

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 29, číslo 4, prosinec 2018

HODNOCENÍ PÁROVÝCH BINÁRNÍCH DAT: CHYBY VE STATISTICKÝCH ANALÝZÁCH A DOPORUČENÉ METODY

EVALUATION OF PAIRED BINARY DATA: ERRORS IN STATISTICAL ANALYSIS AND RECOMMENDED METHODS

Jan Hendl

Adresa: Univerzita Karlova, Fakulta sociálních věd, Smetanovo nábřeží 6,
11001 Praha 1

E-mail: jan.hendl@fsv.cuni.cz

Abstrakt: Uvažuje se situace párových binárních dat z intervenčních výzkumných plánů a základní postupy při statistickém testování, odhadu rozdílu efektivit intervence, podílu efektivit a podílu šancí úspěchu pro tento případ. Dokumentuje se jedna chybně provedená analýza v uveřejněném vědeckém článku.

Klíčová slova: McNemarův test, párová binární data, interval spolehlivosti, chyby ve statistické analýze.

Abstract: We consider paired binary data from intervention studies and basic methods for statistical testing and estimation of effectivity difference, ratio of proportions and odds ratio for this case. Wrong statistical analysis is documented in a published scientific paper.

Keywords: McNemar test, paired binary data, confidence interval, errors in statistics.

1. Úvod

Při hodnocení intervenčního experimentu se často zajímáme o porovnání spárovaných údajů. Páry údajů vznikají tak, že jsme každý prvek výběrového souboru (o rozsahu n) měřili dvakrát, před aplikací a po ní. K párovým datům se dostaneme i pomocí jiných plánů výzkumu. Například můžeme měřit dvojčata, posuzujeme párové orgány nebo použijeme dvě různé metody měření při měření jedné veličiny. Setkáme se s nimi v retrospektivních studiích případů a kontrol nebo v prospektivních studiích. Pokud zjišťujeme u subjektů kvantitativní metrickou proměnnou a pak hodnotíme, zda intervence působila,

použijeme obvykle párový t -test ve tvaru

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_d} \sqrt{n} = \frac{\bar{d}}{s_{\bar{d}}}, \quad (1)$$

kde $\bar{x}_i$, $i = 1, 2$, jsou průměry proměnné před intervencí a po intervenci vypočítané z n údajů, d označuje rozdíly dvojic měření (kterých je n), odhad s_d směrodatné odchylky těchto rozdílů, $\bar{d}$ průměrný rozdíl a $s_{\bar{d}}$ odhad standardní chyby průměrného rozdílu.

Interval spolehlivosti pro očekávaný rozdíl Δ mezi průměry párových proměnných μ_1 a μ_2 má tvar

$$(\bar{d} - ts_{\bar{d}}; \bar{d} + ts_{\bar{d}}), \quad (2)$$

kde t je hodnota relevantního kvantilu t -rozložení o $n - 1$ stupních volnosti. Určujeme jím hladinu spolehlivosti sestrojeného intervalu spolehlivosti. Předpokládáme, že rozdíly měření jsou normálně rozložené. Hodnota $\Delta = \mu_1 - \mu_2$ představuje míru působení intervence.

Hodnotu $s_{\bar{d}}$ lze spočítat podle vzorce

$$s_{\bar{d}} = \sqrt{(s_1^2 + s_2^2 - 2rs_1s_2)/n}. \quad (3)$$

Výběrové rozptyly v tomto vztahu se týkají hodnoty 1., resp. 2. měření. Korelace r se počítá mezi spárovanými řadami měření. Protože je většinou kladná, je zřejmě interval spolehlivosti spočítaný podle vzorce (2) užší, než kdybychom s korelací obou proměnných vůbec nepočítali. To dokumentujeme pro situaci $n = 25$, $\bar{x}_1 = 0$ a $\bar{x}_2 = 0,5$ a směrodatnou odchylkou pro obě řady měření 1,0 při korelaci $r = 0,0$, resp. $r = 0,4$, kdy má testovací t statistika hodnotu 1,768, resp. 2,282. V prvním případě je výsledek statisticky nevýznamný, v druhém případě statisticky významný na 5% hladině významnosti při dvoustranném testu. První případ někdy nastane, jestliže opomineme spárování údajů a použijeme t -test pro dva nezávislé výběry. S obdobným problémem se setkáváme také při hodnocení binárních proměnných.

V našem pojednání nejdříve naznačíme principy tří asymptoticky shodných obecných metod statistické inference. V další části probereme vybrané metody, které používáme při statistické inferenci spárovaných binárních proměnných. Uvedeme výsledek výpočtů pro několik příkladů. V nich používáme dvoustranný test s hladinou významnosti 0,05 a dvoustranné intervaly spolehlivosti s hladinou spolehlivosti 0,95. Často využijeme jako důsledek statistických úvah standardizované normální rozdělení. Hodnota 0,975-kvantilu tohoto rozdělení, kterou budeme využívat při konstrukci intervalů spolehlivosti s hladinou spolehlivosti 0,95, je $z = 1,96$.

2. Metody testování a intervalového odhadu

Existují tři důležité obecné metody tvorby statistických testů a intervalů spolehlivosti. Využíváme je i při hodnocení spárovaných binárních dat [1], [7]. Vycházejí z věrohodnostní funkce $L(\gamma; \mathbf{x})$. V našem zápisu γ znamená odhádovaný parametr a $\mathbf{x}$ jsou data. Postupy statistické inference pomocí těchto metod jsou asymptoticky ekvivalentní, tzn. že při velkých výběrech dávají podobné výsledky. Budeme probírat pouze případ inference o skalárním parametru γ . U uvedených metod existuje i zobecnění pro vektorový parametr [1, ed. 2002, s. 11–15]. Rozlišujeme

- Waldův test a interval spolehlivosti,
- skórový test a interval spolehlivosti,
- test a interval spolehlivosti poměrem věrohodností.

Předpokládáme, že máme maximálně věrohodný odhad $\hat{\gamma}$ parametru γ . Za dosti obecných podmínek je odhad $\hat{\gamma}$ asymptoticky konsistentní, eficientní a normálně rozložený. Platí asymptoticky

$$\hat{\gamma} \sim N(\gamma; V(\hat{\gamma})),$$

kde V značí rozptyl. Tento vztah můžeme bezprostředně využít ke konstrukci intervalového odhadu parametru γ ve tvaru

$$\hat{\gamma} \pm z\sqrt{\hat{V}(\hat{\gamma})},$$

kde z je kvantil standardizovaného normálního rozložení a určujeme jím předpokládanou hladinu spolehlivosti. Místo $V(\hat{\gamma})$ dosazujeme odhad $\hat{V}(\hat{\gamma})$. Takto vytvořenému intervalu říkáme interval spolehlivosti Waldova typu. Má nevýhodu, že vede k symetrickým krajním bodům kolem bodového odhadu a ty navíc nemusejí mít přijatelné hodnoty. Nesymetrické intervaly spolehlivosti však získáme, když maximálně věrohodný odhad není na původní škále parametru. Například v logistické regresi se intervaly spolehlivosti vypočtou pro logaritmus poměru šancí. Interval spolehlivosti pro tento logaritmus je symetrický, ale jestliže ho transformujeme do škály poměru šancí, vzniklý interval je asymetrický.

Waldova testovací statistika pro test nulové hypotézy $H_0 : \gamma = \gamma_0$ má tvar

$$\frac{|\hat{\gamma} - \gamma_0|}{\sqrt{\hat{V}(\hat{\gamma})}}$$

a srovnáváme ji s vhodným kvantilem standardizovaného normálního rozložení. Její druhá mocnina má za platnosti nulové hypotézy asymptoticky rozložení χ^2 o jednom stupni volnosti.

