Prova Métodos não Paramétricos

Márcio Alves

2022-12-20

Introdução

O objetivo deste documento é desenvolver algoritmos com a linguagem R para a solução da prova da disciplina Métodos não Paramétricos.

Questão 1

Nesta questão vamos trabalhar com o teste χ^2_{k-1} para avaliar se as frequências esperadas são iguais às observadas. Vamos construir as hipóteses do teste.

$$H_0: \textit{O modelo genético proposto é adequado para essa população.}$$

$$H_1: \textit{O modelo genético proposto não é adequado para essa população.}$$

Vamos testar com $\alpha = 5$ com $k = 3 - 1$ graus de liberdade. A estatística de teste é constrída da seguinte forma:

$$q = \frac{\sum_{i=1}^{3}(F_{oi} - F_{ei})}{F_{ei}}$$

Vamos construir o teste no código abaixo.

```
F.obs  = c(26,45,29)
v.prob = c(1/4,1/2,1/4)
chisq.test(F.obs, p = v.prob)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  F.obs
## X-squared = 1.18, df = 2, p-value = 0.5543
```

```
qchisq(.05, df=1, lower.tail = FALSE)
```

```
## [1] 3.841459
```

Conclusão: Como a estatística de teste foi $q = 1,18$ e a estatística da tabela foi de $Q = 3,84$, então não se rejeita H_0. Portanto, ao nível de significância $\alpha = 0,05$ o modelo genético proposto é adequado para essa população.

Questão 2

Nesta questão vamos testar normalidade utilizando dois testes: Lilliefors e Shapiro-Wilk.

a) Teste Lilliefors. A estatística do teste é calculada calculando a maior frequência observada da diferença em módulo entre a $F(x)$ e $G(x)$ ou $G(x)^0$. Vamos calcular a média e variância dos dados para construir as hipóteses

```
dados <- sort(c(177, 186, 158, 169, 173, 172, 171, 176, 170, 180))
dados
```

```
##  [1] 158 169 170 171 172 173 176 177 180 186
```

```
media <- mean(dados)
media # média
```

```
## [1] 173.2
```

```
variancia <- var(dados)
variancia # variância
```

```
## [1] 55.28889
```

```
dp <- sd(dados)
dp # desvio padrão
```

```
## [1] 7.43565
```

Agora vamos montar as hipóteses do teste de Lilliefors:

$$H_0 : Os\ dados\ seguem\ uma\ distribuição\ normal\ com\ \bar{X} = 173,2\ e\ s = 7,44.$$

$$H_1 : Os\ dados\ não\ seguem\ uma\ distribuição\ normal.$$

Calculando a estatística de teste.

```
dados <- sort(c(177, 186, 158, 169, 173, 172, 171, 176, 170, 180))
dados
```

```
##  [1] 158 169 170 171 172 173 176 177 180 186
```

```
media <- mean(dados)
media # média
```

```
## [1] 173.2
```

```r
variancia <- var(dados)
variancia # variância
```

```
## [1] 55.28889
```

```r
dp <- sd(dados)
dp # desvio padrão
```

```
## [1] 7.43565
```

```r
z = (dados - mean(dados))/sd(dados)

z
```

```
##  [1] -2.04420599 -0.56484639 -0.43035916 -0.29587192 -0.16138468 -0.02689745
##  [7]  0.37656426  0.51105150  0.91451321  1.72143662
```

```r
Fx = round(pnorm(z, mean = 0, sd = 1, lower.tail = TRUE), 4)
Fx
```

```
##  [1] 0.0205 0.2861 0.3335 0.3837 0.4359 0.4893 0.6468 0.6953 0.8198 0.9574
```

```r
Gx = seq(0.1,1,0.1)
Gx
```

```
##  [1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

```r
Gx_ = seq(0,0.9,0.1)
Gx_
```

```
##  [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
```

```r
abs(Fx-Gx)
```

```
##  [1] 0.0795 0.0861 0.0335 0.0163 0.0641 0.1107 0.0532 0.1047 0.0802 0.0426
```

```r
abs(Fx-Gx_)
```

```
##  [1] 0.0205 0.1861 0.1335 0.0837 0.0359 0.0107 0.0468 0.0047 0.0198 0.0574
```

```r
max(abs(Fx-Gx),abs(Fx-Gx_))
```

```
## [1] 0.1861
```

Conclusão: Na tabela, com $\alpha = 5$, o valor de $d_i = 0,258$. Como o máximo foi de 0,1861 não se rejeita H_0. Neste caso, ao nível de 5%, os dados seguem uma distribuição normal com $\bar{X} = 173,2$ e $s = 7,44$.

b) Teste de Shapiro-Wilk. Neste caso, como o α já está fixado, a estatística do teste é calculada da seguinte forma:

$$T_{s-w} = \frac{b^2}{(n-1)s^2}$$

Em que

$$b = \sum_{i=1}^{k} a_{i,n}[x_{(n-i+1)} - x_{(i)}]$$

E

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n-1}$$

Em que $k = n/2$, $x_{(i)}$ é a observação na i-ésima posição com os dados ordenados. Montando as hipóteses:

$$H_0 : Os\ dados\ seguem\ uma\ distribuição\ normal\ com\ \bar{X} = 173,2\ e\ s = 7,44.$$

$$H_1 : Os\ dados\ não\ seguem\ uma\ distribuição\ normal.$$

Vamos calcular as estatísticas de teste:

