SLOVANSKÉ GYMNÁZIUM V OLOMOUCI
Třída Jiřího z Poděbrad 13, Olomouc, 77900, kraj Olomoucký

SOUTĚŽNÍ PRÁCE SOČ

OBOR 01. MATEMATIKA A STATISTIKA

REGRESNÍ ANALÝZA PRO KOMPOZIČNÍ DATA

Autor: Jaroslava Geletičová (4.D)

Odborný konzultant: RNDr. Karel Hron, Ph.D.,
Přírodovědecká fakulta UP Olomouc,
katedra matematické analýzy a aplikací matematiky

Zadavatel práce: -

Rok vypracování: 2008

ČESTNÉ PROHLÁŠENÍ:

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

V Olomouci dne 11.4. 2008

Podpis:

PŘEDMLUVA

Po loňské a předloňské kladné zkušenosti jsem se i tento rok rozhodla vypracovat soutěžní práci v rámci Středoškolské odborné činnosti. V průběhu letní školy pořádáné Sdružením pro podporu talentované mládeže jsem načerpala mnoho inspirace, kreativních myšlenek a zlepšujích návrhů, a proto jsem se rozhodla radikálně neměnit zkoumanou oblast. Vzhledem k stále rozvíjecím se zkušenostem z předchozích praci jsem přistoupila k ponechání některých kapitol, jež byly patřičně doplněny.

Je mi již známo, jak postupovat při řešení této problematiky, jak pracovat s matematickým textem a jak uplatnit zkušenosti s prací se softwarem R. Navíc jsem tento rok svou práci zpracovala pomocí softwaru Latex, jehož nepoužití mi bylo v minulosti vytýkáno.

Jako téma práce jsem si opět vybrala téma blízce související s oborem statistiky, přičemž jsem mohla využít vlastních experimentálních výsledků. Jelikož se již letos chystám na vysokou školu se zaměřením na obor biofyziky, jejíž významnou součástí je právě statistika, shledávám toto téma i jeho aplikaci za velmi výhodnou.

Práce je kvůli přehlednosti členěna do pěti kapitol. První kapitola je věnována nezbytné matematické teorii, z níž vycházejí základy této práce SOČ. Definice a názorné příklady se zabývají především maticemi, determinanty a nejrůznějšími operacemi s nimi. V druhé kapitole se zaobírám kompozičními daty, speciálním typem mnohorozměrných dat, užívaných pro statistickou anylýzu. Tato relativně nová oblast matematiky je často složitě interpretována, proto jsem se pokoušela tuto teorii zpracovat co nejsrozumitelněji, aby i laik byl schopen porozumět cílům této práce. V rámci této kapitoly jsem se dále zabývala jednou speciální transformací (tzv. isometric logratio transformací, ilr) těchto dat a způsoby, jak s nimi dále pracovat. Ve třetí kapitole se dostávám k ortogonální regresi, jinak nazývané jako metoda totálních čtverců. Pomocí ní a jejích výhodných vlastností jsem posléze vytvořila vhodný model pro regresní analýzu kompozičních dat s využitím ilr transformace. Ukázalo se totiž, že metodu nejmenších čtverců, se kterou jsem pracovala minulý rok, zde není možné použít. I v této kapitole jsem se pokoušela o názornost použitím jednoduchých příkladů. Následující kapitola s názvem "Software R" má za úkol objasnit práci se stejnojmenným statistickým softwarem. Jsou zde popsány základní pracovní příkazy, jejichž použití je názorně předvedeno ve velkém množství příkladů. Ná závěr řeším v této práci příklad z reálného prostředí, jež má ukázat výhodnost popisované metody.

Touto cestou děkuji panu profesorovi Karlu Hronovi. I letos mi byl ochoten poskytnout pomoc při tvoření práce. Především opravil chyby, doporučil literaturu a dal mi možnost diskuze na toto téma. Dále bych chtěla poděkovat panu RNDr. Martinu Kubalovi, Ph.D. z katedry experimentální fyziky PřF UP, který mi poskytl příležitost k provedení experimentů, čímž mne nechal nahlédnout do vědeckého prostředí. Za tuto pomoc jsem velmi vděčná.

1 MATICOVÁ ALGEBRA

Při výpočtu odhadů neznámých parametrů v modelu ortogonální regrese, uvedeného dále, se nevyhneme práci s maticovými výrazy. Proto v této pomocné kapitole, využívajících poznatků z práce [4], uvedeme potřebné operace s maticemi a determinanty a teoretické výsledky budeme prezentovat na vhodných příkladech.

1.1 MATICE

Definice 1: (dle [7], str. 157) *Matice typu (m, n)* je soustava $m \times n$ reálných čísel uspořádaných do m řádků a n sloupců,

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

Čísla $(a_{ij})_{i=1,\dots,m}^{j=1,\dots,n}$, se nazývají prvky matice.

Poznámka 1: (dle [7], str. 157) Matici obvykle označujeme velkým tučným tiskacím písmenem, např. **A**. Chceme-li označit i její prvky, píšeme

$$\mathbf{A}=(a_{ij})_{i=1,\dots,m}^{j=1,\dots,n},$$

nebo zkráceně pouze $\mathbf{A}=(a_{ij})$.

Poznámka 2: Jestliže $m=n$, pak se matice nazývá *čtvercová* a n je řád matice **A**.

Příklad 1: Příklady matic jsou

$$\mathbf{A}=\begin{pmatrix} 1 & 5 & 2 \\ 8 & 3 & 0 \\ 9 & 7 & 6 \end{pmatrix}, \mathbf{B}=\begin{pmatrix} -4 & 1 \\ 3 & -2 \\ 7 & 0 \end{pmatrix}.$$

Matice **A** je typu 3 x 3 a můžeme ji označit jako čtvercovou matici řádu 3. Matice **B** je typu 3 x 2.

Definice 2: Nechť **A** je čtvercová matice, pak prvky $a_{11}, a_{22}, \dots, a_{nn}$ tvoří *hlavní diagonálu* matice **A**.

Definice 3: (dle [7], str. 158) *Nulová matice* libovolného typu *(m, n)* je matice, jejíž všechny prvky jsou rovny nule, značíme ji **O**.

Jednotková matice je čtvercová matice, která má prvky hlavní diagonály rovny jedné a ostatní prvky rovny nule, značíme ji **I**.

Příklad 2: (dle [7], str. 158)

$$\begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Uvedené matice jsou jednotkové matice řádu 1, 2 a 3. Odkud mimo jiné vidíme, že každé reálné číslo je maticí řádu 1.

1.2 DETERMINANTY

Definice 4: *Determinantem čtvercové matice* **A** rozumíme číslo

$$\mathbf{A}= \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{2n} & \dots & a_{nn} \end{vmatrix},$$

ve speciálním případě $n = 2, 3$ získáme det **A** tzv. *Sarrusovým pravidlem.*

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (a_{11}a_{22}) - (a_{21}a_{12}) \, \mathrm{pro} \, n = 2,$$

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$$= (a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}) - (a_{13}a_{22}a_{31} + a_{21}a_{12}a_{33} + a_{11}a_{32}a_{23}) \, \mathrm{pro} \, n = 3.$$

Poznámka 3: Výpočet det **A** pro čtvercové matice vyšších řádů uvádět nebudeme. Příslušný algoritmus je implementován např. ve statistickém softwaru R, který v této práci později popíšeme.

Příklad 3: Vypočítáme determinanty (dle [1], str. 281)

1. $\begin{vmatrix} 6 & 9 \\ 8 & 12 \end{vmatrix} = 72 - 72 = 0$

2. $\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = (0 + 1 + 1) = 2$

3. $\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 5 \\ 7 & 8 & 9 \end{vmatrix} = (45 + 96 + 84) - (105 + 72 + 48) = 225 - 225 = 0$

Příklad 4: Vypočítáme determinant det (2**A**-**B**) pro matice

$\mathbf{A}= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 5 \\ 7 & 8 & 9 \end{pmatrix}$ a $\mathbf{B} = \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{pmatrix}$. Potom postupně obdržíme

$2\mathbf{A}= \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}$, $2\mathbf{A}\text{-}\mathbf{B}= \begin{pmatrix} -7 & -4 & -1 \\ 2 & 5 & 8 \\ 11 & 14 & 17 \end{pmatrix}$, a konečně

$\begin{vmatrix} -7 & -4 & -1 \\ 2 & 5 & 8 \\ 11 & 14 & 17 \end{vmatrix} = (-955 - 352 - 28) - (-55 - 616 - 136) - 975 + 807 = -168.$

1.3 OPERACE S MATICEMI

Definice 5: (dle [2], str. 30) *Součtem* dvou matic

$$\mathbf{A}=\begin{pmatrix} a_{11}, & a_{12}, & \dots & a_{1n} \\ a_{21}, & a_{22}, & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}, & a_{m2}, & \dots & a_{mn} \end{pmatrix} \mathbf{B}=\begin{pmatrix} b_{11}, & b_{12}, & \dots & b_{1n} \\ b_{21}, & b_{22}, & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1}, & b_{m2}, & \dots & b_{mn} \end{pmatrix}$$

téhož typu $(m,\ n)$ rozumíme matici

$$\mathbf{A}+\mathbf{B}=\begin{pmatrix} a_{11}+b_{11}, & a_{12}+b_{12}, & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21}, & a_{22}+b_{22}, & \dots & a_{2n}+b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}+b_{m1}, & a_{m2}+b_{m2}, & \dots & a_{mn}+b_{mn} \end{pmatrix}.$$

Příklad 5: Mějme dány matice **A** a **B**, kde

$\mathbf{A}=\begin{pmatrix} 3 & 2 & 4 \\ 7 & -2 & -4 \\ 1 & 5 & -1 \end{pmatrix}$ $\mathbf{B}=\begin{pmatrix} -1 & 3 & 0 \\ -5 & 6 & 5 \\ 3 & -8 & 9 \end{pmatrix}$. Pak $\mathbf{A}+\mathbf{B}=\begin{pmatrix} 2 & 5 & 4 \\ 2 & 5 & 1 \\ 4 & -3 & 8 \end{pmatrix}$.

Poznámka 4: Sčítání matic má výhodné algebraické vlastnosti. Konkrétně je asociativní a komutativní. Takže pro matice **A**, **B** a **C** příslušných rozměrů platí: (**A**+**B**)+**C**=**A**+(**B**+**C**) a **A+B=B+A**, což lze jednoduše ověřit i pro matice z Příkladu 5.

Definice 6: (dle [2], str. 30) Je-li $r \in R$ libovolné reálné číslo, pak *r-násobkem matice* **A** rozumíme matici

$$r\mathbf{A}=\begin{pmatrix} ra_{11}, & ra_{12}, & \dots & ra_{1n} \\ ra_{21}, & ra_{22}, & \dots & ra_{2n} \\ \dots & \dots & \dots & \dots \\ ra_{m1}, & ra_{m2}, & \dots & ra_{mn} \end{pmatrix}.$$

Příklad 6: Mějme dánu matici **A** a reálné číslo r,

$$r{=}2,\ \mathbf{A}=\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}.\ \text{Potom } r\mathbf{A}=\begin{pmatrix} 2 & 4 \\ 0 & 6 \end{pmatrix}.$$

Příklad 7: Z Definice 5 a 6 plyne speciálně odčítání matic. Rozdíl matic **A**, **B**, který pak označujeme **A-B**, potom definujeme jako **A**+(-**B**); poznamenejme, že na rozdíl od sčítání, odčítání není komutativní.

