KONTINGENČNÍ PARADOXY

Popularizační přednáška o kontingenčních paradoxech.

Dan Pokorný, Ulm

Varování: Článek Kontingenční paradoxy je možno číst až po důkladném prostudování přípravného materiálu, citovaného v literatuře článku Kontingenční paradoxy.

PARADOXY

Jednoho dne měl jistý Kréťan prohlásit: Všichni Kréťané jsou lháři. Výrok jest chápati tak, že Kréťané nelžou pouze příležitostně, ale že jsou zásadoví a charakterní lidé, kteří lžou z principu. Zmíněnou větou o Kréťanech vlastně nic neřekl, ale /snad/ vynalezl pro Evropu paradox. Jakub Malý, mimořádný člen královské české společnosti nauk, ve svém Stručném všeobecném slovníku věcném z roku 1881 uvádí, že "paradoxní /je/ vše, co se příčí obecné domněnce, třeba to bylo pravdivé; paradoxon - tvrzení takové". Stručný etymologický slovník z roku 1968 vysvětluje "paradox /jako/ výrok skutečně nebo zdánlivě protismyslný". Slovo je složeno z řeckých částí "para-", vedle a "doxa", mínění.

Ne vše, co se příčí všeobecně přijatým myšlenkám nebo co je protismyslné, pociťujeme jako paradox. Paradox je něco, nad čím náš "rozum zůstává stát", ba spíše něco, nad čím se náš rozum nemůže zastavit. Uvěříme-li, že všichni Kréťané jsou lháři, musíme v zápětí uvěřit, že jimi nejsou, což nás vede k tomu, že jimi přece jenom jsou atd. ...

Galileo Galilei položil otázku, zda je přirozených čísel 1, 2, 3, 4, 5, ... více nebo stejně než čtverců přirozených čísel 1, 4, 9, 16, 25, ... Na tuto otázku dostal dvě správné a navzájem si odporující odpovědi. Zvláštní stav, do kterého je naše mysl uváděna paradoxem, je snad nejlépe vystižen slovy jednoho z hrdinů Vladimíra Renčína: "Podívejte! A představte si, že úplně stejně jsem rozpolcenej! Vnitřně!"

Paradoxem není nečekané a překvapující tvrzení naším rozumem akceptovatelné. Např.: "Sněžný muž žije". Nebo: "Jako lemma byl v přednášce o exaktních testech předveden deterministicky polynomiálně složitý algoritmus obchodního cestujícího založený na jednoduchém triku. Proto, mimochodem, PN=P". /Pozn.: Algoritmus v přednášce nebyl předveden vzhledem k časové tísni způsobené diskusí ke gnostické teorii/. Paradoxem také není tvrzení, příčící se všeobecné domněnce "1+1=3 /asymptoticky/", které je pouze nesmyslem.

Matematik a filosof matematiky Petr Vopěnka popisuje paradox jako stav, kdy se do rozporu dostanou dva nebo více jevů, všechny pro nás cenné a důležité. Na rozdíl od sporu, antinomie, lze paradox principiálně řešit za cenu zneváženi jednoho z jevů vstupujících do paradoxu. Řešení paradoxu tedy není učené vysvětlení zdánlivého rozporu, ale rozhodnutí třeba i bolestné, křižovatka vědy, ze které je nutno vykročit a na kterou bude možná třeba se i vracet. Paradox Kréťana neřešíme prostým "zavedením pořádku" do formální logiky, ale rezignací na víru v neomezené vyjadřovací schopnosti lidské řeči. Paradox Galileiho lze řešit buď znevážením tisícileté intuice pojmu "množství" nebo znevážením abstrahující představy nekonečna. /Vopěnkovy úvahy tlumočeny volně a svévolně/.

V tomto článku se budu věnovat některým známým i méně známým nepříjemným vlastnostem nejběžnějších měr a statistik souvislosti v dvourozměrných kontingenčních tabulkách. Na umělých příkladech budu demonstrovat, že tyto důvěrně známé postupy mohou statistika přivést před jeho klientem do situací značně delikátních. Čtenář nechť si pak sám promýšlí, jak by se coby statistik z těchto situací vylhal jak před klientem, tak sám před sebou. Co by znevážil - zda běžné statistické postupy, celou matematickou statistiku, sebe jako námezdného klauna této vědy nebo klienta s jeho nábožnou úctou k exaktním vědám.

