

UNIVERZITA PALACKÉHO V OLMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Kompoziční analýza vah
ve vícekritériálním hodnocení variant



Vedoucí bakalářské práce: **doc. RNDr. Karel Hron, Ph.D.**

Vypracoval: **Martin Trefil**

Studijní program: B1103 Aplikovaná matematika

Studijní obor: Matematika–ekonomie se zaměřením na bankovníctví

Forma studia: Prezenční

Rok odevzdání: 2015

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Martin Trefil

Název práce: Kompoziční analýza vah ve vícekriteriálním hodnocení variant

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Karel Hron, Ph.D.

Rok obhajoby práce: 2015

Abstrakt: Logratio metodika kompozičních dat je velmi mladá oblast statistiky. Cílem této bakalářské práce je její aplikace ve vícekriteriálním hodnocení variant při statistickém zpracování souboru vektorů vah kritérií. Práce tak obsahuje teoretické kapitoly o kritériích hodnocení a metodách stanovení jejich vah a o kompozičních datech a jejich specifické geometrii. Následně jsou tyto informace využity v praktickém příkladu při analýze vícekriteriálních vektorů vah získaných dvěma metodami od souboru hodnotitelů.

Klíčová slova: Vícekriteriální hodnocení variant, váha kritéria, logratio transformace, shluková analýza, biplot

Počet stran: 54

Počet příloh: 8

Jazyk: Český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Martin Trefil

Title: Compositional analysis of weights in multi-criteria decision making

Type of thesis: Bachelor's

Department:

Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Karel Hron, Ph.D.

The year of presentation: 2015

Abstract:

Logratio methodology of compositional data is a very young field of statistics. The aim of this bachelor thesis is its application to multi-criteria decision making in statistical analysis of vectors of criteria weights. The thesis contains theoretical sections on evaluation of criteria, on methods of determining their weights and also on compositional data and their specific geometry. These data are subsequently used in a practical example, where multi-criteria vectors of weights obtained by two methods from a sample of evaluators are analyzed.

Key words: Multi-criteria decision making, weight of criterion, logratio transformations, cluster analysis, biplot

Number of pages: 54

Number of appendices: 8

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci vytvořil samostatně pod vedením pana doc. RNDr. Karla Hrona, Ph.D. s použitím uvedených literárních zdrojů.

V Olomouci dne 30. března 2015

Obsah

Úvod	7
1 Kritéria hodnocení a jejich váhy	8
1.1 Úloha vícekriteriálního rozhodování a hodnocení	8
1.2 Kritéria hodnocení	9
1.3 Váhy a metody jejich stanovení	11
1.3.1 Saatyho metoda	12
1.3.2 Postupná Metfesselova alokace	14
2 Kompoziční data	15
2.1 Principy práce s kompozičními daty	16
2.2 Aitchisonova geometrie	17
2.3 Logratio transformace	19
2.3.1 Clr transformace	19
2.4 Popisná statistika	21
2.5 Kompoziční biplot	22
2.5.1 Konstrukce biplotu	23
2.5.2 Interpretace kompozičního biplotu	24
2.6 Hierarchické shlukování	26
3 Praktický příklad	29
3.1 Dotazník	29
3.2 Zpracování datového souboru	30
3.2.1 Datový soubor získaný Metfesselovou alokací	30
3.2.2 Datový soubor získaný Saatyho metodou	34
3.3 Porovnání výstupů použitých metod	38
Závěr	39
Literatura	40
Přílohy	41
Příloha A: Dotazník	41
Příloha B: Metfesselova alokace - Původní datový soubor	44
Příloha C: Metfesselova alokace - Kompoziční číselné charakteristiky	46
Příloha D: Metfesselova alokace - Clr transformovaný datový soubor	47
Příloha E: Saatyho metoda - Původní datový soubor	49
Příloha F: Saatyho metoda - Kompoziční číselné charakteristiky	51
Příloha G: Saatyho metoda - Clr transformovaný datový soubor	52
Příloha H: CD-ROM	54

Poděkování

Děkuji vedoucímu bakalářské práce panu doc. RNDr. Karlu Hronovi, Ph.D. za vstřícnost, cenné rady a za nespočet hodin, které mi věnoval při konzultacích. Dále děkuji své rodině, jež mě během studia podporovala, a Olomouci, která mi byla druhým domovem.

Úvod

Tématem této práce je kompoziční analýza vah ve vícekriteriálním hodnocení variant. Přestože se s vícekriteriálním rozhodováním a hodnocením setkáváme dnes a denně, tak většina z nás o této oblasti nemá ponětí. Já jsem nebyl výjimkou, proto jsem si zvolil toto téma, abych o ní, jakož i o související problematice kompozičních dat, získal vědomosti a pokusil se je předat co nejstručněji čtenářům této práce.

Metody vícekriteriálního hodnocení variant nám slouží k výběru optimální varianty (např. automobilu, nemovitosti nebo také nového zaměstnance) vzhledem k daným kritériím. Přitom příslušné vektory vah kritérií obsahují relevantní informaci především v podílech mezi jejich složkami.

Za cíl jsem si stanovil použití logratio analýzy kompozičních dat při statistickém zpracování vícekriteriálních vektorů vah, které jsem získal od skupiny respondentů dvěma metodami, konkrétně Metfesselovou alokací a Saatyho metodou. Na základě výstupů jejich analýz provedu srovnání, která lépe reflektuje předpoklad, že mužské pohlaví při výběru automobilu preferuje technické vlastnosti před estetickými, přičemž u ženského pohlaví to bývá naopak.

Předložená bakalářská práce se tak dělí na dvě hlavní části, teoretickou a praktickou. V první teoretické kapitole přiblížím úlohu vícekriteriálního hodnocení, pojem kritéria hodnocení včetně jejich vlastností a metod stanovení vah. Ve druhé kapitole se zabývám kompozičními daty, jejich specifickou geometrií, vybranou log-ratio transformací a nástroji, které jsou potřebné k jejich průzkumové statistické analýze (biplot, hierarchické shlukování).

Závěrečnou kapitolu práce věnuji praktickému příkladu, ve kterém všechny předchozí teoretické poznatky aplikuji na reálná data, která jsem získal z mnou vytvořeného dotazníku. Tato data zpracuji ve statistickém softwaru R [11] a popíši jednotlivé numerické a grafické výstupy. Na jejich základě určím metodu, která lépe reflektuje moje předpoklady o preferencích jednotlivých pohlaví při výběru automobilu.

Práci jsem vypracoval pomocí typografického softwaru \TeX [10].

1. Kritéria hodnocení a jejich váhy

Při tvorbě této kapitoly jsme čerpali informace uvedené v [8], [9].

1.1. Úloha vícekriteriálního rozhodování a hodnocení

Lhal by ten, kdo tvrdí, že se nikdy s těmito úlohami nesetkal, protože každý se v nějaké situaci musel rozhodnout. Tyto úlohy jsou součástí našich životů, skrývají se v mnoha činnostech, aniž bychom si to uvědomovali. Například je to klasický rozhodovací problém při výběru televizoru, při koupi nemovitosti nebo automobilu. A právě výběr automobilu bude hrát roli v praktické části této práce. Na první pohled si jistě řeknete, že na tom není nic složitého. Avšak při podrobnějším průzkumu zjistíme, že nalézt ideální řešení není nic triviálního.

Má to své opodstatnění. Zřídka kdy totiž existuje varianta nejlepší ve všech aspektech. V případě televizoru lze hovořit o širokoúhlé SMART TV¹ nejvyšší kvality dostupné za nejnižší cenu na trhu televizorů. Tato varianta by se jistě všem zamlouvala, ovšem v dnešním tržním světě se to zdá býti nereálné. Skutečně je tomu tak. Drtivá většina takovýchto problémů nenabízí optimální řešení (variantu). A proto musíme najít kompromisní řešení, které je dle některých kritérií horší, ale v námi požadovaných kritériích dosahuje uspokojivých výsledků.

Úlohu vícekriteriálního rozhodování lze zjednodušeně definovat jako úlohu, kdy je dána konečná množina n variant, které hodnotíme dle m kritérií. Naším cílem je naleznout variantu, která je dle všech kritérií hodnocena co nejlépe. Tuto variantu považujeme za „optimální“, řekněme alespoň kompromisní. Následně též můžeme varianty seřadit od nejlepší po nejhorší na základě jejich celkového posouzení vzhledem ke kritériím. V našem případě se nebudeme věnovat samotnému výběru optimální varianty, zaměříme se pouze na kritéria hodnocení, jejichž váhy budeme následně dále analyzovat.

¹Televizory vybavené operačním systémem, přístupem na internet a dalšími technologiemi.

1.2. Kritéria hodnocení

Kritéria jsou charakteristiky variant, dle kterých tyto varianty posuzujeme. Při tvorbě souboru kritérií je třeba dbát na několik zásad:

1. Úplnost souboru kritérií

- celkový cíl hodnocení by měl být zcela vyjádřen kritérii, avšak současně se v kritériích nesmějí vyskytovat nadbytečná kritéria, jejichž hodnoty je možné odvodit z hodnot jiných kritérií tohoto souboru.

2. Minimální počet kritérií

- stojí do jisté míry v opozici k požadované úplnosti souboru kritérií.

3. Neredundance kritérií

- je žádoucí, aby kritéria byla nezávislá, vzájemná závislost kritérií totiž vždy komplikuje a ztěžuje vztah mezi dílčími hodnoceními a celkovým hodnocením variant. Více o nezávislosti kritérií pojednává [8].

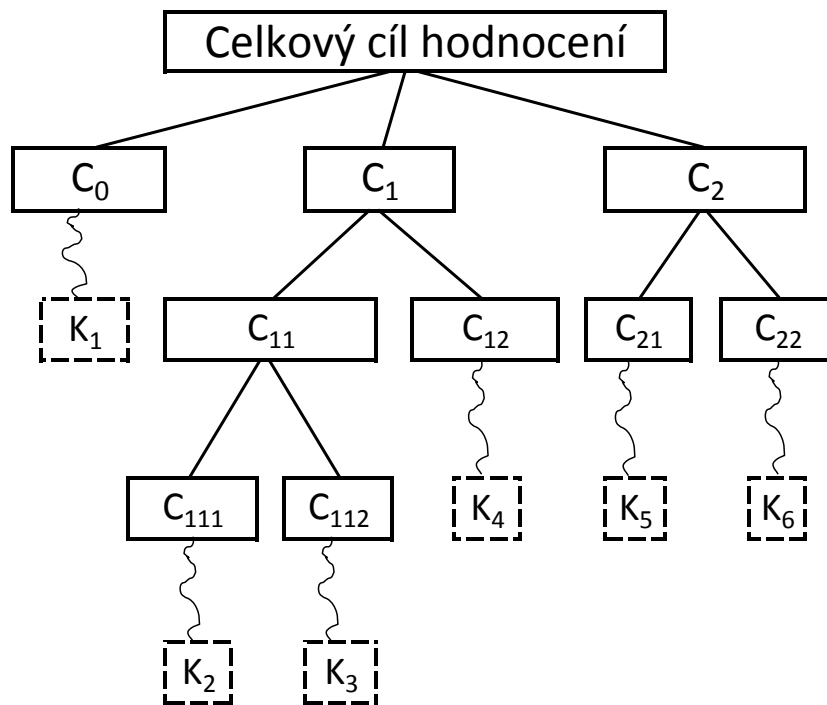
4. Měřitelnost kritérií

- kritéria by měla být měřitelná, abychom byli schopni, ať už jakýmkoliv způsobem (pomocí ordinální či kardinální stupnice), vyjádřit hodnocení variant vzhledem k těmto kritériím.

5. Expertní pomoc

- do souboru kritérií se zahrnují i takové vlastnosti variant, při stanovení jejichž hodnot je zapotřebí expertních znalostí.

K tvorbě souboru kritérií, který splňuje výše dané požadavky, použijeme metodu *stromu dílčích cílů* (obr. 1). Nejvyšší úroveň, zvaná kořen stromu, je celkový cíl hodnocení, který je v další úrovni rozdělen na dílčí cíle (C_0, C_1, C_2), jež jsou dále rozděleny do dílčích cílů ($C_{11}, C_{12}, C_{21}, C_{22}$). Takto pokračujeme, dokud nedospějeme k přímo měřitelným charakteristikám variant, tedy ke koncům větví, které představují jednotlivá kritéria hodnocení ($K_1, K_2, K_3, K_4, K_5, K_6$).



Obrázek 1: Strom dílčích cílů.

Kritéria lze dělit podle jejich charakteru na *kvalitativní*, která vyjadřují rozdílnou kvalitu určité vlastnosti u jednotlivých variant a jejichž hodnoty jsou obvykle zadány slovně, a na *kvantitativní*, která vyjadřují kvantitu určité vlastnosti a jejichž hodnoty jsou zadávány číselně. Dále se dělí podle způsobu vyjádření preferencí, které tato kritéria generují na množině variant, na kritéria *ordinální*, definující na množině variant preferenční relaci, a *kardinální*, která udávají předpoklady pro kvantitativní porovnání rozdílů v hodnocení variant.

1.3. Váhy a metody jejich stanovení

Váhy bývají vymezeny po matematické stránce velmi obecně. Váhami kritérií K_1, K_2, \dots, K_m se rozumí nezáporná reálná čísla v_1, v_2, \dots, v_m , která vyjadřují odlišnou významnost jednotlivých kritérií vzhledem k celkovému hodnocení variant. Jinak řečeno jednotlivé váhy představují kvantitativně vyjádřené relativní příspěvky na celkovém hodnocení. Většina metod pracuje s normovanými váhami, tj. s váhami, pro které platí

$$\sum_{j=1}^m v_j = 1, \quad v_j \geq 0. \quad (1.1)$$

Pokud jsou stanoveny nenormované váhy w_j , $w_j \geq 0$, $j = 1, 2, \dots, m$, pak normované váhy v_j se z nich vypočítají dle vztahu

$$v_j = \frac{w_j}{\sum_{k=1}^m w_k}. \quad (1.2)$$

Základní vlastností, kterou musí množina vah splňovat vzhledem k preferencím na množině kritérií, je následující vlastnost:

$$\begin{aligned} &\forall j, k \in \{1, 2, \dots, m\} : \\ &v_j \geq v_k \Leftrightarrow K_j \text{ je významnější nebo stejně významné jako } K_k. \end{aligned}$$

Váhy lze stanovit několika metodami: metodou párového srovnávání kritérií, Saatyho metodou, Kompenzační metodou, přímou Metfesselovou alokací, postupnou Metfesselovou alokací, atd.

