proLékaře.cz | 29.4.2024

Analýza dat v neurologii

LI. Yuleovo *Q* – užitečný nástroj pro srovnávání různých odhadů poměru šancí

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 seriálu jsme zakončili poznámkou o velkém číselném rozsahu možných hodnot poměru šancí (*OR*), který může být problémem při vzájemném srovnávání výstupů různých studií. Připomeňme, že poměr šancí probíráme z různých pohledů od dílu 35 našeho seriálu jako velmi významný ukazatel míry vztahu „expozice–účinek" v observačních studiích. *OR* počítáme podle vztahu uvedeného pod následující nejjednodušší možnou tabulkou četností 2 × 2 (tab. 1). Jelikož v polích tabulky *a–d* jsou uvedeny četnosti dané vzorkovacím plánem studie, může odhad *OR* nabývat jakékoli hodnoty od 0 do nekonečna. V tomto dílu seriálu se pokusíme odpovědět na otázku, která v této souvislosti jistě mnohé čtenáře napadla. A sice, zda existují nějaké ukazatele umožňující snadné srovnávání odhadů poměru šancí publikovaných různými studiemi, aniž musíme provádět složité statistické testy?

Nespornou výhodou poměru šancí je jistě to, že v rámci jedné práce umožňuje vzájemně srovnat míru asociace více faktorů s cílovým parametrem; číselné vyjádření *OR* včetně intervalu spolehlivosti je velice instruktivní, ať již v tabulkách nebo grafech. O dost obtížněji se však srovnávají odhady *OR* publikované různými studiemi, neboť jejich hodnota závisí na mnoha faktorech, od velikosti vzorku až po statistickou techniku, která je pro odhad použita. Hodnoty *OR* také mohou u problematicky rozložených vzorků nabývat až extrémně vysokých hodnot. Zatímco spodní hranice možných hodnot

Způsob výpočtu poměru šancí (jde o poměr poměrů) vede v různých studiích k číselně velmi heterogenním výstupům. Příklad dokumentuje několik takových situací včetně komentáře příčin daného číselného výsledku.

Riziková skupina	Událost ne	Událost ano
ne	45	15
ano	20	20

$$OR = \frac{20/20}{15/45}$$

OR (95% IS): 3,0 (1,3; 7,0)

Běžný výsledek výpočtu, poměr šancí kvantifikuje vztah mezi rizikovým faktorem a událostí.

Riziková skupina	Událost ne	Událost ano
ne	4	1
ano	1	1

$$OR = \frac{1/1}{1/4}$$

OR (95% IS): 4,0 (0,1; 137,0)

Poměr šancí nabývá číselně běžné hodnoty, nicméně rozsah jeho 95% intervalu spolehlivosti je v důsledku malé velikosti vzorku extrémní. Je zřejmé, že odhad je nestabilní a velmi málo spolehlivý.

Riziková skupina	Událost ne	Událost ano
ne	60	5
ano	3	40

$$OR = \frac{40/3}{5/60}$$

OR (95% IS): 160,0 (36,2; 707,3)

Velmi silný vztah mezi rizikovým faktorem a událostí vede k velmi vysokým hodnotám poměru šancí i hranicím jeho intervalu spolehlivosti.

Riziková skupina	Událost ne	Událost ano
ne	100	1
ano	2	60

$$OR = \frac{60/2}{1/100}$$

OR (95% IS): 3 000,0 (266,3; 33 796,6)

Extrémně silný vztah mezi rizikovým faktorem a událostí vede k extrémním hodnotám poměru šancí i jeho intervalu spolehlivosti.

Příklad 1. Ukázky výsledků výpočtu poměru šancí *(OR)* za různých podmínek.

proLékaře.cz | 29.4.2024

OR je jasně ohraničena nulou, horní hranice prakticky neexistuje.

V příkladu 1 uvádíme několik případů dokládajících číselnou heterogenitu odhadů *OR*.

V praxi takto můžeme čelit úkolu znázornit ve srovnávacím grafu či tabulce hodnoty *OR*, které se mohou lišit až o dva řády. Takové srovnání v grafu samozřejmě postrádá smysl. Odhady *OR* lze sice zlogaritmovat (log*(OR)*) a tak je snáze vykreslit do jednoho grafu, ale takto transformované hodnoty jsou obtížně interpretovatelné a čitelné. Pro srovnání různých odhadů *OR* je možné také využít statistické testy (viz díl 37 seriálu), které sice pomohou určit, zda je rozdíl mezi různými odhady *OR* statisticky významný, ale nenabídnou nám lehce použitelnou škálu pro kvantitativní porovnání různých hodnot.

