

# Estudo de caso: Grupo D 3

*Gilmar Pereira, Maressa Tavares e Victor Ruela*

*30 de Setembro, 2019*

## 1 Summary

O presente trabalho realizou o delineamento e executou os testes estatísticos para avaliar as diferenças no IMC médio entre duas populações de estudantes de pós-graduação em Engenharia elétrica, nos semestres de 2016-2 e 2017-2. As subpopulações masculina e feminina foram analisadas separadamente em todo o trabalho, devido às diferenças inerentes aos dois grupos. Após os testes verificou-se que as médias da subpopulação masculina são iguais, porém para a feminina elas são diferentes, para um intervalo de confiança de 95%.

## 2 Planejamento do experimento

Nesta seção é apresentado o planejamento do experimento, descrevendo os objetivos, determinando as hipóteses a serem testadas e realizando uma análise exploratória dos dados.

### 2.1 Objetivo do experimento

O objetivo deste experimento é estudar as diferenças entre o IMC médio das duas turmas - 2016-2 e 2017-2 - de estudantes de pós-graduação em Engenharia elétrica da UFMG. A análise foi dividida entre as subpopulações de homens e mulheres, visando comparar as subpopulações das duas amostras entre si. Essa separação deve-se às diferenças de IMC entre homens e mulheres, e também ao tamanho amostral.

Para se comparar os valores de IMC, formularam-se as seguintes hipóteses a serem testadas:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Onde  $\mu_1$  e  $\mu_2$  são as turmas de 2016-7 e 2017-2, respectivamente. A mesma hipótese foi testada separadamente para cada subpopulação.

### 2.2 Análise exploratória dos dados

Antes de iniciar a análise exploratória dos dados, foi preciso realizar um pré-processamento dos dados, visto que cada semestre possui formatos ligeiramente diferentes, como por exemplo, para os dados de 2016-2 foi necessário remover as linhas referentes aos alunos de graduação. Ambos os arquivos foram padronizados, para possuírem os mesmos nomes de colunas, e foi adicionada uma nova coluna com o cálculo do IMC. A partir dos dados filtrados e tratados foram criados novos datasets contendo somente os dados referentes a cada sexo e semestre, para auxiliar nas análises.

Realizando uma primeira análise, calculou-se os seguintes dados estatísticos separados por semestre.

```
summary(data.2016.2)
```

```
## Gender      imc
## F: 7   Min.   :17.58
## M:21   1st Qu.:21.27
##        Median :23.52
##        Mean   :23.97
##        3rd Qu.:25.29
##        Max.   :37.55
```

```
summary(data.2017.2)
```

```
## Gender      imc
## F: 4   Min.   :17.36
## M:21   1st Qu.:20.83
##        Median :23.26
##        Mean   :23.35
##        3rd Qu.:25.18
##        Max.   :30.42
```

Analisando os dados observa-se que a quantidade de amostras do sexo feminino é inferior à do masculino, o que pode afetar a potência dos testes a serem executados. Para melhor visualização, um gráfico boxplot foi gerado na figura abaixo.

