proLékaře.cz | 3.7.2025

Analýza dat v neurologii

XXIII. Kontingenční tabulky neslouží jen pro testy hypotézy o nezávislosti znaků

L. Dušek, T. Pavlík,
J. Jarkovský, J. Koptíková
Institut biostatistiky a analýz
Masarykova univerzita, Brno

✉
doc. RNDr. Ladislav Dušek, Dr.
Institut biostatistiky a analýz
Masarykova univerzita, Brno
e-mail: dusek@cba.muni.cz

V předchozích dílech seriálu jsme se věnovali hodnocení hypotézy o nezávislosti dvou kategoriálních znaků. Tyto analýzy jsou postaveny na tabulkách četností obsahujících počty jedinců, u kterých sledujeme výskyt dvou znaků. Znaky mohou být binární (ano/ne) nebo kategoriální a podle toho je výsledkem tabulka velikosti 2 × 2 nebo větší. Jako příklad uveďme hodnocení nezávislosti tří druhů nemocí (znak *A*, kategoriální) a výskytu anémie u pacientů (znak *B*, binární).

Testy hypotézy o nezávislosti dvou znaků (proměnných) jsou nejčastějším využitím kontingenčních tabulek, avšak nikoli jediným. Další možnosti analýz závisí na způsobu, jak a za jakým účelem byly pozorované četnosti sbírány. Test hypotézy o nezávislosti (díl XXI) totiž vyžaduje, aby oba znaky byly sledovány zcela rovnocenně a nezávisle. U každého prvku (jedince) v náhodném výběru o velikosti *N* jsou současně a jednorázově zjišťovány kategorie obou znaků. Při sběru dat za tímto účelem považujeme oba znaky za zcela rovnocenné a jejich hodnocení nepodmiňujeme žádnými předpoklady o jejich vztahu. Ostatně právě jejich vztah je předmětem hodnocení.

Jiným příkladem využití kontingenční tabulky je **tzv. test hypotézy homogenity**. Odborně řečeno, jde o hodnocení podmíněných rozdělení jedné proměnné v rámci kategorií proměnné druhé. V české literatuře někdy též nalezneme termín testování hypotézy o shodnosti struktury. Jedna z proměnných vystupuje jako třídicí (stratifikační) faktor pro druhou proměnnou. Zde tedy nepřistupujeme ke znakům rovnocenně, neboť nás zajímá výskyt kategorií jednoho znaku proti znaku druhému. Třídicí proměnnou již při samotném sběru dat kontrolujeme, neboť v rámci jejích kategorií provádíme sledování proměnné druhé. Příkladem může být sledování, zda se dvě skupiny pacientů (diagnózy A, B) liší věkovou strukturou, kterou vyjádříme pěti kategoriemi (< 35 let, 36–45 let, …, > 65 let). Výsledná tabulka četností 2 × 5 obsahuje počty pacientů ve věkových třídách v rámci dvou skupin daných diagnózami A, B. Nutným předpokladem hodnocení je náhodný výběr každé subpopulace určené třídicím faktorem.

Je evidentní, že testy homogenity pracují s hypotézou rovnosti řádkových anebo sloupcových relativních četností v kontingenční tabulce. Vhodným nástrojem je v tomto případě binomický test, srovnávající dvě relativní četnosti, ovšem hodnocení hypotézy homogenity lze provést i „klasickým" testem pro kontingenční tabulku s využitím statistiky χ^2. O binomickém testu se nezmiňujeme v našem seriálu poprvé, v jeho nejjednodušší podobě jsme jej popsali v díle XIV. Zde v příkladu 1 uvádíme složitější výpočet, který dokumentuje hodnocení homogenity šesti nezávislých výběrů pacientů z různých nemocnic. U pacientů sledujeme výskyt komplikace X (ano/ne) a výsledkem je tedy tabulka četností 6 × 2. Nemocnice vystupuje v roli kontrolovaného, třídicího faktoru. Testujeme, zda se výběry z nemocnic shodují v relativním výskytu znaku X a pokud ne, tak které nemocnice se vzájemně shodují a které nikoli. Obdobně lze řešit i sumarizaci dat z literatury. Studujeme-li výskyt komplikace X u léčených pacientů a získáme výsledky od několika autorů, logicky je každé sledování na jiné populaci o jiné velikosti *N*. Test homogenity potom provádíme stejným postupem jako u nemocnic v příkladu 1. Pokud testem potvrdíme platnost hypotézy homogenity, můžeme příslušné kategorie třídicího znaku sloučit a četnost znaku popisovat na spojeném souboru.

