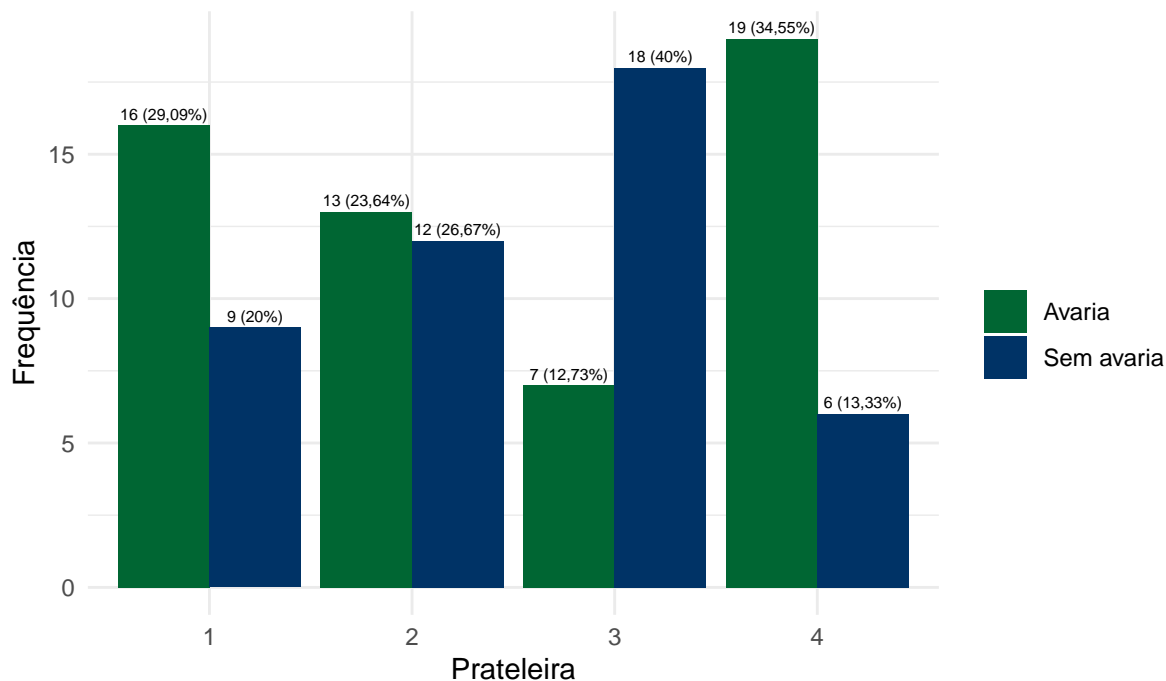
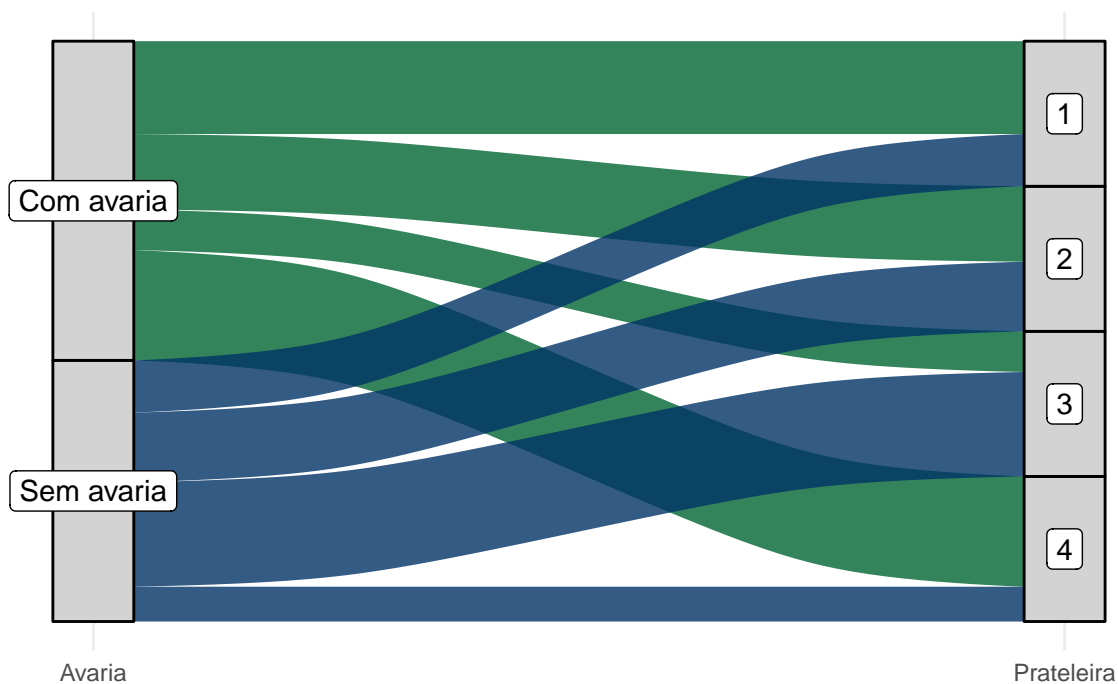


