

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

BAKALÁŘSKÁ PRÁCE

Statistická analýza kompozičních dat pomocí
softwaru CoDaPack



Vedoucí bakalářské práce:
RNDr. Karel Hron, Ph.D.
Rok odevzdání: 2013

Vypracovala:
Lenka Těthalová
ME, III. ročník

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením pana RNDr. Karla Hrona, Ph.D. s použitím uvedené literatury.

V Olomouci dne 18. dubna 2013

Poděkování

Ráda bych poděkovala vedoucímu bakalářské práce panu RNDr. Karlu Hronovi, Ph.D. za spolupráci, čas i neocenitelnou pomoc, kterou mi věnoval při konzultacích.

Obsah

Úvod	4
1 Kompoziční data	5
1.1 Definice a základní vlastnosti	5
1.2 Aitchisonova geometrie na simplexu	6
2 CoDaPack	9
2.1 Práce se soubory	10
2.2 Operace s daty	13
2.3 Statistická analýza	18
2.4 Grafy	22
3 Praktický příklad	30
Závěr	38
Literatura	39
4 Přílohy	40
4.1 Složení volného času dětí ve věku 10 – 11 let	40
4.2 Geochemické složení hornin	42

Úvod

Téma mé bakalářské práce je Statistická analýza kompozičních dat pomocí softwaru CoDaPack. Cílem této práce je podat uživateli jednoduše a srozumitelně informace o tomto softwaru a popsat způsob práce s ním. Tento program byl vytvořen speciálně pro kompoziční data, která nejdou popsat pomocí euklidovské geometrie a nemůžeme na ně tudíž přímo aplikovat standardní statistické metody. Protože jsou kompoziční data poměrně mladý obor, i software CoDaPack byl vytvořen teprve nedávno a stále se vyvíjí.

Práci jsem rozdělila do několika částí, jsou to: Kompoziční data, CoDaPack a Praktický příklad.

Kompoziční data obsahují definice a základní vlastnosti kompozičních dat a popisují, pro mě do této doby neznámou, Aitchisonovu geometrii.

CoDaPack byl pro mě také úplně neznámý software. Po jeho instalaci jsem však zjistila, že není vůbec těžké s ním pracovat. V této kapitole nejprve popíši základní funkce programu a následné statistické operace budu interpretovat na příkladech či přímo v CoDaPacku zpracovávat a ilustrovat pomocí grafů. Statistické metody a jejich grafy budu popisovat na vzorovém „školském“ příkladě, který se typicky užívá při seznamování se s kompozičními daty, tím spíše se softwarem pro jejich zpracování.

Praktický příklad budu prezentovat na reálných datech, která jsem získala z průzkumu prováděného u dětí ve věku 10 – 11 let, jež se zabýval složením jejich volného času. Příklad slouží k demonstraci komplexní statistické analýzy kompozičních dat, kterou provedeme pomocí programu CoDaPack a obdržené výsledky budeme následně vhodně interpretovat.

Práce obsahuje také *Přílohy*, v kterých najdeme tabulky daných pozorování.

1. Kompoziční data

1.1. Definice a základní vlastnosti

Kompoziční data jsou taková mnohorozměrná pozorování, která mají pouze relativní charakter, poskytují pouze relativní informaci o složení daných celků. Znamená to, že z informací, obsažených v datech, budou pro další zpracování smysluplné pouze podíly mezi jednotlivými složkami, na rozdíl od standardních reálných dat, u kterých nás zajímá absolutní informace, tj. absolutní hodnoty jednotlivých složek. Kompoziční data lze reprezentovat pomocí konstantního součtu, a to $\kappa = 100$, jestliže se jedná o procentuální podíly, nebo $\kappa = 1$, pokud se jedná o proporcionální části na celku, a jednotlivé složky kompozičního vektoru (kompozice) jsou vždy kladné. Nyní se zaměříme na jejich definice a vlastnosti.

Tato kapitola vychází z literatury [1], [2], [3], [4], [6], [7].

Jak již bylo zmíněno, kompoziční data lze reprezentovat pomocí kladných vektorů se součtem složek rovným číslu κ , $\kappa \in \mathbf{R}^+$,

$$\mathbf{x} = (x_1, x_2, \dots, x_D), x_i > 0, \sum_{i=1}^D x_i = \kappa.$$

Definice 1.1 *Nechť*

$$\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathbf{R}_+^D$$

je D - složková kompozice.

Pak uzávěrem kompozice $\mathbf{x}$ rozumíme vektor

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

Z této definice vidíme, že i když vynásobíme vektor kladným reálným číslem, informace (podíly mezi složkami) se nezmění. Touto operací, tedy operací uzávěru, kterou značíme $\mathcal{C}$, můžeme libovolně měnit součet složek kompozičního vektoru.

Výběrovým prostorem kompozičních dat je tzv. *simplex*, značíme $\mathcal{S}^D$. Je to množina všech vektorů s D kladnými složkami a konstantním součtem. Matematicky zapsáno

$$\mathcal{S}^D = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_D); x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = \kappa \right\}.$$

S kompozičními daty se setkáváme v běžném životě, aniž bychom si to uvědomovali. Představují výstupy statistických měření v mnoha vědních oborech. Pracujeme tak s nimi například v chemii, geologii, medicíně, ale také v ekonomii či sociologii.

1.2. Aitchisonova geometrie na simplexu

Pro práci s kompozičními daty na simplexu zavádíme Aitchisonovu geometrii. Ta se od standardní euklidovské geometrie v reálném prostoru liší především tím, že jsme v případě kompozic nuceni pracovat pouze s relativní informací, což vyžaduje odlišný geometrický pohled. Abychom s kompozičními daty mohli pracovat obdobně jako v reálném prostoru, zavádíme na simplexu, což je, jak už jsme uvedli, výběrový prostor kompozičních dat, operace vhodné pro kompoziční data, které jsou analogické s operacemi v reálném prostoru. Následující operace nám umožňují zavést vektorový prostor.

Definice 1.2 *Perturbace kompozice $\mathbf{x} = \mathcal{C}(x_1, x_2, \dots, x_D) \in \mathcal{S}^D$ kompozicí $\mathbf{y} = \mathcal{C}(y_1, y_2, \dots, y_D) \in \mathcal{S}^D$ je kompozice*

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, x_2 y_2, \dots, x_D y_D).$$

Pomocí perturbace můžeme také zavést analogii odčítání, která se značí symbolem $\ominus$:

$$\mathbf{y} = \mathbf{x} \oplus \mathbf{p}, \mathbf{p} = \mathbf{y} \ominus \mathbf{x} = \mathbf{y} \oplus ((-1) \odot \mathbf{x}).$$

Definice 1.3 *Mocninná transformace kompozice $\mathbf{x} = \mathcal{C}(x_1, x_2, \dots, x_D) \in \mathcal{S}^D$ reálným číslem $\alpha \in \mathbf{R}$ je kompozice*

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

Perturbace a mocninná transformace dohromady se simplexem tvoří vektorový prostor $(\mathcal{S}^D, \oplus, \odot)$. Z toho mimo jiné vyplývá, že simplex společně s *perturbací* tvoří komutativní grupu.

Definice 1.4 *Aitchisonův skalární součin kompozic $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ definujeme takto:*

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

Definice 1.5 *Aitchisonova norma $\mathbf{x} \in \mathcal{S}^D$ je dána vztahem:*

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

Definice 1.6 *Aitchisonova vzdálenost mezi $\mathbf{x}$ a $\mathbf{y} \in \mathcal{S}^D$ je definována:*

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

Nyní si nadefinujeme transformace, které jsou nedílnou součástí při práci s Aitchisonovou geometrií. Abychom totiž mohli aplikovat na kompoziční data standardní statistické metody, je třeba je převést ze simplexu (omezeného prostoru) do reálného (neomezeného) prostoru. K tomu slouží již zmíněné logratio transformace, a to aditivní logratio transformace, centrovaná logratio transformace a izometrická logratio transformace.

Definice 1.7 *Aditivní logratio transformace, kterou značíme alr_j , $j = 1, \dots, D$, je prostá transformace ze simplexu $\mathcal{S}^D$ do reálného prostoru $\mathbf{R}^{D-1}$, definovaná pro $\mathbf{x} \in \mathcal{S}^D$ jako*

$$\mathbf{y} = alr_j(\mathbf{x}) = \left(x_1^{(j)}, \dots, x_{D-1}^{(j)} \right) = \left(\ln \frac{x_1}{x_j}, \dots, \ln \frac{x_{j-1}}{x_j}, \ln \frac{x_{j+1}}{x_j}, \dots, \ln \frac{x_D}{x_j} \right)$$

Aditivní logratio transformace má jednoduchý zápis pro převedení ze simplexu do reálného prostoru, a také převádí perturbaci na součet vektorů a mocninnou transformaci na násobení skalárem, avšak nezachovává metrické vlastnosti, jako je například vzdálenost. Proto se tato transformace téměř nepoužívá.

Definice 1.8 *Centrovaná logratio transformace, kterou značíme clr , je prostá transformace ze simplexu $\mathcal{S}^D$ do reálného prostoru $\mathbf{R}^D$, definovaná pro $\mathbf{x} \in \mathcal{S}^D$ jako*

$$\mathbf{y} = clr(\mathbf{x}) = (y_1, \dots, y_D) = \left(\ln \frac{x_1}{g_x}, \dots, \ln \frac{x_D}{g_x} \right),$$

kde g_x je geometrický průměr složek $x_1, \dots, x_D$.

Interpretace centrované logratio transformace se nám může jevit také jako jednoduchá, každou složku zde dělíme geometrickým průměrem. Stejně jako aditivní logratio transformace převádí perturbaci a mocninnou transformaci, navíc vzdálenost, skalární součin a norma jsou stejné jak při práci s kompozicemi v Aitchisonově geometrii, tak s jejich clr transformovanými protějšky v euklidovské geometrii. Důvod, proč ani tato transformace není příliš využívána, je ten, že součet clr koeficientů je roven nule, jak plyne z definice, tudíž příslušná varianční matice by byla singulární.

Definice 1.9 *Izometrická logratio transformace, kterou značíme ilr , je prostá transformace ze simplexu $\mathcal{S}^D$ do reálného prostoru $\mathbf{R}^{D-1}$, definovaná pro $\mathbf{x} \in \mathcal{S}^D$ jako*

$$\mathbf{z} = ilr(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{u}_1 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{u}_{D-1} \rangle_a),$$

kde kompozice $\psi_1, \dots, \psi_{D-1}$ tvoří ortonormální bázi na simplexu.

Jednou konkrétní volbou takové báze dostaneme ilr souřadnice

$$\mathbf{z}_i = \sqrt{\frac{D-i}{D-i+1}} \ln \frac{x_i}{\sqrt[D-i]{\prod_{j=i+1}^D x_j}},$$

kde $i = 1, \dots, D-1$.

S obecným postupem volby ilr souřadnic (pomocí tzv. postupného binárního dělení) se seznámíme v další kapitole.

Izometrická logratio transformace má nejnáročnější interpretaci ze všech tří transformací, je však z teoretického hlediska nejvhodnější, a tudíž nejpoužívanější. Hlavní výhodou je zde to, že při výše uvedené volbě ilr transformace první souřadnice reprezentuje veškerou informaci o zvolené kompoziční složce (v tomto případě x_1).

2. CoDaPack

CoDaPack je software navržený pro statistickou analýzu kompozičních dat, vytvořený v návaznosti na logratio přístup (transformace v předchozí kapitole obsahují logaritmy podílů složek kompozice) zavedený na počátku 80. let Johnem Aitchisonem. Jeho používání je velmi jednoduché a intuitivní. Je propojen se snadným a velmi užívaným programem Microsoft Office Excel. S vývojem stále nových verzí Excelu a operačního systému Windows musel být přitom CoDaPack přizpůsoben těmto novým verzím kvůli nekompatibilitě. Software lze neomezeně stáhnout na adrese <http://ima.udg.edu/codapack/>, kde je nutné zvolit si druh operačního systému, do něhož je program stahován. Pokud je dostupná nějaká aktualizace, program se sám přihlásí.

Tato kapitola byla vypracována s využitím zdrojů [2], [4], [5], [6], [8].

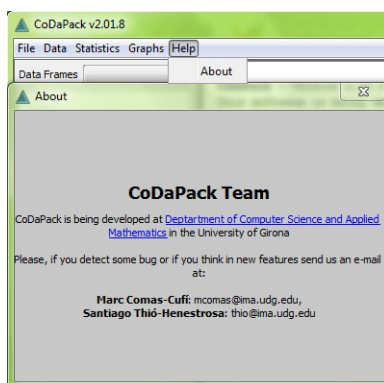
Po spuštění programu se otevře základní okno, které je rozděleno na 3 části (Obrázek 1). Pro ilustraci konkrétních výstupů jsou použita geologická data z databáze konference CoDaWork 2011 (<http://congress.cimne.com/codawork11>), viz. příloha 4.2.