```
dados <- sort(c(177, 186, 158, 169, 173, 172, 171, 176, 170, 180))
dados
```

```
##  [1] 158 169 170 171 172 173 176 177 180 186
```

```
media <- mean(dados)
media # média
```

```
## [1] 173.2
```

```
variancia <- var(dados)
variancia # variância
```

```
## [1] 55.28889
```

```
dp <- sd(dados)
dp # desvio padrão
```

```
## [1] 7.43565
```

```
n = length(dados)
n
```

```
## [1] 10
```

```r
(n-1)*variancia
```

```
## [1] 497.6
```

```r
y2 = dados[1:5]
y2
```

```
## [1] 158 169 170 171 172
```

```r
length(y2)
```

```
## [1] 5
```

```r
y1 = sort(dados[6:10], decreasing = TRUE)
y1
```

```
## [1] 186 180 177 176 173
```

```r
length(y1)
```

```
## [1] 5
```

```r
y1-y2
```

```
## [1] 28 11  7  5  1
```

```r
a.i.n = c(0.3751,0.2574,0.2260,0.2032,0.1847,0.1691,0.1554,0.1430,0.1317,0.1212,0.1113,0.1020,0.0932,0.0
a.i.n
```

```
##  [1] 0.3751 0.2574 0.2260 0.2032 0.1847 0.1691 0.1554 0.1430 0.1317 0.1212
## [11] 0.1113 0.1020 0.0932 0.0846 0.0764 0.0685 0.0608 0.0532 0.0459 0.0386
## [21] 0.0314 0.0244 0.0174 0.0104 0.0035
```

```r
a.i.n*(y1-y2)
```

```
##  [1] 10.5028  2.8314  1.5820  1.0160  0.1847  4.7348  1.7094  1.0010  0.6585
## [10]  0.1212  3.1164  1.1220  0.6524  0.4230  0.0764  1.9180  0.6688  0.3724
## [19]  0.2295  0.0386  0.8792  0.2684  0.1218  0.0520  0.0035
```

```r
b = sum(a.i.n*(y1-y2))
b
```

```
## [1] 34.2842
```

```r
T.S.W = b^2 /((n-1)*variancia)
T.S.W
```

```
## [1] 2.362151
```

Conclusão: Na tabela o valor crítico foi de 0,842, que é menor que 2,362151 e não se rejeita H_0. Portanto, ao nível de $\alpha = 0,05$ os dados seguem uma distribuição normal pelo teste de Shapiro-Wilk.

Questão 3

Neste caso vamos trabalhar com o teste de McNemar. Como $b + c = 4 + 17 > 20$ então podemos trabalhar com a tabela de McNemar. A estatística de teste é calculada da seguinte forma:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

Vamos construir as hipóteses abaixo.

$$H_0 : O\ tratamento\ não\ foi\ eficaz.$$

$$H_1 : O\ tratamento\ foi\ eficaz.$$

Vamos calcular as estatísticas do teste abaixo.

```
library(stats)

questao3 <-matrix(c(45,4,17,34),2,2, byrow = T,
              dimnames = list("Depois" = c("Sem Cefaleia","Com Cefaleia"),
                              "Antes" = c("Sem Cefaleia","Com Cefaleia")
              ))
questao3
```

```
##                 Antes
## Depois          Sem Cefaleia Com Cefaleia
##    Sem Cefaleia           45            4
##    Com Cefaleia           17           34
```

```
# Calculando na mão:

qui_calc <- abs((questao3[1,2]-questao3[2,1]) - 1)/questao3[1,2] + questao3[2,1]
qui_calc
```

```
## [1] 20.5
```

```
# Pacote do R
mcnemar.test(questao3)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  questao3
## McNemar's chi-squared = 6.8571, df = 1, p-value = 0.008829
```

```
mcnemar.test(questao3, correct=FALSE)
```

```
##
##  McNemar's Chi-squared test
##
## data:  questao3
## McNemar's chi-squared = 8.0476, df = 1, p-value = 0.004556
```

Conclusão: O $\chi^2_{crítico}$ foi de 11,35, menor que o $\chi^2_{calculado}$ de 20,5, rejeitando H_0. Portanto, com $\alpha = 0,01$ concluí-se que o tratamento contra a cefaleia foi eficaz.

Questão 4

Vamos construir as hipóteses do teste. Como o teste é para verificar se houve melhora na aprendizagem, será conduzido um teste unilateral à direita com $\alpha = 0,05$

$$H_0 : Não\ houve\ melhora\ na\ aprendizagem.$$

$$H_1 : Houve\ melhora\ na\ aprendizagem.$$

Vamos calcular as estatisticas de teste.

```
paciente <- c(1:12)
nmc_antes <- c(9, 16, 12, 28, 5, 33, 17, 13, 18, 12, 26, 14)
nmc_depois <- c(14, 22, 18, 23, 11, 40, 15, 18, 22, 31, 19, 8)
diferencas <- nmc_antes - nmc_depois
diferencas
```

```
##  [1]  -5  -6  -6   5  -6  -7   2  -5  -4 -19   7   6
```

```
postos <- c(5,6,6,5,6,7,1,5,2,8,7,6)
length(postos)
```

```
## [1] 12
```

```
postos_negativos <- (5+6+6+7+5+2+8+6)
postos_negativos
```

```
## [1] 45
```

```
postos_positivos <- (6+5+1+7)
postos_positivos
```

```
## [1] 19
```

```
Tc <- min(postos_negativos, postos_positivos)
Tc
```

```
## [1] 19
```

Como o $T_c = 19 > T_{\alpha;n}$, então não rejeitamos H_0 conforme as notas de aula. Portanto, não houve diferença na administração da droga em relação à capacidade de aprendizado.