Estudo de Caso 01: Comparação do IMC médio de alunos do PPGEE-UFMG ao longo de dois semestres

Diego Pontes, Elias Vieira, Matheus Bitarães

Janeiro, 2021

Descrição do problema

Neste estudo, deseja-se comparar o Índice de Massa Corpórea (IMC) médio de duas populações de alunos da pós-graduação da Engenharia Elétrica (PPGEE) da UFMG no segundo semestre de 2016 e de 2017. Para isso, foram disponibilizadas duas amostras, sendo uma para cada semestre em questão, onde serão feitas as análises para o estudo já mencionado.

Introdução

Reconhecido internacionalmente pela Organização Mundial da Saúde (OMS), o IMC indica o peso adequado para cada pessoa, fazendo uma relação entre sua massa corpórea (em kg) e sua altura (em m) [1], conforme a Equação 1.

$$IMC = \frac{peso}{altura * altura} \tag{1}$$

Pode-se classificar o valor do IMC conforme apresentado abaixo [1]:

- IMC abaixo de 18,5: Peso abaixo do normal;
- IMC entre 18,5 e 24,9: São pesos considerados normais pela OMS;
- IMC entre 25 e 29,9: Peso em pré-obesidade ou acima do peso;
- IMC entre 30 e 34,9: Este índice indica obesidade grau um;
- IMC acima 35 e 39,9: Indica obesidade grau dois
- IMC acima de 40: Indica obesidade grau três ou mórbida

Design do Experimento

Como já mencionado, deseja-se comparar o IMC dos alunos do PPGEE-UFMG de dois semestres distintos, utilizando as amostras fornecidas. Para tal estudo, serão feitas duas análises, além de testes estatísticos independentes considerando duas subpopulações distintas, sendo uma somente do sexo masculino e outra somente do sexo feminino.

As seguintes hipóteses estatísticas foram definidas:

1) Há evidências de que a média do IMC dos alunos do PPGEE-UFMG de 2/2016 é diferente da média do IMC dos alunos do PPGEE-UFMG de 2/2017? (subpopulação masculina)

2) Há evidências de que a mediana do IMC das alunas do PPGEE-UFMG de 2/2016 é diferente da mediana do IMC das alunas do PPGEE-UFMG de 2/2017? (subpopulação feminina)

No decorrer do estudo ficará claro o motivo pelo qual foi utilizado média e mediana para a subpopulação masculina e feminina, respectivamente.

Dadas as hipóteses estatísticas descritas acima, podem-se definir as seguintes hipóteses de testes, em função da média e mediana do IMC dos alunos do sexo masculino e do sexo feminino, respectivamente:

$$\begin{cases} H_0 : \mu_{m2016} = \mu_{m2017} \\ H_1 : \mu_{m2016} \neq \mu_{m2017} \end{cases}$$

$$\begin{cases} H_0 : \mu_{f2016} = \mu_{f2017} \\ H_1 : \mu_{f2016} \neq \mu_{f2017} \end{cases}$$

Além das hipóteses estatísticas acima, tem-se as seguintes definições:

1) Nível de significância (α) de 0,05. O nível de significância é a probabilidade de ocorrência de um falso positivo em qualquer procedimento de teste de hipótese [2].
2) Potência do Teste (π) = 1 - (β) = 0,80. Onde beta é a probabilidade de ocorrência de um falso negativo em qualquer procedimento de teste de hipótese [2] e, portanto, a potência de teste quantifica a sensibilidade do teste a efeitos que violam sua hipótese nula [2].

Análise Estatística

Importação dos dados

Foram importados os arquivos *imc_20162.csv* e *CS01_20172.csv* para o estudo proposto.

```
# importação dos dados
raw_data_2016 <- read.csv(file = 'imc_20162.csv')
raw_data_2017 <- read.csv(file = 'CS01_20172.csv', sep=';')

head(raw_data_2016)
```

```
##   ID Course Gender Height.m Weight.kg
## 1  1  PPGEE      F     1.57      45.5
## 2  2  PPGEE      F     1.62      53.0
## 3  3  PPGEE      F     1.70      57.0
## 4  4  PPGEE      F     1.62      59.0
## 5  5  PPGEE      F     1.67      63.0
## 6  6  PPGEE      F     1.76      78.0
```

```
head(raw_data_2017)
```

```
##   Weight.kg height.m Sex Age.years
## 1      89.0     1.73   M        23
## 2      72.5     1.64   M        28
## 3      84.0     1.70   M        34
## 4      90.0     1.72   M        27
## 5      60.0     1.70   M        33
## 6      79.0     1.80   M        27
```