Tento problém má naštěstí velmi snadné řešení a tím je transformace odhadu *OR* do statistiky zvané **Yuleovo *Q* (Yule's *Q*)**. Je nazvána po svém autorovi, skotském statistikovi jménem George Udny Yule (1871–1951). Jde v podstatě o transformaci převádějící hodnoty *OR* do jasně čitelné škály od –1 do +1, což z ní činí velmi užitečný nástroj pro vzájemné srovnávání a řazení i číselně velmi odlišných odhadů *OR*.

Tab. 1. Výpočet poměru šancí (*OR*) na základě tabulky četností 2 × 2.

Dělení na případy a kontroly dle výskytu sledovaného jevu (nemoci; u retrospektivní studie případů a kontrol jde o první kritérium výběru osob)	**Dělení dle expozice (vlivu sledované události) jedinců v minulosti** (u retrospektivní studie zjištěno zpětně)	
	expozice (událost): ano	**expozice (událost): ne**
případy	*a*	*b*
kontroly	*c*	*d*

Šance výskytu jevu u exponované a neexponované skupiny: $O_E = a/c$; $O_{NE} = b/d$.
Výpočet poměru šancí provádíme podle vztahu: $OR = ad/bc$.

Yuleovo *Q* vypočítáme přímo z odhadu *OR* velmi jednoduše podle vztahu:

$Q = (OR - 1)/(OR + 1)$,

nebo alternativně přímo z čtyřpolní tabulky četností:

$Q = (bc - ad)/(bc + ad)$.

Yuleovo *Q* se řadí mezi tzv. kontingenční koeficienty, kterým jsme věnovali velký prostor v díle 21 našeho seriálu. Statistika *Q* má řadu zajímavých vlastností. Například nemá žádné modelové statistické rozdělení („*distribution-free statistics*") a nezávisí na marginálních rozděleních četností ve zdrojové tabulce (Awosami et al, 1999; Warrens, 2008). Jeho hlavním omezením je, že znaky dávané do vzájemné asociace musí být povinně binarizovány, neboť Yuleovo *Q* patří mezi ukazatele určené pouze pro tabulky četností 2 × 2. Vynásobíme-li absolutní hodnotu *Q* konstantou 100, můžeme jej interpretovat

Yuleovo *Q* (Yule's *Q*) slouží k transformaci odhadu poměrů šancí (*OR*) do škály s rozsahem –1 až +1, která umožňuje snadné srovnání a interpretaci i velmi rozdílných číselných *OR*. Yuleovo *Q* je vypočteno z poměru šancí (*OR*) pomocí následujícího vztahu:

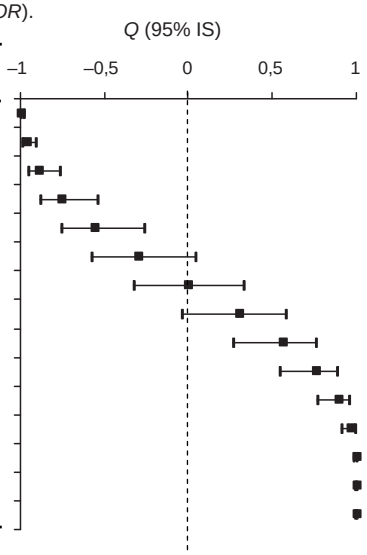
$$Q = \frac{OR - 1}{OR + 1}$$

Alternativní možností je potom výpočet přímo ze čtyřpolní tabulky četností:

$$Q = \frac{bc - ad}{bc + ad}$$

Následující tabulka demonstruje použití Yuleova *Q* pro srovnání číselně heterogenních poměrů šancí (*OR*).

