# 11 Hodnocení kontingenčních tabulek

**Příklad 11.1. Pearsonův $\chi^2$ test**
V roce 1950 zkoumali Yule a Kendall barvu očí a vlasů u 6800 mužů (viz tabulka 1; soubor vlasy_oci.csv).

Tabulka 1: Absolutní četnosti barvy vlasů a barvy očí mužů

| Barva očí | Barva vlasů | | | |
|---|---|---|---|---|
| | světlá | kaštanová | černá | rezavá |
| modrá | 1768 | 807 | 189 | 47 |
| šedá/zelená | 946 | 1387 | 746 | 53 |
| hnědá | 115 | 438 | 288 | 16 |

Na asymptotické hladině významnosti $\alpha = 0.01$ zjistěte, zda mezi barvou očí a vlasů u mužů existuje závislost.

### Řešení příkladu 11.1

- $H_0$ : Barva očí a barva vlasů ................................ stochasticky nezávislé.
- $H_1$ : Barva očí a barva vlasů ................................ stochasticky nezávislé.
- Hladina významnosti $\alpha =$ ............... .

```
data <- read.delim(..., header = T, row.names = 1) # nacteni datoveho souboru
n <- sum(...) # rozsah nahodneho vyberu
```

```
[1] 6800
```

Datový soubor obsahuje údaje o barvě očí a barvě vlasů ......................... mužů.

### Podmínka dobré aproximace
Pearsonův $\chi^2$ test můžeme provést, je-li splněna podmínka dobré aproximace (alespoň 80 % očekávaných četností je větších nebo rovných 5 a zbylých 20 % očekávaný četností neklesne pod hodnotu 2).

```
round(chisq.test(data, correct = F)$expected, 1) # tabulka ocekavanych cetnosti
```

```
            svetla kastanova cerna rezava
modra       1169.5    1088.0 505.6   48.0
seda/zelena 1303.0    1212.3 563.3   53.4
hneda        356.5     331.7 154.1   14.6
```

Podmínka dobré aproximace ................... splněna. Všechny teoretické četnosti jsou ......................... než 5.

### Pearsonův $\chi^2$ test

```
chisq.test(data, correct = F) # Pearsonuv chi-kvadratovy test
alpha <- ... # hladina vyznamnosti alpha
r <- ... # pocet variant znaku X (barva oci)
s <- ... # pocet variant znaku Y (barva vlasu)
q <- qchisq(..., ...) # dolni hranice kritickeho oboru
```

```

        Pearson's␣Chi-squared␣test

data:␣␣data
X-squared␣=␣1073.5,␣df␣=␣6,␣p-value␣<␣2.2e-16
```

```
         q
1 16.81189
```

a) **Test kritickým oborem**
Hodnota testové statistiky $K =$ .................., kritický obor $W$ má tvar ......................................... . Protože ............................., $H_0$ .......................... na hladině významnosti $\alpha =$ ............. .

b) **Test intervalem spolehlivosti**
Pro Pearsonův $\chi^2$ test vynecháváme testování intervalem spolehlivosti.

c) **Test $p$-hodnotou**
$P$-hodnota = .................................. . Protože ........................, $H_0$ .................................. na **asymptotické** hladině významnosti $\alpha =$ .................. .

Pro zjištění míry závislosti v kontingenční tabulce použijeme .................................. koeficient.

```
lsr::cramersV(...) # Crameruv koeficient
```

```
  Crameruv.koeficient
1           0.2809526
```

**Interpretace výsledků:** Mezi barvou očí a barvou vlasů mužů existuje / neexistuje statisticky významná stochastická závislost. Mezi barvou očí a barvou vlasů mužů existuje .................................................... stupeň závislosti ($V =$ ........................).

★

**Příklad 11.2. Test podílem šancí**
Máme k dispozici údaje o frekvenci výskytu vysokého (11.9.7), středního (9.7.5) a nízkého (7.5.5) zakončení tří hlavních dlaňových linií na pravé a levé ruce 50 mužů a 50 žen z populace Mech a 105 mužů a 87 žen z populace Rajbanshi (viz tabulka 2; datový soubor 27-two-samples-probabilities-palmar.txt). Na hladině významnosti alpha = 0.05 testujte hypotézu o nezávislosti výskytu středního zakončení tří hlavních dlaňových linií na levé straně u žen a populace, ze které pocházejí.

Tabulka 2: Četnosti typů zakončení tří hlavních dlaňových linií na levé ruce u jedinců dvou indických populací

| mech-L<br>zakončení | pohlaví<br>muži | <br>ženy |
|---|---|---|
| vysoké | 4 | 5 |
| střední | 16 | 16 |
| nízké | 21 | 19 |
| ostatní | 9 | 10 |

| raj-L<br>zakončení | pohlaví<br>muži | <br>ženy |
|---|---|---|
| vysoké | 30 | 27 |
| střední | 24 | 17 |
| nízké | 38 | 21 |
| ostatní | 13 | 22 |

**Řešení příkladu 11.2**

```
data <- data.frame(mech = c(16, 50 - 16), rajbanshi = c(17, 87 - 17), row.names = c(..., ...))
```

```
        mech rajbanshi
stredni   16        17
ostatni   34        70
```

Datový soubor obsahuje údaje o typu zakončení tří dlaňových linií .................. žen, přičemž ................ žen má střední zakončení a pochází z mechské populace, ................ žen má jiné zakončení a pochází z mechské populace, ................ žen má střední zakončení a pochází z populace Rajbanshi a ................ žen má jiné zakončení a pochází z populace Rajbanshi.