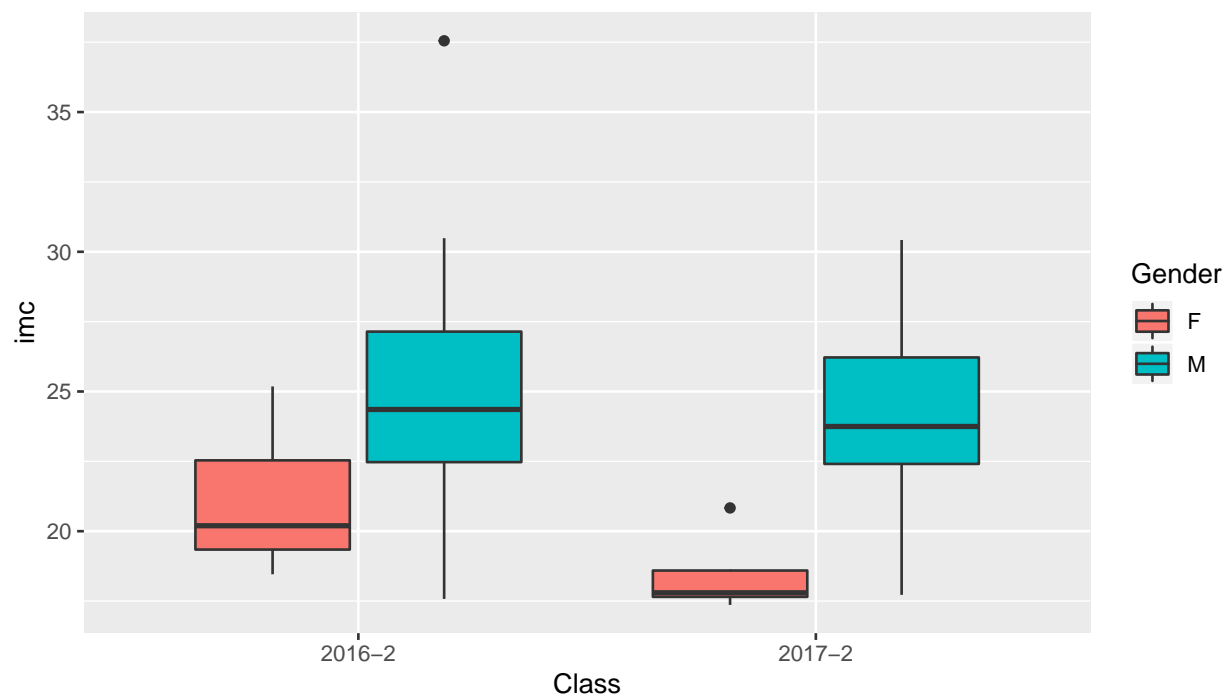


Figure 1: Boxplot dos dados

Analisando os boxplots verifica-se que existem possíveis outliers nos valores de IMC, para a subpopulação dos homens no semestre de 2016-2 e para a subpopulação das mulheres no semestre de 2017-2. Como os valores indicados como outliers estão dentro das faixas de IMC definidas em [1], conclui-se que de fato eles não são outliers e não devem ser removidos da análise.

Devido ao número pequeno de amostras para a população das mulheres, no semestre de 2017-2, o valor fora do padrão pode influenciar nos resultados dos testes. Desta maneira, será desenvolvido nas seções seguintes as análises preliminares das amostras, para constatação dos níveis de significância e potência para os respectivos testes a serem realizados com as populações.

### 2.2.1 Estimativa do tamanho de efeito e da significância

Para os testes de hipóteses, foram especificados um nível de significância  $\alpha = 0.05$  e uma potência de  $\pi = 1 - \beta = 0.8$ , suficientes para uma boa análise das amostras. Como o número de amostras é conhecido e inalterável para cada turma e gênero, utiliza-se a função `pwr.t2n.test` para a estimativa do tamanho de efeito ideal.

```
(sig.est.females <- pwr.t2n.test(n1 = n.females.2016.2,
                                n2= n.females.2017.2,
                                d = NULL,
                                sig.level = 0.05,
                                power = 0.8,
                                alternative = "two.sided"))
```

```
##
##      t test power calculation
##
##          n1 = 4
##          n2 = 7
##          d = 1.974137
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
```

```
(sig.est.males <- pwr.t2n.test(n1 = n.males.2016.2,
                               n2= n.males.2017.2,
                               d = NULL,
                               sig.level = 0.05,
                               power = 0.8,
                               alternative = "two.sided"))
```

```
##
##      t test power calculation
##
##          n1 = 21
##          n2 = 21
##          d = 0.88605
##      sig.level = 0.05
##          power = 0.8
##      alternative = two.sided
```

Portanto, é obtido um tamanho de efeito de 1.9741372 para os dados do gênero feminino e 0.88605 para o masculino. Vale ressaltar que, para uma mesma potência e significância, o tamanho de efeito é maior para subpopulação feminina, uma vez que o baixo número de amostras não permite identificar alterações mais sensíveis nos valores de IMC.

Os valores obtidos pela função `pwr.t2n.test` são normalizados pela variância ponderada, logo, o tamanho de efeito real da diferença das médias é dado por 3.4597032 e 3.7780088, sexo masculino e feminino, respectivamente.

### 3 Validação das premissas

Para realizar as inferências estatísticas sobre o IMC das duas populações foi necessário validar as premissas antes de executar os respectivos testes. Neste caso, como a variância das duas populações é desconhecida, utiliza-se a distribuição  $t$  para o teste de hipóteses e para os intervalos de confiança [2].

