

## Úvod

V současné době se hovoří — v souvislosti s odhalováním příčin malé efektivity sociologických výzkumů — mimo jiné i o nutnosti kritického zhodnocení dosud užívaných metod analýzy. Domníváme se, že v rámci tohoto zhodnocení by se však nemělo zapomenout i na doplňování metodologického aparátu o zatím málo známé metody, které v praxi sociologických výzkumů nebyly ještě použity. Můžeme mezi ně zahrnout i diskriminační analýzu, vhodnou pro řešení klasifikačních problémů. Ukazuje se zřetelně, že je třeba používat nových postupů pro třídění souborů do skupin, zvláště tehdy, máme-li početné soubory, v nichž jednotky jsou charakterizovány několika znaky nebo dokonce desítkami znaků. Klasifikace vyžaduje řadu přesně (nebo alespoň v jasných mezích) definovaných kritérií, kterými rozdělíme předměty našeho výzkumu do určitých skupin neboli tříd. Soubor kritérií musí být zvolen tak — má-li být klasifikace správná a logická — že vzniklé třídy, rozlišené na základě těchto kritérií, se navzájem vylučují a všechny dohromady pokrývají celé pole jevů, které jsou klasifikovány. Nejprostším typem klasifikace je rozdělení nějakého souboru do dvou tříd na základě vhodného výběru kritérií, která splňují již zmíněnou podmínku, že obě třídy se budou vzájemně vylučovat a tvořit uzavřený celek. Tomuto případu věnujeme nyní především pozornost a ukážeme použití multivariační diskriminační analýzy jako vhodného nástroje pro tento typ klasifikace v sociologii. Budeme se tedy zabývat zejména použitím lineárních diskriminačních funkcí. O kvadratických diskriminačních funkcích a vícenásobných diskriminacích, které se hodí pro složitější případy, se zmíníme pouze informativně.

Pro snažší pochopení diskriminačních funkcí jako klasifikační procedury uvedeme v druhé části statí i praktické použití, tak jak bylo provedeno při výzkumu hodnotových orientací mládeže.

## Princíp multivariační diskriminační analýzy

Multivariační diskriminační analýza je statistická metoda umožňující tříditi jednotky do předem definovaných skupin na základě většího počtu znaků, které bereme současně v úvahu. Známe-li dvě skupiny jednotek, které můžeme zařadit do různých populací na podkladě dřívější zkušenosti, může se těchto skupin použít pro vytvoření kritérií klasifikace dalších jednotek. Tím může být současně dále posouzena a testována vhodnost původně navržené klasifikace.

Skupina definovaná  $K$ -znaky může být zobrazena jako shluk bodů v  $k$ -rozměrném prostoru. Druhá skupina, popsaná a určená týmiž  $K$ -znaky, se skládá z druhého shluku bodů. Diskriminační analýza představuje v podstatě výpočetní techniku, kterou určíme ( $K - 1$ ) rozměrnou plochu (nadrovinu) oddělující nejvhodněji dva shluky bodů. Jednotka, o níž nevíme, kam ji zařadit, je klasifikována podle příslušnosti k té či oné skupině na základě toho, na kterou stranu plochy padne. Poloha dvou skupin je popsána  $k$ -rozměrnými souřadnicemi jejich multivariačních průměrů. Stupeň odlišnosti dvou skupin je měřen „vzdáleností“ mezi těmito „průměrnými“ hodnotami skupin. Pro názornost vyjádříme tuto představu pro měření dvou znaků  $A$ ,  $B$ : jednotky, které jsou těmito dvěma znaky charakterizovány, mohou být znázorněny body určenými souřadnicemi  $x$ ,  $y$  v rovině (viz obr. 1).

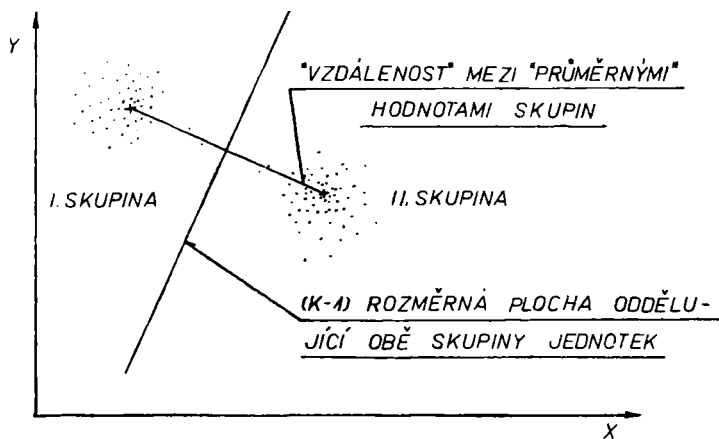
Soubor jednotek charakterizovaný mnoha znaky bude mít potom tvar elipsoidického shluku bodů, jehož rozměry jsou definovány velikostí rozptylu u každého znaku.

