

## O jednom možném přístupu k analýze sociálních rozdílů (diferenciací)

Součástí sociologického studia sociální struktury je problematika sociálních rozdílů, diferenciací. „Sociální strukturou rozumíme dělení společnosti na skupiny a jejich vzájemné vztahy“ píše M. N. Rutkevič<sup>1</sup>.

V tomto stručném vyjádření je shrnuta společná podstata mnohých přístupů k problematice sociální struktury, které na tomto místě vědomě neuvádíme. Protože cílem našeho výkladu bude kvantitativní (nebo lépe řečeno formální) hledisko analýzy sociálních rozdílů, mající svůj nezastupitelný význam pro empirické sledování veškerých aspektů sociální struktury (ale i pro další oblasti sociologického zkoumání), obrátíme naši pozornost nejprve k otázkám abstraktnější povahy, abychom ji později využili při výkladu „obsahově bohatší“ tematiky.

### 1. Formální vymezení diferenciace

V souladu se sociologickou tradicí budeme chápat diferenciaci jako rozdělení zkoumané populace podle určitého kritéria, a to takové rozdělení, podle něž je každý element zkoumané populace začleněn právě do jedné ze skupin vytvořených rozdělením.

Diferenciace tedy, zjednodušeně řečeno, představuje rozdělení odlišujících se elementů a uskupení těch elementů, které se podle daného kritéria neodlišují.

Formálně je diferenciace určitou ekvivalencí na zkoumané populaci. Označme tedy zkoumanou populaci symbolem  $P$ , její elementy pak symbolem  $p_i$ ,  $i \in I$ , tzn.  $P = \{p_i\}_{i \in I}$ . Vzhledem k tomu, že zkoumáme vždy pouze konečné populace, je rovněž možný zápis

$$P = \{p_1, p_2, \dots, p_r\}.$$

Diferenciací  $D$  na množině  $P$  rozumíme tedy takový soubor podmnožin množiny  $P$ , který označíme  $d_1, \dots, d_k$ , pro nějž platí:

$$d_l \cap d_m = \emptyset \quad \text{pro } l \neq m, l, m = 1, \dots, k \quad \text{a} \quad \bigcup_{l=1}^k d_l = P.$$

Počet všech elementů, které leží v  $d_l$ , tj. počet všech takových  $p_i \in d_l$ ,  $i = 1, \dots, n$ , označíme  $|d_l|$ . Zřejmě platí:

$$\sum_{l=1}^k |d_l| = r.$$

Upozorníme v této souvislosti na jednu podstatnou okolnost: to, že jsme elementy diferenciace — tj. soubor množin  $d_1, \dots, d_k$  — očíslovali, neznamená v žádném případě jejich uspořádání. Jde o zjednodušenou formu zápisu  $\{d_l\}_{l \in L}$ , který by byl jistě přesnější. Avšak pro názornost volíme i v tomto případě diferenciace zjednodušený zápis jako v případě základní množiny, tj. zkoumané populace.

Znovu zopakujeme, že proces diferenciace neznamená nic jiného než přechod od základní množiny k určité množině jejích podmnožin, tedy proces seskupení určitých elementů základní množiny do elementů nové množiny, diferenciace dané množiny.

Proto tedy diferenciace v našem významu znamená vlastně „zúžení“ původní diferenciace dané množiny.

Co tím chceme říci?

Uvažujeme takovou diferenciaci  $D$  na původní množině  $P$ , pro niž platí, že každý element diferenciace obsahuje právě jeden element původní množiny. Jak se snadno přesvědčíme, takto vzniklá diferenciace vede opět k původní množině, což vyjádříme tak, že původní množina představuje nejbohatší diferenciaci na sobě samé. Naopak taková diferenciace, jejímž jediným prvkem je celá množina  $P$ , tzn. v dané diferenciaci neodlišujeme žádné elementy výchozí zkoumané populace, představuje nejchudší diferenciaci

<sup>1</sup> M. N. Rutkevič: Tendenci rozvoje sociální struktury sovětského občestva. Moskva, Mysl 1975, s. 3.

na dané populaci. Nejnížší možné zúžení výchozí množiny vede tedy k nejbohatší diferenciaci, naopak její nejvyšší možné zúžení vede k nejchudší diferenciaci. Tyto dva singulární případy, které ohraničují celkový prostor diferenciací zkoumané populace, nejsou zajímavé, neboť kritéria, která je vymezují, jsou vlastně popisem buď každého elementu dané populace nebo populace jako celku.<sup>2</sup>

Množinu všech diferenciací na množině  $P$  označíme  $D(P)$ . Na takto definované množině je možno zavést celou řadu operací. Tak například je-li  $D_1, D_2 \in D(P)$ , přičemž

$$D_1 = \{^1d_1, \dots, ^1d_{l_1}\}, D_2 = \{^2d_1, \dots, ^2d_{l_2}\}$$

a platí, že pro každé  $^1d_s, s = 1, \dots, l_1$ , existuje  $^2d_{m_s}$  takové, že  $^1d_s \subset ^2d_{m_s}$  (ve smyslu množinové inkluze na množině  $P$ ), pak říkáme, že  $D_1$  je zjemněním  $D_2$  a značíme  $D_1 \subset D_2$ . Platí-li  $D_1 \subset D_2, D_2 \subset D_1$ , značíme  $D_1 = D_2$  a nazýváme takové diferenciaci rovné.

Diferenciace  $D_3$ , jejíž elementy vzniknou jako průniky elementů diferenciací  $D_1$  a  $D_2$ , tj.

$$D_3 = \{^1d_1 \cap ^2d_1, ^1d_1 \cap ^2d_2, \dots, ^1d_{l_1} \cap ^2d_{l_2}, \dots, ^1d_{l_1} \cap ^2d_1, \dots, ^1d_{l_1} \cap ^2d_{l_2}\},$$

značíme jako  $D_1 \cap D_2$  a nazýváme průnikem diferenciací  $D_1$  a  $D_2$ . Zřejmě platí, že  $D_1 \cap D_2 \subset D_1, D_1 \cap D_2 \subset D_2$ .

Ve statistice často analyzujeme průnik dvou diferenciací pomocí tzv. kontingenční tabulky, tj. matice<sup>4</sup>

$$\begin{pmatrix} |^1d_1 \cap ^2d_1|, & |^1d_1 \cap ^2d_2|, & \dots, & |^1d_1 \cap ^2d_{l_2}| \\ \vdots & \vdots & & \vdots \\ |^1d_{l_1} \cap ^2d_1|, & |^1d_{l_1} \cap ^2d_2|, & \dots, & |^1d_{l_1} \cap ^2d_{l_2}| \end{pmatrix}$$

<sup>2</sup> Na tomto místě by mohla být položena závažná otázka, zda lze aplikovat kategorii diferenciace na veškeré objekty socialistické analýzy. Z toho, co bylo řečeno, totiž plyne, jak podstatný je výchozí pojem množiny pro označení zkoumané populace. Fakticky jde o situaci, kdy vycházíme z určité diferenciace (z té nejbohatší), dané chápáním zkoumané populace jako množiny odlišených elementů, kterou pouze „ochuzujeme“. Jde tedy ve skutečnosti o určitý druh tautologie: při vymezení diferenciace fakticky předpokládáme určitý typ – navíc nejbohatší – diferenciace. Tento druh „tautologie“ však není nicím neznámým: jeho diskurs bychom však přetvořili rámcem našeho rozboru, který si neklade za cíl analyzovat zásadní ontologické a gnoseologické otázky v tomto směru.

<sup>3</sup> Je jisté, že  $D_1$  je opět diferenciací. Pochopitelně se může stát, že některé z diferenciálních elementů takto vzniklé diferenciace budou prázdné množiny, ty však můžeme vyškrtnout a fakticky je neuvažovat. Přirozeně bychom se tomuto nedostatku mohli při definici průniku dvou diferenciací vy-

Pro další výklad bude velmi užitečné si uvědomit, že každá diferenciace množiny  $P$  generuje diferenciace každého z elementů jiné diferenciace množiny  $P$ . Je-li  $^1d_1$  jeden z elementů diferenciace  $D_1$ , pak soubor podmnožin  $\{^1d_1 \cap ^2d_1, ^1d_1 \cap ^2d_2, \dots, ^1d_1 \cap ^2d_{l_2}\}$  představuje jeho diferenciace, generovanou diferenciací  $D_2$ .

