proLékaře.cz | 6.11.2026

Analýza dat v neurologii

LXI. Závěrečné příklady k analýzám trendů v kontingenčních tabulkách

L. Dušek, T. Pavlík,
J. Jarkovský, J. Koptíková

Institut biostatistiky a analýz
Masarykova univerzita, Brno

✉
doc. RNDr. Ladislav Dušek, Ph.D.
Institut biostatistiky a analýz
MU, Brno
e-mail: dusek@iba.muni.cz

Minulý díl jsme věnovali analýze tzv. trendové složky v tabulkách četností s cílem vysvětlit postupy užitečné při prokazování kauzality vztahu „expozice–účinek". Pokud totiž prokážeme, že s rostoucí úrovní expozice roste i její efekt, půjde s vysokou pravděpodobností o vztah příčinný. Tímto dílem ukončíme téma analýzy složitějších kontingenčních tabulek několika příklady, které ukáží možnost analýzy trendů ještě z jiného pohledu.

Zvláštním typem testu pro hodnocení trendu v tabulkách četností je tzv. Cochranův-Armitageův test. Jeho základní výpočet přibližuje příklad 1. Test byl navržen pro situace, kdy experiment hodnotí binární proměnnou proti proměnné ordinální. Binární proměnná (např. výskyt nemoci ano/ne) typicky představuje odpověď systému na expozici, vyjadřuje tedy následek expozice. Ordinální škála jako druhá dimenze tabulky četností naopak představuje různé úrovně studované expozice, např. dávky léku, dobu trvání působení nějakého faktoru apod. V podstatě zde studujeme, zda se vzájemný poměr hodnot binárního faktoru (poměr relativního výskytu hodnoty „ano" k hodnotě „ne") liší v závislosti na hodnotách dané ordinální škály. Jinými slovy, zda míra expozice ovlivňuje např. relativní riziko vzniku dané nemoci. Nulová hypotéza tohoto testu je neexistence takového trendu. To znamená, že pokud tuto hypotézu nezamítáme, pak jsou relativní četnosti hodnoty „ano" u binární proměnné stejné přes všechny úrovně proměnné ordinální.

Příklad 1 dokládá, že tento test pracuje se systémem vah, jejichž nastavení umožňuje studovat různé hypotetické „lokální" trendy v původní tabulce četností. Rozdělení výsledné testové statistiky se blíží ke standardizovanému normálnímu rozdělení $N(0,1)$, a tudíž s jejími hodnotami lze přímo pracovat jako s kvantily tohoto modelového rozdělení. Možnost nastavení vah je velkou výhodou ve studiích „dávka–odpověď", zejména pokud ordinální proměnná zahrnuje více než tři hodnoty; můžeme takto testovat hypotetické vzájemné relace těchto hodnot.

Cochranův-Armitageův test je velmi často využíván jako standard pro analýzy linearity trendu v genetických asociačních studiích (typicky studie případů a kontrol). Velmi jednoduchou ukázku těchto výpočtů přináší příklad 2, ze kterého je patrné, že možné kombinace alel genu *a, A* zde řadíme

Cochranův-Armitageův test pro trend se používá v analýze kategoriálních dat, kdy je cílem hodnotit vztah mezi binární proměnnou a kategoriální proměnnou tvořenou seřaditelnými (ordinálními) kategoriemi. Test je často používán pro hodnocení trendů v genetických asociačních studiích případů a kontrol. Test má větší sílu testu než Chí-kvadrát test pro prokázání vztahu proměnných v případě, že v datech existuje předpokládaný trend.

