

# EEE933 - Planejamento e Análise de Experimentos

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

Bernardo Bacha\*

Marília Melo†

Gustavo Reis‡

19 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:**  $\text{IMC} = \text{Peso} / \text{Altura}^2$ .
- **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

#### a. 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")

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

---

\*PPGEE/UFMG — bernardobr@ufmg.br

†PPGEE/UFMG — mariliamacedomelo@email.com

‡PPGEE/UFMG — autor3@email.com

```

    IMC = Weight.kg / (Height.m^2))

df_2016 <- df_2016 %>%
  filter(Course == "PPGEE") %>%
  mutate(semester = "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$semester, dados$Gender)

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

```

## b. 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.

```

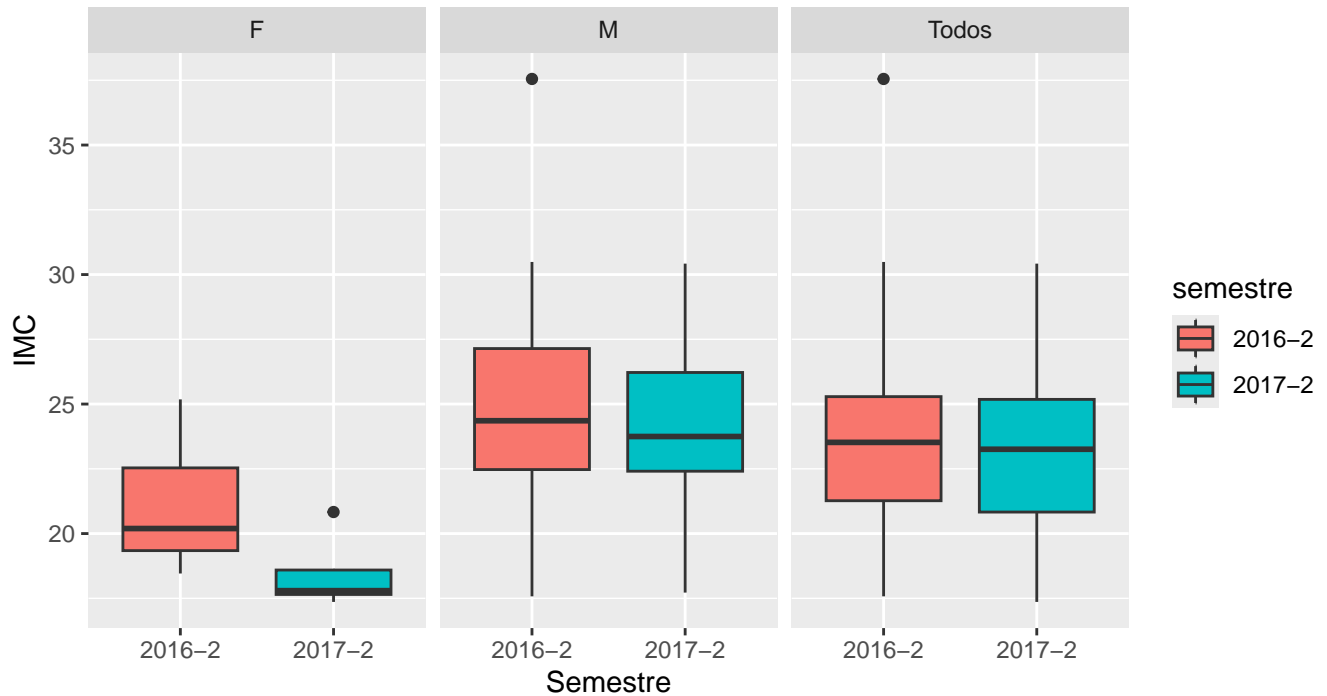
library(ggplot2)

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

# Boxplot
ggplot(dados_global, aes(x = semester, y = IMC, fill = semester)) +
  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)

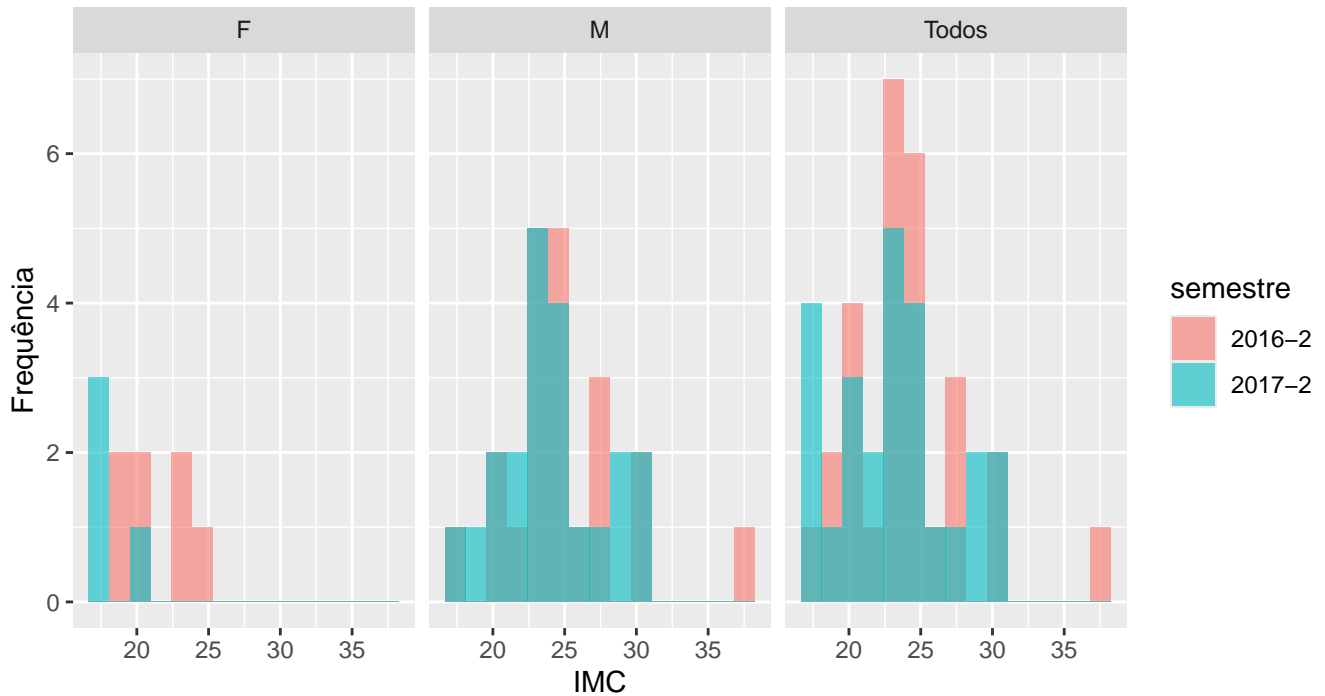


## Histogramas

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

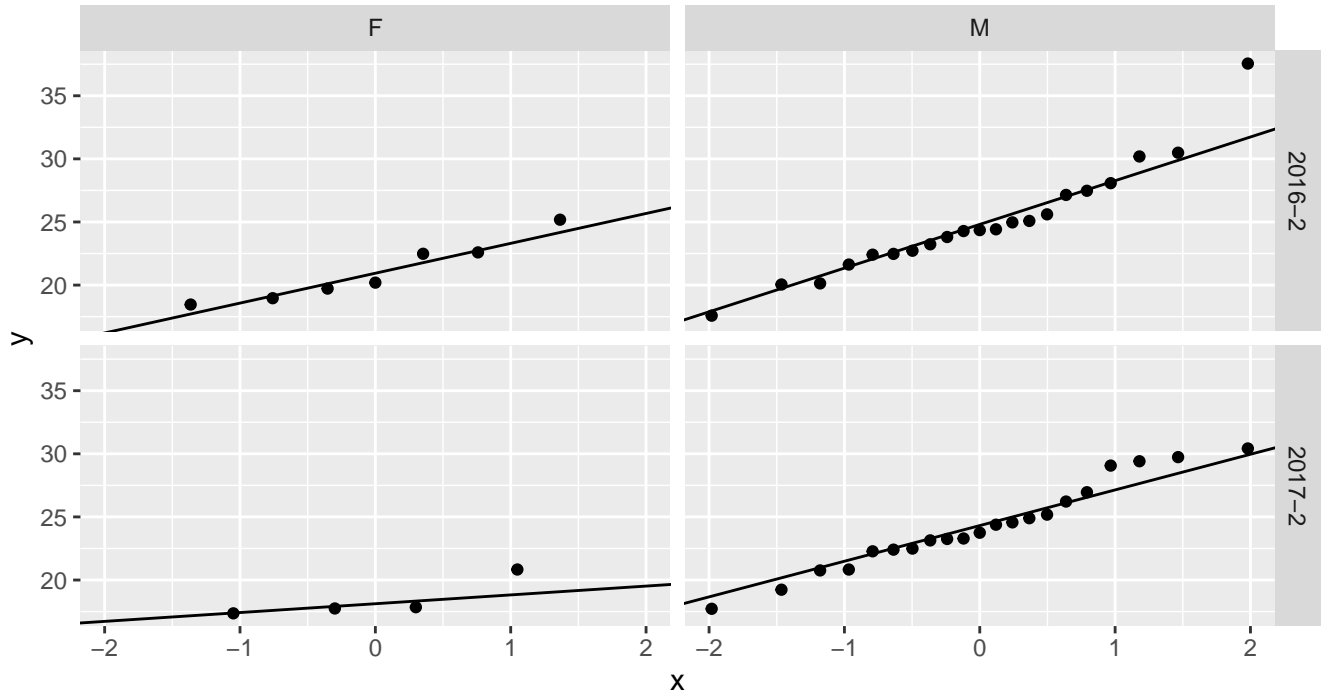
### Validação das Premissas

#### a. Normalidade - Teste de Shapiro-Wilk

Para verificar se os dados seguem uma distribuição normal, combinamos duas abordagens. Primeiro, fizemos uma análise visual com o gráfico QQ-plot para ver se os pontos se distribuíam como o esperado. Em seguida, para uma confirmação estatística, aplicamos o teste de Shapiro-Wilk, que é um método rigoroso e muito recomendado para amostras pequenas.