Nechť jsou tedy zadány matice **A** a **B**,

$$\mathbf{A}=\begin{pmatrix} 4 & 6 \\ 1 & 8 \end{pmatrix} \mathbf{B}=\begin{pmatrix} 4 & 3 \\ -2 & 5 \end{pmatrix}.$$

Pak $\mathbf{A}\text{-}\mathbf{B}=\begin{pmatrix} 0 & 3 \\ 3 & 3 \end{pmatrix}$. Ovšem $\mathbf{B}\text{-}\mathbf{A}=\begin{pmatrix} 0 & -3 \\ -3 & -3 \end{pmatrix}$.

Příklad 8: (dle [10], str. 70) Vypočítáme výraz 2**A**+**B**-3**C** pro

$$\mathbf{A}=\begin{pmatrix} 2 & 3 & -1 & 5 \\ 8 & -4 & 0 & -7 \end{pmatrix}, \mathbf{B}=\begin{pmatrix} -5 & 2 & 1 & -6 \\ 1 & 10 & 2 & -3 \end{pmatrix}, \mathbf{C}=\begin{pmatrix} -4 & 2 & 0 & -1 \\ 3 & 8 & -6 & -4 \end{pmatrix}.$$

Postupným roznásobením dostaneme

$$2\mathbf{A}=\begin{pmatrix} 4 & 6 & -2 & 10 \\ 16 & -8 & 0 & -14 \end{pmatrix}, 3\mathbf{C}=\begin{pmatrix} -12 & 6 & 0 & -3 \\ 9 & 24 & -18 & -12 \end{pmatrix},$$

sečteme $2\mathbf{A}+\mathbf{B}=\begin{pmatrix} -1 & 8 & -1 & 4 \\ 17 & -18 & 2 & -17 \end{pmatrix}$ a nakonec obdržíme

$$2\mathbf{A}+\mathbf{B}\text{-}3\mathbf{C}=\begin{pmatrix} 11 & 2 & -1 & 7 \\ 8 & -42 & 20 & -5 \end{pmatrix}.$$

Definice 7: Nechť je matice **A** typu $m \times n$, kde $\mathbf{A}=(a_{ij})$ a $\mathbf{B}(b_{jk})$ je matice typu $n \times p$, $m, n, p \in R$, pak *součinem matic* **A** a **B** nazveme matici **C**, jejíž *ik-tý* prvek dostaneme součinem prvků *i-tého* řádku matice **A** a *k-tého* sloupce matice **B**.
A to tak, že

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}, i = 1, \ldots, m; j = 1, \ldots, n; k = 1, \ldots, p. \tag{1}$$

Poznámka 5: Vidíme tedy, že matice **A** a **B** můžeme násobit jenom tehdy, je-li počet sloupců matice **A** stejný jako počet řádků matice **B**.

Příklad 9: Nechť jsou dány matice **A** a **B**;

$\mathbf{A}=\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, $\mathbf{B}=\begin{pmatrix} 3 & 4 \\ 3 & 4 \end{pmatrix}$. Potom $\mathbf{AB}=\begin{pmatrix} 9 & 12 \\ 9 & 12 \end{pmatrix}$. Pro

$\mathbf{A}=\begin{pmatrix} 0 & -6 \\ 5 & 0 \\ 9 & 1 \end{pmatrix}$, $\mathbf{B}=\begin{pmatrix} -4 & -5 & 0 \\ 3 & 7 & 1 \end{pmatrix}$ obdržíme $\mathbf{AB}=\begin{pmatrix} -18 & -42 & -6 \\ -20 & -25 & 0 \\ -33 & -38 & 1 \end{pmatrix}$.

Poznámka 6: Násobení matic je asociativní, tzn. platí vztah $(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$, dále je násobení matic distributivní vzhledem ke sčítání, tj. máme-li dány matice **A**, **B** a **C** vhodných rozměrů, pak platí vztah $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$. Násobení matic ovšem není komutativní! A to ani v případě čtvercových matic, jak si ukážeme v následujícím příkladu.

Příklad 10: (dle [7], str. 159)

1. $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
2. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}=\begin{pmatrix} 19 & 20 \\ 43 & 47 \end{pmatrix}$, $\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}=\begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$

Definice 8: *n-rozměrným vektorem* rozumíme matici typu $n \times 1$. Potom hovoříme o tzv. *sloupcovém vektoru.*

Poznámka 7: U vektorů se neužívá spojení typ vektoru, ale hovoříme o rozměru vektoru. Pro označení vektorů se zpravidla užívají malá tučná písmena abecedy.

Příklad 11:

$\mathbf{a}=\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ je trojrozměrný (sloupcový) vektor.

Poznámka 8: Můžeme násobit i matici vektorem, ale pouze, když jde o matici typu m x n a o n-rozměrný vektor. Jak ihned plyne z Definice 7. Dále ještě poznamenejme, že budeme všechny vektory uvažovat jako sloupcové

Příklad 12: Vynásobíme matici **A** vektorem **a**:

$$\mathbf{A}=\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \mathbf{a}=\begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}. \text{ Pak } \mathbf{Aa}=\begin{pmatrix} 7 \\ 2 \\ 4 \end{pmatrix}.$$

Definice 9: (dle [7], str. 161) Nechť $\mathbf{A}=(a_{ij})$ je matice typu $(m,\ n)$, pak matice $\mathbf{A}'=(a_{ji})$ typu $(n,\ m)$ se nazývá *transponovaná matice* k matici **A**.

Poznámka 9: Transponovaná matice $\mathbf{A}'$ k matici **A** vznikne z matice **A** záměnou řádků za sloupce, jde tedy (v případě čtvercové matice) o překlopení prvků kolem hlavní diagonály:

$$\mathbf{A}=\begin{pmatrix} a_{11}, & a_{12}, & \dots & a_{1n} \\ a_{21}, & a_{22}, & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}, & a_{m2}, & \dots & a_{mn} \end{pmatrix}, \mathbf{A}'=\begin{pmatrix} a_{11}, & a_{21}, & \dots & a_{m1} \\ a_{12}, & a_{22}, & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n}, & a_{2n}, & \dots & a_{nm} \end{pmatrix}.$$

Příklad 13: Vytvořte transponovanou matici k matici **A** pro:

1. $\mathbf{A}=\begin{pmatrix} 1 & -7 & 8 \\ 4 & 1 & 2 \\ -5 & 3 & 1 \end{pmatrix}$. Pak transponovaná matice má tvar
   $\mathbf{A}'=\begin{pmatrix} 1 & 4 & -5 \\ -7 & 1 & 3 \\ 8 & 2 & 1 \end{pmatrix}$.

2. $\mathbf{A}=\begin{pmatrix} 1 & -6 & 9 & -2 \\ -10 & 13 & 0 & 5 \\ 5 & 7 & 9 & 3 \end{pmatrix}$. Po transpozici obdržíme
   $\mathbf{A}'=\begin{pmatrix} 1 & -10 & 5 \\ -6 & 13 & 7 \\ 9 & 0 & 9 \\ -2 & 5 & 3 \end{pmatrix}$.

3. Jestliže máme dán sloupcový vektor $\mathbf{a}=\begin{pmatrix} \sqrt{3} \\ 7 \\ 6 \end{pmatrix}$, pak jeho transponováním na $\mathbf{a}'=\left(\sqrt{3},\ 7,\ 6\right)$ získáme vektor řádkový.

Věta 1: Pro operaci transponování platí

1. $(\mathbf{A}')'$ pro každou matici **A**;
2. $(\mathbf{AB})'=\mathbf{B}'\mathbf{A}'$, lze-li matice **A** a **B** násobit.

Důkaz: Důkaz druhého tvrzení lze snadno provést užitím matematické indukce (viz [4]). □

Definice 10: (dle [7], str. 167) Čtvercová matice $\mathbf{A}$ se nazývá *singulární*, jestliže det $\mathbf{A} = 0$.

Definice 11: (dle [7], str. 165) Čtvercová matice $\mathbf{A}$ se nazývá *regulární*, jestliže k ní existuje matice $\mathbf{B}$ tak, že

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

Matice $\mathbf{B}$ se nazývá *inversní* matice k matici $\mathbf{A}$ a značí se $\mathbf{A}^{-1}$.

Důkazy následujících vět jsou uvedeny v [5]:

Věta 2: Inverzní matice $\mathbf{A}^{-1}$ je regulární maticí $\mathbf{A}$ jednoznačně určena a platí $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$.

Věta 3: Čtvercová matice $\mathbf{A}$ je regulární právě tehdy, když je det$\mathbf{A}$ $\neq$0.

Poznámka 10: Jinak řečeno,

$$\forall \mathbf{A}, \det\mathbf{A} \neq 0\, \exists! \mathbf{A}^{-1} : \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Věta 4: Pro inverzní matici k součinu matic $\mathbf{A}$, $\mathbf{B}$ platí vztah

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Poznámka 11: Výpočet inverzní matice provedeme tzv. Jordanovou eliminační metodou, kde sepíšeme danou matici v soustavě s jednotkovou maticí jako pravou stranou soustavy $(\mathbf{A}|\mathbf{I})$ a výsledek je pak tvořen soustavou $\left(\mathbf{I}|\mathbf{A}^{-1}\right)$. Místo podrobného popisu algoritmu si jeho použití ukážeme rovnou na příkladu.

Příklad 14: Ověř existenci a urči matici $\mathbf{A}^{-1}$ k matici $\mathbf{A}$, jestliže $\mathbf{A} = \begin{pmatrix} 2 & -1 & 3 \\ 1 & -1 & 1 \\ -3 & 2 & 1 \end{pmatrix}$. Nejprve si vypočteme det $\mathbf{A}$.

$$\det\mathbf{A} = \begin{vmatrix} 2 & -1 & 3 \\ 1 & -1 & 1 \\ -3 & 2 & 1 \end{vmatrix} =$$

$$= [(2(-1)1 + 1(2)3 + (-3)(-2)(-1))] - [(3(-1)(-3) + 2(-2)2 + +1(-1)1] = -2$$

Ten je různý od nuly, čili inverzní matice $\mathbf{A}^{-1}$ k matici $\mathbf{A}$ dle Věty 4 existuje. Nyní tedy sestavíme soustavu matice $\mathbf{A}$ s jednotkovou maticí $\mathbf{I}$ jako pravou stranou soustavy tj.

