

Faktoranalüüs (*factor analysis*, FA)

Mitmene korrespondentsanalüüs (*multiple correspondence analysis*, MCA)

Tanel Kaart

Mitmemõõtmelise statistika koolitus
Jaanuar 2016, Tartu

Faktoranalüüs (täpsemalt - kirjeldav faktoranalüüs) (*exploratory factor analysis*, FA)

Faktoranalüüsi eesmärgiks on eelkõige

- andmemahu vähendamine info koondamise teel,
- latentsete (otseselt mittemõõdetavate) näitajate/mustrite hindamine,
- üldiste indeksite tekitamine.

Enamasti on peakomponentanalüüs faktoranalüüsi esimene samm.

Uuritavad tunnused peavad olema arvulised.

Faktoranalüüsi näide: Kaie Metsaots, „Ida-Viru küsitlus“.

Vastaja	Värs	Sugu	Rahvus	Haridustas	funktsioon	kaevanduse	informatsioon	kasutusel	ehitised	renoveerimise	osalus	ühetegevus	koostöö	haldusreform	regionaalpoliitika	toetus	rahastamine	tuntus	turismipotentsiaal	Venemaa	meeleolu
respondent	1	2	1	3	5	3	4	5	2	4	5	4	3	3	3	4	5	1			
respondent	2	2	1	5	5	5	5	5	3	4	3	5	5	3	5	5	5	3			
respondent	2	1	1	2	5	3	4	5	5	4	3	3	3	2	4	4	4	2			
respondent	1	2	1	5	5	5	5														
respondent	2	2	1	4	5	4	4														
respondent	2	2	1	4	5	4	5														
respondent	1	2	1	2	5	5	5														
respondent	1	2	1	5	5	5	4														
respondent	1	1,5	1	3	4	2	3														

28. Millise tähtsusega on Teie arvates järgmised tegurid Ida-Virumaa arengu seisukohalt?

Märkige rist vastavas kasti.

	ei ole oluline	väheoluline	neutraalne	oluline	väga oluline
kaevandatud alade uue funktsiooni andmine					
kaevandusprandi väärtustamine ja säilitamine					
rahvale rohkem informatsiooni kaevandusprandi kasutamise võimaluste kohta					
varem kaevandatud alade kasutusselevõtmine					
endiste kaevandusalaehitiste kasutusselevõtmine					
linnade ja asulate välisilme parandamine, sh. hoonete renoveerimine ja avaliku ruumi heakorra parandamine					
avalikkuse osaluse suurendamine otsustusprotsessides kaevandusalaehitiste kasutamise üle					
ühetegevustunde suurendamine rahva ühistegevuste läbi					
kaevandusprandi väärtustamise ja kasutamise osas koostöö parandamine riigi, omavalitsuste, ametite, kaevandusettevõtete, arendajate, turismiasutuste vahel					
haldusreformi elluviimine, omavalitsuste ühendamine					
regionaalpoliitika väljatöötamine ja rakendamine					
EL toetusrahade suurem suunamine kaevandusprandi väärtustamise ja kaevandusalaehitiste kasutamisele					
Ida-Virumaa tuntuse suurendamine					
turismipotentsiaali parem arendamine					
piiri avamine Venemaaga (suurem turistide ligipääs)					
rohkemate igapäevaste vabaaja veetmise- ja meelelahutuskohtade avamine					
uute töökohtade loomine (töötuse vähendamine)					
sotsiaalprobleemidele (kuritegevus, alkoholi liigtarimine, narkomaania jt.) leevenduse leidmine					

WHAT HINDERS AND WHAT FACILITATES GOOD LIVING IN THE OIL SHALE MINING REGION?

KAIE METSAOTS^(a), ANU PRINTSMANN^(b),
TANEL KAART^(c), KALEV SEPP^(d)

