Planejamento e Análise de Experimentos (EEE933)
Estudo de Caso 2

Pedro Vinícius, Samara Silva e Savio Vieira

24 de Agosto de 2020

Introdução

O Índice de Massa Corporal (IMC) é uma medida de gordura corporal baseada na relação entre peso (em kg) e altura (em m) de um indivíduo e é comumente utilizado como uma ferramenta de triagem para indicar se uma pessoa está com um peso saudável para sua altura. Este índice é calculado conforme a Equação 1:

$$IMC = \frac{peso}{(altura)^2} \tag{1}$$

em kg/m^2. Cada faixa de IMC permite classificar o indivíduo em uma das seguintes categorias [1]:

- Baixo peso: $< 18,5$
- Peso normal: $18,5 - 25$
- Sobrepeso: $25 - 30$
- Obesidade: $30 - 35$
- Obesidade mórbida: > 40

Neste estudo de caso deseja-se comparar o IMC médio dos alunos de Pós-Graduação em Engenharia Elétrica (PPGEE) da Universidade Federal de Minas Gerais (UFMG) de dois semestres distintos: 2016/2 e 2017/2. Para tal fim, foram disponibilizadas duas amostras, contendo sexo, altura e peso de alguns alunos [2]. Assim, duas análises estatísticas independentes são propostas: (i) uma sobre o IMC médio dos alunos do sexo masculino e (ii) uma sobre o IMC médio dos alunos do sexo feminino. Para ambos os casos, a condução dos experimentos foram similares, no entanto, alguns testes tiveram que ser adaptados de acordo com as propriedades das distribuições amostrais investigadas.

Planejamento dos Experimentos

As hipóteses estatísticas foram definidas com o intuito de responder às questões propostas abaixo:

- Há evidências de que o IMC médio dos alunos do PPGEE/UFMG de 2016/2 é diferente do IMC médio dos alunos do PPGEE/UFMG de 2017/2 no que se refere ao sexo masculino?
- E quanto ao sexo feminino?

Em concordância com a proposta de comparação do IMC médio entre os alunos de semestres distintos, as hipóteses de teste podem ser formuladas sobre o parâmetro média:

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

onde a hipótese nula H_0 implica na igualdade entre os IMCs médios dos alunos de 2016/2 e 2017/2 e a hipótese alternativa bilateral H_1 na diferença dos IMCs médios e, portanto, em uma potencial diferença dos estilos de vida dos alunos.

Os parâmetros experimentais para realização dos testes são:

- A probabilidade admissível de rejeição da hipótese nula quando ela é verdadeira é de apenas 5%, isto é, o nível de significância do teste é $\alpha = 0,05$;
- A potência do teste é de $\pi = 1 - \beta = 0,8$. Em outras palavras, deseja-se uma probabilidade de falha ao rejeitar a hipótese nula quando ela é falsa de 20%;
- O tamanho de efeito de mínima relevância prática foi definido em $\delta^* = 1$, ou seja, pretende-se detectar, a partir do teste de hipóteses, desvios de 1 kg/m^2.

Pré-Processamento dos Dados

Conforme mencionado anteriormente, as bases de dados imc_20162.csv e CS01_20172.csv foram disponibilizadas [2]. A amostra relativa ao semestre de 2016/2 dispõe dos atributos número de identificação do aluno, visto que a coleta manteve o sigilo dos estudantes, curso (graduação ou pós-graduação), gênero, peso (em kg) e altura (em m). A princípio, foi necessário extrair apenas as informações dos alunos cujo vínculo com a universidade era de discente da pós-graduação e, posteriormente, realizar a fragmentação por gênero, formando duas amostras independentes ($M2016$ e $F2016$). A amostra relativa ao semestre de 2017/2, por sua vez, compreendia os atributos peso (em kg), altura (em m), sexo e idade. Além disso, as observações eram somente de alunos da pós-graduação e, portanto, exigiu apenas a separação por gênero em duas outras amostras ($M2017$ e $F2017$).