Tab. 1 představuje tabulku četností, která je základem pro hodnocení nezávislosti výskytu znaků a hypotézy homogenity. Podstatné je, že oba typy testů lze provést na naprosto stejných datech. A pokud využijeme statistiku χ^2, pak i stejným testovacím postupem. V obou případech je vstupem tabulka četností 2 × 2 nebo větší. Správná volba výpočtu a interpretace výsledku se tedy odvíjí od znalosti daného problému a od způsobu sběru dat. Počítačové programy v tomto příliš nepomohou, zde je úloha analytika dat nezastupitelná.

Jiný přístup k hodnocení tabulek četností představuje **test hypotézy o symetrii** neboli hodnocení binárních proměnných při párovém uspořádání experimentu. V zahraniční literatuře se často používá označení „categorical data observed in matched pairs". Při párovém uspořádání sledujeme u téhož jedince (subjektu) binární proměnnou (ano/ne) dvakrát z nějakého důvodu, který vyplývá z experimentálního plánu. Například:

- sledování výskytu bolesti před operací a po ní, vždy u téhož pacienta,
- sledování výskytu nežádoucích účinků při léčbě X a při léčbě Y, opět u stejného pacienta, kde léčba Y následuje po léčbě X,
- sledování četnosti pozitivního výsledku u dvou laboratorních metod (např. záchyt infekčního agens ve vzorcích krve) při měření na stejné sadě vzorků,
- obecně sledování výskytu nějakého jevu před pokusem a po něm.

Hlavním cílem párového testu je zjistit, zda došlo ke změně výskytu daného znaku. Výstupem měření je opět jed-

dusek.indd 591 12.10.2010 15:36:13

proLékaře.cz | 3.7.2025

Popis problému: Sledujeme relativní četnost pacientů, kteří mají znak X, mezi pacienty šesti nemocnic. Zajímá nás, zda je relativní zastoupení znaku X stejné (homogenní), nebo zda se mezi nemocnicemi liší.

a) Celková tabulka

Nemocnice	Počet pacientů ve vzorku	Četnost pacientů se znakem X	Relativní četnost pacientů se znakem X
1	30	26	0,87
2	16	4	0,25
3	20	5	0,25
4	28	8	0,29
5	17	4	0,24
6	22	5	0,23

b) Tabulka s vyloučením nemocnice s největším příspěvkem k testové statistice χ^2

Nemocnice	Počet pacientů ve vzorku	Četnost pacientů se znakem X	Relativní četnost pacientů se znakem X
2	16	4	0,25
3	20	5	0,25
4	28	8	0,29
5	17	4	0,24
6	22	5	0,23

1. Posuzujeme zastoupení pacientů se znakem X v 6 nemocnicích (tabulka a).
2. Nulová hypotéza: Relativní četnost pacientů se znakem X je ve všech nemocnicích shodná.
3. Pro výpočet využijeme vztah pro posouzení shody řádků a sloupců R × C kontingenční tabulky

$$\chi^2_{(\nu=(r-1)(c-1))} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{i,j}-F_{i,j})^2}{F_{i,j}}$$

kde $f_{i,j}$ jsou pozorované četnosti v buňkách tabulky, $F_{i,j}$ jsou očekávané četnosti v buňkách tabulky.