Objasníme, jak získáme hodnotu $\hat{V}(\hat{\gamma})$. Činíme to pomocí Fisherovy informace o parametru γ . Označíme $l(\gamma; \mathbf{x}) = \ln L(\gamma; \mathbf{x})$. Pozorovaná Fisherova informace je záporně vzatá druhá derivace podle proměnné γ logaritmu věrohodnostní funkce. Hodnota

$$\hat{V}(\hat{\gamma}) = \frac{1}{-l''(\hat{\gamma}; \mathbf{x})}$$

aproximuje rozptyl odhadu $\hat{\gamma}$. Hodnotu $-l''(\hat{\gamma}; \mathbf{x})$ často nahrazujeme funkcí $I(\gamma) = -E[l''(\hat{\gamma}; \mathbf{x})]$. Tuto hodnotu nazýváme očekávaná informace nebo Fisherova informace.

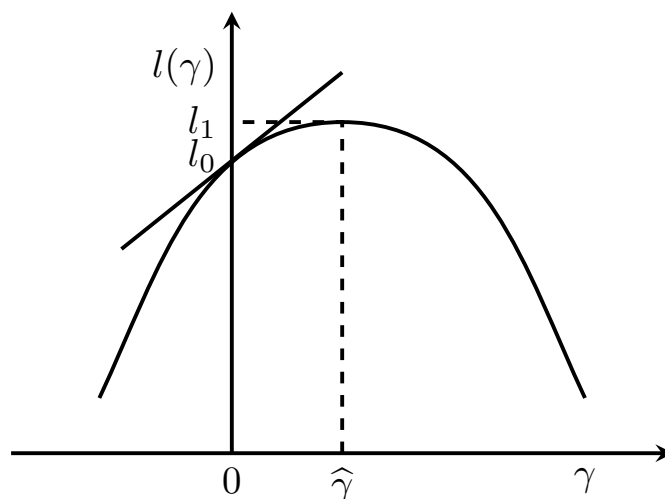
Druhý způsob testu nulové hypotézy $H_0 : \gamma = \gamma_0$ využívá skórovou funkci $u(\gamma) = l'(\gamma)$. Skórová funkce se rovná nule v maximálně věrohodném odhadu $\hat{\gamma}$. Test využívá hodnotu skórové funkce v bodě γ_0 . Testová statistika má tvar

$$\frac{u(\gamma_0)^2}{I(\gamma_0)}.$$

Očekávaná hodnota skórové funkce je nula a její rozptyl se rovná Fisherově informaci. Asymptoticky má skórová funkce standardizované normální rozložení, protože se jedná o součet stejně rozložených nezávislých náhodných proměnných s konečným rozptylem. Proto má uvedená testová statistika asymptoticky rozložení χ^2 o jednom stupni volnosti. Interval spolehlivosti získáme metodou inverze. Tato metoda se provede tak, že hledáme krajní hodnoty parametru γ , kdy uvedená testová statistika nepřekročí kritickou hodnotu. V případě testu Waldova typu inverzní metodou dostaneme interval spolehlivosti Waldova typu.

Třetí způsob využívá porovnání hodnot věrohodnostních funkcí v bodě γ_0 a v maximálně věrohodném odhadu $\hat{\gamma}$. Test metodou poměrem věrohodností srovnává hodnotu $2[l(\hat{\gamma}) - l(\gamma_0)]$ s kritickou hodnotou rozložení χ^2 o jednom stupni volnosti. Interval spolehlivosti pro parametr γ získáme opět inverzní metodou.

Obrázek 1 ilustruje tři popsané principy pro případ $H_0 : \gamma = 0$. Ukazuje schematicky průběh logaritmu věrohodnostní funkce $l(\gamma)$. Statistický test podle Walda využívá chování $l(\gamma)$ v maximálně věrohodném odhadu. Směrodatná chyba odhadu závisí na zakřivení $l(\gamma)$ v tomto bodě. Test pomocí skórové funkce vychází ze směrnice a zakřivení $l(\gamma)$ v bodě $\gamma = 0$. Test pomocí poměru věrohodností kombinuje informaci o $l(\gamma)$ v bodě $\gamma = 0$ a v ma-



Obrázek 1: Průběh logaritmu věrohodnostní funkce a informace použitá ve třech popsaných testech při nulové hypotéze $\gamma = 0$

ximálně věrohodném odhadu parametru ve statistice $-2(l_0 - l_1)$ podle [1]. Tato statistika využívá nejvíce informací ze všech tří typů statistik.

Jako příklad uvedeme konstrukci intervalů spolehlivosti pro parametr π binomického rozložení. Waldova metoda vede k intervalu spolehlivosti

$$\hat{\pi} \pm z\sqrt{\hat{\pi}(1 - \hat{\pi})/n},$$

kde $\hat{\pi} = m/n$ (m znamená četnost úspěšných pokusů z celkem n pokusů). Pomocí metody skórové funkce nalezneme krajní body intervalu spolehlivosti pro π vyřešením rovnice

$$(\hat{\pi} - \pi)\sqrt{\pi(1 - \pi)/n} = \pm z.$$

Musíme vyřešit kvadratickou rovnici v π způsobem, jak popsal Wilson v roce 1927 [14].

Interval spolehlivosti na základě poměru věrohodnostních funkcí je principiálně jednoduchý, ale výpočetně složitější. Hledáme množinu π , pro kterou platí, že dvojnásobek logaritmu poměru věrohodnostních funkcí

$$-2(l_0 - l_1) = 2 \left[(n - m) \ln \frac{1 - \hat{\pi}}{1 - \pi} + m \ln \frac{\hat{\pi}}{\pi} \right],$$

je menší než vhodný kvantil χ^2 rozložení o jednom stupni volnosti.

V dalším výkladu vycházíme z asymptotických vlastností uvedených metod a z větších rozsahů výběrů. Při malých rozsazích výběrů lze použít přesné metody statistické inference.

3. Binární spárovaná data

V případě binárních pozorování zjišťujeme, zda subjekt má (kód = 1), nebo nemá (kód = 0) danou vlastnost. Taková data také získáme, jestliže dichotomizujeme původně kvantitativní metrické údaje určené k hodnocení párovým t -testem. Příslušnou statistickou teorii objasňuje Agresti [1, zvláště kapitoly 10.1 a 14] nebo Anděl [3].

Výsledky popisujeme čtyřpolní tabulkou četností 1.

Tabulka 1: Četnosti spárovaných binárních dat

Četnosti		Po intervenci		Suma
		1	0	
Před intervencí	1	$a = n_{11}$	$b = n_{10}$	n_{1+}
	0	$c = n_{01}$	$d = n_{00}$	n_{0+}
Suma		n_{+1}	n_{+0}	$n = a + b + c + d$

Hodnota a je četnost subjektů s danou vlastností před intervencí i po ní, d je počet subjektů bez vlastnosti před intervencí i po ní, b je počet subjektů s vlastností před intervencí a bez vlastnosti po intervenci, c je počet subjektů bez vlastnosti před intervencí a s vlastností po intervenci. Tabulka 2 popisuje pravděpodobnosti, že získáme pozorování v daném políčku.

Tabulka 2: Pravděpodobnosti jednotlivých kategorií spárovaných binárních dat

Pravděpodobnosti		Po intervenci		Suma
		1	0	
Před intervencí	1	p_{11}	p_{10}	p_{1+}
	0	p_{01}	p_{00}	p_{0+}
Suma		p_{+1}	p_{+0}	1

Uvedme v tabulce 3 příklad výzkumných údajů. Data jsou převzata z knihy [6]. U 237 osob se nejdříve zjišťovalo, zda kouří (kód = 1), nebo ne (kód = 0). Po novinové kampani (intervenci), byly stejné osoby opět otázaný, zda kouří, nebo ne.