Některé z jednotek obou skupin mohou nabývat u některých znaků stejných hodnot. Tím se shluky dvou skupin bodů mohou v některém směru částečně prolínat. Čím jsou oba shluky zřetelněji separovány, tím spolehlivější je přiřazení dalších jednotek k jedné či druhé skupině.

Předpokládáme, že byla vybrána skupina jednotek o rozsahu  $N_1$  a skupina jednotek

o rozsahu  $N_2$ . Obě skupiny byly charakterizovány shodně K-znaky, přičemž symboly  $\bar{x}_{1i}$ ,  $\bar{x}_{2i}$  označují průměrné hodnoty znaků, kde indexy 1 a 2 označují příslušnost k první či druhé skupině a  $i$  označuje příslušný znak.

OBR. 1.



Diference mezi průměry bude  $\bar{x}_{11} - \bar{x}_{21} = d_1$ .

Pro každou skupinu vytvořme multivariační průměr, tedy

$$\lambda_1 \bar{x}_{11} + \lambda_2 \bar{x}_{12} + \dots + \lambda_k \bar{x}_{1k} = \bar{x}_1 \quad (1)$$

$$\lambda_1 \bar{x}_{21} + \lambda_2 \bar{x}_{22} + \dots + \lambda_k \bar{x}_{2k} = \bar{x}_2 \quad (2)$$

kde  $\lambda_i$  jsou konstanty, jejichž smysl či účel bude vyjasněn později. Uvažujme rozdíl těchto multivariačních průměrů

$$\bar{x}_1 - \bar{x}_2 = \lambda_1 d_1 + \lambda_2 d_2 + \dots + \lambda_k d_k = D^2 \quad (3)$$

Rozptyl hodnot  $D^2$  bude

$$s_D^2 = \sum_{i,j} \lambda_i \lambda_j s_{ij} \quad (4)$$

kde  $s_{ij}$  jsou prvky společné kovarianční matice hodnot znaků pro oba soubory.

K posouzení souhrnné vzdálenosti ke všem znakům můžeme zvolit kritérium analogické Studentovu t-rozdělení

$$t_D = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s_D^2 \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}} \quad (5)$$

Nulovou hypotézu  $H_0$ , tj. že tyto dvě skupiny jsou totožné ve všech svých zkoumaných znacích, zamítneme, když hodnota  $t_D$  bude dostatečně velká. Při tom se budeme snažit, aby konstanty  $\lambda_1, \lambda_2 \dots \lambda_k$  byly zvoleny tak,

aby  $t_D$  bylo maximální. Maximalizace charakteristiky  $t_D$  vzhledem ke konstantám  $\lambda_1, \lambda_2 \dots \lambda_k$  se provádí obvyklou matematickou metodou, tj. položíme derivaci rovnou nule a řešíme takto vzniklé rovnice. Získáme tím

soustavu k-rovnic, ze které můžeme určit konstanty  $\lambda_1, \lambda_2 \dots \lambda_k$  tak, aby charakteristika (5) byla maximální.

Rovnice lineární diskriminační funkce  $R$  má potom tvar

$$R = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k, \quad (6)$$

kde symboly  $x_1, x_2, \dots, x_k$  označují znaky, které charakterizují klasifikované skupiny jednotek.

Neznámá jednotka je potom zařazena buďto do první, nebo do druhé skupiny dosazením hodnot K-znaků této jednotky do diskriminační funkce. Rozhodnutí o příslušnosti této neznámé jednotky k první nebo druhé skupině provádíme na základě porovnání její diskriminační hodnoty, získané již zmíněným dosazením hodnot K-znaků do rovnice (6), s hodnotou tzv. diskriminačního indexu  $R_0$ . Diskriminační index  $R_0$ , který tvoří jakousi „hranici“ mezi dvěma skupinami, nalezneme tak, že dosadíme místo hodnot  $x_1, x_2 \dots x_k$  průměry hodnot znaků ze dvou skupin do diskriminační funkce.

$$R_0 = \lambda_1 \frac{\sum x_{11} + \sum x_{21}}{N_1 + N_2} + \lambda_2 \frac{\sum x_{12} + \sum x_{22}}{N_1 + N_2} + \dots + \lambda_k \frac{\sum x_{1k} + \sum x_{2k}}{N_1 + N_2} \quad (7)$$

Dosud nezotříděné jednotky potom zařazujeme podle toho, je-li jejich diskriminační hodnota menší či větší než hodnota  $R_0$ . Je zřejmé, že diskriminační analýzy nelze použít, nejsou-li rozdíly mezi dvěma skupinami, které chceme oddělit, významné nebo jsou-li dokonce součástí téhož souboru. Je proto třeba testovat tento předpoklad rozdílnosti, aby bylo vidět, jak významný je rozdíl mezi průměry mnoha znaků dvou a priori vyčleněných skupin. Test významnosti lze odvodit z Mahalanobisovy zobecněné vzdálenosti  $D^2$ , tzv. míry „vzdálenosti“ mezi multivariačními průměry dvou skupin jednotek, kterou spočteme dosazením příslušných hodnot do rovnice (3). Pro testování významnosti použijeme následující testovací charakteristiky

