

TARTU ÜLIKOOL  
MATEMAATIKA-INFORMAATIKATEADUSKOND  
MATEMAATILISE STATISTIKA INSTITUUT

Kristi Läll

**Mitmemõõtmeline analüüs peptiidide käitumise uurimiseks**

Magistritöö

Juhendajad: Mare Vähi, Anne Selart

TARTU

2013

## Sisukord

Sissejuhatus .....	3
1. Mitmemõõtmeline normaaljaotus, maatriksnormaaljaotus ja Wisharti jaotus .....	5
2. Sõltuvate vaatluste võrdlemise testid .....	7
2.1 Mõõtmistulemuste paaride keskväärtuste võrdlus .....	7
2.2 Ühe tunnuse korduvmõõtmine mitme ravimi korral .....	8
3. Profiilianalüüs .....	10
4. Ühefaktoriline MANOVA.....	14
5. Mitmemõõtmeline korduvmõõtmise mudel .....	17
5.1 Mudeli kujud ja parameetrite hinnangud.....	17
5.2 Hüpoteeside testimine .....	20
6. Andmete analüüs .....	22
6.1 Andmete bioloogiline taust .....	22
6.2 Andmete kirjeldav analüüs.....	25
6.3 Profiilianalüüsi rakendus.....	26
6.4 Peptiidide käitumise võrdlemine .....	28
Summary .....	33
Lisa 1. Ühemõõtmelise Shapiro-Wilk'i testi tulemused .....	34
Lisa 2. Box's M testi väljund R-is iga peptiidi korral, olulisusenivool 0.05: .....	35
Lisa 3. Mitmemõõtmelise normaaljaotusega peptiidide profiilide joonised .....	36
Lisa 4. Peptiidid, mis ei ole mitmemõõtmelise normaaljaotusega .....	39
Lisa 5. Peatükk 5 toodud hinnangute tuletuskäigud.....	44
Lisa 6. R-i kood peatükis 5 esitatud ja peatükis 6.4 rakendatud teooria kasutamiseks.....	49
Kasutatud kirjandus.....	52
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks .....	54

## Sissejuhatus

Tuberkuloosivaktsiini väljatöötamisel on oluline uurida, millised antigeenid vaktsiinis leiduvate antikehadega reageerivad. Samuti on tähtis, et antikehad reageeriks kiiresti ning nõutud koguses pärast vaktsineerimist. Tavaliselt mõõdetakse ühel patisendil vaktsineerimise järel mitmel ajahetkel sadade erinevate antigeenide ehk peptiidide käitumist. Seega on tekkivate andmete maht väga suur ning nende analüüsimiseks ei ole statistika klassikalised meetodid alati sobilikud. Käesolev töö tutvustab üht meetodikat sellise struktuuriga andmete analüüsiks ning uurib selle rakendatavust konkreetset andmestikku.

Esimene peatükk tuletab meelde mitmemõõtmelise normaaljaotust, maatriksnormaaljaotust ja Wisharti jaotust, sest neid kasutatakse hilisemas teoorias korduvalt. Teises peatükis antakse ülevaade kahest lihtsamast uuringudisainist: esimesena sellisest, kus võrreldakse kahte ravimit, mõõtes samadel või sarnastel objektidel  $p$  tunnust ning teisena sellisest, kus omavahel võrreldakse  $q$  ravimit, mõõtes samadel objektidel vaid 1 tunnust. Mõlema juhul on kirjas testitav hüpoteesipaar, võrdlemiseks vajaliku statistiku konstrueerimine ning tema jaotust. Teine peatükk valmistab lugejat ette 5. peatüki jaoks, kus esitatakse nende kahe disaini üks võimalikke üldistusi.

Kolmandas peatükis käsitletakse profiilianalüüsi, mis on mõeldud erinevate populatsioonide keskväärtusvektorite võrdlemiseks. Antakse ülevaade profiilianalüüsi käigust, püstitatavatest hüpoteesidest ja nende kontrollist võrreldes 2 populatsiooni.

Neljandas peatükis tutvustatakse mitmemõõtmelist dispersioonanalüüsi (MANOVAt), mis on vajalik keskväärtuste võrdlemisel, kui võrreldavaid populatsioone on 2-st rohkem. Kirjeldatakse MANOVA kasutamiseks vajalikke eeldusi ning teststatistiku konstrueerimist.

Viiendas peatükis antakse ülevaade mitmemõõtmeliste korduvmõõtmiste tulemusena saadud andmete ühest võimalikust analüüsiviisist ehk mitmemõõtmelisest juhumõjudega korduvmõõtmiste mudelist. Andmed olgu sellise struktuuriga, et igal indiviidil on mõõdetud  $p$  ajahetkel  $m$  tunnust ning võrreldavaid gruppe saab olla  $r$ . Peatükis tuuakse välja juhumõjudega mudeli kuju, mudeli parameetrite hinnangud, mudeli parameetrite kohta käivate hüpoteeside üldine kuju ning nende testimiseks vajalik eeskiri.

Viimases peatükis kirjeldatakse töös kasutatavate andmete bioloogilist tausta ja tekkimise protsessi. Samuti tutvustatakse konkreetset töös kasutatavat andmestikku, rakendatakse profiilianalüüsi ja proovitakse hinnata eelnevas peatükis esitatud juhumõjudega mudel.

Profiilianalüüsi eesmärgiks on uurida, kas leidub peptiide, mille korral antikehade kontsentratsioon ajas ei ole muutunud. Juhumõjudega mudeli hindamise eesmärgiks on peptiidide kasvukõverate kujude võrdlemine, leidmaks käitumiselt sarnaseid peptiide.

Töös on vektorid antud paksus kirjas ja kursiivis ning maatriksid paksus püstkirjas. Töös esitatud praktiline analüüs on läbi viidud tarkvaraga R-2.15.0. (R Development Core Team, 2012). Tööd toetati ETF granti ETF8294 poolt ning magistriõpingute ajal välismaal tudeerimist toetas Euroopa Liidu Euroopa Sotsiaalfond programmi DoRa raames. Programmi DoRa viib ellu Sihtasutus Archimedes.

# 1. Mitmemõõtmeline normaaljaotus, maatriksnormaaljaotus ja Wisharti jaotus

Peatükk 1 põhineb allikatel Johnson & Wichern (2002), Kollo (2012) ning von Rosen & Kollo (2005).

## Normaaljaotus

Juhusliku vektori  $\mathbf{X} : p \times 1$  jaotuseks on mitmemõõtmeline normaaljaotus, kui selle tihedusfunktsioon on kujul

$$f(\mathbf{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left[ -\left(\frac{1}{2}\right) (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right],$$

kus  $\mathbf{x}$  elemendid kuuluvad vahemikku  $(-\infty, \infty)$  ja kovariatsioonimaatriks  $\boldsymbol{\Sigma}$  on astakuga  $p$ . Vektori  $\mathbf{X}$  keskvaartusvektor on  $E\mathbf{X} = \boldsymbol{\mu}$ , kovariatsioonimaatriks  $E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T) = \boldsymbol{\Sigma}$ . Keskvaartusvektori  $\boldsymbol{\mu}$  ja kovariatsioonimaatriksi  $\boldsymbol{\Sigma}$  elemendid on tähistatud :

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \text{ ja } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1p}^2 \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp}^2 \end{pmatrix},$$

kus  $\sigma_{ij}$  on kovariatsioon  $\mathbf{X}$   $i$ -nda ja  $j$ -nda komponendi vahel.

## Maatriksnormaaljaotus

Olgu antud maatriksid  $\boldsymbol{\Sigma} = \boldsymbol{\tau}\boldsymbol{\tau}^T$  ja  $\mathbf{Y} = \boldsymbol{\zeta}\boldsymbol{\zeta}^T$ , kus  $\boldsymbol{\tau}$  on  $p \times r$  maatriks ja  $\boldsymbol{\zeta}$  on  $n \times r$  maatriks. Öeldakse, et maatriks  $\mathbf{X}$ , mis on mõõtmetega  $p \times n$ , on maatriksnormaaljaotusega, parameetritega  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  ja  $\mathbf{Y}$ , kui tal on sama jaotus, nagu maatriksil

$$\boldsymbol{\mu} + \boldsymbol{\tau}\mathbf{U}\boldsymbol{\zeta}^T,$$

kus  $\boldsymbol{\mu}$  on mittejuhuslik  $p \times n$  maatriks ja  $\mathbf{U}$  on  $r \times s$  maatriks, mis sisaldab sõltumatuid juhuslikke vektoreid  $\mathbf{U}_i \sim N_r(\mathbf{0}, \mathbf{I}_r)$ ,  $i = 1, 2, \dots, s$ . Maatriksnormaaljaotust tähistatakse  $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{Y})$ .

## Wisharti jaotus

Olgu  $\mathbf{X}$  juhuslik  $p \times n$  maatriks, mille veerud  $\mathbf{X}_j$ ,  $j = 1, \dots, n$  on sõltumatud ja sama tsentreeritud normaaljaotusega,  $\mathbf{X}_j \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$  ehk  $\mathbf{X} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I}_n)$ , kus  $\mathbf{0}$  tähistab nullide

maatriksit. Juhusliku maatriksi  $\mathbf{V} = \mathbf{X}\mathbf{X}^T$  jaotust nimetatakse Wisharti jaotuseks parameetritega  $\mathbf{\Sigma}$  ja  $n$  ja tähistatakse  $\mathbf{V} \sim W_p(\mathbf{\Sigma}, n)$ . Maatriksi  $\mathbf{V}$  keskvaärtuseks on  $E\mathbf{V} = n\mathbf{\Sigma}$  ja dispersiooniks  $D\mathbf{V} = n(\mathbf{I}_{p^2} + \mathbf{K}_{p,p})(\mathbf{\Sigma} \otimes \mathbf{\Sigma})$ , kus  $\mathbf{K}_{p,p}$  tähistab kommuteerimismaatriksit ning  $\mathbf{I}_{p^2}$  tähistatakse ühikmaatriksit mõõtmetega  $pp \times pp$ .

Omadus 1. Olgu  $\mathbf{R}$  sümmeetriline idempotentne maatriks mõõtmetega  $n \times n$  ja  $\mathbf{X}\mathbf{X}^T \sim W_p(\mathbf{\Sigma}, n)$ . Siis  $\mathbf{X}\mathbf{R}\mathbf{X}^T \sim W_p(\mathbf{\Sigma}, r)$ , kus  $r$  on maatriksi  $\mathbf{R}$  astak.

Omadus 2. Olgu  $\mathbf{X} \sim N_{p,n}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{Y})$ . Iga  $\mathbf{R}: w \times p$  ja  $\mathbf{Q}: t \times n$  korral

$$\mathbf{R}\mathbf{X}\mathbf{Q}^T \sim N_{w,t}(\mathbf{R}\boldsymbol{\mu}\mathbf{Q}^T, \mathbf{R}\mathbf{\Sigma}\mathbf{R}^T, \mathbf{Q}\mathbf{Y}\mathbf{Q}^T).$$

## 2. Sõltuvate vaatluste võrdlemise testid

Peatükk 2 põhineb allikal Johnson & Wichern (2002).

### 2.1 Mõõtmistulemuste paaride keskväärtuste võrdlus

Kahe ravimi võrdlemiseks võib kasutada järgmist lähenemist: määrata mõlemad ravimid samade või sarnastele objektidele. Seejärel saab võrrelda paaride mõõtmistulemusi, arvutades mõõtmistulemuste vahe. Võttes paari sama inimese 2 mõõtmistulemust või kahe sarnase objekti mõõtmistulemused, elimineeritakse enamik paarisolevate objektide sisesest varieeruvusest.

Ühemõõtmelisel juhul saame kahe ravimi mõõtmistulemusi võrrelda, kasutades paaride  $t$ -testi. Olgu  $X_{j1}$  mõõtmistulemus pärast esimese ravimi manustamist (või mõõtmistulemus enne ravimi manustamist) ja  $X_{j2}$  mõõtmistulemus pärast teise ravimi manustamist (või mõõtmistulemus pärast ravimidoosi)  $j$ -ndal paaril. Seega,  $(X_{j1}, X_{j2})$  võivad olla mõõtmised  $j$ -ndal objektil või  $j$ -ndal sarnaste objektide paaril. Olgu mõõtmistulemuste erinevus defineeritud  $D_j = X_{j1} - X_{j2}$ ,  $j = 1, 2, \dots, n$ . See vahe peaks väljendama ainult ravimite efektide erinevust. Eeldame, et  $D_j$  on sõltumatud vaatlused normaaljaotusest  $D_j \sim N(\mu, \sigma_d^2)$ . Defineerime  $t$ -statistiku:

$$t = \frac{\bar{D} - \mu}{\frac{s_d}{\sqrt{n}}},$$

kus

$$\bar{D} = \frac{1}{n} \sum_{j=1}^n D_j \text{ ja } s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})^2.$$

$T$ -statistik on  $t$ -jaotusega vabadusastmetega  $n - 1$ . Testimaks, kas  $\mu = 0$  olulisusenivool  $\alpha$ , peaks  $t$ -statistiku absoluutväärtust võrdlema  $\bar{t}_{n-1}(\frac{\alpha}{2})$  täiendkvantiiliga. Juhul, kui  $t$ -statistiku väärtus on suurem, võetakse vastu alternatiivne hüpotees.