Tabulka 3: První soubor dat o působení novinové kampaně na kouření

Četnosti		Po intervenci		Suma
		1	0	
Před intervencí	1	$a = 80$	$b = 25$	105
	0	$c = 12$	$d = 120$	132
Suma		92	145	237

Jestliže hodnoty měření v párech souhlasí, mluvíme někdy o konkordantních párech, v našem příkladu jich je $a + d = 200$. V tabulce je $b + c = 37$ diskordantních párů měření, kdy měření mají rozdílné hodnoty.

Párová binární data z menšího výběru zachycuje tabulka 4. Jde o údaje převzaté z práce [7]. Byly zjišťovány plicní komplikace u dětí (přítomné komplikace kód = 1) před transplantací kmenových buněk (intervence) a po ní.

Tabulka 4: Druhý soubor dat o plicních komplikacích u dětí před transplantací kmenových buněk a po ní

Četnosti		Po intervenci		Suma
		1	0	
Před intervencí	1	$a = 1$	$b = 1$	2
	0	$c = 7$	$d = 12$	19
Suma		8	13	21

4. Jednoduché procedury

Označme p_{1+} , resp. p_{+1} pravděpodobnost měření s hodnotou 1 před intervencí, resp. po ní.

Pak rozdíl $\Delta = p_{1+} - p_{+1}$ vyjadřuje účinek intervence. O velikosti Δ (zda intervence je účinná) rozhodují diskordantní měření, což je patrné z bodového odhadu $\hat{\Delta}$. Bodový odhad $\hat{\Delta}$ je dán rozdílem

$$\hat{\Delta} = \hat{p}_{1+} - \hat{p}_{+1} = \frac{n_{1+}}{n} - \frac{n_{+1}}{n} = \frac{b - c}{n}.$$

Poznamenejme, že kovariance proměnných $\hat{p}_{+1}$ a $\hat{p}_{1+}$ má podle [1, kap. 10.1.1] hodnotu $(p_{11} \cdot p_{00} - p_{10} \cdot p_{01})/n$.

Pro náš první příklad (tab. 3) dostaneme

$$\hat{\Delta} = \frac{b - c}{n} = \frac{25 - 12}{237} = 0,054.$$

Pro druhý příklad (tab. 4) má tento odhad hodnotu

$$\hat{\Delta} = \frac{b - c}{n} = \frac{1 - 7}{21} = -0,286.$$

McNemarův [9] podmíněný asymptotický test nulové hypotézy $\Delta = p_{1+} - p_{+1} = 0$, totiž že intervence nemá žádný účinek, se opírá o testovací statistiku

$$\chi^2 = \frac{(b - c)^2}{(b + c)}. \quad (4)$$

Ta má za platnosti nulové hypotézy $H_0 : \Delta = 0$ asymptoticky rozložení χ^2 o jednom stupni volnosti. Jedná se o nejpoužívanější test této nulové hypotézy. Podmínka zde znamená fixaci součtu četností $b + c$.

Přesnou verzi testu stejné hypotézy uvádí Anděl [3]. Jde o přesný podmíněný binomický test pro rozdělení $\text{Bi}(b + c; 0,5)$. Jeho asymptotický tvar vede k (4), viz [1] nebo [3].

Pro náš první příklad vychází McNemarova testovací statistika

$$\chi^2 = \frac{(25 - 12)^2}{(25 + 12)} = 4,57.$$

V druhém příkladu má McNemarova statistika hodnotu

$$\chi^2 = \frac{(1 - 7)^2}{(1 + 7)} = 4,50.$$

Jelikož příslušná kritická hodnota pro 5% hladinu významnosti je 3,84, můžeme nulovou hypotézu o neexistenci účinku intervence v obou případech na této hladině významnosti zamítnout.

Pro první příklad mají bodové odhady $\hat{p}_{+1}$, $\hat{p}_{1+}$ a $\hat{\Delta}$ hodnoty $\hat{p}_{+1} = 0,388$, $\hat{p}_{1+} = 0,443$ a $\hat{\Delta} = 0,055$. Pro druhý příklad pak $\hat{p}_{+1} = 0,380$, $\hat{p}_{1+} = 0,095$ a $\hat{\Delta} = -0,286$.

Asymptoticky platný statistický test W nulové hypotézy $H_0 : \Delta = 0$ podle Walda a příslušný interval spolehlivosti pro Δ v případě binárních párových

dat mají tvar (viz například kapitola 10.1.1 v [1])

$$\begin{aligned}
 W &= \frac{(\hat{p}_{+1} - \hat{p}_{1+}) \sqrt{n}}{\sqrt{s_{+1}^2 + s_{1+}^2 - 2rs_{+1}s_{1+}}}, \\
 \hat{p}_{+1} - \hat{p}_{1+} &\pm \frac{z}{\sqrt{n}} \sqrt{s_{+1}^2 + s_{1+}^2 - 2rs_{+1}s_{1+}}, \\
 s_{+1}^2 &= \hat{p}_{+1}(1 - \hat{p}_{+1}), \quad s_{1+}^2 = \hat{p}_{1+}(1 - \hat{p}_{1+}),
 \end{aligned} \tag{5}$$

kde r označuje odhad korelace mezi $\hat{p}_{+1}$ a $\hat{p}_{1+}$ a z je kvantil standardizovaného normálního rozdělení pro zvolenou hladinu spolehlivosti. Odhad korelačního koeficientu získáme na základě úvah v [1, ed. 2002, s. 410].

Užitečnější pro vytvoření intervalu spolehlivosti pro Δ jsou vyjádření

$$\begin{aligned}
 \frac{b-c}{n} \pm \frac{z}{n} \sqrt{b+c - \frac{(b-c)^2}{n}}, \text{ resp.} \\
 \hat{p}_{01} - \hat{p}_{10} \pm z \sqrt{[\hat{p}_{01} + \hat{p}_{10} - (\hat{p}_{01} - \hat{p}_{10})^2]/n},
 \end{aligned} \tag{6}$$

kde $\hat{p}_{ij} = n_{ij}/n$.

Směrodatná chyba ve výrazu (6) je počítána podle Waldovy procedury bez předpokladu platnosti nulové hypotézy $H_0 : \Delta = 0$.

Pro naše dva příklady mají intervalové odhady Δ hodnotu $0,054 \pm 0,049 = (0,005; 0,103)$, resp. $-0,286 \pm 0,223 = (-0,508; -0,062)$.

Ukázali jsme, že data pro McNemarův test mohou být přepsána pomocí dvou proměnných X a Y s hodnotami 0 a 1. Na tato data lze použít standardní párový t -test. Získáme hodnotu testovací statistiky t . Nechť T je hodnota McNemarovy testovací statistiky. Lze ukázat, že

$$t^2 = \frac{(n-1)T}{n-T} \quad \text{a} \quad T = \frac{nt^2}{n-1+t^2}.$$

Protože obě funkce jsou rostoucí vzhledem ke svým nezávisle proměnným, jsou ekvivalentní i příslušné statistické testy se stejnou hladinou významnosti a vhodně zvolenými kritickými mezemi.

Někdy se zajímáme o poměr pravděpodobností $\varphi = p_{1+}/p_{+1}$. Bodový odhad nalezneme pomocí výpočtu

$$\hat{\varphi} = \frac{n_{1+}}{n_{+1}}. \tag{7}$$

Jako míra efektu intervencí může být poměr pravděpodobností úspěchu v některých situacích vhodnější než jejich rozdíl, zvláště když tyto pravděpodobnosti jsou velmi malé.

