Vícenásobná a parciální asociace v kontingenčních tabulkách

JAN ŘEHÁK
BLANKA ŘEHÁKOVÁ
Geografický ústav ČSAV, Praha

Při zpracování mnohorozměrných kontingenčních tabulek řešíme úlohy, které vznikají rozšířením známých situací z tabulek druhého stupně třídění:

a) *testování hypotéz nezávislosti* mezi proměnnými a skupinami proměnných a za různých podmínek (rozšíření testu nezávislosti v dvourozměrné tabulce);
b) *testování a odhad jednotlivých parametrů* určujících specifické projevy závislosti (rozšíření postupu znaménkového schématu a odhadu reziduálních parametrů);
c) *měření stupně závislosti* mezi jednotlivými proměnnými a skupinami proměnných a za různých podmínek (rozšíření postupů měření síly asociace v dvourozměrné tabulce).

Testování hypotéz a využití reziduí se provádí pomocí logaritmicko-lineárních modelů (viz např. G. J. G. Upton [1978], Y. M. M. Bishop, S. E. Fienberg, P. W. Holland[1975], T. Havránek [1980], P. Matějů [1985], nebo na základě Pearsonových testů dobré shody chí-kvadrát (viz např. J. Řehák, B. Řeháková [1985], popřípadě dalšími postupy (např. S. J. Haberman [1982]).

Měřením asociace ve vícerozměrných kontingenčních tabulkách se zabývali L. A. Goodman, W. H. Kruskal [1963], (parciální Guttmanovo lambda), L. N. Gray, J. S. Williams [1981] (parciální Wallisovo tau), J. Řehák, B. Řeháková [1984], B. Řeháková [1985]. Kromě toho lze použít pro modelování lineárních, resp. monotónních asociačních vztahů celou teorii parciálních korelací Pearsonových nebo Kendallových, popřípadě (při přijetí jistých dohod) i Spearmanových (viz M. G. Kendall [1962], J. Řehák, B. Řeháková [1985]), obdobně jako u spojitých proměnných lze využít parciální korelační poměr η^2 (viz C. R. Rao [1978]).

V této práci uvádíme *model vícenásobné a parciální asociace* pro situaci, v níž sledujeme vlivy několika nominálních proměnných (faktorů) na jednu závislou proměnnou, která může být obecného typu. Obecný typ proměnné byl definován v rámci D-modelu, který vyvinuli J. Řehák a B. Řeháková (viz J. Řehák [1976], J. Řehák, B. Řeháková [1979], [1982], [1984], [1985], B. Řeháková [1982], [1985]).

Typ proměnné **A** s kategoriemi $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_S$ je určen maticí $\mathbf{D} = \|d_{st}\|$ o rozměrech $S \times S$, jejíž každý prvek d_{st} vyjadřuje stupeň nepodobnosti, obsahové odlišnosti nebo vzdálenosti mezi s-tou kategorií $\boldsymbol{A}_S$ a t-tou kategorií $\boldsymbol{A}_t$. Specifikací matice **D**[1] dostaneme nejen běžně známé případy nominálního, ordinálního a kardinálního znaku, ale též například *znak s geografickými relacemi* mezi kategoriemi (vhodný pro zkoumání výskytů jevů plošně rozložených), *typologie*, které jsou mnohorozměrně numericky ohodnoceny (např. výsledky seskupovací analýzy), *kombinované dichotomie*,[2] *zobecněný ordinální znak*, *znak k-násobné volby* ap.

V práci shrneme nejprve pojem statistického měření asymetrické asociace $\mathbf{B} \rightarrow \mathbf{A}$, kde **B** je faktor a **A** je závislá proměnná a rozšíříme jej na *pojem vícenásobné asociace*,

[1] Požadavky na skóry d_{st} jsou uvedeny v dodatku.

[2] Příklad kombinované dichotomie je uveden dále v textu.

pomocí něhož měříme stupeň ovlivnění variability závislé proměnné **A** celou skupinou faktorů $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_M$. Ve druhé části zavedeme *koeficienty parciální asociace*, které ukazují na čistý příspěvek jedné proměnné v rámci působení nějaké skupiny, ve třetí části uvádíme praktické příklady.[3] Dodatek obsahuje vzorce pro výpočet *přibližné výběrové chyby koeficientů*, které odvodila B. Řeháková [1982], [1985].

1. Asymetrická statistická asociace: dvourozměrná tabulka a zobecnění na koeficienty mnohonásobné asociace

O koeficientech asymetrické asociace bylo v Sociologickém časopise již pojednáno několikrát (J. Řehák [1976], J. Řehák, B. Řeháková [1973], [1975], [1979]), a proto zde provedeme pouze stručné shrnutí.

Model statistické závislosti v této části zahrnuje dvě proměnné: nominální faktor **B**, který ovlivňuje závislou proměnnou **A** obecného typu. Vztah označujeme $\mathbf{B} \rightarrow \mathbf{A}$. V rámci systému (**A**, **B**) je tedy **B** *považována* za příčinu a **A** za důsledek.[4] Označme si $\mathbf{B} \times \mathbf{A}$ kontingenční tabulku, jejíž řádky jsou určeny hodnotami proměnné $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_R\}$ a sloupce hodnotami proměnné[5] $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_S; \mathbf{D}\}$. Typ tabulky je tak $R \times S$, R je počet řádků a S je počet sloupců.

Asociační model vychází z jednoduché představy: jestliže proměnná **B** ovlivňuje proměnnou **A**, projeví se to růzností rozdělení četností jednotlivých řádků tabulky.[6] Čím více se liší, tím silnější je statistická závislost **A** na **B**. Jestliže jde o různost v zastoupení jednotlivých kategorií, mluvíme o asociaci nominální, lze však specifikovat (a to i za pomoci matice **D**) jiné vlastnosti rozdělení, jejichž rozdílnost v řádcích zkoumáme a tak dostáváme i jiné typy statistické asociace.[7]

K měření stupně asociace se využívá koeficientů, které jsou konstruované na nejrůznějších principech. Nejlogičtějším se zdá být *princip redukce predikční chyby* (L. A. Goodman, W. H. Kruskal [1954] a s ním souběžný princip redukce variability (J. Řehák [1976], J. Řehák, B. Řeháková [1979]).[8] Koeficient takto vytvořený má význam *relativního podílu vysvětlené variability proměnné* **A** *proměnnou* **B**.

---

[3] Příklady jsou pouze ilustrační, vyřazeny z kontextu, nemohou být zobecňovány.

[4] To ovšem neznamená, že **B** je skutečnou příčinou, neboť $\mathbf{B} \rightarrow \mathbf{A}$ je zjednodušený sociologický model, který vyjadřuje směr působení. Přidání dalších proměnných může ukázat *zprostředkování vztahu* nebo *nepravé asociace*. I směr je určen z *vnějších* (teoretických, analytických, hypotetických) úvah a ve většině případů neplatí jednoznačně (většinou jde o převažující směr vlivu v situaci, kdy působí zpětné vazby). Asymetrická statistická asociace je tedy modelem, který pro danou analýzu přijímáme, jehož platnost se ověřuje především konzistencí všech analytických závěrů a teoretických úvah a jenž může být těmito úvahami též vyvrácen. O problematice vztahu statistické závislosti a kauzality viz B. Řeháková [1982].

[5] Symbolem **D** u seznamu kategorií proměnné **A** vyjadřujeme fakt, že typ proměnné **A** je určen pomocí skórů $\mathbf{D} = \|d_{st}\|$. Matice skórů vytvářejících typ nemá přímý vliv na vytváření tabulky, hodnoty d_{st} se využívají až ve výpočtech. Je však výhodné uspořádat sloupce podle určitých relací, které **D** vyjadřuje: například podle pořadí kategorií u ordinálního a kardinálního znaku.

[6] Připomeňme, že případ statistické nezávislosti v kontingenční tabulce nastává tehdy, když všechna řádková rozdělení jsou totožná, tj. když hodnoty proměnné **B** nemají vliv na rozdělení **A**.

[7] Např. u číselných dat (**A** je kardinální) je místo úplné nezávislosti zkoumána statistická nezávislost vzhledem k vlivu **B** na průměr **A** v řádcích. *Kardinální nezávislost* se tedy projeví shodou řádkových průměrů v tabulce, zatímco závislost je chápána jako rozdíl průměrů — čím větší rozdílnost průměrů, tím vyšší stupeň kardinální asociace.

[8] Jiným známým principem je *normování chi-kvadrátové vzdálenosti* tabulky od hypotetické tabulky konstruované za předpokladu nezávislosti a princip *Danielsova koeficientu korelace*. *Princip redukce variability* je znám z teorie regrese a lineární korelace (koeficient determinace η^2).

