UNIVERZITA PALACKÉHO V OLOMOUCI

PEDAGOGICKÁ FAKULTA

Katedra technické a informační výchovy

Bakalářská práce

Jiří Chovanec

Vícerozměrné statistické metody v systému STATISTICA a možnosti jejich využití při zpracování výzkumných dat

Olomouc 2018 vedoucí práce: Doc. PhDr. Miroslav Chráska, Ph.D.

Prohlášení

Prohlašuji, že jsem bakalářskou práci s názvem *Vícerozměrné statistické metody v systému STATISTICA a možnosti jejich využití při zpracování výzkumných dat* vypracoval samostatně, a že veškerá použitá literatura je uvedena v závěru práce.

V Olomouci dne: 20. 4. 2018

Jiří Chovanec

Poděkování

Děkuji panu doc. PhDr. Miroslavu Chráskovi, Ph.D. za veškeré poskytnuté materiály, rady, konzultace a především za odborné vedení mé bakalářské práce. Dále bych chtěl poděkovat ředitelům základních škol, na kterých mi bylo umožněno provést dotazníkový průzkum. V neposlední řadě bych chtěl poděkovat své přítelkyni a panu Radimu Badurovi za neobyčejnou psychickou podporu během tvorby mé práce.

Obsah

Úvod

Zpracování výzkumných dat pomocí statistických metod je jednou z obzvláště složitých oblastí statistiky, avšak je využitelné v nesmírně velkém množství vědních oborů od informačních technologií, přes medicínu až k fyzice, či chemii. Přes tuto širokou možnost využití však zůstává schopnost zpracovávat statistická data v dosti uzavřené společnosti statistiků a není příliš známá veřejnosti. V současné době již existuje velké množství statistických programů, proto si myslím, že by se tato schopnost mohla rozšířit i mezi veřejnost z jiných vědních oborů, než je matematika. Nejblíž k tomuto tématu má podle mého názoru informatika.

Hlavním důvodem výběru tohoto tématu byla velká spojitost s matematikou, kterou také studuji. Současně schopnost porozumět složitějším výstupům ze zpracovaných dat mi přijde jako velká výhoda při snaze učitele vyvinout nové a efektivnější učební metody. Povolání učitele, které bych chtěl v budoucnu vykonávat, se bez obnov učebních metod neobejde a myslím si, že touto schopností by bylo značně ulehčeno tomuto cíli.

Práci tvoří část teoretická a část praktická. V teoretické části jsem popsal jedny z nejvyužívanějších vícerozměrných statistických metod a principy, na kterých jsou tyto metody založeny a z kterých vychází. V praktické části jsem provedl dotazníkový výzkum v okrese Opava. Tématem tohoto výzkumu bylo podvádění žáků na základní škole a na těchto datech jsem následně aplikoval jednotlivé metody. Šlo mi především o možnost ukázat, jak vypadají výsledky daných metod v prostředí programu STATISTICA 13.3 EN a jak je možné výstupní data chápat a interpretovat.

Cíle bakalářské práce

Hlavním cílem bakalářské práce je představit postupy a možnosti využití vícerozměrných statistických metod při zpracovávání výzkumných dat.

Cílem teoretické části této práce je popsat jednotlivé nejpoužívanější vícerozměrné metody, aby čtenář pochopil, jak dané statistické metody fungují a jaké jsou principy zpracování dat za pomocí těchto metod. Současně by také po přečtení teoretické části mělo být jasné, v jakém formátu musí být vstupní data, aby mohly být dané metody použity.

Cílem praktické části bude názorně ukázat, jaká je možnost využití zpracování výzkumných dat v systému STATISTICA 13.3 EN a jak zpracovaná data chápat. Především půjde o popsání jednotlivých výstupních dat po aplikaci statistických metod a předvedení, jak je možné dané metody použít. Současně budeme porovnávat výsledky našeho výzkumu s výsledky práce, která byla pro náš výzkum předlohou. Výzkum bude proveden na téma podvádění žáků na vyšším stupni základních škol a bude realizován na základních školách v okrese Opava.

1 Teoretická část

V bakalářské práci budou popsány jen nejdůležitější a nejčastěji používané vícerozměrné statistické metody jako je shluková analýza, metoda hlavních komponent (někdy zařazovaná do faktorové analýzy) a klasická explorativní faktorová analýza.

1.1. Shluková analýza

1.1.1. Cíle shlukové analýzy

Hlavním cílem shlukové analýzy je podle Hebáka a kol. (2005) rozčlenění objektů vstupních dat do malého počtu skupin s podobnými proměnnými. Tato metoda je efektivní především v takových případech, kdy mají objekty sklony shlukovat se přirozeně. Sebera (2012b) definuje výstup faktorové analýzy jako dendogram[1], ve kterém jsou objekty s podobnými, či stejnými proměnnými ve shluku blízkém, nebo stejném. Naopak shluky s objekty o různých proměnných jsou od sebe vzdáleny.

V případě vícerozměrné analýzy máme vstupní data v matici X typu n × p, přičemž n určuje počet objektů a p je počet proměnných. Shlukovou analýzou se snažíme o rozložení množiny objektů *n* do *k* shluků, přičemž objekty v témže shluku by měly být maximálně podobné a objekty z jiných shluků by měly mít podobnost minimální (Hebák a kol. (2005).

Shlukování je také možné provádět pro jednotlivé vlastnosti objektů. Jsou- li si vlastnosti podobné, může je nahradit jediná vlastnost, jako je tomu například u Metody hlavních komponent, či Faktorové analýzy (viz kapitoly 1.2 a 1.3). Další možnosti využití shlukové analýzy viz Hebák a kol. (2005), Meloun a kol. (2005).

1.1.2. Míry podobnosti

Podobnost je u shlukové analýzy zásadní, protože podle ní určíme jednotlivé shluky objektů. Meloun a kol. (2005) udává, že prvním krokem je zvolit si vlastnost udávající podobnost, jež se následně sdružují do podobnostních měr. Následně je možné takto srovnávat objekty.

[1] stromový diagram znázorňující kroky shlukové analýzy

Míry podobnosti máme zpravidla tři. Jedná se o korelační míry, míry vzdálenosti a míry asociace. Každá skupina má specifický pohled na podobnost a závisí na typu vstupních dat. První dvě zmíněné skupiny jsou určeny především pro metrická data, míry asociace jsou vhodnější pro data nemetrická.

1.1.2.1. Korelační míry

Obecným měřítkem podobnosti objektů nebo proměnných je podle Melouna a kol. (2005) Pearsonův párový korelační koeficient r, či Spearmanův korelační koeficient. Oba tyto koeficienty ve své práci blíže specifikuje Meloun (2018). Pro oba tyto koeficienty platí, že čím větší je párový korelační koeficient, tím jsou si dva objekty podobnější. Tyto *mí* jsou však v praxi méně užívané.

1.1.2.2. Míry vzdálenosti

Pro kvantitativní data platí, že splňují- li vlastnosti symetrie[2] a trojúhelníkové nerovnosti[3], můžeme o nich mluvit jako o metrikách. Mezi nejznámější typy mír vzdálenosti patří euklidovská vzdálenost d_E, Čebyševova d_C a manhattanská d_H, avšak existuje jich daleko víc (Meloun a kol. (2005), Hebák a kol. (2005)).

$$d_E(x_i, x_{i'}) = \sqrt{\sum_{j=1}^{p}(x_{ij}, x_{i'j})^2}, \quad (1.1)$$

$$d_C(\mathrm{x_i}, \mathrm{x_{i'}}) = \mathrm{max_j}\left|\mathrm{x_{ij}} - \mathrm{x_{i'j}}\right|, \quad (1.2)$$

$$d_H(x_i, x_{i'}) = \sum_{j=1}^{p} |x_{ij} - x_{i'j}|. \quad (1.3)$$

Euklidovská vzdálenost v podstatě měří vzdálenost dvou bodů tak, jak ji známe z geometrie. Manhhatanská vzdálenost měří vzdálenost tak, jako bychom se pohybovali vždy pouze vodorovně, nebo jen svisle. V podstatě měří vzdálenost tím stylem, jakým se v šachu pohybuje figurka věže. U Čebyševovy vzdálenosti měříme rozdíl souřadnic

---

[2] jsou li data v matici stejnolehlá podle diagonály; d(x,y) = d(y,x)

[3] součet velikostí dvou stran trojúhelníka je vždy větší než velikost třetí strany; d(x,y) ≤ d(x,z) + d(y,z)

pouze v jednom směru (odečteme tedy pouze hodnoty na ose x). V následující praktické části, bude použito pouze měření euklidovské vzdálenosti.

1.1.2.3. Míry asociace

Pro binární data (nemetrického charakteru) se používají míry asociace. Na příkladu, ve kterém budeme hodnotit míru souhlasu dvou objektů, si nejlépe ukážeme míru asociace. V kontingenční tabulce bude hodnota „1" udávat souhlas a hodnota „0" nesouhlas (Meloun a kol. (2005).

Kontingenční tabulka			
		Objekt O_i	
		1	0
Objekt O_j	1	*a*	*b*
	0	*c*	*d*

Tabulka 1: Kontingenční tabulka souhlasného koeficientu

Jak můžeme vidět v tabulce, popsali jsme si všechny možnosti, které nám mohou nastat. Znak *a* nám znázorňuje stav, kdy oba objekty zvolily hodnotu „1", což je tzv. pozitivní shoda. Znaky *b* a *c* udávají stavy, kdy jeden z objektů volil souhlas a druhý nesouhlas. Znak d uvádí možnost, kdy oba objekty mají hodnotu „0". V tomto případě mluvíme o tzv. negativní shodě. Pomocí takto určených znaků vypočítáme koeficienty podobnosti, mezi které řadíme koeficient jednoduché shody S_{SM}, Jaccardův koeficient S_J, Diceův S_D a jiné (Meloun a kol., 2005).

$$S_{SM} = \frac{a+d}{a+b+c+d}, \quad (1.4)$$

$$S_J = \frac{a}{a+b+c}, \quad (1.5)$$

$$S_D = \frac{2a}{2a+b+c}. \quad (1.6)$$

1.1.3. Způsoby shlukování

1.1.3.1. Hierarchické shlukování

Podstatou tohoto shlukování je postupné seřazení shluků, či objektů podle podobnosti. Shlukování podle hierarchie má podle směru přístupu dva postupy. Jedná se o aglomerační shlukování a divizní shlukování – např. Meloun a kol. (2005).

Aglomerační shlukování je založeno na tvorbě nových shluků na základě největší podobnosti. Najdeme dva objekty s největší podobností a vytvoříme z nich shluk. Následně vytvoříme novou matici dat s původními objekty, mimo nově shluklé objekty. Takto se postupuje, dokud není dosaženo požadovaného počtu shluků. Divizní shlukování má přesně opačný postup. V prvním kroku si všechny objekty spojíme do jediného shluku a následně v každém dalším kroku shluk dělíme, až dojdeme do fáze, kdy máme jednotlivé objekty samostatné (Meloun a kol., 2005), Kučera (2018).

Hierarchické shlukování má tu výhodu, že nepotřebujeme znát ideální počet shluků. Ten si určujeme až během shlukování. Nevýhodou ovšem je vyjádření podobnosti a problém při výběru vhodné shlukovací procedury (Meloun a kol., 2005), Hebák a kol. (2005), Kučera (2018).

1. *Metoda nebližšího souseda*

   Tato metoda se také nazývá *metodou jednoduchého spojení*. Její princip za založený na spojování dvou nejbližších objektů. Ty následně vytvoří shluk. V dalším kroku k nově vzniklému shluku připojíme třetí objekt, který má nejmenší vzdálenost. V podstatě bereme dva nejbližší objekty ze dvou shluků. Tento typ shlukování není vždy nejvýhodnější, protože se často i velmi vzdálené objekty mohou potkat ve stejném shluku.

2. *Metoda nejvzdálenějšího souseda*

   Jedná se o podobnou metodu, jako byla předchozí, avšak s tím rozdílem, že nyní spojujeme shluky pomocí jejich nejvzdálenějších objektů. Metoda je tedy postavena na maximální vzdálenosti shluků. Této metodě se též říká *metoda úplného propojení*. Nevýhoda výše zmíněné metody v tomto případě odpadá.

3. *Metoda průměrné vzdálenosti*

Vzdálenost dvou shluků je v případě této metody vypočtena jako průměrná vzdálenost všech objektů jednoho shluku ke všem objektům shluku druhého. Tato metoda nebere v potaz extrémní hodnoty (nejbližší nebo nejvzdálenější objekt), ale průměrné, proto je vznik shluku závislý na všech objektech.

4. *Wardova metoda*

Měřítkem této metody již není vzdálenost shluků, ale součet druhých mocnin odchylek jednotlivých objektů od těžiště shluku, do kterého spadají. Dojde ke sloučení těch dvou shluků, které mají daný součet nejmenší.

5. *Metoda těžiště*

Princip této metody je postaven na velikosti euklidovské vzdálenosti těžišť dvou shluků. Těžiště shluku se určí pomocí průměrů hodnot pro jednotlivé znaky. Po každém kroku shlukování se těžiště počítá znovu. Velkou výhodou této metody oproti ostatním je minimální ovlivnitelnost samostatnými body.

6. *Metoda mediánová*

Tato metoda je jakousi obdobou předchozí metody. Rozdíl je ve snaze zanedbat rozdílnou velikost shluků. Ta totiž silně ovlivňuje umístění těžiště, a tudíž i vzdálenosti shluků.