Nechť koeficient EF má hodnotu

$$EF = \exp \left(z \sqrt{\frac{b+c}{n_{1+} \cdot n_{+1}}} \right).$$

Interval spolehlivosti pro $\varphi = p_{1+}/p_{+1}$ dostaneme ve tvaru

$$(\hat{\varphi}/EF; \hat{\varphi} \cdot EF).$$

Pro naše dva příklady dostaneme bodové a intervalové odhady φ jako 1,141 (1,011; 1,287), resp. 0,250 (0,062; 0,992).

Dále se věnujeme poměru šancí. Hlavní využití poměru šancí (tuto hodnotu v souhlase s literaturou označujeme symbolem θ) je v retrospektivních studiích případů a kontrol, které mají za cíl objasnit etiologii nemoci. Pak není obvykle možné odhadnout rozdíl pravděpodobností úspěchu. Agresti [1] rozebírá dva různé modely pro hodnocení šancí v případě spárovaných dat. Rozlišuje dva případy: a) marginální šance a b) podmíněné párové šance. Nechť (Y_1, Y_2) označuje pár měření před intervencí a po ní pro náhodně vybraného jedince, přičemž pro obě proměnné platí, že „1“ označuje kategorii „úspěch“ a „0“ kategorii „neúspěch“. Jestliže vyjdeme z logistického regresního modelu, pak případ a) je popsán rovnicí

$$\text{logit}[P(Y_t = 1)] = a + \ln(\theta)x_t,$$

kde $x_t = 0$, nebo $x_t = 1$ podle toho, zda jedinec byl měřen před, nebo po intervenci. Parametr θ je podíl šancí. Tento model se zaměřuje na marginální rozložení odpovědí pro dvě pozorování (před intervencí a po ní).

Maximálně věrohodný odhad hodnoty θ je v tomto modelu dán vztahem

$$\hat{\theta} = \frac{\hat{p}_{1+}\hat{p}_{+0}}{\hat{p}_{+1}\hat{p}_{0+}} \quad (8)$$

s odhadem asymptotického rozptylu

$$\hat{V}[\ln(\hat{\theta})] = \left\{ \frac{1}{\hat{p}_{1+}\hat{p}_{0+}} + \frac{1}{\hat{p}_{+1}\hat{p}_{+0}} - 2(\hat{p}_{11}\hat{p}_{00} - \hat{p}_{10}\hat{p}_{01})/(\hat{p}_{1+}\hat{p}_{0+}\hat{p}_{+1}\hat{p}_{+0}) \right\} / n.$$

Nesymetrický interval spolehlivosti pro θ dostaneme použitím Waldovy metody vztahem

$$\ln(\hat{\theta}) \pm z \sqrt{\hat{V}[\ln(\hat{\theta})]}.$$

Na rozdíl od případu a), případ b) uvažuje běžnější situaci, ve které se parametr a mění mezi jedinci. Nechť (Y_{i1}, Y_{i2}) označuje měření i -tého jedince před intervencí a po ní.

Logistický regresní model má pak tvar

$$\text{logit}[P(Y_{it} = 1)] = a_i + \ln(\theta_{\text{cond}})x_t.$$

Efekt intervence je vyjádřen pomocí pro všechny jedince společným parametrem $\ln(\theta_{\text{cond}})$, přitom povolujeme jisté interindividuální rozdíly pomocí parametrů a_i . Parametr $\ln(\theta_{\text{cond}})$ odhadujeme za podmínky, že postačující statistiky pro parametry a_i (zde vystupují jako rušivé parametry [3, s. 187]) mají dané hodnoty [1, s. 416–417]. Pro každý subjekt šanci události s kódem 1 je dostaneme jako šanci události s kódem 0 vynásobenou výrazem $\exp(\theta_{\text{cond}})$ [7]. Za uvedeného předpokladu vyjadřujeme θ_{cond} vztahem $\theta_{\text{cond}} = p_{10}/p_{01}$, tedy podílem pravděpodobností diskordandních párů. Index cond označuje, že jde při výpočtu podílu šancí o případ b).

Podmíněný maximálně věrohodný odhad θ_{cond} za podmínky fixního součtu $b + c$ má v případě modelu b) tvar

$$\hat{\theta}_{\text{cond}} = \frac{c}{b}. \quad (9)$$

Intervalový odhad pro $\hat{\theta}_{\text{cond}}$ se vypočte z hodnoty $\hat{\theta}_{\text{cond}}$ upravené přepočítávací konstantou EFF

$$(\hat{\theta}_{\text{cond}}/EFF; \hat{\theta}_{\text{cond}} \cdot EFF), \quad (10)$$

přičemž EFF má hodnotu $EFF = \exp(z\sqrt{1/b + 1/c})$.

Výraz $\sqrt{1/b + 1/c}$ je asymptotický odhad rozptylu veličiny $\ln(\hat{\theta}_{\text{cond}})$.

Hodnota z odpovídá kvantilu standardizovaného normálního rozložení a určuje hladinu spolehlivosti. Pro naše dva příklady dostaneme pro případ b) odhad θ_{cond} jako 2,083 (1,050; 4,132), resp. 0,143 (0,017; 1,149). Druhý interval pokrývá hodnotu 1!

5. Doporučené návrhy hodnocení párových binárních hodnot pomocí intervalů spolehlivosti

Pro situaci párových binárních dat jsme uvedli McNemarův test a jednoduché bodové odhady a intervalové odhady hodnot Δ , φ a θ , které vycházely z Waldova přístupu. Existuje mnoho vylepšení a variant příslušných odhadů. Jejich přehled uvádějí Fagerland, Lydersen a Laake [7]. V [7] jsou intervaly spolehlivosti posuzovány podle toho, zda jde o odhad Waldova typu nebo na základě Wilsonova způsobu odhadu [14] pomocí skórové funkce, zda pro výpočty existuje software nebo zda existuje vzorec pro interval spolehlivosti v uzavřené

formě. Samozřejmě je rozhodující dobrá kvalita intervalového odhadu (jako pravděpodobnost pokrytí neznámého parametru, šířka intervalového odhadu a jeho poloha, necitlivost k extrémním podmínkám).

Popíšeme varianty intervalových odhadů parametrů Δ , φ a θ doporučené v práci [7]. Pro obě modelové tabulky dat (3 a 4) uvádíme vypočtené příslušné bodové a intervalové odhady v tabulkách 5–7 spolu s dříve spočítanými jednoduchými odhady Waldova typu.

5.1. Intervalový odhad rozdílu pravděpodobností

Tango [12] navrhl asymptoticky platný interval spolehlivosti pro rozdíl párových relativní četností získaný iterativním výpočtem (viz dále). Vychází ze skórové funkce. F označuje celou čtveřici četností v hodnocené tabulce. Pro Δ_0 z intervalu $(-1, 1)$ definujeme statistiku

$$T(F; \Delta_0) = \frac{b - c - n\Delta_0}{\sqrt{n[2\tilde{p}_{01} + \Delta_0(1 - \Delta_0)]}},$$

kde $\tilde{p}_{01}$ je maximálně věrohodný odhad p_{01} za předpokladu, že $\Delta = \Delta_0$, tedy

$$\tilde{p}_{01} = \frac{-B + \sqrt{B^2 - 4AC}}{2A},$$

kde $A = 2n$, $B = -b - c + (2n - b + c)\Delta_0$, $C = -c\Delta_0(1 - \Delta_0)$.

