které četnosti rovné nule, což dělá určitou potíž při některých jiných technikách.)

Koeficient souvislosti znaku u_1 se znakem u_2 se značí r (2, 1) a může nabývat hodnot od 0 do 1. Pro dva znaky, jež spolu nesouvisí, platí r (2, 1) = 0. Jestliže znak u_1 souvisí se znakem u_2 v maximálně možné míře, platí r (2, 1) = 1. O souměrnosti míry souvislosti viz dodatek 3.

2

Příklad 4

Kromě znaku u_1 (postoj k práci) a znaku u_2 (kvalifikace) sledujeme též nominální alternativní znak u_3 . . . práce . . . náročná
nenáročná.

Tabulka T 5

Relativní četnost p		Práce u_3		
		náročná	nenáročná	součet
Postoj k práci u_1	dobrý	0,25	0,25	0,50
	špatný	0,25	0,25	0,50
	součet	0,50	0,50	1,00
Kvalifikace u_2	vyšší	0,25	0,25	0,50
	nižší	0,25	0,25	0,50
	součet	0,50	0,50	1,00

Jsou-li známy prvotní četnosti pouze druhého stupně, platí kromě tabulky T 3 i tabulka T 5.

Nejsou-li k dispozici jiné údaje, mohlo by se u tabulky T 5 usuzovat, že postoj k práci nesouvisí ani s kvalifikací, ani s náročností práce. Tento závěr by byl správný, kdyby prvotní četnosti třetího stupně odpovídaly tabulce T 6.

Příklad 5

Četnosti druhého stupně podle tabulky T 5 mohou však odpovídat i četnostem třetího stupně podle tabulky T 7.

I z názoru je patrné, že v tomto případě jde o velmi intenzívní souvislost postoje k práci s kvalifikací i s náročností práce. Rozdíl proti příkladu 4 je v tom, že například pro každou fixní (nezměněnou, stálou) hodnotu znaku kvalifikace je velmi silná souvislost zbývajících dvou znaků. Lze si představit myšlenkový pokus, podobný pokusu uvedenému, při němž máme předpovědět postoj k práci náhodně vybraného jedince, a to jednou se znalostí jeho kvalifikace, jednou bez této znalosti; pokus provádíme odděleně pro náročnou i nenáročnou práci. Výsledky dílčích pokusů budou opět záviset na tom, jaké množství informace o postoji jedince k práci je obsaženo ve znalosti jeho kvalifikace, za předpokladu konstantní náročnosti práce. I této úvaze odpovídá matematicky zpracovaný pojem u teorie informací, a to tzv. podmíněná informace, od níž lze odvodit

Tabulka T 6

Relativní četnost p				Práce u_3		
				náročná	nenáročná	součet
Postoj k práci u_1	dobrý	kvalifikace u_2	vyšší	0,125	0,125	0,250
			nižší	0,125	0,125	0,250
			součet	0,250	0,250	0,500
	špatný		vyšší	0,125	0,125	0,250
			nižší	0,125	0,125	0,250
			součet	0,250	0,250	0,500
	součet		vyšší	0,250	0,250	0,500
			nižší	0,250	0,250	0,500
			součet	0,500	0,500	1,000

Tabulka T 7

Relativní četnost p				Práce u_3		
				náročná	nenáročná	součet
Postoj k Práci u_1	dobrý	kvalifikace u_2	vyšší	0,250	0,000	0,250
			nižší	0,000	0,250	0,250
			součet	0,250	0,250	0,500
	špatný		vyšší	0,000	0,250	0,250
			nižší	0,250	0,000	0,250
			součet	0,250	0,250	0,500
	součet		vyšší	0,250	0,250	0,500
			nižší	0,250	0,250	0,500
			součet	0,500	0,500	1,000

míru souvislosti uvedeným postupem. Tuto míru jsme nazvali koeficientem parciální souvislosti a označili (pro daný příklad) r (2, 1 | 3), což čteme „koeficient souvislosti znaku u_1 se znakem u_2 při fixním znaku u_3".

3

Pro sociologii je typická nutnost zpracování velkého počtu znaků. Je to nutné předně proto, že některé jevy bývají popsány větším množstvím znaků; za druhé proto, že bývá často velký počet jevů, jejichž vzájemné souvislosti je třeba zjistit. Popisovaný způsob měření souvislosti umožňuje, aby místo každého znaku, jímž jsme se dosud zabývali, se sledovala libovolně velká skupina znaků. Myšlenkový pokus, ukazující význam takového výpočtu, je zcela obdobný pokusům dosavadním. Matematický postup je založen na pojmu „informačního obsahu složeného pokusu" a nečiní potíží. Mírou souvislosti je opět koeficient souvislosti, který pro uvedený příklad 4 má tvar r (2, 1 3) = 0 a čte se „koeficient souvislosti znaku u_1 se znaky u_2 a u_3". Skutečný postup při návrhu vyhodnocení vypadá potom obecně takto:

Množinu π všech znaků, sledovaných při téže akci, rozdělíme do čtyř množin.

Množinu π_y tvoří tzv. znaky výsledné, jejichž souvislost s jinými znaky chceme sledovat (v příkladě 4 to byl znak u_1 . . . postoj k práci).

Množinu π_x tvoří tzv. znaky průvodní, u nichž chceme sledovat souvislost (v příkladě 4 to byl znak u_2 . . . kvalifikace jedince).

Množinu π_z tvoří tzv. znaky fixní, které jsme při myšlenkových pokusech předpokládali neměnné (v příkladě 4 to byl znak u_3 . . . náročnost práce).

Množinu π_0 tvoří znaky ostatní, při tomto měření nesledované. Potom lze naznačeným způsobem vypočíst r (π_x, π_y | π_z), tj. koeficient souvislosti znaků výsledných se znaky průvodními při znacích fixních. Mezi jednotlivými koeficienty, měřícími souvislosti některé skupiny znaků, platí jednoduché matematické vztahy. Počet možných koeficientů souvislosti je velmi velký; tak při třiceti znacích dosahuje řádově počtu 10^{20}.