Konstrukce koeficientu na *principu redukce variability*:

1. určíme míru variability pro klasifikaci $\mathbf{A} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_S; \mathbf{D}\}$ podle jejího typu (matice **D**), např. nomvar, dorvar, rozptyl, zobecněný rozptyl,[9]
2. pro celkovou variabilitu souboru a variability v jednotlivých řádcích (podsouborech) platí, že[10]

(1) *celková variabilita souboru = průměr variabilit v řádcích + průměr čtverců vzdáleností mezi řádky*,[11]

3. *koeficient explanace*[12] zavedeme jako

$$\delta_{A|B} = \frac{\text{průměr čtverců vzdáleností mezi řádky}}{\text{celková variabilita souboru}} = \tag{2}$$

$$= 1 - \frac{\text{průměr variabilit v řádcích}}{\text{celková variabilita souboru}} . \tag{3}$$

Pro explicitní vyjádření této formule si zavedeme následující značení:

$$\boldsymbol{f}_{(r)} = (f_{1|r}, f_{2|r}, \ldots, f_{S|r}) =$$

$$= \textit{rozdělení relativních četností v r-tém řádku tabulky}$$

$$(r = 1, \ldots, R;\ \sum_s f_{s|r} = 1) ,$$

$$\boldsymbol{f} = \sum_s f_{r+} \boldsymbol{f}_{(r)} = (f_1, f_2, \ldots, f_S) =$$

$$= \textit{marginální rozdělení (celý soubor)},$$

$$f_{r+} = \sum_s f_{rs},\ (f_{11}, f_{12}, \ldots, f_{RS}) =$$

$$= \textit{rozdělení relativních četností pro celou tabulku},$$

$$\sum_r \sum_s f_{rs} = \sum_r f_{r+} = 1 ,$$

$$\mathrm{Gvar}\, \boldsymbol{f}_{(r)} = \textit{míra variability v r-tém řádku } (r = 1, \ldots, R),$$

$$\mathrm{Gvar}\, \boldsymbol{f} = \textit{míra variability v celém souboru}.$$

---

[9] Uvedené míry rozptýlení jsou postupně vhodné pro nominální (nomvar), ordinální (dorvar), kardinální (rozptyl) a obecný typ proměnné. Vzorce pro výpočet jsou uvedeny v tab. A a v dodatku.

[10] Ke každé matici **D**, pro niž má model smysl, je určena nejen míra rozptylu, ale také míra rozdílnosti řádků tabulky, kterou je *vzdálenost distribucí*. Předpoklady pro platnost vzorce (1) a existenci vzdáleností řádků jsou natolik volné, že zahrnují jednak běžné typy znaků (nominální, ordinální, kardinální), ale také znaky geografické (areální), kombinované dichotomie, metrizované typologie, vícerozměrné ordinální a kardinální znaky apod. U těch typů, jejichž matice **D** požadavkům teorie nevyhovují, je možno provést matematickou úpravu, která skóry d_{st} změní tak, aby předpoklady byly splněny, a typ se přitom změnil jen nepatrně.

[11] Poslední člen vzorce (1) lze napsat také jako *„průměr čtverců vzdálenosti řádků od marginálního řádku"*. Přitom průměr je v celém vzorci chápán jako vážený. Vzorec (1) je analogický k analýze rozptylu: *TSS = WSS + BSS*, tj. *celkový součet čtverců = součet čtverců uvnitř skupin + součet čtverců mezi skupinami*.

[12] Plný název tohoto typu koeficientu asociace je *koeficient explanační síly rozkladu* **B** *pro proměnnou* **A**.

Koeficient explanace teď můžeme vyjádřit jednoduše jako

$$\delta_{A|B} = 1 - \frac{\sum_{r=1}^{R} f_{r+} \operatorname{Gvar} f_{(r)}}{\operatorname{Gvar} f}. \tag{4}$$

Tento obecný vzorec [13] zahrnuje širokou škálu speciálních případů, z nichž nejběžnější jsou specifikovány v tabulce A.

Koeficienty δ mají tyto vlastnosti:

1. jsou definovány vždy, je-li $\operatorname{Gvar} f > 0$ (tj. existuje nenulová variabilita, kterou chceme vysvětlit;)
2. $0 \leqq \delta \leqq 1$;
3. $\delta = 1$, právě když variability v jednotlivých řádcích zcela vymizí — buď se všechny jednotky v řádku tabulky soustředí v jedné kategorii proměnné A, nebo kategorie, v nichž se nalézají řádkové četnosti, nejsou z hlediska typu rozlišovány (skóry jejich nepodobnosti jsou nulové, z hlediska modelu splývají);
4. $\delta = 0$, právě když průměrný součet čtverců vzdáleností mezi řádky je nulový, to znamená, že buď se sobě všechna rozdělení rovnají (případ statistické nezávislosti, který platí např. u nominálního nebo ordinálního typu), nebo všechna rozdělení mají určité stejné dílčí vlastnosti dané typem (např. u kardinálního znaku je to rovnost průměrů, u vícerozměrného kardinálního znaku je to shoda těžišť);[14]
5. $\delta_1 > \delta_2$ znamená, že rozdílnosti mezi řádky jsou v prvním případě větší než ve druhém (měřeno relativně k celkovému rozptylu dat).

Příklady na výpočet explanačního koeficientu pro nominální a ordinální typ proměnné A může čtenář nalézt v pracech J. Řeháka a B. Řehákové [1973] (koeficient τ) a J. Řeháka [1976] (koeficient β). Kardinální typ vede na běžný koeficient determinace (korelační poměr η^2). Proto jako příklad uvedeme výpočet koeficientu explanace pro obecný typ.

Příklad 1. Kombinovaná dichotomie

Ve výzkumu přání a návrhů dělníků v jednom podniku byly (za pomoci faktorové analýzy a kategorizace spojitých škál) určeny dvě trichotomické ordinální proměnné: *orientace respondentů na sociální a na výrobně investiční opatření*. Po vynechání středních kategorií nevyhraněného názoru (a příslušné redukci souboru) vznikly dvě dichotomické proměnné

$$\mathbf{S} = (S, \bar{S}) = (\textit{vyžaduje opatření sociálního typu, nevyžaduje}),$$
$$\mathbf{V} = (V, \bar{V}) = (\textit{vyžaduje opatření výrobně investičního typu, nevyžaduje}).$$

Každý z těchto znaků můžeme analyzovat zvlášť, jejich kombinací však přijímáme komplexnější pohled na respondenta, hodnotíme současně oba aspekty. Kombinace

[13] Obecné vyjádření má několik důležitých významů: 1. jednotnost interpretace koeficientů pro různě zvolené prvky matice **D**, 2. jednotnost výpočetních procedur pro různé typy, 3. možnost pracovat s takovými typy znaků, které odpovídají danému analytickému požadavku, tj. lze určovat prvky **D** případ od případu jako *model konkrétní situace*, 4. úspora místa a zvýšení přehlednosti při psaní vzorců, které nemusí být opakovány pro každý jednotlivý případ zvlášť (výhoda při programování).

[14] Odlišení obou situací v matematickém smyslu: jestliže matice skórů $\mathbf{D} = \|d_{st}\|$ vytváří metriku, pak jde o *úplnou nezávislost*, pokud vytváří polometriku, pak jde o **D**-*nezávislost*, jež je charakterizována tím, že všechna řádková rozložení patří do téže třídy ekvivalence vytvářené polometrikou.

Tabulka A. Speciální případy koeficientu explanační síly rozkladu proměnné **B** *pro proměnnou* **A**

Typ **A**	Skóry d_{st}	Vzorec Gvar	Vzorec koeficientu δ
Nominální	$d_{st} = 1,\ s \neq t$ $d_{tt} = 0$	$\text{nomvar}\, f = 1 - \sum_{s=1}^{S} f_s^2$	$\tau = \dfrac{\sum_{r=1}^{R} f_{r+} \sum_{s=1}^{S} f_{s\|r}^2 - \sum_{s=1}^{S} f_s^2}{1 - \sum_{s=1}^{S} f_s^2}$
Ordinální	$d_{st} = \|s - t\|$	$\text{dorvar}\, f = 2 \sum_{s=1}^{S-1} F_s(1 - F_s)$ $F_s = \sum_{t=1}^{s} f_t$	$\beta = 1 - \dfrac{\sum_{r=1}^{R} f_{r+} \sum_{s=1}^{S-1} F_{s\|r}(1 - F_{s\|r})}{\sum_{s=1}^{S-1} F_s(1 - F_s)}$ $F_{s\|r} = \sum_{t=1}^{s} f_{t\|r}, \quad F_s = \sum_{t=1}^{s} f_t$
Kardinální $(x_1, \ldots, x_S)$	$d_{st} = (x_s - x_t)^2$	$\text{Gvar}\, f = 2 \sum_{s=1}^{S} f_s(x_s - \bar{X})^2$ $\bar{X} = \sum_{s=1}^{S} f_s x_s$	$\eta^2 = 1 - \dfrac{\sum_{r=1}^{R} f_{r+} \sum_{s=1}^{S} f_{s\|r}(x_s - \bar{X}_{(r)})^2}{\sum_{s=1}^{S} f_s(x_s - \bar{X})^2}$ $\bar{X}_{(r)} = \sum_{s=1}^{S} f_{s\|r} x_s, \quad \bar{X} = \sum_{s=1}^{S} f_s x_s$
Obecný	d_{st}	$\text{Gvar}\, f = \sum_{s=1}^{S} \sum_{t=1}^{S} f_s f_t d_{st}$	$\delta = 1 - \dfrac{\sum_{r=1}^{R} f_{r+} \text{Gvar}\, f_{(r)}}{\text{Gvar} f}$

$\mathbf{S} \times \mathbf{V}$ je čtyřhodnotový znak **OPAT** = $(\boldsymbol{SV}, \boldsymbol{S\bar{V}}, \boldsymbol{\bar{S}V}, \boldsymbol{\bar{S}\bar{V}})$, v němž spojení dvou písmen značí současný výskyt nebo nevýskyt obou požadavků. Tento znak není čistě nominální, i když v běžných analýzách jej za takový považujeme. Podobnost kategorií $\boldsymbol{SV}$ a $\boldsymbol{S\bar{V}}$ je větší než $\boldsymbol{SV}$ a $\boldsymbol{\bar{S}V}$, resp. $\boldsymbol{S\bar{V}}$ a $\boldsymbol{\bar{S}\bar{V}}$. Abychom tuto skutečnost zahrnuli do analýzy dat, určíme u znaku **OPAT** (opatření) skóry d_{st} vytvářející jeho typ, a to zcela přirozeně jako počet neshod mezi dvěma kategoriemi:

$$d(\boldsymbol{SV}, \boldsymbol{S\bar{V}}) = d(\boldsymbol{SV}, \boldsymbol{\bar{S}V}) = d(\boldsymbol{S\bar{V}}, \boldsymbol{\bar{S}\bar{V}}) = d(\boldsymbol{\bar{S}V}, \boldsymbol{\bar{S}\bar{V}}) = 1$$

(kategorie se liší vždy jen v jednom směru),

$$d(\boldsymbol{SV}, \boldsymbol{\bar{S}\bar{V}}) = d(\boldsymbol{S\bar{V}}, \boldsymbol{\bar{S}V}) = 2$$

(kategorie se liší v obou směrech).

