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 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 disciplica 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 levam em considerações 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.

```r
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 PPGEE students
```

```r
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:

```r
# 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
```

```r
turma1.f$year = '2016'
turma2.f$year = '2017'
t1.f <- rbind(turma1.f, turma2.f)
```

```r
turma1.m$year = '2016'
turma2.m$year = '2017'
t1.m <- rbind(turma1.m, turma2.m)
```

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.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \not\equiv 0 \end{cases}$$

Para esse teste, definiu-se um nível de significância de $\alpha = 0.05$, um efeito de relevância mínimo $\delta^* = 4$. Tendo em vista que os valores já foram coletados para ambos os testes, definiu-se $N_{m1} = N_{m2} = 21$ e $N_{f1} = 7; N_{f2} = 4$. Com isso, pode-se calcular a potência do teste para detecção do efeito de relevância mínimo:

```r
t1.alpha = 0.05
t1.delta = 4
nm <- 21
## desvio padrão da turma 1 é maior que da turma 2
sd <- sd(turma1.m$IMC)

power.t.test(n = nm, sd = sd(turma1.m$IMC), sig.level = t1.alpha, delta = t1.delta,
             type = "two.sample",
             alternative = "two.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 21
##           delta = 4
##              sd = 4.323356
##       sig.level = 0.05
##           power = 0.832817
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#t.test(t1.m$IMC ~ t1.m$year, alternative = 'two.sided', mu = 0, conf.level = 0.99)
```

Os valores de média e variância de cada população serão obtidos através dos dados coletados.

Assim, fazendo uso do teste de Welch, usando uma amostra de tamanho $N = 21$, teremos uma potência de aproximadamente 83% para detectar o efeito de relevância mínimo desejado. Por nossa hipótese H_1 ser bidirecional, a região crítica do teste T pode ser determinada como:

$$P(-t_{\alpha/2,n_1+n_2-2} \leq t_0 \leq t_{\alpha/2,n_1+n_2-2} \mid H_0 \text{ seja verdadeira})$$

Isto é, para que a hipótese nula seja rejeitada com um nível de confiança de 95%, é preciso que $-t_{\alpha/2,n_1+n_2-2} \leq t_0 \leq t_{\alpha/2,n_1+n_2-2}$.

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 próximo 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')
```