Mitmemõõtmelise juhu, kus uuritakse ravimite mõju paljudele näitajatele, defineerimiseks on tarvis vahet teha  $p$  tunnusel, kahel ravimil ja  $n$  objektil. Tähistame  $j$ -ndal objektil tehtud mõõtmisi järgmiselt:

$$X_{1j1} = \text{tunnuse 1 väärtus ravimi 1 korral}$$

$X_{1j_2} = \text{tunnuse } 2 \text{ väärtus ravimi } 1 \text{ korral}$

...

$X_{1j_p} = \text{tunnuse } p \text{ väärtus ravimi } 1 \text{ korral}$

$X_{2j_1} = \text{tunnuse } 1 \text{ väärtus ravimi } 2 \text{ korral}$

...

$X_{2j_p} = \text{tunnuse } p \text{ väärtus ravimi } 2 \text{ korral}$

Olgu  $p$  paaride vahet defineeritud kui sama tunnuse mõõtmistulemuste vahe erinevate ravimite korral:

$$D_{j_1} = X_{1j_1} - X_{2j_1}, D_{j_2} = X_{1j_2} - X_{2j_2}, \dots, D_{j_p} = X_{1j_p} - X_{2j_p}.$$

Tähistagu  $\mathbf{D}_j^T$  vektorit kujul  $\mathbf{D}_j^T = (D_{j_1}, D_{j_2}, \dots, D_{j_p})$  ning kehtigu  $j = 1, 2, \dots, n$  korral  $E(\mathbf{D}_j^T) = \boldsymbol{\mu}^T = (\mu_1, \mu_2, \dots, \mu_p)$  ja  $Cov(\mathbf{D}_j) = \boldsymbol{\Sigma}_d$ . Kui saab eeldada, et  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n$  on sõltumatud juhuslikud suurused jaotusega  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_d)$ , siis vahede keskvaertusvektori kohta käivaid järeldusi tehakse kasutades Hotellingu  $T^2$ -statistikut. See on esitatud järgmiselt:

$$T^2 = n(\bar{\mathbf{D}} - \boldsymbol{\mu})^T \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\mu}) \sim T^2(p, n - 1),$$

kus

$$\bar{\mathbf{D}} = \frac{1}{n} \sum_{j=1}^n \mathbf{D}_j \text{ ja } \mathbf{S}_d = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{D}_j - \bar{\mathbf{D}})(\mathbf{D}_j - \bar{\mathbf{D}})^T. \quad (2.1.1)$$

Statistiku  $T^2$  teisendus on  $F$ -jaotusega:

$$T^2 \frac{(n-p)}{(n-1)p} = \frac{(n-p)}{(n-1)p} n(\bar{\mathbf{D}} - \boldsymbol{\mu})^T \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\mu}) \sim F_{p, n-p}.$$

Hüpoteesi  $H_0: \boldsymbol{\mu} = \mathbf{0}$  testimine tähendab sisuliselt väite, et kahe ravimi vahel pole keskmiselt mingit erinevust, kontrollimist. Alternatiivne hüpotees  $\boldsymbol{\mu} \neq \mathbf{0}$  võetakse vastu, kui statistiku  $\frac{(n-p)}{(n-1)p} T^2$  väärtus ületab kriitilist väärtust  $\bar{F}_{p, n-p}(\alpha)$ , mis on  $F_{p, n-p}$  jaotuse  $\alpha$ -täiendkvantiil.

## 2.2 Ühe tunnuse korduvmõõtmine mitme ravimi korral

Ühe võimaliku ühemõõtmelise paaride  $t$ -statistiku mitmemõõtmelise üldistuse saab teha olukorras, kus võrreldakse  $q$  ravimit, mõõtes ühte kindlat tunnust. Iga indiviid manustab igat

ravimit teatud ajavahemiku järel. Olgu  $\mathbf{X}_j = (X_{j1}, X_{j2}, \dots, X_{jq})^T, j = 1, 2, \dots, n$ , kus  $X_{j1}$  tähistab  $j$ -nda objekti mõõtmistulemust esimese ravi korral,  $X_{j2}$   $j$ -nda objekti mõõtmistulemust teise ravi korral jne. Olgu  $\boldsymbol{\mu} = E(\mathbf{X}_j)$ . Võrdlemise eesmärgil vaatame keskväärtusvektorite elementide vahesid, mida on võimalik esitada kujul

$$\begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_q \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} = \mathbf{C}_1 \boldsymbol{\mu}$$

või

$$\begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_q - \mu_{q-1} \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} = \mathbf{C}_2 \boldsymbol{\mu}.$$

Nii  $\mathbf{C}_1$  kui ka  $\mathbf{C}_2$  on kontrastimaatriksid, sest nende  $q - 1$  rida on lineaarselt sõltumatud ja iga rida on kontrastivektor. Kuna kõigile indiviididele on manustatud kõiki ravimeid ja katse on korraldatud randomiseeritud plaani alusel, siis ravimite võrdlemisel ei mõjuta indiviidi omapära tulemust. Kui uuritava tunnuse keskväärtus on iga ravimi kasutamisel sama, siis  $\mathbf{C}_1 \boldsymbol{\mu} = \mathbf{C}_2 \boldsymbol{\mu} = \mathbf{0}$ . Hüpooteesiks on, et ravimid ei erine üksteisest, ning selle kontrollimisel pole oluline, kumba kontrastimaatriksit kasutada.

Ravimite mõju kontrollimiseks korduvmõõtmiste korral eeldatakse, et populatsioon on pärit jaotusest  $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  ning olgu kontrastimaatriks tähistatud  $\mathbf{C}$ -ga. Testimaks hüpoteesipaari  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$  versus  $H_1: \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}$  olulisusenivool  $\alpha$ , kasutatakse otsustusreeglit

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})^T (\mathbf{C}\mathbf{S}\mathbf{C}^T)^{-1} (\mathbf{C}\bar{\mathbf{x}}) > \frac{(n-1)(q-1)}{(n-q+1)} \bar{F}_{q-1, (n-q+1)}(\alpha),$$

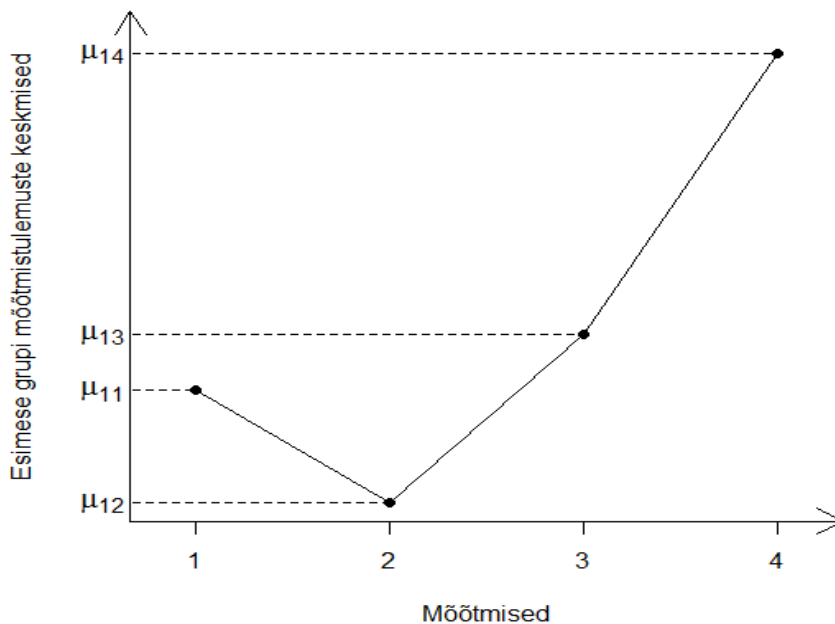
kus  $\bar{\mathbf{x}}$  on valimi keskväärtusvektor,  $\mathbf{S}$  valimi dispersioonimaatriks ja  $\bar{F}_{q-1, (n-q+1)}(\alpha)$  on  $F$ -jaotuse (vabadusastmetega  $q - 1$  ja  $n - q + 1$ )  $\alpha$ -täiendkvantiil. Valimi keskväärtusvektor ja dispersioonimaatriks on defineeritud analoogselt valemitele (2.1.1):

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \text{ ja } \mathbf{S} = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T.$$

### 3. Profiilianalüüs

Profiilianalüüsi kasutatakse situatsioonides, kus kahe või enama grupi objektidel on tehtud  $p$  mõõtmist, testi jne. Kõik tulemused peaksid olema esitatud võrreldavates skaalades. Eeldatakse, et mõõtmistulemused ühes grupis ei sõltu mõõtmistulemustest teises grupis. Tavaliselt on uurimisküsimuseks see, kas populatsioonide keskväärtusvektorid on võrdsed. Profiilianalüüs uurib gruppide keskväärtusvektorite võrdsust lähtudes kolmest aspektist: keskväärtuste profiilide paralleelsus, kokkulangevus ja horisontaalsus. Peatükk põhineb allikatel Johnson & Wichern (2002) ja Kollo (2012).

Profiili mõiste selgitamiseks vaadatakse grupis tehtud 4 erinevat mõõtmist. Olgu grupi keskväärtusvektor esitatud kujul  $\mu_1^T = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14})$ , kus  $\mu_{1j}$ ,  $j = 1, \dots, 4$  on  $j$ -nda mõõtmise tulemuste keskmine. Sageli esitatakse need neli keskmist graafikul, kus nad on pideva joonega ühendatud. Profiiliks nimetataksegi murdjoont, mille  $j$ -nda tipu vertikaalkoordinaadiks on  $\mu_{1j}$ , horisontaalkoordinaadiks aga  $j$ :



Joonis 1. Esimese grupi profiil nelja mõõtmisega

Edasi esitatakse teooria üldisel kujul. Olgu sõltumatud valimid suurusega  $n_i$ ,  $i = 1, 2, \dots, g$  erinevatest populatsioonidest ehk gruppidest. Eeldatakse, et igas grupis on mõõdetud vektor normaaljaotusega, keskväärtusvektoriga  $\mu_i$ ,  $i = 1, 2, \dots, g$  ja kovariatsioonimaatriksiga  $\Sigma_1 = \dots = \Sigma_g$ . Profiili on võimalik leida igale vaadeldavale grupile. Kõigepealt uuritakse

ainult 2 gruppi. Olgu  $\boldsymbol{\mu}_1^T = (\mu_{11}, \mu_{12}, \dots, \mu_{1p})$  ja  $\boldsymbol{\mu}_2^T = (\mu_{21}, \mu_{22}, \dots, \mu_{2p})$  kahe grupi keskväärtusvektorid  $p$  mõõtmisega. Soovitakse kontrollida hüpoteesi  $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  ehk kahe grupi keskväärtusvektorite võrdsust. Selle hüpoteesi kehtivust uuritakse sammhaaval:

1. Kas profiilid on paralleelsed?  
Ehk kas hüpotees  $H_{01}: \mu_{1i} - \mu_{1i-1} = \mu_{2i} - \mu_{2i-1}, i = 2, 3, \dots, p$  kehtib  $\forall i$  korral?
2. Eeldades, et profiilid on paralleelsed, kas saab öelda, et nad on ühtivad?  
Ehk kas hüpotees  $H_{02}: \mu_{1i} = \mu_{2i}, i = 1, 2, \dots, p$  kehtib?
3. Eeldades, et profiilid on ühtivad, kas saab öelda, et kõik keskväärtused on võrdsed?  
Ehk kas hüpotees  $H_{03}: \mu_{11} = \mu_{12} = \dots = \mu_{1p} = \mu_{21} = \mu_{22} = \dots = \mu_{2p}$  kehtib?

