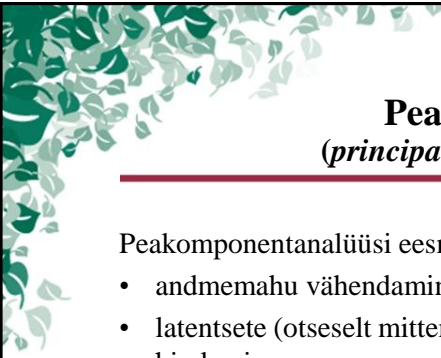




**Mitmemõõtmelise statistika  
koolitus**

**Peakomponent- ja faktoranalüüs**

Eesti Maaülikool  
20.-24. jaanuar 2014  
Tanel Kaart



**Osa 1**

**Peakomponentanalüüs**  
*(principal component analysis, PCA)*

---

Peakomponentanalüüsi eesmärgiks on

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

---

## Näide: Ulvi Moor jt, „õunauuring“.

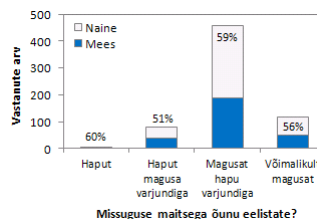
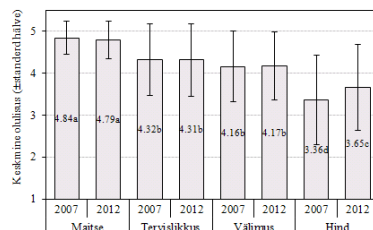
- Ankeetküsitlus aastatel 2007 ja 2012;
- kokku 668 inimest, mehed-naised, viis vanuseklassi, neli haridustaset;
- küsimused kujul
  - „Kas eelistate kodumaist või importõuna?“,
  - „Kui olulised on õuna puhul Teile kui tarbijale ...“ jne
- küsimuste vastused 5-palliskaalal (1-5)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Jrknr	Aasta	Sugu	Vanuse,Haridus	X1	X5_1	X5_2	X5_3	X5_4	X8	X9	X10	X11	X13	X14	
2	201	2007	2	4	4	5	2	4	5	5	5	3	3	4	3	3
3	202	2007	2	1	2	3		4	5	4	4	4	3	3	1	1
4	203	2007	2	4	4	4	4	4	5	5	5	4	3	4	3	3
5	204	2007	2	3	3	4	3	4	4	2	4	2	3	4	4	2
6	205	2007	2	4	4	4	5	4	5	5	4	2	3	4	2	2
7	206	2007	2	3	4	4	4	4	5	5	2	3	1	4	3	2
8	207	2007	2	2	4	4	3	4	5	5	4	4	3	2	4	3
9	208	2007	2	4	4	5	4	4	5	5	4	3	2	4	3	4
10	209	2007	2	4	4	4	2	4	5	2	2	3	3	5	4	2
11	210	2007	2	4	4	4	2	4	5	2	2	3	3	4	1	2

## „Õunauuring“ – kuidas uurida seoseid ja erinevusi?

- Tavaline lahendus:
  - sagedustabelid ja vastavad joonised

### ◦ keskmised

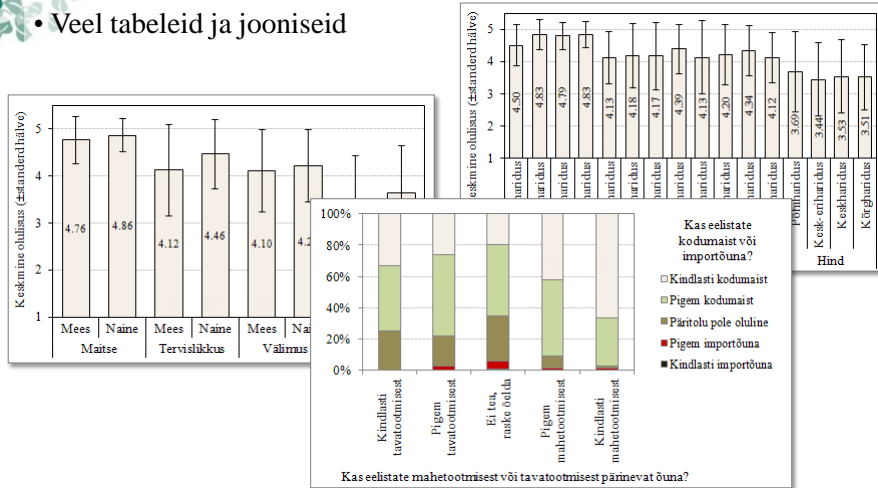


### ◦ korrelatsioonid

	Kodumaisus	Hind	Välimus	Maitse	Tervislikkus	Maheus	Magusus
Kodumaisus	1						
Hind	0.0008808	1					
Välimus	-0.0786273	0.1235804	1				
Maitse	0.0185697	-0.0942401	0.2282795	1			
Tervislikkus	0.3669458	0.1049978	0.0228645	0.1102184	1		
Maheus	0.3176403	-0.0465162	0.0144182	0.0326808	0.2969469	1	
Magusus	-0.0318697	0.0810915	0.1678925	0.1122004	-0.0066489	-0.0069751	1

## „Õunauuring“ – kuidas uurida seoseid ja erinevusi?

- Veel tabelleid ja jooniseid

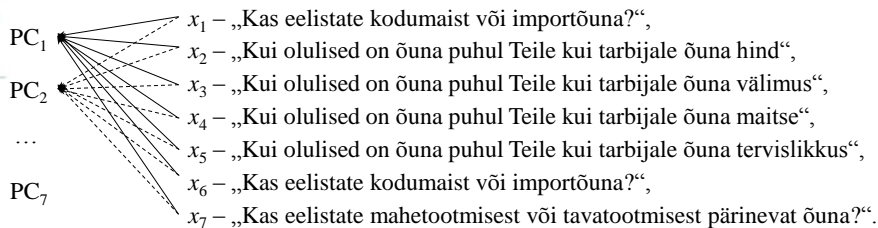


Probleem:

erinevate tabelite ja jooniste arv väga suur, mistõttu on raske aduda terviklikku kompleksset pilti ja olulisemaid seaduspärasid!

## „Õunauuring“ – peakomponentanalüüs (PCA)

- Eesmärgiks leida algsete tunnuste lineaarkombinatsioonid, mis kirjeldaksid võimalikult hästi ära algseis tunnustes sisalduva info:



- Taolisi lineaarkombinatsioone nimetatakse **peakomponentideks** või **faktoriteks** (*principal components, factors*):

$$PC_1 = b_{11}x_1 + b_{21}x_2 + \dots + b_{71}x_7$$

$$PC_2 = b_{12}x_1 + b_{22}x_2 + \dots + b_{72}x_7$$

$$\dots$$

$$PC_7 = b_{17}x_1 + b_{27}x_2 + \dots + b_{77}x_7$$

- Peakomponentanalüüs moodustab lineaarkombinatsioonid nii, et

- esimene peakomponent kirjeldab ära võimalikult suure osa kõigi alg tunnuste koguvarieeruvusest,
- teine võimalikult suure osa alles jäänud varieeruvusest, jne.

