

Estudo de Caso 2: Planejamento e Análise de Experimentos

Matheus Marzochi, Mayra Mota, Rafael Ramos e Victor Magalhães

30 de Setembro de 2019

Resumo

O experimento disponibilizado para este trabalho consiste em avaliar se o estilo de vida entre duas populações de estudantes se alterou ao longo do tempo. Nessa avaliação duas amostras foram utilizadas, uma contendo informações (peso e altura dos alunos) referentes ao ano de 2016 e a outra com informações referentes a 2017.

Neste trabalho o IMC (Índice de Massa Corporal) é usado como parâmetro de avaliação do estilo de vida. Vale ressaltar que esse indicador possui limitações como pode ser verificado em [1][2]. Além disso, as duas populações foram subdivididas por gênero. A motivação desta divisão é a consideração de que, em média, o IMC masculino pode ser diferente do feminino.

Papéis Desempenhados

A divisão de tarefas no grupo segue a descrição da *Declaração de Políticas de Equipe*. Estando aqui organizada da seguinte forma:

- Matheus: Verificador
- Mayra: Monitora
- Rafael: Coordenador
- Victor: Revisor

Planejamento do Experimento

A hipótese nula (H_0) usada neste experimento é de que a diferença das médias de IMC das duas populações é nula. A hipótese alternativa (H_1), por sua vez, afirma que existe sim uma alteração entre as médias dos dois semestres.

$$\begin{cases} H_0 : \mu_{2016/2} - \mu_{2017/2} = 0 \\ H_1 : \mu_{2016/2} - \mu_{2017/2} \neq 0 \end{cases}$$

Como a comparação é relativa a duas amostras, em função da influência do gênero no valor do IMC, as análises são realizadas de forma independente para cada sexo. Como consequência dessa independência, neste trabalho são realizados testes de comparação simples entre as populações.

Coleta dos Dados

O procedimento de coleta de dados foi baseado na rotina presente abaixo.

```
data_2017 <- read.csv(file='CS01_20172.csv', sep=';') #Dados ano 2016
data_2016<-read.csv("imc_20162.csv"); #Dados ano 2017
```

```

PPGEE_dados=data_2016[data_2016[2]=='PPGEE',];

#Dados população feminina e masculina do ano de 2016.
Dados_Masculino_2016=PPGEE_dados[PPGEE_dados[3]=='M',];
Dados_Feminino_2016=PPGEE_dados[PPGEE_dados[3]=='F',];

Height_Masc_2016=Dados_Masculino_2016[,4];
Heigh_Feminino_2016=Dados_Feminino_2016[,4];
Weight_Masculino_2016=Dados_Masculino_2016[,5];
Weight_Feminino_2016=Dados_Feminino_2016[,5];

#Calculo do IMC masculino 2016
IMC_masculino_2016=(Weight_Masculino_2016/((Height_Masc_2016)*(Height_Masc_2016)));
#Calculo do IMC feminino 2016
IMC_Feminino_2016=(Weight_Feminino_2016/((Heigh_Feminino_2016)*(Heigh_Feminino_2016)));

#Dados população feminina e masculina do ano de 2017.
Dados_Masculino_2017=data_2017[data_2017[3]=='M',];
Dados_Feminino_2017=data_2017[data_2017[3]=='F',];

Height_Masculino_2017=Dados_Masculino_2017[,2];
Height_Feminino_2017=Dados_Feminino_2017[,2];
Weight_Masc_2017=Dados_Masculino_2017[,1];
Weight_Feminino_2017=Dados_Feminino_2017[,1];

#Calculo do IMC masculino 2017
IMC_masculino_2017=(Weight_Masc_2017/((Height_Masculino_2017)*(Height_Masculino_2017)));
#Calculo do IMC feminino 2017
IMC_Feminino_2017=(Weight_Feminino_2017/((Height_Feminino_2017)*(Height_Feminino_2017)));