A seguir são apresentados os testes realizados para validar as premissas exigidas pelo teste  $t$ : análise da normalidade dos dados, homocedasticidade e independência. Para facilitar as análises optou-se por separar o grupo na população feminina e masculina.

#### 3.1 Subpopulação feminina

##### A - Normalidade:

Para avaliar a normalidade dos dados das duas subpopulações, utilizou-se o gráfico quantil-quantil e o teste de Shapiro Wilk com  $\alpha = 0.05$ . Os resultados são apresentados a seguir.

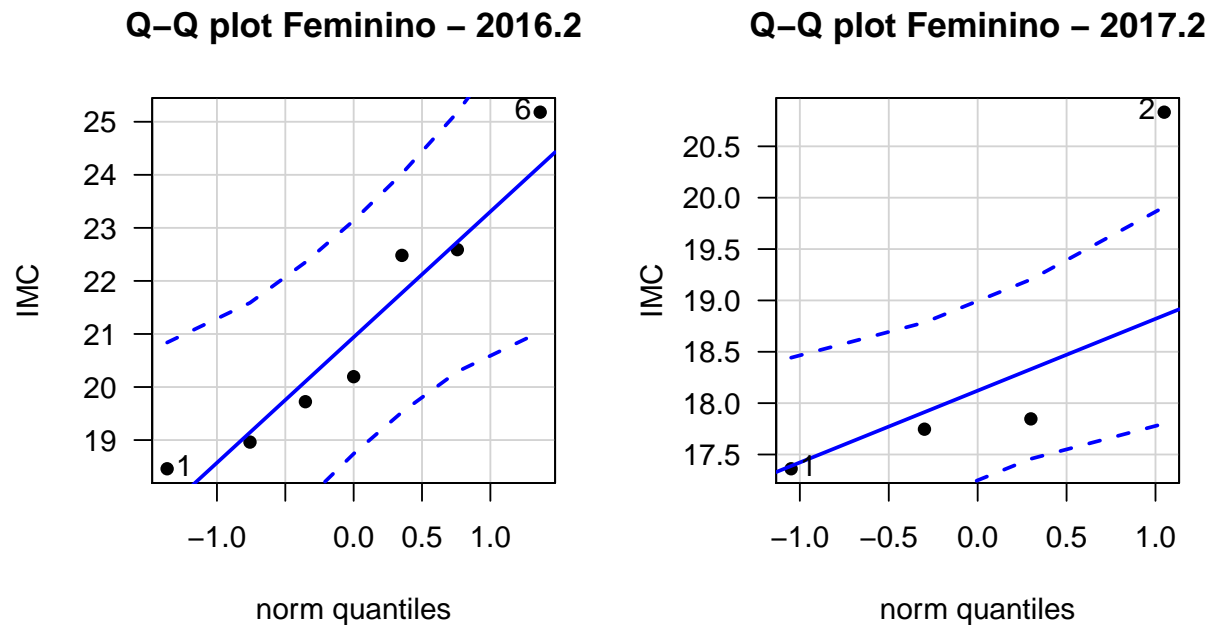


Figure 2: Gráfico quantil-quantil das populações - Feminino

```
shapiro.test(data.2016.2.Females$imc)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: data.2016.2.Females$imc  
## W = 0.91974, p-value = 0.4674
```

```
shapiro.test(data.2017.2.Females$imc)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  data.2017.2.Females$imc  
## W = 0.7475, p-value = 0.03659
```

Pela análise do teste de Shapiro-Wilk e do gráfico conclui-se que não há evidências para rejeitar a normalidade dos dados de 2016-2, pois o p-valor do teste de Shapiro Wilk foi maior que 0.05. Por outro lado, os dados de 2017-2 apresentam um p-valor inferior a 0.05, por isso, há evidências para rejeitar a normalidade dos dados.

Diante dos resultados, foi necessário fazer uma transformação logarítmica nos dados de 2017-2 visando normalizá-los, e possibilitar a realização dos testes estatísticos.