Matici **D** tak dostáváme[15] jako

$$\mathbf{D} = \begin{array}{c} \\ \boldsymbol{SV} \\ \boldsymbol{S\bar{V}} \\ \boldsymbol{\bar{S}V} \\ \boldsymbol{\bar{S}\bar{V}} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{SV} & \boldsymbol{S\bar{V}} & \boldsymbol{\bar{S}V} & \boldsymbol{\bar{S}\bar{V}} \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Tabulka 1 zachycuje třídění druhého stupně mezi řádkovým znakem $\mathbf{B} \equiv$ aktivnost respondenta (aktivní, pasivní), který byl odvozen z jiných údajů dotazníku a znakem **OPAT**.

Tabulka 1. Návrhy na opatření podle aktivity respondentů
(souboru dělníků zkoumaného podniku)

	Respondent navrhuje opatření				
	sociální i výrobně investiční	sociální	výrobně investiční	žádné	
Aktivní typ	73 (68 %)	15 (14 %)	14 (13 %)	5 (5 %)	107 (100 %)
Pasívní typ	8 (8 %)	8 (8 %)	10 (11 %)	69 (73 %)	95 (100 %)
Celkem	81 (40 %)	23 (11 %)	24 (12 %)	74 (37 %)	202 (100 %)

[15] Obdobně bychom určili matici **D** pro kombinaci většího počtu dichotomií. Pro $\mathbf{B} = (\boldsymbol{B}, \boldsymbol{\bar{B}})$, $\mathbf{C} = (\boldsymbol{C}, \boldsymbol{\bar{C}})$, $\mathbf{D} = (\boldsymbol{D}, \boldsymbol{\bar{D}})$ dostaneme znak o osmi (2^3) kategoriích **KOMB** = $(\boldsymbol{BCD}, \boldsymbol{BC\bar{D}}, \boldsymbol{B\bar{C}D}, \boldsymbol{\bar{B}CD}, \boldsymbol{B\bar{C}\bar{D}}, \boldsymbol{\bar{B}C\bar{D}}, \boldsymbol{\bar{B}\bar{C}D}, \boldsymbol{\bar{B}\bar{C}\bar{D}})$. Skóry nepodobnosti jsou

$$d(\boldsymbol{BCD}, \boldsymbol{BC\bar{D}}) = d(\boldsymbol{BCD}, \boldsymbol{B\bar{C}D}) = d(\boldsymbol{BCD}, \boldsymbol{\bar{B}CD}) = d(\boldsymbol{BC\bar{D}}, \boldsymbol{B\bar{C}\bar{D}}) =$$
$$d(\boldsymbol{BC\bar{D}}, \boldsymbol{\bar{B}C\bar{D}}) = d(\boldsymbol{B\bar{C}D}, \boldsymbol{B\bar{C}\bar{D}}) = d(\boldsymbol{B\bar{C}D}, \boldsymbol{\bar{B}\bar{C}D}) = d(\boldsymbol{\bar{B}CD}, \boldsymbol{\bar{B}C\bar{D}}) =$$
$$d(\boldsymbol{\bar{B}CD}, \boldsymbol{\bar{B}\bar{C}D}) = d(\boldsymbol{B\bar{C}\bar{D}}, \boldsymbol{\bar{B}\bar{C}\bar{D}}) = d(\boldsymbol{\bar{B}C\bar{D}}, \boldsymbol{\bar{B}\bar{C}\bar{D}}) = 1,$$
$$d(\boldsymbol{BCD}, \boldsymbol{\bar{B}\bar{C}\bar{D}}) = d(\boldsymbol{BC\bar{D}}, \boldsymbol{\bar{B}C\bar{D}}) = d(\boldsymbol{B\bar{C}D}, \boldsymbol{\bar{B}\bar{C}D}) = d(\boldsymbol{\bar{B}CD}, \boldsymbol{B\bar{C}\bar{D}}) = 3,$$

zbytek jsou dvojky (na diagonále ovšem nuly). Pro kombinaci K dichotomií dostaneme znak s 2^K hodnotami a skóry od nuly do K.

Z dat tabulky 1 a skórů d_{st} snadno spočteme míry variability a koeficient explanace (výpočet zobecněné variance podle vzorce (14), δ podle (4)):

$$\text{Gvar}\, f_{(1)} = \frac{2}{107^2}\,(73 \times 15 + 73 \times 14 + 15 \times 5 + 14 \times 5 + \\ + 2 \times 73 \times 5 + 2 \times 15 \times 14) = 0.596,$$

$$\text{Gvar}\, f_{(2)} = 0.587,\ \text{Gvar}\, f = 0.999,\ f_{1+} = 0.53,\ f_{2+} = 0.47,$$

$$\delta = 1 - \frac{0.596 \times 0.53 + 0.587 \times 0.47}{0.999} = 0.408.$$

Znak aktivnost tedy vysvětluje 41 % variability znaku **OPAT** (velmi silná asociace).

Pro srovnání uvedeme také výsledky pro případ, že znak **OPAT** považujeme za nominální:

$$\text{nomvar}\, f_{(1)} = 0.496, \qquad \text{nomvar}\, f_{(2)} = 0.447,$$

$$\text{nomvar}\, f = 0.678, \qquad \tau = 0.302,$$

což ukazuje na 30 % vysvětlené variance. To zasluhuje komentář:

1. různé typy variancí nelze přímo číselně srovnávat, neboť skóry z matice **D** přinášejí do výpočtů numerický posun v délce škály pro hodnoty míry. Maximální hodnota pro nomvar čtyřhodnotového znaku je 0.75, pro kombinovanou dichotomii s uvedenou maticí **D** je 1.00. Pro přímou komparaci lze obě míry normalizovat poměrem k dosažitelnému maximu (v našem příkladě lze porovnávat nomvar/0.75 a Gvar);
2. δ je vyšší než τ, což znamená, že závislost (rozdílnost v četnostech řádků) se projevuje především u kategorií, které se od sebe více liší. V našem případě je to u dvojice $\boldsymbol{SV}$, $\bar{\boldsymbol{S}}\boldsymbol{V}$ a nebo u dvojice $\bar{\boldsymbol{S}}\boldsymbol{V}$, $\boldsymbol{S}\bar{\boldsymbol{V}}$. I když je nominální asociace vysoká, je ještě posílena přijetím modelu kombinované dichotomie. Naopak pokles hodnoty δ by znamenal, že rozdíly se projevují ve významem bližších kategoriích. Skóry d_{st} tak vyjadřují váhy důležitosti pro přesun četností ve zkoumání asociací.[16]

Nyní přejdeme velmi jednoduše k měření vícenásobné asociace. Pro tabulku třídění třetího stupně $\mathbf{B}_1 \times \mathbf{B}_2 \times \mathbf{A}$, v níž **A** považujeme za vysvětlovanou a $\mathbf{B}_1$, $\mathbf{B}_2$ za vysvětlující proměnné, zavedeme *koeficient vícenásobné asociace*[17] následovně:

a) Určíme kombinovaný znak $\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2$, který má $R = R_1 \times R_2$ hodnot vzniklých kombinací všech R_1 hodnot znaku $\mathbf{B}_1$ se všemi R_2 hodnotami znaku $\mathbf{B}_2$.

b) Provedeme třídění druhého stupně $\mathbf{B} \times \mathbf{A}$, které dává tabulku o R řádcích a S sloupcích (je to zároveň třídění třetího stupně $\mathbf{B}_1 \times \mathbf{B}_2 \times \mathbf{A}$).

c) Pro vzniklou tabulku zjistíme koeficient explanace $\delta_{\mathbf{A}/\mathbf{B}}$ a prohlásíme jej za koeficient vícenásobné asociace mezi dvojicí[18] ($\mathbf{B}_1$, $\mathbf{B}_2$) a **A**, takže

$$\delta_{\mathbf{A}/\mathbf{B}_1,\,\mathbf{B}_2} = \delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} = \delta_{\mathbf{A}/\mathbf{B}}\,. \tag{5}$$

[16] Typickým dalším případem je geografický znak, v němž kategorie $\boldsymbol{A}_s$ jsou ohraničená území nebo body v ploše (na mapě, v krajině) a skóry **D** jsou nějaké zvolené míry vzdálenosti mezi nimi (např. vzdušná vzdálenost mezi středy území nebo její čtverec, vzdálenost po silnici ap.). Obdobné rozdíly v asociaci jsou zde velmi názorné: stupeň asociace určuje rozdíl v zastoupení kategorií a vzdálenost kategorií, u nichž dochází k rozdílu.

[17] Lze jej také nazvat koeficientem explanační síly kombinovaného rozkladu $\mathbf{B}_1 \times \mathbf{B}_2$. Používá se též termín koeficient násobné nebo mnohonásobné asociace.

[18] Obojí značení ve vzorci (5) je možné. Používáme však spíše $\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2}$, aby se tím i symbolicky zdůraznilo, že do modelu vstupuje kombinace $\mathbf{B}_1$ a $\mathbf{B}_2$.

