Estatística para Ciência de Dados
Resolução do trabalho 05

Eduardo Façanha Dutra
Giovanni Brígido

1 Enunciado

Objetivos: Analisar o teste de hipótese T-test, usando o R Saber reportar o resultado

1. Aplique o T-test para esta situação de Teste A/B usando o arquivo meet_file. Objetiva-se saber se existe diferença significativa no tempo de acesso que os usuários tiveram ao usar os dois Meet: Zoom e Hangout. O meet que tem mais tempo de acesso foi o mais preferido.

```
library(readr)
meet_file <- read_csv("Dados/meet-file.csv",
     col_types = cols(Subject = col_skip(),
         Meet = col_factor(levels = c("Zoom",
             "Hangout"))))


t.test(Tempo ~ Meet, data=meet_file, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Tempo by Meet
## t = -3.0889, df = 38, p-value = 0.003745
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -274.874  -57.226
## sample estimates:
##    mean in group Zoom mean in group Hangout
##                302.10                468.15
```

1.1. Qual a hipótese NULA?

H0 - Não existe diferença significatva entre as médias dos tempos de uso de cada plataforma

1.2. Qual a relação da H0 com a hipótese alternativa?

A hipótese alternativa pode ser considerada quando houver evidências suficientes para rejeitar estatísticamente a hipótese nula. Caso não haja evidências suficientes a hipótese nula pode ser aceita. No teste observa-se que o p-value de 0.003745 nos permite rejeitar a hipótese nula, portanto, há diferença significativa entre as médias. O hangout, por possuir a maior média, é a ferramenta preferida.

2. Aplique novamente o T-test para esta situação de Teste A/B, agora considerando a variável logtempo. Faça uma análise da diferença entre com o item 1.

```
meet_file$logTime = log(meet_file$Tempo)
t.test(logTime ~ Meet, data=meet_file, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  logTime by Meet
## t = -3.3121, df = 38, p-value = 0.002039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6276133 -0.1514416
## sample estimates:
##    mean in group Zoom mean in group Hangout
##              5.666118              6.055645
```

Com o p-value de 0.002039, menor do que o teste passado, ainda pode-se rejeitar a hipótese de que há diferença entre o logaritmo do tempo de uso das plataformas, portanto, a plataforma Hangout ainda parece ser a mais utilizada entre as duas.

3. Agora usando o arquivo Tempoporsite, com 600 entradas, conduza novamente o t-test no Time por Site. Assuma variâncias iguais. Tempo está em segundos. Objetiva-se saber se existe diferença significativa no tempo de acesso que os usuários tiveram ao usar os dois sites A e B. O meet que tem mais tempo de acesso foi o mais preferido.

```
 tempoporsite <- read_csv("Dados/tempoporsite.csv",
     col_types = cols(Subject = col_skip(),
         Site = col_factor(levels = c("A",
           "B"))))
t.test(Time ~ Site, data=tempoporsite, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Time by Site
## t = 2.2889, df = 598, p-value = 0.02243
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.807652 23.659015
## sample estimates:
## mean in group A mean in group B
##        360.2933        347.5600
```

3.1. Quantos graus de liberdade resultaram da t-test?

df= 598

3.2. Para o décimo milésimo mais próximo (quatro dígitos), qual foi o valor p para este teste?

p-value = 0.0224

3.3. Considere esta forma de descrever o resultado do t-test. Descreva o que ele quer dizer T(598)= 2.28,p<0.05

O valor entre parenteses (598) representa o número de graus de liberdade utilizados para obter a função densidade de probabilidade t-student para o teste aplicado.

O valor de 2.28 representa o valor t do teste, que deve ser comparado com a tabela T, juntamente aos graus de liberade, para se obter o valor p do teste.

O valor p<0.05 representa que o teste aplicado obteve um valor p menor que 0.05 e, dependendo do nível de rigor do pesquisador e do processo subjacente, pode-se rejeitar a hipótese nula do teste em questão.

3.4. Descreva desta mesma forma, o resultado do exemplo do item 1.

T(38)= -3.088, p-value < 0.05

3.5. Qual o site foi mais usado?

Para um valor p de 0.0224 pode-se rejeitar a hipótese nula e considerar que a amostra com maior média foi a do site mais utilizado, no caso o Site A, mesmo que a diferença seja pequena