

Estudo de Caso 01: Comparação do IMC médio de alunos do PPGE-UFMG ao longo de dois semestres

Diego Pontes, Elias Vieira, Matheus Bitarões

Janeiro, 2021

Descrição do problema

... A descrição é o seguinte

Design do Experimento

...

Análise Estatística

...

Importação dos dados

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

```
# 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  PPGE    F      1.57      45.5
## 2  2  PPGE    F      1.62      53.0
## 3  3  PPGE    F      1.70      57.0
## 4  4  PPGE    F      1.62      59.0
## 5  5  PPGE    F      1.67      63.0
## 6  6  PPGE    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
```

Podemos observar que os arquivos possuem algumas diferenças estruturais. Alguns nomes de colunas estão diferentes e o arquivo de 2016 contém dados de alunos de fora do PPGEE. Estes dados devem portanto ser tratados para que fiquem similares.

```
# 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"
```

Agora devemos criar uma nova coluna com o calculo do IMC e separar os dados entre masculino e feminino

```
# 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
```

Informação dos dados

Após amostras estarem tratadas e o imc calculado, seguem alguns dados estatísticos

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  144.8   203.5   230.7   238.4   272.5   352.2
```

```
summary(imc_m_2017)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  146.0   187.8   208.9   220.4   256.0   311.0
```

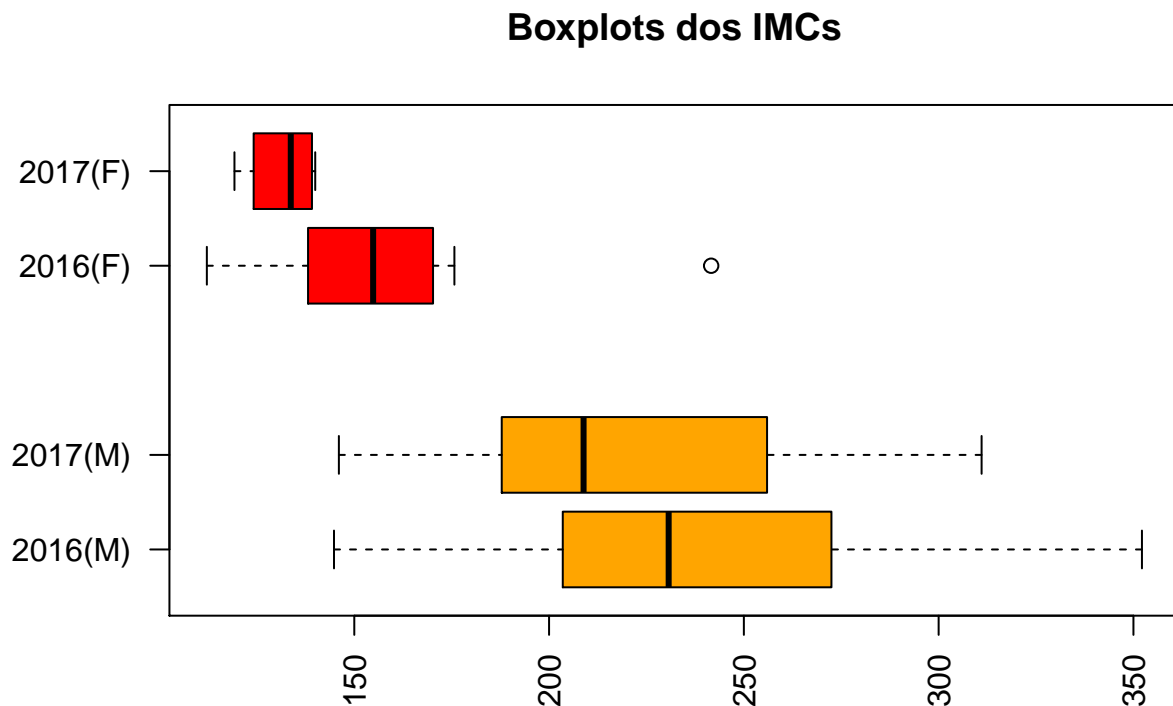
```
summary(imc_f_2016)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  112.2   138.1   154.8   160.8   170.2   241.6
```

```
summary(imc_f_2017)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    119.2   126.6   133.7   131.7   138.7   140.0
```

```
# 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
)
```



obs: tem muito pouca amostra feminina. Acredito que precisamos falar sobre o impacto disso nas análises

Agora precisa verificar se amostras podem ser analisadas como distribuições normais (teste de Shapiro-Wilk)

```
# teste de shapiro-wilk. Se o p-valor < 0.05, não podemos assumir que as amostras apresentam normalidad
shapiro.test(imc_m_2016)
```

```
##
```

```
## Shapiro-Wilk normality test
##
## data:  imc_m_2016
## W = 0.97371, p-value = 0.813
```

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

```
##
## Shapiro-Wilk normality test
##
## data:  imc_m_2017
## W = 0.95917, p-value = 0.4995
```

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

```
##
## Shapiro-Wilk normality test
##
## data:  imc_f_2016
## W = 0.90281, p-value = 0.3483
```

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

```
##
## Shapiro-Wilk normality test
##
## data:  imc_f_2017
## W = 0.90875, p-value = 0.4758
```

depois, caso sim, pode-se realizar o teste t:

```
# teste t de student
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 = 1.1384, df = 36.682, p-value = 0.2623
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -14.01230  49.92465
## sample estimates:
## mean of x mean of y
## 238.3947 220.4385
```

```
t.test(imc_f_2016, imc_f_2017, "two.sided", mu=0, conf.level = 0.95)
```

```
##
## Welch Two Sample t-test
```

```
##
## data:  imc_f_2016 and imc_f_2017
## t = 1.7856, df = 7.0567, p-value = 0.117
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.37583 67.58559
## sample estimates:
## mean of x mean of y
## 160.7570 131.6521
```

Discussão e Conclusão

...

Atividades dos membros

...

Referencias