Zcela obdobně postupujeme pro K vysvětlujících proměnných $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K$. Definujeme kombinovanou proměnnou $\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2 \times \ldots \times \mathbf{B}_K$ a koeficient vícenásobné asociace jako

$$\delta_{\mathbf{A}/\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_k} = \delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2 \times \cdots \times \mathbf{B}_k} = \delta_{\mathbf{A}/\mathbf{B}}. \tag{6}$$

Koeficient vícenásobné asociace má obdobné vlastnosti jako prostý koeficient asociace (je stejně definován). Navíc má ještě tu vlastnost, že *přidáním další proměnné se jeho hodnota nezmenší*, tj.

$$\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} \geqq \delta_{\mathbf{A}/\mathbf{B}_1}, \quad \delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} \geqq \delta_{\mathbf{A}/\mathbf{B}_2} \tag{7}$$

a obecně

$$\delta_{\mathbf{A}/\mathbf{B}_1 \times \cdots \times \mathbf{B}_k} \geqq \delta_{\mathbf{A}/\mathbf{B}_1 \times \cdots \times \mathbf{B}_{k-1}} \tag{8}$$

2. Parciální asociace ve víceozměrné kontingenční tabulce

Pojem parciální asociace se vyskytuje v souvislosti se dvěma analytickými otázkami:

a) při působení skupiny proměnných (faktorů) na závislou proměnnou **A** se ptáme na *čistý vlastní přínos každé proměnné v rámci dané skupiny*;

b) při sledování vztahů a budování kauzálních modelů se ptáme, zda vztah dvou proměnných je *přímý*, či *zprostředkovaný* (resp. *částečně zprostředkovaný*) další proměnnou nebo skupinou proměnných. Přidáme-li ke vztahu $\mathbf{C} \rightarrow \mathbf{A}$ další faktor **B** a čistý přínos **C** v rámci působení dvojice **B**, **C** vymizí, znamená to, že proměnná **B** pohltila statistickou asociaci mezi **C** a **A** a *asociace se řetězí*[19] $\mathbf{C} \rightarrow \mathbf{B} \rightarrow \mathbf{A}$, vazba **C** a **A** je *zprostředkovaná* proměnnou **B**.

Obě situace budeme demonstrovat na příkladu z praxe.

Příklad 2. Zájmové orientace

Proměnná zájmové orientace (**A**) vznikla kombinací tří dichotomií (má tedy 8 hodnot). Při studiu vlivu sociodemografických znaků, které jsou pro danou populaci mezi sebou značně závislé, sledujeme, které asociace jsou odrazem skutečného ovlivňování a které jsou jen *nepravé*, způsobené nikoli příčinnými vazbami, ale pouze strukturními souvislostmi. V analýze dat sociologických výzkumů obvykle zjišťujeme asociaci mezi **A** a znaky: pohlaví, věková skupina, velikost místa bydliště, vzdělání, socioprofesní skupina ap. Přímá asociace může však být zavádějící, a proto je vhodné určit čistý vliv (parciální asociaci) každého z těchto znaků na **A** v rámci jejich celku.[20]

Dále se můžeme ptát, zda asociace zájmových orientací a pohlaví je způsobena rozdílnými vlastnostmi a přístupy u mužů a žen či zda je zprostředkována jejich možnostmi, rozdílnými povinnostmi ap. To zjistíme zkoumáním platnosti vztahu: *pohlaví* → *povinnosti v rodině* → *zájmy volného času* (proměnná povinnosti v rodině byla konstruována z několika dílčích znaků).

Prostředkem k řešení uvedených úloh je *koeficient parciální asociace*. Jeho význam, interpretace i využitelnost jsou zřejmé ze zavedení, které provedeme v postupných jednoduchých logických krocích. Omezíme se na třírozměrnou tabulku $\mathbf{B}_1 \times \mathbf{B}_2 \times \mathbf{A}$, tj. na případ dvou faktorů ovlivňujících **A**.

---

[19] K přijetí modelu $\mathbf{C} \rightarrow \mathbf{B} \rightarrow \mathbf{A}$ je ovšem nutné, aby byla prokázána statistická asociace $\mathbf{C} \rightarrow \mathbf{B}$ a aby směr šipek dával logický věcný smysl.

[20] Postup je analogický k určení koeficientů vícenásobné lineární regrese v případě číselných proměnných.

1. Přímou asociaci $\mathbf{B}_1 \to \mathbf{A}$ vyjádříme koeficientem $\delta_{\mathbf{A}/\mathbf{B}_1}$.
2. Přímou asociaci $(\mathbf{B}_1, \mathbf{B}_2) \to \mathbf{A}$ vyjádříme koeficientem $\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2}$.
3. Přírůstek asociace zaznamenaný přidáním $\mathbf{B}_2$ je

$$\Delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} = \delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} - \delta_{\mathbf{A}/\mathbf{B}_1} .$$

4. Takto definovaná charakteristika je vždy nezáporná, ale má horní hranici $1 - \delta_{\mathbf{A}/\mathbf{B}_1}$ (to vyjadřuje stupeň nevysvětlené variability proměnné **A** před vstupem $\mathbf{B}_2$ do modelu, ta se však mění podle hodnoty $\delta_{\mathbf{A}/\mathbf{B}_1}$, a tudíž je nesrovnatelná pro různé tabulky i různé proměnné v jedné vícerozměrné tabulce.
5. Aby byla zajištěna srovnatelnost, definujeme koeficient parciální asociace $\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)}$ normalizací jako podíl k maximu:

$$\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} = \frac{\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} - \delta_{\mathbf{A}/\mathbf{B}_1}}{1 - \delta_{\mathbf{A}/\mathbf{B}_1}} . \tag{9}$$

Takto definovaný *koeficient parciální asociace 1. řádu* má vlastnosti:

1. je definován kdykoliv $\delta_{\mathbf{A}/\mathbf{B}_1} < 1$, tj. pokud není variabilita A vysvětlena první proměnnou $\mathbf{B}_1$ úplně (nutnou podmínkou pro celou analýzu je ovšem existence $\delta_{\mathbf{A}/\mathbf{B}_1}$, tj. Gvar $\boldsymbol{f} > 0$);
2. $0 \leqq \delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} \leqq 1$;
3. $\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} = 1$ právě když $\mathbf{B}_2$ vysvětluje veškerou zbylou variabilitu **A**, tj. $\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} = 1$;
4. $\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} = 0$ právě když přírůstek Δ je nulový, tj. když $\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2} = \delta_{\mathbf{A}/\mathbf{B}_1}$, $\mathbf{B}_2$ nepřináší žádnou novou redukci variability;
5. $\delta_1 > \delta_2$ znamená, že v první situaci nastává vyšší podíl redukce variability dodatečnou proměnnou než v druhém;
6. má *interpretaci relativního úbytku dosud nevysvětlené variability.*

Výpočet koeficientu parciální asociace 1. řádu $\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)}$ se provádí dvěma způsoby:

1. způsob:
   - spočítáme $\delta_{\mathbf{A}/\mathbf{B}_1}$ z tabulky $\mathbf{B}_1 \times \mathbf{A}$
   - spočítáme $\delta_{\mathbf{A}/\mathbf{B}_1 \times \mathbf{B}_2}$ z tabulky $\mathbf{B} \times \mathbf{A}$, kde $\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2$ je kombinace faktorů $\mathbf{B}_1$ a $\mathbf{B}_2$
   - dosadíme do vzorce (9).
2. způsob:
   - spočítáme řádkové míry rozptýlení Gvar $\boldsymbol{f}_{(r)}$, $r = 1, 2, \ldots, R_1$, v tabulce $\mathbf{B}_1 \times \mathbf{A}$ a marginální rozdělení $(f_{1++}, f_{2++}, \ldots, f_{R_1++})$ znaku $\mathbf{B}_1$
   - spočítáme řádkové míry rozptýlení Gvar $\boldsymbol{f}_{(rq)}$, $r = 1, \ldots, R_1$, $q = 1, \ldots, R_2$ v tabulce $\mathbf{B} \times \mathbf{A}$ typu $R \times S$, kde $R = R_1 \times R_2$ a $\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2$ je kombinace obou faktorů a určíme marginální rozdělení[21] $(f_{11+}, \ldots, f_{rq+}, \ldots, f_{R_1R_2+})$ znaku **B**
   - získané hodnoty dosadíme do vzorce

$$\delta_{\mathbf{A}/\mathbf{B}_2(\mathbf{B}_1)} = 1 - \frac{\sum_{r=1}^{R_1} \sum_{q=1}^{R_2} f_{rq+} \operatorname{Gvar} \boldsymbol{f}_{(rq)}}{\sum_{r=1}^{R_1} f_{r++} \operatorname{Gvar} \boldsymbol{f}_{(r)}} \tag{10}$$

[21] Dvojice indexu „rq„ odpovídá řádku, v němž je uvedeno rozložení ke kombinaci r-té hodnoty znaku $\mathbf{B}_1$ a q-té hodnoty znaku $\mathbf{B}_2$.

Výraz pro výpočet *přibližné směrodatné chyby* je uveden v dodatku.