1.1.3.2. Nehierarchické shlukování

Nehierarchické shlukování na rozdíl od hierarchického vstupní množinu dat dělí do podmnožin. Tento typ shlukování hledá optimální rozklad již nějakého předem existujícího rozkladu. Tento druh shlukování funguje na základě těchto tří postupů (Meloun a kol., 2005):

a) *Sekvenční práh*

Metoda funguje na základě volby jednoho prvotního shluku, do kterého jsou zařazeny všechny objekty v předem definované vzdálenosti. Následně je vybrán další prvotní shluk a do něj přiřazeny objekty z dané vzdálenosti. Objekty, které jsou již součástí některého ze shluků, již následně nejsou shlukovány. Tento proces se po té pořád opakuje. Metoda je výhodná při zpracovávání objemných datových souborů.

b) *Paralelní práh*

Rozdíl v této metodě nastává ihned na začátku. Prvotní shluky jsou zvoleny na začátku a objekty ze zvolené vzdálenosti jsou současně přidělovány do prvotních shluků. Tato vzdálenost se během procesu shlukování může změnit, aby přiřadila více, či méně objektů do shluku, avšak některé objekty mohou zůstat nepřiřazeny do shluku kvůli velké vzdálenosti.

c) *Optimalizace*

Metoda funguje obdobně jako obě předešlé, avšak objekty, které již byly zařazeny do nějakého shluku, mohou být zařazeny do jiného, pokud jsou k němu blíže.

Při používání nehierarchických metod je nutné znát předem počet shluků, i když se jejich počet může během procedury změnit. Dojde- li ke změně počtu prvotních shluků, mluvíme o tzv. nehierarchických metodách s optimalizovaným počtem shluků. V opačném případě mluvíme o nehierarchických metodách s konstantním počtem shluků. Prvotním shlukem je následně vždy první úplný objekt v matici vstupních dat. Dalším prvotním shlukem je také úplný objekt, avšak musí být za předem definovanou vzdáleností. Po vzniku všech prvotních shluků začne program, ve kterém data zpracováváme tvořit opravdové shluky. Hlavním problémem nehierarchických metod je zvolení prvotních shluků. Změníme- li totiž pořadí objektů v datové matici, můžeme dostat velmi rozdílné výsledky. Z tohoto důvodu se často doporučuje sestavovat datovou matici náhodně (Meloun a kol., 2005).

Nehierarchické shlukování sestává ze tří druhů principů shlukování (Kučera, 2018). Prvním principem je *metoda nejbližších těžišť*, též zvaná *K- Means*. Jak již naznačuje název, princip je založen na vzdálenosti těžišť jednotlivých shluků. Metoda vytvoří pouze jedno řešení pro zadaný počet shluků.

Dalším z principů je *shlukování metodou optimálních středů*. V tomto případě je optimální střed shluku takový objekt, který má minimální průměrnou vzdálenost od ostatních objektů v daném shluku. Počet optimálních středů je shodný s počtem shluků.

Posledním z principů je fuzzy shlukování, které je výjimečné tím, že objekty mohou být zařazeny do více shluků. Tato skutečnost je hlavní výhodou, avšak nevýhodou je velký nárůst informací, které se musí vysvětlit. Velmi obtížné je také ve fuzzy shlukování nalezení správného počtu shluků.

Všechny tři principy využívají ještě další metody. Bohužel je však nemožné je všechny vysvětlovat zde a ani to není náplní této práce. Do větší hloubky se jednotlivými principy zabývají Hebák a kol. (2005), Meloun a kol. (2005) a také Kučera (2018).

1.2. Analýza hlavních komponent

1.2.1. Podstata metody hlavních komponent

Hlavním cílem metody hlavních komponent spočívá především ve zjednodušení popisu skupiny vzájemně korelovaných znaků. Jde tedy o zmenšení počtu proměnných, kterými jsou popsány vlastnosti objektu, ale nesmí dojít ke ztrátě informací. Je tedy zřejmé, že tato metoda se hodí především pro zpracování velmi objemných dat a může předcházet aplikaci jiných vícerozměrných statistických metod (např. faktorová analýza, shluková analýza) - viz (Sebera, 2012a).

Principem analýzy hlavních komponent je tedy nahradit nějaké množství proměnných výrazně menším počtem skrytých proměnných, které lépe vystihují skoro celou proměnlivost původních znaků a vzájemně se neovlivňují. Tyto nové proměnné jsou ony hlavní komponenty a jedná se o lineární kombinace původních neznámých. První hlavní komponenta nahrazuje největší část rozptylu původních dat, druhá hlavní komponenta nahrazuje největší část rozptylu původních dat, která nejsou obsažena v první hlavní komponentě. Hlavní komponenty jsou tedy seřazeny od největšího rozptylu po nejmenší (Meloun a kol., 2005).

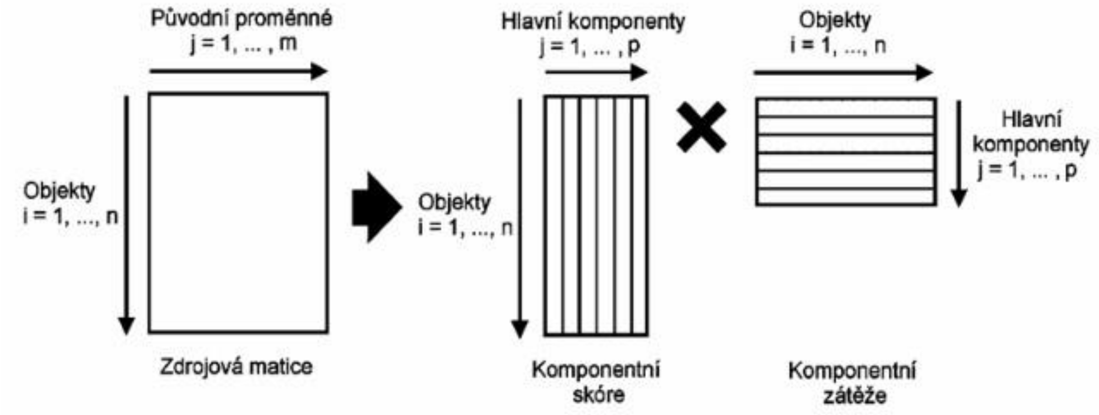
Protože analýza hlavních komponent dosti souvisí i s jinými vícerozměrnými metodami a v mnoha statistických programech je začleněna přímo do těchto metod, nebudu se jí v mé práci tolik věnovat, pouze nastíním její princip.

1.2.2. Cíl metody hlavních komponent

Jako u všech vícerozměrných dat je vstupem zdrojová matice **X** o *n* řádcích (objektech) a *m* sloupcích (znacích). Podstatné je, aby určitý znak byl známý u všech objektů pozorování. Správné složení matice X je velmi závislé na povaze úlohy, pro kterou analýzu hlavních komponent děláme (Meloun a kol., 2005).

Přes velké množství pozorovaných znaků *m* by množství hlavních komponent mělo být dosti malé. Nedílnou součástí analýzy hlavních komponent je vysvětlení užitných komponent, vysvětlení vztahů k původním znakům a jejich pojmenování.

Metoda hlavních komponent tedy nahrazuje vstupní datovou matici **X** její mírně upravenou a odhadnutou maticí (viz obr. 1). Aproximace má řadu výhod v interpretaci dat. Nejde zde jen o změnu uspořádání souřadnic, ale hlavně o nalezení a vypuštění šumu. Cílem analýzy hlavních komponent je tedy transformace do nového uspořádání os a zmenšení objemu dat v úloze vytvořením několika prvních hlavních komponent (Meloun a kol., 2005).



Obrázek č. 1: Schéma maticových výpočtů v analýze hlavních komponent - převzato z (Meloun a kol., 2005, s. 61)

1.2.2.1. Počet hlavních komponent

Minimální počet hlavních komponent není nijak nastaven, ale měl by být takový, aby nám řešení dávalo smysl. Naopak maximální počet hlavních komponent je určen jako počet objektů mínus jedna (n-1), nebo počet znaků (m) v závislosti na tom, které z daných čísel je menší. Účinný počet hlavních komponent je rovný hodnosti zdrojové matice **X**[4] (Meloun a kol., (2005).

1.3. Faktorová analýza

Faktorová analýza je podobně jako analýza hlavních komponent zaměřena na vytvoření nových proměnných a redukci velikosti dat, přičemž se snaží o minimální ztrátu dat. Tato metoda zkoumá vztah proměnných a zjišťuje, zda je možné je rozdělit do skupin, kde by proměnné více vzájemně souvisely. Rovněž se snaží vytvořit nové proměnné (faktory) ve snaze lépe interpretovat vstupní data (Sebera, 2012a), Meloun a kol. (2012).

Na rozdíl od metody hlavních komponent se při faktorové analýze snažíme vysvětlit závislost znaků. Nevýhodou faktorové analýzy je potřeba zvolit si množství sdílených faktorů před začátkem vlastní analýzy (Meloun a kol., (2012).

1.3.1. Podstata faktorové analýzy

Vstupní matice musí být standardizovaná, čehož se dosáhne pomocí vztahu (3.1).

$$x_{ij} = (x_{ij} - \bar{x}_j)/s_j \tag{3.1}$$

„X_{ij} jsou prvky původní zdrojové matice a vzhledem ke standardizaci mají střední hodnotu 0 a rozptyl roven 1. Předpokládejme, že x_i^T = $(x_1, x_2, \ldots, x_m)^T$ je obecný i-tý objekt pro znaky s korelační maticí R. Tento objekt je vyjádřen vztahy (Meloun a kol., 2005, s. 97):

$$x_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1p}F_p + e_1 \,,$$

$$x_2 = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2p}F_p + e_2 \,,$$

[4] Počet nezávislých sloupců nebo řádků v matici

$$\vdots$$

$$x_m = a_{m1}F_1 + a_{m2}F_2 + \ldots + a_{mp}F_p + e_m,$$

kde $F_1, F_2, \ldots, F_p$ je p vybraných společných faktorů, které vyvolávají korelace mezi m původními znaky.“

Střední hodnota faktorů je rovna nule a $e_1, e_2, \ldots, e_m$ jsou chybové faktory.

Ve tvaru matice by vztah vypadal takto:

$$X = F\,A^T + E^*. \tag{3.2}$$

Faktorová analýza je obdobná jako metoda hlavních komponent, avšak je mírně rozdílná. Počet hlavních komponent je stejný jako počet vstupních znaků, i když takový to počet komponent se využívá jen málokdy. Počet faktorů je naopak již od začátku menší než počet vstupních znaků (Meloun a kol., 2005).

1.3.2. Faktorová rotace

Principem rotace je vyvození lehce vysvětlitelných a pojmenovatelných společných faktorů. Rotace transformuje původní faktory tak, aby byly lépe srozumitelné a vysvětlitelné. V ideálním případě máme zvolené zátěže blízko hodnot *1* nebo *0*, čímž bude korelační koeficient původního znaku a faktoru buď vysoký, nebo nízký. Faktorová rotace má především dvě známé rotace (Meloun a kol., 2005), Sebera (2012a).

a) *Quartimax*

Tato metoda minimalizuje počet faktorů potřebných k popisu původních znaků a vytváří obecný faktor. Kritériem tétu metody je součet čtvrtých mocnin faktorových zátěží.

b) *Varimax*

V této metodě jde o snahu co nejvíce zmenšit množství znaků, které mají vysokou zátěž faktoru. Jedná se o nejpoužívanější rotaci.

1.3.3. Grafické pomůcky faktorové analýzy

Pro snadnější zjištění vnitřní struktury, je lepší použít grafickou diagnostiku. Jedná se o názornější metodu, než hledání hodnot v tabulce.

1.3.3.1. Graf faktorových zátěží

Jde o 2D nebo 3D- graf, který nám znázorní úspěšnost otočení faktorů. Na osách grafu jsou pak jednotlivé faktory. Dosáhli- li jsme při rotaci jednoduché struktury, v grafu se nám zobrazí shluky znaků. Znaky, které jsou vzdáleny od souřadnicových os, se nazývají *znaky faktorově nečisté*. Pokud je takových to znaků v grafu málo, jedná se o jednoduchou strukturu (Meloun a kol., 2005).

1.3.3.2. Graf faktorového skóre

Chceme- li si určit výjimečné objekty, je vhodné si vypočíst faktorové skóre pro některé faktory a následně z grafu faktorového skóre je určit. Faktorové skóre se vypočítá podle vztahu:

$$\hat{F}_{jk} = \sum_{i=1}^{m} w_{ji}\ x_{ik}\ , \qquad (3.3)$$

kde x_{ij} je standardizovaná hodnota pro *i*-tý znak *k*-tého objektu. W_{ij} udává faktorové skóre *j*-tého faktoru pro *i*-tý znak. Proměnná *m* značí celkový počet všech původních znaků (Meloun a kol., 2005).

1.3.4. Etapy faktorové analýzy

Při aplikaci faktorové analýzy bychom měli postupovat podle následujících čtyř kroků. Jejich dodržením se zvyšuje šance na úspěšné použití této metody (Sebera (2012a).

1. Zvolíme si počet faktorů (například metodou hlavních komponent).
2. Určíme faktorové zátěže mezi původními proměnnými a faktory.
3. Pomocí rotací se snažíme dojít k nejlepšímu podání faktorů.
4. Určíme si faktorové skóre pro všechny objekty.