$\left(\begin{array}{rrr|rrr} 2 & -1 & 3 & 1 & 0 & 0 \\ 1 & -1 & 7 & 0 & 1 & 0 \\ -3 & 2 & 1 & 0 & 0 & 1 \end{array}\right)$. Pak postupnou eliminací prvků matice $\mathbf{A}$

(Gaussovou eliminační metodou) získáme tzv. trojúhelníkový tvar matice

$\left(\begin{array}{rrr|rrr} 2 & -1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 7 & 1 & -2 & 0 \\ 0 & 1 & 11 & 3 & 0 & 2 \end{array}\right) \approx \left(\begin{array}{rrr|rrr} 2 & -1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 7 & 1 & -2 & 0 \\ 0 & 0 & 4 & 2 & 2 & 2 \end{array}\right)$. Pokračujeme dále v eliminaci, dokud nevytvoříme na levé straně soustavy jednotkovou matici,

$$\left(\begin{array}{ccc|ccc} 2 & -1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 7 & 1 & -2 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right) \approx \left(\begin{array}{ccc|ccc} 2 & -1 & 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{5}{2} \\ 0 & 1 & 0 & -\frac{5}{2} & -\frac{11}{2} & -\frac{7}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right) \approx$$

$$\approx \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{3}{2} & -\frac{7}{2} & -\frac{5}{2} \\ 0 & 1 & 0 & -\frac{5}{2} & -\frac{11}{2} & -\frac{7}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{array}\right).$$

Tedy výsledná matice $\mathbf{A^{-1}}=\begin{pmatrix} -\frac{3}{2} & -\frac{7}{2} & -\frac{5}{2} \\ -\frac{5}{2} & -\frac{11}{2} & -\frac{7}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

Definice 12: (dle [7], str. 167) Čtvercová matice $\mathbf{A}$ řádu n se nazývá *symetrická*, jestliže $\mathbf{A}=\mathbf{A}'$, tj. pro prvky $(a_{ij})_{i=1,\dots,m}^{j=1,\dots,n}$ matice $\mathbf{A}$ platí, že $a_{ij}=a_{ji}$.

Příklad 15: Matice

$$\mathbf{A}=\begin{pmatrix} 1 & -2 & 3 & 5 \\ -2 & 0 & 8 & -4 \\ 3 & 8 & 5 & 9 \\ 5 & -4 & 9 & 3 \end{pmatrix}$$

je symetrická.

1.4 VLASTNÍ ČÍSLA A VLASTNÍ VEKTORY MATIC

Definice 13: Nechť je dána čtvercová matice $\mathbf{A}$ řádu n. Pak kořeny algebraické rovnice stupně n (proměnné λ),

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0,$$

nazýváme *vlastní čísla* matice $\mathbf{A}$, kde $\mathbf{I}$ je jednotková matice řádu n.

Příklad 16: Vypočítáme vlastní čísla matice $\mathbf{A}$,

$$\mathbf{A}=\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Nejprve vypočteme determinant

$$\det\left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} 1-\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right| = \left| \begin{matrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{matrix} \right] =$$
$$= -2\lambda + \lambda^2 = -2\lambda + \lambda^2 = 0$$

a obdržíme kvadratickou rovnici, kterou budeme řešit standartním způsobem:

tedy $\lambda(-2+\lambda) = 0, \lambda_1 = 0 \vee \lambda_2 = 2$, jsou vlastní čísla matice $\mathbf{A}$.

Definice 14: (dle [7], str. 168) Nechť $\mathbf{A}=(a_{ij})$ je čtvercová matice řádu n. Nenulový vektor $\mathbf{x}$ se nazývá *vlastní vektor matice* $\mathbf{A}$ příslušný vlastnímu číslu λ, jestliže platí

$$\mathbf{Ax} = \lambda\mathbf{x}$$

pro nějaké (reálné nebo komplexní) vlastní číslo λ.

Poznámka 12: Pro pozdější použití si povšimněme, že soustavu lineárních rovnic, např.

$$2x - 3y = -1,$$
$$x - y = -4,$$

můžeme také zapsat pomocí matice a vektorů jako
$\begin{pmatrix} 2 & -3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ -4 \end{pmatrix}.$

Takto tedy můžeme každou soustavu lineárních rovnic zapsat maticově.

Věta 5: Nechť λ_0 je vlastní číslo matice $\mathbf{A}$, které není násobné (tzv. jednoduchý kořen), potom jediný vlastní vektor matice $\mathbf{A}$, příslušný vlastnímu číslu λ_0 nalezneme řešením soustavy lineárních rovnic zapsané maticově: $(\mathbf{A} - \lambda_0 \mathbf{I})\,\mathbf{x} = \mathbf{0}$, kde $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ je nulový vektor.

Tato soustava je pro n=2 a $\mathbf{x}$=$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ve tvaru

$$(a_{11} - \lambda_0)\,x_1 + a_{12}x_2 = 0,$$
$$a_{21}x_1 + (a_{22} - \lambda_0)\,x_2 = 0.$$

Poznámka 13: Opravdu, rozepsáním

$$\begin{pmatrix} a_{11} - \lambda_0 & a_{12} \\ a_{21} & a_{22} - \lambda_0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

dostaneme

$$(a_{11} - \lambda_0)\,x_1 + a_{12}x_2 = 0,$$
$$a_{21}x_1 + (a_{22} - \lambda_0)\,x_2 = 0.$$

Poznámka 14: Získaný vektor (řešení soustavy rovnic) $\mathbf{x}$ má pak očekávanou vlastnost

$$\mathbf{A}\mathbf{x} = \lambda_0 \mathbf{x}.$$

Příklad 17: Vypočítáme vlastní čísla a vlastní vektory matice $\mathbf{A}$

a) $\mathbf{A}$=$\begin{pmatrix} 2 & -6 \\ -1 & 3 \end{pmatrix}$; potom $\mathbf{I}$=$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ a
$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 2-\lambda & -6 \\ -1 & 3-\lambda \end{vmatrix} = (2-\lambda)(3-\lambda) - 6 = \big(6 + \lambda^2 - 2\lambda - 3\lambda - 6\big| =$
$=\big(\lambda^2 - 5\lambda\big)\ \lambda^2 - 5\lambda = 0 \Rightarrow \lambda_1 = 0 \vee \lambda_2 = 5$. Například, pro $\lambda = 5$ odbržíme soustavu lineárních rovnic

$$(2-5)\,x_1 - 6x_2 = 0,$$
$$-x_1 + (3-5)\,x_2 = 0.$$

Takto získanou soustavu rovnic řešíme standartním postupem, přičemž kořeny x_1 a x_2 jsou

$$x_1 = -2x_2 \wedge x_2 = t, \text{ tedy}$$
$$x_1 = -2t \wedge x_2 = t, t \in R,$$

kde t je reálný parametr. Zvolíme např. $t = -1$ dostaneme vlastní vektor příslušný vlastnímu číslu λ=5, $\mathbf{x}$=$\begin{pmatrix} 2 \\ -1 \end{pmatrix}$. Pro kontrolu dosadíme do vzorce dle Poznámky 14, což ověříme tak, že porovnáme levou a pravou stranu rovnosti

$$\mathbf{Ax} = \lambda \mathbf{x}.$$

$$\mathrm{L} = \begin{pmatrix} 2 & -6 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 10 \\ -5 \end{pmatrix}, \mathrm{P} = 5 \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \text{ dohromady L=P}$$

Z toho vyplývá, že vlastní vektor matice $\mathbf{A}$ příslušný vlastnímu číslu rovnému 5 je (až na nenulový násobek) vskutku $\mathbf{x}$=$\begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

b)

$$\mathbf{A}=\begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}. \text{ Pak } |\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 2-\lambda & -2 \\ 1 & -2-\lambda \end{vmatrix} = \lambda^2 - 4 + 3 = \lambda^2 - 1$$

Položíme rovno nule a získáme $\lambda_1 = 1 \vee \lambda_2 = -1$. Například, pro λ=1, dostaneme soustavu

$$\begin{aligned} x_1 - 3x_2 &= 0 \\ x_1 - 3x_2 &= 0, \end{aligned}$$

s řešením a volbou parametru

$$\begin{aligned} x_1 &= 3x_2, \\ x_2 &= t \end{aligned}$$

tedy

$$\begin{aligned} x_1 &= 3t \\ x_2 &= t,\ t \in R. \end{aligned}$$

Například pro $t = 2$ získáme $\mathbf{x}$=$\begin{pmatrix} 6 \\ 2 \end{pmatrix}$.

Dosazením do $\mathbf{Ax}$=$\lambda\mathbf{x}$ ověřme

$$\begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix} = 1 \begin{pmatrix} 6 \\ 2 \end{pmatrix}$$

Až na násobek je tedy vlastní vektor matice $\mathbf{A}$ příslušný vlastnímu číslu $\lambda = 1$ roven $\mathbf{x} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}$.

2 KOMPOZIČNÍ DATA

V této kapitole se seznámíme se speciálním druhem mnohorozměrných dat (vektorů), jejichž výzkum v současné době prochází velkým rozvojem, s četnými aplikacemi od geologie až po archeologii a sociologii. Uvedeme si stručně jejich vlastnosti včetně speciální transformace, které nám později pro tato data umožní výpočet regresní přímky metodou tzv. ortogonální regrese.

Sloupcový vektor $\mathbf{x} = (x_1, \ldots, x_D)'$ nazveme D - složkovou kompozicí, jestliže všechny jeho složky jsou kladná reálná čísla nesoucí pouze relativní informaci, což znamená, že příslušná informace je osažena pouze v podílech mezi složkami kompozice. Jinak řečeno, jestliže c je kladné reálné číslo, pak kompozice $(x_1, \ldots, x_D)'$ a $(c\,x_1, \ldots, c\,x_D)'$ nám sdělují v podstatě shodnou informaci. Jednou z nejsnažších cest, jak následně pracovat s vektory kompozičních dat, je prezentovat je v jejich tzv. uzavřené formě, tj. jako kladné vektory, jejichž součet složek je roven konstantě k (tato konstanta je nejčastěji volena rovna 1 nebo 100 v případě interpretace složek kompozic jako procentuálních podílů nějakého celku). Dále v této práci budeme uvažovat kompozice se součtem složek 1, tedy

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

Převedeme-li kompozice (kompoziční data) s daným součtem složek do grafické podoby v D-rozměrném reálném prostoru, vytvoří v něm tato data (D-1) - rozměrnou podmnožinu v tomto prostoru. Tuto podmnožinu nazýváme *simplex* (značíme $\mathbf{S}^D$). Simplex je tedy množina všech kompozičních dat se součtem k, respektive 1.

Poznámka 15: V našem případě budeme nejčastěji uvažovat pouze trojrozměrný prostor, a tedy 3-rozměrná kompoziční data.

V případě trojrozměrných kompozičních dat můžeme totiž s výhodou využít jejich specifický charakter a zobrazit je v tzv. ternárním diagramu. Ternární diagram je graf podoby rovnostranného trojúhelníku, jehož vrcholy A, B, C mají následující souřadnice v prostoru:

$$\text{A } [1,0,0], \text{ B } [0,1,0], \text{ C } [0,0,1].$$

Vnitřní body tohoto trojúhelníku potom odpovídají jednotlivým kompozicím. Následně mějme dán vnitřní bod P se souřadnicemi $[p_a, p_b, p_c]$, kde p_a je vzdálenost bodu P od protější strany vrcholu A, p_b vzdálenost bodu P od protější strany vrcholu B a analogicky pro p_c. Potom tedy P odpovídá nějaké kompozici $\mathbf{x}$ se složkami p_a, p_b, p_c.

Jak plyne již z definice kompozičních dat, nelze s nimi bohužel pracovat jako s obyčejnými vektory. Proto byly zavedeny operace odpovídající jejich charakteru.

2.1 OPERACE NA SIMPLEXU

U kompozičních dat nelze užít klasické sčítání vektorů ani násobení vektoru reálným číslem, neboť výsledným vektorem by nemusela být opět kompozice.

Proto se pod souhrným názvem tzv. Aitchisonova geometrie na simplexu zavádí speciální operace, perturbace a mocninná transformace, které již tuto vlastnost mají (viz. [9]).

