Základy biostatistiky

(MD710P09)
ak. rok 2008/2009

Karel Zvára

karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 5. května 2009)

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \dots, A_r$
- nominální znak s hodnotami $B_1, \dots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** pstí jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \ldots, A_r$
- nominální znak s hodnotami $B_1, \ldots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** pstí jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \dots, A_r$
- nominální znak s hodnotami $B_1, \dots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** pstí jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \ldots, A_r$
- nominální znak s hodnotami $B_1, \ldots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** pstí jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \dots, A_r$
- nominální znak s hodnotami $B_1, \dots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** psti jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

nezávislost **nominálních** znaků

- nominální znak s hodnotami $A_1, \dots, A_r$
- nominální znak s hodnotami $B_1, \dots, B_c$
- N_{ij} kolikrát současně A_i a B_j (**sdružené četnosti**)
- **marginální** četnosti

$$N_{i\bullet} = \sum_{j=1}^{c} N_{ij} \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}$$

- **nezávislost** znaků: pro všechny dvojice i, j platí

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j)$$

- charakteristika nezávislosti: z **marginálních** pstí jevů A_i, B_j dokážeme rekonstruovat **sdružené psti** jevů $A_i \cup B_j$

test nezávislosti dvou kvalitativních znaků

- teoretické četnosti (protějšek N_{ij}) –
  četnosti, které **v průměru očekáváme**, platí-li hypotéza

$$o_{ij} = n \cdot \widehat{\mathrm{P}(A_i \cap B_j)} = n \cdot \widehat{\mathrm{P}(A_i)} \cdot \widehat{\mathrm{P}(B_j)} = n \cdot \frac{N_{i\bullet}}{n} \cdot \frac{N_{\bullet j}}{n} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

- H_0 : znaky jsou **nezávislé**

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - o_{ij})^2}{o_{ij}}$$

- nezávislost se zamítá pokud $X^2 \geq \chi^2_{(r-1)(c-1)}(\alpha)$
- stupně volnosti $n - 1 - q = r \cdot c - 1 - (r-1) - (c-1) = r \cdot c - r - c + 1 = (r-1)(c-1)$
- musí být $o_{ij} \geq 5 \ \forall \ (i, j)$ (tj. pro všechny dvojice)

test nezávislosti dvou kvalitativních znaků

- teoretické četnosti (protějšek N_{ij}) –
četnosti, které **v průměru očekáváme**, platí-li hypotéza

$$o_{ij} = n \cdot \widehat{\mathrm{P}(A_i \cap B_j)} = n \cdot \widehat{\mathrm{P}(A_i)} \cdot \widehat{\mathrm{P}(B_j)} = n \cdot \frac{N_{i\bullet}}{n} \cdot \frac{N_{\bullet j}}{n} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

- H_0 : znaky jsou **nezávislé**

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - o_{ij})^2}{o_{ij}}$$

- nezávislost se zamítá pokud $X^2 \geq \chi^2_{(r-1)(c-1)}(\alpha)$
- stupně volnosti $n - 1 - q = r \cdot c - 1 - (r-1) - (c-1) =$
$r \cdot c - r - c + 1 = (r-1)(c-1)$
- musí být $o_{ij} \geq 5 \ \forall \ (i, j)$ (tj. pro všechny dvojice)

test nezávislosti dvou kvalitativních znaků

- teoretické četnosti (protějšek N_{ij}) –
  četnosti, které **v průměru očekáváme**, platí-li hypotéza

$$o_{ij} = n \cdot \widehat{\mathrm{P}(A_i \cap B_j)} = n \cdot \widehat{\mathrm{P}(A_i)} \cdot \widehat{\mathrm{P}(B_j)} = n \cdot \frac{N_{i\bullet}}{n} \cdot \frac{N_{\bullet j}}{n} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

- H_0 : znaky jsou **nezávislé**

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - o_{ij})^2}{o_{ij}}$$

- nezávislost se zamítá pokud $X^2 \geq \chi^2_{(r-1)(c-1)}(\alpha)$
- stupně volnosti $n - 1 - q = r \cdot c - 1 - (r-1) - (c-1) = r \cdot c - r - c + 1 = (r-1)(c-1)$
- musí být $o_{ij} \geq 5 \ \forall \ (i, j)$ (tj. pro všechny dvojice)

test nezávislosti dvou kvalitativních znaků

- teoretické četnosti (protějšek N_{ij}) –
  četnosti, které **v průměru očekáváme**, platí-li hypotéza

$$o_{ij} = n \cdot \widehat{\mathrm{P}(A_i \cap B_j)} = n \cdot \widehat{\mathrm{P}(A_i)} \cdot \widehat{\mathrm{P}(B_j)} = n \cdot \frac{N_{i\bullet}}{n} \cdot \frac{N_{\bullet j}}{n} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

