Aalto-yliopisto
Perustieteiden
korkeakoulu

# Poikkeavien havaintojen tunnistaminen pääkomponenttianalyysin menetelmin (valmiin työn esittely)

*Kalle Alaluusua*

15.10.2018

Ohjaaja: *TkT Lauri Viitasaari*
Valvoja: *Apulaisprofessori Pauliina Ilmonen*



Systeemianalyysin
laboratorio

# Tavoite

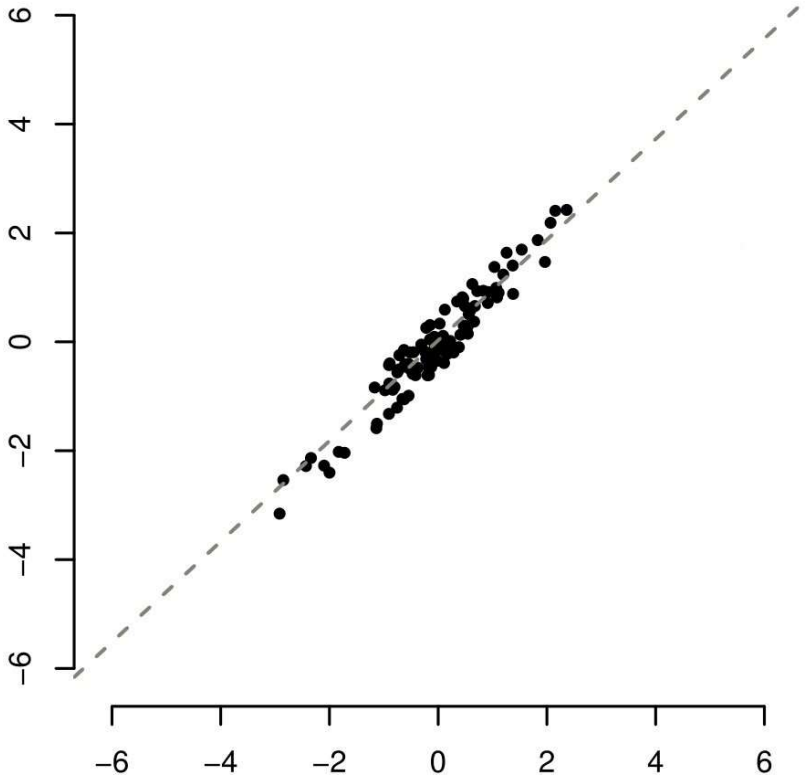
- Tunnistetaan poikkeavat havainnot epäsymmetrisestä havaintojoukosta tuotantotekniikan alalta

Idea:

- Laaditaan malli, joka kuvaa tyypillisten havaintojen ominaispiirteitä
- Tarkastellaan, kuinka hyvin kukin havainto sopii tähän malliin

# Pääkomponenttianalyysi

- Laaditaan joukko keskenään korreloimattomia uusia muuttujia alkuperäisten muuttujien lineaarikombinaatioina

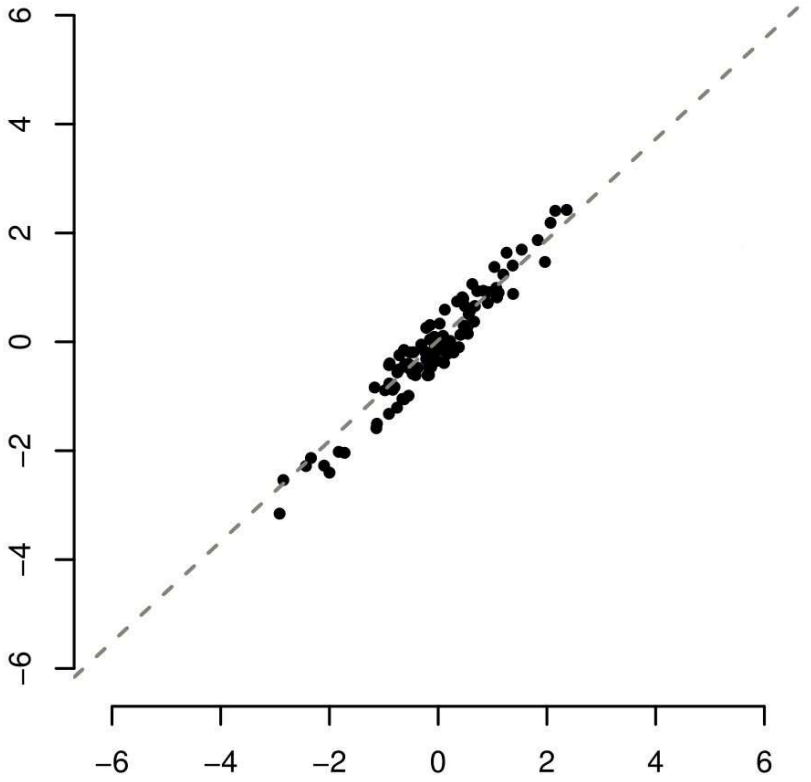


# Pääkomponenttianalyysi

- Laaditaan joukko keskenään korreloimattomia uusia muuttujia alkuperäisten muuttujien lineaarikombinaatioina