Oil Shale. 2015, Vol. 32, No. 4, pp. 373–392
doi: 10.3176/oil.2015.4.06

ISSN 0208-189X
© 2015 Estonian Academy Publishers

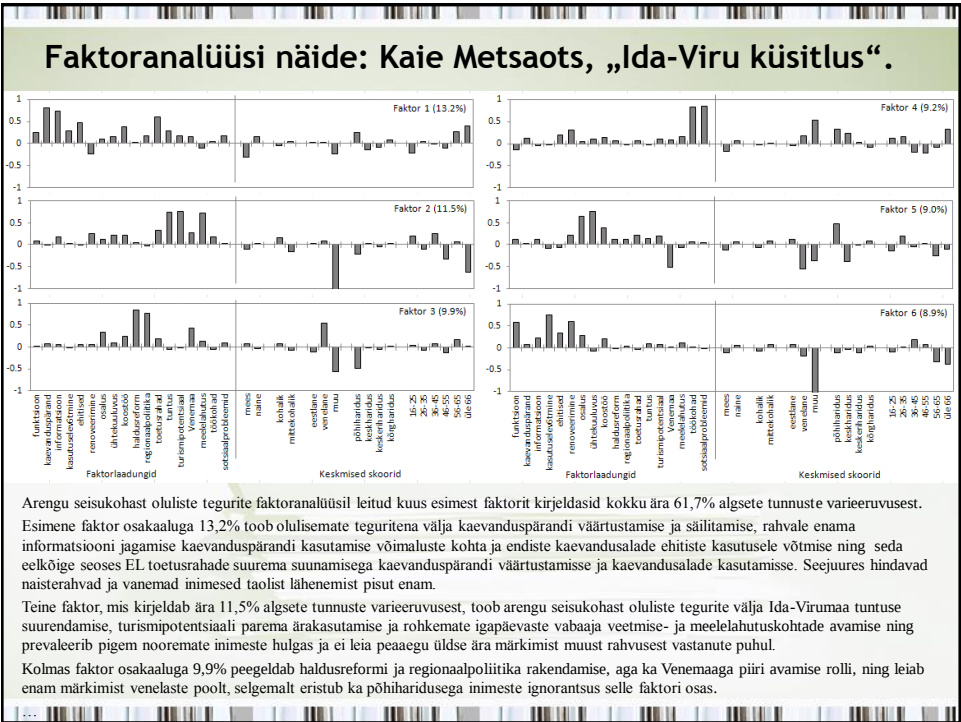
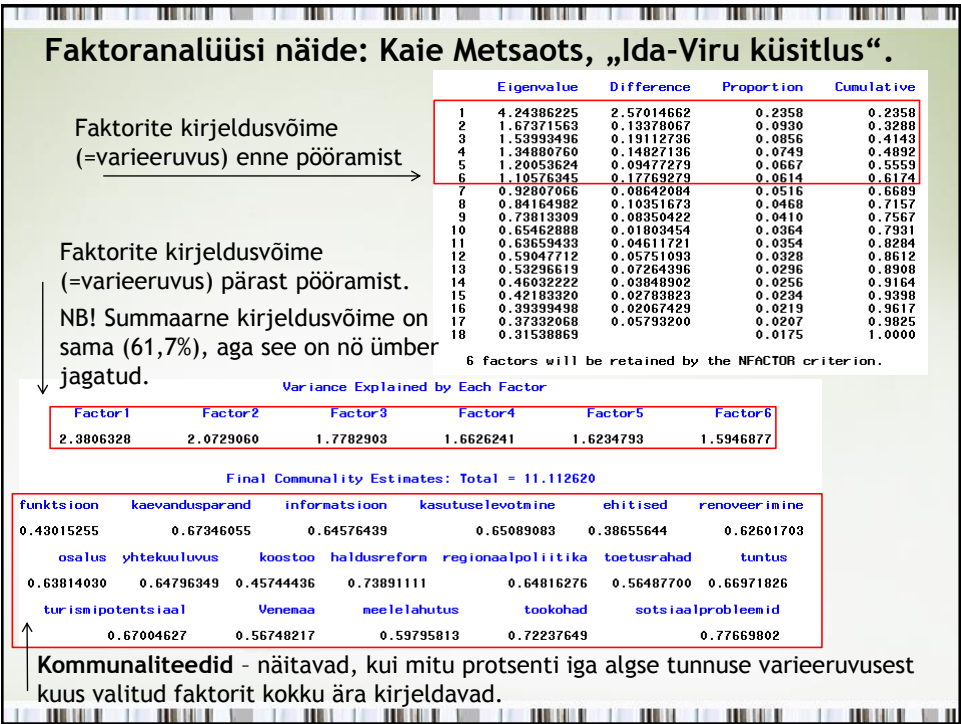
Faktoranalüüsi näide: Kaie Metsaots, „Ida-Viru küsitlus“.

