Mitmemõõtmelise statistika koolitus

Korrespondents-, klaster- ja diskriminantanalüüs ning mitmemõõtmeline skaleerimine

Eesti Maaülikool
20.-24. jaanuar 2014
Tanel Kaart

Klasteranalüüs
[cluster analysis]

Klasteranalüüsi eesmärgiks on kas tunnuste või uuritavate objektide rühmitamine. Kaks peamist algoritmi:

- hierarhiline klasterdamine;
- k-keskmiste klasterdamine.

Hierarhiline klasterdamine

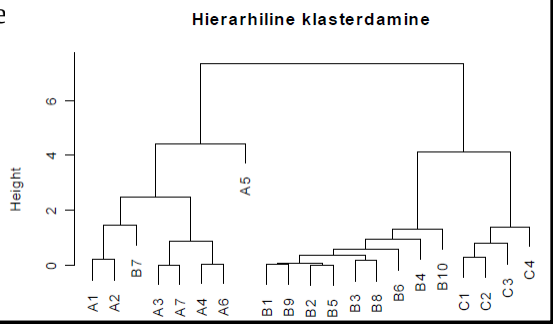
• Loodusuurija sõitis Asustamata Saarele ja avastas seal suure hulga seni teadusele tundmatuid putukaid.

• Ta mõõtis oma lühikese saarelviibimise jooksul tervel hunnikul seninägematutel putukatel igasuguseid näitajaid (igatsorti pikkuseid ja mustrielementide arvu ja paljut muudki).

• Järgmiseks soovis loodusuurija määratleda, mitmesse alamliiki leitud putukad võiksid kuuluda.

• Et saada esimest ligikaudset lähendit, kust oma uurimistööga pihta hakata, soovis ta leida sarnaste putukate rühmad – kes oleksid siis alamliikide kanditaatideks.

• Selleks söötis ta oma andmed klasteranalüüsi teostavasse programmi, mis joonistas järgmise pildi:



Märt Mölsi loengukonspektist

Hierarhiline klasterdamine

... on hästi kasutatav siis, kui meil on suhteliselt vähe objekte või kui on oodata, et klastrid suhteliselt selgelt üksteisest eristuvad.

Hierarhiline klasteranalüüs põhineb väga lihtsal algoritmil:

samm-sammult hakatakse omavahel kokku panema kõige sarnasemaid objekte. Näiteks, kui leidub kaks täpselt ühesuguste tulemustega objekti, siis liidetakse nad esimesel sammul üheks klastriks, peale seda võrreldakse kõiki üksikobjekte ja juba tekkinud klastreid ja liidetakse jälle kõige sarnasemad omavahel jne.

Vaatluste omavahelise kauguse määramine:

Eukleidese kaugus:

Manhattani kaugus:

…

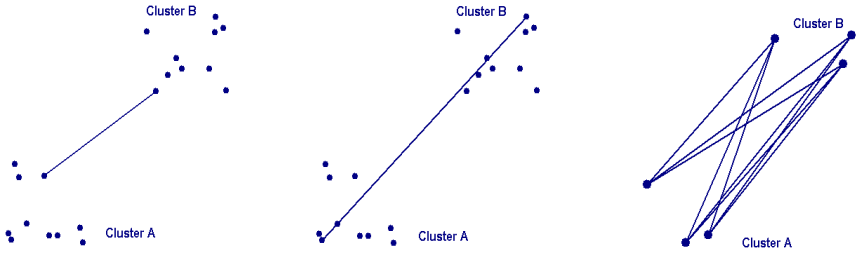
$$d\ (x_1, y_1), (x_2, y_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$d\ (x_1, y_1), (x_2, y_2) = |x_1 - x_2| + |y_1 - y_2|$$

Hierarhiline klasterdamine

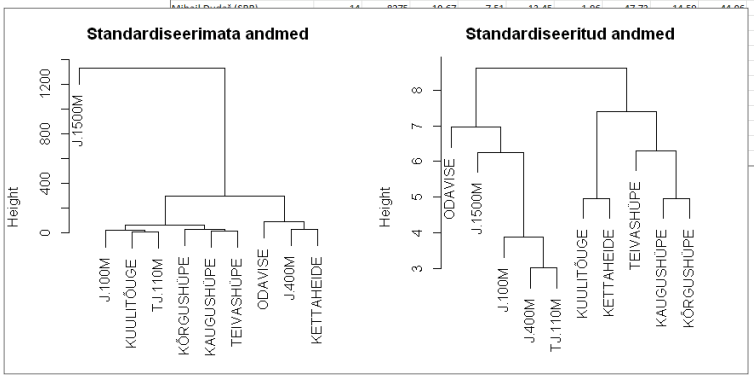
Klastritevahelise kauguse määramine:

"single" "complete" "average"



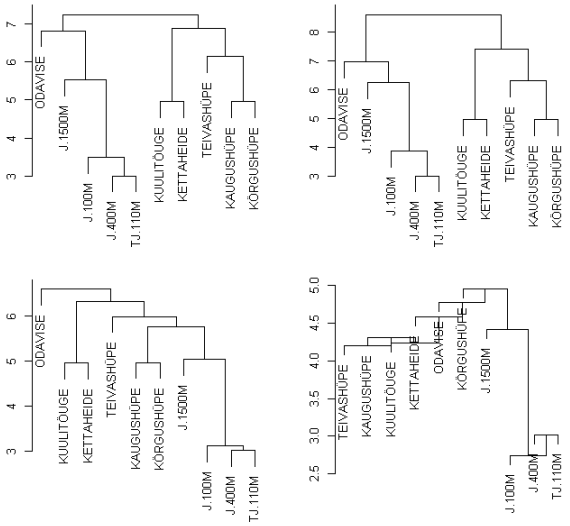
Kergejõustiku MM 2013, meeste 10-võistlus