Sample	SiO2	TiO2	Al2O3	MnO	MgO	CaO	Na2O	K2O
1 RT1-1	73.40	0.22	14.19	0.04	0.48	1.04	4.33	4.18
2 RT1-1	73.86	0.26	14.00	0.04	0.51	1.14	4.32	3.95
3 RT1-1	73.85	0.30	13.85	0.04	0.55	1.15	4.31	3.82
4 RT1-1	78.12	0.32	11.36	0.04	0.57	1.02	3.41	3.16
5 RT1-1	78.77	0.39	10.29	0.06	0.86	1.07	2.81	2.92
6 RT1-1	74.43	0.54	11.88	0.07	0.94	2.22	2.95	3.20
7 RT1-1	63.09	0.73	15.99	0.12	2.24	2.17	3.10	4.58
8 RT1-1	60.27	0.85	18.05	0.11	2.23	2.10	3.47	5.05
9 RT1-1	59.15	1.01	19.10	0.09	1.87	2.32	3.77	4.71
10 RT1-1	54.62	0.85	19.19	0.16	3.69	1.32	2.52	6.05
11 RT1-1	49.49	1.03	18.63	0.22	4.02	1.62	2.07	5.40
12 RT1-10	71.66	0.14	15.34	0.02	0.37	0.91	4.82	4.76
13 RT1-10	71.22	0.17	15.92	0.03	0.39	1.02	5.24	4.49
14 RT1-10	73.51	0.19	14.45	0.03	0.36	0.97	4.78	3.97
15 RT1-10	80.51	0.22	10.45	0.03	0.37	0.82	3.20	2.70

Obrázek 1: Základní okno s již přidávanými daty.

V levé části se nám po zvolení zobrazí složky, se kterými budeme pracovat, respektive záhlaví datové tabulky, jež je aktualizováno vždy se sloupcem nových dat, které jsou výsledkem zvolených operací. Do pravé horní části se automaticky zapisují jednotlivé činnosti a operace, které provádíme, a my k nim můžeme doplňovat svoje poznámky. Poslední část, vpravo dole, která je zobrazena formou tabulky, slouží k práci s daty. Zapisují se zde výpočty zvolených operací a s těmito výpočty můžeme dále pracovat. Maximální množství dat, které můžeme použít v CoDaPacku, je 200 složek, a celkem 6000 pozorování.

V hlavním menu tohoto programu najdeme 5 základních záložek, a to PRÁCE SE SOUBORY (*File*), OPERACE S DATY (*Data*), STATISTICKÁ ANALÝZA (*Statistics*), GRAFY (*Graphs*) a NÁPOVĚDA (*Help*). V záložce NÁPOVĚDA však nejsou žádné rady pro práci s programem, tudíž je, nejen pro naše účely, nepotřebná (Obrázek 2).

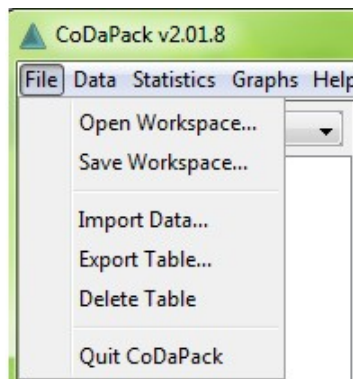


Obrázek 2: Náповěda.

Ostatními záložkami a tím, co nabízí, se budeme zabývat postupně v následujících podkapitolách.

2.1. Práce se soubory

V záložce *File* najdeme základní příkazy pro práci se souborem, jak je známe z práce v programu Excel (Obrázek 3).

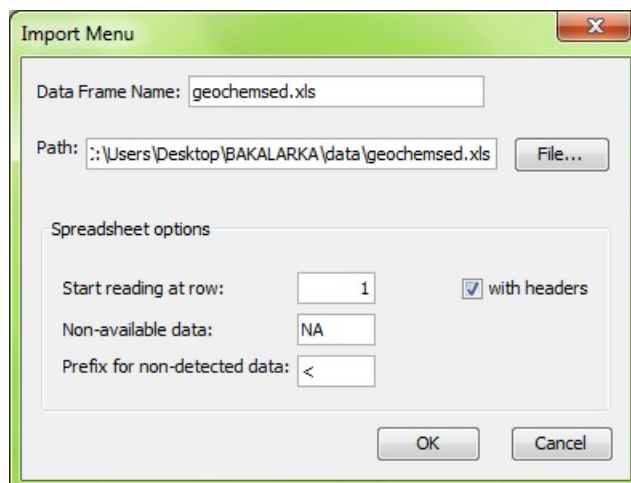


Obrázek 3: Menu pro práci se souborem.

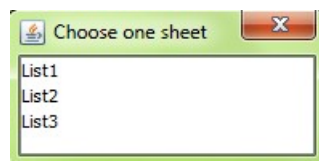
Otevřít soubor (*Open workspace*) – slouží k otevření již dříve zpracovaných dat v programu CoDaPack spolu s provedenými operacemi. Příslušný soubor má příponu .cdp.

Uložit soubor (*Save workspace*) – pomocí tohoto příkazu ukládáme zpracovaná data, soubor je uložen s již zmíněnou příponou .cdp do zvolené složky.

Vložit data (*Import data*) – slouží k importu, můžeme použít data vytvořená v jiných programech, v tomto případě tedy v Microsoft Office Excel. Pokud chceme tato data importovat, zobrazí se okno (Obrázek 4), které si žádá název a umístění souboru, v dialogovém okně také můžeme nastavit číslo prvního čteného řádku *Start reading at row*, zda mají zůstat popisky složek *with headers*, jak označovat chybějící hodnoty v datech *Non-available data* (označení volíme sami) nebo nastavit limit pro nejnižší možné hodnoty v datech, se kterými chceme pracovat *Prefix for non-detected data*. Po kliknutí na tlačítko OK se data zobrazí v tabulce a my můžeme pracovat s různými funkcemi. Pokud má excelovský soubor více listů, program se nás po výběru souboru dotáže, který z listů chceme použít a ten otevřeme dvojitým kliknutím (Obrázek 5).



Obrázek 4: Vložení dat.



Obrázek 5: Výběr listu.

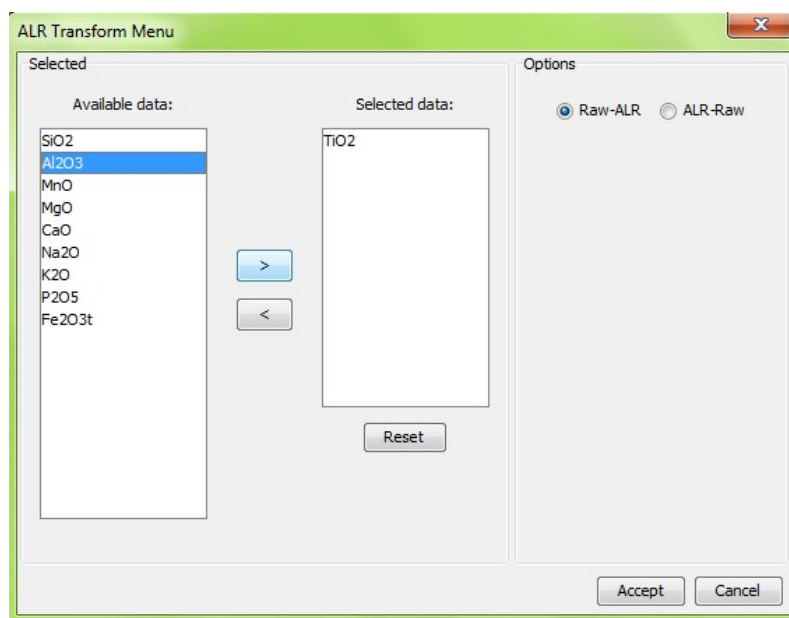
Exportovat tabulku (*Export table*) – tento příkaz automaticky přepisuje, pokud nezvolíme jiné cílové místo, zpracovaná či upravená data do původního souboru, a to v programu Excel, tedy s příponou .xls, ve kterém můžeme jednotlivé hodnoty upravovat, což v samotném CoDaPacku nejde.

Smazat tabulku (*Delete table*) – příkaz slouží k vymazání celého datového souboru, následně můžeme vložit jiná data. Zachovává však část s výpočty a tabulkami, do které můžeme vpisovat poznámky.

Zavřít CoDaPack (*Quit CoDaPack*) – použijeme k úplnému ukončení programu CoDaPack. Soubor se však neukládá automaticky, při každém zavření musíme tedy použít *Save workspace* nebo *Export table*. Před ukončením se nás ještě CoDaPack dotáže, zda chceme tuto operaci opravdu provést.

2.2. Operace s daty

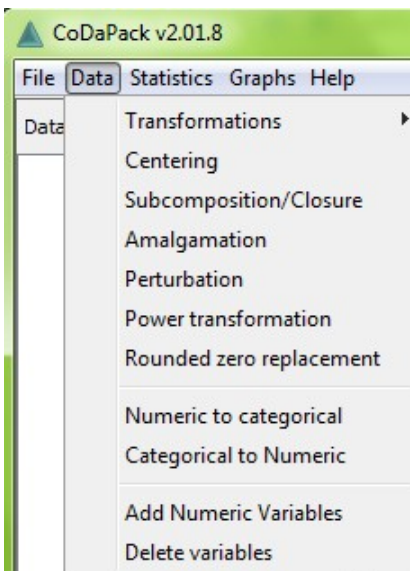
Záložka *Data* slouží k práci s vybranými kompozičními daty. Z nabídky si zvolíme operaci, kterou chceme s daty provést. Po zvolení operace se zobrazí dialogové okno (Obrázek 6), které je rozděleno na 2 části. První je **Výběr** (*Selected*), kde vybíráme složky celku, jejichž výstup nás zajímá, pro něž tedy chceme danou operaci použít. Prvky vybíráme dvojitým kliknutím na danou složku nebo označením složky a následným použitím tlačítka se šipkou, které leží mezi seznamy dostupných kompozičních složek (*Available data*) a vybraných složek (*Selected data*). Pokud si zvolíme nesprávné či nepotřebné složky, stejným způsobem je dostaneme zpět na původní místo. Je zde také tlačítko *Reset*, použitelné v případě zrušení celého výběru. Druhou částí dialogového okna jsou **Možnosti** (*Options*), kde se podle zvolené operace zobrazí možnost nastavení parametrů.



Obrázek 6: Dialogové okno.

Tento postup je stejný pro záložky *Data*, *Statistické údaje* i *Grafy*.

V záložce *Data* máme k dispozici tyto operace (Obrázek 7):



Obrázek 7: Data.

Transformace (*Transformations*) - jak už je uvedeno výše, transformace slouží k převodu dat ze simplexu do reálného prostoru, každá má jinou interpretaci a možnosti využití. Po výběru transformace se otevře dialogové okno a v části **Možnosti** volíme, zda chceme použít transformaci ze simplexu do reálného prostoru nebo příslušnou inverzní transformaci.

Alr - Aditivní logratio transformace – v tomto případě logaritmujeme podíl složek vektoru, kdy právě jedna ze složek je dělitelem a dělí všechny ostatní, tato složka je libovolná dle našeho výběru (nemusí to být vždy ta poslední, viz příklad).

Clr - Centrovaná logratio transformace – zde logaritmujeme podíl, kdy dělitelem je geometrický průměr složek a dělencem jednotlivé složky kompozičního vektoru.

Ilr - Izometrická logratio transformace – při této transformaci musíme nejprve provést postupné binární dělení, abychom mohli následně zkonstruovat ilr souřadnice s příslušnou interpretací. Mějme D –složkovou kompozici na simplexu, rozdělíme složky na dvě skupiny, kdy všechny složky v první skupině označíme $+$ a všechny složky v druhé skupině $-$, jednu ze skupin si zvolíme a složky z ní opět rozdělujeme stejným způsobem. V dělení budeme pokračovat do té doby, než bude v každé skupině pouze jedna složka $+$ a jedna $-$. To samé provedeme i s pů-

vodní druhou skupinou. Složky, které se právě neúčastní binárního dělení, jsou označeny 0. Detailnější popis postupného binárního dělení v programu CoDaPack najdeme v kapitole 2.4. - Bilanční dendrogram.

Pro provedení všech kroků potřebujeme $D - 1$ kroků. Výsledná ilr transformace pak bude ve tvaru

$$\mathbf{z}_i = \sqrt{\frac{rs}{r+s}} \ln \frac{x_{i_1} \dots x_{i_r}^{\frac{1}{r}}}{x_{j_1} \dots x_{j_s}^{\frac{1}{s}}},$$

kde r značí počet $+$, s značí počet $-$ v i -tém kroku dělení, $1 \leq i_1 < \dots < i_r \leq D$ a $1 \leq j_1 < \dots < j_s \leq D$.