Interval spolehlivosti $(L; U)$ pro Δ vypočítáme iterativně jako řešení rovnice

$$T(F; L) = -z \quad \text{a} \quad T(F; U) = z.$$

Poznamenejme, že Bonett a Price [5] navrhli úpravu Waldova přístupu tak, že při vytváření intervalu spolehlivosti podle (6) dosadí $\hat{p}_{ij} = (n_{ij} + 1)/m$, kde $m = n + 2$ (odhad podle Laplace). Místo dělení číslem n ve vzorci (6) použijí číslo m . Bonett a Price ukazují, že tento odhad se chová skoro stejně dobře jako odhad podle Tango [12]. Přitom je výpočetně jednodušší, viz také [4], [10].

5.2. Intervalový odhad poměru pravděpodobností

Využijeme postup podle Tang [11] a výpočet intervalu spolehlivosti iterativním algoritmem pomocí skórové funkce, viz také [4], [10].

F opět označuje celou čtveřici četností v hodnocené tabulce.

Nechť $\varphi = \varphi_0$. Definujeme statistiku

$$T(F; \varphi_0) = \frac{n_{1+} - n_{+1}\varphi_0}{\sqrt{n(1 + \varphi_0)\tilde{p}_{01} + (a + b + c)(\varphi_0 - 1)}},$$

kde $\tilde{p}_{01} = (-B + \sqrt{B^2 - 4AC})/2A$, přičemž

$$A = n(1 + \varphi_0), \quad B = (a + b)\varphi_0^2 - (a + b + 2c), \quad C = c(1 - \varphi_0)(a + b + c)/n.$$

Interval spolehlivosti $(L; U)$ pro φ vypočítáme iterativním vyřešením rovnic

$$T(F; L) = -z \quad \text{a} \quad T(F; U) = z.$$

5.3. Intervalový odhad podílu šancí

Jestliže odhadujeme intervalově podíl šancí θ_{cond} , pak je vhodné uvážit aplikaci Wilsonovy metody pomocí skórové funkce [7]. Vychází se z odpovídajícího intervalového odhadu pro $O = p_{01}/(p_{10} + p_{01})$. Nejdříve získáme interval spolehlivosti $(L^*; U^*)$ pro O , který transformujeme do výsledného intervalu spolehlivosti pro θ_{cond} . Použijeme přitom vztah $\theta_{\text{cond}} = O/(1 - O)$.

Provádíme následující výpočty.

Nejdříve vypočteme konstanty A , B a C . Konstanta z opět určuje hladinu spolehlivosti.

Označme $n_d = b + c$,

$$A = z\sqrt{z^2 + 4b\left(1 - \frac{b}{n_d}\right)}, \quad B = 2b + z^2, \quad C = 2(n_d + z^2).$$

Pak $L^* = (B - A)/C$, $U^* = (A + B)/C$.

Interval spolehlivosti pro náš parametr θ_{cond} dostaneme ve tvaru $(L; U)$, kde $L = L^*/(1 - L^*)$ a $U = U^*/(1 - U^*)$.

Tabulky 5–7 obsahují bodové a intervalové odhady pro oba modelové příklady dat (tabulky 3 a 4) popsané v předchozím textu.

6. Nutný počet pozorování u McNemarova testu

McNemarův test patří mezi nejpoužívanější metody v kontextu spárovaných binárních dat. K určení počtu pozorování v případě aplikace McNemarova testu při síle 0,80 použijeme tabulku 8. Síla testu $1 - \beta$ pro danou alternativu je pravděpodobnost, že test indikuje statisticky významný výsledek, kde β je podmíněná pravděpodobnost chyby druhého druhu za podmínky, že platí

Tabulka 5: Tabulka odhadů rozdílu Δ dle jednoduché a doporučené metody

Název metody \ Odhad	bodový (tab. 3)	intervalový (tab. 3)		bodový (tab. 4)	intervalový (tab. 4)	
		dolní mez	horní mez		dolní mez	horní mez
Metoda podle Walda [4]	0,054	0,005	0,103	-0,286	-0,508	-0,062
Doporučená metoda: Tango [12]	0,054	0,005	0,107	-0,286	-0,517	-0,026

Tabulka 6: Tabulka odhadů φ dle jednoduché a doporučené metody

Název metody \ Odhad	bodový (tab. 3)	intervalový (tab. 3)		bodový (tab. 4)	intervalový (tab. 4)	
		dolní mez	horní mez		dolní mez	horní mez
Jednoduchá metoda podle Walda [7]	1,141	1,011	1,287	0,25	0,062	0,992
Doporučená metoda: Tang [11]	1,141	1,011	1,300	0,25	0,065	0,907

Tabulka 7: Tabulka odhadů θ_{cond} dle doporučených metod

Název metody \ Odhad	bodový (tab. 3)	intervalový (tab. 3)		bodový (tab. 4)	intervalový (tab. 4)	
		dolní mez	horní mez		dolní mez	horní mez
Jednoduchá metoda podle Walda [7]	2,083	1,047	4,132	0,143	0,017	1,149
Transf. odhad metodou skóru [7]	2,083	1,060	4,093	0,143	0,023	0,890

Tabulka 8: Určení rozsahů výběru pro dosažení síly McNemarova testu 0,80

Síla = 0,8		p_{01}			
		0,05	0,10	0,15	0,20
p_{10}	0,10	357			
	0,15	113	610		
	0,20	60	179	860	
	0,25	39	90	243	1109
	0,30	28	57	119	305
	0,35	22	40	73	148
	0,40	18	30	51	90
	0,45	15	24	38	61
	0,50	13	20	30	45
	0,55	11	17	24	35
	0,60	10	14	20	29

daná alternativní hypotéza. Příslušné vzorce a zjištěné požadované rozsahy výběru uvádějí Vuolo et al. [13].

Tabulka 8 obsahuje rozsahy výběru nutné k dosažení síly 0,80, jestliže pravděpodobnosti p_{10} , resp. p_{01} mají určité hodnoty.

7. Příklad nesprávného hodnocení párových hodnot

Nejobvyklejší chyba při hodnocení párových dat je opominutí jejich párového charakteru (podobně jako u t -testu). Pak se použije χ^2 test nezávislosti pro kontingenční tabulky, přičemž se počítá s marginálními četnostmi tabulky 1. McNemarův test bývá často špatně pochopen. Ale setkáme se i s kurióznějšími případy. Herout ve svém článku [8] hodnotil kompetence pracovníků veřejné správy v zacházení s kancelářským softwarem, přičemž si všímal rozdílů mezi deklarovanou dovedností resp. pozorovanou dovedností u sedmi dovedností. Dotazníkem P nejdříve u 224 respondentů zjišťoval subjektivní ocenění vlastních dovedností. Pozorovanou dovednost odhadl pomocí praktického testu T dovedností. Zjistil tak četnosti pro deklarované (P) a prokázané dovednosti (T). Autor prezentoval tabulku relativních četností

Tabulka 9: Četnosti a relativní četnosti pro deklarované (D) a prokázané (T) dovednosti informatickým testem ve vzorku 224 jedinců spolu se statistikou $H\chi^2$ a z -statistikou podle Walda (umocněná na druhou) pro korelační koeficient 0,0, resp. 0,4 mezi $\hat{p}_{+1}$ a $\hat{p}_{1+}$. Tučné číslice signalizují statistickou nevýznamnost výsledků.

