Estudo de Caso 02: Comparação do IMC médio de alunos do PPGEE-UFMG ao longo de dois semestres

Diego Pontes, Elias Vieira, Matheus Bitarães

Janeiro, 2021

Descrição do problema

Neste estudo, deseja-se comparar o IMC médio de duas populações de alunos da pós-graduação da Engenharia Elétrica (PPGEE) da UFMG no segundo semestre de 2016 e de 2017. Para este estudo, foram disponibilizadas duas amostras, sendo uma para cada semestre em questão, onde serão feitas as análises para o estudo já mencionado.

Introdução

Reconhecido internacionalmente pela Organização Mundial da Saúde (OMS), o IMC indica o peso adequado para cada pessoa, fazendo uma relação entre sua massa corpórea (em kg) e sua altura (em m) [1], conforme a Equação 1.

$$IMC = \frac{peso}{altura * altura} \tag{1}$$

Pode-se classificar o valor do IMC conforme apresentado abaixo [1]:

- IMC abaixo de 18,5: Peso abaixo do normal;
- IMC entre 18,5 e 24,9: São pesos considerados normais pela OMS;
- IMC entre 25 e 29,9: Peso em pré-obesidade ou acima do peso;
- IMC entre 30 e 34,9: Este índice indica obesidade grau um;
- IMC acima 35 e 39,9: Indica obesidade grau dois
- IMC acima de 40: Indica obesidade grau três ou mórbida

Design do Experimento

Como já mencionado, deseja-se comparar o IMC dos alunos do PPGEE-UFMG de dois semestres distintos, conforme mostras fornecidas. Para tal estudo, serão feitas duas análises e testes estatísticos independentes considerando duas subpopulações distintas, uma considerando somente o sexo masculino e outra para o sexo feminino.

As seguintes hipóteses estatísticas foram definidas:

1) Há evidências de que a média do IMC dos alunos do PPGEE-UFMG de 2/2016 é diferente da média do IMC dos alunos do PPGEE-UFMG de 2/2017? (subpopulação masculina)

2) Há evidências de que a mediana do IMC das alunas do PPGEE-UFMG de 2/2016 é diferente da mediana do IMC das alunas do PPGEE-UFMG de 2/2017? (subpopulação feminina)

No decorrer do estudo ficará claro o motivo pela qual foi utilizado média e mediana para a subpopulação masculina e feminina, respectivamente.

Dadas as hipóteses estatísticas descritas acima, podem-se definir as seguintes hipóteses de testes, em função da média e mediana do IMC dos alunos do sexo masculino e do sexo feminino, respectivamente:

$$\begin{cases} H_0 : \mu_{m2016} = \mu_{m2017} \\ H_1 : \mu_{m2016} \neq \mu_{m2017} \end{cases}$$

$$\begin{cases} H_0 : \mu_{f2016} = \mu_{f2017} \\ H_1 : \mu_{f2016} \neq \mu_{f2017} \end{cases}$$

Além das hipóteses estatísticas acima, tem-se as seguintes definições:

1) Nível de significância (α) de 0,05. O nível de significância é a probabilidade de ocorrência de um falso positivo em qualquer procedimento de teste de hipótese [2].
2) Potência do Teste (π) = 1 - (β) = 0,80. Onde beta é a probabilidade de ocorrência de um falso negativo em qualquer procedimento de teste de hipótese [2] e, portanto, a potência de teste quantifica a sensibilidade do teste à efeitos que violam sua hipótese nula [2].

Análise Estatística

Importação dos dados

Foram importados os arquivos *imc_20162.csv* e *CS01_20172.csv* para o estudo proposto.

```
# importação dos dados
raw_data_2016 <- read.csv(file = 'imc_20162.csv')
raw_data_2017 <- read.csv(file = 'CS01_20172.csv', sep=';')

head(raw_data_2016)
```

```
##   ID Course Gender Height.m Weight.kg
## 1  1  PPGEE      F     1.57      45.5
## 2  2  PPGEE      F     1.62      53.0
## 3  3  PPGEE      F     1.70      57.0
## 4  4  PPGEE      F     1.62      59.0
## 5  5  PPGEE      F     1.67      63.0
## 6  6  PPGEE      F     1.76      78.0
```

```
head(raw_data_2017)
```

```
##   Weight.kg height.m Sex Age.years
## 1      89.0     1.73   M        23
## 2      72.5     1.64   M        28
## 3      84.0     1.70   M        34
## 4      90.0     1.72   M        27
## 5      60.0     1.70   M        33
## 6      79.0     1.80   M        27
```