- H_0 : znaky jsou **nezávislé**

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - o_{ij})^2}{o_{ij}}$$

- nezávislost se zamítá pokud $X^2 \geq \chi^2_{(r-1)(c-1)}(\alpha)$
- stupně volnosti $n - 1 - q = r \cdot c - 1 - (r-1) - (c-1) = r \cdot c - r - c + 1 = (r-1)(c-1)$
- musí být $o_{ij} \geq 5 \ \forall \ (i,j)$ (tj. pro všechny dvojice)

test nezávislosti dvou kvalitativních znaků

- teoretické četnosti (protějšek N_{ij}) –
  četnosti, které **v průměru očekáváme**, platí-li hypotéza

$$o_{ij} = n \cdot \widehat{\mathrm{P}(A_i \cap B_j)} = n \cdot \widehat{\mathrm{P}(A_i)} \cdot \widehat{\mathrm{P}(B_j)} = n \cdot \frac{N_{i\bullet}}{n} \cdot \frac{N_{\bullet j}}{n} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

- H_0 : znaky jsou **nezávislé**

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - o_{ij})^2}{o_{ij}}$$

- nezávislost se zamítá pokud $X^2 \geq \chi^2_{(r-1)(c-1)}(\alpha)$
- stupně volnosti $n - 1 - q = r \cdot c - 1 - (r-1) - (c-1) = r \cdot c - r - c + 1 = (r-1)(c-1)$
- musí být $o_{ij} \geq 5 \ \forall \ (i,j)$ (tj. pro všechny dvojice)

příklad: kouření u mužů

data: lchs

empirické sdružené a marg. četnosti

vzdělání	zákl.	odb.	mat.	VŠ	celk.
nekuřák	14	55	55	73	197
bývalý k.	11	28	44	42	125
kuřák	14	24	24	17	79
silný k.	78	189	175	106	548
celkem	117	296	298	238	949

očekávané sdružené a marg. četnosti

vzdělání	zákl.	odb.	mat.	VŠ	celk.
nekuřák	24,3	61,4	61,9	49,4	197
bývalý k.	15,4	39,0	39,3	31,3	125
kuřák	9,7	24,6	24,8	19,8	79
silný k.	67,6	170,9	172,1	137,4	548
celkem	117	296	298	238	949

$$\chi^2 = \frac{(14 - 24{,}3)^2}{24{,}3} + \dots + \frac{(106 - 137{,}4)^2}{137{,}4} = 38{,}68$$

$$f = (4-1)(4-1) = 9$$

$$p < 0{,}0001$$

[chisq.test(matrix(c(14,11,14,8,55,28,24,189,55,44,24,175,73,42,17,106),nr=4,nc=4))]

závislost jsme na 5% hladině prokázali

příklad **Baden**

barva očí	barva vlasů				celkem
	světlá	hnědá	černá	ryšavá	
modrá	1 768	807	189	47	2 811
šedá/zelená	946	1 387	746	53	3 132
hnědá	115	438	288	16	857
celkem	2 829	2 632	1 223	116	6 800

- barva očí $r = 3$, barva vlasů $c = 4$, $n = 6800$
- $o_{11} = 2811 \cdot 2829/6800 = 1169\ldots$
- $o_{34} = 116 \cdot 857/6800 = 14{,}62 \geq 5$

$$\chi^2 = \frac{(1768 - 1169)^2}{1169} + \frac{(807 - 1088)^2}{1088} + \ldots = 1073{,}5$$

$$> \chi^2_6(0{,}05) = 12{,}5916$$

$$p < 0{,}0001$$

závislost je na každé rozumné hladině **prokázána**

příklad **Baden**

barva očí	barva vlasů				celkem
	světlá	hnědá	černá	ryšavá	
modrá	1 768	807	189	47	2 811
šedá/zelená	946	1 387	746	53	3 132
hnědá	115	438	288	16	857
celkem	2 829	2 632	1 223	116	6 800