Figura 6: Diagrama de Sankey da proporção de livros avariados pela prateleiras



Testaremos a hipótese de que a quantidade de avarias difere entre as prateleiras.

$$\begin{cases} H_0 : \text{As médias de livros avariados das prateleiras são iguais} \\ H_1 : \text{Existe pelo menos uma prateleira com média diferente} \end{cases}$$

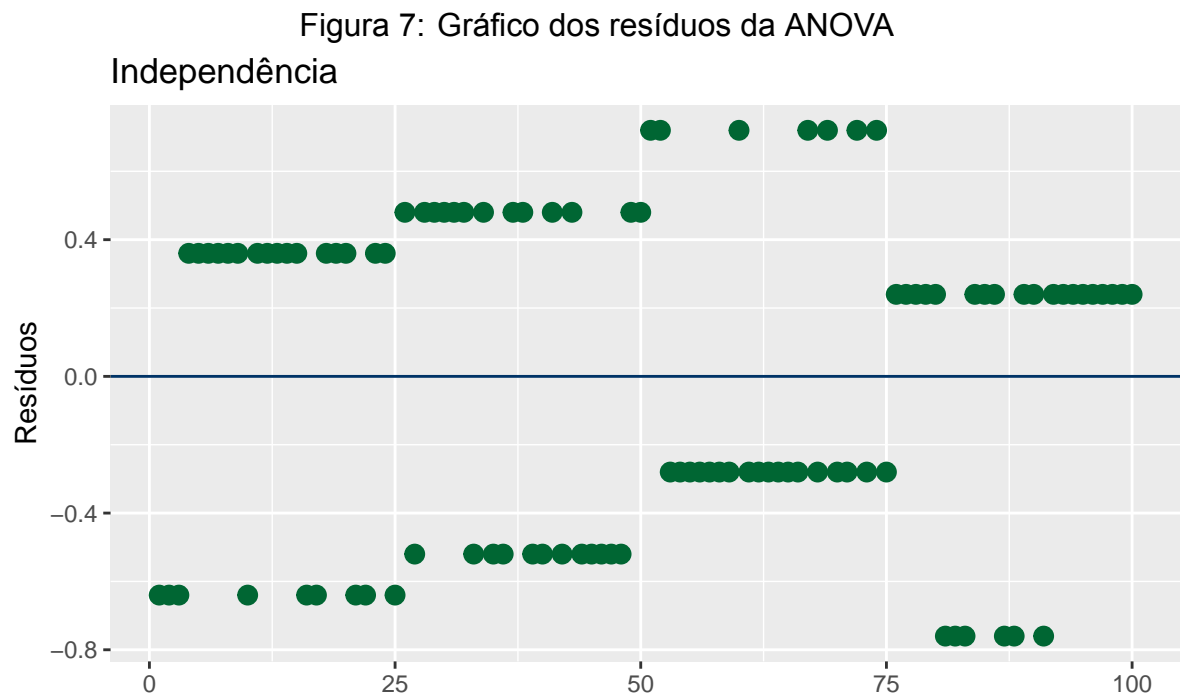
Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	Estatística F	P-valor
Prateleiras	3	3,15	1,05	4,667	0,00435
Resíduos	96	21,6	0,225		
Total	99	24,75			

Sob um nível de significância  $\alpha = 5\%$ , rejeitamos a hipótese nula  $H_0$  de igualdade de médias de avarias nas prateleiras. Ou seja, ao menos uma prateleira difere em relação a quantidade de livros avariados.

Devemos verificar os pressupostos do teste ANOVA.

#### 4.1.2.1 Independência

Testaremos a independência pelo gráfico de dispersão dos resíduos.



Por este gráfico, não podemos concluir pela independência dos resíduos, pois estes formam padrões lineares no gráfico de dispersão.