Esimesel sammul on võimalik hüpotees sõnastada kasutades maatrikskuju:

$$H_{01}: \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2,$$

kus  $\mathbf{C}: (p - 1) \times p$  on kontrastimaatriks

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \quad (3.1)$$

Esimest hüpoteesi saab kontrollida, moodustades tunnused  $\mathbf{C}\mathbf{x}_{1j}, j = 1, 2, \dots, n_1$  ja  $\mathbf{C}\mathbf{x}_{2j}, j = 1, 2, \dots, n_2$ , kus  $\mathbf{x}_{1j} = (x_{1j1}, x_{1j2}, \dots, x_{1jp})^T$ . Nende keskväärtusvektorid on vastavalt  $\mathbf{C}\bar{\mathbf{x}}_1$  ja  $\mathbf{C}\bar{\mathbf{x}}_2$  ning ühine kovariatsioonimaatriks on  $\mathbf{C}\mathbf{S}_{\text{ühine}}\mathbf{C}^T$ , kus

$$\mathbf{S}_{\text{ühine}} = \frac{\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)^T + \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)^T}{(n_1 + n_2 - 2)}.$$

Kasutades omadust 6.2 raamatust Johnson & Wichern (2002), mis defineerib  $T^2$ -statistiku jaotuse, saab paralleelsuse testimiseks järgmise eeskirja:  $H_{01}: \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$  lükatakse ümber olulisusenivool  $\alpha$  kui

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{C}^T \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{C}\mathbf{S}_{\text{ühine}}\mathbf{C}^T \right]^{-1} \mathbf{C} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > c^2,$$

kus

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} \bar{F}_{p-1, n_1+n_2-p}(\alpha).$$

Kui profiilid on paralleelsed, siis kehtib üks kahest võimalusest: kas esimene asub y- telje mõttes kõrgemal kui teine ( $\mu_{1i} > \mu_{2i} \forall i$ ) või vastupidi. Sellise omaduse kehtides saavad profiilid olla ühtivad vaid siis, kui  $\mu_{11} + \mu_{12} + \dots + \mu_{1p} = \mathbf{1}^T \boldsymbol{\mu}_1$  ja  $\mu_{21} + \mu_{22} + \dots + \mu_{2p} = \mathbf{1}^T \boldsymbol{\mu}_2$  on võrdsed. Seega saab teisel sammul kontrollitava hüpoteesi kirja panna järgmiselt:

$$H_{02}: \mathbf{1}^T \boldsymbol{\mu}_1 = \mathbf{1}^T \boldsymbol{\mu}_2.$$

Kahe normaaljaotusega populatsiooni korral öeldakse, et  $H_{02}: \mathbf{1}^T \boldsymbol{\mu}_1 = \mathbf{1}^T \boldsymbol{\mu}_2$  ei kehti olulisusnivool  $\alpha$  kui

$$\begin{aligned} T^2 &= \mathbf{1}^T (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{1}^T \mathbf{S}_{\text{ühine}} \mathbf{1} \right]^{-1} \mathbf{1}^T (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ &= \left( \frac{\mathbf{1}^T (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left( \frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{1}^T \mathbf{S}_{\text{ühine}} \mathbf{1}}} \right)^2 > \bar{t}_{n_1+n_2+2}^2 \left( \frac{\alpha}{2} \right) = \bar{F}_{1, n_1+n_2-2}(\alpha). \end{aligned}$$

Ühtivate profiilide korral on  $\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1}$  ja  $\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2}$  vaatlused pärit samast normaaljaotusest. Viimase sammuna saab uurida, kas kõik keskvaartused on võrdsed.

Kui hüpoteesid  $H_{01}$  ja  $H_{02}$  kehtivad, siis hinnatakse kahe valimi pealt ühine keskvaartusvektor  $\boldsymbol{\mu}$ , kasutades kõiki  $n_1 + n_2$  vaatlust:

$$\bar{\mathbf{x}} = \frac{1}{n_1 + n_2} \left( \sum_{j=1}^{n_1} \mathbf{x}_{1j} + \sum_{j=1}^{n_2} \mathbf{x}_{2j} \right) = \frac{n_1}{n_1 + n_2} \bar{\mathbf{x}}_1 + \frac{n_2}{n_1 + n_2} \bar{\mathbf{x}}_2.$$

Kui ühtivad profiilid on x-teljega paralleelsed, siis peavad  $\mu_1 = \mu_2 = \dots = \mu_p$  ja nullhüpotees on esitatav kujul  $H_{03}: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , kus on  $\mathbf{C}$  on maatriks, defineeritud valemiga 3.1. Hüpoteesi  $H_{03}$  kontrollimiseks sobib järgmine eeskiri.

Kahe normaaljaotusest pärit populatsiooni korral kummutatakse hüpotees  $H_{03}: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$  olulisusenivool  $\alpha$ , kui

$$(n_1 + n_2) \bar{\mathbf{x}}^T \mathbf{C}^T [\mathbf{C}\mathbf{S}\mathbf{C}^T]^{-1} \mathbf{C}\bar{\mathbf{x}} > c^2,$$

kus  $\mathbf{S}$  on valimi kovariatsioonimaatriks, mis on arvatud kogu vaatluste hulga  $n_1 + n_2$  pealt

$$\text{ja } c^2 = \frac{(n_1+n_2-1)(p-1)}{n_1+n_2-p+1} \bar{F}_{p-1, n_1+n_2-p+1}(\alpha).$$

Juhul, kui on vaja võrrelda rohkem kui 2 grupi profiile, tuleb Hotellingu  $T^2$ -testide asemel kasutada mitmemõõtmelise dispersioonanalüüsi meetodit, mis on toodud järgnevas peatükis.

## 4. Ühefaktoriline MANOVA

Mitmemõõtmelise dispersioonianalüüsi ehk MANOVA kirjeldus põhineb allikatel Fujikoshi, Ulyanov, & Shimizu (2010) ja Johnson & Wichern (2002).

Sageli on tarvis võrrelda rohkem kui 2 populatsiooni, millel on tehtud  $p$  mõõtmist. Juhuslikud valimid, mis pärinevad  $g$  populatsioonist, olgu järjestatud järgmiselt:

populatsioon 1:  $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$

populatsioon 2:  $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$

⋮

populatsioon  $g$ :  $\mathbf{X}_{g1}, \mathbf{X}_{g2}, \dots, \mathbf{X}_{gn_g}$ ,

kus  $\mathbf{X}_{lj} = (X_{lj1}, \dots, X_{ljp})^T$ . Ühefaktorilise MANOVA läbiviimisel eeldatakse, et kehtivad järgnevad tingimused:

1. Olgu  $\mathbf{X}_{l1}, \mathbf{X}_{l2}, \dots, \mathbf{X}_{ln_l}$  juhuslikud valimid suurusega  $n_l$  populatsioonist, mille keskvärtusvektoriks on  $\boldsymbol{\mu}_l$ ,  $l = 1, 2, \dots, g$ . Valimid erinevatest populatsioonidest on sõltumatud. Olgu  $n = n_1 + n_2 + \dots + n_l$ .
2. Kõikidel populatsioonidel on sama kovariatsioonimaatriks  $\boldsymbol{\Sigma}$ .
3. Kõik populatsioonid on mitmemõõtmelise normaaljaotusega  $N_p(\boldsymbol{\mu}_l, \boldsymbol{\Sigma})$ .

MANOVA mudel  $g$  populatsiooni keskvärtusvektorite võrdlemiseks on järgmine:

$$\mathbf{X}_{lj} = \boldsymbol{\mu} + \boldsymbol{\alpha}_l + \boldsymbol{\varepsilon}_{lj} = \boldsymbol{\mu}_l + \boldsymbol{\varepsilon}_{lj}, \quad j = 1, 2, \dots, n_l \text{ ja } l = 1, 2, \dots, g$$

kus  $\boldsymbol{\varepsilon}_{lj}$  on sõltumatud, jaotusega  $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ . Vektori  $\boldsymbol{\varepsilon}_{lj}$  komponendid on omavahel korreleeritud. Mudelis olev vektor  $\boldsymbol{\mu}$  tähistab üldkeskmist ning  $\boldsymbol{\alpha}_l$  tähistab  $l$ -inda grupi (näiteks ravi) efekti. Ühese lahendi saamiseks kasutatakse veel kitsendust  $\sum_{l=1}^g n_l \boldsymbol{\alpha}_l = \mathbf{0}$ . Sellise kitsenduse kasutamisel on  $\boldsymbol{\mu}$  defineeritud kui  $\boldsymbol{\mu} = \frac{1}{n} \sum_{l=1}^g n_l \boldsymbol{\mu}_l$ . Mitmemõõtmelise dispersioonianalüüsi korral on eesmärgiks kontrollida, kas nullhüpootees  $H_0: \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \dots = \boldsymbol{\alpha}_g = \mathbf{0}$ , mis on samaväärne  $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_g$  võiks kehtida. Selleks esitatakse kõigepealt üks vaatlusvektor lahtikirjutatult, kasutades eelpool toodud MANOVA mudelit:

$$\mathbf{x}_{lj} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)$$

ehk

vaatlustvektor = üldkeskmine  $\hat{\boldsymbol{\mu}}$  + grupi hinnatud efekt  $\hat{\boldsymbol{\alpha}}_l$  + jääk  $\hat{\boldsymbol{\epsilon}}_{lj}$ .

Sarnaselt ühemõõtmelisele juhule, saab hälvete ruutude summa esitada kujul

$$\begin{aligned} (\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})^T &= [(\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)][(\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)]^T \\ &= (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)^T + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)^T \\ &\quad + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T. \end{aligned}$$

Summeerides üle  $j$  nähakse, et 2 keskmist liiget on avaldises nullvektorid, sest  $\sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) = \mathbf{0}$ . Seega summeerides üle  $l$  ja  $j$  saame järgmise tulemuse:

$$\begin{aligned} &\underbrace{\sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})^T}_{\text{SST}} \\ &= \underbrace{\sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T}_{\text{SSB}} + \underbrace{\sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)^T}_{\text{SSW}}, \end{aligned}$$

kus võrrandi vasakpoolne osa on summaarne hälvete ruutude summa ehk **SST** (ing k. *Total sum of squares and products*) ning parempoolne osa jaguneb grupi efekti hindavaks hälvete ruutude summaks ehk **SSB** (ing k. *Between sum of squares and products*) ja grupi siseste hälvete ruutude summaks **SSW** (ing k. *Within sum of squares and products*). Maatriks **SSW** on ka esitatav kujul  $\mathbf{SSW} = (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \dots + (n_g - 1)\mathbf{S}_g$ , kus  $\mathbf{S}_l$  on  $l$ -inda valimi kovariatsioonimaatriks.

Maatriksitest **SSB** ja **SSW** saab konstrueerida statistiku, millega eelpool nimetatud nullhüpoteesi kontrollida. Alternatiivne hüpotees leiab tõestust, kui üldistatud dispersioonide suhe

$$A = \frac{|\mathbf{SSW}|}{|\mathbf{SSW} + \mathbf{SSB}|}$$

on väike. Wilks'i lambda ehk  $\Lambda$  jaotus sõltub võrreldavate gruppide ja mõõdetud tunnuste arvust. Juhul, kui võrreldavaid gruppe ja tunnuseid on palju (grupe rohkem kui 3, tunnuseid rohkem kui 1), siis teisendatakse  $\Lambda$  teguriga  $-\left(n - 1 - \frac{(p+g)}{2}\right)$ , mille on kirja pannud Bartlett (Bartlett, 1938). Ta on näidanud, et juhul, kui  $H_0$  kehtib ning  $n$  on suur, siis modifitseeritud statistik

$$-\left(n - 1 - \frac{(p + g)}{2}\right) \ln \Lambda = -\left(n - 1 - \frac{(p + g)}{2}\right) \ln \left( \frac{|\mathbf{SSW}|}{|\mathbf{SSW} + \mathbf{SSB}|} \right)$$

on ligilähedasel hii-ruut jaotusega, vabadusastmete arvuga  $p(g - 1)$ . Seega lükatakse nullhüpotees ümber olulisusenivool  $\alpha$ , kui

$$-\left(n - 1 - \frac{(p + g)}{2}\right) \ln \left( \frac{|\mathbf{SSW}|}{|\mathbf{SSW} + \mathbf{SSB}|} \right) > \bar{\chi}_{p(g-1)}^2(\alpha),$$

kus  $\bar{\chi}_{p(g-1)}^2(\alpha)$  on hii-ruut jaotuse  $\alpha$ -täiendkvantil.

Juhul, kui gruppe või tunnuseid on väga vähe, saab  $\Lambda$ -le leida ka täpse jaotuse raamatust Johnson & Wichern (2002) või vaadata tema kriitilisi väärtusi raamatust Rencher (2002).

Veel üks võimalus Wilks'i lamdat esitada on kasutada maatriksi  $\mathbf{SSW}^{-1}\mathbf{SSB}$  omaväärtusi  $\lambda_1, \lambda_2, \dots, \lambda_s$ . Siis esitub statistik kujul  $\Lambda = \prod_{i=1}^s \left( \frac{1}{1+\lambda_i} \right)$ , kus  $s = \min(p, g - 1)$  on maatriksi  $\mathbf{SSB}$  astakuks.