A partir dos pesos e alturas disponíveis, os índices de massa corporal foram calculados para cada aluno, conforme a Equação 1. Por fim, as observações de interesse foram compiladas em uma única estrutura de dados. Os 10 primeiros IMCs de cada amostra podem ser visualizados abaixo, onde os valores "NA" presentes nas amostras femininas ($F2016$ e $F2017$) indicam que ambas possuem tamanho amostral $N < 10$, isto é, 7 e 4 observações, respectivamente. As amostras masculinas ($M2016$ e $M2017$), no entanto, apresentam 21 observações cada uma.

```r
# 10 primeiras observações de cada amostra
show(IMCs[c(1:10),])
```

```
##      M2016    M2017    F2016    F2017
## 1  24.96801 29.73704 18.45917 17.36111
## 2  23.23346 26.95568 20.19509 20.83253
## 3  28.07504 29.06574 19.72318 17.84652
## 4  37.55102 30.42185 22.48133 17.74623
## 5  22.40879 20.76125 22.58955       NA
## 6  24.28098 24.38272 25.18079       NA
## 7  27.14304 23.74764 18.96193       NA
## 8  24.41928 22.49135       NA       NA
## 9  22.47121 24.89814       NA       NA
## 10 21.62630 22.40818       NA       NA
```

Análise Exploratória de Dados

```r
# Histogram
ggplot(IMCs, aes(value)) + xlab(expression("IMC (kg/"*"m"^2*")")) +
    ylab("Frequência") + geom_histogram(bins = 20) +
    facet_wrap(~variable, scales = 'free')
```

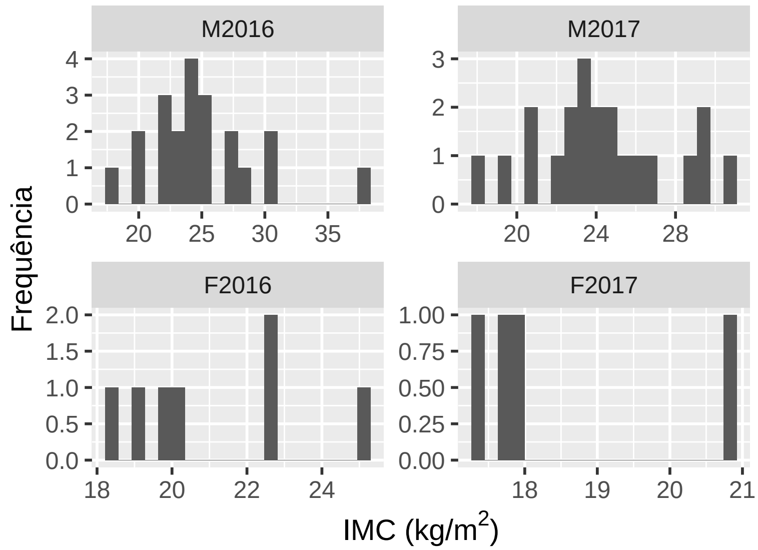

Figura 1: Histogramas.

```
# Boxplot
ggplot(data = IMCs, aes(y = "", x = value)) + xlab(expression("IMC (kg/"*"m"^2*")")) +
    ylab("") + geom_boxplot(lwd = 0.3) +
    facet_wrap(~variable, scales = 'free') + coord_flip()
```