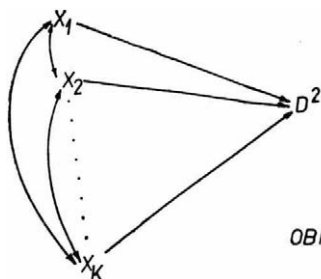
$$F = \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 - k - 1)}{k(N_1 + N_2)} \cdot D^2, \\ \frac{N_1 + N_2 - 2}{N_1 + N_2 - 2}$$

kde  $k$  je počet znaků charakterizujících skupiny jednotek. Tato testová charakteristika má za předpokladu, že obě zkoumané populace jsou v těchto znacích identické,  $F$  – rozdělí o  $k$  a  $(N_1 + N_2 - k - 1)$  stupňů volnosti.

O tom, do jaké míry přispívají jednotlivé znaky k celkové vzdálenosti mezi multivariačními průměry, se můžeme do jisté míry informovat použitím vzorců

$$\frac{\lambda_1 d_1}{D^2} \cdot 100$$

Toto procentuální vyjádření tedy umožňuje posoudit příspěvek zkoumaných znaků k velikosti  $D^2$ . Takto stanovené příspěvky jednotlivých znaků k celkové vzdálenosti mezi multivariačními průměry zahrnuje nejen přímý vliv jednotlivých znaků na zmíněnou hodnotu  $D^2$ , ale i vliv nepřímý, který se projevuje prostřednictvím ostatních znaků. Graficky si tuto situaci můžeme znázornit takto:



OBR. 2.

Pokud uvedeného způsobu použijeme k vyřazení nepodstatných znaků z celku použitých kritérií klasifikace, je situace komplikována často tím, že neznáme míru nezávislosti jednotlivých znaků navzájem. K tomu, do jaké míry je určitý znak významný pro rozlišení souboru, lze konstruovat i přesný test.

Někdy je vhodné se pokusit o přiřazení osob k a priori definovaným skupinám ne pomocí diskriminační funkce typu, který je popsán vztahem (6), ale pomocí funkce jiného typu.

Lineární diskriminační funkci nazýváme takovou funkci, která je lineární v hledaných koeficientech  $\lambda_i$ . Stanovení koeficientů u takové funkce vždy vede k řešení systému lineárních rovnic. Zvolení nelineární funkce, pokud jde o hledané koeficienty  $\lambda_i$ , vede k řešení nelineárního systému rovnic, což představuje po stránce numerické značné nesnáze.

Nicméně v literatuře někteří autoři hovoří o kvadratické diskriminační funkci. Tato funkce je lineární v koeficientech  $\lambda_i$ , ale nelineární v některých zkoumaných znacích  $x_i$ , podle kterých se provádí zařazování do předem zvolených skupin. Pro dva znaky  $x_1$  a  $x_2$  má taková diskriminační funkce tento tvar

$$R = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_1 x_2 + \lambda_4 x_1^2 + \lambda_5 x_2^2 \quad (8)$$

Je patrné, že podobných funkcí, které budou nelineární ve znacích  $x_i$ , je možno konstruovat libovolné množství. O tom, do jaké míry bude lépe sloužit k diskriminaci zvolená funkce, např. kvadratická diskriminační funkce, nás bude informovat do jisté míry charakteristika  $D^2$ . Čím větší je charakteristika  $D^2$ , tím více je k ní příslušná diskriminační funkce vyhovující.

Ještě se stručně zmíníme o diskriminační analýze pro více souborů. V literatuře [1], [2] jsou popsány dva způsoby, kdy v [1] se řeší přiřazení dalších jednotek k a priori definovaným skupinám. V [2] je řešena tatáž úloha, ale kromě toho se ještě řeší spojení dvou či více skupin. Podrobnější popis těchto metod přesahuje rámec stati.

Ještě než přejdeme k praktickým ukázkám, je nutné upozornit na jeden z předpokladů aplikace diskriminační analýzy, kdy klasifikační kritéria musí být vyjádřena znaky, u nichž je smyslupné pracovat s charakteristikami, jako je průměr a rozptyl.

## Konkrétní aplikace diskriminační analýzy

Použité příklady jsou převzaty, jak již bylo v úvodu poznamenáno, z výzkumu hodnotových orientací mládeže, kde diskriminační analýza byla použita při řešení dvou typů úloh:

1. Ověřit, zda verbálně vyjádřený postoj zkoumaných osob může být východiskem pro odhad jejich konkrétního jednání.
2. Porovnat sociální skupiny mládeže tak, abychom současně mohli stanovit společné a specifické znaky porovnávaných skupin.

První typ úlohy převedeme na klasifikační problém, kdy známe dvě skupiny, které na podkladě „dřívejších zkušeností“ můžeme zařadit do různých populací a které použijeme pro posouzení vhodnosti navržených klasifikačních kritérií.