```

Análise Exploratória dos Dados

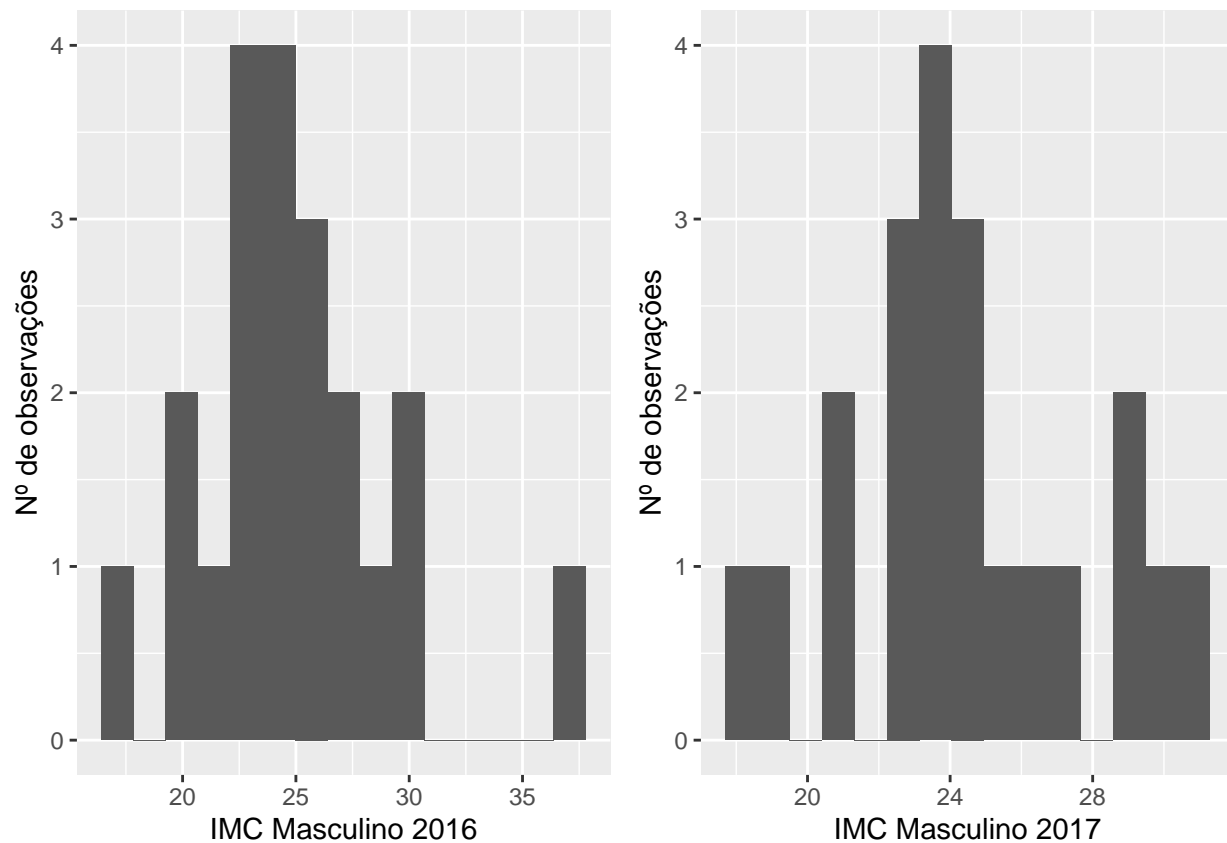
Antes da análise estatística, os dados são avaliados de forma qualitativa na análise exploratória com o objetivo de se extrair informações úteis. Entre as ferramentas usadas para se explorar os dados existentes está o Histograma.