Jednotlivé transformace můžeme ilustrovat na jednoduchém příkladu: Mějme dány dvě kompozice $\mathbf{x} = [80, 15, 5] \in \mathcal{S}^3$, $\mathbf{y} = [20, 30, 40] \in \mathcal{S}^3$, a reálná čísla $\alpha = 2$, $\kappa = 3$. Tyto nadefinované proměnné budeme používat i dále v názorných příkladech následujících operací. Nejprve si ukážeme aplikace alr a clr transformace,

$$alr(\mathbf{x}) = alr[80, 15, 5] = \left[\ln \frac{80}{5}, \ln \frac{15}{5} \right] = [\ln(16), \ln(3)] = [2.77, 1.10],$$

$$\begin{aligned} clr(\mathbf{x}) &= clr[80, 15, 5] = \left[\ln \frac{80}{(80 \cdot 15 \cdot 5)^{\frac{1}{3}}}, \ln \frac{15}{(80 \cdot 15 \cdot 5)^{\frac{1}{3}}}, \ln \frac{5}{(80 \cdot 15 \cdot 5)^{\frac{1}{3}}} \right] = \\ &= \left[\ln \frac{80}{(6000)^{\frac{1}{3}}}, \ln \frac{15}{(6000)^{\frac{1}{3}}}, \ln \frac{5}{(6000)^{\frac{1}{3}}} \right] = \left[\ln \frac{80}{18 \cdot 17}, \ln \frac{15}{18 \cdot 17}, \ln \frac{5}{18 \cdot 17} \right] = \\ &= [\ln(4.40), \ln(0.83), \ln(0.28)] = [1.48, -0.19, -1.29]. \end{aligned}$$

Ilr transformaci kompozice $\mathbf{x}$ vypočteme pomocí vztahu z první kapitoly

$$\begin{aligned} ilr(\mathbf{x}) &= ilr[80, 15, 5] = \left[\frac{1}{\sqrt{6}} \ln \frac{80^2}{15 \cdot 5}, \frac{1}{\sqrt{2}} \ln \frac{15}{5} \right] = \\ &= \left[\frac{1}{\sqrt{6}} \ln(85, 333), \frac{1}{\sqrt{2}} \ln(3) \right] = [1.82, 0.78]. \end{aligned}$$

Centrování (*Centering*) - je speciální případ perturbace, centruje soubor vybraných kompozic pomocí jejich centra, výběrové charakteristiky centra distribuce dat. Používá se zejména v situacích, kdy je obtížné zobrazit a interpretovat datovou strukturu.

Uvedeme si zde celý průběh centrování kompoziční datové matice (v jejich řád-

cích jsou uvedeny jednotlivé kompozice).

Máme danu matici $\mathbf{X}$,

$$\mathbf{X} = [X_1, X_2, X_3] = \begin{bmatrix} 0.45 & 46.41 & 53.14 \\ 4.73 & 11.15 & 84.12 \\ 8.02 & 6.75 & 85.23 \\ 12.39 & 4.19 & 83.42 \\ 37.99 & 0.74 & 61.28 \end{bmatrix}.$$

Vypočteme geometrické průměry jednotlivých sloupců:

$$g(X_1) = (0.45 \cdot 4.73 \cdot 8.02 \cdot 12.39 \cdot 37.99)^{\frac{1}{5}} = 6.05;$$

$$g(X_2) = (46.41 \cdot 11.15 \cdot 6.75 \cdot 4.19 \cdot 0.74)^{\frac{1}{5}} = 6.40;$$

$$g(X_3) = (53.14 \cdot 84.12 \cdot 85.23 \cdot 83.42 \cdot 61.28)^{\frac{1}{5}} = 72.09.$$

Výsledky přepočteme pomocí operace uzávěru, aby součet složek byl 100 a tím dostaneme odhad centra:

$$cen(\mathbf{X}) = \mathcal{C}[6.05, 6.40, 72.09] = [7.15, 7.58, 85.27].$$

Následně vypočteme inverzi k získaným hodnotám:

$$\ominus cen(\mathbf{X}) = \mathcal{C}\left[\frac{1}{7.15}, \frac{1}{7.58}, \frac{1}{85.27}\right] = [49.30, 46.56, 4.14].$$

V závěrečném kroku perturbujeme původní matici vypočteným inverzním vektorem; dostaneme matici, kterou je však třeba ještě přepočítat, aby součet složek v řádku byl 100 (kvůli lepší interpretaci), poté dostaneme výslednou vycentrovanou datovou matici:

$$\begin{aligned} & \begin{bmatrix} 0.45 & 46.41 & 53.14 \\ 4.73 & 11.15 & 84.12 \\ 8.02 & 6.75 & 85.23 \\ 12.39 & 4.19 & 83.42 \\ 37.99 & 0.74 & 61.28 \end{bmatrix} \oplus \begin{pmatrix} 49.30 \\ 46.56 \\ 4.14 \end{pmatrix} = \mathcal{C} \begin{bmatrix} 0.45 \cdot 49.30 & 46.41 \cdot 46.56 & 53.14 \cdot 4.14 \\ 4.73 \cdot 49.30 & 11.15 \cdot 46.56 & 84.12 \cdot 4.14 \\ 8.02 \cdot 49.30 & 6.75 \cdot 46.56 & 85.23 \cdot 4.14 \\ 12.39 \cdot 49.30 & 4.19 \cdot 46.56 & 83.42 \cdot 4.14 \\ 37.99 \cdot 49.30 & 0.74 \cdot 46.56 & 61.28 \cdot 4.14 \end{bmatrix} = \\ & = \mathcal{C} \begin{bmatrix} 22.19 & 2160.85 & 219.00 \\ 233.19 & 519.14 & 348.26 \\ 395.39 & 314.28 & 352.85 \\ 610.83 & 195.09 & 345.36 \\ 1872.91 & 34.45 & 253.70 \end{bmatrix} = \begin{bmatrix} 0.93 & 89.92 & 9.15 \\ 21.20 & 47.18 & 31.62 \\ 37.21 & 29.60 & 33.19 \\ 53.09 & 16.93 & 29.98 \\ 86.68 & 1.59 & 11.73 \end{bmatrix}. \end{aligned}$$

Uzávěr pro podkompozici (*Subcomposition/Closure*) – slouží k úpravě součtu složek ve vektoru vedoucí ke zjednodušení interpretace

Jednotlivé složky vektoru zde upravujeme pomocí zvoleného reálného čísla κ .

Pro $\mathbf{x} = [80, 15, 5]$, $\kappa = 3$ složky vektoru postupně násobíme reálným číslem a dělíme celkovým součtem složek:

$$\mathbf{x} = \left[\frac{80 \cdot 3}{100}, \frac{15 \cdot 3}{100}, \frac{5 \cdot 3}{100} \right] = [2.4, 0.45, 0.15].$$

Amalgamace (*Amalgamation*) - slouží k sečtení zvolených kompozičních složek; tím zmenšíme počet složek v kompozici a usnadníme její interpretaci, ale naopak ztratíme informaci o původních složkách. Jako příklad budeme v matici $\mathbf{X}$ amalgamovat neboli sčítat sloupce odpovídající první a druhé kompoziční složce:

$$\mathbf{X} = \begin{bmatrix} 0.45 & 46.41 & 53.14 \\ 4.73 & 11.15 & 84.12 \\ 8.02 & 6.75 & 85.23 \\ 12.39 & 4.19 & 83.42 \\ 37.99 & 0.74 & 61.28 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.45 + 46.41 & 53.14 \\ 4.73 + 11.15 & 84.12 \\ 8.02 + 6.75 & 85.23 \\ 12.39 + 4.19 & 83.42 \\ 37.99 + 0.74 & 61.28 \end{bmatrix} = \begin{bmatrix} 46.86 & 53.14 \\ 15.88 & 84.12 \\ 14.77 & 85.23 \\ 16.58 & 83.42 \\ 38.73 & 61.28 \end{bmatrix}.$$

Perturbace (*Perturbation*) - používá se pro ni symbol $\oplus$, je obdobou operace sčítání vektorů v reálném prostoru. Dle příslušného vztahu z první kapitoly perturbujeme zvolené vektory:

$$\mathbf{x} \oplus \mathbf{y} = [80, 15, 5] \oplus [20, 30, 40] = \mathcal{C}[80 \cdot 20, 15 \cdot 30, 5 \cdot 40] = \mathcal{C}[1600, 450, 200].$$

Mocninná transformace (*Power transformation*) – symbolické označení této operace je $\odot$, jedná se o obdobu násobení vektoru reálným číslem.

Zde původní kompozici $\mathbf{x}$ mocníme zvoleným reálným číslem, v tomto případě $\alpha = 2$:

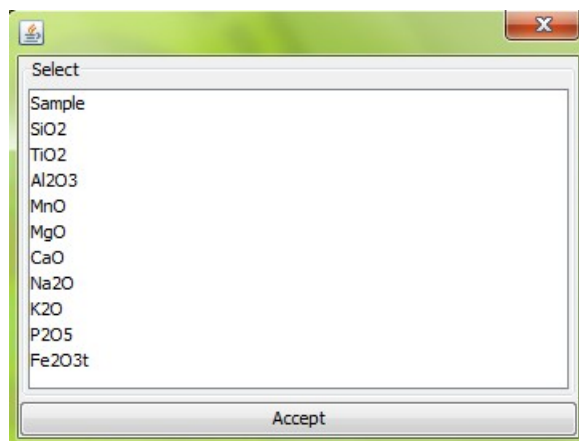
$$\alpha \odot \mathbf{x} = \mathcal{C}[80^2, 15^2, 5^2] = \mathcal{C}[6400, 225, 25].$$

Číselné na kategoriální (*Numeric to categorical*),

Kategoriální na číselné (*Categorical to numeric*)

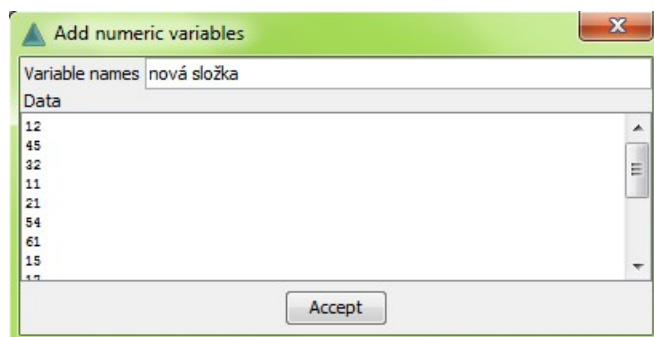
- pomocí těchto dvou příkazů můžeme měnit formát dat, dle potřeby práce s daty. Převádí se zde celé vybrané sloupce datových hodnot. Převody se však nezapisují do nových sloupců, ale přepíše se vybraný sloupec (Obrázek 8).

Přidat číselné proměnné (*Add numeric variables*) – používáme v případě,



Obrázek 8: Převod hodnot z číselných na kategoriální nebo naopak.

že chceme doplnit další sloupec dat přímo v softwaru CoDaPack, tzn. název příslušné proměnné bude zobrazen v záhlaví datové tabulky (Obrázek 9).



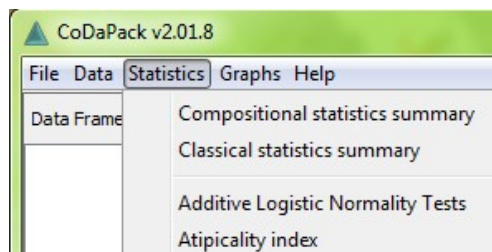
Obrázek 9: Manuální přidávání nového sloupce hodnot. Do *Variable names* napíšeme název nového sloupce a do *Data* píšeme pod sebe hodnoty dané proměnné.

Smazat proměnné (*Delete variables*) – využijeme při odstranění celého sloupce dat, například při nepotřebnosti dříve vypočítaných sloupců.

2.3. Statistická analýza

Výstupem jsou číselné charakteristiky datového souboru (Obrázek 10). Ty můžeme vypočítat pomocí standardních či kompozičních statistických metod. Mezi standardní statistické metody patří například aritmetický průměr, směrodatná odchylka, varianční matice a kvartily, mezi kompoziční statistické metody patří centrum, variační matice a rozptyly clr transformovaných kompozic. Použití

a interpretace zmíněných metod uvedeme vzhledem k omezenému rozsahu této práce v části věnované praktickému příkladu.



Obrázek 10: Statistické údaje.

Pokud chceme znát výstupní hodnoty, klikneme na jedno ze zvolených statistických shrnutí a objeví se dialogové okno. V pravé části **Možnosti** je více „zatržíték“, pomocí nichž si můžeme vybrat z nabídky statistických metod. Zatržítka se v rámci jednotlivých příkazů mění.

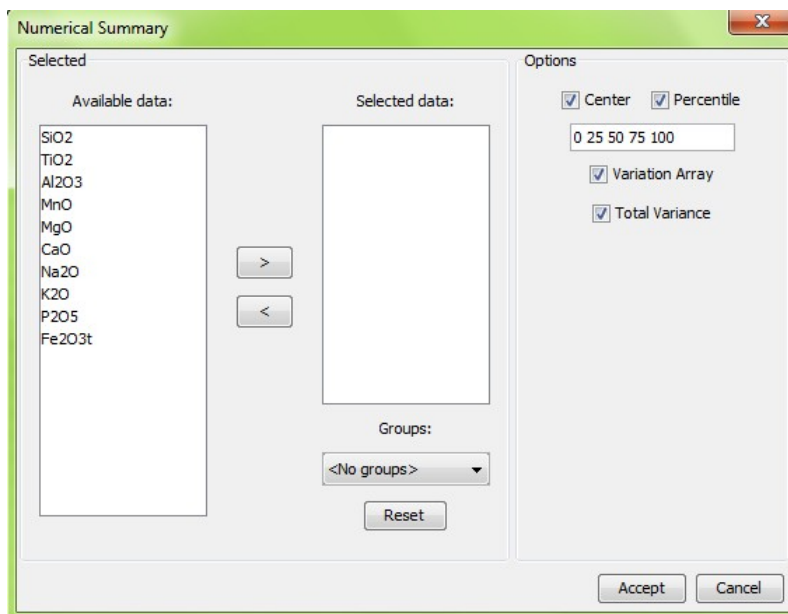
Funkce dostupné v této záložce:

Kompoziční statistické shrnutí (*Compositional statistics summary*) – výstupem je centrum (tvořené geometrickými průměry složek), minima, maxima a kvartily, dále CLR rozptyly (rozptyly výsledných clr proměnných), variační matice a celkový rozptyl datového souboru (Obrázek 11).