Dovednost	D ano	D(%)	T ano	T(%)	$H\chi^2$	Waldova t. s.	
						$r = 0,0$	$r = 0,4$
Formátování dokumentu	208	92,9	185	82,6	1,346	11,330	18,029
Automatický text	138	61,6	45	20,1	47,262	97,121	159,921
Formátování tabulky	182	81,3	153	68,3	2,510	10,266	16,941

obou typů výsledků pro sedm dovedností a porovnával je. V tabulce 9 uvádíme první a poslední řádek z tabulek pro sedm dovedností a řádek, který podle autora [8] nesvědčí o velkém rozdílu deklarovaných a prokázaných dovedností. Ani první řádek tabulky neproказuje statisticky významný rozdíl, posuzovaný v [8] vypočtenou testovací statistikou. Testovací statistiku počítal Herout dle vzorce

$$H\chi^2 = \frac{(\text{expected} - \text{observed})^2}{(\text{expected} + \text{observed})}.$$

Autor dosazuje do vzorce pro $H\chi^2$ za *observed* četnosti prokázaných dovedností a za *expected* četnosti deklarovaných dovedností v souboru 224 pracovníků. Autor přitom chybně neuvažuje přechody mezi rozdílným hodnocením praktickým testem a deklarovanou dovedností. Testovací statistika $H\chi^2$ je sice podobná McNemarově testovací statistice, avšak ve svém postupu dosazuje autor [8] za *observed* a *expected* četnosti jiné hodnoty (ne četnosti b , resp. c dle tabulky 1).

Statistika $H\chi^2$ v tabulce 9 je statistika, kterou použil Herout pro hodnocení nulové hypotézy, že není rozdíl mezi deklarovanou a prokázanou informatickou dovedností. Podle autora [8] má za platnosti nulové hypotézy rozdělení χ^2 o jednom stupni volnosti. V tabulce 9 také uvádíme hodnoty testovací statistiky podle Walda (viz výraz 3), která má asymptoticky standardizované normální rozdělení, pro případy, že korelační koeficient mezi statistikami $\hat{p}_{+1}$ a $\hat{p}_{1+}$ je 0,0, resp. 0,4 (tyto hodnoty jsme zvolili shodně s příkladem z úvodu článku). Waldovu statistiku jsme umocnili na druhou, aby srovnání bylo adekvátní.

Z údajů v Heroutově článku je patrné, že se má jednat o hodnocení spárovaných binárních údajů. Autor však chybně porovnává pouze údaje o četnostech z tabulky 9. Neuvažuje přechody mezi rozdílným hodnocením praktickým testem (prokázaná dovednost) a vlastním hodnocením v dotazníku (přiznaná dovednost).

Pokud předpokládáme korelaci mezi přiznanou a pozorovanou dovedností 0,0, resp. 0,4, pak výsledky použitím asymptotické Waldovy testovací statistiky pomocí výrazu (3) by měly pro uvažované řádky Heroutovy tabulky hodnoty v posledních dvou sloupcích tabulky 9. Je patrné, že ve všech řádcích jsou rozdílnosti takto hodnocené statisticky významné (což je v rozporu s tvrzením v [8]).

8. Závěry

Uvedli jsme základní vzorce pro hodnocení spárovaných binárních údajů. Podobně jako u testu homogenity pro dva nezávislé výběry binárních dat, i v tomto případě bylo navrženo několik alternativ, jak ukázali například autoři v [7]. Hlavním doporučeným prostředkem v analýze je McNemarův statistický test nulové hypotézy $H_0 : \Delta = 0$. Pro tento test jsme uvedli potřebný počet pozorování pro dosažení síly testu 0,80. Také jsme popsali příklad jedné chybně provedené analýzy popisované formy dat [8].

Na závěr upozorňujeme, že popsané metody se také využívají při komparaci binárních diagnostických laboratorních testů, pokud vede experiment ke spárovaným údajům a zajímá nás senzitivita a specificita testu.

Výpočty v příkladech byly provedeny programem NCSS 2007 a pomocí připravených vzorců pro Excel editor v knížce [10, soubor K10649]. Také lze použít balíčky v R, např. PropCIs¹, ExactCIdiff² nebo Alan Agresti: A.2 R and S-Plus examples³.

Poděkování

Autor děkuje J. Klaschovi za nepočítaně trpělivě vedených konzultací a anonymním recenzentům za zásadní připomínky.

¹<https://cran.r-project.org/web/packages/PropCIs/index.html>

²<https://cran.r-project.org/web/packages/ExactCIdiff/index.html>

³<http://users.stat.ufl.edu/~aa/cda/software.html>, resp.

<http://users.stat.ufl.edu/~aa/cda/R.web.pdf>

Reference

- [1] Agresti A.: *Categorical Data Analysis*, 2. a 3. ed. John Wiley & Sons: Hoboken, New Jersey, 2002, 2013 (zvláště 10.1.1 a 14, vyd. 2002) .
- [2] Agresti A., Min Y.: Simple improved confidence intervals for comparing matched proportions. *Statistics in Medicine*, 24, 729–740, 2005.
- [3] Anděl J.: *Základy matematické statistiky*. Matfyzpress, Praha, 2007.
- [4] Bilder R., Loughin, T.M.: *Analysis of Categorical Data with R*. Boca Raton. CRC Press, 2015.
- [5] Bonett D. G., Price R. M.: Confidence intervals for a ratio of binomial proportions based on paired data. *Statistics in Medicine*, 25, 3039–3047, 2006.
- [6] Bortz J., Schuster S.: *Statistik für Human- und Sozialwissenschaftler*, 7. ed., Springer Verl., Berlin, 2010.
- [7] Fagerland M. W., Lydersen S., Laake P.: Recommended tests and confidence intervals for paired binomial proportions. *Statistics in Medicine*, 33, 2850–2875, 2014.
- [8] Herout L.: Effective Use of Digital Technologies in the Public Administration. *Socioekonomické a humanitní studie*, 7(1), 40–58, 2017. URL: http://sehs.educast.cz/subdom/sehs/wp-content/uploads/2017/09/SEHS_1-2017.pdf
- [9] McNemar Q.: Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153–157, 1947.
- [10] Newcombe R. G.: *Confidence Intervals for Proportions and Related Measures of Effect Size*. CRC Press, London, 2013.
- [11] Tang N. S., Tang M. L., Chan I. S. F.: On tests of equivalence via non-unity relative risk for matched-pair design. *Statistics In Medicine*, 22, 1217–1233, 2003.
- [12] Tango T.: Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine*, 17, 891–908, 1998.
- [13] Vuolo M., Uggen Ch., Lageson S.: Statistical power in experimental audit studies: cautions and calculations for matched tests with nominal outcomes. *Sociological Methods and Research*, February, 1–44, 2015.
- [14] Wilson E. B.: Probable inference, the law of succession, and statistical inference. *J. Amer. Statist. Assoc.*, 22, 209–212, 1927.

NEPÁROVÁ NÁHRAŽKA MCNEMAROVA TESTU? POZNÁMKA K ČLÁNKU J. HENDLA AN UNPAIRED SURROGATE FOR MCNEMAR TEST? A NOTE TO PAPER BY J. HENDL

Jan Klaschka

Adresa: Ústav informatiky AV ČR, v.v.i., Pod Vodárenskou věží 271/2,
182 07 Praha 8

E-mail: klaschka@cs.cas.cz

Abstrakt: J. Hendl poukázal na chybnou statistickou analýzu párových binárních dat v publikovaném článku L. Herouta. Ukazuje se, že autor ne-správné analýzy „vynalezl“, nejspíš nevědomky, potenciálně užitečný nástroj – statistiku, která je svou povahou nepárová, ale poskytuje dolní odhad McNemarova χ^2 . Metoda v původní podobě ovšem fungovala dobře jen v některých případech a musela být lépe domyšlena, aby dávala vždy nejpřesnější dolní mez pro McNemarovu statistiku.