### Q-Q plot Feminino – 2017.2

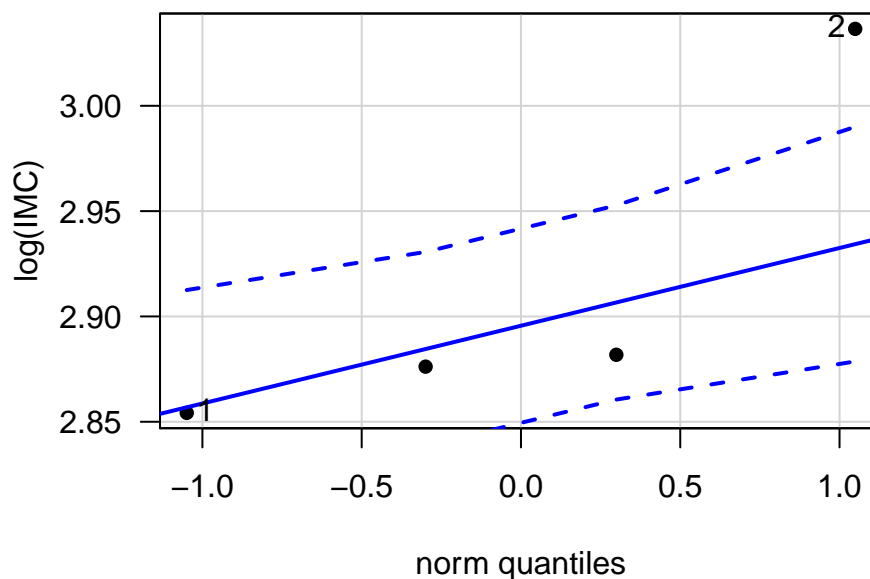


Figure 3: Gráfico quantil-quantil para transformação logarítmica dos dados

```
shapiro.test(log(data.2017.2.Females$imc))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  log(data.2017.2.Females$imc)  
## W = 0.75727, p-value = 0.04506
```

Pela análise do gráfico quantil-quantil e do teste de Shapiro Wilk após a transformação logarítmica, os dados de 2017.2 continuam sem apresentar uma distribuição normal, pois o p-valor permanece inferior a 0.05.

Como não foi possível obter uma distribuição normal após a transformação dos dados, torna-se necessário a realização de outras análises. Pela análise de boxplot na figura 1, verificamos que o valor extremo da amostra não pode ser considerado um outlier, porém, como a amostra da população é muito pequena esse valor está influenciando significativamente na normalidade dos dados. Por esse motivo, o valor extremo não pode ser excluído dos dados, pois isso caracterizaria “Cherry-Pick”.

Diante disso, como descrito na seção 4, optou-se por realizar o teste paramétrico e não-paramétrico (bootstrapping) nesse caso, e comparar os resultados, pois não foi possível validar as premissas.

### B- Igualdade de Variâncias:

A segunda premissa a ser avaliada é a igualdade de variâncias das duas populações, homocedasticidade. Para tanto, utilizou-se o teste F com a função `var.test` e  $\alpha = 0.05$ , considerando como hipótese nula a igualdade da variância de duas amostras.

```
var.test(data.2016.2.Females$imc,data.2017.2.Females$imc)

##
## F test to compare two variances
##
## data: data.2016.2.Females$imc and data.2017.2.Females$imc
## F = 2.2688, num df = 6, denom df = 3, p-value = 0.5353
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1539783 14.9715344
## sample estimates:
## ratio of variances
##          2.268827
```

Pelo teste F conclui-se que não há evidências para rejeitar a hipótese nula de igualdade das variâncias das duas populações, portanto, as amostras são consideradas homocedásticas.

### C - Independência:

Considerando que as populações referem-se a dois grupos distintos da pós-graduação, sabe-se a priori, que as amostras são independentes. Porém, para concluir com precisão em relação à independência realizou-se o teste Qui-quadrado com  $\alpha = 0.05$ .

```
data.chi.Female <- c(data.2016.2.Females$imc, data.2017.2.Females$imc)
chisq.test(data.chi.Female)

##
## Chi-squared test for given probabilities
##
## data: data.chi.Female
## X-squared = 3.0047, df = 10, p-value = 0.9813
```

Como era esperado, o teste Qui-quadrado reafirma a independência entre as duas amostras, como o p-valor muito alto (0.9813).

Após a análise das premissas para o grupo de mulheres percebe-se que, provavelmente, o tamanho pequeno das amostras de IMC feminino pode ter influenciado nas conclusões em relação às premissas. Portanto, esse fator pode influenciar também nos testes de hipóteses que serão apresentados nas próximas seções.