```
# 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. Portanto, embora o teste de Shapiro-Wilk sugira uma violação da premissa de normalidade, esse resultado deve ser interpretado com cautela devido à baixa amostra. Essa sensibilidade do teste com  $n=4$  reforça a necessidade de complementar a análise com abordagens não-paramétricas ou testes robustos como o de Welch, que é menos sensível a desvios da normalidade.

#### b. Homogeneidade de Variâncias - Teste de Fligner-Killeen

Para avaliar a premissa de homogeneidade de variâncias, foi utilizado o teste de Fligner-Killeen. A escolha deste teste se justifica por sua robustez a possíveis desvios da normalidade, uma consideração importante, visto que a normalidade da amostra do grupo feminino foi questionada na análise anterior. A hipótese nula do teste é de que as variâncias do IMC são iguais entre os semestres.

```
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:  $p = 0.3991$

Os resultados não indicaram evidências para rejeitar essa hipótese, com um p-valor de 0.7735 para o grupo masculino e 0.3991 para o feminino. Portanto, conclui-se que as variâncias podem ser consideradas homogêneas entre os semestres para ambos os sexos.

#### c. Teste de Hipóteses - Testes t - Welch e Student

Para comparar as médias de IMC entre os semestres **2016-2** e **2017-2**, separadamente para homens e mulheres, o método principal adotado foi o **teste t de Welch para duas amostras independentes**, com um nível de significância de  $\alpha = 0,05$ . A preferência por este teste segue as recomendações da literatura estatística para garantir a robustez da análise. Conforme demonstrado em estudos clássicos como o de Moser e Stevens (1992), a prática de realizar um teste preliminar de variâncias para então decidir qual teste t utilizar pode aumentar a taxa de erro do tipo I. Desta forma, o teste de Welch se apresenta como uma alternativa mais segura. Ainda assim, para fins de uma análise comparativa completa, o teste t de Student (`var.equal = TRUE`) também foi executado para que os resultados de ambos os métodos pudessem ser avaliados.