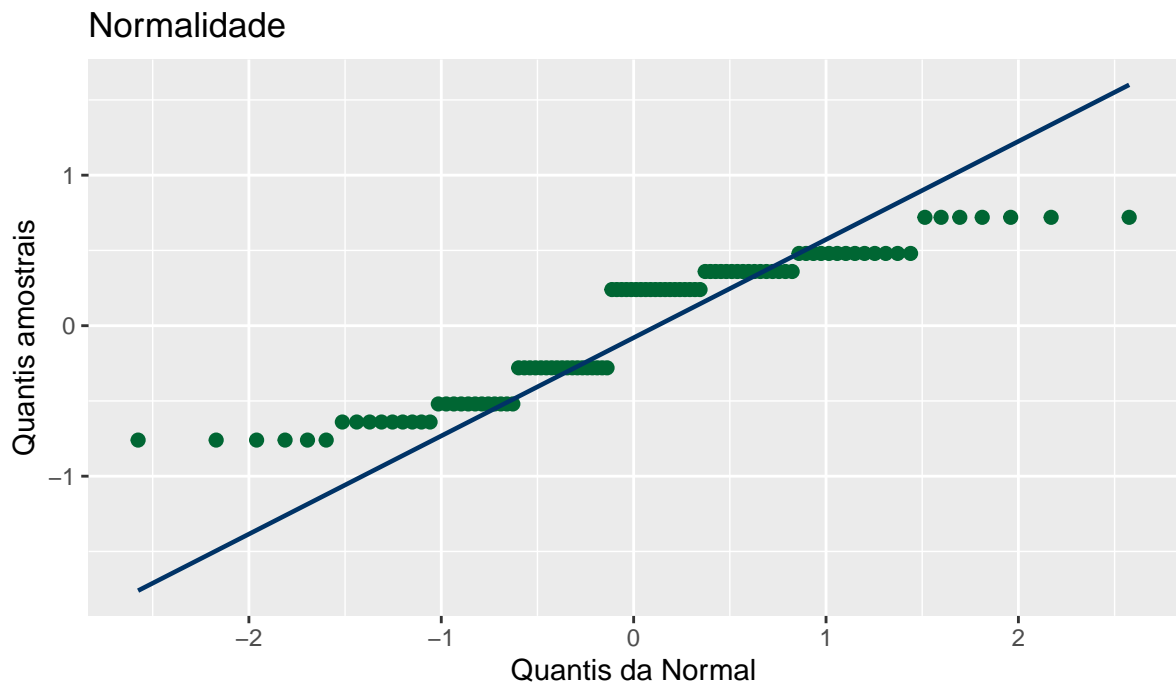
#### 4.1.2.2 Normalidade

$$\begin{cases} H_0 : \text{Os resíduos seguem distribuição normal} \\ H_1 : \text{Os resíduos não seguem distribuição normal} \end{cases}$$

Quadro 1: P-valor do teste de Shapiro-Wilk para normalidade dos resíduos

Variável	Teste Shapiro-Wilk	Decisão do teste
Resíduos ANOVA	<0,001	Rejeita $H_0$

Figura 8: Gráfico Q-Q dos resíduos da ANOVA



Pelo teste de Shapiro-Wilk e pelo Gráfico Q-Q, concluímos que os resíduos não seguem distribuição normal

#### 4.1.2.3 Homocedasticidade

Como os resíduos não seguem distribuição normal, faremos o teste de Levene para homocedasticidade, em detrimento do teste de Bartlett, este muito sensível a normalidade.

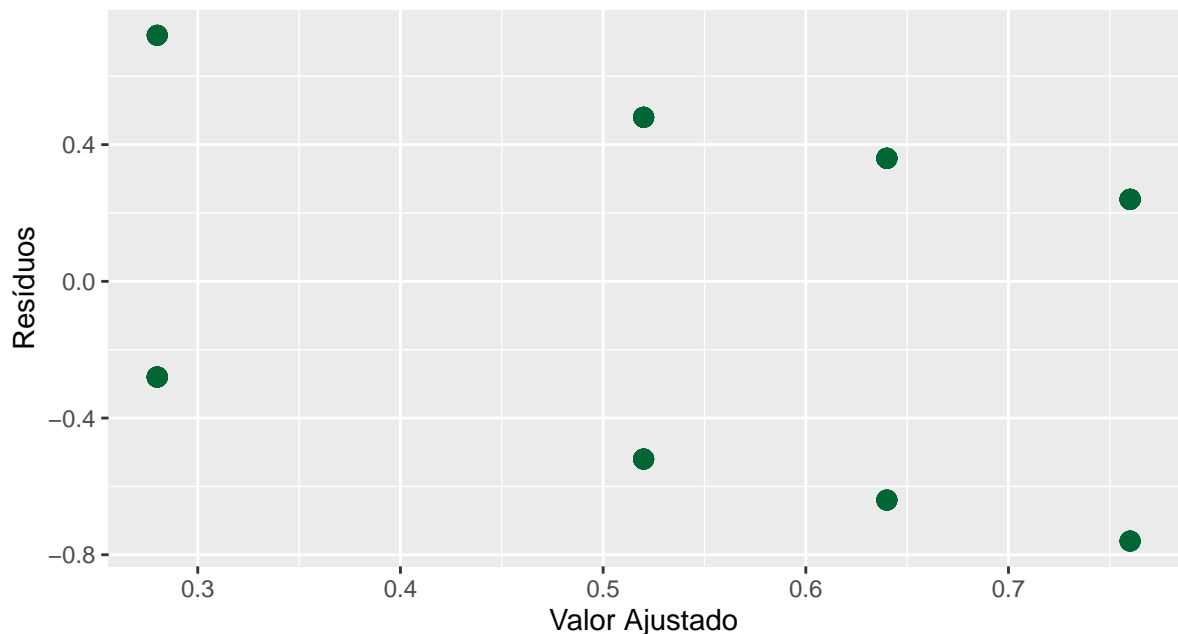
$$\begin{cases} H_0 : \text{As variâncias das prateleiras são homogêneas} \\ H_1 : \text{Ao menos uma prateleira contém variância heterogênea} \end{cases}$$