NIMI	KOHT	KOKKU	J 100M	KAUGUSH	KUULITÕU	KÕRGUSH	J 400M	TJ 110M	KETTAHEI	TEIVASHÜ	ODAVISE	J 1500M
Ashton Eaton (USA)	1	8809	10.35	7.73	14.39	1.93	46.02	13.72	45	5.2	64.83	269.8
Michael Schrader (GER)	2	8670	10.73	7.85	14.56	1.99	47.66	14.29	46.44	5	65.67	265.4
Damian Warner (CAN)	3	8512	10.43	7.39	14.23	2.05	48.41	13.96	44.13	4.8	64.67	270
Kevin Mayer (FRA)	4	8446	11.23	7.5	13.76	2.05	49.53	14.21	45.37	5.2	66.09	265
Eelco Sintnicolaas (NED)	5	8391	10.85	7.65	14.08	2.02	48.25	14.18	39.21	5.3	56.75	264.6
Carlos Chinin (BRA)	6	8388	10.78	7.54	14.49	1.96	48.8	14.05	45.84	5.1	59.98	276
Rico Freimuth (GER)	7	8382	10.6	7.22	14.8	1.99	48.05	13.9	48.74	4.9	56.21	277.8
Ilya Shkurenev (RUS)	8	8370	10.97	7.35	13.88	2.05	48.39	14.34	44.06	5.4	59.46	277
Willem Coertzen (RSA)	9	8343	10.95	7.44	13.88	2.05	48.32	14.3	43.25	4.5	69.35	264.6
Leonel Suárez (CUB)	10	8317	11.07	7.33	14.2	1.9	48.21	14.62	46.41	4.9	68.61	263.9
Pascal Behrenbruch (GER)	11	8316	10.95	7.19	15.86	1.99	48.4	14.46	45.66	4.7	67.07	277.2
Andrei Krauchanka (BLR)	12	8314	11.19	7.39	14.84	2.11	49.65	14.44	46.12	5.1	59.98	279.6
Gunnar Nixon (USA)	13	8312	10.84	7.8	14.68	2.14	48.56	14.57	42.38	4.6	57.97	275.8
										4.9	59.06	266.6
										5.1	65.31	277.9
										5	61.83	273
										5	62.89	273.2
										4.6	50.74	273.7
										4.9	59.63	288.3
										4.5	69.38	286.4
										5.1	64.38	275.9
										4.9	67.65	278.9
										4.5	57.3	278.5
										4.5	54.86	260.1



Kergejõustiku MM 2013, meeste 10-võistlus

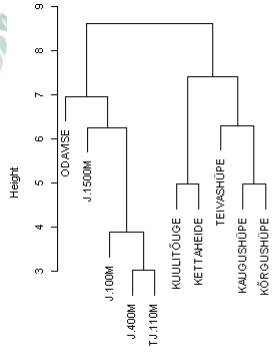
Erinevad klastrite vahelise kauguse määramise meetodid võivad anda erinevad tulemused!



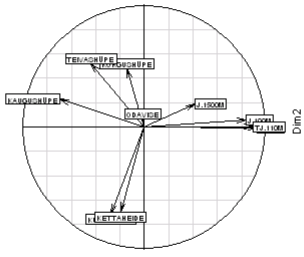
Kergejõustiku MM 2013, meeste 10-võistlus

Grupeerimiseks/klasterdamiseks võib kasutada ka teisi meetodeid.

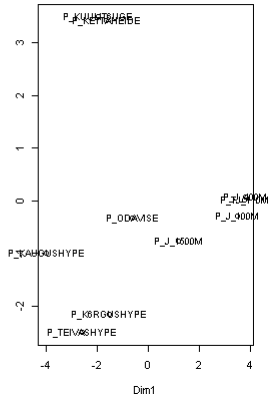
Hierarhiline klasterdamine



Peakomponentanalüüs

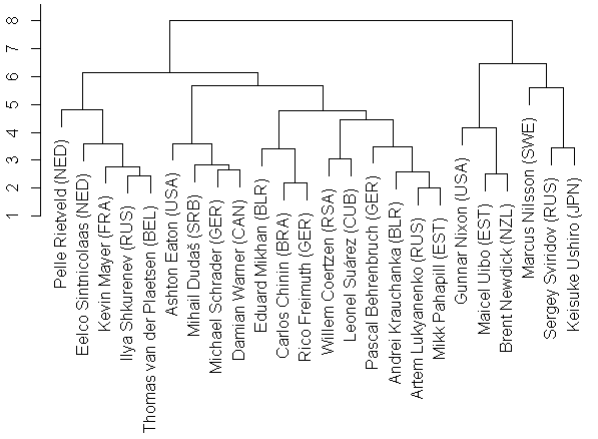


Mitmemõõtmeline skaleerimine



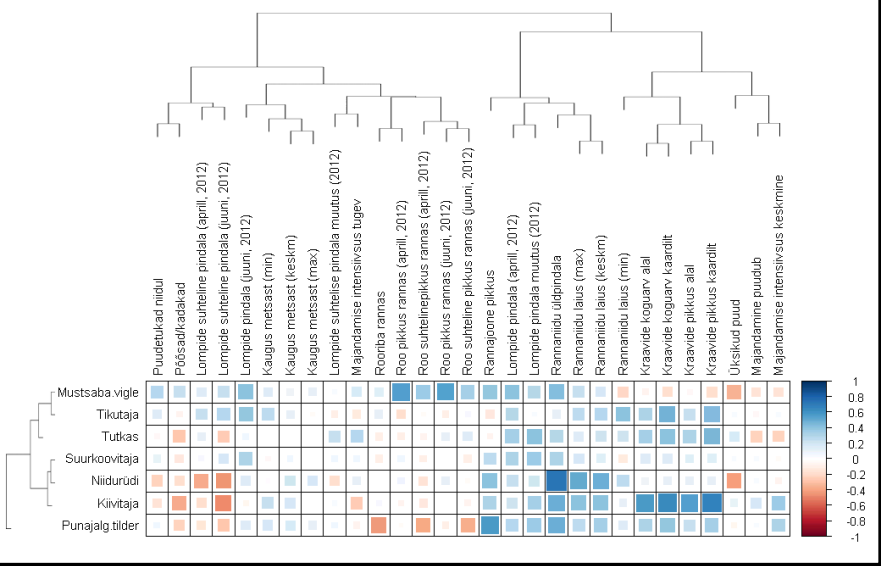
Kergejõustiku MM 2013, meeste 10-võistlus

Klasterdada võib ka andmebaasi ridu.



Klasteranalüüs tunnuste järjestamismeetodina

Eesmärgiks tuua visuaalselt selgemalt välja andmetes peituvad sarnasusmustrid.