Průnik dvou diferenciací  $D_1 \cap D_2$  lze tedy analyzovat také tak, že představuje vlastně soubor  $l_1$  diferenciací množin  $^2d_1, ^2d_2, \dots, ^2d_{l_2}$  či  $l_2$  diferenciací množin  $^1d_1, ^1d_2, \dots, ^1d_{l_1}$ . Vzniká tedy přirozeně otázka, nakolik diferencovaně působí diferenciace na celé množině a na jednotlivých elementech jiné diferenciace, jejíž zodpovězení by mohlo poskytnout cennou informaci o vztahu daných diferenciací. Tím se však dostáváme k další části výkladu.

## 2. Měření diferenciací a jejich vztahů

Nejprve dokážeme pomocné tvrzení.

*Lemma 1:* Nechť  $A$  je množina všech nezáporných  $n$ -členných posloupností takových, že součet jejich členů je roven 1. Označme dva obecné prvky této množiny

$$a = \{a_1, \dots, a_n\}, b = \{b_1, \dots, b_n\}$$

tzn.  $a_i \geq 0, b_i \geq 0, i = 1, \dots, n$

$$a \quad \sum_{i=1}^n a_i = \sum_{i=1}^n b_i = 1$$

Potom funkce definovaná vztahem

$$\varrho(a, b) = 1 - \sum_{i=1}^n \min(a_i, b_i)$$

je metrikou na  $A$ <sup>5</sup>, jinými slovy  $A$  je metrickým prostorem s metrikou  $\varrho$ .

hnout tím, že bychom podmínku neprázdnoti nově vzniklých diferenciálních elementů vytýčili jako definiční podmínku.

<sup>4</sup> Přitom připouštíme, že některá z čísel mohou být rovna nule – viz předchozí poznámku. Avšak platí, že v každém řádku i v každém sloupci musí být za předpokladu kladnosti všech hodnot

$$|^1d_s|, |^2d_m|,$$

$s = 1, \dots, l_1, m = 1, \dots, l_2$  vždy alespoň jedno z čísel nenulové.

<sup>5</sup> Metrikou rozumíme takovou reálnou funkci, pro níž platí

- a)  $\varrho(a, b) = 0 \Leftrightarrow a = b$
- b)  $\varrho(a, b) = \varrho(b, a)$
- c)  $\varrho(a, b) + \varrho(b, c) \geq \varrho(a, c)$

dále poznamenejme, že  $a = b$  znamená  $a_i = b_i$  pro  $i = 1, \dots, n$ .

Dokažme nejprve, že  $\varrho(a, b) = 0 \Leftrightarrow a = b$ .  
Vzhledem k tomu, že

$$1 - \sum_{i=1}^n \min(a_i, b_i) = 0 \Rightarrow \sum_{i=1}^n \min(a_i, b_i) = 1,$$

a dále, že

$$\sum_{i=1}^n \min(a_i, b_i) \leq \sum_{i=1}^n a_i = 1$$

a

$$\sum_{i=1}^n \min(a_i, b_i) \leq \sum_{i=1}^n b_i = 1,$$

musí zřejmě

$\min(a_i, b_i) = a_i$  a  $\min(a_i, b_i) = b_i$ ,  
tedy platí  $a_i = b_i$  pro  $i = 1, \dots, n$ , tedy  
 $a = b$ . V opačném případě je důkaz evi-  
dentní.

Vztah  $\varrho(a, b) = \varrho(b, a)$  plyne přímo z de-  
finice metriky. Nyní přejdeme k důkazu  
tzv. trojúhelníkové nerovnosti. Nechť  
 $a = \{a_1, \dots, a_n\}$ ,  $b = \{b_1, \dots, b_n\}$ ,  $c =$   
 $= \{c_1, \dots, c_n\}$  splňují podmínky dané větou,  
tj. jsou elementy  $A$ . Dokažme nejprve, že  
platí

$$\begin{aligned} \sum_{i=1}^n (\min(a_i, b_i) + \min(b_i, c_i) - \\ - \min(a_i, c_i)) \leq \sum_{i=1}^n b_i. \end{aligned}$$

Označme symboly  $J_1, \dots, J_6$  množiny přiro-  
zených čísel od 1 do  $n$ , pro něž platí:

$$J_1: a_i \leq b_i \leq c_i$$

$$J_2: a_i \leq c_i \leq b_i$$

$$J_3: b_i \leq a_i \leq c_i$$

$$J_4: b_i \leq c_i \leq a_i$$

$$J_5: c_i \leq a_i \leq b_i$$

$$J_6: c_i \leq b_i \leq a_i \quad 6$$

Pro  $i \in J_1$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = a_i + b_i - a_i = b_i. \end{aligned}$$

Pro  $i \in J_2$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = a_i + c_i - a_i = c_i \leq b_i \end{aligned}$$

(poslední nerovnost plyne z toho, že pro  
 $i \in J_2$  platí  $c_i \leq b_i$ ).

Pro  $i \in J_3$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = b_i + b_i - a_i = 2b_i - a_i \leq b_i \end{aligned}$$

(poslední nerovnost plyne z toho, že pro  
 $i \in J_3$  platí  $b_i \leq a_i$  neboli  $2b_i \leq a_i + b_i$ ).

Pro  $i \in J_4$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = b_i + b_i - c_i = 2b_i - c_i \leq b_i \end{aligned}$$

(poslední nerovnost plyne z toho, že pro  
 $i \in J_4$  platí  $b_i \leq c_i$  neboli  $2b_i \leq c_i + b_i$ ).

Pro  $i \in J_5$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = a_i + c_i - c_i = a_i \leq b_i \end{aligned}$$

(poslední nerovnost plyne z toho, že pro  
 $i \in J_5$  platí  $a_i \leq b_i$ ).

Pro  $i \in J_6$  platí:

$$\begin{aligned} \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) = \\ = b_i + c_i - c_i = b_i. \end{aligned}$$

Dokázali jsme, že pro libovolné  $i$  platí  
 $\min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) \leq b_i$   
a tedy

$$\begin{aligned} \sum_{i=1}^n \min(a_i, b_i) + \min(b_i, c_i) - \min(a_i, c_i) \leq \\ \leq \sum_{i=1}^n b_i = 1 \end{aligned}$$

Úpravou poslední nerovnosti (přičtením 2  
k oběma stranám nerovnosti a jednoduchým  
převodem jednotlivých členů nerovnosti)  
dostáváme:

$$\begin{aligned} 1 - \sum_{i=1}^n \min(a_i, b_i) + 1 - \sum_{i=1}^n \min(b_i, c_i) \geq \\ \geq 1 - \sum_{i=1}^n \min(a_i, c_i), \end{aligned}$$

čímž je důkaz proveden.

Dokázané lemma tedy umožňuje studovat  
prostor  $n$ -členných posloupností nezáporných  
čísel, jejichž součet je roven jedné, jako tzv.  
metrický prostor. Užitečné pro další úvahy  
bude ukázat, jak lze interpretovat metriku  
zavedenou lemmatem 1. Nechť  $a \in A$ . Nechť  
je dána další  $n$ -členná posloupnost reálných  
čísel

$$x = \{x_1, \dots, x_n\} \text{ taková, že } \sum_{i=1}^n x_i = 0$$

a  $a_i + x_i \geq 0$  pro  $i = 1, \dots, n$ .

Potom

$$b = a + x = \{a_1 + x_1, \dots, a_n + x_n\} \in A.$$

Důkaz tohoto tvrzení je evidentní.