## „Õunauuring“ – peakomponentanalüüs

- Matemaatiliselt baseerub peakomponentanalüüs uuritavate tunnuste korrelatsiooni- või kovariatsioonimaatriksi **A** omaväärtusanalüüsil:
  - lahendatakse maatriksvõrrand  $(A - \lambda I)\mathbf{b} = \mathbf{0}$ , kus
  - $\lambda$  on maatriksi **A** **omaväärtus** (*eigenvalue*), mis peakomponentanalüüsi kontekstis näitab faktori poolt kirjeldatavat varieeruvust,
  - vektor **b** on korrelatsioonimaatriksi **omavektor** (*eigenvector*),
    - mis peakomponentanalüüsi kontekstis sisaldab korrelatsioonikordajaid algsete tunnuste ja moodustatud peakomponendi vahel ja
    - mille elemendid on ka algsete tunnuste (standardiseeritud) väärtuste kordajaks peakomponendi väärtust määravas võrrandis – nn **faktorlaadungid** (*factor loadings*);
    - kõik omavektorid on omavahel ortogonaalsed, millest tuleneb ka, et vastavad peakomponendid on sõltumatud (nende vaheline korrelatsioon on null).

## „Õunauuring“ – peakomponentanalüüs

Andmed

Aasta	Sugu	Vanus	Hariidus	KodumaisHind	Valimus	Maitse	Tervislikk	Mahesus	Magusus
2007	Naine	36-45	Kõrghariidus	5	2	4	5	5	5
2007	Naine	36-45	Kõrghariidus	4	4	4	5	5	4
2007	Naine	31-35	Keskhariidus	4	3	4	4	2	4
2007	Naine	36-45	Kõrghariidus	4	5	4	5	5	4
2007	Naine	31-35	Kõrghariidus	4	4	4	5	5	2
2007	Naine	26-30	Kõrghariidus	4	3	4	5	5	4
2007	Naine	36-45	Kõrghariidus	5	4	4	5	5	4

Korrelatsiooni-  
maatriks

Correlation Matrix

	Kodumaisus	Hind	Valimus	Maitse	Tervislikkus	Mahesus	Magusus
Kodumaisus	1.0000	0.0024	-0.1168	0.0390	0.3672	0.3150	-0.0336
Hind	0.0024	1.0000	0.0969	-0.0955	0.0923	-0.0532	0.0803
Valimus	-0.1168	0.0969	1.0000	0.2488	-0.0026	0.0009	0.1644
Maitse	0.0390	-0.0955	0.2488	1.0000	0.1175	0.0463	0.1065
Tervislikkus	0.3672	0.0923	-0.0026	0.1175	1.0000	0.3010	-0.0253
Mahesus	0.3150	-0.0532	0.0009	0.0463	0.3010	1.0000	-0.0137
Magusus	-0.0336	0.0803	0.1644	0.1065	-0.0253	-0.0137	1.0000

Omaväärtused

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	1.68647373	0.31979660	0.2409	0.2409
2	1.36667713	0.27516894	0.1952	0.4362
3	1.09150819	0.21761178	0.1559	0.5921
4	0.87389641	0.11057473	0.1248	0.7169
5	0.76332168	0.13469642	0.1090	0.8260
6	0.62862526	0.03912765	0.0898	0.9158
7	0.58949761		0.0842	1.0000

Omavektorid

Eigenvectors

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Kodumaisus	0.586188	-0.102402	0.049232	0.130467	-0.232729	-0.646768	0.392403
Hind	0.003733	0.150112	0.853858	-0.279199	-0.090420	0.223958	0.334805
Valimus	-0.064445	0.651881	0.012185	-0.320255	0.466492	-0.473779	-0.142776
Maitse	0.138294	0.550956	-0.441186	-0.227814	-0.451197	0.329749	0.344188
Tervislikkus	0.582203	0.080325	0.150378	-0.173669	-0.237079	0.136212	-0.725984
Mahesus	0.539543	0.009267	-0.085812	0.152815	0.669466	0.415805	0.238774
Magusus	-0.054988	0.481603	0.209123	0.834896	-0.109897	0.034131	-0.104925

## „Õunauuring“ – peakomponentanalüüs

### Omväärtused

	Eigenvalue	Difference	Proportion	Cumulative
1	1.68647373	0.31979660	0.2409	0.2409
2	1.36667713	0.27516894	0.1952	0.4362
3	1.09150819	0.21761178	0.1559	0.5921
4	0.87389641	0.11057473	0.1248	0.7169
5	0.76332168	0.13469642	0.1090	0.8260
6	0.62862526	0.03912765	0.0898	0.9158
7	0.58949761		0.0842	1.0000

- Korrelatsioonimaatriksi **omaväärtused** näitavad peakomponendi poolt kirjeldatavat varieeruvust ja selle osa koguvareeruvusest.
- Et algsed tunnused analüüsi eel **standardiseeritakse**, on nende dispersioon võrdne ühega ja sestap omavad üksiktunnustest suuremat kirjeldusvõimet vaid peakomponendid, millele vastav omaväärtus > 1.

## „Õunauuring“ – peakomponentanalüüs

### Omväärtused

- Erinevad programmid väljastavad tulemusi erineval kujul!

	Eigenvalue	Difference	Proportion	Cumulative
1	1.68647373	0.31979660	0.2409	0.2409
2	1.36667713	0.27516894	0.1952	0.4362
3	1.09150819	0.21761178	0.1559	0.5921
4	0.87389641	0.11057473	0.1248	0.7169
5	0.76332168	0.13469642	0.1090	0.8260
6	0.62862526	0.03912765	0.0898	0.9158
7	0.58949761		0.0842	1.0000

```

> apple.pc.cor <- princomp(appledata[,c(-1:-4)], cor=TRUE, scores=TRUE)
> summary(apple.pc.cor)
Importance of components:
      Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7
Standard deviation  1.2960971  1.1713805  1.0450556  0.9372917  0.8733405  0.78392800  0.7746460
Proportion of Variance  0.2399811  0.1960189  0.1560202  0.1255022  0.1089605  0.08779187  0.0857252
Cumulative Proportion  0.2399811  0.4360000  0.5920202  0.7175224  0.8264829  0.91427480  1.0000000
> apple.pc.cor$sdev^2 # peakomponentide dispersioonid (omaväärtused)
      Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7
1.6798677  1.3721324  1.0921411  0.8785157  0.7627236  0.6145431  0.6000764
    
```

SAS-i protseduur *princomp*

```

> apple.pc.cor2 = dudi.pca(appledata[,c(-1:-4)], scale=T, scannf=F)
> summary(apple.pc.cor2)
Class: pca dudi
Call: dudi.pca(df = appledata[, c(-1:-4)], scale = T, scannf = F)

Total inertia: 7

Eigenvalues:
  Ax1    Ax2    Ax3    Ax4    Ax5
 1.6799  1.3721  1.0921  0.8785  0.7627

Projected inertia (%):
  Ax1    Ax2    Ax3    Ax4    Ax5
 24.00  19.60  15.60  12.55  10.90

Cumulative projected inertia (%):
  Ax1  Ax1:2  Ax1:3  Ax1:4  Ax1:5
 24.00  43.60  59.20  71.75  82.65
    
```