Po zadání zvolených operací dostaneme následující výsledky (Obrázek 12):

Nejprve se nám zobrazí tabulka, ve které najdeme hodnoty geometrických průměrů složek (*Center*), minima a maxima (0 a 100) a jednotlivých kvartilů (25, 50, 75), vždy reprezentované v proporcích.

Další tabulka obsahuje výběrové rozptyly a průměry logaritmů podílů jednotlivých dvojic složek kompozice, rozptyly clr proměnných (*CLR variances*) a součet těchto rozptylů, reprezentující celkový rozptyl datového souboru (*Total variance*).



Obrázek 11: Dialogové okno kompozičního statistického shrnutí.

Sample size: 87		Variation array:						
Statistics		Variance $\ln(X_i/X_j)$						
	Center	0	25	50	75	100		
SiO2	0.8228	0.6561	0.7678	0.8286	0.8621	0.9009		
TiO2	0.0040	0.0012	0.0024	0.0034	0.0073	0.0140		
Al2O3	0.1647	0.0911	0.1301	0.1660	0.1946	0.2572		
MnO	0.0005	0.0002	0.0003	0.0004	0.0008	0.0030		
MgO	0.0079	0.0017	0.0040	0.0059	0.0148	0.0741		

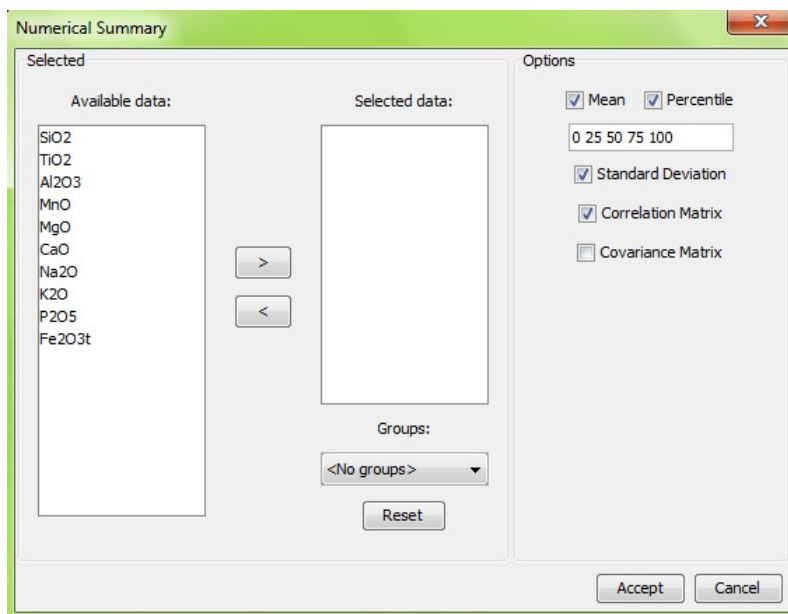
Xi\Xj	SiO2	TiO2	Al2O3	MnO	MgO	clr variances
SiO2		0.4916	0.1144	0.5239	0.9896	0.2119
TiO2	-5.3240		0.2809	0.0799	0.2451	0.1097
Al2O3	-1.6083	3.7158		0.2900	0.6133	0.1299
MnO	-7.3698	-2.0458	-5.7615		0.1176	0.1011
MgO	-4.6432	0.6808	-3.0350	2.7266		0.1966
Mean $\ln(X_i/X_j)$						0.7493 Total Variance

Obrázek 12: Výsledky operací kompozičního statistického shrnutí.

Standardní statistické shrnutí (*Classical statistics summary*) – zde je výstupem aritmetický průměr, směrodatná odchylka, korelační matice, varianční matice a dále minimum, maximum pro jednotlivé složky a příslušné kvartily (Obrázek 13).

Výsledky opět v následujících tabulkách (Obrázek 14):

Zde nám první tabulka nabízí aritmetický průměr (*Mean*), směrodatnou odchylku (*Standard deviation*), opět minima a maxima (0 a 100) a kvartily (25,50,75). Druhá tabulka představuje korelační matici a třetí tabulka varianční matici původních dat.



Obrázek 13: Dialogové okno standardního statistického shrnutí.

Sample size:
87

Statistics

	Mean	Std.Dev	0	25	50	75	100
SiO2	70.8082	8.1797	49.4900	64.4500	72.7900	77.1800	83.6900
TiO2	0.4124	0.2569	0.1000	0.2200	0.3000	0.6200	1.0300
Al2O3	14.4280	3.1251	8.4700	11.6900	14.4900	16.4400	20.2900
MnO	0.0551	0.0427	0.0180	0.0270	0.0380	0.0660	0.2190
MgO	1.0568	1.1712	0.1600	0.3600	0.5200	1.2400	5.6600

Correlation:

	SiO2	TiO2	Al2O3	MnO	MgO
SiO2	1.0000	-0.8005	-0.9244	-0.8223	-0.8289
TiO2	-0.8005	1.0000	0.6668	0.8238	0.7648
Al2O3	-0.9244	0.6668	1.0000	0.5848	0.6037
MnO	-0.8223	0.8238	0.5848	1.0000	0.9488
MgO	-0.8289	0.7648	0.6037	0.9488	1.0000

Covariance:

	SiO2	TiO2	Al2O3	MnO	MgO
SiO2	66.9069	-1.6820	-23.6284	-0.2873	-7.9410
TiO2	-1.6820	0.0660	0.5353	0.0090	0.2301
Al2O3	-23.6284	0.5353	9.7661	0.0781	2.2098
MnO	-0.2873	0.0090	0.0781	0.0018	0.0475
MgO	-7.9410	0.2301	2.2098	0.0475	1.3718

Obrázek 14: Výsledky operací klasického statistického shrnutí.

Aditivní test normality (*Additive Logistic Normality Tests*) – tato funkce umožňuje provést test normality na celé kompozici o D složkách prostřednictvím její alr transformace. Mezi nejvhodnější testy pro kompoziční data patří *Anderson-Darling test*, *Cramer-von Mises test* a *Watson test*. Testy zahrnují testování jednorozměrných marginálních rozdělání ($D - 1$ testů), dvojrozměrných rozdělání $((D - 1)(D - 2)$ testů) a celkový test alr transformované kompozice (Obrázek 15).

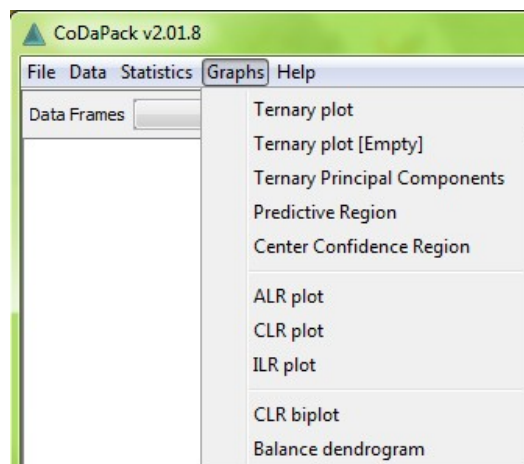
Normality Tests:						
	Anderson-Darling		Cramer-von Mises		Watson	
	λ^2*	P	W^2*	P	U^2*	P
Marginals						
alr(SiO ₂ ,Al ₂ O ₃)	0.6441	[0.15, 0.10]	0.0985	[0.15, 0.10]	0.0983	[0.10, 0.05]
alr(TiO ₂ ,Al ₂ O ₃)	1.4191	<0.01	0.2028	<0.01	0.1853	<0.01
Bivariate angle						
alr(SiO ₂ ,Al ₂ O ₃)	1.7844	[0.15, 0.10]	0.2640	>0.15	0.2282	[0.025, 0.01]
alr(TiO ₂ ,Al ₂ O ₃)						
Radius						
	2.7001	[0.05, 0.025]	0.4789	[0.05, 0.025]	0.2685	<0.01

Obrázek 15: Test logistického normálního rozdělání pro prvky SiO_2 , TiO_2 , Al_2O_3 .

Odlehlá pozorování (*Atypical index*) – tato operace slouží k detekci odlehlých pozorování a zobrazení jejich indexů za předpokladu aditivního logistického normálního rozdělání (tzn. normálního rozdělání alr transformovaných dat) u vybraného datového souboru. Za účelem provedení příslušného testování je třeba určit hladinu testu α . Přednastavená je hodnota $0.95 = 1 - \alpha$ a je to také nejčastěji udávaná hodnota.

2.4. Grafy

Možným výstupem analýzy kompozičního datového souboru jsou též grafy popisující strukturu sledovaných dat. K dispozici tu máme zobrazení alr, clr a ilr transformací, clr biplotu, ternárního diagramu a bilančního dendrogramu (Obrázek 16).



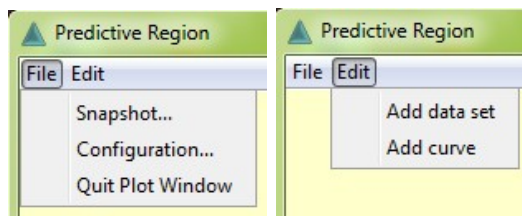
Obrázek 16: Grafy

Vždy se aktivuje již známé dialogové okno rozdělené na **Výběr**, tato část je vždy stejná, a **Možnosti**, tato část se mění podle typu grafu, v tomto případě je však u většiny grafů prázdná.

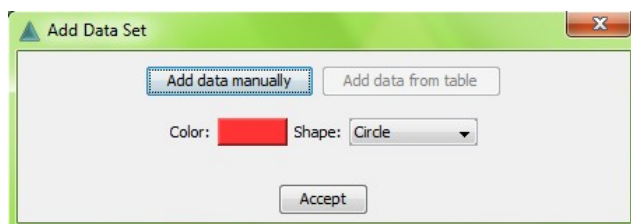
Výstup se nám zobrazuje v novém okně, kde můžeme s grafem dále pracovat. Na ose s názvem *Zoom* přibližujeme či oddalujeme celý objekt, můžeme také zatrhnout *Center* a tím data v grafu vycentrovat. Pokud je graf 2D (tedy rovinový graf), máme k dispozici zatržítko *Grid*, díky němuž se nám v grafu zobrazí pomocná mřížka. Dále pomocí tlačítek



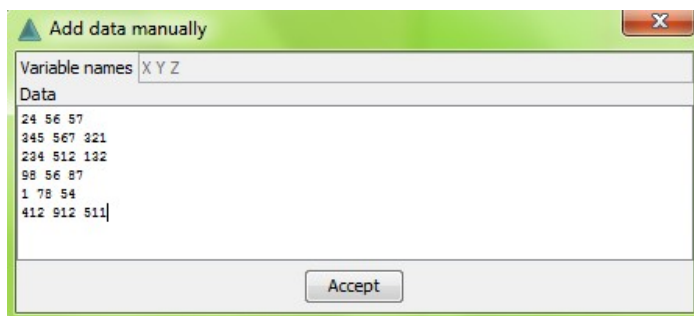
vyměňujeme postavení jednotlivých os nebo otáčíme celým obrázkem. Pokud je graf 3D (prostorový), pomocí levého tlačítka myši rotujeme celým objektem dle potřeby. Příkaz *File* → *Configuration* slouží k upravování barevnosti grafu (pozadí, osy, popisky), především pro přehlednost. Grafy lze také samostatně ukládat, a to pomocí příkazu *File* → *Snapshot*, k dispozici máme přípony jpeg, eps, png a bmp (Obrázek 17). U některých typů grafů máme ještě nabídku *Edit* → *Add data set*, pomocí níž přidáváme nová data, která potřebujeme zakreslit do již vytvořeného grafu (Obrázky 18, 19) nebo *Add curve*, čímž zakreslíme křivku ze zadaných dat při vytvoření grafu *Predictive region* či *Center confidence region*.



Obrázek 17: Menu pro práci s grafem.



Obrázek 18: Výběr označení nových dat, v nabídce je kolečko, čtverec, hvězda a trojúhelník špičkou nahodu či dolů.

