JAROSLAV KAPR
ZDENĚK ŠAFÁŘ

socio logie nebo zdravý rozum?

Praxe sociologického průzkumu

MLADÁ FRONTA
PRAHA
1969

Část čtvrtá

INTERPRETACE DAT

1. Nesprávná interpretace
2. Spolehlivost údajů
3. Závěrečná zpráva
4. Vztah sociologie a praxe
5. Základní orientace v literatuře

zde je pouze vybraná část o **Statistické významnosti, kontingencních tabulkách, testování hypotéz a Chíkvadrát testu** (s. 180-193).

Celou kapitolu naleznete v souboru „**kap. 4. Interpretace dat**“ [Kapr&Safar1969_k4_Interpretace_dat.pdf]

Dejme tomu, že jsme v průzkumu dvou závodů zjistili takováto data:
Na otázku: Jak jsou v práci využity vaše schopnosti – odpovědělo:

	ZÁVOD A		ZÁVOD B	
1 – plně využity	24	5,8 %	31	10,2 %
2 – spíše využity	125	30 %	74	24 %
3 – nemohu posoudit	104	25 %	103	34 %
4 – spíše nevyužity	142	34 %	79	26 %
5 – nejsou vůbec využity	22	5,3 %	17	5,6 %
celkem	147	100 %	304	100 %

Elementárním poznatkem bude, že nelze dělat závěr z toho, že na obou závodech odpověděl téměř stejný počet osob alternativou 3, protože v obou závodech neodpovídal stejný počet osob.

Podobně problematický by byl ovšem i již uvedený komentář, který by se vyjadřoval na základě tohoto procentuálního rozložení ke kádrové práci na závodech.

Úvahy, které jsou obvykle vyvozovány z takové situace, mlčky předpokládají, že na obou závodech by mělo být *stejné* procentuální rozložení odpovědí, např. na alternativu 1. Není-li tomu tak a odpovídá-li v našem příkladě na závodě B téměř dvakrát tolik lidí, že „jsou jejich schopnosti využity“, přímo se vnucuje komentář, že na tomto závodě bude lepší kádrová práce. Položme si nyní otázku. Je taková úvaha oprávněná?

Představme si, že rozdáváme každému ze 4 hráčů kanasty po čtrnácti kartách. V balíčku je 108 karet, 54 modrých a 54 červených. Musí každý hráč dostat 7 červených a 7 modrých

karet? Mohu usuzovat, že když dostanu 9 modrých a 5 červených karet, že v balíčku bylo *více* modrých karet?

Anebo ještě jeden příklad. Představme si, že máme urnu a v ní dobře promícháno 200 kuliček. 100 modrých a 100 červených. Potom by někdo se zavázanýma očima vytahoval kuličky a každou čtvrtou by dal na stranu do misky. Po roztřídění bychom měli v misce 50 kuliček. Kdyby v ní bylo 20 červených kuliček a 30 modrých, mohli bychom podezírat toho, kdo je vybíral, že nás podváděl a díval se?

statistická významnost

Jak čtenář jistě již pochopil, jsou tyto příklady zaměřeny k tvrzení, *že takto empiricky zjištěný rozdíl, výkyv, může být náhodný* a je pak nepřípustné z něho dělat nějaké závěry.

To nás ovšem staví před další otázku. Jak poznáme, že je tento rozdíl nenáhodný nebo, jak to nejčastěji říkáme, *statisticky významný?*

Vraťme se proto ještě na chvíli k našim kuličkám. Vzpomeňme si na misku s 50 kuličkami. Kdybychom pokus s rozřaďováním opakovali 100krát, dostali bychom v misce vždy různé zastoupení červených a modrých kuliček. Např. v 75 % by bylo v misce 22 až 28 červených kuliček, v 10 % 18 až 22, v 7 % 15 až 18 atd.

Takové empirické pokusy byly několikrát důkladně opakovány a máme k dispozici spolehlivé tabulky, tj. model s náhodným rozložením výsledků různých podobných pokusů.

Co to vlastně při těchto pokusech děláme? Všimněme si, že rozložení 22 až 28 červených kuliček se ocitlo v misce velmi často, kdežto rozložení 15 až 17 již poměrně zřídka.

