

# Estudo de Caso 02: Desempenho de uma nova versão de Software

*André Boechat(Checker), Mateus Pongelupe(Coordinator), Samuel Leite(Recorder)*

*17 de Outubro de 2018*

## Resumo - finalizar

Este relatório é o segundo dos Estudos de Casos na disciplina de Planejamento e Análise de Experimentos. O problema inicial consiste em duas populações, sendo elas: uma turma que cursou uma disciplina de Engenharia Elétrica em 2016 e outra que cursou a mesma disciplina em 2017. O teste feito girou em torno do IMC dessas populações, tentando de alguma forma testar o estilo de vida dos alunos com base nesse índice.

**INSERIR AQUI COMENTÁRIOS SOBRE O RESULTADO DO EXPERIMENTO**

## Planejamento do Experimento

Nesse experimento, está sendo avaliado o IMC de duas populações diferentes. Cada população é um grupo de alunos que cursou uma disciplina de Engenharia Elétrica. Deseja-se comparar os IMC's dessas duas populações.

O Índice de Massa Corporal, ou IMC, é um valor calculado à partir da altura e massa de cada indivíduo, procurando estimar se sua altura e peso estão de acordo com padrões saudáveis. Porém, existem vários estudos mostrando que não é possível relacionar diretamente o IMC de uma pessoa com estar dentro ou fora do peso, uma vez que esse cálculo não leva em consideração fatores como alimentação ou percentual de gordura. O cálculo de IMC é dado pela equação:

$$IMC = \frac{w}{h^2}$$

em que  $w$  equivale ao peso do indivíduo em **kg** e  $h$  à sua altura em **m**.

O experimento realizado consiste na comparação das médias dos IMCs de cada uma das populações, separando-as entre seus gêneros, masculino e feminino. Dessa forma, serão realizados dois testes: entre as populações feminino e masculino de cada semestre dos alunos de Engenharia Elétrica.

## Coleta de Dados e Tratamento de Dados - em andamento

Os dados foram coletados à partir de dois arquivos enviados. Nesses arquivos, estão contidos os valores para a altura e a massa de cada aluno. Os arquivos devem ser lidos em uma variável e então separados de acordo com o seu gênero e descartados de acordo com sua turma, caso aplicável. Esse último caso se refere à parcela de alunos na turma 2016/2 que não fazem parte do curso de Engenharia Elétrica.

```
input_file1 <- '20162.csv'
input_file2 <- '20171.csv'

# Importing files
turma1<- read.csv(file = input_file1, header = T)
turma2<- read.csv(file = input_file2, header = T)

# Filtering PPGE students
```

```

turma1.ele <- turma1[ turma1$Course == 'PPGEE', c('Gender', 'Height', 'Weight')]

# Separating values on genders
turma1.f <- turma1.ele[ turma1.ele$Gender == 'F', c('Height', 'Weight')]
turma1.m <- turma1.ele[ turma1.ele$Gender == 'M', c('Height', 'Weight')]

turma2.f <- turma2[ turma2$Gender == 'F', c('Height', 'Weight')]
turma2.m <- turma2[ turma2$Gender == 'M', c('Height', 'Weight')]

```

Tendo selecionado os dados de interesse, isto é, apenas dos alunos de Engenharia Elétrica, os mesmos foram separados por gênero para efetuação das comparações. Em seguida, dado a fórmula do cálculo de IMC já descrito anteriormente, calcula-se o mesmo para as populações:

```

# Definition of IMC function
imc <- function(h, w) {
  return( w / ( h^2 ) )
}

applyIMC <- function(df) {
  df$IMC <- apply(df, 1, function (x) imc(x['Height'], x['Weight']))
  df$Height <- NULL
  df$Weight <- NULL
  return(df)
}

##Calculating IMCs of populations
turma1.f <- applyIMC(turma1.f)
turma1.m <- applyIMC(turma1.m)
turma2.f <- applyIMC(turma2.f)
turma2.m <- applyIMC(turma2.m)

turma1.f.var <- var(turma1.f$IMC)
turma2.f.var <- var(turma2.f$IMC)
turma1.m.var <- var(turma1.m$IMC)
turma2.m.var <- var(turma2.m$IMC)

cat("\n",
    "Populations' Variances:\n",
    "2016-MASC: ", turma1.m.var, "\n",
    "2017-MASC: ", turma2.m.var, "\n",
    "2016-FEM: ", turma1.f.var, "\n",
    "2017-FEM: ", turma2.f.var, "\n"
)

##
## Populations' Variances:
## 2016-MASC: 18.69141
## 2017-MASC: 11.80097
## 2016-FEM: 5.83963
## 2017-FEM: 2.573854