- Kukin pääkomponentti maksimoi tälle projektoidun datan varianssin

- Kuvaa havaintojoukon kovarianssirakennetta



# Pääkomponenttianalyysi

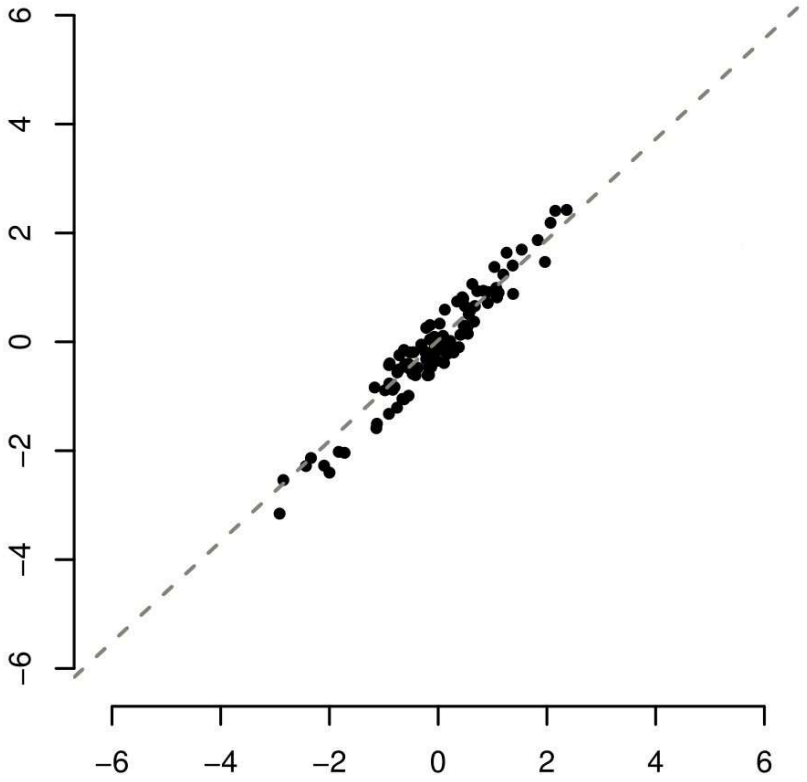
- Laaditaan joukko keskenään korreloimattomia uusia muuttujia alkuperäisten muuttujien lineaarikombinaatioina

- Pääkomponentit **P** kovarianssimatriisin **S** ominaisvektoreita

$$\mathbf{S} = \mathbf{P}_{p,k}\mathbf{L}_{k,k}\mathbf{P}'_{p,k},$$

ja projektio **T** siten

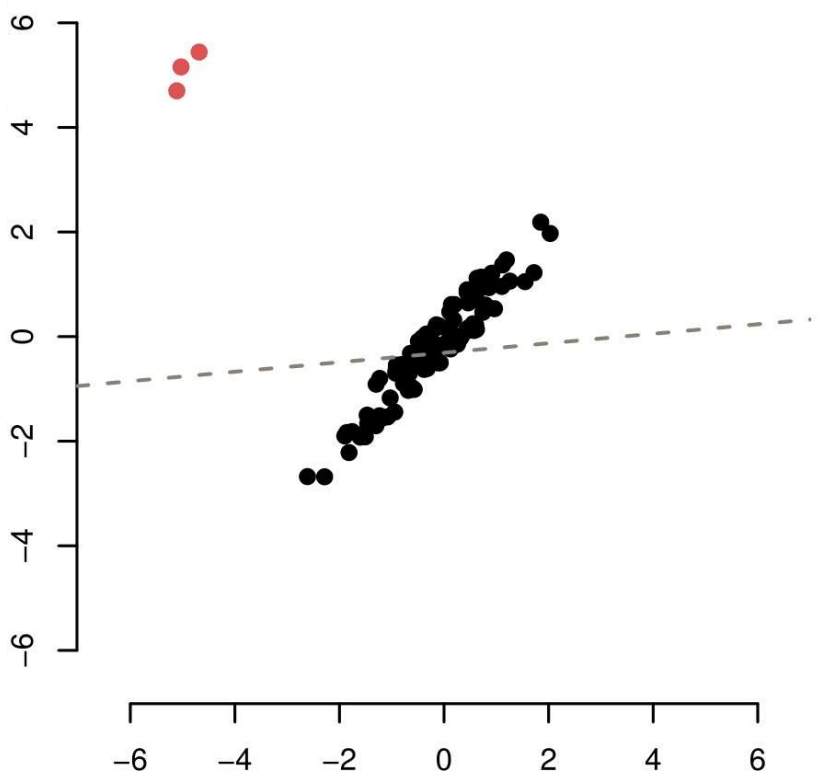
$$\mathbf{T}_{n,k} = (\mathbf{X}_{n,p} - \mathbf{1}_n\hat{\boldsymbol{\mu}}')\mathbf{P}_{p,k}$$



# Pääkomponenttianalyysi

- Laaditaan joukko keskenään korreloimattomia uusia muuttujia alkuperäisten muuttujien lineaarikombinaatioina

- Herkkä poikkeaville havainnoille!