Quadro 2: P-valor do teste de Levene para homocedasticidade

Variável	Teste de Levene	Decisão do teste
Variâncias das prateleiras	0,298	Não rejeita $H_0$

Pelo teste de Levene e Gráfico dos resíduos pelos valores ajustados, concluímos pela não rejeição de  $H_0$ , ou seja, as variâncias são homogêneas.

Figura 9: Gráfico dos resíduos pelos resíduos ajustados da ANOVA Homocedasticidade



Rejeitados alguns dos pressupostos, devemos portanto utilizar uma abordagem não paramétrica para testar a hipótese da diferença das médias de avarias nas prateleiras. Utilizaremos o teste de Kruskal-Wallis.

$$\begin{cases} H_0 : \text{As medianas de livros avariados das prateleiras são iguais} \\ H_1 : \text{Existe pelo menos uma prateleira com mediana diferente} \end{cases}$$

Quadro 3: P-valor do teste de Kruskal-Wallis

Variável	Teste Kruskal-Wallis	Decisão do teste
Resíduos ANOVA	0.005	Rejeita $H_0$

Pelo teste de Kruskal-Wallis, concluímos que existem diferenças entre as medianas de avarias nas prateleiras.

4.1.3 Tipos de avaria

Figura 10: Gráfico de setores do tipo de avaria

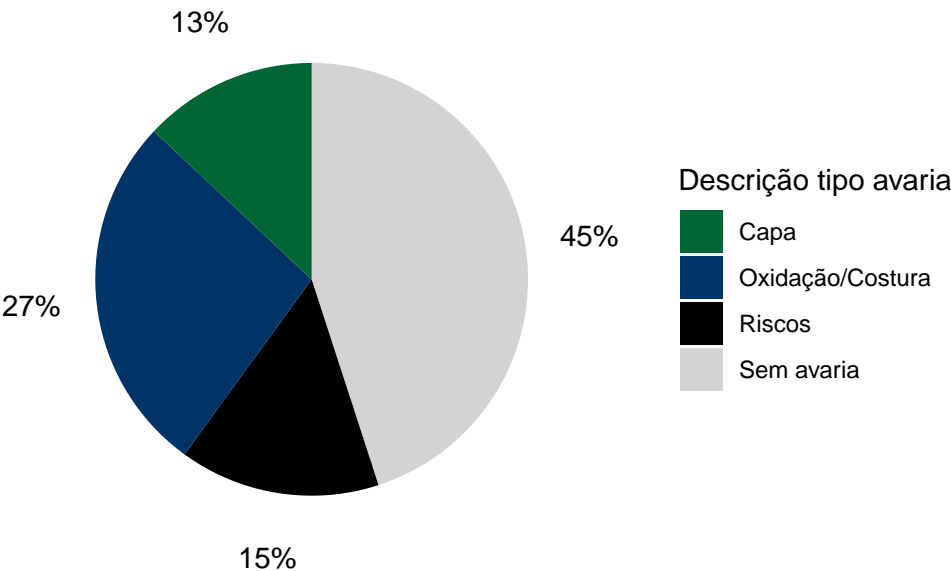
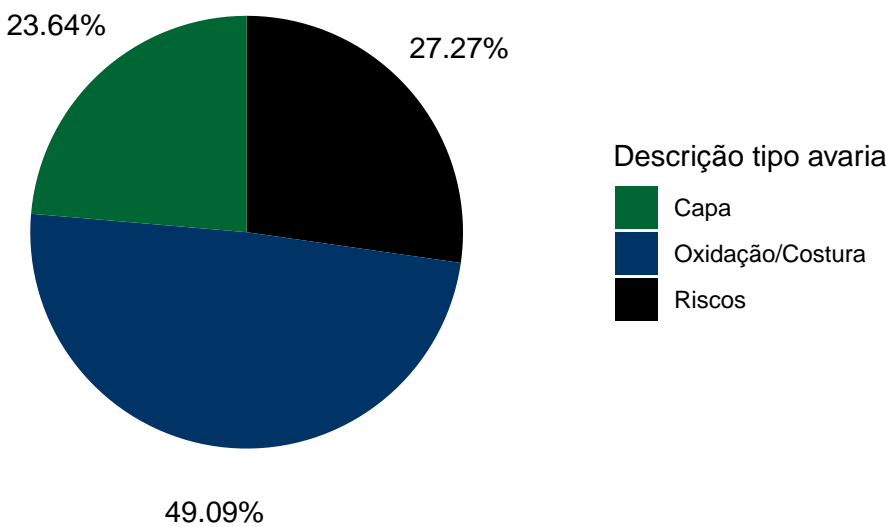


