

EEE933 - Planejamento e Análise de Experimentos

Estudo de Caso 01: Comparação do IMC Médio de Alunos do PPGEE-UFMG

Bernardo Bacha* Autor 2 (placeholder)[†] Autor 3 (placeholder)[‡] Autor 4 (placeholder)[§]

15 de setembro de 2025

I. Descrição do Problema

Este estudo busca comparar o IMC médio dos alunos do PPGEE/UFMG entre os semestres **2016-2** e **2017-2**. Além da análise geral, também será feita a comparação separada por sexo (masculino e feminino).

II. Desenho Experimental

- **População de interesse:** alunos do PPGEE/UFMG.
- **Variável resposta:** $IMC = \text{Peso} / \text{Altura}^2$.
- **Fatores:** semestre (2016-2 vs 2017-2) e sexo (M/F).
- **Hipóteses de teste (bicaudais):**
 - $H_0: \mu_{2016-2} = \mu_{2017-2}$
 - $H_1: \mu_{2016-2} \neq \mu_{2017-2}$

III. Desenvolvimento

Importação e Organização dos Dados

Primeiro, importamos os dados dos dois semestres (2016-2 e 2017-2), ajustamos o formato, filtramos apenas a pós-graduação e calculamos o IMC de cada estudante. Depois unimos tudo em um único dataframe.

```
library(dplyr)

# Ler os dados
df_2017 <- read.csv("data/CS01_20172.csv", sep = ";")
df_2016 <- read.csv("data/imc_20162.csv")
```

*PPGEE/UFMG — bernardobr@ufmg.br

[†]PPGEE/UFMG — autor2@email.com

[‡]PPGEE/UFMG — autor3@email.com

[§]PPGEE/UFMG — autor4@email.com

```

# Ajustar e calcular IMC
df_2017 <- df_2017 %>%
  rename(Weight.kg = Weight.kg,
         Height.m = height.m,
         Gender = Sex) %>%
  mutate(semestre = "2017-2",
         IMC = Weight.kg / (Height.m^2))

df_2016 <- df_2016 %>%
  filter(Course == "PPGEE") %>%
  mutate(semestre = "2016-2",
         IMC = Weight.kg / (Height.m^2))

# Unir as duas bases
dados <- bind_rows(df_2016, df_2017)

# Resumo rápido
summary(dados$IMC)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  17.36   20.83   23.26   23.68   25.18   37.55

```

```
table(dados$semestre, dados$Gender)
```

```

##
##           F  M
##  2016-2   7 21
##  2017-2   4 21

```

Análise Exploratória

Agora vamos explorar os dados. O objetivo é entender como está a distribuição do IMC entre os alunos, comparando os dois semestres e separando por sexo. Para complementar, também apresentamos os resultados no grupo total.

Boxplots

Os boxplots permitem observar a mediana, a dispersão e possíveis outliers do IMC em cada grupo.

```

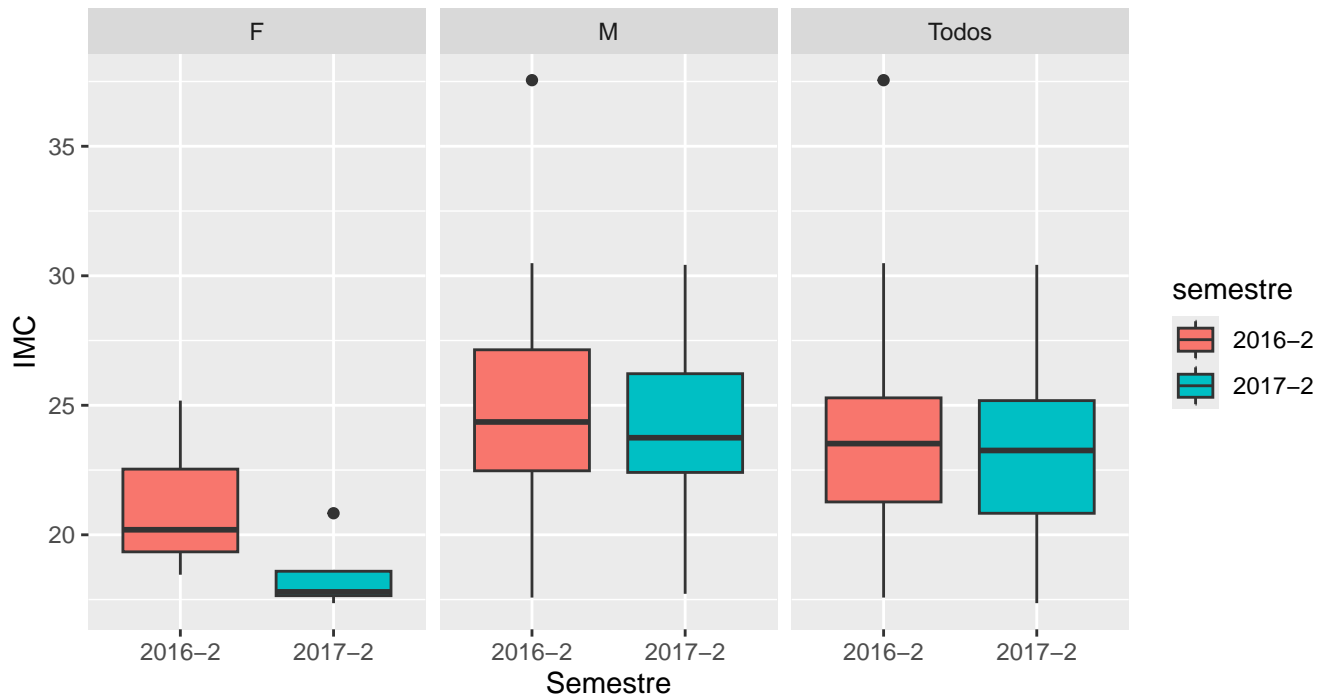
library(ggplot2)

# Adicionar categoria "Todos"
dados_global <- dados %>%
  mutate(Gender = as.character(Gender)) %>%
  bind_rows(dados %>% mutate(Gender = "Todos"))

# Boxplot
ggplot(dados_global, aes(x = semestre, y = IMC, fill = semestre)) +
  geom_boxplot() +
  facet_wrap(~ Gender) +
  labs(title = "Distribuição do IMC por semestre (global e por sexo)",
       x = "Semestre", y = "IMC")