```

## Teste da média - em andamento

Para avaliar se o IMC das populações é o mesmo, serão realizados testes de hipótese para a população feminina e outro para a população masculina. Para ambos os casos, será realizado um teste em que a hipótese nula consiste na igualdade das médias das populações. A hipótese alternativa será bilateral, isto é, estamos interessados em detectar quaisquer diferenças entre as médias das populações.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

Para esse teste, definiu-se um nível de significância de  $\alpha = 0.01$ , um efeito de relevância mínimo  $\delta^* = e$  e uma potência desejada de  $\pi = 1 - \beta = 0.8$ . Os valores de média e variância de cada população serão obtidos através dos dados coletados.

```
h0.mean = 50
h0.sd = sqrt(100)

t1.alpha = 0.01
t1.delta = 4
t1.beta = 0.2
t1.power = 1 - t1.beta
```

CONTINUAR ? PARTIR DAQUI Assumindo que a hipótese nula  $H_0$  se comporte com uma distribuição de média populacional  $\mu = 50$  e variância populacional  $\sigma^2 = 100$ , pode-se calcular o número de amostras mínimo a partir do teste Z, haja vista que a variância da hipótese nula é conhecida. Para fazer esse cálculo, foi usado o pacote *asbio*:

```
library(asbio)

## Loading required package: tcltk

n <- power.z.test(power = t1.power, alpha = t1.alpha, effect = t1.delta,
                  sigma = h0.sd, test = "one.tail")$n
t1.N <- ceiling(n)
cat("N: ", t1.N)
```

```
## N: 63
```

Assim, fazendo uso do teste Z, precisaremos de uma amostra de tamanho  $N = 63$  para executar o nosso teste com uma potência  $\pi = 0.8$ . Por nossa hipótese  $H_1$  ser unidirecional, a região crítica do teste Z pode ser determinada como:

$$P(z_\alpha \leq Z_0 \mid H_0 \text{ seja verdadeira})$$

Isto é, para que a hipótese nula seja rejeitada com um nível de confiança de 99% é preciso que  $z_\alpha > Z_0$ .

## Análise de Dados exploratória

Antes de proceder com o teste de hipóteses, é interessante fazer uma análise exploratória dos dados. Dessa forma, foram gerados os gráficos Q-Q para as populações masculinas e femininas das turmas de 2016 e 2017.

```
par(mfrow=c(2,2), mai = c(0.5, 0.75, 0.5, 0.5))

library(car)
car::qqPlot(turma1.m$IMC ,
            pch=16,
            cex=1.5,
```

```

    las=1,
    main = '2016-MASC',
    ylab = 'IMC')

car::qqPlot(turma2.m$IMC,
  pch=16,
  cex=1.5,
  las=1,
  main = '2017-MASC',
  ylab = 'IMC')

car::qqPlot(turma1.f$IMC ,
  pch=16,
  cex=1.5,
  las=1,
  main = '2016-FEM',
  ylab = 'IMC')

car::qqPlot(turma2.f$IMC,
  pch=16,
  cex=1.5,
  las=1,
  main = '2017-FEM',
  ylab = 'IMC')

```

Observando os gráficos, verifica-se que há mais observações nas populações masculinas e que as mesmas tendem a respeitar uma distribuição próxima à normal, exceto por alguns efeitos de cauda. No caso das populações femininas, verifica-se a presença de poucas observações e uma fuga da normalidade para a população do ano de 2017. Isso é confirmado pelos histogramas de cada população, presentes a seguir.

```

library(cowplot, warn.conflicts = FALSE)

theme_set(theme_cowplot(font_size=12))

plot.hist1 <- ggplot(turma1.m, aes(x=IMC)) +
  geom_histogram(colour="black", fill="#ff9999", bins = 10) +
  background_grid(major = 'xy') +
  ggtitle('2016-MASC')

plot.hist2 <- ggplot(turma2.m, aes(x=IMC)) +
  geom_histogram(colour="black", fill="#ff9999", bins = 10) +
  background_grid(major = 'xy') +
  ggtitle('2017-MASC')

plot.hist3 <- ggplot(turma1.f, aes(x=IMC)) +
  geom_histogram(colour="black", fill="#ff9999", bins = 10) +
  background_grid(major = 'xy') +
  ggtitle('2016-FEM')

plot.hist4 <- ggplot(turma2.f, aes(x=IMC)) +
  geom_histogram(colour="black", fill="#ff9999", bins = 10) +
  background_grid(major = 'xy') +
  ggtitle('2017-FEM')

