Kordamisteemad, IFI7071 mitmemõõtmeline statistika 2015

1. Põhimõisted

1.1. Selgita, mida tähendab mitmemõõtmeline statistika.
1.2. Mis on objekt, muutuja, tunnus, väärtus? Too näiteid.
1.3. Too näiteid sõltuvatest ja sõltumatutest tunnustest.
1.4. Milliseid tegevusi tuleb rakendada andmestiku analüüsiks ettevalmistamisel?
1.5. Millised on enamlevinud võimalused puuduvate väärtuste asendamisel? Kuidas mõjutab jaotuse keskväärtust puuduvate väärtuste asendamine keskväärtustega? Kuidas mõjutab sama asendamine jaotuse variatiivsust ja see omakorda tunnuste vahelisi seoseid?
1.6. Millal (millises olukorras) on tunnuste grupeerimine vajalik? Nimeta 2 põhiolukorda.
1.7. Nimeta kaks põhilist moodust, kuidas saab gruppide väärtuseid kokku panna. Selgita nende erinevusi (skaala, puuduvad väärtused).
1.8. Too näide, kuidas tunnuste kokkupanemisel tunnuse tüüp muutub (algtunnused olid … tüüpi, kokkupandud tunnuse tüüp on ...). Kuidas mõjutab selline muutus andmete analüüsi?

2. Klasteranalüüs. Tunnuste grupeerimine

2.1. Millal kasutatakse klasteranalüüsi?
2.2. Klasteranalüüs võimaldas grupeerida tunnuseid ja objekte. Too näiteid nimetatud olukordadest.
2.3. Mis oli hierarhilise klasteranalüüsi kui meetodi põhimõte? Mis tüüpi peavad olema tunnused, et klasteranalüüsi läbi viia?
2.4. Tunnuste grupeerimisel oli hierarhilise klasteranalüüsi juures oluline arvutilt tellida (selgita näite põhjal, mida antud valikud võimaldavad):
2.4.1. Agglomeration schedule ja *Dendrogram*
2.4.2. Proximitry matrix
2.4.3. Cluster Method: *Between-groups linkage*
2.4.4. Measure: *Pearson correlation*
2.5. Milliseid tabeleid/jooniseid kasutad otsuste/gruppide tegemiseks. Mis on kriteerium, mille põhjal otsustad, et rohkemate gruppide moodustamine ei ole mõistlik.
2.6. Tõlgenda analüüsil saadud tulemusi ja otsusta, millised tunnused oleks mõistlik kokku panna:
2.6.1. sisulise sobivuse põhjal
2.6.2. statistilise sobivuse põhjal
2.7. Selgita, miks peaks loodud klastritele tegema veel reliaabluse kontrolli? Mida võimaldas öelda võimalus *Scale if Item Deleted*?
2.8. Kui oled klasteranalüüsi tulemusena saanud, et tunnused k1, k5, k6 ja k7 moodustavad ühe klastri, siis kuidas nendest tunnustest saab moodustada ühe tunnuse? Kas klasteranalüüs pakub selleks võimalusi?

3. Klasteranalüüs. Objektide grupeerimine

3.1. Selgita hierarhilise klasterdamise ja k-keskmiste klasterdamise meetodite sisulist erinevust. Too välja kriteerium(id), mille alusel tuleks sobiv meetod valida.
3.2. Mis on objektide klasterdamisel sarnasuse mõõduks e mille põhjal sarnaseid objekte kokku pannakse?
3.3. Kas objektide klasterdamisel on vajalik ka tunnuste väärtuseid standardiseerida? Miks?
3.4. Milline on optimaalne arv tunnuseid, mille järgi oleks mõistlik tüpoloogiad luua. Miks rohkem tunnuseid ei kaasata?
3.5. Tunnuste grupeerimisel oli k-keskmiste klasteranalüüsi juures oluline arvutilt tellida (selgita näite põhjal, mida antud valikud võimaldavad):
3.5.1. *Number of Clusters*
3.5.2. *Iterate/Use running means*
3.5.3. *Maximum Iterations*
3.6. Milliseid tabeleid/jooniseid kasutad otsuste/gruppide tegemiseks. Mille põhjal otsustad, et rohkemate gruppide moodustamine ei ole mõistlik.

4. Faktoranalüüs

4.1. Millal kasutatakse faktoranalüüsi? Too näide olukorrast, kus oleks mõistlik kasutada faktoranalüüsi.
4.2. Millised eeldused seab faktoranalüüs andmetele? Mida peaks silmas pidama (tugevalt) asümmeetriliste tunnuste puhul.
4.3. Mida tehakse andmetega vaikimisi enne faktoranalüüsi läbiviimist? Miks?
4.4. Mida lubab andmete kohta öelda KMO test?
4.5. Selgita näite kaudu faktoranalüüsi läbiviimise põhisamme
4.5.1. Method: *Principal Components*
4.5.2. Extract: *Based on Eigenvalue …*
4.5.3. Rotation: *Varimax*
4.5.4. Options: *Coefficient Display Format – Sorted by size ja Supress small coefficients*
4.5.5. Scores: *Save as variables*

4.6. Tõlgenda näite põhjal järgmiseid tabeleid/jooniseid.
   4.6.1. Descriptive Statistics
   4.6.2. Communalities
   4.6.3. Total Variance Explained
   4.6.4. Rotated Components Matrix
4.7. Millised on kaks võimalust summatunnuste väljaarvutamiseks faktoranalüüsi korral? Kirjelda nende erinevust ja seda, millises olukorras, millist võimalust kasutada tuleks.