Como pode ser visto, há diferenças estruturais entre os arquivos, como colunas com nomes diferentes, além de dados de alunos que não pertencem ao PPGEE-UFMG no arquivo de dados de 2016. Portanto, estes dados foram tratados para que ficassem com mesma estrutura, conforme códigos abaixo.

```
# Filtra dados apenas de estudantes do ppgee (necessario apenas em 2016)
raw_data_2016 <- subset(raw_data_2016, Course=="PPGEE")

# renomeia coluna de 2016
names(raw_data_2016)[names(raw_data_2016) == "Gender"] <- "Sex"

# renomeia coluna de 2017
names(raw_data_2017)[names(raw_data_2017) == "height.m"] <- "Height.m"
```

Na sequência, deve-se criar uma nova coluna com o calculo do IMC e separar os dados entre masculino e feminino, conforme códigos abaixo.

```
# cria coluna com calculo do IMC
raw_data_2016$IMC = raw_data_2016$Weight.kg / (raw_data_2016$Height.m * raw_data_2016$Height.m)
raw_data_2017$IMC = raw_data_2017$Weight.kg / (raw_data_2017$Height.m * raw_data_2017$Height.m)

# separa entre masculino e feminino e armazena apenas o IMC
imc_m_2016 <- subset(raw_data_2016, Sex=="M")$IMC
imc_f_2016 <- subset(raw_data_2016, Sex=="F")$IMC
imc_m_2017 <- subset(raw_data_2017, Sex=="M")$IMC
imc_f_2017 <- subset(raw_data_2017, Sex=="F")$IMC
```

Dados estatísticos

Com as amostras tratadas e separadas conforme proposta inicial e em posse do IMC destas subpopulações, tem-se os seguintes dados estatísticos:

```
# imprime um sumario com as principais informações estatisticas dos IMCs
summary(imc_m_2016)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.58   22.47   24.36   24.94   27.14   37.55
```

```
summary(imc_m_2017)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.72   22.41   23.75   24.29   26.22   30.42
```

```
summary(imc_f_2016)
```

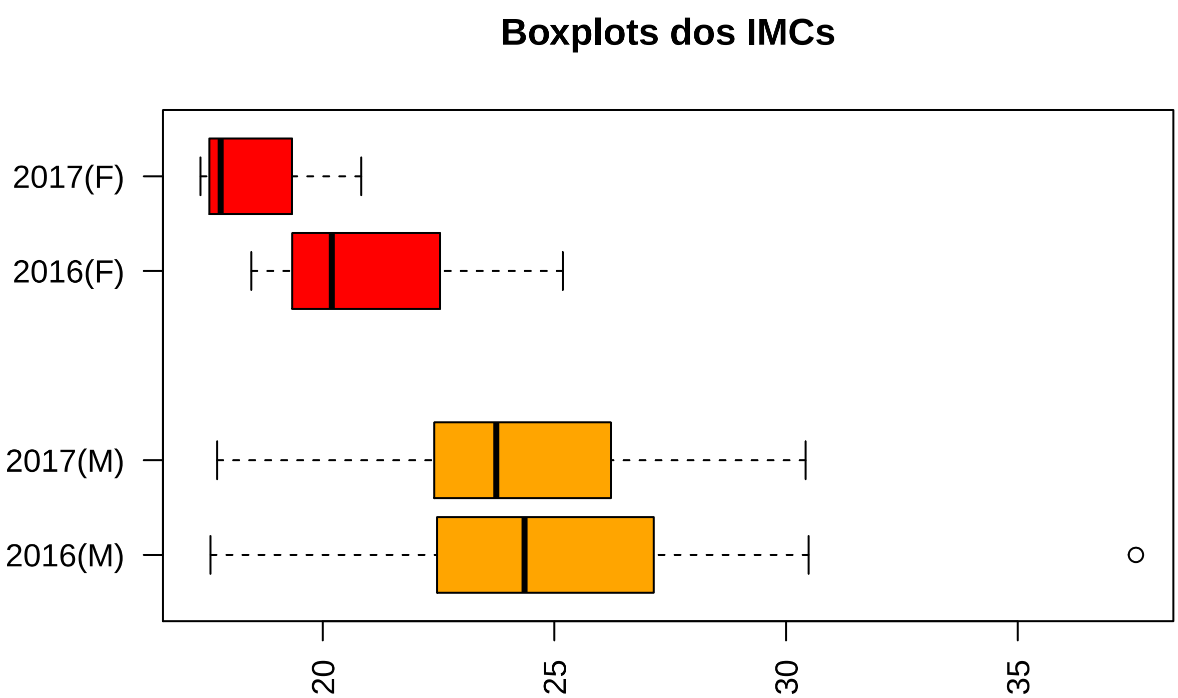
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.46   19.34   20.20   21.08   22.54   25.18
```

