Relatório - Avaliação Prática 3

Marco Túlio Alves de Barros

2022-12-03

Ex 1: Um estudo foi realizado para verificar uma possível diferença entre as intenções de compra entre dois grupos distintos. Ao final do estudo foi observado que no G1, dos 150 entrevistados, cerca de 45% tinham intenção de comprar; do G2, dos 230 entrevistados, cerca de 50% tinham intenção de compra. Com base nos resultados a seguir diga se existe diferença entre os grupos ao nível de 5%, defina e hipótese e conclua.

Definindo a Hipótese Nula (H0) como os dois grupos sendo iguais e a H1 representando a diferença entre os grupos podemos realizar o teste de proporção.

```r
nG1 = 0.45 * 150 # defindo a proporção correta do grupo 1
nG2 = 0.5 * 230 # defindo a proporção correta do grupo 2
intencao = c(nG1, nG2)
grupos = c(150, 230)
prop.test(intencao, grupos, correct=F)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  intencao out of grupos
## X-squared = 0.90931, df = 1, p-value = 0.3403
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.15253734  0.05253734
## sample estimates:
## prop 1 prop 2
##   0.45   0.50
```

Conclusão: Tendo em vista que o p-valor obtido foi de 0.3404, valor maior que a referência de 0.05, H0 NÃO deve ser rejeitada e portanto não temos evidências o suficiente para afirmar que existe diferença entre os grupos. Outro elemento que indica a não diferença dos grupos é a presença do 0 no intervalo de confinça obtido.

Ex 2: Um estudo foi realizado para verificar um determinado treinamento foi eficaz para isso foram feitas avaliações antes e depois de cada pessoa. Os resultados são apresentados no conjunto de dados treinamento.xlsx. Pode-se dizer ao nível de 5% de significância que o treinamento foi eficiente. Para resolver o exercício retire uma amostra de tamanho 15 utilizando o número de matrícula como semente.

```
require(readxl)
```

```
## Carregando pacotes exigidos: readxl
```

```
## Warning: package 'readxl' was built under R version 4.2.2
```

```
dados_brutos = read_excel("C:/Users/marco/OneDrive/Área de Trabalho/UEL/Statistics/Relatórios/DADOS/tre
set.seed(560105);
antes = sample(dados_brutos$antes, 15);
depois = c(149.72, 252.64, 201.1, 223.95, 196.39, 241.17, 307.3, 255.75, 288.61, 254.66, 227.89, 144.54
data.frame(as.double(antes), depois)
```

```
##    as.double.antes. depois
## 1            240.10 149.72
## 2            176.85 252.64
## 3            256.30 201.10
## 4            156.45 223.95
## 5            255.39 196.39
## 6            149.91 241.17
## 7            259.72 307.30
## 8            181.15 255.75
## 9            221.59 288.61
## 10           194.00 254.66
## 11           239.70 227.89
## 12           190.91 144.54
## 13           132.65 188.03
## 14           183.44 281.59
## 15           264.53 244.59
```

```
## Como são duas colunas correlacionadas, retira-se primeiro a amostra dos valores
## 'antes' e em seguida completo a coluna 'depois' com os valores adequados
## referentes a cada valor de 'antes'
```

```r
## realizando teste de normalidade para garantir o teste de hipótese
shapiro.test(diff(as.double(antes) - depois))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  diff(as.double(antes) - depois)
## W = 0.93806, p-value = 0.3941
```

No teste de normalidade de Shapiro p-valor foi maior que 0.05 e portanto H0 não será rejeitada e podemos afirmar que existe uma distribuição normal nos dados

```r
## realizando teste de hipótese para valor pareados
t.test(depois, as.double(antes), alternative='two.side', mu=0, paired=T)
```

```
##
##  Paired t-test
##
## data:  depois and as.double(antes)
## t = 1.4464, df = 14, p-value = 0.1701
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -11.43533  58.80066
## sample estimates:
## mean difference
##        23.68267
```

Nesse caso, H0 é afirmar que não o treinamento não surtiu efeito positivo no grupo. Dado um p-valor de 0.1701, chegamos aa conclusão de que a hipótese nula não deve ser rejeitada e portanto o treinamento não foi eficiente.

Ex 3: Um levantamento de dados foi realizado para verificar se o gasto médio com compras e um site era de 500 reais. Com base nos resultados do conjunto de dados gastos.xlsx faça o intervalo de confiança e em seguida verifique se os gastos são de 500 reais. Para resolver o exercício retire uma amostra de tamanho 20 utilizando o núumero de matrícula como semente.

```r
gastos_brutos = read_excel("C:/Users/marco/OneDrive/Área de Trabalho/UEL/Statistics/Relatórios/DADOS/ga
set.seed(560105)

amostra_gastos = as.double(sample(gastos_brutos$gastos, 20)); amostra_gastos
```

```
##  [1] 523.31 470.32 606.52 578.75 434.50 496.10 424.13 560.45 614.52 604.95
## [11] 612.38 483.51 578.06 426.73 524.02 451.22 524.54 394.54 641.62 499.12
```

```r
## garantindo distribuição normal dos dados
shapiro.test(amostra_gastos)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  amostra_gastos
## W = 0.95071, p-value = 0.378
```

```r
t.test(amostra_gastos, mu=500)
```

```
##
##  One Sample t-test
##
## data:  amostra_gastos
## t = 1.3462, df = 19, p-value = 0.1941
## alternative hypothesis: true mean is not equal to 500
## 95 percent confidence interval:
##  487.5363 557.3927
## sample estimates:
## mean of x
##  522.4645
```

