Pravděpodobnost a statistika

12. cvičení

(vytvořeno 12. prosince 2020)

Úloha 1 (χ^2-test dobré shody se známým rozdělením)**.** Hodili jsme stokrát mincí. Tabulka udává, kolikrát padl rub než padl první líc. Otestujte na hladině významnosti 5% hypotézu, že naměřené hodnoty pochází z geometrického rozdělením k koeficientem (pravděpodobnost rubu) $q = 0,5$.

hodnota	0	1	2	3	4	5	6
naměřená četnost	57	24	10	6	2	0	1

Úloha 2 (χ^2-test dobré shody se známým rozdělením)**.** Máme podezření že hrací kostka je falešná a padá na ní šestka častěji, než by měla. Otestujte na hladině významnosti 5% hypotézu, že naměřené hodnoty pochází z roznoměrného rozdělení.

hodnota	1	2	3	4	5	6
naměřená četnost	7	5	5	12	6	15

Úloha 3 (χ^2-test dobré shody dvou neznámých rozdělení)**.** Dvě skupiny studentů psaly test, ze kterého bylo možné získat maximálně 4 body. Následující tabulka uvádí počty studentů, kteří dosáhli daného bodového ohodnocení. Na hladině významnosti 5% otestujte hypotézu, že obě skupiny mají stejné rozdělení pravděpodobnosti (tj., že není statisticky významný rozdíl mezi tím, jak dopadl test v první a ve druhé skupině).

počet bodů	0	1	2	3	4
první skupina	12	8	4	16	20
druhá skupina	8	7	1	9	15

Úloha 4 (Test nezávislosti)**.** Následující tabulka udává naměřené četnosti lidí podle barvy očí a světlosti vlasů. Otestujte na hladině významnosti 5% hypotézu, že barva očí nezávisí na světlosti vlasů.

vlasy ╲ oči	modré	šedé	hnědé
tmavé	10	10	40
světlé	20	10	10

Úloha 5 (Test korelace)**.** Změřili jsme výšku a váhu čtyř mužů. Otestujte na hladině významnosti 5% hypotézu, že výška a váha nejsou korelovány.

výška	205	155	185	155
váha	95	55	65	85

Řešení

Řešení 1:

- nulová hypotéza H_0: naměřené hodnoty pochází z geometrického rozdělením k koeficientem $q = 0,5$
- alternativní hypotéza H_1: naměřené hodnoty nepochází z geometrického rozdělením k koeficientem $q = 0,5$
- testovací statistika:

$$T = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \quad \dots \quad \chi^2(k-1) \quad (\text{pro } n \to \infty)$$

- realizace testovací statistiky: $t = 2.5$

hodnota	i	0	1	2	3 a více	Σ
naměřená četnost	n_i	57	24	10	9	
teoretická pravděpodobnost	p_i	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	
teoretická četnost	$n \cdot p_i$	50	25	12.5	12.5	
	$n_i - n\,p_i$	7	-1	-2.5	-3.5	
	$(n_i - n\,p_i)^2$	49	1	6.25	12.25	
	$\frac{(n_i - n\,p_i)^2}{n\,p_i}$	0.98	0.04	0.5	0.98	2.5

- kvantily: $q_{\chi^2(3)}(0.025) \doteq 0.216$, $q_{\chi^2(3)}(0.975) \doteq 9.35$
- Nulovou hypotézu H_0 nezamítáme.

Řešení 2:

- nulová hypotéza H_0: naměřené hodnoty pochází z roznoměrného rozdělení
- alternativní hypotéza H_1: naměřené hodnoty nepochází z roznoměrného rozdělení (kostka je falešná)
- realizace testovací statistiky: $t = 6.4$

hodnota	i	1 až 5	6	Σ
naměřená četnost	n_i	35	15	
teoretická pravděpodobnost	p_i	$\frac{5}{6}$	$\frac{1}{6}$	
teoretická četnost	$n \cdot p_i$	$\frac{250}{6} = \frac{125}{3}$	$\frac{50}{6} = \frac{25}{3}$	
	$n_i - n\,p_i$	$-\frac{20}{6}$	$\frac{20}{6}$	
	$(n_i - n\,p_i)^2$	$\frac{400}{9}$	$\frac{400}{9}$	
	$\frac{(n_i - n\,p_i)^2}{n\,p_i}$	$\frac{400}{375}$	$\frac{400}{75}$	$\frac{400+2000}{375} = \frac{32}{5} = 6.4$

- kvantily: $q_{\chi^2(1)}(0.025) \doteq 0.001$, $q_{\chi^2(1)}(0.975) \doteq 5.02$

α	0.025	0.5	0.95	0.975
$q_{\chi^2(1)}(\alpha)$	0.001	0.0039	3.84	5.02

- Nulovou hypotézu H_0 zamítáme, kostka je falešná.