```
summary(imc_f_2017)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.36   17.65   17.80   18.45   18.59   20.83
```

A fim de avaliar a distribuição dos dados obtidos, foi feita uma análise gráfica a partir do boxplot das subpopulações separadas pelo ano.

```r
# boxplot
boxplot(imc_m_2016, imc_m_2017, imc_f_2016, imc_f_2017,
main = "Boxplots dos IMCs",
at = c(1,2,4,5),
names = c("2016(M)", "2017(M)", "2016(F)", "2017(F)"),
las = 2,
col = c("orange","orange", "red", "red"),
horizontal = TRUE,
notch = FALSE
)
```



Necessário uma análise mais detalhada do boxplot(diego)

Para uma avaliação estatística sobre a validação da hipótese de uma distribuição normal para as amostras, pode-se usar o teste Shapiro-Wilk. Para este teste, temos:

$$\begin{cases} H_0 : \text{A amostra provém de uma população com distribuição normal} \\ H_1 : \text{A amostra não provém de uma população com distribuição normal} \end{cases}$$

```r
# teste de Shapiro-Wilk
shapiro.test(imc_m_2016)
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  imc_m_2016
## W = 0.92833, p-value = 0.1275
```

```
shapiro.test(imc_m_2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_m_2017
## W = 0.96494, p-value = 0.6206
```

```
shapiro.test(imc_f_2016)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_f_2016
## W = 0.91974, p-value = 0.4674
```

```
shapiro.test(imc_f_2017)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  imc_f_2017
## W = 0.7475, p-value = 0.03659
```

Como interpretação do teste, temos que se p-valor < 0.05 (α), deve-se rejeitar a hipótese nula, ou seja, os dados não possuem distribuição normal[3], caso contrário, não é possível concluir que os dados não seguem uma distribuição normal. Portanto, analisando os resultados dos testes dispostos acima, tem-se uma confirmação das suposições de normalidade feitas anteriormente, com exceção da da subpopulação feminina de 2017, a qual não podemos assumir uma distribuição normal.

Com estas informações, justifica-se as hipóteses estatísticas definidas, pois a média é usada para distribuições numéricas normais, que têm uma baixa quantidade de valores discrepantes, enquanto a mediana é geralmente utilizada para retornar a tendência central para distribuições numéricas distorcidas[6]. **[matheus]não entendi este paragrafo**

Em posse dessas conclusões, o estudo será dividido em duas partes (masculino e feminino), devido às diferenças no processamento dos dados que ocorrerão.

1) Subpopulação masculina

Em posse do resultado do teste de Shapiro-Wilk, podemos assumir normalidade para as amostras da subpopulação masculina de 2016 e 2017. Pretende-se utilizar o Teste T para comparação das amostras e, para isto, é necessário que as variâncias possam ser consideradas iguais (homocedasticidade). Para a análise da variância dos dados, pode-se usar o Teste F, onde tem-se:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

```r
# Teste F
var.test(imc_m_2016, imc_m_2017, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  imc_m_2016 and imc_m_2017
## F = 1.5839, num df = 20, denom df = 20, p-value = 0.3119
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6426853 3.9034665
## sample estimates:
## ratio of variances
##           1.583888
```

Analisando o p-valor, temos que p-valor>0,05 (α) e, portanto, não há evidência estatistica forte o suficiente que indique que as variâncias não são iguais. Portanto, será considerada a homocedasticidade entre as subpopulações.

Para a hipótese estatística proposta para as médias do IMC masculino, pode-se usar o teste t de student, visto que a distribuição normal dos valores já foi validada, assim como a variância constante dos erros experimentais para observações distintas. Neste teste, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \mu 1 = \mu 2 \\ H_1 : \mu 1 \neq \mu 2 \end{cases}$$