- barva očí $r = 3$, barva vlasů $c = 4$, $n = 6800$
- $o_{11} = 2811 \cdot 2829/6800 = 1169\ldots$
- $o_{34} = 116 \cdot 857/6800 = 14{,}62 \geq 5$

$$\chi^2 = \frac{(1768 - 1169)^2}{1169} + \frac{(807 - 1088)^2}{1088} + \ldots = 1073{,}5$$

$$> \chi^2_6(0{,}05) = 12{,}5916$$

$$p < 0{,}0001$$

závislost je na každé rozumné hladině **prokázána**

příklad **Baden**

barva očí	barva vlasů				celkem
	světlá	hnědá	černá	ryšavá	
modrá	1 768	807	189	47	2 811
šedá/zelená	946	1 387	746	53	3 132
hnědá	115	438	288	16	857
celkem	2 829	2 632	1 223	116	6 800

- barva očí $r = 3$, barva vlasů $c = 4$, $n = 6800$
- $o_{11} = 2811 \cdot 2829/6800 = 1169\ldots$
- $o_{34} = 116 \cdot 857/6800 = 14{,}62 \geq 5$

$$\chi^2 = \frac{(1768 - 1169)^2}{1169} + \frac{(807 - 1088)^2}{1088} + \ldots = 1073{,}5$$

$$> \chi^2_6(0{,}05) = 12{,}5916$$

$$p < 0{,}0001$$

závislost je na každé rozumné hladině **prokázána**

test **homogenity**

- hodnoty znaku $B_1, \dots, B_c$
- r **nezávislých** výběrů z různých populací
- H_0 : populace se **neliší**
- dál stejně jako pro nezávislost
- příklad **krevní skupiny**

populace	skupina				celkem
	0	A	B	AB	
C	121	120	79	33	353
D	118	95	121	30	364
celkem	239	215	200	63	717

$$\chi^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + \dots = 11{,}742 > \chi^2_3(0{,}05) = 7{,}815$$

nejm. teoretická četnost: $353 \cdot 63/717 = 31{,}02 > 5$, $p = 0{,}8\ \%$

test **homogenity**

- hodnoty znaku $B_1, \ldots, B_c$
- r **nezávislých** výběrů z různých populací
- H_0 : populace se **neliší**
- dál stejně jako pro nezávislost
- příklad **krevní skupiny**

populace	skupina				celkem
	0	A	B	AB	
C	121	120	79	33	353
D	118	95	121	30	364
celkem	239	215	200	63	717

$$\chi^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + \ldots = 11{,}742 > \chi^2_3(0{,}05) = 7{,}815$$

nejm. teoretická četnost: $353 \cdot 63/717 = 31{,}02 > 5$, $p = 0{,}8\ \%$

test **homogenity**

- hodnoty znaku $B_1, \ldots, B_c$
- r **nezávislých** výběrů z různých populací
- H_0 : populace se **neliší**
- dál stejně jako pro nezávislost
- příklad **krevní skupiny**

populace	skupina				celkem
	0	A	B	AB	
C	121	120	79	33	353
D	118	95	121	30	364
celkem	239	215	200	63	717

$$\chi^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + \ldots = 11{,}742 > \chi^2_3(0{,}05) = 7{,}815$$

nejm. teoretická četnost: $353 \cdot 63/717 = 31{,}02 > 5$, $p = 0{,}8\ \%$

test **homogenity**

- hodnoty znaku $B_1, \ldots, B_c$
- r **nezávislých** výběrů z různých populací
- H_0 : populace se **neliší**
- dál stejně jako pro nezávislost
- příklad **krevní skupiny**

populace	skupina				celkem
	0	A	B	AB	
C	121	120	79	33	353
D	118	95	121	30	364
celkem	239	215	200	63	717

$$\chi^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + \ldots = 11{,}742 > \chi^2_3(0{,}05) = 7{,}815$$

nejm. teoretická četnost: $353 \cdot 63/717 = 31{,}02 > 5$, $p = 0{,}8\ \%$

test **homogenity**

- hodnoty znaku $B_1, \ldots, B_c$
- r **nezávislých** výběrů z různých populací
- H_0 : populace se **neliší**
- dál stejně jako pro nezávislost
- příklad **krevní skupiny**

populace	skupina				celkem
	0	A	B	AB	
C	121	120	79	33	353
D	118	95	121	30	364
celkem	239	215	200	63	717

$$\chi^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + \ldots = 11{,}742 > \chi^2_3(0{,}05) = 7{,}815$$

nejm. teoretická četnost: $353 \cdot 63/717 = 31{,}02 > 5$, $p = 0{,}8\ \%$

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \ldots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j} \sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \dots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j}\sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \ldots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j}\sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \ldots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j} \sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \dots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j}\sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \ldots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j} \sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

McNemarův test (test symetrie)

nezaměňovat s testem nezávislosti!