<sup>6</sup> Pochopitelně, že rozklad indexů, který jsme  
uvodili, nemusí být disjunktní, tzn. že například  
platí-li  $a_i = b_i = c_i$ , potom  $i \in J_1$  i  $i \in J_2$ . Stejně  
tak některé z množin indexů mohou být prázdné,

což však není pro náš důkaz podstatné. Podstatné  
je, že každý index leží alespoň v jedné z uvedených  
množin indexů.

Naopak dokážeme, že je-li dána jedna  $a \in A$ , potom libovolný element  $b \in A$  lze vyjádřit jako  $a + x$ , kde  $x$  je  $n$ -členná posloupnost reálných čísel s výše uvedenými vlastnostmi. Platí totiž, že  $b_i = a_i + b_i - a_i$ ,  $i = 1, \dots, n$ ,

označme  $x_i = b_i - a_i$ . Zřejmě  $\sum_{i=1}^n x_i = 0$  a

$$a_i + x_i = b_i \geq 0. \text{ Proto lze psát } \varrho(a, b) = 1 - \sum_{i=1}^n \min(a_i, b_i) = 1 - \sum_{i=1}^n \min(a_i, a_i + x_i),$$

kde  $\sum_{i=1}^n x_i = 0$ ,  $a_i + x_i \geq 0$ . Označme  $J_1$

množinu všech indexů takových, že  $x_i \geq 0$ ,  $J_2$  pak množinu všech takových, že  $x_i < 0$ . Zřejmě  $J_1 \cap J_2 = \emptyset$  a  $J_1 \cup J_2 = \{1, \dots, n\}$ . Lze tedy psát:

$$\begin{aligned} 1 - \sum_{i=1}^n \min(a_i, a_i + x_i) &= \\ = 1 - \sum_{i \in J_1} a_i - \sum_{i \in J_2} (a_i + x_i) &= \\ = 1 - \sum_{i=1}^n a_i - \sum_{i \in J_2} x_i = - \sum_{i \in J_2} x_i. \end{aligned}$$

Vzhledem k tomu, že  $\sum_{i=1}^n x_i = 0$ , platí, že

$$\begin{aligned} \sum_{i \in J_1} x_i &= - \sum_{i \in J_2} x_i, \text{ tzn. lze psát} \\ \varrho(a, b) &= \sum_{i \in J_1} x_i = - \sum_{i \in J_2} x_i. \end{aligned}$$

$$\begin{aligned} \text{Zřejmě však } x_i &= b_i - a_i \text{ (viz výše), tudíž} \\ \varrho(a, b) &= \sum_{i \in J_1} (b_i - a_i) = - \sum_{i \in J_2} (b_i - a_i) = \\ &= \sum_{i: b_i \geq a_i} (b_i - a_i) = - \sum_{i: b_i < a_i} (b_i - a_i) = \\ &= \sum_{i: b_i < a_i} (a_i - b_i). \end{aligned}$$

(Z posledních vztahů mimo jiné vyplývá, že

$$\varrho(a, b) = \frac{1}{2} \sum_{i=1}^n |a_i - b_i|;$$

toto vyjádření metriky je „obvyklejší“ a představuje metriku na širším prostoru, než který jsme uvažovali, tj. na množině všech  $n$ -členných posloupností reálných čísel. V této formě bývá také tato metrika někdy užívána i v analýzách sociálních dat.) Metrika  $\varrho(a, b)$  znamená, o kolik celkem musíme zmenšit (či zvětšit) jednotlivé elementy  $a_i$ , pro něž  $a_i > b_i$  ( $a_i < b_i$ ), abychom „změnili  $a$  na  $b$ “ (pochopitelně oč zmenšíme či zvětšíme příslušné elementy  $a_i$ , o to musíme zbylé elementy  $a_i$  zvětšit či zmenšit).

Nechť například

$$a = (\frac{1}{2}, \frac{1}{2}), \quad b = (\frac{1}{4}, \frac{3}{4}).$$

Potom

$$\varrho(a, b) = 1 - (\frac{1}{4} + \frac{1}{2}) = \frac{1}{4}.$$

Abychom „přešli“ od  $a$  k  $b$ , musíme od  $a_1$  odečíst  $\frac{1}{4}$ , kterou pochopitelně přičteme k  $a_2$ :

$$(\frac{1}{2} - \frac{1}{4}, \frac{1}{2} + \frac{1}{4}) = (\frac{1}{4}, \frac{3}{4}).$$

Zde je význam uvedené metriky zcela patrný.

Je zřejmé, že nejvyšší hodnoty, které může  $\varrho(a, b)$  dosáhnout, je 1. Je-li například  $a = (1, 0)$ ,  $b = (0, 1)$ , potom  $\varrho(a, b) = 1$ . Zřejmě lze veškeré dosavadní úvahy provést analogicky pro prostor  $n$ -členných nezáporných reálných posloupností takových, že součet jejich členů je roven kladnému číslu  $s$ . V takovém případě má příslušná metrika tvar

$$\begin{aligned} \varrho(a, b) &= s - \sum_{i=1}^n \min(a_i, b_i) = \\ &= s \left[ 1 - \sum_{i=1}^n \min\left(\frac{a_i}{s}, \frac{b_i}{s}\right) \right]. \end{aligned}$$

Zpravidla budeme dané metriky používat v těch případech, kdy  $n$ -členné posloupnosti představují distribuci (procentuální zastoupení výskytu jednotlivých hodnot) určitého znaku na zkoumané populaci. V takových případech pak bude mít daná metrika tvar

$$\varrho(a, b) = 100 - \sum_{i=1}^n \min(a_i, b_i).$$

Nechť  $D_1, D_2$  jsou dvě diferenciace populace  $P$  takové, že mají stejný počet elementů a navíc jsou tyto elementy uspořádány, nebo alespoň existuje jistá korespondence (jednoznačné zobrazení množiny elementů  $D_1$  na množinu elementů  $D_2$ ) dovolující je uspořádat, tzn. lze psát

$$D_1 = \{d_1^1, \dots, d_n^1\}, \quad D_2 = \{d_1^2, \dots, d_n^2\}.$$

Potom veličinou

$$1 - \sum_{i=1}^n \min\left(\frac{|d_i^1|}{r}, \frac{|d_i^2|}{r}\right),$$

resp.

$$100 - \sum_{i=1}^n \min\left(100 \cdot \frac{|d_i^1|}{r}, 100 \cdot \frac{|d_i^2|}{r}\right)$$

lze ohodnotit vztah těchto dvou diferenciací, tj. jejich vzdálenost, kterou značíme

$$\varrho(D_1(P), D_2(P)).$$

Odtud je tedy zřejmé, že danou vzdálenost lze užít k měření vztahu dvou diferenciací na téže populaci pouze za určitých doplňujících podmínek.

Poněkud příznivější situace je v případě, kdy chceme srovnat působení „téže“ diferenciace na dvou populacích. Pod pojmem „táž“ diferenciace rozumíme totožnost obsahových kritérií (vlastně jistých predikátů, majících smysl na obou zkoumaných populacích).

Označme dané populace  $P_1$  a  $P_2$ , počty jejich elementů  $r_1$  a  $r_2$ . Necht  $D(P_1) = \{^1d_1, \dots, ^1d_n\}$ ,  $D(P_2) = \{^2d_1, \dots, ^2d_n\}$ .<sup>7</sup>

Potom veličina

$$\varrho(D(P_1), D(P_2)) = \\ = 1 - \sum_{i=1}^n \min \left( \frac{|^1d_i|}{r_1}, \frac{|^2d_i|}{r_2} \right),$$

případně v procentuálním vyjádření

$$\varrho(D(P_1), D(P_2)) = \\ = 100 - \sum_{i=1}^n \min \left( 100 \cdot \frac{|^1d_i|}{r_1}, 100 \cdot \frac{|^2d_i|}{r_2} \right),$$

měří „rozdíllost diferenciace  $D$ “ na populacích  $P_1$  a  $P_2$ .