Při zjištění statistické významnosti vlastně hledáme procento případů, kolikrát by se určitý typ rozdělení četností objevil v nějakém konstruovaném náhodném modelu, třeba jako nyní v příkladu s kuličkami. Čím je toto procento vyšší, čím vícekrát se dané rozložení v pokusu objevuje, tím je větší pravděpodobnost, že k němu došlo náhodně. Čím je nižší, tím větší je pravděpodobnost, že k němu nedošlo *náhodně*, ale že bylo způsobeno nějakými příčinami, nám třeba ještě neznámými.

To, co nás nyní zajímá, bude *mez, hranice*, kdy výsledky ve sledovaných skupinách jsou tak odlišné, že musíme důvodně předpokládat, že nevznikly náhodně, ale že jsou způsobeny jinými vlivy. Složitost sociálních jevů je taková, že to nikdy nemůžeme říci s jistotou, ale snažme se stanovit *míru pravděpodobnosti*, se kterou můžeme tvrdit, že dosažená distribuce hodnot je nenáhodná.

Hranici takového nenáhodného rozdělení stanovujeme sami podle toho, s jak velkou pravděpodobností chceme nějakou distribuci potvrzovat. Nejčastěji se stanovuje tzv. 5% hladina významnosti. Není to nic jiného než hranice, která přibližně odpovídá takovému rozložení, které se ocitlo v našem pokusu jen 5× ze 100 opakování, nebo méněkrát. Takové rozložení se vyskytne v pokusu náhodně tak *zřídka*, že objeví-li se, můžeme s určitou pravděpodobností (95 %) tvrdit, že není náhodné.

K tomu, abychom mohli zjistit *hladinu statistické významnosti*, pro určité tvrzení, slouží celá řada statistických testů. O možnostech použití těchto testů, o tzv. jejich parametrech (tj. vlastnostech, hranicích použitelnosti, vhodnosti pro určitý typ dat) je nutná konzultace se statistikem.

Účelem takové konzultace nebude, aby se každý naučil jak matematické operace potřebné k zdůvodnění a výpočtům testu, tak i různé praktické způsoby výpočtů, ale daleko spíše by tu mělo jít o upozornění na slabiny a nevýhody toho kterého testu, o výklad, který by vedl k jasnému pochopení logického smyslu testu.

verifikace hypotéz

Předpokládejme, že jsme si nechali otestovat významnost výkyvů v odpovědích na otázku Jsou v práci využity vaše

schopnosti na závodech A a B. Máme potom k dispozici následující tabulku:

	ZÁVOD abs. čet.	A %	Hl. význ. α	ZÁVOD abs. čet.	B %
Odpověď 1	24	5,8 %	0,1 %	31	10,2 %
Odpověď 2	125	30 %	7 %	74	24 %
Odpověď 3	104	25 %	.	103	34 %
Odpověď 4	145	34 %	.	79	26 %
Odpověď 5	22	5,3 %	40 %	17	5,6 %

V prvním sloupci máme absolutní četnosti, ve druhém relativní. Prostřední sloupec, označený α (alfa) udává hladinu statistické významnosti pro rozdíl v relativních četnostech obou závodů. Zjišťujeme, že výkyvy v odpovědi na alternativu 1 *by se náhodně* mohl vyskytnout asi jednou z tisíce pokusů, kdežto výkyv v odpovědích na alternativu 2 by se náhodně mohl vyskytnout již sedmkrát ze sta pokusů a rozdíl v odpovědích na alternativu 5 dokonce asi čtyřicetkrát ze sta pokusů.

Již jsme uvedli, že se obvykle považuje za konvenčně dohodnotou pětiprocentní hladina významnosti. To znamená, že všechny rozdíly, jejichž předpokládaný „náhodný" výskyt vypočteme jako menší nebo rovný 5 %, budeme považovat za statisticky významné. Ostatní budeme považovat za nevýznamné.

Komentář uvedené tabulky by byl asi takovýto:

Na závodě A je statisticky významně méně odpovědí alternativou 1 a 3, významně více odpovědí 4. Vcelku lze tvrdit, že na závodě A je statisticky významná tendence (ve srovnání se závodem B) odpovídat, že „schopnosti nejsou využity".

Náš komentář je ovšem prohřeškem proti úspornosti vyjadřování, protože se dá vyčíst z tabulky.

Ověřování hypotéz, o kterých jsme mluvili v projektu výzkumu a při sestavování dotazníku, se děje tímto neznačeným způsobem, tj. testováním. Verifikace projektem takto stanovených hypotéz je poměrně jednoduchou interpretační úlohou, jsou-li ovšem hypotézy postaveny tak, aby sebraná data na ně umožňovala poskytnout okamžitou odpověď.

