Estudo 1

Jonatan Almeida and Helbert Paulino

2023-09-27

Resumo

Este estudo de caso é uma comparação de coletados dados de alunos da UFMG nos semestres de 2016/2 e 2017/2. Os dados são compostos de:

- altura
- idade
- sexo
- peso
- curso (PPGEE ou ENGSIS) - aplica-se apenas para 2016/2

A pergunta de interesse é a seguinte:

> Existe alteração no estilo de vida entre os alunos do PPGEE de um semestre para outro?

Para isso, um dos estimadores pontuais que podem ser utilizados para responder a essa pergunta é o IMC (Índice de Massa Corporal), cuja relação matemática é dada por:

$$IMC = \frac{peso}{altura^2}$$

Tendo em vista que há a possibilidade de haver diferenças nos valores médios do IMC para homens e mulheres, a analise será feita por subgrupos, masculino e feminino.

Design experimental

A pergunta de interesse nos leva a definir os seguintes testes de hipoteses:

$$\begin{cases} H_0 : \mu_{2016} = \mu_{2017} \\ H_1 : \mu_{2016} \neq \mu_{2017} \end{cases}$$

Onde o parametro μ sigfica o IMC médio de cada turma. A hipotese H_0 significa que não houve altereção no estilo de vida entre os alunos e a hipotese H_0 significa que houve alteração, ou seja, as médias de IMC são diferentes entre os alunos.

Para o IMC, existe as seguintes classificações:

- **IMC** $<$ 18,5kg/mm^2 - baixo peso
- **IMC** $>$ 18,5 até 24,9kg/mm^2 - eutrofia (peso adequado)
- **IMC** $\geq$ 25 até 29,9kg/mm^2 - sobrepeso
- **IMC** $>$ 30,0kg/m2 até 34,9kg/mm^2 - obesidade grau 1
- **IMC** $>$ 35kg/m2 até 39,9kg/mm^2 - obesidade grau 2
- **IMC** $>$ 40kg/m2 - obesidade extrema

Nota-se que a alteração é sempre de 5 em 5 kg/m^2. Logo um valor interessante para o efeito minimo relevante (δ^*) é uma alteração de 5 entre as médias ou uma alteração na classificação da média do IMC da turma.

Para o teste estatistico será divido em duas análises, uma para o sexo masculino e uma para o sexo feminino. Então serão dois teste de hipoteses distintos, um para cada sexo.

Como a variância da população não é conhecida, utilizaremos o teste t com um $\alpha = 0{,}5$.

Description of the data collection

TBA

Análise exploratória dos dados

Carregando os dados:

```
data2016 = read.csv('https://raw.githubusercontent.com/fcampelo/Design-and-Analysis-of-Experiments/maste
data2017 = read.csv('https://raw.githubusercontent.com/fcampelo/Design-and-Analysis-of-Experiments/maste
```

Já foi mencionado no *Resumo* que existem dados de alunos de graduação (ENGSIS) nos dados de 2016. O primeiro passo é expurgar estes dados para não contaminarem nossa amostra.

```
ppgeeStudents1 = subset(data2016, Course=='PPGEE')
```

Além disso, é de grande importância que os dados dos alunos sejam separados por ano e por sexo. Dessa forma, a separação em masculino e feminino se deu por:

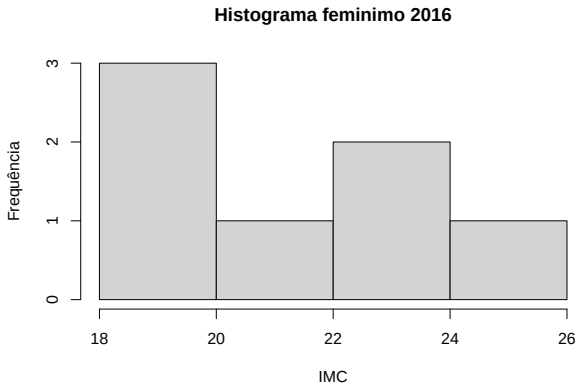
```
female2016 = subset(ppgeeStudents1, Gender=='F')
male2016 = subset(ppgeeStudents1, Gender=='M')
female2017 = subset(data2017, Sex=='F')
male2017 = subset(data2017, Sex=='M')
```