Data pro výpočet testu mají tvar následující kontingenční tabulky:

	B = 1	**B = 2**	**B = 3**	**Celkem**
A = 1	N_{11}	N_{12}	N_{13}	R_1
A = 2	N_{21}	N_{22}	N_{23}	R_2
Celkem	C_1	C_2	C_3	N

Proměnná *B* představuje ordinální škálu seřazenou dle rostoucí hodnoty a proměnná *A* je binární odpovědí systému na vliv proměnné *B*. Testová statistika testu je definována následovně:

$$T = \sum_{i=1}^{k} t_i \left(N_{1i} R_2 - N_{2i} R_1 \right),$$

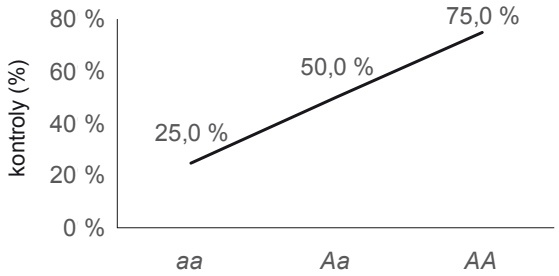
kde t_i je váha nastavovaná dle testovaného předpokládaného trendu. Např. pokud předpokládáme, že $B = 1$ a $B = 2$ mají podobné četnosti a $B = 3$ odlišné, nastavíme váhy na $t = (1, 1, 0)$, v případě předpokládaného lineárního trendu nastavíme $t = (0, 1, 2)$. Nastavením vah tak můžeme studovat různé hypotetické „lokální" trendy v původní tabulce četností a výsledek testu tak vypovídá o tom, jak se tyto varianty liší od reality naměřené ve zdrojové tabulce četností.

Příklad 1. Cochranův-Armitageův test pro trend v kontingenční tabulce.

proLékaře.cz | 6.11.2026

Typickou aplikaci Cochranova-Armitageova testu dokládá příklad genetické asociační studie případů a kontrol. Relativní četnost případů (nemocných určitou chorobou) je sledována ve srovnání s kontrolním ramenem ve vazbě na tři varianty genotypu určitého genu.

	Genotyp		
Rameno studie	***aa***	***Aa***	***AA***
kontroly	10	20	30
případy	30	20	10
kontroly	25,0 %	50,0 %	75,0 %
případy	75,0 %	50,0 %	25,0 %



V analýze předpokládáme lineární trend s nastavením t = (0, 1, 2). Po dosazení do rovnice získáme:

$$T = \sum_{i=1}^{k} t_i \left(N_{1i}R_2 - N_{2i}R_1\right) = t_1\left(N_{11}R_2 - N_{21}R_1\right) + t_2\left(N_{12}R_2 - N_{22}R_1\right) + t_3\left(N_{13}R_2 - N_{23}R_1\right) =$$

$$= 0 \times (10 \times 60 - 30 \times 60) + 1 \times (20 \times 60 - 20 \times 60) + 2 \times (30 \times 60 - 10 \times 60) = 2\,400 \Rightarrow \mathbf{p < 0{,}001}$$

Výsledek Cochranova-Armitageova testu: $p < 0{,}001$, a tedy byl prokázán statisticky významný trend ve vztahu obou proměnných. Váhy t vysvětlené v příkladu 1 zde lze využít pro testování různých pravděpodobnostních modelů vzájemných vztahu alel *a*, *A* daného genu.

Příklad 2. Aplikace Cochranova-Armitageova testu pro trend v kontingenční tabulce z genetické asociační studie.

Kontingenční tabulky s alespoň jednou ordinální proměnnou s více než dvěma kategoriemi můžeme využít pro výpočet analýzy trendu. Jedním z dobře interpretovatelných přístupů je srovnání hodnocených kategorií s kategorií referenční (typicky hodnotově nebo nejméně riziková kategorie, např. v tomto příkladu nejmladší pacienti).

V našem příkladu hodnotíme závislost výskytu onemocnění na kategorii věku pacienta. Jako referenční kategorie jsou použiti nejmladší pacienti, použitou statistikou je poměr šancí (Odds Ratio; *OR*).

A) Jednoznačný nárůst výskytu onemocnění s věkem

	Věková kategorie		
Onemocnění	**20–40**	**40–60**	**> 60**
ne	90	60	35
ano	10	20	35
%	10 %	25 %	50 %
OR	referenční	3,0	9,0

Závěr: S vyššími věkovými kategoriemi narůstá i šance na výskyt onemocnění oproti referenční kategorii 20–40 let; v kategorii 40–60 let je *OR* = 3,0; v kategorii > 60 let je *OR* = 9,0.