Rozšíření na parcializaci M-tého řádu, při níž sledujeme čistý vliv proměnné $\mathbf{B}_{M+1}$ po odečtení vlivu M proměnných $\mathbf{B}_1, \ldots, \mathbf{B}_M$ je snadné. *Koeficient parciální asociace M-tého řádu* definujeme

$$\delta_{\mathbf{A}/\mathbf{B}_{M+1}(\mathbf{B}_1, \cdots \mathbf{B}_M} = \delta_{\mathbf{A}/\mathbf{B}_{M+1}(\mathbf{B}^*)}\,, \tag{11}$$

kde $\mathbf{B}^* = \mathbf{B}_1 \times \ldots \times \mathbf{B}_M$ je kombinace M proměnných. Výpočet se provádí podle předchozích postupů tak, že místo $\mathbf{B}_2$ dosadíme $\mathbf{B}_{M+1}$, místo $\mathbf{B}_1$ vystupuje $\mathbf{B}^*$ a $\mathbf{B} = \mathbf{B}_1 \times \ldots \times \mathbf{B}_{M+1}$. Tak je výpočet *rekurentní*, vícerozměrný případ je převeden na třírozměrný a není nutné uvádět složité explicitní vzorce pro obecný případ. V kontextu úvah o vícerozměrné kontingenční tabulce se přímý koeficient asociace nazývá také *parciální asociační koeficient nultého řádu.*[22]

Zcela obdobně je možno zavést *koeficient parciální asociace skupiny proměnných při odečtení vlivu jiné skupiny.* Například v modelu $(C_1, C_2, C_3, C_4) \rightarrow \mathbf{A}$, můžeme určit koeficient $\delta_{\mathbf{A}/\mathbf{C}_3\times\mathbf{C}_4(\mathbf{C}_1,\mathbf{C}_2)}$ tak, že do vzorce (9) dosadíme $\mathbf{B}_1 = \mathbf{C}_1 \times \mathbf{C}_2$, $\mathbf{B}_2 = \mathbf{C}_3 \times \mathbf{C}_4$. Tento postup umožňuje podrobné a úplné analýzy asociací ve vícerozměrných tabulkách.

Mezi koeficienty asociace platí užitečné vztahy, které jsou známé u parciálních korelací a korelačních poměrů:

$$\begin{aligned} 1 - \delta_{\mathbf{A}/\mathbf{B}\times\mathbf{C}} &= (1 - \delta_{\mathbf{A}/\mathbf{B}})\,(1 - \delta_{\mathbf{A}/\mathbf{C}(\mathbf{B})}) \\ &= (1 - \delta_{\mathbf{A}/\mathbf{C}})\,(1 - \delta_{\mathbf{A}/\mathbf{B}(\mathbf{C})}), \end{aligned} \tag{12}$$

$$\begin{aligned} 1 - \delta_{\mathbf{A}/\mathbf{B}\times\mathbf{C}\times\mathbf{D}} &= (1 - \delta_{\mathbf{A}/\mathbf{B}\times\mathbf{C}})\,(1 - \delta_{\mathbf{A}/\mathbf{D}(\mathbf{B}\times\mathbf{C})}) \\ &= (1 - \delta_{\mathbf{A}/\mathbf{B}})\,(1 - \delta_{\mathbf{A}/\mathbf{C}(\mathbf{B})})\,(1 - \delta_{\mathbf{A}/\mathbf{D}(\mathbf{B}\times\mathbf{C})})\,. \end{aligned} \tag{13}$$

3. Příklady využití parciální asociace

Model přímé, vícenásobné a parciální asociace umožňuje provádět jednak výběr přímo působících faktorů na závislou proměnnou a jednak konstruovat modely kauzálního řetězení.

Příklad 3. Zájem o čtení rubrik

Populace:čtenáři týdeníku ve věku do 30 let; *závislá proměnná*: kombinovaná dichotomie **ZAJMY** spojující volbu tří rubrik, které byly obsahově zaměřeny k témuž účelu a pro mladé čtenáře.[23]

Faktor	Symbol	Koeficient	
věková skupina	**VEK**	0.356	(přímá asociace)
vzdělání	**VZD**	0.472	(přímá asociace)
věková skupina v kombinaci se vzděláním	**VEK** × **VZD**	0.516	(vícenásobná asociace)

[22] Koeficienty $\delta_{\mathbf{A}|\mathbf{B}_1}$, $\delta_{\mathbf{A}|\mathbf{B}_2}$ se mohou poněkud numericky lišit při výpočtech z dvourozměrných tabulek $\mathbf{B}_1 \times \mathbf{A}$ a $\mathbf{B}_2 \times \mathbf{A}$ a při výpočtu z trojrozměrné tabulky $\mathbf{B}_1 \times \mathbf{B}_2 \times \mathbf{A}$, neboť u té může nastat vyšší redukce vynechávaných hodnot.

[23] Při studiu asociací se ukázal nepatrný zájem u žen a proto údaje byly dále analyzovány jen u mužské části výběrového souboru ve věku 15—30 let.

Tabulka B. Souhrnná tabulka vztahů mezi třemi kategorizovanými proměnnými — asymetrická asociace s jednou závislou proměnnou (**A**) *a dvěma faktory* (**B**$_1$, **B**$_2$)

Vztahy mezi proměnnými	Asociace					
Model	(1) $(B_1 \times B_2) \rightarrow A$	(2) $B_1 \rightarrow A$	(3) $B_2 \rightarrow A$	(4) $B_1 \leftrightarrow B_2$	(5) $B_1 \rightarrow A(B_2)$	(6) $B_2 \rightarrow A(B_1)$
1 **B**$_1$ a **B**$_2$ nejsou s **A** asociovány	0	0	0	~	0	0
2 Dvě nezávislé příčiny	+	+	+	0	■ (2)	■ (3)
3 Dvě nezávislé příčiny s interakčním efektem	+	+	+	0	> (2)	> (3)
4 Dvě vzájemně závislé příčiny	+	+	+	+	+	+
5 Dvě závislé příčiny s interakčním efektem	+	+	+	+	> (2)	> (3)
6 Skrytá interakční asociace dvou proměnných	+	0	0	~	+	+
7a Řetězení $B_1 \rightarrow B_2 \rightarrow A$	+	+	+	+	0	+
7b Řetězení $B_2 \rightarrow B_1 \rightarrow A$	+	+	+	+	+	0
8a Nepravá asociace $B_1 \rightarrow A$	+	+	+	+	0	+
8b Nepravá asociace $B_2 \rightarrow A$	+	+	+	+	+	0

Pokračování tab. B

Vztahy mezi proměnnými	Asociace					
Model	(1) $(B_1 \times B_2) \to A$	(2) $B_1 \to A$	(3) $B_2 \to A$	(4) $B_1 \leftrightarrow B_2$	(5) $B_1 \to A(B_2)$	(6) $B_2 \to A(B_1)$
9a Asociace $B_1 \to A$ je částečně zprostředkována B_2	+	+	+	+	+ < (2)	+
9b Asociace $B_2 \to A$ je částečně zprostředkována B_1	+	+	+	+	+	+ < (3)
10a Isolovaná proměnná B_1	+	0	+	0	0	+ ■ (3)
10b Isolovaná proměnná B_2	+	+	0	0	+ ■ (2)	0

Použité symboly:
+ = existence asociačního vztahu (statistická závislost)
0 = absence asociačního vztahu (statistická nezávislost)
$>$ (*s*) = hodnota výrazně vyšší než ve sloupci (*s*)
$<$ (*s*) = hodnota výrazně nižší než ve sloupci (*s*)
■ (*s*) = hodnota přibližně stejná jako ve sloupci (*s*)
~ = libovolná hodnota koeficientu

Poznámky:
a) Asymetrie v modelech a typ **A** určujeme podle obsahu proměnných a podle vnějších hledisek
b) Modifikace modelu 6 vzniká v případě, že jedna ze dvou asociací (2), (3) je kladná a při parcializaci se zvýší (5) > (2), resp. (6) > (3)
c) Model 8a ukazuje na dvě závislé proměnné B_1 a **A** se společnou příčinou B_2
Model 8b ukazuje na dvě závislé proměnné B_2 a **A** se společnou příčinou B_1

U mladých čtenářů lze předpokládat dosti silné statistické asociace mezi věkem a vzděláním, a tudíž i překrývání vlivu na zájem o rubriky. Parciální asociační koeficienty spočteme podle (9)

$$\delta_{\mathbf{ZAJMY} \mid \mathbf{VEK\ (VZD)}} = \frac{0.516 - 0.472}{1 - 0.472} = 0.083\,,$$

$$\delta_{\mathbf{ZAJMY} \mid \mathbf{VZD\ (VEK)}} = \frac{0.516 - 0.356}{1 - 0.356} = 0.248\,.$$

Silná redukce v prvním případě naznačuje, že věk nepůsobí přímo, ale že zájmy jsou ovlivněny dosaženým stupněm vzdělání, a to je statisticky závislé na věku. Zprostředkovanost, resp. nepravost asociace není však úplná, věk má 8 % vlastního čistého vlivu na proměnlivost zájmů. Hypotetický model **VEK** → **VZD** → **ZAJMY** se potvrzuje, je však nutno jej doplnit o přímý vliv **VEK** → **ZAJMY**. Druhá parciální asociace ukazuje na podstatně nižší redukci, z čehož plyne důležitost vzdělání jako samostatného faktoru zájmu o rubriky.

Přidáním proměnné *způsob čtení časopisu* (**CT**), která má pět hodnot, můžeme ilustrovat ještě využitelnost parciálních asociací druhého řádu. Přehled všech přímých, násobných a parciálních asociačních koeficientů podává tabulka 2.