Figura 11: Gráfico de setores do tipo de avaria



#### 4.1.4 Tipo de avaria por prateleira

Figura 12: Gráfico de barras do tipo de avaria pela prateleira

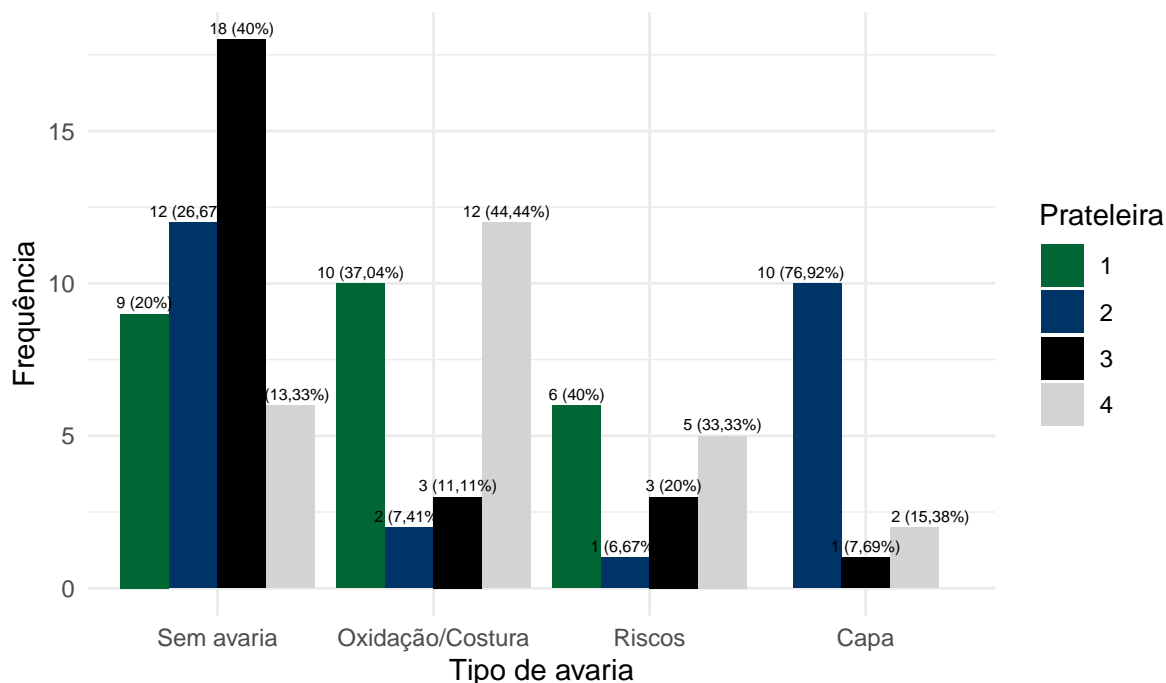


Tabela 1: Frequências dos tipos de avaria pela prateleira

Tipos de avaria	Prateleira 1	Prateleira 2	Prateleira 3	Prateleira 4	Total
Capa	0	10	1	2	13
Oxidação/Costura	10	2	3	12	27
Riscos	6	1	3	5	15
<b>Total</b>	<b>16</b>	<b>13</b>	<b>7</b>	<b>19</b>	<b>42</b>

Faremos um teste para testar a hipótese do tipo de avaria ter relação com a prateleira no qual o livro se encontra.

$$\begin{cases} H_0 : \text{O tipo de avaria é independente da prateleira ao qual o livro se encontra} \\ H_1 : \text{Existe dependência do tipo de avaria à prateleira em que o livro se encontra} \end{cases}$$

Quadro 4: P-valor do teste Qui-Quadrado de independência

Variável	Teste Qui-quadrado	Decisão do teste
Tipo de avaria	<0,0001	Rejeita $H_0$



Pelo teste Qui-quadrado, concluímos que existe relação entre o tipo de avaria e a prateleira em que o livro se encontra.