Samm 1: peakomponentanalüüs
(iga faktor püüab olla võimalikult tugevalt seotud kõigi algsete tunnustega)

	Factor1	Factor2	Factor3
funktsioon	0.39696	-0.23243	-0.32288
kaevandusprandi	0.55598	-0.36915	-0.14738
informatsioon	0.64122	-0.35385	-0.25786
kasutusselevõtmine	0.40399	-0.24898	-0.43860
ehitised	0.44346	-0.18667	-0.19649
renoveerimine	0.37570	0.33993	-0.09857
osalus	0.55523	-0.06224	0.23932
ühetegevus	0.46312	0.11942	0.20080
koostöö	0.64627	-0.07624	0.06731
haldusreform	0.36118	-0.20778	0.70670
regionaalpoliitika	0.39371	-0.36148	0.56794
toetus	0.65860	-0.12374	0.06791
tuntus	0.59357	0.23981	-0.25331
turismipotentsiaal	0.59447	0.38295	-0.15306
Venemaa	0.21345	-0.07070	0.17632

Samm 2: faktorite pööramine
(faktoreid nõ pööratakse saavutamaks nende maksimaalset korreleeritust mõnedega algannustest (nii on tulemusi lihtsam tõlgendada)).

	Factor1	Factor2	Factor3
funktsioon	0.24613	0.08632	0.01530
kaevandusprandi	0.80350	-0.01380	0.08010
informatsioon	0.73634	0.16170	0.06493
kasutusselevõtmine	0.28170	0.03112	-0.01398
ehitised	0.47759	-0.00017	0.05460
renoveerimine	-0.22690	0.24754	0.05081
osalus	0.09963	0.11629	0.33969
ühetegevus	0.15784	0.20475	0.10193
koostöö	0.38156	0.21069	0.24020
haldusreform	0.02774	0.04094	0.84781
regionaalpoliitika	0.18111	-0.02757	0.77336
toetus	0.60620	0.32335	0.18452
tuntus	0.29540	0.74347	-0.04672
turismipotentsiaal	0.17953	0.76198	-0.01686



Mitmene korrespondentsanalüüs

(multiple correspondence analysis, MCA)

