

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 sub-populações masculina e feminina foram analisadas separadamente. Após os testes verificou-se que ...

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:

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

Onde μ_1 e μ_2 são as turmas de 2016-7 e 2017-2, respectivamente. A mesma hipótese será testada separadamente para cada sub-população.

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 e todos os dados, para auxiliar nas análises.

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.

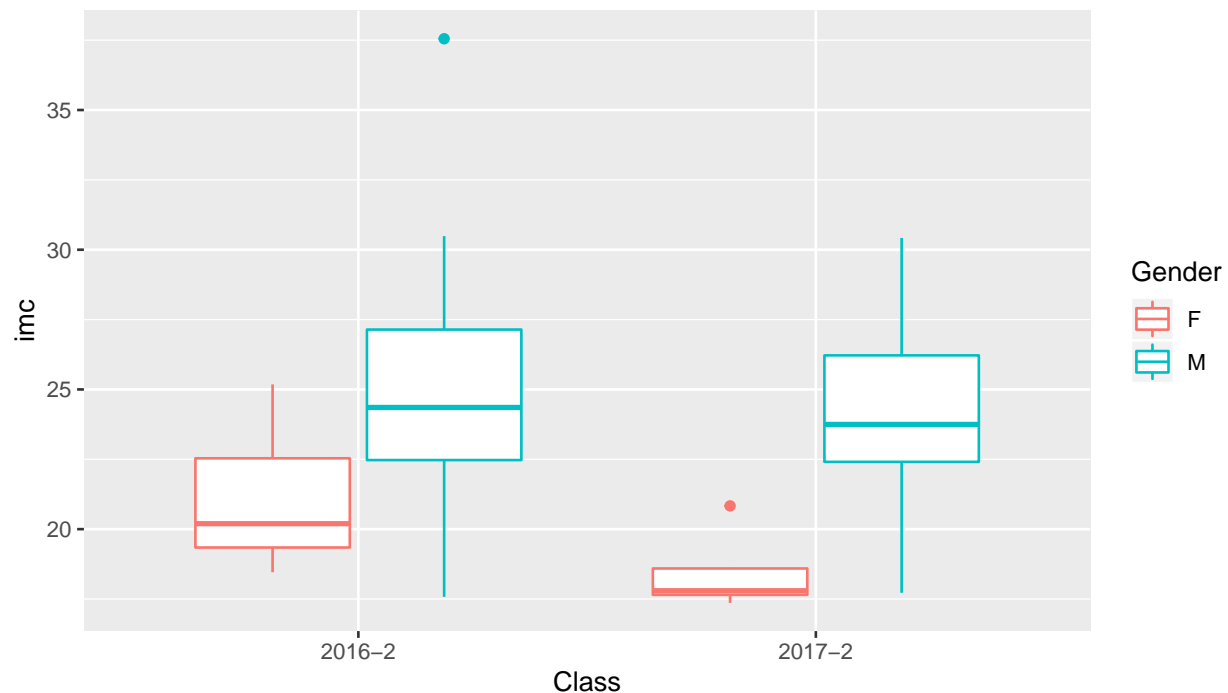


Figure 1: Boxplot dos dados

2.2.1 Estimativa da significância e potência

Um nível de significância $\alpha = 0.05$ é estabelecido para uma potência de $\pi = 1 - \beta = 0.8$. Dado que o número de amostras é conhecido para cada turma e gênero, avalia-se a seguir se a quantidade de amostras é suficiente para atender os parâmetros especificados.

```
(sig.est.males <- pwr.t.test(d = d.cohen.males,
  n = n.males,
  sig.level = NULL,
  type = "two.sample",
  alternative = "two.sided",
  power = pi))
```

```
##
## Two-sample t test power calculation
```

```
##
##           n = 21
##           d = 1.280529
##       sig.level = 0.002185197
##           power = 0.8
##       alternative = two.sided
##
## NOTE: n is number in *each* group
(sig.est.females <- pwr.t.test(d = d.cohen.females,
  n = n.females,
  sig.level = NULL,
  type = "two.sample",
  alternative = "two.sided",
  power = pi))
```

```
##
##       Two-sample t test power calculation
##
##           n = 4
##           d = 2.437794
##       sig.level = 0.04525425
##           power = 0.8
##       alternative = two.sided
##
## NOTE: n is number in *each* group
```

Portanto, é possível ver que há amostras suficientes para a realização dos testes de cada sub-população. É importante notar que, dado que há mais amostras de estudantes do sexo masculino, a significância mínima estimada foi bem menor. Um outro resultado interessante, causado pela diferença do número de amostras, é que a potência do teste para alunos do sexo feminino será significativamente menor em relação à do sexo masculino. Assumindo $\alpha = 0.05$, a potência é estimada a seguir.

```
(pwr.est.females <- pwr.t.test(d = d.cohen.females,
  n = n.females,
  sig.level = 0.05,
  type = "two.sample",
  alternative = "two.sided",
  power = NULL))
```

```
##
##       Two-sample t test power calculation
##
##           n = 4
##           d = 2.437794
##       sig.level = 0.05
##           power = 0.8177638
##       alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
(pwr.est.males <- pwr.t.test(d = d.cohen.males,
  n = n.males,
  sig.level = 0.05,
  type = "two.sample",
  alternative = "two.sided",
```

```
power = NULL))
```

```
##  
##      Two-sample t test power calculation  
##  
##              n = 21  
##              d = 1.280529  
##      sig.level = 0.05  
##              power = 0.9815966  
##      alternative = two.sided  
##  
## NOTE: n is number in *each* group
```

Conforme esperado, a potência do teste masculino é 17% maior que o feminino, indicando uma maior possibilidade de erros do tipo 2.

Referências

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