## 4.2 Amostragem

Com base nas fórmulas referenciadas no referencial teórico, estimamos a verdadeira proporção de livros avariados na população.

Com estatística pontual  $p = 0,55$  e erro padrão  $EP = 0,0497$ , inferimos sobre o intervalo de confiança  $\alpha = 5\%$  assintótico para a proporção em:

Quadro 5: Intervalo de confiança para a proporção de livros avariados na população

Estatística pontual	Intervalo de Confiança (95%)
0,56	0,4525    0,6475

Aqui, não estamos fazendo correção de população finita, que deve ser posteriormente realizado caso o parâmetro  $N$  seja conhecido.

## 5 Conclusão

## 6 Códigos Computacionais

Confira na íntegra no Github.

```
# 1.0 Pacotes ----
```

```
if (!require("pacman")) install.packages("pacman")
```

```
pacman::p_load(  
  tidyverse, data.table,  
  readxl, readr, ggcorrplot, cowplot,  
  RColorBrewer, scales, nortest, xlsx,  
  skimr, xtable  
)
```

```
windowsFonts(Arial=windowsFont("sans"))
```

```
options(scipen=999)
```

```
# Definindo paleta de cores da UnB
```

```
cores_unb <- c("#006633", "#003366", "#000000", "lightgray")
```

```
percent <- function(absolute, digits = 2) {  
  return(round(100 * absolute / sum(absolute), digits))  
}
```

```
# Definindo função que retorna banco de dados com frequências  
# relativas e absolutas de uma variável categórica
```

```
vector_frequencies <- function(vector) {  
  frequency <- vector %>%  
    table() %>%  
    as_tibble() %>%
```

```

    mutate(
      rel = n %>%
        percent() %>%
        paste("%", sep = " ")
    )
  colnames(frequency) <- c("groups", "absolute", "relative")
  return(frequency)
}

```

# 1.1 Dados ----

```

df <- read_excel("banco/grupo3.xlsx",
  col_types = c("skip",
    "text", "text", "text", "text",
    "text", "text", "text", "text"))

```

# 1.2 ETL ----

```

colnames(df)
df$Classe <- factor(df$Classe)
df$Avaria <- as.numeric(df$Avaria)
df$Descrição_avaria <- factor(df$Descrição_avaria)
df$Tipo_avaria <- factor(df$Tipo_avaria)
df$Descrição_tipo_avaria <- factor(df$Descrição_tipo_avaria)
df$Prateleira <- factor(df$Prateleira)

```

# 2 Análises ----

# 2.0 Exploratória ----

# 2.0.1 Tabela completa em LaTeX ----

```

p_load(xtable)
xtable(df)

```

# 2.0.2 Tabela de contingência do tipo de avaria pela prateleira - LaTeX

```

table(df$Descrição_tipo_avaria, df$Prateleira)

```

# 2.0.3 Gráfico: Tipo de avaria pela prateleira ----

```
df$Descrição_tipo_avaria <- as.character(df$Descrição_tipo_avaria)
df %>%
  select(Descrição_tipo_avaria , Prateleira) %>%
  mutate(Descrição_tipo_avaria = ifelse(is.na(Descrição_tipo_avaria), "",
  group_by(Descrição_tipo_avaria , Prateleira) %>%
  summarise(freq = n()) %>%
  mutate(
    freq_relativa = freq %>% percent(),
    porcentagens = str_c(freq_relativa , "%") %>% str_replace("\\.", " ,",
    legendas = str_squish(str_c(freq , " (", porcentagens , ")"))
  ) %>%
  ggplot() +
  aes(
    x = fct_reorder(Descrição_tipo_avaria , freq , .desc = T),
    y = freq ,
    fill = Prateleira ,
    label = legendas
  ) +
  geom_col(position = position_dodge2(preserve = "single", padding = 0)
  geom_text(
    position = position_dodge(width = .9),
    vjust = -0.5, hjust = 0.5,
    size = 2) +
  scale_fill_manual(values = cores_unb)+
  labs(x = "Tipo de avaria", y = "Frequência") +
  theme_minimal()
#ggsave("resultados/grafico1.pdf", width = 158, height = 93, units = "m
```