4. Očekávané četnosti jsou kalkulovány podle vztahu

$$F_{i,j} = \frac{\sum f_{i,\cdot} \cdot \sum f_{\cdot,j}}{N}$$

5. Po dosazení do vztahu v bodě 3 získáme

$$\chi^2_{(\nu=5)} = 37{,}02$$

6. Kritická hodnota testové statistiky χ^2 pro $\alpha = 0{,}05$ je $\chi^2_{0,95(5)} = 11{,}07$
7. Hodnota $\chi^2 = 37{,}02$ je větší než nalezená kritická hodnota $\chi^2_{0,95(5)} = 11{,}07$. Nulovou hypotézu tedy zamítáme.
8. Na hladině významnosti $\alpha = 0{,}05$ jsme zamítli hypotézu o shodě zastoupení pacientů se znakem X v jednotlivých nemocnicích. Datové soubory z nemocnic není možné na základě tohoto závěru sloučit.
9. Po vyřazení nemocnice 1 (tabulka b) s největším příspěvkem k celkové hodnotě statistiky χ^2 získáme následující výsledek

$$\chi^2_{(\nu=4)} = 0{,}27$$

10. Závěr: Nyní je hodnota $\chi^2 = 0{,}27$ menší než kritická hodnota $\chi^2_{0,95(4)} = 9{,}48$; analýza umožňuje sloučení datových souborů z analyzovaných nemocnic na základě potvrzené homogenity relativních četností pacientů se znakem X.

Příklad 1. Test homogenity relativních četností ve vícenásobných nezávislých výběrech.

Tab. 1. Příklad tabulky četností dvou binárních proměnných nabývajících hodnot 1/0.

		Proměnná II		Řádkové součty	
		1	0		Absolutní četnosti (počty jedinců, subjektů) a–d jsou získány prostým náhodným výběrem. Z tabulky lze počítat sloupcové a řádkové relativní četnosti, např.: $p_a = a/(a + c)$ nebo $p_a = a/(a + b)$
proměnná I	1	a	b	a + b	
	0	c	d	c + d	
sloupcové součty		a + c	b + d	N = a + b + c + d	

Hodnocená hypotéza, předpoklady a popis testu		Test
nezávislost proměnných I a II	Prostý náhodný výběr *N* jedinců ze sledované populace. Obě proměnné jsou sledovány rovnocenně a nezávisle.	χ^2 test
homogenita relativních četností	Jedna z proměnných vystupuje jako třídicí faktor. V rámci skupin daných jejími hodnotami (1/0) je sledován výskyt hodnot druhé proměnné prostým náhodným výběrem. Testujeme shodu rozdělení této druhé proměnné mezi skupinami danými první proměnnou jako třídicím faktorem.	binomický test χ^2 test

dusek.indd 592 12.10.2010 15:36:14

proLékaře.cz | 3.7.2025

Popis problému: Sledujeme výsledek stanovení infekčního agens dvěma rozdílnými kultivačními metodami na sadě *N* vzorků. Obě metody jsou nasazovány na stejné vzorky, sledování tedy nejsou nezávislá. Výsledek hodnotíme jako binární (dichotomickou) proměnnou +/-. Testujeme hypotézu shody výsledků obou metod ve dvou různých příkladech: 2-A, 2-B.

Příklad 2-A.

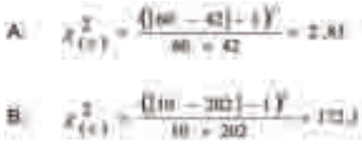
Metoda 1	Metoda 2	Četnost
+	+	202 = a
+	–	60 = b
–	+	42 = c
–	–	10 = d

Příklad 2-B.

Metoda 1	Metoda 2	Četnost
+	+	40 = a
+	–	10 = b
–	+	202 = c
–	–	60 = d

1. Srovnáváme výsledky dvou metod stanovení infekčního agens kultivací vzorků krve; výsledek hodnotíme jako binární proměnnou (+/-).
2. Pomocí McNemarova testu testujeme shodu rozdělení (H_0: metoda 1 je stejně úspěšná při stanovení infekčního agens jako metoda 2).
3. McNemarův test je založen na vztahu

kde *b* a *c* jsou pozorované četnosti kombinací +/- v buňkách tabulky.
4. Po dosazení do vztahu vypočítáme hodnotu testové statistiky:



5. Vypočtenou hodnotu testové statistiky porovnáme s kritickou hodnotou testu pro $\alpha = 0,05$ $\chi^2_{(1-\alpha)}{}^{(v=1)} = 3,84$.
6. Výsledné hodnoty statistické významnosti jsou:
   A: $\chi^2_{(c)} < \chi^2_{(1-\alpha)}{}^{(v=1)}$, $p > 0,05$ – nezamítáme H_0.
   B: $\chi^2_{(c)} > \chi^2_{(1-\alpha)}{}^{(v=1)}$, $p < 0,05$ – zamítáme H_0.