4

Postup numerického výpočtu koeficientu souvislostí (který tato práce nepopisuje) je pro obvyklé případy jednoduchý a nevyžaduje znalostí z teorie informace. Pro případy složité a rozsáhlé je vhodné užít samočinného číslicového počítače. Tak při užití programu PROSA stačí napsat instrukci, např. „r (20, 10 | 30)" a počítač si sám naprogramuje a provede výpočet koeficientu (a dalších charakteristik) ze znaků u_{10}, u_{20} a u_{30}, počínaje tříděním a konče vytištěním protokolu. Naproti tomu rozhodnutí, který koeficient (a které další charakteristiky) se má počítat a jak je ho třeba interpretovat, je obtížným sociologickým neformalizovatelným problé-

mem. Dosavadní výsledky při vyhodnocování konkrétních sociologických výzkumných akcí ukazují, že měření souvislostí pomocí informační entropie by mohlo být účelné. Definitivní zhodnocení a vymezení vhodných oblastí aplikace si však vyžádá určitý čas.

5

Užití koeficientu souvislosti se osvědčilo při řešení několika problémů sociologického výzkumu. Předně je to postup při *verifikování pracovních hypotéz* celé výzkumné akce; k tomuto postupu byl výklad v předchozích odstavcích převážně zaměřen.

Každá pracovní hypotéza (jejíž potvrzení, zamítnutí či upravení je cílem výzkumné akce) musí být verifikovatelná některým druhem souhrnné informace a) až d), uvedené v dodatku této práce. Malou část hypotéz lze verifikovat sledováním jediného znaku (např. hypotézu „Většina jedinců ve vzorku kouří"). U ostatních hypotéz je verifikace možná pouze sledováním vztahu mezi několika znaky, tj. postupy podle b) nebo d); musí proto každé z těchto hypotéz odpovídat některý koeficient souvislosti. Jako příklad 6 uvedeme sledování souvislosti znaků

u_1 . . . volba zábavních programů
u_2 . . . pracoviště
u_3 . . . vzdělání

(podle výzkumu kulturních zájmů horníků — [13]).

Byly nalezeny koeficienty:

$$r(2, 1) = 0{,}83$$
$$r(2, 1 \mid 3) = 0{,}16$$

Z koeficientu r (2, 1) je patrná silná souvislost volby programů s pracovištěm. Závěry, že pracoviště přímo volbu programů ovlivňuje, by však byly nesprávné; z koeficientu r (2, 1 | 3) vidíme, že lidé se stejným vzděláním volí programy v průměru velmi podobně, ať pracují kdekoli. Tímto způsobem lze porovnat souvislost volby programů s různými skupinami znaků a seřadit jednotlivé souvislosti podle intenzity. U silných souvislostí následovala podrobnější analýza.

Výpočtový postup lze (jako variantu) upravit tak, že měří souvislost jedné určité třídy znaku. Tak bylo vyšetřováno, zda volba programů není podstatně odlišná zejména u jedinců fárajících pod zem, ve srovnání s jedinci patřícími do všech ostatních tříd podle pracoviště.

6

Jiným případem užití koeficientu souvislosti je tzv. „item analysis". Jako příklad 7 uveďme problém: Sleduje se jen zájem o kulturu, jehož dostatečným ukazatelem je znak u_1. Tento znak bylo možno sledovat pouze při intenzívním pilotním výzkumu, zároveň se znaky u_2 až u_{10}.

Při hlavním dotazníkovém výzkumu nelze znak u_1 sledovat vůbec, ze znaků u_2 až u_{11} lze (vzhledem k rozsahu dotazníku) sledovat nanejvýš pět znaků. Úkolem je vybrat pět nejvhodnějších znaků a odhadnout, jak přesně budou znak u_1 reprezentovat. K dispozici jsou četnosti podle znaku u_1 a dalších kombinací znaků. Z vypočtených koeficientů souvislosti se odhadne, která kombinace znaků je nejvýhodnější, jaký by mohl být koeficient souvislosti mezi znakem u_1 a navrženou kombinací a jaké další třídění by mohlo odhad zlepšit (třídění vysokých stupňů mezi velkým počtem znaků bývají obtížná). Je to nejlepší odhad, jakého lze dosáhnout na podkladě hotových třídění. Dále se buď tento návrh realizuje, nebo se provedou další navrhovaná třídění a podle nich další odhad.

7

Konečně lze navrhované míry souvislostí užít (jako kvalitativního myšlenkového modelu nebo jako výpočtového postupu) při některých transformacích a redukcích informace v poznávacím řetězci, zejména při volbě sledovaných jevů, analýze vztahu jevu a znaku, kategorizaci a škálování. Ve většině problémů je třeba výpočet míry souvislosti považovat za prvý krok analýzy, vhodný zejména pro komplexní problémy s velkým počtem znaků a jevů. Často následuje podrobná analýza vybraných souvislostí, buď neformalizovaná nebo s použitím některého ze známých (např. statistických) postupů.

DODATEK 1

Techniky užívané k měření souvislosti nominálních znaků

Pro dva nominální alternativní znaky s četnostmi danými tabulkou

Absolutní četnosti	Třídění podle prvého znaku	
Třídění podle druhého znaku	a	b
	c	d

lze vypočíst několik různých koeficientů [14].

Hodnoty Yuleových koeficientů jsou

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

$$r = \frac{1 - \sqrt{\frac{bc}{ad}}}{1 + \sqrt{\frac{bc}{ad}}}$$

a nejčastěji

$$V = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Dalších dvou koeficientů lze užít pro měření souvislosti dvou množných znaků (jeden ze znaků má e tříd a druhý f < e tříd).

Z tabulky absolutních četností se obvyklými metodami vypočte veličina χ^2 a z ní potom

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

$$C' = \frac{C}{\sqrt{\frac{f - 1}{f}}}$$

Uvedené koeficienty Q, r, V, C, C′ se užívají při výpočtech v uvedeném počtu znaků a počtu tříd. Jejich gnoseologická interpretace není zcela jasná. Kromě toho se někdy považuje za míru souvislosti i zjištěná hladina významnosti při testování hypotéz. Tento postup je nebezpečný a bez podrobného statistického rozboru může vést ke zcela chybným výsledkům (např. není-li vhodně respektován rozsah vzorku.)

Ke sledování dvou alternativních znaků se někdy užívá i postupů odvozených od korelace, o jejichž oprávněnosti jsou však pochybnosti.