Odtud okamžitě plyne, že uvedená metrika poskytuje velmi jednoduchý, přitom však smysluplně interpretovatelný prostředek komparace téže diferenciace na jistém souhrnu zkoumaných populací.

Avšak navíc ukazuje tato skutečnost cestu, jak užít dosavadních výsledků k vytvoření velmi přirozené míry závislosti dvou diferenciací na téže populaci (bez přidavných podmínek, jak tomu bylo výše), o níž jsme se zmínili na konci úvodní části. Než však přistoupíme k výkladu tohoto postupu, dokážeme ještě jedno pomocné tvrzení.

**Lemma 2:** Necht je dána uspořádaná posloupnost  $m \cdot n$  nezáporných čísel, resp. matice

$$\begin{pmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} \end{pmatrix} \text{ taková, že}$$

součet všech jejích elementů je roven jedné,

$$\text{tj.} \quad \sum_{i=1}^m \sum_{j=1}^n a_{ij} = 1.$$

Dále označme symbolem

$$a_{i.} = \sum_{j=1}^n a_{ij}, \quad a_{.j} = \sum_{i=1}^m a_{ij}$$

a předpokládejme, že

$$a_{i.} \neq 0, \quad a_{.j} \neq 0 \quad \text{pro } i = 1, \dots, m, \\ j = 1, \dots, n.$$

Potom platí:

$$\sum_{i=1}^m a_{i.} \left[ 1 - \sum_{j=1}^n \min \left( a_{.j}, \frac{a_{ij}}{a_{i.}} \right) \right] = \\ = \sum_{j=1}^n a_{.j} \left[ 1 - \sum_{i=1}^m \min \left( a_{i.}, \frac{a_{ij}}{a_{.j}} \right) \right] = \\ = 1 - \sum_{i=1}^m \sum_{j=1}^n \min(a_{ij}, a_{i.} \cdot a_{.j}).$$

*Důkaz:*

Zřejmě platí:

$$\sum_{i=1}^m a_{i.} \left[ 1 - \sum_{j=1}^n \min \left( a_{.j}, \frac{a_{ij}}{a_{i.}} \right) \right] = \\ = \sum_{i=1}^m \left[ a_{i.} - \sum_{j=1}^n a_{i.} \cdot \min \left( a_{.j}, \frac{a_{ij}}{a_{i.}} \right) \right] = \\ = \sum_{i=1}^m a_{i.} - \sum_{i=1}^m \sum_{j=1}^n \min(a_{i.} \cdot a_{.j}, a_{ij}) = \\ = 1 - \sum_{i=1}^m \sum_{j=1}^n \min(a_{ij}, a_{i.} \cdot a_{.j}).$$

(Využili jsme podmínky lemmatu, tj. že

$$\sum_{i=1}^m a_{i.} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} = 1).$$

Analogicky bychom dokázali, že

$$\sum_{j=1}^n a_{.j} \left[ 1 - \sum_{i=1}^m \min \left( a_{i.}, \frac{a_{ij}}{a_{.j}} \right) \right] = \\ = 1 - \sum_{i=1}^m \sum_{j=1}^n \min(a_{ij}, a_{i.} \cdot a_{.j}).$$

Tím je důkaz pomocného tvrzení proveden.

Z uvedeného lemmatu přímo plyne postup pro zavedení možné míry vztahu dvou diferenciací, které oproti předchozímu nyní jednodušeji označíme jako

<sup>7</sup> V tomto případě jsou zaručeny — vzhledem k obsahové totožnosti kritérií dané diferenciace — jak shodný počet elementů  $D(P_1)$  a  $D(P_2)$  tak

i jejich vzájemná jednoznačná korespondence, kterou vyjadřujeme příslušným očíslováním jednotlivých elementů  $D(P_1)$  a  $D(P_2)$ .

$D_1 = \{^1d_1, \dots, ^1d_m\}$ ,  $D_2 = \{^2d_1, \dots, ^2d_n\}$ .  
Symbolem  $D_1 : ^2d_k$  označme diferenciace  $^2d_k$   
(tedy elementu diferenciace  $D_2$ ,  $k = 1, \dots$   
 $\dots, n$ ), která je vytvořena jako souhrn pod-  
množin  $\{^1d_1 \cap ^2d_k, ^1d_2 \cap ^2d_k, \dots, ^1d_m \cap ^2d_k\}$ ;  
analogicky pak definujeme

$$D_2 : ^1d_l = \\ = \{^1d_l \cap ^2d_1, ^1d_l \cap ^2d_2, \dots, ^1d_l \cap ^2d_n\}, \\ l = 1, \dots, m.$$

Průnik dvou diferenciací tedy můžeme chápat také tak, že je dána soustava diferenciací jedné diferenciace na elementech druhé diferenciace a naopak: jde buď o soustavu  $\{D_1 : ^2d_k\}$ ,  $k = 1, \dots, n$  nebo o soustavu  $\{D_2 : ^1d_l\}$ ,  $l = 1, \dots, m$ . Vzniká zcela přirozená otázka, jak diferencovaně působí jedna diferenciace na elementech druhé diferenciace, „jak diferencuje jedna diferenciace druhou diferenciací“.

Jde o to srovnat „charakter“ diferenciace  $D_1$  a soustavy diferenciací  $D_1 : ^2d_k$ ,  $k = 1, \dots, n$ , či diferenciace  $D_2$ , a soustavy diferenciací  $D_2 : ^1d_l$ ,  $l = 1, \dots, m$ .

Diferenciace  $D_1$  a každá ze soustavy diferenciací  $D_1 : ^2d_k$  splňují podmínku našich předchozích úvah, týkajících se aplikace výše uvedené metriky na komparaci působení „těže“ diferenciace na různých populacích, neboť  $D_1$  a  $D_1 : ^2d_k$  jsou v tomto smyslu „týž diferenciace“.

Jejich vzdálenost lze měřit veličinou

$$\varrho(D_1, D_1 : ^2d_k) = \\ = 1 - \sum_{i=1}^m \min \left( \frac{|^1d_i|}{r}, \frac{|^1d_i \cap ^2d_k|}{|^2d_k|} \right),$$

kde  $r$  značí počet prvků výchozí zkoumané populace. Tato metrika měří vlastně „diferenci“ v diferenciacích  $D_1$  a  $D_1 : ^2d_k$ .

Je přirozené, že za celou „diferenci“ diferenciací  $D_1$  a  $D_2$  budeme považovat součet „diferencí“ diferenciací  $D_1$  a  $D_1 : ^2d_k$ , vážený hodnotami  $\frac{|^2d_k|}{r}$ ,  $k = 1, \dots, n$ , či součet „diferencí“ diferenciací  $D_2$  a  $D_2 : ^1d_l$ , váže-

ný hodnotami  $\frac{|^1d_l|}{r}$ ,  $l = 1, \dots, m$ , tedy veličinami

$$\sum_{k=1}^n \frac{|^2d_k|}{r} \cdot \varrho(D_1, D_1 : ^2d_k) = \\ = \sum_{k=1}^n \frac{|^2d_k|}{r} \cdot \left[ 1 - \sum_{i=1}^m \min \left( \frac{|^1d_i|}{r}, \frac{|^1d_i \cap ^2d_k|}{|^2d_k|} \right) \right],$$

$$\text{či} \quad \sum_{l=1}^m \frac{|^1d_l|}{r} \cdot \varrho(D_2, D_2 : ^1d_l) = \\ = \sum_{l=1}^m \frac{|^1d_l|}{r} \cdot \left[ 1 - \sum_{i=1}^n \min \left( \frac{|^2d_i|}{r}, \frac{|^1d_l \cap ^2d_i|}{|^1d_l|} \right) \right].$$

Poslední lemma ukazuje, že oba výrazy jsou si rovné a lze je vyjádřit ve tvaru

$$\Delta(D_1, D_2) = \\ = 1 - \sum_{l=1}^m \sum_{k=1}^n \min \left( \frac{|^1d_l \cap ^2d_k|}{r}, \frac{|^1d_l| \cdot |^2d_k|}{r^2} \right)^8.$$