# 2.0.4 Proporção avaria ----

```

contagem2 <- df %>%
  mutate(Descrição_avaria = ifelse(Descrição_avaria == "Sem avaria", "S
  group_by(Descrição_avaria) %>%
  summarise(Freq = n()) %>%
  mutate(Prop = round(100 * (Freq / sum(Freq)), 2)) %>%
  arrange(desc(Descrição_avaria)) %>%
  mutate(posicao = cumsum(Prop) - 0.5 * Prop,
         ymax = cumsum(Prop),
         ymin = c(0, head(ymax, n=-1)))

ggplot(contagem2) +
  aes(
    x = factor(""),
    y = Prop,
    fill = factor(Descrição_avaria)
  ) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  scale_fill_manual(values = cores_unb, name = "") +
  theme_void() +
  geom_text(
    aes(x = 1.8, y = posicao, label = paste0(Prop, "%")),
    color = "black"
  )
#ggsave("resultados/grafico2.pdf", width = 158, height = 93, units = "m

```

# 2.0.5 Proporção tipo de avaria ----

```

contagem <- df %>%
  mutate(Descrição_tipo_avaria = ifelse(is.na(Descrição_tipo_avaria), "S
  group_by(Descrição_tipo_avaria) %>%

```

```

summarise(Freq = n()) %>%
mutate(Prop = round(100 * (Freq / sum(Freq)), 2)) %>%
arrange(desc(Descrição_tipo_avaria)) %>%
mutate(posicao = cumsum(Prop) - 0.5 * Prop,
       ymax = cumsum(Prop),
       ymin = c(0, head(ymax, n=-1)))

ggplot(contagem) +
  aes(
    x = factor(""),
    y = Prop,
    fill = factor(Descrição_tipo_avaria)
  ) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  scale_fill_manual(values = cores_unb, name = "Descrição tipo avaria") +
  theme_void() +
  geom_text(
    aes(x = 1.8, y = posicao, label = paste0(Prop, "%")),
    color = "black"
  )
#ggsave("resultados/grafico3.pdf", width = 158, height = 93, units = "mm")

# 2.0.6 Proporção tipo de avaria - tirando "sem avaria" ----
contagem3 <- df %>%
  na.omit() %>%
  group_by(Descrição_tipo_avaria) %>%
  summarise(Freq = n()) %>%
  mutate(Prop = round(100 * (Freq / sum(Freq)), 2)) %>%
  arrange(desc(Descrição_tipo_avaria)) %>%
  mutate(posicao = cumsum(Prop) - 0.5 * Prop,
       ymax = cumsum(Prop),

```

```

      ymin = c(0, head(ymax, n=-1)))

ggplot(contagem3) +
  aes(
    x = factor(""),
    y = Prop,
    fill = factor(Descrição_tipo_avaria)
  ) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  scale_fill_manual(values = cores_unb, name = "Descrição tipo avaria") +
  theme_void() +
  geom_text(
    aes(x = 1.8, y = posicao, label = paste0(Prop, "%")),
    color = "black"
  )
#ggsave("resultados/grafico4.pdf", width = 158, height = 93, units = "mm")

# 2.1 Proporção estimada na população, com intervalo de confiança; esta
p_load(samplingbook)
Sprop(y=df$Avaria)

# 2.1.1 Mesmo, porém "chutando" um valor para N ----
# N = População de livros na Classe 2 - Religião - na BCE.
Sprop(y=df$Avaria, N=197+2*1500)

# 2.2.1 Verificando se a avaria pode ser explicada pelo tipo da avaria
summary(aov(Avaria ~ Tipo_avaria + Prateleira, data=df)) # Não significa

# 2.2.2 Verificando se a avaria pode ser explicada por qual prateleira
df %>%

```

```

select(Descrição_avaria , Prateleira ) %>%
group_by(Descrição_avaria , Prateleira ) %>%
summarise(freq = n()) %>%
mutate(
  freq_relativa = freq %>% percent() ,
  porcentagens = str_c(freq_relativa , "%") %>% str_replace("\\.", " ,")
  legendas = str_squish(str_c(freq , " (", porcentagens , ")"))
) %>%
ggplot() +
aes(
  x = fct_reorder(factor(Prateleira), freq , .desc = T),
  y = freq ,
  fill = Descrição_avaria ,
  label = legendas
) +
labs(fill = '') +
geom_col(position = position_dodge2(preserve = "single", padding = 0)
geom_text(
  position = position_dodge(width = .9),
  vjust = -0.5, hjust = 0.5,
  size = 2) +
scale_fill_manual(values = cores_unb)+
labs(x = "Prateleira", y = "Frequência") +
theme_minimal()
#ggsave("resultados/grafico5.pdf", width = 158, height = 93, units = "m

anova = aov(Avaria ~ Prateleira ,data=df)
summary(anova) # O teste anova indica que a prateleira em que o livro f
xtable(summary(anova))
# Pressupostos do teste