## 5. Mitmemõõtmeline korduvmõõtmise mudel

Eelnevalt on töös antud ülevaade olukordadest, kus võrreldavaid gruppe (ravimeid) on 2 ning mõõdetud on  $p$  tunnust või võrreldavaid ravimeid on  $q$  ning mõõdetud tunnuseid on vaid 1. Siinkohal antakse ülevaade olukorrast, kus igal indiviidil mõõdetakse rohkem kui ühte tunnust mitmel erineval ajahetkel ning võrreldavaid gruppe (näiteks ravimeid) võib olla  $r$ . Ehk siis on tegemist mitmemõõtmelise korduvmõõtmisega. Lihtsustamiseks eeldatakse, et andmeteale sobib mitmemõõtmeline juhumõjudega korduvmõõtmiste mudel. Peatükk põhineb allikal Reinsel (1982) ja töö autor on kasutanud samu tähistusi, kui artiklis kasutati. Töös esitatud parameetrite hinnangute tuletamised on lisas 5.

### 5.1 Mudeli kujud ja parameetrite hinnangud

Olgu igal indiviidil  $p$  erineval ajahetkel mõõdetud  $m$  tunnust. On lubatud, et kõikidel ajahetkedel ei pruugi katsetingimused olla ühesugused. Eeldatakse, et katses osalevad  $N$  indiviidi randomiseeritakse katsesse nii, et erinevate indiviidide mõõtmised on sõltumatud. Tähistagu  $y_{ijk}$   $i$ -nda tunnuse  $j$ -ndat mõõtmist  $k$ -ndal indiviidil ( $i = 1, \dots, m$ ,  $j = 1, \dots, p$ ,  $k = 1, \dots, N$ ) ja olgu  $\mathbf{y}_{jk} = (y_{1jk}, y_{2jk}, \dots, y_{mjk})^T$ . Eeldatakse, et mõõtmistulemuste vektori  $\mathbf{y}_{jk}$  jaoks on mudel kirjutatav kujul

$$\mathbf{y}_{jk} = \boldsymbol{\theta}_k \mathbf{X}_j + \mathbf{u}_{jk}, \quad j = 1, \dots, p, \quad k = 1, \dots, N,$$

kus  $\mathbf{X}_j$ -s on kirjas ajale vastavad kontrastid ning  $\boldsymbol{\theta}_k$  on tundmatute parameetrite maatriks mõõtmetega  $m \times p$ . Mudel  $k$ -nda indiviidi kõigile  $p$  ajahetkele on esitatav maatrikskujul

$$\mathbf{Y}_k = \boldsymbol{\theta}_k \mathbf{X}^T + \mathbf{U}_k, \quad k = 1, \dots, N, \quad (5.1.1)$$

kus  $\mathbf{X}^T = (\mathbf{X}_1, \dots, \mathbf{X}_p)$  on teadaolev  $q \times p$  maatriks, mis on täisastakuga  $q \leq p$ . Edaspidiseks kasutamiseks tuuakse sisse maatriksalgebras kasutatav *vec*-operaator. Olgu  $\mathbf{y}_k = \text{vec}(\mathbf{Y}_k) = (\mathbf{y}_{1k}^T, \dots, \mathbf{y}_{pk}^T)^T$ , siis saab (5.1.1) esitada vektorkujul

$$\mathbf{y}_k = (\mathbf{X} \otimes \mathbf{I}_m) \boldsymbol{\phi}_k + \mathbf{u}_k, \quad k = 1, \dots, N \quad (5.1.2)$$

kus  $\mathbf{u}_k = \text{vec}(\mathbf{U}_k)$ ,  $\boldsymbol{\phi}_k = \text{vec}(\boldsymbol{\theta}_k)$  ja kasutatud on *vec* – operaatori omadust  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ . Edaspidi eeldatakse, et vektorid  $\boldsymbol{\phi}_k$  on sellised, et  $(\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N) = \mathbf{B} \mathbf{A}^T$ , kus  $\mathbf{B}$  on  $mq \times r$  maatriks tundmatutest parameetritest ja  $\mathbf{A}^T = (\mathbf{a}_1, \dots, \mathbf{a}_N)$  on  $r \times N$  täisastakuga ( $r < N$ ) maatriks teada olevatest konstantidest, mis määrab uuringu disaini. Tähistades  $\mathbf{Y}^T = (\mathbf{y}_1, \dots, \mathbf{y}_N)$  ja  $\mathbf{U}^T = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ , saab mudeli (5.1.2) avaldada järgmiselt:

$$\mathbf{Y}^T = (\mathbf{X} \otimes \mathbf{I}_m) \mathbf{B} \mathbf{A}^T + \mathbf{U}^T.$$

Eeldatakse, et maatriksi  $\mathbf{U}^T$  veerud,  $\mathbf{u}_k = \text{vec}(\mathbf{U}_k)$  on sõltumatud jaotusega  $N_{mp}(\mathbf{0}, \mathbf{\Omega})$ , kus  $\mathbf{\Omega}$  on tundmatu  $mp \times mp$  kovariatsioonimaatriks.

Paljudel praktilistel juhtudel võib dimensioon  $mp$  olla valimimahuga  $N$  võrreldes liiga suur. Sellisel juhul tuleks  $\mathbf{\Omega}$  struktuuri kohta teha täiendavaid eeldusi, et tagada nii parameetermaatriksi  $\mathbf{B}$  kui ka kovariatsioonimaatriksi parameetrite täpsemad hinnangud. Üks võimalikke struktuure, mida saab teatud olukordades pidada sobilikuks, on juhuslike mõjudega mudel:

$$\mathbf{y}_{jk} = \boldsymbol{\theta}_k \mathbf{X}_j + \boldsymbol{\lambda}_k + \boldsymbol{\varepsilon}_{jk} \quad (5.1.3)$$

kus  $\boldsymbol{\lambda}_k$  on  $m \times 1$  vektor  $k$ -nda indiviidi juhuslikest mõjudest, jaotusega  $N_m(\mathbf{0}, \mathbf{\Sigma}_\lambda)$  ning  $\boldsymbol{\varepsilon}_{jk}$  on juhuslikud vead jaotusega  $N_m(\mathbf{0}, \mathbf{\Sigma}_e)$ , mis on sõltumatud  $\boldsymbol{\lambda}_k$ -dest. Sellist struktuuri kasutades saab mudel (5.1.1) kujul

$$\mathbf{Y}_k = \boldsymbol{\theta}_k \mathbf{X}^T + \boldsymbol{\lambda}_k \mathbf{1}^T + \mathbf{E}_k, \quad (5.1.4)$$

kus  $\mathbf{1}$  tähistab ühtede vektorit mõõtmetega  $p \times 1$  ja  $\mathbf{E}_k = (\boldsymbol{\varepsilon}_{1k}, \dots, \boldsymbol{\varepsilon}_{pk})$  ning mudel (5.1.2) saab kujul

$$\mathbf{y}_k = (\mathbf{X} \otimes \mathbf{I}_m) \boldsymbol{\phi}_k + (\mathbf{1} \otimes \mathbf{I}_m) \boldsymbol{\lambda}_k + \mathbf{e}_k, \quad (5.1.5)$$

kus  $\mathbf{e}_k = \text{vec}(\mathbf{E}_k)$  ning  $\text{cov}(\mathbf{y}_k) = \mathbf{\Omega} = (\mathbf{I}_p \otimes \mathbf{\Sigma}_e) + (\mathbf{1} \mathbf{1}^T \otimes \mathbf{\Sigma}_\lambda)$ . Kuna mudel (5.1.3) sisaldab vabaliiget iga  $m$  tunnuse jaoks, siis eeldatakse, et maatriksi  $\mathbf{X}$  esimene veerg koosneb ühtedest, ehk  $\mathbf{X} = (\mathbf{1} \mathbf{Z})$ , kus  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_p)^T$  on  $p \times (q - 1)$  maatriks, mille korral kehtib  $\mathbf{1}^T \mathbf{Z} = \mathbf{0}^T$ . Seega  $\mathbf{1}^T \mathbf{X} = p \mathbf{v}_1^T$ , kus  $\mathbf{v}_1^T = (1 \ 0 \ \dots \ 0)$  on mõõtmetega  $1 \times q$ .

Mudelit (5.1.5) on võimalik esitada kujul

$$\mathbf{Y}^T = (\mathbf{X} \otimes \mathbf{I}_m) \mathbf{B} \mathbf{A}^T + \mathbf{U}^T, \quad (5.1.6)$$

kus  $\mathbf{U}^T = (\mathbf{1} \otimes \mathbf{I}_m) \boldsymbol{\Lambda}^T + \mathbf{E}^T$ ,  $\boldsymbol{\Lambda}^T = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_N)$  ning  $\mathbf{E}^T = (\mathbf{e}_1, \dots, \mathbf{e}_N)$ . Esitades mudeli (5.1.6) vektorkujul, saadakse mudel kujul

$$\mathbf{y} = (\mathbf{A} \otimes \mathbf{X} \otimes \mathbf{I}_m) \boldsymbol{\beta} + \mathbf{u}, \quad (5.1.7)$$

kus  $\mathbf{y} = \text{vec}(\mathbf{Y}^T)$ ,  $\mathbf{u} = \text{vec}(\mathbf{U}^T)$  ja  $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ . Kovariatsioonimaatriks  $\mathbf{V}$  avaldub kujul

$$\begin{aligned}\mathbf{V} = cov(\mathbf{u}) &= (\mathbf{I}_N \otimes \mathbf{\Omega}) = \mathbf{I}_N \otimes \left( (\mathbf{I}_p \otimes \mathbf{\Sigma}_e) + (\mathbf{1}\mathbf{1}^T \otimes \mathbf{\Sigma}_\lambda) \right) \\ &= (\mathbf{I}_N \otimes \mathbf{I}_p \otimes \mathbf{\Sigma}_e) + (\mathbf{I}_N \otimes \mathbf{1}\mathbf{1}^T \otimes \mathbf{\Sigma}_\lambda).\end{aligned}$$

Eelpool toodud eeldusi kasutades on võimalik näidata, et üldistatud vähimruutude hinnang  $\boldsymbol{\beta}$ -le on sama, mis vähimruutude hinnang (Ravishanker & Dey, 2002). Seega saab kasutada suurima tõepära hinnangut  $\boldsymbol{\beta}$ -le

$$\hat{\boldsymbol{\beta}} = ((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \otimes (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \otimes \mathbf{I}_m) \mathbf{y} \quad (5.1.8)$$

ning suurima tõepära hinnang  $\mathbf{B}$ -le on

$$\hat{\mathbf{B}} = ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}. \quad (5.1.9)$$

Seega on  $\hat{\boldsymbol{\beta}}$  mitmemõõtmelise normaaljaotusega, parameetritega  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$  ja kovariatsioonimaatriksiga

$$cov(\hat{\boldsymbol{\beta}}) = ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{X}^T \mathbf{X})^{-1} \otimes \mathbf{\Sigma}_e) + ((\mathbf{A}^T \mathbf{A})^{-1} \otimes \mathbf{v}_1 \mathbf{v}_1^T \otimes \mathbf{\Sigma}_\lambda). \quad (5.1.10)$$

Maatriksi  $\mathbf{B}$ , mis on mõõtmetega  $m q \times r$ , saab jagada kaheks:  $\mathbf{B} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\Gamma} \end{pmatrix}$ , kus  $\boldsymbol{\mu}$  on  $m \times r$  ja  $\boldsymbol{\Gamma}$  on  $m(q-1) \times r$ . Maatriksi  $\boldsymbol{\mu}$  asuvad elemendid vastavad vabaliikmetele mudelis (5.1.6). Kui jagada maatriks  $\hat{\mathbf{B}}$  sarnaselt maatriksiga  $\mathbf{B}$ , siis on  $\boldsymbol{\Gamma}$  hinnang

$$\hat{\boldsymbol{\Gamma}} = ((\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}$$

ja vektor  $\hat{\boldsymbol{\gamma}} = vec(\hat{\boldsymbol{\Gamma}})$  on jaotusega  $N_{m(q-1)r}(\boldsymbol{\gamma}, (\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \mathbf{\Sigma}_e)$ , kus  $\boldsymbol{\gamma} = vec(\boldsymbol{\Gamma})$ . Vabaliikmete maatriksi hinnang avaldub

$$\hat{\boldsymbol{\mu}} = \frac{1}{p} (\mathbf{1}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} = \bar{\mathbf{Y}}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1},$$

kus  $\bar{\mathbf{Y}}^T = (\bar{\mathbf{y}}_{\cdot 1}, \dots, \bar{\mathbf{y}}_{\cdot N})$  on indiviidide keskväärtusvektorite maatriks ning  $\bar{\mathbf{y}}_{\cdot k}$  on esitatud

$$\bar{\mathbf{y}}_{\cdot k} = \frac{1}{p} \sum_{j=1}^p \mathbf{y}_{jk}, k = 1, \dots, N \quad (5.1.11)$$

ja  $vec(\hat{\boldsymbol{\mu}})$  on jaotusega  $N_{mr} \left( vec(\boldsymbol{\mu}), \frac{1}{p} (\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{\Sigma}_e + p \mathbf{\Sigma}_\lambda) \right)$ .