Platí-li

$$\frac{|^1d_l \cap ^2d_k|}{r} = \frac{|^1d_l| \cdot |^2d_k|}{r^2}$$

$$\text{pro } l = 1, \dots, m, \\ k = 1, \dots, n,$$

pak zřejmě  $\Delta(D_1, D_2) = 0$  (jde vlastně o klasický případ statistické nezávislosti dvou jevů).<sup>9</sup>

$\Delta(D_1, D_2)$  tedy měří vzdálenost diferenciace  $D_1 \cap D_2$  od jisté obecně hypotetické diferenciace (ve smyslu poznámky 8 diferenciace  $D_1 \cdot D_2$ ), pro jejíž elementy platí

$$\widetilde{d}_{ik} = \frac{|^1d_i| \cdot |^2d_k|}{r}$$

<sup>8</sup> Předpokládáme-li, že existuje taková diferenciace výchozí populace  $P$ , jejíž elementy jsou uspořádány tak, že ji lze zapsat ve tvaru

$$\widetilde{D}_{1,2} = \{\widetilde{d}_{1,1}, \dots, \widetilde{d}_{1,m}, \widetilde{d}_{2,1}, \dots, \widetilde{d}_{2,n}\}$$

(existuje tedy korespondence mezi elementy diferenciace  $D_1 \cap D_2$  a  $\widetilde{D}_{1,2}$ ), a platí-li

$$\frac{|\widetilde{d}_{ik}|}{r} = \frac{|^1d_i| \cdot |^2d_k|}{r^2},$$

pak takovou diferenciaci označíme  $D_1 \cdot D_2$ . Zřejmě pak platí  $\Delta(D_1, D_2) = \varrho(D_1 \cap D_2, D_1 \cdot D_2)$ , neboli vztah dvou diferenciací  $D_1$  a  $D_2$  měříme vzdáleností průniku a součinu těchto dvou diferenciací.

<sup>9</sup> Ve smyslu předchozí poznámky lze v takovém případě psát  $D_1 \cap D_2 = D_1 \cdot D_2$ .

(označení opět ve smyslu poznámky 8). Jeho výhodou je, že jde o veličinu smysluplně interpretovatelnou: udává vlastně minimální nezbytně nutný počet změn (pohybů) v diferenciaci  $D_1 \cap D_2$ , aby se „změnila“ v onu hypotetickou diferenciaci charakterizovanou výše uvedenou podmínkou (tj. v diferenciaci  $D_1 \cdot D_2$ ).

Ve smyslu výše uvedených úvah lze říci, že  $\Delta(D_1, D_2)$  je vzdáleností  $D_1 \cap D_2$  od „jejich hypotetické nezávislosti“.<sup>10</sup> Okamžitě se objevuje otázka, jakých maximálních hodnot může  $\Delta(D_1, D_2)$  nabýt, tzn. pro jaké  $D_1$  a  $D_2$  je  $D_1 \cap D_2$  nejvíce vzdáleno od „hypotetické diferenciace jejich nezávislosti“. Je zřejmé, že nemůže převýšit hodnotu 1 (plyne to přímo z definice metriky  $\rho$  v lemmatu 1). Je však jasné, že za předpokladu daných  $|d_l|$ ,  $|d_k|$ ,  $l = 1, \dots, m$ ,  $k = 1, \dots, n$ , které jsou vesměs nenulové, nemůže  $\Delta(D_1, D_2)$  nabýt hodnoty 1. Abychom určili tuto maximální hodnotu, uvedeme ještě další lemma.

**Lemma 3:** Necht' je dána matice  $A$  nezáporných čísel

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

taková, že

$$\sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} a_{ij} = 1.$$

Dále označme

$$a_i = \sum_{j=1}^n a_{ij}, \quad a_j = \sum_{i=1}^m a_{ij}$$

a předpokládejme, že

$$a_i \neq 0, \quad a_j \neq 0 \quad \text{pro} \quad i = 1, \dots, m, \\ j = 1, \dots, n.$$

Označme dále symbolem  $L(A)$  množinu všech nezáporných matic

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$$

takových, že

$$\sum_{j=1}^n x_{ij} = a_i, \quad \sum_{i=1}^m x_{ij} = a_j;$$

obecný prvek  $L(A)$  značíme  $X$ . Dále označme symbolem  $A^*$  matici

$$\begin{pmatrix} a_{1.} \cdot a_{.1} & a_{1.} \cdot a_{.2} & \dots & a_{1.} \cdot a_{.n} \\ a_{2.} \cdot a_{.1} & a_{2.} \cdot a_{.2} & \dots & a_{2.} \cdot a_{.n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m.} \cdot a_{.1} & a_{m.} \cdot a_{.2} & \dots & a_{m.} \cdot a_{.n} \end{pmatrix}$$

a symbolem  $S$  množinu všech matic reálných čísel

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \dots & C_{mn} \end{pmatrix}$$

takových, že

$$C_{kn} = - \sum_{j=1}^{n-1} C_{kj}, \quad k = 1, \dots, m-1$$

$$C_{ml} = - \sum_{i=1}^{m-1} C_{il}, \quad l = 1, \dots, n-1$$

$$C_{mn} = \sum_{\substack{i=1, \dots, m-1 \\ j=1, \dots, n-1}} C_{ij},$$

přičemž  $C_{ij} \geq -a_i \cdot a_j$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . Obecný prvek  $S$  budeme značit  $C$ .

Necht' dále  $\tilde{C} \in S$  je takový, že

$$\sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} |\tilde{C}_{ij}| = \max.$$

Pro jeho výpočet je možné užít i druhý způsob jeho vyjádření, který plyne z výše uvedených úvah. Platí totiž, že

$$\Delta(D_1, D_2) = \frac{1}{2} \sum_{l=1}^m \sum_{k=1}^n \left| \frac{|d_l \cap d_k|}{r} - \frac{|d_l| \cdot |d_k|}{r^2} \right|$$

(zde poznamenejme, že symboly  $|\dots|$  značí jednak absolutní hodnotu v běžném smyslu [(je-li výraz, na nějž je symbol aplikován, číslem), jednak počet prvků [(je-li výraz, na nějž je symbol aplikován, označením množiny)]).

<sup>10</sup> Index  $\Delta(D_1, D_2)$  nabývá své nulové hodnoty v téže situaci jako celá řada dalších statisticky tradičních charakteristik vztahu dvou diferenciací (korelační koeficient, index kontingence atd.). Jeho výhoda spočívá nejen v tom, že je definován pro „nejobecnější“ typ sociologických proměnných (tzv. kvalitativních znaků), ale především v tom, že je velmi přirozeně interpretovatelný: hodnota, jíž nabývá, udává minimálně nutný počet změn v diferenciaci  $D_1 \cap D_2$ , aby se  $D_1 \cap D_2$  „změnila“ v  $D_1 \cdot D_2$ . Technická přednost pak spočívá v jeho snadné vyčíslitelnosti.