```



```
# 2.2.2.1 Normalidade dos resíduos ----
```

```
shapiro.test(anova$residuals) # Não são normais
```

```
qqnorm(anova$residuals)
```

```
qqline(anova$residuals)
```

```
ggplot(anova, aes(sample=.resid)) +
```

```
  stat_qq(colour=cores_unb[1], size = 2) +
```

```
  stat_qq_line(size = 0.8, colour = cores_unb[2]) +
```

```
  labs(x="Quantis da Normal", y="Quantis amostrais", title = "Normalidade")
```

```
#ggsave("resultados/grafico6.pdf", width = 158, height = 93, units = "mm")
```

```
# 2.2.2.2 Independência ----
```

```
plot(anova$residuals)
```

```
plot(anova$residuals~anova$fitted.values)
```

```
# Não aparentam ser independentes
```

```
ggplot(anova, aes(x=c(1:length(anova$residuals)),y=.resid)) +
```

```
  geom_point(colour=cores_unb[1], size=3) +
```

```
  geom_hline(yintercept=0,colour=cores_unb[2]) +
```

```
  labs(x="", y="Resíduos", title = "Independência")
```

```
#ggsave("resultados/grafico7.pdf", width = 158, height = 93, units = "mm")
```

```
# 2.2.2.3 Homocedasticidade ----
```

```
pacman::p_load(car)
```

```
leveneTest(y=df$Avaria ,group=df$Prateleira)
```

```
# Variâncias homogêneas.
```

```
residuos <- anova$residuals
```

```
valores_ajustados <- anova$fitted.values
```

```

ggplot(data = data.frame(fitted.values = valores_ajustados , residuos =
      mapping = aes(x = valores_ajustados , y = residuos)) +
  geom_point(colour=cores_unb[1], size=3) +
  scale_x_continuous() +
  labs(x="Valor Ajustado",
      y="Resíduos", title = "Homocedasticidade")
#ggsave("resultados/grafico8.pdf", width = 158, height = 93, units = "m

```

# 2.2.3 Teste não paramétrico - Kruskal-Wallis ----

```

kruskal.test(Avaria ~ Prateleira ,data=df) # Pelo teste não paramétrico

```

# 2.2.4 Verificando se o tipo de avaria é homogêneo entre as prateleiras

```

p_load(stats)
chisq.test(df$Tipo_avaria ,df$Prateleira
          #           , simulate.p.value = TRUE,B=10000
)

```

# O teste qui-quadrado indica que O tipo de avaria é independente da pr

# 2.2.5 Diagrama de Sankey: Proporção de livros avariados/não avariados

```

p_load(ggalluvial)
prop <- df |>
  select(Descrição_avaria , Prateleira) |>
  count(Descrição_avaria , Prateleira) |>
  mutate(proptot = prop.table(n),
         Descrição_avaria = ifelse(Descrição_avaria == "Avaria", "Com av
ggplot(as.data.frame(prop),
      aes(y = proptot , axis1 = factor(Descrição_avaria), axis2 = facto
  geom_alluvium(aes(fill = factor(Descrição_avaria)), width = 1/12,alpha
  geom_stratum(width = 1/12, fill = cores_unb[4], colour = cores_unb[3]

```

```

geom_label(stat = "stratum", infer.label = TRUE) +
scale_x_discrete(limits = c("Avaria", "Prateleira"),
                  expand = c(.05, .05),
                  labels = c("Avaria", "Prateleira")) +
scale_fill_manual(values = cores_unb) +
scale_y_continuous(labels = NULL,
                   name = NULL,
                   breaks = NULL) +
theme_minimal()
#ggsave("resultados/grafico9.pdf", width = 158, height = 93, units = "m

```

## Referências

- [1] W.G. COCHRAN. *Sampling Techniques*. Wiley Series in Probability and Statistics. Wiley, 1977. ISBN: 9780471162407.
- [2] Rand Corporation. *A Million Random Digits with 100,000 Normal Deviates*. Free Press, 1955. ISBN: 9780029257906.
- [3] João Renato Falcão. *IMPLEMENTAÇÃO DE ALGORITMO COMPUTACIONAL PARA ANÁLISE DE DADOS AMOSTRAIS COMPLEXOS*. Relatório. Universidade de Brasília, 2013.