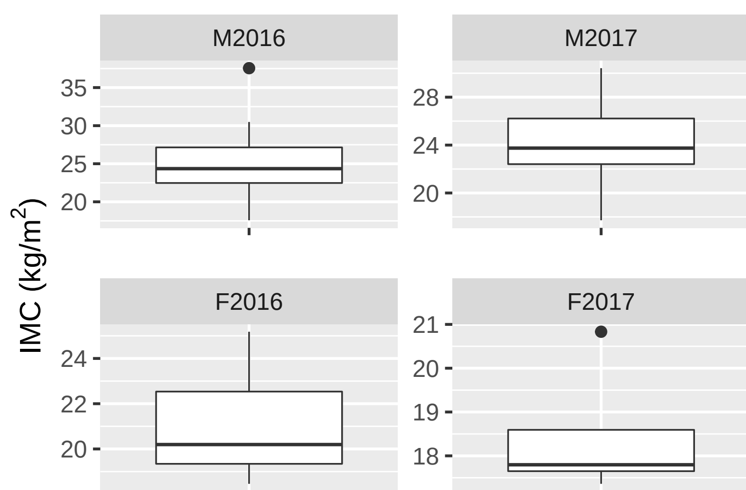

Figura 2: Boxplots.

```
# QQ-Plots
ggplot(data = IMCs, aes(sample = value)) +
  facet_wrap(~variable, scales = "free") +
  stat_qq() + stat_qq_line() + scale_y_continuous(name = 'Quantis da Amostra') +
  scale_x_continuous(name = 'Quantis Teóricos Normais')
```

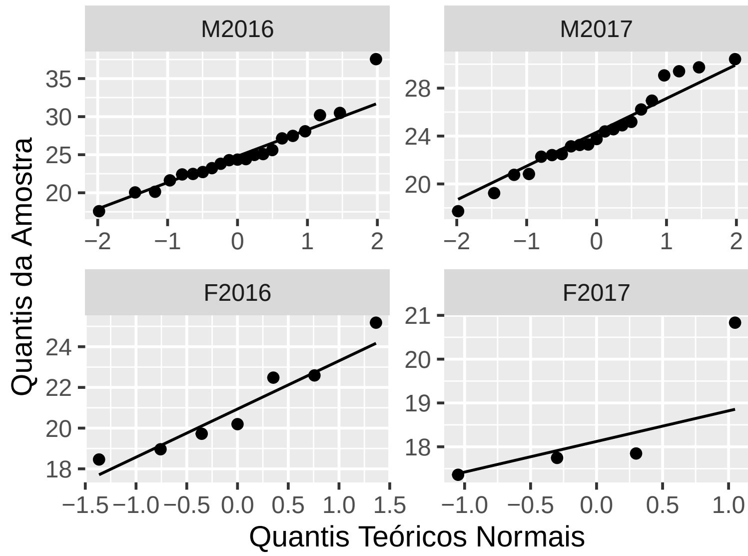

Figura 3: QQ-Plots

Análise Estatística

```
(t_test <- t.test(x = M2016,
                  y = M2017,
                alternative = "two.sided",
                conf.level = 0.95))
```

```
##
##  Welch Two Sample t-test
##
## data:  M2016 and M2017
## 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
```

```
cat('Intervalo de confiança:', t_test$conf.int[1:2])
```

```
## Intervalo de confiança: -1.788823 3.089716
```

```
wilcox.test(x = F2016, y = F2017, alternative = "two.sided")
```

```
##
##  Wilcoxon rank sum test
##
## data:  F2016 and F2017
## W = 24, p-value = 0.07273
## alternative hypothesis: true location shift is not equal to 0
```

Validação de Premissas

```r
shapiro.test(M2016)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  M2016
## W = 0.92833, p-value = 0.1275
```

```r
shapiro.test(M2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  M2017
## W = 0.96494, p-value = 0.6206
```

```r
shapiro.test(F2016)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  F2016
## W = 0.91974, p-value = 0.4674
```

```r
shapiro.test(F2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  F2017
## W = 0.7475, p-value = 0.03659
```

Conclusões

Discussão de Melhorias

Atividades Desempenhadas

Referências

[1] Alexandra M. N. Borba, Juliane H. Wolff Wolff, and Rafaela Liberali. Avaliação do Perfil Antropométrico e Alimentar de Idosos Institucionalizados em Blumenau-Santa Catarina. *RBONE-Revista Brasileira de Obesidade, Nutrição e Emagrecimento*, 1(3), 2007.

[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.