Segue abaixo o trecho de código usado para gerar o histograma da população masculina dos anos de 2016 e 2017.

```

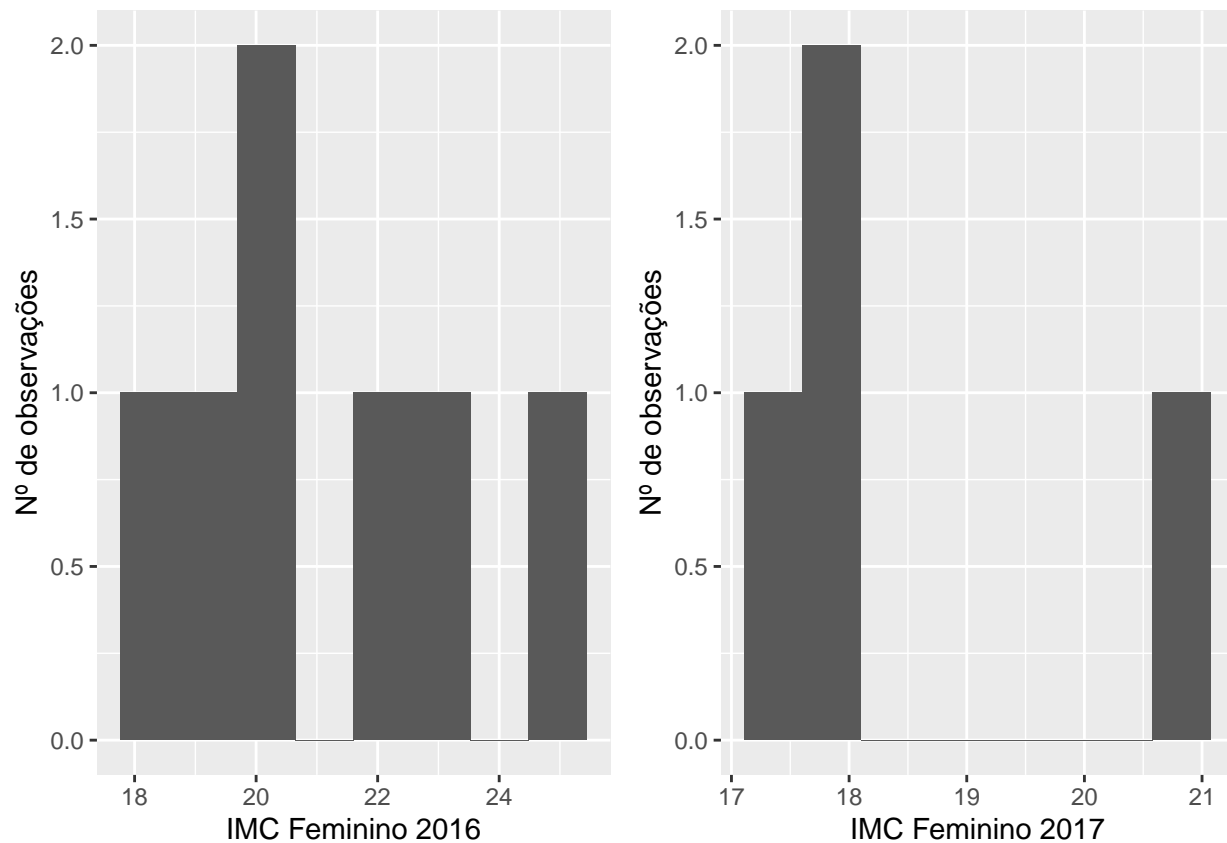
p1 <- ggplot(as.data.frame(IMC_masculino_2016), aes(x = IMC_masculino_2016))
p1 <- p1 +geom_histogram(bins = 15)+xlab("IMC Masculino 2016")+ylab("Nº de observações")
p2 <- ggplot(as.data.frame(IMC_masculino_2017), aes(x = IMC_masculino_2017))
p2 <- p2 + geom_histogram(bins = 15)+xlab("IMC Masculino 2017")+ylab("Nº de observações")
ggarrange(p1, p2, nrow = 1, ncol = 2)

```



Segue abaixo o trecho de código usado para gerar o histograma da população feminina dos anos de 2016 e 2017.

```
p1 <- ggplot(as.data.frame(IMC_Feminino_2016), aes(x = IMC_Feminino_2016))
p1 <- p1 + geom_histogram(bins = 8) + xlab("IMC Feminino 2016") + ylab("Nº de observações")
p2 <- ggplot(as.data.frame(IMC_Feminino_2017), aes(x = IMC_Feminino_2017))
p2 <- p2 + geom_histogram(bins = 8) + xlab("IMC Feminino 2017") + ylab("Nº de observações")
ggarrange(p1, p2, nrow = 1, ncol = 2)
```



Através da análise exploratória dos dados é possível observar que o número de amostras fornecidos é pequeno, especialmente no caso feminino. Para a população masculina apesar do número de amostras não ser elevado, a distribuição das amostras leva há indícios que a distribuição pode ser normal. No caso feminino há indícios de que seja necessário o uso de estatística não paramétrica. Essas hipótes são analisadas mais profundamente na análise estatística.

Análise estatística

Conclusões

Bibliografia

- [1] <https://www.nytimes.com/interactive/projects/cp/summer-of-science-2015/latest/how-often-is-bmi-misleading>
- [2] <https://science.sciencemag.org/content/341/6148/856.summary>