1.3.5. Předpoklady faktorové analýzy

Důležitým předpokladem aplikace faktorové analýzy jsou vzájemné vztahy mezi jednotlivými znaky. Pokud by tato korelace neexistovala, nebylo by možné vytvořit společné faktory. Ke zjištění, zda tyto vztahy existují, nám slouží (Meloun a kol., 2005):

a) *Bartlettův test sféričnosti*

Jedná se o statický korelační test původních znaků. Tento test zkoumá existenci vzájemných vztahů mezi znaky.

b) *Parciální korelační koeficienty*

Tyto hodnoty nám vyjadřují, jak silný je vztah mezi znaky. Pokud jsou koeficienty malé, znaky mají společný faktor. V případě, kdy jsou koeficienty velké, měli bychom zvážit, zda je vhodné faktorovou analýzu vůbec aplikovat.

c) *Anti- image korelační matice*

Po provedení faktorové analýzy si vytvoříme matici parciálních korelačních koeficientů. Ta nám udává, jak moc je povaha znaků vysvětlena jednotlivými faktory.

d) Kaiserova- Meyerova- Olkinova (KMO) míra

Tato míra nám porovnává velikosti experimentální korelační koeficienty a parciální korelační koeficienty podle vzorce:

$$KMO = \frac{\sum_{i \neq j}^{m} \sum_{j \neq i}^{m} r_{ij}^2}{\sum_{i \neq j}^{m} \sum_{j \neq i}^{m} r_{ij}^2 + \sum_{i \neq j}^{m} \sum_{j \neq i}^{m} a_{ij}^2} \qquad (3.4)$$

kde r_{ij} je párový korelační koeficient v rozmezí *i*- tého a *j*- tého znaku. A_{ij} označuje korelační koeficient v témže rozmezí. Hodnota KMO se pohybuje mezi 0 a 1. Pokud jsou hodnoty *KMO* v rozmezí *0 – 0,5* je vztah mezi znaky nepřijatelná a aplikace faktorové analýzy by byla neúčinná. Při hodnotách do *0,6* je korelace slabá, do hodnoty *0,7* je korelace střední, do hodnoty *0,8* se jedná o průměrnou korelaci a do hodnoty *0,9* jde o dobrou korelaci. Je- li hodnota do *1*, korelace je výborná a faktorová analýza bude mít efektivní výsledky.

e) *Čtverec vícenásobného korelačního koeficientu R^2*

Jde o další měřítko vzájemných vztahů mezi znaky. Malá hodnota R2 udává, že tento znak by měl být z původního souboru odstraněn. Podmínkou tohoto kritéria je homogenita dat. Například nemůžeme mít ve vstupních datech obě pohlaví, pokud by to mohlo znaky silně ovlivnit.

1.3.6. Interpretace výsledků

Jedním z výstupů faktorové analýzy je *faktorová zátěž*. Ta nám udává, jak velký je vzájemný vztah mezi daným faktorem a původním znakem. Při hledání nejlepšího výsledku musíme často zvážit, zda se určitý faktor vůbec vyplatí zvolit (Meloun a kol., 2005).

1.3.6.1. Praktický význam

Má- li faktorová zátěž hodnotu větší než ±0,30 jedná se o minimální přijatelnou hladinu. Důležitější jsou zátěže větší než ±0,40 a teprve zátěže větší než ±0,50 jsou považovány za významné. Hodnota faktorové zátěže umocněná na druhou nám udává počet procent rozptylu původního znaku, který faktor vysvětluje. V podstatě zátěž 0,30 nám udává $0{,}30^2 = 0{,}09$. Faktor nám tedy vysvětlí přibližně 10 % rozptylu původního znaku. Proto, aby faktor vysvětlil alespoň 50 % rozptylu původního znaku, je potřeba mít faktorovou zátěž alespoň ±0,70 ($0{,}70^2 = 0{,}49$) – viz (Meloun a kol., 2005).

Hodnoty, které by byly větší, než ±0,8 nejsou příliš časté, avšak pokud jich dosáhneme, jsou ideální. Tyto poznatky mají především praktický význam a platí především pro úlohy, kde jsme měli více než 100 objektů. Při různém množství objektů mohou být hodnoty mírně modifikovány (Meloun a kol., 2005).

1.3.6.2. Statistická významnost

Faktorové zátěže mívají větší standardní chybu než korelace. Z tohoto důvodu je kritérium jejich posuzování mnohem přísnější. Například pro 120 objektů je odpovídající zátěž 0,5. Při množství 85 objektů, je významná zátěž 0,60. Pro výpočet hladiny významnosti se užívá stejný test, jakým je možno vypočítat významnost korelačního koeficientu (Meloun a kol. (2005)).

2 Praktická část – analýza výzkumu podvádění studentů na základních školách

Pro praktickou část své práce, ve které budou představeny možnosti zpracování výzkumných dat pomocí vícerozměrných statistických metod pomocí systému STATISTICA 13.3 EN, jsem si vybral téma podvádění žáků na základních školách. Tímto tématem se ve své disertační práci již zabývala také Vrbová (2013), proto jsem převzal její dotazník a ten jsem mírně modifikoval. Dotazník se skládal ze dvou částí. V první části byly otázky, které zkoumaly příčiny podvádění (příloha 1) a ve druhé části otázky na jednotlivé druhy podvádění (příloha 2). Dotazník byl ještě doplněn o volbu pohlaví, navštěvované třídy a základní školu, kterou žák navštěvuje. Celkově měl tedy formulář 68 otázek. Do první části spadaly otázky č. 1 - 47 a žáci měli možnost odpovídat na ně mírou souhlasu na škále od 1 do 4. Otázky 48 - 65 souvisely s druhou částí a respondenti zde vybírali odpovědi v míře četnosti výskytu ve škále od 1 do 5. Zbylé 3 otázky byly doplňující. Dotazník byl plně anonymní a pokládal jsem jej v elektronické podobě žákům 6. – 8. ročníku na dvou Základních školách. Celkem dotazník vyplnilo 185 respondentů, přičemž 93 dotazovaných bylo chlapců a 92 dívek. Žáků 6. ročníků vyplnilo dotazník 73, žáků 7. ročníků 58 a žáků navštěvujících 8. ročník bylo 54.

2.1. Cíle praktické části

Náplní této části bude ukázat možnou práci s daty v programu STATISTICA pomocí vícerozměrných statistických metod a především vyhodnocení výstupních dat při aplikaci vícerozměrných metod. Bude se jednat především o Shlukovou analýzu a Faktorovou analýzu. Po přečtení této části by měl být čtenář schopný porozumět výstupům shlukové a faktorové analýzy v programu STATISTICA. Ke zpracování dat jsem využil verzi STATISTICA 13.3 EN.

2.2. Užití shlukové analýzy

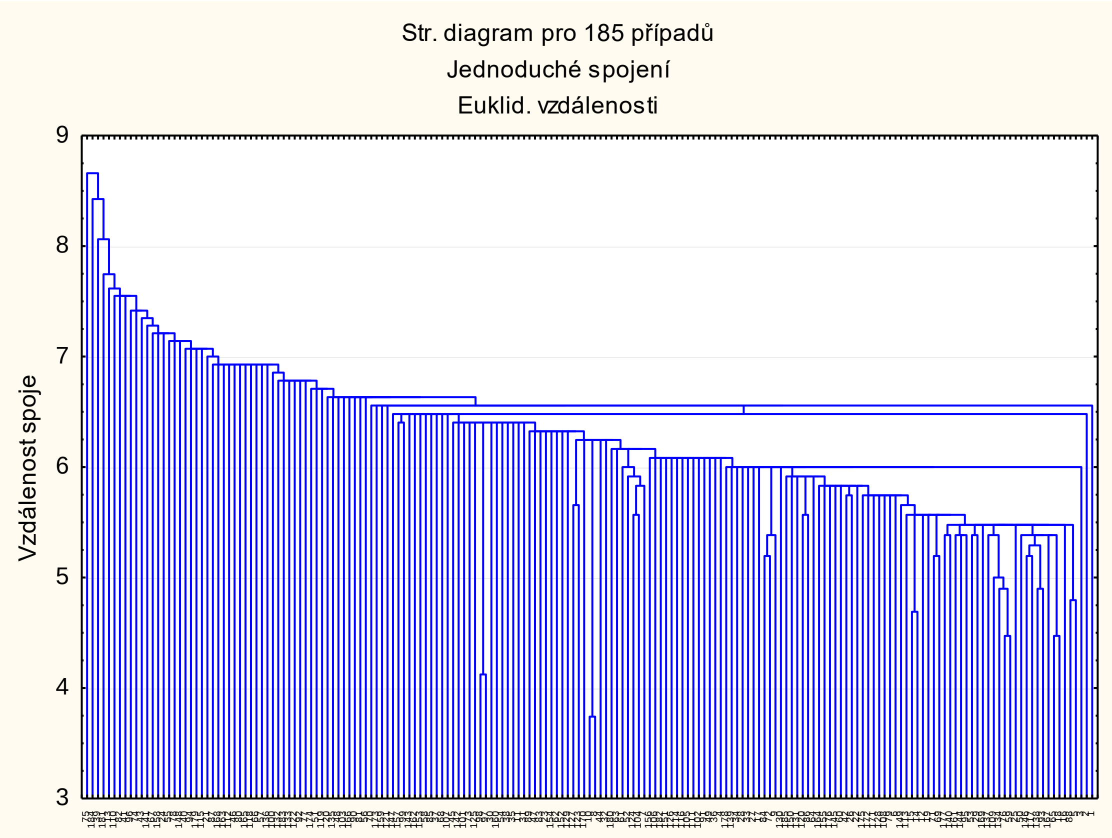
Pro názorné ukázky práce se shlukovou analýzou budu využívat pouze první část dotazníku obohacenou o údaje o pohlaví tázaného žáka. Nejdříve představím možnost hierarchického shlukování pro shlukování objektů, následně nehierarchického pro shlukování objektů výzkumu.

2.2.1. Hierarchická shluková analýza

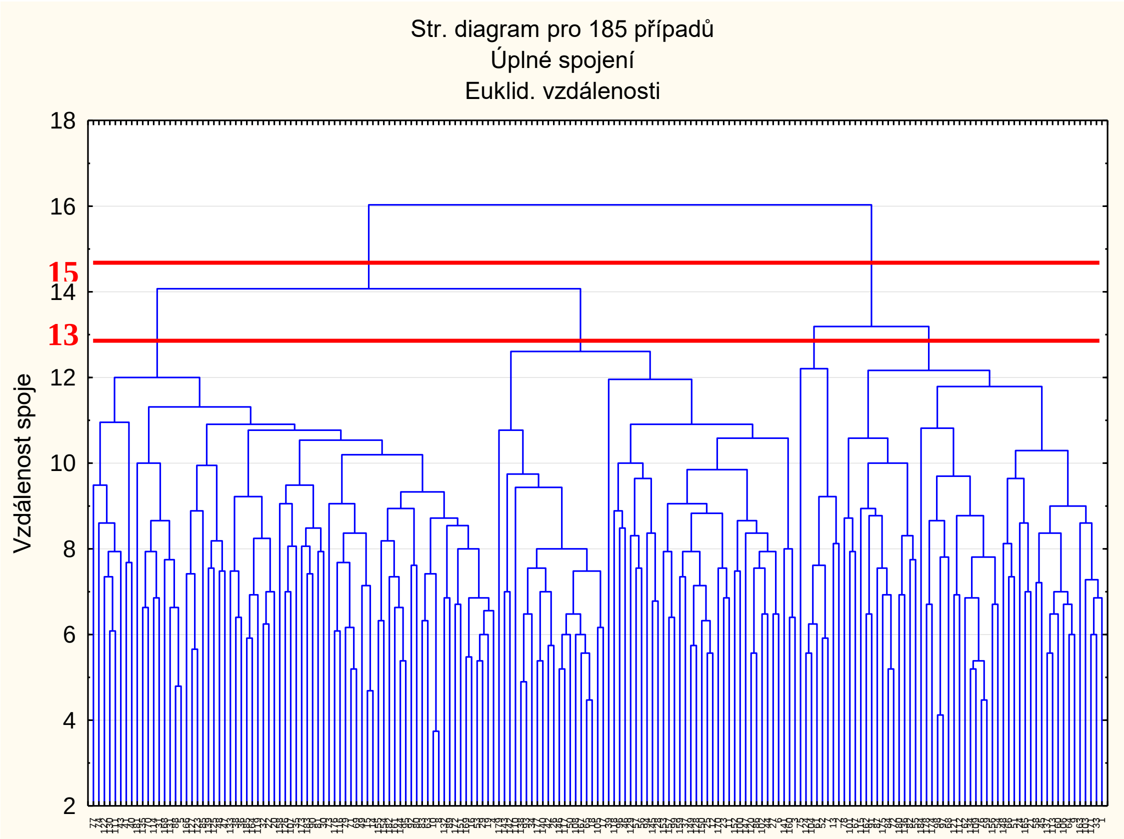
V prvním případě si ukážeme stromový graf hierarchického jednoduchého spojení za použití měření euklidovských vzdáleností.

Výstupem shlukové analýzy je v tomto případě dendogram. Na ose y máme hodnoty vzdálenosti, které nám říkají, v jaké vzdálenosti se jednotlivé objekty spojily. Na ose x jsou jednotlivé objekty seřazeny tak, aby odpovídaly dendogramu. Jak je na první pohled vidět na obrázku č. 2, tento typ shlukování nám o možných shlucích v množině znaků moc nenapoví, ačkoli vidíme, že ve vzdálenosti cca 8,5 můžeme vidět existenci dvou shluků. Tento typ shlukování se nepoužívá příliš často, i přes to však může být někdy užitečný.

Protože nám toho použití úplného spojení neřeklo, použijeme tedy opět euklidovskou metriku, ale nyní zvolíme úplné spojení. Z obrázku č. 3 již vidíme větší počet shluků. Například ve vzdálenosti 15 vidíme dva shluky, kdežto bude-li pro nás podstatná vzdálenost 13, vidíme, že zde jsou již 4 shluky. Tato vzdálenost není až tak velká, proto se dá předpokládat, že odpovědi jednotlivých objektů výzkumu se nelišily v takové míře. Můžeme tedy říct, že ve vzdálenosti 13 se respondenti dotazníků podle podobností svých odpovědí dělí do 4 shluků.

