

# Bi8600: Vícerozměrné metody

## 5. cvičení

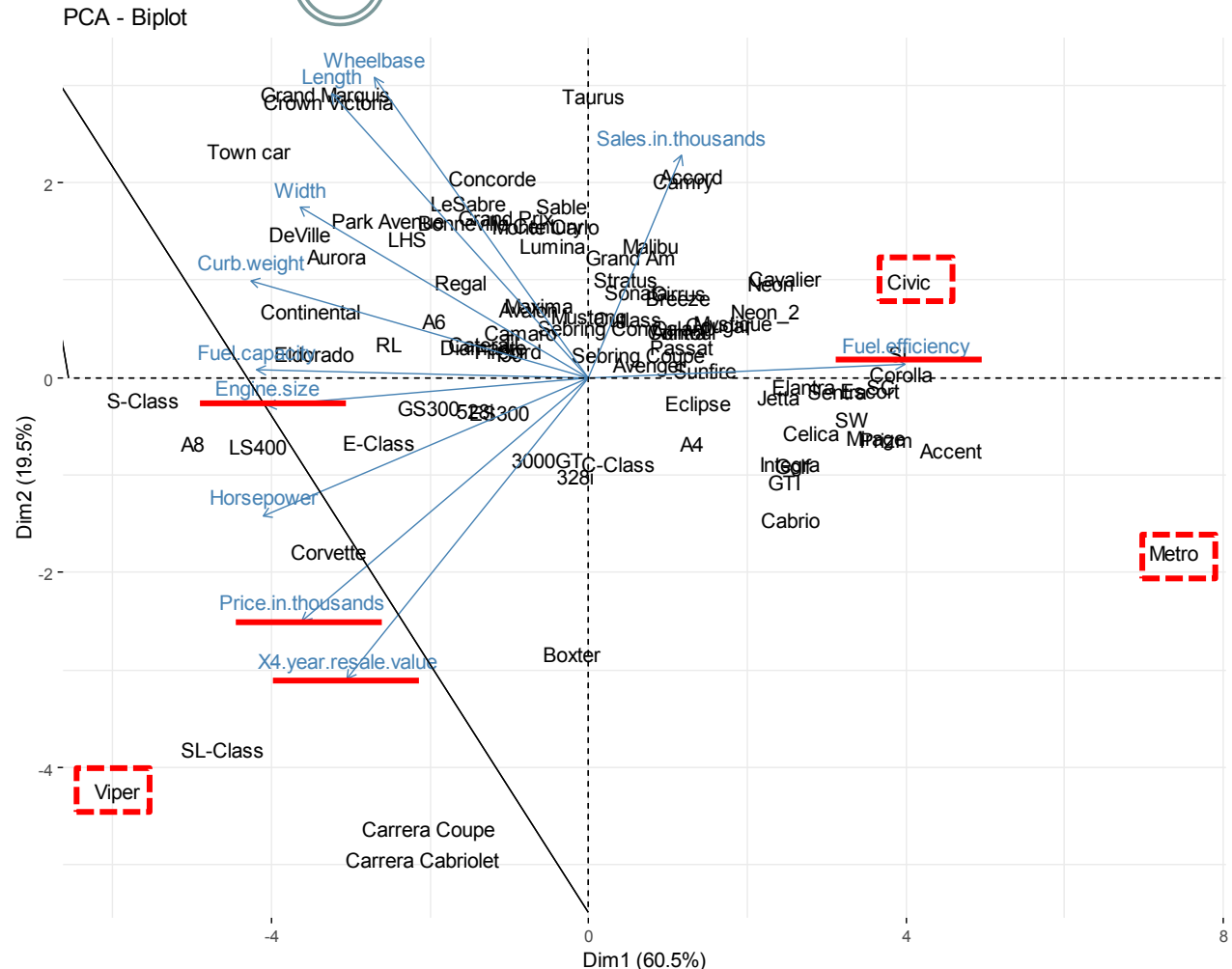


Opakování  
Ordinační metody  
Korespondenční analýza  
Nemetrické škálování

# PCA



- Jaké procento rozptylu popisují první dvě osy PCA?
- Jaká je korelace efektivity paliva a velikosti motoru?
- Jaká je efektivita paliva a velikost motoru modelu Viper?
- Jaká je efektivita paliva a velikost motoru modelu Metro?
- Víme-li, že prodejní cena modelu Civic je po 4 letech nízká, jaká bude cena (Price.in.thousands) tohoto modelu?