V každém poněkud obsáhlejším dotazníku (kolem 50 otázek) lze postavit tisíce takových elementárních hypotéz, které mohou být ověřovány. Proto stále více hledáme cesty, jak zpracovávaný materiál lépe utřídit, abychom si usnadnili interpretaci.

Třídění, kterým jsme se zabývali dosud, se nazývá tříděním prvního stupně. Daleko častější formou třídění materiálu, který je nám předložen k interpretaci, a který lze také srovnávat s uvedenými způsoby testování, je tzv. třídění druhého stupně.

čtyřpolní tabulky — Chi² test

Představme si, že údaje některých otázek shrneme do této tabulky:

Čtyřpolní tabulka:

	Ženy	Muži	Celkem	0_1	0_2
1. Méně	391 a	131 b	522	333,9	188,1
2. Více	2009 c	1221 d	3230	2066,1	1163,9
Celkem	2424	1372	3752 = n		

CHI = 31.475 $\alpha < 0{,}01\,\%$

Tuto formu tabulky nazýváme ***čtyřpolní*** a je výsledkem zmíněného třídění druhého stupně.

V políčku „a" máme totiž osoby z našeho souboru, které mají současně dva znaky (vlastnosti), a to „muž" a „návštěva kina méně než dvakrát měs.". Ve sloupcích „celkem" jsou potom výsledky třídění prvního stupně. Celková suma „n" označuje počet osob, které odpověděly na obě otázky.

Z této tabulky můžeme, jak čtenář jistě poznal, testovat hypotézu, zda *více* navštěvují kino ženy či muži. Mohli bychom to udělat třeba testováním podle relativních četností pro sloupce muži a ženy, ale nejčastěji to děláme tzv. Chi² testem, jehož princip je v podstatě stejný, jako u ostatních již zmíněných testů. Jeho výhodou je, že se dá snadno programovat a může sloužit jako univerzální kritérium existence *závislosti* mezi dvěma dichotomickými znaky.

V našem případě je vedle tabulky uvedena suma Chi² a hladina statistické významnosti α.

Tabulku lze potom interpretovat takto: *Mezi návštěvou kina a pohlavím osob existuje statisticky významná závislost.* O směru této závislosti jsme informováni tzv. očekávanými četnostmi,[1]) které často do tabulky vpisujeme. Tyto očekávané četnosti vlastně znamenají *model* rozložení četností za předpokladu, že by mezi sledovanými vlastnostmi *nebyl žádný vztah.* Na základě porovnání políčka „očekávané" a skutečné četnosti zjišťujeme, že je více žen s vlastností (znakem) 1 (chodí do kina méně než dvakrát týdně) a méně mužů (viz očekávaná četnost) s touto vlastností. Interpretace není složitá. Můžeme říci, že muži navštěvují významně více kino než ženy.

kontingenční tabulka

Nejčastěji se v sociologii setkáváme s typem tabulek, které mají více než čtyři políčka, tj. třídíme spolu znaky o více než dvou alternativách. Nejčastější název pro tento typ tabulek je tabulky *kontingenční* nebo tabulky se „dvojím vchodem".

Následující tabulka ukazuje vztah mezi údaji o „vzdělání mládeže" kategorizované jako 1 = základní vzdělání — 2 = středoškolské — 3 = vysokoškolské a

[1]) Někdy používáme názvu „teoretické četnosti".

údaji o příjmu: 1 — příjem do 800 Kčs
(čistý příjem) 2 — příjem do 1100 Kčs
3 — příjem do 1500 Kčs
4 — příjem do 1800 Kčs
5 — příjem nad 1800 Kčs

Vzdělání

	1	2	3	suma	0_1	0_3	0_2			
1	293	156	3	452	251	185	15	+++	——	———
2	876	609	34	1519	845	623	50	0	0	——
3	712	667	68	1447	806	565	47	———	+++	+++
4	173	112	18	303	168	125	10	0	0	++
5	67	25	2	94	52	39	3	++	——	0
Σ	2121	1569	125	3815						

Σ Chi^2 = 73,294 α < 0,01 %

Levá část tabulky obsahuje absolutní četnosti získané ze třídění II. stupně znaků „vzdělání" (o třech hodnotách) a „příjem" (o 5 hodnotách). V tříděném souboru je 3815 jedinců (průsečík sloupců pod „suma", tyto sloupce často nazýváme marginály).