1. **perturbace:** Mějme dány kompozice $\mathbf{x}=(x_1,\ldots,x_D)'$ a $\mathbf{y}=(y_1,\ldots,y_D)'$. Pak perturbaci kompozic $\mathbf{x}$ a $\mathbf{y}$ vypočítáme pomocí vztahu

$$\mathbf{x}\oplus\mathbf{y}=\left(\frac{x_1y_1}{\sum_{i=1}^{D}x_iy_i},\ldots,\frac{x_Dy_D}{\sum_{i=1}^{D}x_iy_i}\right)'.$$

   Perturbace je analogií součtu dvou vektorů v reálném prostoru.

   Pro perturbaci kompozic odpovídajících rozměrů platí:

   (a) komutativní zákon:
$$\mathbf{x}\oplus\mathbf{y}=\mathbf{y}\oplus\mathbf{x}$$

   (b) asociativní zákon:
$$(\mathbf{x}\oplus\mathbf{y})\oplus\mathbf{z}=\mathbf{x}\oplus(\mathbf{y}\oplus\mathbf{z})$$

2. **mocninná transformace:** Mějme dánu kompozici $\mathbf{x}=(x_1,\ldots,x_D)'$ a reálné číslo $\alpha\in R$. Potom je mocninná transformace definována jako

$$\alpha\odot\mathbf{x}=\left(\frac{x_1^\alpha}{\sum_{i=1}^{D}x_i^\alpha},\ldots,\frac{x_D^\alpha}{\sum_{i=1}^{D}x_i^\alpha}\right)',\ \alpha\in\mathbf{R}.$$

Mocninná transformace je analogií násobku vektoru číslem.

Pro kompozice $\mathbf{x}$, $\mathbf{y}$ příslušného rozměru a reálná čísla α,β platí:

1. asociativní zákon:
$$\alpha\odot(\beta\odot\mathbf{x})=(\alpha\cdot\beta)\odot\mathbf{x}$$

2. distributivní zákon vzhledem ke sčítání:

$$\alpha\odot(\mathbf{x}\oplus\mathbf{y})=(\alpha\odot\mathbf{x})\oplus(\alpha\odot\mathbf{y})$$
$$(\alpha+\beta)\odot\mathbf{x}=(\alpha\odot\mathbf{x})\oplus(\beta\odot\mathbf{x}).$$

V případě trojsložkových kompozic potom obdržíme výsledek výše uvedených operací pro kompozice $\mathbf{x}=(x_1,x_2,x_3)'$ a $\mathbf{y}=(y_1,y_2,y_3)'$ ve tvaru:

$$\mathbf{x}\oplus\mathbf{y}=\left(\frac{x_1y_1}{x_1y_1+x_2y_2+x_3y_3},\frac{x_2y_2}{x_1y_1+x_2y_2+x_3y_3},\frac{x_3y_3}{x_1y_1+x_2y_2+x_3y_3}\right)',$$

$$\alpha\odot\mathbf{x}=\left(\frac{x_1^\alpha}{(x_1^\alpha+x_2^\alpha+x_3^\alpha)},\frac{x_2^\alpha}{(x_1^\alpha+x_2^\alpha+x_3^\alpha)},\frac{x_3^\alpha}{(x_1^\alpha+x_2^\alpha+x_3^\alpha)}\right)'$$

Příklad 18: Mějme dány kompozice $\mathbf{x}$=$(0,2;0,5;0,3)'$ a $\mathbf{y}$=$(0,15;0,45;0,4)'$. Zřejmě součet složek složek obou z nich je roven 1 a mohou tak reprezentovat např. procentuální zastoupení prvků v hornině. Pak perturbaci těchto kompozic určíme takto pomocí výše uvedeného vztahu:

$$\mathbf{x} \oplus \mathbf{y} \left(\frac{x_1y_1}{x_1y_1 + x_2y_2 + x_3y_3}, \frac{x_2y_2}{x_1y_1 + x_2y_2 + x_3y_3}, \frac{x_3y_3}{x_1y_1 + x_2y_2 + x_3y_3} \right)' =$$

$$= \left(\frac{0,2*0,15}{0,2*0,15+0,5*0,45+0,3*0,4}, \frac{0,5*0,45}{0,2*0,15+0,5*0,45+0,3*0,4}, \right.$$

$$\left. , \frac{0,3*0,4}{0,2*0,15+0,5*0,45+0,3*0,4} \right)' =$$

$$= \left(\frac{0,03}{0,375}, \frac{0,225}{0,375}, \frac{0,12}{0,375} \right)' = (0,08;0,6;0,32)'$$

Výsledná kompozice je tedy $\mathbf{x} \oplus \mathbf{y}$=$(0,08;0,6;0.32)'$. Lze snadno ověřit, že součet jejích složek je opět 1.

Příklad 19: Mějme dáno číslo $\alpha = 2$ a kompozici $\mathbf{x}$=$(0,35;0,23;0,42)'$. Následně určíme mocninou transformaci $\alpha \odot \mathbf{x}$ jako:

$$2 \odot \mathbf{x} = \left(\frac{x_1^2}{x_1^2 + x_2^2 + x_3^2}, \frac{x_2^2}{x_1^2 + x_2^2 + x_3^2}, \frac{x_3^2}{x_1^2 + x_2^2 + x_3^2} \right)' =$$

$$= \left(\frac{0,35^2}{0,35^2 + 0,23^2 + 0,42^2}, \frac{0,23^2}{0,35^2 + 0,23^2 + 0,42^2}, \frac{0,42^2}{0,35^2 + 0,23^2 + 0,42^2} \right)' =$$

$$= (0,348;0,150;0,502)'.$$

Výsledná kompozice je $\alpha \odot \mathbf{x}$= $(0,348;0,150;0,502)'$.

2.2 TRANSFORMACE DAT

Použití standartních statistických metod na zpracování kompozičních dat vede zpravidla k zavádějícím a nesmyslným výsledkům, protože kompozice se nacházejí na simplexu, zatímco většina statistických metod předpokládá (vícerozměrný) reálný prostor. Tento problém řeší skupina tzv. logratio transformací z $\mathbf{S}^D$ prostoru do (*D-1*) - rozměrného reálného prostoru. V našem případě se blíže budeme zabývat pouze tzv. isometric logratio transformací (zkráceně ilr).

Poznamenejme, že v této kapitole se přidržíme původního anglického názvosloví, protože jejich české ekvivalenty ještě nejsou ustáleny a jejich použití by mohlo vést k nedorozumněním.

Poznámka 16: Protože dále budeme pracovat jen s trojrozměrnými kompozicemi, jsou již tomuto uzpůsobeny obecné rovnice pro ilr transformaci.

2.2.1 ISOMETRIC LOGRATIO (ILR) TRANSFORMACE

Ilr transformace nám umožňuje pracovat s transformovanými kompozicemi jako s obyčejnými vektory v rovině a má i další výhodné vlastnosti, co se Aitchisonovy geometrie týče. Výsledkem pro libovolnou kompozici $\mathbf{x}$ z trojrozměrného simplexu ($\mathbf{S}^3$) bude rovinný vektor $\mathbf{z} \in \mathbf{R}^2$. Jedná se vlastně o transformaci z podmnožiny trojrozměrného prostoru (simplexu $\mathbf{S}^3$) do dvojrozměrného prostoru, tedy z prostoru do roviny.

Transformační vzorce jsou nejčastěji uváděny jako

$$\operatorname{ilr}(\mathbf{x}) = \mathbf{z} = (z_1, z_2)', z_1 = \frac{\sqrt{2}}{2} \ln \frac{x_1}{x_2}, z_2 = \frac{\sqrt{6}}{3} \ln \frac{\sqrt{x_1 x_2}}{x_3}.$$

Inverzní transformaci ilr, označenou ilr^{-1}, kterou výsledné rovinné vektory zobrazíme zpět do simplexu, získáme využitím následujících vztahů, které využívají hlubších teoretických poznatků o skupině logratio transformací. Takto tedy pro $\mathbf{z} \in \mathbf{R}^2$ obdržíme kompozici $\mathbf{x} = (x_1, x_2, x_3)'$, kde složky x_1, x_2, x_3 jsou dány vztahy

$$x_1 = \frac{\exp(y_1)}{\exp(y_1) + \exp(y_2) + \exp(y_3)},$$

$$x_2 = \frac{\exp(y_2)}{\exp(y_1) + \exp(y_2) + \exp(y_3)},$$

$$x_3 = \frac{\exp(y_3)}{\exp(y_1) + \exp(y_2) + \exp(y_3)},$$

do kterých dosadíme následující výrazy

$$\begin{aligned} y_1 &= \frac{z_1\sqrt{2}}{2} + \frac{\sqrt{6}}{6} z_2, \\ y_2 &= \frac{\sqrt{6}}{6} z_2 - \frac{\sqrt{2}}{2} z_1, \\ y_3 &= -\frac{\sqrt{6}}{3} z_2. \end{aligned}$$

Jak již bylo naznačeno, platí žřejmě pro kompozice $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^2$ a jejich transformace $\mathbf{z}_1 = \operatorname{ilr}(\mathbf{x}_1)$, $\mathbf{z}_2 = \operatorname{ilr}(\mathbf{x}_2)$ následující vztahy:

$$\begin{aligned} \operatorname{ilr}(\mathbf{x}_1 \oplus \mathbf{x}_2) &= \operatorname{ilr}(\mathbf{x}_1) + \operatorname{ilr}(\mathbf{x}_2) = \mathbf{z}_1 + \mathbf{z}_2, \\ \operatorname{ilr}(c \odot \mathbf{x}_2) &= c.\operatorname{ilr}(\mathbf{x}_1) = c \cdot \mathbf{z}_1. \end{aligned}$$

Při výpočtech s kompozičními daty lze tedy tato data nejprve transformovat ilr transformací, provést analýzu (např. statistické výpočty), výsledky tranformovat zpět a následně již provést jejich interpretaci přímo na simplexu.

Příklad 20: V tomto příkladu se nejprve pokusíme zadané vektory transformovat ilr transformací, určit jejich průměr a následně výsledek zobrazit zpět a interpretovat na simlexu. Mějme tedy dány kompozice $\mathbf{x}_1 = (0,27; 0,30; 0,43)'$, $\mathbf{x}_2 = (0,15; 0,54; 0,31)'$ a $\mathbf{x}_3 = (0,62; 0,26; 0,12)'$. Podle výše uvedených vztahů provedeme ilr transformaci pro všechny zadané vektory:

$\mathrm{ilr}\,\mathbf{x}_1 = \mathbf{z}_1 = (z_{11},\ z_{12})'$,

$$z_{11} = \frac{\sqrt{2}}{2}\ln\frac{x_{11}}{x_{12}} = \frac{\sqrt{2}}{2}\ln\frac{0,27}{0,3} = -0,075;$$

$$z_{12} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{x_{11}x_{12}}}{x_{13}} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{0,27 * 0,3}}{0,43} = -0,337.$$

$\mathrm{ilr}\,\mathbf{x}_2 = \mathbf{z}_2 = (z_{21},\ z_{22})'$,

$$z_{21} = \frac{\sqrt{2}}{2}\ln\frac{x_{21}}{x_{22}} = \frac{\sqrt{2}}{2}\ln\frac{0,15}{0,54} = -0,906;$$

$$z_{22} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{x_{21}x_{22}}}{x_{23}} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{0,15 * 0,54}}{0,31} = -0,070.$$

$\mathrm{ilr}\,\mathbf{x}_3 = \mathbf{z}_3 = (z_{31},\ z_{32})'$,

$$z_{31} = \frac{\sqrt{2}}{2}\ln\frac{x_{31}}{x_{32}} = \frac{\sqrt{2}}{2}\ln\frac{0,62}{0,26} = 0,615;$$

$$z_{32} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{x_{31}x_{32}}}{x_{33}} = \frac{\sqrt{6}}{3}\ln\frac{\sqrt{0,32 * 0,26}}{0,12} = 0,986.$$

Takto obdržíme transformovaná kompoziční data v dvojrozměrném prostoru jako vektory

$\mathbf{z}_1 = (-0,075;\ -0,337)'$ $\mathbf{z}_2 = (-0,906;\ -0,070)'$ $\mathbf{z}_3 = (0,615;\ 0,986)'$

Následně určíme průměr těchto transformovaných dat, a to zprůměrovaním příslušných složek. Tento vzniklý rovinný vektor označíme $\bar{\mathbf{z}}$, tedy

$$\bar{\mathbf{z}} = \frac{1}{3}\left(z_1 + z_2 + z_3\right);$$

$\bar{\mathbf{z}} = (-0,112;\ 0,193)$ a budeme jej dále transformovat zpět na simplex.