Potom platí, že

$$1 - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \min(a_{ij}, a_i \cdot a_j) \leq \\ = \max_{X \in L(A)} (1 - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \min(X_{ij}, a_i \cdot a_j)) = \\ = 1 - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \min(a_i \cdot a_j + \tilde{C}_{ij}, a_i \cdot a_j).$$

**Důkaz:** Omezíme se pouze na základní body důkazu. Předně platí, že  $A^* \in L(A)$ . Uvědomme si, že jak  $A$ , tak  $A^*$  představují dvě řešení soustavy  $m + n$  rovnic

$$\sum_{j=1}^n y_{ij} = a_i, \quad \sum_{i=1}^m y_{ij} = a_j, \\ i = 1, \dots, m, \quad j = 1, \dots, n.$$

$L(A)$  pak představuje soustavu všech nezáporných řešení daných rovnic (hodnota této soustavy rovnic je rovna  $(\min(m, n) - 1)$ ). Snadno si ověříme, že soustava  $\bar{S}$ , která vznikne rozšířením soustavy  $S$  pominutím podmínky  $C_{ij} \geq -a_i \cdot a_j$  představuje soustavu všech řešení rovnic:

$$\sum_{j=1}^n y_{ij} = 0, \quad \sum_{i=1}^m y_{ij} = 0, \\ i = 1, \dots, m, \quad j = 1, \dots, n,$$

která je příslušnou homogenní soustavou výše uvedené soustavy rovnic. Na základě elementární věty řešení lineárních rovnic, totiž že obecné řešení nehomogenní soustavy lineárních rovnic je rovno součtu obecného řešení příslušné homogenní soustavy, tj. v našem případě  $\bar{S}$ , a jednoho partikulárního řešení nehomogenní soustavy, na něž zvolíme  $A^*$ , lze obecné řešení výše uvedené soustavy psát ve tvaru  $\tilde{C} + A^*$ , kde  $\tilde{C} \in \bar{S}$ . Podmínky  $C_{ij} \geq -a_i \cdot a_j$  vybírají z  $\bar{S}$  pouze nezáporná řešení. Platí tedy, že  $L(A) = A^* + C$ , kde  $C \in S$ . Lze tedy psát

$$1 - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \min(x_{ij}, a_i \cdot a_j) = \\ = 1 - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \min(a_i \cdot a_j + C_{ij}, a_i \cdot a_j) =$$

$$= \frac{1}{2} \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} |a_i \cdot a_j + C_{ij} - a_i \cdot a_j| = \\ = \frac{1}{2} \sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} |C_{ij}|, \quad \text{kde } C \in S.$$

Je-li  $\tilde{C} \in S$  takové, že

$$\sum_{\substack{i=1, \dots, m \\ j=1, \dots, n}} |\tilde{C}_{ij}| = \max,$$

tvrzení lemmatu je zřejmé.

Nalezení  $\tilde{C} \in S$  není věc obecně jednoduchá: je nutné ji řešit postupy matematického programování. Poznamenejme konečně, že toto řešení nemusí být jediné. Diskuse těchto otázek by však překročila rámec našich úvah. Z lemmatu je okamžitě zřejmé, že maximální hodnota  $\Delta(D_1, D_2)$  je rovna

$$\frac{1}{2} \sum_{\substack{i=1, \dots, m \\ h=1, \dots, n}} |q_{ik}|,$$

kde

$$q_{in} = - \sum_{j=1}^{n-1} q_{ij}, \quad i = 1, \dots, m-1$$

$$q_{mj} = - \sum_{i=1}^{m-1} q_{ij}, \quad j = 1, \dots, n-1$$

$$q_{mn} = \sum_{\substack{i=1, \dots, m-1 \\ j=1, \dots, n-1}} q_{ij},$$

přičemž

$$q_{ik} \geq \frac{-|d_i| \cdot |d_k|}{r^2}$$

$$\text{a} \quad \sum_{\substack{l=1, \dots, m \\ k=1, \dots, n}} |q_{lk}|$$

je maximální.

Označme tuto hodnotu  $\max \Delta(D_1, D_2)$ .

Definujeme dále vztahem

$$\delta(D_1, D_2) = \frac{\Delta(D_1, D_2)}{\max \Delta(D_1, D_2)}$$

tzv. index vztahu diferenciací  $D_1$  a  $D_2$ .

Tento index má následující vlastnosti:

- je definován pro kvalitativní znaky;
- nabývá hodnot mezi nulou a jednou, přičemž je roven nule, jsou-li  $D_1$  a  $D_2$  nezávislé (ve statistickém smyslu) a roven jedné, jsou-li  $D_1$  a  $D_2$  maximálně závislé (ve smyslu maximální vzdálenosti od případu statistické nezávislosti);<sup>11</sup>

<sup>11</sup> V uvedeném smyslu je možno fakticky užít uvedeného postupu pro přirozenou definici minimálně statistické závislosti daných dvou diferen-

ciací  $D_1$  a  $D_2$ . Tyto úvahy by však opět překročily stanovený rámec našich úvah.

– je smysluplně interpretovatelný: jeho hodnota označuje minimálně nezbytný počet změn v diferenciaci  $D_1 \cap D_2$ , jimiž „přejde“ tato diferenciace v hypotetickou diferenciaci  $D_1 \cdot D_2$  (charakterizující statistickou nezávislost diferenciací  $D_1$  a  $D_2$ ), relativizovanou maximální hodnotou těchto změn;

– jeho jistou nevýhodou je, že pro svůj výpočet obecně vyžaduje poměrně náročné prostředky matematického programování.

Uvědomme si konečně, že v určitých případech bude možné užít pouze indexu difference diferenciací  $D_1$  a  $D_2$ , totiž  $\Delta(D_1, D_2)$ . Nevystačíme s ním však tehdy, budeme-li chtít porovnat vztahy diferenciací  $D_1$  a  $D_2$  na jedné straně a diferenciací  $D_3$  a  $D_4$  na straně druhé. Může se stát například, že

$$\Delta(D_1, D_2) > \Delta(D_3, D_4),$$

ale

$$\delta(D_1, D_2) < \delta(D_3, D_4)$$

apod. V takových případech pak je nutné užívat indexu vztahu diferenciací  $\delta$ .

„Procentuální“ vyjádření  $\delta$  lze snadno odvodit a nebudeme ho tudíž explicitně uvádět.

### 3. Příklady

Ukažme si nejprve, jak se výše uvedený postup projevuje v nejjednodušším případě tzv. dichotomických diferenciací  $D_1$  a  $D_2$ , tj. platí

$$D_1 = ({}^1d_1, {}^1d_2), \quad D_2 = ({}^2d_1, {}^2d_2).$$

Pro jednoduchost zápisu označme

$$\frac{|{}^1d_1|}{r} = a_1, \quad \frac{|{}^1d_2|}{r} = a_2, \\ \frac{|{}^2d_1|}{r} = a_{.1}, \quad \frac{|{}^2d_2|}{r} = a_{.2}$$

a dále analogicky

$$\frac{|{}^1d_1 \cap {}^2d_1|}{r} = a_{11}, \quad \frac{|{}^1d_1 \cap {}^2d_2|}{r} = a_{12}, \\ \frac{|{}^1d_2 \cap {}^2d_1|}{r} = a_{21}, \quad \frac{|{}^1d_2 \cap {}^2d_2|}{r} = a_{22}.$$

Hodnotu  $\Delta(D_1, D_2)$  lze určit přímým dosazením do výše uvedeného vzorce:

$$\Delta(D_1, D_2) = 1 - \sum_{\substack{i=1,2 \\ j=1,2}} \min(a_{ij}, a_{i.} \cdot a_{.j}) = \\ = \frac{1}{2} \sum_{\substack{i=1,2 \\ j=1,2}} |a_{ij} - a_{i.} \cdot a_{.j}|.$$

Výpočet  $\max \Delta(D_1, D_2)$  není v tomto případě obtížný.

