

# Analiza multivariata a datelor

# Analiza multivariata

- Presupune utilizarea unui grup de metode statistico-matematice cu ajutorul carora se pot cerceta simultan legaturile de asociere existente intre mai mult de doua variabile.

# Analiza multivariata

- Cele mai utilizate tehnici:
  - ➔ Analiza regresiei multiple
  - ➔ Analiza discriminantului
  - ➔ Analiza multivariata a variatiei
  - ➔ Analiza canonica
  - ➔ Analiza factoriala
  - ➔ Analiza grupurilor
  - ➔ Scalarea multidimensională

# Clasificare

- Criterii de clasificare
  - Existența unor relații de dependență (cauzalitate) și/sau asociere
  - Numărul de variabile dependente
  - Tipul de scală utilizat

# Clasificare

- O singura variabila dependenta:
  - ▶ Scale parametrice: analiza discriminantului liniar multiplu
  - ▶ Scale nonparametrice: regresia multipla
- Mai multe variabile dependente:
  - ▶ Scale parametrice: analiza multivariata a variatiei
  - ▶ Scale nonparametrice: analiza canonica

# Clasificare

- Variabile independente:
  - ▶ Scale parametrice:
    - ▶ Analiza factoriala
    - ▶ Analiza grupurilor
    - ▶ Scalare multidimensionala parametrica
  - ▶ Scale nonparametrice:
    - ▶ Scalare multidimensionala nonparametrica
    - ▶ Analiza structurilor latente

# Metoda regresiei multiple

- Permite analiza relatiei liniare dintre o variabila dependenta si una sau mai multe variabile independente
- **Obiectiv:** explicarea si previziunea variatiei variabilei dependente in functie de covarianta ei cu variabilele independente.

$$\hat{Y} = a + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_i X_i + \dots + \hat{\beta}_n X_n$$

- Utilizeaza metoda celor mai mici patrate
- **Ex.:** cererea de bunuri/servicii (dependenta) in functie de factori determinanti (venituri, cifra de afaceri, pret, etc.)

# Discriminantul liniar multiplu

- Estimarea relatiei dintre o variabila dependenta masurata cu ajutorul unei scale nonparametrice si combinatii liniare ale mai multor variabile independente (masurate pe o scala parametrica)
- Presupune definirea a doua sau mai multe grupuri, omogene interne si eterogene intre ele.
- **Obiectiv:** maximizarea raportului dintre dispersie dintre grupuri si dispersia interna



# Discriminantul liniar multiplu

## ■ Utilizari:

- ▶ Determinarea contributiei fiecarei variabile explicative la discriminare
- ▶ Determinarea celei mai bune combinatii liniare a variabilelor explicative care maximizeaza raportul intre dispersia dintre grupuri si dispersiile din interiorul grupurilor
- ▶ Stabilirea procedurilor de predictie legata de apartenenta componentelor la diversele grupuri (segmentare)
- ▶ Analiza existentei diferentelor semnificative statistic intre profilurile grupurilor definite

# Analiza multivariata a variatiei

- Cuprinde un grup de metode: proiectari factoriale, patratele latine, patratele greco-latine.
- **Obiectiv:** separarea si testarea semnificatiei efectelor cauzate de actiunea simultana a mai multor factori

# Analiza multivariata a variatiei

- Cerinte asupra datelor:
  - ▶ Datele provin de la grupuri experimentale, la nivelul unei populatii normale, pe baza unei esantionari aleatoare.
  - ▶ Eroarea experimentală are o abatere standard constantă.
  - ▶ Variatia aleatoare este independentă de timp.
  - ▶ Efectele diferitelor surse de variație este aditiv, nu multiplicativ.

# Analiza canonica

- Nu depinde de tipul scalei utilizate.
- **Obiectiv**: studierea relatiei liniare intre un grup de variabile dependente si un grup de variabile independente.
- Numarul maxim de functii utilizat este egal cu numarul variabilelor din grupul cel mai mic.

# Analiza canonica

## ■ Utilizari:

- ▶ Determinarea directiei, intensitatii si semnificatiei corelatiei dintre cele doua seturi de variabile.
- ▶ Estimarea coeficientilor celor doua seturi de variabile, in conditiile maximizarii corelatiei dintre seturi.
- ▶ Explicarea si previzionarea variatiei setului dependent, pe baza covariantei acestuia cu setul independent.
- ▶ Determinarea contributiei relative a fiecarei variabile in cadrul functiilor canonice.

# Analiza factoriala

- Realizata cu ajutorul mai multor tehnici: metoda componentelor principale, criteriul varimax, criteriul quartimax, rotatia axelor.
- **Obiectiv:** cercetarea legaturilor de interdependenta dintre mai multe variabile, in scopul determinarii setului minim de explicarea a unui anumit fenomen.
- Presupune definirea unor **factori** care sa **condenseze** o cat mai mare parte din informatia (variabilele) initiala.

# Analiza factoriala

## ■ Utilizari:

- ▶ Identificarea setului de dimensiuni latente existente in variabila initiala (**analiza factoriala de tip R**).
- ▶ Combinarea sau condensarea componentelor unui populatii statistice si crearea unor grupuri distincte (**analiza factoriala de tip Q**).
- ▶ Concentrarea (minimizarea) variabilelor.

# Analiza grupurilor

- Realizata cu ajutorul mai multor tehnici: **distanța euclidiană, gruparea pe baza coliniarității și gruparea de proximitate.**
- **Obiectiv:** clasificarea componentelor unei multimi eterogene în grupuri omogene, pe baza unor criterii obiective.
- Varianta 1: abordare top-down.
- Varianta 2: abordare bottom-up.



# Scalarea multidimensională

- Obiectele sau fenomenele analizate sunt reprezentate sub forma de puncte într-un spațiu perceptual multidimensional.
- **Obiectiv:** estimarea importanței relative a variabilelor și a relațiilor perceptuale dintre ele.
- Permite construirea unui spațiu perceptual în care se reprezintă obiectele sau fenomenele, pe baza percepțiilor respondenților.