Tabulka 2. Asociační koeficienty ve vztahu mezi věkem (**VEK**), *vzděláním* (**VZD**), *čtenářským typem* (**CT**) *a zájmem o skupinu rubrik* (**ZAJMY**)

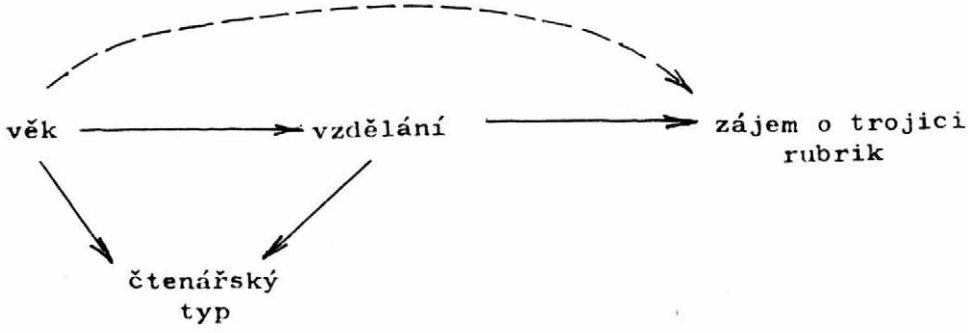
Faktory	Podmínka v parciální asociaci	Hodnota koeficientu	Typ asociace
VEK	—	0.356	přímá
VZD	—	0.472	
CT	—	0.232	
VEK × **VZD**	—	0.516	násobná
VEK × **CT**	—	0.392	
VZD × **CT**	—	0.491	
VEK × **VZD** × **CT**	—	0.527	násobná
VEK	**VZD**	0.083	parciální prvního řádu
VZD	**VEK**	0.248	
CT	**VEK**	0.056	
VEK	**CT**	0.208	
VZD	**CT**	0.337	
CT	**VZD**	0.036	
VEK	**VZD** × **CT**	0.071	parciální druhého řádu
VZD	**VEK** × **CT**	0.222	
CT	**VEK** × **VZD**	0.023	

Parciální asociace se spočtou podle (9) a (11), např.

$$\delta_{\mathbf{ZAJMY}\,|\,\mathbf{VEK}\,(\mathbf{VZD},\,\mathbf{CT})} = \frac{\delta_{\mathbf{ZAJMY}\,|\,\mathbf{VEK}\times\mathbf{VZD}\times\mathbf{CT}} - \delta_{\mathbf{ZAJMY}\,|\,\mathbf{VZD}\times\mathbf{CT}}}{1 - \delta_{\mathbf{ZAJMY}\,|\,\mathbf{VZD}\times\mathbf{CT}}} =$$

$$= \frac{0.527 - 0.491}{1 - 0.491} = 0.071\,.$$

Pokusme se interpretovat výsledky tabulky 2. Ze tří koeficientů parciální asociace druhého řádu je vidět, že **CT** v rámci skupiny (**CT**, **VEK**, **VZD**) má na **ZAJMY** jen nepatrný vliv. Po jeho vynechání přecházíme k analýze (**VEK**, **VZD**) → **ZAJMY**, která byla již provedena výše. Další parciální asociace ukazují na řetězení **CT** ← ← **VEK** → **ZAJMY** a **CT** ← **VZD** → **ZAJMY**, tedy na to, že způsob čtení se mění s věkem i se vzděláním a že přímá asociace se zájmy je nepravá. Výsledek naznačíme v obr. 1.



Obrázek 1. Vlivy na zájem o trojici rubrik.
Vzdělání má přímý silný vliv; věk působí vedle převážně zprostředkovaného vlivu dosaženým vzděláním také slabě přímo. Čtenářský typ se vyvíjí a mění s věkem a vzděláním, ale na vlastní zájem (preferenci ve struktuře rubrik) přímý vliv nemá

Příklad 4. Volba nejlepšího pořadu

Ve výzkumu televizních zájmů byl respondentům předložen úkol vybrat nejlepší pořad z pěti nabídnutých v seznamu. Souvislosti tohoto nominálního znaku s ostatními ukázaly mimo jiné koeficienty asociace

$$\tau_{\mathbf{PORAD}\,|\,\mathbf{POHL}} = 0.051\,,\quad \tau_{\mathbf{PORAD}\,|\,\mathbf{VEK}} = 0.078\,,$$

$$\tau_{\mathbf{PORAD}\,|\,\mathbf{POHL}\times\mathbf{VEK}} = 0.235\,.$$

Volba pořadu je z 5 % ovlivňována pohlavím a z 8 % věkem (čtyři věkové skupiny). To je asociace, která v kontextu analýzy dat může být důležitá, nicméně není nikterak výrazná. Násobná asociace však představuje skok, který ukazuje, že *v kombinaci obou faktorů vzniká nová kvalita vzhledem k výběru pořadu*. Zatímco muži a ženy ani věkové skupiny se od sebe ve volbě neliší, skupiny vzniklé kombinací (tj. mladší muži, muži středního věku, starší muži, mladší ženy atd.) se od sebe ve výběru liší velmi značně. Říkáme, že vzniká *interakční efekt dvou proměnných v ovlivňování závislé proměnné*. Asociace tohoto typu se z třídění druhého stupně nezjistí, proto jí říkáme *skrytá asociace*. Parciální koeficienty asociace za této situace vykazují značný

přírůstek

$$\tau_{\text{PORAD | VEK (POHL)}} = \frac{0.235 - 0.051}{1 - 0.051} = 0.195\,,$$

$$\tau_{\text{PORAD | POHL (VEK)}} = \frac{0.235 - 0.078}{1 - 0.078} = 0.170\,.$$

Analýza těchto dat pokračovala tak, že byly expertně oceněny skóry vytvářející typ ($\mathbf{D} = \|d_{st}\|$) proměnné **PORAD**. Při tomto chápání typu se přímé asociace zvýšily na 0.150, 0.170 a pro kombinaci na 0.354. Parciální asociace pro model D jsou

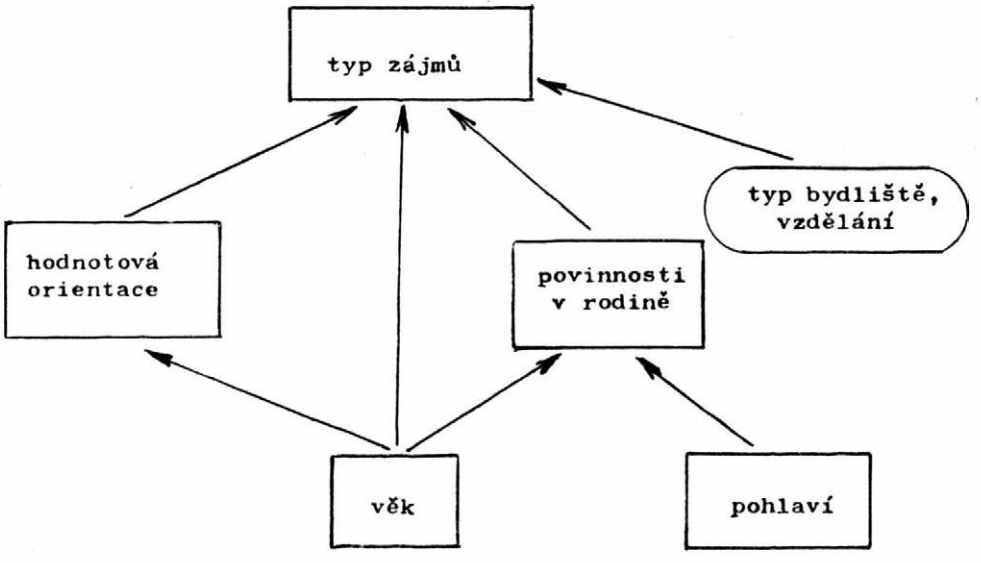
$$\delta_{\text{PORAD | VEK (POHL)}} = \frac{0.354 - 0.150}{1 - 0.150} = 0.240\,,$$

$$\delta_{\text{PORAD | VEK (POHL)}} = \frac{0.345 - 0.170}{1 - 0.170} = 0.222\,.$$

Závěr o interakčním působení se potvrzuje, značně vyšší koeficienty však ukazují na to, že rozdíly ve volbách mezi skupinami se projevují především u rozdílných pořadů. Pokles hodnot koeficientů by naopak ukázal na to, že rozdíly mezi skupinami se projevují u podobných pořadů, které by tak měly zastupitelnou roli.

Příklad 5. Faktory zájmových orientací

Vyšetřování vlivu šesti faktorů na zájmové orientace předpokládá třídění sedmého stupně. Výsledek na obr. 2 dává již složitější strukturu vztahů, lze však k němu dojít i studiem a porovnáváním trojic a čtveřic faktorů. Proměnná *typ zájmů* vznikla jako pětihodnotová typologie ze seskupovací analýzy, skóry nepodobnosti $\mathbf{D} = \|d_{st}\|$ byly určeny jako čtverce vzdálenosti centroidů skupin.



Obrázek 2. Síť vztahů vytvářejících typy zájmů ve volném čase.
Přímý vliv má hodnotová orientace, povinnosti v rodině, kombinace vzdělání a typu bydliště. Věk působí přímo i nepřímo, pohlaví nepřímo

Obr. 2 vyjadřuje tyto asociační vztahy:

a) Pohlaví (**POHL**), věk, kombinace typu bydliště a vzdělání jsou nezávislé.

b) Kombinace typu bydliště a vzdělání působí na typ zájmů (**Z**).

c) Věk působí na typ zájmů přímo, ale také prostřednictvím hodnotových orientací **HODOR** a povinností v rodině (**POVROD**) (obojí se s věkem mění). K tomu platí $\delta_{\mathbf{Z}\,|\,\mathbf{VEK\,(HODOR\times POVROD)}} > 0$, na věku statisticky závisí typ zájmů, hodnotové orientace i povinnosti v rodině, $\delta_{\mathbf{Z}\,|\,\mathbf{HODOR}} > 0$, $\delta_{\mathbf{Z}\,|\,\mathbf{POVROD}} > 0$, $\delta_{\mathbf{Z}\,|\,\mathbf{VEK(HODOR)}} > \delta_{\mathbf{Z}\,|\,\mathbf{VEK\,(HODOR\times POVROD)}}$, $\delta_{\mathbf{Z}\,|\,\mathbf{VEK\,(POVROD)}} > \delta_{\mathbf{Z}\,|\,\mathbf{VEK\,(HODOR\times POVROD)}}$.

d) Pohlaví je statisticky nezávislé na hodnotových orientacích, ale přímo statisticky závislé na povinnostech v rodině i na typu zájmů; přitom vymizí parciální asociace $\delta_{\mathbf{Z}\,|\,\mathbf{POHL\,(POVROD)}}$, a tudíž vztah pohlaví a typu zájmů je zprostředkován.

e) Vzájemná statistická závislost hodnotových orientací a povinností v rodině se ukázala jako nepravá, při podmínění věkem vymizí.