```

Distribuição do IMC por semestre (global e por sexo)

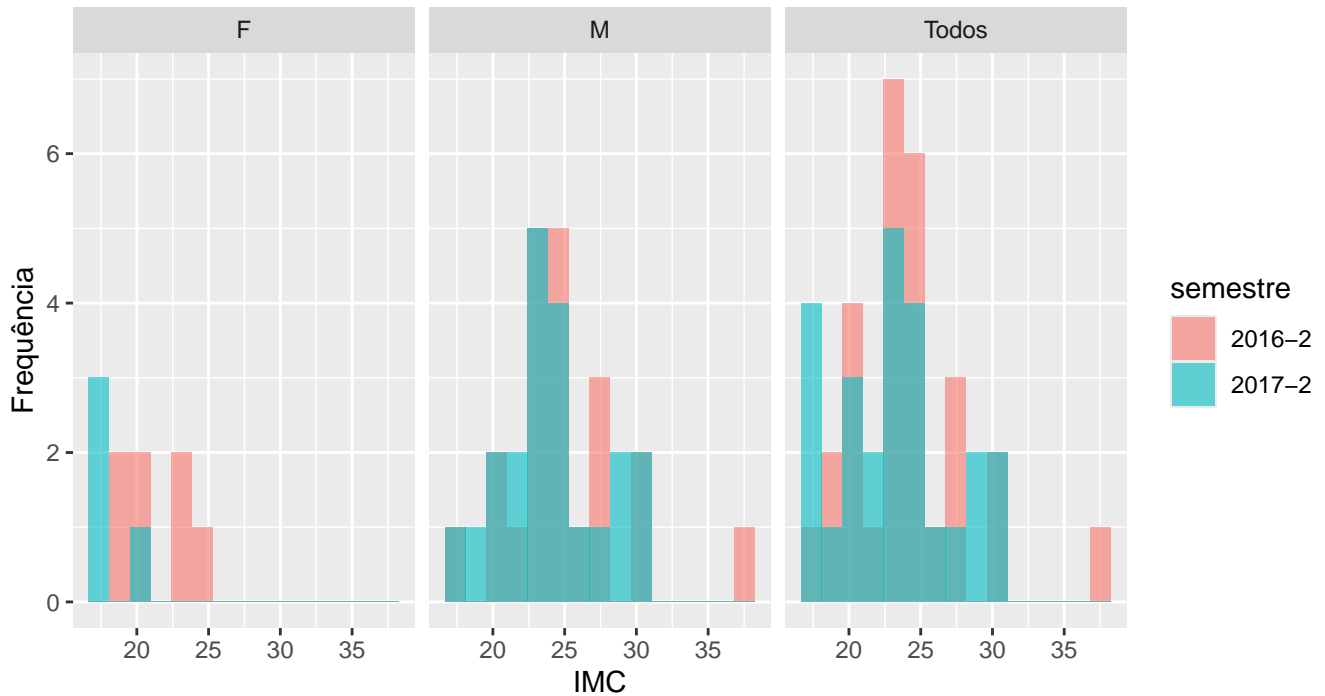


Histogramas

Já os histogramas mostram como os valores de IMC estão distribuídos dentro de cada grupo, ajudando a verificar se os dados seguem um padrão próximo da normalidade.

```
ggplot(dados_global, aes(x = IMC, fill = semestre)) +  
  geom_histogram(alpha = 0.6, position = "identity", bins = 15) +  
  facet_wrap(~ Gender) +  
  labs(title = "Distribuição do IMC por semestre (global e por sexo)",  
        x = "IMC", y = "Frequência")
```