```

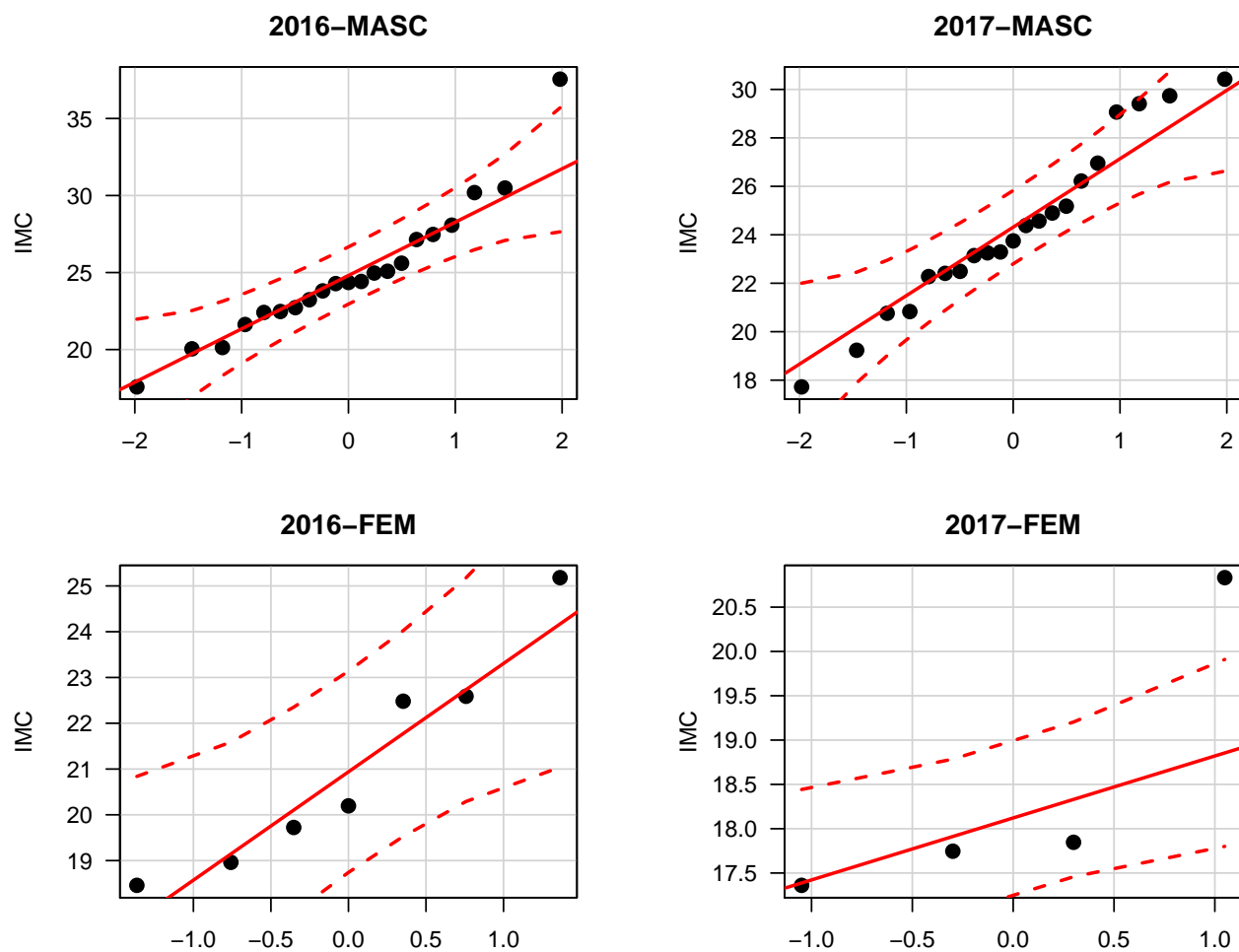


Figura 1: Comparação dos quantis das quatro populações com os quantis de uma distribuição normal

```
plot_grid(plot.hist1, plot.hist2, plot.hist3, plot.hist4, labels = c('A', 'B', 'C', 'D'), ncol = 2)
```