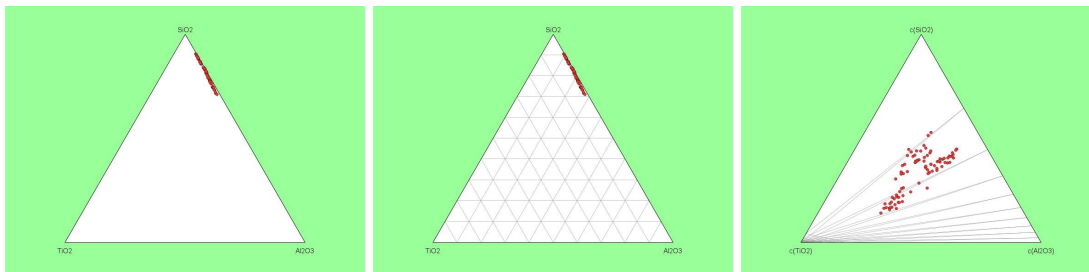


Obrázek 19: Tabulka pro vložení nových dat, kde jsou již předepsány názvy složek, tedy vrcholy trojúhelníku (ternárního diagramu) X, Y, Z. Pozorování řadíme pod sebe a jejich číselné hodnoty píšeme do řádků a oddělujeme mezerou.

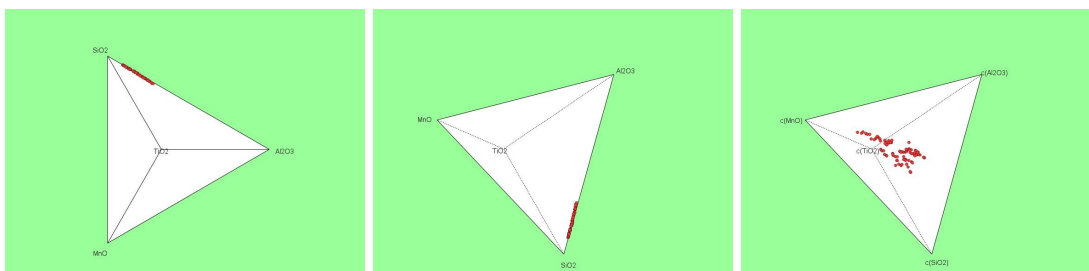
Ternární diagram (*Ternary plot*) – je tvořen ze tří nebo čtyř vybraných složek kompozic. V případě výběru tří složek je výstupem ternární diagram (Obrázek 20).

V případě výběru čtyř složek je grafem tetrahedron (Obrázek 21).

Prázdný ternární diagram (*Ternary plot [Empty]*) - při výběru této nabídky se nás program neptá na žádná data, automaticky zobrazí prázdný ternární diagram, do něhož data můžeme následně vkládat pomocí již zmíněného *Add data set*. Další funkce jsou stejné jako u ostatních grafů.



Obrázek 20: Vybraná pozorování zobrazená v 2D grafu (ternárním diagramu), vpravo jsou zobrazena centrovaná data.

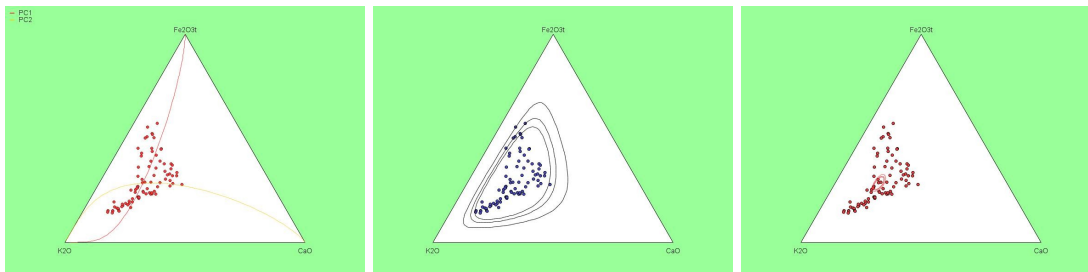


Obrázek 21: Vybrané prvky zobrazené v 3D grafu (tetrahedronu), vpravo jsou zobrazena centrovaná data.

Ortogonální regrese pro trojsložková kompoziční data (*Ternary Principal Components*) – pomocí této funkce program vykresluje v ternárním diagramu nebo tetrahedronu regresní přímky (či křivky), které mají data charakterizovat. Neznámé parametry přímek zde odhadujeme pomocí ortogonální regrese (Obrázek 22).

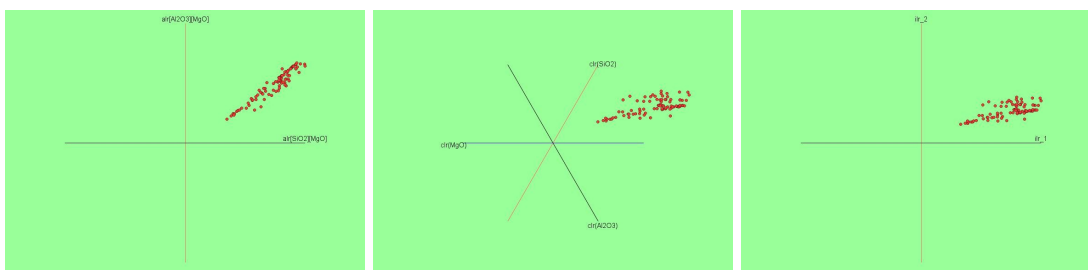
Oblast výskytu datového souboru (*Predictive Region*) – tato funkce nám v ternárním diagramu zobrazuje plošné oblasti, ve kterých se nachází pozorovaná data. K dispozici máme 3 plochy, jejichž velikost lze nastavit v dialogovém okně v části **Možnosti** (*Predictive level*). Například při nastavení 0.90, 0.95, 0.99 se nám zobrazí plochy, které obsahují 90%, 95% a 99% pozorovaných výsledků (Obrázek 22).

Oblast spolehlivosti pro centrum kompozičního datového souboru (*Center Confidence Region*) – stejně jako předchozí funkce i tato se nám zobrazuje v ternárním diagramu. Vykreslí pro centrum datového souboru oblasti spolehlivosti, které dle potřeby volíme analogickým způsobem jako v předchozím grafu (Obrázek 22).



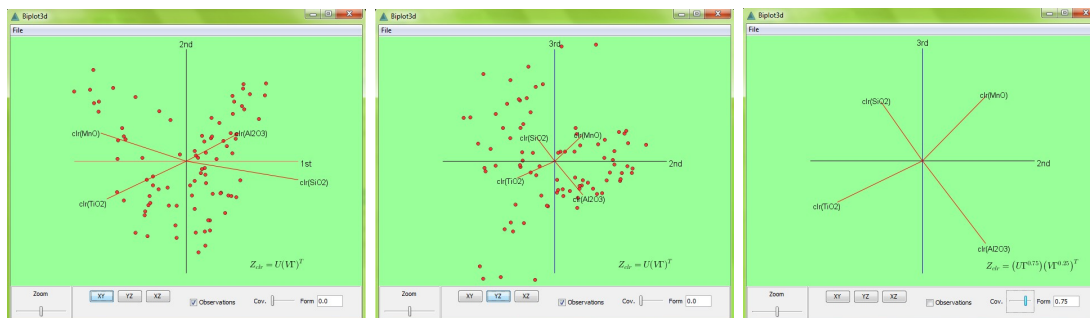
Obrázek 22: Data zobrazená funkcemi: *Ternary Principal Components*, *Predictive Region* a *Center Confidence Region*

Grafy ALR, CLR a ILR – jsou grafickými výstupy logratio transformací (Obrázek 23).



Obrázek 23: Zobrazení alr transformace, clr transformace a ilr transformace.

CLR Biplot – představuje nám grafické zobrazení vztahů mezi pozorovanými složkami datového souboru pomocí metody hlavních komponent. V dialogovém okně vybíráme složky pro zpracování. S grafem zobrazeným v novém okně můžeme dále pracovat. Máme k dispozici tlačítka pro zobrazení preferované osy, můžeme odstranit data a nechat zobrazené pouze polopřímky reprezentující jednotlivé clr proměnné (Obrázek 24). V hlavním okně se nám zobrazí příslušné číselné výstupy (Obrázek 25). Protože metoda hlavních komponent přesahuje tématicky výrazně zaměření této práce, budeme se zabývat až konkrétní interpretací grafu v praktické části práce.



Obrázek 24: Různá zobrazení grafu biplotu.

Biplot generated:
Data: SiO2 TiO2 Al2O3 MnO

Principal Components:

	clr(SiO2)	clr(TiO2)	clr(Al2O3)	clr(MnO)	Cum.Prop.Exp.
PC1	0.6581	-0.4662	0.3110	-0.5029	0.8270
PC2	-0.3187	-0.6530	0.4840	0.4876	0.9225
PC3	0.4641	-0.3260	-0.6473	0.5092	1.0000

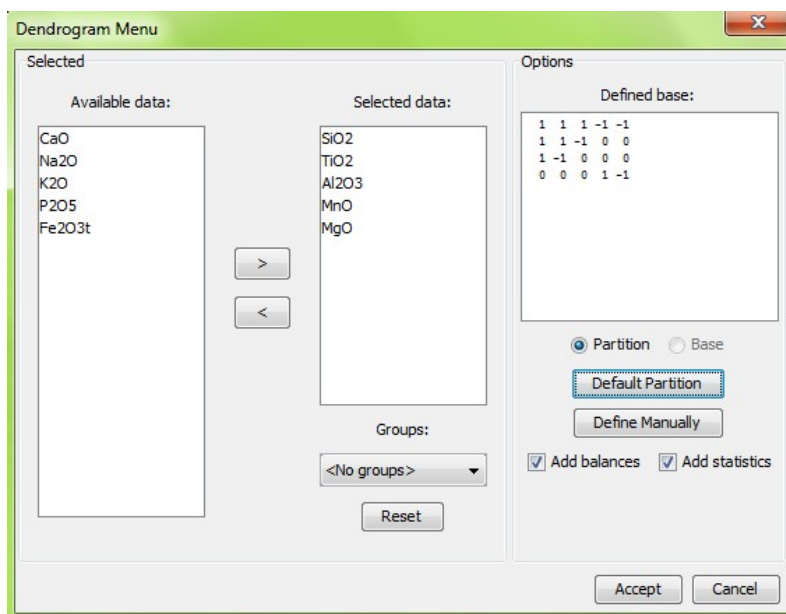
Obrázek 25: Číselné výstupy biplotu.

Bilanční Dendrogram (*Balance Dendrogram*) – pomocí něj můžeme zobrazit postupné binární dělení, poměry, rozptyly a další číselné charakteristiky výsledných ilr souřadnic (bilancí).

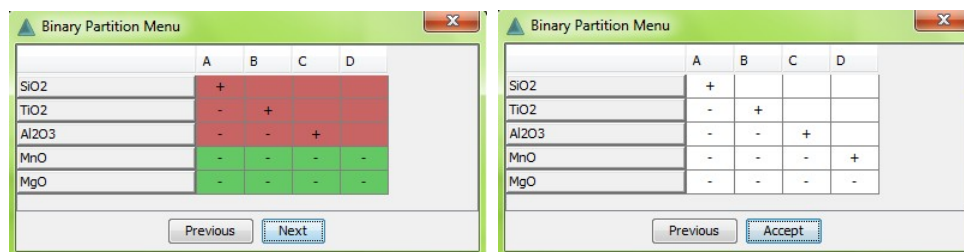
Po zvolení tohoto grafu z nabídky se otevře dialogové okno (Obrázek 26), vybereme složky, se kterými chceme pracovat v části **Výběr**, a v části **Možnosti** si buď sami nadefinujeme postupné binární dělení (*Define manually*), nebo si ho necháme nadefinovat stisknutím tlačítka *Default partition*, ten postupně odděluje vždy první složku. Dále si zde můžeme vybrat, zda chceme vypočítat všechny bilance a všechny statistiky.

Při vlastním definování postupného binárního dělení se nám zobrazí tabulka, do které budeme přiřazovat složkám znaménka + a – (podle vlastní preference) (Obrázek 27). V celém prvním sloupečku jsou dopředu nadefinovaná mínus, ta změníme na plus pouhým kliknutím v daném řádku. Po vyplnění prvního řádku musíme kliknout na tlačítko *Next*, tím se přesuneme o řádek níž. Nedefinované řádky jsou značeny zelenou barvou, po nadefinování jednotlivých řádků se barva

změní na červenou. Až je vyplněna celá tabulka, klikneme opět na *Next* a barva tabulky se změní na bílou. O jeden či více kroků zpět se můžeme posunout pomocí tlačítka *Previous*. Pokud jsme tabulku nadefinovali podle požadavků, klikneme opět na *Accept*, které vznikne z tlačítka *Next* po dokončení tabulky.



Obrázek 26: Nastavení bilančního dendrogramu.

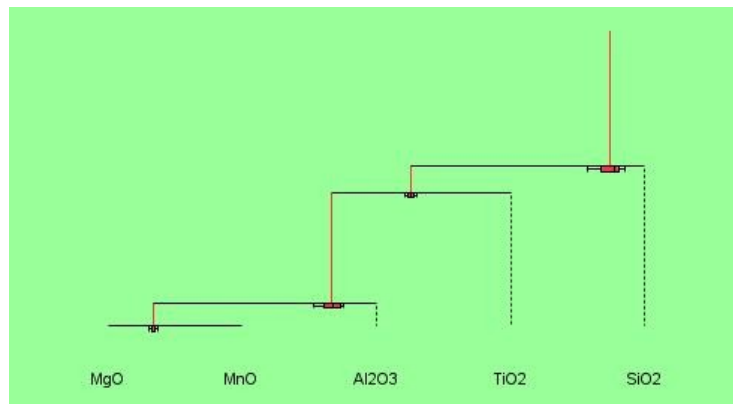


Obrázek 27: Tabulka pro vlastní definování postupného binárního dělení.