V našem konkrétním případě jsme pracovali vždy se dvěma skupinami osob, které se lišily způsobem jednání, přičemž operacionálním vyjádřením způsobu jednání byl údaj, zda sledované osoby vykonávají či nevykonávají určitou činnost. Ověřovali jsme dále, nakolik verbální projevy osob týkající se příslušné činnosti odpovídají jejich konkrétnímu jednání. Jinými slovy, zda můžeme vypovídat o způsobu jednání, známe-li pouze verbálně vyjádřené postoje zachycené v dotazníku. Z hlediska klasifikace jsou skupiny osob, které se liší výkonem činnosti, zmiňovanými skupinami, o nichž víme, že je můžeme zařadit do různých populací na podkladě „dřívejších zkušeností“; testovanými klasifikačními kritérii jsou navrhované znaky (otázky), kterými zjišťujeme verbálně vyjádřené postoje.

### Příklad 1:

Soubor učňů z jednoho pražského strojírenského závodu jsme rozdělili do dvou skupin podle toho, zda sledované osoby vykonávaly či nevykonávaly sportovní činnosti.<sup>1</sup> Obě skupiny byly shodně charakterizovány klasifikačními kritérii, tj. touto četveřící znaků:

X<sub>1</sub>: Kdykoli sám(a) sportuji, pocituji radost z uspokojení svých zájmů.<sup>2</sup>

X<sub>2</sub>: Považuji sportovní činnost z rozumných důvodů za velmi významnou, i když vždy neuspokojuje mé zájmy a při jejím výkonu nepociťuji bezpodmínečnou radost.

X<sub>3</sub>: Za všech okolností usiluji, abych sám(a) sportoval(a).

X<sub>4</sub>: Pokládám za velmi žádoucí získávat osobně též ostatní pro sportování.

Dále byla provedena klasifikace osob a stanoveny tyto charakteristiky:

- a) koeficienty  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ;
- b) diskriminační index  $R_0$ ;
- c) míra rozdílu mezi multivariačními průměry obou podskupin, tj. hodnota Mahalanobisovy vzdálenosti  $D^2$ , která byla testována;
- d) hodnoty relativních parametrů.

Pro obě skupiny jsme vypočtené charakteristiky získali v následujícím tvaru:

### Řešení soustavy rovnic

$$X(1) = -0,5400$$

$$X(2) = 0,1744$$

$$X(3) = -0,2728$$

$$X(4) = -0,7827$$

### Mahalanobisova vzdálenost

$$1,013464$$

### Diskriminační index

$$RO = -2,8009671$$

$$F = 1,596205$$

Stupně volnosti: 4 45

### Relativní parametry<sup>3</sup>

$$0,2695619$$

$$-0,0594055$$

$$0,1922530$$

$$0,5975905$$

Hodnoty prvních čtyř řádků nám umožňují sestavit rovnici diskriminační funkce

$$R = -0,5400 X_1 + 0,1744 X_2 - 0,2728 X_3 - 0,7827 X_4$$

<sup>1</sup> Tento fakt jsme zjišťovali prostřednictvím dichotomické otázky: Věnujete se pravidelně aktivnímu sportování? — ano — ne

Jelikož jsme i v tomto případě odkázali na výpověď respondenta (jejíž pravdivost nemůžeme kontrolovat), musíme si uvědomit, že v méně příznivém případě budeme zjišťovat pouze konzistentnost výpovědi respondenta.

<sup>2</sup> Dotazovaní vyjadřovali svůj souhlas či nesouhlas odpověďmi: 1 — zcela souhlasím, 2 — souhlasím, 3 — nevím, nemám názor, 4 — nesouhlasím, 5 — zcela nesouhlasím.

<sup>3</sup> Součet hodnot relativních parametrů (s ohledem na znaménka) musí být vždy roven 1.

pomocí níž bychom mohli zařadit dosud neroztříděné jedince buď do jedné, nebo do druhé podskupiny.

Museli jsme se však ještě přesvědčit, zda rozdíl mezi multivariačními průměry  $D^2$  je významný.

Porovnali jsme proto vypočtenou hodnotu testovací charakteristiky  $F$  s tabulkovou hodnotou pro  $\alpha = 0,05$  a stupně volnosti  $\gamma_1 = 4$  a  $\gamma_2 = 45$ . Potom  $F = 1,596205 < F_{0,05; 4; 45} = 2,58$ ; nemůžeme tedy zamítnout nulovou hypotézu. To znamená, že rozdíl mezi multivariačními průměry není významný, nemůžeme proto použít diskriminační funkce pro zařazování osob do daných skupin a nemá ani smysl na základě hodnot relativních parametrů posuzovat podíl jednotlivých znaků na zjištěném rozdílu obou multivariačních průměrů.

Navrhovaná kritéria tedy nevyhovují požadavkům klasifikace. Představíme-li si osoby charakterizované danými kritérii jako body ve čtyřrozměrném prostoru, tvoří navzájem se prolínající shluky bodů, které není možno statisticky oddělit tak, abychom mohli rozhodnout, kteří jedinci sportovní činnost vykonávají a kteří ne.