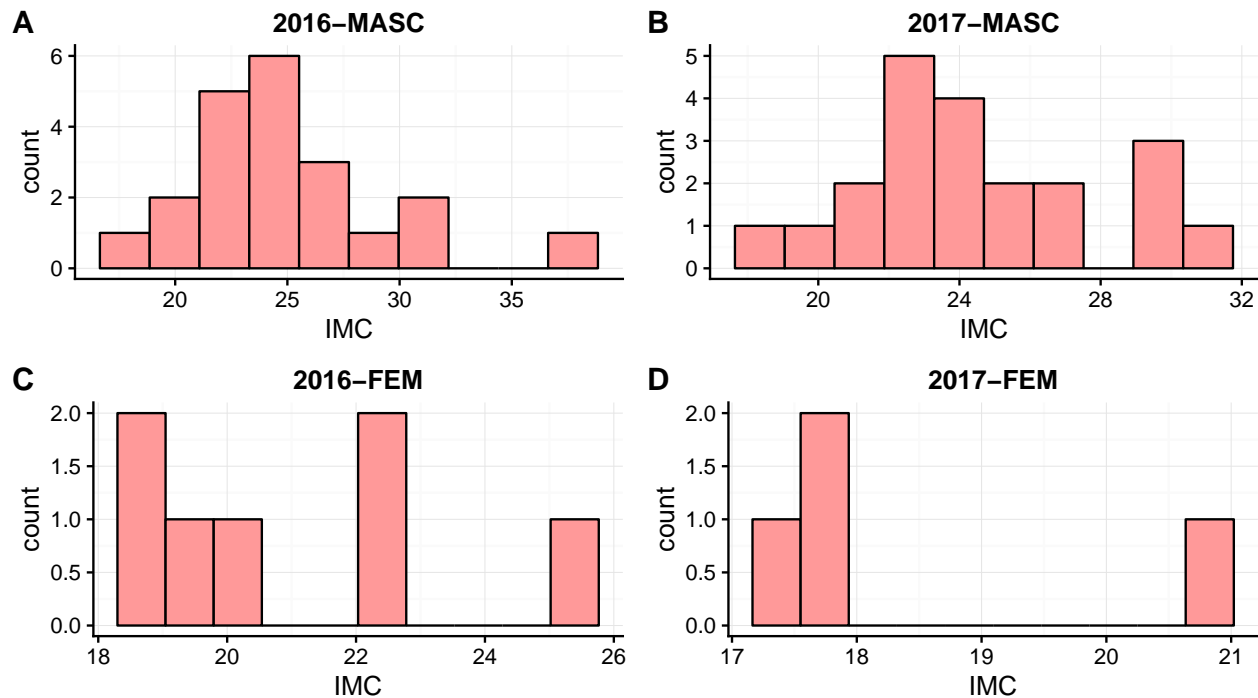


Figura 2: Histogramas para as populações avaliadas

Segue abaixo a codificação dessas funções, bem como um exemplo da chamada de *generate\_n\_samples* para  $N = 10$ :

## Análise Estatística

### Teste da Média

Dados os parâmetros definidos na seção *Planejamento do Experimento* para o teste da média, foram recolhidas  $N = 63$  amostras e o teste foi executado nas linhas abaixo. O intervalo de confiança também foi calculado, considerando uma distribuição normal cuja variância populacional  $\sigma^2 = 100$  é conhecida.

Como  $Z_\alpha < Z_0$ , conclui-se que não há evidências suficientes para rejeitar  $H_0$  a um nível de confiança de 99%.

### Teste da Variância

Com os dados coletados e armazenados na variável *t1.samples*, é possível verificar se o novo software irá gerar dados com uma variância menor ou maior que aquela resultada no processo original.

O teste foi executado conforme explicado na seção *Teste da Variância*.

Como o valor de  $P$  calculado é menor do que o valor de  $\alpha$ , é possível afirmar que a hipótese nula está negada e a variância do novo teste é, portanto, inferior à variância do processo original.

## Avaliando suposições do modelo

A validação das suposições de um experimento é um passo importante de uma análise de experimento. Não apenas permite verificá-las e como também identificar possíveis efeitos nos resultados encontrados, decorrentes de violações das premissas do planejamento experimental.

Ao fazer o teste da média, foi suposta uma distribuição normal das amostras. Para avaliar essa suposição, o teste de Shapiro-Wilk é uma boa alternativa. Trata-se de um teste de normalidade que assume uma hipótese nula de que a distribuição de um conjunto de dados é normal. O resultado do teste fornece um valor  $p$  que, se menor que o nível  $\alpha$  desejado, permite rejeitar a hipótese nula. Para o método padrão disponível no R, o valor de  $p < 0,05$  indica que não é uma distribuição normal.

Outro indicador interessante é o *qqplot* que é um gráfico em que se compara os quantis da distribuição das amostras aos quantis de uma distribuição normal. Ele fornece um bom indicativo do comportamento da distribuição das amostras em relação a uma normal, permitindo avaliar o quão próximo é de uma normal. Ambos indicadores foram calculados para as amostras colhidas, bem como um histograma e um gráfico de densidade.