### 3.2 Subpopulação masculina

Assim como na validação das premissas da subpopulação feminina, foram realizados os testes para a subpopulação masculina, como descrito a seguir.

#### A - Normalidade:

Para validar a normalidade dos dados utilizou-se o gráfico quantil-quantil e o teste de Shapiro Wilk com  $\alpha = 0.05$ .

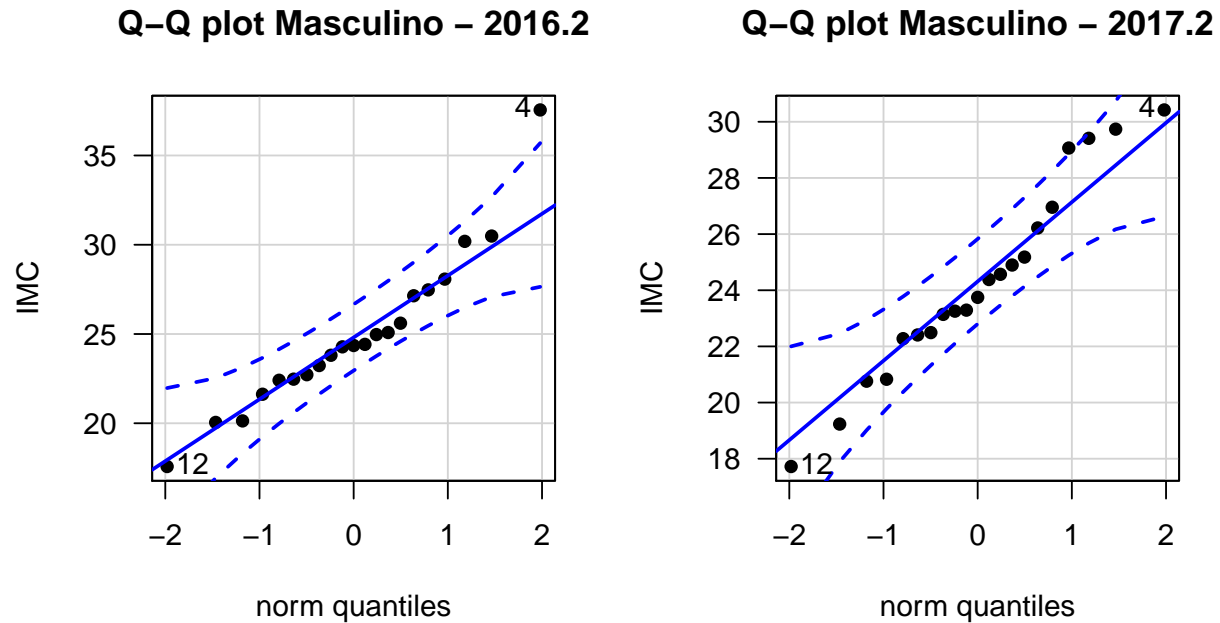


Figure 4: Gráfico quantil-quantil das populações - Feminino

```
shapiro.test(data.2016.2.Males$imc)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: data.2016.2.Males$imc  
## W = 0.92833, p-value = 0.1275
```

```
shapiro.test(data.2017.2.Males$imc)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: data.2017.2.Males$imc  
## W = 0.96494, p-value = 0.6206
```

Pela análise do teste de Shapiro-Wilk e dos gráficos conclui-se que não há evidências para rejeitar a hipótese nula para as duas amostras, pois em ambas o p-valor do teste foi superior a 0.05. Portanto, para a subpopulação masculina dos dois semestres os dados estão normalmente distribuídos.

### B - Igualdade de Variâncias:

Para validar a igualdade de variâncias das duas amostras utilizou-se o teste F com  $\alpha = 0.05$ . A hipótese nula considera a igualdade da variância das duas amostras.

```
var.test(data.2016.2.Males$imc,data.2017.2.Males$imc)

##
## F test to compare two variances
##
## data: data.2016.2.Males$imc and data.2017.2.Males$imc
## 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
```

Pelo teste F conclui-se que as populações são homocedásticas, pois o p-valor foi superior a 0.05, logo, não há evidências para rejeitar a hipótese nula.