Parameetrite maatriksi või vektori kohta käivate hüpoteeside testimiseks on vaja leida kovariatsioonimaatriksite  $\mathbf{\Sigma}_e$  ja  $\mathbf{\Sigma}_\lambda$  hinnangud. Jääkide ruutude summa maatriksi (ingl k. *error sum of squares*) saab avaldada järgmiselt:

$$\mathbf{S}_e = \sum_{k=1}^N \sum_{j=1}^p (\mathbf{y}_{jk} - \bar{\mathbf{y}}_{.k} - (\mathbf{Z}_j^T \otimes \mathbf{I}_m) \hat{\boldsymbol{\Gamma}} \mathbf{a}_k) (\mathbf{y}_{jk} - \bar{\mathbf{y}}_{.k} - (\mathbf{Z}_j^T \otimes \mathbf{I}_m) \hat{\boldsymbol{\Gamma}} \mathbf{a}_k)^T, \quad (5.1.12)$$

mis on Wisharti jaotusega  $W_m(\boldsymbol{\Sigma}_e, N(p-1) - r(q-1))$  ning sõltumatu  $\hat{\boldsymbol{\Gamma}}$ -st (Ravishanker & Dey, 2002). Grupi keskmiste hälbumuste ruutude summa maatriks (ingl k. *between-groups sum of squares*) on avaldatav

$$\mathbf{S}_\lambda = p \sum_{k=1}^N (\bar{\mathbf{y}}_{.k} - \hat{\boldsymbol{\mu}} \mathbf{a}_k) (\bar{\mathbf{y}}_{.k} - \hat{\boldsymbol{\mu}} \mathbf{a}_k)^T \quad (5.1.13)$$

ning see on Wisharti jaotusega, parameetritega  $(\boldsymbol{\Sigma}_e + p\boldsymbol{\Sigma}_\lambda, N-r)$  ja sõltumatu  $\hat{\boldsymbol{\mu}}$ -st. Seega, nihketa hinnangud  $\boldsymbol{\Sigma}_e$  ja  $\boldsymbol{\Sigma}_\lambda$  jaoks on

$$\hat{\boldsymbol{\Sigma}}_e = \mathbf{S}_e / [(N(p-1) - r(q-1))] \quad (5.1.14)$$

ja

$$\hat{\boldsymbol{\Sigma}}_\lambda = \frac{1}{p} [\mathbf{S}_\lambda / (N-r) - \hat{\boldsymbol{\Sigma}}_e]. \quad (5.1.15)$$

## 5.2 Hüpoteeside testimine

Artiklis tuuakse välja  $\boldsymbol{\mu}$  ja  $\boldsymbol{\Gamma}$  kohta käivate lineaarsete hüpoteeside kontrollimiseks eraldi testid. Et hilisemas analüüsis ei olda huvitatud  $\boldsymbol{\mu}$  kohta käivate hüpoteeside testimisest, siis tuuakse siin peatükis välja vaid  $\boldsymbol{\Gamma}$  kohta käivate hüpoteeside kontrollimiseks vajalik test. Üldiselt saaks mõlema parameetri kohta käivaid teste kontrollida, kasutades MANOVAt.

Parameetri  $\boldsymbol{\Gamma}$  kohta käivate hüpoteeside testimiseks tuleb esmalt  $\boldsymbol{\Gamma}$  jaotada  $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1^T, \dots, \boldsymbol{\Gamma}_{q-1}^T)^T$ , kus  $\boldsymbol{\Gamma}_i$ ,  $i = 1, \dots, q-1$  on  $m \times r$  maatriks, sisaldades regressiooni-koefitsiente  $i$ -ndatele komponentidele  $\mathbf{Z}_j$ -st. Lihtsuse huvides esitatakse koefitsientide maatriksi kujul  $\boldsymbol{\Psi} = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \dots, \boldsymbol{\Gamma}_{q-1})$ . Lineaarsete hüpoteeside kontrollimiseks antakse test kujul

$$H_0: \mathbf{M}\boldsymbol{\Psi}\mathbf{C}^T = \mathbf{0},$$

kus  $\mathbf{M}$  on  $c \times m$  maatriks astakuga  $c \leq m$  ja  $\mathbf{C}^T$  on maatriks mõõtmetega  $r(q-1) \times g$  ning astakuga  $g \leq r(q-1)$ . Nüüd  $\text{vec}(\hat{\boldsymbol{\Psi}}) \sim N(\text{vec}(\boldsymbol{\Psi}), (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes (\mathbf{A}^T \mathbf{A})^{-1} \otimes \boldsymbol{\Sigma}_e)$  ning ta on sõltumatu maatriksist  $\mathbf{S}_e$ . Seega,  $H_0$  kehtides on  $\hat{\boldsymbol{\Delta}} = \mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T$  sellise jaotusega, et  $\text{vec}(\hat{\boldsymbol{\Delta}}) = (\mathbf{C} \otimes \mathbf{M})\text{vec}(\hat{\boldsymbol{\Psi}})$  jaotuseks on  $N_{cg}(\mathbf{0}, \mathbf{C}((\mathbf{Z}^T \mathbf{Z})^{-1} \otimes (\mathbf{A}^T \mathbf{A})^{-1})\mathbf{C}^T \otimes \mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T)$ . Kontrollimaks püstitatud  $H_0$ , kasutatakse maatriksit  $\mathbf{H}$ , milleks on

$$\mathbf{H} = \mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T (\mathbf{C}((\mathbf{Z}^T \mathbf{Z})^{-1} \otimes (\mathbf{A}^T \mathbf{A})^{-1})\mathbf{C}^T)^{-1} \mathbf{C}\hat{\boldsymbol{\Psi}}^T \mathbf{M}^T,$$

kus  $\mathbf{H}$  on Wisharti jaotusega, parameetritega  $(\mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T, g)$ , olles sõltumatu maatriksist  $\mathbf{G} = \mathbf{M}\mathbf{S}_e\mathbf{M}^T$ , mis on Wisharti jaotusega  $(\mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T, N(p-1) - r(q-1))$ . Seega, nullhüpoteesi kontrollimiseks võib kasutada Wilks'i lambda statistikut, nagu toodud peatükis 4:

$$\Lambda = \frac{|\mathbf{G}|}{|\mathbf{G} + \mathbf{H}|}.$$

## 6. Andmete analüüs

### 6.1 Andmete bioloogiline taust

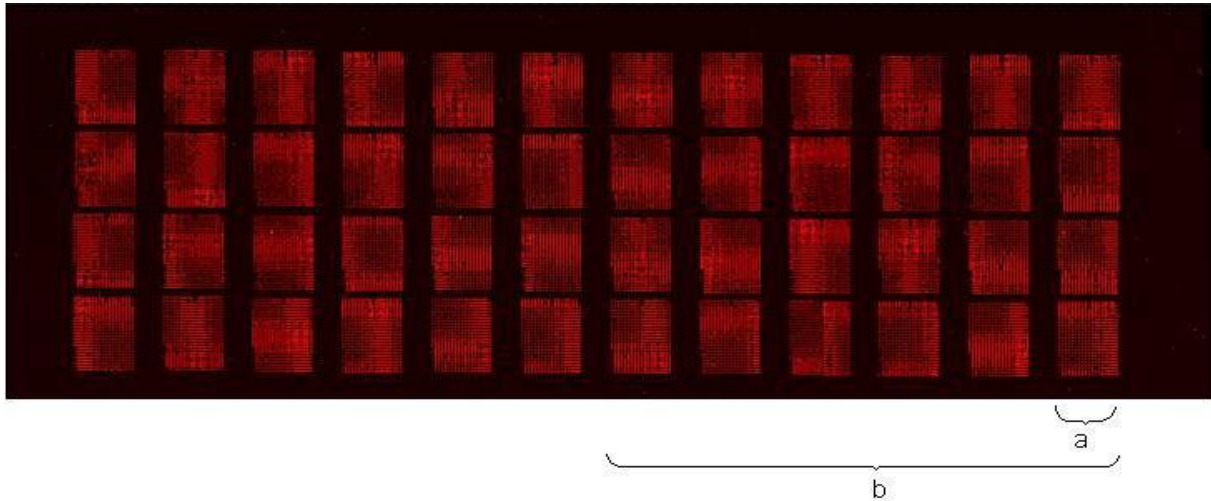
Järgnevas ülevaates andmete bioloogilisest taustast ja peptiidide mikrokiibi tehnoloogiast põhineb bioloogiline osa Sikut (2001) loengukonspektil, Siegrist (2008) raportil, Valge, (2012) magistritööl ning mikrokiipide osa artiklil Brus (2009) ning Nahtman *et al.* (2007).

Inimene puutub igapäevaselt kokku mitmete haigustekitajatega (viirused, mikroobid), mis kutsuvad organismi sattudes esile immuunvastuse. Antigeen on organismi potentsiaalselt kahjustav kehavõõras aine, mille sissetungimine organismi põhjustab nende vastu suunatud antikehade tekke. Antigeenid võivad organismi tungida naha, limaskestade, hingamis- ja seedetrakti kaudu. Esmaselt puutuvad antigeenid kokku naha ja limaskestadega, mis takistavad suurel osal neist organismi sattumast. Kui nad aga esmasest kaitsebarjäärist läbi tungivad, hakkavad tööle immunoloogilised kaitsemehhanismid. Organismi esimene kaitse võõrorganismide vastu toimub fagotsüütide abil. Need rakud on spetsialiseerunud ebasoovitavate materjalide sissevõtmiseks ja nende lagundamiseks. Peamised fagotsüteerivad rakud on neutrofiilid ja makrofaagid. Rakulist immuunsust kannavad veel ka nn T-lümfotsüüdid, mille pinnal on retseptorid, mis tunnevad ära viiruste või mikroobide poolt nakatatud rakke. Naturaalkillerid on samuti organismis üks oluline rakuline immuunsuse kandja. T-lümfotsüüdid ja naturaalkiller-rakud täiendavad teineteist: need nakatatud rakud, mis jäävad märkamatuks T-rakkude poolt, äratavad naturaalkiller-rakkude tähelepanu, ja vastupidi.

Peaaegu iga võõras makromolekulaarne ühend kutsub organismi sattudes esile antikehade sünteesi ja vastavate T-rakkude proliferatsiooni ehk T-rakkude vohamise. Immuunvastus on spetsiifiline, s.t. et tekkinud antikehad ja T-lümfotsüüdid tunnevad spetsiifiliselt ära selle aine, mis kutsus esile nende sünteesi. See aga võtab aega. Kuna osad patogeenid on organismile nii ohtlikud, et ei saa oodata, kuni tekib spetsiifiline immuunvastus, siis selleks on olemas mittespetsiifilised kaitstesüsteemid, mis lähevad käiku kohe. Siia kuuluvad eelpoolmainitud fagotsüüdid, komplemendi valgud ja naturaalkiller-rakud. Sellisele mittespetsiifilisele immuunvastusele järgneb spetsiifiline immuunvastus.