Užitím příslušných vztahů provedeme inverzní ilr transformaci

$\mathbf{c} = \mathrm{ilr}^{-1}\left(\bar{\mathbf{z}}\right)$:

$$y_1 = z_1\frac{\sqrt{2}}{2} + z_2\frac{\sqrt{6}}{6} = (-0,112)\,\frac{\sqrt{2}}{2} + 0,193\frac{\sqrt{6}}{6} = 0,004;$$

$$y_2 = z_2\frac{\sqrt{6}}{6} - z_1\frac{\sqrt{2}}{2} = 0,193\frac{\sqrt{6}}{6} + 0,112\frac{\sqrt{2}}{2} = 0,158;$$

$$y_3 = -z_2\frac{\sqrt{6}}{3} = (-0,193)\,\frac{\sqrt{6}}{3} = -0,157.$$

Výsledné hodnoty dosadíme do vztahů pro složky x_1, x_2, x_3 kompozice $\mathbf{c}$, které jsme uvedli v této kapitole.

$$x_1 = \frac{\exp(y_1)}{\exp(y_1) + \exp(y_2) + \exp(y_3)} = \frac{1,004}{3,03} = 0,33;$$
$$x_2 = \frac{\exp(y_2)}{\exp(y_1) + \exp(y_2) + \exp(y_3)} = \frac{1,17}{3,03} = 0,39;$$
$$x_3 = \frac{\exp(y_3)}{\exp(y_1) + \exp(y_2) + \exp(y_3)} = \frac{0,855}{3,03} = 0,28.$$

Obdržená kompozice $\mathbf{c}$ se nazývá centrum a odpovídá průměrné kompozici vzhledem k Aitchisonově geometrii

$$\mathbf{c} = (0,33;\ 0,39;\ 0,28)\,.$$

Poznamenejme, že vzhledem k vlastnostem ilr transformace je $\mathbf{c} = \frac{1}{3} \odot (x_1 \oplus x_2 \oplus x_3)$ a výpočet jsme tak v tomto případě mohli provést i přímo na simplexu. U většiny statistických metod (včetně dále uvedené metody) je ovšem transformace kompozic nutnou součástí analýzy.

Pro ověření numerické správnosti provedeme zpětnou transformaci i u rovinných vektorů $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$.

$\mathrm{ilr}^{-1}(\mathbf{z}_1) = (-0,075;\ -0,337)\,,$

$$y_{11} = z_{11}\frac{\sqrt{2}}{2} + z_{12}\frac{\sqrt{6}}{6} = (-0,075)\,\frac{\sqrt{2}}{2} - 0,337\frac{\sqrt{6}}{6} = -0,190;$$
$$y_{12} = z_{12}\frac{\sqrt{6}}{6} - z_{11}\frac{\sqrt{2}}{2} = (-0,337)\,\frac{\sqrt{6}}{6} + 0,075\frac{\sqrt{2}}{2} = -0,085;$$
$$y_{13} = -z_{12}\frac{\sqrt{6}}{3} = 0,337\frac{\sqrt{6}}{3} = 0,275.$$

$$x_{11} = \frac{\exp(y_{11})}{\exp(y_{11}) + \exp(y_{12}) + \exp(y_{13})} = \frac{0,827}{3,062} = 0,27;$$
$$x_{12} = \frac{\exp(y_{12})}{\exp(y_{11}) + \exp(y_{12}) + \exp(y_{13})} = \frac{0,427}{3,062} = 0,30;$$
$$x_{13} = \frac{\exp(y_{13})}{\exp(y_{11}) + \exp(y_{12}) + \exp(y_{13})} = \frac{1,317}{3,062} = 0,43;$$

$\mathrm{ilr}^{-1}(\mathbf{z}_2) = (-0,906;\ -0,07)\,,$

$$y_{21} = z_{21}\frac{\sqrt{2}}{2} + z_{22}\frac{\sqrt{6}}{6} = (-0,906)\,\frac{\sqrt{2}}{2} - 0,07\frac{\sqrt{6}}{6} = -0,670;$$

$$y_{22} = z_{22}\frac{\sqrt{6}}{6} - z_{21}\frac{\sqrt{2}}{2} = (-0,07)\,\frac{\sqrt{6}}{6} + 0,906\frac{\sqrt{2}}{2} = 0,612;$$
$$y_{23} = -z_{22}\frac{\sqrt{6}}{3} = (0,07)\,\frac{\sqrt{6}}{3} = 0,057.$$

$$x_{21} = \frac{\exp(y_{21})}{\exp(y_{21}) + \exp(y_{22}) + \exp(y_{23})} = \frac{0,612}{3,41} = 0,15;$$
$$x_{22} = \frac{\exp(y_{22})}{\exp(y_{21}) + \exp(y_{22}) + \exp(y_{23})} = \frac{1,844}{3,41} = 0,54;$$
$$x_{23} = \frac{\exp(y_{23})}{\exp(y_{21}) + \exp(y_{22}) + \exp(y_{23})} = \frac{1,059}{3,412} = 0,31.$$

$\mathrm{ilr}^{-1}(\mathbf{z}_3) = (0,615;\, 0,986)$,

$$y_{31} = z_{31}\frac{\sqrt{2}}{2} + z_{32}\frac{\sqrt{6}}{6} = 0,615\frac{\sqrt{2}}{2} + 0,986\frac{\sqrt{6}}{6} = 0,84;$$
$$y_{32} = z_{32}\frac{\sqrt{6}}{6} - z_{31}\frac{\sqrt{2}}{2} = 0,986\frac{\sqrt{6}}{6} - 0,0,615\frac{\sqrt{2}}{2} = -0,032;$$
$$y_{33} = -z_{32}\frac{\sqrt{6}}{3} = -0,986\frac{\sqrt{6}}{3} = -0,08.$$

$$x_{31} = \frac{\exp(y_{31})}{\exp(y_{31}) + \exp(y_{32}) + \exp(y_{33})} = \frac{2,316}{3,734} = 0,62;$$
$$x_{32} = \frac{\exp(y_{32})}{\exp(y_{31}) + \exp(y_{32}) + \exp(y_{33})} = \frac{0,969}{3,734} = 0,26;$$
$$x_{33} = \frac{\exp(y_{33})}{\exp(y_{31}) + \exp(y_{32}) + \exp(y_{33})} = \frac{0,449}{3,734} = 0,12.$$

Obdržením původních kompozic $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ je ověření úspěšně dokončené.

3 ORTOGONÁLNÍ REGRESE A KOMPOZIČNÍ DATA

Při řešení praktických úloh v rovině se často dostáváme do situace, kdy chceme danými daty proložit přímku (či obecně jinou křivku), která tato data co nejlépe charakterizuje. Jinak řečeno, odhadnout neznámé parametry přímky, kterou proložíme data a to ve smyslu, který je v dané situaci nejvhodnější. Pro odhad neznámých parametrů lze užít například tzv. metodu nejmenších čtverců (viz [4]), která ovšem není vhodná v situacích, kdy jsou rovinná data výsledkem měření s možností chyb v obou složkách. V takové situaci je dobré užít tzv. ortogonální regresi (metodu totálních čtverců). Při určování přímky metodou totálních čtverců se snažíme minimalizovat součet druhých mocnin (neboli čtverců) vzdáleností rovinných dat od odhadované přímky. Při výpočtech významně využijeme vlastností maticové algebry, kterou jsme popsali v první kapitole.

V případě trojrozměrných kompozičních dat uvedená situace vzniká, když chceme tato data na simplexu proložit křivkou, která je co nejlépe charakterizuje. Jestliže totiž transformujeme trojrozměrná data pomocí ilr transformace, obdržíme právě dvojrozměrná rovinná data.
Ilr transformací n trojsložkových kompozic $\mathbf{x}_1, \ldots, \mathbf{x}_n$ získáme datovou matici $(\mathbf{a}_1, \mathbf{a}_2)$ tvořenou sloupcovými vektory $\mathbf{a}_1, \mathbf{a}_2$, v jejichž řádcích budou transformované kompozice $\mathbf{z}_1, \ldots, \mathbf{z}_n$ a označíme matici $\mathbf{X} = (\mathbf{1}_n, \mathbf{a}_1)$, kde $\mathbf{1}_n$ je n-rozměrný vektor čísel 1. Pro tuto matici rovněž předpokládáme, že počet řádků je větší nebo roven počtu sloupců této matice. V případě trojrozměrných kompozičních dat pak tedy musí platit $n \geq 2$.
Následně zvolíme matici $\mathbf{A}= (\mathbf{X}, \mathbf{a}_2)$, pro kterou platí, že počet řádků je větší než počet sloupců této matice. Potom vztah pro výpočet odhadů neznámých parametrů $\hat{\boldsymbol{\beta}}$ metodou ortogonální regrese je dána jako

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} - \lambda_3 \mathbf{I})^{-1}\, \mathbf{X}'\mathbf{a}_2,$$

kde λ_3 je nejmenší vlastní číslo (symetrické) matice $\mathbf{A}'\mathbf{A}$ (viz. [6]).

Příklad 21: Mějme dány kompozice dle příkladu 20. Tzn.
$\mathbf{x}_1 = (0,27; 0,30; 0,43)'$, $\mathbf{x}_2 = (0,15; 0,54; 0,31)'$ a $\mathbf{x}_3 = (0,62; 0,26; 0,12)'$.

Poznámka 17: Pro určení vlastních čísel, které je v následujícím příkladu nutné, byl užit statstického softwaru R, jímž se podrobněji zabývá následující kapitola.

Transformací těchto kompozic získáme data v dvojrozměrném prostoru se souřadnicemi

$$\mathbf{z}_1 = (-0,075; -0,337)\ \mathbf{z}_2 = (-0,906; -0,070)\ \mathbf{z}_3 = (0,615; 0,986)$$

Nejprve vytvoříme datovou matici $(\mathbf{a}_1, \mathbf{a}_2) = \begin{pmatrix} -0,075 & -0,337 \\ -0,906 & -0,070 \\ 0,615 & 0,986 \end{pmatrix}$.
Dále vytvoříme matici $\mathbf{X}$,

$$\mathbf{X} = \begin{pmatrix} 1 & -0,075 \\ 1 & -0,906 \\ 1 & 0,615 \end{pmatrix},$$

a matici $\mathbf{A}$, pro kterou následně vytvoříme $\mathbf{A}'\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} 1 & -0,075 & -0,337 \\ 1 & -0,906 & -0,070 \\ 1 & 0,615 & 0,986 \end{pmatrix}, \mathbf{A}'\mathbf{A} = \begin{pmatrix} 3 & -0,336 & 0,579 \\ -0,336 & 1,205 & 0,695 \\ 0,579 & 0,695 & 1,091 \end{pmatrix}$$

Pomocí softwaru R určíme vlastní čísla matice $\mathbf{A}'\mathbf{A}$. Tato čísla jsou

$$\lambda_1 = 3,178 \ \lambda_2 = 1,835 \ \lambda_3 = 0,283$$

přičemž pro naši další práci je důležité pouze vlastní číslo λ_3.