### C - Independência:

Assim como na subpopulação feminina, sabe-se a priori, que as amostras são independentes. Para concluir com precisão em relação à independência realizou-se o teste Qui-quadrado com  $\alpha = 0.05$ .

```
data.chi.male <- c(data.2016.2.Males$imc,data.2017.2.Males$imc)
chisq.test(data.chi.male)

##
## Chi-squared test for given probabilities
##
## data: data.chi.male
## X-squared = 24.96, df = 41, p-value = 0.9772
```

Como era esperado, o teste Qui-quadrado reafirma a independência entre as duas amostras, com o p-valor = 0.9772.

Ao contrário do que aconteceu com as amostras da subpopulação feminina, nesse caso o tamanho das amostras foram maiores, e por isso, não houve nenhum problema em relação às premissas do teste.

## 4 Resultados

Nesta seção são apresentados os resultados do teste para as hipóteses estabelecidas anteriormente com  $\alpha = 0.05$ . Os resultados estão subdivididos entre a população feminina e masculina devido às diferenças de IMC que existem entre eles.



## 4.1 Teste de hipóteses subpopulação feminina

Embora as premissas do teste paramétrico não tenham sido validadas completamente, foi realizado o teste t para verificar a igualdade das duas médias amostrais, como apresentado a seguir.

```
#teste para custo médio:
(mean.t.teste <- t.test(data.2016.2.Females,
                        data.2017.2.Females,
                        var.equal = TRUE, conf.level = 1 - sig_level))

##
## Two Sample t-test
##
## data: data.2016.2.Females and data.2017.2.Females
## t = 1.9308, df = 9, p-value = 0.08556
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.4527037  5.7283762
## sample estimates:
## mean of x mean of y
## 21.08443 18.44660
```

Analisando os resultados, verifica-se que o p-valor retornado pelo teste é superior ao nível de significância, desta maneira não se possui evidências suficientes para a rejeição da hipótese nula. Além disso, verifica-se que a diferença entre as médias,  $d = 2.6378362$ , está contido no intervalo de confiança calculado.

Devido ao baixo número de amostras da subpopulação e à não validação das premissas, realizou-se também o teste não paramétrico. Para tanto, realizou-se uma remostragem utilizando-se a técnica de bootstrapping como forma de obter uma melhor aproximação do intervalo de confiança, conforme mostrado a seguir. teste foi realizado tanto para a média como para mediana.

```
means.diff <- function(x, i){
  sample.2016.2 <- sample((x %>% filter(Class == '2016-2'))$imc,replace=TRUE)
  sample.2017.2 <- sample((x %>% filter(Class == '2017-2'))$imc,replace=TRUE)

  return((mean(sample.2016.2) - mean(sample.2017.2)))
}

median.diff <- function(x, i){
  sample.2016.2 <- sample((x %>% filter(Class == '2016-2'))$imc,replace=TRUE)
  sample.2017.2 <- sample((x %>% filter(Class == '2017-2'))$imc,replace=TRUE)

  return((median(sample.2016.2) - median(sample.2017.2)))
}

set.seed(5021)
data.diff.means.boot <- boot(data.all.Females, statistic = means.diff, R=1000)
data.diff.median.boot <- boot(data.all.Females, statistic = median.diff, R=1000)

par(mfrow=c(1,2))
hist(data.diff.means.boot$t)
hist(data.diff.median.boot$t)
```