- **párový** test pro nominální veličinu s hodnotami $B_1, \ldots, B_k$
- zjišťujeme hodnoty nominálního znaku na **stejných** objektech za **dvojích** okolností (před ošetřením, po ošetření)
- N_{ij} počet objektů, u nichž první měření B_i a druhé měření B_j
- **nulová hypotéza**: pravděpodobnosti možných hodnot znaku jsou **stejné** za obojích okolností (před ošetřením i po něm)

$$X^2 = \sum_{i<j}\sum \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}}$$

- hypotézu zamítneme při $X^2 \geq \chi^2_{k(k-1)/2}(\alpha)$
- výrazy ve jmenovateli musí být kladné!
- nezávisí na počtu objektů, kdy vyšly oba výsledky stejně

příklad **stromy**

1994	1995			celkem
	1	2	3	
1	4	3	3	10
2	7	21	11	39
3	1	15	35	51
celkem	12	39	49	100

- stav týchž stromů ve dvou sezónách
- celkem 100 stromů

$$\chi^2 = \frac{(3-7)^2}{3+7} + \frac{(3-1)^2}{3+1} + \frac{(11-15)^2}{11+15} = 3{,}215$$

- $\chi^2_3(0{,}05) = 7{,}8147, \quad p = 36{,}0\ \%$
- rozdíl mezi sezónami jsme neprokázali
- [mcnemar.test(matrix(c(4,7,1,3,21,15,3,11,35),3,3))]

příklad **stromy**

1994	1995			celkem
	1	2	3	
1	4	3	3	10
2	7	21	11	39
3	1	15	35	51
celkem	12	39	49	100

- stav týchž stromů ve dvou sezónách
- celkem 100 stromů

$$\chi^2 = \frac{(3-7)^2}{3+7} + \frac{(3-1)^2}{3+1} + \frac{(11-15)^2}{11+15} = 3{,}215$$

- $\chi^2_3(0{,}05) = 7{,}8147, \quad p = 36{,}0\ \%$
- rozdíl mezi sezónami jsme neprokázali
- [mcnemar.test(matrix(c(4,7,1,3,21,15,3,11,35),3,3))]

příklad **stromy**

1994	1995			celkem
	1	2	3	
1	4	3	3	10
2	7	21	11	39
3	1	15	35	51
celkem	12	39	49	100

- stav týchž stromů ve dvou sezónách
- celkem 100 stromů

$$\chi^2 = \frac{(3-7)^2}{3+7} + \frac{(3-1)^2}{3+1} + \frac{(11-15)^2}{11+15} = 3{,}215$$

- $\chi^2_3(0{,}05) = 7{,}8147, \quad p = 36{,}0\ \%$
- rozdíl mezi sezónami jsme neprokázali
- [mcnemar.test(matrix(c(4,7,1,3,21,15,3,11,35),3,3))]

příklad **stromy**

1994	1995			celkem
	1	2	3	
1	4	3	3	10
2	7	21	11	39
3	1	15	35	51
celkem	12	39	49	100

- stav týchž stromů ve dvou sezónách
- celkem 100 stromů

$$\chi^2 = \frac{(3-7)^2}{3+7} + \frac{(3-1)^2}{3+1} + \frac{(11-15)^2}{11+15} = 3{,}215$$

- $\chi^2_3(0{,}05) = 7{,}8147, \quad p = 36{,}0\ \%$
- rozdíl mezi sezónami jsme neprokázali
- [mcnemar.test(matrix(c(4,7,1,3,21,15,3,11,35),3,3))]

příklad **stromy**

1994	1995			celkem
	1	2	3	
1	4	3	3	10
2	7	21	11	39
3	1	15	35	51
celkem	12	39	49	100

- stav týchž stromů ve dvou sezónách
- celkem 100 stromů

$$\chi^2 = \frac{(3-7)^2}{3+7} + \frac{(3-1)^2}{3+1} + \frac{(11-15)^2}{11+15} = 3{,}215$$

- $\chi^2_3(0{,}05) = 7{,}8147, \quad p = 36{,}0\ \%$
- rozdíl mezi sezónami jsme neprokázali
- [mcnemar.test(matrix(c(4,7,1,3,21,15,3,11,35),3,3))]