Pak dosazením do vztahu pro odhad neznámých parametrů $\hat{\boldsymbol{\beta}}$ získáme

$$\hat{\boldsymbol{\beta}} = \left[\begin{pmatrix} 1 & 1 & 1 \\ -0,075 & -0,906 & 0,615 \end{pmatrix} \begin{pmatrix} 1 & -0,075 \\ 1 & -0,906 \\ 1 & 0,615 \end{pmatrix} - 0,283 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]^{-1} *$$

$$* \begin{pmatrix} 1 & 1 & 1 \\ -0,075 & -0,906 & 0,615 \end{pmatrix} \begin{pmatrix} -0,337 \\ -0,070 \\ 0,986 \end{pmatrix} = \begin{pmatrix} 0,332 \\ 0,886 \end{pmatrix}$$

tedy odhady neznámých parametrů jsou

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0,332 \\ 0,886 \end{pmatrix}.$$

Přitom $\hat{\beta}_1$ je odhad absolutního členu regresní přímky a $\hat{\beta}_2$ parametru u lineárního členu této přímky. Následně jsme schopni zapsat obecnou rovnici této přímky. Tato obecná rovnice je $y - 0,886x - 0,332 = 0$, kde x a y jsou souřadnice v rovině.

Určíme rovněž aproximace rovinných bodů $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ pomocí regresní přímky, kterou jsme získali metodou ortogonální regrese, jejich kolmou projekcí na uvedenou přímku.

Určit tyto projekce znamená řešit vždy soustavu dvou lineárních rovnic o dvou neznámých, kde jednou z těchto rovnic je obecná rovnice regresní přímky a druhou je obecná rovnice kolmice procházející daným bodem. V tomto případě pro transformovanou kompozici $\mathbf{z}_1$ to znamená řešit soustavu

$$\begin{gathered} -0,886x + y - 0,332 = 0, \\ x + 0,886y - 0,030 = 0, \end{gathered}$$

jejíž výsledkem je bod $\hat{\mathbf{z}}_1$ o souřadnicích $[-0,148; 0,201]$.

Pro transformovanou kompozici $\mathbf{z}_2$ má druhá z rovnic tvar $x + 0,886y - 0,733$. Výsledkem aproximace je bod $\hat{\mathbf{z}}_2$ se souřadnicemi $[0,246; 0,550]$.

Pro $\mathbf{z}_3$ je obecná rovnice kolmice ve tvaru $x + 0,886y + 0,441$. Jako aproximaci získáme bod $\hat{\mathbf{z}}_3$ se souřadnicemi $[-0,337; 0,033]$. Takto obdržené aproximace

mohou být po transformaci zpět na simplex základem pro další analýzu.

Spočtěme ještě nakonec vzdálenosti transformovaných kompozic $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ od svých aproximací.

Např. pro $\hat{\mathbf{z}}_1\,[-0,148; 0,201]$

$\hat{\mathbf{x}}_1 = \mathrm{ilr}^{-1}\,(\mathbf{z}_1)$:

$$y_{11} = z_1\frac{\sqrt{2}}{2} + z_2\frac{\sqrt{6}}{6} = (-0,148)\,\frac{\sqrt{2}}{2} + 0,201\frac{\sqrt{6}}{6} = -0,023;$$

$$y_{12} = z_2\frac{\sqrt{6}}{6} - z_1\frac{\sqrt{2}}{2} = 0,201\frac{\sqrt{6}}{6} + 0,148\frac{\sqrt{2}}{2} = 0,187;$$

$$y_{13} = -z_2\frac{\sqrt{6}}{3} = -\,(0,201)\,\frac{\sqrt{6}}{3} = -0,164;$$

.

$$x_{11} = \frac{\exp(y_1)}{\exp(y_1)+\exp(y_2)+\exp(y_3)} = \frac{0,977}{3,032} = 0,322;$$

$$x_{12} = \frac{\exp(y_2)}{\exp(y_1)+\exp(y_2)+\exp(y_3)} = \frac{1,206}{3,032} = 0,398;$$

$$x_{13} = \frac{\exp(y_3)}{\exp(y_1)+\exp(y_2)+\exp(y_3)} = \frac{0,848}{3,032} = 0,280;$$

Výsledná aproximace $\hat{\mathbf{x}}_1$ kompozice $\mathbf{x}_1$ má složky $[0,322; 0,398; 0,280]$. U dalších aproximovaných kompozic postupujeme analogicky, přičemž získáme složky $[0,469; 0,331; 0,200]$ a $[0,261; 0,421; 0,318]$.
Vzdálenost bodu od regresní přímky je dána obecným vztahem

$$|Ap| = \frac{|a\,p_1 + b\,p_2 + c|}{\sqrt{a^2+b^2}},$$

kde bod P má souřadnice p_1, p_2 a přímka je dána obecnou rovnicí ve tvaru $ax + by + c = 0$. Po dosazení získáme pro vzdálenost $\mathbf{z}_1$ od své aproximace $\hat{\mathbf{z}}_1$ na regresní přímce

$$\frac{|(-0,148)\,(-0,886) + (0,201)\cdot 1 - 0,332|}{\sqrt{(-0,148)^2 + (0,201)^2}} = \frac{1,28\cdot 10^{-4}}{1,506} = 5,128\cdot 10^{-4}.$$

Výsledná vzdálenost bodu $\mathbf{z}_2$ od $\hat{\mathbf{z}}_2$ je pak 0,452. Vzdálenost $\mathbf{z}_3$ od regresní přímky je $1,24\cdot 10^{-3}$.

4 SOFTWARE R

Pro úspěšné řešení každého rozsáhlejšího numerického problému je zapotřebí užít odpovídající software. Vzhledem ke kladným zkušenostem z minulosti je v této práci opět použit statistický software R, přičemž byly vybrány a doplněny odpovídající příkazy z dřívějších prací viz. [3] a [4].

R provádí širokou škálu statistických (lineární a nelineární modelování, klasické statistické testy, analýzy,. . .) a grafických operací. Jeho nespornou výhodnou je, že se jedná o volně šiřitelný program (tzv. freeware). Mezi jeho další výhody patří podpora pod systémy Windows, Unixem, MacOS či všemi verzemi Linuxu; je rovněž lehce šiřitelný, přizpůsobivý. Získat jej můžeme na `www.r-project.org` .

Je na místě zmínit i jeho nevýhody, ovšem nutno podotknout, že jich není mnoho. Uveďme tedy, že dokáže jen pomalu, nebo vůbec pracovat s velkým objemem dat (v řádech milionů) a postrádá přátelštější uživatelské prostředí.
Nyní Vám předkládáme komentované příklady početních operací, které jsme posléze použili při řešení našeho numerického problému.

Větší množství příkazů lze získat na [8].

Zadání objektů

Jednotlivým objektům je možné přiřadit určitou hodnotu použitím symbolů `<` a `-` právě v tomto pořadí či symbolem `=`. Přičemž zadáním objektu zjistíme jeho hodnotu.

Poznámka 18: Jelikož tento program rozlišuje velká a malá písmena, musíme tedy rozlišovat např. `x` a `X`.

Příklad 22: Přiřadíme `a` hodnotu 14 a ověříme vypsáním

```
> a = 14
> a
[1] 14
```

Poznámka 19: V tomto programu se při součtu užívá stejná symbolika, a to `+`. U operací jako je sčítání, odčítání a násobení jsou znaménka stejná, ale při dělení se nepoužívá `:`, nýbrž `/`.

Příklad 23: Vypočítáme hodnotu součtu x+X pro hodnoty x = 5 a X = 1/2.

```
> x = 5
> X = 1/2
> x + X
[1] 5.5
```

Hodnota tohoto výrazu je tedy 5,5.

Příklad 24: Vypočítáme výraz a+A-b-B/C+c, pro hodnoty a = 2, A = 6, b = 1/7, B = 7, c = 5, C = 15.

```
> a = 2
> A = 6
> b = /7
> B = 7
> c = 5
```

$>$ `C` $=$ `15`
$>$ `a` $+$ `A` $-$ `b` $-$ `B/c` $+$ `C`
`[1] 21.45714`
Hodnota tohoto výrazu je přibližně 21,45714.

Poznámka 20: Již existující objekt lze smazat příkazem `ls (objekt, který má být smazán)`.

Matice a vektory

Matice i vektory v tomto programu jsou základními datovými strukturami.

Vektory

Vektor vytvoříme z čísel pomocí příkazu `b` $< -$c(`1`, `2`, `3`, `4`, `5`), kde `b` je označení vektoru a hodnoty v závorce jsou jednotlivé prvky vektoru.

Příklad 25: Vytvoříme vektor **a**, jehož prvky budou 1, 9, 4, 6

$>$ `a` $=$ `c`(`1`, `9`, `4`, `6`)
$>$ `a`
`[1] 1 9 4 6`

Matice

V R je možné matice tvořit dvěma základními způsoby:
- vytvoření matice pomocí sloupců příkazem `A=cbind(c,(1,10),c(77,2))`, kde **A** je označení matice, `c` je označení každého sloupce a v závorce jsou uvedené jednotlivé prvky sloupců matice **A**.

Příklad 26: Vytvoření matice **B** pomocí sloupců

$>$ `A` $=$ `rbind`(`c`(`1`, `2`, `3`, `4`, `5`), `c`(`5`, `6`, `7`, `8`, `9`), `c`(`9`, `8`, `7`, `6`, `5`), `c`(`4`, `1`, `2`, `5`, `6`), `c(7,4,1,2,5))`
$>$ `A`

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]
[1,]	1	2	3	4	5
[2,]	5	6	7	8	9
[3,]	9	8	7	6	5
[4,]	4	1	2	5	6
[5,]	7	4	1	2	5

- Matici je také možné vytvořit pomocí řádků příkazem
`B=rbind(c(5,7),c(6,8))`, kde `c` je označení jednotlivých řádků matice **B** a v závorce jsou jednotlivé prvky řádků matice **B**.

Příklad 27: Vytvoření matice **A** pomocí řádků

$>$ `B` $=$ `cbind`(`c`(`1`, `2`, `3`, `4`, `5`), `c`(`5`, `6`, `7`, `8`, `9`), `c`(`9`, `8`, `7`, `6`, `5`), `c`(`4`, `1`, `2`, `5`, `6`), `c(7,4,1,2,5))`
$>$ `B`

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]
[1,]	1	5	9	4	7
[2,]	2	6	8	1	4
[3,]	3	7	7	2	1
[4,]	4	8	6	5	2
[5,]	5	9	5	6	5

Lze též vypsat také pouze určitý řádek matice **A** příkazem `A[číslo řádku , ]` popřípadě je možné vypsat pouze určitý sloupec, a to příkazem `A[, číslo sloupce]`.