DODATEK 2

Výpočty

Zadání problému

Sledujeme soubor N jedinců, a to prvého, druhého, . . . n-tého, N-tého. Počet jedinců N je velmi velký. O každém jedinci jsou dány hodnoty znaků

$$u_1 \ldots, u_2 \ldots, u_j \ldots, u_J .$$

U každého znaku jsou předem známy stavy, kterých může tento znak nabývat (jde tedy o kódované znaky). Znaky mohou být nominální, ordinální nebo kardinální (znaky kardinální a ordinální nazýváme souhrnně kvantitativní). Veškeré údaje, které jsou o souboru známy, lze shrnout do tabulky T 8.

Tabulka T 8: Hodnoty znaků

Jedinců	Znaky $u_1 \ldots\ldots u_j \ldots\ldots u_J \ldots\ldots$		
1	u_1 (1)	u_j (1)	u_J (1)
:			
n	u_1 (n)	u_j (n)	u_J (n)
:			
N	u_1 (N)	u_j (N)	u_J (N)

Prvotní tabulka

Údaje v tabulce nazveme prvotními údaji.

Hodnota znaku u_j, kterého nabyl n-tý jedinec, je označena u_j (n).

Hraje-li při sledování souboru úlohu čas, může být (analogicky s ostatními proměnnými) respektován dvojím způsobem: Buď předpokládáme, že všem prvotním údajům přísluší tentýž časový okamžik, nebo považujeme čas za jeden ze sledovaných znaků.

Hledáme způsob měření souvislostí mezi znaky $u_1 \ldots u_j \ldots u_J$, buď mezi všemi nebo mezi některými, dělenými libovolně do skupin. Hledaný způsob by měl mít jednotný gnoseologický i formální podklad pro nominální i kvantitativní znaky a měl by být vhodný pro automatizovaný výpočet. Výstupem operace „měření souvislostí" by měla být nová alespoň ordinální proměnná (slovo „alespoň" znamená, že by byla výhodnější proměnná kardinální, kterou však není lehké v použitelném stavu odvodit). Nejprve se řeší případ měření souvislostí při vyčerpávajícím šetření. Neuvažují se tedy souvislosti uvnitř rozsáhlejší populace, z níž je sledovaný soubor výběrovým vzorkem.

Měření souvislostí nominálních znaků

Uvažujeme nejprve případ, kdy všech J znaků je kvalitativních. U každého znaku u_j přiřadíme každému možnému stavu tohoto znaku libovolné reálné číslo, odlišné od čísla přiřazeného jinému stavu téhož znaku (lze např. užít řady přirozených čísel).

Zavedeme pravděpodobnostní J-rozměrný prostor, ve kterém každý znak u_j je přiřazen jedné dimenzi. V prostoru je definována pravděpodobnostní funkce

$$p(1, \ldots j, \ldots J) \qquad [1]$$

Její hodnota v bodě o souřadnicích

$$u_1 = A, \quad u_j = B, \quad u_J = C$$

je pravděpodobností, že náhodně vybraný jedinec souboru bude popsán uvažovanými stavy, tedy

$$P[(u_1 = A) \wedge \ldots \wedge (u_J = C)] \qquad [2]$$

Zřejmě platí

$$\sum_{u_1} \ldots \sum_{u_j} \ldots \sum_{u_J} p(1, \ldots j, \ldots J) = 1$$

Z pravděpodobností funkce J proměnných lze odvodit pravděpodobnostní funkci (J − a) proměnných v tvaru

$$p(a + 1, \ldots J)$$

vztahem

$$p(a + 1, \ldots J) = \sum_{u_1} \ldots \sum_{u_a} p(1, \ldots a, a + 1, \ldots J) \qquad [3]$$

založeným na marginálních četnostech vzhledem k prvním a proměnným.

Předpokládejme, že známe pravděpodobnostní funkci $p(1, \ldots b, \ldots c, \ldots d)$, kde

$$1 < b \leqq c \leqq d \leqq J$$

Proměnné $u_1, \ldots u_b$ označíme názvem „výsledné proměnné" a jejich množinu označíme π_y.

Proměnné $u_{b+1}, \ldots u_c$ označíme názvem „průvodní proměnné" a jejich množinu označíme π_x.

Proměnné $u_{c+1}, \ldots u_d$ označíme názvem „fixní proměnné" a jejich množinu označíme π_z.

Proměnné $u_{d+1}, \ldots u_J$, uvažované v rovnici [1], nebudeme sledovat.

Označení množin budeme užívat podle potřeby jako argumentů některých funkcí místo vyjmenování jednotlivých znaků. Zavedeme dále pojmy entropie a množství informace, známé z teorie informací. Nepodmíněná entropie $H(\pi_y)$ výsledných znaků

$$H(\pi_y) = -\Sigma p(1, \ldots b) \log p(1, \ldots b) \qquad [4]$$

kde se součet provádí přes všechny hodnoty pravděpodobnosti podle znaků množiny π_y (Podobně je vyjádřena nepodmíněná entropie pro jiné množiny znaků). O podmíněné entropii výsledných znaků při známých hodnotách průvodních znaků $H(\pi_y \mid \pi_x)$ platí

$$H(\pi_y \mid \pi_x) = H(\pi_y \cup \pi_x) - H(\pi_x) \qquad [5]$$

kde $\pi_y \cup \pi_x$ značí spojení obou množin. Množství informace o výsledných znacích π_y, obsažené v průvodních znacích π_x

$$I(\pi_x, \pi_y) = H(\pi_y) - H(\pi_y \mid \pi_x) = H(\pi_y) + H(\pi_x) - H(\pi_y \cup \pi_x) \qquad [6]$$

Množství informace o výsledných znacích π_y, obsažené v průvodních znacích π_x při konstantních fixních znacích π_z

$$I(\pi_z, \pi_y \mid \pi_z) = H(\pi_y \mid \pi_z) - H(\pi_y \mid \pi_x \cup \pi_z) = H(\pi_x \cup \pi_z) + H(\pi_y \cup \pi_z) - H(\pi_y \cup \pi_x \cup \pi_z) \qquad [7]$$

Za míru souvislosti znaků je možno užít buď přímo množství informace I (které lze, vzhledem k jeho vlastnostem, považovat za kardinální proměnnou), nebo jeho vhodné funkce. Jako základní požadavek na míru souvislostí jsme stanovili, že má být ordinální proměnnou. Vhodná funkce, která by transformovala množství informace na míru souvislosti, musí být zřejmě monotónní, má-li být míra souvislostí ordinální. Zavedeme koeficient $q(\pi_x, \pi_y)$ jako míru souvislosti výsledných znaků π_y s průvodními znaky π_x vztahem

$$q(\pi_x, \pi_y) = \varphi[I(\pi_x, \pi_y)] \qquad [8]$$

a koeficient $q(\pi_x, \pi_y \mid \pi_z)$ jako míru souvislosti výsledných znaků π_y s průvodními znaky π_x při konstantních fixních znacích π_z vztahem

$$q(\pi_x, \pi_y \mid \pi_z) = \varphi[I(\pi_x, \pi_y \mid \pi_z)] \qquad [9]$$

Dále budeme předpokládat, že funkce φ je rostoucí a v celé úvaze stejná. Nejvhodnější tvar funkce určíme později při analýze kvantitativních znaků.