Heatmap

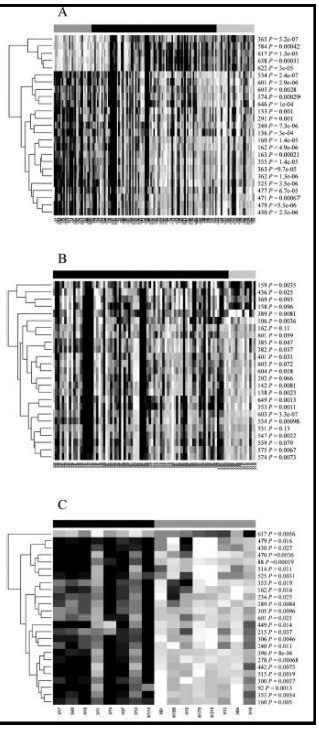


Figure 3. (A) Heat map of noncoagulating (dark gray bar on top left), intermediate (black bar on top center), and well-coagulating (light gray bar on top right) milk samples. Statistical comparison with Student *t*-test was calculated between well-coagulating and noncoagulating groups. The 25 most significantly different m/z values with respective *P*-values can be seen in the right-hand side. Rows are arranged into clusters (dendrogram on left-hand side) based on pattern similarities. The lighter the shading of a data point, the more intense the signal is compared with the mean value for the respective m/z among all analyses. Gray points are close to the mean and dark tones indicate signal intensities below the mean; (B) heat map of milks with poor (black bar) and good (gray bar) coagulation ability; (C) heat map of noncoagulating (black bar) and well-coagulating (gray bar) milks.

J. Dairy Sci. TBC:1–8
doi:10.3168/jds.2011-4468

Comparison of the metabolic profiles of noncoagulating and coagulating bovine milk

H. Harzia,*†‡[1] K. Kilk,† I. Jõudu,*‡ M. Henno,* O. Kärt,* and U. Soomets†
*Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, Kreutzwaldi 46, Tartu 51006, Estonia
†Department of Biochemistry, Medical Faculty, University of Tartu, the Centre of Excellence for Translational Medicine, Ravila 19, 50411 Tartu, Estonia
‡Bio-Competence Centre of Healthy Dairy Products, Kreutzwaldi 1, 51014 Tartu, Estonia

k-keskmiste klasterdamine

... sobib grupeerimise meetodiks siis, kui objekte on nii palju, et hierarhilise klasteranalüüsi tulemus muutub ebaülevaatlikuks, kuid ka siis kui me oskame meile sobivat klastrite arvu ligilähedaselt ennustada ning ühtlasi soovime saada ka tekkivate klastrite kirjelduse nende tunnuste osas, mis on grupeerimise aluseks.

Algoritm:

- kõigepealt tuleb määrata klastrite arv, siis
- jagada objektid esialgsetesse klastritesse,
- arvutada välja klastrite keskpunktid ning
- hakata võrdlema igat objekti kõigi klastrite keskpunktidega; kui osutub, et objekti kaugus mõne muu klastri keskpunktist on väiksem kui selle klastri keskpunktist,milles ta parasjagu asub, siis tuleb objekt teise klastrisse ümber tõsta;
- peale objekti ümbertõstmist tuleb pöörduda uuesti sammu 3 juurde ja jätkata protsessi niikaua kui kõik objektid on klastris, mille keskpunktile nad kõige lähemal asuvad.

Näide: puude klasterdamine

ID	KUIV	PE	H100	KKT	ARENGUK	D	H	A
410057	0	MA	15.9	MS	K	14	12	55
410082	0	KU	23.6	JM	V	28	23	90
410111	0	KS	16.3	TR	K	14	13	60
410151	0	MA	14.8	MS	L	12	11	55
410173	0	MA	17.5	KN	L	6	4	29
410179	0	MA	22	MS	K	26	22	100
410202	0	MA	17.2	MS	K	16	15	70
410230	0	MA	24	JM	V	26	23	85
410248	1	MA	19.5	KM	K	24	20	110
410251	1	MA	17.5	SN	N	4	5	26

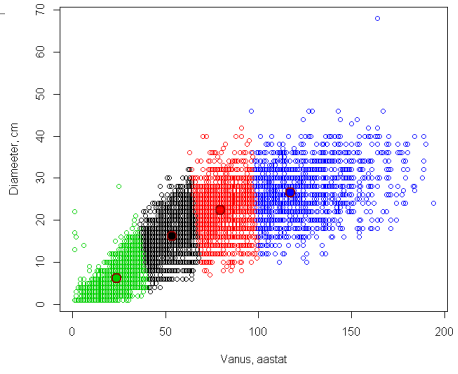
- 29539 puud,
- eesmärk jagada puud klastritesse diameetri (D) ja vanuse (A) alusel.

Näide: puude klasterdamine

```
> table(puu_kmean1$cluster) # klastrite suurused

   1    2    3    4
8964 7493 5941 3026
> # klasterdamise aluseks olnud tunnuste keskmised klastrite kaupa
> puu_kmean1$centers
       [,1]     [,2]
1  53.52443 16.21586
2  79.33298 22.34472
3  23.54385  6.14543
4 117.20787 26.56345
```

Miks tundub, et klastrid on moodustatud eelkõige puu vanuse järgi?

Vastus – puude vanust märkivad arvud (eelkõige nende varieeruvus) on suuremad, kui diameetrit märkivad arvud.
Soovides, et klasterdamisalgoritm peaks puu diameetrit sama oluliseks kui vanust, tuleb klasterdamiseks kasutatavad tunnused eelnevalt standardiseerida.

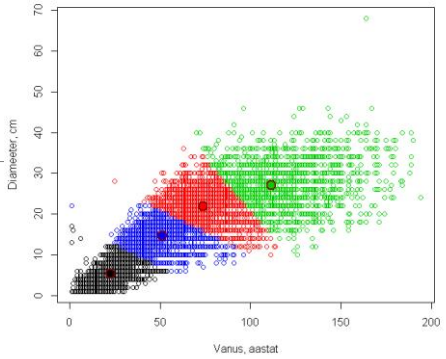


Näide: puude klasterdamine

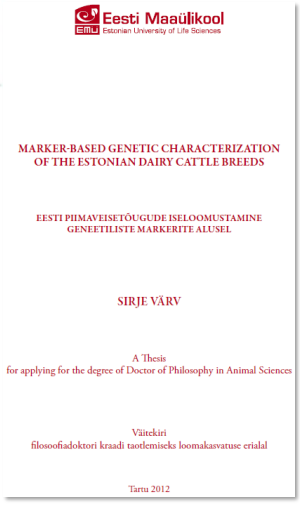
Klasterdamise tulemused peale vanuse ja diameetri standardiseerimist

```
> table(puu_kmean2$cluster) # klastrite suurused

   1    2    3    4
5346 8295 3836 7947
>
> # klasterdamise aluseks olnud tunnuste keskmised klastrite kaupa
> puu_kmean2$centers
       [,1]       [,2]
1 -1.2838316 -1.4539809
2  0.3980414  0.6267682
3  1.6430013  1.2736169
4 -0.3449028 -0.2908839
>
>  # Mis tüüpi puud mingitesse klastritesse kuuluvad?
> table(puu_kmean2$cluster, puud1$ARENGUKL)

      A    K    L    N    S    V    Y
  1   1  148 2480 2566    3  110   38
  2   0 5003    1    0    0 1592 1699
  3   0  895    0    1    3  697 2240
  4   0 5825 1529    0    1  358  234
```

Kuidas on puud klasterdunud sõltuvalt arenguklassist (A – lage, N – noorendikud, L – latimets, K – keskealised, V – valmiv, Y – küps, S – selguseta)?