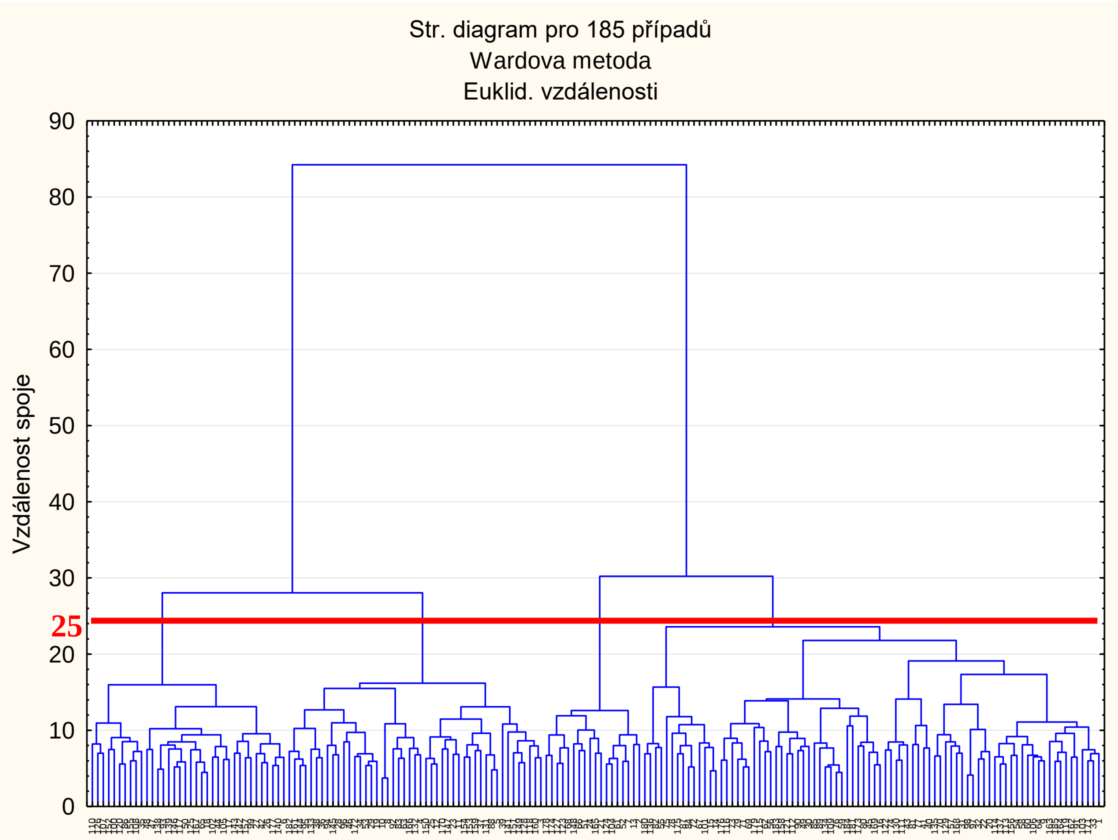


Obrázek č. 2: Shluková analýza - jednoduché spojení



Obrázek č. 3: Shluková analýza – úplné spojení

Ve třetím případě opět ponechám euklidovskou metriku a aplikuji Wardovu metodu shlukování. Na první pohled je vidět, že existují 2 diametrálně rozdílné shluky, protože k jejich spojení dochází až ve vzdálenosti přibližně 85 (viz obrázek č. 4). Ve vzdálenosti 25 vidíme, že jsou zde čtyři shluky, které ovšem nejsou stejného charakteru. V případě čtvrtého shluku, který se nachází nejvíce vpravo, totiž vidíme, že oproti ostatním třem shlukům je tvořen mnohem větším počtem shluků. Tato skutečnost je opodstatněna tím, že v tomto shluku existují objekty, které mají shodnější odpovědi na některé otázky.



Obrázek č. 4: Shluková analýza – Wardova metoda

Dalším částí výstupu shlukové analýzy je tabulka vzdáleností objektů. I pro tento případ jsem ponechal asi nejpoužívanější euklidovskou metriku. Výřez z tabulky vzdáleností (obrázek č. 5) nám poskytuje možnost vidět, jak jsou od sebe jednotlivé objekty vzdáleny. Vzdálenosti se pohybují mezi hodnotami 3,7 a 14. Hodnotu 3,7 mají objekty 8 a 10, což znamená, že tito dva žáci odpovídali nejpodobněji. Naopak vzdálenost 14 mají žáci s pořadovými čísly třináct a dvanáct. Ti naopak odpovídali nejrozdílněji.

Případ	Euklid. vzdálenosti (DATA Chovanec)																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0,0	9,7	11,0	7,7	8,8	9,9	9,2	9,4	7,8	9,8	8,8	10,9	10,4	7,9	7,9	7,9	8,7	8,5	8,9	10,1	7,9	8,0
2	9,7	0,0	13,3	11,3	11,5	12,2	6,5	11,2	8,9	10,9	11,9	12,0	8,1	10,1	10,4	10,1	11,1	10,3	10,8	10,7	8,2	10,8
3	11,0	13,3	0,0	7,7	8,2	7,9	11,5	7,6	10,9	8,2	7,4	7,5	13,7	9,8	8,8	8,4	8,0	8,3	7,1	9,7	11,7	7,4
4	7,7	11,3	7,7	0,0	7,3	8,5	9,4	7,6	8,6	7,2	7,3	7,9	12,8	8,7	8,2	6,9	6,6	7,1	6,2	8,9	9,7	6,6
5	8,8	11,5	8,2	7,3	0,0	7,7	10,2	8,9	8,6	9,2	7,3	7,4	13,0	8,2	7,5	7,4	7,6	6,9	6,7	8,2	9,7	7,0
6	9,9	12,2	7,9	8,5	7,7	0,0	10,2	8,7	10,1	8,7	8,2	6,8	13,2	9,2	9,2	7,2	7,0	6,6	9,2	9,4	9,3	9,1
7	9,2	6,5	11,5	9,4	10,2	10,2	0,0	10,2	6,9	9,7	11,0	10,6	7,6	9,2	9,2	8,4	9,3	8,0	8,6	9,3	8,5	9,5
8	9,4	11,2	7,6	7,6	8,9	8,7	10,2	0,0	9,1	3,7	7,1	8,8	11,8	8,8	8,5	7,1	7,9	7,5	7,4	8,8	9,9	8,2
9	7,8	8,9	10,9	8,6	8,6	10,1	6,9	9,1	0,0	9,2	9,2	10,3	9,5	6,6	7,5	8,4	8,9	8,9	8,1	9,9	10,1	8,5
10	9,8	10,9	8,2	7,2	9,2	8,7	9,7	3,7	9,2	0,0	8,2	8,9	11,7	9,2	9,0	7,4	7,9	7,3	7,1	8,5	10,0	8,3
11	8,8	11,9	7,4	7,3	7,3	8,2	11,0	7,1	9,2	8,2	0,0	7,3	13,2	8,4	7,9	7,3	8,2	7,9	7,6	9,7	9,6	7,6
12	10,9	12,0	7,5	7,9	7,4	6,8	10,6	8,8	10,3	8,9	7,3	0,0	14,0	9,2	9,3	8,0	7,7	6,0	7,6	8,3	10,1	8,4
13	10,4	8,1	13,7	12,8	13,0	13,2	7,6	11,8	9,5	11,7	13,2	14,0	0,0	9,8	10,0	10,7	12,1	11,2	11,3	12,1	10,1	11,4
14	7,9	10,1	9,8	8,7	8,2	9,2	9,2	8,8	6,6	9,2	8,4	9,2	9,8	0,0	4,7	7,9	8,7	8,1	8,1	10,6	9,7	8,9
15	7,9	10,4	8,8	8,2	7,5	9,2	9,2	8,5	7,5	9,0	7,9	9,3	10,0	4,7	0,0	7,6	8,8	8,2	7,1	9,6	9,8	8,5
16	7,9	10,1	8,4	6,9	7,4	7,2	8,4	7,1	8,4	7,4	7,3	8,0	10,7	7,9	7,6	0,0	7,4	5,5	6,5	8,3	7,7	6,9
17	8,7	11,1	8,0	6,6	7,6	7,0	9,3	7,9	8,9	7,9	8,2	7,7	12,1	8,7	8,8	7,4	0,0	6,6	7,3	7,7	9,0	7,9
18	8,5	10,3	8,3	7,1	6,9	6,6	8,0	7,5	8,9	7,3	7,9	6,0	11,2	8,1	8,2	5,5	6,6	0,0	6,9	6,9	7,1	7,2
19	8,9	10,8	7,1	6,2	6,7	9,2	8,6	7,4	8,1	7,1	7,6	7,6	11,3	8,1	7,1	6,5	7,3	6,9	0,0	7,4	10,2	6,3
20	10,1	10,7	9,7	8,9	8,2	9,4	9,3	8,8	9,9	8,5	9,7	8,3	12,1	10,6	9,6	8,3	7,7	6,9	7,4	0,0	9,7	7,0
21	7,9	8,2	11,7	9,7	9,7	9,3	8,5	9,9	10,1	10,0	9,6	10,1	10,1	9,7	9,8	7,7	9,0	7,1	10,2	9,7	0,0	9,9
22	8,0	10,8	7,4	6,6	7,0	9,1	9,5	8,2	8,5	8,3	7,6	8,4	11,4	8,9	8,5	6,9	7,9	7,2	6,3	7,0	9,9	0,0
23	9,1	10,1	8,7	7,1	7,9	9,8	10,0	8,5	9,1	8,4	6,9	7,7	12,4	9,2	8,8	8,1	8,4	7,5	7,4	8,1	9,1	7,4
24	8,4	9,2	11,7	9,7	9,4	11,4	9,1	9,7	8,4	10,2	9,1	10,2	11,6	8,9	8,9	8,9	10,5	8,8	8,8	9,7	8,3	8,5
25	8,5	10,8	7,2	6,6	6,8	6,9	9,2	7,5	8,9	7,5	6,4	5,6	12,5	8,7	7,4	6,7	6,0	5,9	6,2	6,6	8,7	7,1

Obrázek č. 5: Shluková analýza – hierarchické shlukování – euklidovské vzdálenosti

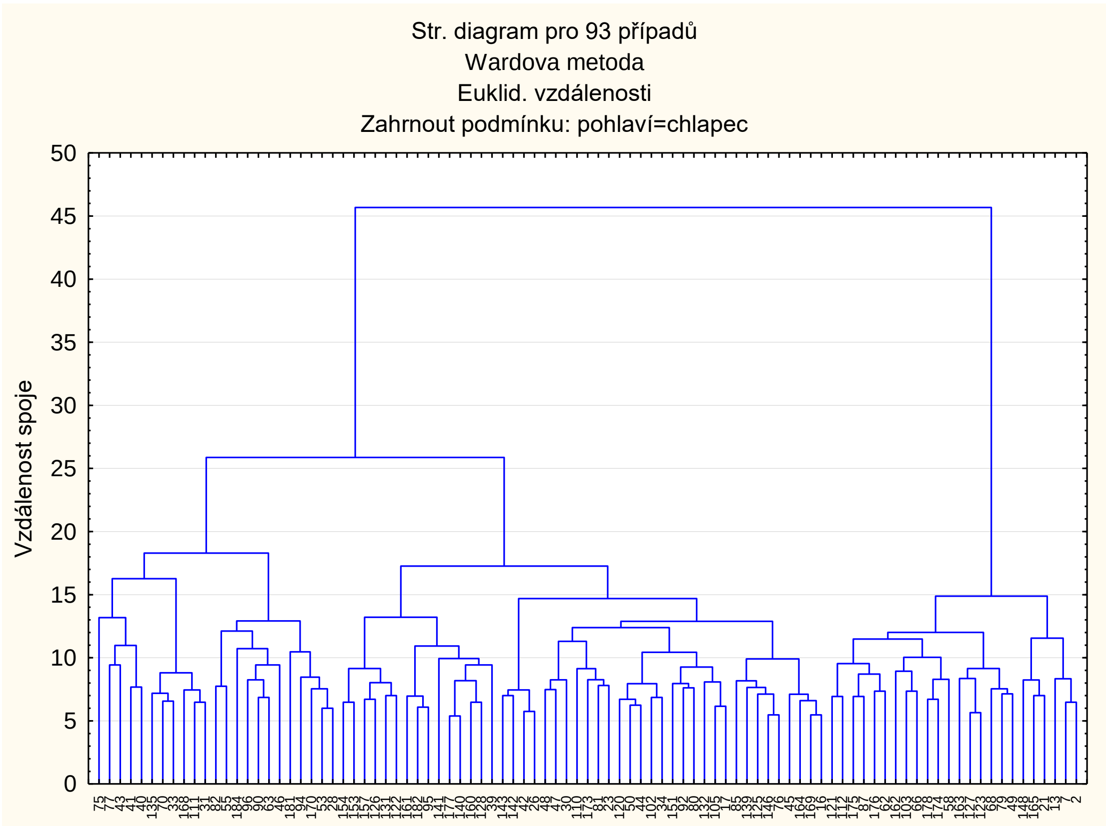
Případ	Průměry a směrodat. od			Případ	Průměry a směrodat. od	
	Průměr	Sm.Odch.			Průměr	Sm.Odch.
1	2,829787	1,203602		21	2,191489	1,055798
2	2,617021	1,243298		22	3,000000	1,063219
3	2,851064	1,141675		23	2,680851	1,044788
4	2,936170	1,091554		24	2,936170	1,205139
5	2,893617	1,005075		25	2,808511	0,850532
6	2,574468	1,078338		26	2,361702	0,870421
7	2,617021	1,074470		27	2,702128	1,019695
8	2,680851	1,124942		28	2,893617	1,088158
9	3,063830	0,894531		29	3,000000	0,859727
10	2,595745	1,173245		30	2,808511	1,135175
11	2,829787	1,089857		31	3,085106	1,039017
12	2,787234	1,020149		32	2,893617	1,088158
13	2,361702	1,342054		33	2,936170	0,894531
14	2,829787	0,842335		34	2,787234	0,931016
15	2,872340	0,849988		35	3,021277	1,112954
16	2,659575	1,089008		36	2,680851	0,980381
17	2,680851	0,934982		37	2,723404	1,280321
18	2,489362	0,929524		38	3,000000	1,083473
19	2,957447	0,907875		39	2,723404	1,440141
20	2,680851	1,144104		40	3,170213	0,939916