B) Výskyt onemocnění s věkem neroste

	Věková kategorie		
Onemocnění	**20–40**	**40–60**	**> 60**
ne	90	60	35
ano	28	20	10
%	24 %	25 %	22 %
OR	referenční	1,1	0,9

Závěr: Ve vyšších věkových kategoriích je šance na výskyt onemocnění obdobná jako v referenční kategorii 20–40 let; v kategorii 40–60 let je *OR* = 1,1; v kategorii > 60 let je *OR* = 0,9.

Příklad 3. Ukázka komplexního posouzení trendové složky v asociační studii zkoumající vztah mezi věkem a výskytem nemoci.

do ordinální škály *aa*–*Aa*–*AA*, která následně vstupuje do testu. I tady můžeme pracovat se systémem vah, které byly vysvětleny v příkladu 1 (váhy označeny jako t_i, kde i nabývá hodnot od 1 do počtu kategorií ordinální škály). Například dominanci vlivu alely *a* na výskyt nemoci by jako dílčí test ověřila sada vah t = (1, 1, 0), která vyřadí z testování kombinaci *AA*. Naopak dominanci alely *A* by

proLékaře.cz | 6.11.2026

testovala sada vah $t = (0, 1, 1)$, kodominanci obou alel potom optimálně sada $t = (0, 1, 2)$. Tato možnost dílčích testování je i u genetických studií velmi vítaná, avšak u plánovaných a protokolárně vedených studií nesmí být aplikována dle volby nebo zdání analytika „ad hoc". Testované hypotézy a odpovídající váhy musí být nastaveny předem v protokolu studie a vlastní hodnocení dat musí tomuto plánu odpovídat.

Cochranův-Armitageův test v našem seriálu představuje další ilustrativní příklad nestárnoucí historie biostatistiky. Ačkoli jde o test velmi starý (publikovaný v letech 1954–1955), jeho význam nejenže neupadá, ale s rostoucí četností pozorovacích studií zahrnujících genetické markery dokonce narůstá. Autory testu také v našem seriálu neuvádíme poprvé. Jméno W. G. Cochrana (1909–1980) nese řada významných biostatistických postupů, např. z oblasti analýz tabulek četností je to Cochranův Q test nebo Cochranova-Mantelova-Haenszelova statistika. Rovněž dosud žijící P. Armitage (nar. 1924) patří mezi významné biostatistiky minulého století; mimo jiné působil i jako prezident vlivné mezinárodní společnosti pro klinickou biostatistiku.

Posledním příkladem k problematice trendu v kontingenčních tabulkách dokumentujeme postup výpočtu využívající odhad poměru šancí (*OR*). Studie uvedená v příkladu 3 ověřuje, zda výskyt zkoumané nemoci trendově závisí na věku pacientů. Sbíraná data odpovídají běžné tabulce četností 2×3, věk zde vystupuje jako proměnná o třech kategoriích a výskyt nemoci je zaznamenáván v nejjednodušším módu, tedy binárně. Platí-li, že se riziko výskytu nemoci s věkem zvyšuje, pak bychom pro věkové kategorie měli pozorovat gradient hodnot *OR* vztažených k určité referenční kategorii (jako referenci příklad kalkuluje nejmladší zkoumanou věkovou kategorii). Existující gradient v hodnotách poměru šancí potom logicky potvrzuje statistický test trendové složky.

Tímto dílem uzavíráme výklad testů trendů v kontingenčních tabulkách a zároveň i problematiku asociačních studií. V následujících dílech se budeme krátce věnovat problematice tzv. ekologických (korelačních) studií a dále potom analýze korelací spojitých faktorů.

Literatura

Cochran WG. Some methods of strengthening the common c2 tests. Biometrics 1954;10:417–51.
Armitage P. Tests for linear trends in proportions and frequencies. Biometrics 1955;11:375–86.
Agresti A. Categorical Data Analysis. New York: John Wiley and Sons 2002.