5. Skaleerimine (MDS)

5.1. Millal kasutatakse mitmemõõtmelist skaleerimist.
5.2. Too näiteid (erinevatest) andmetest, mille analüüsimiseks skaleerimine sobib.
5.3. Mida tähendab, kui skaleerimise tulemusena on loodud 2D mudel? Vastus: see tähendab, et rohkem kui kaks tunnust on taandatud ….
5.4. Selgita, kui oluline on mudeli headuse hindamisel:
   5.4.1. sisuline tõlgendatavus
   5.4.2. statistiline tõlgendatavus (stressi näitaja)

6. Faktoriaalne Anova

6.1. Mille poolest erinevad järgmised meetodid: ühemõõtmeline anova, faktoriaalne anova, ancova, manova, mancova
6.2. Milliste andmete korral ja millises olukorras (millise ülesande/probleemi korral) kasutatakse faktoriaalset anovat?
6.3. Faktoriaalse anova eeldused, nende kontroll.
6.4. Tõlgenda näite põhjal faktoriaalse anova tulemusi (kogu mudeli statistiline olulisus, pea- ja koosmõju statistiline olulisus, E^2).
6.5. Selgita post-hoc testide ja kontrastide sisulist erinevust (mitte väga detailselt). Too näiteid olukordadest, kus kasutada post-hoc teste ja kus kasutada kontraste. Too näiteid, millist post-hoc testi millises olukorras kasutada. Tõlgenda näite põhjal post-hoc testide ja kontrastide tulemusi.

7. Manova

7.1. Millal kasutatakse manova testi? Milliste testide alternatiivtest on manova? Miks eelistada manova testi nimetatud olukorras?
7.2. Milliseid tunnuseid (DV ja IV lõikes) ja kui suurel hulgal manovas kasutada saab?
7.3. Millised on manova eeldused? Kuidas seda/neid kontrollitakse?
7.4. Millise väärtuse põhjal otsustad, kas erinevus on statistiliselt oluline? Tõlgenda näite põhjal manova testi tulemusi (k.a. effekti suurust).
7.5. Tõlgenda näite põhjal post-hoc testide ja kontrastide tulemusi.

8. Diskriminantanalüüs

8.1. Millal kasutatakse diskriminantanalüüsi? Mis tüüpi peavad olema DA-s kasutatavad andmed?
8.2. Eeldused DA-s ja nende kontrollimine.
8.3. Tulemuste tõlgendamine näite põhjal (mudeli eristuse kontroll, väärtuste prognoosimine).

9. Lineaarne regressioon

9.1. Millal kasutatakse lineaarset regressioonanalüüsi? Mil viisil erineb regressioonanalüüs korrelatsioonanalüüsist?
9.2. Millised nõudmised esitab lineaarne regressioonanalüüs andmete tüüpidele?
9.3. Kuidas on võimalik kontrollida seose kuju lineaarsust? Miks seda üldse on vaja kontrollida?
9.4. Tõlgenda näite põhjal järgmiseid tabeleid/jooniseid:
   9.4.1. Model Summary
   9.4.2. Anova
   9.4.3. *Coefficients* (ole valmis kirjutama regressioonvõrrandit kahel viisil ning esitama kahte moodi prognoose: objekti väärtuste põhiselt ning suurenemise/vähenemise kaudu)
9.5. Selgita binaarsete tunnuste komplekti (*dummy variables*) loomise vajadust ning sisulist loogikat (mitte tehnilist tegemist). Kuidas tõlgendatakse selliselt mudelisse lisatud tunnuseid.

10. Logistiline regressioon

10.1. Millal kasutatakse logistilist regressiooni? Selgita logistilise regressiooni erinevust lineaarsest regressioonist.
10.2. Too näide olukorrast (tunnustest), mille korral võiks kasutada logistilist regressioonanalüüsi.
10.3. Kõige tavalisemas logistilise regressiooni analüüsis on sõltuv tunnus binaarset tüüpi. Too näide, kuidas tõlgendatakse analüüsi tulemusi (kasuta taustarühma ja riskirühma mõisteid).
10.4. Selgita, kuidas erineb tunnuse tüübi järgi sõltumatute tunnuste analüüsi lisamine.
10.5. Milliste väärtuste põhjal ja kuidas tõlgendatakse logistilise regresioonanalüüsi tulemusi.