Řešení 3:

- nulová hypotéza H_0: naměřené hodnoty pochází ze stejného rozdělení
- alternativní hypotéza H_1: naměřené hodnoty nepochází ze stejného rozdělení
- testovací statistika:

$$
\begin{aligned}
T &= \sum_{i=1}^{k} \frac{(m_i - m\,p_i)^2}{m\,p_i} + \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \\
p_i &= \frac{m_i + n_i}{m + n}
\end{aligned}
$$

- realizaci t testujeme na $\chi^2(k-1)$

- realizace testovací statistiky:

i	1	$2,3$	4	5	
m_i	14	12	16	18	$m = 60$
n_i	10	6	10	14	$n = 40$
$m_i + n_i$	24	18	26	32	$m + n = 100$
$p_i = \frac{m_i+n_i}{m+n}$	0.24	0.18	0.26	0.32	
$m\,p_i$	14.4	10.8	15.6	19.2	
$m_i - m\,p_i$	-0.4	1.2	0.4	-1.2	
$(m_i - m\,p_i)^2$	0.16	1.44	0.16	1.44	
$\frac{(m_i-m\,p_i)^2}{m\,p_i}$	0.1111	0.1333	0.0103	0.075	0.3297
$n\,p_i$	9.6	7.2	10.4	12.8	
$n_i - n\,p_i$	0.4	-1.2	-0.4	1.2	
$(n_i - n\,p_i)^2$	0.16	1.44	0.16	1.44	
$\frac{(n_i-n\,p_i)^2}{n\,p_i}$	0.1667	0.2	0.0154	0.1125	0.4946

$$t = 0.3297 + 0.4946 = 0.8243$$

- kvantil: $q_{\chi^2(3)}(0.95) \doteq 7.81$
- Nulovou hypotézu H_0 nezamítáme.

Řešení 4:

- nulová hypotéza H_0: jevy jsou nezávislé
- alternativní hypotéza H_1: jevy nejsou nezávislé
- testovací statistika:

$$T = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(n_{ij} - n\,p_i\,q_j)^2}{n\,p_i\,q_j} \quad \dots \quad \chi^2(km-1) \quad (\text{pro } k \to \infty,\, m \to \infty)$$

$$p_i = \frac{1}{n}\sum_{j=1}^{m} n_{ij}$$

$$q_j = \frac{1}{n}\sum_{i=1}^{k} n_{ij}$$

- realizaci t ale testujeme na $\chi^2\Big((k-1)(m-1)\Big)$
- realizace testovací statistiky:

n_{ij}: $\quad i \diagdown j$	1	2	3	Σ
1	10	10	40	60
2	20	10	10	40
Σ	30	20	50	100

$p_i\,q_j$: $\quad i \diagdown j$	1	2	3	p_i
1	0.18	0.12	0.3	0.6
2	0.12	0.08	0.2	0.4
q_j	0.3	0.2	0.5	1

$n\,p_i\,q_j$: $\quad i \diagdown j$	1	2	3
1	18	12	30
2	12	8	20

$n_{ij} - n\,p_i\,q_j$: $\quad i \diagdown j$	1	2	3
1	-8	-2	10
2	-8	2	-10

$(n_{ij} - n\,p_i\,q_j)^2$: $\quad i \diagdown j$	1	2	3
1	64	4	100
2	64	4	100

$\frac{(n_{ij}-n\,p_i\,q_j)^2}{n\,p_i\,q_j}$: $\quad i \diagdown j$	1	2	3
1	64/18	4/12	100/30
2	64/12	4/8	100/20

$$t = 1 + \frac{1}{18} \doteq 18.056$$

- kvantil: $q_{\chi^2(2)}(0.95) \doteq 5.992$
- Nulovou hypotézu H_0 zamítáme.

Řešení 5:

- nulová hypotéza H_0: naměřené hodnoty nejsou korelovány (korelace je nulová)
- alternativní hypotéza H_1: naměřené hodnoty jsou korelovány
- testovací statistika:

$$T = \frac{R_{(X,Y)}\sqrt{n-2}}{\sqrt{1-R^2_{(X,Y)}}} \quad \dots \ t(n-2)\mathrm{t}(n-2)$$

$$R_{(X,Y)} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

- realizace testovací statistiky:

	i	1	2	3	4	
výška	x_i	205	155	185	155	$\bar{x} = 175$
váha	y_i	95	55	65	85	$\bar{y} = 75$
	$x_i - \bar{x}$	30	-20	10	-20	
	$y_i - \bar{y}$	20	-20	-10	10	
	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	600	400	-100	-200	$\Sigma = 700$
	$(x_i - \bar{x})^2$	900	400	100	400	$\Sigma = 1800$
	$(y_i - \bar{y})^2$	400	400	100	100	$\Sigma = 1000$

$$r_{(X,Y)} = \frac{700}{\sqrt{1800 \cdot 1000}} \doteq 0.5217$$

$$t = \frac{0.5217 \cdot \sqrt{2}}{\sqrt{1 - 0.5217^2}} \doteq 0.8649$$

- kvantily: $q_{\mathrm{t}(2)}(0.025) \doteq -4.3$, $q_{\mathrm{t}(2)}(0.975) \doteq 4.3$
- Nulovou hypotézu H_0 nezamítáme.