KONTINGENČNÍ REPETITORIUM

Budeme sledovat vlastnosti statistik v RxC kontingenční tabulce

$$X^2 = \Sigma\, \Sigma\, (a_{ij} - e_{ij})^2 / e_{ij}^{\ 2},$$

$$Y^2 = 2\, \Sigma\, \Sigma\, a_{ij} \ln (a_{ij} / e_{ij}),$$

mající asymptoticky rozdělení χ^2 s (R - 1) (C - 1) stupni volnosti.

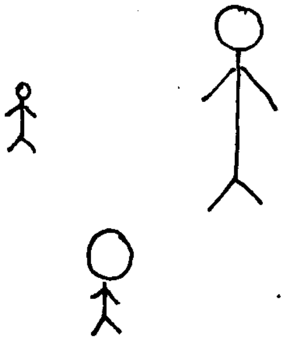
Pro 2x2 tabulky se budeme věnovat následujícím mírám a statistikám:
- Fisherův test,
- logaritmická integrační statistika, navržená Goodmanem: $W^2 = (\ln (a_{11}\, a_{22} / a_{12}\, a_{21}))^2 / (1/a_{11} + 1/a_{12} + 1/a_{21} + 1/a_{22})$, kde za $a_{ij} = 0$ dosazujeme hodnotu 1/2
- CROSS = $(a_{11}\, a_{22}) / (a_{12}\, a_{21})$
- YULE Q = $(a_{11}\, a_{22} - a_{12}\, a_{21}) / (a_{11}\, a_{22} + a_{12}\, a_{21}) =$
  $= (\text{CROSS} + 1) / (\text{CROSS} - 1)$.

PODOBENSTVÍ O KRITICKÝCH HODNOTÁCH

Obecně platí, že čím větší je počet stupňů volnosti d, tím větší je kritická hodnota K_d. Pro velká d roste příslušná hodnota K_d přibližně lineárně. Čím větší jsou dimense tabulky R, C, tím větší musí být hodnota statistiky /např.: X^2/, aby test prokázal hypotézu souvislosti v tabulce. Čím více údajů sčítáme, tím jsou přísnější naše kriteria. Lze zavést následující podobenství:

d, počet stupňů volnosti	...	výška člověka
K_d, kritická hodnota	...	největší průměr hlavy, který u člověka výšky d ještě považujeme za úměrný,
X^2, hodnota testové statistiky	...	průměr hlavy jistého člověka
testové kriterium $X^2 \geq K_d$	...	"ten člověk má velkou hlavu".

Příklady:



jedinci s nevelkou hlavou
/nevýznamné/

jedinci s velkou hlavou
/významné/

Všimněme si hodnot K_d /pro hladinu 0,05/ a poměrů K_d / d:

d	K_d	K_d/d
1	3.84	3.84
2	5.99	3.00
3	7.81	2.60
4	9.49	2.37
5	11.07	2.21

Vidíme, že analogie platí i potud, že "dítě má relativně větší hlavu".

Úvahu o hodnotách K_d jsem provedl proto, abych čtenáři vsugeroval, že "přidáme-li do tabulky něco, co přispívá souvislosti", může se výsledek testového kriteria - vzhledem ke změněným stupňům volnosti - změnit všelijak, ale hodnota statistiky by neměla klesnout. Pokud čtenář k této myšlence nezíská patřičný citový vztah, pak některéz následujících příkladů mu nepřipadnou paradoxní.

PARADOX O ZATOULANÉ KRYSE

Dějství 1

Badatel: Léta pracuji s krysami a najednou mě napadlo, že sice ne všechny samice kojí mláďata, ale přece jenom:

Samice kojí mláďata častěji, než samci.

Tak jsem to začal pozorovat a tady máte protokol:

	KOJÍ	NEKOJÍ
SAMICE	11	13
SAMEC	0	11

Myslíte, že to mohu publikovat?