Konkrétnější interpretace vah je vždy rozdílná v závislosti na použité metodě jejich stanovení. V jednoduchých metodách je postačující výše uvedené ordinální pojetí vah, v Saatyho metodě váhy vyjadřují poměry významnosti kritérií a v Metfesselově alokaci pak mají podobu procentuálního podílu dílčího cíle odpovídajícího danému kritériu na celkovém cíli hodnocení. Jednotlivé váhy tak vyjadřují kvantitativně vyjádřené relativní příspěvky na celkovém hodnocení.

V této práci se podrobněji zmíníme o Saatyho metodě a postupné Metfesselově alokaci.

1.3.1. Saatyho metoda

Při jejím použití zadáváme matici intenzit preferencí \mathbf{S} . Její prvky $s_{i,j}$ vždy vyjadřují námi stanovenou relativní významnost i -tého kritéria vzhledem k j -tému (tzn. vyjadřují, kolikrát je i -té kritérium významnější než j -té). Na diagonále leží samé 1.

\mathbf{S}	K_1	K_2	K_3	K_4	K_5	K_6
K_1	1					
K_2		1				
K_3			1			
K_4				1		
K_5					1	
K_6						1

Při zadávání hodnot prvků matice \mathbf{S} používáme základní pětibodovou stupnici intenzit preferencí, jejíž součástí jsou i příslušné jazykové deskriptory.

$$s_{i,j} = \begin{cases} 1 & \text{jsou-li obě kritéria stejně významná,} \\ 3 & \text{je-li } i\text{-té kritérium slabě významnější než } j\text{-té,} \\ 5 & \text{je-li } i\text{-té kritérium dosti významnější než } j\text{-té,} \\ 7 & \text{je-li } i\text{-té kritérium prokazatelně významnější než } j\text{-té,} \\ 9 & \text{je-li } i\text{-té kritérium absolutně významnější než } j\text{-té.} \end{cases} \quad (1.3)$$

Naopak, pokud je i -té kritérium méně významné než j -té, píšeme

$$s_{i,j} = \frac{1}{s_{j,i}}.$$

Nenormované váhy získáme jako vlastní vektor matice \mathbf{S} odpovídajícímu jejímu největšímu vlastnímu číslu λ_{max} , avšak pro dosažení relevantního výsledku musí být matice \mathbf{S} konzistentní, tedy v ideálním případě by mělo platit, že

$$\forall i, j, k \in \{1, 2, \dots, m\} : s_{i,k} = s_{i,j} \cdot s_{j,k}.$$

Proto se zavedl tzv. *koefficient nekonzistence CI*.

$$CI = \frac{\lambda_{max} - m}{m - 1}, \quad (1.4)$$

kde m je počet kritérií. Čím blíže je tento koeficient k nule, tím je matice \mathbf{S} konzistentnější. Jelikož neznáme přesnou hranici, kdy je koeficient nekonzistence ještě blízky nule a kdy už ne, tak zavádíme tzv. *podílový koeficient nekonzistence* CR . Ten vypočítáme dle vztahu

$$CR = \frac{CI}{RI(m)}, \quad (1.5)$$

kde $RI(m)$ je tzv. náhodný koeficient nekonzistence. Koeficienty $RI(m)$ závisí na řádu matice \mathbf{S} (počet kritérií), proto byla vytvořena tabulka těchto koeficientů.

m	3	4	5	6	7	8	9
$RI(m)$	0,525	0,882	1,115	1,252	1,341	1,404	1,452
m	10	11	12	13	14	15	16
$RI(m)$	1,484	1,513	1,535	1,555	1,57	1,583	1,595

Tabulka 1: Hodnoty koeficientů $RI(m)$

Koeficient CR nabývá hodnot z intervalu $\langle 0, 1 \rangle$. Nula vyjadřuje úplnou konzistenci, jednička absolutní náhodnost prvků v matici \mathbf{S} (nekonzistenci). Matici \mathbf{S} považujeme za konzistentní tehdy, když je koeficient $CR \leq 0,1$. V případě nekonzistence je možné vytvořit novou matici \mathbf{S}' tak, že

$$s'_{ij} = \frac{w_i}{w_j},$$

kde w_i, w_j jsou prvky vlastního vektoru z původní matice \mathbf{S} . Poté vypočítáme nové maximální vlastní číslo matice \mathbf{S}' a příslušný vlastní vektor.

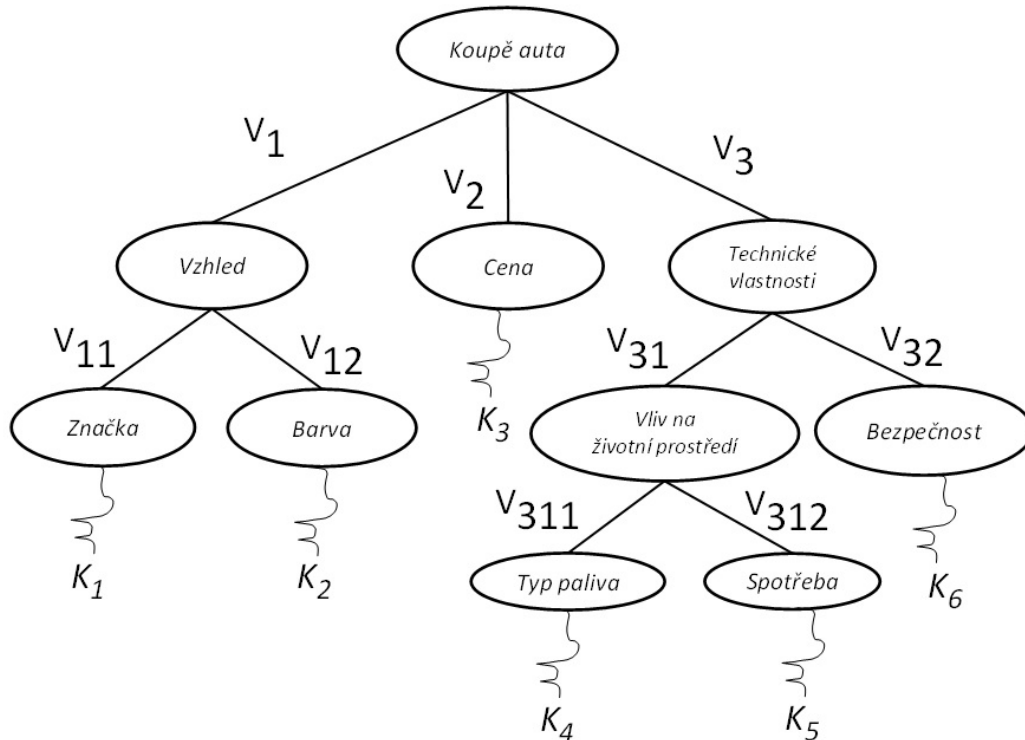
1.3.2. Postupná Metfesselova alokace

V této metodě využijeme strom dílčích cílů. Naším úkolem je přiřadit váhu každému dílčímu cíli, přičemž se vždy ptáme, jaký podíl má dílčí cíl na cíli ve vyšší úrovni. V každém větvení na libovolné úrovni musí být součet vah roven 1 a žádná váha nesmí být nulová, jelikož by pak všechna kritéria vedoucí přes tuto váhu byla nepoužitelná (nulová). Takto postupujeme do té doby, než dojdeme k jednotlivým kritériím hodnocení.

Váhu kritéria hodnocení dostaneme tak, že vynásobíme váhy, které se nacházejí na větvích vedoucích k tomuto kritériu. Situaci budeme ilustrovat na konkrétním příkladu stanovení vah při výběru automobilu. Váhy jednotlivých kritérií z obrázku 2 jsou:

$$v_{K_1} = v_{11} \cdot v_1, \quad v_{K_2} = v_{12} \cdot v_1, \quad v_{K_3} = v_2, \quad v_{K_4} = v_{311} \cdot v_{31} \cdot v_3,$$

$$v_{K_5} = v_{312} \cdot v_{31} \cdot v_3, \quad v_{K_6} = v_{32} \cdot v_3.$$



Obrázek 2: Schéma postupné Metfesselovy alokace.

2. Kompoziční data

V této kapitole popíšeme, co jsou kompoziční data vůbec zač a uvedeme jejich základní vlastnosti. Vycházíme z literatury [1], [2], [4], [7].

Tato oblast matematiky je poměrně mladá, přestože se s těmito daty setkáváme v praxi velmi často. Za zakladatele této oblasti je považován skotský statistik John Aitchison, jelikož v 80. letech 20. století právě on přišel s možností statisticky zpracovat kompoziční data (kompozice) pomocí tzv. logratio transformací. Kompoziční data jsou vícerozměrná data, jejichž neobvyklost tkví v tom, že jedinou relevantní informaci v nich poskytují podíly jednotlivých složek na celku. Složky tak typicky mohou být vyjádřeny v proporcích či v procentech. Nejčastěji se s kompozičními daty můžeme setkat v přírodních vědách jako je biologie, geologie, geografie, chemie a dále v archeologii, medicíně, socio-ekonomické oblasti a jiných. Pro názornější představu se jedná např. o koncentrace chemických prvků v hornině, poměr národnostních menšin v ČR, podíl každodenních výdajů na rozpočtu domácnosti atd.

Definice 2.1. Sloupcový vektor $\mathbf{x} = (x_1, x_2, \dots, x_D)^T$ se nazývá D -složková kompozice, jsou-li všechny jeho složky $x_i > 0$, $i = 1, \dots, D$ a nesou pouze relativní informaci.

Kompozice většinou reprezentujeme pomocí konstantního součtu složek κ . Často se $\kappa = 1$ (proporce) nebo $\kappa = 100$ (procenta), ovšem záleží na typu úlohy. Někdy konstantní součet totiž určit nejde, jelikož složky kompozice jsou v jednotkách koncentrace (např. mg/cm^3) apod. Reprezentace kompozičních dat pomocí libovolného součtu složek κ přitom nemá na jejich zpracování pomocí logratio transformací vliv.

Definice 2.2. Výběrový prostor kompozic se nazývá *simplex* definovaný jako

$$S^D = \{\mathbf{x} = (x_1, x_2, \dots, x_D)^T \mid x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = \kappa\}, \quad (2.1)$$

kde κ je libovolná pevně zvolená konstanta. Simplex je $(D - 1)$ - rozměrný podprostor R^D .

Pro formální vyjádření kompozic jako dat s konstantním součtem složek slouží následující definice.

Definice 2.3. Pro každý D -složkový kladný vektor $\mathbf{x} = (x_1, x_2, \dots, x_D)^T \in R_+^D$ je uzávěr kompozice \mathbf{x} definován jako

$$C(\mathbf{x}) = \left(\frac{\kappa \cdot x_1}{\sum_{i=1}^D x_i}, \frac{\kappa \cdot x_2}{\sum_{i=1}^D x_i}, \dots, \frac{\kappa \cdot x_D}{\sum_{i=1}^D x_i} \right). \quad (2.2)$$

Tato operace nám umožní bez ztráty informace převést součet složek kompozice na zvolenou konstantu κ .

Definice 2.4. Nechť je dána kompozice \mathbf{x} . Užitím operace uzávěru na podvektor $(x_{i_1}, x_{i_2}, \dots, x_{i_p})^T$ vektoru \mathbf{x} získáme podkompozici \mathbf{x}_p o p složkách, ($p < D$). Indexy i_1, i_2, \dots, i_p , $1 \leq i_1 < \dots < i_p \leq D$, označují ty složky, které jsou zahrnuty do podkompozice. Nemusí to nutně být prvních p složek.

2.1. Principy práce s kompozičními daty

Abychom mohli aplikovat na kompoziční data standardní statistické metody, je zapotřebí, aby tyto metody vyhovovaly třem podmínkám. Dle [1], [7] je to *škálová neměnnost, neměnnost při změně pořadí a soudržnost podkompozice*.

Škálová neměnnost

Informace v kompozici nezávisí na jednotkách, ve kterých je kompozice vyjádřena, jelikož kompoziční data nesou pouze relativní informaci.

Definice 2.5. Dva vektory \mathbf{x} , \mathbf{y} s D kladnými reálnými složkami jsou kompozičně ekvivalentní, jestliže existuje konstanta $\lambda \in R^+$ taková, že

$$\mathbf{x} = \lambda \cdot \mathbf{y}$$

a tedy

$$C(\mathbf{x}) = C(\mathbf{y}).$$

Přitom nám záleží na tom, abychom dostali totožné výsledky statistické analýzy nezávisle na konstantě λ .

Definice 2.6. Funkce $f(\cdot)$ je škálově neměnná, jestliže pro každou konstantu $\lambda \in R^+$ a pro každou kompozici $\mathbf{x} \in S^D$ funkce splňuje rovnost

$$f(\lambda\mathbf{x}) = f(\mathbf{x}).$$

To znamená, že získáme stejné výsledky pro každý kompozičně ekvivalentní vektor.

Neměnnost při změně pořadí

Permutace složek kompozice nezmění informaci obsaženou v kompozičním vektoru. Následně, funkce (statistická metoda) je invariantní vůči změně pořadí, poskytuje-li stejné výsledky, i když změním uspořádání jednotlivých složek v kompozici.

Soudržnost podkompozice

Informace zprostředkovaná kompozicí o D složkách by neměla být ve sporu s výsledky obdrženy z podkompozice o p složkách, $p < D$. Speciálně vzdálenost, kterou naměříme mezi dvěma kompozicemi, musí být větší nebo rovna vzdálenosti, kterou naměříme mezi jakýmkoliv jejich podkompozicemi. Tato vlastnost se nazývá podkompoziční dominance.

2.2. Aitchisonova geometrie

Práce s kompozičními daty je mnohem složitější než s klasickým datovým souborem využívajícím principy euklidovské geometrie. Proto se zavádí Aitchisonova geometrie s takovými operacemi, aby poskytovaly smysluplné informace o kompozicích při respektování principů práce s kompozičními daty. Mezi základní operace patří perturbace a mocninná transformace, které indukují vektorový prostor na simplexu a vyžadují operaci uzávěru, která je uvedena v definici 2.3.

Perturbace je analogií pro operaci sčítání v reálném prostoru, mocninná transformace je analogií pro násobení skalárem.