```

# --- Grupo Masculino ---
print("--- Homens: Teste de Welch (var.equal = FALSE) ---")

## [1] "--- Homens: Teste de Welch (var.equal = FALSE) ---"
t.test(IMC ~ semestre, data = subset(dados, Gender == "M"), var.equal = FALSE)

##
## 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
print("--- Homens: Teste de Student (var.equal = TRUE) ---")

## [1] "--- Homens: Teste de Student (var.equal = TRUE) ---"
t.test(IMC ~ semestre, data = subset(dados, Gender == "M"), var.equal = TRUE)

##
## Two Sample t-test
##
## data: IMC by semestre
## t = 0.53979, df = 40, p-value = 0.5923
## alternative hypothesis: true difference in means between group 2016-2 and group 2017-2 is not equal to 0
## 95 percent confidence interval:
## -1.784943 3.085836
## sample estimates:
## mean in group 2016-2 mean in group 2017-2
## 24.93595 24.28551

# --- Grupo Feminino ---
print("--- Mulheres: Teste de Welch (var.equal = FALSE) ---")

## [1] "--- Mulheres: Teste de Welch (var.equal = FALSE) ---"
t.test(IMC ~ semestre, data = subset(dados, Gender == "F"), var.equal = FALSE)

##
## 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
print("--- Mulheres: Teste de Student (var.equal = TRUE) ---")

## [1] "--- Mulheres: Teste de Student (var.equal = TRUE) ---"

```

```
t.test(IMC ~ semestre, data = subset(dados, Gender == "F"), var.equal = TRUE)

##
## Two Sample t-test
##
## data: IMC by semestre
## t = 1.9308, df = 9, p-value = 0.08556
## alternative hypothesis: true difference in means between group 2016-2 and group 2017-2 is not equal to 0
## 95 percent confidence interval:
## -0.4527037 5.7283762
## sample estimates:
## mean in group 2016-2 mean in group 2017-2
## 21.08443 18.44660
```

### Resultados - Testes de hipóteses:

A análise dos testes de hipóteses não apontou diferença estatisticamente significativa no IMC médio para nenhum dos grupos ao nível de  $\alpha = 0,05$ . Para o grupo masculino, os resultados do teste de Welch ( $p = 0.5925$ ) e do teste de Student ( $p = 0.5923$ ) foram praticamente idênticos, não indicando alteração entre os semestres. Já para o grupo feminino, os p-valores divergiram: o teste de Welch apresentou um resultado marginalmente significativo ( $p = 0.0595$ ), mais próximo do limiar de significância que o do teste de Student ( $p = 0.08556$ ). Essa comparação reforça a robustez do teste de Welch que, apesar de não atingir a significância formal, capturou uma tendência mais forte nos dados do grupo feminino.

## d. Tamanho do Efeito e Poder Estatístico

Após a análise de hipóteses, que avalia a significância estatística, a investigação foi aprofundada em duas frentes para uma interpretação completa dos resultados. A primeira foi quantificar a magnitude prática da diferença observada, o que é feito através do cálculo do **tamanho do efeito** (utilizando o g de Hedges, por ser mais preciso para amostras pequenas conforme demonstrado por Hedges e Olkin (1985)). A segunda foi avaliar a capacidade do estudo em detectar um efeito, caso ele realmente exista, o que é medido pela análise de **poder estatístico**. A análise conjunta dessas duas métricas é o que permite contextualizar os p-valores e entender cenários complexos, como um efeito de grande magnitude que não atinge significância estatística devido ao baixo poder do teste.