Distribuição do IMC por semestre (global e por sexo)



IV. Análise Estatística

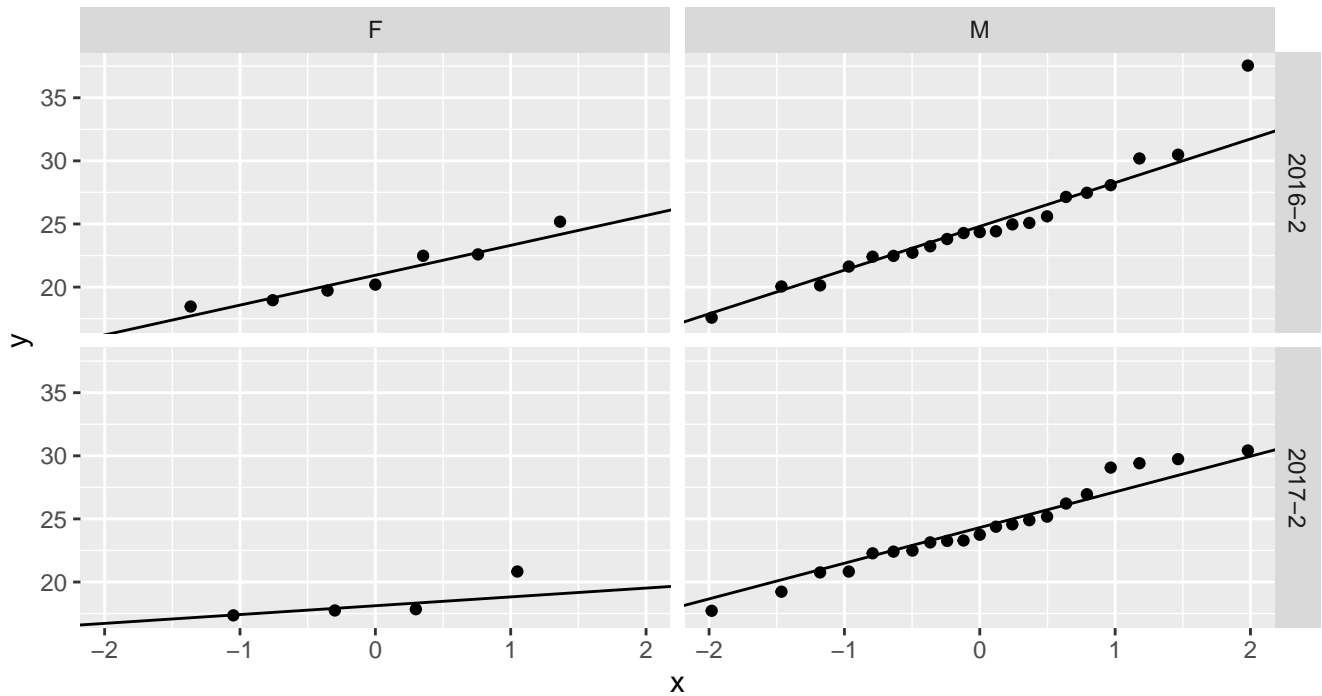
Validação das Premissas

Normalidade

```
library(ggplot2)

# QQ-plots por semestre x sexo
ggplot(dados, aes(sample = IMC)) +
  stat_qq() +
  stat_qq_line() +
  facet_grid(semestre ~ Gender) +
  labs(title = "QQ-plots do IMC por semestre e sexo")
```

QQ-plots do IMC por semestre e sexo



```
# Shapiro-Wilk por grupo (semestre x sexo)
grupos <- split(dados$IMC, list(dados$semestre, dados$Gender), drop = TRUE)
lapply(grupos, function(x) {
  if (length(x) >= 3 && length(x) <= 5000) shapiro.test(x) else NA
})
```

```
## $'2016-2.F'
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.91974, p-value = 0.4674
##
##
## $'2017-2.F'
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.7475, p-value = 0.03659
##
##
## $'2016-2.M'
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.92833, p-value = 0.1275
##
##
## $'2017-2.M'
```

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

Resultados (Shapiro–Wilk):

- 2016-2 Feminino: $p = 0.4674$
- 2017-2 Feminino: $p = 0.0366$
- 2016-2 Masculino: $p = 0.1275$
- 2017-2 Masculino: $p = 0.6206$

Nota: o grupo feminino de 2017-2 tem $n = 4$. Com tamanhos tão pequenos, testes de normalidade ficam instáveis e sensíveis a um único valor. O resultado é registrado, mas será complementado com análise não-paramétrica na próxima subseção.