Příklad 6. Výběr oboru studia

Při analýze výběru oboru vysokoškolského studia studenty gymnázií se ukázaly takovéto asociace:

faktor	přímá asociace	parciální asociace (na všechny ostatní proměnné)
nejoblíbenější předmět	0.231	0.174
velikost místa bydliště	0.254	0.002
přání rodičů	0.278	0.073
zaměstnání a vzdělání rodičů	0.325	0.051
typ hodnotové orientace	0.371	0.320
pohlaví	0.457	0.423

Analýza vychází ze třídění sedmého stupně, proto kategorie u jednotlivých faktorů byly silně slučovány. Na základě přímých koeficientů explanace se faktory jeví jako relativně vyrovnané. Při parcializaci šestého řádu se však ukazuje, že velikost místa bydliště nemá žádný vliv a u přání rodičů a jejich vzdělání a zaměstnání se v asociaci ukazuje vysoká redukce. V rámci celé soustavy se tak faktory už vyrovnaně nejeví. Bezprostřední silný vliv na výběr mají pohlaví, typ hodnotové orientace a též i nejoblíbenější předmět, slabý vliv se ukazuje ze strany rodičů.

Závěr

Parciální a mnohonásobná asociace, tak jak zde byla uvedena, značně rozšiřuje možnosti statistické analýzy, jednak pro sám pojem parcializace a pro možnost hledat řetězené kauzální modely, přímé a nepravé vazby i skryté závislosti interakčního charakteru, jednak pro rozšíření pojmu na případ obecného typu proměnných a možnost zavést do analytického postupu další relevantní rysy zkoumaných jevů. Obecný typ proměnné zahrnuje pomocí skórů $\mathbf{D} = \|d_{st}\|$ běžné nominální, ordinální a kardinální případy, ale i kombinované dichotomie, vícerozměrně metrizované typologie, územně rozložené kategorie (enklávy) a celou řadu dalších sociologicky relevantních situací.

Obdobné úlohy parcializace a sestavování modelů lze řešit též pomocí logaritmicko-lineárních modelů (popřípadě jiných) a v mnohém bude výsledná interpretace shodná.

Tabulka C. Koeficienty asymetrické asociace δ pro K rozměrné kontingenční tabulky

	Nezávislá proměnná) (faktor)	Závislá proměnná[1])	Podmínka v parcializaci	Poznámka
Přímá asociace $\delta_{\mathbf{A}\mid\mathbf{B}}$				
1. základní koeficient explanační síly **B** pro **A**	**B**	**A**	—	**B** je chápána nominálně
2. vícenásobný koeficient explanace	$\mathbf{B} = (\mathbf{B}_1 \times \mathbf{B}_2 \times \ldots \times \mathbf{B}_M)$ $M \geqq 1$	**A**	—	Kombinace **B** je chápána nominálně
3. explanace vícerozměrné závislé proměnné	$\mathbf{B} = (\mathbf{B}_1 \times \mathbf{B}_2 \times \ldots \times \mathbf{B}_M)$ $M \geqq 1$	$\mathbf{A} = (\mathbf{A}_1 \times \mathbf{A}_2 \times \ldots \times \mathbf{A}_L)$ $L \geqq 1$	—	Matice **D** pro **A** může být odvozena z $\mathbf{D}_1, \mathbf{D}_2 \ldots \mathbf{D}_L$ **B** je nominální
Parciální asociace $\delta_{\mathbf{A}\mid\mathbf{B}(\mathbf{C})}$				
4. základní koeficient parciální asociace (1. řádu)	**B**	**A**	**C**	**B, C** nominální
5. koeficient parciální asociace **E**-tého řádu	**B**	**A**	$\mathbf{C} = (\mathbf{C}_1 \times \mathbf{C}_2 \times \ldots \times \mathbf{C}_E)$ $\mathbf{E} \geqq 2$	**B, C** nominální
6. skupinový koeficient parciální asociace **E**-tého řádu	$\mathbf{B} = (\mathbf{B}_1 \times \mathbf{B}_2 \times \ldots \times \mathbf{B}_M)$ $M \geqq 1$	$\mathbf{A} = (\mathbf{A}_1 \times \mathbf{A}_2 \times \ldots \times \mathbf{A}_L)$ $L \geqq 1$	$\mathbf{C} = (C_1 \times \mathbf{C}_2 \times \ldots \times \mathbf{C}_E)$ $\mathbf{E} \geqq 2$	matice **D** pro **A** může být odvozena z $\mathbf{D}_1, \mathbf{D}_2 \ldots \mathbf{D}_L$ **B, C** nominální

Poznámky: 1. U proměnné **A** lze zavést typ pomocí matice skórů $\mathbf{D} = \|d_{rl}\|$
2. $K = M + L + \mathbf{E}$, pro přímou asociaci je $\mathbf{E} = 0$

Přístup parciálních asociací má tyto *výhody*:

1. měří stupeň asociace, nejen její existenci;
2. shrnuje informaci z mnoha polí mnohorozměrné kontingenční tabulky do několika statistik, jejichž stabilitu můžeme určit pomocí intervalů spolehlivosti;
3. umožňuje pracovat s různými typy závislého znaku podle reálné situace a umožňuje srovnávat různé přístupy.

Má také *nevýhody*:

1. odhaluje pouze stupeň globální asociace, ale nespecifikuje interakce v polích, tj. vazby mezi jednotlivými kategoriemi;
2. netestuje hypotézy;
3. nemá v současné době tak dalece rozvinutou teorii ani programové prostředky.

V analýze dat bude vhodné kombinovat obě tyto metodiky, parciální asociační koeficienty i testování hypotetických závislostních modelů.

Obdobně jako u postupů testování hypotéz, nelze ani při používání parciálních asociací libovolně zvyšovat stupeň třídění. Počet proměnných, které vstupují do analýzy, je nutně omezen vzhledem k limitům na počet polí mnohorozměrné kontingenční tabulky, který roste geometrickou řadou. Proto v praxi často sestavujeme kauzální řetězení proměnných postupně analýzou segmentů souboru proměnných, trojic, čtveřic, případně i pětic. Jen málokdy si můžeme (i u velkých souborů dat) dovolit třídění vyšších stupňů. Takto sestavované sítě nemohou být úplně přesné a výsledky i postupy postrádají eleganci přesně aplikovaných matematických modelů vcelku. Nicméně *pro praktické použití* postup zcela vyhovuje.

V obdobné situaci s číselnými proměnnými, u nichž předpokládáme lineární vztahy, používáme ke konstrukci modelů kauzálního řetězení regresní vícenásobnou lineární analýzu a parciální koeficienty lineární korelace; v takovém případě můžeme vytvářet sítě s desítkami proměnných. V lineární analýze se omezujeme na lineární vztahy (i po případné transformaci proměnných). U kontingenčních tabulek vyšetřujeme obecný typ závislosti, uvažujeme různé interakční efekty projevující se v působení kombinací znaků, platíme za to však rozměrem úlohy.

Teorie parciálních asociací ve spojení s obecným typem závislého znaku se jeví jako jeden z nejužitečnějších přístupů odhalování příčinných vztahů.

DODATEK

Obecný typ kategorizované proměnné $\mathbf{A} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_S; \mathbf{D}\}$ s kategoriemi $\boldsymbol{A}_k$, $k = 1, 2, \ldots, S$ je určen maticí $\mathbf{D} = \|d_{kj}\|$, pro jejíž prvky platí: 1. $d_{kk} = 0$, 2. $d_{kj} \geqq 0$, 3. $d_{kj} = d_{jk}$, 4. $d_{kj} > d_{pq}$ znamená, že dvojice kategorií $\boldsymbol{A}_k, \boldsymbol{A}_j$ si je méně obsahově podobná než dvojice $\boldsymbol{A}_p, \boldsymbol{A}_q$.