Poznámka 21: Matici můžeme také vytvořit pomocí příkazu `matrix`, kde pomocí parametrů `nrow` a `ncol` předdefinujeme počet řádků, respektive sloupců této matice.

Operace s maticemi

Jak jsme již uvedli v první kapitole, je možno s maticemi dále pracovat. Project R nabízí efektivní zjednodušení těchto často složitých algoritmů. Proto zde uvedeme formou komentovaných příkladů některé příkazy, jejichž užití jsme se této práci nevyhnuli.

Skalární násobek matice (tj. násobek matice reálným číslem) lze jednoduše provést pomocí `*` jako symbolu součinu.

Příklad 28: Příklad výpočtu skalárního násobku matice

$> \mathtt{A} = \mathtt{cbind}(\mathtt{c}(-2,4,9), \mathtt{c}(20,-1,4), \mathtt{c}(5,9,4))$
$> \mathtt{b} = 5$

$> \mathtt{Z} = \mathtt{b} * \mathtt{A}$
$> \mathtt{Z}$

	[,1]	[,2]	[,3]
[1,]	−10	100	25
[2,]	20	−5	45
[3,]	45	20	20

Příklad 29: Sečteme matici **B** s toutéž maticí
$> \mathtt{B}$

	[,1]	[,2]	[,3]	[,4]
[1,]	1	0	−1	2
[2,]	8	2	1	−1
[3,]	−4	5	2	3

$> \mathtt{Z} = \mathtt{B} + \mathtt{B}$
$> \mathtt{Z}$

	[,1]	[,2]	[,3]	[,4]
[1,]	2	0	−2	4
[2,]	16	4	2	−2
[3,]	−8	10	4	6

Tento program umožňuje snadné násobení matic (nesmíme opomenout fakt, že počet sloupců první matice se musí rovnat počtu řádků druhé matice; viz. předchozí kapitola). Pro tento příkaz je nutné použít % jako oddělovač znaménka početní operace.

Příklad 30: Vynásobíme matice **A** a **B**

$> \mathtt{A} = \mathtt{cbind}(\mathtt{c}(5,3), \mathtt{c}(9,5), \mathtt{c}(1,4))$
$> \mathtt{B} = \mathtt{cbind}(\mathtt{c}(2,5,3), \mathtt{c}(1,2,9))$

$> \mathtt{A}\% * \%\mathtt{B}$

	[,1]	[,2]
[1,]	58	32
[2,]	43	49

Transpozici matice určíme pomocí příkazu `t(označení matice)`

Příklad 31: Transponujeme matici **X**

`t(X)`

> `X = cbind(c(2,3,4), c(9,8,7), c(5,6,1))`

	[,1]	[,2]	[,3]
[1,]	2	9	5
[2,]	3	8	6
[3,]	4	7	1

`t(X)`

[1,]	2	3	4
[2,]	9	8	7
[3,]	5	6	1

Pro určení inverzní matice v softwaru R užíváme příkaz `solve(název matice)`

Příklad 32: Určíme inverzní matici k matici **W**

> `W = cbind(c(2,3,4), c(9,8,7), c(5,6,1))`

> `solve(W)`

	[,1]	[,2]	[,3]
[1,]	−0.5151515	0.3939394	0.21212121
[2,]	0.3181818	−0.2727273	0.04545455
[3,]	−0.1666667	0.3333333	−0.16666667

Příkaz `I=diag(w)` nám vytvoří jednotkovou matici řádu n, jestliže `w` je vektorem jedniček příslušného řádu.

Příklad 33: Vytvoříme jednotkovou matici řádu 4

```
> w = c(1,1,1,1)
> w
[1] 1 1 1 1
> I = diag(w)
> I
```

	[,1]	[,2]	[,3]	[,4]
[1,]	1	0	0	0
[2,]	0	1	0	0
[3,]	0	0	1	0
[4,]	0	0	0	1

Je potřeba na tomto místě podotknout, že uvedené operace lze dále zefektivnit. Pro příklady menšího rozsahu jsou ale výše zavedené zcela postačující.

Vlastní čísla a vlastní vektory matice

K výpočtu vlastních čísel (uspořádáných dle absolutní hodnoty od největšího po nejmenší) a vektorů matice slouží příkaz `eigen(A)`, kde **A** je označení čtvercové matice.

Příklad 34: Vypočítáme vlastní čísla a vlastní vektory matice z Příkladu 26:

$>$ `A` $=$ `rbind(c(1,2,3,4,5),c(5,6,7,8,9),c(9,8,7,6,5),c(4,1,2,5,6),`
`c(7,4,1,2,5))`
$>$ `A`

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	2	3	4	5
[2,]	5	6	7	8	9
[3,]	9	8	7	6	5
[4,]	4	1	2	5	6
[5,]	7	4	1	2	5

$>$ `eigen(A)`

`values`

`[1] 2.311548e+01 -4.464461e+00 3.522235e+00 1.826744e+00`
`-7.213179e-16`

`vectors`

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	−0.2557492	−0.4646280	−0.05281898	−0.04652443	−0.1703886
[2,]	−0.6122836	−0.4318125	0.23373735	0.14792358	0.4543695
[3,]	0.6355867	0.5466666	0.76920980	0.53264444	−0.6247580
[4,]	−0.2590424	−0.1968263	−0.58066630	−0.79396519	0.5679618
[5,]	−0.2977040	0.5099743	−0.11713216	0.24872329	−0.2271847

Poznámka 22: Protože program je v anglickém jazyku, musíme uvést, že `values` označuje vlastní čísla matice **A** a `vectors` příslušné sloupcové vlastní vektory téže matice.

Zadávání funkcí

V některých případech jsme nuceni pracovat se složitějšími funkcemi, které by při zadání až v příkazové řádce mohly znepřehlednit celkové řešení daného problému. V takovýchto případech je výhodné užít možnosti tuto funkci si předdefinovat. Příkladem takovéto funkce, zde reprezentující ilr transformaci kompozičních dat, je

```
ilr = function(w1,w2,w3) {
dat = cbind((1/sqrt(2)) * log(w1/w2),(2/sqrt(6)) * log(sqrt(w1 * w2)/w3))
return(dat)
}.
```

Nyní popíšeme jednotlivé části takovéto funkce.

- `ilr`... název funkce
- `w1,w2,w3`... parametry funkce
- `dat = cbind((1/sqrt(2)) * log(w1/w2),(2/sqrt(6))*`
  `*log(sqrt(w1 * w2)/w3))`... tělo funce
- `return(dat)`... vrácená hodnota (v tomto případě matice)

Dále pro nás bude výhodné načítat data z již vytvořených datových souborů s příponou .txt do matice. Pro tuto operaci, konkrétně pro načtení souboru do matice o třech sloupcích, slouží příkaz
`d=matrix(scan("název-souboru.txt"),ncol=3,byrow=T)`, kde parametr

`byrow=T` (T - true) nám umožňuje načítat jednotlivá data po řádcích. Jestliže chceme data načíst po sloupcích, užijeme příkaz `byrow=F` (F - false).

Příklad 35: Načteme datovou matici z textového dokumentu a vypíšeme ji:

```
> d = matrix(scan("databiofyzika.txt"), ncol = 3, byrow = T)
Read30items
> d
```

	[,1]	[,2]	[,3]
[1,]	1.051564	45.30266	53.645778
[2,]	1.419082	56.40097	42.179952
[3,]	2.028647	66.92659	31.044764
[4,]	2.386431	70.84039	26.773181
[5,]	2.925925	77.79628	19.277792
[6,]	3.311921	80.12847	16.559606
[7,]	3.742433	85.32624	10.931329
[8,]	4.141881	87.45318	8.404935
[9,]	4.510034	88.54524	6.944725
[10,]	5.235592	93.04998	1.714426

Výše popsané příkazy reprezentují pouze úvod do práce se statistickým softwarem R. Proto tento popis v komentáři ke zpracovanému reálnému problému, uvedenému v další kapitole, ještě dále rozšíříme.

5 APLIKACE V BIOFYZICE

Pro demonstraci uvedené teorie jsem se rozhodla si zvolit příklad z reálného prostředí, konkrétně jsem analyzovala experimentální biofyzikální data získaná z Katedry experimentální fyziky Univerzity Palackého v Olomouci. Jednotlivá data nám dávají informace o příspěvku tří různých komponent (složek) podílející se na celkové fluorescenci. Tato data byla obdržena ve formě počtů molekul přispívajících k emisi záření na určité vlnové délce, která nebyla blíže specifikována; přitom příspěvek jednotlivých komponent se v závislosti na vlnové délce liší. Vzhledem k podstatě problému, převedením těchto počtů na procentuální vyjádření příspěvků jednotlivých komponent k fluorescenci, neztratíme žádnou informaci, v datech obsaženou. Data v této podobě jsou uvedena jako prvky vypsané matice `d` v minulé kapitole. Úkolem bude zjistit, jak se budou měnit poměry mezi jednotlivými komponenty vyvíjet v závislosti na jejich relativních hodnotách.

Při další práci s těmito daty jsem použila sofware R pro vyčíslení jednotlivých parametrů ortogonální regrese, vlatních čísel datové matice popřípadě k určení transformací. Pro úplnost zde uvedeme použité příkazy, které následně okomentujeme:

```
d=matrix(scan("databiofyzika.txt"),ncol=3,byrow=T)

library(compositions)

Id=diag(c(1,1))

ilr=function(w1,w2,w3){
dat=cbind((1/sqrt(2))*log(w1/w2),(2/sqrt(6))*log(sqrt(w1*w2)/w3))
return(dat)
}

invilr=function(x,y){
y1=x/sqrt(2)+y/sqrt(6)
y2=-x/sqrt(2)+y/sqrt(6)
y3=-y*sqrt(2/3)
rat=apply(cbind(exp(y1),exp(y2),exp(y3)),1,sum)
dat=cbind(exp(y1)/rat, exp(y2)/rat, exp(y3)/rat)
return(dat)
}

d1=ilr(d[,1],d[,2],d[,3])

postscript("fig1.eps",paper="special",width=10,height=5,horizontal=FALSE)
plot(d1,type="n",xlim=c(-3,-1.5),ylim=c(-3,3))
text(d1,labels=1:nrow(d))
X=cbind(rep(1,10),d1[,1])
A=cbind(X,d1[,2])
lambda=eigen(t(A)%*%A)$values[3]
b = solve(t(X)% * %X − lambda * Id)% * %t(X)% * %d1[, 2]
abline(b[1],b[2],col=3)
Rc=cbind(1:20000,1:20000)
for(i in 1:20000){
Rc[i,1]=-10+i/1000
```

```
Rc[i,2]=b[1]+b[2]*(-10+i/1000)
}
Rc1=invilr(Rc[,1],Rc[,2])
title("biofyzika")
dev.off()

postscript("fig2.eps",paper="special",width=10,height=5,
horizontal=FALSE)

plot.acomp(d[1,],pch="1")
for (i in 2:9){
plot.acomp(d[i,],pch=48+i,add=T)
}
plot.acomp(d[10,]+c(d[10,1]+3,d[10,2]-3,d[10,3]),pch=49,add=T)
plot.acomp(d[10,]+c(d[10,1]-3,d[10,2]+3,d[10,3]),pch=48,add=T)

plot.acomp(Rc1,pch='.',col='red',add=TRUE)
title("biofyzika")
dev.off()
```