## Histogram of data.diff.means.boot\$      Histogram of data.diff.median.boot\$

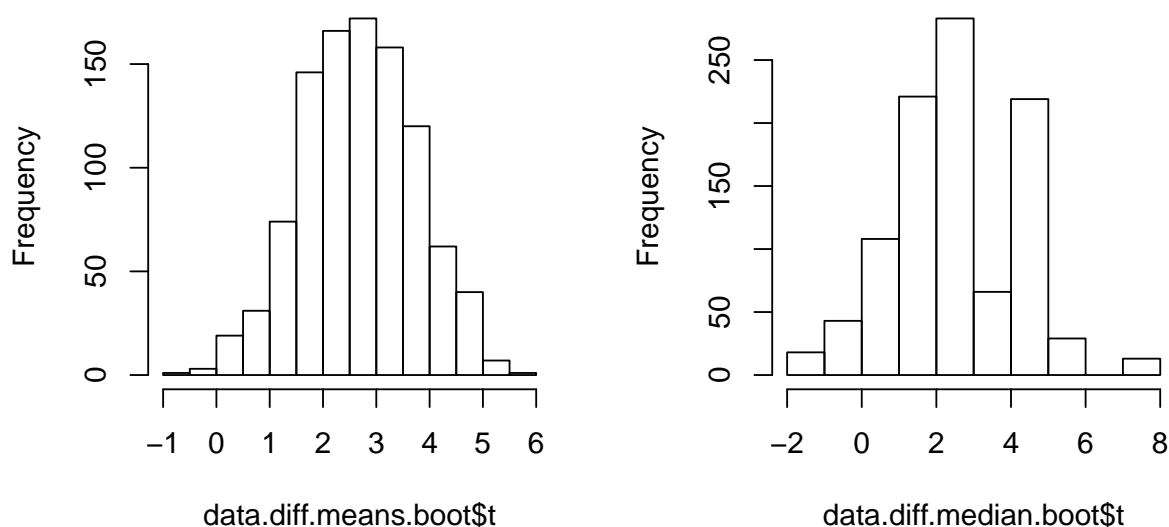


Figure 5: Histograma do bootstrapping da média e mediana da subpopulação feminina

```
ci_sd <- 1 - 2 * sig_level
(test.boot.var <- boot.ci(data.diff.means.boot, conf = ci_sd, type = "bca"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = data.diff.means.boot, conf = ci_sd, type = "bca")
##
## Intervals :
## Level      BCa
## 90%      ( 1.233,  4.727 )
## Calculations and Intervals on Original Scale
```

```
(test.boot.var <- boot.ci(data.diff.median.boot, conf = ci_sd, type = "bca"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = data.diff.median.boot, conf = ci_sd, type = "bca")
##
## Intervals :
## Level      BCa
## 90%      (-1.109,  4.739 )
## Calculations and Intervals on Original Scale
```

Analisando o histograma gerado, pode-se observar a normalização dos dados para subpopulação após a reamostragem, justificado pelo teorema do limite central para médias amostrais. Analisando o intervalo de confiança resultante verifica-se que, dado que a hipótese nula assume que as médias populacionais são iguais ( $\mu_1 - \mu_2 = 0$ ), a hipótese nula é rejeitada para o teste da média, pois está fora do intervalo de confiança, porém, ela não pode ser rejeitada para o teste da mediana.

## 4.2 Teste de hipóteses subpopulação masculina

Assim como para a subpopulação feminina, utilizou-se o teste t para análise dos dados. O teste está especificado abaixo.

```
#teste para custo médio:
(mean.t.teste.male <- t.test(data.2016.2.Males,
                             data.2017.2.Males,
                             var.equal = TRUE, conf.level = 1 - sig_level))

##
## Two Sample t-test
##
## data: data.2016.2.Males and data.2017.2.Males
## 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
```

O teste t para as subpopulações masculinas resultou em um p-valor superior ao nível de significância de 0.05, como especificado. Desta forma não se possui evidências suficientes para rejeitar a hipótese nula de igualdade das médias. Verifica-se também que o valor das diferenças entre as médias está dentro do valor especificado.

Para melhor avaliar os resultados, a técnica de bootstrapping também foi aplicada para a subpopulação masculina.

```
set.seed(5021)
data.diff.means.boot.males <- boot(data.all.Males, statistic = means.diff, R=1000)
data.diff.median.boot.males <- boot(data.all.Males, statistic = median.diff, R=1000)

par(mfrow=c(1,2))
hist(data.diff.means.boot.males$t)
hist(data.diff.median.boot.males$t)
```