R-i funktsioon *princomp*

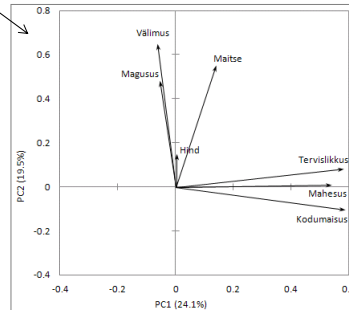
R-i funktsioon *dudi.pca*

## „Õunauuring“ – peakomponentanalüüs

### Omavektorid

	Eigenvectors						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Kodumaisus	0.586188	-0.102402	0.049232	0.130467	-0.232729	-0.646768	0.392403
Hind	0.003733	0.150112	0.853858	-0.279199	-0.090420	0.223958	0.334805
Välimus	-0.064445	0.651881	0.012185	-0.320255	0.466492	-0.473779	-0.142776
Maitse	0.138294	0.550956	-0.441186	-0.227814	-0.451197	0.323749	0.344188
Tervislikkus	0.582203	0.080325	0.150378	-0.173669	-0.237079	0.138212	-0.725984
Mahesus	0.539543	0.009267	-0.085812	0.152815	0.669466	0.415805	0.238774
Magusus	-0.054988	-0.481603	0.209123	0.834896	-0.109897	0.034131	-0.104925

- Korrelatsioonimaatriksi **omavektorid** (*eigenvectors*) kujutavad enesest korrelatsioonikordajaid algsete tunnuste ja moodustatud peakomponentide vahel.
- Samuti on omavektori elemendid algsete tunnuste standardiseeritud väärtuste kordajajaks peakomponenti väärtust määravas võrrandis – nn **faktorlaadungid** (*factor loadings*).



Näiteks avaldub esimese peakomponenti väärtus kujul

$$PC_1 = 0,586 * st(Kodum.) + 0,004 * st(Hind) - 0,064 * st(Välimus) + \dots,$$

siin st() tähendab standardiseeritud tunnust.

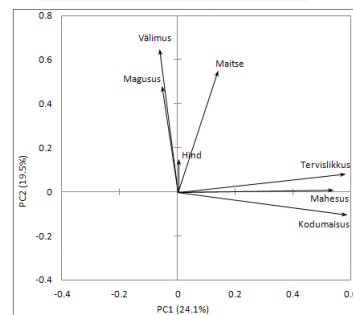
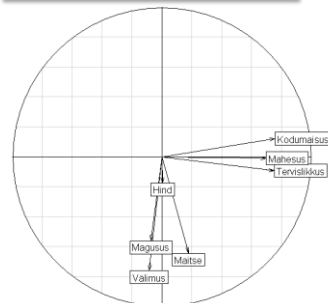
## „Õunauuring“ – peakomponentanalüüs

### Omavektorid

- Erinevad programmid võivad anda tulemuseks vastupidise märgiga omavektorid!
- Näiteks antud juhul  $PC2_R = -PC2_{SAS}$ .

```
R: > apple.pc.cor2$c1
      CS1      CS2
Kodumaisus 0.5861880488 0.1024020777
Hind        0.003732834 -0.1501116355
Välimus    -0.064444901 -0.5518807222
Maitse     0.138293607 -0.5509556677
Tervislikkus 0.582203193 -0.0803254600
Mahesus    0.539542991 -0.0092668799
Magusus    -0.054987774 -0.4816031922
```

```
SAS:
      PC1      PC2
Kodumaisus 0.586188 -0.102402
Hind        0.003733 0.150112
Välimus    -0.064445 0.651881
Maitse     0.138294 0.550956
Tervislikkus 0.582203 0.080325
Mahesus    0.539543 0.009267
Magusus    -0.054988 0.481603
```



## „Õunauuring“ – peakomponentanalüüs

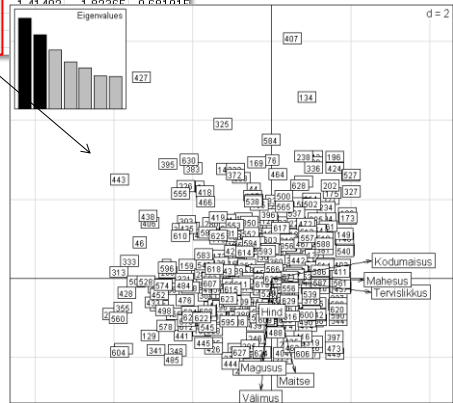
### Skoorid

- Peakomponentide väärtuseid nimetatakse ka **skoorideks** (*scores*).

Skoorid on välja arvatavad iga andmebaasi rea tarvis:

Kodum	Hind	Vaiumus	Maitse	Tervisli	Mahe	Magus	PCA1	PCA2	PCA3	PCA4	PCA5
5	2	4	5	5	5	3	1.901704	0.264187	0.399534	-0.32087	-0.52952
3	4	4	5	4	4	4	0.771246	1.260386	0.168522	1.385335	-0.0863
4	4	4	5	5	5	4	-0.38148	-3.14741	-0.01897	-0.00201	1.475698
4	4	4	4	2	4	2	-0.38591	0.426845	-1.41409	-1.69265	0.661015
4	5	4	5	5	4	2	-0.78525	-0.05342			
4	4	4	5	5	2	3	1.011838	0.40264			
4	3	4	5	5	4	4	1.666824	0.256119			

- NB! Kui mõned uurimisobjektid on mõne algse tunnuse väärtus puudu, jääb taoline objekt vaikimisi analüüsist välja (st, et ei arvestata ka tema teisi, olemasolevaid väärtuseid).

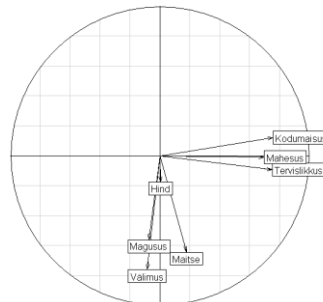
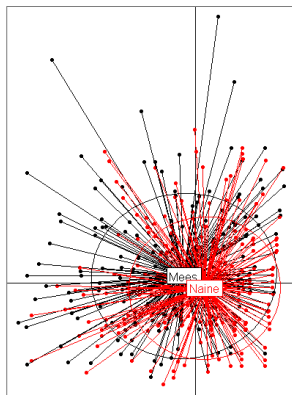