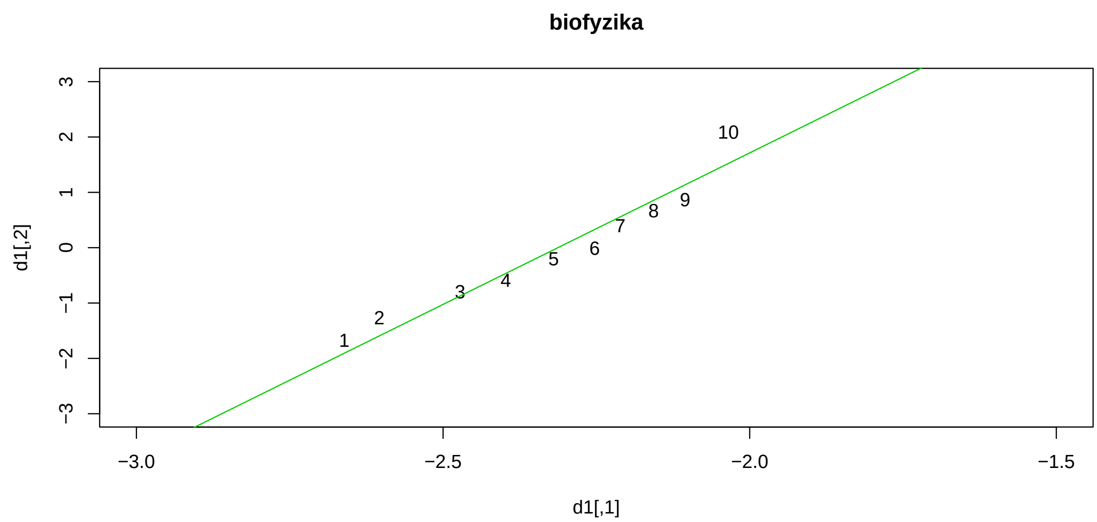
Nyní užité příkazy popíšeme.

- `d=matrix(scan("databiofyzika.txt"),ncol=3,byrow=T)`...načtení datové matice ze souboru databiofyzika.txt,
- `library(compositions)`...načtení příslušného balíku z knihovny softwaru,
- `Id=diag(c(1,1))`...vytvoření jednotkové matice řádu 2,

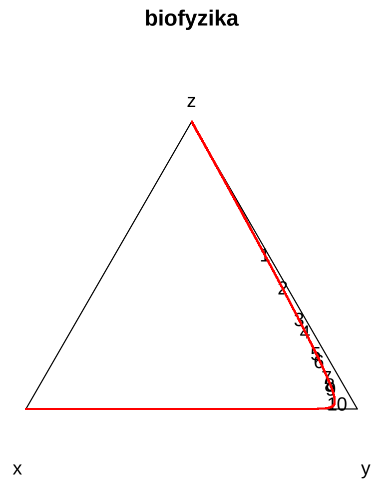
  ```
  ilr=function(w1,w2,w3){
  dat=cbind((1/sqrt(2))*log(w1/w2),(2/sqrt(6))*log(sqrt(w1*w2)/w3))
  return(dat)}
  ```
  ...zadání funkce pro ilr transformaci,
- ```
  invilr=function(x,y){
  y1=x/sqrt(2)+y/sqrt(6)
  y2=-x/sqrt(2)+y/sqrt(6)
  y3=-y*sqrt(2/3)
  rat=apply(cbind(exp(y1),exp(y2),exp(y3)),1,sum)
  dat=cbind(exp(y1)/rat, exp(y2)/rat, exp(y3)/rat)
  return(dat)}
  ```
  ...zadání funkce pro inverzní ilr transformaci,
- `postscript("fig1.eps",paper="special",width=10,height=5,horizontal=FALSE)` ...formát a parametry obrázku,
- `plot(d1)`...zobrazení souřadnic v rovině,
- `type="n"`...body grafu nebudou popsány,
- `xlim=c(-3,-1.5),ylim=c(-3,3))`...ohraničení rozsahu grafu,
- `(rep(1,10),d1[,1])`...matice, kde první sloupec je získán pomocí příkazu `rep`. První parametr označuje, jaké číslo se má opakovat, a druhé označuje, kolikrát se má toto opakování provést,
  ```

- `lambda=eigen(t(A)%*%A)$values[3]`...určení nejmenšího vlastního čísla matice $\mathbf{A}'\mathbf{A}$ řádu 3,
- $\texttt{b} = \texttt{solve(t(X)}\% * \%\texttt{X} - \texttt{lambda} * \texttt{Id)}\% * \%\texttt{t(X)}\% * \%\texttt{d1}[, 2]$...výpočet odhadů parametrů přímky pomocí ortogonální regrese,
- `abline(b[1],b[2],col=3)`...do již existujícího grafu bude zakreslena přímka s absolutním členem a parametrem u lineárního členu, jejichž odhady jsme určili ortogonální regresí
- ```
  Rc=cbind(1:20000,1:20000)
  for(i in 1:20000){
  Rc[i,1]=-10+i/1000
  Rc[i,2]=b[1]+b[2]*(-10+i/1000)
  }
  ```
  ...zápis hodnot x-ových a y-ových souřadnic přímky do matice, abychom ji po inverzní ilr transformaci mohli vykreslit na simplexu
- `Rc1=invilr(Rc[,1],Rc[,2])` ...vytvoří body na simplexu
- `title("biofyzika")`...název grafu
- `dev.off()`...uzavření obrázku
- `plot.acomp`...zobrazí ternární diagram
- `pch=48`...označí jednotlivé body od 1 do 10
- `add=T`...přidání dalších údajů do již existujícího grafu
- `plot.acomp(Rc1,pch='.',col='red',add=TRUE)`...zobrazení matice, transformovaných bodů přímky, získané metodou totálních čtverců v ternárním diagramu,

Následně uvedeme grafy pro ortogonální regresi u transformovaných kompozic,


  ```

a ternární diagram, zobrazující transformované výsledky na simplexu.



Již při pohledu na ternární diagram je zřejmé, že podíl první složky na celkové fluorescenci je víceméně stabilní, zatímco podíly ostatních dvou se budou mezi sebou měnit. Odhadnutá přímka potom vyjadřuje, jak bude vývoj závislosti probíhat. Můžeme pomocí ní následně například zjistit, jaké hodnoty ostatních složek můžeme očekávat při dané hodnotě jedné z nich. To můžeme následně využít jako podklad pro další analýzu.

ZÁVĚR

Hlavním předmětem mé soutěžní práce SOČ byla problematika regresní analýzy pro kompoziční data. Podařilo se mi přitom rozvinout postup, který ještě pro řešení této úlohy nebyl aplikován a zdá se být, při porovnání s jinými obdobnými procedurami, efektivní a přitom relativně jednoduchý pro použití. Navíc, jako při tvorbě prací minulých, i nyní jsem se snažila propojit matematickou teorii s jejím praktickým využitím a výsledkem je příklad z oblasti biofyziky. Ráda bych se k jednotlivým kapitolám práce vrátila ještě trochu podrobněji.

První kapitola se zabývala maticemi a determinanty. Především se zde nacházejí definice a vlastní názorné příklady, které lépe vysvětlují některé pro někoho na první pohled nejasné problémy. Se zkušenostmi z minulých let mi již její zpracování nedělalo větší problémy a naopak jsem přišla na mnohé další souvislosti.

Následující kapitola měla za úkol seznámit s kompozičními daty. Uvádím zde, jak s těmito daty pracovat popřípadě, jak tuto práci co nejvíce zjednodušit. Narozdíl od kapitoly předchozí byla pro mě tato problematika zcela nová a její nastudování a pochopení si vyžádalo nemalé úsilí.

Třetí kapitola se zabývá ortogonální regresí. Tuto statistickou metodu jsem vybrala z toho důvodu, že je výhodná pro situace, kdy výsledkem měření mohou být rovinná data s možností chyby v obou složkách. Byla to pro mě po metodě nejmenších čtverců již druhá zkušenost s regresní analýzou.

Ve čtvrté kapitole jsem pracovala s matematickým softwarem R, který mi velmi ulehčil práci (určil za mne vlastní číslo datové matice, odhady neznámých parametrů a následně jsem s jeho pomocí byla schopná vytvořit grafické zobrazení výsledků regresní analýzy). S tímto softwarem jsem pracovala již dříve, a proto bylo pro mne výhodné využít tyto praktické zkušenosti. Tento software je volně dostupný a je hojně využíván všude ve světě.

V poslední kapitole nazvané Aplikace v biofyzice jsem použila postupy popsané v předchozích kapitolách. Tato kapitola je praktickou částí a zároveň i určitým vrcholem této práce. Samotné užití teorie jsem přitom zaměřila na oblast, která je pro mne osobně zajímavá. Vzhledem k mému rozhodnutí tento obor studovat na vysoké škole bych chtěla i v budoucnu tuto problematiku dále rozvíjet.

ABSTRAKT

Kompoziční data jsou mnohorozměrná data, kde jediná relativní informace je obsažena pouze v podílech mezi jejich složkami. Jejich příkladem jsou procentuální podíly na určitém celku. Při statistické analýze těchto dat je většinou potřeba provést speciální transformaci, aby bylo následně možné aplikovat standartní metody pro jejich zpracování. Předpokládaná soutěžní práce se zabývá možností regresní analýzy trojsložkových kompozičních dat aplikací metody totálních čtverců (ortogonální regrese); této metody doposud pro řešení daného problému nebylo použito. Potřebné výpočty byly provedeny s využitím statistického softwaru R. Teorie je doplněna praktickým příkladem z oblasti biofyziky.

SEZNAM LITERATURY

[1] BICAN,L. *Lineární algebra v úlohách*. Praha: Státní pedagogické nakladatelství, 1979.

[2] BICAN, L. *Lineární algebra a geometrie*. Praha: Academia, 2000. ISBN 80 - 200 - 0843 - 8.

[3] GELETIČOVÁ, J. *Výběrové řízení pro nákup výpočetní techniky užitím matematických metod rozhodování*. Slovanské gymnázium Olomouc, 2006, 28 s.

[4] GELETIČOVÁ, J., PETERA, M. *Dvouepochový lineární model a jeho aplikace*. Slovanské gymnázium Olomouc, 2007, 25 s.

[5] HRON, K., FIŠEROVÁ, E., *Total Least Squares Solution for Compositional Data using Calibration Line*. Journal of Applical Statistics, odesláno.

[6] HORT, D., RACHŮNEK, J. *Algebra 1*. Olomouc: VUP, 2003. ISBN 80 - 244 - 0631 - 4.

[7] KOJECKÁ, J., ZÁVODNÝ, M., KOJECKÝ, T. *Úvod do matematiky pro geoinformatiky* . Olomouc: VUP, 2004.

[8] KULICH, M. *Dokumentace a manuály k R [online]*. 2005 [cit. 2005 - 17 - 2]. < http >: //www.karlin.mff.cuni.cz/ >.

[9] PAWLOWSKY-GLAHN, V. *Lecture Notes on Compositional Data Analysis[online]*. Universitat de Girona, 2004 [cit. 2008 - 6 - 4]. < http >: //diobma.udg.edu : 8080/dspace/bitstream/10256.1/297/1/ CoDa − book.pdf/ .

[10] RACHŮNEK, J. *Algebra a teoretická aritmetika*. Olomouc: Přírodovědecká fakulta UP 1988. ISBN 80 - 7290 - 90 - 190 - 7.

Obsah