Statistik: Dobrá, vaši hypotézu o souvislosti pohlaví a kojení budeme testovat logaritmickou interakční statistikou ... okamžik ...
$W^2 = (\ln ((11 \times 11) / (13 \times 0.5)))^2 / (1/0.5 + 1/11 + 1/11 + 1/13)$ ... čili ...
$W^2 = 3.79 < K_1 = 3.84$ ..., ... nevýznamné ...
... lituji, vaše hypotéza, že mláďata kojí samice se neprokázala.

Badatel: /odchází, brblaje cosi o užitečnosti vědy statistické/.

Dějství 2

Badatel: Našli jsme pod kotcem jeden ztracený záznam, tak vám ho tu pro úplnost nesu, opravte výpočty.
Nebude to asi k ničemu, představte si:

jeden samec kojil!

Statistik: No dobrá, tak tedy to máme tabulku:

	KOJÍ	NEKOJÍ
SAMICE	11	13
SAMEC	1	11

... $W^2 = 3.95 > K_1 = 3.84$

No vidíte, teď je to významné.

Samice kojí častěji než samci. Gratuluji!

Badatel /hne brvou/: Ale jak je to možné?!

Statistik /nehne/: Vy jste neslyšel o vyjímce, která potvrzuje pravidlo?
/Pro sebe mumlá odbornicky/: ... někdy jsou kontaminovaná data dokonce lepší.

Zneváženi: V tomto cvičném paradoxu je volba snadná; jedná se o vlastnost, kterou má statistika W^2 a žádná jiná z výše popsaných měr a statistik. Znevážena je proto statistika W^2 pro 2x2 tabulky.

POMOCNÝ PARADOX V PODTABULCE

Vybereme-li z tabulky podtabulku, např.

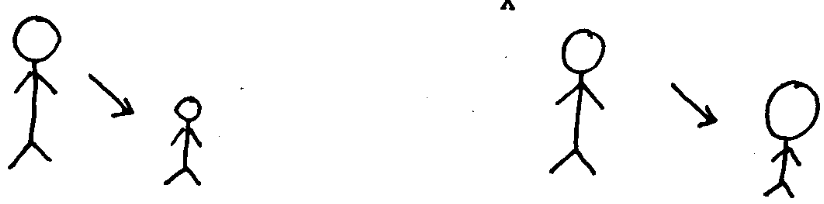
$$\boxed{\begin{array}{cc} 3 & 0 \\ 0 & 1 \\ 3 & 2 \end{array}} \quad d=2 \qquad\qquad \boxed{\begin{array}{cc} 3 & 0 \\ 0 & 1 \end{array}} \quad d=1$$

říká nám intuice, že souvislost v podtabulce může být více i méně významná, protože v podtabulce je "méně závislosti", ale i méně stupňů volnosti /zde 2, resp. 1/. "Méně závislosti" v této intuici odpovídá menší hodnotě statistiky. V našem podobenství, v podtabulce čekáme menší bytost s menší hlavou. Statistika X^2 nás zklame: její hodnoty v 3x2 tabulce a 2x podtabulce jsou

$X^2 = 3.6$ $\qquad\qquad$ $X^2 = 4.0$

Obecně "slušné" chování Y^2 a možné chování X^2 lze znázornit:

Y^2 $\qquad\qquad$ X^2



PARADOX APOLINÁŘSKÝ

Mějme zdravotnické zařízení Z s dvěmi klinikami V a A /venerická a protialkoholní/. Na tyto kliniky přicházejí pacientky s různým stupněm V nebo A obtíží, některé z nich jsou po vyléčení posílány k doléčení komplikací na sousední kliniku. Úkolem statistikovým bylo zjistit souvislosti mezi stupněm obtíží V a A. Obě veličiny byly sledována na škále: 1-těžké, 2-střední, 3-lehké, 4-žádné obtíže.

V tabulce 4x4 byly četnosti:

četnosti

	V1	V2	V3	V4	
A1	0	0	0	3	3
A2	2	2	2	0	6
A3	3	3	3	0	9
A4	5	5	5	13	28
	10	10	10	16	46

adjustovaná residua

	V1	V2	V3	V4
A1	-.9	-.9	-.9	+2.4
A2	.7	.7	.7	-1.9
A3	.9	.9	.9	-2.4
A4	-.8	-.8	-.8	+2.1

Test X^2 ukázal, že souvislost není významně prokázána:

$$X^2 = 15.3 < K_9 = 16.92$$

Sděliv to zklamaným lékařským výzkumníkům, zamyslel se statistik nad četností 13 v políčku A4-V4. V diskusi se ukázalo, že nejde o léčbu zdravých pacientek, ale omylem započítané sestřičky obou klinik.