Příslušná soustava matic  $S$  má v tomto konkrétním případě tvar

$$\begin{pmatrix} C & -C \\ -C & C \end{pmatrix},$$

přičemž

$$\max(-a_{1.} \cdot a_{.1}, -a_{2.} \cdot a_{.2}) \leq C \leq \\ \leq \min(a_{1.} \cdot a_{.2}, a_{2.} \cdot a_{.1}).$$

Je tedy zřejmé, že

$$\sum_{\substack{i=1,2 \\ j=1,2}} |C_{ij}|$$

bude maximální právě tehdy, bude-li

$$C = \max(\min(a_{1.} \cdot a_{.2}, a_{2.} \cdot a_{.1}), \\ -\max(-a_{1.} \cdot a_{.1} - a_{2.} \cdot a_{.2})) = \\ = \max(\min(a_{1.} \cdot a_{.2}, a_{2.} \cdot a_{.1}), \\ \min(a_{1.} \cdot a_{.1}, a_{2.} \cdot a_{.2})).$$

Je tedy jasné, že v tomto speciálním případě

$$\max \Delta(D_1, D_2) = \\ = 2 \cdot \max(\min(a_{1.} \cdot a_{.2}, a_{2.} \cdot a_{.1}), \\ \min(a_{1.} \cdot a_{.1}, a_{2.} \cdot a_{.2})).$$

Odtud je zřejmý obecný tvar  $\delta(D_1, D_2)$  pro čtyřpolní tabulku:

$$\delta(D_1, D_2) = \\ = \frac{1 - (\min(a_{11}, a_{1.} \cdot a_{.1}) + \min(a_{12}, a_{1.} \cdot a_{.2} + \\ + \min(a_{21}, a_{2.} \cdot a_{.1}) + \min(a_{22}, a_{2.} \cdot a_{.2}))}{2 \cdot \max(\min(a_{1.} \cdot a_{.2}, a_{2.} \cdot a_{.1}), \\ \min(a_{1.} \cdot a_{.1}, a_{2.} \cdot a_{.2}))}$$

V konkrétním příkladě, kde

$$A = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{6} \end{pmatrix},$$

tzn.

$$a_{1.} = \frac{1}{2}, \quad a_{2.} = \frac{1}{2}, \quad a_{.1} = \frac{7}{12}, \quad a_{.2} = \frac{5}{12},$$

platí, že

$$\Delta(D_1, D_2) = \frac{1}{12},$$

$$\max \Delta(D_1, D_2) = \frac{5}{12},$$

a tudíž

$$\delta(D_1, D_2) = \frac{1}{5}.$$

Přes zdánlivě složitý tvar výpočtu  $\delta(D_1, D_2)$  není v tomto speciálním případě výpočet indexu vztahu dvou diferenciací nikterak obtížný.<sup>12</sup>

Ukažme v dalším, jaké zajímavé a svým způsobem nové pohledy zpracování dat týkajících se sociální diferenciace přinášejí výše uvedené úvahy (omezíme se nyní pouze na aplikace uvedené metriky  $Q$ ).

V krátkosti ukážeme, jak lze těchto úvah využít k určitému typu rozboru vztahu věkové a sociální diferenciace tak, jak je poskytlují data ze Sčítání lidu ČSSR v r. 1970. Budeme vycházet z tabulky udávající sociálně věkovou diferenciaci ekonomicky aktivní populace.<sup>13</sup>

Na základě tabulky 1 lze řešit celou řadu úloh. Jednou z nich je analyzovat vzdálenosti sociálního složení jednotlivých věkových kohort od sociálního složení celé ekonomicky aktivní populace.

Převedením výše uvedené tabulky na tabulku sociálního složení jednotlivých věkových kohort (tj. vydělením každého elementu řádku uvedené tabulky celkovým procentem daného řádku), kterou zde explicitně neuvádíme, a určením vzdáleností  $Q$  takto vzniklých distribucí získáme tabulku 2.

Nebudeme na tomto místě provádět rozbor situace, kterou uváděná data vyjadřují. Omezíme se pouze na zdůraznění toho, že velmi výrazně odrážejí postupné historické proměny sociálního složení ekonomicky aktivní populace. Zároveň naznačují i důležité aspekty reprodukčního cyklu sociální struktury: starší generace jsou vzdáleny od průměru sociální struktury (tj. celku), neboť jsou nositeli minulých forem sociální diferenciace, zatímco odehlyly sociálního složení mladých věkových skupin od celkového sociálního složení vyjadřují, že jde o nositele nových tvarů, forem a hodnot sociální struktury, tedy budoucnosti sociálně třídního složení společnosti. Přítomnost v tomto aspek-

Tabulka 1. Věkově sociální diferenciace ekonomicky aktivní populace ČSSR v r. 1970 (%)

| Věková skupina | Sociální skupina | Dělníci mimo zemědělství | Zemědělníci dělníci | Ostatní zaměstnanci | Družstevní rolníci | Ostatní družstevníci | Jednotlivé hospodářci rolníci | Ostatní | Celkem |
|----------------|------------------|--------------------------|---------------------|---------------------|--------------------|----------------------|-------------------------------|---------|--------|
| 15-20          |                  | 5,1                      | 0,2                 | 0,7                 | 1,1                | 0,2                  | 0                             | 0       | 7,3    |
| 20-30          |                  | 15,1                     | 0,7                 | 8,2                 | 1,1                | 0,5                  | 0                             | 0       | 25,6   |
| 30-40          |                  | 10,5                     | 0,6                 | 10,0                | 1,4                | 0,4                  | 0,1                           | 0       | 23     |
| 40-50          |                  | 11,4                     | 0,7                 | 7,4                 | 2,3                | 0,5                  | 0,2                           | 0,1     | 22,6   |
| 50-60          |                  | 6,7                      | 0,4                 | 7,3                 | 1,7                | 0,3                  | 0,2                           | 0       | 16,6   |
| 60+            |                  | 2,1                      | 0,2                 | 0,9                 | 1,4                | 0,1                  | 0,2                           | 0       | 4,9    |
| Celkem         |                  | 50,9                     | 2,8                 | 34,5                | 9                  | 2                    | 0,7                           | 0,1     | 100 %  |

<sup>12</sup> Nebudeme se na tomto místě zabývat diskusí o vztahu různých indexů těsnosti vztahu dvou dichotomických znaků a námi zavedeným indexem vztahu dvou diferenciací  $Q$ . Poznamenejme, pouze, že v uvedeném konkrétním případě se Yule-ův index

$Q = -\frac{2}{3}$ , jeho interpretace je však (na rozdíl od

hodnoty  $Q = \frac{1}{5}$ ) těžko přirozeně interpretovatelná.

<sup>13</sup> Zdůrazňujeme, že nám v tomto okamžiku nejde o diskusi zvolených hodnot sociální diferenciace (struktury), která je determinována znaky, jichž bylo v daném censu užito. Takové úvahy by totiž vybočily z rámce úvah, který jsme si vytkli.

**Tabulka 2. Vzdálenosti sociálního složení jednotlivých věkových kohort od sociálního složení ekonomicky aktivní populace ČSSR v r. 1970 (v %)**

|       | $\rho$ |
|-------|--------|
| 15–20 | 25,1   |
| 20–30 | 8,1    |
| 30–40 | 9      |
| 40–50 | 2,3    |
| 50–60 | 11,2   |
| 60+   | 24,2   |

tu reprezentují střední věkové skupiny obyvatelstva.

Nikoliv bez zajímavosti bude i tabulka uvádějící vzdálenosti sociálního složení jednotlivých věkových kohort, které plasticky dokreslí výše uvedená data (zde jde opět o velmi jednoduchou aplikaci metriky  $\rho$  zavedené v lemmatu 1, aplikovanou na jednotlivé řádky příslušně upravené matice, resp. tabulky 1).

I analýzou této tabulky dostaneme zajímavé údaje o historických proměnách sociál-

ního složení ekonomicky aktivního obyvatelstva, které zdaleka nejsou patrné z tabulky. Vzdálenosti sociálního složení jednotlivých věkových skupin totiž odrážejí shody či rozdíly sociální historie v životních drahách jednotlivých věkových skupin.