Como pode ser visto, há diferenças estruturais entre os arquivos, como colunas com nomes diferentes, além de dados de alunos que não pertencem ao PPGEE-UFMG no arquivo de dados de 2016. Portanto, estes dados foram tratados para que ficassem com mesma estrutura, conforme códigos abaixo.

```
# Filtra dados apenas de estudantes do ppgee (necessario apenas em 2016)
raw_data_2016 <- subset(raw_data_2016, Course=="PPGEE")

# renomeia coluna de 2016
names(raw_data_2016)[names(raw_data_2016) == "Gender"] <- "Sex"

# renomeia coluna de 2017
names(raw_data_2017)[names(raw_data_2017) == "height.m"] <- "Height.m"
```

Na sequência, deve-se criar uma nova coluna com o calculo do IMC e separar os dados entre masculino e feminino, conforme códigos abaixo.

```
# cria coluna com calculo do IMC
raw_data_2016$IMC = raw_data_2016$Weight.kg / (raw_data_2016$Height.m * raw_data_2016$Height.m)
raw_data_2017$IMC = raw_data_2017$Weight.kg / (raw_data_2017$Height.m * raw_data_2017$Height.m)

# separa entre masculino e feminino e armazena apenas o IMC
imc_m_2016 <- subset(raw_data_2016, Sex=="M")$IMC
imc_f_2016 <- subset(raw_data_2016, Sex=="F")$IMC
imc_m_2017 <- subset(raw_data_2017, Sex=="M")$IMC
imc_f_2017 <- subset(raw_data_2017, Sex=="F")$IMC
```

Dados estatísticos

Com as amostras tratadas e separadas conforme proposta inicial e em posse do IMC destas subpopulações, tem-se os seguintes dados estatísticos:

```
# imprime um sumario com as principais informações estatisticas dos IMCs
summary(imc_m_2016)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.58   22.47   24.36   24.94   27.14   37.55
```

```
summary(imc_m_2017)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.72   22.41   23.75   24.29   26.22   30.42
```

```
summary(imc_f_2016)
```

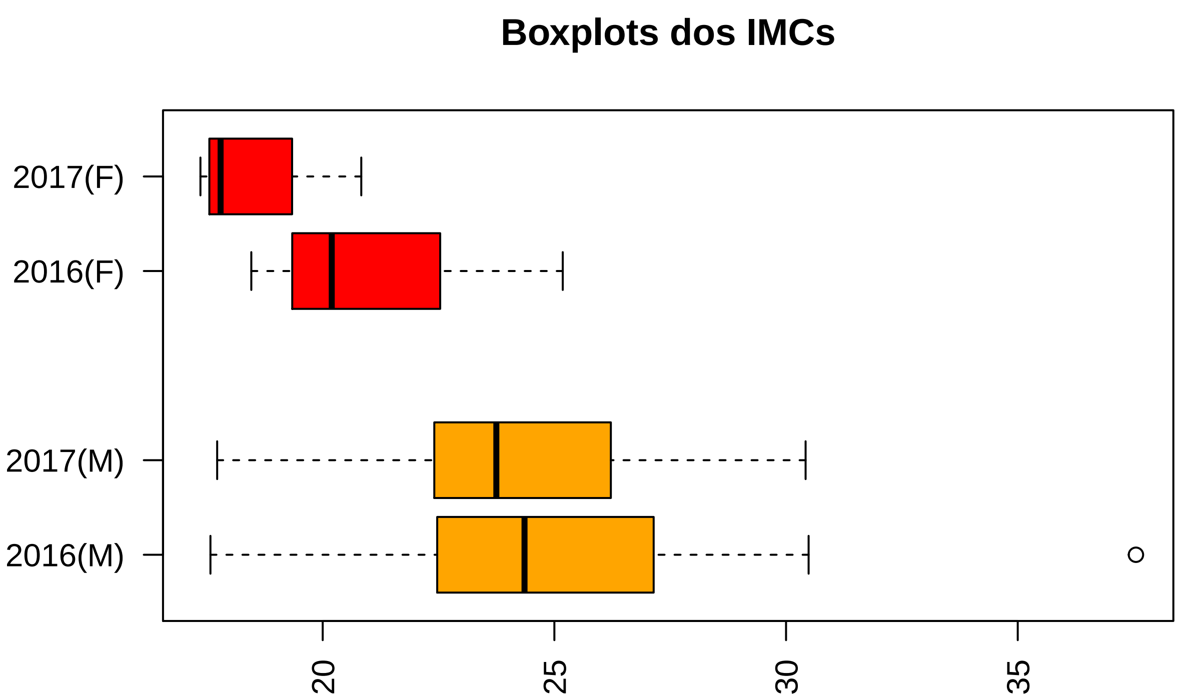
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.46   19.34   20.20   21.08   22.54   25.18
```

