

# Užití kompozičního biplotu při analýze medicínských dat

Alžběta Kalivodová

Katedra matematické analýzy a aplikací matematiky  
Univerzita Palackého v Olomouci

Laboratoř dědičných metabolických poruch  
Ústav molekulární a translační medicíny  
UPOL a FNOL

Robust 2012  
9. – 14. září 2012, Němčičky

# Obsah

## 1 Úvod

## 2 Konstrukce biplotu

## 3 Kompoziční data

- Clr transformace

## 4 Příklady

- Příklad - Úmrtí v důsledku nemoci
- Příklad - Metabolomická data

- Biplot je v současnosti hojně užívaný grafický nástroj mnohorozměrné statistické analýzy.
- Zobrazuje objekty a proměnné do jednoho grafu.
- Často se aplikuje také při statistické analýze speciálních typů dat, tzv. kompozičních dat, nesoucích pouze relativní informaci (speciálně procenta, proporce).
- Aplikace standardního a kompozičního biplotu na data z oblasti lékařství.

# Konstrukce biplotu

- Je založena na singulárním rozkladu datové matice  $\mathbf{X}$ .
- Nahradíme tuto matici její nejlepší aproximací (maticí o hodnosti rovné dvěma) z pohledu minimalizace součtu čtverců odchylek jejích prvků od příslušných prvků aproximované matice  $\mathbf{X}_{(2)}$ .
- Následně matici  $\mathbf{X}_{(2)}$  rozdělíme na součin matic skóru, reprezentujících pozorování, a zátěží, reprezentujících proměnné.
- Současným zobrazením těchto matic do jednoho grafu získáme biplot.

# Kompoziční data

- Vícerozměrný statistický soubor, jehož složky představují kvantitativně vyjádřené části celku velikosti  $k$  (100 nebo 1).

## Definice

*Sloupcový vektor  $\mathbf{x} = (x_1, x_2, \dots, x_D)^T$  se nazývá  $D$ -složková kompozice, jestliže jsou všechny jeho prvky kladná reálná čísla nesoucí pouze relativní informaci.*

- Informace obsažená v kompozici se změnou měřítka nezmění.
- Požadujeme invarianci vůči permutaci - aby se při změně pořadí složek vektoru nezměnily podíly mezi nimi.
- Podkompoziční soudružnost - podíly mezi složkami v podkompozici jsou vždy stejné jako podíly v rámci celé kompozice.
- Řídí se tzv. Aitchisonovou geometrií.

# Kompoziční data - Clr transformace

- Škála kompozic je relativní.
- Na reálném euklidovském prostoru nelze s kompozičními daty pracovat přímo, proto byly sestrojeny tzv. logratio transformace.
- Clr transformace (Centred logratio transformation) - zachovává vzdálenosti mezi daty, vede k singulární varianční matici.
- Clr transformace tvoří souřadnice vzhledem ke generujícímu systému, řídí se Aitchisonovou geometrií.
- Clr transformace vyjádřená po složkách:

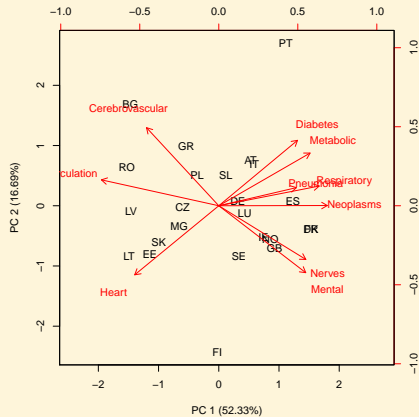
$$\text{clr}(\mathbf{x}) = \left( \ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right)^T,$$

kde  $g(\mathbf{x}) = \sqrt[D]{\prod_{i=1}^D x_i}$  značí geometrický průměr složek kompozice  $\mathbf{x}$ .