Visto que o intervalo de confiança obtido na amostra foi de 487.5363 - 557.3927 e com uma média de 522.4645. Conclui-se, analisando o p-valor de 0.1941, para a hipótese nula em que a média de gastos seria de 500 reais não deve ser recusada. Tal decisão pode ser baseada pela presença do valor 500 na região crítica e pelo p-valor superior a 0.05

Ex 4: Com o interesse em verificar se dois grupos são similares entre si em relação ao tempo para executar uma tarefa foi realizado um levantamento de dados. Com base nos resultados do conjunto de dados tempo.xlsx verifique se existe diferença entre as variâncias e em seguida verifique se existe diferença entre os grupos. Para resolver o exercício retire uma amostra de tamanho 20 para o grupo 1 e uma amostra de tamanho 30 para grupo 2, utilizando o número de matrícula como semente.

```r
tempo = read_excel("C:/Users/marco/OneDrive/Área de Trabalho/UEL/Statistics/Relatórios/DADOS/tempo.xlsx
set.seed(560105)

grupo1 = as.double(sample(tempo$Grupo1, 20))
```

```r
grupo2 = as.double(sample(tempo$Grupo2, 30))

## realizando testes de shapiro para garantir distribuição normal nos dados
shapiro.test(grupo1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  grupo1
## W = 0.89485, p-value = 0.03306
```

```r
shapiro.test(grupo2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  grupo2
## W = 0.94339, p-value = 0.1122
```

```r
#var1 = var(grupo1); var1
#var2 = var(grupo2); var2
#f = var2 / var1; f
#var.test(grupo1, grupo2)
#t.test(grupo1, grupo2, var.equal=T)
```

Apesar dos testes de variância, pelo teste de normalidade de Shapiro obtemos p-valor inferior a 0.05 para os dados do grupo 1 e portanto não faz sentido realizar testes de hipótese para dados que não seguem uma distribuição normal.

Ex 5: Com o interessem em verificar se dois grupos são similares entre si em relação ao gasto médio com compras foi realizado um levantamento de dados. Com base nos resultados do conjunto de dados compras2.xlsx verifique se

existe diferença entre as variâncias e em seguida verifique se existe diferença entre os grupos. Para resolver o exercício retire uma amostra de tamanho 15 para o grupo 1 e uma amostra de tamanho 15 para grupo 2, utilizando o número de matrícula como semente.

```r
gastos_brutos2 = read_excel("C:/Users/marco/OneDrive/Área de Trabalho/UEL/Statistics/Relatórios/DADOS/g
set.seed(560105)

gastos_g1 = as.double(sample(gastos_brutos2$Grupo1, 15))
gastos_g2 = as.double(sample(gastos_brutos2$Grupo2, 15))

## realizando testes de shapiro para garantir distribuição normal nos dados
shapiro.test(gastos_g1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  gastos_g1
## W = 0.879, p-value = 0.04585
```

```
shapiro.test(gastos_g2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  gastos_g2
## W = 0.96301, p-value = 0.7445
```

```
var_g1 = var(gastos_g1); var_g1
```

```
## [1] 3227.329
```

```
var_g2 = var(gastos_g2); var_g2
```

```
## [1] 2519.479
```

```
f = var_g1 / var_g2; f
```

```
## [1] 1.280951
```

```
var.test(gastos_g1, gastos_g2)
```

```
##
##  F test to compare two variances
##
## data:  gastos_g1 and gastos_g2
## F = 1.281, num df = 14, denom df = 14, p-value = 0.6495
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4300532 3.8154250
## sample estimates:
## ratio of variances
##           1.280951
```

Realizando o teste das variâncias para garantir que são homogêneas chegamos a um p-valor de 0.4385. Dessa forma, concluimos que as variâncias são homogêneas e os dados apresentam distribuição normal (devido ao teste de Shapiro).

```
t.test(gastos_g1, gastos_g2, var.equal=T)
```

```
##
##  Two Sample t-test
##
## data:  gastos_g1 and gastos_g2
## t = -0.81301, df = 28, p-value = 0.4231
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -56.00776  24.18109
## sample estimates:
## mean of x mean of y
##  563.1100  579.0233
```

Em suma, pelo teste de hipótese com variâncias homogêneas, obtemos p-valor de 0.3291 em um intervalo de confiança de -570.42 - 19.79 e concluimos que não há evidências o suficiente para afirmar que os gastos dos grupos são diferentes, afinal o p-valor é superior a 0.05 não rejeito H0 e também pela presença de 0 na região crítica.