# Opakování I.

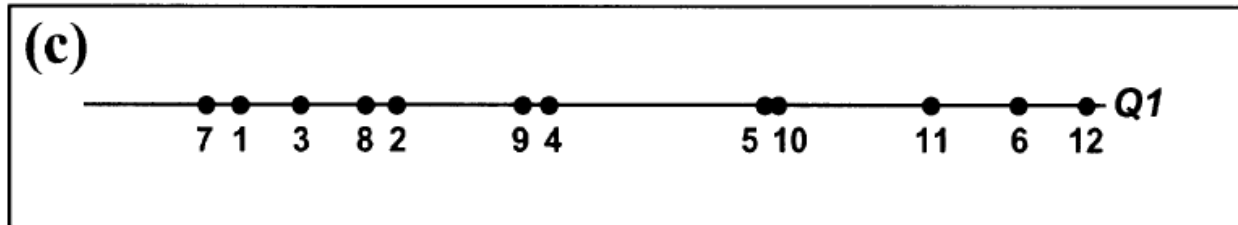
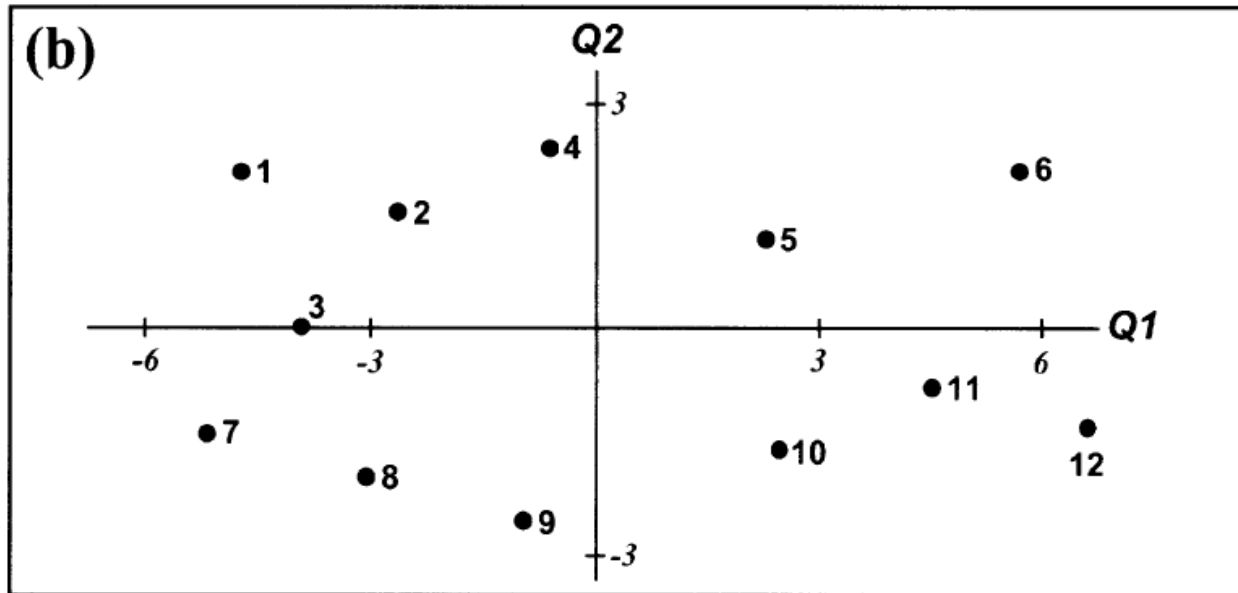