čtyřpolní tabulka
znovu test nezávislosti či homogenity

a c	b d	$a+b$ $c+d$
$a+c$	$b+d$	n

- speciální případ kontingenční tabulky pro $r = c = 2$
- test nezávislosti i test homogenity
  statistiku lze upravit na pohodlnější vyjádření

$$X^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

zamítá se pro $X^2 \geq \chi_1^2(\alpha) = z(\alpha/2)^2$

čtyřpolní tabulka

znovu test nezávislosti či homogenity

a c	b d	$a+b$ $c+d$
$a+c$	$b+d$	n

- speciální případ kontingenční tabulky pro $r = c = 2$
- test nezávislosti i test homogenity

  statistiku lze upravit na pohodlnější vyjádření

$$X^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

zamítá se pro $X^2 \geq \chi_1^2(\alpha) = z(\alpha/2)^2$

případ malých četností

- je-li některá očekávaná četnost malá, pak lze u čtyřpolní tabulky použít upravený postup: **Yatesova korekce**

$$X_Y^2 = \frac{n(|ad - bc| - n/2)^2}{(a+c)(b+d)(a+b)(c+d)}$$

- **Fisherův exaktní** test počítá přímo dosaženou hladinu p
- pro tabulku s velkými četnostmi je výpočet Fisherova testu výpočetně náročný (paměťové nároky, trvání výpočtu)
- existuje zobecnění Fisherova testu i pro větší tabulky, než je čtyřpolní

případ malých četností

- je-li některá očekávaná četnost malá, pak lze u čtyřpolní tabulky použít upravený postup: **Yatesova korekce**

$$X_Y^2 = \frac{n(|ad - bc| - n/2)^2}{(a+c)(b+d)(a+b)(c+d)}$$

- **Fisherův exaktní** test počítá přímo dosaženou hladinu p
- pro tabulku s velkými četnostmi je výpočet Fisherova testu výpočetně náročný (paměťové nároky, trvání výpočtu)
- existuje zobecnění Fisherova testu i pro větší tabulky, než je čtyřpolní

případ malých četností

- je-li některá očekávaná četnost malá, pak lze u čtyřpolní tabulky použít upravený postup: **Yatesova korekce**

$$X_Y^2 = \frac{n(|ad - bc| - n/2)^2}{(a+c)(b+d)(a+b)(c+d)}$$

- **Fisherův exaktní** test počítá přímo dosaženou hladinu p
- pro tabulku s velkými četnostmi je výpočet Fisherova testu výpočetně náročný (paměťové nároky, trvání výpočtu)
- existuje zobecnění Fisherova testu i pro větší tabulky, než je čtyřpolní

případ malých četností

- je-li některá očekávaná četnost malá, pak lze u čtyřpolní tabulky použít upravený postup: **Yatesova korekce**

$$X_Y^2 = \frac{n(|ad - bc| - n/2)^2}{(a+c)(b+d)(a+b)(c+d)}$$

- **Fisherův exaktní** test počítá přímo dosaženou hladinu p
- pro tabulku s velkými četnostmi je výpočet Fisherova testu výpočetně náročný (paměťové nároky, trvání výpočtu)
- existuje zobecnění Fisherova testu i pro větší tabulky, než je čtyřpolní

příklad **hraboš**

Frenkelia spp.	*Sarcocystis spp.*		celkem
	+	−	
+	4	27	31
−	11	473	484
celkem	15	500	515

- souvisí spolu nákazy dvěma cizopasníky?
- nulová hypotéza: **nezávislost**

$$\chi^2 = \frac{515(4 \cdot 473 - 11 \cdot 27)^2}{15 \cdot 500 \cdot 31 \cdot 484} = 11{,}643, \quad p = 0{,}06\ \%$$

- [chisq.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

Frenkelia spp.	*Sarcocystis spp.*		celkem
	+	−	
+	4	27	31
−	11	473	484
celkem	15	500	515

- souvisí spolu nákazy dvěma cizopasníky?
- nulová hypotéza: **nezávislost**

$$\chi^2 = \frac{515(4 \cdot 473 - 11 \cdot 27)^2}{15 \cdot 500 \cdot 31 \cdot 484} = 11{,}643, \quad p = 0{,}06\ \%$$

- [chisq.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

Frenkelia spp.	*Sarcocystis spp.*		celkem
	+	−	
+	4	27	31
−	11	473	484
celkem	15	500	515