Zobecněná variance pro rozložení relativních četností $\boldsymbol{f} = (f_1, \ldots, f_S)$ je pro proměnnou $\mathbf{A}$ definována jako

$$\text{(14)} \qquad \operatorname{Gvar} \boldsymbol{f} = \boldsymbol{f}'\mathbf{D}\boldsymbol{f} = \sum_{j}^{S} \sum_{k}^{S} f_j f_k d_{jk} .$$

Při značení uvedeném v článku a definici koeficientu explanace podle (4) je jeho *asymptotická (přibližná) výběrová směrodatná chyba* (za předpokladu prostého náhodného výběru nebo nezávislého procesu při generování dat a nenulového skutečného koeficientu explanace) dána vzorcem

$$\text{(15)} \qquad s_{\mathbf{A}|\mathbf{B}} = \left\{ \frac{1}{n\xi^4} \sum_{b=1}^{R} \sum_{a=1}^{S} f_{ba} [v(2d_a^* - \xi) - \xi(2d_{a|b}^* - \operatorname{Gvar} \boldsymbol{f}_{(b)})]^2 \right\}^{\frac{1}{2}} ,$$

kde

$$v = \sum_{r=1}^{R} f_{r+} \operatorname{Gvar} \boldsymbol{f}_{(r)} , \quad \xi = \sum_{i=1}^{S} \sum_{j=1}^{S} f_{+i} f_{+j} d_{ij} ,$$

$$d_{a|b}^* = \sum_{j=1}^{S} f_{j|b} d_{ja} , \quad d_a^* = \sum_{j=1}^{S} f_{+j} d_{aj} ,$$

$$f_{+j} = \sum_{r=1}^{R} f_{rj} , \quad f_{r+} = \sum_{j=1}^{S} f_{rj} .$$

Při značení z článku a definici koeficientu parciální asociace podle (10) je jeho *asymptotická směrodatná chyba* (za předpokladu prostého náhodného výběru nebo nezávislého procesu při generování dat a nenulového skutečného koeficientu parciální asociace) dána vzorcem

$$\text{(16)} \qquad s_{\mathbf{A}|\mathbf{B}_2(\mathbf{B}_1)} = \left\{ \frac{1}{n\xi^4} \sum_{b=1}^{R_1} \sum_{c=1}^{R_2} \sum_{a=1}^{S} f_{bca} [v(2d_{a|b}^* - \operatorname{Gvar} \boldsymbol{f}_{(b)}) - \right.$$
$$\left. - \xi(2d_{a|bc}^* - \operatorname{Gvar} \boldsymbol{f}_{(bc)})]^2 \right\}^{\frac{1}{2}} ,$$

kde

$$v = \sum_{r=1}^{R_1} \sum_{q=1}^{R_2} f_{rq+} \operatorname{Gvar} \boldsymbol{f}_{(rq)} , \qquad \xi = \sum_{r=1}^{R_1} f_{r++} \operatorname{Gvar} \boldsymbol{f}_{(r)} ,$$

$$d_{a|bc}^* = \sum_{j=1}^{S} \frac{f_{bcj}}{f_{bc+}} d_{ja} , \qquad d_{a|b}^* = \sum_{j=1}^{S} \frac{f_{b+j}}{f_{b++}} d_{ja} ,$$

$$f_{bc+} = \sum_{a=1}^{S} f_{bca} , \qquad f_{b+j} = \sum_{c=1}^{R_2} f_{bcj} ,$$

$$f_{b++} = \sum_{c=1}^{R_2} \sum_{a=1}^{S} f_{bca} .$$

Intervaly spolehlivosti (řekněme od velikosti výběru $n \geqq 50$) lze stanovit jako

$$\text{(17)} \qquad \delta_{\mathbf{A}|\mathbf{B}} \pm z_\alpha s_{\mathbf{A}|\mathbf{B}} , \quad \delta_{\mathbf{A}|\mathbf{B}_2(\mathbf{B}_1)} \pm z_\alpha s_{\mathbf{A}|\mathbf{B}_2(\mathbf{B}_1)} ,$$

kde z_α je $(1 - \frac{\alpha}{2})$ 100%ní kvantil standardního normálního rozdělení ($z_{0.95} = 1.96$, $z_{0.99} = 2.56$).

Literatura

Bishop, Y. M. M.—Fienberg, S. E.—Holland, P. W.: *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, The MIT Press 1975.
Goodman, L. A.—Kruskal, W. H.: *Measures of Association for Cross Classifications.* J. Amer. Statist. Assoc. 49, 1954, s. 732—764.
Goodman, L. A.—Kruskal, W. H.: *Measures of Association for Cross Classifications III: Approximate Sampling Theory.* J. Amer. Statist. Assoc. 58, 1963, s. 310—364.
Gray, L. N.—Williams, J. S.: *Goodman and Kruskal's Tau b, Multiple and Partial Analogs.* Sociological Methods and Research 10, 1981, s. 50—62.
Haberman, S. J.: *Analysis of Dispersion of Multinomial Responses.* J. Amer. Statist. Assoc. 77, 1982, s. 568—580.
Havránek, T.: *Třídění třetího a čtvrtého stupně — úlohy a metody.* In: Řehák, J. (red.): Statistická analýza dat a kauzální závěry. Praha, ČSSS při ČSAV 1980, s. 87—121.
Kendall, M. G.: *Rank Correlation Methods.* London, Ch. Griffin 1962.
Matějů, P.: *K využití logaritmicko-lineárních modelů v analýze sociální mobility.* Sociologický časopis XXI, 1985, s. 53—70.
Rao, C. R.: *Lineární metody statistické indukce a jejich aplikace.* Praha, Academia 1978.
Řehák, J.: *Základní deskriptivní míry pro rozložení ordinálních dat.* Sociologický časopis XII., 1976, s. 416—431.
Řehák, J. (red.): Statistická analýza dat a kauzální závěry. Praha, ČSSS při ČSAV 1980.
Řehák, J.—Řeháková, B.: *Měření statistické závislosti nominálních znaků.* Sociologický časopis IX, 1973, s. 404—417.
Řehák, J.—Řeháková, B.: *Míry statistické závislosti pro ordinální znaky.* Sociologický časopis XI, 1975, s. 75—92.
Řehák, J.—Řeháková, B.: *Základní charakteristiky proměnných s konečným počtem hodnot a distanční analýza jejich rozložení.* Sociologický časopis XV, 1979, s. 214—233.
Řehák, J.—Řeháková, B.: *Distanční přístup k analýze kategorizovaných dat a jeho aplikace na problém shody.* In: Antoch, J.—Jurečková, J. (red.) Robust 82, s. 76—80,Jednota československých matematiků a fyziků 1982.
Řehák, J.—Řeháková, B.: *Parciální asociační koeficienty v kontingenčních tabulkách.* In: Antoch, J.—Jurečková, J. (red.) Robust 84, s. 105—108, Jednota československých matematiků a fyziků 1984.
Řehák, J.—Řeháková, B.: *Analýza kategorizovaných dat v sociologii.* Praha, Academia 1986.
Řeháková, B.: *Příčinné závislosti a jejich statistický odraz.* Sociologický časopis XVIII, 1982, s. 532—543.
Řeháková, B.: *Koeficienty parciální asociace pro obecný typ kategorizované proměnné.* Práce k aspirantskému minimu. Praha, MFF UK 1982.
Upton, G. J. G.: *The Analysis of Cross-tabulated Data.* New York, John Wiley 1978.

Резюме

Я. Ржегак - Б. Ржегакова: Множественная и частная ассоциация в таблицах сопряженности

Теория множественной и частной ассоциации разработана для ситуаций, в которых встречается одна зависимая переменная общего типа. Остальные переменные, вступающие в анализ, понимаются как номинальные. Общий тип определяется с помощью соотношений, отражающих степень несходства или удаленности категорий: он включает такие специальные случаи, как обычные номинальные, ординальные и кардинальные переменные, географические переменные, многомерные квантификации категорий, комбинированные дихотомии, категоризации, у которых учитываются отношения соседства, многомерный ординальный случай, к-выборочные переменные и много других полезных типов. Тем самым достигнут высокий ступень всеобщности.

Сначала обобщена модель асимметрической статистической зависимости в двухмерной таблице сопряженности, в которой выступает столбцовая переменная общего типа, а строчная переменная понимается номинально. Затем определяется коэффициент экспланации на основе принципа редукции вариабильности. Модель просто распространена на множественную ассоциацию. Общий подход иллюстрируется на примере комбинированной дихотомии. Затем вводится и объясняется понятие частной ассоциации и выводится модель для ее измерения, включая приведение свойств итогового коэффициента. Примеры из прикладной практики включают иллюстрацию сцепления ассоциаций, определение чистого влияния отдельных переменных, выявление скрытых ассоциаций и порядок составления причинных сетей.

В заключении обобщаются преимущества и недостатки метода, являющегося параллелью к методологии логарифмическо-линейных моделей. В дополнении приводятся основные определения и формулы для асимптотических стандартных ошибок введенных коэффициентов. В таблицах приводятся формулы для коэффициентов экспланации у номинальной, ординальной и кардинальной зависимых переменных (таблица А), суммируются ситуации отношений между тремя переменными с характеристикой нумерических значений коэффициентов (таблица Б) и показывается (таблица Ц), как можно обобщить прямой и простой частный коэффициент ассоциации, применив его к различным случаям многодименсиональных таблиц сопряженности.

Summary

J. Řehák — B. Řeháková: Multiple and Partial Association in Contingency Tables

The theory of multiple and partial association is developed for a situation with one dependent variable, which may be of general type. Other variables (independent and intermediating) are considered nominal. A general type of variable is defined by a matrix of scores that reflects the relations of dissimilarity among categories. It includes such special cases as common nominal, ordinal and cardinal variables, geographical variables, multidimensional quantification of categories, combined dichotomies, i.e.: categories with which we consider relations of neighbourhood, multidimensional ordinal variables, k-choice variables and many other useful types. In this way, the model reflects a high degree of generalization.

First, a model for asymmetric association in two way contingency tables having a column variable of general type and a nominal row variable is considered. The coefficient of explanatory power is derived from the principle of variability reduction. A simple generalization to multiple association is also given. In an example, a comelined dichotomy is used as an illustration of a non-routine variable type. Next, a notion of partial association is explained and a model for its measurement is derived: also properties of the resulting coefficient are listed. Examples from practical applications show the chaining of associations, determining of a clean influence of individual variables on the variability of the dependent one, discovery of hidden associations and a procedure for a causal network construction. Some advantages and disadvantages of the method are listed and the Appendix contains basic definitions and the formulas for asymptotic standard errors of coefficients. The tables contain coefficients of explanatory power for nominal, ordinal and cardinal variables (Table A), various possibilities of relations among three variables with the characterization of the numerical values of coefficients (Table B) and demonstration how the direct and simple partial association can be generalized to various multidimensional cases (Table C).