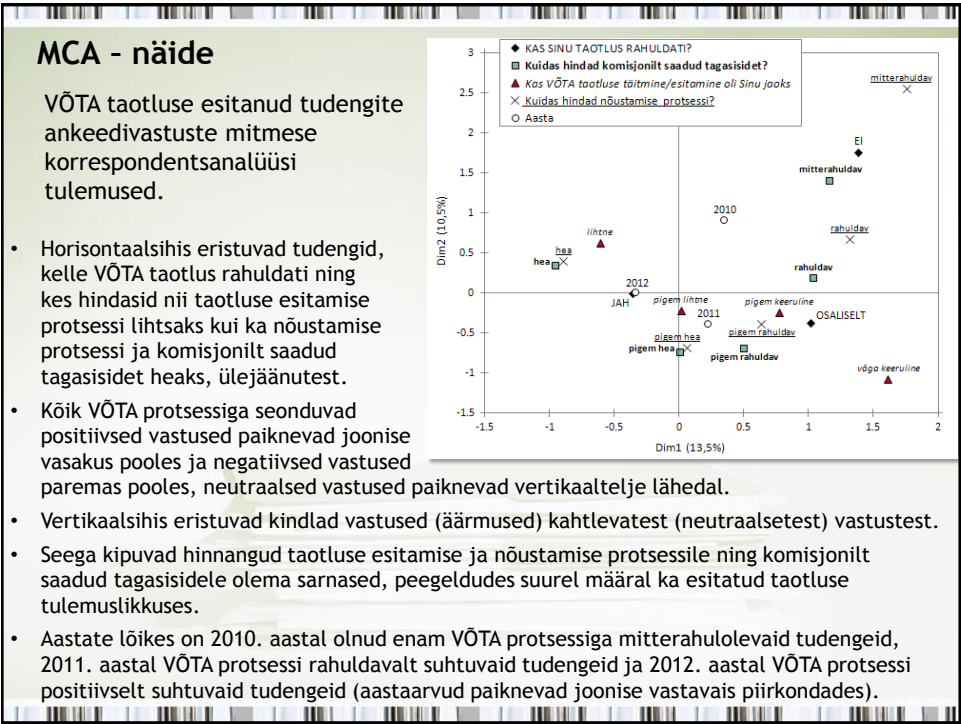
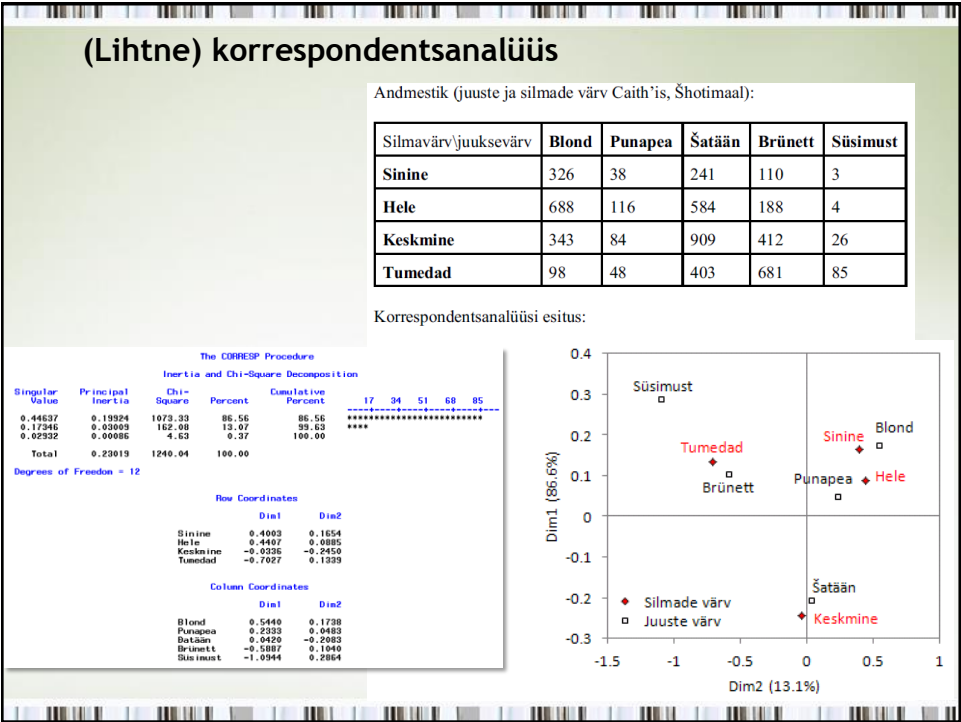
Korrespondentsanalüüsi eesmärgiks on

- andmemahu vähendamine info koondamise teel,
- latentsete (otseselt mittemõõdetavate) näitajate/mustrite hindamine,
- üldiste indeksite tekitamine,
- objektide/gruppide klasterdamine,
- ...

Uuritavad tunnused peavad olema mittearvulised või diskreetsed (võivad olla ka pideva arvtunnuse väärtuste alusel moodustatud grupid).

Mitmene korrespondentsanalüüsi olemus

- Mitmene korrespondentsanalüüs (*multiple correspondence analysis, MCA*) on nn andmete tihendamise meetod, mis võimaldab hulga mittearvuliste tunnuste (näiteks küsitluse tulemuste) mitmemõõtmeliste sagedustabelite kujul esitatavad seosed projitseerida madalamamõõtmelisse (enamasti kahemõõtmelisse) ruumi.
- Analüüsi tulemused esitatakse enamasti joonise kujul, kus teatud väärtuste (küsimuste vastuste) lähestikku paiknemine viitab nende sagedasemale koosesinemisele.
- Seeläbi annab mitmemõõtmeline korrespondentsanalüüs võimaluse uurida korraga mitmete tunnuste ja nende väärtuste omavahelisi (mittelineaarseid) seoseid ning tuvastada võimalikke mustreid vastanute seas.
- Projitseerimise tarvis arvutatavad nn dimensioonid e skoorid on edasi kasutatavad arvulisi tunnuseid eeldavates analüüsides.



Mitmene korrespondentsanalüüs - näide

Raaperi, K., Bougeard, S., Alekseev, A., Orro, T., Viltrop, A., 2012. ASSOCIATION OF HERD BRSV AND BHV-1 SEROPREVALENCE WITH RESPIRATORY DISEASE AND REPRODUCTIVE PERFORMANCE IN ADULT DAIRY CATTLE. Acta Veterinaria Scandinavica 54:4.

