

Cvičení 10 – Hodnocení kontingenčních tabulek

Zavedení kontingenční tabulky: Pořídíme náhodný výběr $(X_1, Y_1), \dots, (X_n, Y_n)$ rozsahu n z rozložení, kterým se řídí dvourozměrný diskrétní náhodný vektor (X, Y) , kde X má r variant a Y má s variant. Zjištěné absolutní simultánní četnosti n_{jk} dvojice variant $(x_{[j]}, y_{[k]})$ uspořádáme do kontingenční tabulky:

	y	Y _[1]	...	Y _[s]	n _{j.}
x	n _{jk}				
X _[1]		n ₁₁	...	n _{1s}	n _{1.}
...	
X _[r]		n _{r1}	...	n _{rs}	n _{r.}
n _{.k}		n _{.1}	...	n _{.s}	n

$n_{j.} = n_{j1} + \dots + n_{js}$ je marginální absolutní četnost varianty $x_{[j]}$

$n_{.k} = n_{1k} + \dots + n_{rk}$ je marginální absolutní četnost varianty $y_{[k]}$

Testování hypotézy o nezávislosti

H_0 : X, Y jsou stochasticky nezávislé náhodné veličiny proti alternativě H_1 : X, Y nejsou stochasticky nezávislé náhodné veličiny.

Testová statistika:
$$K = \sum_{j=1}^r \sum_{k=1}^s \frac{\left(n_{jk} - \frac{n_{j.} n_{.k}}{n} \right)^2}{\frac{n_{j.} n_{.k}}{n}} \approx \chi^2((r-1)(s-1)), \text{ když } H_0 \text{ platí.}$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}((r-1)(s-1)), \infty \rangle.$

Podmínky dobré aproximace: Teoretické četnosti $\frac{n_{j.} n_{.k}}{n}$ aspoň v 80 % případů nabývají hodnoty větší nebo rovné 5 a ve zbylých 20 % neklesnou pod 2. Není-li splněna podmínka dobré aproximace, doporučuje se slučování některých variant.

Cramérův koeficient: $V = \sqrt{\frac{K}{n(m-1)}}$, kde $m = \min\{r, s\}$, $V \in (0, 1)$. Čím blíže je V k 1, tím

je závislost mezi X a Y těsnější, čím blíže je V k 0, tím je tato závislost volnější.

Význam hodnot Cramérova koeficientu:

mezi 0 až 0,1 ... zanedbatelná závislost, mezi 0,1 až 0,3 ... slabá závislost, mezi 0,3 až 0,7 ... střední závislost, mezi 0,7 až 1 ... silná závislost.

Čtyřpolní tabulky: Necht' $r = s = 2$. Používáme označení: $n_{11} = a, n_{12} = b, n_{21} = c, n_{22} = d$.

Testová statistika:
$$K = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \approx \chi^2(1), \text{ když } H_0 \text{ platí.}$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}(1), \infty \rangle.$

Podíl šancí ve čtyřpolní tabulce: $OR = \frac{ad}{bc}$ - výběrový podíl šancí, je odhadem neznámého

teoretického podílu šancí $op = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}$. Interval spolehlivosti pro podíl šancí op :

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

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

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

Kategorie pracovníků	Hlavní příčina nespokojenosti				
	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.

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.

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	
	ano	ne
ano	15	9
ne	7	19

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.