# Vakaa pääkomponenttianalyysi

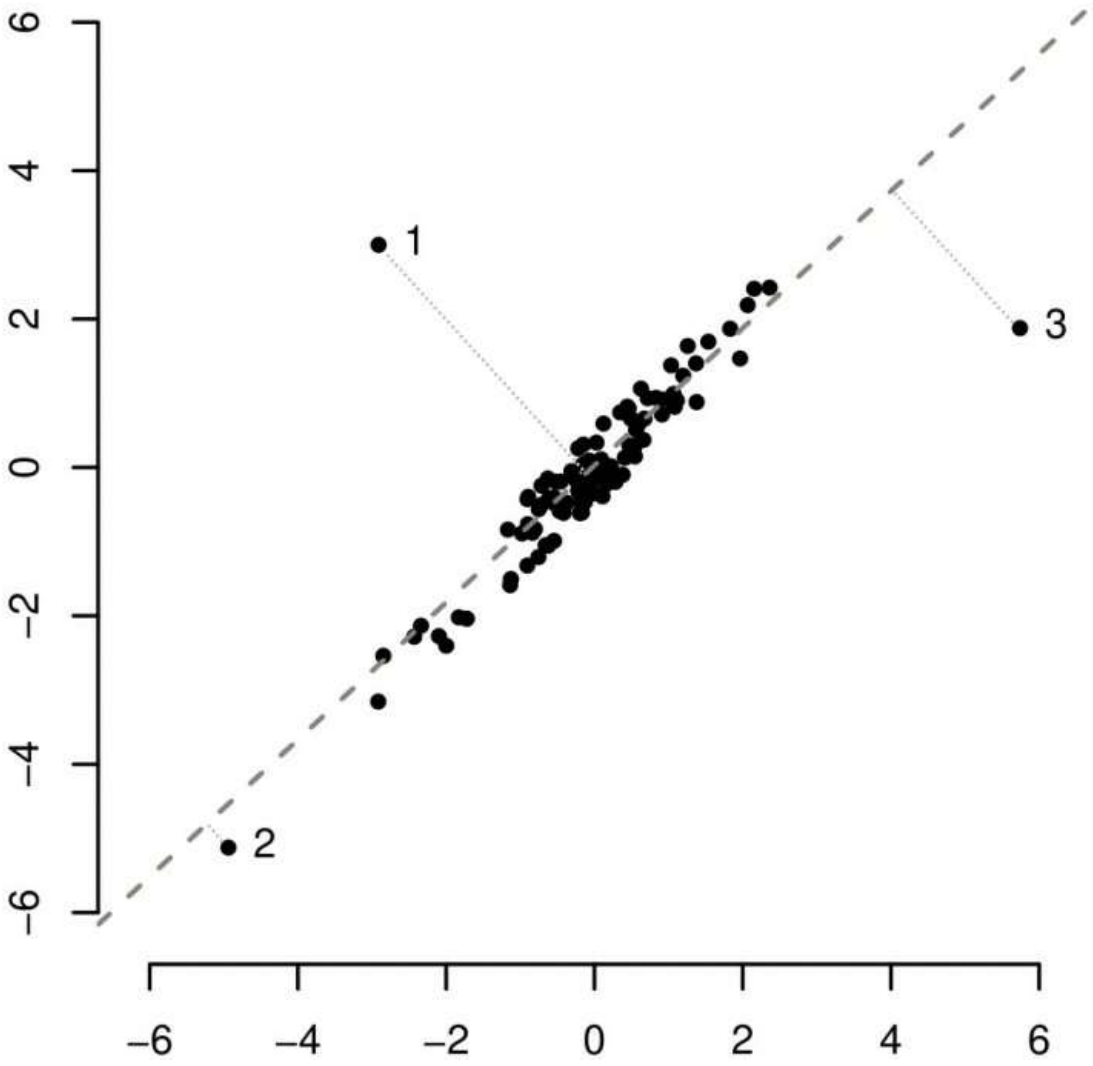
Kaksi koulukuntaa:

1. Korvataan klassinen kovarianssimatriisi vakaalla estimaatilla

2. Haetaan pääkomponenttien suunnat iteratiivisesti maksimoimalla vakaata hajontaestimaattia

Ideaalitilanne:

- Poikkeavat havainnot kaukana havaintojoukon keskipisteestä sekä pääkomponenttien virittämästä tasosta