Příklad 2. McNemarův test pro párové srovnání dichotomických proměnných.

Tab. 2. Tabulka četností dvou párově sledovaných binárních proměnných nabývajících hodnot 1/0.

Kombinace možných výsledků párových sledování*		Absolutní četnost	Testování hypotézy symetrie, předpoklady
Sledování I	Sledování II		Test symetrie hodnotí hypotézu shody rozdělení (zde výskytu binárního znaku) mezi dvěma párově spojenými proměnnými (zde označeno jako sledování I a II). Zásadní význam mají četnosti b, c, neboť právě rozdíl mezi nimi určuje rozdíl mezi proměnnými I a II.
1	1	a	
1	0	b	
0	1	c	
0	0	d	

*Sledování nejsou nezávislá. Mohou představovat výsledky metod I a II (hodnoceno jako 1/0), které vzájemně srovnáváme na stejné sadě vzorků. Jiným příkladem může být opakované hodnocení v čase (I → II), kdy u *N* pacientů sledujeme výskyt nějakého znaku (1/0). I zde lze hodnotit dílčí relativní četnosti, např. podíly kladných výsledků (kódované jako 1): $p_I = (a + b)/N$, $p_{II} = (a + c)/N$. Nebo relativní četnost výsledků, kde se obě sledování (proměnné) shodují: $(a + d)/N$.

noduchá tabulka četností 2 × 2. Při sledování párových dichotomických proměnných ji zapisujeme způsobem, který obecně uvádí tab. 2. Při párovém uspořádání ale nelze provést χ^2 test, neboť pozorování zaznamenaná v tabulce četností nejsou nezávislá. Sledujeme v podstatě rozdíl ve výskytu daného znaku (pozitivita testu, výskyt bolesti) mezi dvěma vzájemně provázanými měřeními. Využijeme-li zápis v tab. 2 jako příklad, pak nás logicky zajímají pouze četnosti *b*, *c*, které jedině rozhodují o rozdílu mezi měřeními. Testujeme tedy, zda rozdíl v četnostech *b*, *c* mezi měřeními je náhodný, nebo zda je statisticky významný, což signalizuje změnu výskytu znaku. Klasickým postupem je zde McNemarův test, jehož výpočet přibližuje příklad 2. Tento test řadíme mezi neparametrické metody a jeho výpočet lze aplikovat pouze na tabulky o velikosti 2 × 2. Hovoříme o testování rozdílů u párově sbíraných nominálních dat.

Rozšířením McNemarova testu pro více než dva závislé výběry při sledování binární proměnné je tzv. Cochranův test (neboli Q-test dle Cochrana). Jde o neparametrické zobecnění McNemarova testu, kde sledovanou binární proměnnou tří-

dusek.indd 593 12.10.2010 15:36:14

proLékaře.cz | 3.7.2025

díme podle více vzájemně nějak závislých kritérií (skupin) a sledujeme, zda se výskyt sledovaného znaku mění. Tak jako u McNemarova testu je sledovaná proměnná vždy binární (ano/ne). Příkladem využití mohou být tyto experimentální situace:

- srovnáváme výskyt pozitivního výsledku u více laboratorních metod a testujeme je závisle vždy na stejném vzorku; máme tedy *N* opakování (vzorků) a hypotézou je shoda mezi *k* metodami, kde $k > 2$,
- sledujeme výskyt určitého jevu (bolesti, nežádoucího účinku) u stejných pacientů opakovaně v čase; experiment o *N* opakováních (pacienti) srovnávající výskyt bolesti v *k* opakovaných měřeních, opět $k > 2$,
- sledujeme výskyt nějaké reakce v chování pokusných zvířat, která postupně podrobujeme různým pokusným zásahům; zde je počet opakování dán počtem zvířat v experimentu, srovnáváme různé pokusné zásahy.

Vždy se tedy jedná o situaci, kdy je jedinec (subjekt) vícenásobně sledován a je hodnocen výskyt určitého znaku. Provedená pozorování opět nejsou nezávislá a nelze je hodnotit χ^2 testem. Výpočet Cochranova testu je uveden v příkladu 3.