```
summary(imc_f_2017)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.36   17.65   17.80   18.45   18.59   20.83
```

A fim de avaliar a distribuição dos dados obtidos, foi feita uma análise gráfica a partir do boxplot das subpopulações separadas pelo ano.

```
# boxplot
boxplot(imc_m_2016, imc_m_2017, imc_f_2016, imc_f_2017,
main = "Boxplots dos IMCs",
at = c(1,2,4,5),
names = c("2016(M)", "2017(M)", "2016(F)", "2017(F)"),
las = 2,
col = c("orange","orange", "red", "red"),
horizontal = TRUE,
notch = FALSE
)
```



Necessário uma análise mais detalhada do boxplot(diego)

Se quiserem fazer alguma análise aqui, acredito que pode ser algo como: "Através da análise gráfica é possível perceber alguns traços que serão confirmados por testes posteriormente. Nota-se que os dados das subpopulações femininas são bastante distintos de um ano para outro, inclusive o máximo do ano de 2017 é muito próximo do mínimo de 2016 o que sugere uma diferença grande entre as populações (aqui seria interessante se pudermos relacionar com algum indicador, como por exemplo, distribuição normal, mas nao sei se podemos). Em contra partida, as subpopulações masculinas apresentam dados mais semelhantes o que sugere pouca distinção entre elas (mesma coisa, seria interessante relacionar com algum indicador)." (Foi Elias que escreveu aqui e estou manjando menos que vocês, se acharem que está muito papagaiado pode editar/remover)

Em posse dos dados iniciais, o estudo será dividido em duas partes (masculino e feminino), devido às diferenças no processamento dos dados que ocorrerão.

1. Subpopulação masculina

a. Verificação de premissas para o teste de hipóteses

A intenção inicial é realizar um Teste T para a comparação destas amostras. Para isso, precisamos garantir que as amostras apresentam uma distribuição normal e que suas variâncias possam ser consideradas iguais (homocedasticidade).

Verificação de Normalidade: Para uma avaliação estatística sobre a validação da hipótese de uma distribuição normal para as amostras, pode-se usar o teste Shapiro-Wilk. Para este teste, temos:

$$\begin{cases} H_0 : \text{A amostra provém de uma população com distribuição normal} \\ H_1 : \text{A amostra não provém de uma população com distribuição normal} \end{cases}$$

```r
# teste de Shapiro-Wilk
shapiro.test(imc_m_2016)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_m_2016
## W = 0.92833, p-value = 0.1275
```

```r
shapiro.test(imc_m_2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_m_2017
## W = 0.96494, p-value = 0.6206
```

Como interpretação do teste, temos que se p-valor < 0.05 (α), deve-se rejeitar a hipótese nula, ou seja, os dados não possuem distribuição normal [3], caso contrário, não é possível concluir que os dados não seguem uma distribuição normal. Portanto, analisando os resultados dos testes dispostos acima, podemos considerar que as amostras masculinas seguem distribuição normal.

Verificação de homocesdaticidade: Para a análise da variância dos dados, pode-se usar o Teste F, onde tem-se:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