Esimerkkitapauksia poikkeavista havainnoista:
1) ortogonaalinen, 2) ”hyvä vipupiste”, 3) ”huono vipupiste”

# Menetelmät ja havaintojoukko

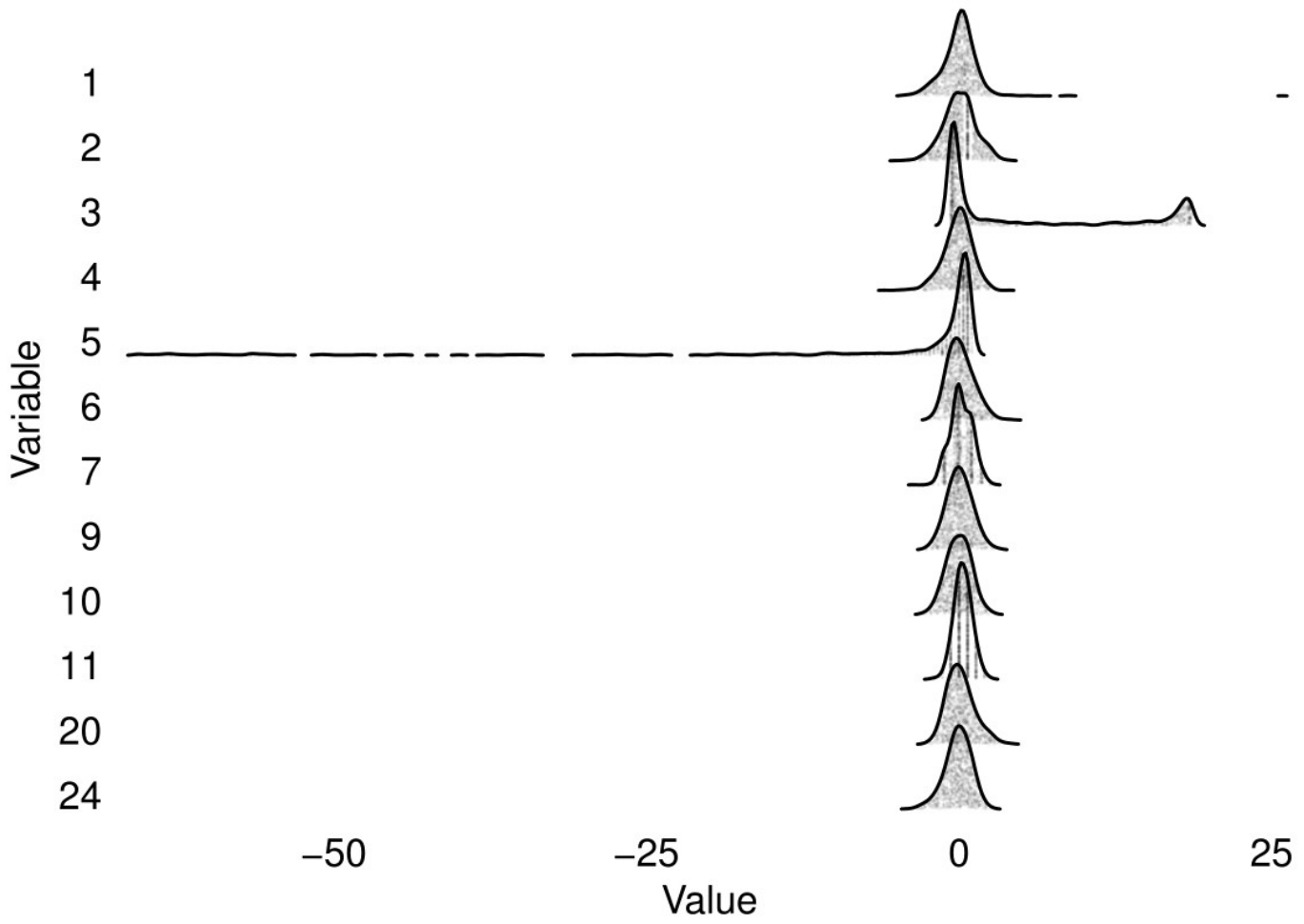
**ROBPCA**

- Hyödyntää kumpaakin esiteltyä vakauttamisperiaatetta
- Menetelmästä kehitetty muunnelma epäsymmetrisen datan analyysiin

Notaatio:

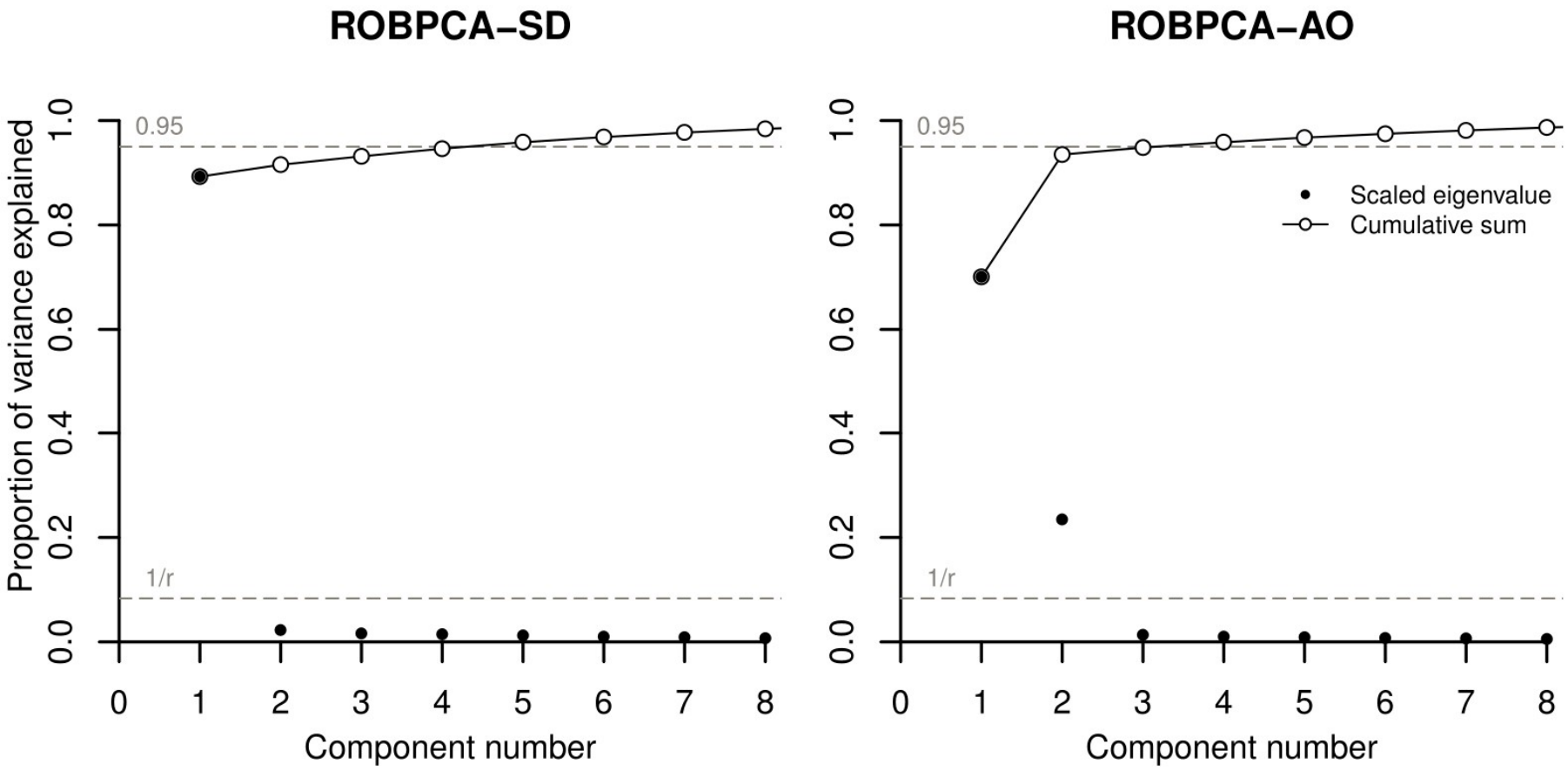
ROBPCA-SD: normaalisuusoletus tyypillisille havainnoille