V případě, že bychom získali pouze tuto strukturu tabulky, bez dalších statistických výpočtů, je zřejmé, že by tato tabulka pro čtenáře byla „mrtvá" — nedala by se přímo interpretovat. Jako minimální interpretační pomůcka by mohly být jednotlivé absolutní četnosti v políčkách tabulky převedeny na procenta — v této podobě se také s kontingenčními tabulkami často setkáváme (absolutní četnosti v tabulce můžeme převést na procenta v podstatě trojím způsobem: 100 % = sumace v řádku, 100 % = suma ve sloupcích a 100 % = celkové „n" v tabulce).

Interpretace procent uvnitr tabulky, tj. podmíněných pravdepodobností výskytu v kategoriích závislé (vysvetlované) promenné pokud patrí do urcitých kategorií nezávislé promenné **je NEZBYTNÁ. Samotný test celkové homogenity cetností polí tabulky** (napr. zde Chikv. "dobré shody") **nestací!**

Snadnější pro interpretaci takovéhoto typu tabulek je použít opět již zmíněného testovacího kritéria Chi^2, které také uvádíme v naší tabulce, a jemu odpovídající hladinu významnosti alfa. Z těchto údajů můžeme okamžitě učinit pro uváděnou tabulku jednoduchý interpretační závěr: *odmítáme* hypotézu (na 0,01 % hladině), že velikost příjmu v souboru se *neliší* podle typu vzdělání. Tj. velikost příjmu *je* nějakým způsobem *závislá* na vzdělání.

Směr této závislosti (tj. ve kterých kategoriích vzdělání existuje vyšší — nižší příjem) vyčteme z porovnání tzv. očekávaných četností (uváděno v matici vedle absolutních četností, pod 0_1, 0_2, 0_3) a četností absolutních.

V našem příkladu je tento rozdíl zřetelnější z jednoduchého znaménkového schematu, které uvádíme v tabulce: kde např. „+++" značí „absolutní četnost je značně vyšší než četnost očekávaná" a podobně „—" značí „absolutní četnost je ještě významně menší než četnost očekávaná". Z takto označeného přehledu[1]) lze vyslovit např. tyto interpretační výroky:

Vysokoškoláků je více ve vyšších příjmových kategoriích. Pracovníci se základním vzděláním jsou v nižších příjmových kategoriích, ale na druhé straně řada z nich je v nejvyšší apod.

Uveďme si ještě jeden příklad interpretace podobné tabulky:

Sledujeme odpovědi na otázku č. 1 — Zajímá vás práce, kterou vykonáváte?

1 — velmi mě zajímá
2 — více mě zajímá než nezajímá
3 — zajímá i nezajímá (tak 50 % : 50 %)
4 — spíše mě nezajímá
5 — vůbec mě nezajímá

[1]) Hodnocení rozdílu mezi absolutními a očekávanými četnostmi v kontingenční tabulce není vhodné ponechávat subjektivnímu odhadu, je opět výhodnější testovat velikost těchto rozdílů.

→ Adjustovaná residua a z nich **Znaménkové schéma**

tříděné s otázkou č. 3 — Kdybyste měl možnost jiného vhodného zaměstnání, tak
1 — jistě bych neodešel(a)
2 — asi bych neodešel(a)
3 — těžko říci, jak bych se rozhodl(a)
4 — asi bych odešel(a)
5 — jistě bych odešel(a)

Otázka č. 1

		1	2	3	4	5	Suma	O_1	O_2
Otázka č. 3	1	66	42	23	6	1	138	36	102
	2	129	206	75	2	1	413	52	148
	3	148	508	292	25	3	976	123	349
	4	90	449	601	136	20	1296	163	464
	5	54	180	467	207	140	1048	132	375
	Σ	487	1385	1458	376	165	3871		

$\Sigma\ Chi^2 = 1127{,}906 \quad \alpha < 0{,}001\ \%$

O_3	O_4	O_5	
108	28	12	+ + + 0 − − − − − − − − −
155	40	17	+ + + + + + − − − − − − − − −
368	95	42	+ + + + + − − − − − − − − −
488	126	55	− − − 0 + + + 0 − − −
394	102	45	− − − − − − + + + + + + + + +

Z tabulky vyplývá, že mezi odpověďmi na obě otázky je opět velmi silný vztah v logickém smyslu — čím více „práce zajímá“, tím méně je ochoten pracovník odejít.

