

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.

2 Planejamento do experimento

2.1 Objetivo do experimento

O objetivo é estudar as diferenças entre o IMC médio entre duas turmas de estudantes de pós-graduação em Engenharia elétrica na UFMG, para dados coletados nos semestres de 2016-2 e 2017-2. A análise será dividida entre as duas sub-populações (homens e mulheres), uma vez que é esperado diferenças no ICM e também no tamanho amostral. Portanto, para verificar estas diferenças, as seguintes hipóteses serão testadas:

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

A mesma hipótese será testada para ambas sub-populações. Os testes serão realizados para um tamanho de efeito $\delta^* = 5$, visto que essa é a diferença de valor entre os níveis de classificação IMC encontrados na literatura [1].

2.2 Análise Exploratória dos Dados

Antes de iniciar a análise exploratória dos dados, é preciso realizar um breve pré-processamento dos dados, visto que os dados de cada ano possuem formatos ligeiramente diferentes. Por exemplo, para os dados de 2016-2, é necessário remover as linhas referentes aos alunos de graduação. Ambos os arquivos foram padronizados para possuir o mesmo nome de coluna, e uma nova coluna com o cálculo do IMC foi adicionada em cada. Novos datasets também foram criados contendo somente os dados de cada sexo, para auxiliar nas análises.

```
imc <- function(height, weight){  
  return(weight/(height^2))  
}  
  
data.2016.2 <- read.csv('imc_20162.csv', header = T, sep = ",")  
data.2017.2 <- read.csv('CS01_20172.csv', header = T, sep = ";")  
data.2017.2 <- data.2017.2 %>% rename(Gender = Sex)  
  
# calculate the IMC  
data.2016.2$imc <- imc(data.2016.2$Height.m, data.2016.2$Weight.kg)  
data.2017.2$imc <- imc(data.2017.2$height.m, data.2017.2$Weight.kg)  
  
# remove undergraduate students for 2016.2 data  
data.2016.2 <- data.2016.2 %>% filter(Course == 'PPGEE')  
  
# remove unnecessary columns and rename
```

```

data.2016.2 <- data.2016.2 %>% select(Gender, imc)
data.2017.2 <- data.2017.2 %>% select(Gender, imc)

# split the 2016.2 data by gender and remove undergraduate students
data.2016.2.Females <- data.2016.2 %>% filter(Gender == 'F') %>% select(imc)
data.2016.2.Males <- data.2016.2 %>% filter(Gender == 'M') %>% select(imc)

# split the 2017.2 data by gender
data.2017.2.Females <- data.2017.2 %>% filter(Gender == 'F') %>% select(imc)
data.2017.2.Males <- data.2017.2 %>% filter(Gender == 'M') %>% select(imc)

# create a single dataset
data.all <- data.2016.2 %>% mutate(Class = '2016-2')
data.all <- data.all %>% bind_rows(data.2017.2 %>% mutate(Class = '2017-2'))

```

Para ter uma ideia inicial dos dados, suas estatísticas básicas são calculadas:

```
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

```

É possível fazer as seguintes observações que a quantidade de amostras do sexo feminino é bem menor do que a do masculino, o que pode afetar a potência dos testes a serem executados.

Para melhor visualização, um gráfico boxplot é gerado na figura abaixo. Nele é possível ver que há alguns outliers nos valores de IMC em alguns anos, que serão removidos posteriormente.



Figure 1: Boxplot dos dados

Referências

[1] *Tabela de imc*. <http://www.calculoimc.com.br/tabela-de-imc/>.