ROBPCA-AO: huomioi tyypillisten havaintojen vinouden



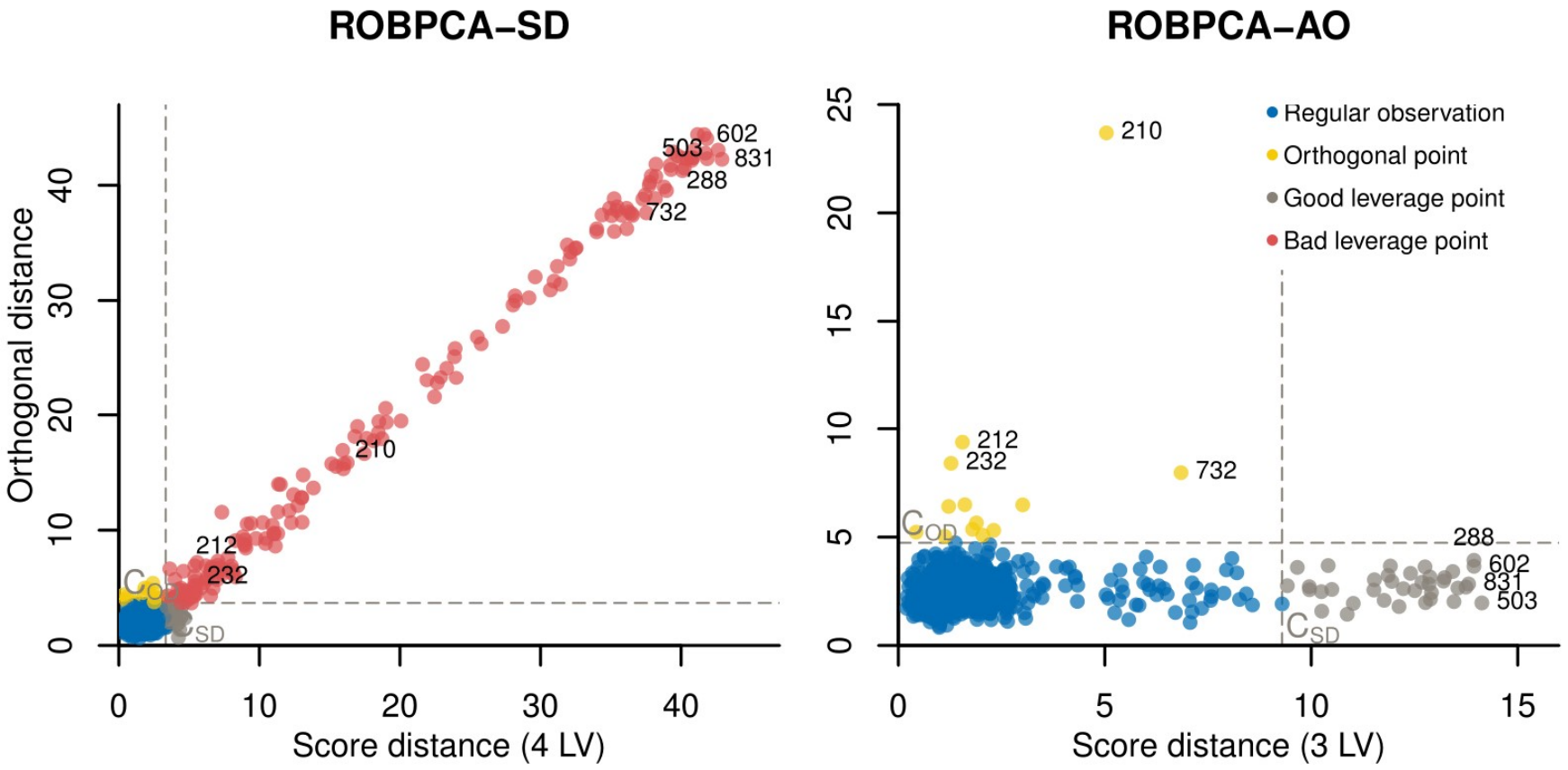

Tiheyskuvaaja vakaasti keskitetystä havaintojoukosta **X**, joka sisältää 851 havaintoa 12 muuttujasta

