

Kursus: Mitmemõõtmeline statistika

Seminar VII: Tunnuste grupeerimine

väärtuste summeerimine üle tunnuste;
suurema korrelatsiooni tee;
hierarhiline klasteranalüüs;
reliaabluskoefitsient Cronbach'i α

Õppejõud: Katrin Niglas
PhD, dotsent
informaatika instituut



Tunnuste grupeerimine

Erinevad eesmärgid/olukorrad, miks **tunnuseid koondada**:

Testide korral (üksikküsimus, alamosa, kogu test):

- õigete/valede vastuste arvu leidmine
- punktide summeerimine

Ankeetide korral (sisuliselt haakuvate küsimuste plokk):

- teatud tüüpi vastuste arvu leidmine (nt "jah", vastamata,...)
- teatud küsimusteploki põhjal koondtunnuse arvutamine (nt keskmine hinnang, summaarne kulu, ...)
- **latentsete** e peidetud muutujate/tunnuste leidmine (nt hoiakud fenomeni eri aspektide suhtes, psühhosomaatilised seisundid, ...)



Tunnuste grupeerimine – näide "test"

Fail: Eksam_Inglise_keel.sav

Test, millel on 5 alamosa
testiküsimused ja –ülesanded hinnatud punktidega 0 kuni ...

ÜI1: Leida alamosade ja kogu testi punktisummad!

Transform/Compute...

NB! Funktsioon *sum(...)* arvutab summa ka puuduvate väärtuste korral
Liitmistehe "+" märgi abil puuduvate väärtuste korral vastust ei anna

ÜI2: Leida täiesti valede vastuste arv kolmandas alamosas

Transform/Count...



Tunnuste grupeerimine – näide "tagasiside"

Fail: tagasiside.sav

Tagasisideküsitlus ainekursuste kohta,
11 hinnangut igale kursusele 5-palli skaalal

ÜI1: Leida iga hindaja koondhinnang!

Transform/Compute...

NB! Funktsioon *mean(...)* arvutab keskmise ka puuduvate väärtuste korral
Arvutustehe "+" ja "/" märgi abil puuduvate väärtuste korral vastust ei anna

ÜI2: Leida puuduvate hinnangute arv iga vastaja jaoks

Transform/Count...



Tunnuste grupeerimine – näide “keelehoiakud”

Fail: Keelehoiakud.sav

Ankeetküsitlus koolinoorte keelehoiakute analüüsimiseks,
Keelekäitumise küsimused 5-palli skaalal (8 küsimust)
Keelehoiakute küsimused 4-palli skaalal (orig 35 küsimust)

Ü1: Leida keelehoiakute erinevad aspektid ja moodusta iga aspekti kohta koondtunnus/summamuutuja!

? Millised tunnused omavahel kokku panna?

Sisuline sobivus + statistiline sobivus (väljendub seoste tugevuses)



Tunnuste grupeerimine – latentsed muutujad

Latentsete e peidetud muutujate/tunnuste leidmine – sisuliselt otsitakse ühisosa omavaid alg tunnuseid ja vajadusel moodustatakse omavahel sobivatest tunnustest koondtunnus.

Eelsamm: Uuri seoseid (seosekordajad, diagrammid, risttabelid, ...)

Alternatiiv1: Moodusta suurima korrelatsiooni tee

Alternatiiv2: Kasuta hierarhilist klasteranalüüsi

Alternatiiv3: Kasuta faktoranalüüsi

Järelsamm: Kontrolli tunnuste_grupi/koondtunnuse usaldusväärsust (Cronbach'i α)



Tunnuste grupeerimine – klasteranalüüs

Hierarhilise klasteranalüüsi idee seisneb selles, et samm-sammult hakatakse kokku panema neid tunnuseid, mis on omavahel kõige sarnasemad st millel on omavahel kõige suurem seos.

Kauguse/sarnasuse mõõduks

- *Measure/Interval: Pearson correlation*

Klastrite ühendamise meetod väljal *Cluster Method:*

- *Between-groups linkage* klastrite vaheline sarnasus arvutatakse üksiktunnuste seosekordajate keskväärtusena
- *Nearest neighbor* klastrite vaheline sarnasus arvutatakse erinevates klastrites olevate omavahel kõige enam seotud tunnuste vahelise seosekordajana. See meetod annab tulemuseks nn **suurima korrelatsiooni tee**

=> Esimesel sammul iga tunnus omaette – viimasel sammul kõik koos st üks klaster/grupp – põhiküsimus: millal kokkupanemine lõpetada!?

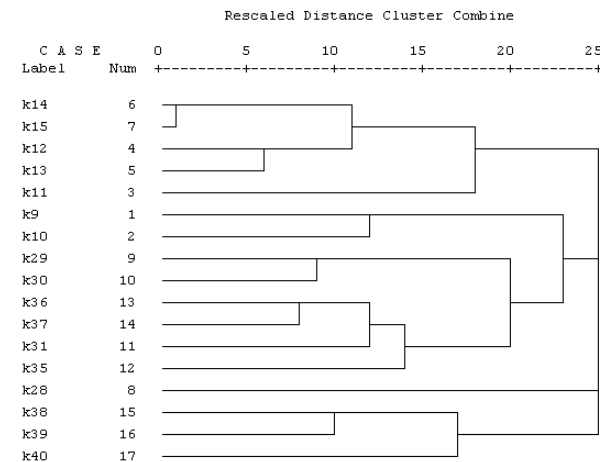


TALLINNA ÜLIKOO

Tunnuste grupeerimine – klasteranalüüs

*****HIERARCHICAL CLUSTER ANALYSIS*****

Dendrogram using Average Linkage (Between Groups)



TALLINNA

Tunnuste grupeerimine – Cronbach'i α

Usaldusväarsuse kontrollimise idee antud kontekstis:
kui hästi antud alg tunnuste komplekt moodab
ühte ühemõõtmelist latentset nähtust?

Matemaatiliselt: võrreldakse alg tunnuste variatiivsuse ühisosa
alg tunnuste üldise variatiivsusega. Mida suurem
variatiivsuse ühisosa, seda kõrgem usaldusväarsus.

Standardiseeritud tunnuste korral (kui alg tunnuste variatiivsus on
võrdne) on valemil lihtne kuju:

$$\alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

PS! Mida suuremad korrelatsioonid (ühisosa), seda suurem α
Mida rohkem tunnuseid, seda suurem α