Symmetric variable plot (axes F1 and F2: 80,39 %)

Figure 1 Graphical display of MCA for high incidence of respiratory disease symptoms in cows and pregnant heifers (100 herds).

Table 2 Descriptive characteristics of the variables included in the models (100 herds in Model 1 and 77 herds in Model 2)

Variable	Definition of the categories of the variable	Model 1	Model 2
Nasal discharge ("red nose") in cows and/or pregnant heifers (NASCOV)	0 - not present at all or was shown only as single cases at some point during the last two years 1 - present in more than just single cases at some point during the last two years	82	18
in	0 - not present at all or was shown only as single cases at some time point during the last two years 1 - present in more than just single cases at some point during the last two years	80	20
fers	0 - not present at all or was shown only as single cases at some point during the last two years 1 - present in more than just single cases at some point during the last two years	88	12
idor	0 - less than two respiratory disease symptoms were present in more than a single case at some time during the last two years 1 - at least two respiratory disease symptoms were present in more than a single case at some time during the last two years	81	19
or	0 - < 1.3% in a herd (median for cut-off value) 1 - ≥ 1.3% in a herd	37	40
rs	0 - < 1.9 in a herd (median for cut-off value) 1 - ≥ 1.9 in a herd	38	39
rs	1 - 20-99 cows 2 - 100-199 cows 3 - 200-399 cows 4 - > 400 cows	40	25
rs	0 - no 1 - yes	19	21
rs	0 - no 1 - yes	18	14
rs	0 - no 1 - yes	77	23
rs	0 - no 1 - yes	28	72
rs	0 - no 1 - yes	44	61
rs	0 - no 1 - yes	39	51
rs	0 - together 1 - in separate building from 6 months until pregnancy	49	23
rs	1 - lost 2 - loose	30	47
rs	3 - some period of life test, some period loose 1 - lost 2 - loose	71	29
rs	0 - no 1 - yes	47	53
rs	0 - no 1 - yes (at least one animal tested positive)	77	20
rs	0 - negative 1 - 1 - 49%	46	34
rs	2 - 450%	40	31
rs	0 - negative 1 - 1 - 49%	14	10
rs	2 - 450%	37	28
rs	0 - negative 1 - positive	15	28
rs	0 - negative 1 - positive	56	40
rs	0 - negative 1 - positive	44	37

Mitmene korrespondentsanalüüs - näide

Phenotypic plasticity in a generalist insect herbivore with the combined use of direct and indirect cues

Oikos 122: 1626–1635, 2013
doi: 10.1111/oikos.12266
© 2013 The Authors. Oikos © 2013 Nordic Society Oikos
Subject Editor: Jessica Abbott. Accepted 1 April 2013

Siiri-Liis Sandre, Ants Kaasik, Ute Eulitz and Toomas Tammaru

Table 1. Partial contributions of the values of the patterning and colouration traits of *E. atornaria* larvae to the inertias of the two dimensions, results of the MCA analysis of data pooled over all the reported experiments.

Dimension	Dorsal spot		Dorsal horizontal line			Lateral horizontal line			Colour of lateral line					Dominant colour				
	No spot	Spot	No line	Light points	White line	Yellow line	Continuous line	Intermediate line	Dashed line	Yellow	Dark brown	Light brown	Pink	Grey	Green	Light brown	Grey	Dark brown
Dimension 1 (patterning)	0.007	0.116	0.002	0.011	0.175	0.07	0.047	<0.001	0.081	0.072	0.003	<0.001	0.042	0.128	0.075	<0.001	0.170	0.001
Dimension 2 (darkness)	0.001	0.018	0.132	0.018	0.032	0.128	<0.001	<0.001	0.001	0.122	0.147	0.034	<0.001	0.025	0.129	0.039	0.017	0.156

Figure 1. Correspondence of larval appearance to the two dimensions obtained from the MCA which was applied to a multidimensional data set of discrete variables describing different aspects of coloration and patterning from all experiments. Crosses (X) denote the exact position of the depicted larva in the parameter space.

Figure 2. Frequencies of different larval phenotypes in the 2009 plant experiment in the parameter space formed by the two principal dimensions of the MCA analysis. The area of the circle is proportional to the number of individuals with combinations of values falling into the part of the character space defined by respective hexagon. The colors denote the proportions of larvae fed on bilberry (blue) or heather (green). The data are weighted to correct for the different numbers of larvae in the two host plant treatments.