- Popiš vícerozměrná data? Jaký je rozdíl mezi jednorozměrnou a vícerozměrnou analýzou?
- V jaké situaci byste před analýzou standardizovali data? Popište, jak byste provedli.
- Jaký je rozdíl mezi standardizací a transformací? Uveďte příklady transformací.
- Jaký je cíl ordinačních metod? Které ordinační metody znáte?
- Jaký vztah mezi sebou mají nové osy z PCA?
- Čemu je roven součet vlastních čísel u PCA (zvláště pro PCA s kovarianční a korelační maticí na vstupu)?

# Opakování II.



- Na kterém obrázku dochází k redukci vícerozměrného prostoru – 4b nebo 4c?  
Bude v tomto prostoru možné odlišit objekty 4 a 9?



Kenkel et al. (2002)

# Korespondenční analýza - otázky



- Korespondenční analýza je nástroj pro hodnocení vztahů mezi ... a ... datové matice.
- Co popisuje vlastní číslo v korespondenční analýze?
- Co značí vysoká hodnota inercie? V jaké situaci bude hodnota inercie nízká?
- Vyberte, co lze interpretovat z biplotu korespondenční analýzy:
  - 1) vztah objektů
  - 2) vztah proměnných
  - 3) vztah objektů a proměnných
- Jaký maximální počet nových os může vzniknout?

# Korespondenční analýza



- Analogie k PCA
- Vstupní data = agregované údaje objektů/vzorků (průměry, počty)
- Výpočet = analýza vlastních čísel na matici chi-kvadrát hodnot.

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^n [X_{ij} - E_{ij}]^2 / E_{ij}$$

- CA přerozděluje inercii, vysoká inercie – silná vazba mezi řádky a sloupci
- Využití: nejčastěji data abundancí (ekologii), dotazníkové studie
- Nevýhoda: upřednostňuje unikátní málo četné kombinace

## R packages

- `CA()` [*FactoMineR* package],
- `ca()` [*ca* package],
- `dudi.coa()` [*ade4* package],
- `corresp()` [*MASS* package],
- and `epCA()` [*ExPosition* package]

# Korespondenční analýza – interpretace biplotu I.



- Pozice objektů (vzorky, v obrázku plná kolečka) a proměnných (druhy, prázdné čtverečky) v biplotu korespondenční analýzy interpretujeme následujícím způsobem:

2. Druhy, které se vyskytovaly spolu ve vzorcích, budou v ordinačním diagramu umístěny poblíž sebe (C, D).

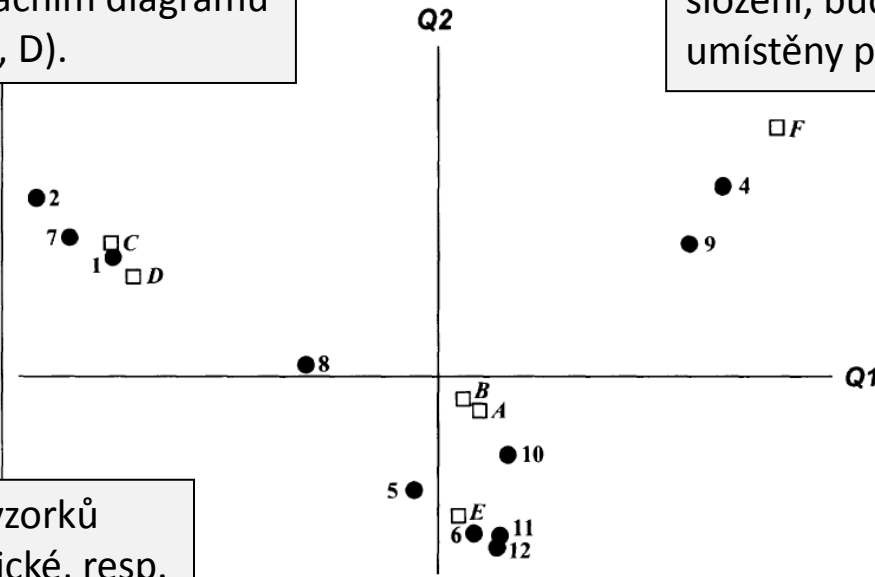
Druhy, které se vyskytovaly v jiných vzorcích, budou v diagramu umístěny dále od sebe (E, F).