# Mallin dimensionaalisuuden valinta

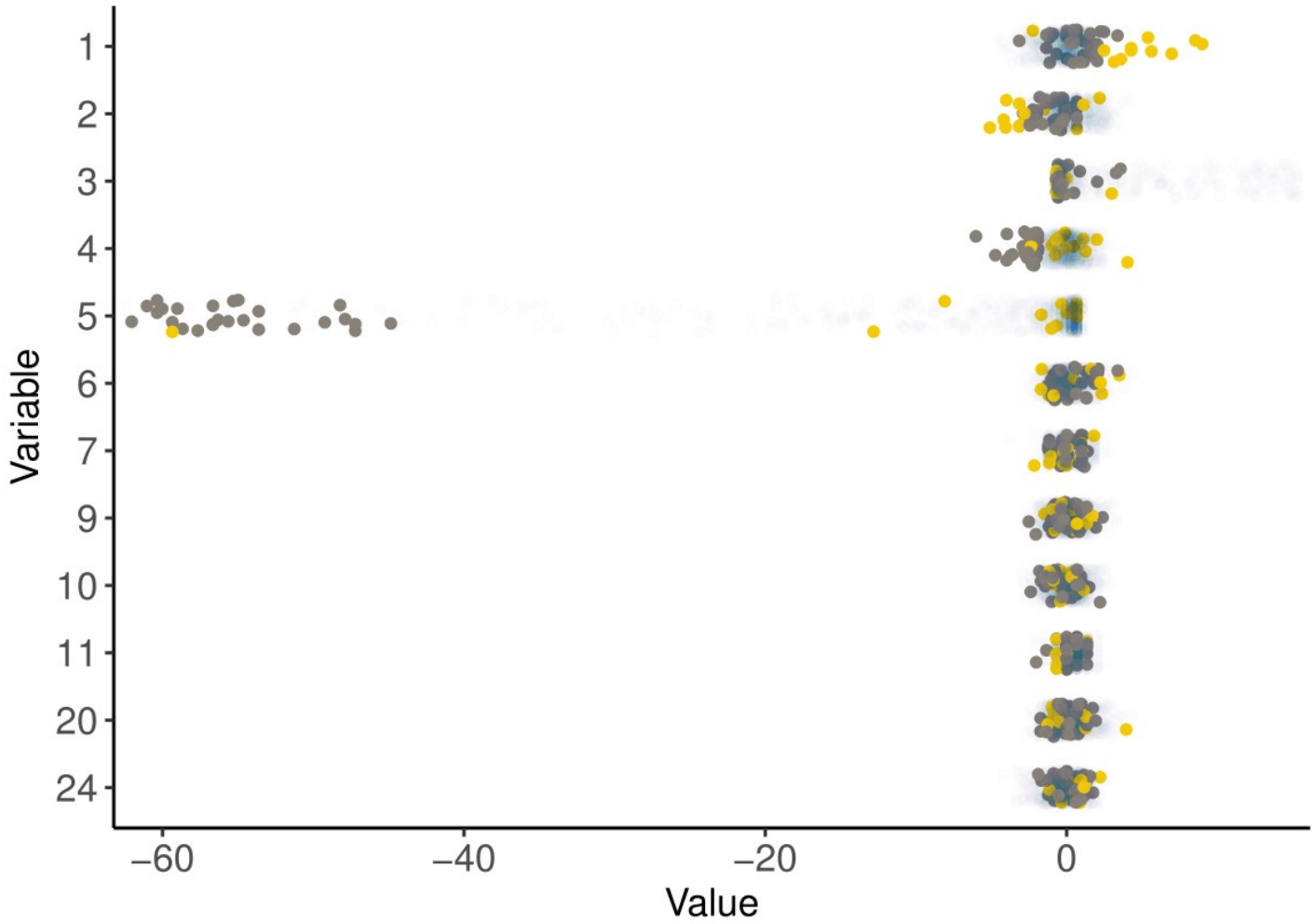



Selitetään 95% havaintojoukon varianssista

# Poikkeavat havainnot kartalla




$$OD_i = \|\mathbf{x}_i - \hat{\boldsymbol{\mu}} - \mathbf{P}_{p,k}\mathbf{t}_i'\|, \; SD_i = \sqrt{\sum_{j=1}^{k} \frac{t_{ij}^2}{l_j}}$$




Väritetty nauhakuvaaja (ortogonaaliset havainnot keltaisella, hyvät vipupisteet harmaalla)