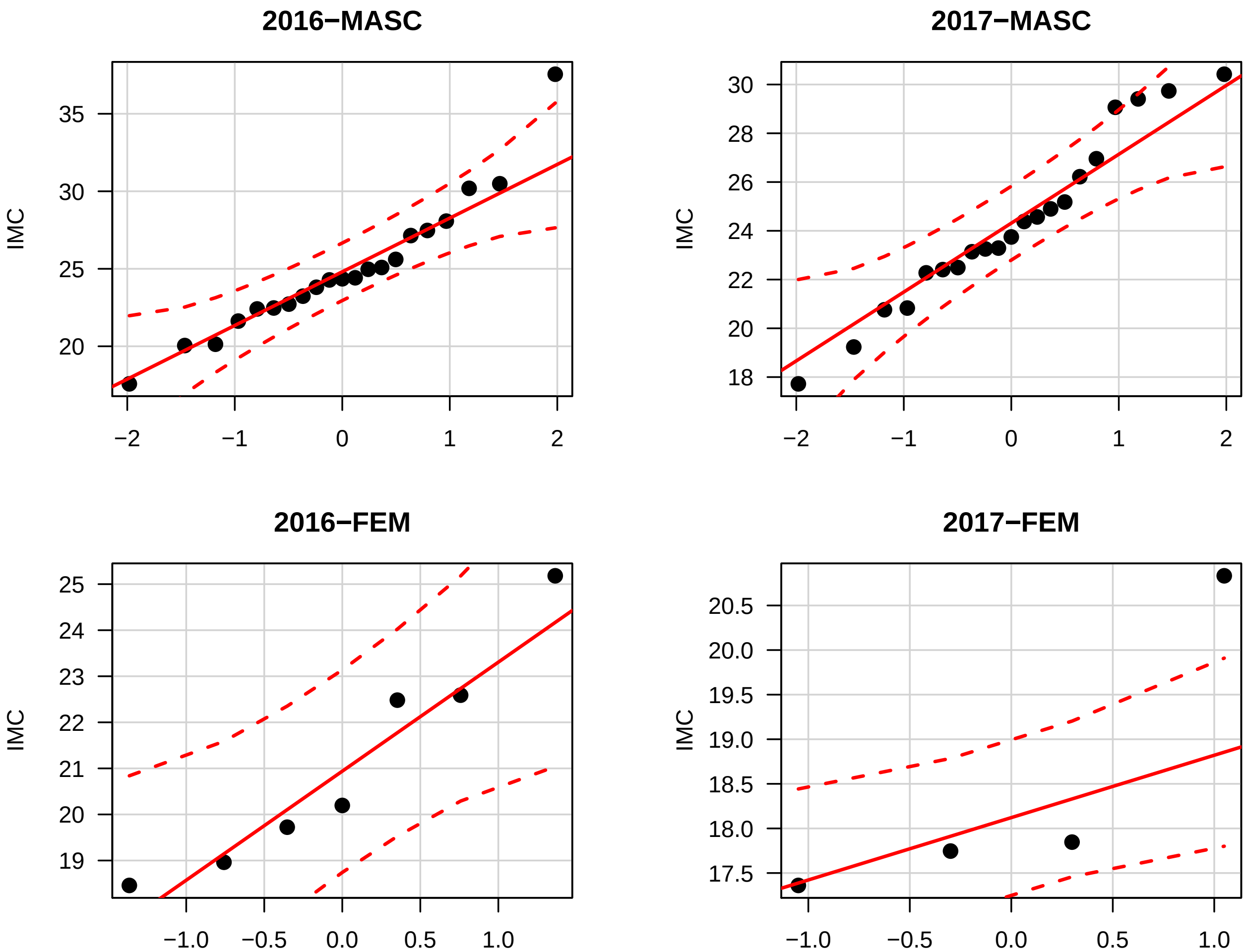

Figura 1: Comparação dos quantis das quatro populações com os quantis de uma distribuição normal

```r
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')

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

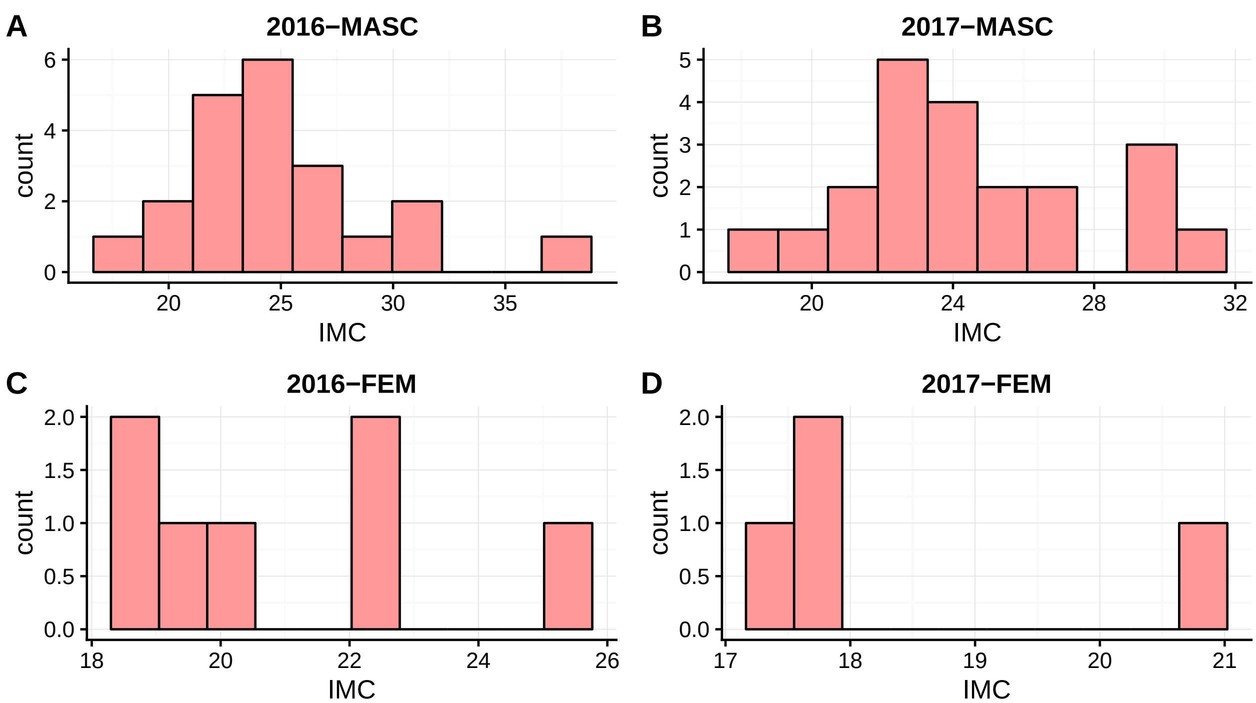

Figura 2: Histogramas para as populações avaliadas

Segue abaixo a codifica??o dessas fun??es, bem como um exemplo da chamada de *generate_n_samples* para $N = 10$:

An?lise Estat?stica

Teste da M?dia

Dados os par?metros definidos na se??o *Planejamento do Experimento* para o teste da m?dia, foram recolhidas $N = 63$ amostras e o teste foi executado nas linhas abaixo. O intervalo de confian?a tamb?m foi calculado, considerando uma distribui??o normal cuja vari?ncia populacional $\sigma^2 = 100$? conhecida.

Como $Z_\alpha < Z_0$, conclui-se que n?o h? evid?ncias suficientes para rejeitar H_0 a um n?vel de confian?a de 99%.

Teste da Vari?ncia

Com os dados coletados e armazenados na vari?vel *t1.samples*, ? poss?vel verificar se o novo software ir? gerar dados com uma vari?ncia menor ou maior que aquela resultada no processo original.

O teste foi executado conforme explicado na se??o *Teste da Vari?ncia*.

Como o valor de P calculado ? menor do que o valor de α, ? poss?vel afirmar que a hip?tese nula est? negada e a vari?ncia do novo teste ?, portanto, inferior ? vari?ncia do processo original.

Avaliando suposi??es do modelo

A valida??o das suposi??es de um experimento ? um passo importante de uma an?lise de experimento. N?o apenas permite verific?-las e como tamb?m identificar poss?veis efeitos nos resultados encontrados, decorrentes de viola??es das premissas do planejamento experimental.

Ao fazer o teste da m?dia, foi suposta uma distribui??o normal das amostras. Para avaliar essa suposi??o, o teste de Shapiro-Wilk ? uma boa alternativa. Trata-se de um teste de normalidade que assume uma hip?tese nula de que a distribui??o de um conjunto de dados ? normal. O resultado do teste fornece um valor p que, se menor que o n?vel α desejado, permite rejeitar a hip?tese nula. Para o m?todo padr?o dispon?vel no R, o valor de $p < 0,05$ indica que n?o ? uma distribui??o normal.

Outro indicador interessante ? o *qqplot* que ? um gr?fico em que se compara os quantis da distribui??o das amostras aos quantis de uma distribui??o normal. Ele fornece um bom indicativo do comportamento da distribui??o das amostras em rela??o a uma normal, permitindo avaliar o qu?o pr?ximo ? de uma normal. Ambos indicadores foram calculados para as amostras colhidas, bem como um histograma e um gr?fico de densidade.

Observando os resultados do teste de Saphiro-Wilk, verifica-se que $p = 4,22 \times 10^{-6} < 0,05$, isto ?, o teste indica que a distribui??o das amostras n?o segue uma distribui??o normal. Isso tamb?m ? observ?vel no *qqplot*, em que ? percept?vel que os quantis da distribui??o das amostras n?o est?o pr?ximos dos quantis normais em todo o intervalo. Contudo, uma boa parte dos quantis est? em uma regi?o quase normal, sendo que passa a fugir de um comportamento de uma normal quando o custo supera 60.