```r
# teste t de student
t.test(imc_m_2016, imc_m_2017, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  imc_m_2016 and imc_m_2017
## t = 0.53979, df = 40, p-value = 0.5923
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.784943  3.085836
## sample estimates:
## mean of x mean of y
##  24.93595  24.28551
```

Como pode ser visto, o teste t de student retornou um p-valor igual a 0,5923, cujo valor é maior que o nível de significâcia adotado (0,05). Portanto, com 95% de confiança não é possível rejeitar a hipótese nula do estudo proposto de que as médias das populações masculinas dos dois anos são iguais. Sendo a hipótese alternativa H1 bilateral, o intervalo de confiança para a diferença das médias é [-1.784943 3.085836].

Um outro teste pode ser feito para validação do resultado que é o Teste T de Welch. Percebe-se que ele também falhou em rejeitar a hipótese nula do estudo.

```r
# teste t de Welch
t.test(imc_m_2016, imc_m_2017, "two.sided", mu=0, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  imc_m_2016 and imc_m_2017
## 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
```

Embora o nível de significância tenha sido definido inicialmente, é interessante encontrar o tamanho de efeito, sendo este a medida da importância prática dos resultados de eventuais diferenças encontradas entre duas ou mais médias ou variâncias [4]. Existem várias maneiras de se fazer isto tais como: o Teste de Cohen, Teste de Glass, Teste de Hedges, Teste Psi, dentre outros [4]. O Teste de Cohen, por exemplo, foi desenhado para ser utilizado quando os escores das duas populações que estão sendo comparadas são contínuos e de distribuição normal[5]. Dado a distribução normal e a igualdade de variâncias assumidas de acordo com os testes de Shapiro-Wilk e F, respectivamente, pode-se usar o d de Cohen como estimativa do tamanho de efeito dos dados masculinos.

```r
# d de Cohen
# library(effsize)
#cohen.d(imc_m_2016, imc_m_2017)
```

O tamanho de efeito retornado pela função foi de d = 0,1665.

Para se calcular a Potência do Teste, pode-se utilizar a função power.t.test. Nela deve ser inserido o parâmetro o número de observações por grupo (n = 21), delta (d de cohen = 0,1665831), desvio padrão conjugado (sd = 3,9046). O desvio padrão conjugado pode ser obtido pela Equação 2, onde s_1 e s_2 podem ser obtidos pela função sd.

```r
# Desvio padrão
sd(imc_m_2016)
```

```
## [1] 4.323356
```

```r
sd(imc_m_2017)
```

```
## [1] 3.435254
```

$$sd = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{2}$$

```r
# Potência do Teste
power.t.test(21,0.1665831, 3.9046, sig.level = 0.05)
```

```
##
##      Two-sample t test power calculation
##
##               n = 21
```

```
##           delta = 0.1665831
##              sd = 3.9046
##       sig.level = 0.05
##           power = 0.03400086
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Observa-se que a potencia de teste obtida foi de 3,4% que é muito menor que a potência desejada (80%). Mantendo o tamanho de efeito calculado, para que a Potência do Teste fosse a desejada, seriam necessárias 8626 amostras para cada ano, o que não é possível.

```r
# Número de amostras para o Potência do Teste de 80%
power.t.test(NULL, 0.1665831, 3.9046, sig.level = 0.05, power = 0.80)
```

```
##
##      Two-sample t test power calculation
##
##               n = 8625.36
##           delta = 0.1665831
##              sd = 3.9046
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

2) Subpopulação feminina

Diferente da subpopulação masculina, não podemos tratar a amostra feminina em uma distribuição normal. Para a análise da proposta estatística da população feminina, pode-se usar o teste não-paramétrico de Wilcoxon com as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As medianas das populações são iguais} \\ H_1 : \text{As medianas das populações não são iguais} \end{cases}$$

```r
# teste de Wilcoxon
wilcox.test(imc_f_2016, imc_f_2017, alternative = "two.sided",conf.int = TRUE)
```