# Muuttujien vaikutus poikkeavuuteen

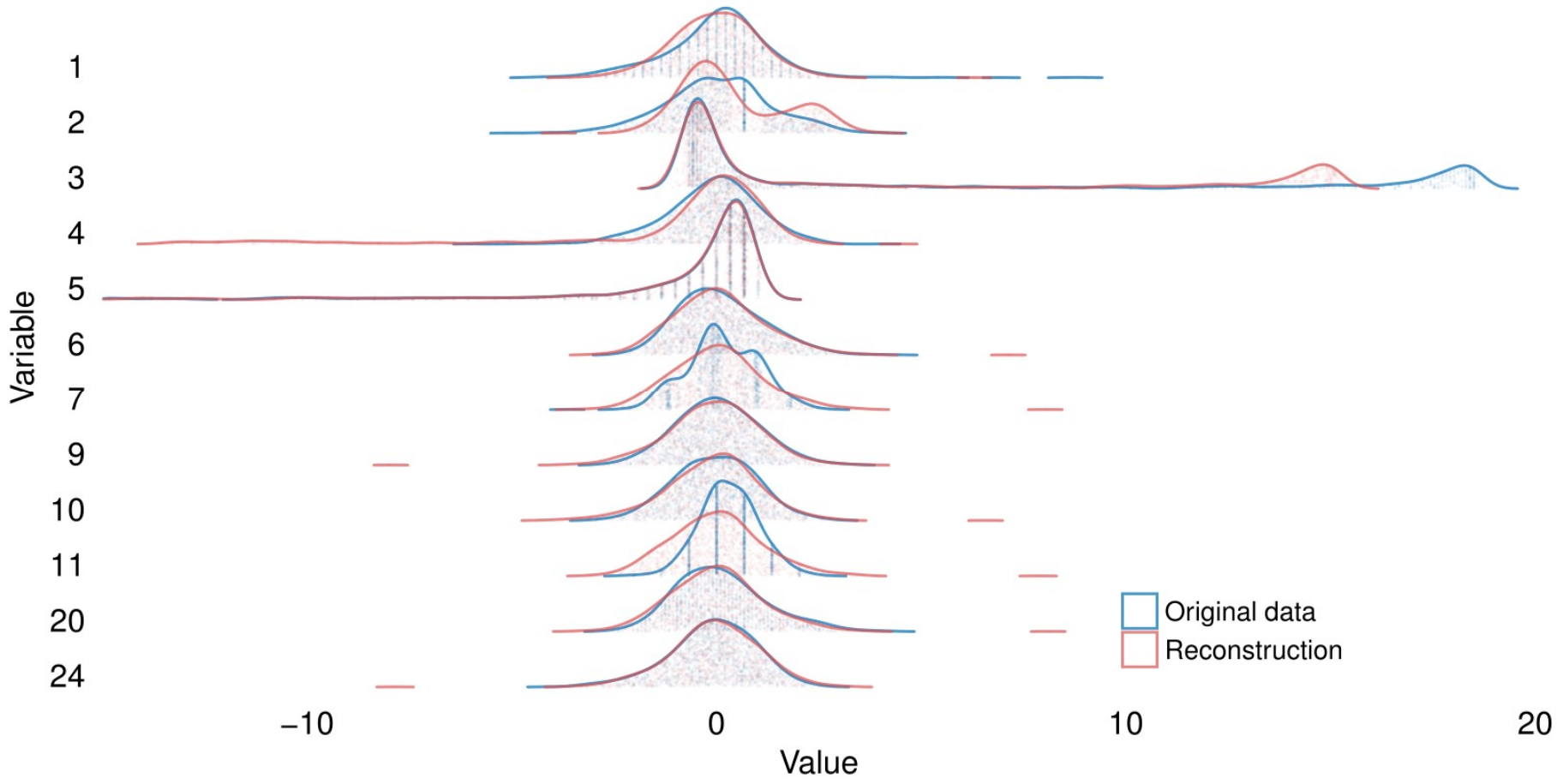
| Observation | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X9 | X10 | X11 | X20 | X24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53 | 6.2 | 0.3 | 0 | 0.1 | 0 | 1.7 | 5.2 | 0.2 | 1.1 | 0.1 | 12.3 | 0.1 |
| 125 | 5.6 | 12.9 | 0.3 | 0 | 0 | 0.5 | 2.5 | 0.4 | 0 | 1.4 | 1.2 | 3.5 |
| 172 | 8 | 5.4 | 0.2 | 0 | 0 | 3.8 | 0.3 | 3.4 | 1 | 1.1 | 0 | 5.5 |
| 210 | 436.3 | 1.4 | 1.4 | 3.1 | 0 | 0.9 | 6.7 | 7.8 | 4.8 | 19.3 | 14 | 65.6 |
| 212 | 67.6 | 1.7 | 0.2 | 2.5 | 0 | 0 | 0.7 | 0.2 | 1.3 | 2.6 | 8 | 3.3 |
| 232 | 53.9 | 0.1 | 0.2 | 1 | 0 | 1.8 | 0.1 | 0.8 | 2 | 2.4 | 1.6 | 6.7 |
| 277 | 16.9 | 5.2 | 0.2 | 0.5 | 0 | 0.1 | 0.1 | 1.2 | 0 | 1.2 | 3.4 | 3 |
| 359 | 12.8 | 12.7 | 0.4 | 1.2 | 0 | 3.5 | 0.6 | 0 | 2 | 1.3 | 3.9 | 2.9 |
| 362 | 17.7 | 5.8 | 0.2 | 2.8 | 0 | 1.7 | 0.3 | 4.5 | 0.7 | 0.7 | 4.1 | 3.6 |
| 369 | 0.1 | 4.7 | 0 | 20.4 | 0 | 10.3 | 0.5 | 1.4 | 0.7 | 0.3 | 3 | 0.6 |
| 379 | 2.1 | 15.1 | 0.3 | 0.3 | 0 | 5.1 | 0.4 | 0 | 0.1 | 0 | 0.7 | 1.8 |
| 518 | 3.7 | 13.4 | 0.3 | 0.6 | 0 | 1 | 0.6 | 0.6 | 1.1 | 1 | 0.9 | 2.1 |
| 732 | 44.4 | 0.6 | 0.1 | 0.1 | 0 | 2.3 | 2.6 | 0.9 | 0.2 | 1.3 | 6.2 | 5 |

Muuttujien vaikutus neliöityyn ortogonaaliseen etäisyyteen

$\mathrm{SPE} \equiv ||\tilde{\mathbf{x}}||^2 = ||(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\mathbf{x}||^2$ (Squared Prediction Error)

# Mallin arviointi rekonstruktion avulla

- Rekonstruoidaan havaintojoukko viemällä projektoitu havaintojoukko alkuperäiseen koordinaatistoon: $\hat{\mathbf{X}}_{n,p} = \mathbf{T}_{n,k}\mathbf{P}_{p,k}^{\intercal}$
- Jos mallin dimensio olisi valittu samaksi kuin havaintojoukon, $\mathbf{P}_{p,k}\mathbf{P}_{p,k}^{\intercal}$ on identiteetti ja rekonstruktio täydellinen
- Haluttiin tarkastella etäisyyttä mallista




Keskitetty ja skaalattu havaintojoukko sekä rekonstruktio

# Lopputulos

- Lopullinen malli huomioi havaintojoukon erityispiirteet
- Tunnistettiin poikkeavat havainnot ja systemaattisuus näiden takana
- Kuva havaintojoukon rakenteesta selkeytyi

Jatkossa voitaisiin tutkia mm. logistisin regressiomenetelmin, johtaako poikkeavuus heikompaan lopputuotteeseen.