11 Hodnocení kontingenčních tabulek

Příklad 11.1. 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	180	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. Simultánní četnosti znázorněte graficky.

Podmínka dobré aproximace

```
##                svetla kastanova    cerna   rezava
## modra       1167.2593  1085.976 500.9024 47.86217
## seda/zelena 1304.7310  1213.875 559.8952 53.49904
## hneda        357.0097   332.149 153.2025 14.63879
```

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

Pearsonův χ^2 test

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 1088.1, df = 6, p-value < 2.2e-16
```

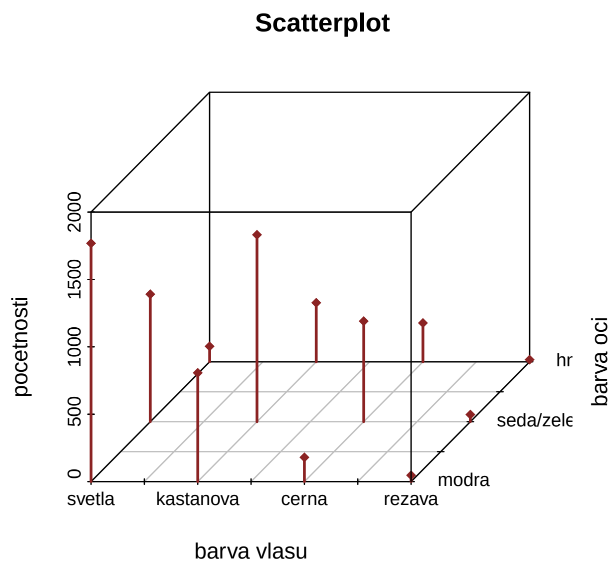
Hodnota testovací statistiky K =..........................., počet stupňů volnosti df =........................... Protože p-hodnota = ... je .... než $\alpha = 0.05$, nulovou hypotézu H_0 o nezávislosti barvy očí a barvy vlasů .... na **asymptotické** hladině významnosti α =................................. .

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

```
## [1] 0.2830494
```

Hodnota Cramérova koeficientu je ...., což svědčí o .... závislosti barvy očí a vlasů.

Grafické znázornění četností



Příklad 11.2. Otevřete si soubor ped_hodnost.txt. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu o nezávislosti pedagogické hodnosti a pohlaví. Dále vypočtěte Cramérův koeficient vyjadřující intenzitu závislosti pedagogické hodnosti na pohlaví. Data v souboru mají následující tvar:

pohlaví	pedagogická hodnost		
	odb. asistent	docent	profesor
muž	32	15	8
žena	34	8	3

Podmínka dobré aproximace

```
##      odb.asistent docent profesor
## muz          36.3  12.65     6.05
## zena         29.7  10.35     4.95
```

Podmínky dobré aproximace .... splněny. Všechny teoretické četnosti až na jednu jsou .... než 5.

Pearsonův χ^2 test

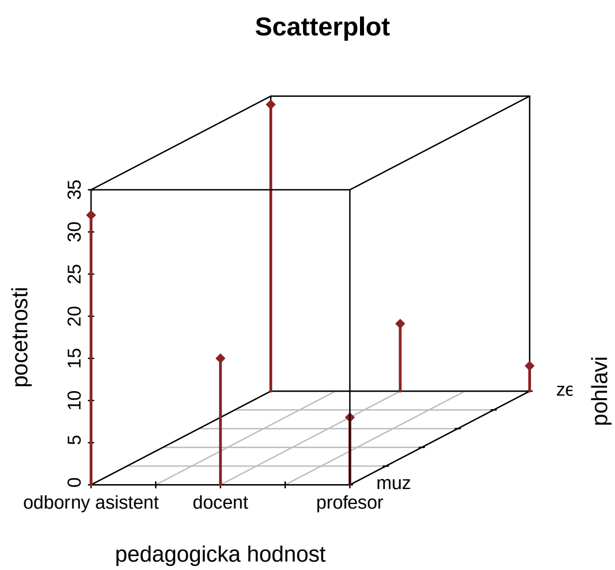
```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 3.4988, df = 2, p-value = 0.1739
```

Hodnota testovací statistiky K =..........................., počet stupňů volnosti df =.......................... . Protože p-hodnota = ... je .... než $\alpha = 0.05$, nulovou hypotézu H_0 o nezávislosti pedagogické hodnosti a pohlaví .... na **asymptotické** hladině významnosti α =................................ .