Tím jsme prokázali, že v daném případě nemůžeme považovat verbálně vyjádřené postoje za východisko pro odhad konkrétního jednání sledovaných osob.<sup>4</sup>

#### Příklad 2:

Při rozdělení téhož souboru na jedince, kteří se zabývají činnostmi v zájmových kroužcích, a na jedince, kteří se jí nezabývají, jsme použili těchto klasifikačních kritérií.

- $x_1$ : Kdykoli se sám(a) podílím na klubovní činnosti, pracuji-li v zájmovém kroužku, pocituji radost z uspokojení svých zájmů.
- $x_2$ : Považuji z rozumových důvodů podíl na klubových činnostech a práci v zájmových kroužcích za velmi významné, i když vždy neuspokojují mé zájmy a nepocituji při nich bezpodmínečnou radost.
- $x_3$ : Za všech okolností usiluji, abych se účastnil(a) klubových činností nebo pracoval(a) v některých zájmových kroužcích a organizacích.
- $x_4$ : Pokládám za velmi žádoucí získávat osobně též ostatní pro účast na klubových činnostech či pro práci v zájmových kroužcích i organizacích.

Získali jsme tyto výsledky:

#### Řešení soustavy rovnic

$$\begin{aligned}x(1) &= -,2391 \\x(2) &= -,1836 \\x(3) &= -,7126 \\x(4) &= -,3201\end{aligned}$$

#### Mahalanobisova vzdálenost

$$1,178905$$

#### Diskriminační index

$$\begin{aligned}RO &= -3,2185364 \\F &= 3,365406\end{aligned}$$

$$\text{Stupně volnosti:} \quad 4 \quad 45$$

#### Relativní parametry

$$\begin{aligned}&,0948977 \\&,1050792 \\&,6630475 \\&,1369756\end{aligned}$$

Nejdříve se podíváme, zda rozdíl mezi multivariačními průměry je významný. Porovnáme-li opět hodnoty  $F$  a  $F_{0,05; 4; 45}$ , potom

$$F = 3,365406 > F_{0,05; 4; 45} = 2,58$$

tzn., že můžeme zamítnout nulovou hypotézu a považujeme tedy rozdíl mezi průměry, reprezentovaný hodnotou  $D^2$ , za statisticky významný. Přistoupíme proto k další analýze uvedeného příkladu. Dejme tomu, že máme dvě osoby, o kterých dosud nevíme, kam je přiřadit, a kteří dosáhli těchto výsledků:

| Osoba A   | Osoba B   |
|-----------|-----------|
| $x_1$ : 1 | $x_1$ : 5 |
| $x_2$ : 2 | $x_2$ : 5 |
| $x_3$ : 1 | $x_3$ : 4 |
| $x_4$ : 1 | $x_4$ : 4 |

Hodnoty dosažených výsledků dosadíme do lineární diskriminační funkce

$$\begin{aligned}R &= -0,2391 x_1 - 0,1836 x_2 - \\&- 0,7126 x_3 - 0,3201 x_4\end{aligned}$$

Osoba A:

$$\begin{aligned}R_A &= -0,2391 \cdot 1 - 0,1836 \cdot 2 - \\&- 0,7126 \cdot 1 - 0,3201 \cdot 1 = -1,6390\end{aligned}$$

Osoba B:

$$\begin{aligned}R_B &= -0,2391 \cdot 5 - 0,1836 \cdot 5 - \\&- 0,7126 \cdot 4 - 0,3201 \cdot 4 = -6,2443\end{aligned}$$

<sup>4</sup> V dalším postupu bychom tedy byli nuceni hledat takové formulační vyjádření otázek, které by

nám umožňovalo provádět odhady konkrétního jednání sledovaných osob.

Srovnáme-li diskriminační hodnoty  $R_A$ ,  $R_B$  s hodnotou diskriminačního indexu, potom

$$R_A = -1,6390 > R_0 = -3,2185364$$

$$R_B = -6,2443 < R_0 = -3,1285364$$

Vidíme, že leží na opačných stranách indexu  $R_0$ . Každá z obou osob tedy přísluší do jiné skupiny. Konkrétně by byly zařazeny do té skupiny, jejíž diskriminační hodnota<sup>5</sup> leží na stejné straně diskriminačního indexu jako diskriminační hodnota zařazované osoby. Podíváme-li se na hodnotu relativních parametrů, je zřejmé, že nejvýznamnější podíl na zjištěném rozdílu mezi multivariačními průměry mají výpovědi o aspiracích, tzn. výpovědi týkající se znaku  $x_3$ . Vyjádříme-li podíl na velikosti rozdílu v procentech, můžeme hovořit o 66,5 % ( $0,6650475 \cdot 100 = 66,5$  %) podílu výpovědi o aspiracích. Podíly zbývajících výpovědí nepovažujeme za významné.

V tomto případě tedy navrhovaná kritéria vyhovují požadavkům klasifikace (zejména kritérium reprezentované znakem  $x_3$ ), neboť představíme-li si opět osoby charakterizované danými kritérii jako body ve čtyřrozměrném prostoru, získáme takové shluky bodů, které budou statisticky oddělitelné tak, že je možno s jistou pravděpodobností rozhodnout, které z osob vykonávají činnosti v zájmových kroužcích a které ne. Můžeme proto příslušné verbálně vyjadřované postoje považovat za východisko pro odhad konkrétního jednání sledovaných osob.