Riziková skupina	ne		ano			
Událost	ne	ano	ne	ano	***OR* (95% IS)**	***Q* (95% IS)**
Studie 1	4	61	64	1	0,001 (0,000; 0,009)	–0,998 (–1,000; –0,982)
Studie 2	9	56	59	6	0,016 (0,005; 0,049)	–0,968 (–0,990; –0,907)
Studie 3	14	51	54	11	0,056 (0,023; 0,134)	–0,894 (–0,955; –0,764)
Studie 4	19	46	49	16	0,135 (0,062; 0,293)	–0,762 (–0,883; –0,547)
Studie 5	24	41	44	21	0,279 (0,135; 0,576)	–0,563 (–0,762; –0,269)
Studie 6	29	36	39	26	0,537 (0,268; 1,078)	–0,301 (–0,577; 0,038)
Studie 7	34	31	34	31	1,000 (0,502; 1,990)	0,000 (–0,332; 0,331)
Studie 8	39	26	29	36	1,862 (0,928; 3,738)	0,301 (–0,037; 0,578)
Studie 9	44	21	24	41	3,579 (1,736; 7,382)	0,563 (0,269; 0,761)
Studie 10	49	16	19	46	7,414 (3,408; 16,131)	0,762 (0,546; 0,883)
Studie 11	54	11	14	51	17,883 (7,436; 43,009)	0,894 (0,763; 0,955)
Studie 12	59	6	9	56	61,185 (20,452; 183,044)	0,968 (0,907; 0,989)
Studie 13	64	1	4	61	976,000 (106,084; 8 979,479)	0,998 (0,981; 1,000)
Studie 14	69	1	2	66	2 277,000 (201,645; 25 712,134)	0,999 (0,990; 1,000)
Studie 15	74	1	1	71	5 254,000 (322,411; 85 618,998)	1,000 (0,994; 1,000)



Číselný rozsah odhadů poměrů šancí mezi jednotlivými studiemi je extrémní (od téměř 0 do více než 5 000) a v původních hodnotách vylučuje smysluplnou vizualizaci. Použitím Yuleova *Q* poměry šancí transformujeme do škály od –1 do +1 a umožníme vizuální srovnání.

Příklad 2. Yuleovo *Q* jako nástroj transformace heterogenních odhadů poměru šancí *(OR)*.

proLékaře.cz | 29.4.2024

Příklad popisuje výpočet Yuleova *Q* a jeho statistické významnosti pomocí *Z* testu. V souboru pacientů je zjišťován vztah mezi výskytem rizikové skupiny pacientů a výskytem události.

Riziková skupina	Událost ne	Událost ano
ano	a	b
ne	c	d

Riziková skupina	Událost ne	Událost ano
ano	24	41
ne	44	21

Poměr šancí (*OR*) a Yuleovo *Q* jsou spočítány standardním způsobem.

$$OR = \frac{b/a}{d/c} = \frac{41/24}{21/44} = 3{,}579 \quad \Rightarrow \quad Q = \frac{OR-1}{OR+1} = \frac{bc-ad}{bc+ad} = 0{,}563$$

OR (95% IS): 3,579 (1,736; 7,382) *Q* (95% IS): 0,563 (0,269; 0,761)

Statistická významnost koeficientu *Q* je počítána pomocí Z testu s nulovou hypotézou *Q* = 0. V prvním kroku je koeficient *Q* převeden na Z statistiku se standardizovaným normálním rozdělením podle následujícího vztahu:

$$Z = \frac{Q}{\sqrt{0{,}25*(1-Q^2)^2\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}} = \frac{0{,}563}{\sqrt{0{,}25*(1-0{,}563^2)^2\left(\frac{1}{24}+\frac{1}{41}+\frac{1}{44}+\frac{1}{21}\right)}} = 6{,}090$$

V druhém kroku je vypočtená hodnota Z statistiky srovnána s kvantilem standardizovaného normálního rozdělení, postup je shodný jako při jiných parametrických testech, kdy v případě hladiny statistické významnosti α = 0,05 je jako hranice pro statistickou významnost použita hodnota $z_{1-\alpha/2}$ = 1,96.

Výsledná hodnota Z statistiky ve výše uvedeném výpočtu je za touto kritickou hranicí, proto zamítáme nulovou hypotézu, že *Q* = 0, čímž potvrzujeme statisticky významný vztah mezi charakteristikou pacientů a výskytem události. Skutečná statistická významnost je pro dané Z menší než 0,001.

Příklad 3. Hodnocení statistické významnosti Yuleova *Q*.

v procentech. Takto zjednodušeně řečeno vyjadřuje relativní redukci chyby, kterou děláme, když z jednoho znaku v tabulce četností predikujeme hodnoty znaku druhého (Baddie a Fred, 1995).