Observando os resultados do teste de Shapiro-Wilk, verifica-se que  $p = 4,22 \times 10^{-6} < 0,05$ , isto é, o teste indica que a distribuição das amostras não segue uma distribuição normal. Isso também é observável no *qqplot*, em que é perceptível que os quantis da distribuição das amostras não estão próximos dos quantis normais em todo o intervalo. Contudo, uma boa parte dos quantis está em uma região quase normal, sendo que passa a fugir de um comportamento de uma normal quando o custo supera 60.

Os gráficos da figura seguinte, o histograma e o gráfico de densidade ajudam a ressaltar isso. No gráfico de densidade, percebe-se um comportamento próximo de uma normal até o custo atingir 60. A partir desse valor, a função de densidade apresenta dois picos que prejudicam bastante a premissa de normalidade.

## Conclusões e Recomendações

O estudo conduzido nesse trabalho mirou avaliar o desempenho de uma nova versão de um software em comparação a sua versão anterior, cujo custo de execução é bem representado por uma distribuição populacional de média  $\mu = 50$  e variância  $\sigma^2 = 100$ . Para tal, foram empregados métodos estatísticos provenientes das aulas da disciplina de Planejamento e Análise de Experimentos em ensaios acerca da média e da variância do custo de execução do novo software. No teste da média foi empregado o teste Z, haja vista a premissa que o comportamento da função era normal com variância semelhante ao da função comparada. Foi empregado o Teste de Hipóteses com base em uma distribuição  $\chi^2$  para analisar a alteração da variância no novo software.

Com relação ao teste da média, ele falhou em refutar a hipótese nula, isto é, ele foi incapaz de afirmar ao nível de significância de 99% que a nova versão do software possui um custo médio de execução mais baixo. Portanto, o teste executado não suporta a hipótese de que a nova versão do software possui um desempenho superior à versão anterior em termos da média do custo de execução. Apesar disso, a partir dos dados colhidos, foi estimado o intervalo de confiança para a média  $\mu_1$ , com um grau de confiança de 99%:  $\mu_1 \in [47.53, 54.04]$ . Quanto a esse intervalo, observa-se que o seu centro é um pouco superior a 50 e que os limites do intervalo batem, aproximadamente, com o efeito de relevância mínima que este teste buscou detectar.

Posteriormente ao teste, em um momento de análise das premissas, verificou-se que a distribuição das amostras não era normal. A violação dessa premissa explica um pouco como o resultado do experimento pode ter sido distorcido, talvez pelo uso de procedimentos não adequados para o caso. Entre esses procedimentos, pode-se citar o teste Z, em que foi considerada a variância populacional da versão anterior, e o procedimento de seleção da amostra/cálculo da média.

Tendo ciência disso, uma alternativa seria a execução do teste T, que considera a variância da amostra retirada. Outra alternativa seria um tratamento/descarte de amostras espúrias ou que estão nas “pontas” da distribuição das amostras, de forma a atenuar o efeito que essas observações têm na distribuição das amostras.

O Teste de Vari ncia foi realizado seguindo-se a premissa de que a distribui o segue uma tend ncia  $\chi^2$ . Apesar de o teste afirmar que a vari ncia da amostra pode ser considerada como inferior   vari ncia original  $\sigma^2 = 100$ , a premissa de que a distribui o amostral pode ser considerada como normal foi refutada. Portanto, a modelagem dessa distribui o em um padr o qui-quadr tico fica tamb m contestada. Ainda, os intervalos de confian a calculados est o bem distante do valor da vari ncia da amostra e esse valor   bem inferior ao valor original.

## Refer ncias

- R Man Pages - asbio package - <https://rdr.io/cran/asbio/man/power.z.test.html>
- R Man Pages - car package - <https://rdr.io/cran/car/man/qqPlot.html>
- Statistics R Tutorial - <https://www.cyclismo.org/tutorial/R/confidence.html>
- Montgomery, Douglas C. - Applied statistics and probabiliy for engineers (3  Edi o) - Cap tulos 8,9
- Notas de Aula - <https://github.com/fcampelo/Design-and-Analysis-of-Experiments>
- Notas - [https://edisciplinas.usp.br/pluginfile.php/2063723/mod\\_resource/content/0/Aula11-2016.pdf](https://edisciplinas.usp.br/pluginfile.php/2063723/mod_resource/content/0/Aula11-2016.pdf)