Na těchto dvou příkladech (tabulka 2 a 3) jsme chtěli alespoň v náznavu ukázat, jaký cenný (a přitom nikterak složitý) nástroj analýzy vztahu dvou diferenciací uvedené charakteristiky respektive postupy představují. Přitom jde vesměs o sekundární kvantitativní charakteristiky, které mají zcela jasný „sociální obsah“ a v mnohých případech mohou sloužit jako specifický typ indikátorů sociálního vývoje. V tom také spočívá jejich přednost před tradičními prostředky statistiky.<sup>14</sup>

#### 4. Závěrečné poznámky

Uvedený přístup k analýze sociálních diferenciací vychází z nového — nebo lépe řečeno ne tak uživaného — pohledu na danou tematiku. Nechápe diferenciaci absolutně, nechce ji vyhlašovat za univerzální makro-

**Tabulka 3. Vzdálenosti sociálního složení jednotlivých kohort ekonomicky aktivní populace ČSSR v roce 1970**

| 15–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60+  |
|-------|-------|-------|-------|-------|------|
| 15–20 | 22,4  | 34,3  | 24,9  | 35,6  | 27,7 |
|       | 20–30 | 13,7  | 8,6   | 19,1  | 29,7 |
|       |       | 30–40 | 10,8  | 5,5   | 27,9 |
|       |       |       | 40–50 | 11,6  | 22,5 |
|       |       |       |       | 50–60 | 25,6 |

<sup>14</sup> Chceme na tomto místě poukázat na to, že celá řada možností dalšího rozpracování zvoleného postupu zůstala nevyužita. Předně je nutné zdůraznit, že veškeré naše dosavadní úvahy lze rozšířit na libovolný konečný počet diferencí (myslíme zde především zavedení  $\Delta (D_1, D_2, D_3, \dots, D_n)$  či  $\sigma (D_1, D_2, D_3, \dots, D_n)$  po konečné  $n$ ).

Stejně tak zůstala prozatím stranou celá „statistika“ příslušných charakteristik  $\Delta$  a  $\sigma$ . Poznamenáme dále, že veličinu  $\sigma$  lze s úspěchem použít při konstrukci tzv. kauzálních modelů kvalitativních sociologických proměnných (viz Charvát F.: *Model kauzality v soustavě kvalitativních sociologických*

*proměnných*, Sociologický časopis 1, 1975).

S určitou adaptací lze zvolený postup užít i pro analýzu tzv. mobilních tabulek (nový způsob zavedení tzv. strukturální a čisté mobility, a to „sociálně přirozený“ způsob zavedení).

Příslušná metrika  $\rho$  pak může sloužit jako základní prostředek pro aplikaci tzv. multidimenzionálního škálování či typologických procedur v případě komparace řady distribucí (diferenciací na různých populacích) apod.

K těmto otázkám se vrátíme v dalších pracích, věnovaných rozpracování daného tematického postupu.

sociologickou kategorií, jako to činí tzv. stratifikační přístup, ale chápe ji dialekticky jako určitý projev integračních sociálních tendencí. Pochopitelně nemůže ve svém rámci „řešit“ výchozí princip marxisticko-leninské analýzy společenského vývoje soudobých historických podmínek, totiž rozhodující úlohu tříd v sociálně ekonomickém vývoji soudobých společností. Může však k jeho empirickému vyjádření velmi konkrétně přispět, a tedy z něj vycházet a rozvíjet jej v empirických formách vědeckého odrazu sociální skutečnosti.

Zdánlivě jednoduchá myšlenka charakterizující průnik dvou diferenciací jako systém diferenciací daných jednou celkovou diferenciací v elementech druhé diferenciace nejenže naplňuje obsahem nových pohledů běžné, rutinní analýzy tzv. kontingenčních tabulek, třídění druhého stupně sociologických znaků, ale kontrastuje s určujícími teoreticko-empirickými postuláty stratifikačních koncepcí (syntetizace dílčích statusů, skládání dílčích stratifikací v jediný sociální status ať již užitím elementárních aritmetických či typologických a jiných procedur). Umožňuje totiž nejen konstruovat přirozené charakteristiky vztahu dvou či více diferenciací, ale bezprostředně iniciuje i závažné problémy a otázky meritorního rázu, jako například otázka jaký je podíl dělnické třídy na proměnách sociální struktury dané společnosti, k jakým závěrům vede srovnání podílu jednotlivých tříd a sociálních skupin na procesech reprodukce sociálně třídní struktury,<sup>15</sup> v jakých konkrétních podobách a měřítech probíhá základní proces sociálně třídního vývoje socialistické společnosti, totiž sblížení tříd a sociálních skupin, a celou řadu dalších.

Proto je logickým závěrem této dílčí informace požadavek vytvoření teorie sociální diferenciace, marxisticko-leninské teorie, která by svými poznatky vydatně přispěla k řízení složitých procesů výstavby socialistické a komunistické společnosti.

## Резюме

Харват Ф.: Об одном из возможных подходов к анализу социальных различий (дифференций)

В статье прежде всего вводится т. наз. «расстояние между двумя упорядоченными диффе-

ренциациями с одинаковым числом элементов» как

$$\varrho(a, b) = 1 - \sum_{i=1}^m \min(a_i, b_i),$$

где  $a = \{a_i\}_1^m$ ,  $b = \{b_i\}_1^m$ ,  $a_i \geq 0$ ,  $b_i \geq 0$ ,

$$\sum_{i=1}^m a_i = \sum_{i=1}^m b_i = 1.$$

Отношение двух общих дифференциаций (качественных признаков) можно, естественно, измерять величиной, которая выражает, как дифференцированно проявляется одна дифференциация в элементах второй дифференциации.

Показано, что мерой этого отношения можно считать величину

$$\Delta(D_1, D_2) = 1 - \sum_{i=1}^m \sum_{j=1}^n \min(a_{ij}, a_i \cdot a_j),$$

где  $(a_{ij})$  контингентальная таблица дифференциаций  $D_1$  и  $D_2$  и  $a_i$ ,  $a_j$  являются соответствующими суммами по строкам и столбцам,  $i = 1, \dots, m$ ;  $j = 1 \dots n$ . Далее показано, что максимальная возможная величина  $\Delta(D_1, D_2)$  при данных  $a_i$ ,  $a_j$  может быть исчислена лишь средствами математического программирования. Вводится т. наз. индекс отношения двух дифференциаций в виде величины

$$\delta(D_1, D_2) = \frac{\Delta(D_1, D_2)}{\max \Delta(D_1, D_2)}$$

которая обладает естественными и предполагаемыми свойствами.

Заключение статьи посвящено определению  $\delta$  для четырехмерной таблицы и приводятся образчики использования  $\varrho$  расстояния для анализа данных социальной и возрастной дифференциации.

## Summary

Charvát F.: Analysis of the Structure of Social Differences — One of the Possible Approaches

The paper begins with introducing the so-called "distance between two ordered differentiations having the same number of elements" as

$$\varrho(a, b) = 1 - \sum_{i=1}^m \min(a_i, b_i), \text{ where}$$

$$a = \{a_i\}_1^m, \quad b = \{b_i\}_1^m, \quad a_i \geq 0, \quad b_i \geq 0,$$

$$\sum_{i=1}^m a_i = \sum_{i=1}^m b_i = 1.$$

sociální struktury socialistické společnosti: K některým aspektům obecných a zvláštních momentů její vedoucí úlohy (Sociologický časopis 6/1976.).

<sup>15</sup> K obecné teoretické formulaci těchto závažných otázek dynamiky sociálně třídní struktury viz F. Charvát: Dělnická třída v procesech reprodukce

The relation between two common differentiations (qualitative variables) can naturally be measured in terms of the value indicating the differentiated operation of one differentiation in the elements of another differentiation.

It has been proved that the quantity

$$\Delta(D_1, D_2) = 1 - \sum_{i=1}^m \sum_{j=1}^n \min(a_{ij}, a_{i.}, a_{.j}),$$

where  $(a_{ij})$  contingency table of differentiations  $D_1$  and  $D_2$  and  $a_{i.}$ ,  $a_{.j}$  are the respective sums of rows and columns,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ , can be taken for the measure of this relation.

The paper goes on to prove that, while  $a_{i.}$ ,  $a_{.j}$  are given, the maximum possible value  $D_1$ ,  $D_2$  is generally calculable only by applying the means of mathematical programming. The so-called index of the relation between two differentiations is introduced as the quantity

$$\delta(D_1, D_2) = \frac{\Delta(D_1, D_2)}{\max \Delta(D_1, D_2)},$$

having natural and expected properties.

In conclusion the paper is concerned with the determination of  $\delta$  for a four-field table and with examples of the application of  $\delta$  of distance in analyzing data of social and age differentiation.