10 Hodnocení kontingenčních tabulek

Příklad 10.1. Testování hypotézy o nezávislosti, měření síly závislosti
V roce 1950 zkoumali Yule a Kendall barvu očí a vlasů u 6800 mužů. Výsledky zkoumání jsou uvedeny v následující tabulce a v souboru vlasy_oci.csv.

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 $\alpha = 0.05$ testujte hypotézu o nezávislosti barvy očí a barvy vlasů. Vypočtěte Cramérův koeficient.

Řešení příkladu 10.1

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

Podmínka dobré aproximace

```
            svetla kastanova cerna rezava
modra       1167.3    1086.0 500.9   47.9
seda/zelena 1304.7    1213.9 559.9   53.5
hneda        357.0     332.1 153.2   14.6
```

Podmínky dobré aproximace .... splněny. Všechny teoretické četnosti jsou .... než 5.

Pearsonův χ^2 test

```
X-squared
 1088.149
```

```
[1] 12.59159
```

a) Test pomocí kritického oboru
Hodnota testovací statistiky K je .... Kritický obor má tvar ....
Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

b) Test pomocí p-hodnoty

```
[1] 7.645911e-232
```

P-hodnota vyšla .... Protože p-hodnota = ... α, H_0 .... na hladině významnosti $\alpha =$............................ .

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

```
[1] 0.2830494
```

Hodnota Cramérova koeficientu je ....

Interpretace výsledků: Znaky barva očí a barva vlasů jsou / nejsou stochasticky nezávislé. Mezi barvou očí a barvou vlasů existuje .... stupeň závislosti.

Příklad 10.2. Fisherův faktoriálový test
100 náhodně vybraných mužů a žen bylo dotázáno, zda dávají přednost nealkoholickému nápoji A či B. Údaje jsou uvedeny ve čtyřpolní kontingenční tabulce.

pref. nápoj	pohlaví	
	muž	žena
A	20	30
B	30	20

Na hladině významnosti $\alpha = 0.05$ testujte pomocí Fisherova faktoriálového testu hypotézu, že preferovaný typ nápoje nezáleží na pohlaví respondenta.

Řešení příkladu 10.2

- H_0 : Znaky pohlaví a preference .... stochasticky nezávislé.
- H_1 : Znaky pohlaví a preference .... stochasticky nezávislé.
- Hladina významnosti $\alpha =$ ....

Fisherův faktoriálový test

```

        Fisher's Exact Test for Count Data

data:  data
p-value = 0.07134
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1846933 1.0640121
sample estimates:
odds ratio
 0.4481632
```

P-hodnota vyšla .... Protože p-hodnota $=$ .... α, H_0 .... na hladině významnosti $\alpha =$............................ .

Interpretace výsledků: Znaky pohlaví a preference jsou / nejsou stochasticky nezávislé.

Příklad 10.3. Podíl šancí
Pro údaje z příkladu č.3 vypočtěte podíl šancí a sestrojte 95 % asymptotický interval spolehlivosti pro logaritmus podílu šancí. Pomocí tohoto intervalu spolehlivosti testujte na asymptotické hladině významnosti $\alpha = 0.05$ hypotézu, že preferovaný typ nápoje nezáleží na pohlaví respondenta.

Řešení příkladu 10.3

- H_0 : .... $\rightarrow$ ....
- H_1 : .... $\rightarrow$ ....
- Hladina významnosti $\alpha =$ ....

Podmínka dobré aproximace

Podmínky dobré aproximace .... splněny. Všechny teoretické četnosti jsou .... než 5.

```
  muz zena
A  25   25
B  25   25
```

Výpočet (logaritmu) podílu šancí

```
[1] 0.4444444
```

```
[1] -0.8109302
```

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

a) Testování pomocí kritického oboru

```
[1] -1.986365
```

```
[1] -1.959964
```

```
[1] 1.959964
```

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar ....
Protože ...., H_0................................ na hladině významnosti α =........................... .

b) Test pomocí intervalu spolehlivosti
Proti .... alternativě postavíme .... IS.

```
[1] -1.611082
```

```
[1] -0.01077827
```

Interval spolehlivosti má tvar .... Protože ...., H_0................................ na hladině významnosti α =........................... .

c) Test pomocí p-hodnoty

```
[1] 0.04699278
```

P-hodnota vyšla .... Protože p-hodnota = ... α, H_0 .... na hladině významnosti α =........................... .

Interpretace výsledků: Znaky pohlaví a preference jsou / nejsou stochasticky nezávislé. Muži preferují nápoj A .... krát častěji než ženy, resp. ženy preferují nápoj A .... krát častěji než muži.

Poznámka: Uvedený výsledek je v rozporu s výsledkem, ke kterému dospěl Fisherův faktoriálový (přesný) test. Je to způsobeno tím, že test pomocí asymptotického intervalu spolehlivosti je pouze přibližný. Ke stejnému závěru, jaký jsme dostali u testování pomocí podílu šancí, dospějeme, pokud použijeme Pearsonův chí-kvadrát test o nezávislosti.

```

        Pearson's Chi-squared test

data:  data
X-squared = 4, df = 1, p-value = 0.0455
```

Ve funkci chisq.test() však můžeme zadat parametr correct=T, který provede korekci Pearsonova testu pro kontingenční tabulky typu 2×2. Výsledek takto provedeného testu je již v souladu s Fisherovým faktoriálovým testem.

```

        Pearson's Chi-squared test with Yates' continuity correction

data:  data
X-squared = 3.24, df = 1, p-value = 0.07186
```