Dále najdeme maximální hodnotu množství informace $I_{max}(\pi_x)$, které vůbec lze dosáhnout pro danou pravděpodobnostní funkci znaků $\pi_y \ldots p(1, \ldots a)$.

Jak známo, platí

$$I_{max}(\pi_y) = H(\pi_y) \qquad [10]$$

Z transformace pomocí funkce φ plyne maximální hodnota koeficientu $q(\pi_x, \pi_y \mid \pi_z)$, kterou označíme

$$Q(\pi_y) = \varphi[H(\pi_y)] \qquad [11]$$

Konečně definujeme koeficient souvislosti

$$r(\pi_x, \pi_y \mid \pi_z) = \frac{q(\pi_x, \pi_y \mid \pi_z)}{Q(\pi_y)} \qquad [12]$$

respektive

$$r(\pi_x, \pi_y) = \frac{q(\pi_x, \pi_y)}{Q(\pi_y)} \qquad [13]$$

V této souvislosti může být zajímavé si povšimnout i podobnosti s redundancí, která je vyjádřena podobně a kterou lze považovat též za míru souvislosti.

Účelnost jejího užití v sociologickém výzkumu se prozatím nepodařilo prokázat.

Měření souvislosti kvantitativních znaků

Předpokládejme nyní, že znak $u_1 = x$ je kvantitativní a spojitý. Pravděpodobnostní funkce p (1), odvozená vztahem [3] z prvotní tabulky, udává pak četnosti tříděných hodnot znaku x. Předpokládaný průběh skutečné hodnoty, spojitého znaku x, lze popsat funkcí hustoty pravděpodobnosti $f_1(x)$ s tímto obvyklým významem: Rozdělme celý obor znaku x na velmi malé úseky o délce ε a aproximujme

$$f_1(x) = f_1(n\varepsilon)$$

pro

$$n\varepsilon \leqq x < (n+1)\,\varepsilon \qquad [14]$$

Funkce $f_1(n\varepsilon)$ je pak dána vztahem

$$P\{n\varepsilon \leqq x < (n+1)\,\varepsilon\} = \varepsilon f_1(n\varepsilon) \qquad [15]$$

$$\sum_{n=-\infty}^{+\infty} \varepsilon f_1(n\varepsilon) = 1 \qquad [16]$$

Další postup plyne z nemožnosti bezprostředního položení $\varepsilon \to 0$.

Všechny součty Σ jsou od $n = -\infty$ do $+\infty$. Podle [4] je pak entropie znaku x (při dělení oboru jeho hodnot na velmi malé úseky ε)

$$H_1 = -\Sigma \varepsilon f_1(n\varepsilon) \log [\varepsilon f_1(n\varepsilon)] =$$
$$= -\log \varepsilon \Sigma \varepsilon f_1(n\varepsilon) - \varepsilon \Sigma f_1(n\varepsilon) \log f_1(n\varepsilon) =$$
$$= -\log \varepsilon - \Sigma f_1(n\varepsilon) \log f_1(n\varepsilon)$$

Uvažujme hustotu pravděpodobnosti $f_2(x)$ podmíněného rozdělení znaku x při známé hodnotě jiného znaku; jestliže platí

$$f_2(x) = \frac{1}{k} f_1(kx),$$

pak

$$P\{n\varepsilon \leqq x\,(n+1)\,\varepsilon\} = \frac{1}{k} f_1(kx) \qquad [15]$$

$$\Sigma \frac{\varepsilon}{k} f_1(kx) = 1 \qquad [16]$$

Analogicky s [17] je entropie znaku x při hustotě pravděpodobnosti $f_2(x)$

$$H_2 = -\Sigma \frac{\varepsilon}{k} f_1(kn\varepsilon) \log \left[\frac{\varepsilon}{k} f_1(kn\varepsilon)\right] =$$
$$= -\log \varepsilon \Sigma \frac{\varepsilon}{k} f_1(kn\varepsilon) + \log k \Sigma \frac{\varepsilon}{k} f_1(kn\varepsilon) -$$
$$- \varepsilon \Sigma \frac{1}{k} f_1(kn\varepsilon) \log f_1(kn\varepsilon) =$$
$$= -\log \varepsilon + \log k - \varepsilon \Sigma \frac{1}{k} f_1(kn\varepsilon) \log f_1(kn\varepsilon).$$

Potom

$$H_2 - H_1 = \log k + \varepsilon [\Sigma f_1(n\varepsilon) \log f_1(n\varepsilon) -$$
$$- \Sigma \frac{1}{k} f_1(kn\varepsilon) \log f_1(kn\varepsilon)] \qquad [20]$$

Jestliže $\varepsilon \to 0$, přejde výraz v hranaté závorce v rozdíl dvou konečných integrálů a celý třetí člen pravé strany se blíží nule, takže

$$H_2 - H_1 = \log k \qquad [21]$$

Je patrné, že rovnice [21] platí i pro obecnou lineární transformaci proměnné x

$$f_2(x) = \frac{1}{k} f_1(kx + m) \qquad [22]$$

pro jakékoli pevné konečné m.

Sledujme nyní případ, kdy jediný výsledný znak u_1 i jediný průvodní znak u_2 jsou kardinální. Na podkladě znalosti rozložení p(1) aproximujme je nejprve (jako příklad) rozložením formálním (aniž bychom něco o skutečném rozložení předpokládali). Je tedy

$$u_1 \ldots N\ [M(1),\ S(1)]$$

kde M(1) je aritmetický průměr a $S^2(1)$ rozptyl. Podobně platí

$$u_2 \ldots N\ [M(2),\ S(2)]$$

na podkladě znalosti p(2).