Näide: Sirje Värv, veiste klasterdamine geneetiliste markerite alusel


Eesti Maaülikool
Estonian University of Life Sciences

MARKER-BASED GENETIC CHARACTERIZATION OF THE ESTONIAN DAIRY CATTLE BREEDS

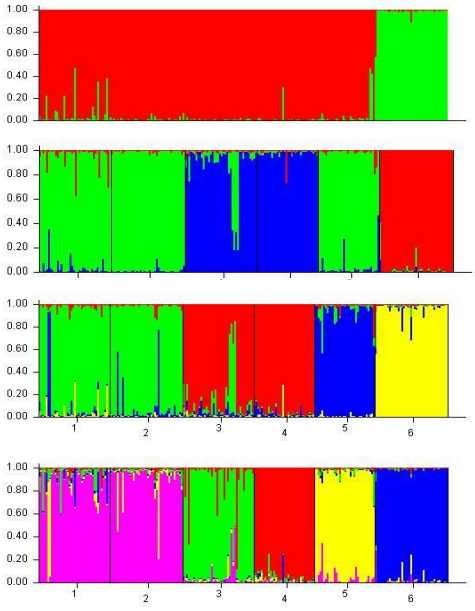
EESTI PIIMAVEISETÕUGUDE ISELOOMUSTAMINE GENEETILISTE MARKERITE ALUSEL

SIRJE VÄRV

A Thesis
for applying for the degree of Doctor of Philosophy in Animal Sciences

Väitekiri
filosoofiadoktori kraadi taotlemiseks loomakasvatuse erialal

Tartu 2012

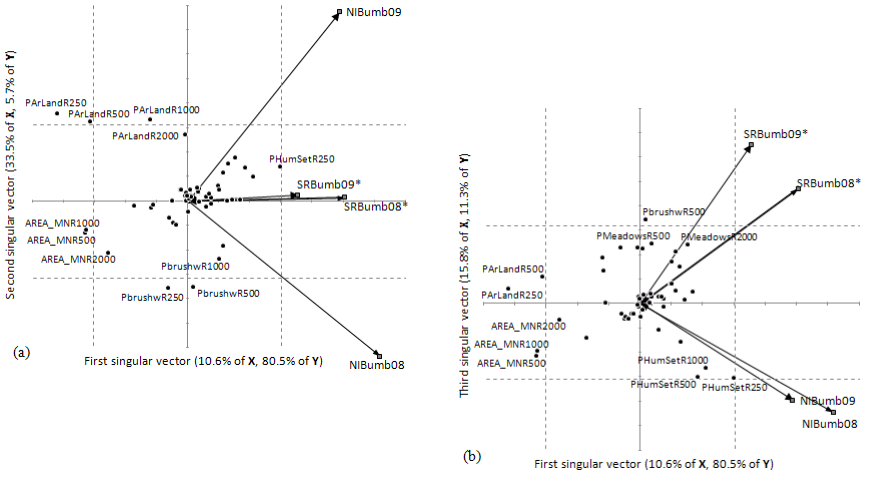


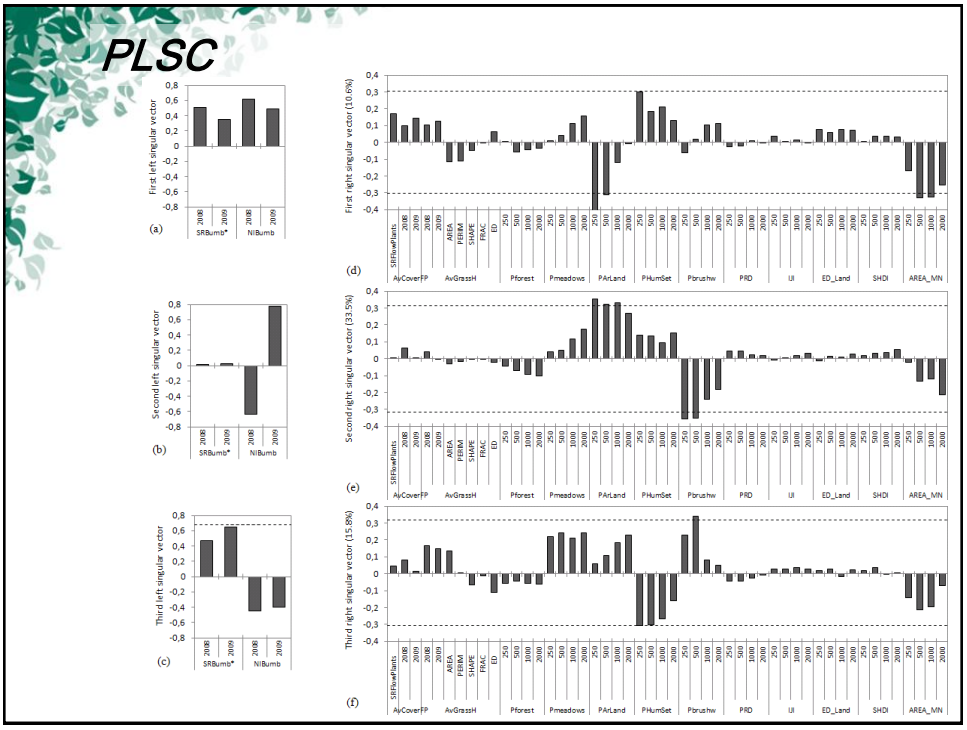
1 – EN,
2 – WFC,
3 – ER,
4 – EHF,
5 – SwRP,
6 – DkJer

Osavähimruutude regressioon ja korrelatsioon

[*Partial Least Square Regression and Corelation, Reduced Rank Regression, Principal Component Regression, …*]

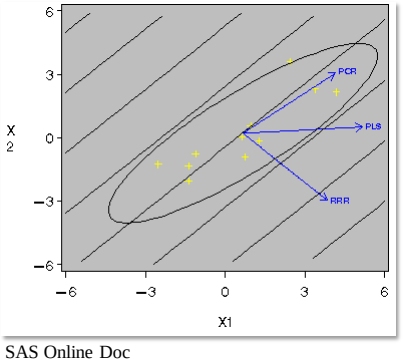
Osavähimruutude korrelatsioon (*partial least square correlation*, *PLSC*)





Mustrid (*patterns*)?