Tabulka 2: Shluková analýza- hierarchické shlukování - Průměry a směr. odchylky

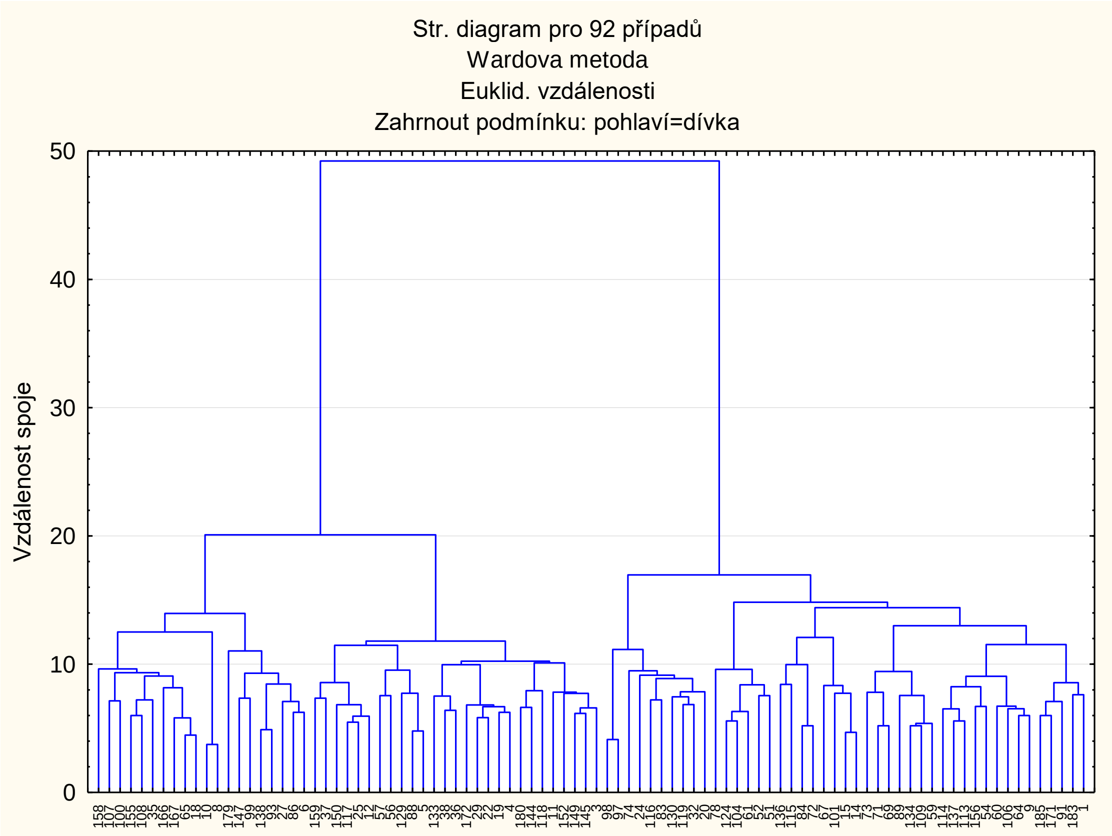
V tabulce č. 3 vidíme průměrnou hodnotu odpovědí na jednotlivé otázky a směrodatnou odchylku od daného průměru pro prvních čtyřicet dotazovaných. Protože tázaní odpovídali na otázky mírou souhlasu na škále 1- 4, tudíž hranice mezi souhlasem a nesouhlasem byla v hodnotě 2,5. Žáci, u kterých je tedy průměr menší než hodnota 2,5 odpovídali na otázky spíše záporně, naopak žáci s hodnotami blížícími se hodnotě 4 spíš odpovídali souhlasně. Odchylka nám udává hodnotu, jak moc se lišili odpovědi na jednotlivé otázky od průměrné hodnoty.

Na posledním případě hierarchického shlukování si porovnáme shlukování zvlášť pro chlapce a zvlášť pro dívky. Na obrázku č. 6 vidíme graf chlapců a na obrázku č. 7 graf dívek. Porovnáme-li oba obrázky, můžeme vidět, že ke spojení dvou shluků, dochází u chlapců ve vzdálenosti přibližně 46, kdežto u dívek dojde ke shluknutí ve vzdálenosti 49. Dále vidíme, že u shlukování chlapců se vytvoří tři dosti podobné shluky. Pokud porovnali shluky dva, vidíme, že shluk nacházející se více vlevo je značně větší a je v něm mnohem větší počet menších shluků. Z obrázku č. 7 vidíme, že dívky se dělí spíše do čtyř stejných shluků při vzdálenosti přibližně 15. Pokud bychom se zajímali o stav shluků ve vzdálenosti 22, je zde možné vidět dva relativně porovnatelné shluky. Můžeme tedy říct, že mezi chlapci existují tři skupiny, se stejným názorem na dané téma. Mezi dívkami naopak existují takové skupiny rovnou čtyři.

Použitím hierarchického shlukování si můžeme pomoci před použitím shlukování nehierarchického. Stromové grafy nám mohou napovědět, kolik shluků si máme zvolit při použití meto k- means, avšak tato nápověda nemusí být vždy správná.



Obrázek č. 6: Shluková analýza – Wardova metoda – dendogram chlapců



Obrázek č. 7: Shluková analýza – Wardova metoda – dendogram dívek

2.2.2. Nehierarchická shluková analýza

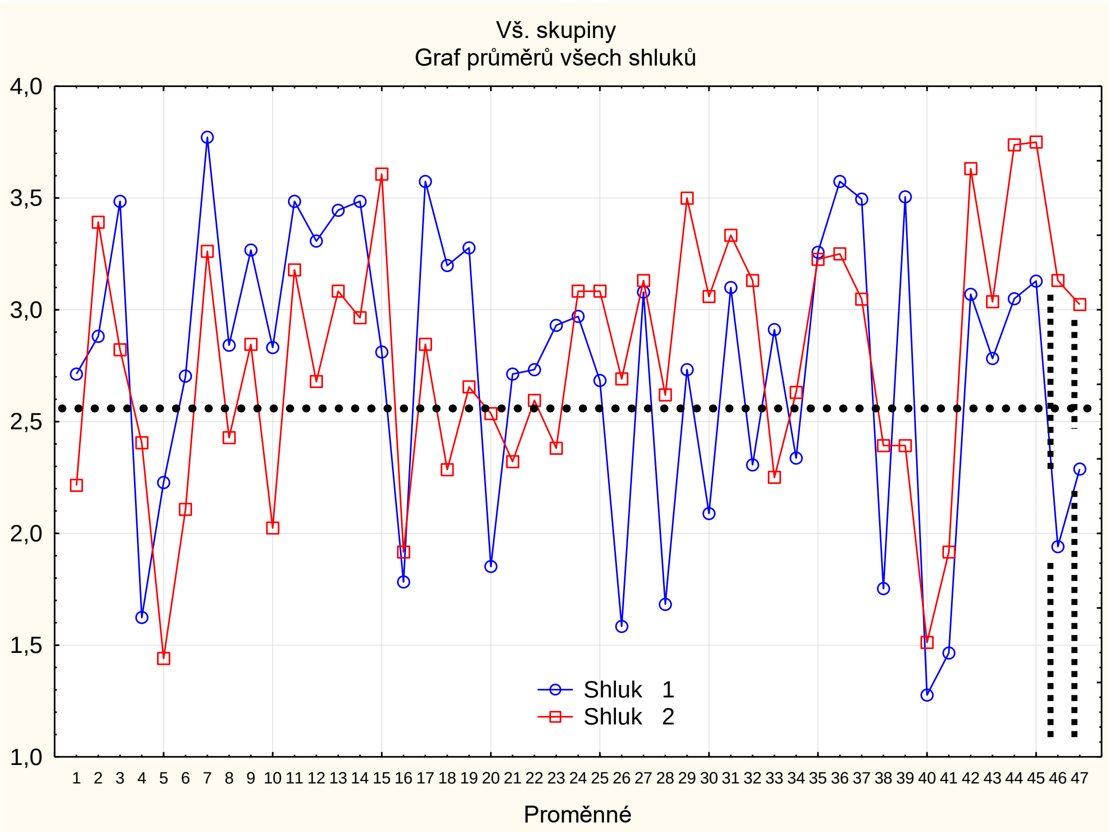
V následujícím případě (opět jde o shlukování 185 objektů pozorování) si představíme zpracování dat pomocí nehierarchické metody. Pro metodu jsme si nastavili měření vzdáleností pomocí euklidovské metriky, a zvolili jsme metodu K- prům. shlukování. Počet shluků byl nastaven na 2, což bylo odvozeno od dendogramů vzniklých v předešlém příkladě.

Proměnná	Prům. shluků (DATA Chovanec)		
	Shluk	Shluk	p signifikance
Učením strávím každý den hodně času (víc než hodinu):	2,71	2,21	0,001
Učím se, jen když musím:	2,88	3,39	0,000
Na zkoušení či na písemku se vždycky připravuji:	3,49	2,82	0,000
Úkoly píšu většinou až ve škole:	1,62	2,40	0,000
Učení mne baví:	2,23	1,44	0,000
Někdy toho vím víc, než kolik je napsáno v učebnici:	2,70	2,11	0,000
Mám dobrý pocit, když chápu věci, o kterých se učíme:	3,77	3,26	0,000
S rodiči nebo kamarády si povídám o věcech, které probíráme ve škole:	2,84	2,43	0,003
Je pro mě velmi důležité, aby si ostatní nemysleli, že jsem hloupý/á:	3,27	2,85	0,001
Často se hlásím, že si chci opravit známku:	2,83	2,02	0,000
Dobré známky jsou pro mne nejdůležitější:	3,49	3,18	0,007
Důležité je dělat věci do školy tak, aby byli učitelé spokojeni:	3,31	2,68	0,000
Naši učitelé chtějí, abychom učivo chápali a neučili se ho jenom nazpaměť:	3,45	3,08	0,005
Učitelům nevadí, když se jich ptáme na to, čemu nerozumíme:	3,49	2,96	0,000
Naši učitelé nás často porovnávají s jinými třídami:	2,81	3,61	0,000
Učitelé často zlepšují známky, jen aby bylo ve třídě hodně vyznamenání:	1,78	1,92	0,322
Kdybych měl nějaký problém, můžu jít za učiteli a říct jim to:	3,57	2,85	0,000
Většiny učitelů v naší škole si vážím a jsou pro mne vzorem:	3,20	2,29	0,000
Naše škola nabízí kroužky a akce, na které rád chodím:	3,28	2,65	0,000
Myslím, že ostatní školy jsou lepší, než ta naše:	1,85	2,54	0,000
Jsem si jistý, že na střední škole nebudu mít s učením problémy:	2,71	2,32	0,006
Když mi něco nejde, je to jenom proto, že se dost nesnažím:	2,73	2,60	0,289
Když si dám nějaké předsevzetí, nebo se o něco snažím, dosáhnu toho:	2,93	2,38	0,000

Tabulka 3: Shluková analýza – nehierarchické shlukování – průměry shluků

V tabulce č. 3 vidíme průměrné hodnoty zvolených odpovědí, které odpovídaly míře souhlasu na stupnici 1- 4. Signifikace nám současně určuje hladinu významnosti, pro jednotlivé otázky. Jsou-li v tomto případě hodnoty signifikace větší než hodnota 0.05, nejsou odpovědi na danou otázku v obou shlucích statisticky významně odlišné. V použitém výřezu jde vidět, že se jedná pouze o dvě otázky (červeně zvýrazněné hodnoty). V tabulce také můžeme vidět, se kterou otázkou jednotlivé shluky průměrně souhlasily nejvíc (žlutě zvýrazněné).

Nejvýznamnějším a nejnázornějším výstupem této metody je graf průměrných hodnot jednotlivých shluků (graf č. 1). V grafu vidíme, že na ose y jsou hodnoty míry souhlasu (1- nesouhlasím, 4- souhlasím) a na ose x jsou jednotlivé proměnné. Průměrná hodnota souhlasu je hodnota 2,5. Pokud se tedy graf pohybuje nad hodnotou 2,5, objekty v daném shluku spíše vyjádřily souhlas na danou otázku. Při porovnání obou křivek vidíme, že se občas jejich hodnoty značně liší, avšak pokud jsou obě hodnoty nad průměrnou hodnotou 2,5 (resp. pod průměrnou hodnotou) nejedná se o nic významného. V tomto případě se liší jen míra souhlasu (resp. nesouhlasu). Aby se názor na danou otázku diametrálně lišil, je nutné, aby byl bod grafu prvního shluku nad průměrnou hodnotou a bod grafu druhého shluku pod průměrnou hodnotou (nebo naopak) jako tomu je např. u otázky 46 a otázky č. 47 (zvýrazněné).