To je první krok úvahy (jejíž zobecnění je uvedeno dále).

V druhém kroku úvahy vyjdeme z rozložení p(1, 2) a předpokládáme lineární regresi při konstantním α, β v tvaru

$$M(1 \mid 2) = \alpha u_2 + \beta \qquad [23]$$

$$S(1 \mid 2) = \sqrt{1 - r_{12}^2}\ S(1) \qquad [24]$$

kde r_{12} je koeficient lineární korelace, M (1 | 2) je odhad průměrné hodnoty znaku u_1 podle u_2 a $S^2(1 \mid 2)$ je rozptyl hodnot znaku u_1 při předpovědi podle znaku u_2. Má-li být možno považovat korelační počet za speciální případ měření souvislostí pomocí informační entropie, musí existovat jednoznačný vztah mezi koeficientem korelace a koeficientem souvislosti.

Tvrdíme, že

$$|r_{12}| = |r_{21}| = r(2, 1)$$

jestliže funkci φ v rovnici [8] zvolíme ve tvaru

$$q(2, 1) = \sqrt{1 - 10^{-2I(2,1)}} \qquad [25]$$

Skutečně, po dosažení

$$k = \frac{S(1)}{S(1 \mid 2)} \qquad [26]$$

podle [21], [6]

$$I(2, 1) = H_2 - H_1 = \log k$$

podle [25]

$$q(2, 1) = \sqrt{1 - 10^{-2\log k}} = \sqrt{1 - \frac{1}{k^2}} =$$
$$= \sqrt{1 - \frac{S^2(1 \mid 2)}{S^2(2)}} = r_{12}$$

a protože $Q(2) \to 1$ platí i

$$q(2,1) = r(2,1) = |r_{12}| \qquad [27]$$

Důkaz nezávisel na vztahu [23] a platí tedy pro regresi na libovolnou funkci.

Dále lze výpočet provést obecněji pro

$\pi_y = \{u_1\}$
$\pi_x = \{u_2, \ldots u_c\}$
$\pi_z = \{u_{c+1}, \ldots u_d\}$

$$k = \frac{S(\pi_y \mid \pi_x \cup \pi_z}{S(\pi_y \mid \pi_z)}$$

Výsledky pro typické varianty jsou shrnuty v tabulce T 9.

jeden výsledný znak. I toto zobecnění je možné užitím známého vztahu

$$I(\pi_x, \pi_{y1} \cup \pi_{y2}) = I(\pi_x, \pi_{y1}) + I(\pi_x, \pi_{y2} \mid \pi_{y1}) \qquad [28]$$

Jiným zjednodušeným případem je měření souvislostí výsledné kardinální proměnné u_1 s průvodními nominálními proměnnými π_x při fixních nominálních proměnných π_z.

Výpočet koeficientu souvislosti podle rovnic [7], [25], [26], [27] a jeho interpretace obvyklým způsobem jsou bez obtíží možné. Porovnání výsledků s výsledky analýzy rozptylu není proveditelné, protože obě techniky

Tabulka T 9

Var	π_y	π_x	π_z	$I(\pi_x, \pi_y \mid \pi_z)$	$r(\pi_x, \pi_y \mid \pi_z)$
A	u_1	u_2		$\log S(1) - \log S(1 \mid 2)$	$\sqrt{1 - \frac{S^2(1 \mid 2)}{S^2(1)}}$
B	u_1	u_2	u_3	$\log S(1 \mid 3) - \log S(1 \mid 2, 3)$	$\sqrt{1 - \frac{S^2(1 \mid 2, 3)}{S^2(1 \mid 3)}}$
C	u_1	u_2 u_3		$\log S(1) - \log S(1 \mid 2, 3)$	$\sqrt{1 - \frac{S^2(1 \mid 2, 3)}{S^2(1)}}$
D	u_1	u_2 u_3	u_4 u_5	$\log S(1 \mid 4, 5) -$ $-\log S(1 \mid 2, 3, 4, 5)$	$\sqrt{1 - \frac{S^2(1 \mid 2, 3, 4, 5)}{S^2(1 \mid 4, 5)}}$

Souvislosti kvantitativních znaků

Ve variantě A se koeficient souvislosti $r(2, 1)$ rovná koeficientu prosté korelace mezi proměnnými u_1 a u_2.

Ve variantě B se koeficient $r(2, 1 \mid 3)$ rovná koeficientu parciální korelace mezi proměnnými u_1 a u_2 při kontrolované proměnné u_3.

Ve variantě C se koeficient $r(2, 1\ 3)$ rovná koeficientu mnohonásobné korelace proměnné u_1 na proměnných u_2 a u_3.

Variantu D lze považovat za poněkud zobecnělý případ korelace, která je zároveň parciální i mnohonásobná. Její koeficient souvislosti se značí $r(2\ 3, 1 \mid 4\ 5)$. Dále lze platnost vztahu [25] rozšířit i na jiná rozložení než normální, a to na taková, která splňují vztah [22] odpovídající lineární transformaci. Použití jiných rozdělení — kromě rozdělení pravoúhlého — při měření souvislostí v sociologii nám není t.č. známo.

Podobně nebylo dosud použito možnosti měření pro případ, že sledujeme více než

zpracovávají sice stejnou vstupní informaci, ale sledují cíle poněkud odlišné. Spíše zde jde o určitou podobnost v principu i postupech. Podobná zjednodušení jsou možná pro další kombinace nominálních kardinálních znaků. Tam, kde jde o nominální výsledný znak a o kardinální znak průvodní, je vhodné užít známého vztahu

$$I(\pi_y, \pi_x \mid \pi_z) = I(\pi_x, \pi_y \mid \pi_z) \qquad [29]$$

Obtížnější je zpracování znaků ordinálních, které se v některých výzkumných akcích užívají ve značné míře, a to zejména při větším počtu jedinců. Většinou není zcela jasné, jak a zda lze charakterizovat shodu odhadu se skutečností konečným počtem reálných čísel. Jedním z mnohých způsobů je najít pro každý ordinální znak u_j funkci $\psi(u_j)$, jíž lze znak u_j převést na jiný znak $u_k = \psi(u_j)$, který má pravoúhlou distribuci. V takovém případě je smysl znaku u_k popsán myšlenkovým pokusem: Seřadíme-li všechny jedince do řady podle hodnoty znaku u_j, pak je hod-

nota znaku u_k u každého jedince rovna jeho pořadovému číslu v řadě, dělenému počtem jedinců. Se znakem u_k lze pak zacházet jako se znakem kardinálním. Souvislost ordinálních znaků lze měřit i přímo; s interpretací tohoto způsobu však dosud nejsou praktické zkušenosti.