Statistik vysvětlil, že příspěvek políčka A4-V4 k hodnotě X^2 je neoprávněný; že však ani nelze do tohoto políčka dosadit četnost 0, která by k celkové hodnotě X^2 také přispívala. Je nutno se držet modelu tzv. kvasinezávislosti a do políčka A4-V4 dosadit takovou hodnotu, aby rozdíl této hodnoty a nově vypočtené očekávané hodnoty byl nulový. V případě pouze jednoho políčka, které je strukturální nulou, t. j. neobsazené již z definice výběru, se tato vypočítá snadno:

$$a_{44}^{*} = (r_4 - a_{44}) \cdot (k_4 - a_{44}) / (m - r_4 - k_4 + a_{44}) = 3.$$

Statistika by měla být menší; počet stupňů volnosti se zmenší o 1.

Řekl, učinil:

$$X^2 = 18 > K_8 = 15.51$$

Lékařům řekl:

$$X^2 = 18 < K_8 < K_9 < X_9^2 = 15.3,$$

původní tabulka byla nevýznamná, vyloučili jsme vliv políčka A4-V4 s velkým residuálem +2.1, takže tabulka je teď významná.

Pak lékaři začali pít a statistik přestal nebo naopak, já už nevím.

PARADOX O KONTROLNÍ SKUPINĚ

Onehdy zase za statistikem přišli psychologové s výsledky pilotáže k výzkumu upřímnosti. Žádali doporučení pro plán hlavního experimentu. Pokus byl prostý: osobám blbým, navedeným a kontrolním položili otázku: "Jste blbej nebo navedenej?" Blbí odpovídali, že blbí, navedení, že navedení, kontrolní osoby přiznávaly v polovině případů to, v polovině ono.

		odpověď B	odpověď N
osoba	B	4	0
	N	0	2
	K	1	1

Hodnocení X^2 testem:

$$X^2 = 5.866 > K_2 = 5.84$$

Statistik doporučil: Je to významné na 5ti procentech, ale jen o fous.
Proto doporučuji: a/ experiment provést,
b/ zdvojnásobit kontrolní vzorek.
Vše pak klaplo, zcela přesně jako v pilotáži:

		odpověď B	odpověď N
osoba	B	4	0
	N	0	2
	K	2	2

Hodnocení X^2 testem:

$$X^2 = 5.833 < K_2 = 5.84$$

... nevýznamné!

Statistik sdělil negativní výsledek psychologům a přidal vlastní myšlenku, že od nynějška "snižováním rozsahu souboru získáváme na průkaznosti a šetříme celospolečenské náklady".

Psychologové mu zcela spontánně položili otázku, kterou předtím kladli probandům.

Znevážení: Poslední tři paradoxy se týkaly vlastnosti statistiky X^2, o statistice Y^2 je naopak známo, že tyto nežádoucí vlastnosti nemá. Historicky starší statistika X^2 /navržená začátkem století K. Pearsonem/ je výpočetně nenáročná, její vzorec je velmi "průhledný" a výsledky snadno interpretovatelné. Jak upozornil v diskusi I. Vajda, na rozdíl od Y^2 nenabývá statistika X^2 svého minima při volbě $e_{ij} = r_i k_j / m$, ale při hodnotách poněkud odlišných. Tato vlastnost stojí v pozadí demonstrovaných paradoxů. Můžeme proto ne zatratit, ale poněkud znevážit Pearsonovu X^2 statistiku. /Ostatně to už možná učinila před námi paní Pearsonová/. Z trochu obecnějšího pohledu můžeme znevážit intuitivně "dobré a rozumné předpoklady", které se v matematice často a snadno dostávají do sporu s jinými předpoklady, intuitivně stejně dobrými a rozumnými.