3. Druhy umístěny poblíž vzorků byly pro tyto vzorky typické, resp. se vyskytovaly pouze v nich (1-C).

Když se druh v daném vzorku nevyskytoval, budou od sebe druh a vzorek v ordinačním diagramu vzdáleny (1-F).

1. Vzorky, které mají podobné druhové složení, budou v ordinačním diagramu umístěny poblíž sebe (4, 9).

Vzorky, které nemají společné druhy, budou v ordinačním diagramu umístěny dále od sebe (1, 9).



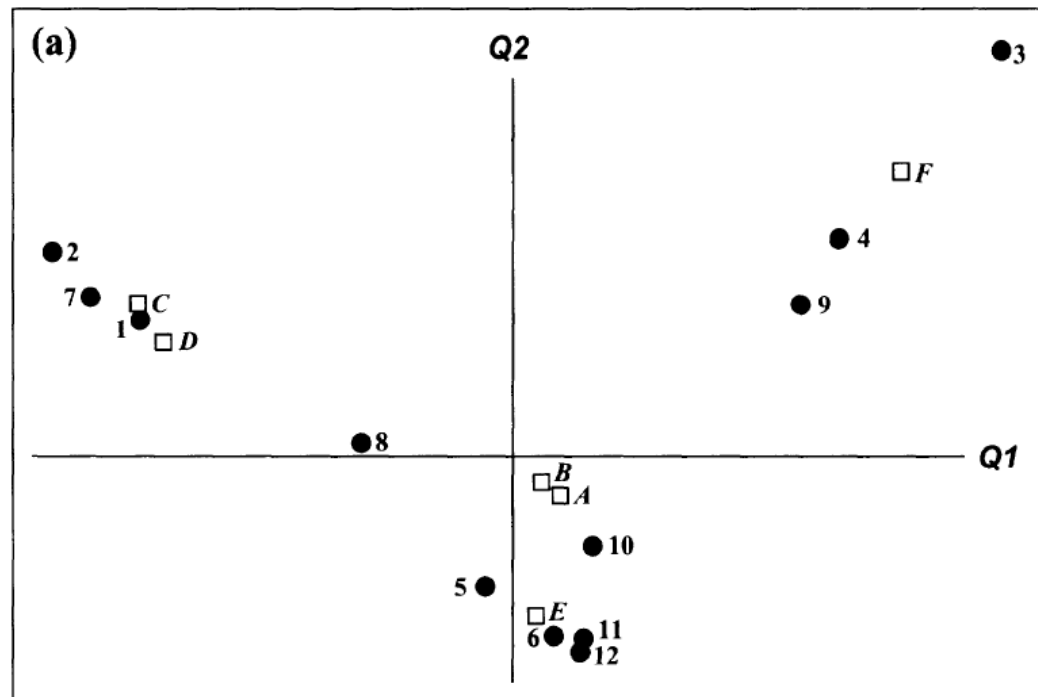
Kenkel et al. (2002)

Body poblíž středu ordinačního diagramu nemají výrazný profil (B, A).

# Korespondenční analýza – interpretace biplotu II.



- Interpretujte biplot z korespondenční analýzy:
  - a) Vztah vzorku 2 vs. 7 a 2 vs. 9.
  - b) Které druhy se vyskytovaly ve stejných a které v odlišných vzorcích?
  - c) Ve kterém vzorku je nejvíce přítomný druh E a C?



Kenkel et al. (2002)



# Nemetrické škálování (NMDS)



- Jaký je princip a základní výstup ne/metrického škálování?
- Jaký je rozdíl mezi metrickým a nemetrickým škálováním?
- Jaké jsou předpoklady NMDS?