**Řešení příkladu 11.2**

- $H_0$ : Typ zakončení a populace ..................... stochasticky nezávislé. → ............................ → ............................
- $H_1$ : Typ zakončení a populace ..................... stochasticky nezávislé. → ............................ → ............................
- Hladina významnosti $\alpha =$ ............... .

**Podmínka dobré aproximace**
Nejprve musíme ověřit, zda je splněna podmínka dobré aproximace (alespoň 80 % očekávaných četností je větších nebo rovných 5 a zbylých 20 % očekávaný četností neklesne pod hodnotu 2).

```
round(chisq.test(...)$expected, ...) # tabulka ocekavanych cetnosti
```

```
        mech rajbanshi
stredni   12        21
ostatni   38        66
```

Podmínka dobré aproximace .................... splněna. Všechny teoretické četnosti jsou .......................... než 5. Protože je podmínka dobré aproximace splněna, můžeme otestovat hypotézu ze zadání pomocí **testu podílem šancí**. Pokud by podmínka dobré aproximace splněna nebyla, museli bychom použít Fisherův faktoriálový (exaktní) test.

### Výpočet (logaritmu) podílu šancí

```
a <- 16; b <- 17; c <- 34; d <- 70
OR <- ... / ... # podil sanci ad / bc
lnOR <- log(...) # logaritmus podilu sanci
```

```
        OR      lnOR
1 1.937716 0.6615101
```

Podíl šancí $OR$ =.................................. . Logaritmus podílu šancí $\ln(OR)$ =.................................. .

### Test podílem šancí

```
source('Sbirka-AS-I-2018-funkce-II.R')
odds.ratio.test(..., conf.level = ...)
alpha <- ... # hladina vyznamnosti
qnorm(...) # hranice kritickeho oboru
qnorm(...) # hranice kritickeho oboru
```

```
        OR      lnOR       t0         dh       hh         p
1 1.937716 0.6615101 1.628422 -0.1346817 1.457702 0.1034355
```

```
         q1       q2
1 -1.959964 1.959964
```

a) **Test kritickým oborem**
Hodnota testové statistiky $t_0$ je .......................... Kritický obor má tvar ..................................................... .
Protože ........................, $H_0$.................................. na **asymptotické** hladině významnosti $\alpha$ =......................... .

b) **Test intervalem spolehlivosti**
Interval spolehlivosti má tvar ......................................... . Protože ............................., $H_0$.................................
na **asymptotické** hladině významnosti $\alpha$ =......................... .

c) **Test $p$-hodnotou**
$P$-hodnota = .................................. . Protože ........................, $H_0$ .................................. na **asymptotické** hladině významnosti $\alpha$ = .................. .

**Interpretace výsledků:** Mezi typem zakončení tří hlavních dlaňových linií u žen a populací, ze které pochází, existuje / neexistuje statisticky významná stochastická závislost. U žen mechské / rajbanshské populace je 1.94 krát větší šance na výskyt středního zakončení tří hlavních dlaňových linií než u žen mechské / rajbanshské populace.

★

**Příklad 11.3. Fisherův faktoriálový test**
Máme k dispozici údaje o frekvenci výskytu vysokého (11.9.7), středního (9.7.5) a nízkého (7.5.5) zakončení tří hlavních dlaňových linií na pravé a levé ruce 50 mužů a 50 žen z populace Mech a 105 mužů a 87 žen z populace Rajbanshi (viz tabulka 2; datový soubor 27-two-samples-probabilities-palmar.txt). Na hladině významnosti alpha = 0.05 testujte hypotézu o nezávislosti výskytu středního zakončení tří hlavních dlaňových linií na levé straně u žen a populace, ze které pocházejí. Zadanou hypotézu otestujte Fisherovým faktoriálovým testem.

### Řešení příkladu 11.3

Fisherův faktoriálový test (jinak zvaný též Fisherův přesný (exaktní) test) používáme primárně, pokud podmínka dobré aproximace není splněna. Můžeme jej ale použít i v případě splnění podmínky dobré aproximace jako alternativní test k testu podílem šancí.

### Fisherův faktoriálový test

- $H_0$ : Typ zakončení tří hlavních dlaňových linií a populace ................................ stochasticky nezávislé.
- $H_1$ : Typ zakončení tří hlavních dlaňových linií a populace ................................ stochasticky nezávislé.

- Hladina významnosti $\alpha =$ ............... .

```
fisher.test(data, alternative = ..., conf.level = ...) # Fisheruv faktorialovy test
```

```

        Fisher's␣Exact␣Test␣for␣Count␣Data

data:␣␣data
p-value␣=␣0.1455
alternative␣hypothesis:␣true␣odds␣ratio␣is␣not␣equal␣to␣1
95␣percent␣confidence␣interval:
␣0.8047687␣4.6240301
sample␣estimates:
odds␣ratio
␣␣1.927942
```

a) **Test kritickým oborem**
Pro Fisherův faktoriálový test vynecháváme testování kritickým oborem.

b) **Test intervalem spolehlivosti**
Pro Fisherův faktoriálový test vynecháváme testování intervalem spolehlivosti.

c) **Test $p$-hodnotou**
$P$-hodnota = .......................... . Protože ............................, $H_0$ ................................. na hladině významnosti $\alpha =$ .................. .

**Interpretace výsledků:** Mezi typem zakončení tří hlavních dlaňových linií u žen a populací, ze které pocházejí, existuje / neexistuje statisticky významná stochastická závislost. ★