Po nastavení všech parametrů v části **Možnosti** klikneme na *Accept* a zobrazí se nám bilanční dendrogram v novém okně a v horní části základního okna budou výběrové průměry a rozptyly bilancí a do tabulky v dolní části se doplní jednotlivé hodnoty bilancí.

Samotný graf reprezentuje vypočítané hodnoty (Obrázek 29) pomocí horizontálních a vertikálních úseček (průměry bilancí, resp. jejich rozptyly), zobrazeny

jsou též příslušné boxploty pro jednotlivé bilance, viz Obrázek 28.



Obrázek 28: Bilanční dendrogram.

ILR binary partition:

sport	čtení	TV	PC	hudba	venku
1	-1	-1	-1	-1	1
1	0	0	0	0	-1
0	1	-1	-1	-1	0
0	0	-1	-1	1	0
0	0	-1	1	0	0

Mean:

Balance 1	Balance 2	Balance 3	Balance 4	Balance 5
0.6464	-0.7006	-0.9421	-1.3964	-0.1962

Variance:

Balance 1	Balance 2	Balance 3	Balance 4	Balance 5
0.8672	0.5506	1.1299	1.3029	0.6375

Obrázek 29: Číselné hodnoty bilančního dendrogramu, konkrétně aritmetický průměr a rozptyl.

3. Praktický příklad

Jako praktický příklad jsem si zvolila analýzu složení volného času dětí ve věku 10 – 11 let. Znovu si ukážeme některé funkce programu CoDaPack a také se zaměříme na interpretaci grafických výstupů programu.

Za účelem sběru dat jsem vytvořila dotazník pro děti mladšího školního věku, tedy pro 5. třídu, zaměřený na jejich volný čas (viz příloha 4.1). Počtem hodin či minut měly děti kvantifikovat šest zájmových činností, které byly zaměřeny jak na aktivní, tak pasivní zájmy. Patřil mezi ně *sport*, *čtení*, *sledování televize*, *čas strávený u počítače* (jak hraním her, tak chatováním či prohlížením internetu), *hudba* (jak aktivní, což je například hra na hudební nástroj nebo zpěv, tak pasivní, kam můžeme zařadit poslech hudby), a poslední *čas strávený venku*. Tato data jsem získala od 30 respondentů.

Jako první se zaměříme na kompoziční statistické metody. Můžeme přitom uvést, že pro kompoziční přístup je formát dat libovolný. Pokud však pracujeme se standardními statistickými metodami, je potřeba pracovat s procentuální reprezentací kompozic. Využitím funkce *Compositional statistic summary* jsme dostali dvě tabulky, první obsahuje postupně ve sloupečcích (Obrázek 30) geometrické průměry, které nám ukazují, jaký podíl času v průměru respondenti tráví danou činností (vzhledem k celkové době volného času), dále minima, jednotlivé kvartily a maxima.

Statistics						
	Center	0	25	50	75	100
sport	0.1124	0.0038	0.0511	0.0849	0.1663	0.3158
čtení	0.0466	0.0034	0.0176	0.0343	0.0894	0.2174
TV	0.2811	0.0303	0.1588	0.2466	0.3684	0.7019
PC	0.2129	0.0100	0.1109	0.2069	0.3347	0.6061
hudba	0.0442	0.0034	0.0156	0.0346	0.0900	0.2553
venku	0.3028	0.0381	0.1671	0.2493	0.3871	0.5958

Obrázek 30: Kompoziční statistické shrnutí – geometrické průměry, minima, kvartily, maxima.

Ve druhé tabulce (Obrázek 31) najdeme průměry logaritmů podílů jednotlivých složek (dolní diagonála čtvercové tabulky) a rozptyly logaritmů podílů složek (horní diagonála čtvercové tabulky), které jsou prvky symetrické, tzv. variační matice. Čím je menší hodnota rozptylu, tím je stabilnější podíl (a tedy těsnější vztah) mezi příslušnými kompozičními složkami. Takto vidíme, že spolu nejtěsněji souvisí sport a venkovní aktivity.

Variation array:

Xi\Xj	Variance $\ln(X_i/X_j)$						clr variances
	sport	čtení	TV	PC	hudba	venku	
sport		2.2601	1.3156	1.8263	1.9917	1.1011	0.7079
čtení	-0.8803		2.1056	2.5770	1.7772	1.4043	0.8437
TV	0.9164	1.7967		1.2750	2.0381	1.2158	0.6625
PC	0.6389	1.5192	-0.2775		2.5081	1.7296	0.8263
hudba	-0.9326	-0.0523	-1.8490	-1.5715		1.8024	0.8431
venku	0.9908	1.8711	0.0744	0.3519	1.9234		0.6044
Mean $\ln(X_i/X_j)$							4.4880 Total Variance

Obrázek 31: Kompoziční statistické shrnutí – rozptyly a průměry logaritmů podílů složek, CLR rozptyly, celkový rozptyl

Druhým krokem budou standardní statistické metody, které dostaneme využitím funkce *Classical statistic summary*. Jak je uvedeno výše, je potřeba převést data na proporcionální podíly složek. Pokud tak neučiníme, dostaneme výsledky, které nelze porovnat s výstupy kompozičních metod. Ukážeme si obě varianty. Pokud data neupravíme, dostaneme tyto výsledky: v první tabulce (Obrázek 32) máme k dispozici (opět postupně ve sloupečcích) aritmetické průměry, které by měly vyjadřovat průměrné hodnoty složek pro tento soubor a příslušné směrodatné odchylky. Dále jsou uvedena minima, kvartily a maxima a druhá tabulka (Obrázek 33) zobrazuje korelační matici. Jak víme ze základního kurzu statistiky, korelační matice je symetrická a prvky na hlavní diagonále jsou rovny jedné. Prvky, které neleží na diagonále, náleží do intervalu $[-1, 1]$ a hodnoty blíže krajním bodům intervalu značí těsnější lineární vztah mezi proměnnými. Můžeme si všimnout, že se závěry získané pomocí korelační matice budou lišit od těch, obdržených užitím variační matice. Je zřejmé, že interpretaci kompozičních metod

zde dáváme přednost. Výsledky pro upravená (proporcionální) data jsou následující: první tabulka (Obrázek 34) zobrazuje opět aritmetické průměry, tyto však nyní můžeme porovnávat s geometrickými (Obrázek 30) a vidíme, že se průměrné hodnoty složek pomocí obou přístupů liší přece jen méně než v případě korelací (u předchozího srovnání i v případě hodnot z Obrázku 35).

Statistics							
	Mean	Std.Dev	0	25	50	75	100
sport	283.5000	191.2901	10.0000	180.0000	290.0000	360.0000	900.0000
čtení	151.6667	151.7308	10.0000	60.0000	60.0000	240.0000	600.0000
TV	692.5000	400.1390	60.0000	420.0000	720.0000	900.0000	1680.0000
PC	615.6667	449.0337	30.0000	240.0000	600.0000	960.0000	1680.0000
hudba	154.1667	166.0576	10.0000	50.0000	75.0000	180.0000	600.0000
venku	762.8333	560.6078	120.0000	300.0000	600.0000	1020.0000	2400.0000

Obrázek 32: Standardní statistické shrnutí – aritmetické průměry, směrodatné odchylky, minima, kvartily, maxima.

Correlation:							
	sport	čtení	TV	PC	hudba	venku	
sport	1.0000	0.0859	-0.1385	0.0400	-0.1732	0.1360	
čtení	0.0859	1.0000	-0.2372	-0.2380	0.2305	0.2601	
TV	-0.1385	-0.2372	1.0000	0.2357	-0.0280	-0.0813	
PC	0.0400	-0.2380	0.2357	1.0000	-0.1680	-0.0409	
hudba	-0.1732	0.2305	-0.0280	-0.1680	1.0000	0.0024	
venku	0.1360	0.2601	-0.0813	-0.0409	0.0024	1.0000	

Obrázek 33: Standardní statistické shrnutí – korelace.

	Mean	Std.Dev	0	25	50	75	100
clo(sport)	0.1148	0.0796	0.0038	0.0511	0.0849	0.1663	0.3158
clo(čtení)	0.0602	0.0593	0.0034	0.0176	0.0343	0.0894	0.2174
clo(TV)	0.2635	0.1547	0.0303	0.1588	0.2466	0.3684	0.7019
clo(PC)	0.2225	0.1457	0.0100	0.1109	0.2069	0.3347	0.6061
clo(hudba)	0.0614	0.0643	0.0034	0.0156	0.0346	0.0900	0.2553
clo(venku)	0.2776	0.1513	0.0381	0.1671	0.2493	0.3871	0.5958

Obrázek 34: Standardní statistické shrnutí po úpravě na proporcionální části – aritmetické průměry, směrodatné odchylky, minima, kvartily, maxima.

	clo(sport)	clo(čtení)	clo(TV)	clo(PC)	clo(hudba)	clo(venku)
clo(sport)	1.0000	-0.0267	-0.3628	-0.2221	-0.0809	0.1038
clo(čtení)	-0.0267	1.0000	-0.2685	-0.2428	0.1721	0.0573
clo(TV)	-0.3628	-0.2685	1.0000	-0.1552	-0.0920	-0.5383
clo(PC)	-0.2221	-0.2428	-0.1552	1.0000	-0.2394	-0.4911
clo(hudba)	-0.0809	0.1721	-0.0920	-0.2394	1.0000	-0.1251
clo(venku)	0.1038	0.0573	-0.5383	-0.4911	-0.1251	1.0000

Obrázek 35: Standardní statistické shrnutí po úpravě – korelace.

Následující metodou bude test normality. Příkaz pro spuštění je *Statistic* → *Additive Logistic Normality Tests*. Obecně se testuje nulová hypotéza proti alternativě. V tomto případě zní nulová hypotéza: náhodný výběr pochází z mnohorozměrného normálního rozdělení (alr transformovaných dat) a alternativní hypotéza: náhodný výběr z mnohorozměrného normálního rozdělení nepochází. V našem případě se přitom používá p -hodnota místo porovnávání realizace testovacího kritéria s kritickými hodnotami. Uvedená p -hodnota je pravděpodobnost, že vypočítaná hodnota testovacího kritéria má právě realizovanou hodnotu či hodnotu ještě více odporující nulové hypotéze. Pokud je p -hodnota menší nebo rovna volené hladině testu α , pak nulovou hypotézu (náhodný výběr pochází z mnohorozměrného normálního rozdělení) zamítáme na hladině testu α . Nulovou hypotézu na dané hladině nemůžeme zamítnout, pokud je p -hodnota větší než α . V tomto případě nulovou hypotézu nezamítáme, jelikož p -hodnoty nejsou ani v jednom případě menší než $\alpha = 0.05$. Tedy nemůžeme zamítnout, že data pochází z normálního rozdělení (Obrázek 36).

Normality Tests:						
	Anderson-Darling		Cramer-von Mises		Watson	
	A^2*	p	W^2*	p	U^2*	p
Marginals						
alr(sport,venku)	0.6306	[0.15, 0.10]	0.0628	>0.15	0.0534	>0.15
alr(čtení,venku)	0.2135	>0.15	0.0262	>0.15	0.0262	>0.15
alr(TV,venku)	0.2744	>0.15	0.0356	>0.15	0.0333	>0.15
alr(PC,venku)	0.2420	>0.15	0.0293	>0.15	0.0269	>0.15
alr(hudba,venku)	0.6415	[0.15, 0.10]	0.0900	>0.15	0.0884	[0.15, 0.10]

Bivariate angle						
alr(sport,venku)	0.4083	>0.15	0.0589	>0.15	0.0678	>0.15
alr(čtení,venku)						
alr(sport,venku)	0.1044	>0.15	0.0079	>0.15	0.0175	>0.15
alr(TV,venku)						
alr(sport,venku)	0.7512	>0.15	0.1480	>0.15	0.1095	>0.15
alr(PC,venku)						
alr(sport,venku)	0.9250	>0.15	0.1024	>0.15	0.0095	>0.15
alr(hudba,venku)						
alr(čtení,venku)	0.3002	>0.15	0.0435	>0.15	0.0518	>0.15
alr(TV,venku)						
alr(čtení,venku)	0.2641	>0.15	0.0323	>0.15	0.0420	>0.15
alr(PC,venku)						
alr(čtení,venku)	0.3753	>0.15	0.0679	>0.15	0.0729	>0.15
alr(hudba,venku)						
alr(TV,venku)	0.2804	>0.15	0.0545	>0.15	0.0641	>0.15
alr(PC,venku)						
alr(TV,venku)	0.6376	>0.15	0.0691	>0.15	-0.0009	>0.15
alr(hudba,venku)						
alr(PC,venku)	0.8525	>0.15	0.0841	>0.15	-0.0229	>0.15
alr(hudba,venku)						
Radius						
	0.7812	>0.15	0.0837	>0.15	0.0826	>0.15

Obrázek 36: Test normality.