PARADOX MARGINÁLNÍ

Organizátoři konference ROBUST si chtěli ověřit oblibu této akce u odborné veřejnosti. Každému z 8 540 respondentů položili tři otázky:

/1/ Jste muž nebo žena?
/2/ Jste statistický expert nebo laik? /Do kategorie laik byli zahrnuti experti v něčem jiném, experti v ničem a experti ve všem/.
/3/ Preferujete konferenci ROBUST nebo COMPSTAT?

Hypotéza, kterou organizátoři rádi prokázali, hovořila o souvislosti expertství ve statistice a preferencí ROBUSTu. Vyhodnocení výsledků z obou sledovaných souborů bylo proto velkým zklamáním.

MUŽI:

	ROBUST	COMPSTAT	
expert	1000	240	1240
laik	2500	500	3000
	3500	740	4240

$X^2 = 4.4$
$Y^2 = 4.3$
$Q = -0.091$

ŽENY:

	ROBUST	COMPSTAT	
expertka	200	600	800
laička	1000	2500	3500
	1200	3100	4300

$X^2 = 4.1$
$Y^2 = 4.2$
$Q = -0.091$

Téměř shodnými hodnotami statistik vyšlo v obou souborech:

EXPERTI PREFERUJÍ COMPSTAT
EXPERTKY PREFERUJÍ COMPSTAT

Pro publikaci byla připravena jen sumarizující tabulka, bez členění na podsoubory:

LIDÉ:

	ROBUST	COMPSTAT	
expert	1200	840	2040
laik	3500	3000	6500
	4700	3840	8540

$X^2 = 15.5$
$Y^2 = 15.6$
$Q = 0.101$

Hodnoty statistik jsou vzhledem k většímu vzorku významnější, souvislost vyjádřená Yuleovým Q má zhruba stejnou intenzitu ... a opačnou orientaci, experti/tky preferují ROBUST.

Naše zkoumání lze shrnout do dvou vět:

EXPERTI I EXPERTKY PREFERUJÍ COMPSTAT
EXPERTI A EXPERTKY PREFERUJÍ ROBUST.

Zneváženi: I když víme, že parciální a margiální asociace ve vícerozměrné kontingenční tabulce se mohou podstatně lišit, přesto nás /nebo alespoň mne/ tento příklad uvádí v úžas. Můžeme detailně studovat kteroukoliv jednotlivost příkladu; to, s čím se těžko smiřujeme, je celek. Situace tím připomíná známý obrázek do sebe uzavřeného a do nekonečna vedoucího schodiště.

"Marginální paradox" není - narozdíl od předchozích - závislý na volbě konkrétní míry či statistiky; to co zde selhává, je naše zřejmě zjednodušená intuice samotného pojmu souvislosti.

BAJKA SIMULAČNÍ

Na Lutherově universitě wittenbergské žili, pracovali a o kontingenčních tabulkách přemýšleli prof. Helmuth Enke a jeho mladý aspirant, dr. Malte Bismarck.

Prof. Enke byl přítelem jednoduchých výpočetních postupů, nepotřebujících počítače ani jiná díla ďáblova. Ve snaze osvobodit statistiky od úmorných výpočtů Fisherova exaktního testu a zklamán mnoha nectnostmi Pearsonovy X^2 statistiky, upnul své naděje k maximálně věrohodnostní Y^2 statistice. Dr. Bismarckovi uložil, aby ověřil hypotézu, že výsledky Fisherova a Y^2-testu vycházejí souhlasně tak zhusta, že oba postupy jsou prakticky vzájemně nahraditelné. Dr. Bismarck bádal a s úzkostí si všímal všech situací, kdy oba testy vycházejí nesouhlasně a zdráhal se hypotézu prof. Enkeho potvrdit. Protože se aspirantura Dr. Bismarcka zvolna chýlila ke konci, rozhodli se, že otázku rozřeší jednou provždy simulačním pokusem.

Praktický výzkum, jehož výsledky právě analyzovali, jim posloužil jako vzor pro design simulačního pokusu:
/a/ Bude generováno 80 objektů s dvěma nezávislými dichotomickými

veličinami;

/b/ Vzniklá čtyřpolní tabulka bude testována na hladině 5 procent Fisherovým a Y^2-testem;

/c/ Takových čtyřpolních tabulek bude 22.