Graf 1: Shluková analýza – K- means – průměry shluků

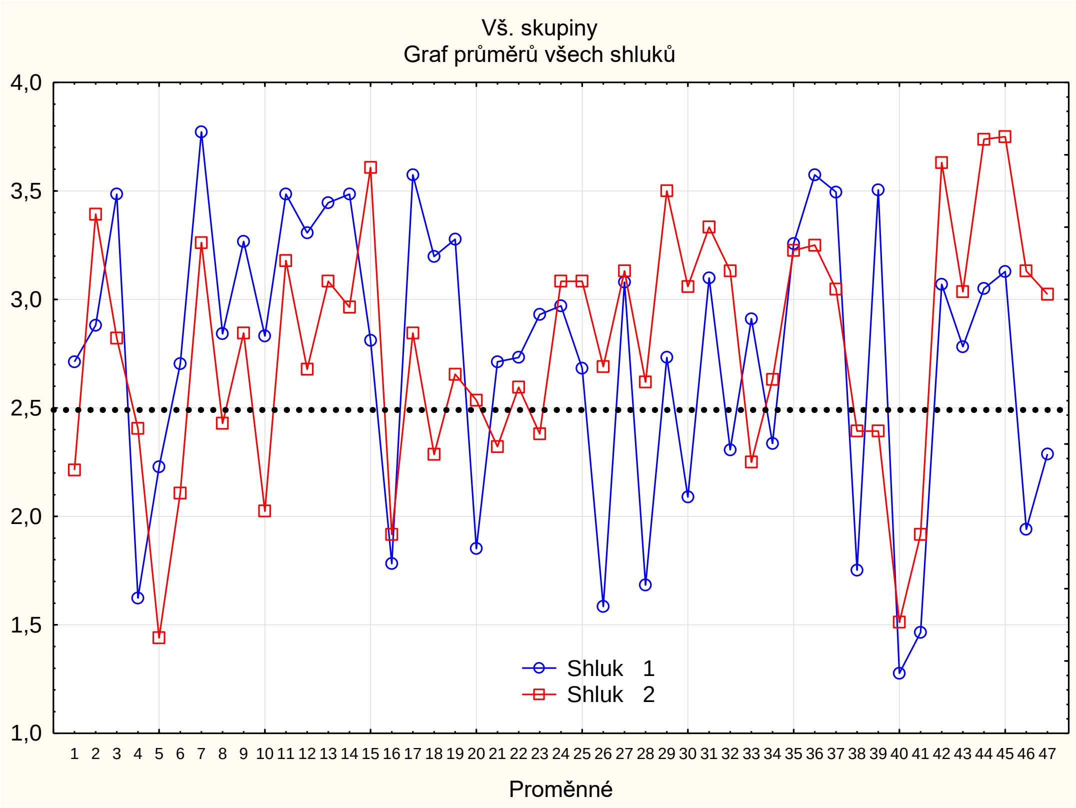
Posledním velmi užitečným výstupem jsou tabulky s výpisem jednotlivých členů daných shluků, jejich vzdáleností od středu shluku, do kterého patří a jejich počtem.

Tato skutečnost může být užitečná při zpracovávaní dat, která nejsou anonymní a je potřeba si doplnit informace o některých objektech. Jak vidíme v tabulce č. 4, do shluku č. 1 spadají objekty s pořadovým číslem 3, 4, 5, 6, 8 atd. Současně se dozvídáme, že shluk č. 1 obsahoval 101 respondentů a shluk č. 2 pouze 84.

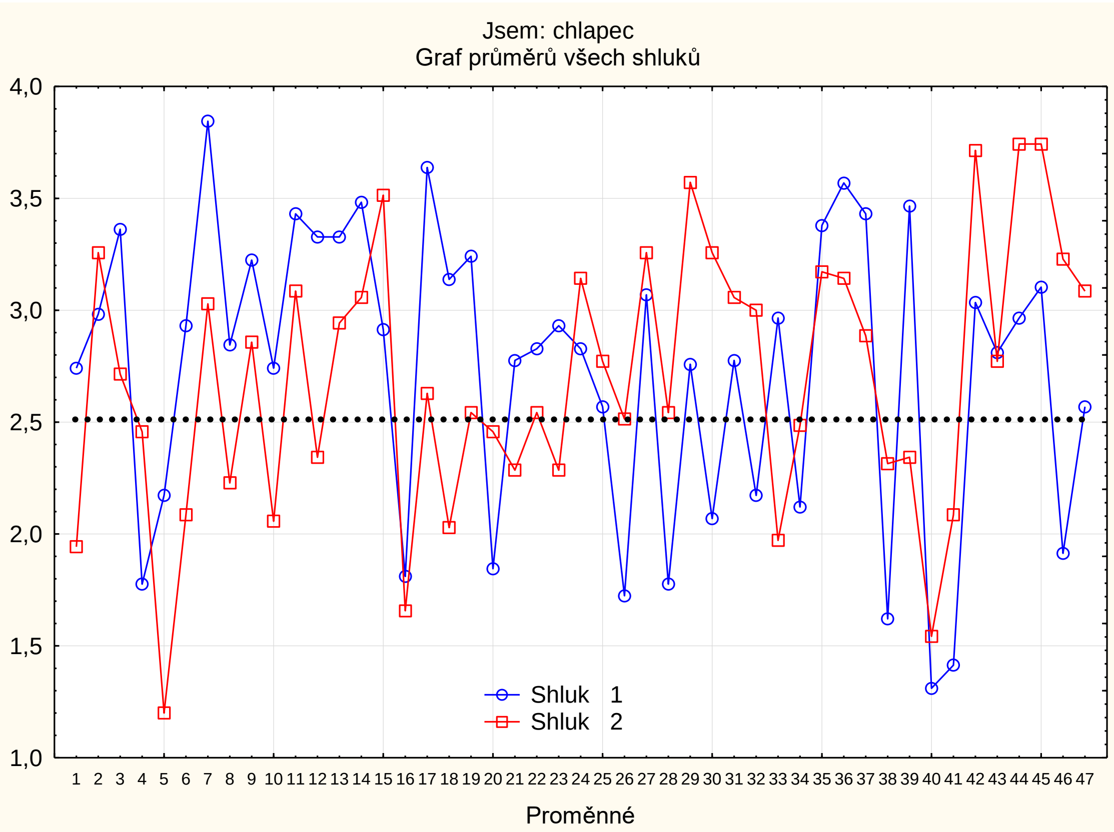
	Členy shluku číslo 1 (DATA Chovanec) a vzdálenosti od přislušného středu shluku Shluk obsahuje 101 příp.		Členy shluku číslo 2 (DATA Chovanec) a vzdálenosti od přislušného středu shluku Shluk obsahuje 84 příp.
	Vzdálen.		Vzdálen.
3	0,78939	1	0,898172
4	0,788322	2	1,082338
5	0,744056	7	0,853936
6	0,829597	9	0,697895
8	0,887384	13	1,231027
10	0,879036	14	0,736394
11	0,855105	15	0,784685
12	0,764307	21	1,049546
16	0,788723	24	0,972378
17	0,72716	31	0,900988
18	0,616641	32	0,84319
19	0,71001	33	0,704038
20	0,93726	35	0,974719
22	0,816285	40	1,095133
23	0,89941	45	0,888817

Tabulka 4: Shluková analýza – K- means – počty objektů ve shluku

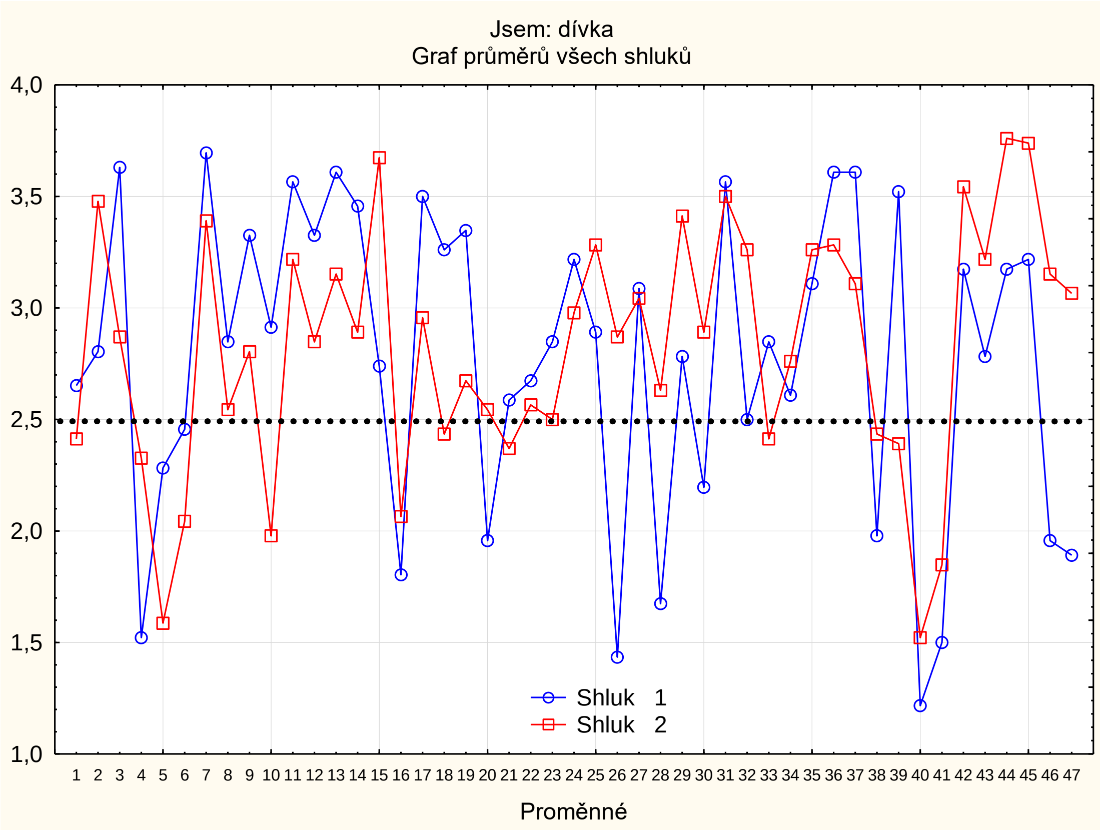
Poslední názornou ukázkou u shlukové analýzy bude porovnání grafů průměrných odpovědí pro dva shluky u všech objektů, u chlapců a u dívek.



Graf 2: Shluková analýza – průměr shluků – všechny objekty



Graf 3: Shluková analýza – průměr shluků – chlapci



Graf 4: Shluková analýzy – průměr shluků - dívky

Při porovnání grafů pro všechny tři případy, vidíme, že u otázky č. 47 se odpovědi diametrálně liší v případě všech objektů a dívek. V případě chlapců oba shluky průměrně souhlasí, ačkoli shluk č. 2 měl míru souhlasu větší. Naopak při pohledu na grafy u otázky č. 1 vidíme, že ve všech třech případech se odpovědi shluků diametrálně liší, ale v případě chlapců byl rozdíl značně větší než v případě dívek. Tímto způsobem je možné porovnávat vzniklé shluky u jednotlivých otázek a tyto poznatky využít při výzkumu.

Shluková analýza má nespočetné množství možností aplikací, a je tedy jen na nás, jakou metodu, vzdálenost, či způsob budeme zpracovávat. Důležité je pouze to, aby nám vyhodnocená data byla co možná nejvíce k užitku a měli bychom dbát na jejich správnost.

2.3. Užití faktorové analýzy

Při praktických příkladech užití faktorové analýzy budu data zpracovávat stejně jako Vrbová (2013) abych mohl své výsledky porovnat s jejími. Výsledky se samozřejmě nemusí vždy shodovat, což může být ovlivněno řadou důvodů, o kterých se zmíním u jednotlivých příkladů.

2.3.1. Faktorové zátěže pro nečestné chování

V tomto případě zvolíme stejné zadání jako Vrbová (2013, s. 60) a faktorové zátěže zvolíme pouze pro 13 hodnot z druhé části dotazníku. Aby mohly být výsledky porovnány, zvolím si také 2 faktory a metodu hlavních komponent. Po aplikaci faktorové analýzy na faktory ještě aplikuji rotaci. Tímto jsou data zpracována jako v případě Vrbové. Hlavním výstupem v této chvíli budou faktorové zátěže, které následně porovnám.

Tabulka 2. Faktorové zátěže pro13 dotazníkových položek – část nečestné chování. (Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 5 iterations.)

Dotazníková položka	**Komponenta**	
	1	2
51. udělám si tahák, a když můžu, tak ho použiji	**0,819**	−0,100
52. když neumím, snažím se opsat od spolužáka nebo chci, aby mi napověděl	**0,800**	−0,101
50. před písemkou si půjčím tahák od kamaráda	**0,737**	
48. když si zapomenu napsat domácí úkol, opíši ho ve škole od spolužáků	**0,713**	
53. když neumím, opisuji rovnou z knížky nebo ze sešitu	**0,671**	
49. když si zapomenu napsat domácí úkol, vymluvím se, že jsem si zapomněl doma sešit	**0,588**	
57. ptám se spolužáků, kteří už písemku psali, jaké byly otázky	**0,553**	
59. když si sami opravujeme písemku (diktát, úkol) opravím si, co jsem napsal špatně, nebo nahlásím méně chyb	0,409	0,252
60. když mám vypracovat referát, tak ho opíši téměř doslova z nějaké knížky nebo z internetu	0,392	0,204
64. když mám v žákovské špatnou známku, „ztratím ji“, nebo mám dvě žákovské	−0,166	**0,869**
65. když mají rodiče podepsat úkol nebo známky, podepíši si je sám, nebo mi je podepíše kamarád		**0,690**
58. když mám být zkoušený, nebo psát písemku, zůstanu doma	0,186	**0,478**
55. při rozdělení do skupin (A, B) si vyměníme písemky tak, abychom ti, co máme stejnou skupinu, seděli se vedle sebe	0,297	0,299
% vysvětlené variance	34	10

Obrázek č. 8: Faktorové zátěže – nečestné chování – Vrbová (převzato z (Vrbová (2013, s. 61)