Další je grafické zobrazení dat sestrojením biplotu pro clr transformovaná kompoziční data. Body na obrázku zobrazují jednotlivá pozorování a směrodatné odchylky jsou aproximovány pomocí délek šipek, reprezentujících jednotlivé clr proměnné, potažmo původní kompoziční složky. Ještě můžeme pozorovat vzdálenost mezi vrcholy dvou šipek, která aproximuje rozptyl příslušného logaritmu podílu.

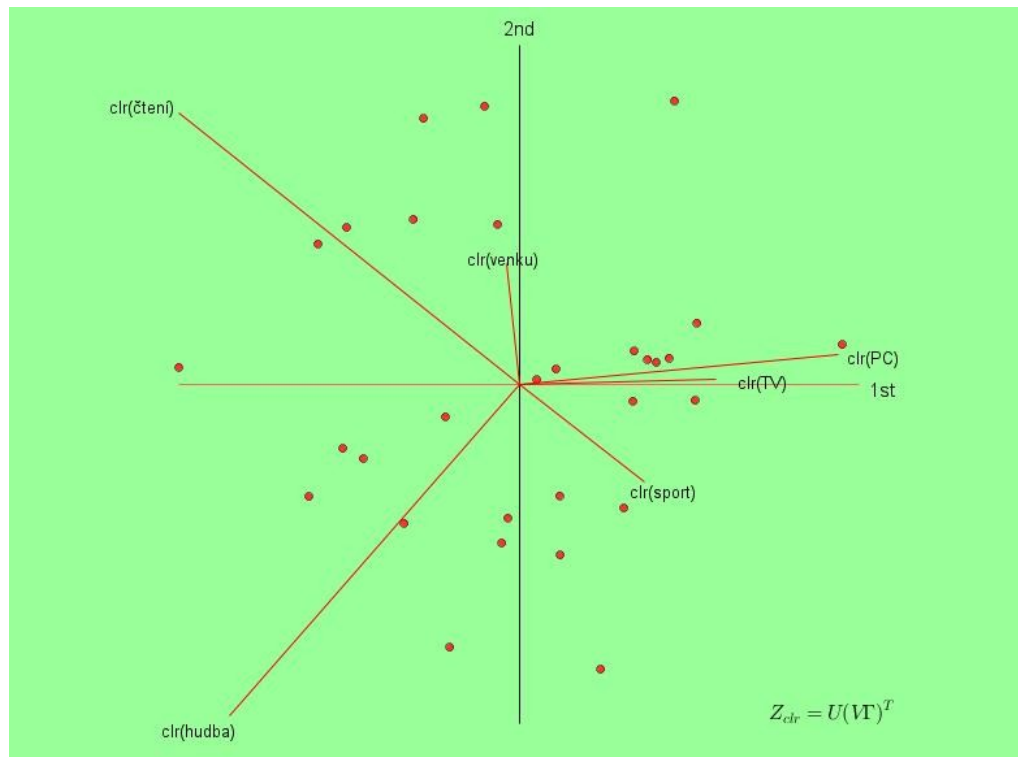
Číselný výstup biplotu (Obrázek 37) vypovídá o tom, kolik procent variability je vysvětleno hlavními komponentami, proměnnými, které (konkrétně první dvě) jsou pro konstrukci biplotu využity. Například hlavní komponenta číslo 1 popisuje 36 % celkové variability datového souboru. Společně s druhou hlavní komponentou tak vysvětlují přes 57 % celkové variability.

Na grafickém výstupu (Obrázek 38) můžeme pozorovat, že největší směrodatnou odchylku mají složky **čtení** a **hudba**, naopak nejmenší je u složky **venku**. Tyto složky tak mají na výsledné zobrazení dat v biplotu největší vliv. Co se týká

rozptylu logaritmu podílu složek, ukazuje se, že silný je např. vztah složek **TV** a **PC**, což také odpovídá tomu, že obě dvě činnosti jsou převážně pasivní.

Principal Components:							
	clr(sport)	clr(čtení)	clr(TV)	clr(PC)	clr(hudba)	clr(venku)	Cum.Prop.Exp.
PC1	0.2101	-0.5698	0.3309	0.5351	-0.4846	-0.0217	0.3604
PC2	-0.2133	0.5958	0.0125	0.0649	-0.7246	0.2648	0.5717
PC3	-0.6590	0.1475	0.0307	0.5987	0.2390	-0.3569	0.7801
PC4	0.3310	0.0497	-0.8380	0.4284	-0.0162	0.0450	0.9038
PC5	0.4470	0.3598	0.1434	-0.0278	-0.1266	-0.7958	1.0000

Obrázek 37: Biplot – výstupy metody hlavních komponent.

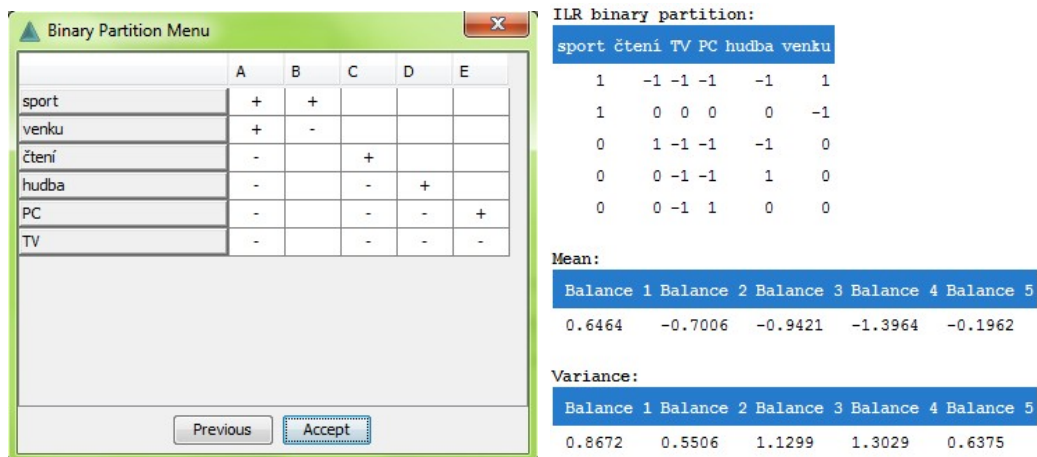


Obrázek 38: Biplot.

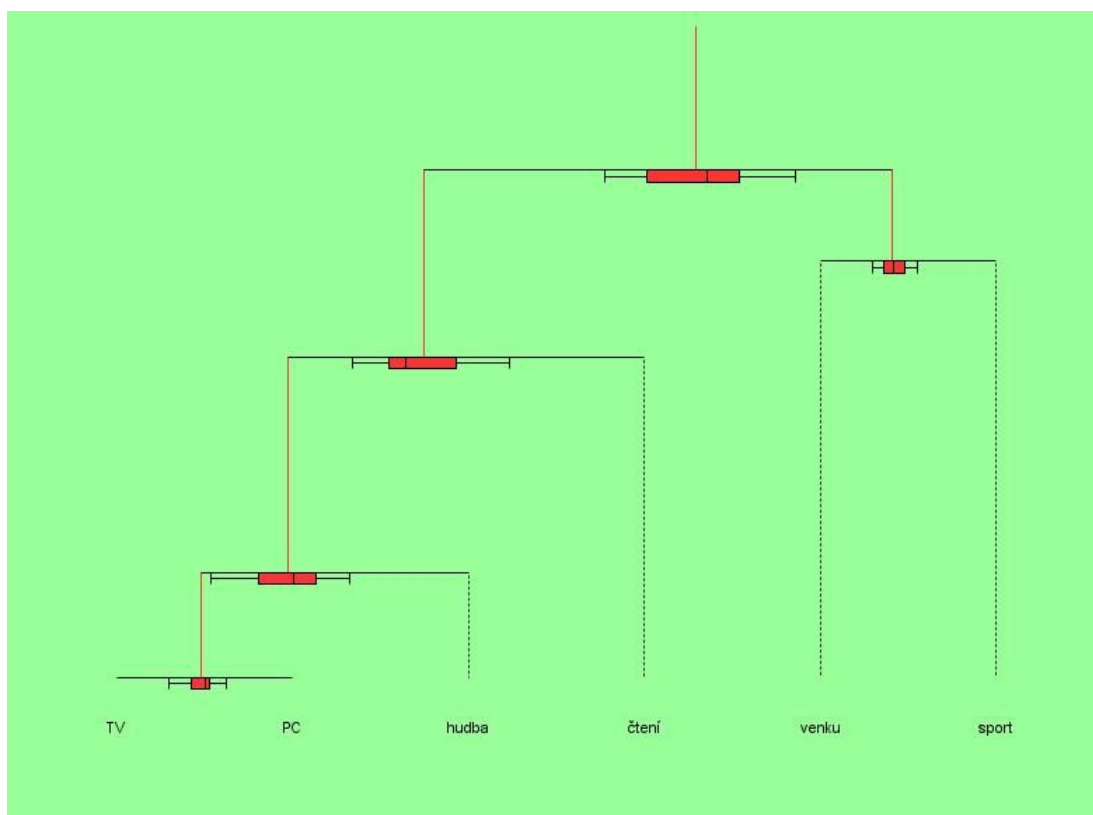
Poslední funkcí, kterou si ukážeme a popíšeme, bude bilanční dendrogram. K jeho vytvoření potřebujeme nadefinovat postupné binární dělení. To jsme udělali následujícím způsobem (Obrázek 39): první dělení vytvoří skupinu složek odpovídajících aktivně trávenému volnému času, zahrnující **sport** a **venku**, a skupinu složek pasivně tráveného volného času, zahrnující **čtení**, **hudba**, **TV** a **PC**.

Skupina aktivního trávení času se v následujícím kroku rozdělí na činnost rozvíjející osobnost (**sport**) a nikoli (**venku**). Tím dělení aktivně tráveného času skončilo. Přejdeme tedy k dělení pasivně tráveného času. Zde se budeme opět zabývat rozvojem osobnosti, od ostatních se ale oddělí pouze **čtení**. V následujícím kroku budeme opět posuzovat míru rozvoje a znaménko + dostane **hudba**. Poslední dvojici **TV** a **PC** můžeme rozdělit libovolně. V tomto případě jsme zvolili + u složky **PC**.

Po dokončení nastavení v dialogovém okně se zobrazí výstup jak číselný (Obrázek 39), tak grafický (Obrázek 40). V číselné tabulce dostaneme hodnoty výběrového průměru a rozptylu jednotlivých bilancí. Tyto hodnoty jsou rovněž zaznamenány v dendrogramu. Největší rozptyl najdeme u čtvrté bilance, poměry mezi složkami **TV** a **PC** na jedné straně a složkou **hudba** se tedy v rámci datového souboru výrazně liší. **TV**, **PC** a **hudba** sice patří do pasivních činností, ale nemůžeme slučovat hru na hudební nástroj a „válení“ se u televize. Naopak nejmenší rozptyl má druhá bilance (**sport**, **venku**), to znamená, že podíl mezi příslušnými složkami je zde stabilní a vztah těsný. O složkách tedy můžeme říct, že přesto, že sport je řízená a více rozvíjející činnost, dochází u obou aktivit k pohybu a spalování energie a platí, že kdo se věnuje jedné aktivitě, bude se pravděpodobně věnovat i druhé z nich. Jen o málo větší rozptyl než u druhé bilance najdeme u páté, kde **TV** a **PC** jsou činnosti tématicky velmi podobné.



Obrázek 39: Tabulka s vlastním binárním dělením a tabulky číselných výstupů dendrogramu.



Obrázek 40: Bilanční dendrogram.

Závěr

V této práci jsem chtěla ukázat, jak pracovat se softwarem CoDaPack. Popsala jsem jeho jednotlivé funkce a většinu z nich interpretovala na příkladech.

Při tvorbě této práce jsem se zdokonalila také v práci s programem $\text{\TeX}$, objevila jsem některé jeho možnosti a také jsem si některé potřebné příkazy zautomatizovala. Zajímavá pro mě taky byla zkušenost v souvislosti s prováděním průzkumu se školními dětmi a s jejich bezprostředními reakcemi při vyplňování dotazníků. Nové znalosti oceňuji a doufám, že mi budou do dalších studijních let přínosem. V softwaru CoDaPack mě práce bavila, ale následná interpretace výsledků již tak přitažlivá pro svoji obtížnost nebyla. Seznamovala jsem se s novými grafickými zobrazeními jako biplot či bilanční dendrogram a také překlady z angličtiny mi zabraly hodně času. Doufám tedy, že pro nové uživatele softwaru CoDaPack bude tato práce užitečná a zjednoduší jim seznámení s jednotlivými jeho funkcemi.