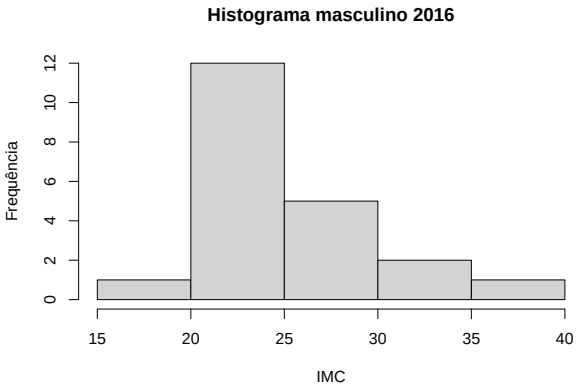
O parametro de interesse é o IMC, cujo valor não está explícito nos dados. Porém, tendo em vista sua conhecida fórmula, foram combinados os valores da massa corporal e da altura dos alunos para calcular o IMC e este foi inserido na tabela de dados original.

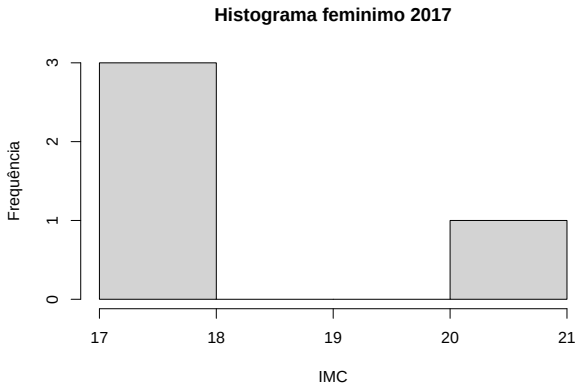
```
female2016$imc <- (female2016$Weight.kg / (female2016$Height.m*female2016$Height.m))
male2016$imc <- (male2016$Weight.kg / (male2016$Height.m*male2016$Height.m))
female2017$imc <- (female2017$Weight.kg / (female2017$height.m*female2017$height.m))
male2017$imc <- (male2017$Weight.kg / (male2017$height.m*male2017$height.m))
```

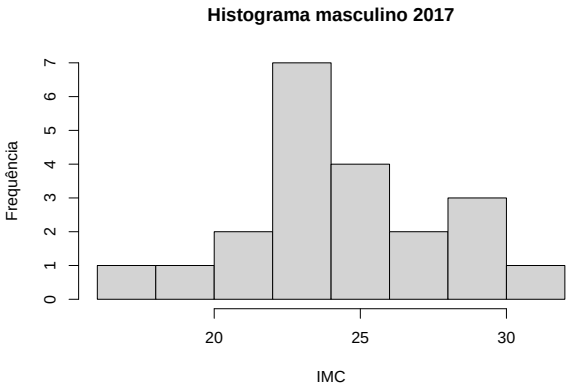
Plotar a distribuição dos dados é uma boa forma de entender qual o padrão dos dados para definir os testes a serem aplicados. Segue os histogramas dos dados de interesse.

```
hist(female2016$imc, xlab = 'IMC', ylab = 'Frequência', main = "Histograma feminimo 2016")
hist(male2016$imc, xlab = 'IMC', ylab = 'Frequência', main = "Histograma masculino 2016")
hist(female2017$imc, xlab = 'IMC', ylab = 'Frequência', main = "Histograma feminimo 2017")
hist(male2017$imc, xlab = 'IMC', ylab = 'Frequência', main = "Histograma masculino 2017")
```









