Cvičení 10 – příklady u tabule

Příklad 1. (viz př. 11.5.1. ze skript): Na hladině významnosti 0,05 testujte hypotézu o nezávislosti pedagogické hodnosti a pohlaví a vypočtěte Cramérův koeficient, jsou-li k dispozici následující údaje:

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

Řešení: Tabulku doplníme o marginální četnosti.

pohlaví	pedagogická hodnost			$n_{j.}$
	odb. asistent	docent	profesor	
muž	32	15	8	55
žena	34	8	3	45
$n_{.k}$	66	23	11	n=100

Vypočteme teoretické četnosti:

$$\frac{n_{1.}n_{.1}}{n} = \frac{55 \cdot 66}{100} = 36{,}3, \frac{n_{1.}n_{.2}}{n} = \frac{55 \cdot 23}{360} = 12{,}65, \frac{n_{1.}n_{.3}}{n} = \frac{55 \cdot 11}{100} = 6{,}05,$$

$$\frac{n_{2.}n_{.1}}{n} = \frac{45 \cdot 66}{100} = 29{,}7, \frac{n_{2.}n_{.2}}{n} = \frac{45 \cdot 23}{100} = 10{,}35, \frac{n_{2.}n_{.3}}{n} = \frac{45 \cdot 11}{100} = 4{,}95.$$

Vypočteme testovou statistiku: $K = \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{\left(n_{jk} - \frac{n_{j.}n_{.k}}{n}\right)^2}{\frac{n_{j.}n_{.k}}{n}}$, r = 2, s = 3, tedy

$K = \frac{(32-36{,}3)^2}{36{,}3} + \frac{(15-12{,}65)^2}{12{,}65} + \ldots + \frac{(3-4{,}95)^2}{4{,}95} = 3{,}5$, $\chi^2_{0,95}(2) = 5{,}991$. Protože K < 5,991, hypotézu o nezávislosti pohlaví a pedagogické hodnosti nezamítáme na asymptotické hladině významnosti 0,05.

Cramérův koeficient: $V = \sqrt{\frac{K}{n(m-1)}}$, kde m = min{r,s}, tedy $V = \sqrt{\frac{3{,}5}{100 \cdot 1}} = 0{,}187$.

Příklad 2. (viz př. 11.5.5. ze skript): 400 náhodně vybraných pracovníků potravinářského podniku bylo dotázáno na příčiny nespokojenosti na pracovišti.

kategorie	hlavní příčina nespokojenosti				
pracovníků	pracovní prostředí	špatné vztahy	organizace práce	výdělek	jiné
dělníci	80	50	75	40	55
THP	10	10	25	30	25

Na hladině významnosti 0,05 testujte hypotézu, že hlavní příčina nespokojenosti nezávisí na kategorii, do níž je pracovník zařazen. Vypočtěte Cramérův koeficient.

Řešení:

r = 2, s = 5, n = 400

Ověření podmínek dobré aproximace: teoretické četnosti jsou 67,5 45 75 52,5 60 22,5 15 25 17,5 20. Podmínky dobré aproximace jsou tedy splněny. Vypočteme realizaci testového kritéria:

$K = \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{\left(n_{jk} - \frac{n_{j.} n_{.k}}{n}\right)^2}{\frac{n_{j.} n_{.k}}{n}} = 25{,}053$, kritický obor $W = \langle \chi^2_{0,95}(4), \infty) = \langle 9{,}4877; \infty)$. Protože testové kritérium se realizuje v kritickém oboru, hypotézu o nezávislosti zamítáme na asymptotické hladině významnosti 0,05.

Cramérův koeficient: $V = \sqrt{\frac{K}{n(m-1)}} = \sqrt{\frac{25{,}053}{400}} = \sqrt{0{,}6262} = 0{,}25$

Příklad 3.: 18 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	5	3
ne	6	4

Vypočtěte podíl šancí a sestrojte 95% asymptotický interval spolehlivosti pro teoretický podíl šancí. Pomocí tohoto intervalu spolehlivosti testujte na asymptotické hladině významnosti 0,05 hypotézu, že přežití nezávisí na léčení.

Řešení: Podíl šancí $OR = \frac{ac}{bd} = \frac{5 \cdot 4}{3 \cdot 6} = \frac{20}{18} = \frac{10}{9} = 1{,}\bar{1}$. Protože podíl šancí je větší než 1, je zřejmě výhodnější se nechat léčit.

Dolní a horní mez 100(1-α)% asymptotického intervalu spolehlivosti intervalu spolehlivosti pro oρ vypočteme podle vzorců

$$d = \exp\left(\ln OR - \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\, u_{1-\alpha/2}\right), h = \exp\left(\ln OR + \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\, u_{1-\alpha/2}\right)$$

Výsledek: 0,1645 < oρ < 7,506 s pravděpodobností přibližně 0,95. Protože tento interval spolehlivosti obsahuje 1, nelze na asymptotické hladině významnosti 0,05 zamítnout hypotézu, že přežití nezávisí na léčení.

Příklad 4.: V náhodném výběru 50 studentů gymnázia bylo zjišťováno, zda se účastní biologické olympiády a zda sledují novinky v biologickém výzkumu. Veličina X – účast v olympiádě, veličina Y – sledování novinek ve výzkumu. Výsledky průzkumu jsou uvedeny v kontingenční tabulce:

X	Y	
	ne	ano
ne	19	7
ano	9	15

Na hladině významnosti 0,05 testujte hypotézu o nezávislosti náhodných veličin X a Y. Použijte zjednodušený vzorec pro výpočet testové statistiky K. Vypočtěte Cramérův koeficient.

Řešení: Tabulku doplníme o marginální četnosti.

X	Y		$n_{j.}$
	ne	ano	
ne	19	7	26
ano	9	15	24
$n_{.k}$	28	22	50

Vypočteme teoretické četnosti:

$$\frac{n_{1.}n_{.1}}{n} = \frac{26 \cdot 28}{50} = 14{,}56, \quad \frac{n_{1.}n_{.2}}{n} = \frac{26 \cdot 22}{50} = 11{,}44, \quad \frac{n_{2.}n_{.1}}{n} = \frac{24 \cdot 28}{50} = 13{,}44, \quad \frac{n_{2.}n_{.2}}{n} = \frac{24 \cdot 22}{50} = 10{,}56 .$$

Podmínky dobré aproximace jsou splněny.

Testovou statistiku vypočteme podle vzorce $K = \dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \dfrac{50(19 \cdot 15 - 7 \cdot 9)^2}{26 \cdot 24 \cdot 28 \cdot 22} = 6{,}4108$

Stanovíme kritický obor $W = \langle \chi^2_{1-\alpha}(1), \infty) = \langle \chi^2_{0,95}(1), \infty) = \langle 3{,}841, \infty)$

Testová statistika se realizuje v kritickém oboru, tedy hypotézu o nezávislosti veličin X a Y zamítáme na asymptotické hladině významnosti 0,05.

Cramérův koeficient: $V = \sqrt{\dfrac{K}{n(m-1)}} = \sqrt{\dfrac{6{,}4108}{50}} = 0{,}3581$