Definice 2.7. Perturbace kompozice $\mathbf{x} \in S^D$ kompozicí $\mathbf{y} \in S^D$ je kompozice

$$\mathbf{x} \oplus \mathbf{y} = C(x_1y_1, x_2y_2, \dots, x_Dy_D)^T. \quad (2.3)$$

Definice 2.8. Mocninná transformace kompozice $\mathbf{x} \in S^D$ konstantou $\alpha \in R$ je dána vztahem

$$\alpha \odot \mathbf{x} = C(x_1^\alpha, x_2^\alpha, \dots, x_D^\alpha)^T. \quad (2.4)$$

Takto obdržíme vektorový prostor (S^D, \oplus, \odot) , který má následující vlastnosti:

Vlastnost 2.2.1 (S^D, \oplus) je komutativní grupa, tedy pro $\mathbf{x}, \mathbf{y}, \mathbf{z} \in S^D$ platí

- komutativita: $\mathbf{x} \oplus \mathbf{y} = \mathbf{y} \oplus \mathbf{x}$;
- asociativita: $(\mathbf{x} \oplus \mathbf{y}) \oplus \mathbf{z} = \mathbf{x} \oplus (\mathbf{y} \oplus \mathbf{z})$;
- existence neutrálního prvku: $\mathbf{n} = C(1, 1, \dots, 1)^T = (\frac{1}{D}, \frac{1}{D}, \dots, \frac{1}{D})^T$;
- existence inverzního prvku: $\mathbf{x}^{-1} = C(x_1^{-1}, x_2^{-1}, \dots, x_D^{-1})^T$, kde $\mathbf{x} \oplus \mathbf{x}^{-1} = \mathbf{n}$.

Vlastnost 2.2.2 V případě mocninné transformace pro $\mathbf{x}, \mathbf{y} \in S^D$ a $\alpha, \beta \in R$ platí

- asociativita: $(\alpha \cdot \beta) \odot \mathbf{x} = \alpha \odot (\beta \odot \mathbf{x})$;
- distributivita: $\alpha \odot (\mathbf{x} \oplus \mathbf{y}) = (\alpha \odot \mathbf{x}) \oplus (\alpha \odot \mathbf{y})$; $(\alpha + \beta) \odot \mathbf{x} = (\alpha \odot \mathbf{x}) \oplus (\beta \odot \mathbf{x})$;
- existence neutrálního prvku: $1 \odot \mathbf{x} = \mathbf{x}$.

Vektorový prostor doplníme operacemi *skalární součin*, *norma* a *vzdálenost mezi kompozicemi*, čímž obdržíme euklidovský vektorový prostor.

Definice 2.9. Aitchisonův skalární součin kompozic $\mathbf{x}, \mathbf{y} \in S^D$ je definován jako

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}. \quad (2.5)$$

Díky vlastnostem logaritmu podílu lze skalární součin zapsat také jako

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}.$$

Definice 2.10. Aitchisonova norma kompozice $\mathbf{x} \in S^D$ je definována vztahem

$$\|\mathbf{x}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2} = \langle \mathbf{x}, \mathbf{x} \rangle_a. \quad (2.6)$$

Definice 2.11. Aitchisonova vzdálenost mezi \mathbf{x} a $\mathbf{y} \in S^D$ je dána jako

$$d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}, \quad (2.7)$$

kde $\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus \mathbf{y}^{-1}$.

2.3. Logratio transformace

Z předchozí části již víme, že ve standardním euklidovském prostoru nelze s kompozičními daty pracovat kvůli jejich odlišné geometrii. A proto byla zavedena v kapitole 2.2 Aitchisonova geometrie. Abychom mohli použít běžné statistické metody pro analýzu kompozic, použijeme tzv. logratio transformace. Aditivní logratio (alr) transformace nezachovává metrické vlastnosti dat (není izometrická), a proto tuto transformaci v další části práce nepoužijeme a nebudeme se jí dále zabývat. Centrovaná logratio (clr) transformace tuto vlastnost (izometrii) má, vyhovuje typu této práce, proto ji podrobněji popíšeme. Vlastnosti předchozích transformací vedly k myšlence zavedení třetí transformace, izometrické logratio (ilr) transformace, která je izometrií mezi S^D a R^{D-1} . Ilr transformaci taktéž jako alr transformaci nevyužijeme, proto ji nebudeme dále popisovat.

2.3.1. Clr transformace

V praktickém příkladu nám postačí pouze tato transformace, která zachovává metrické vzdálenosti dat. Kompoziční data jsou v S^D formulovány pomocí

kanonické báze $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_D\}$ v R^D . Tudíž můžeme každý vektor vyjádřit jako

$$\mathbf{x} = x_1(1, 0, \dots, 0)^T + x_2(0, 1, \dots, 0)^T + \dots + x_D(0, 0, \dots, 1)^T = \sum_{i=1}^D x_i \cdot \mathbf{e}_i.$$

Výše zmíněná kanonická báze bohužel není generujícím systémem ani bází vzhledem k Aitchisonově geometrii na S^D . Místo ní tak použijeme následující generující systém

$$\mathbf{w}_i = C(\exp(\mathbf{e}_i))^T = C(1, 1, \dots, e, \dots, 1)^T, \quad i = 1, 2, \dots, D.$$

Definice 2.12. Nechť je dána kompozice $\mathbf{x} \in S^D$. Clr koeficienty jsou složky vektoru $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_D)^T = \text{clr}(\mathbf{x})$, který splňuje

$$\mathbf{x} = \text{clr}^{-1}(\boldsymbol{\xi}) = C(\exp(\boldsymbol{\xi}))^T, \quad \sum_{i=1}^D \xi_i = 0, \quad (2.8)$$

i -tý clr koeficient je ve tvaru

$$\xi_i = \ln \frac{x_i}{g(\mathbf{x})}. \quad (2.9)$$

Kompozici $\mathbf{x} \in S^D$ lze tak vyjádřit vztahem

$$\mathbf{x} = \bigoplus_{i=1}^D \ln \frac{x_i}{g(\mathbf{x})} \odot \mathbf{w}_i = \ln \frac{x_1}{g(\mathbf{x})} \odot (e, 1, \dots, 1)^T \oplus \dots \oplus \ln \frac{x_D}{g(\mathbf{x})} \odot (1, 1, \dots, e)^T, \quad (2.10)$$

kde $g(\mathbf{x}) = \sqrt[D]{\prod_{i=1}^D x_i} = \exp\left(\frac{1}{D} \sum_{i=1}^D \ln x_i\right)$ je geometrický průměr složek kompozice \mathbf{x} . Centrovaná logratio (clr) transformace je vyjádřena pravou stranou této rovnosti a lze ji zapsat i po složkách jako vektory

$$\text{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right)^T = \boldsymbol{\xi}. \quad (2.11)$$

Inverze produkující koeficienty kanonické báze reálného prostoru následně vypadá takto

$$\text{clr}^{-1}(\boldsymbol{\xi}) = C(\exp(\xi_1), \exp(\xi_2), \dots, \exp(\xi_D))^T = \mathbf{x}. \quad (2.12)$$

Neměli bychom vynechat převod operací a metriky ze simplexu do reálného prostoru. K tomu si nejdříve označíme euklidovskou vzdálenost, normu a skalární součin v R^D pomocí $d(\cdot, \cdot)_e$, $\|\cdot\|_e$ a $\langle \cdot, \cdot \rangle_e$.

Věta 2.1. *Nechť jsou dány kompozice $\mathbf{x}_1, \mathbf{x}_2 \in S^D$ a reálné konstanty α, β , pak platí*

$$\begin{aligned} \text{clr}(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) &= \alpha \cdot \text{clr}(\mathbf{x}_1) + \beta \cdot \text{clr}(\mathbf{x}_2); \\ \langle \mathbf{x}_1, \mathbf{x}_2 \rangle_a &= \langle \text{clr}(\mathbf{x}_1), \text{clr}(\mathbf{x}_2) \rangle_e; \\ \|\mathbf{x}_1\|_a &= \|\text{clr}(\mathbf{x}_1)\|_e; \quad d_a(\mathbf{x}_1, \mathbf{x}_2) = d_e(\text{clr}(\mathbf{x}_1), \text{clr}(\mathbf{x}_2))_e. \end{aligned} \tag{2.13}$$

Clr transformace umožňuje interpretaci jednotlivých koeficientů jako proměnných vysvětlujících všechnu relativní informaci o dané složce kompozice, pohybujících se v čitateli příslušného koeficientu. To zapřičiňuje možnost použití jmen pro jednotlivé proměnné odpovídající názvům původních složek kompozice v popisu výstupů dané statistické analýzy. Jak se zmíníme později, tuto transformaci využijeme i při konstrukci biplotu. Nevýhodou clr transformace je, že výsledné proměnné jsou kolineární, jelikož platí rovnost $\sum_{i=1}^D \xi_i = 0$. To je důvod, proč clr transformované kompozice nelze zpracovat metodami, které jsou závislé na plné hodnosti datové matice.

2.4. Popisná statistika

Tato kapitola je napsaná na základě informací čerpaných z [2], [4], [7]. V této kapitole používáme datovou matici \mathbf{X} o n řádcích a D sloupcích se složkami x_{ij} , obsahující v řádcích kompozice, které pozorujeme. Před samotnou analýzou datového souboru musíme zkontrolovat, zda-li v datech nejsou chyby, soubor neobsahuje nuly a zda-li mají data logický význam. Zajímají nás také odlehle hodnoty, jejich počet a vzdálenost od ostatních dat v souboru.

Z důvodu nevhodnosti přímého použití standardních statistických metod v Aitchisonově geometrii byly zavedeny specifické číselné charakteristiky, např. centrum, variační matice či celkový rozptyl.

Definice 2.13. Charakteristika polohy pro kompozice se nazývá centrum. Pro sou-

bor o rozsahu n je definováno jako

$$\mathbf{g} = C(g_1, g_2, \dots, g_D)^T, \quad \text{kde } g_i = \left(\prod_{j=1}^n x_{ij} \right)^{\frac{1}{n}}, \quad i = 1, 2, \dots, D. \quad (2.14)$$

Definice 2.14. Rozptyl souboru kompozic je vyjádřen pomocí variační matice, matice rozptylů logaritmu podílů (resp. ortonormálních souřadnic) jednotlivých dvojic složek (x_{ik}, x_{jk}) , $i, j = 1, 2, \dots, D$; $k = 1, 2, \dots, n$,

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1D} \\ t_{21} & t_{22} & \dots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & \dots & t_{DD} \end{pmatrix}, \quad \text{kde } t_{ij} \text{ je rozptyl souboru } \left\{ \ln \frac{x_{ik}}{x_{jk}}, k = 1, 2, \dots, n \right\}, \quad (2.15)$$

nebo pomocí normované variační matice

$$\mathbf{T}^* = \begin{pmatrix} t_{11}^* & t_{12}^* & \dots & t_{1D}^* \\ t_{21}^* & t_{22}^* & \dots & t_{2D}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1}^* & t_{D2}^* & \dots & t_{DD}^* \end{pmatrix}, \quad \text{kde } t_{ij}^* \text{ je rozptyl souboru } \left\{ \frac{1}{\sqrt{2}} \ln \frac{x_{ik}}{x_{jk}}, k = 1, 2, \dots, n \right\}. \quad (2.16)$$

Pro t_{ij} a t_{ij}^* platí: $t_{ij}^* = \text{var}\left(\frac{1}{\sqrt{2}} \ln \frac{x_{ik}}{x_{jk}}\right) = \frac{1}{2} t_{ij}$, tedy $\mathbf{T}^* = \frac{1}{2} \mathbf{T}$. Tyto matice jsou symetrické a nuly tvoří jejich hlavní diagonály.

Definice 2.15. Charakteristikou celkové variability souboru kompozic je celkový rozptyl

$$\text{totvar}(\mathbf{X}) = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D t_{ij} = \frac{1}{D} \sum_{i=1}^D \sum_{j=1}^D t_{ij}^*. \quad (2.17)$$

2.5. Kompoziční biplot

V této části jsme informace čerpali zejména z [2], [5], [6]. Biplot zavedl v roce 1971 matematik K. R. Gabriel. John Aitchison později biplot přizpůsobil kompozičním datům a prokázal, že tento nástroj může být užitečným při redukci dimenze a interpretaci mnohorozměrné struktury kompozičních dat.

2.5.1. Konstrukce biplotu

Uvažujme kompoziční datovou matici \mathbf{X} s n řádky a D sloupce a předpokládáme, že již byla centrována (ve smyslu Aitchisonovy geometrie). Vytvoříme clr-transformovanou centrovanou matici $\mathbf{Z} = \text{clr}(\mathbf{X})$ pomocí clr koeficientů uvedených v definici 2.12. Předpokládáme, že transformovaná datová matice \mathbf{Z} má hodnotu s (obvykle $s = D - 1$). Tedy na matici \mathbf{Z} lze použít standardní statistické postupy, zejména pak fakt, že nejlepší aproximace matice \mathbf{Z} maticí \mathbf{Y} z pohledu nejmenšího součtu čtverců odchylek je poskytována singulárním rozkladem matice \mathbf{Z} .

Singulární rozklad matice clr koeficientů získáme z matice vlastních vektorů $\mathbf{Z}\mathbf{Z}^T$ s označením \mathbf{L} , matice vlastních vektorů $\mathbf{Z}^T\mathbf{Z}$ označené \mathbf{M} a diagonální matice odmocnin s kladných vlastních čísel $\lambda_1, \lambda_2, \dots, \lambda_s$ libovolné z matic $\mathbf{Z}\mathbf{Z}^T$ nebo $\mathbf{Z}^T\mathbf{Z}$ s označením \mathbf{K} . Následně, pro $k_i = \sqrt{\lambda_i}$, to lze zapsat maticově jako

$$\mathbf{Z} = \mathbf{L} \begin{pmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & k_s \end{pmatrix} \mathbf{M}', \quad (2.18)$$

kde tzv. singulární hodnoty k_1, k_2, \dots, k_s jsou v sestupném pořadí dle velikosti.

Obě matice \mathbf{L} a \mathbf{M} jsou ortogonální. Matici aproximace \mathbf{Y} o hodnotě 2 lze tedy vyjádřit jako

$$\begin{aligned} \mathbf{Y} &= (\mathbf{l}_1 \ \mathbf{l}_2) \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \end{pmatrix} = \\ &= \begin{pmatrix} l_{11} & l_{21} \\ l_{12} & l_{22} \\ \vdots & \vdots \\ l_{1n} & l_{2n} \end{pmatrix} \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \times \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1D} \\ m_{21} & m_{22} & \dots & m_{2D} \end{pmatrix}. \end{aligned} \quad (2.19)$$

Podíl celkové variability datového souboru zachovaný touto aproximací se vypočítá jako $\frac{\lambda_1 + \lambda_2}{\sum_{i=1}^s \lambda_i}$.