Literatura

- [1] Aitchison, J., *The statistical analysis of compositional data*. Chapman and Hall, London, 1986.
- [2] Buccianti, A., Mateu-Figureas, G., Pawlowsky-Glahn, V., *Compositional data analysis in the geosciences: From theory to practice*. The Geological Society, London, 2006.
- [3] Hron, K., Elementy statistické analýzy kompozičních dat. In *Informační bulletin České statistické společnosti*, 2010, roč. 21, č. 3, s. 41 – 48.
- [4] Hron, K., Kunderová, P., *Základy počtu pravděpodobnosti a metod matematické statistiky*. Olomouc: Vydavatelství Univerzity Palackého, 2013.
- [5] Manuál programu CoDaPack [online], dostupné z: <http://ima.udg.edu/codapack/assets/codapack-manual.pdf>, [citováno 20. 9. 2012]
- [6] Pawlowsky-Glahn, V., Buccianti, A., *Compositional data analysis: Theory and applications*. Wiley, Chichester, 2011.
- [7] Pawlowsky-Glahn, V., Egozcue, J. J., Tolosana-Delgado, R., *Lecture notes on compositional data analysis* [online], dostupné z: <http://dugi-doc.udg.edu/bitstream/10256/297/1/CoDa-book.pdf> [citováno 17. 11. 2012]
- [8] Thió-Henestrosa, S., Comas, M., CoDaPack 2.0: A stand-alone, multi-platform compositional software [online], dostupné z: congress.cimne.com/codawork11/Admin/Files/FilePaper/p28.pdf, [citováno 20. 9. 2012]

4. Přílohy

4.1. Složení volného času dětí ve věku 10 – 11 let

Data jsou uvedena v minutách a interpretují podíly na celkovém složení volného času dětí.

	sport	čtení	TV	PC	hudba	venku
1	300	420	720	420	35	2400
2	900	600	420	90	60	780
3	180	60	840	310	180	180
4	720	70	840	1260	35	840
5	480	60	900	600	50	1320
6	60	300	600	60	60	300
7	10	240	420	660	55	1260
8	300	60	120	60	10	540
9	180	30	900	120	365	525
10	390	120	720	720	90	720
11	90	10	1260	120	15	300
12	180	35	420	600	30	420
13	300	420	300	30	270	1680
14	330	60	780	840	60	420
15	390	60	840	600	180	210
16	240	10	1440	600	10	600
17	600	25	300	1200	300	420
18	60	300	60	1200	60	300
19	120	210	420	280	600	720
20	300	60	600	600	300	1800
21	360	120	360	240	120	600
22	180	300	1200	420	560	840
23	120	420	600	720	540	300
24	360	20	180	240	75	1290
25	135	180	1080	1200	10	240
26	290	60	75	180	135	255
27	300	60	1020	1380	60	1020
28	240	60	600	1680	60	745
29	210	60	1680	960	120	120
30	180	120	1080	1080	180	1740

JAK TRÁVÍM SVŮJ VOLNÝ ČAS

Napište, kolik hodin nebo minut za týden věnujete těmto zájmům, pokud zrovna nejste ve škole nebo nepomáháte rodičům.

Např.: Hra na hudební nástroj 45 min.
 Sport – sport s rodiči 0
 Sport – sportovní kroužky 1 h 30 min.

HRA NA HUDEBNÍ NÁSTROJ	_____
SPORT – SPORT S RODIČI	_____
SPORT – SPORTOVNÍ KROUZKY	_____
ČTENÍ	_____
SLEDOVÁNÍ TELEVIZE	_____
HRANÍ NA POČITAČI	_____
POSLOUCHÁNÍ HUDBY	_____
ZÁJMOVÉ KROUZKY (dramatik, výtvarná, ručičák, elektro-kroužek)	_____
VENKU S KAMARADY	_____
NICNEDĚLÁNÍ	_____

Další:

_____	_____
_____	_____
_____	_____

4.2. Geochemické složení hornin

	SiO ₂	TiO ₂	Al ₂ O ₃	MnO	MgO	CaO	Na ₂ O	K ₂ O	P ₂ O ₅	Fe ₂ O _{3t}
RT1-1	73,4	0,22	14,19	0,037	0,48	1,04	4,33	4,18	0,06	1,74
RT1-1	73,86	0,26	14	0,038	0,51	1,14	4,32	3,95	0,07	1,72
RT1-1	73,85	0,3	13,85	0,041	0,55	1,15	4,31	3,82	0,07	1,82
RT1-1	78,12	0,32	11,36	0,043	0,57	1,02	3,41	3,16	0,06	1,81
RT1-1	78,77	0,39	10,29	0,061	0,86	1,07	2,81	2,92	0,07	2,54
RT1-1	74,43	0,54	11,88	0,073	0,94	2,22	2,95	3,2	0,29	2,96
RT1-1	63,09	0,73	15,99	0,121	2,24	2,17	3,1	4,58	0,28	5,49
RT1-1	60,27	0,85	18,05	0,108	2,23	2,1	3,47	5,05	0,24	5,13
RT1-1	59,15	1,01	19,1	0,094	1,87	2,32	3,77	4,71	0,26	4,49
RT1-1	54,62	0,85	19,19	0,16	3,69	1,32	2,52	6,05	0,13	7,58
RT1-1	49,49	1,03	18,63	0,219	4,02	1,62	2,07	5,4	0,22	9,36
RT1-10	71,66	0,14	15,34	0,025	0,37	0,91	4,82	4,76	0,05	1,11
RT1-10	71,22	0,17	15,92	0,028	0,39	1,02	5,24	4,49	0,05	1,18
RT1-10	73,51	0,19	14,45	0,026	0,36	0,97	4,78	3,97	0,05	1,13
RT1-10	80,51	0,22	10,45	0,027	0,37	0,82	3,39	2,79	0,05	1,14
RT1-10	83,69	0,3	8,47	0,036	0,43	0,92	2,66	2,14	0,07	1,24
RT1-16	72,47	0,14	14,84	0,025	0,33	0,71	4,44	5,14	0,05	1,05
RT1-16	72,82	0,15	14,5	0,024	0,27	0,8	4,56	4,68	0,04	0,98
RT1-16	72,53	0,25	14,66	0,029	0,28	0,97	4,82	4,36	0,05	1,1
RT1-16	77,12	0,26	11,73	0,029	0,22	0,78	3,88	3,54	0,04	0,97
RT1-16	83,62	0,15	8,81	0,022	0,22	0,43	2,85	2,81	0,03	0,78
RT1-16	82,23	0,22	9,91	0,024	0,16	0,87	3,25	2,93	0,1	0,76
RT1-16	76,44	0,28	12,48	0,033	0,36	1,17	3,94	3,53	0,14	1,31
RT1-16	71,65	0,3	14,61	0,032	0,41	1,13	4,72	3,93	0,12	1,34
RT1-16	65,25	0,42	19,82	0,033	0,41	1,35	4,87	3,94	0,15	1,47
RT1-16	66,96	0,42	16,44	0,076	1,21	1,17	3,87	4,8	0,16	3,36
RT1-16	62,28	0,53	17,33	0,107	1,87	1,24	3,78	5,3	0,14	4,63
RT1-17	72,26	0,13	14,51	0,022	0,29	0,72	4,37	5,11	0,04	0,98
RT1-17	73,26	0,15	14,78	0,023	0,27	0,81	4,69	4,73	0,04	0,98
RT1-17	72,79	0,21	14,59	0,026	0,26	0,9	4,82	4,41	0,05	1,01
RT1-17	77,57	0,24	11,84	0,026	0,19	0,73	3,97	3,62	0,04	0,84
RT1-17	83,63	0,17	8,93	0,022	0,16	0,48	2,95	2,79	0,03	0,67
RT1-19	71,76	0,16	14,49	0,027	0,34	1,04	4,82	3,9	0,05	1,21
RT1-19	73,97	0,18	14,49	0,024	0,3	1,03	4,95	3,79	0,05	1,11

RT1-19	74,64	0,2	14,13	0,024	0,26	0,98	4,91	3,69	0,05	1,02
RT1-19	78,65	0,3	11,69	0,028	0,24	0,99	4,13	2,88	0,05	1,03
RT1-19	80,42	0,31	10,15	0,038	0,44	0,94	3,36	2,61	0,05	1,44
RT1-19	75,49	0,42	12,16	0,047	0,49	2,11	3,64	2,99	0,24	1,89
RT1-19	67,93	0,56	15,31	0,067	1,02	2,44	4,17	3,78	0,29	2,98
RT1-19	64,58	0,71	17,21	0,061	0,99	2,48	4,8	3,99	0,24	2,81
RT1-19	61,14	1,01	20,09	0,063	0,86	2,96	4,89	3,94	0,31	2,72
RT1-19	63,42	0,64	17,8	0,078	1,58	1,72	4,33	4,78	0,18	3,72
RT1-19	54,02	0,88	18,83	0,189	3,43	1,54	3,02	5,72	0,19	7,93
RT1-20	72,25	0,1	13,75	0,018	0,23	0,72	4,58	4,15	0,04	0,8
RT1-20	74,34	0,11	14,27	0,018	0,23	0,83	4,91	3,9	0,04	0,84
RT1-20	74,31	0,13	14,39	0,018	0,22	0,83	5,08	3,84	0,04	0,82
RT1-20	77,97	0,22	11,73	0,023	0,24	0,85	4,19	2,89	0,04	0,9
RT1-20	79,55	0,25	10,16	0,039	0,58	0,75	3,28	2,77	0,05	1,63
RT1-3	73,47	0,26	13,97	0,04	0,58	1,2	4,21	3,99	0,08	1,84
RT1-3	73,02	0,3	13,99	0,042	0,57	1,17	4,3	3,93	0,08	1,86
RT1-3	73,03	0,29	13,52	0,039	0,52	1,13	4,25	3,78	0,07	1,73
RT1-3	77,9	0,33	11,6	0,04	0,51	1,06	3,56	3,16	0,06	1,69
RT1-3	78,97	0,41	10,2	0,056	0,69	1,25	2,88	2,75	0,08	2,21
RT1-7A	71,11	0,22	15,86	0,031	0,51	1,43	5,47	3,53	0,07	1,5
RT1-7A	70,63	0,26	16,14	0,033	0,53	1,52	5,72	3,44	0,08	1,55
RT1-7A	71,5	0,31	15,7	0,033	0,46	1,5	5,67	3,28	0,08	1,44
RT1-7A	78,92	0,31	11,6	0,027	0,34	1,14	4,2	2,4	0,06	1,07
RT1-7A	82,43	0,3	9,24	0,029	0,36	1,01	3,27	1,93	0,07	1,07
RT1-7A	77,18	0,46	11,37	0,043	0,47	2,13	3,69	2,29	0,28	1,7
RT1-7A	68,06	0,58	15,3	0,065	1,19	2,69	4,2	3,3	0,32	3,02
RT1-7A	64,21	0,76	17,48	0,062	1,24	2,78	4,67	3,78	0,3	3,08
RT1-7A	63,7	0,85	18,52	0,054	1,06	2,65	5,07	3,79	0,29	2,71
RT1-7A	58,54	0,74	18,93	0,103	3,06	1,78	3,59	5,11	0,18	5,33
RT1-7A	54,52	0,88	19,06	0,154	4,12	1,74	3,19	5,51	0,17	7,32
RT1-7B	71,11	0,25	15,62	0,033	0,63	1,32	5,31	3,59	0,07	1,59
RT1-7B	64,45	0,26	14,21	0,031	0,64	1,63	6,33	3,15	0,08	1,44
RT1-7B	70,9	0,32	15,66	0,034	0,52	1,52	5,6	3,25	0,08	1,49
RT1-7B	78,47	0,33	11,59	0,029	0,41	1,18	4,2	2,39	0,06	1,16
RT1-7B	78,56	0,32	8,89	0,03	0,43	1,03	3,3	1,88	0,08	1,17

RT1-8	72,83	0,21	14,24	0,028	0,54	1,1	5,32	2,71	0,06	1,32
RT1-8	70,76	0,18	15,65	0,026	0,46	1,12	5,76	3,33	0,06	1,22
RT1-8	71,71	0,19	15,29	0,023	0,45	1,07	5,73	3,11	0,06	1,15
RT1-8	78,44	0,21	11,51	0,022	0,42	0,86	4,33	2,3	0,05	1
RT1-8	81,37	0,27	9,7	0,03	0,63	0,88	3,46	1,96	0,08	1,34
RT1-8	76,02	0,46	12,06	0,045	0,76	1,95	3,93	2,39	0,28	2,04
RT1-8	66,3	0,62	15,82	0,071	1,63	2,37	4,27	3,44	0,31	3,51
RT1-8	64,14	0,7	17,75	0,059	1,45	2,24	5,01	3,77	0,26	3,1
RT1-8	60,97	0,85	20,29	0,058	1,28	2,45	5,18	3,61	0,32	2,96
RT1-8	57,58	0,73	19,05	0,107	3,41	1,7	3,47	5,16	0,18	5,63
RT1-8	51,85	0,89	19,25	0,162	4,69	1,59	2,82	5,68	0,19	7,8
RT1-9	75,01	0,3	11,14	0,071	2	0,59	2,56	3,52	0,04	3,38
RT1-9	78,5	0,29	9,97	0,057	1,5	0,66	2,54	2,93	0,05	2,69
RT1-9	75,25	0,47	11,69	0,066	1,2	1,74	3,03	3,07	0,25	2,91
RT1-9	63,75	0,67	16,08	0,109	2,61	2,07	3,3	4,4	0,26	5,22
RT1-9	62,37	0,92	17,47	0,087	1,83	2,89	4,21	4,07	0,35	4,1
RT1-9	55,69	0,67	18,59	0,138	4,34	1,17	2,67	5,84	0,13	7,04
RT1-9	50,13	0,79	19,65	0,179	5,66	1,1	1,98	6,2	0,14	9,04