Další zlepšení by zde mohlo vyjít od gnoseologického vyjasnění požadavků na formální techniku, tj. navržení vhodného principu interpretace hledané míry souvislostí [15].

Navrhovaný postup byl aplikován na určitý počet problémů; jeho užití pro výběrové šetření se teprve rozpracovává. Před definitivním zhodnocením jeho účelnosti bude ještě třeba vyčkat, jak se zdaří další teoretická práce i jak výhodná bude jeho aplikace z hlediska praktické interpretace.

Poznámky:

Vhodný postup při výpočtu informačního obsahu a entropie pro jednotlivé koeficienty souvislostí při vyhodnocení konkrétní sociologické výzkumné akce, spolu s užitím základních vztahů teorie informace, může vést k značné časové úspoře. Rozsáhlejší výpočty je výhodné svěřit samočinnému číslicovému počítači.

Sledujeme-li komplexně n znaků, lze je rozdělit do čtyř množin, tj. do množiny π_x, π_y, π_z, nebo do žádné z nich. Takových kombinací existuje 4^n. Každé z nich odpovídá jeden koeficient souvislosti, kromě případů, kdy alespoň jedna z množin π_x, π_y je prázdná.

Celkem tedy existuje

$$m = 4^n - 2 \cdot 3^n + 2^n \qquad [33]$$

koeficientů souvislosti, což lze pro velký počet znaků aproximovat jako

$$m = 10^{\frac{2n}{3}} \qquad [34]$$

DODATEK 3

Souměrnost míry souvislosti

Popisovaný postup zavádí dvě míry souvislostí. Jedna z nich je souměrná,

$$q(b, a) = q(a, b)$$
$$q(b, a \mid c) = q(a, b \mid c)$$

druhá je nesouměrná, tj. nemusí vždy platit:

$$r(b, a) = r(a, b)$$
$$r(b, a \mid c) = r(a, b \mid c)$$

Praxe denního života často postuluje intuitivně požadavek souměrnosti. Přesnější rozbor při sestrojování modelů komplexních vztahů však naprosto nevede k odmítnutí nesouměrných měr. Jako ilustraci uveďme zjednodušený fiktivní příklad vztahu pracovitosti rodičů a dětí ve dvou variantách (viz variantu A, variantu B).

Varianta A

Relativní četnosti p		Pracovití rodiče ano	ne	součet
Pracovité děti	ano	0,8	0,0	0,8
	ne	0,1	0,1	0,2
	součet	0,9	0,1	1,0

Varianta B

Relativní četnosti p		Pracovití rodiče ano	ne	součet
Pracovité děti	ano	0,8	0,1	0,9
	ne	0,0	0,1	0,1
	součet	0,8	0,2	1,0

Jakákoli souměrná míra souvislostí má stejnou hodnotu pro obě varianty. Stavba modelu nebo logický rozbor však nemusí souvislost v obou variantách interpretovat jako stejně silnou.

Volba vhodného koeficientu záleží zřejmě vždy na předpokládané interpretaci.

DODATEK 4

Při vyhodnocení kvantitativní sociologické výzkumné akce se obvykle vychází z provedeného třídění podle některých sledovaných znaků. Vstupní informací operace „vyhodnocení" jsou tedy absolutní četnosti prvého nebo vyšších řádů, získané tříděním kódovaných údajů o jedincích vzorku. Cílem vyhodnocení bývá nejčastěji zjištění tří druhů souhrnných informací o populaci:

a) Vyšetření relativních četností prvního a vyšších řádů podle různých kombinací znaků. Výpočet je poměrně jednoduchý. Určité potíže působí odhad konfidenčních intervalů pro komplikované techniky výběru vzorku (dobrý přehled podává Kahuda) [1].

b) Výběr optimálního rozhodnutí (z předem dané množiny možných rozhodnutí) podle daného kritéria optimálnosti. Nejdelší vývoj má zde za sebou teorie testování statistických hypotéz, která je tč. nejpropracovanější [2], [3], [4], [5]. Rychlý rozvoj v této oblasti v poslední době souvisí s úspěchy teorie statistického rozhodování a strategických her [6].

Volba vhodné rozhodovací funkce pro určitou sociologickou akci a interpretace výsledků v sociologickém výzkumu je (kromě užití několika elementárních technik testování nulové hypotézy) často obtížná, protože gnoseologický model není v mnohých případech buď jasný, nebo pro sociologický výzkum použitelný. Při prováděných výzkumných akcích jsou chybné interpretace běžné.

c) Návrh statistických distribučních funkcí jednoho znaku jako modelu k jeho dalšímu zpracování. V sociologické praxi se užívá takřka výhradně malého počtu nejjednodušších funkcí podrobně zpracovaných matematickou statistikou; jiné typy funkcí praxe dosud nevyžaduje.

d) Zjišťování souvislosti znaků. Jedním z problémů souvislosti znaků je problém intenzity souvislosti znaků charakterizovaných otázkou, „jak silně spolu tyto dva znaky souvisí". Obvykle se hledá vhodná míra souvislostí, kterou se každé intenzitě souvislostí přiřazuje určité reálné číslo; jde tedy o kvantitativní proměnnou. Přehled běžně užívaných technik je v Dodatku 1.

DODATEK 5

Jakákoliv charakteristika statistického souboru má být [Yule, 7]: objektivně definovaná; závislá na všech pozorováních; pochopitelně vysvětlitelná; jednoduchá k výpočtu; použitelná pro výběrová šetření; použitelná pro další výpočet.

K tomu doplňujeme požadavky na míru souvislostí v sociologii: Užitelnost pro nominální a kvantitativní znaky. Možnost zpracování velkého počtu znaků. Gnoseologická jednotnost. Návaznost na osvědčené techniky (např. na korelační počet). Možnost automatizace výpočtu.