- Peakomponentide regressioon (*principal components regression*, *PCR*)
- Osavähimruutude korrelatsioon ja regressioon (*partial least squares correlation*, *PLSC*), kanooniline korrelatsioon (*canonical correlation*)
- *Reduced rank regression* (*RRR*)



SAS Online Doc

Klasteranalüüs

Klasteranalüüsi eesmärgiks on kas tunnuste või uuritavate objektide rühmitamine. Kaks peamist algoritmi:

- hierarhiline klasterdamine;
- k-keskmiste klasterdamine.

Kasutusala: kus iganes (geneetika, ökoloogia, sotsioloogia, meditsiin, majandus, …)

Klasteranalüüs

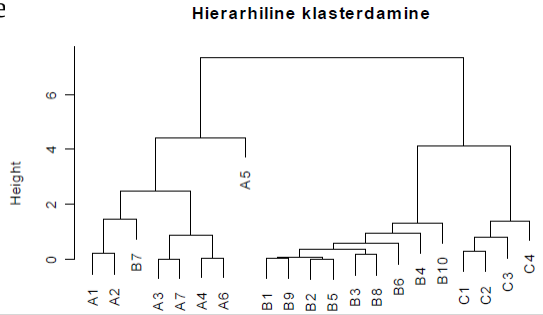
• Loodusuurija sõitis Asustamata Saarele ja avastas seal suure hulga seni teadusele tundmatuid putukaid.

• Ta mõõtis oma lühikese saarelviibimise jooksul tervel hunnikul seninägematutel putukatel igasuguseid näitajaid (igatsorti pikkuseid ja mustrielementide arvu ja paljut muudki).

• Järgmiseks soovis loodusuurija määratleda, mitmesse alamliiki leitud putukad võiksid kuuluda.

• Et saada esimest ligikaudset lähendit, kust oma uurimistööga pihta hakata, soovis ta leida sarnaste putukate rühmad – kes oleksid siis alamliikide kanditaatideks.

• Selleks söötis ta oma andmed klasteranalüüsi teostavasse programmi, mis joonistas järgmise pildi:



Märt Mölsi loengukonspektist

Klasteranalüüs

Hierarhiline klasterdamine

... on hästi kasutatav siis, kui meil on suhteliselt vähe objekte või kui on oodata, et klastrid suhteliselt selgelt üksteisest eristuvad.

Hierarhiline klasteranalüüs põhineb väga lihtsal algoritmil:

samm-sammult hakatakse omavahel kokku panema kõige sarnasemaid objekte. Näiteks, kui leidub kaks täpselt ühesuguste tulemustega objekti, siis liidetakse nad esimesel sammul üheks klastriks, peale seda võrreldakse kõiki üksikobjekte ja juba tekkinud klastreid ja liidetakse jälle kõige sarnasemad omavahel jne.

Vaatluste omavahelise kauguse määramine:

Eukleidese kaugus: $d\ (x_1, y_1), (x_2, y_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

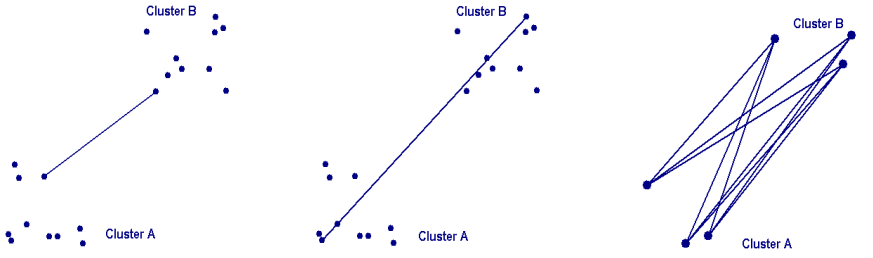
Manhattani kaugus: $d\ (x_1, y_1), (x_2, y_2) = |x_1 - x_2| + |y_1 - y_2|$

…

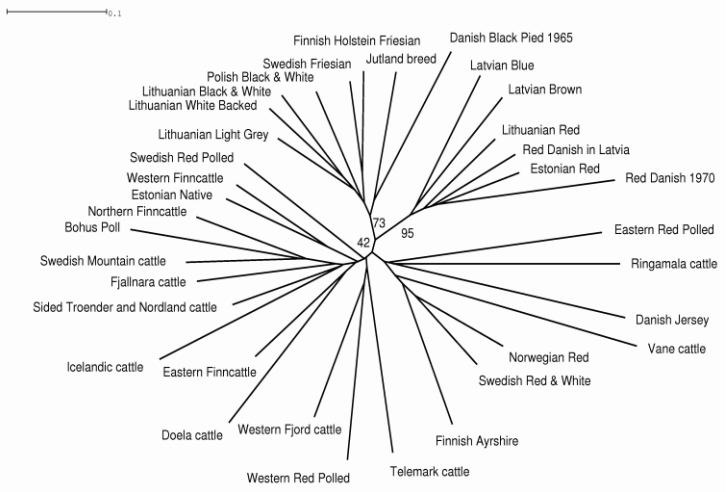
Klasteranalüüs

Klastritevahelise kauguse määramine:

“single” “complete” “average”



Hierarhiline klasterdamine – näide



Tapio et al 2006. Prioritization for Conservation of Northern European Cattle Breeds Based on Analysis of Microsatellite Data. Conservation Biology 20, 1768-1779.

Klasteranalüüs

k-keskmiste klasterdamine

... sobib grupeerimise meetodiks siis, kui objekte on nii palju, et hierarhilise klasteranalüüsi tulemus muutub ebaülevaatlikuks, kuid ka siis kui me oskame meile sobivat klastrite arvu ligilähedaselt ennustada ning ühtlasi soovime saada ka tekkivate klastrite kirjelduse nende tunnuste osas, mis on grupeerimise aluseks.