```
## [1] "Crameruv koeficient: V= 0.187"
```

Hodnota Cramérova koeficientu je ...., což svědčí o .... závislosti mezi pedagocickou hodností a pohlavím.

Grafické znázornění četností



Příklad 11.3. 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.

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
```

Protože p-hodnota=...................................... je .... než $\alpha = 0.05$, nulovou hypotézu H_0 o nezávislosti preferovaného typu nápoje na pohlaví .... na hladině významnosti $\alpha =$.............................. .

Příklad 11.4. 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.

Podmínka dobré aproximace

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

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

Podíl šancí $OR =$.................................. .

95% interval spolehlivosti pro $\ln o\rho =$ .... Protože ...., nulovou hypotézu H_0 o nezávislosti preferovaného typu nápoje na pohlaví respondenta .... na asymptotické hladině významnosti $\alpha =$..................... .

```
## [1] 0.4444444
## [1] -1.611082
## [1] -0.01077827
```

Tento výsledek je v rozporu s výsledkem, ke kterému dospěl Fisherův 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 přesným testem.

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 3.24, df = 1, p-value = 0.07186
```

Příklad 11.5. 36 mužů onemocnělo určitou chorobou. Někteří z nich se léčili, jiní ne. Někteří se uzdravili, jiní zemřeli. Údaje jsou uvedeny ve čtyřpolní kontingenční tabulce.

přežití	léčení	
	ano	ne
ano	10	6
ne	12	8

Vypočtěte a interpretujte podíl šancí. Pomocí intervalu spolehlivosti pro logaritmus podílu šancí testujte na asymptotické hladině významnosti $\alpha = 0.05$ hypotézu, že přežití nezávisí na léčení, proti tvrzení, že léčení zvyšuje šance na přežití.

```
##          muz     zena
## A  9.777778 6.222222
## B 12.222222 7.777778
```

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

```
## [1] "OR= 1.1111"
## [1] "dolni hranice IS: -1.0283"
```

Podíl šancí $OR =$.................................. .

95% interval spolehlivosti pro $\ln o\rho =$ .... Protože ...., nulovou hypotézu H_0 .... na asymptotické hladině významnosti $\alpha =$..................... .

Příklad 11.6. V průzkumu o kuřáctví bylo dotázáno 92 osob. Z 64 mužů jich kouří 19 a z 28 žen jich kouří 6.

a) Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že kouření se vyskytuje stejně často u mužů a žen. Použijte Pearsonův chi-kvadrát test i Fisherův přesný test.

b) Vypočtěte a interpretujte podíl šancí a stanovte meze 95% intervalu spolehlivosti pro podíl šancí.

Pearsonův χ^2 test

```
## [1] "Podminka dobre aproximace"
##              muz      zena
## kurak   17.3913  7.608696
## nekurak 46.6087 20.391304
```

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

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 0.31889, df = 1, p-value = 0.5723
```

Hodnota testovací statistiky $K =$..........................., počet stupňů volnosti $df =$.......................... . p-hodnota........................... .
Protože p-hodnota je .... než $\alpha = 0.05$, nulovou hypotézu H_0 nezávislosti kouření a pohlaví ....
na **asymptotické** hladině významnosti $\alpha =$................................ .

Fisherův přesný test

```
##
##  Fisher's Exact Test for Count Data
##
## data:  data
## p-value = 0.4576
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
```

```
##  0.498056 5.398695
## sample estimates:
## odds ratio
##    1.54109
```

Protože p-hodnota=...................................... je .... než $\alpha = 0.05$, nulovou hypotézu H_0 o nezávislosti kouření na pohlaví .... na hladině významnosti $\alpha =$.............................. .

Podíl šancí:

```
## [1] "OR= 1.5481"
## [1] "dolni hranice IS: 0.5418"
## [1] "horni hranice IS: 4.4239"
```

Podíl šancí $OR =$................................... .

95% interval spolehlivosti pro $\ln o\rho =$ .... Protože ...., nulovou hypotézu H_0 .... na asymptotické hladině významnosti $\alpha =$..................... .