Základem navrhovaného postupu je vzájemný vztah pojmu souvislosti a informačního obsahu, specifikovaný Shannonem [8] a navržený Perezem [9] a Watanabem [10]. Tento článek je tedy aplikací principů teorie informací na sociologii. Další práce, popisující podobné postupy, jsou uvedeny v seznamu literatury [16] až [21].

Je patrné, že míru souvislostí (její striktní definici, postup výpočtu a pravidla pro interpretaci) nelze jednoznačně odvodit logickou úvahou z pojmu „souvislost", který sám není jednoznačně definován. Lze pouze požadovat, aby míra souvislostí, plnící uvedené požadavky, nebyla v rozporu s intuitivním významem pojmu "souvislost" tak, jak je obvykle užíván, a aby se v praxi osvědčila.

Literatura

[1] Kahuda F.: *Výzkumné metody v sociologii.* (Teorie a praxe sociologických výzkumů), SPN, Praha 1965.

[2] Cohen L.: *Statistical Methods for Social Scientists,* New York 1954.

[3] Mc Collough C. — Van Atta L.: *Statistical Concepts: A Program for Self-Instruction,* 1963.

[4] Dore R. P.: *Function and Cause,* American Sociological Review, vol. 26, 1963, č. 6, s. 843—853.

[5] Hagood H. J.: *Statistics for Sociologists*

[6] Blackwell D. — Girschick M. A.: *Theory of Games and Statistical Decisions,* New York 1955.

[7] Yule, C. U. — Kendall M. G.: *An Introduction to the Theory of Statistics,* Shanghai 1950.

[8] Shannon, C. E. — Weaver W.: *The Mathematical Theory of Communication,* Urbana 1949.

[9] Nikl L. — Perez A.: *Aplikace metod teorie informace při studiu závislostí v biologických systémech.* Čs. filosofie 10, 1961, č. 5, s. 454—60.

[10] Watanabe S.: *Information Theoretical Analysis of Multivariate Correlation,* IBM Journal, 1960 January, s. 66—82.

[11] Feinstein A.: *Foundations of Information Theory,* New York — Toronto — London 1958.

[12] Jaglom A. M. — Jaglom I. M.: *Pravděpodobnost a informace,* Praha, Nakladatelství ČSAV 1964.

[13] Perglerová Z.: *Kvantitativní vyhodnocení předvýzkumu akce „Zájmy",* Archiv Osvětového ústavu Praha, 23. 12. 1964.

[14] Malý V.: *Statistická analýza kvantitativních dat,* Acta Universitatis Carolinae Medica 1962, č. 1, s. 33—34.

[15] Somers R. H.: *A New Asymetric Measure of Association for Ordinal Variables,* 1962, American Sociological Review, č. 6, s. 799—812.

[16] Luce D. R. ed: *Reading in Mathematical Psychology,* Vol. 1, Wiley and Sons, 1963

[17] Gill W. J.: *Multivariate Information Transmission,* in [16], s. 89—103.

[18] Gill W. J.: *Multivariate Information Transmission,* Psychometrika, 1954, 19, 97—116.

[19] Kullback S.: *Information Theory and Statistics,* Willey and Sons, 1959.

[20] Coleman J. S.: *Introduction to Mathematical Sociology,* Glencoe 1964.

[21] Lazarsfeld P. F.: *Latent Structure Analysis,* in: Mathematics and Social Sciences, Mouton et Co., Paris 1965.

Poznámky k metodologickým problémům sociometrie

IVO NEJEZCHLEB
Pedagogická fakulta UP, Olomouc

Dosavadní teoretické práce o sociometrii[1], [11], napsané marxistickými sociology, se koncentrují většinou na kritiku sociometrie jako sociologické teorie výkladu společnosti. Výjimkou v tomto směru jsou studie uvažující o sociometrii ve dvou rovinách analýzy — jako o teorii sociálního systému a jako o metodě sociálního měření[22].

Statí zabývajících se pokusy o ospravedlnění oddělení sociometrické metody od sociometrické teorie výkladu sociální skutečnosti máme v marxistické sociologické a sociálně psychologické literatuře zatím poskrovnu[16], [28]. Autoři zde pracují s daty starších sociometrických výzkumů a zaměřují se většinou jen na problém aplikability a validity sociometrického testu v sociálním výzkumu.

Cílem tohoto článku je pokus o konfrontaci dosavadních nemarxistických teoretických prací o sociometrii jako metodě s některými pracemi marxisticky orientovaných sociologů a sociálních psychologů, s přihlédnutím k novější sociometrické metodologii.

Předmět sociometrie

Vyjdeme-li z Morenovy definice sociometrie jako metody zkoumající lidské afektivní vztahy k ostatním bytostem [19][2], mohli bychom za vlastní předmět sociometrických šetření považovat afektivní stránku sociálního vztahu. Tato redukce předmětu sociometrických šetření na afektivní zřetel vztahu mezi individui plně odpovídá celkové Morenově orientaci na psychologii (L. Wiese jej v tomto směru klade do protikladu k Durkheimovi), která se pokouší vykládat sociální dynamiku „doktrinou spontaneity a tvořivé síly".

Moreno rozumí pod pojmem interpersonálního vztahu vzájemné souvislosti mezi tzv. teleprvky[1]. Studium těchto vztahů pak má umožnit poznání struktury tzv. sociálních atomů[2].

Sociální vztahy se z tohoto hlediska konstituují z prvků citových vztahů lidí, kteří jsou ve vzájemném kontaktu. Zúčastněné osoby tedy vzájemně (explicite či implicite) vyjadřují určité preference nebo odmítání. Podle L. Wieseho[29] a s ohledem na kategorie jeho „Beziehungslehre" by bylo možno hovořit o procesech sbližování (Zueinander) a oddělování (Auseinander). Základem těchto preferencí či odmítání jsou podle Morena citové faktory evokované čistě spontánně, vycházející z podstaty individua, z jeho specifického sociálního pudu. Tato „teorie" sociálních vztahů (i když se v poslední době objevuje u Morena a některých jeho následovníků snaha o neomezování sociálního atomu jen na souvislosti emocionální) jednostranně akcentuje jistě významný afektivní zřetel na úkor neméně důležitých aspektů sociálních.

Při koncipování pojmu sociálního vztahu, který by respektoval jak sociální, tak psychologické faktory, se jeví v poslední době jako nejpřiměřenější použití kategorie interakce.