```r
# Teste F
var.test(imc_m_2016, imc_m_2017, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  imc_m_2016 and imc_m_2017
## F = 1.5839, num df = 20, denom df = 20, p-value = 0.3119
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
##  0.6426853 3.9034665
## sample estimates:
## ratio of variances
##           1.583888
```

Analisando o p-valor, temos que p-valor > 0,05 (α) e, portanto, não há evidência estatística forte o suficiente que indique que as variâncias não são iguais. Portanto, será considerada a homocedasticidade entre as subpopulações.

b. Tamanho de efeito e potência do teste

Tamanho de efeito: Embora o nível de significância tenha sido definido inicialmente, é interessante encontrar o tamanho de efeito, sendo este a medida da importância prática dos resultados de eventuais diferenças encontradas entre duas ou mais médias ou variâncias [4]. Existem várias maneiras de se fazer isto tais como: o Teste de Cohen, Teste de Glass, Teste de Hedges, Teste Psi, dentre outros [4]. O Teste de Cohen, por exemplo, foi desenhado para ser utilizado quando os escores das duas populações que estão sendo comparadas são contínuos e de distribuição normal[4]. Dada a distribução normal e a igualdade de variâncias assumidas de acordo com os testes de Shapiro-Wilk e F, respectivamente, pode-se usar o d de Cohen como estimativa do tamanho de efeito dos dados masculinos.

```
# d de Cohen
# install.packages("ggplot2") caso a biblioteca "effsize" não esteja instalada
library(effsize)
cohen.d(imc_m_2016, imc_m_2017)
```

```
##
## Cohen's d
##
## d estimate: 0.1665831 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.4582151  0.7913813
```

O tamanho de efeito retornado pela função foi de d = 0,1665.

Potencia do teste: Para se calcular a Potência do Teste, pode-se utilizar a função power.t.test. Nela deve ser inserido o parâmetro do número de observações por grupo (n = 21), delta (d de cohen = 0,1665831), desvio padrão conjugado (sd = 3,9046). O desvio padrão conjugado pode ser obtido pela Equação 2, onde s_1 e s_2 podem ser obtidos pela função sd.

```
# Desvio padrão
sd(imc_m_2016)
```

```
## [1] 4.323356
```

```
sd(imc_m_2017)
```

```
## [1] 3.435254
```

$$sd = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{2}$$

```r
# Potência do Teste
power.t.test(21, 0.1665831, 3.9046, sig.level = 0.05)
```

```
##
##      Two-sample t test power calculation
##
##               n = 21
##           delta = 0.1665831
##              sd = 3.9046
##       sig.level = 0.05
##           power = 0.03400086
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Observa-se que a potência de teste obtida foi de 3,4% que é muito menor que a potência desejada (80%). Isso significa que, para conseguirmos identificar diferenças de 0,1665831 nas médias padronizadas, com uma potência de 0,80, iremos precisar de um número maior de amostras (considerando o mesmo desvio padrão).

Mantendo o tamanho de efeito calculado, para que a Potência do Teste fosse a desejada, seriam necessárias 8626 amostras para cada ano, conforme descrito abaixo. Como não é possivel matricular mais alunos nas classes, a possibilidade de maior número de amostras é infactível.

```r
# Número de amostras para o Potência do Teste de 80%
power.t.test(NULL, 0.1665831, 3.9046, sig.level = 0.05, power = 0.80)
```

```
##
##      Two-sample t test power calculation
##
##               n = 8625.36
##           delta = 0.1665831
##              sd = 3.9046
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

c. Teste de hipóteses

Para a hipótese estatística proposta para as médias do IMC masculino, pode-se usar o Teste T, visto que a distribuição normal dos valores já foi validada, assim como a variância constante dos erros experimentais para observações distintas. Neste teste, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \mu 1 = \mu 2 \\ H_1 : \mu 1 \neq \mu 2 \end{cases}$$

```
# teste t de student
t.test(imc_m_2016, imc_m_2017, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  imc_m_2016 and imc_m_2017
## t = 0.53979, df = 40, p-value = 0.5923
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.784943  3.085836
## sample estimates:
## mean of x mean of y
##  24.93595  24.28551
```

Como pode ser visto, o teste t de student retornou um p-valor igual a 0,5923, cujo valor é maior que o nível de significâcia adotado (0,05). Portanto, com 95% de confiança não é possível rejeitar a hipótese nula do estudo proposto de que as médias das populações masculinas dos dois anos são iguais. Sendo a hipótese alternativa H1 bilateral, o intervalo de confiança para a diferença das médias é [-1.784943 3.085836].

Um outro teste pode ser feito para validação do resultado que é o Teste T de Welch. Percebe-se que ele também falhou em rejeitar a hipótese nula do estudo.