	Proměnná	Faktor. zátěže (Varimax normaliz.) (DATA Chovanec) Extrakce: Hlavní komponenty (Označené zatěže jsou >,700000)		
		Faktor	Faktor	Komunalita
48.	Když si zapomenu napsat domácí úkol, opíši ho ve škole od spolužáků:	0,746986	0,201899	**0,60**
49.	Když si zapomenu napsat domácí úkol, vymluvím se, že jsem si zapomněl doma sešit:	0,462899	0,422057	**0,39**
50.	Před písemkou si půjčím tahák od kamaráda:	0,536428	0,589000	**0,63**
51.	Udělám si tahák, a když můžu, tak ho použiji:	0,758355	0,311875	**0,67**
52.	Když neumím, snažím se opsat od spolužáka nebo chci, aby mi napověděl:	0,841563	0,046262	**0,71**
53.	Když neumím, opisuji rovnou z knížky nebo ze sešitu:	0,512382	0,440116	**0,46**
55.	Při rozdělení do skupin (A, B) si vyměníme písemky tak, abychom ti, co máme stejnou skupinu, seděli se vedle sebe:	0,044775	0,694769	**0,48**
57.	Ptám se spolužáků, kteří už písemku psali, jaké byly otázky:	0,738445	0,089376	**0,55**
58.	Když mám být zkoušený, nebo psát písemku, zůstanu doma:	0,197067	0,754487	**0,61**
59.	Když si sami opravujeme písemku (diktát, úkol) opravím si, co jsem napsal špatně, nebo nahlásím méně chyb:	0,628336	0,266490	**0,47**
60.	Když mám vypracovat referát, tak ho opíši téměř doslova z nějaké knížky nebo z internetu:	0,586023	0,296478	**0,43**
64.	Když mám v žákovské špatnou známku, „ztratím ji“, nebo mám dvě žákovské :	0,141412	0,101405	**0,03**
65.	Když mají rodiče podepsat úkol nebo známky, podepíši si je sám, nebo mi je podepíše kamarád:	0,194306	0,577507	**0,37**
	Výkl.roz	3,988005	2,421581	
	Prp.celk	0,306770	0,186275	

Obrázek č. 9: Faktorové zátěže – nečestné chovaná - Chovanec

Ve výzkumu Vrbové zastupuje první faktor celkem sedm otázek a druhý faktor tři otázky. Z obrázku č. 9 je vidět, že v mém případě zastupuje první faktor šest otázek, v případě že vezmeme faktory s hodnotou větší než ±0,55. Druhý faktor zastupují při stejném kritériu 3 otázky. Toto je jen první rozdíl. V případě Vrbové je první faktor sycený otázkami č. 48, 49, 50, 51, 52, 53 a 57. V mém případě první faktor sytí otázky č. 48, 51, 52, 57, 59 a 60. V tomto případě se tedy 4 otázky shodují, ale zbylé ne. Druhý faktor je u Vrbové vyjádřen otázkami č. 58, 64 a 65. V mém případě se jedná o otázky č. 50, 55, 58. Zde se již shoduje pouze jedna otázka. Tyto neshody mohou být zapříčiněny velkým množstvím důvodů, ať už se jedná o rozdílný počet respondentů (Vrbová měla asi dvojnásobný počet respondentů), jejich věk, či typ školy. Samozřejmě může být tato odlišnost způsobena špatnou aplikací faktorové analýzy ze statistického hlediska. Vrbová (2013, s. 60) totiž udává, že první faktor čerpá 34 % variance a druhý faktor 10 %, což dohromady dává 44 %. Toto v podstatě znamená, že tyto dva faktory vysvětlují 44 % znaků vstupních dat, což je ze statistického hlediska dosti málo. Toto pozorování, však není náplní mé práce a proto jej zde nebudu již dále rozvíjet. Pro zajímavost, v mém případě čerpá přibližně 40 % a druhý faktor přibližně 9 %, v součtu tedy celkem 49 %. Toto je o něco málo lepší výsledek, avšak ani ten by se nedal ze striktně statistického hlediska obhájit. Současně je na obrázku č. 9 vidět, sloupeček komunalita. Ten udává % objasněného rozptylu dané otázky pro oba faktory. Ideální hodnota komunality by se měla pohybovat nad hodnotou 0,7. U otázky č. 64 (Když mám v žákovské špatnou známku, „ztratím ji“, nebo mám dvě žákovské) tedy vidíme, že komunalita je přibližně 3 % a tato otázka je tedy oběma faktory absolutně neobjasněmná. Myslím si, že to může být způsobeno tím, že v současné době jsou již na základních školách, kde jsem výzkum prováděl, elektronické žákovské knížky, a žáci tedy nevidí důvod, aby žákovskou „ztráceli“, protože známky jsou přístupné online. Je tedy vidět, že ačkoli obě řešení byly značně podobné, jejich výsledky by se daly po statistické stránce interpretovat velmi rozdílně.

Pro zajímavost, jsem pro stejné zadání provedl faktorovou analýzu metodou maximální věrohodnosti, která by měla mít na výstupu vždy statisticky nejvhodnější výsledky. Z tabulky č. 5 vidíme, že podle hodnocení programu Statistica (ze statistického hlediska je hodnocení správné) je faktor č. 1 sycený pouze otázkami č. 48 a 52. Druhý faktor je sycený pouze 50. otázkou. Protože by v tomto případě nedošlo k efektivnímu nahrazení původních znaků novým faktorem, nebylo řešení s dvěma faktory v tomto případě vhodné, avšak posuzování správného užití daných metod opět není náplní mé práce.

Proměnná	Faktor	Faktor 2
48.	0,700705	0,255343
49.	0,367950	0,465450
50.	0,399150	0,735006
51.	0,659039	0,443969
52.	0,836483	0,112992
53.	0,443708	0,446844
55.	0,173206	0,362725
57.	0,606553	0,248745
58.	0,236746	0,534306
59.	0,506580	0,354850
60.	0,529861	0,292198
64.	0,032685	0,214710
65.	0,177125	0,487487
Výkl.roz	3,140383	2,205593
Prp.celk	0,241568	0,169661

Tabulka 5: Faktorové zátěže – metoda maximální věrohodnosti

2.3.2. Faktorové zátěže pro faktory ovlivňující podvádění žáků

Opět budu volit stejné postupy pro zpracování dat jako Vrbová (2013, s. 66) aby mohly být výsledky porovnány. Je tedy zvolena metoda hlavních komponent, a faktory jsou následně rotovány metodou Varimax.

Při pohledu na výstupní tabulku v programu STATISTICA je ihned jasné, že v mém případě není ani jeden z faktorů statisticky nijak významně dotovaný (STATISTICA automaticky barevně vyznačí významné položky). Abych se dostal k podobným výsledkům jako Vrbová, musely by být za významné považovány i položky s dotací menší než ±0,7. Stejně jako v případě nečestného chování (kapitola 2.3.1) může tyto výsledky ovlivňovat velké množství příčin.

Dalším výstupem faktorové analýzy je reziduální matice. Hodnoty této matice by se měly při správném zastoupení faktorů pohybovat co nejblíže 0 s maximálním rozmezím ±0,03. Když se na výřez z této matice podíváme (příloha III) vidíme zde hodnoty, které přesahují hodnotu ±0,1 (například hodnoty ±0,16 u proměnných 4 a 8).

Pokud takovéto hodnoty v matici najdeme, nejedná se o korektní řešení a měli bychom zvolit jiné řešení.

Proměnná	Faktor 1	Faktor 2	Faktor 3	Faktor 4
1.	0,568079	-0,066152	0,042675	0,129098
2.	-0,309616	0,118427	0,051063	0,264957
3.	0,560883	0,012001	0,392494	-0,045412
4.	-0,231341	0,130980	-0,275179	0,378747
5.	0,420899	-0,611678	0,087135	0,016330
6.	0,285003	-0,443254	0,034490	0,092030
7.	0,415470	-0,253828	0,265974	0,020994
8.	0,343670	-0,131907	0,042383	-0,118524
9.	0,401549	-0,038310	0,097444	-0,065194
10.	0,519190	-0,064588	0,051942	-0,282534
11.	0,538636	0,069932	0,184317	-0,007006
12.	0,480192	-0,096725	0,476837	-0,067238
13.	0,192984	-0,146736	0,406340	0,123978
14.	0,021117	-0,173486	0,500099	0,004486
15.	0,040972	0,259984	-0,307516	0,384361
16.	-0,035754	-0,094381	0,156364	0,472741
17.	0,144760	-0,135967	0,636880	0,008416
18.	0,188998	-0,363402	0,546408	0,026610
19.	0,086200	-0,041328	0,387923	-0,107146
20.	0,023823	-0,072535	-0,477159	0,456371
21.	0,075047	-0,460731	-0,022493	0,179344
22.	0,180512	0,014559	0,339056	0,335994
23.	0,246453	-0,269467	0,128413	0,042990
24.	0,267722	0,336759	-0,093365	-0,044808
25.	0,244763	0,330802	0,128214	0,478129
26.	-0,281105	0,258996	-0,203675	0,571382
27.	-0,413018	0,009933	-0,016319	-0,105071
28.	-0,347374	0,391103	-0,011741	0,465910
29.	-0,299297	0,452211	-0,203627	0,147823
30.	-0,364522	0,566613	-0,061499	0,205968
31.	0,384152	0,430220	0,183823	0,247810
32.	-0,039561	0,366381	-0,187980	0,212804
33.	0,248562	-0,095062	0,499192	0,127679
34.	0,211098	0,280634	-0,055291	0,333976
35.	0,510046	0,121229	-0,244560	-0,149137
36.	0,396505	-0,004600	0,130193	0,093764
37.	-0,070207	0,092260	0,528863	-0,255362
38.	0,210207	0,181759	-0,257987	0,349956
39.	0,315878	-0,137689	0,321859	-0,275921
40.	-0,097519	-0,090498	-0,045508	0,498642
41.	-0,287484	0,029629	-0,103670	0,206610
42.	0,120190	0,590346	-0,025525	0,247827
43.	-0,137896	0,074505	-0,019690	-0,006082
44.	0,042456	0,480430	-0,263189	0,054961
45.	-0,078282	0,596146	-0,258297	0,102550
46.	0,006072	0,359421	-0,458415	0,201671
47.	-0,061904	0,310099	-0,139179	0,255025
Výkl.roz	4,175049	3,904318	3,677222	2,949192
Prp.celk	0,088831	0,083071	0,078239	0,062749

Tabulka 6: Faktorové zátěže pro faktory ovlivňující podvádění žáků

3 Diskuse

Na základě dat získaných z dotazníkového průzkumu provedeného na základních školách v okrese Opava s celkovým počtem 185 respondentů jsem měl možnost provést jak shlukovou analýzu, tak analýzu faktorovou. Faktorovou analýzu jsem byl současně schopný porovnat s výsledky výzkum Vrbové (2013), z jejíž práce můj výzkum vycházel. V kapitole 2.3.1 proběhlo šetření typů podvádění, přičemž výsledky mnou zpracovaných dat jevily obdobné řešení jako v práci Vrbové (2013, s. 61). Naše výsledky sdílely stejný počet faktorů, lišily se pouze v zastoupení otázek v jednotlivých faktorech. Největší rozdíl nastal u otázky č. 64 (Když mám v žákovské špatnou známku, „ztratím ji“, nebo mám dvě žákovské), kdy v mém řešení tato otázka neměla žádný význam ve vztahu k faktorům, kdežto v řešení Vrbové byla klíčovým znakem druhého faktoru. Rozdílnost v zastoupení otázek může být ovlivněno věkem dotazovaných žáků (Vrbová pokládala dotazník v 9. třídách; můj dotazník byl pokládán v 6. – 8. třídách), množstvím odpovídajících respondentů (Vrbová měla 401 respondentů, v mém případě se jednalo o 185 respondentů) a také typem škol. Vrbová zkoumala také žáky víceletých gymnázií, oproti tomu můj výzkum byl proveden pouze na základních školách. Dalším možným aspektem rozdílných řešení může být současný vývoj dětí. Vrbová zkoumala tuto problematiku v již v roce 2010, čímž nám vznikl osmiletý skok mezi jednotlivými výzkumy. Dlouhá prodleva mezi oběma výzkumy dává také dostatečný prostor k vysvětlení diametrálně rozdílného významu otázky č. 64. Během těchto osmi let byla na mnoha základních školách po celé ČR zavedena elektronická žákovská knížka (Papírové žákovské knížky nahradí elektronické, 2009). Shodou okolností na základních školách, kde jsem prováděl výzkum je už elektronická žákovská knížka součástí administrativy. Protože jsou nyní známky pro rodiče dostupné on-line, nevidí již žáci základních škol důvod, proč papírovou žákovskou knížku ztrácet. Bohužel tato teorie je pouze mým subjektivním názorem, a není nijak podložená, protože to nebylo cílem mé práce.

Druhá možnost srovnání výsledků se naskytla v kapitole 2.3.2. V tomto případě jsou však výsledky výzkumu Vrbové a mého výzkumu diametrálně rozdílné. Zatímco Vrbová při zvoleném řešení čtyř faktorů našla jednotlivé otázky, které dané faktory sytily, v případě mého zpracování dat se mi neprokázala platnost ani jednoho faktoru. Řešení by možná bylo také možno najít, bylo by však nutné přijmout méně významné znaky sytící dané faktory, což by bylo protichůdné proti výsledkům interpretovaným programem STATISTICA a tyto výsledky (tj. počet faktorů) by nebyly statisticky významné. Je také otázka, zda by práce Vrbové nebyla ze statistického hlediska napadnutelná z důvodu malého procentuálního vysvětlení znaků jednotlivými faktory. Jedná se však pouze o mou domněnku a nebylo cílem mé práce takováto tvrzení podkládat.

Závěr

Bakalářská práce na téma „Vícerozměrné statistické metody v systému STATISTICA a možnost jejich využití při zpracování výzkumných dat“ byla vytvořena z důvodu větší propagace a objasnění problematiky statistického zpracování dat a bližšímu porozumění principu těchto metod.

V teoretické části byly představeny jednotlivé nejdůležitější vícerozměrné metody a popsány jejich základní principy, jejichž znalost byla nutná k možnému praktickému využití daných metod v systému STATISTICA. V praktické části bylo provedeno zpracování výzkumu, založeného na tématu podvádění žáků na základní škole. Tento výzkum byl proveden na 185 respondentech formou on-line anonymního dotazníku. Zpracováním výzkumných dat pomocí shlukové a faktorové analýzy bylo názorně ukázáno a vysvětleno, jak pracovat s jednotlivými metodami a jak provádět interpretaci jejich výsledků a grafických výstupů.