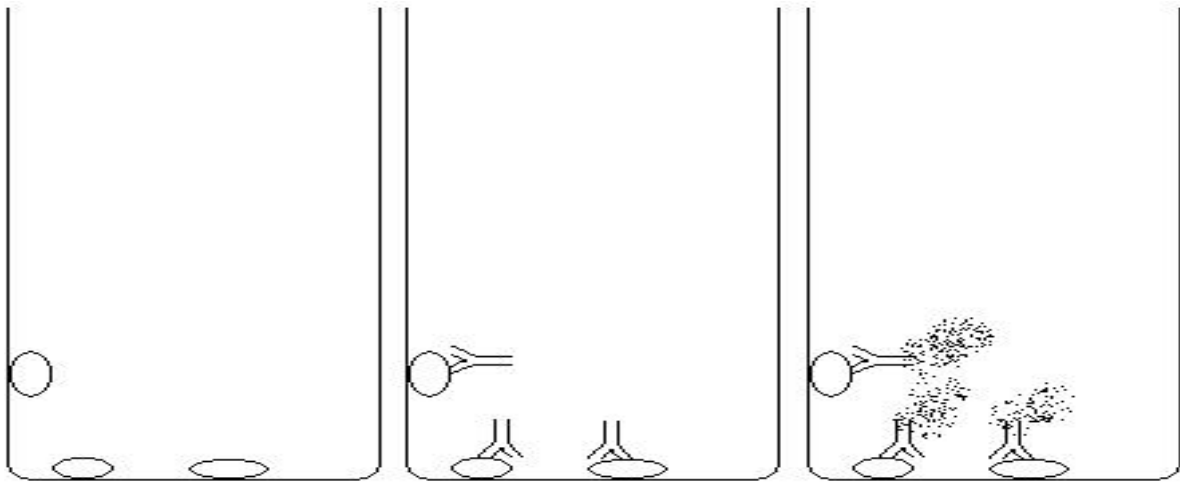
Haiguse/viiruse leviku kontrollimiseks või täielikuks kadumiseks on vajalik, et teatud osa populatsioonist omaks selle vastu immuunsust. Sellise olukorra saavutamiseks on tavaliselt tarvis tehisklikke abivahendeid, millel on võime tekitada organismi pikaajaliselt kestev kaitse. Immuunsuse kestvus sõltub paljudest teguritest, näiteks vaktsiini liigist, vaktsineerimise

vanusest ja konkreetsetest antigeenidest. Teada on, et passiivne immuunsus tuberkuloosi vastu (emalt saadud antikehad või kehasse süstitud antikehad) ei ole efektiivne. Seetõttu peab tuberkuloosivastased antikehad iga organism tekitama ise. Selleks, et otsustada, kui tihedalt ja millise vaktsiiniga peaks inimesi vaksineerima, on tarvis teada, kuidas erinevad antigeenid antikehade tekkimist mõjutavad ning kui pikalt antikehad inimese organismis püsivad. Antikehade-antigeenide kontsentratsiooni mõõtmiseks on kasutusele võetud peptiidide mikrokiibid.



**Joonis 2. Klaasplaadi näidis. Klaasplaadil on kaks alamplaati (osa b), mõlemad sisaldavad 24 plokki (osa a), millest igaüks koosneb 18x18 täpist. (Brus, 2009)**

Järgnevalt kirjeldatakse peptiidide mikrokiipi tööpõhimõtet (vt joonis 2). Mikrokiip on põhimõtteliselt klaasplaat, milles on tuhandeid süvendeid. Iga klaasplaat koosneb tihti mitmest alamplaadist (ing *sub-array*), mis on omakorda jagatud plokkideks (ing *block*). Plokk koosneb täppide maatriksist, kus igale täpile on määratud kindel peptiid. Kordusmõõtmiste eesmärgil võib mitmel täpil esineda sama peptiid. Veel on kiibil osad täpid jäetud kontrollmõõtmisteks, et näha, kas kiip töötab korrektselt. Negatiivsed kontrollid on need, mis mõõtmisel ei tohiks anda mingit signaali ning positiivsed kontrollid täpid, mis peaksid kindlasti andma tugeva signaali. Kontrolltäppe andmete analüüsis ei kasutata.



**Joonis 3. Vasakul asuvas süvendisse on asetatud fikseeritult antigeenid. Keskmises süvendis on antigeenid antikehadega seostunud. Parempoolses süvendis on antikehad reageerinud fluorestseeriva lahusega, olles nüüd valgussignaalina mõõdetavad. (Brus, 2009)**

Antikehade olemasolu testimiseks kaetakse mikrokiibil asuvad täpid antigeenidega, mis peaks uuritavad antikehad esile kutsuma (vt joonis 3). Seejärel kaetakse mikrokiip seerumiga. Kui seerum sisaldab antigeenidele vastavaid antikehi, siis nad seostuvad omavahel. Seostumata antikehad pestakse plaadilt maha. Seejärel pannakse mikrokiibile fluorestseeriv lahus, mis seostub antikehadega ning üleliigne lahus eemaldatakse jällegi plaadi pesemisega. Viimaks mõõdetakse spetsiaalse skanneriga igast täpist ja tema ümbrusest antikehade kontsentratsiooni, kasutades täpi helendamise intensiivsust.

Täpi helendamise intensiivsust, mida mõõdetakse täpist endast, nimetatakse esiplaani mõõtmiseks. Täppi ümbritseva ala helendamise intensiivsust nimetatakse tagaplaani mõõtmiseks. Artiklis Nahtman *et al.* (2007) pakutakse välja, et reageerinud antikehade kontsentratsiooni hindamiseks võiks kasutada valemit:

$$indeks = \log_2 \left( \frac{\text{esiplaani mõõtmine}}{\text{tagaplaani mõõtmine}} \right) \quad (6.1.1)$$

sest selline teisendus eristas hästi negatiivseid kontrolle, positiivseid kontrolle ja tegelikke mõõtmisi ning aitas vähendada algsete andmete positiivset asümmeetriat. Indeksi väärtuseks on null, kui nii esiplaani kui ka tagaplaani mõõtmine on võrdsed. Selline olukord tekib, kui seerumis antigeenidele vastavaid antikehasid polnud. Kui seerumis leidis antigeenidele vastavaid antikehasid, siis peaks esiplaani mõõtmine olema suurem (täpist mõõdetav valgussignaal on intensiivsem) kui tagaplaani mõõtmine.

## 6.2 Andmete kirjeldav analüüs

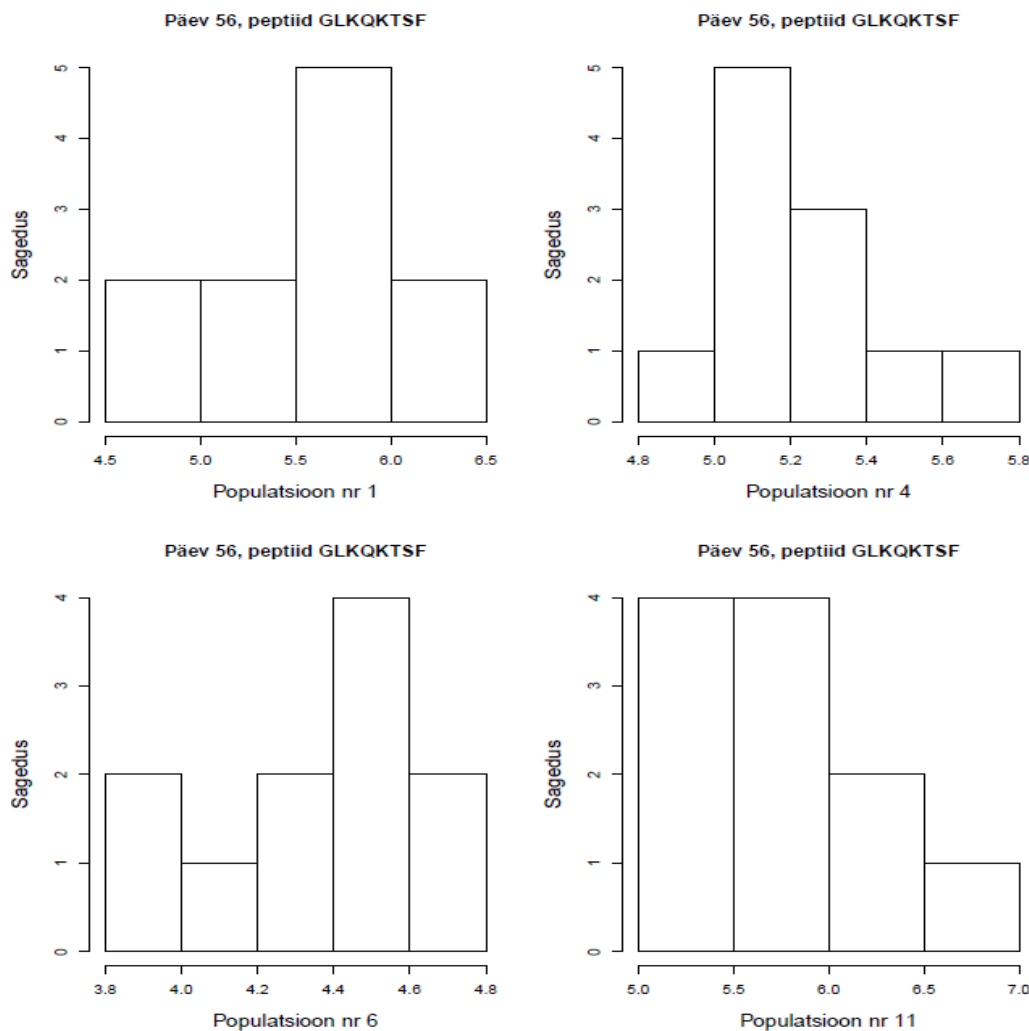
Praktiline osa viiakse läbi Karolinska instituudis kogutud andmestikuga, mis sisaldab tuberkuloosivaktsiini manustanud patsientide mõõtmistulemusi. Andmestik koosneb 4 tunnusest. Tunnus „patient“ näitab patsiendi identifitseerimisnumbrit. Patsientidel on identifitseerimisnumbrid väärtustega 1,101, 102, ..., 110, 4, 401, 402, ..., 410, 6,601, 602, ..., 610 ja 11, 1101, 1102,..., 1110. Tähistused on selliselt valitud, sest algselt oli andmestikus vaid 4 patsienti, identifitseerimisnumbriga 1, 4, 6 ja 11. Nende mõõdetud väärtustest genereeriti juurde 40 indiviidile mõõtmistulemused nii, et nad on üksteisest sõltumatud. Patsiendist 1 genereeriti identifitseerimisnumbrid 101, 102, jne. Järgnevates peatükkides on patsiendist 1 genereeritud indiviididele viidatud kui populatsioonile 1, patsient 4 genereeritud indiviididele kui populatsioonile 4, jne. Seda tehakse ühel põhjusel: uurides, kas mõne peptiidi korral indeksi väärtused on ajas keskmiselt konstantsed, keskmistatakse peptiidi indeksi väärtused eraldi igas populatsioonis. Algsete patsientide kohta pole infot, mis lubaks eeldada, et nende mõõtmistulemused mõnel ajahetkel ei võiks omavahel oluliselt erineda. Kui üks algsetest patsientidest on teistest mingil hetkel üsna erineva käitumisega, ei pruugi üle kõikide indiviidide arvatud indeksi keskmine väljendada üldist tendentsi. Samas pole ühegi algse patsiendi puhul alust väita, et tegemist on erindiga, mille peaks analüüsist välja jätma. Tunnus „ID“ näitab kasutatud peptiidi nime. Kokku on algses andmestikus 1033 erinevat peptiidi. Tunnus „index“ näitab igale peptiidile vastavat mõõtmistulemust, mis on saadud, kasutades valemit (6.1.1). Mida kõrgem on peptiidile vastav indeks, seda rohkem antikehi selle peptiidiga reageeris. Tunnus „day“ näitab mõõtmiste tegemiste aega. Algses andmestikus mõõdeti patsiente 4 päeval: 0, 56, 112 ja 252 päeval. Päev 182 genereeriti 112. ja 252. päeva vaatluste põhjal. Nullinda päeva mõõtmine on tehtud vahetult enne vaktsineerimist.

Andmestiku suur puudus on väga palju puuduvaid väärtusi. Vaid 14 peptiidil oli mõõdetud indeksi väärtused kõigil 5 ajahetkel kõigi 44 patsiendi jaoks. Järgnev analüüs viiaksegi läbi kasutades neid 14 peptiidi, sest vastasel korral oleks valimimaht võrreldes mõõdetud tunnuste arvuga liiga väike. Samuti on kaasatud ainult 3 esimest ajahetke, sest meid huvitab peptiidide esmane käitumine pärast vaktsineerimist. Kirjeldamiseks kõikide valitud peptiidi käitumist, esitatakse nende profiilid kõigis neljas populatsioonis (vt lisa 3 ja lisa 4). Graafikuid vaadates märgati, et patsiendist number 4 genereeritud indiviidide käitumine on ülejäänud valimist mõnevõrra erinev - kõikide mõõdetud peptiidide korral on indeksi väärtus teiseks ajahetkeks

jäänud enam-vähem samaks esimese ajahetkega võrreldes või veidi tõusnud ning siis 3-ndaks ajahetkeks järsult langenud. Teiste populatsioonide puhul erines indeksi käitumine peptiiditi.