V našich příkladech si čtenář všimne, že odpovědi na sledované otázky jsou uspořádány určitým systémem, a to tak, aby měly ordinální charakter, tj. alternativa jedna je příznivější než alternativa dvě apod. Výhodou takovéhoto uspořádání je to, že máme-li rozdíly mezi očekávanými a absolutními četnostmi uspořádány určitým způsobem, tj. ve směru *diagonály* kontingenční tabulky (viz znaménka v našem schématu), můžeme celou tabulku interpretovat *jediným* výrokem, jako v našem případě.

Máme-li v kontingenční tabulce kombinaci dvou znaků, z nichž jeden je ordinální a druhý nominální, potom zpravidla interpretujeme obsah tabulky řadou výroků — vážících se k počtu řádků nebo sloupců tabulky, tj. pro každou hodnotu nominálního znaku vyslovujeme jeden interpretační výrok.

Při kombinaci dvou znaků nominálních se často neobejdeme bez popisu každého *políčka* tabulky.[1])

Interpretaci kontingenčních tabulek s různými typy znaků by si měl každý čtenář osvojit — je to prozatím nejfrekventovanější „řeč“ sociologického průzkumu.

Důležitým omezením, které si čtenář musí uvědomit např. při použití Chi^2 testu je to, že špatně rozlišuje „výkyvy“ mezi *malými* četnostmi (menšími než 10 a zvláště menšími než 5), které se často vyskytují zvláště při vyšších stupních třídění. V takovýchto tabulkách je formálně správně vypočtená hladina významnosti problematická — čtenáře odkazujeme, aby interpretaci těchto tabulek konzultoval se statistikem.

složitější interpretační postupy s kontingenčními tabulkami

Obtíže analýzy narůstají tehdy, když používáme vyšších

[1]) Interpretace dat v kontingenční tabulce v tomto případě naráží na formální překážky vyplývající z „modelu“ Chi^2 testu. Zpravidla se proto interpretační závěry dělají z procentuálního rozložení četností v kontingenční tabulce (podle řádků či sloupců) — ale i zde záleží na zkušenostech interpreta, který nesmí pouštět ze zřetele již zmíněné principy srovnávání různých procentuálních údajů.

stupňů třídění nebo když se snažíme obsáhnout úsporně v textu řadu tabulek.

Např. zjišťujeme:

Studenti, kteří udávají, že více čtou beletrii, mají lepší prospěch.
Studenti, kteří udávají, že více chodí do divadla, mají lepší prospěch.
Studenti, kteří udávají, že více navštěvují výstavy a koncerty, mají lepší prospěch. ← trídení 3. stupne
Studenti, kteří udávají, že provozují aktivní kulturní činnost, mají lepší prospěch.
Studenti, kteří udávají, že chodí více do kina a více poslouchají televizní pořady, mají statisticky nevýznamně lepší prospěch.

Závěr z těchto šesti výroků, které jsou podloženy šesti kontingenčními tabulkami, lze vyjádřit jedním výrokem:

Studenti s většími kulturními zájmy mají zpravidla i lepší prospěch.

Často ovšem interpretace není takto jednoznačná. Máme např. opět sérii výroků, tentokráte ze čtyřpolních tabulek:

Kolejáci propadají významně více než nekolejáci.
Muži propadají významně méně než ženy.
Ženy bydlí více na koleji než muži.

Propadají kolejáci významně více proto, že zde působí faktor bydlení nebo proto, že je zde více žen?

I na tuto otázku lze dát jistou (i když podmíněnou) odpověď. Zjistíme si totiž, (vyššími stupni třídění), že kolejáci ženy propadají jen nevýznamně více než nekolejáci ženy. Zdá se tedy, že je vzhledem k „propadání" silnější faktor „pohlaví" než „bydlení".

Naše interpretace je ovšem velmi podmíněná. Schéma naší úvahy by totiž bylo asi následující:

Propadání:
Významně působí (podle našeho zjištění)
a, b, c, d, e
Významně působí (předpokládáme)
f, g, h . . .
Nepůsobí
x, y, z, k, l . . .