## „Õunauuring“ – peakomponentanalüüs

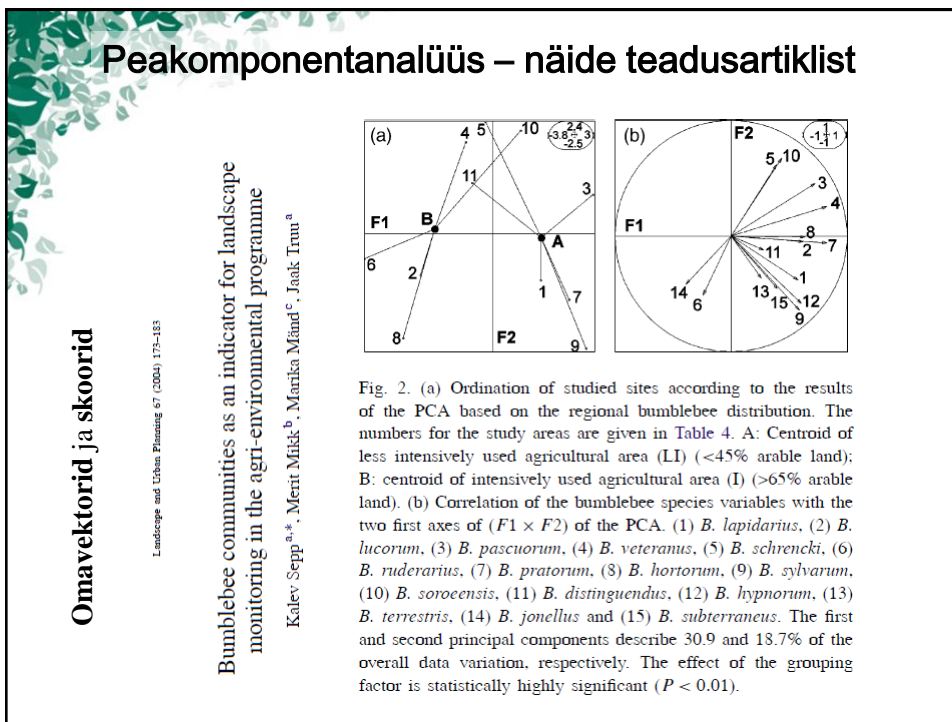
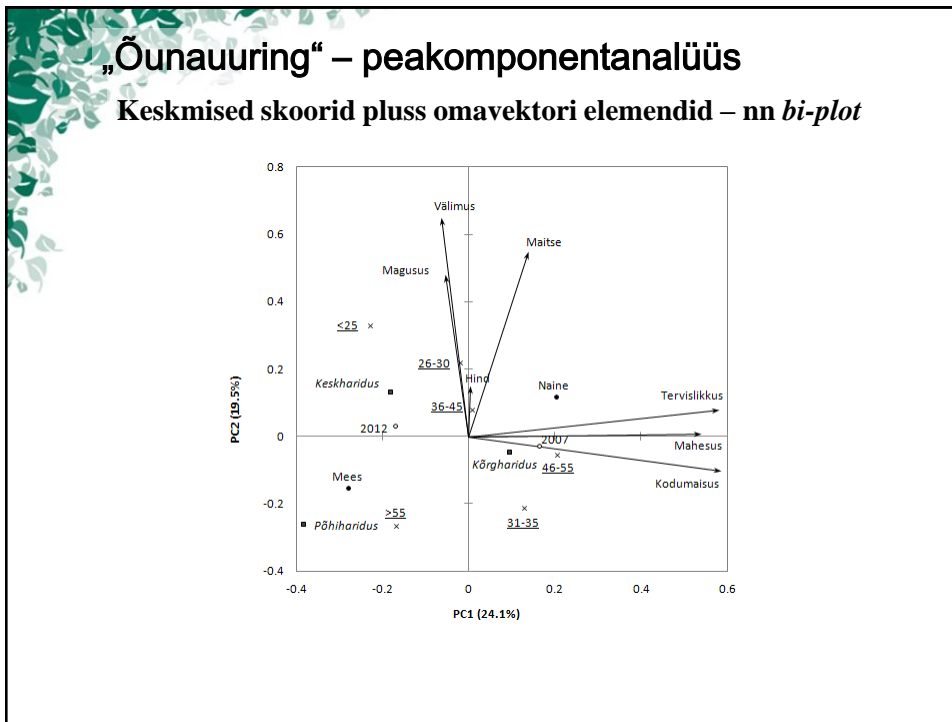
### Skoorid

- Peakomponentide väärtused e skoorid võimaldavad peakomponente seostada teiste andmebaasi kuuluvate tunnustega ja teha seeläbi üldisemaid järeldusi andmetes leiduvate mustrite/struktuuride kohta.

Samas ei oma peakomponentide väärtused mingit sisulist tähendust arvestamata nende seotust algsete tunnustega.