Algoritm:

• kõigepealt tuleb määrata klastrite arv, siis

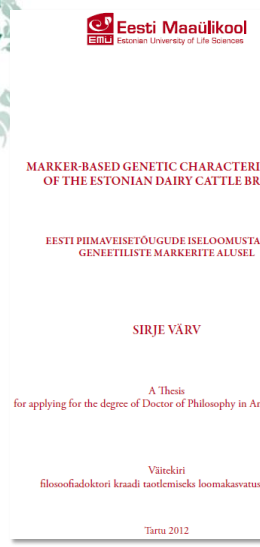
• jagada objektid esialgsetesse klastritesse,

• arvutada välja klastrite keskpunktid ning

• hakata võrdlema igat objekti kõigi klastrite keskpunktidega; kui osutub, et objekti kaugus mõne muu klastri keskpunktist on väiksem kui selle klastri keskpunktist,milles ta parasjagu asub, siis tuleb objekt teise klastrisse ümber tõsta;

• peale objekti ümbertõstmist tuleb pöörduda uuesti sammu 3 juurde ja jätkata protsessi niikaua kui kõik objektid on klastris, mille keskpunktile nad kõige lähemal asuvad.

K-keskmiste klasteranalüüs – näide


Eesti Maaülikool
Estonian University of Life Sciences

MARKER-BASED GENETIC CHARACTERIZATION OF THE ESTONIAN DAIRY CATTLE BREEDS

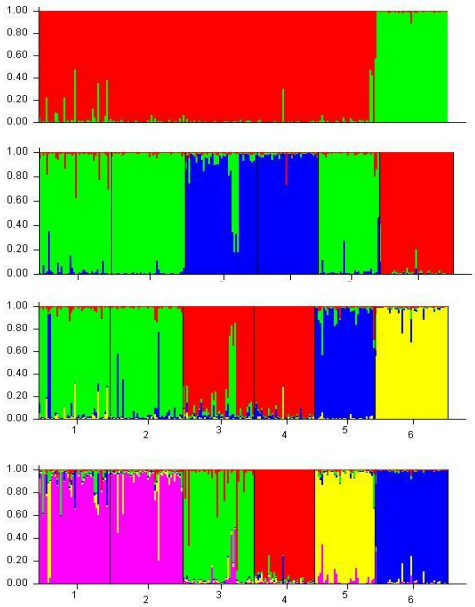
EESTI PIIMAVEISETÕUGUDE ISELOOMUSTAMINE GENEETILISTE MARKERITE ALUSEL

SIRJE VÄRV

A Thesis
for applying for the degree of Doctor of Philosophy in Animal Sciences

Väitekiri
filosoofiadoktori kraadi taotlemiseks loomakasvatuse erialal

Tartu 2012



1 – EN,
2 – WFC,
3 – ER,
4 – EHF,
5 – SwRP,
6 – DkJer

Diskriminantanalüüs

Diskriminantanalüüsi eesmärgiks on objektide rühmitamine nendel mõõdetud tunnuste alusel.

Seejuures on objektide klassidesse kuuluvus enne analüüsi teada (erinevalt peakomponent või klasteranalüüsist).

Diskriminantanalüüs

Objektidel mõõdetud tunnuste alusel kostrueeritakse nn diskremineeriv funktsioon, mis eristaks gruppe võimalikult selgelt:

$$d = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$

Näiteks
2,3×"jalgade pikkus" + 6,7×"saba pikkus" – 2,8×"kere pikkus" + 5,1×"noka pikkus"
Kui saadud väärtus <10,2, siis on ilmselt tegu isase isendiga.

Täpsemalt öeldes hinnatakse tunnuste vektori **x** tihedusfunktsioon $f_t(\mathbf{x})$ igas grupis t ning arvutatakse iga objekti mingisse gruppi kuulumise tõenäosus Bayesi valemist kujul

$$\mathrm{P}(t \mid \mathbf{x}) = \frac{q_t f_t(\mathbf{x})}{\sum_u q_u f_u(\mathbf{x})},$$

misjärel määratakse iga objekt tõenäoliseimasse gruppi
(suurus q_t eelnevas valemis on objekti gruppi t kuulumise algtõenäosus).

Diskriminantanalüüs – näide

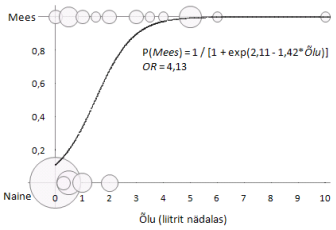
Jaanika Hõimra

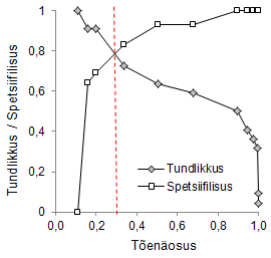
Morfomeetriliste tunnuste varieeruvuse sõltuvus keskkonnatingimustes kortslehe (*Alchemilla* L.) viiel mikroliigil eksperimendi tingimustes

Tabel 8. Klassifitseeriv diskriminantanalüüs 18 tunnuse alusel (2001). Ridades on antud empiiriliselt määratud liigid ja tulpades prognoos.

	%	KAREDA-KARVANE	VÄIKE	TERAVA-HÕLMINE	KÜÜT
KAREDAKARVANE	93,2	**262**	19	0	0
VÄIKE	99,7	1	**286**	0	0
TERAVAHÕLMINE	91,5	0	0	**259**	24
KÜÜT	87,4	0	0	36	**250**
KOKKU	92,9	**263**	**305**	**295**	**274**

Diskriminantanalüüs *vs* logistiline regressioon





Logistilise regressioonanalüüsi tulemus:
nii mehed kui ka naised identifitseeritakse õigesti 80%-lise tõenäosusega.

Et diskriminantanalüüs loeb objekti kuuluvaks suurima tõenäosusega gruppi, ei pruugi saadav klassifitseerimiseeskiri olla optimaalseim.

Diskriminantanalüüsi tulemus:
meestest identifitseeritakse õigesti 59,1%, naistest 92,9%.

Number of Observations and Percent Classified into SUGU

From SUGU	0	1	Total
0	39 92.86	3 7.14	42 100.00
1	9 40.91	13 59.09	22 100.00
Total	48 75.00	16 25.00	64 100.00

Korrespondentsanalüüs
(*correspondence analysis*)

Korrespondentsanalüüs võimaldab graafiliselt kirjeldada sagedustabelite kujul esitatud tunnuste vahelisi seoseid.