- souvisí spolu nákazy dvěma cizopasníky?
- nulová hypotéza: **nezávislost**

$$\chi^2 = \frac{515(4 \cdot 473 - 11 \cdot 27)^2}{15 \cdot 500 \cdot 31 \cdot 484} = 11{,}643, \quad p = 0{,}06\ \%$$

- [chisq.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187$ $p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187 \qquad p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187$ $p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187$ $p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187$ $p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad **hraboš**

- nejmenší očekávaná četnost: $15 \cdot 31/515 = 0{,}9 < 5$
- **Yates**: $\chi^2 = 8{,}187$ $p = 0{,}42\ \%$
  [chisq.test(matrix(c(4,11,27,473),2,2))]
- **Fisherův test**: $p = 0{,}92\ \%$
  [fisher.test(matrix(c(4,11,27,473),2,2))]
- na 5% hladině závislost **prokázána**
- **vyskytují se dvojí cizopasníci se stejnou pstí?**
  (zcela jiná otázka, než na nezávislost)
- odpověď dá McNemarův test:

$$\chi^2 = \frac{(11-27)^2}{11+27} = 6{,}7368, \qquad p = 0{,}94\ \%$$

[mcnemar.test(matrix(c(4,11,27,473),2,2),correct=FALSE)]

příklad: barva květů a tvar pylových zrnek
(jiný postup)

- připomeňme data

barva	purpurová	červená	celkem
oválný tvar	296	27	323
kulatý tvar	19	85	104
celkem	315	112	427

- kdybychom neznali předem teoretické poměry u barvy a tvaru, použijeme běžný postup pro čtyřpolní tabulku

$$\chi^2 = \frac{427 \cdot (296 \cdot 85 - 19 \cdot 27)^2}{315 \cdot 112 \cdot 323 \cdot 104} = 218{,}9$$

- porovnat s $\chi^2_1(0{,}05) = 3{,}84$ a nikoliv s $\chi^2_3(0{,}05) = 7{,}81$
- nyní marginální psti odhadujeme, kdežto v 11. přednášce (obrázek 203) jsme je znali

příklad: barva květů a tvar pylových zrnek
(jiný postup)

- připomeňme data

barva	purpurová	červená	celkem
oválný tvar	296	27	323
kulatý tvar	19	85	104
celkem	315	112	427

- kdybychom neznali předem teoretické poměry u barvy a tvaru, použijeme běžný postup pro čtyřpolní tabulku

$$\chi^2 = \frac{427 \cdot (296 \cdot 85 - 19 \cdot 27)^2}{315 \cdot 112 \cdot 323 \cdot 104} = 218{,}9$$

- porovnat s $\chi_1^2(0{,}05) = 3{,}84$ a nikoliv s $\chi_3^2(0{,}05) = 7{,}81$
- nyní marginální psti odhadujeme, kdežto v 11. přednášce (obrázek 203) jsme je znali

příklad: barva květů a tvar pylových zrnek
(jiný postup)

- připomeňme data

barva	purpurová	červená	celkem
oválný tvar	296	27	323
kulatý tvar	19	85	104
celkem	315	112	427

- kdybychom neznali předem teoretické poměry u barvy a tvaru, použijeme běžný postup pro čtyřpolní tabulku

$$\chi^2 = \frac{427 \cdot (296 \cdot 85 - 19 \cdot 27)^2}{315 \cdot 112 \cdot 323 \cdot 104} = 218{,}9$$

- porovnat s $\chi^2_1(0{,}05) = 3{,}84$ a nikoliv s $\chi^2_3(0{,}05) = 7{,}81$
- nyní marginální psti odhadujeme, kdežto v 11. přednášce (obrázek 203) jsme je znali

příklad: barva květů a tvar pylových zrnek
(jiný postup)

- připomeňme data

barva	purpurová	červená	celkem
oválný tvar	296	27	323
kulatý tvar	19	85	104
celkem	315	112	427

- kdybychom neznali předem teoretické poměry u barvy a tvaru, použijeme běžný postup pro čtyřpolní tabulku

$$\chi^2 = \frac{427 \cdot (296 \cdot 85 - 19 \cdot 27)^2}{315 \cdot 112 \cdot 323 \cdot 104} = 218{,}9$$

- porovnat s $\chi^2_1(0{,}05) = 3{,}84$ a nikoliv s $\chi^2_3(0{,}05) = 7{,}81$
- nyní marginální psti odhadujeme, kdežto v 11. přednášce (obrázek 203) jsme je znali