Klíčová slova: párová binární data, McNemarův test, náhražka testu, chyby ve statistické analýze.

Abstract: J. Hendl pointed to a wrong statistical analysis of paired binary data in a published paper by L. Herout. It is demonstrated that the author of the erroneous analysis “invented”, most likely unconsciously, a potentially useful tool – a statistic of unpaired nature yielding a lower bound of McNemar χ^2 . The method as is, however, works well only in some cases, and is further elaborated in the present text, so as to make the lower bound of the McNemar statistic universally tight.

Keywords: binary paired data, McNemar test, surrogate for a test, errors in statistical analysis.

1. Jak se něco nepovedlo

J. Hendl v článku [2] uvádí příklad chybné analýzy dat z práce [3] L. Herouta. Rád bych ukázal, že metoda, již L. Herout použil, je při vší své hrubé nekorektnosti i svým způsobem inspirativní.

Autor práce [3] v situaci, kdy by kontingenční tabulku

$$\begin{array}{cc|c}
 a & b & n_{1+} \\
 c & d & n_{0+} \\
 \hline
 n_{+1} & n_{+0} & n
 \end{array} \tag{1}$$

bylo namístě hodnotit McNemarovým testem s testovou statistikou

$$\chi^2 = \frac{(b - c)^2}{b + c},$$

použil opticky podobný vzorec, ale dosadil do něj místo b a c marginální četnosti n_{1+} a n_{+1} . Dostal tak statistiku

$$H\chi^2 = \frac{(n_{1+} - n_{+1})^2}{n_{1+} + n_{+1}} = \frac{((a + b) - (a + c))^2}{(a + b) + (a + c)} = \frac{(b - c)^2}{2a + b + c},$$

která se od McNemarovy statistiky liší jen jmenovatelem vyšším o $2a$.

L. Herout, jak vidno, „vynalezl“ statistiku, která je dolním odhadem¹ McNemarova χ^2 , k jejímuž výpočtu ale není třeba se zabývat tím, jak jsou data spárována, a určovat četnosti b a c . Překročí-li $H\chi^2$ kritickou hodnotu McNemarova testu na zvolené hladině významnosti, překročí ji v každém případě i McNemarův χ^2 . V tomto smyslu bychom tedy postup použitý v [3] mohli chápat jako jakousi *nepárovou náhražku* McNemarova testu pro situace, kdy nechceme nebo nemůžeme zjišťovat četnosti b , c diskordantních párů.

Měl-li snad (o čemž mám ovšem pochybnosti) autor [3] něco takového v úmyslu, muselo by se vzápětí dodat, že se věci mohl zhostit lépe. Vysvětlím.

2. Jak se to mohlo nepověst o něco méně

Tabulku (1) při známých marginálních četnostech, ale neznámých četnostech a , b , c , d lze zapsat pomocí parametru x jako

$$\begin{array}{cc|c} x & n_{1+} - x & n_{1+} \\ n_{+1} - x & n - (n_{1+} + n_{+1}) + x & n_{0+} \\ \hline n_{+1} & n_{+0} & n \end{array}. \quad (2)$$

Nazvěme *přípustnými* takové hodnoty x , pro které jsou všechny četnosti v tabulce nezáporné. McNemarovu statistiku

$$\chi^2 = \frac{(n_{1+} - n_{+1})^2}{n_{1+} + n_{+1} - 2x}$$

vypočtenou z tabulky (2) minimalizuje, jak je zřejmé, nejmenší přípustné x .

Je-li $n_{1+} + n_{+1} \leq n$, je evidentně nejmenší přípustnou hodnotou parametru $x = 0$ a nejmenší možný McNemarův χ^2 se rovná $H\chi^2$.

¹Nepleťme si se statistickým odhadem – prostě $H\chi^2 \leq \chi^2$.

Pokud však $n_{1+} + n_{+1} > n$, jsou přípustné až hodnoty počínaje $x = n_{1+} + n_{+1} - n$. McNemarova statistika pro takové x je pak

$$H'\chi^2 = \frac{(n_{1+} - n_{+1})^2}{2n - (n_{1+} + n_{+1})}.$$

Protože jmenovatel je roven $n_{0+} + n_{+0}$ a rozdíl $n_{1+} - n_{+1}$ v čitateli je až na opačné znaménko stejný jako $n_{0+} - n_{+0}$, můžeme výraz dále přepsat jako

$$H'\chi^2 = \frac{(n_{0+} - n_{+0})^2}{n_{0+} + n_{+0}},$$

což je „skoro totéž“ jako $H\chi^2$ až na to, že jsme marginální četnosti týkající se kategorie 1 vyměnili za četnosti kategorie 0. Vyjádření téhož vzorce pomocí četností a, b, c, d je

$$H'\chi^2 = \frac{(b - c)^2}{2d + b + c}.$$

Můžeme tedy shrnout, že nejmenší možný McNemarův χ^2 při daných marginálních četnostech je roven větší ze statistik $H\chi^2$ a $H'\chi^2$, což je $H\chi^2$, pokud $n_{1+} + n_{+1} \leq n$, a $H'\chi^2$, pokud $n_{1+} + n_{+1} > n$ (nebo i $n_{1+} + n_{+1} = n$, tehdy $H\chi^2 = H'\chi^2$).

Pokud bychom použití statistiky $H\chi^2$ v práci [3] chápali jako autorovu snahu nahradit McNemarův test co nejpřesněji něčím, co může spočítat z marginálních četností, museli bychom konstatovat, že ne vždy „sázel na správného koně“. Podíváme-li se do Tab. 9 v práci [2], kde jsou uvedeny vybrané výsledky z [3], vidíme hned, že ve dvou případech ze tří by statistika $H'\chi^2$ posloužila lépe než $H\chi^2$. Konkrétně v 1. řádku (týká se dovednosti *formátování dokumentu*) z dat

$a = ?$	$b = ?$	208
$c = ?$	$d = ?$	16
185	39	224

vyšlo $H\chi^2 = (208 - 185)^2 / (208 + 185) = 23^2 / 393 = 1,346$, ale $H'\chi^2$ vychází jako $(16 - 39)^2 / (16 + 39) = 23^2 / 55 = 9,618$. (Hodnoty 1,346, uvedené v Tab. 9, by McNemarův χ^2 dosáhl při $x = 0$, tj. $a = 0, b = 208, c = 185, d = -169$, což jaksí nejde.) Obdobně na 3. řádku Tab. 9 (dovednost *formátování tabulky*) by bylo výhodnější místo čísel 182 a 153 použít jejich doplňky do 224, tedy 42 a 71: Místo $H\chi^2 = 2,510$ bychom pak dostali $H'\chi^2 = 7,442$. Lze tedy konstatovat, že v obou uvedených případech se z marginálních četností dá vydedukovat, že ať jsou neznámé četnosti a, b, c, d jakékoli, McNemarův

test by musel vyjít statisticky významně na 5% hladině (kritická hodnota je 3,841). Hodnoty $H\chi^2 = 1,346$, resp. $H\chi^2 = 2,510$, k nimž došel Herout v [3], takovou jistotu nedávají.