### 6.3 Profiilianalüüsi rakendus

Profiilianalüüsi eesmärgiks on leida need peptiidid, mis kõigis neljas populatsioonis käituvad ajas konstantselt ehk mille indeksi väärtus erinevatel ajahetkedel oluliselt ei muutu. Et võrreldavaid populatsioone on 4 ning mõõdetuid tunnuseid 3 (ajahetked 0, 56, 112), siis tuleb profiilianalüüsi rakendamisel kasutada MANOVAt. Enne MANOVA rakendamist kontrollitakse, kas selle meetodi kasutamiseks on eeldused täidetud (vt peatükk 4). Eeldatakse, et erinevad populatsioonid on sõltumatud. Teisena veenduti eelduses, et uuritavad 14 peptiidi on kõikides populatsioonides normaaljaotusega. Selleks vaadeldi esmalt kõikide peptiidide jaoks igas populatsioonis igal ajahetkel tehtud mõõtmiste histogrammi (vt joonis 4).



Joonis 4. Näide ühe peptiidi histogrammidest eri populatsioonides

Histogrammidelt nähtu põhjal võis aimata, et kõik peptiidid ei ole igal ajahetkel normaaljaotusega. Seega oli oodatav, et kõik peptiidid ei ole mitmemõõtmelise normaaljaotusega. Mitmemõõtmelise normaaljaotuse kontrolliks rakendati kõikidele peptiididele mitmemõõtmelist Shapiro-Wilk'i test. Testi tulemuste tõlgendamisel tuli arvesse võtta, et tegemist on mitmese testimisega – sooviti, et peptiid oleks korraga kõigis neljas populatsioonis normaaljaotusega. Mitmesel testimisel otsustati kasutada Benjamini & Hochberg meetodit. Benjamini & Hochberg meetod kontrollib valesi ümber lükatud nullhüpoteeside oodatavat osakaalu (ingl k. *False discovery rate*). Meetodi miinuspooleks on, et esimest liiki vea tõenäosus suureneb. (Wikipedia, 2013).

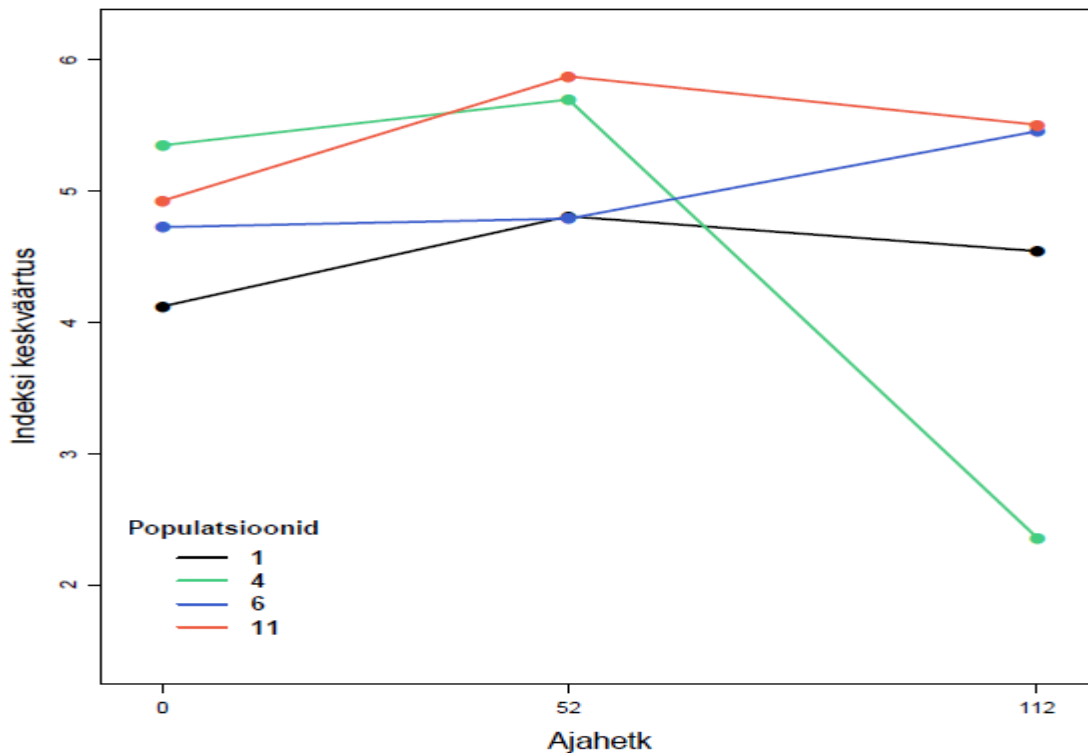
Testimisel lükati mitmemõõtmelise normaaljaotuse sobivus ümber 9 peptiidi korral. Peptiide, millele sobis mitmemõõtmeline normaaljaotus kõigis neljas populatsioonis, oli 5. Kindluse mõttes testiti, kas nende 5 peptiidi korral kõik 3 komponenti (indeksi väärtus igal ajahetkel) on normaaljaotusega. Mitmesel testimisel rakendati ka siin Benjamini & Hochberg meetodit. Tulemused näitasid, et kõigi 5 peptiidi korral saab jääda ka komponenthaaval normaaljaotuse eelduse juurde. (vt lisa 1).

Kolmandana kontrollitakse, kas iga mitmemõõtmelise normaaljaotusega peptiidi kovariatsioonimaatriks võiks kõikides populatsioonides olla sama. Kuigi on teada, et olukorras, kus võrreldavate populatsioonide valimimahud on võrdsed, kovariatsioonimaatriksite homogeensust tavaliselt ei kontrollita (Macedo & Waterson, 2008), siis korrektsuse mõttes tehakse seda töö autori poolt siiski. Kovariatsioonimaatriksite homogeensuse testimiseks kasutatakse Box's  $M$ -statistikut (Timm, 2002), mis põhineb tõepärasuhte testil ning on tuletatud Box'i poolt 1949. aastal. Juhul, kui valimimaht on väike või keskmine, kasutatakse  $F$ -jaotust, suure valimi korral hii-ruut jaotust. Nullhüpoteesiks on, et kovariatsioonimaatriksid on homogeensed. Testimiseks kasutatakse  $R$ -i funktsiooni Box's  $M$ -test (Liaw, 2004). Selle funktsiooni korral loetakse valimit suureks juhul, kui ühes populatsioonis on rohkem kui 19 vaatlust. Tulemuseks oli, et kõigi 5 peptiidi korral võib kovariatsioonimaatriksid lugeda võrdseteks kõigis populatsioonides (vt lisa 2).

Pärast MANOVA tingimuste täidetuse kontrollimist asuti profiilianalüüsi läbi viima. Esiteks sooviti kontrollida kõigi 5 peptiidi kohta hüpoteesi profiilide paralleelsusest populatsiooniti. Selleks testiti iga peptiidi korral hüpoteesi (Maitra, 2013):

$$H_0: \mathbf{C}_{3 \times 4} \boldsymbol{\beta}_{4 \times 3} \mathbf{M}_{3 \times 2} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Tulemuseks saadi kõigi 5 peptiidi korral, et nende profiilid populatsioonides ei ole paralleelsed (vt joonis 5 ja lisa 3).



**Joonis 5. Populatsioonide keskvaartused 3 ajahetkel peptiidil MLRGIASG**

Seega saab kõik 5 peptiidi hilisemasse analüüsi kaasata. Juhul, kui profiilid mõne peptiidi korral oleksid olnud paralleelsed, saanuks edasi kontrollida, kas nad on kõik ka  $x$ -teljega paralleelsed. Kui  $x$ -teljega paralleelne peptiid leidunuks, oleks saanud ta hilisemast analüüsist välja jätta, sest keskmine antikehade kontsentratsioon ajas ei muutunud.

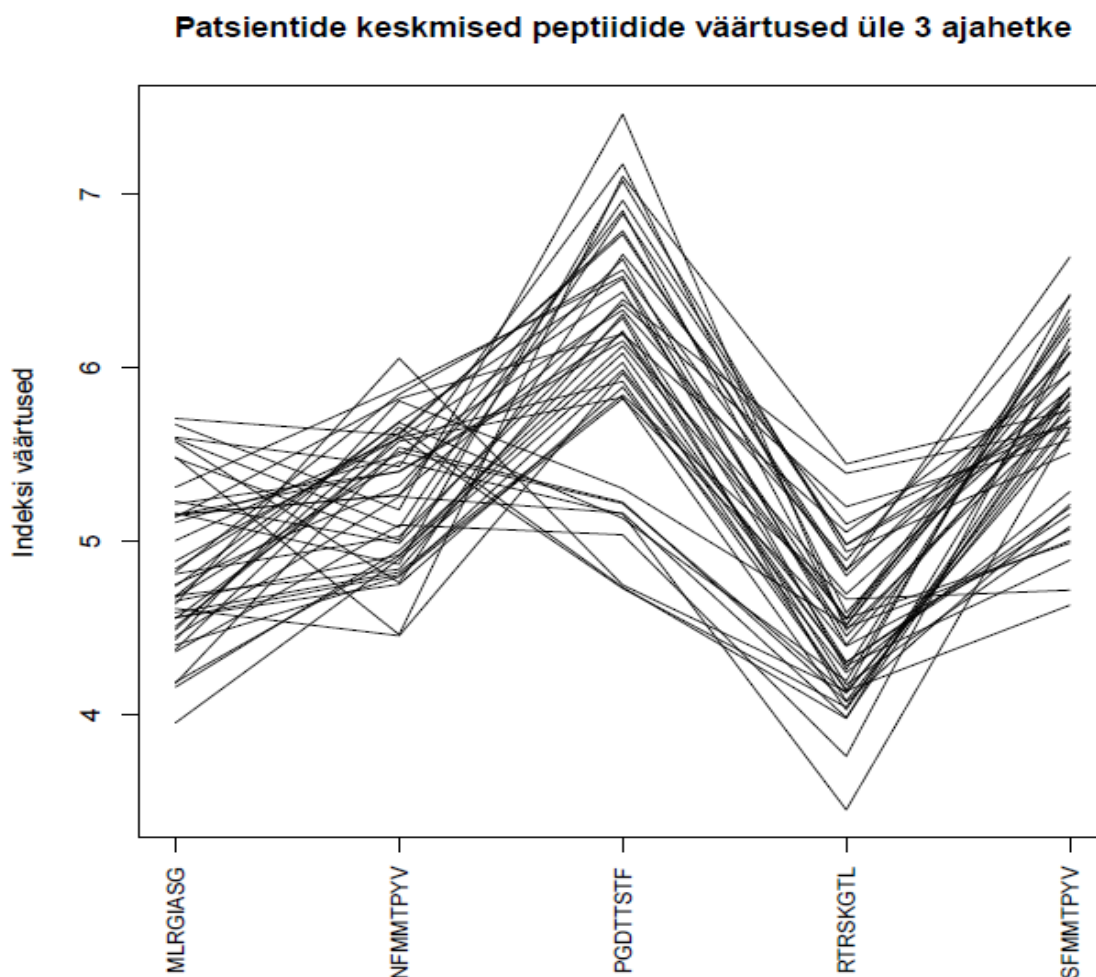
## 6.4 Peptiidide käitumise võrdlemine

Peptiidide käitumise hindamiseks kõiki kasutatakse 44 patsienti ja 5 peptiidi, mille jaotus vastab mitmemõõtmelisele normaaljaotusele. Siinkohal sooviks töö autor rõhutada, et algsete patsientide kohta ei ole teada kirjeldavaid tunnuseid nagu näiteks sugu ning vanus. Kui vaktsiinile reageerimine on mõne sellise tunnuse poolt mõjutatud, siis tuleks need kindlasti arvesse võtta ning väljatoodud tulemused peptiidide käitumise kohta võivad muutuda.

Hinnangute vektori  $\boldsymbol{\beta}$  ja kovariatsioonimaatriksite hindamiseks vajalikud maatriksid on defineeritud järgnevalt: Maatriks  $\mathbf{A}$  on  $44 \times 1$  maatriks, mis sisaldab ainult ühtesid, sest

eeldatakse, et kõik patsiendid on ühest populatsioonist. Matriks  $\mathbf{B}$  on  $(5 * 3) \times 1$  matriks, mis koosneb mudeli vabaliikme ja 2 parameetri hinnangutest. Matriks  $\mathbf{X}$  on täisastakuga  $3 \times 3$  matriks, mille esimene veerg koosneb ühtedest ning mille kaks ülejäänud veergu on valitud nii, et iga veeru elementide summa oleks 0 ning nad oleks omavahel ortogonaalsed. Teine veerg  $\mathbf{X}$ -st väljendab aja lineaarset mõju ja kolmas aja ruutmõju.

Esimese parameetrina hinnatakse  $\mathbf{B}$ . Selle arvutamiseks on vajalik kõikide indiviidide keskväärtusvektorite leidmine üle kõikide ajahetkede (valem 5.1.11). Indiviidide keskväärtusvektoreid iseloomustab järgmine joonis.