Os gr?ficos da figura seguinte, o histograma e o gr?fico de densidade ajudam a ressaltar isso. No gr?fico de densidade, percebe-se um comportamento pr?ximo de uma normal at? o custo atingir 60. A partir desse valor, a fun??o de densidade apresenta dois picos que prejudicam bastante a premissa de normalidade.

Conclus?es e Recomenda??es

O estudo conduzido nesse trabalho mirou avaliar o desempenho de uma nova vers?o de um software em compara??o a sua vers?o anterior, cujo custo de execu??o ? bem representado por uma distribui??o populacional de m?dia $\mu = 50$ e vari?ncia $\sigma^2 = 100$. Para tal, foram empregados m?todos estat?sticos provenientes das aulas da disciplina de Planejamento e An?lise de Experimentos em ensaios acerca da m?dia e da vari?ncia do custo de execu??o do novo software. No teste da m?dia foi empregado o teste Z, haja vista a premissa que o comportamento da fun??o era normal com vari?ncia semelhante ao da fun??o comparada. Foi empregado o Teste de Hip?teses com base em uma distribui??o χ^2 para analisar a altera??o da vari?ncia no novo software.

Com rela??o ao teste da m?dia, ele falhou em refutar a hip?tese nula, isto ?, ele foi incapaz de afirmar ao n?vel de signific?ncia de 99% que a nova vers?o do software possui um custo m?dio de execu??o mais baixo. Portanto, o teste executado n?o suporta a hip?tese de que a nova vers?o do software possui um desempenho superior ? vers?o anterior em termos da m?dia do custo de execu??o. Apesar disso, a partir dos dados colhidos, foi estimado o intervalo de confian?a para a m?dia μ_1, com um grau de confian?a de 99%: $\mu_1 \in [47.53, 54.04]$. Quanto a esse intervalo, observa-se que o seu centro ? um pouco superior a 50 e que os limites do intervalo batem, aproximadamente, com o efeito de relev?ncia m?nima que este teste buscou detectar.

Posteriormente ao teste, em um momento de an?lise das premissas, verificou-se que a distribui??o das amostras n?o era normal. A viola??o dessa premissa explica um pouco como o resultado do experimento pode ter sido distorcido, talvez pelo uso de procedimentos n?o adequados para o caso. Entre esses procedimentos, pode-se citar o teste Z, em que foi considerada a vari?ncia populacional da vers?o anterior, e o procedimento de sele??o da amostra/c?lculo da m?dia.

Tendo ci?ncia disso, uma alternativa seria a execu??o do teste T, que considera a vari?ncia da amostra retirada. Outra alternativa seria um tratamento/descarte de amostras esp?rias ou que est?o nas "pontas" da distribui??o das amostras, de forma a atenuar o efeito que essas observa??es t?m na distribui??o das amostras.

O Teste de Vari?ncia foi realizado seguindo-se a premissa de que a distribui??o segue uma tend?ncia χ^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://rdrr.io/cran/asbio/man/power.z.test.html
- R Man Pages - car package - https://rdrr.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