Přejdeme nyní k ukázkám klasifikace v rámci druhého typu úlohy, kdy jsme chtěli porovnat sociální skupiny mládeže tak, abychom mohli zároveň stanovit společné a specifické znaky porovnávaných skupin.

Převědeme si tuto úlohu opět na klasifikační problém a budeme testovat významnost mezi a priori definovanými skupinami, které chceme prostřednictvím souboru klasifikačních kritérií oddělit, a současně posoudíme podíly jednotlivých kritérií na zjištěném rozdílu (bude-li ovšem rozdíl významný). Mírou rozdílu mezi skupinami je hodnota  $D^2$ , tzv. Mahalanobisova vzdálenost, pro jejíž testování použijeme charakteristiky F. Podkladem pro posouzení příspěvků z jednotlivých kritérií k hodnotě  $D^2$  jsou hodnoty relativních parametrů

Pracovali jsme s třemi a priori definovanými skupinami, tj. se skupinou středo-

školáků, učňů a dělníků. Klasifikační kritéria byla reprezentována 16 znaky  $x_1, x_2 \dots x_{16}$ . Za specifické znaky u porovnávaných skupin jsme považovali ty znaky, které podstatně přispívají ke zjištěným rozdílům mezi těmito skupinami. Ostatní znaky, které nejsou pro oddělování skupin podstatné a které můžeme z rovnice diskriminační funkce vyloučit, jsme zahrnuli mezi znaky společné.

Stejně jako v předcházejícím typu úlohy jsme získali následující charakteristiky:

- koeficienty  $\lambda_1, \lambda_2 \dots \lambda_{16}$
- diskriminační index  $R_0$
- míra rozdílu mezi multivariačními průměry obou skupin, tj. hodnota Mahalanobisovy vzdálenosti  $D^2$ , která byla testována
- hodnoty relativních parametrů.

### Příklad 3:

V rámci tohoto příkladu provedeme posouzení rozdílu mezi skupinami učňů a dělníků z jednoho pražského strojírenského závodu a zároveň stanovení specifických a společných znaků obou skupin, které byly charakterizovány 16 znaky  $x_1, x_2 \dots x_{16}$ . V následující tabulce jsou zahrnuty pouze hodnoty relativních parametrů a hodnota Mahalanobisovy vzdálenosti  $D^2$ , neboť z těchto údajů jsme vycházeli při interpretaci výsledků a nepovažujeme již za nutné uvádět vypočtené údaje ve formě výstupů, tak jako v příkladech č. 1 a 2.

Při rozdělování znaků na společné a specifické jsme museli stanovit mezní hodnotu relativních parametrů tak, abychom mohli oddělit podstatné a nepodstatné znaky z hlediska jejich příspěvků k hodnotě  $D^2$ . S tímto problémem jsme se již setkali v rámci příkladu č. 2, kde byla situace ovšem mnohem jednodušší, neboť znak  $x_3$  se opravdu výrazně podílel na zjištěném rozdílu mezi multivariačními průměry.

Nepoužije-li výzkumník přesného testu pro řešení daného problému, je to otázka zkušeností, kdy posouzením většího počtu případů se získá představa o vhodné velikosti mezní hodnoty. Je vhodná i konzultace s odborníkem. Při tomto v podstatě arbitrárním postupu, kdy nejsou jasně vymezena pravidla rozhodování, mohou samozřejmě dojít dva výzkumníci k poněkud rozdílným závěrům. Přesto ale považujeme tento způsob posouzo-

<sup>5</sup> Diskriminační hodnotu skupiny lze získat tak, že do diskriminační funkce dosadíme průměrné

hodnoty jednotlivých znaků, tzn. buď hodnoty  $X_{11}$  nebo  $X_{21}$ .