Abychom získali hledaný biplot, je nejprve nutné zapsat matici \mathbf{Y} jako součin dvou matic \mathbf{GH}^T , kde \mathbf{G} je matice o rozměrech $n \times 2$ a \mathbf{H} je matice o rozměrech $D \times 2$. Existují různé možnosti pro získání takového rozkladu, z nichž jeden je

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} \sqrt{n-1}l_{11} & \sqrt{n-1}l_{21} \\ \sqrt{n-1}l_{12} & \sqrt{n-1}l_{22} \\ \vdots & \vdots \\ \sqrt{n-1}l_{1n} & \sqrt{n-1}l_{2n} \end{pmatrix} \times \\ &\times \begin{pmatrix} \frac{k_1 m_{11}}{\sqrt{n-1}} & \frac{k_1 m_{12}}{\sqrt{n-1}} & \dots & \frac{k_1 m_{1D}}{\sqrt{n-1}} \\ \frac{k_2 m_{21}}{\sqrt{n-1}} & \frac{k_2 m_{22}}{\sqrt{n-1}} & \dots & \frac{k_2 m_{2D}}{\sqrt{n-1}} \end{pmatrix} = \\ &= \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_n \end{pmatrix} (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_D). \end{aligned} \quad (2.20)$$

Biplot tak zjednodušeně představuje společné znázornění $n + D$ vektorů $\mathbf{g}_j, j = 1, 2, \dots, n$, a $\mathbf{h}_i, i = 1, 2, \dots, D$, v rovině. Vektory $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ se přitom nazývají skóry a vektory $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_D$ se nazývají zátěže.

2.5.2. Interpretace kompozičního biplotu

Biplot převádí aproximaci matice \mathbf{Z} pomocí matice \mathbf{Y} o hodnoti dva dané singulárním rozkladem do grafického zobrazení. Kompoziční biplot se skládá z:

- počátku O , který představuje centrum kompozičního datového souboru,
- vrcholu na pozici \mathbf{h}_i pro každou z D složek,
- ukazatele na pozici \mathbf{g}_j pro každé z n pozorování.

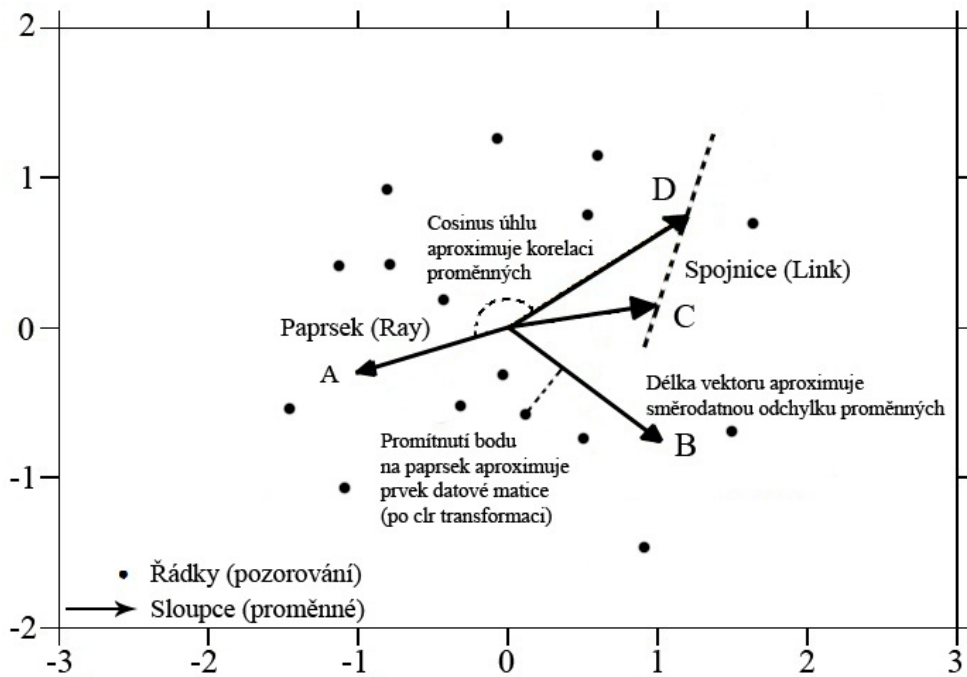
Propojení počátku O s vrcholem \mathbf{h}_i se nazývá paprsek $\overline{O\mathbf{h}_i}$ a spojení dvou vrcholů \mathbf{h}_i a \mathbf{h}_k se značí jako spojnice $\overline{\mathbf{h}_i\mathbf{h}_k}$. Tyto rysy tvoří základní charakteristiky biplotu s následujícími vlastnostmi pro výklad kompoziční variability.

1. Paprsky a spojnice poskytují informaci o relativní variabilitě v kompozičním datovém souboru vyjádřené jako

$$|\overline{\mathbf{h}_i\mathbf{h}_k}|^2 \approx \text{var} \left[\ln \frac{x_i}{x_k} \right] \quad \text{a} \quad |\overline{O\mathbf{h}_i}|^2 \approx \text{var} [Z_i] = \left[\ln \frac{x_i}{g(\mathbf{x})} \right]. \quad (2.21)$$

Nicméně je nutné dát pozor při interpretaci paprsků, které nemohou být ztotožněny ani s rozptylem $\text{var}[x_i]$, ani s rozptylem $\text{var}[\ln x_i]$, protože závisí na celé kompozici skrze $g(\mathbf{x}_1), g(\mathbf{x}_2), \dots, g(\mathbf{x}_n)$ a mění se, jestliže se pracuje pouze s podkompozicí.

- Jestliže vrcholy i a k splývají (téměř splývají), znamená to, že $\text{var}\left[\ln \frac{x_i}{x_k}\right]$ je roven nule (téměř roven nule), takže podíl $\frac{x_i}{x_k}$ je konstantní (téměř konstantní), a složky x_i a x_k můžeme považovat za vzájemně nahraditelné.



Obrázek 3: Osy biplotu skrze paprsky a spojnice.

2.6. Hierarchické shlukování

Informace zde uvedené jsme získali v [3], [6].

Cílem shlukové analýzy je nalezení skupin podobných objektů ve vícerozměrném prostoru. Obsahuje nespočet přístupů a metod, kterými lze tohoto cíle dosáhnout. Uplatnění nalézá tam, kde se množina objektů rozpadá do skupin, tzn. objekty se kupí v jednotlivých přirozených shlucích. Následně je nutné charakterizovat vzniklé skupiny.

Výše zmíněné metody a přístupy se liší např. měřením vzdálenosti mezi objekty, algoritmem spojování objektů do shluků nebo interpretací výstupů. V případě kompozičních dat by takovou vhodnou vzdáleností byla zřejmě Aitchisonova vzdálenost, ekvivalentně pak můžeme pracovat s clr transformovanými daty a standardní euklidovskou vzdáleností.

Dle [3] lze postup při shlukování objektů popsat tak, že je dána datová matice \mathbf{X} typu $n \times D$, kde n je počet objektů a D je počet proměnných (složek kompozice). Dále uvažujeme různé rozklady $S^{(k)}$ množiny n objektů do k shluků. Hledá se takový rozklad, který by byl v určitém smyslu nejvýhodnější. Cílem je dosažení stavu, kdy objekty uvnitř shluku jsou si co nejvíce podobné a zároveň co nejméně podobné s objekty z různých shluků.

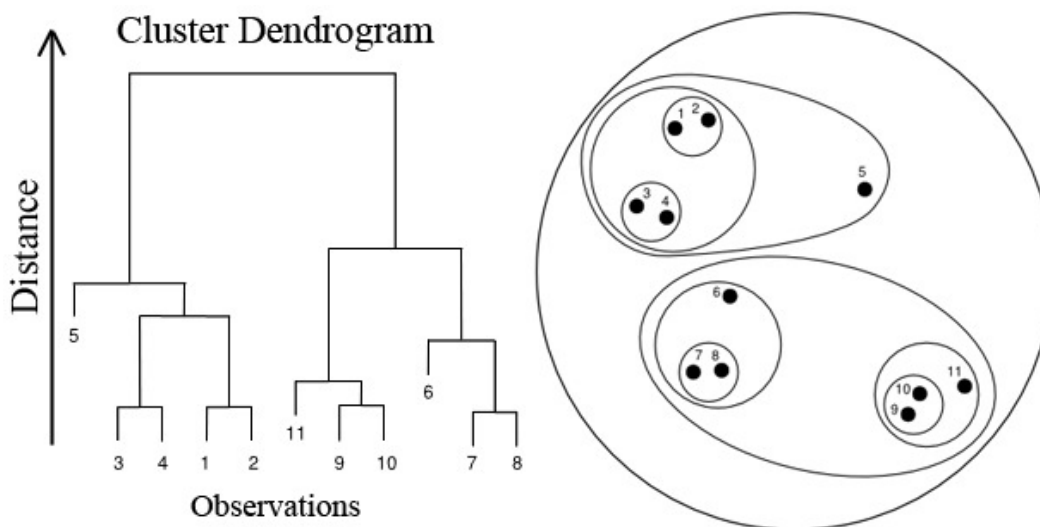
Při hierarchickém shlukování, kterým se budeme dále zabývat, se snažíme vytvořit vhodnou hierarchii shluků. Existují dva postupy tohoto shlukování, tj. *aglomerativní* a *divizivní*. Divizivní hierarchický postup je méně užívaný. Při něm se vychází z jediného shluku $S^{(1)}$ a v každém kroku jeden ze shluků rozdělíme na dva, takže na konci tohoto postupu dostaneme $S^{(n)}$ obsahující jednotlivá pozorování. Aglomerativní hierarchický postup lze dle [3] popsat následovně:

1. Výpočet matice \mathbf{D} vhodných měř vzdáleností.
2. Proces začne od rozkladu $S^{(n)}$, tzn. od n shluků, z nichž každý obsahuje jeden objekt.
3. Vzhledem k symetrii matice \mathbf{D} projdeme pouze horní nebo dolní trojúhelníkovou část a najdeme dva shluky (h -tý a h' -tý), jejichž vzdálenost $D_{hh'}$ je maximální.

4. Spojíme h -tý a h' -tý shluk do nového g -tého shluku. V matici \mathbf{D} odstraníme h -tý a h' -tý řádek i sloupec a nahradíme je řádkem i sloupcem pro nový shluk. Řád matice \mathbf{D} se tak snížil o jedničku.
5. Poznamenejme pořadí cyklu $v = 1, 2, \dots, n - 1$, identifikaci spojených objektů h, h' a hladinu pro spojení $d_1 = D_{hh'}$.
6. Jestliže proces vytváření rozkladů ještě neskončil spojením všech objektů do jediného shluku $S^{(1)}$, tak pokračujeme krokem 3.

Dendrogram

Je to stromový diagram, který se s výhodou používá ke grafickému znázornění výsledků hierarchických shlukovacích metod. Obrázek 4 má dvě části. Vpravo je zobrazení obecného postupu shlukování a vlevo je příslušný dendrogram. Tvoří jej dvě osy – na vodorovné ose jsou jednotlivé objekty a na svislé ose je hladina, na které jsou jednotlivé objekty spojovány.



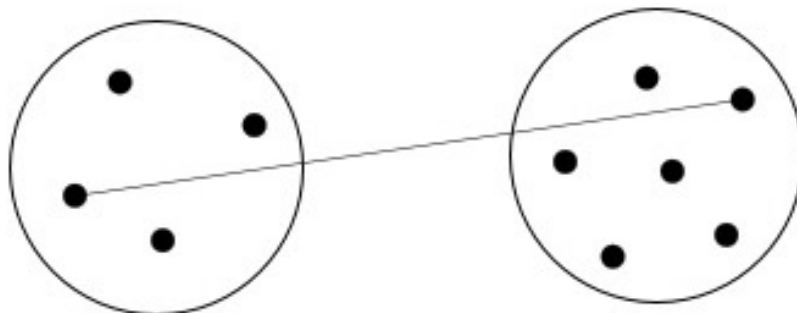
Obrázek 4: Dendrogram a postup shlukování, převzato z [12].

Výsledky výše popsaného algoritmu se liší dle metody hodnotící vzdálenost mezi shluky. Vzdáleností pro dva spojované shluky existuje celkem $q = n_h n_{h'}$.

Metod hodnotících vzdálenosti mezi shluky existuje hned několik, mezi nejužívanější patří právě výše popsaná *metoda nejvzdálenějšího souseda*; další metody (*metoda nejbližšího souseda*, která vede na řetězové dendrogramy, dále *metoda průměrné vazby*, *centroidní metoda* a *Wardova metoda*) v praktickém příkladu nepoužijeme, tudíž je nebudeme více konkretizovat, zájemce proto odkazujeme na [3], [6].

Metoda nejvzdálenějšího souseda

Tato metoda se občas nazývá jako *metoda úplné vazby*, či *úplného spojení* (complete linkage). Oba shluky, které chceme spojit, jsou zastoupeny objekty, mezi kterými je vzdálenost maximální (obr. 5). Touto vzdáleností D_{hh} , tedy rozumíme maximum ze všech q mezishlukových vzdáleností objektů.



Obrázek 5: Spojování shluků v metodě nejvzdálenějšího souseda, převzato z [12].

3. Praktický příklad

V této kapitole aplikujeme výše zmíněné teoretické poznatky na vlastním příkladu, týkajícím se výběru automobilu. Jako vstupní hypotézu přitom předpokládáme, tak jako široká veřejnost, že při výběru automobilu ženy preferují estetické vlastnosti před technickými vlastnostmi, kdežto u mužů to bývá naopak.

Ke stanovení vah jsme využili dvě metody – *Metfesselovu alokaci* a *Saatyho metodu*. I když jsou postupy zpracování obou datových souborů ve statistickém softwaru R analogické, věnujeme každému datovému souboru vlastní podkapitolu, kde popíšeme postup zpracování, numerické a grafické výstupy. Nakonec se pokusíme obě metody porovnat na základě dosažených výsledků a určit tu metodu, která lépe reflektuje vstupní předpoklad.

K výpočtu jednotlivých vah byly užity softwary Matlab a Microsoft Excel. Další numerické výpočty a grafická znázornění jsou provedeny ve statistickém softwaru R [11].

3.1. Dotazník

Abychom mohli uskutečnit výpočty a grafická znázornění, bylo zapotřebí mít dostatečné množství dat. Nejdříve jsme tedy vytvořili pomocí softwaru Microsoft Excel příslušný dotazník (Příloha A). Dotazník se skládá ze tří stran a obsahuje obě metody stanovení vah. První strana uvádí údaje o pohlaví a věku dotazovaného (respondenta) a přibližnou instrukci ke správnému vyplnění; druhá strana obsahuje stromový graf (*Metfesselova alokace*); třetí strana obsahuje matici intenzit preferencí (*Saatyho metoda*).