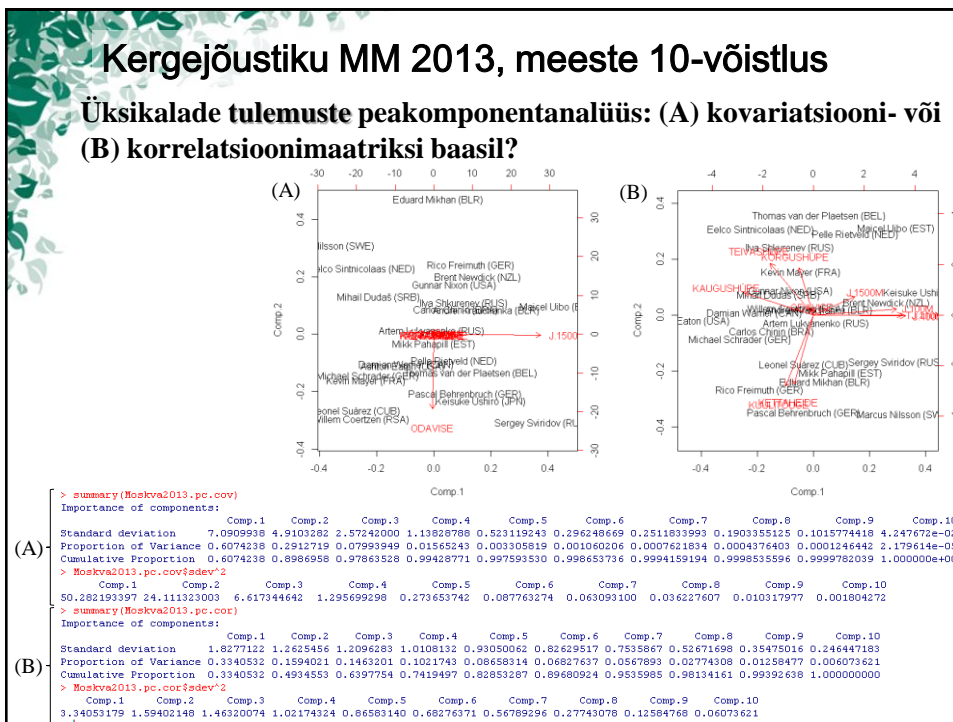






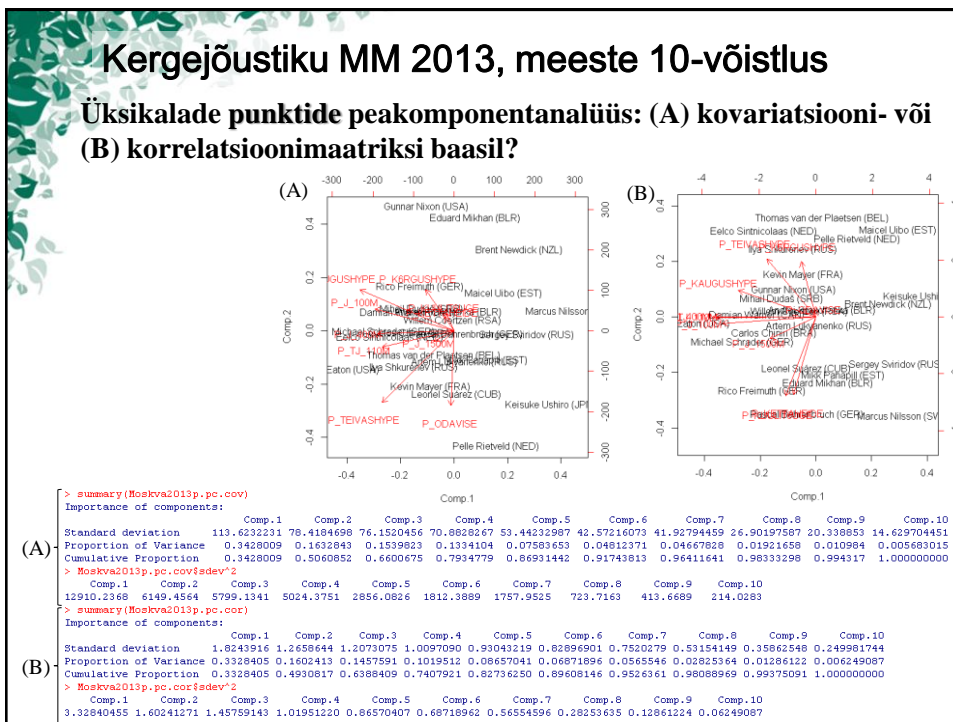
## Kergejõustiku MM 2013, meeste 10-võistlus

Üksikalade tulemuste peakomponentanalüüs: (A) kovariatsiooni- või (B) korrelatsioonimaatriksi baasil?



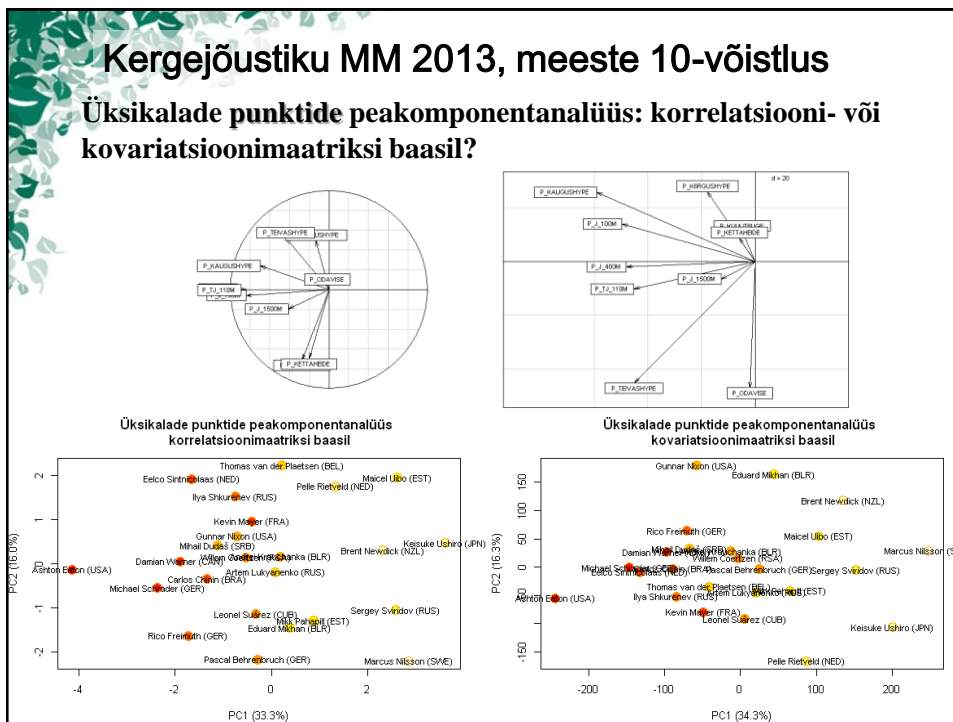
## Kergejõustiku MM 2013, meeste 10-võistlus

Üksikalade punktide peakomponentanalüüs: (A) kovariatsiooni- või (B) korrelatsioonimaatriksi baasil?

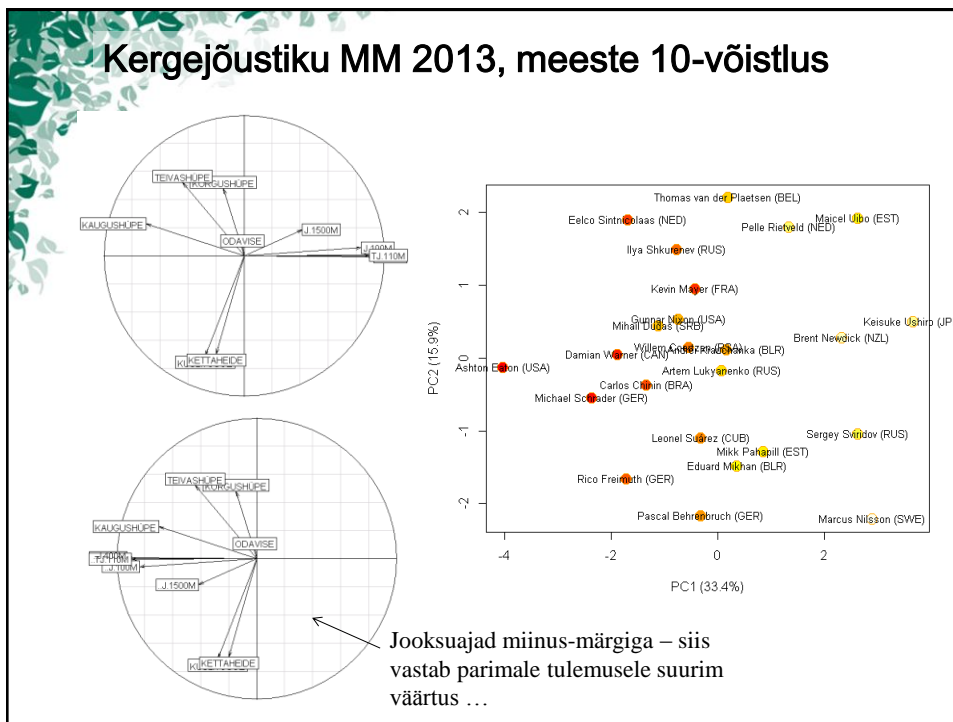


## Kergejõustiku MM 2013, meeste 10-võistlus

Üksikalade punktide peakomponentanalüüs: korrelatsiooni- või kovariatsioonimaatriksi baasil?