| Klasifikační kritéria                                                                                                                                                | Hodnoty<br>relativních<br>parametrů | Charakteristika<br>znaků |            | Hodnota<br>Mahalano-<br>bisovy vzdá-<br>lenosti D |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|--------------------------|------------|---------------------------------------------------|
|                                                                                                                                                                      |                                     | společný                 | specifický |                                                   |
| x <sub>1</sub> : Získávat osobně též ostatní pro sportování po-<br>kládám za velmi žádoucí                                                                           | 0,2171589                           |                          | ×          |                                                   |
| x <sub>2</sub> : Osobně získávat též ostatní pro uměleckou čin-<br>nost pokládám za velmi žádoucí                                                                    | −0,0281317                          | ×                        |            |                                                   |
| x <sub>3</sub> : Získávat osobně též ostatní pro vykonávání růz-<br>ných zábavných, rekreačních a odpočinkových<br>činností pokládám za velmi žádoucí                | 0,0541241                           | ×                        |            |                                                   |
| x <sub>4</sub> : Získávat osobně též ostatní pro tvorbu předmětů<br>a pro sběratelství pokládám za velmi žádoucí                                                     | 0,0645566                           | ×                        |            |                                                   |
| x <sub>5</sub> : Získávat osobně též ostatní pro účast na klubov-<br>ých činnostech či pro práci v zájmových krouž-<br>cích i organizacích pokládám za velmi žádoucí | 0,0368444                           | ×                        |            |                                                   |
| x <sub>6</sub> : Získávat osobně též ostatní pro vážný vztah ke<br>kamarádství pokládám za velmi žádoucí                                                             | −0,0282756                          | ×                        |            |                                                   |
| x <sub>7</sub> : Získávat osobně též ostatní pro vážný vztah<br>k milenečce lásce pokládám za velmi žádoucí                                                          | −0,0204441                          | ×                        |            |                                                   |
| x <sub>8</sub> : Získávat osobně též ostatní k účtě vůči společ-<br>nému životu s rodiči a sourozenci pokládám za<br>velmi žádoucí                                   | 0,0349467                           | ×                        |            |                                                   |
| x <sub>9</sub> : Získávat osobně též ostatní k vážnosti vůči práci<br>a ke studiu vlastního oboru pokládám za velmi<br>žádoucí                                       | −0,0544879                          | ×                        |            |                                                   |
| x <sub>10</sub> : Získávat osobně též ostatní pro zvyšování kvali-<br>fikace a prohlubování znalostí ve studijním obo-<br>ru pokládám za velmi žádoucí               | −0,0444141                          | ×                        |            |                                                   |
| x <sub>11</sub> : Osobně získávat též ostatní, aby přispěli k pra-<br>covnímu zlepšení ve svém oboru či zaměstnání,<br>pokládám za velmi žádoucí                     | 0,1186223                           |                          | ×          |                                                   |
| x <sub>12</sub> : Osobně získávat též ostatní pro účast na různých<br>pracovních a odborných soutěžích pokládám za<br>velmi žádoucí                                  | 0,2924604                           |                          | ×          |                                                   |
| x <sub>13</sub> : Přesvědčovat osobně též ostatní, že je důležité<br>získávat informace týkající se každodenního<br>politického dění, pokládám za velmi žádoucí      | −0,1269813                          |                          | ×          |                                                   |

|                                                                                                                                                    |           |   |   |          |
|----------------------------------------------------------------------------------------------------------------------------------------------------|-----------|---|---|----------|
| x <sub>14</sub> : Přesvědčovat osobně i ostatní, že je důležité hlouběji pochopit ideologické problémy současného světa, pokládám za velmi žádoucí | 0,0588662 | × |   |          |
| x <sub>15</sub> : Přesvědčovat osobně i ostatní, že je důležité zúčastňovat se na činnosti mládežnických organizací, pokládám za velmi žádoucí     | 0,3710196 |   | × |          |
| x <sub>16</sub> : Osobně přesvědčovat i ostatní, že je důležité zúčastňovat se na činnosti celospolečenských organizací, pokládám za velmi žádoucí | 0,0511354 | × |   |          |
|                                                                                                                                                    |           |   |   | 5,651573 |

Porovnáním hodnot  $F$  a  $F_{0,05;16;83}$ , kdy  $F > F_{0,05;16;83}$ , jsme se přesvědčili o statistické významnosti Mahalanobisovy vzdálenosti  $D^2$ . Jsme tedy oprávněni pokračovat v rozboru výsledků.

vání podílu jednotlivých znaků na hodnotě Mahalanobisovy vzdálenosti  $D$  za vyhovující.

V našem případě jsme považovali za specifické ty znaky, jejichž podíl na „vzdálenosti“

$D^2$  byl větší než 10 %, tzn.  $\frac{\lambda_1 d_1}{D^2} \cdot 100 > 10 \%$ . Rozdělení znaků do příslušných kategorií je zachyceno v tab. č. 1 (sloupec „Charakteristika znaků“).

Vidíme, že pouze menší část klasifikačních kritérií představuje specifické znaky porovnávaných skupin ( $x_1, x_{11}, x_{12}, x_{13}$  a  $x_{15}$ ), přičemž můžeme hovořit o 21,7 % podílu znaku  $x_1$ , 11,9 % podílu znaku  $x_{11}$ , 29,3 % podílu znaku  $x_{12}$ , 12,7 % podílu znaku  $x_{13}$  a o 37,1 % podílu znaku  $x_{15}$  na zjištěném rozdílu mezi skupinami. Dosažených výsledků bychom mohli použít i při klasifikaci dosud nezařazených osob. Z rovnice diskriminační funkce můžeme nyní vyloučit ty znaky, které podstatně nepřispívají k hodnotě  $D^2$ , tzn. společné znaky. Získáme tak následující tvar diskriminační funkce:

$$R = \lambda_1 x_1 + \lambda_{11} x_{11} + \lambda_{12} x_{12} + \lambda_{13} x_{13} + \lambda_{15} x_{15}$$

Prostřednictvím kritérií uvedených v tab. č. 1 jsme provedli klasifikaci ještě pro další dvojice předem vymezených skupin, tj. středoškoláci—učni a středoškoláci—dělníci. Zabývali jsme se otázkou, které ze sociálních skupin jsou při globálním pohledu relativně nejvíce a nejméně odlišné.