Homogeneidade de Variâncias

```
# Fligner-Killeen (robusto à não normalidade) entre semestres, por sexo
fligner.test(IMC ~ semestre, data = subset(dados, Gender == "M"))
```

```
##
## Fligner-Killeen test of homogeneity of variances
##
## data:  IMC by semestre
## Fligner-Killeen:med chi-squared = 0.082824, df = 1, p-value = 0.7735
```

```
fligner.test(IMC ~ semestre, data = subset(dados, Gender == "F"))
```

```
##
## Fligner-Killeen test of homogeneity of variances
##
## data:  IMC by semestre
## Fligner-Killeen:med chi-squared = 0.71101, df = 1, p-value = 0.3991
```

Resultados (Fligner–Killeen):

- Feminino: $p = 0.7735$
- Masculino: $p = 0.3991$

Boa! O problema é que alguns caracteres como $-$ (traço longo), \pm , μ e afins não estão sendo renderizados no LaTeX. A solução é escrever tudo no formato que o LaTeX entende: $\backslash pm$, $\backslash mu$, intervalos com colchetes [...] e traços simples.

Aqui está sua seção reescrita com código **antes** dos resultados e todo o texto em formato compatível com LaTeX:

Teste de Hipóteses

Nesta etapa, comparamos as médias de IMC entre os semestres **2016-2** e **2017-2**, separadamente para homens e mulheres. O teste utilizado foi o **t de Welch para duas amostras independentes**. O nível de significância adotado é $\alpha = 0,05$.

```
# Homens
t.test(IMC ~ semestre, data = subset(dados, Gender == "M"),
       var.equal = FALSE, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  IMC by semestre
## t = 0.53979, df = 38.057, p-value = 0.5925
## alternative hypothesis: true difference in means between group 2016-2 and group 2017-2 is not equal to 0
## 95 percent confidence interval:
##  -1.788823  3.089716
## sample estimates:
## mean in group 2016-2 mean in group 2017-2
##           24.93595           24.28551

# Mulheres
t.test(IMC ~ semestre, data = subset(dados, Gender == "F"),
       var.equal = FALSE, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  IMC by semestre
## t = 2.17, df = 8.5966, p-value = 0.0595
## alternative hypothesis: true difference in means between group 2016-2 and group 2017-2 is not equal to 0
## 95 percent confidence interval:
##  -0.1318507  5.4075232
## sample estimates:
## mean in group 2016-2 mean in group 2017-2
##           21.08443           18.44660
```

Resultados:

- **Homens:** $p = 0.5925$, intervalo de confiança 95% = $[-1.79 ; 3.09]$. Não houve evidência de diferença significativa entre os semestres. As médias foram 24.94 (2016-2) e 24.29 (2017-2).
- **Mulheres:** $p = 0.0595$, intervalo de confiança 95% = $[-0.14 ; 5.41]$. A diferença não atingiu significância ao nível de 5%, mas ficou próxima do limiar. As médias foram 21.08 (2016-2) e 18.45 (2017-2).

Tamanho do Efeito (Cohen's d)

Além do p-valor, avaliamos a magnitude da diferença usando **Cohen's d** (Hedges' g com correção para amostras pequenas).

```
library(effsize)

# Homens
cohen.d(IMC ~ semestre, data = subset(dados, Gender == "M"),
        hedges.correction = TRUE)

##
## Hedges's g
##
## g estimate: 0.16344 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.4495299  0.7764100

# Mulheres
cohen.d(IMC ~ semestre, data = subset(dados, Gender == "F"),
        hedges.correction = TRUE)

##
## Hedges's g
##
## g estimate: 1.106459 (large)
## 95 percent confidence interval:
##      lower      upper
## -0.2786636  2.4915825
```

Resultados:

- **Homens:** $g = 0.16$ (intervalo $[-0.45 ; 0.78]$) → efeito desprezível.
- **Mulheres:** $g = 1.11$ (intervalo $[-0.28 ; 2.49]$) → efeito grande, mas com intervalo de confiança muito amplo devido ao tamanho amostral reduzido.

V. Discussão e Conclusões

Espaço reservado para:

- Interpretação dos resultados;
- Conclusões sobre diferenças de IMC entre semestres e sexos;
- Limitações do estudo.

VI. Referências

Serão incluídas referências às notas de aula e bibliografia principal.