

Testes em dados categóricos e testes não paramétricos

Francisco Caramelo - fcaramelo@fmed.uc.pt

Anteriormente estudámos testes sobre variáveis quantitativas que obdeciam ao pressuposto de normalidade. Agora vamos analisar testes sobre variáveis nominais e testes não paramétricos.

Teste qui-quadrado de independência

Para estudar o teste χ^2 vamos fazer o uso dos dados contidos em *survey* que pertencem à livraria *MASS*

```
library(MASS)
head(survey,6)
```

```
##      Sex Wr.Hnd NW.Hnd W.Hnd  Fold Pulse  Clap Exer Smoke Height
## 1 Female  18.5  18.0 Right R on L   92   Left Some Never 173.00
## 2 Male   19.5  20.5 Left  R on L  104   Left None Regul 177.80
## 3 Male   18.0  13.3 Right L on R   87 Neither None Occas    NA
## 4 Male   18.8  18.9 Right R on L   NA Neither None Never 160.00
## 5 Male   20.0  20.0 Right Neither  35   Right Some Never 165.00
## 6 Female  18.0  17.7 Right L on R   64   Right Some Never 172.72
##      M.I    Age
## 1  Metric 18.250
## 2 Imperial 17.583
## 3    <NA> 16.917
## 4  Metric 20.333
## 5  Metric 23.667
## 6 Imperial 21.000
```

A variável *Sex* indica o género dos alunos e a variável *Smoke* a carga tabágica (*help(survey)* permite obter a descrição da base de dados). Considerando estas duas variáveis pode construir-se uma tabela de contigência a partir dos dados.

```
library(knitr)
tb = table(survey$Sex,survey$Smoke)
kable(tb)
```

	Heavy	Never	Occas	Regul
Female	5	99	9	5
Male	6	89	10	12

Para responder à questão se existe associação entre o género e os hábitos tabágicos devemos realizar um teste de χ^2 .

```
chisq.test(tb)
```

```
##
```

```
## Pearson's Chi-squared test
##
## data:  tb
## X-squared = 3.5536, df = 3, p-value = 0.3139
```

Observa-se que não existe associação estatisticamente significativa ($\chi(3) = 3.55, p = 0.3139$) entre o género e os hábitos tabágicos.

Exercício 1

1. Indique qual é a hipótese nula para o teste de χ^2 .
2. Verifique se existe associação entre o género e a forma como os alunos dobram (*Fold*) os braços. Reporte o resultado indicando a estatística de teste, o número de graus de liberdade e o valor p.

Testes não paramétricos

Os testes não paramétricos são realizados sobre variáveis quantitativas quando existe violação dos pressupostos de normalidade e sobre variáveis ordinais. Para exemplo de aplicação vamos fazer uso dos dados constantes na base de dados *mtcars* da livreria *datasets*.

```
library('datasets')
head(mtcars)
```

```
##           mpg  cyl  disp  hp  drat    wt   qsec vs  am  gear  carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46 0   1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 0   1    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61 1   1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44 1   0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 0   0    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22 1   0    3    1
```

Vamos começar por criar duas variáveis referentes ao peso do carro em que uma descreve o peso dos carros com até 4 cilindros e a outra o peso dos restantes carros e vamos determinar o tamanho de cada variável. Esta divisão de variáveis tem por objectivo comparar o peso dos carros.

```
weightLess4 <- mtcars[mtcars$cyl<=4,]$wt
print(length(weightLess4))
```

```
## [1] 11
```

```
weightMore4 <- mtcars[mtcars$cyl >4,]$wt
print(length(weightMore4))
```

```
## [1] 21
```

Vamos aplicar o teste de Shapiro-Wilk para verificar se as variáveis seguem uma distribuição normal.

```
shapiro.test(weightLess4)
```

```
##
## Shapiro-Wilk normality test
##
## data: weightLess4
## W = 0.9441, p-value = 0.5704
```

```
shapiro.test(weightMore4)
```

```
##
## Shapiro-Wilk normality test
##
## data: weightMore4
## W = 0.841, p-value = 0.00299
```

Uma vez que pelo menos uma das variáveis não verifica o pressuposto da normalidade vamos aplicar o teste não paramétrico correspondente (teste de Mann-Whitney) para comparar as medianas. (Nota: o teste de Mann-Whitney também é conhecido como teste de Wilcoxon rank sum)

```
wilcox.test(weightLess4,weightMore4)
```

```
## Warning in wilcox.test.default(weightLess4, weightMore4): cannot compute
## exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: weightLess4 and weightMore4
## W = 9, p-value = 2.581e-05
## alternative hypothesis: true location shift is not equal to 0
```

Exercício 2

2.1 Assumindo uma violação ao pressuposto da normalidade verifique se existem diferenças estatisticamente significativas no consumo (*mpg*) e o tipo de transmissão (*am*). Reporte os resultados indicando a estatística de teste e o valor p.