```
# teste t de Welch
t.test(imc_m_2016, imc_m_2017, "two.sided", mu=0, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  imc_m_2016 and imc_m_2017
## t = 0.53979, df = 38.057, p-value = 0.5925
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.788823  3.089716
## sample estimates:
## mean of x mean of y
##  24.93595  24.28551
```

2. Subpopulação feminina

a. Verificação de premissas para o teste de hipóteses

Assim como foi feito com a subpopulação masculina, a intenção inicial é realizar o Teste T para a comparação das amostras da subpopulação feminina. Para isso, precisamos garantir que as amostras apresentam uma distribuição normal e que suas variâncias possam ser consideradas iguais (homocedasticidade).

Verificação de Normalidade: Assim como na análise da subpopulação masculina, será usado o teste Shapiro-Wilk.

```
# teste de Shapiro-Wilk
shapiro.test(imc_f_2016)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_f_2016
## W = 0.91974, p-value = 0.4674
```

```r
shapiro.test(imc_f_2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_f_2017
## W = 0.7475, p-value = 0.03659
```

Como interpretação do teste, temos que se p-valor < 0.05 (α), deve-se rejeitar a hipótese nula, ou seja, os dados não possuem distribuição normal [3], caso contrário, não é possível concluir que os dados não seguem uma distribuição normal. Portanto, analisando os resultados dos testes dispostos acima, temos evidencia para inferir que as amostras de 2017 não seguem uma distribuição normal.

Portanto, não podemos seguir com o Teste T para comparação destas amostras. Será necessario utilizar um método não paramétrico de comparação de amostras. Para este trabalho, escolhemos o metodo de Wilcoxon.

b. Tamanho de efeito e potência do teste

Tamanho de efeito: Uma forma de se calcular o tamanho de efeito neste caso é utilizar a Medida A, de Vargha-Delaney. Esta é uma forma de se medir o tamanho de efeito sem levar em conta o tipo da distruibuição dos dados.

```r
# Tamanho de efeito
(vd_a <- VD.A(na.omit(imc_f_2016), na.omit(imc_f_2017)))
```

```
##
## Vargha and Delaney A
##
## A estimate: 0.8571429 (large)
```

```r
cat('Magnitude:', vd_a$magnitude)
```

```
## Magnitude: 4
```

A medida A pertence ao domínio $[0,\ 1]$ com as seguintes características:

- Se a medida $A = 0,5$, pode-se afirmar que não há diferenças evidentes na estatística que se estuda;
- Se a medida $A < 0,5$, a primeira população apresenta valores menores da estatística de interesse
- Se a medida $A > 0,5$, a segunda população apresenta valores menores da estatística de interesse.

Desta forma, pode-se concluir que, em 85% das vezes, alunas de 2/2016 possuem maior IMC mediano que as alunas do 2/2017, podendo encontrar diferenças maiores ou iguais a 4.

Abordagem do Potência do Teste

A menina me passou o que colocou nessa parte, mas acho que devemos encontrar algum teste para justificar isso

A ausência de normalidade da distribuição feminina do segundo semestre de 2017 impossibilita a aplicação da função `power.t.test` para análise do poder do teste, uma vez que esse cálculo de potência é específico para o teste t de Student. Nesse caso, o apropriado é calcular o poder do teste a partir de variações dessa função para o teste de Wilcoxon. No R existem algumas possibilidades dentro do pacote `wmwpow` [5], como o cálculo do poder usando a abordagem de Shieh [6] e outros dois baseados no método de Monte Carlo. Todavia, tais abordagens requerem que o tipo de distribuição seja passado por parâmetro, e portanto, conhecido previamente. Em vista disso e principalmente pelo baixo tamanho amostral que impossibilita a estimativa empírica da distribuição amostral, o cálculo do poder não pôde ser realizado. No entanto, dado que os tamanhos amostrais são distintos ($n_{F2016} \neq n_{F2017}$) e extremamente pequenos, é bem provável que a potência para efeitos maiores ou iguais a 4 kg/m^2 seja substancialmente menor que a desejada ($\pi = 0,80$). **deixando texto acima temporariamente como backup**

Potência do teste: A ausência de normalidade da distribuição feminina do segundo semestre de 2017 impossibilita a aplicação da função `power.t.test` para análise do poder do teste, uma vez que esse cálculo de potência é específico para o teste t de Student. Nesse caso, o apropriado é calcular o poder do teste a partir de uma variação dessa função para o teste de Wilcoxon. O que pode ser feito é