G. Homans[12] definuje interakci jako společenskou účast dvou nebo více osob v jednom sociálním procesu. Tento autor rozlišil základní roviny sociálního vztahu:

1. aktivita (jednání v nejširším slova smyslu);
2. interakce (specifické formy sociálního chování mezi jedinci);
3. city;
4. normy (chování, které skupina od svých členů v určitých konkrétních situacích očekává);

[1] „Teleprvek" (teleelement) je podle Morena nejjednodušší jednotkou citů (sociální elektron), která je předávána jedním individuem druhému. Je to elementární vztah, který může existovat mezi jedinci (nebo jedinci a předměty) a v člověku se od jeho narození proměňuje v pocit vztahu mezi lidmi.

[2] „Sociální atom" je základním prvkem sociometrické struktury, který již není možno rozdělit na samostatné celky. Podle Morena je vybudován na psychologických vztazích spojujících individuum s jinými individui. Sociální atom jako komplex projektovaných a rejektovaných sociálních elektronů určitého individua.

5. hodnoty (určité představy, které je možno pro specifický sociální systém považovat za víceméně obecné a které určují chování jedinců).

Každý typ sociálních vztahů[3] tedy nutně musí vedle psychologického zřetele (bývá v definicích vyjádřen obyčejně v pojmech postoje) akcentovat zřetel normativní (hodnoty, normy, role). Jako příklad můžeme uvést Zaborowského[32] pojetí sociálních vztahů[4] jako určitými rolemi a sociálními vzory normované postoje osob k sobě navzájem, vytvořené během vědomého, bezprostřednějšího a trvalejšího vzájemného působení.

Sociální vztahy se tedy vytvářejí v činnosti, jsou spoluutvářeny obsahem a účelem interakce a normativními standardy členských (a vztažných) mikroskupin a makroskupin. Teoreticky se dá předpokládat, že co určitá činnost, to specifická struktura sociálních vztahů ve skupině (nebo aspekt globální struktury).

Znalost hierarchie skupinových hodnot a norem umožní stanovit vedle těchto dílčích struktur sociálních vztahů ve skupině (podle jednotlivých činností a v závislosti na zvoleném kritériu hodnocení) také tzv. globální strukturu, která udává komplexní obraz sociálních vztahů ve skupině. Ze sociologického hlediska jde o zjištění, která sociální role (a jí odpovídající činnost) má pro skupinu větší význam než ostatní.

Psychologický pohled na strukturu sociálních vztahů ve skupině akcentuje – na rozdíl od sociologického přístupu, při němž jde o šetření hierarchie pozic a rolí – subjektivní vztah individua k zaujímané pozici, k jiným osobám a k sobě. Žák může zaujímat relativně nízkou pozici (např. v hierarchii rangů) a přesto s ní může být spokojen. Tato „míra spokojenosti" závisí na povaze psychosociálních potřeb žáka (na jejich subjektivním významu) a na stupni jejich uspokojování skupinou. Důležitou roli v této souvislosti hraje aspirační úroveň žáka a stupeň sebeakceptace. Tento psychologický aspekt sociálních vztahů, ve kterem (podle M. Pilkiewicze[25]) vedle elementů oceňování a hodnocení[5] hrají roli elementy emocionální adaptace (příchylnost, láska, nenávist) a motivační prvky (touha po navázání kontaktů, jejich udržení, přerušení), je sociometriky nadměrně a jednostranně zdůrazňován a považován za výchozí pro koncipování sociálního vztahu vůbec; to jim v poslední instanci umožní redukování vztahu na dimenzi sympatie-antipatie.

Nadměrné akcentování psychologických obsahů sociálního vztahu vede většinu sociometriků a jejich následovníků k pojímání předmětu sociometrických šetření ve smyslu „osobních vztahů", „vnitřní struktury" skupiny, „psychologické struktury", (psychogrupa) „infrastruktury", „neformální struktury" apod. Ani někteří marxisticky orientovaní sociologové a sociální psychologové se nevyhnuli těmto tradovaným formulacím (Zaborowski, Kolominskij, Rozov, Maslow ...) předmětu sociometrických šetření. H. H. Jenningsová[15] se naopak snažila dokázat, že sociometrickými technikami se měří v závislosti na použitém kritériu jak struktury „osobních" vztahů („psychogrupa"), tak vztahy vyplývající ze společenské činnosti („sociogrupa").

M. Vorwegovi[28] se podařilo prokázat, že ani tzv. „skupinové vztahy" (kolektivní, vyplývající ze skupinové činnosti), ani „osobní vztahy"[6] nemohou existovat nezávisle na sobě. Kdyby tomu tak bylo, mohla by sociometrie (podle Vorwega)

a) odkrýt určitou skupinovou strukturu nezávislou na konkrétních vztazích;
b) výsledky všech použitých otázek by se daly v poslední instanci redukovat na dimenzi sympatie-antipatie (při faktorové analýze by měl být účinný jen faktor sympatie).

Během několikaletých sociometrických experimentů v dětských skupinách, s použitím statisticky signifikantních technik, prokázal Vorweg neudržitelnost uvedených předpokladů.

[3] Sociální vztahy možno rozlišit na: a) interpersonální, b) meziskupinové, c) vztahy jedince ke skupině.

[4] Pojem sociálního vztahu nutno rozlišit od pojmu sociálního styku (volná, krátkodobá a nepravidelná interakce) a kontaktu (trvalejší a častější interakce).

[5] Podle Engelmayera [7] jsou sociometrické volby akty hodnocení. Jde při nich o soudy o sobě samém, o vlastním místě, vlastním poměru ke skupině, ale také jde o soudy o ostatních.

[6] „Osobní vztahy" v tom smyslu, jak je chápou někteří sociometrikové, tj. identické se sociometrickým pletivem a nezávislé na aktuálních strukturách vztahů ve skupině. Toto pojetí „osobních" vztahů jako sociometrického pletiva nemá nic společného s jednou z možných rovin klasifikace sociálních vztahů ve skupině na převážně osobní a převážně věcné, jejichž základem je sociální interakce vyplývající z provádění a organizace činností, často skupinově zaměřených. Sociometrická šetření odkrývají specifické aspekty struktur obou těchto typů vztahů na základě zvoleného kritéria sociometrické volby.