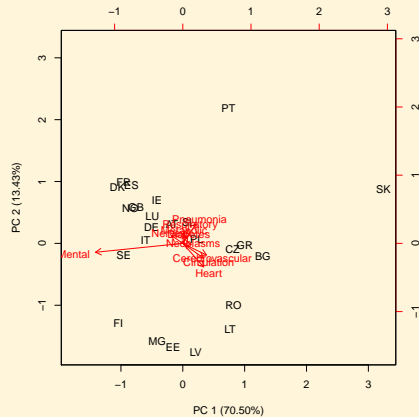
# Příklad - Úmrtí v důsledku nemoci

- Jsou zkoumány podíly nemocí na úmrtnost v každém z vybraných 24 států Evropy.
- **Úkol** - vzájemný vztah mezi jednotlivými státy a také mezi jednotlivými nemocemi.
- Porovnání standardního a kompozičního biplotu (s užitím clr transformace).
- Pro výpočty byl použit software R.

# Příklad - Úmrtí v důsledku nemoci



standardní biplot



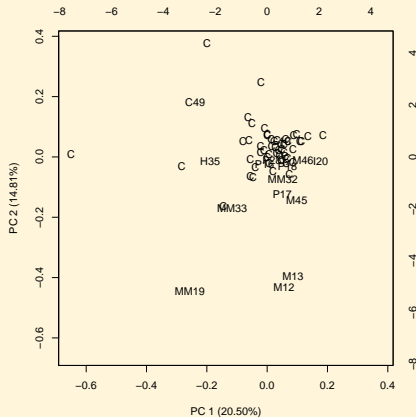
kompoziční biplot



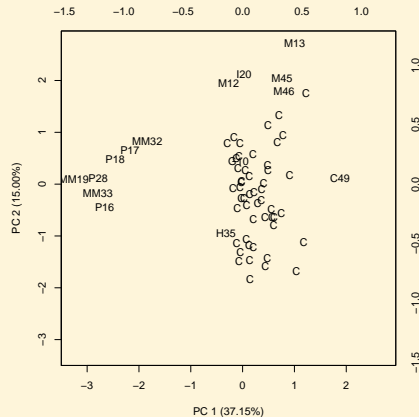
# Příklad - Metabolomická data

- Hodnoty naměřené v rámci výzkumu v Laboratoři dědičných metabolických poruch Ústavu molekulární a translační medicíny ve Fakultní nemocnici Olomouc.
- Hladiny acylovaných karnitinů a aminokyselin u pacientů, jež trpí poruchami souvisejícími s těmito látkami. Dále jsou zahrnuty kontrolní vzorky.
- Protože byla data srovnávána na základě velkého počtu proměnných, byly v grafickém výstupu vynechány paprsky reprezentující tyto proměnné a zůstanou zobrazena pouze pozorování (pacienti a kontroly) v podobě bodů.
- Cílem je vyšetřit datovou strukturu, zejména výskyt shluků.

# Příklad - Metabolomická data (acylované karnitiny)

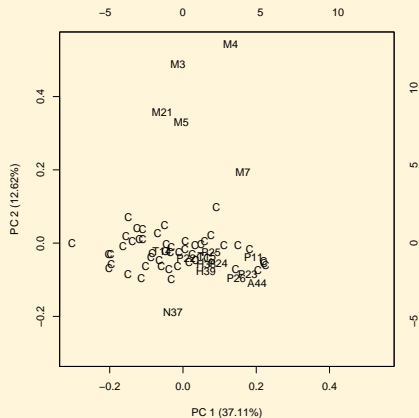


skóry standardního biplotu

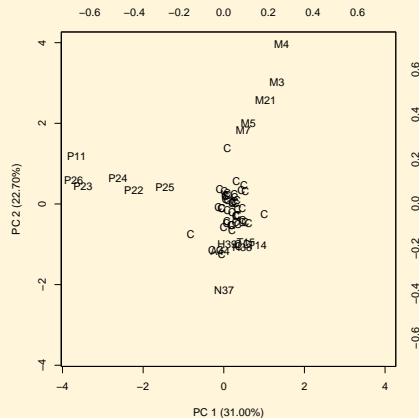


skóry kompozičního biplotu

# Příklad - Metabolomická data (aminokyseliny)






skóry standardního biplotu



skóry kompozičního biplotu

# Literatura

-  J. Aitchison (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
-  J. Aitchison, M. Greenacre (2002) *Biplots of compositional data*. Journal of the Royal Statistical Society, 51 (4), 375–392.
-  K. R. Gabriel (1971) *The biplot graphic display of matrices with application to principal component analysis*. Biometrika, 58 (3), 453–467.