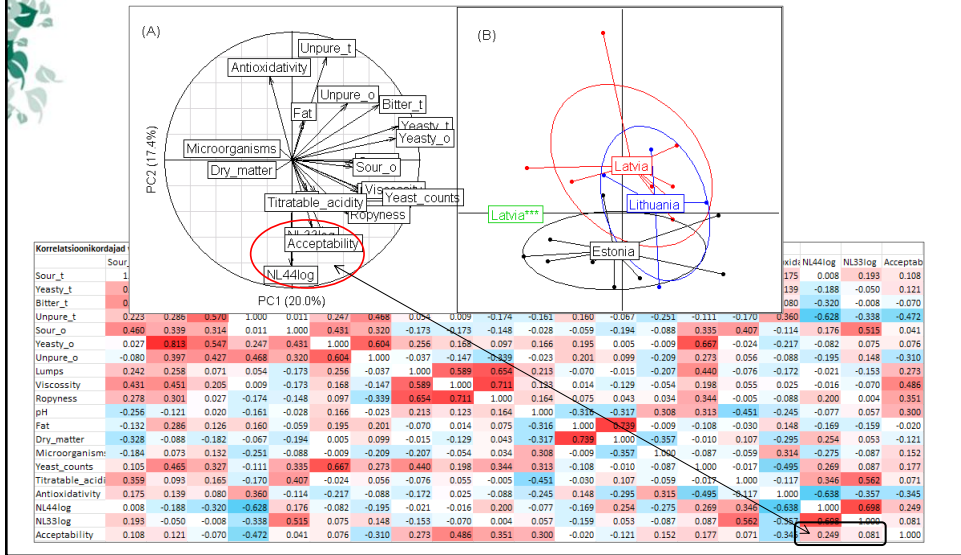


## Kergejõustiku MM 2013, meeste 10-võistlus



### Näide: Piret Raudsepp jt, „keefiride uuring“.

Peakomponentanalüüsi tulemus ei pruugi ühtida korrelatsioonanalüüsi tulemustega!



### Näide: Henn Raave, „heintaimede saagikus“.

Peakomponentide väärtuste (e skooride) kasutamine modelleerimisel (näiteks *principal component regression*)

N_norm	Niidete_arv	Katseaaast	Taimiku_v	Sademed	Sademetes	Temp	Sademed_Kerahein	Aasnurmil	Teised	Kogusaak
0	5	1971	1	219.5	12.2	13.9	15.8	0.97	0.04	0.08
0	5	1972	2	316.5	14.5	15.3	20.6	0.23	0.38	0.58
0	5	1973	3	340.1	18.9	14.1	24.1	0.38	0.32	0.78
0	5	1974	4	291.4	14.8	13.3	22	0.43	0.71	0.02
0	5	1975	5	219.2	12.3	15.1	14.5	0.44	0.68	0
0	5	1976	6	231.8	13.5	12.7	18.3	0.24	0.3	0.13
0	5	1977	7	284.9	18.5	13.2	21.5	0.16	0.12	0.14
0	5	1978	8	522.5	20.2	13.8	41.8	0.46	0.6	0.06

```
lm(formula = Aasnurmikas ~ N_norm + Niidete_arv + Taimiku_vanus,
data = saagikus)

Residuals:
Min      1Q  Median      3Q      Max
-4.9784 -0.9711 -0.0953  0.8703  4.8951

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5093323  0.3808963  1.337  0.1819
N_norm       0.0075063  0.0004018  18.681 <2e-16 ***
Niidete_arv -0.0159360  0.0803638  -0.198  0.8429
Taimiku_vanus 0.0305137  0.0096053  3.177  0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.619 on 402 degrees of freedom
Multiple R-squared:  0.4718,    Adjusted R-squared:  0.4679
F-statistic: 119.7 on 3 and 402 DF,  p-value: < 2.2e-16
```

Aga ilmastik?

## Näide: Henn Raave, „heintaimede saagikus“.

```
lm(formula = Aasnurmikas ~ N_norm + Niidete_arv + Taimiku_vanus +
Temp + Sademed + Sademed_temp + Sademete_standardhälve, data = saagikus)
```

Residuals:

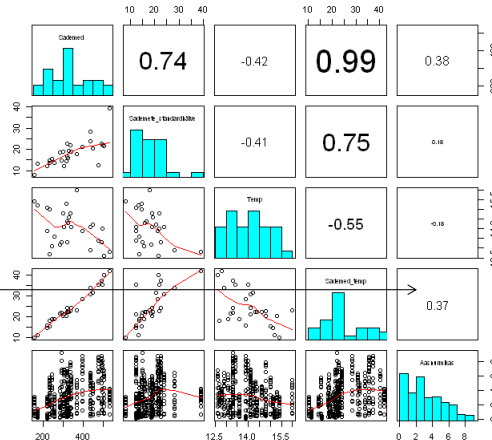
Min	1Q	Median	3Q	Max
-4.1404	-0.8614	-0.0436	0.8641	4.8396

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	12.5231821	3.1827921	3.93	0.0000000 ***
N_norm	0.0075063	0.0003282	22.87	<.0000000 ***
Niidete_arv	-0.0159360	0.0656346	-0.24	0.8140000
Taimiku_vanus	0.0274527	0.0082193	3.34	0.0008000 ***
Temp	-1.0140230	0.2266325	-4.47	<.0000000 ***
Sademed	0.0466797	0.0093545	4.99	<.0000000 ***
Sademed_temp	-0.4948229	0.1298193	-3.81	0.0002000 ***
Sademete_standardhälve	-0.0794528	0.0169599	-4.68	<.0000000 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'  
Residual standard error: 1.323 on 398 degrees of freedom  
Multiple R-squared: 0.6512, Adjusted R-squared: 0.6487  
F-statistic: 106.1 on 7 and 398 DF, p-value: < 2.2e-16

Aga ilmastik?



Ongi need mõjud negatiivsed?

Korrelatsioonanalüüs väidab vastupidist!

Miks siis mudelis teistpidine seos tuleb?

Vastus: multikollinearsus!

## Näide: Henn Raave, „heintaimede saagikus“.

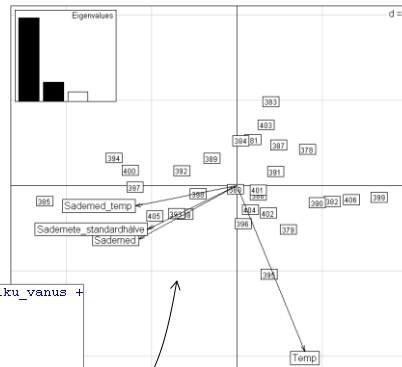
Mida teha?

- Mitte pöörata tähelepanu ilmastikunäitajate mõju olemusele – eesmärgiks ei ole hinnata ilmastiku mõju, vaid elimineerida ilmastiku mõju tegelikult huvi pakkuvate faktorite mõjust ja/või hinnata, kuivõrd kogu faktorite komplekt saagikust üldiselt prognoosida võimaldab.
- Valida välja vaid üks ilmastikunäitaja ja konstrueerida mudel sellega või konstrueerida mitu alternatiivset mudelit, igas erinev ilmastikunäitaja.
- Rakendada omavahel seoses olevatele ilmastikunäitajatele peakomponentanalüüsi ning kasutada saagikust prognoosivas mudelis esimest (ja vajadusel ka teist-kolmandat) peakomponenti; järeldused ilmastikunäitajate mõjust saagikusele tulenevad siis peakomponenti seoses algsete tunnustega „vaadatuna läbi“ peakomponentidele vastava regressioonikordaja mudelis.