Korrespondentsanalüüs

Andmestik (juuste ja silmade värv Caith'is, Šhotimaal):

Silmavärv\juuksevärv	Blond	Punapea	Šatään	Brünett	Süsimust
Sinine	326	38	241	110	3
Hele	688	116	584	188	4
Keskmine	343	84	909	412	26
Tumedad	98	48	403	681	85

Korrespondentsanalüüsi esitus:

The CORRESP Procedure

Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	17 34 51 68 85
0.44637	0.19924	1073.33	86.56	86.56	**********
0.17346	0.03009	162.08	13.07	99.63	****
0.02932	0.00086	4.63	0.37	100.00	
Total	0.23019	1240.04	100.00		

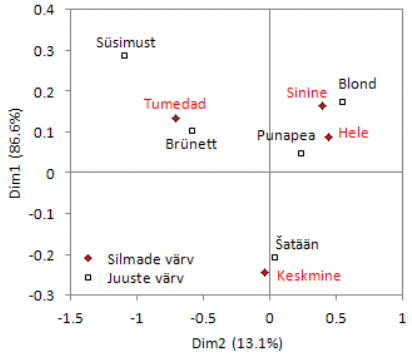
Degrees of Freedom = 12

Row Coordinates

	Dim1	Dim2
Sinine	0.4003	0.1654
Hele	0.4407	0.0885
Keskmine	-0.0336	-0.2450
Tumedad	-0.7027	0.1339

Column Coordinates

	Dim1	Dim2
Blond	0.5440	0.1738
Punapea	0.2333	0.0483
Batään	0.0420	-0.2083
Brünett	-0.5887	0.1040
Süsimust	-1.0944	0.2864

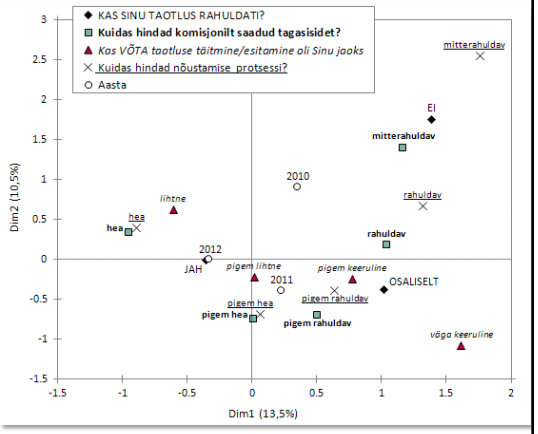


Mitmene korrespondentsanalüüs

- Mitmemõõtmeline 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 küsimuste ja vastuste omavahelisi seoseid ning tuvastada võimalikke mustreid vastanute seas.
- Projitseerimise tarvis arvutatavad nn dimensioonid on edasi kasutatavad juba arvulisi tunnuseid eeldavates analüüsides.

Mitmene korrespondentsanalüüs

VÕTA taotluse esitanud tudengite ankeedivastuste mitme-mõõtmelise korrespondents-analüüsi tulemused.



- Horisontaalsihis eristuvad tudengid, kelle VÕTA taotlus rahuldati ning kes hindasid nii taotluse esitamise protsessi lihtsaks kui ka nõustamise protsessi ja komisjonilt saadud tagasisidet heaks, ülejäänutest.
- Kõik VÕTA protsessiga seonduvad positiivsed vastused paiknevad joonise vasakus pooles ja negatiivsed vastused paremas pooles, neutraalsed vastused paiknevad vertikaaltelje lähedal.
- Vertikaalsihis eristuvad kindlad vastused (äärmused) kahtlevatest (neutraalsetest) vastustest.
- Seega kipuvad hinnangud taotluse esitamise ja nõustamise protsessile ning komisjonilt saadud tagasisidele olema sarnased, peegeldudes suurel määral ka esitatud taotluse tulemuslikkuses.
- Aastate lõikes on 2010. aastal olnud enam VÕTA protsessiga mitterahulolevaid tudengeid, 2011. aastal VÕTA protsessi rahuldavalt suhtuvaid tudengeid ja 2012. aastal VÕTA protsessi positiivselt suhtuvaid tudengeid (aastaarvud paiknevad joonise vastavais piirkondades).

Mitmene korrespondentsanalüüs

Raaperi, K., Bougeard, S., Aleksejev, A., Orro, T., Viltrop, A., 2012. ASSOCIATION OF HERD BRSV AND BHV-I SEROPREVALENCE WITH RESPIRATORY DISEASE AND REPRODUCTIVE PERFORMANCE IN ADULT DAIRY CATTLE.

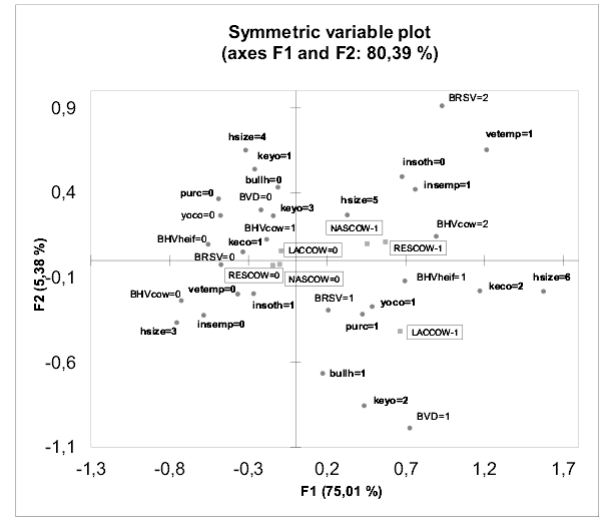


Figure 1 Graphical display of MCA for high incidence of respiratory disease symptoms in cows and pregnant heifers (100 herds).

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

Variable	Definition of the categories of the variable	Number of herds	
		Model I	Model II
Nasal discharge ("red nose") in cows and/or pregnant heifers (NASCOW)	0 - not present at all or was shown only as single cases at some point during the last two years	82	
	1 - present in more than just single cases at some point during the last two years	18	
	0 - not present at all or was shown only as single cases at some time point during the last two years	80	
	1 - present in more than just single cases at some point during the last two years	20	
	0 - not present at all or was shown only as single cases at some point during the last two years	88	
	1 - present in more than just single cases at some point during the last two years	12	
	0 - less than two respiratory disease symptoms were present in more than a single case at some time during the last two years	81	
	1 - at least two respiratory disease symptoms were present in more than a single case at some time during the last two years	19	
	0 - < 1.9% in a herd (median for cut-off value)		37
	1 - ≥ 1.9% in a herd		40
	0 - < 19 in a herd (median for cut-off value)		38
	1 - ≥1.9 in a herd		39
	3 - 20-99 cows	40	25
	4 - 100-199 cows	19	17
	5 - 200-399 cows	23	21
	6 - > 400 cows	18	14
	0 - no	77	
	1 - yes	23	
	0 - no	28	
	1 - yes	72	
	0 - no	56	
	1 - yes	44	
	0 - no	61	
	1 - yes	39	
	0 - together	51	
	1 - in separate building from 6 months until pregnancy	49	
	1 - tied	23	
	2 - loose	30	
	3 - some period of life tied, some period loose	47	
	1 - tied	71	
	2 - loose	29	
	0 - no	47	
	1 - yes	53	
	0 - no	77	57
	1 - yes (at least one animal tested positive)	23	20
	0 - negative	46	34
	1 - 1 - 49%	40	33
	2 - ≥50%	14	10
	0 - negative	37	25
	1 - 1 - 49%	28	24
	2 - ≥50%	35	28
	0 - negative	56	40
	1 - positive	44	37