Cílovou skupinou byli studenti Přírodovědecké fakulty Univerzity Palackého v Olomouci ve věku 21-25 let. Dotazník byl vyvěšen na sociální síti Facebook ve skupině sdružující tyto studenty, dále byl rozdáván v budově fakulty a v přílehlých univerzitních stravovacích zařízeních. Takto bylo získáno 77 dotazníků, bohužel některé z nich nebyly zcela správně vyplněné. Proto bylo k analýze použito 70 dotazníků, v nichž jsou rovnoměrně zastoupena obě pohlaví.

3.2. Zpracování datového souboru

Jak již bylo zmíněno, výpočty a grafická zobrazení jsou prováděna ve statistickém programu R. Než v něm začneme pracovat, je zapotřebí nastavit zdrojovou složku v daném počítači, k čemuž použijeme příkaz:

```
> setwd("C:/bakalarska prace").
```

Pro práci s kompozičními daty je třeba otevřít příslušné knihovny:

```
> library(robCompositions), > library(compositions).
```

Tyto knihovny obsahují různé statistické metody pro kompoziční data. Kompoziční datové soubory jsou vyjádřeny ve stejných "jednotkách", proto nebude třeba škálování, které by jinak odstranilo jejich možný vliv na výsledky použitých statistických metod.

3.2.1. Datový soubor získaný Metfesselovou alokací

Na počátku načteme data, která jsou uložena v textovém souboru s názvem `Metfessel.txt`, jako matici o 70 řádcích (respondenti) a 6 sloupcích (kritéria):

```
> M=matrix(scan("Metfessel.txt"),ncol=6,byrow=T).
```

Dále pojmenujeme řádky, přičemž "+" značí studenty a "o" značí studentky:

```
> rownames(M)=c(rep("+",35),rep("o",35)),
```

i sloupce:

```
> colnames(M)=c("Znacka","Barva","Cena","Palivo","Spotreba",  
"Bezpecnost").
```

Tato tabulka tvoří Přílohu B.

Základní kompoziční číselné charakteristiky (centrum, variační matice a celkový rozptyl) jsou pak obsaženy v Příloze C.

Centrum získáme příkazem:

```
> g=mean(acomp(M)).
```

Lze jej zobrazit koláčovým grafem, použijeme k tomu následující příkaz:

```
> pie(g,main="Centrum-Metfessel",col=rainbow(6),cex.main=2,cex=1.5).
```

Variační matici vytvoříme příkazem:

```
> T=variation(acomp(M)).
```

K získání celkového rozptylu užijeme příkaz:

```
> totvar=sum(T)/(2*ncol(T)).
```

Nyní provedeme centrovanou logratio transformaci datového souboru:

```
> y=cenLR(M)$x.c1r.
```

Vzniklá tabulka tvoří Přílohu D.

Biplot v R nebudeme konstruovat přes singulární rozklady jak bylo popsáno výše, ale ekvivalentně použitím tzv. metody hlavních komponent, pro kterou R obsahuje velmi elegantní implementaci. Další zadaný příkaz:

```
> summary(princomp(y))
```

tak ukazuje, kolik procent celkové variability (po vynásobení 100) je vysvětleno pomocí jednotlivých hlavních komponent. Ty tvoří nové proměnné (skóry), použité při konstrukci biplotu. V případě biplotu nás zajímají pouze první dva číselné sloupce, které vyjadřují (z hlediska variability nejvýznamnější) hlavní komponenty. V našem případě tak získáme tabulku:

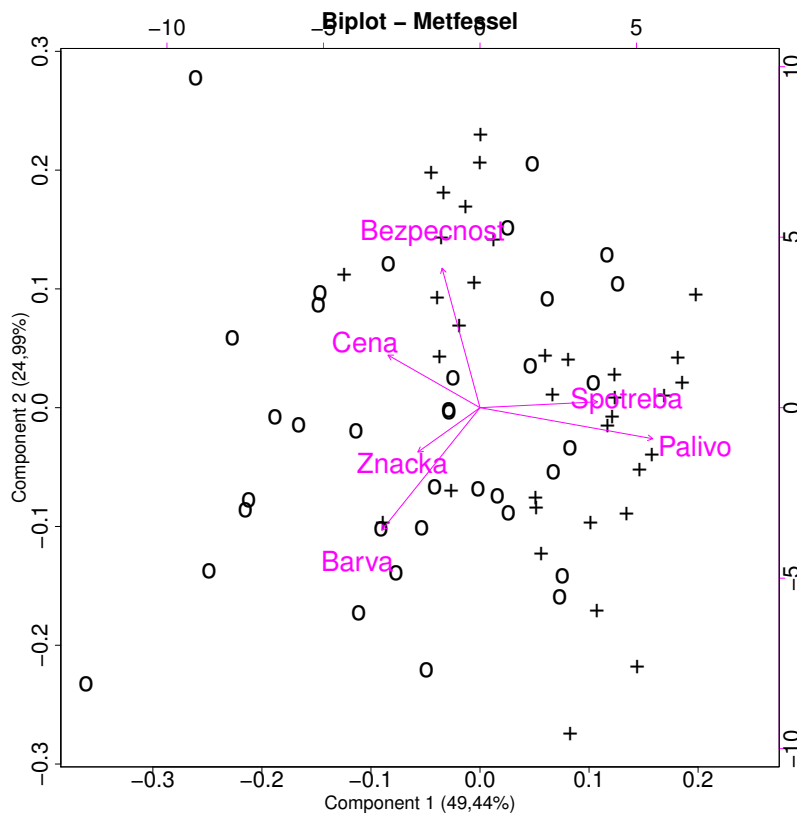
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.2352	0.8783	0.6330	0.4768	0.4011	0.0000
Proportion of Variance	0.4944	0.2499	0.1298	0.0737	0.0522	0.0000
Cumulative Proportion	0.4944	0.7443	0.8741	0.9478	1.0000	1.0000

V prvním řádku `Standard deviation` jsou uvedeny směrodatné odchylky hlavních komponent neboli singulární hodnoty k_i . Řádek `Proportion of Variance` říká, že první hlavní komponenta `Comp.1` vysvětluje přibližně 49,44% a druhá hlavní komponenta `Comp.2` dalších 24,99% celkové variability, v součtu to dává 74,43%. Tuto hodnotu lze vyčíst z řádku `Cumulative Proportion` ze sloupce s názvem `Comp.2` (číslo 0.7443), což lze považovat za velmi dobrý výsledek. A tedy biplot dobře znázorní reálnou strukturu vícerozměrného datového souboru.

Vykreslení tohoto biplotu zajistí příkaz:

```
> biplot(princomp(y), main="Biplot-Metfessel", col=c(1,14),  
xlab="Component 1 (49,44\\%)", ylab="Component 2 (24,99\\%)",  
xlim=c(-0.36,0.25), ylim=c(-0.28,0.28) cex.main=2.5, cex=3).
```

Nyní se podíváme, jak můžeme příslušný biplot interpretovat. Při pohledu



Obrázek 6: Kompoziční biplot.

do levé části biplotu (obr. 6) lze říci, že silně převažují ženy. Okolo konců paprsků *barva*, *značka* a *cena* vidíme shluk studentek, lze tedy říci, že u těchto osob převažují očekávané preference barvy, značky a ceny před technickými vlastnostmi. U konců paprsků *palivo* a *spotřeba* se nachází shluk studentů, což vyjadřuje, že preferují palivo a spotřebu před ostatními kritérii.

Co se týče směrů paprsků, tak je zřetelné, že paprsky (*značka-barva*) směřují opačně než paprsky (*palivo-spotřeba*), není tedy mezi nimi prakticky žádný vztah (ve smyslu variability podílů složek). Můžeme navíc říci, že čím více kupující preferuje estetické vlastnosti (barva, značka), tím více snižuje preference technických vlastností (spotřeba a typ paliva). Ve spojení s tím, že zleva doprava ubývá žen, odpovídá to předpokladu, že ženy preferují estetické vlastnosti a cenu před technickými vlastnostmi. Zaměříme se na délku spojnic a uvidíme,

že nejkratší spojnice mají dvojice (*spotřeba-palivo* a *značka-barva*). Znamená to, že jejich podíl (x_i/x_j) bude skrze celý soubor docela stabilní, existuje mezi nimi určitá "závislost" (přímá úměrnost).

Délky paprsků aproximují směrodatnou odchylku proměnných (clr transformovaných). Vidíme, že paprsky *bezpečnost*, *barva*, *palivo* jsou delší než-li paprsky *cena*, *spotřeba*, *značka*. A tedy mají nepatrně větší rozptyl hodnot.

Zadáním příkazu

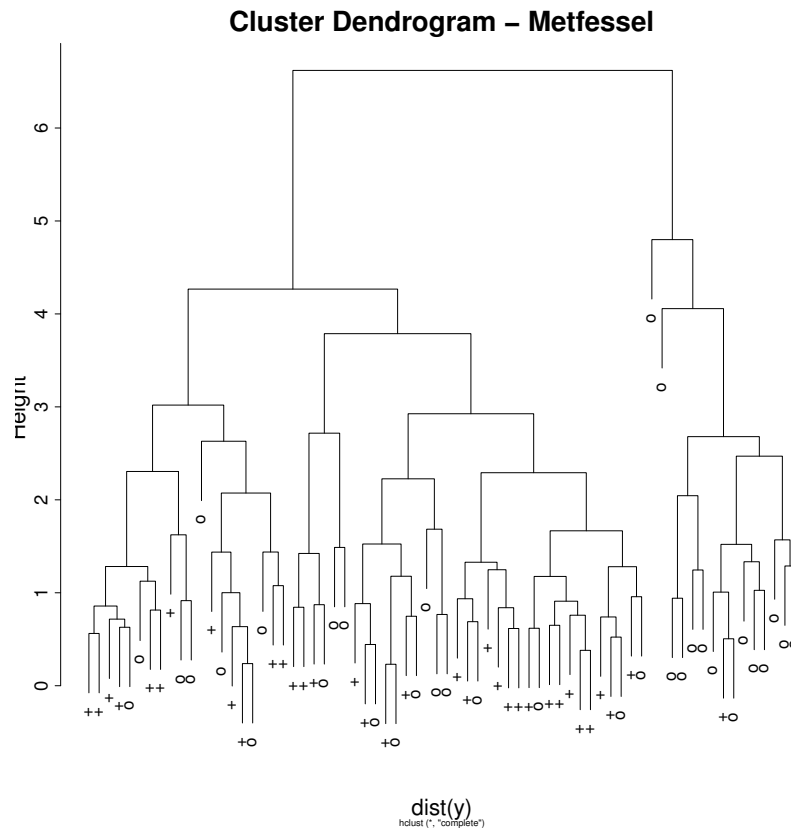
```
> hclust(dist(y)) získáme tabulku,
```

```
Cluster method   : complete
Distance         : euclidean
Number of objects: 70
```

jejíž první řádek obsahuje název použité shlukové metody (*complete*), druhý řádek uvádí typ vzdálenosti (*euclidean*) a třetí řádek udává počet objektů (70). Chceme-li shlukování zobrazit graficky a získat tak dendrogram, použijeme příkaz:

```
> plot(hclust(dist(y)), cex=1.4, cex.main=3, cex.axis=2, cex.lab=2.4,
main="Cluster Dendrogram-Metfessel", xlab=NULL, ylab="Height").
```

Při pohledu na dendrogram (obr. 7) jsou zřetelné dva shluky – vlevo velký shluk studentů a studentek dohromady a v pravé části menší shluk studentek s jedinou mužskou výjimkou.



Obrázek 7: Shlukový dendrogram.

3.2.2. Datový soubor získaný Saatyho metodou

Nejdříve načteme data uložená v textovém souboru s názvem `Saaty.txt` jako matici o 70 řádcích (respondenti) a 6 sloupcích (kritéria):

```
> S=matrix(scan("Saaty.txt"),ncol=6,byrow=T).
```

Dále pojmenujeme řádky i sloupce, kde "+" značí studenty a "o" značí studentky:

```
> rownames(S)=c(rep("+",35),rep("o",35)),
```

```
> colnames(S)=c("Znacka","Barva","Cena","Palivo","Spotreba",
"Bezpecnost").
```

Tato tabulka tvoří Přílohu E.

Základní kompoziční číselné charakteristiky (centrum, variační matice a celkový rozptyl) jsou pak shrnuty v Příloze F.

Centrum získáme příkazem:

```
> g=mean(acompl(S)).
```

Lze jej zobrazit koláčovým grafem, použijeme k tomu následující příkaz:

```
pie(g,main="Centrum-Saaty",col=rainbow(6),cex.main=2,cex=1.5).
```

Variační matici vytvoříme příkazem:

```
> T=variation(acompl(S)).
```

K získání celkového rozptylu užijeme příkaz:

```
> totvar=sum(T)/(2*ncol(T)).
```

Nyní provedeme centrovanou logratio transformaci datového souboru:

```
> y=cenLR(S)$x.clr.
```

Vzniklá tabulka tvoří Přílohu G.

Další zadaný příkaz:

```
> summary(princomp(y))
```

ukazuje, kolik procent celkové variability (po vynásobení 100) vysvětlují jednotlivé hlavní komponenty. V případě biplotu nás opět zajímají pouze první dva číselné sloupce, které vyjadřují (z hlediska variability nejvýznamnější) hlavní komponenty. V našem případě tak získáme tabulku:

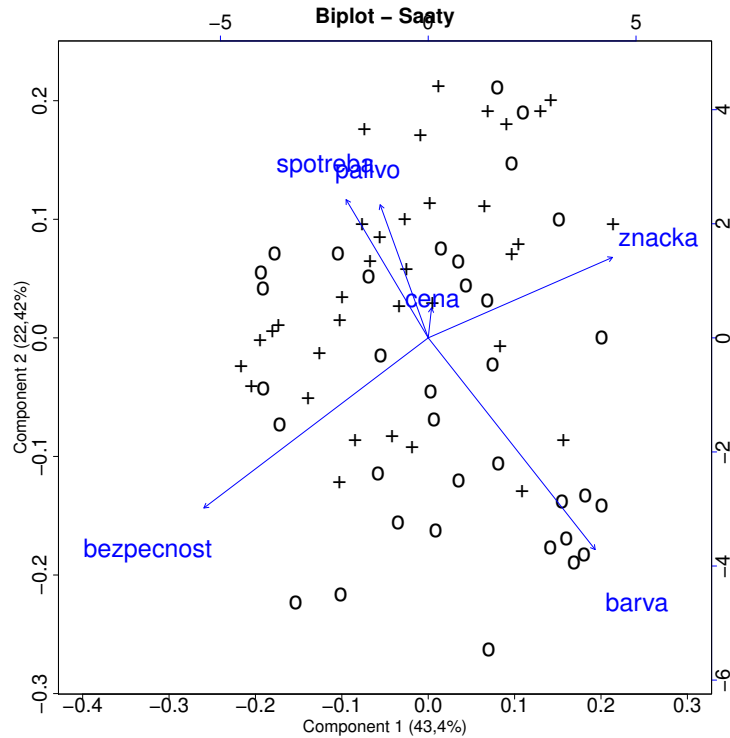
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.2516	0.8996	0.8259	0.6029	0.4337	0.0000
Proportion of Variance	0.4340	0.2242	0.1890	0.1007	0.0521	0.0000
Cumulative Proportion	0.4340	0.6582	0.8472	0.9479	1.0000	1.0000

V prvním řádku `Standard deviation` jsou uvedeny směrodatné odchylky hlavních komponent. Řádek `Proportion of Variance` nám říká, že první hlavní komponenta `Comp.1` vysvětluje přibližně 43,4% a druhá hlavní komponenta `Comp.2` dalších 22,42% celkové variability, v součtu to dává 65,82% celkové variability. Tuto hodnotu lze lehce vyčíst z třetího řádku `Cumulative Proportion` ze sloupce s názvem `Comp.2` (číslo 0.6582), což není špatný výsledek, ale je nižší než hodnota u předchozí metody, proto příslušný biplot bude hůře zobrazovat reálnou strukturu vícerozměrného datového souboru.