```
# potencia do teste de wilcoxon

# não consegui chegar em uma conclusão, olhei as referencias abaixo mas nao tive tempo de aprofundar ho

# https://www.rdocumentation.org/packages/MKpower/versions/0.5/topics/sim.power.wilcox.test ??
# rdocumentation.org/packages/litteR/versions/0.7.0/topics/power.wilcoxon ??
# https://stats.stackexchange.com/questions/15540/how-to-compute-the-power-of-wilcoxon-test??
# https://cran.r-project.org/web/packages/wmwpow/wmwpow.pdf??
```

c. Teste de hipóteses

Para a análise da proposta estatística da população feminina, será utilizado o teste não-paramétrico de Wilcoxon com as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As medianas das populações são iguais} \\ H_1 : \text{As medianas das populações não são iguais} \end{cases}$$

```
# teste de Wilcoxon
wilcox.test(imc_f_2016, imc_f_2017, alternative = "two.sided",conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test
##
## data:  imc_f_2016 and imc_f_2017
## W = 24, p-value = 0.07273
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.6374374  5.2284403
## sample estimates:
## difference in location
##               2.162763
```

Como pode ser visto, o teste de Wilcoxon retornou um p-valor igual a 0,07273, cujo valor é maior que o nível de significância adotado (0,05). Portanto, com 95% de confiança não é possível rejeitar a hipótese nula do estudo proposto de que as médias das populações femininas dos dois anos são iguais. Sendo a hipótese alternativa H1 bilateral, o intervalo de confiança para a diferença das médias é [-0, 6374374; 5, 22844].

Discussão e Conclusão

Lembrando que precisamos comentar ou incrementar o que foi comentado sobre:

- **Derivação de conclusões e recomendações.**

- **Discussão sobre a potência do teste (se aplicável).** Aqui para o masculino podemos retomar a discussão que o Diego fez na parte de potência do teste e enfatizar a necessidade de mais amostras.

- **Discussão sobre possíveis formas de melhorar este experimento.** Tomar cuidado para nao sermos repetitivos mas a forma de melhorar esse experimento é principalmente aumentar o numero de amostras, principalmente para as subpopulações fmeininas.

Atividades dos membros

...

Referências Bibliográficas

[1] Aline André Mesquita Holanda. Aprenda agora a calcular o imc e descubra seu peso ideal. http://www.unimedfortaleza.com.br/blog/cuidar-de-voce/como-calcular-imc. Acesso em 18 de Janeiro de 2021.

[2] Felipe Campelo. Lecture Notes on Design and Analysis of Experiments. http://git.io/v3Kh8, 2018. Version 2.12; Creative Commons BY-NC-SA 4.0.

[3] Como realizar teste de normalidade no r? https://rpubs.com/paternogbc/46768. Acesso em 18 de Janeiro de 2021.

[4] Juliana Dal-Ri Lindenau and Luciano Santos Pinto Guimarães. Calculando o tamanho de efeito no SPSS. *Revista HCPA*, 2012.

[5] Camden Bay Ilana Trumble, Orlando Ferrer. Package wmwpow. https://cran.r-project.org/web/packages/wmwpow/wmwpow.pdf. Acesso em 20 de Agosto de 2020.

[6] Gwowen Shieh, Show-li Jan, and Ronald H Randles. On Power and Sample Size Determinations for the Wilcoxon–Mann–Whitney test. *Journal of Nonparametric Statistics*, 18(1):33–43, 2006.