```
##
##  Wilcoxon rank sum test
##
## data:  imc_f_2016 and imc_f_2017
## W = 24, p-value = 0.07273
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.6374374  5.2284403
## sample estimates:
## difference in location
##               2.162763
```

Como pode ser visto, o teste de Wilcoxon retornou um p-valor igual a 0,07273, cujo valor é maior que o nível de significância adotado (0,05). Portanto, com 95% de confiança não é possível rejeitar a hipótese nula do estudo proposto de que as médias das populações masculinas dos dois anos são iguais. Sendo a hipótese alternativa H1 bilateral, o intervalo de confiança para a diferença das médias é [-0, 6374374; 5, 22844].

Abordagem do tamanho de efeito

```
# Tamanho de efeito
# (vd_a <- VD.A(na.omit(imc_f_2016), na.omit(imc_f_2017)))

# cat('Magnitude:', vd_a$magnitude)
```

A medida A pertence ao domínio $[0,\ 1]$ com as seguintes características:

- Quando a medida $A = 0,5$, pode-se afirmar que não há diferenças evidentes na estatística que se estuda;
- Quando a medida $A < 0,5$, a primeira população apresenta valores menores da estatística de interesse
- Quando a medida $A > 0,5$, a segunda população apresenta valores menores da estatística de interesse.

Desta forma, pode-se concluir que, em 85% das vezes, alunas de 2/2016 possuem maior IMC mediano que as alunas do 2/2017, podendo encontrar diferenças maiores ou iguais a 4.

Abordagem do Potência do Teste

A menina me passou o que colocou nessa parte, mas acho que devemos encontrar algum teste para jsutificar isso

A ausência de normalidade da distribuição feminina do segundo semestre de 2017 impossibilita a aplicação da função `power.t.test` para análise do poder do teste, uma vez que esse cálculo de potência é específico para o teste t de Student. Nesse caso, o apropriado é calcular o poder do teste a partir de variações dessa função para o teste de Wilcoxon. No R existem algumas possibilidades dentro do pacote `wmwpow` [**?**], como o cálculo do poder usando a abordagem de Shieh [**?**] e outros dois baseados no método de Monte Carlo. Todavia, tais abordagens requerem que o tipo de distribuição seja passado por parâmetro, e portanto, conhecido previamente. Em vista disso e principalmente pelo baixo tamanho amostral que impossibilita a estimativa empírica da distribuição amostral, o cálculo do poder não pôde ser realizado. No entanto, dado que os tamanhos amostrais são distintos ($n_{F2016} \neq n_{F2017}$) e extremamente pequenos, é bem provável que a potência para efeitos maiores ou iguais a 4 kg/m^2 seja substancialmente menor que a desejada ($\pi = 0,80$).

Discussão e Conclusão

Lembrando que precisamos comentar ou incrementar o que foi comentado sobre:

- **Derivação de conclusões e recomendações.**
- **Discussão sobre a potência do teste (se aplicável).**
- **Discussão sobre possíveis formas de melhorar este experimento.**

Atividades dos membros

...

Referencias

[1] https://www.unimedfortaleza.com.br/blog/cuidar-de-voce/como-calcular-imc

[2] Notas de aula

[3] https://rpubs.com/paternogbc/46768

[4] http://www.cpaqv.org/estatistica/tamanho_do_efeito.pdf

[5] https://www.lume.ufrgs.br/bitstream/handle/10183/158388/000944613.pdf?sequence=1#:~:text=Cohen%20introduziu%

[6] https://support.zendesk.com/hc/pt-br/articles/228989407-M%C3%A9dia-versus-Mediana#:~:text=A%20m%C3%A9dia%

[7] https://www.researchgate.net/profile/Helena_Espirito_Santo/publication/273143169_Calcular_e_apresentar_tamanhos_do_efeito_em_trabalhos_cientificos_1_As_limitacoes_do_p_005_na_analise_de_diferencas_de_medias_de_dois_grupos_Calculating_and_reporting_effect_sizes_on_scientific_papers/links/54fa10290cf2040df21b1b1c/Calcular-e-apresentar-tamanhos-do-efeito-em-trabalhos-cientificos-1-As-limitacoes-do-p-0-05-na-analise-de-diferencas-de-medias-de-dois-grupos-Calculating-and-reporting-effect-sizes-on-scientific-pap.pdf