## Näide: Henn Raave, „heintaimede saagikus“.

```
> pca.saagikus$co
              Comp1      Comp2
Sademed      -0.9385820 -0.24593248
Sademete_standardhälve -0.8550646 -0.20018672
Temp          0.6474827 -0.76170716
Sademed_temp -0.9739194 -0.09363388

> pca.saagikus$eig
[1] 2.979824467 0.689522617 0.328559334 0.002093582
> 100*pca.saagikus$eig/sum(pca.saagikus$eig)
[1] 74.49561167 17.23806544  8.21398335  0.05233955
```



```
lm(formula = Aasnurmikas ~ N_norm + Niidete_arv + Taimiku_vanus +
    PCA1, data = saagikus)

Residuals:
    Min       1Q   Median       3Q      Max
-4.3101 -0.8323 -0.0670  0.7625  4.9816

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.6021396   0.3411591   1.765  0.07833 .
N_norm       0.0075063   0.0003598  20.864 < 2e-16 ***
Niidete_arv -0.0159360   0.0719533  -0.221  0.82483
Taimiku_vanus 0.0243266   0.0086222   2.821  0.00502 **
PCA1        -0.4188815   0.0417899  -10.024 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.45 on 401 degrees of freedom
Multiple R-squared:  0.5776,    Adjusted R-squared:  0.5734
F-statistic: 137.1 on 4 and 401 DF,  p-value: < 2.2e-16
```

Mida rohkem sademeid, suurem sademete varieeruvus, madalam temperatuur ja suurem sademete temperatuuri suhe, seda kõrgem saagikus.

## Alterations in milk metabolome and coagulation ability during the lactation of dairy cows

H. Harzia,<sup>††</sup> A. Ilves,<sup>††</sup> M. Ots,<sup>††</sup> M. Henno,<sup>†</sup> I. Jõudu,<sup>†</sup> T. Kaart,<sup>†</sup> K. Ling,<sup>†</sup> O. Kärt,<sup>†</sup> K. Klík,<sup>†</sup> and U. Soomets<sup>‡</sup>

J. Dairy Sci. 96:6440–6448  
<http://dx.doi.org/10.3168/jds.2013-6808>  
 © American Dairy Science Association, 2013.

## Näide: Hedi Harzia jt, piima metabooloomi muutus laktatsiooni kestel

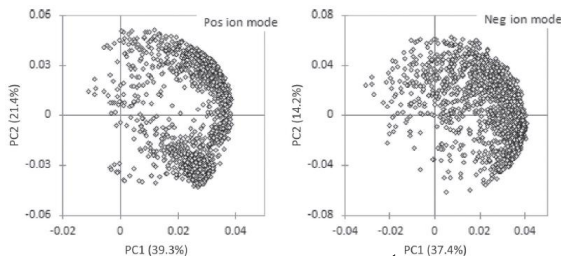


Figure 3. Loading plots of principal component (PC) analyses and their account of total variances of MS signals measured in positive (pos) and negative (neg) ion mode.

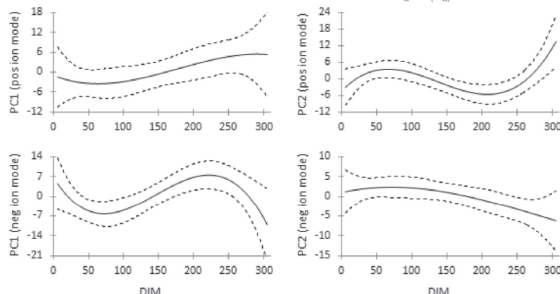


Figure 2. Results of principal component (PC) analyses of mass/charge ratio ( $m/z$ ) in positive (pos) and negative (neg) ion mode. Change in the first 2 principal component values (with 95% CI) during lactation estimated with third-order polynomial according to the linear mixed model considering additionally fixed effects of breed, parity (prim- and multiparous), diet and milking time (morning and evening) and random effect of cow are presented.

$$y_{ijklm} = \mu + B_i + P_j + D_k + M_l + b_1 \times LDIM_{ijklm} + b_2 \times LDIM_{ijklm}^2 + b_3 \times LDIM_{ijklm}^3 + C_m + \epsilon_{ijklm}$$


## Osa 2

### Faktoranalüüs (factor analysis, FA)

Faktoranalüüsi eesmärgiks on eelkõige

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

Peakomponentanalüüs on faktoranalüüsi esimene samm.



### Näide: Kaie Metsaots, „Ida-Viru küsitlus“.

Vastaja	Venus	Sugu	Rahvus	Häiduste	funktsioon	kaevandu	informatsioon	kasutusel	ehitised	renoveerim	osalus	ühitekuulu	koostöö	haldusreform	regionaal	toetusrahe	tuntus	turismipot	Venemaa	meelale
Respondent	1	2	1	3	5	3	4	5	2	4	4	5	4	4	3	3	4	5	1	
Respondent	2	2	1	5	5	5	5	5	5	5	5	5	5	3	5	5	5	5	3	
Respondent	2	2	1	5	5	5	5	5	3	4	3	5	4	1	2	5	5	5	3	
Respondent	2	1	1	2	5	3	4	5	5	4	3	3	3	2	4	4	4	4	2	
Respondent	1	2	1	5	5	5	5	5	5	5	4	5	4	4	4	4	5	5	3	
Respondent	2	2	1	4	5	4	4	4	3	5	3	5	4	2	2	4	5	5	1	
Respondent	2	2	1	4	5	4	4	4	3	5	3	5	4	2	2	4	5	5	1	
Respondent	1	2	1	2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
Respondent	1	2	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
Respondent	1	1,5	1	3	4	2														

**28. Millise tähtsusega on Teie arvates järgmised tegurid Ida-Virumaa arengu seisukohalt?**  
Märkige rist vastavasse kasti.

	ei ole oluline	vähe-oluline	neutraalne	oluline	väga oluline
kaevandatud aladele uue funktsiooni andmine					
kaevandusparandi väärtustamine ja säilitamine					
rahvale rohkem informatsiooni kaevandusparandi kasutamise võimaluste kohta					
varem kaevandatud alade kasutuselevõtmine					
endiste kaevandusalade ehitiste kasutusele võtmine					
linna- ja asulate välisilme parandamine, sh. hoonete renoveerimine ja avaliku ruumi heakorra parandamine					
avalikkuse osaluse suurendamine otsustusprotsessides kaevandusala hilisema kasutamise üle					
ühitekuuluvustunde suurendamine rahva ühistevõtte läbi					
kaevandusparandi 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 kaevandusparandi väärtustamise ja kaevandusala kasutamise					
Ida-Virumaa tuutuse suurendamine					
turismipotentsiaali parem ärakasutamine					
piiri avamine Venemaaga (suurem turistide ligipääs)					
rohkemate igapäevaste vabaaja veetmise- ja meelelahutuskohade avamine					
uute töökohade loomine (tööstuse vähendamine)					
sotsiaalprobleemidele (kuritegevus, alkoholi liigtarbimine, narkomaania jt.) leevenduse leidmine					



## Näide: Kaie Metsaots, „Ida-Viru küsitlus“.

### Samm 1: peakomponentanalüüs

(iga faktor püüab olla võimalikult tugevalt seotud kõigi algsete tunnustega)

	Factor Pattern		
	Factor1	Factor2	Factor3
funktsioon	0.39696	-0.23243	-0.32288
kaevandusparand	0.55598	-0.36915	-0.14738
informatsioon	0.64122	-0.35385	-0.25786
kasutuselevotmine	0.40339	-0.24899	-0.43860
ehitised	0.44346	-0.18667	-0.13649
renoveerimine	0.37570	0.33993	-0.09857
osalus	0.55523	-0.06224	0.23932
yhtekuuluvus	0.48312	0.11942	0.20080
koostoo	0.64627	-0.07624	0.06731
haldusreform	0.36118	-0.20778	0.70670
regionaalpoliitika	0.39371	-0.36148	0.56794
toetusrahad	0.65860	-0.12374	0.00791
tuntus	0.59357	0.23981	-0.25331
turismiipotentsiaal	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 alg tunnustest (nii on tulemusi lihtsam tõlgendada).