3. Jak se s tím (možná) dá pracovat dál

Mám ještě několik poznámek:

- McNemarův χ^2 se dá odhadnout také shora. (Může to být k něčemu dobré? Třeba, co já vím, by si někdo rád ušetřil práci se zjišťováním četností b a c , když už z marginálních četností je zřejmé, že signifikantního výsledku by se tak nedobral.) Zde poslouží *největší* přípustná hodnota x , již evidentně je menší z četností n_{1+} a n_{+1} . McNemarův χ^2 pro takové x se rovná $|n_{1+} - n_{+1}|$. Je ovšem zapotřebí upozornit, že ačkoli kritická hodnota McNemarova testu na 5% hladině významnosti je menší než 4, je lepší „začít si dělat naděje“ až při $|n_{1+} - n_{+1}| \geq 6$. Při $|n_{1+} - n_{+1}| \leq 5$ totiž četnosti c , d mohou být v nějakém pořadí 0 a 4 nebo 5. Při tak malých četnostech nelze spoléhat na asymptotiku (např. dle [1] je aproximace rozdělení McNemarovy statistiky rozdělením χ^2 dobrá až pro $c + d \geq 8$) a exaktní varianta McNemarova testu (oboustranný exaktní binomický test hypotézy, že oba typy diskordantních párů – dvojice (0; 1) a (1; 0) – jsou stejně pravděpodobné) vychází nevýznamně.
- Nepárovou náhražku může mít i Studentův párový t -test pro data $(x_1, y_1), \dots, (x_n, y_n)$. Stačí ve vzorci

$$t = \frac{\bar{x} - \bar{y}}{s_d} \sqrt{n}$$

nahradit ve jmenovateli výběrovou směrodatnou odchylku s_d rozdíly $d_1 = x_1 - y_1, \dots, d_n = x_n - y_n$ součtem $s_x + s_y$ výběrových směrodatných odchylek dat $x_1, \dots, x_n$ a $y_1, \dots, y_n$. Nahradíme-li s_d výrazem $|s_x - s_y|$, dostaneme *horní* mez pro t . Takto zkonstruované meze představují minimální a maximální možnou hodnotu statistiky t při daných průměrech $\bar{x}$, $\bar{y}$ a směrodatných odchylkách s_x , s_y (a n), nejsou-li o datech žádné další informace. Pokud je známo něco víc, např. jsou k dispozici data $x_1, \dots, x_n$ a $y_1, \dots, y_n$, jen nevíme, jak mají být správně spárovaná, reálné meze pro t mohou být užší.

- Mám představu i o nepárové náhražce neparametrického znaménkového testu. Potřeboval bych k tomu ale více místa, než se v těchto rádobách stručných poznámkách hodí. Snad tedy o tom někdy jindy.

Poděkování

Děkuji autorům [3] a [2] za inspiraci a RVO:67985807 za podporu.

Reference

- [1] Anděl J.: *Matematická statistika*. SNTL, Praha, 1978.
- [2] Hendl J.: Hodnocení párových binárních dat: Chyby ve statistických analýzách a doporučené metody. *Informační bulletin České statistické společnosti*, 29(4), 1–18, 2018.
- [3] Herout L.: Effective Use of Digital Technologies in the Public Administration. *Socioekonomické a humanitní studie*, 7(1), 40–58, 2017. URL: http://sehs.educast.cz/subdom/sehs/wp-content/uploads/2017/09/SEHS_1-2017.pdf

ČSÚ OTEVÍRÁ VÝZKUMNÍKŮM NOVÉ SAFECENTRUM

CZECH STATISTICAL OFFICE OPENS NEW SAFECENTRUM FOR RESEARCHERS

Jan Cieslar

Adresa: Český statistický úřad, Na padesátém 3268/81, 100 82 Praha 10

E-mail: jan.cieslar@czso.cz

Český statistický úřad výrazným způsobem rozšiřuje nabídku poskytovaných služeb v oblasti mikrodát. Od 15. ledna 2019 bude pracovníkům výzkumných a vědeckých institucí umožněno samostatně pracovat v nově zřízeném SafeCentru v prostorách Ústřední statistické knihovny. Přístup bude možný za přesně vymezených podmínek s důrazem na vysokou ochranu dat.

Obdobné centrum zřídil i Eurostat a k poskytování mikrodát podobným způsobem se připojují další statistické úřady včetně mimoevropských. „Zřízením SafeCentra sledujeme současné trendy světové i evropské statistiky. Příímým impulzem a inspirací pro nás byla aktivita našich maďarských kolegů, kteří jsou v tomto směru průkopníky ve středoevropském regionu,“ vysvětluje Marek Rojíček, předseda Českého statistického úřadu.

Přístup výzkumníků do centra bude možný pouze pro vědecké účely a na základě předchozího uzavření smlouvy a podpisu slibu mlčenlivosti. „Český

statistický úřad vychází vstříc zájmu akademické obce o práci s mikrodaty. Díky SafeCentru budou moci výzkumníci s těmito podklady samostatně pracovat na svém výzkumu a dále tak posunovat poznání v mnoha oborech. Důraz je přitom kladen na bezpečnost námi spravovaných statistických dat a na zajištění anonymizace údajů,“ upozorňuje Jaroslav Sixta, místopředseda Českého statistického úřadu.

Prostor SafeCentra je důkladně monitorován a není z něj bez kontroly dovoleno cokoli odnášet v písemné nebo elektronické podobě, používat e-mail a internet a pořizovat jakékoliv záznamy. S ohledem na bezpečnost není výzkumníkům do centra umožněno vnášet například mobilní telefon.

Výsledky vědeckého bádání v SafeCentru jsou dále posouzeny odbornými útvary ČSÚ tak, aby nemohly obsahovat informace ohrožující ochranu dat a splňovaly příslušná ustanovení zákona o státní statistické službě. Výzkumník má také povinnost i po skončení své vědecké práce zachovávat mlčenlivost o důvěrných statistických údajích, se kterými se v průběhu výzkumné práce v SafeCentru seznámil. Veškeré zásady jsou podrobně popsány na webu ČSÚ, více viz https://www.czso.cz/csu/czso/data_pro_ucely_vedeckehe_vyzkumu.

„Těší mne, že SafeCentrum odpovídající moderním standardům samostatné a bezpečné práce s mikrodaty otevíráme právě v době stoletého výročí vzniku moderní statistické služby v naší zemi. Chápu ho jako další krok k modernizaci oficiální české statistiky,“ uzavírá předseda ČSÚ Marek Rojíček.

Provoz SafeCentra v současné době probíhá v pilotním režimu.

Obsah

Vědecké a odborné články

Jan Hendl

Hodnocení párových binárních dat: Chyby ve statistických analýzách a doporučené metody 1

Jan Klaschka

Nepárová náhražka McNemarova testu?
Poznámka k článku J. Hendla 19

Zprávy a informace

Jan Cieslar

ČSÚ otvírá výzkumníkům nové SafeCentrum 23

Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214.

The Information Bulletin of the Czech Statistical Society is published quarterly.
The contributions in the journal are published in English, Czech and Slovak languages.

Předseda společnosti: RNDr. Marek MALÝ, CSc., Státní zdravotní ústav, Šrobárova 48, Praha 10, 100 42, e-mail: mmaly@szu.cz.

Redakce: prof. RNDr. Gejza DOHNAL, CSc. (šéfredaktor), prof. RNDr. Jaromír ANTOCH, CSc., doc. Ing. Jozef CHAJDIK, CSc., doc. RNDr. Zdeněk KARPÍŠEK, CSc., RNDr. Marek MALÝ, CSc., doc. RNDr. Jiří MICHÁLEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. Ing. Iveta STANKOVIČOVÁ, PhD., doc. Ing. Josef TVRDÍK, CSc., Mgr. Ondřej VENCÁLEK, Ph.D.

Redaktor časopisu: Mgr. Ondřej VENCÁLEK, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, <http://www.statapol.cz/>.

DOI: 10.5300/IB, <http://dx.doi.org/10.5300/IB>
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.