Mitmemõõtmeline skaleerimine

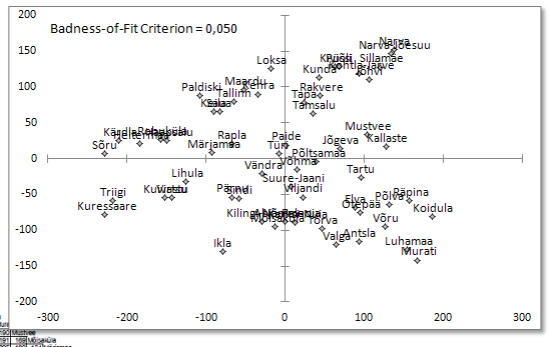
(*multidimentional scaling*)

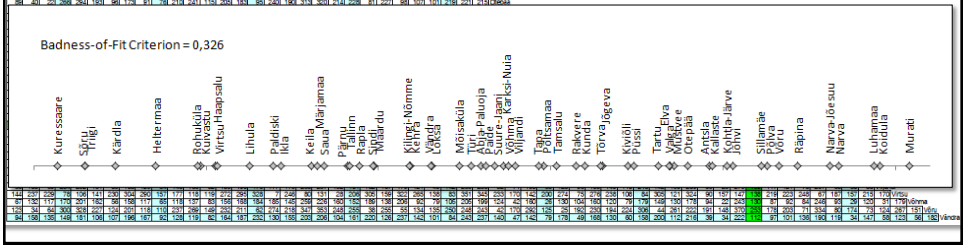
Mitmedimensionaalse skaleerimise eesmärgiks on objektide erinevuste või sarnasuste teisendamine distantsiks mitmemõõtmelises ruumis ja saadu esitamine kaardina objektide omavahelisest paiknemisest.

Kasutusala: kus iganes (geneetikas, ökoloogias, majandusteaduses, sotsioloogias, …)

Mitmedimensionaalne skaleerimine

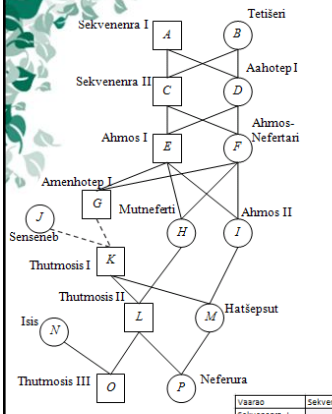
Eesti linnadevahelised kaugused

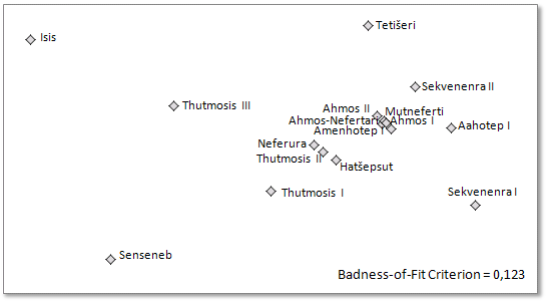




Mitmedimensionaalne skaleerimine

Egiptuse vaaraode 18. dünastia sugupuu algus





Vaarao	Sekvenenr	Tetiseri	Sekvenenr	Aahotep	Ahmos_I	Ahmos_Ne	Amenhote	Mutneferi	Ahmos_II	Senseneb	Thutmosis	Thutmosis	Hatshepsu	Isis	Thutmosis	Neferura
Sekvenenra_I	1	0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0,25	0,375	0,375	0	0,1875	0,37
Tetiseri	0	1	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0,25	0,375	0,375	0	0,1875	0,37
Sekvenenra_II	0,5	0,5	1	0,5	0,75	0,75	0,75	0,75	0,75	0	0,375	0,5625	0,5625	0	0,28125	0,562
Aahotep	0,5	0,5	0,5	1	0,75	0,75	0,75	0,75	0,75	0	0,375	0,5625	0,5625	0	0,28125	0,562
Ahmos_I	0,5	0,5	0,75	0,75	1,25	0,75	1	1	1	0	0,5	0,75	0,75	0	0,375	0,7
Ahmos_Nefertari	0,5	0,5	0,75	0,75	0,75	1,25	1	1	1	0	0,5	0,75	0,75	0	0,375	0,7
Amenhotep_I	0,5	0,5	0,75	0,75	1	1	1,375	1	1	0	0,6875	0,84375	0,84375	0	0,421875	0,843
Mutneferti	0,5	0,5	0,75	0,75	1	1	1	1,375	1	0	0,5	0,9375	0,75	0	0,46875	0,843
Ahmos_II	0,5	0,5	0,75	0,75	1	1	1	1	1,375	0	0,5	0,75	0,9375	0	0,375	0,843
Senseneb	0	0	0	0	0	0	0	0	0	1	0,5	0,25	0,25	0	0,125	0,2
Thutmosis_I	0,25	0,25	0,375	0,375	0,5	0,5	0,6875	0,5	0,5	0,5	1	0,75	0,75	0	0,375	0,7
Thutmosis_II	0,375	0,375	0,5625	0,5625	0,75	0,75	0,84375	0,9375	0,75	0,25	0,75	1,25	0,75	0	0,625	
Hatshepsut	0,375	0,375	0,5625	0,5625	0,75	0,75	0,84375	0,75	0,9375	0,25	0,75	0,75	1,25	0	0,375	
Isis	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0,5	
Thutmosis_III	0,1875	0,1875	0,28125	0,28125	0,375	0,375	0,421875	0,46875	0,375	0,125	0,375	0,625	0,375	0,5	1	0
Neferura	0,375	0,375	0,5625	0,5625	0,75	0,75	0,84375	0,84375	0,84375	0,25	0,75	1	1	0	0,5	1,37