Nota-se que claramente os dados masculinos de 2016 e 2016 tendem a seguir uma distrubuição Normal, onde há indicios que o teste proposto anteriormente (t test) é adequado. O QQ plot, plot de quantis, é um bom gráfico para entender a distribuição dos dados, segue abaixo:

```
library(car)
```

```
## Carregando pacotes exigidos: carData
```

```
qqPlot(female2016$imc)
```

```
## [1] 6 1
```

```
qqPlot(male2016$imc)
```

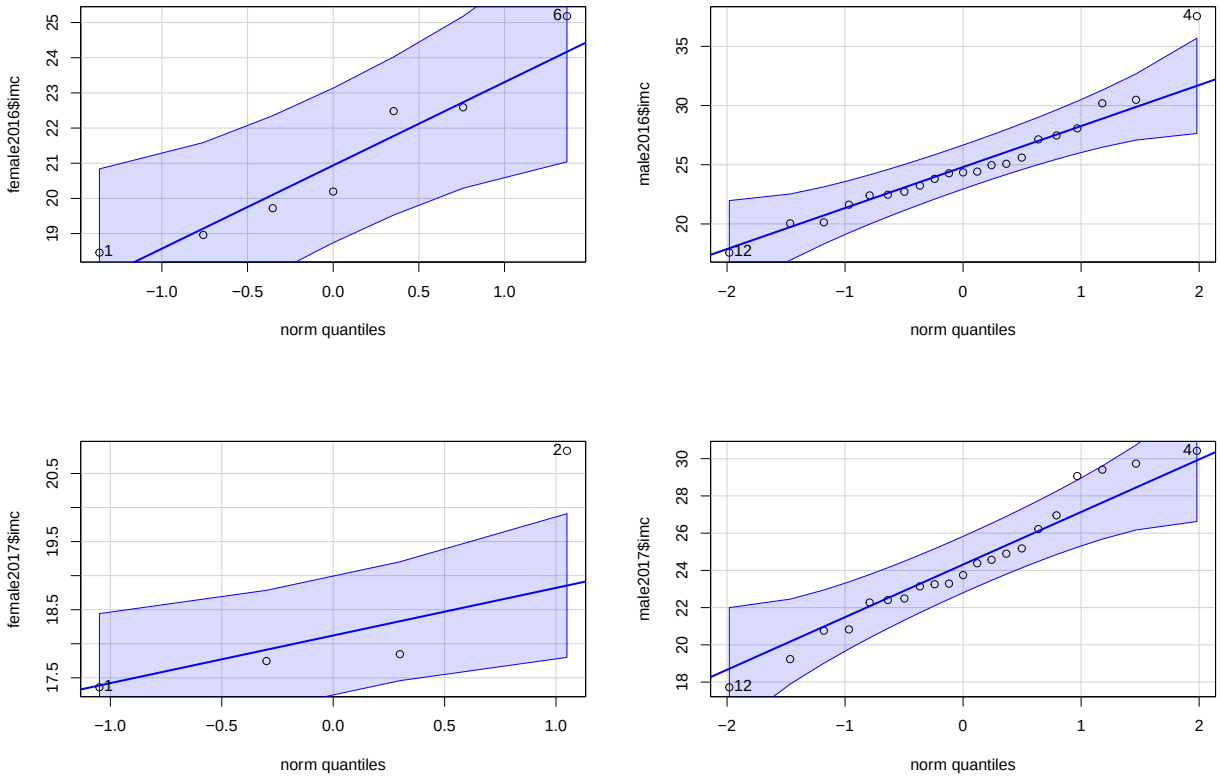
```
## [1]  4 12
```

```
qqPlot(female2017$imc)
```

```
## [1] 2 1
```

```
qqPlot(male2017$imc)
```

```
## [1] 12  4
```



Em relação a interpretação do QQ Plot, caso os pontos se concentrem em torno de uma reta, existe indicios que é uma distrubuição Normal. Neste caso temos mais um indicio que os dados masculinos seguem uma Normal e podemos notar que os dados femininos também se concetral em torno de uma reta, indicando indicios de uma Normal.