## Histogram of data.diff.means.boot.maHistogram of data.diff.median.boot.ma

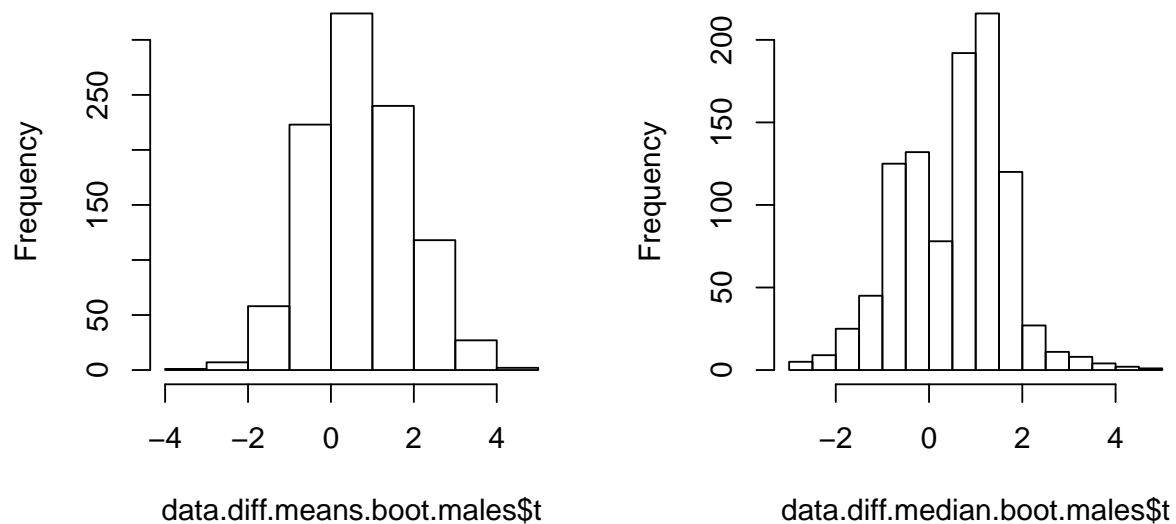


Figure 6: Histograma do bootstrapping da média e mediana da subpopulação masculina

```
ci_sd <- 1 - 2 * sig_level
(test.boot.mean.males <- boot.ci(data.diff.means.boot.males, conf = ci_sd, type = "bca"))

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = data.diff.means.boot.males, conf = ci_sd,
##         type = "bca")
##
## Intervals :
## Level      BCa
## 90%      (-0.287,  3.656 )
## Calculations and Intervals on Original Scale
## Some BCa intervals may be unstable

(test.boot.median.males <- boot.ci(data.diff.median.boot.males, conf = ci_sd, type = "bca"))

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = data.diff.median.boot.males, conf = ci_sd,
##         type = "bca")
##
## Intervals :
```

```
## Level          BCa
## 90%    (-2.3313,  1.6768 )
## Calculations and Intervals on Original Scale
```

Note que, neste caso, tanto para a média quanto para a mediana, não há evidências para rejeitar a hipótese nula. Isso reforça ainda mais os resultados obtidos para o teste t.

## 5 Discussão e Conclusões

Neste trabalho foi feito um estudo estatístico da diferença das médias de IMC entre duas turmas de alunos de pós-graduação. Os dados foram divididos entre uma subpopulação masculina e feminina, as quais foram analisadas separadamente. Para a masculina, todas as premissas são atendidas, logo o teste t foi utilizado e o p-valor de 0.5923338 obtido comprova que não é possível rejeitar a hipótese nula, ou seja, as médias não são diferentes. Já para a subpopulação feminina, a premissa de normalidade não é respeitada para os dados de 2017-2, o que fez com que fosse necessário utilizar técnicas não paramétricas para estimativa do intervalo de confiança. A técnica de bootstrapping foi aplicada à média das diferenças de IMC entre as turmas, e obteve-se que a hipótese nula é rejeitada para o teste da média. Além disso, a técnica paramétrica também leva a concluir que as médias são diferentes para as mulheres.

Como os dados da subpopulação feminina de 2017-2 possui uma instância fora do padrão das demais, embora não seja um outlier, executou-se também o bootstrapping para a mediana das diferenças. Como a mediana é mais robusta à outliers, conclui-se que para este estimador a hipótese nula não pode ser rejeitada, de acordo com o intervalo de confiança resultante. Como forma de comprovar os resultados obtidos com o teste t, o bootstrapping também foi executado para a subpopulação masculina. Os intervalos de confiança obtidos mostram que, tanto para a média quanto para a mediana, a hipótese nula não pode ser rejeitada, reforçando os resultados do teste t.

## 6 Divisão das Atividades

- Victor - Verificador e Monitor
- Maressa - Relatora
- Gilmar - Coordenador

## Referências

- [1] *Tabela de imc*. <http://www.calculoimc.com.br/tabela-de-imc/>.
- [2] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers, (with cd)*. John Wiley & Sons, 2007.