V diskusi s kolegy prof. Enke a dr. Bismarck vyjádřili zásadní ochotu měnit počet objektů i předpoklad nezávislosti veličin; na čemž však kategoricky trvali, byl celkový počet tabulek 22. Jaký má smysl, ptali se, simulacemi modelovat analýzy desetitisíců tabulek, když tak rozsáhlé problémy v praxi stejně nikdo neřeší.

U každé z 22 tabulek zaznamenali výsledky obou testů a vše shrnuli do jedné čtyřpolní tabulky:

		Y^2 VÝZNAMNĚ	Y^2 NEVÝZNAMNĚ	
FISHER	VÝZNAMNĚ	1	0	1
	NEVÝZNAMNĚ	1	20	21
		2	20	22

Prof. Enke analyzoval tabulku Y^2-testem a pomyslel si: "Y^2 = 5.36, ale souvislost mezi testy se prokázala, je tedy lepší užívat výpočetně nenáročný Y^2".

Dr. Bismarck analyzoval tabulku Fisherovým testem a pomyslel si: "Fisher = 0.0909, hle, souvislost mezi testy se neprokázala, je tedy lepší používat přesný Fisherův test".

Prof. Enke a dr. Bismarck si podali ruce a řekli nahlas: "A je to jasný"!

* PRVNÍ KONEC BAJKY*

V jiném z možných světů to dopadlo právě naopak: Y^2-test souvislost mezi testy neprokázal a Fisherův test ano. Katastrofálními důsledky, které to vzápětí vyvolalo, se musela na wittenbergské universitě zabývat řada odborných a společenských organizací:

- společnost pro matematiku a fyziku,
  Gesselschaft für Mathematik und Physik,
- podspolečnost pro matematiku,
  Untergesselschaft für Mathematik,
- nadspolečnost pro vědu a výzkum,
  Forschungswissenschaftliche Obergesselschaft,
- statistické bratrstvo,
  Statistische Brüderschaft,
- dataanalytické sesterstvo,
  Dataanalytische Schwesterschaft,
- odborový svaz pracovníků v kontingenčních tabulkách,
  Kontingenzgewerschaftsbund,
- zemědělské družstvo, které se do aféry zamíchalo omylem,
  Landwirtschaftlich Produktionasgennossenschaft Sluschwitz,

- a další.

Konečné slovo musel říci sám rektor.

$\boxed{{}^{x}{}_{x}}$ Prof. Enke analyzoval tabulku svým oblíbeným Y^2-testem a řekl si: "Hle, souvislost mezi testy se neprokázala, dr. Bismarck měl pravdu. Budu používat Fishera"!

Současně dr. Bismarck analyzoval tabulku svým oblíbeným Fisherovým testem a řekl si: "Hle, souvislost mezi testy se prokázala, prof. Enke měl pravdu. Budu používat Y^2-test".

Řekli, učinili.

Prof. Enke reanalyzoval tabulku svým nově oblíbeným Fisherem a řekl si: "Hle, souvislost mezi testy se prokázala, přece jenom jsem měl pravdu já. Budu používat Y^2-test"!

Současně dr. Bismarck reanalyzoval tabulku Y^2-testem a řekl si: "Hle, atakdále, budu používat Fishera"! /Čtenář nyní bude pokračovat ve čtení od značky $\boxed{{}^{x}{}_{x}}$, i když se proto nikdy nedozví, co nakonec řekl rektor/.

A tak prof. Enke a dr. Bismarck počítali a počítali, zatahovali do sporu další kolegy, z nichž ti nejpilnější měnili názor i dvacetkrát denně.

Enke-Bismarckův problém nakonec universitě pohltil veškeré kapacity lidské i výpočetní a hrozil, že provalí starobylé universitní zdi a ochromí celé město.

V této situaci společenské organizace moudře rozhodly přestěhovat zbytky university z Wittenbergu do Halle. Poslední slovo ovšem musel říci sám rektor: "Nařizuji, že napříště se na jednou sebraná data smí použít pouze jeden statistický test a to pouze jednou".

LITERATURA

Pokorný, O.: Kontingenční paradoxy. Popularizační přednáška o kontingenčních paradoxech. In: Sborník ROBUST 1984