#### Příklad 4:

Chtěli jsme získat informaci o „stupni“ odlišnosti jednotlivých sociálních skupin mládeže. Vycházeli jsme z hodnot  $D^2$ , které jsme potom porovnávali.

|                       |                     |
|-----------------------|---------------------|
| Středoškoláci—učni    | $D_{1N} = 2,684505$ |
| Středoškoláci—dělníci | $D_{2N} = 1,364499$ |
| Dělníci—učni          | $D_{3N} = 5,651573$ |

Porovnáním získáme nerovnost  $D_{2N} < D_{1N} < D_{3N}$ . Relativně největší rozdíl byl tedy prostřednictvím klasifikačních kritérií zjištěn mezi skupinami dělníků a učňů. Relativně nejmenší rozdíl byl naopak zjištěn u skupiny středoškoláků a dělníků.

#### Závěr

Ukázali jsme si jeden z možných způsobů řešení klasifikačních problémů při sociologických výzkumech. Diskriminační analýza není samozřejmě jediným nástrojem schopným řešit dané problémy. Existují i jiné klasifikační procedury, jako např. faktorová analýza (Q-technika), postupy numerické taxonomie ap. Praktické ukázky řešení mají potom především umožnit pochopení principů diskriminační analýzy a nikoliv sloužit jako striktní návod k její aplikaci.

Oblast možného použití této v sociologii dosud nepříliš známé metody analýzy je mnohem širší. Výzkumník se při analýze dat s problémy klasifikace často setkává, přičemž soubor osob, se kterým pracuje, je většinou charakterizován značným počtem znaků,

кdy tedy právě použití diskriminační analýzy může být oprávněné.

Je nutné si ale i uvědomit, že klasifikační problémy implikují i řadu dalších metodologických problémů, např. operacionalizace klasifikačních kritérií, vymezení skupin, do kterých budeme osoby zařazovat, atd. Tyto je třeba vyřešit pomocí i jiných analytických nástrojů. Rozhodne li se tedy výzkumník pro použití diskriminační analýzy, musí ovšem vzít v úvahu, že bude nucen se zabývat i jinými, zdánlivě s klasifikací nesouvisejícími úlohami, jejichž vyřešením se teprve vytvoří předpoklady pro vlastní klasifikaci.

## Literatura

1. Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*. New York, John Wiley 1958.
2. Rao, C. R.: *Advanced Statistical Methods in Biometrics*. Research, New York, John Wiley 1952.

## Резюме

Гольда Д. - Тошовский Я.: Дискриминационный анализ

В своей статье авторы показывают пример, как можно использовать дискриминационный анализ в решении классификационных проблем в рамках социологических исследований. Первая часть статьи является введением в теорию этого до сих пор малоизвестного в социологии метода. Указаны принципы дискриминационного анализа и в общем наброшены возможности его применения в оценке полученных данных. Прежде всего уделяется внимание линейным дискриминационным функциям

$$R = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k,$$

тогда как квадратические дискриминационные функции и многократные дискриминации, годные к применению в более сложных случаях, обсуждаются только в информационном порядке. Употребленные примеры были заимствованы из исследования ценностных ориентаций учащейся и трудящейся молодежи, в котором дискриминационный анализ был использован в решении двух типов задач:

1. Проверить, может ли вербально выраженное отношение исследуемых лиц быть исходным моментом их конкретного поведения.

2. Сопоставить социальные группы молодежи таким образом, чтобы можно было в одно и то же время установить общие и специфические признаки сравниваемых групп.

В заключение статьи авторы подчеркивают, что дискриминационный анализ, конечно, не является единственным аналитическим орудием, способным решать данные проблемы, но в то же время приведенные примеры не представляют всю сферу возможностей его использования.

## Summary

Holda D. — Tošovský J.: Discrimination Analysis

In their paper, the authors present an example of using the discrimination analysis in solving classification problems within the scope of sociological research.

The first part of the paper contains an introduction to the theory of this method which, so far, has been little known in sociology. It discusses the principles of the discrimination analysis and indicates the possibilities of its application in evaluating the data obtained. Particular attention is paid to the linear discrimination functions

$$R = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k,$$

while the quadratic discrimination functions and multiple discriminations which may be adequately applied in more complicated cases are dealt with only for the sake of information. The examples introduced here were taken over from the research into the value orientations of studying and working youth, wherein the discrimination analysis was applied in solving the following two types of tasks:

1. to verify whether the subjects' verbally expressed attitude may serve as a basis for evaluating their concrete behaviour,

2. to compare the social groups of youth in a way enabling us simultaneously to determine the general and the specific features of the groups subjected to comparison.

In concluding their paper, the authors emphasize the fact that the discrimination analysis is by no means the only analytic instrument capable of solving the given problems and, simultaneously, that the examples presented do not cover the whole area of its application possibilities.