Stanovené cíle práce byly beze zbytku splněny a na příkladech byly názorně ukázány možnosti využití vícerozměrných statistických metod. Při porovnání výsledků dvou výzkumů došlo k jedné shodě s částečnými modifikacemi a k jednomu výsledku značně rozdílného charakteru. Vzhledem k problematice daného tématu a rozsahu práce si myslím, že je tato práce určena spíše pro studenty, učitele a laiky, kteří nemají se zpracováváním dat ještě tolik zkušeností. Pro zkušené statistiky by tato práce nemusela být, z čistě odborného pohledu, přínosná.

Seznam bibliografických citací

Cluster analysis. Statsoft.com. 2010. [Online] Statsoft, 2010. [Citace: 4. 2 2018.] http://www.statsoft.com/Textbook/Cluster-Analysis.

Hebák, Petr, a další. 2005. *Vícerozměrné statistické metody [3].* Praha : Informatorium, 2005. 80-7333-039-3.

Kučera, Jiří. neuvedeno. Shluková analýza. Masarykova univerzita. [Online] neuvedeno. [Citace: 14. 4 2018.] https://is.muni.cz/th/172767/fi_b/5739129/web/web/main.html.

Meloun, Milan. 2018. Milan Meloun. Meloun.upce.cz. [Online] 2018. [Citace: 10. 4 2018.] https://meloun.upce.cz/docs/research/chemometrics/methodology/7metody.pdf.

Meloun, Milan, Militký, Jiří a Hill, Martin. 2005. *Počítačová analýza vícerozměrných dat v příkladech.* Praha : Academia, 2005. 80-200-1335-0.

Papírové žákovské knížky nahradí elektronické. 2009. Denik.cz. [Online] Denik.cz, 19. 4 2009. [Citace: 18. 4 2018.] https://www.denik.cz/z_domova/papirove-zakovske-knizky-nahradi-20090419.html.

Řezanková, Hana. 2018. Klasifikace pomocí shlukové analýzy. Hana Řezanková. Docplayer.cz. [Online] 2018. [Citace: 14. 4 2018.] http://docplayer.cz/2550328-Klasifikace-pomoci-shlukove-analyzy-hana-rezankova.html.

Sebera, Martin. 2012a. Analýza hlavních komponent a faktorová analýza. Vícerozměrné statistické metody. [Online] 2012. [Citace: 26. 3 2018.] http://www.fsps.muni.cz/~sebera/vicerozmerna_statistika/pca.html.

Sebera, Martin. 2012b. Shluková analýza. Vícerozměrné statistické metody. [Online] 2012. [Citace: 10. 4 2017] http://www.fsps.muni.cz/~sebera/vicerozmerna_statistika/shlukova.html.

Statistica Help. 2017. Statsoft.com. [Online] Statsoft, 2017. [Citace: 16. 1 2018.] http://documentation.statsoft.com/STATISTICAHelp.aspx?path=common/AboutSTATISTICA/ElectronicManualIndex.

SPOHNEROVÁ, Kateřina. 2010. Shluková analýza. Olomouc, 2010. Diplomová práce. Univerzita Palackého v Olomouci. Vedoucí práce Jaroslav Marek

TONHAUSEROVÁ, Zuzana. 2013. Metoda hlavních komponent a její aplikace. Olomouc, 2013. Diplomová práce. Univerzita Palackého v Olomouci. Vedoucí práce Ondřej Vencálek

VRBOVÁ, Jana. 2013. Školní podvádění starších žáků: od explorativního výzkumu k strukturnímu modelu: Pilotní studie. České Budějovice, 2013. Disertační práce. Jihočeská univerzita v Českých Budějovicích. Vedoucí práce I., Stuchlíková.

CIGÁŇOVÁ, Veronika. 2006. *Faktorová analýza v systému STATISTICA*. Brno, 2006. Bakalářská práce. Masarykova univerzita. Vedoucí práce Marie Budíková.

Seznam tabulek

Seznam grafů

Seznam obrázků

Seznam příloh

Příloha 1: Dotazník 1. část

		ANO	Spíše ANO	Nevím	Spíše NE	NE
1.	Učením strávím každý den hodně času (víc než hodinu):					
2.	Učím se, jen když musím:					
3.	Na zkoušení či na písemku se vždycky připravuji:					
4.	Úkoly píšu většinou až ve škole:					
5.	Učení mne baví:					
6.	Někdy toho vím víc, než kolik je napsáno v učebnici:					
7.	Mám dobrý pocit, když chápu věci, o kterých se učíme:					
8.	S rodiči nebo kamarády si povídám o věcech, které probíráme ve škole:					
9.	Je pro mě velmi důležité, aby si ostatní nemysleli, že jsem hloupý/á:					
10.	Často se hlásím, že si chci opravit známku:					
11.	Dobré známky jsou pro mne nejdůležitější:					
12.	Důležité je dělat věci do školy tak, aby byli učitelé spokojeni:					
13.	Naši učitelé chtějí, abychom učivo chápali a neučili se ho jenom nazpaměť:					

14.	Učitelům nevadí, když se jich ptáme na to, čemu nerozumíme:					
15.	Naši učitelé nás často porovnávají s jinými třídami:					
16.	Učitelé často zlepšují známky, jen aby bylo ve třídě hodně vyznamenání:					
17.	Kdybych měl nějaký problém, můžu jít za učiteli a říct jim to:					
18.	Většinu učitelů v naší škole si vážím a jsou pro mne vzorem:					
19.	Naše škola nabízí kroužky a akce, na které rád chodím:					
20.	Myslím, že ostatní školy jsou lepší, než ta naše:					
21.	Jsem si jistý, že na střední škole nebudu mít s učením problémy:					
22.	Když mi něco nejde, je to jenom proto, že se dost nesnažím:					
23.	Když si dám nějaké předsevzetí, nebo se o něco snažím, dosáhnu toho:					
24.	Je hodně věcí ve škole, které nedokážu změnit, ať se snažím sebevíc:					
25.	Často se bojím, že nebudu rozumět nové látce:					
26.	Kdybych neopisoval, tak bych měl horší známky, než když opisuji:					
27.	Myslím, že jsou důležitější věci, než známky:					
28.	Není pro mne důležité, abych se dobře učil, hlavně že projdu:					

29.	Většinu věcí, které se ve škole učíme, stejně v životě nebudu potřebovat:					
30.	Nejradši bych do školy vůbec nechodil:					
31.	Písemek či zkoušení se bojím, mám strach, že je zkazím:					
32.	Často mne bolí hlava a cítím se unavený/á:					
33.	Bojím se opisovat, aby mne učitel nenachytal:					
34.	Když se mi něco ve škole nepovede, bojím se to říct doma:					
35.	Moji rodiče říkají, že dobré známky jsou důležité:					
36.	Moji rodiče říkají, že se neučím kvůli nim, ale kvůli sobě:					
37.	Moji rodiče ocení každý můj úspěch, třeba i malý:					
38.	Moji rodiče mne porovnávají se spolužáky nebo s dětmi svých známých:					
39.	Moji rodiče říkají, že se ve škole nemá lhát a podvádět:					
40.	Moji rodiče kvůli mne lžou, jen abych neměl ve škole problémy:					
41.	Moji rodiče se se mnou o škole nebaví:					
42.	Na některé předměty prostě nemám hlavu:					
43.	Abych něco pochopil, důležitější než předmět je učitel, který ho učí:					

44.	Abych se na písemku stíhal/a připravit, neměli bychom psát víc, než jednu za den:					
45.	Někteří učitelé toho po nás chtějí zbytečně moc:					
46.	Někdy mám pocit, že si na mne učitel zasedl a dává mi horší známky než ostatním spolužákům:					
47.	Ti co opisují, dostanou z úkolu lepší známku, než ti co ho neopíší:					

Příloha 2: Dotazník 2. část

		Velmi často	Často	Občas	Vyjímečně	Nikdy
48.	Když si zapomenu napsat domácí úkol, opíši ho ve škole od spolužáků:					
49.	Když si zapomenu napsat domácí úkol, vymluvím se, že jsem si zapomněl doma sešit:					
50.	Před písemkou si půjčím tahák od kamaráda:					
51.	Udělám si tahák, a když můžu, tak ho použiji:					
52.	Když neumím, snažím se opsat od spolužáka nebo chci, aby mi napověděl:					
53.	Když neumím, opisuji rovnou z knížky nebo ze sešitu:					
54.	Dám opsat spolužákovi, když to potřebuje:					
55.	Při rozdělení do skupin (A, B) si vyměníme písemky tak, abychom ti, co máme stejnou skupinu, seděli se vedle sebe:					
56.	Když vím, že bude písemka, raději nejdu do školy a naučím se až na opravnou písemku:					
57.	Ptám se spolužáků, kteří už písemku psali, jaké byly otázky:					
58.	Když mám být zkoušený, nebo psát písemku, zůstanu doma:					
59.	Když si sami opravujeme písemku (diktát, úkol) opravím si, co jsem napsal špatně, nebo nahlásím méně chyb:					
60.	Když mám vypracovat referát, tak ho opíši téměř doslova z nějaké knížky nebo z internetu:					
61.	Viděl jsem, jak moji spolužáci podvádí (za poslední rok):					
62.	Když učitel vidí, že někdo podvádí, potrestá ho:					

63.	Učitel se tváří, že nevidí, když někdo opisuje nebo našeptává:					
64.	Když mám v žákovské špatnou známku, „ztratím ji“, nebo mám dvě žákovské:					
65.	Když mají rodiče podepsat úkol nebo známky, podepíši si je sám, nebo mi je podepíše kamarád:					
66.	Ročník, který navštěvuji:					
67.	Pohlaví (chlapec x dívka):					
68.	Má základní škola (např. ZŠ Kobeřice)					

Příloha 3: Reziduální matice

Proměnná	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	0,65	-0,12	0,18	0,01	-0,11	-0,09	-0,05	0,01	-0,06	-0,01
2.	-0,12	0,82	-0,06	-0,08	-0,04	0,06	0,21	-0,01	-0,02	0,03
3.	0,18	-0,06	0,53	-0,03	-0,05	-0,01	-0,02	0,01	-0,08	0,01
4.	0,01	-0,08	-0,03	0,71	0,03	0,03	-0,02	0,16	0,11	-0,03
5.	-0,11	-0,04	-0,05	0,03	0,44	-0,09	-0,05	-0,02	0,04	-0,04
6.	-0,09	0,06	-0,01	0,03	-0,09	0,71	0,01	0,15	0,02	-0,05
7.	-0,05	0,21	-0,02	-0,02	-0,05	0,01	0,69	-0,05	0,15	0,01
8.	0,01	-0,01	0,01	0,16	-0,02	0,15	-0,05	0,85	-0,06	-0,12
9.	-0,06	-0,02	-0,08	0,11	0,04	0,02	0,15	-0,06	0,82	-0,02
10.	-0,01	0,03	0,01	-0,03	-0,04	-0,05	0,01	-0,12	-0,02	0,64
11.	0,03	0,09	0,03	0,04	0,06	-0,02	0,07	-0,13	-0,01	-0,09
12.	-0,07	0,09	-0,06	0,05	-0,01	-0,07	-0,01	0,05	-0,06	-0,02
13.	-0,07	-0,08	-0,04	0,05	0,01	0,04	-0,04	0,08	0,03	-0,01
14.	0,02	-0,07	-0,01	0,12	-0,02	0,08	-0,09	0,01	-0,02	0,02
15.	-0,04	-0,07	0,05	-0,11	0,01	0,02	0,04	-0,02	0,03	-0,01
16.	-0,01	0,04	-0,05	-0,03	-0,04	0,01	0,06	0,01	0,06	0,12
17.	0,13	0,00	0,01	0,03	-0,01	-0,04	0,03	0,00	-0,07	-0,06
18.	-0,04	-0,06	-0,05	-0,11	0,07	0,01	-0,04	-0,10	-0,01	0,01
19.	-0,01	-0,07	0,04	0,07	0,04	0,06	-0,10	0,09	-0,04	0,03
20.	0,02	-0,04	-0,03	-0,05	-0,06	-0,04	0,04	0,01	0,06	0,08

ANOTACE

Jméno a příjmení:	Jiří Chovanec
Katedra:	Katedra technické a informační výchovy
Vedoucí práce:	doc. PhDr. Miroslav Chráska, Ph.D.
Rok obhajoby:	2018

Název práce:	Vícerozměrné statistické metody v systému STATISTICA a možnosti jejich využití při zpracování výzkumných dat
Název v angličtině:	Multivariate statistical methods in STATISTICA possibility of their used in processing research data.
Anotace práce:	Bakalářská práce se zabývá využitím vícerozměrných statistických metod v programu STATISTICA při zpracování dat. V teoretické části jsou popsány nejpoužívanější vícerozměrné metody a principy, na kterých jsou dané metody založeny. V praktické části byl proveden výzkum mezi žáky vyššího stupně základních škol na téma podvádění (n=185). Následně byly interpretovány možné výstupy shlukové a faktorové analýzy, jejichž pomocí byla získaná data analyzována. Výzkum byl proveden formou on-line dotazníku, který byl po domluvě s řediteli základních škol publikováno žákům v době výuky informační výchovy.
Klíčová slova:	Faktorová analýza, shluková analýza, vícerozměrné statistické metody, STATISTICA
Anotace v angličtině:	This thesis is focused on the usage of multidimensional statistical methods while processing data in the „STATISTICA“ programme. The theoretical part describes the most used multidimensional methods and principles they are based on. In practical part the research on the topic of cheating was conducted among the middle

	school students (n=185). As following the possible outcomes of cluster and factor analysis were interpreted to be used during analysis of acquired data. The research was conducted in form of on-line survey which was as based on the agreement with elementary shool directors distributed to students during the classes of IT.
Klíčová slova v angličtině:	Factor analysis, cluster analysis, multivariate statistical methods, STATISTICA
Přílohy vázané v práci:	Příloha 1: Dotazník 1. část Příloha 2: Dotazník 2. část Příloha 3: Reziduální matice
Rozsah práce:	47 stran, 7 stran příloh
Jazyk práce:	Český jazyk