	Rotated Factor Pattern		
	Factor1	Factor2	Factor3
funktsioon	0.24613	0.08632	0.01530
kaevandusparand	0.80350	-0.01380	0.08010
informatsioon	0.73594	0.16770	0.06493
kasutuselevotmine	0.28170	0.03112	-0.01398
ehitised	0.47753	-0.00017	0.05460
renoveerimine	-0.22630	0.24754	0.05081
osalus	0.09963	0.11629	0.33969
yhtekuuluvus	0.15784	0.20475	0.10133
koostoo	0.38156	0.21069	0.24020
haldusreform	0.02774	0.04094	0.84781
regionaalpoliitika	0.18111	-0.02757	0.73335
toetusrahad	0.60620	0.32935	0.10452
tuntus	0.29540	0.74347	-0.04672
turismiipotentsiaal	0.17953	0.76198	-0.01686

## Näide: Kaie Metsaots, „Ida-Viru küsitlus“.

Faktorite kirjeldusvõime (=varieeruvus) enne pööramist

	Eigenvalue	Difference	Proportion	Cumulative
1	4.24386225	2.57014662	0.2358	0.2358
2	1.67371563	0.13378067	0.0930	0.3288
3	1.5393496	0.19112736	0.0856	0.4143
4	1.34880760	0.14027136	0.0749	0.4892
5	1.20053624	0.09477279	0.0667	0.5559
6	1.10576345	0.17769279	0.0614	0.6174
7	0.92807066	0.08642084	0.0516	0.6689
8	0.84164982	0.10351673	0.0468	0.7157
9	0.73813309	0.08350422	0.0410	0.7567
10	0.65462888	0.01803454	0.0364	0.7931
11	0.63659433	0.04611721	0.0354	0.8284
12	0.59047712	0.05751093	0.0328	0.8612
13	0.53296619	0.07264396	0.0296	0.8908
14	0.46032222	0.03948902	0.0256	0.9164
15	0.42183320	0.02783823	0.0234	0.9398
16	0.39399498	0.02067429	0.0219	0.9617
17	0.37332068	0.05793200	0.0207	0.9825
18	0.31538869		0.0175	1.0000

Faktorite kirjeldusvõime (=varieeruvus) pärast pööramist.

NB! Summaarne kirjeldusvõime on sama (61,7%), aga see on ümber jagatud.

6 factors will be retained by the NFACTOR criterion.

#### Variance Explained by Each Factor

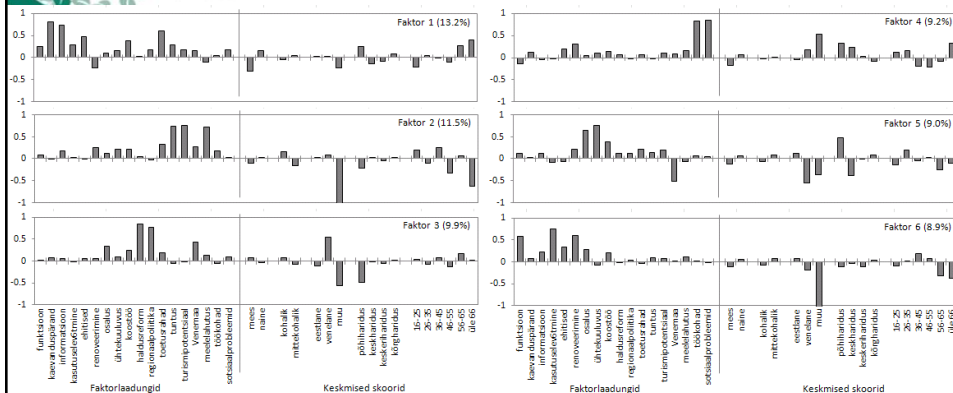
Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
2.3806328	2.0729060	1.7782903	1.6626241	1.6234793	1.5946877

#### Final Communality Estimates: Total = 11.112620

funktsioon	kaevandusparand	informatsioon	kasutuselevotmine	ehitised	renoveerimine	
0.43015255	0.67346055	0.64576439	0.65089083	0.38655644	0.62601703	
osalus	yhtekuuluvus	koostoo	haldusreform	regionaalpoliitika	toetusrahad	tuntus
0.63814030	0.64796349	0.45744436	0.73891111	0.64816276	0.56487700	0.66971826
turismiipotentsiaal	Venemaa	neelelahutus	tookhad	sotsiaalprobleemid		
0.67004627	0.56748217	0.53795813	0.72237649	0.77669802		

**Kommunaliteetid** – näitavad, kui mitu protsenti iga algse tunnuse varieeruvusest kuus valitud faktorit kokku ära kirjeldavad.

### Näide: Kaie Metsaots, „Ida-Viru küsitlus“.



Arengu seisukohast oluliste tegurite faktoranalüüsil leitud kuus esimest faktorit kirjeldasid kokku ära 61,7% algsete tunnuste varieeruvusest.

Esimene faktor osakaaluga 13,2% toob olulisemate teguritena välja kaevanduspärandi väärtustamise ja säilitamise, rahvale enama informatsiooni jagamise kaevanduspärandi kasutamise võimaluste kohta ja endiste kaevandusalade ehitiste kasutusele võtmise ning seda eelkõige seoses EL toetusrahade suurema suunamisega kaevanduspärandi väärtustamise ja kaevandusalade kasutamisse. Seejuures hindavad naisterahvad ja vanemad inimesed taolist lähenemist pisut enam.

Teine faktor, mis kirjeldab ära 11,5% algsete tunnuste varieeruvusest, toob arengu seisukohast oluliste tegurite välja Ida-Virumaa tuntuks suurendamise, turismpotentsiaali parema ärakasutamise ja rohkemate igapäevaste vabaaja veetmise- ja meelelahutuskohtade avamise ning prevaleerib pigem nooremate inimeste hulgas ja ei leia peaaegu üldse ära märkimist muust rahvusest vastanute puhul.

Kolmas faktor osakaaluga 9,9% peegeldab haldusreformi ja regionaalpoliitika rakendamise, aga ka Venemaaga piiri avamise rolli, ning leiab enam märkimist venelaste poolt, selgemalt eristub ka põhiharidusega inimeste ignorantsus selle faktori osas.

...