Tak jako u mnoha jiných statistických testů sahá původ těchto metod hluboko do minulého století. Nicméně jde o výpočty stále platné a hojně využívané. Quinn McNemar (1900–1986) byl profesorem psychologie a statistiky univerzity ve Stanfordu a test publikoval v časopise Psychometrika v roce 1947. Tento test patří k jeho nejvýznamnějším dílům, mimo jiné také výrazně přispěl k revizi a vývoji testů IQ. Cochranův Q test je pojmenován po významném skotském statistikovi Williamu G. Cochranovi (1909–1980).

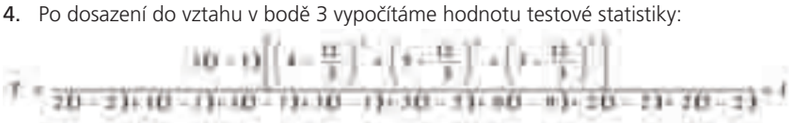
Tento díl seriálu opět zakončíme již téměř tradičním mentorováním o roli člověka ve věku počítačů. Je totiž zřejmé, že na numericky stejných datech (tabulka četností) lze bez problémů současně provést rozdílné testy, které ale hodnotí zcela rozdílné situace a vyžadují specifickou interpretaci výsledku. Hodnocení nezávislosti výskytu znaků nelze zaměnit za zkoumání změn ve výskytu znaku po určitém podnětu. Nicméně ale, když do statistických programů vložíme tabulku četností, například 2 × 2, nabídne nám menu všechny zde probrané testy naráz. Počítač nepřemýšlí, pouze koná. Je na experimentátorovi, lékaři, aby si uvědomil, co testuje a adekvátně zvolil vhodný test. Snad tomu náš seriál aspoň trochu napomáhá :-).

Popis problému: Srovnáváme vliv tří diet na četnost překročení určité hranice obsahu glukózy v krvi u N = 8 diabetiků. Testované diety jsou aplikovány u všech pacientů s určitým odstupem, hodnocení tedy není nezávislé. Testujeme shodu rozdělení u tří proměnných (výsledek hodnocení tří diet), které jsou vzájemně provázány osobou pacienta. Hodnocená proměnná je binární (dichotomická: 1/0).

	Dieta 1	Dieta 2	Dieta 3	Součet řádků
pacient 1	0	1	1	2
pacient 2	0	1	0	1
pacient 3	1	0	0	1
pacient 4	0	1	0	1
pacient 5	1	1	1	3
pacient 6	0	0	0	0
pacient 7	1	0	1	2
pacient 8	1	1	0	2
Součet sloupců	**4**	**5**	**3**	**12**

1. Srovnáváme vliv různých druhů diety u pacientů s diabetem na překročení určité hladiny glukózy v krvi. Sledovanou proměnnou je binární kód (1 = překročení hladiny, 0 = hladina není překročena).
2. Pomocí Cochranova testu srovnáme rozdělení sledované proměnné mezi pokusnými zásahy (dietami)
   (H_0: diety se v četnosti překročení hladiny glukózy v krvi neliší)
3. Cochranův test je dán vztahem

   kde *k* je počet diet, *N* je počet pacientů v experimentu, $X_{.j}$ je součet hodnot „1" u j-té diety, $X_{i.}$ je četnost hodnot „1" u i-tého pacienta, *b* je celkový součet četností hodnot „1".
4. Po dosazení do vztahu v bodě 3 vypočítáme hodnotu testové statistiky:



5. Vypočtená hodnota testové statistiky je porovnána s kritickou hodnotou testu pro $\alpha = 0{,}05$ $\chi^2_{(1-\alpha)}{}^{(v=k-1)} = 5{,}99$.
6. Výsledná hodnota statistické významnosti je:
   $T < \chi^2_{(1-\alpha)}{}^{(v=k-1)}$, $p > 0{,}05$ – nezamítáme H_0, testované diety tedy mají srovnatelný vliv na překročení určité hladiny krevní glukózy.

Příklad 3. Cochranův Q test pro srovnání dichotomických proměnných ve více než dvou závislých pokusných variantách.

Literatura

Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley 1981.
McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947; 12(2): 153–157.

dusek.indd 594 12.10.2010 15:36:15