Jak již bylo řečeno, hlavním přínosem Yuleova *Q* je standardizace hodnot *OR* do intervalu < −1; +1 >. Hodnoty *Q* blízké nule znamenají malou závislost asociovaných znaků a naopak hodnoty blízké −1 nebo +1 znamenají velmi silnou závislost. V literatuře můžeme nalézt i podrobnější dělení škály pro Yuleovo *Q*, které mohou být interpretačně velmi užitečné. Jako příklad uvádíme škálu pro *Q* publikovanou Bohrnstedtem a Knokem v roce 1988:

- hodnoty *Q* 0–0,24: žádný nebo zanedbatelný vztah proměnných,
- hodnoty *Q* 0,25–0,49: slabý vztah,
- hodnoty *Q* 0,50–0,74: středně silný vztah,
- hodnoty *Q* 0,75–1,00: velmi silný vztah.

Velkou výhodou výpočtu *Q* je, že jej lze aplikovat na již publikované odhady *OR*. Tedy bez znalosti primárních dat a bez dalších předpokladů o rozdělení zapojených statistik. Ukázku transformace i velmi rozdílných hodnot *OR* pomocí Yuleova *Q* přináší příklad 2. Je patrné, že aplikace tohoto jednoduchého nástroje značně zvyšuje čitelnost a usnadňuje interpretaci rozdílných odhadů *OR*.

I u statistiky *Q* se můžeme opřít o hodnocení její významnosti, což vzhledem k jejím hodnotám znamená testovat platnost hypotézy *Q* = 0 (tedy hypotézy neexistence závislosti „expozice–účinek" v tabulce četností). Statistický test využívá standardizovaného normálního rozdělení (testová statistika *Z*). Příklad 3 je ukázkou tohoto testu.

Je zajímavé, jak se statistický ukazatel starší více než 100 let stále užitečně uplatňuje v prezentování výsledků současného výzkumu. Nadto i v relativně soudobé literatuře nalezneme metodické práce rozšiřující informační potenciál nebo usilující o nové aplikace Yuleova *Q*. Jako příklad můžeme uvést práci Lipsitze a Fitzmaurice z roku 1994 zaměřenou na uplatnění *Q* statistiky při analýze vícerozměrných binárních dat. Statistika *Q* dále nalézá velké uplatnění v moderní analýze sociologických průzkumů a v hodnocení edukačních programů, kde je zkoumání vlivu různých prediktorů na efektivitu vzdělávání velmi časté (Adeyemi, 2003).

Pozn.: Určitou modifikací statistiky *Q* je tzv. **Yuleovo *Y***, které počítáme dle následujícího vztahu: $Y = (\sqrt{OR} - 1)/(\sqrt{OR} + 1)$. Statistika *Y* je také nazývána koeficient koligace, zatímco statistika *Q* je nazývána koeficient asociace. Rozdíl je zde v uplatnění druhé odmocniny odhadu *OR*. Rovněž škála statistiky *Y* nabývá hodnot od −1 do +1, v tomto smyslu není při jejím uplatnění žádný rozdíl od *Q*. Yuleovo *Y* transformuje hodnoty *OR* konzervativněji než statistika *Q*.

Literatura

Adeyemi TO. Statistical Techniques for educational research. Lagos: Universal Publisher 2003.

Awosami RO, Olopoenia S, Adeyefa O. Economics for senior secondary school certificate examinations. Ibadan: Spectrum Books Ltd. 1999.

Baddie E, Fred H. Advantages in social research: data analysis using SPSS for Windows the thousands oaks. California: Pine Forge Press a Sage Publications Company 1995.

Bohrnstedt GW, Knoke D. Statistics for social data analysis. 2nd ed. Itasca, Illinois: F.E. Peacock Publishers, Inc. 1988.

Lipsitz RS, Fitzmaurice G. An extension of Yule's Q to multivariate binary data. Biometrics 1994; 50(3): 847–852.

Yule GU. On the methods of measuring association between two attributes. Journal of the Royal Statistical Society 1912; LXXV: 579–652.

Warrens M. On association coefficients for 2 × 2 tables and properties that do not depend on the marginal distributions. Psychometrika 2008; 73(4): 777–789.