Além das analises gráficas, pode-se obter alguns estimadores pontuais e isso foi feito da seguinte forma:

Cálculo do IMC médio

```
meanFemIMC2016 = mean(female2016$imc)
meanMaleIMC2016 = mean(male2016$imc)
meanFemIMC2017 = mean(female2017$imc)
meanMaleIMC2017 = mean(male2017$imc)
```

Cálculo do desvio padrão

```
sdFemIMC2016 = sd(female2016$imc)
sdMaleIMC2016 = sd(male2016$imc)
sdFemIMC2017 = sd(female2017$imc)
sdMaleIMC2017 = sd(male2017$imc)
```

Análise Estatística

Os dados obtidos para os alunos variam em tamanho da amostra, sendo $N < 30$ e cuja variância é desconhecida. Dessa forma, o teste t é o indicado para a análise estatística. Para a avaliação dos dados, definimos os seguintes parâmetros:

- $\alpha = 0{,}5$
- $\delta^* = 5$ kg/m^2

Além disso, tendo em vista que queremos avaliar se houve mudanças no IMC médio da turma, realizamos o teste bilateral, com intervalo de confiança 1 - $\alpha = 0.95$. Dessa forma, obtivemos os seguintes resultados:

Comparando homens entre 2016 e 2017

```
t.test(male2017$imc, alternative="two.sided", mu=meanMaleIMC2016, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  male2017$imc
## t = -0.86769, df = 20, p-value = 0.3959
## alternative hypothesis: true mean is not equal to 24.93595
## 95 percent confidence interval:
##  22.72180 25.84921
## sample estimates:
## mean of x
##  24.28551
```

Como se pode perceber, o valor médio do IMC dos homens de 2017 está dentro de um intervalo de confiança ($22.72180 < \mu$=24.93595 < 25.84921$) esperado, quando se comparado à média dos homens de 2016. Isso também fica explícito pelo valor de p (0.3959), que é significativamente maior que o índice de significância.

Comparando mulheres entre 2016 e 2017

```
t.test(female2017$imc, alternative="two.sided", mu=meanFemIMC2016, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  female2017$imc
## t = -3.2884, df = 3, p-value = 0.04613
## alternative hypothesis: true mean is not equal to 21.08443
## 95 percent confidence interval:
##  15.89376 20.99943
## sample estimates:
## mean of x
##   18.4466
```

Diferentemente do caso dos homens, o valor médio do IMC das mulheres de 2017 está fora do um intervalo de confiança $[15.89376, 20.99943] < \mu = 21.08443$ esperado, quando se comparado à média das mulheres de 2016. Isso também fica explícito pelo valor de p (0.04613), que é menor que o índice de significância escolhido. Deve-se, no entanto, levar em consideração que entre esses dois grupos há uma diferença no tamanho da amostra, sendo que em 2016 tinhamos 7 mulheres e em 2017 tinhamos 4, uma amostra que possui tamanho pequeno, causando impactos na análise.

Além da análise do IMC, tendo em vista que as mudanças no estilo de vida tem uma probabilidade maior de afetar o peso corporal das pessoas do que em suas alturas, avaliamos, também, o peso dos alunos. Dessa forma, temos:

(INSERINDO)

Checking Model Assumptions

The assumptions of your test should also be validated, and possible effects of violations should also be explored.

```
#par(mfrow=c(2,2), mai=.3*c(1,1,1,1))
#plot(model,pch=16,lty=1,lwd=2)
```

Conclusions and Recommendations

The discussion of your results, and the scientific/technical meaning of the effects detected, should be placed here. Always be sure to tie your results back to the original question of interest!