Vykreslení tohoto biplotu zajistí příkaz:

```
> biplot(princomp(y),main="Biplot-Saaty",cex.main=2.5,cex.lab=2,
```

```
xlab="Component 1 (43,4\%)",ylab="Component 2 (22,42\%)",cex=3,
cex.axis=2.4,xlim=c(-0.4,0.3),ylim=c(-0.28,0.23),col=c(1,4)).
```



Obrázek 8: Kompoziční biplot.

Při pohledu na spodní část biplotu (obr. 8) lze říci, že tu převažuje ženské pohlaví, především vidíme jednoznačný shluk studentek u konce paprsku *barva*. Lze tedy říci, že u těchto osob silná preference barvy dosti potlačuje ostatní preference. Jinak jsou ovšem obě pohlaví, narozdíl od předchozího biplotu, docela promíchaná.

Co se týče směrů paprsků, tak je zřetelné, že paprsky *palivo* a *spotřeba* směřují opačně než paprsek *barva*, což znamená, že mezi nimi není prakticky žádný proporcionálně stabilní vztah.

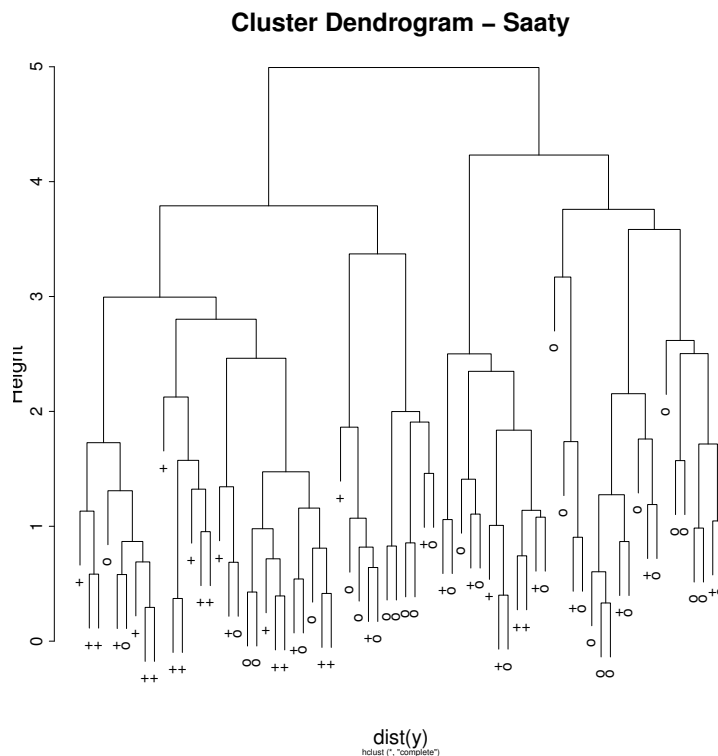
Dále lze říci, že, čím více kupující preferuje *barvu*, tím více se snižuje preference technických vlastností (*spotřeba* a *typ paliva*). Podobně pak můžeme ohodnotit též vztah složek *znacka* a *bezpečnost*.

Zaměříme se na délku spojnic a uvidíme, že nejkratší spojnici má dvojice (*spotřeba-palivo*). Znamená to, že jejich podíl (x_i/x_j) bude skrze celý soubor docela stabilní, existuje mezi nimi zřejmá přímá úměrnost.

Délky paprsků aproximují směrodatnou odchylku proměnných (clr transformovaných). Jednoznačně vidíme, že paprsky *bezpečnost* a *barva* jsou delší než-li paprsky *palivo*, *spotřeba* a *značka*. A tedy mají větší rozptyl hodnot a mají větší vliv na uspořádání pozorování v biplotu.

Celkově je ovšem nutno podotknout, že struktura pozorování a vztahy mezi proměnnými v biplotu získaném pomocí Saatyho metody se zdají poněkud méně intuitivní než v případě Metfesselovy alokace. To se ostatně projevuje i ve výsledcích shlukové analýzy. Příslušný dendrogram získáme opět příkazem:

```
> plot(hclust(dist(y)), cex.main=3, cex.lab=2.4, cex.axis=2, cex=1.8,
main="Cluster Dendrogram-Saaty", xlab=NULL, ylab="Height").
```



Obrázek 9: Shlukový dendrogram.

Při pohledu na tento dendrogram (obr. 9) vidíme, že shlukování vektorů vah dle pohlaví je jistě méně zřetelné než u předchozí metody.

3.3. Porovnání výstupů použitých metod

Shrňme si ještě jednou přehledně rozdíly mezi oběma metodami konstrukce vah kritérií. Při srovnání tabulek vysvětlené variability pomocí hlavních komponent si všimneme, že u datového souboru *Metfessel* první dvě hlavní komponenty vysvětlují 74,43% variability, přičemž u datového souboru *Saaty* 65,82%, tedy o 8,61% méně. To způsobí, že biplot první metody kvalitněji zobrazí reálnou strukturu dat.

	Datový soubor <i>Metfessel</i>		Datový soubor <i>Saaty</i>	
	Comp.1	Comp.2	Comp.1	Comp.2
Standard deviation	1.2352	0.8783	1.2516	0.8996
Proportion of Variance	0.4944	0.2499	0.4340	0.2242
Cumulative Proportion	0.4944	0.7443	0.4340	0.6582

Co se týká shlukové analýzy, podíváme-li se na dendrogram *Metfessel*, uvidíme početný shluk studentek v pravé části a několik méně početných, ovšem již ne tak přehledných, shluků jednotlivých pohlaví. U dendrogramu *Saaty* vidíme nepřehledné shluky.

Při pohledu na biploty jsou zřetelnější odlišnosti preferencí estetických vlastností (*značka, barva*) a *ceny* před technickými vlastnostmi (*bezpečnost, spotřeba, palivo*) a viditelnější shluky u paprsků předpokládaných preferencí studentek a studentů u biplotu souboru *Metfessel*. Dle mého názoru tento biplot vyjadřuje skutečné preference respondentů, naopak biplot souboru *Saaty* nevyjadřuje preference dotazovaných přesně tak, jak je skutečně cítí. Domníváme se, že je to způsobeno intuitivnějším zadáváním vah při *Metfesselově* alokaci než při *Saatyho* metodě. Při *Saatyho* metodě je zadávání složitější, poněvadž se tu "váhy" zadávají nepřímo přes hodnoty vyjadřující slovní hodnocení.

Závěr

V teoretické části této práce bylo mým cílem, aby se čtenáři seznámili s nejzákladnějšími teoretickými pojmy vícekritériálního rozhodování a hodnocení, zahrnující kritéria hodnocení a dvě metody stanovení jejich vah, a dále se specifickou geometrií a prací s kompozičními daty. Za cíl praktické části jsem si stanovil využít nastudované teorie při analýze dat získaných z dotazníku oběma metodami stanovení vah a co nejintuitivněji interpretovat jejich výstupy, které jsem získal ze statistického softwaru *R*.

Při tvorbě této práce pro mě bylo nejtěžší proniknout do teorie, porozumět odlišné geometrii a principům práce s kompozičními daty a dále pak práce s cizojazyčnou literaturou. Neméně složité byly počátky práce se dvěma softwary, typografickým systémem \TeX , ve kterém je tato práce vysázena, a statistickým softwarem *R*. Ale věřím, že tyto zkušenosti zúročím v dalším studiu nebo v budoucí pracovní praxi.

Jelikož o sobě vím, že se stydím vystupovat před více lidmi, tak jsem měl obavy před sbíráním dat zahrnující oslovování studentů. S každým dotázaným studentem a studentkou jsem se tohoto studu zbavoval a ke konci jsem si uvědomil, že již nemám problém s výstupem před publikem, což považuji za osobní vítězství.

Doufám, že svého cíle v teoretické i praktické části jsem dosáhl. Mým přáním je, aby tato práce probudila alespoň v hrstce čtenářů zájem dozvědět se více o této mladé, velmi zajímavé a v praxi široce využitelné oblasti statistické analýzy.

Literatura

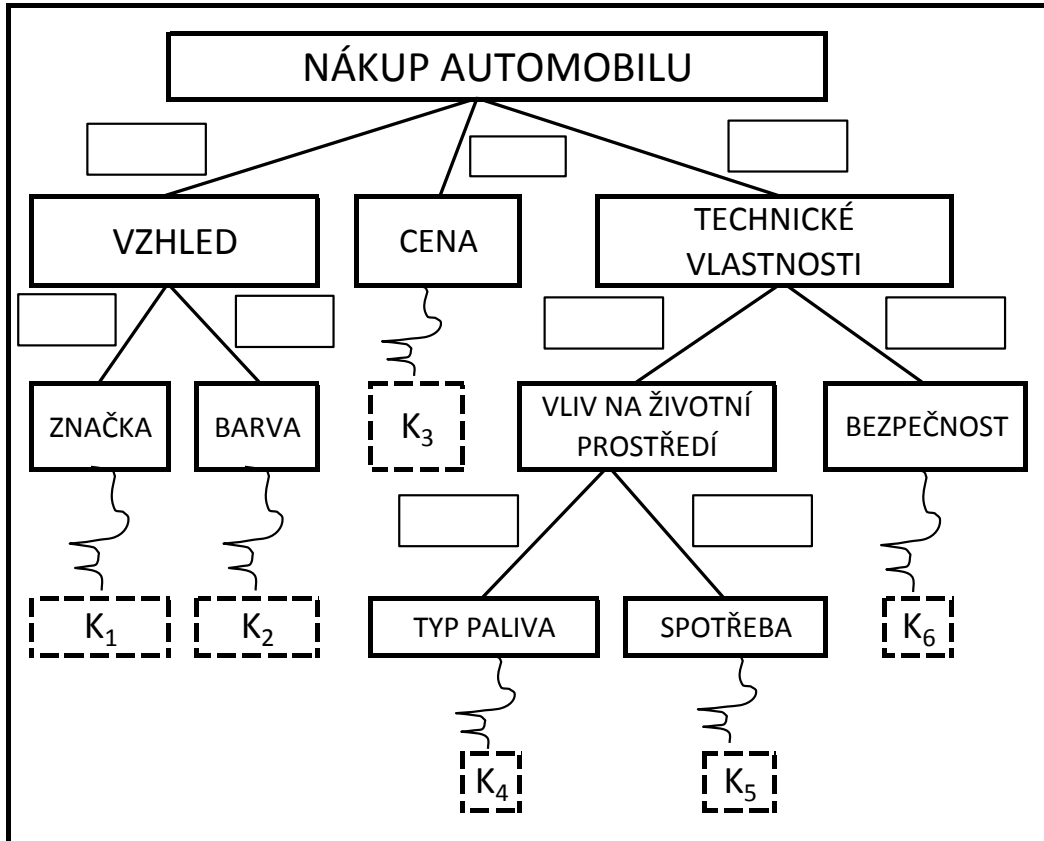
- [1] Aitchison, J.: The statistical analysis of compositional data. Chapman&Hall, London, 1986.
- [2] Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V.: Compositional data analysis in the geosciences. From theory to practice. Geological Society, London, 2006.
- [3] Hebák, P., Hustopecký, J., Pecáková, I., Průša, M., Řezanková, H., Svobodová, A., Vlach, P.: Vícerozměrné statistické metody 3. Informatorium, Praha, 2005.
- [4] Hron, K.: Elementy statistické analýzy kompozičních dat. Informační bulletin České statistické společnosti, 41-46, 2010.
- [5] Kalivodová, A.: Diplomová práce - Kompoziční biplot. Přf UPOL, Olomouc, 2012.
- [6] Meloun, M., Militký, J., Hill, M.: Statistická analýza vícerozměrných dat v příkladech. Academia, Praha, 2012.
- [7] Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R.: Lecture notes on compositional data analysis. Universitat de Girona, 2007.
- [8] Talašová, J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování. Univerzita Palackého, Olomouc, 2003.
- [9] Triantaphyllou, E.: Multi-criteria decision making methods: a comparative study. Kluwer Academic Publishers, Dordrecht, 2000.
- [10] T_EX Users Group, <http://www.tug.org> [online 14. 12. 2014].
- [11] The R project for statistical computing, <http://www.r-project.org> [online 20. 12. 2014].
- [12] <http://cs.jhu.edu/~razvanm/fs-expedition/tux3.html> [online 12. 1. 2015].

Přílohy

Příloha A - Dotazník

1. strana

Dotazník k bakalářské práci Kompoziční analýza vah ve vícekritériálním hodnocení variant	
<p>Dotazník obsahuje dvě metody stanovení vah. Prosím, přečtěte si pozorně zadání každé metody. Dotazník je anonymní, pouze uveďte pohlaví a věk. Děkuji za Váš čas, který věnujete vyplnění tohoto dotazníku.</p>	
Pohlaví:	<input type="text"/>
Věk:	<input type="text"/>
Kritéria	
K1 - Značka	Škoda, Volkswagen, Peugeot, Citroen, BMW, atd.
K2 - Barva	Barva karoserie - modrá, bílá, červená, fialová, atd.
K3 - Cena	Cena vozidla v Kč
K4 - Palivo	Benzín, nafta, elektromotor
K5 - Spotřeba	Spotřeba paliva
K6 - Bezpečnost	Bezpečnost vozidla - ABS, počet airbagů, typ karoserie, atd.
Metfesselova alokace	
Vepište do okének NENULOVÉ váhy dle vašich preferencí. Součet vah každého větvení každé úrovně musí být roven 1!!!	
Saatyho metoda	
Stačí vyplnit horní, respektive dolní trojúhelník matice S . Doplňte dle vašich preferencí hodnoty z tabulky s jazykovými deskriptory.	