K tomu, abychom zjistili, zdali opravdu faktor „a" působí na propadání, museli bychom zajistit sledování dvou souborů jedinců:

S_1 (jedinci, mající vlastnost „a"),
S_2 (jedinci nemající vlastnost „a"),

ovšem v obou souborech by musely být ostatní faktory (vlastnosti) zastoupeny *stejně*. Jinak by se nám totiž mohlo stát, že v souboru S_1 bude více jedinců např. s vlastností „b" a my budeme experimentálně vyšší výskyt vlastnosti „propadání" připisovat faktoru „a", i když ve skutečnosti byl způsoben faktorem „b".

Čistou konstrukci této úlohy zřejmě nejsme s to zajistit, nikdy si nebudeme jisti, jestli do námi sledovaných struktur nezasahují vlastnosti, které jsme třeba nemohli sledovat nebo které jsme považovali pro daný problém za nepodstatné.

Postupem k vyšším stupňům třídění dokážeme kontrolovat alespoň zčásti potřebné (nezkreslené) složení souboru — interpretační postup je ovšem značně složitý a nepřehledný.

K tomu, abychom se orientovali ve značném množství kontingenčních tabulek, jejichž počet v průzkumu o cca 100 otázkách často přesahuje 1000, bývá výhodné použít určitých schémat — např.:

Znaky: (otázky)	a	b	c	d	e
a	■	++	+	0	0
b	++	■	++	0	++
c	+	++	■	0	+
d	0	+	0	■	0
e	0	++	+	0	■
.		+		0	
.					

Z takto uspořádaného přehledu můžeme poměrně snadno „vyčíst“ základní souvislosti mezi znaky (odpověďmi na otázky) našeho souboru.

Každé políčko této tabulky totiž představuje vlastně 1 kontingenční tabulku odpovídající kombinaci průsečíku znaků ve sloupci a příslušném řádku. Jednotlivá znaménka pak označují: „0“ = v kontingenční tabulce pro kombinaci těchto dvou znaků není nalezen mezi těmito znaky statisticky významný vztah (tabulka je statisticky nevýznamná), „+“ = tabulka odpovídající kombinaci znaků je statisticky významná na 5% hladině významnosti, „++“ = tabulka je významná na méně než 1 % hladině významnosti.

Tak např. potom, máme-li ve sloupci pod znakem „d“ samé nuly, můžeme z toho soudit, že příslušný znak v daném souboru znaků „nepůsobí“, kdežto naopak, u znaku „b“, kde vidíme kumulaci „++“, můžeme soudit, že jeho přítomnost či nepřítomnost ovlivňuje všechny další charakteristiky souboru.

Je zřejmé, že orientace v takovémto množství informací, i když reprezentovaném tabulkami i se statistickými výpočty, nepostrádá subjektivismus — na druhé straně však ta osoba, která přeměňuje tyto informace v interpretační text, musí podat text logicky konsistentní, tj. neodporující zjištěným souvislostem.

Interpretace je proto značně namáhavá a časově náročná práce, kladoucí ovšem také nároky na formulační zručnost interpreta a na jeho schopnost vystihnout „jádro“ problémů a maximálně úspornou a srozumitelnou cestou svá vysvětlení předávat.

Pro ty pracovníky, kteří ještě nezačali vlastní interpretaci materiálu ve formě kontingenčních tabulek, upozorňujeme, že je nevýhodné slovně *popisovat* každou získanou tabulku a domnívat se, že z těchto slovních popisů získáme přehlednou informaci, ze které budeme vytvářet hlavní text. Ukazuje se, že orientace v těchto popisech je tak pracná a složitá, že je výhodnější vytvářet okamžitě *vlastní* „publikační“ text bez uvedeného mezistupně — přímo z tabulek nebo přehledných informací. Za důležité vodítko pro vlastní text slouží projektem formulované *pracovní hypotézy*, které zpravidla tvoří jádro průzkumu. Odpovědi na tyto hypotézy by měly tvořit jádro závěrečné zprávy — spolu se zjištěními, které průzkumný tým považuje za obecně zajímavá a cenná.

Nedomnívejme se, že stačíme převést informace, které jsou ve formě tabulek do souvislého textu během týdne — interpretace v tomto případě při plném soustředění často trvá měsíce.

míry závislosti

Z panelového průzkumu máme před sebou 2 čtyřpolní tabulky, které nám umožňují testovat hypotézu, zdali existuje statisticky významný vztah mezi frekvencemi návštěvnosti kina a pohlavím — jednodušeji řečeno, zdali muži chodí do kina více než ženy nebo naopak.