```
# Carregar pacotes necessários
# Se não tiver instalado, rode no console: install.packages(c("effsize", "pwr"))
library(effsize)
library(pwr)

# ---- Tamanho do Efeito (g de Hedges) ----

print("--- Tamanho do Efeito | Homens ---")

## [1] "--- Tamanho do Efeito | 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

print("--- Tamanho do Efeito | Mulheres ---")
```



```
## [1] "--- Tamanho do Efeito | 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
# ---- Análise de Poder Estatístico ----

# Poder para o grupo MASCULINO (n1=21, n2=21, d=0.16)
print("--- Poder Estatístico | Homens ---")

## [1] "--- Poder Estatístico | Homens ---"
pwr.t2n.test(n1 = 21, n2 = 21, d = 0.16, sig.level = 0.05, alternative = "two.sided")

##
##      t test power calculation
##
##          n1 = 21
##          n2 = 21
##          d = 0.16
##      sig.level = 0.05
##          power = 0.07982448
##      alternative = two.sided
# Poder para o grupo FEMININO (n1=7, n2=4, d=1.11)
print("--- Poder Estatístico | Mulheres ---")

## [1] "--- Poder Estatístico | Mulheres ---"
pwr.t2n.test(n1 = 7, n2 = 4, d = 1.11, sig.level = 0.05, alternative = "two.sided")

##
##      t test power calculation
##
##          n1 = 7
##          n2 = 4
##          d = 1.11
##      sig.level = 0.05
##          power = 0.3534739
##      alternative = two.sided
# ---- Cálculo de amostra para potência de 80%

power.t.test(power = 0.8,      # potência desejada
             delta = 0.5,      # tamanho do efeito esperado (diferença entre médias)
             sd = 1,           # desvio padrão estimado
             sig.level = 0.05,
             type = "two.sample", # tipo do teste
             alternative = "two.sided") # bilateral

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

```
##          n = 63.76576
##          delta = 0.5
##          sd = 1
##          sig.level = 0.05
##          power = 0.8
##          alternative = two.sided
##
## NOTE: n is number in *each* group
```

### Resultados - Tamanho do Efeito e Poder Estatístico:

A análise do tamanho do efeito e do poder estatístico revela cenários distintos para os dois grupos. Para o grupo masculino, foi observado um efeito de magnitude desprezível ( $g = 0.16$ ), com um poder estatístico de apenas 8% para detectá-lo, o que é consistente com o resultado não-significativo do teste t. Em contraste, para o grupo feminino, a magnitude do efeito foi classificada como grande ( $g = 1.11$ ). No entanto, a análise de poder indicou que o estudo tinha apenas 35% de chance de detectar um efeito dessa magnitude como estatisticamente significativo. Isso explica o resultado marginal do teste t ( $p = 0.0595$ ): embora a diferença observada seja grande na prática, o estudo não teve poder suficiente para confirmá-la estatisticamente devido à amostra reduzida.

## V. Discussão e Conclusões

- Tanto para amostras menores (gênero feminino), quanto para amostras maiores (gênero masculino), os testes apresentaram baixa potência, inferior a 80%, o que sugere que seriam fracos para detectar diferenças reais e consequentemente possuem baixa confiabilidade de rejeitar corretamente a hipótese nula. Para testes de potência superior a 80%, cada grupo deveria possuir ao menos 64 amostras.
- Conclusões sobre diferenças de IMC entre semestres e sexos;
- Limitações do estudo.

## VI. Referências

MOSER, B. K.; STEVENS, G. R. Homogeneity of Variance in the Two-Sample Means Test. **The American Statistician**, v. 46, n. 1, p. 19-22, fev. 1992.

HEDGES, Larry V.; OLKIN, Ingram. **Statistical methods for meta-analysis**. Orlando: Academic Press, 1985.