Saatyho metoda						
S	K_1 - značka	K_2 - barva	K_3 - cena	K_4 - palivo	K_5 - spotřeba	K_6 - bezpečnost
K_1 - značka	1					
K_2 - barva		1				
K_3 - cena			1			
K_4 - palivo				1		
K_5 - spotřeba					1	
K_6 - bezpečnost						1
s_{ij}	Jazykový deskriptor					
1	K_i je stejně významné jako K_j					
3	K_i je slabě významnější než K_j					
5	K_i je dosti významnější než K_j					
7	K_i je prokazatelně významnější než K_j					
9	K_i je absolutně významnější než K_j					
Pokud je K_j významnější než K_i , tak přiřadte obrácenou hodnotu dle vztahu $s_{ij} = \frac{1}{s_{ji}}$						

Příloha B

Datový soubor vah získaných Metfesselovou alokací.
 Studenti jsou označení (+) a studentky (o).

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
+	0.32	0.08	0.20	0.08	0.12	0.20
+	0.125	0.125	0.35	0.036	0.084	0.28
+	0.24	0.06	0.40	0.027	0.063	0.21
+	0.14	0.06	0.60	0.012	0.028	0.16
+	0.1625	0.0875	0.35	0.08	0.12	0.20
+	0.12	0.08	0.30	0.015	0.135	0.35
+	0.20	0.30	0.30	0.024	0.056	0.12
+	0.16	0.04	0.40	0.024	0.056	0.32
+	0.24	0.06	0.30	0.072	0.168	0.16
+	0.15	0.05	0.40	0.042	0.078	0.28
+	0.17	0.03	0.50	0.027	0.063	0.21
+	0.15	0.05	0.50	0.0225	0.1275	0.15
+	0.12	0.03	0.35	0.025	0.075	0.40
+	0.06	0.04	0.50	0.04	0.04	0.32
+	0.24	0.06	0.25	0.18	0.18	0.09
+	0.2925	0.1575	0.15	0.162	0.198	0.04
+	0.42	0.18	0.15	0.09	0.135	0.025
+	0.36	0.09	0.30	0.0875	0.0875	0.075
+	0.26	0.14	0.40	0.044	0.066	0.09
+	0.16	0.04	0.30	0.20	0.20	0.10
+	0.27	0.03	0.20	0.15	0.225	0.125
+	0.315	0.135	0.15	0.112	0.168	0.12
+	0.20	0.05	0.30	0.072	0.288	0.09
+	0.18	0.12	0.40	0.135	0.135	0.03
+	0.12	0.03	0.25	0.135	0.315	0.15
+	0.21	0.14	0.30	0.0525	0.21	0.0875
+	0.225	0.225	0.20	0.0735	0.1715	0.105
+	0.15	0.10	0.30	0.126	0.189	0.135
+	0.15	0.15	0.20	0.12	0.28	0.10
+	0.11	0.09	0.20	0.168	0.252	0.18
+	0.195	0.105	0.40	0.0315	0.0735	0.195
+	0.12	0.03	0.55	0.0135	0.0765	0.21
+	0.15	0.10	0.25	0.105	0.245	0.15
+	0.165	0.135	0.15	0.154	0.231	0.165
+	0.08	0.12	0.20	0.078	0.312	0.21

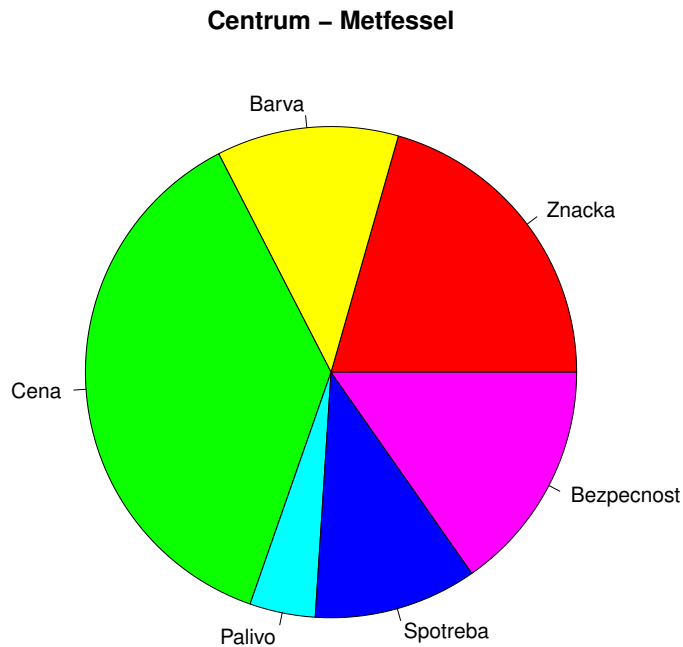
	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
o	0.18	0.02	0.40	0.032	0.128	0.24
o	0.21	0.09	0.30	0.06	0.14	0.20
o	0.18	0.12	0.40	0.036	0.084	0.18
o	0.09	0.06	0.25	0.12	0.12	0.36
o	0.16	0.04	0.30	0.09	0.21	0.20
o	0.0225	0.1275	0.60	0.0125	0.05	0.1875
o	0.2334	0.10	0.3333	0.015	0.0183	0.30
o	0.08	0.12	0.20	0.072	0.108	0.42
o	0.2333	0.10	0.3333	0.0067	0.06	0.2667
o	0.40	0.40	0.15	0.0035	0.0015	0.045
o	0.14	0.21	0.35	0.0045	0.0255	0.27
o	0.16	0.24	0.35	0.015	0.06	0.175
o	0.20	0.20	0.20	0.036	0.084	0.28
o	0.0975	0.0525	0.80	0.005	0.02	0.025
o	0.32	0.08	0.50	0.006	0.024	0.07
o	0.09	0.01	0.80	0.001	0.009	0.09
o	0.20	0.20	0.25	0.084	0.196	0.07
o	0.16	0.24	0.50	0.006	0.024	0.07
o	0.275	0.275	0.25	0.028	0.042	0.13
o	0.42	0.28	0.20	0.003	0.027	0.07
o	0.22	0.33	0.30	0.0045	0.0405	0.105
o	0.16	0.04	0.40	0.032	0.128	0.24
o	0.16	0.04	0.70	0.032	0.048	0.02
o	0.195	0.105	0.40	0.105	0.105	0.09
o	0.18	0.12	0.60	0.018	0.072	0.01
o	0.135	0.315	0.45	0.018	0.042	0.04
o	0.24	0.16	0.40	0.04	0.06	0.10
o	0.20	0.20	0.30	0.12	0.12	0.06
o	0.10	0.30	0.40	0.024	0.096	0.08
o	0.25	0.25	0.30	0.014	0.126	0.06
o	0.175	0.175	0.15	0.0825	0.1925	0.225
o	0.1575	0.1925	0.25	0.02	0.18	0.20
o	0.20	0.30	0.10	0.064	0.096	0.24
o	0.10	0.10	0.30	0.07	0.28	0.15
o	0.20	0.20	0.20	0.028	0.252	0.12

Příloha C - Číselné charakteristiky souboru Metfessel

Charakteristika polohy pro kompozice se nazývá **centrum**:

$$\mathbf{g} = C(0.2058, 0.1198, 0.3709, 0.0432, 0.1075, 0.1527)^T$$

Grafické znázornění centra:



Variabilita souboru kompozic je charakterizována **variační maticí** rozptylů logaritmů podílů jednotlivých dvojic složek:

$$\mathbf{T} = \begin{pmatrix} 0.0000 & 0.6239 & 0.5717 & 1.4840 & 1.1409 & 1.0342 \\ 0.6239 & 0.0000 & 1.0943 & 2.0592 & 1.6058 & 1.4737 \\ 0.5717 & 1.0943 & 0.0000 & 1.9957 & 1.3672 & 0.8619 \\ 1.4840 & 2.0592 & 1.9957 & 0.0000 & 0.4740 & 1.7794 \\ 1.1409 & 1.6058 & 1.3672 & 0.4740 & 0.0000 & 1.2200 \\ 1.0342 & 1.4737 & 0.8619 & 1.7794 & 1.2200 & 0.0000 \end{pmatrix}$$

Celkový rozptyl souboru:

$$\text{totvar}(\mathbf{X}) = 3.1311$$

Příloha D

Clr transformovaný datový soubor vah získaných Metfesselovou alokací.

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
+	0.7822	-0.6040	0.3122	-0.6040	-0.1985	0.3122
+	-0.0323	-0.0323	0.9973	-1.2770	-0.4297	0.7741
+	0.7552	-0.6310	1.2660	-1.4295	-0.5822	0.6216
+	0.5541	-0.2931	2.0093	-1.9026	-1.0553	0.6876
+	0.1093	-0.5097	0.8765	-0.5993	-0.1938	0.3169
+	0.0633	-0.3420	0.9796	-2.0160	0.1811	1.1338
+	0.5155	0.9209	0.9209	-1.6047	-0.7574	0.0046
+	0.4539	-0.9323	1.3702	-1.4431	-0.5958	1.1471
+	0.5215	-0.8647	0.7446	-0.6824	0.1648	0.1160
+	0.2367	-0.8618	1.2175	-1.0362	-0.4171	0.8609
+	0.5461	-1.1884	1.6249	-1.2937	-0.4464	0.7574
+	0.3257	-0.7728	1.5296	-1.5714	0.1631	0.3257
+	0.1917	-1.1945	1.2621	-1.3768	-0.2782	1.3957
+	-0.4296	-0.8351	1.6906	-0.8351	-0.8351	1.2443
+	0.4836	-0.9026	0.5244	0.1959	0.1959	-0.4972
+	0.7095	0.0905	0.0417	0.1187	0.3193	-1.2799
+	1.2289	0.3816	0.1993	-0.3114	0.0939	-1.5924
+	0.9943	-0.3919	0.8120	-0.4201	-0.4201	-0.5742
+	0.7327	0.1137	1.1635	-1.0437	-0.6382	-0.3280
+	0.1302	-1.2560	0.7588	0.3533	0.3533	-0.3397
+	0.6729	-1.5243	0.3728	0.0851	0.4906	-0.0971
+	0.7028	-0.1444	-0.0391	-0.3312	0.0742	-0.2622
+	0.4060	-0.9802	0.8115	-0.6155	0.7707	-0.3924
+	0.3290	-0.0764	1.1275	0.0413	0.0413	-1.4627
+	-0.1089	-1.4952	0.6250	0.0088	0.8561	0.1141
+	0.3850	-0.0203	0.7417	-1.0012	0.3850	-0.4903
+	0.3783	0.3783	0.2605	-0.7404	0.1068	-0.3837
+	-0.0398	-0.4453	0.6533	-0.2142	0.1912	-0.1452
+	-0.0472	-0.0472	0.2404	-0.2703	0.5769	-0.4526
+	-0.3570	-0.5576	0.2408	0.0664	0.4719	0.1354
+	0.4498	-0.1691	1.1683	-1.3731	-0.5258	0.4498
+	0.3232	-1.0630	1.8456	-1.8615	-0.1269	0.8828
+	-0.0398	-0.4453	0.4709	-0.3965	0.4507	-0.0398
+	0.0047	-0.1959	-0.0905	-0.0642	0.3412	0.0047
+	-0.6037	-0.1982	0.3125	-0.6290	0.7572	0.3613

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
o	0.5298	-1.6673	1.3283	-1.1973	0.1889	0.8175
o	0.3662	-0.4810	0.7229	-0.8864	-0.0391	0.3174
o	0.3297	-0.0757	1.1282	-1.2796	-0.4323	0.3297
o	-0.4296	-0.8351	0.5920	-0.1419	-0.1419	0.9566
o	0.1396	-1.2466	0.7682	-0.4357	0.4115	0.3628
o	-1.2248	0.5097	2.0585	-1.8126	-0.4263	0.8954
o	0.9218	0.0742	1.2780	-1.8228	-1.6240	1.1728
o	-0.5291	-0.1236	0.3871	-0.6344	-0.2290	1.1291
o	0.8774	0.0303	1.2341	-2.6727	-0.4805	1.0112
o	2.2483	2.2483	1.2675	-2.4903	-3.3376	0.0635
o	0.5269	0.9324	1.4432	-2.9105	-1.1759	1.1837
o	0.3450	0.7504	1.1277	-2.0221	-0.6358	0.4346
o	0.3743	0.3743	0.3743	-1.3404	-0.4931	0.7107
o	0.7382	0.1192	2.8430	-2.2321	-0.8458	-0.6226
o	1.5044	0.1181	1.9507	-2.4721	-1.0858	-0.0153
o	1.1358	-1.0614	3.3206	-3.3640	-1.1667	1.1358
o	0.2857	0.2857	0.5088	-0.5817	0.2655	-0.7640
o	0.7437	1.1491	1.8831	-2.5396	-1.1534	-0.0829
o	0.8347	0.8347	0.7393	-1.4498	-1.0443	0.0854
o	1.7708	1.3654	1.0289	-3.1707	-0.9735	-0.0208
o	0.9343	1.3397	1.2444	-2.9552	-0.7580	0.1946
o	0.3161	-1.0701	1.2324	-1.2932	0.0930	0.7216
o	0.8005	-0.5857	2.2764	-0.8088	-0.4034	-1.2789
o	0.3186	-0.3003	1.0371	-0.3003	-0.3003	-0.4545
o	0.8851	0.4796	2.0890	-1.4174	-0.0311	-2.0052
o	0.3912	1.2385	1.5952	-1.6236	-0.7763	-0.8251
o	0.6580	0.2525	1.1688	-1.1337	-0.7282	-0.2174
o	0.3033	0.3033	0.7088	-0.2074	-0.2074	-0.9006
o	-0.1323	0.9663	1.2539	-1.5594	-0.1731	-0.3554
o	0.8020	0.8020	0.9843	-2.0803	0.1168	-0.6250
o	0.0932	0.0932	-0.0608	-0.6587	0.1885	0.3445
o	0.1714	0.3720	0.6334	-1.8922	0.3049	0.4103
o	0.3297	0.7352	-0.3633	-0.8096	-0.4041	0.5121
o	-0.3628	-0.3628	0.7357	-0.7195	0.6667	0.0426
o	0.3743	0.3743	0.3743	-1.5918	0.6054	-0.1365

Příloha E

Datový soubor vah získaných Saatyho metodou.

Studenti jsou označení (+) a studentky (o).

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
+	0.2676	0.1759	0.1665	0.1098	0.1703	0.1099
+	0.0578	0.0257	0.1423	0.0805	0.2430	0.4507
+	0.2788	0.0202	0.1205	0.0960	0.1795	0.3050
+	0.0579	0.0264	0.4357	0.0541	0.1174	0.3085
+	0.4137	0.1087	0.2434	0.0243	0.0425	0.1674
+	0.0618	0.0605	0.2752	0.0323	0.1430	0.4272
+	0.0765	0.0700	0.2808	0.0594	0.1415	0.3718
+	0.1642	0.0525	0.3838	0.0493	0.0459	0.3043
+	0.2434	0.0242	0.4137	0.0426	0.1674	0.1087
+	0.0531	0.0216	0.2028	0.0879	0.1608	0.4738
+	0.0254	0.0397	0.1387	0.1949	0.2148	0.3865
+	0.1190	0.0464	0.4922	0.0220	0.2238	0.0966
+	0.1627	0.0441	0.3325	0.0245	0.0971	0.3391
+	0.0743	0.0760	0.1055	0.3701	0.2416	0.1325
+	0.3402	0.0314	0.3238	0.1269	0.1269	0.0508
+	0.3646	0.0820	0.0820	0.1538	0.2892	0.0284
+	0.4894	0.1312	0.1712	0.0625	0.1238	0.0219
+	0.4201	0.0405	0.2988	0.0657	0.1508	0.0241
+	0.2806	0.0405	0.3921	0.0380	0.1858	0.0630
+	0.1374	0.0295	0.4303	0.0971	0.2362	0.0695
+	0.2892	0.0236	0.2369	0.0966	0.2819	0.0718
+	0.4373	0.0604	0.1981	0.0331	0.1981	0.0730
+	0.2204	0.0217	0.3860	0.0380	0.1886	0.1453
+	0.1825	0.0455	0.4693	0.0664	0.2090	0.0273
+	0.0860	0.0223	0.2915	0.0766	0.4314	0.0922
+	0.2315	0.0661	0.4279	0.0270	0.2048	0.0427
+	0.2567	0.2567	0.2567	0.0241	0.1424	0.0634
+	0.1018	0.0536	0.3562	0.0473	0.3273	0.1138
+	0.0835	0.0691	0.3619	0.0594	0.2931	0.1330
+	0.0816	0.0380	0.2346	0.0400	0.3462	0.2596
+	0.0400	0.0199	0.1424	0.0529	0.1594	0.5854
+	0.0366	0.0187	0.1203	0.0584	0.1929	0.5731
+	0.0949	0.0435	0.2182	0.0521	0.4310	0.1603
+	0.0839	0.0335	0.3060	0.0477	0.2908	0.2381
+	0.0313	0.0626	0.1846	0.0475	0.3862	0.2878

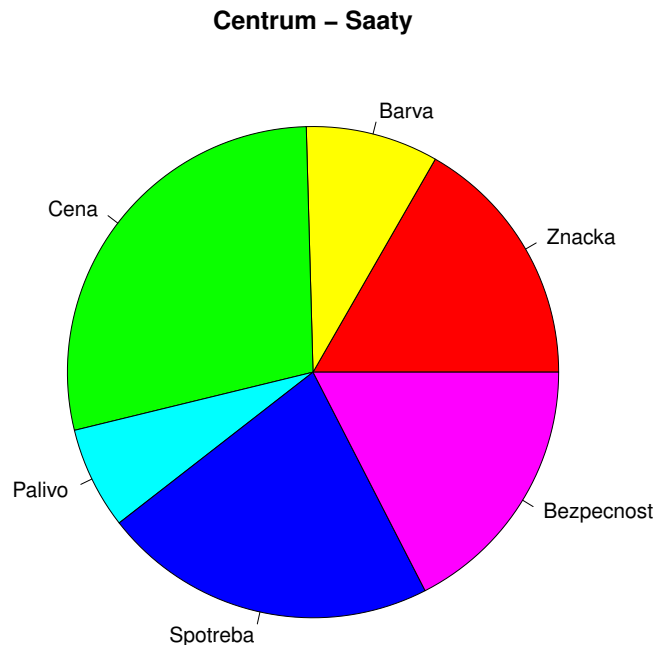
	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
o	0.0471	0.0176	0.1664	0.0648	0.3289	0.3752
o	0.0398	0.1090	0.2033	0.0492	0.0704	0.5283
o	0.0992	0.0347	0.3858	0.0485	0.2530	0.1788
o	0.0405	0.0259	0.1419	0.1769	0.2166	0.3982
o	0.2552	0.4619	0.0254	0.1489	0.0430	0.0656
o	0.0234	0.0665	0.2084	0.0251	0.0870	0.5896
o	0.0602	0.0214	0.1342	0.0232	0.1735	0.5875
o	0.0425	0.0301	0.0920	0.2337	0.2877	0.3140
o	0.0333	0.0427	0.0572	0.1570	0.1985	0.5113
o	0.3456	0.3456	0.1199	0.0594	0.0275	0.1020
o	0.0684	0.0962	0.4925	0.0449	0.2518	0.0462
o	0.1229	0.2778	0.4547	0.0197	0.0547	0.0702
o	0.2661	0.2661	0.0374	0.0549	0.2157	0.1598
o	0.1572	0.0446	0.3759	0.0270	0.3102	0.0851
o	0.3170	0.0440	0.4082	0.0232	0.1140	0.0936
o	0.1950	0.0438	0.4427	0.0192	0.1040	0.1953
o	0.2615	0.2615	0.2706	0.0514	0.1268	0.0282
o	0.1778	0.2637	0.4202	0.0205	0.0467	0.0711
o	0.2610	0.2610	0.2983	0.0285	0.0652	0.0860
o	0.4791	0.2210	0.1620	0.0186	0.0466	0.0727
o	0.1984	0.3237	0.3326	0.0205	0.0475	0.0773
o	0.0809	0.0231	0.3821	0.0444	0.2710	0.1985
o	0.4098	0.0355	0.2965	0.0722	0.1505	0.0355
o	0.3852	0.0458	0.1032	0.2100	0.2100	0.0458
o	0.1423	0.0737	0.2995	0.0409	0.4190	0.0246
o	0.0396	0.3615	0.1180	0.0396	0.3615	0.0798
o	0.2958	0.0850	0.0986	0.0986	0.2958	0.1262
o	0.2213	0.2213	0.0864	0.2213	0.2213	0.0284
o	0.0557	0.1976	0.4204	0.0284	0.1351	0.1628
o	0.1307	0.1307	0.2466	0.0343	0.2740	0.1837
o	0.1019	0.1019	0.1187	0.0287	0.2917	0.3571
o	0.0652	0.1338	0.3235	0.0213	0.1941	0.2621
o	0.1870	0.3412	0.0712	0.0246	0.0827	0.2933
o	0.0583	0.0583	0.4076	0.0261	0.3049	0.1448
o	0.1735	0.1643	0.1735	0.0289	0.3763	0.0835

Příloha F - Číselné charakteristiky souboru Saaty

Charakteristika polohy pro kompozice se nazývá **centrum**:

$$\mathbf{g} = C(0.1671, 0.0873, 0.2837, 0.0671, 0.2202, 0.1746)^T$$

Grafické znázornění centra:



Variabilita souboru kompozic je charakterizována **variační maticí** rozptylů logaritmů podílů jednotlivých dvojic složek:

$$\mathbf{T} = \begin{pmatrix} 0.0000 & 1.0704 & 1.0748 & 1.2994 & 1.4635 & 2.6537 \\ 1.0704 & 0.0000 & 1.5205 & 1.6603 & 1.7923 & 2.3000 \\ 1.0748 & 1.5205 & 0.0000 & 1.3973 & 0.8430 & 1.4979 \\ 1.2994 & 1.6603 & 1.3973 & 0.0000 & 0.6732 & 1.4404 \\ 1.4635 & 1.7923 & 0.8430 & 0.6732 & 0.0000 & 1.2851 \\ 2.6537 & 2.3000 & 1.4979 & 1.4404 & 1.2851 & 0.0000 \end{pmatrix}$$

Celkový rozptyl souboru:

$$\text{totvar}(\mathbf{X}) = 3.6621$$

Příloha G

Clr transformovaný datový soubor vah získaných Saatyho metodou.

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
+	0.5211	0.1015	0.0466	-0.3697	0.0691	-0.3687
+	-0.6519	-1.4624	0.2490	-0.3206	0.7841	1.4018
+	0.8133	-1.8114	-0.0254	-0.2527	0.3730	0.9032
+	-0.5908	-1.3761	1.4274	-0.6586	0.1160	1.0821
+	1.3136	-0.0228	0.7832	-1.5209	-0.9619	0.4089
+	-0.5992	-0.6205	0.8942	-1.2481	0.2396	1.3340
+	-0.5257	-0.6145	0.7745	-0.7787	0.0892	1.0552
+	0.3586	-0.7815	1.2077	-0.8444	-0.9159	0.9756
+	0.7835	-1.5248	1.3139	-0.9593	0.4092	-0.0225
+	-0.7068	-1.6063	0.6331	-0.2028	0.4011	1.4817
+	-1.5065	-1.0599	0.1910	0.5311	0.6284	1.2158
+	0.1311	-0.8106	1.5509	-1.5569	0.7628	-0.0773
+	0.3776	-0.9278	1.0923	-1.5156	-0.1385	1.1120
+	-0.6227	-0.6001	-0.2721	0.9829	0.5564	-0.0442
+	1.0510	-1.3317	1.0015	0.0648	0.0648	-0.8506
+	1.1052	-0.3868	-0.3868	0.2420	0.8735	-1.4471
+	1.4843	0.1678	0.4339	-0.5736	0.1098	-1.6223
+	1.4030	-0.9361	1.0623	-0.4523	0.3784	-1.4552
+	0.9177	-1.0178	1.2523	-1.0816	0.5054	-0.5760
+	0.1473	-1.3911	1.2888	-0.1998	0.6890	-0.5342
+	0.8701	-1.6357	0.6706	-0.2264	0.8445	-0.5231
+	1.3224	-0.6572	0.5305	-1.2586	0.5305	-0.4677
+	0.6813	-1.6367	1.2417	-1.0765	0.5255	0.2646
+	0.5366	-0.8523	1.4811	-0.4744	0.6722	-1.3632
+	-0.2395	-1.5893	0.9811	-0.3553	1.3731	-0.1699
+	0.7667	-0.4866	1.3811	-1.3819	0.6442	-0.9235
+	0.7255	0.7255	0.7255	-1.6401	0.1363	-0.6728
+	-0.1873	-0.8287	1.0651	-0.9538	0.9805	-0.0758
+	-0.4429	-0.6322	1.0235	-0.7835	0.8126	0.0225
+	-0.3635	-1.1278	0.6924	-1.0765	1.0816	0.7937
+	-0.8195	-1.5176	0.4502	-0.5399	0.5630	1.8639
+	-0.8998	-1.5713	0.2901	-0.4325	0.7623	1.8512
+	-0.2483	-1.0284	0.5841	-0.8480	1.2648	0.2758
+	-0.3495	-1.2676	0.9444	-0.9142	0.8934	0.6935
+	-1.2693	-0.5762	0.5052	-0.8522	1.2433	0.9492

	Značka	Barva	Cena	Palivo	Spotřeba	Bezpečnost
o	-0.7692	-1.7536	0.4928	-0.4502	1.1742	1.3059
o	-1.0010	0.0064	0.6297	-0.7890	-0.4307	1.5847
o	-0.1862	-1.2366	1.1719	-0.9018	0.7499	0.4028
o	-1.0405	-1.4876	0.2132	0.4337	0.6361	1.2450
o	0.8986	1.4919	-1.4086	0.3599	-0.8821	-0.4597
o	-1.3068	-0.2624	0.8798	-1.2367	0.0063	1.9198
o	-0.3584	-1.3927	0.4432	-1.3119	0.7000	1.9197
o	-1.0073	-1.3523	-0.2350	0.6971	0.9050	0.9925
o	-1.1428	-0.8941	-0.6018	0.4078	0.6424	1.5885
o	1.0951	1.0951	0.0365	-0.6658	-1.4359	-0.1251
o	-0.4675	-0.1264	1.5066	-0.8884	0.8357	-0.8599
o	0.1794	0.9949	1.4876	-1.6513	-0.6300	-0.3806
o	0.7100	0.7100	-1.2521	-0.8682	0.5000	0.2001
o	0.3472	-0.9125	1.2190	-1.4144	1.0269	-0.2664
o	1.0965	-0.8781	1.3493	-1.5182	0.0738	-0.1233
o	0.6031	-0.8902	1.4229	-1.7149	-0.0255	0.6046
o	0.7572	0.7572	0.7914	-0.8695	0.0334	-1.4698
o	0.5265	0.9207	1.3866	-1.6336	-0.8103	-0.3899
o	0.7630	0.7630	0.8966	-1.4515	-0.6240	-0.3471
o	1.5537	0.7800	0.4694	-1.6949	-0.7765	-0.3317
o	0.6059	1.0955	1.1226	-1.6638	-0.8235	-0.3366
o	-0.3009	-1.5543	1.2515	-0.9009	0.9079	0.5966
o	1.3256	-1.1204	1.0020	-0.4105	0.3239	-1.1204
o	1.1315	-0.9979	-0.1855	0.5249	0.5249	-0.9979
o	0.3059	-0.3519	1.0501	-0.9408	1.3859	-1.4492
o	-1.0359	1.1755	0.0559	-1.0359	1.1755	-0.3352
o	0.7160	-0.5310	-0.3826	-0.3826	0.7160	-0.1358
o	0.4989	0.4989	-0.4415	0.4989	0.4989	-1.5541
o	-0.7620	0.5041	1.2591	-1.4356	0.1239	0.3104
o	-0.0629	-0.0629	0.5719	-1.4007	0.6772	0.2774
o	-0.1985	-0.1985	-0.0459	-1.4656	0.8531	1.0554
o	-0.6140	0.1048	0.9877	-1.7327	0.4769	0.7772
o	0.4597	1.0610	-0.5058	-1.5686	-0.3561	0.9098
o	-0.6175	-0.6175	1.3271	-1.4211	1.0368	0.2922
o	0.3006	0.2461	0.3006	-1.4916	1.0748	-0.4306

Příloha H

Obsah **CD-ROM**:

- Vyplněné dotazníky.
- Výpisy provedených výpočtů v softwaru R.
- Saatyho matice od respondentů.