**Joonis 6. Patsientide keskväärtusvektorid, iseloomustamaks peptiidide keskväärtusi üle 3 ajahetke**

Jooniselt 6 nähakse, et mõningatel indiviididel on peptiidile PGDTTSTF vastavate indeksite väärtused teistest madalamal. Lisainformatsiooni saamisel võiks uurida, mille poolest need indiviidid teistest erinevad.

Kasutades valemit 5.1.9, leitakse  $\mathbf{B}$  hinnang  $\hat{\mathbf{B}} = \begin{pmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\Gamma}} \end{pmatrix}$ , mille elemendid on toodud tabelis 1.

**Tabel 1. Hinnatud parameetrid maatriksis  $\hat{\mathbf{B}}$**

Parameeter	Hinnang	Peptiid
$\mu_1$	4,8475	MLRGIASG
$\mu_2$	5,2724	NFMMPYV
$\mu_3$	6,1089	PGDTTSTF
$\mu_4$	4,4773	RTRSKGTL
$\mu_5$	5,7136	SFMMPYV
$\Gamma_{11}$	-0,1571	MLRGIASG
$\Gamma_{12}$	-0,4622	NFMMPYV
$\Gamma_{13}$	-0,088	PGDTTSTF
$\Gamma_{14}$	-0,0487	RTRSKGTL
$\Gamma_{15}$	-0,2932	SFMMPYV
$\Gamma_{21}$	0,2229	MLRGIASG
$\Gamma_{22}$	0,4482	NFMMPYV
$\Gamma_{23}$	0,0146	PGDTTSTF
$\Gamma_{24}$	0,1889	RTRSKGTL
$\Gamma_{25}$	0,2781	SFMMPYV

Järgmisena hinnatakse jääkide kovaratsioonimaatriks  $\Sigma_e$ , kasutades valemit 5.1.14. Hinnang on toodud tabelis 2.

**Tabel 2. Jääkide kovaratsioonimaatriksi hinnang**

	MLRGIASG	NFMMPYV	PGDTTSTF	RTRSKGTL	SFMMPYV
MLRGIASG	0,6631	-0,2724	1,0362	0,6421	0,2302
NFMMPYV	-0,2724	0,6093	-0,5809	-0,2995	0,1741
PGDTTSTF	1,0362	-0,5809	2,4272	1,2477	0,1917
RTRSKGTL	0,6421	-0,2995	1,2477	0,9160	0,2037
SFMMPYV	0,2302	0,1741	0,1917	0,2037	0,4175

Kasutades valemit 5.1.15, hinnati kovaratsioonimaatriks  $\Sigma_\lambda$ , mille elemendid on esitatud tabelis 3.

**Tabel 3. Kovaratsioonimaatriksi  $\Sigma_\lambda$  hinnang**

	MLRGIASG	NFMMPYV	PGDTTSTF	RTRSKGTL	SFMMPYV
MLRGIASG	0,0000	0,0952	-0,1189	-0,0843	0,0035
NFMMPYV	0,0952	0,0000	0,1111	0,0649	0,0754
PGDTTSTF	-0,1189	0,1111	0,0000	-0,2107	-0,0587
RTRSKGTL	-0,0843	0,0649	-0,2107	0,0000	-0,0507
SFMMPYV	0,0035	0,0754	-0,0587	-0,0507	0,0928

Algselt olid  $\Sigma_\lambda$  diagonaalil 0-de asemel negatiivsed hinnangud. Et dispersioonid negatiivsed olla ei saa, asendati need nullidega. Kuigi teoorias on dispersioonid alati mittenegatiivsed, siis nende hinnangud valimilt võivad tulla negatiivsed (Ravishanker & Dey, 2002).

Enne analüüsi tegemist on tarvis veenduda, et valimi pealt hinnatud kovariatsioonimaatriks on liitsümmeetria struktuuriga kovariatsioonimaatriks. Kõigepealt arvutatakse  $\mathbf{S}$ :

$$\mathbf{S} = \frac{1}{N - r} \mathbf{Y}^T (\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T) \mathbf{Y}$$

ning siis  $\mathbf{\Omega} = (\mathbf{I}_p \otimes \mathbf{\Sigma}_e) + (\mathbf{1}\mathbf{1}^T \otimes \mathbf{\Sigma}_\lambda)$ . Seejärel kasutatakse teststatistikut (Timm, 2002)

$$K = (N - r)(\ln(|\mathbf{\Sigma}_0|) - \ln(|\mathbf{S}|) + \text{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1}) - pm),$$

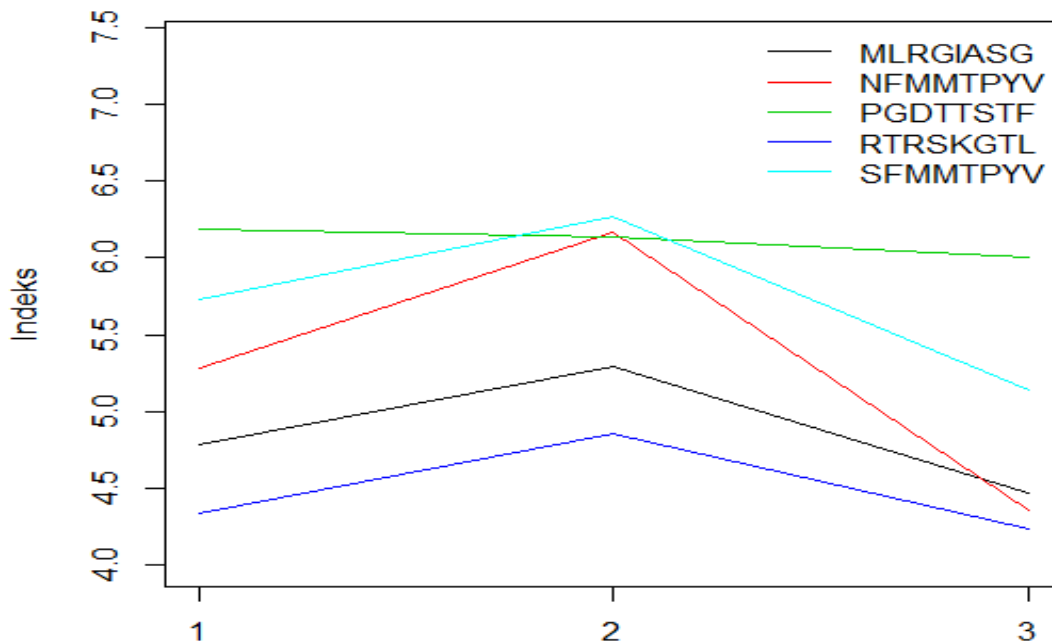
et kontrollida hüpoteesi  $H_0: \mathbf{\Sigma} = \mathbf{\Sigma}_0$ , kus  $\mathbf{\Sigma}_0$  on teadaolev maatriks, hetkel siis  $\mathbf{\Omega}$ . Korrutades statistikut  $K$  suurusega  $(1 - LC)$ , kus

$$LC = (2 * (mp)^2 + 3mp - 1) / (6 * (N - r) * (mp + 1)),$$

saadakse hii-ruut jaotusega statistik. Tõestatuks loetakse  $H_1$  juhul, kui  $K(1 - LC)$  väärtus ületab kriitilist väärtust  $\bar{\chi}^2\left(\alpha, pm\left(\frac{pm+1}{2}\right)\right)$ , mis on hii-ruut jaotuse  $\chi^2\left(pm\left(\frac{pm+1}{2}\right)\right)$   $\alpha$  –täiendkvantiil. Kontrollides hüpoteesi  $\mathbf{\Omega}$  ja  $\mathbf{S}$  võrdsusest, jõuti järeldusele, et need ei ole formaalselt võrdsed (teststatistiku väärtus 715 > kriitiline väärtus 147). Kuid põhjusel, et see test on üsna tundlik (Kollo, suuline konsultatsioon, 2013), jätkatakse analüüsiga. Siiski usub autor, et järgnevalt esitatud tulemustesse tuleks suhtuda ettevaatlikkusega.

Kõigepealt kontrolliti, kas kõikide peptiidide kasvukõverate kujud võiksid olla samad. Seda tehakse, arvutades Wilks'i lambda. Kasutatava testi puhul võetakse vastu sisukas hüpotees, kui teststatistik Wilks'i lambda on väiksem kui kriitiline väärtus. Kriitiline väärtus on pärit tabelist (Rencher, 2002). Tulemuseks saadi, et kõikide peptiidide kasvukõverate kujud ei ole võrdsed (teststatistik 0.729, kriitiline 0.83). Jooniselt 7 vaadati, et peptiidide MLRGIASG,

RTRSKGTL ja SFMMTPYV kasvukõverate kujud võiksid olla samad. Testimise tulemusena jäädakse nullhüpoteesi juurde (teststatistik 0.911, kriitiline väärtus 0.897), seega võib öelda, et need peptiidid käituvad sarnaselt.



**Joonis 7. Peptiidide profiilid**

Andmestiku edasiseks analüüsimiseks võiks katsetada üldisemat kovariatsioonimaatriksi struktuuri eeldust, kus lubatakse peale juhusliku vabaliikme ka juhuslikkust aja mõju liikmetel. Sellise struktuuriga andmete analüüsimisest on kirjutanud Reinsel (1982). Samuti võiks kaaluda ka nende peptiidide analüüsi kaasamist, millel on mõõdetud 3 esimest ajahetke, kuid 4 ja 5 ajahetke mõõtmistulemus võib olla puuduv.

# “Multivariate analysis of peptides’ behavior”

Master thesis

Kristi Läll

## Summary

While developing the vaccine against tuberculosis, it is important to investigate which antigens react with the antibodies found in the vaccine. Also, less significant isn't the information about how fast and in what concentration does the reaction take place. During vaccine trials, in a common situation several characteristics are measured on each individual at several points in time. Therefore, the size of the data can be quite large and classical statistical methods may become unfit to develop a suitable analysis.

Two methods are applied in this thesis. First, profile analysis is performed using MANOVA to explore the behavior of peptides' over time and to find out if any of them can be excluded from the following analysis. Secondly, Reinsel (1982) has introduced a method for analysing multiple-response repeated measurements under a simplifying assumption of a multivariate random-effects model. He gives the estimates and distributions for parameters in the model and for the covariance matrices. Also, he proposes a test for hypothesis checking about the parameters. This test is used to compare peptides' growth curves with each other.

As a result, five peptides could be used after profile analysis and 3 similarly behaving peptides (MLRGIASG, RTRSKGTL, SFMMTPYV) were found while comparing both linear and quadratic trend in means at the same time. It can be, that there are some important covariates (e. g. age and gender), which were not included in the current analysis and which can lead to different conclusions while added. Also, it might be that due to the huge amount of missing data in the original dataset, some similarly acting peptides were just not incorporated in the analysis at the moment.

## Lisa 1. Ühemõõtmelise Shapiro-Wilk'i testi tulemused

Peptiid	Ajahetk 0	Ajahetk 56	Ajahetk 112	Populatsioon
MLRGIASG	0,88	0,88	1	1
NFMMPYV	1	1	1	1
PGDTTSTF	1	1	1	1
RTRSKGTL	1	1	0,56	1
SFMMPYV	0,48	0,44	0,44	1
MLRGIASG	1	0,59	0,59	4
NFMMPYV	1	1	1	4
PGDTTSTF	1	0,91	0,91	4
RTRSKGTL	0,7	0,7	0,7	4
SFMMPYV	1	1	1	4
MLRGIASG	1	1	1	6
NFMMPYV	1	1	1	6
PGDTTSTF	0,25	0,82	0,99	6
RTRSKGTL	1	1	1	6
SFMMPYV	0,31	1	1	6
MLRGIASG	0,59	1	0,18	11
NFMMPYV	1	1	1	11
PGDTTSTF	0,73	1	0,27	11
RTRSKGTL	1	1	1	11
SFMMPYV	0,42	1	1	11

Tabel 4. Ühemõõtmelise Shapiro-Wilk'i testi korrigeeritud p-väärtused igas populatsioonis. Sinisega on tähistatud populatsioon number 1, punasega populatsioon number 4, rohelisega populatsioon number 6 ja oranžiga populatsioon number 11.

**Lisa 2. Box's M testi väljund R-is iga peptiidi korral, olulisusenivool 0.05:**

Mbox	F	df1	df2	P
21.2763	1.0182	18	5653	0.4345
Covariance matrices are not significantly different.				

**Tabel 1. Peptiid MLRGIASG testimise tulemus**

Mbox	F	df1	df2	P
22.5113	1.0773	18	5653	0.3685
Covariance matrices are not significantly different.				

**Tabel 2. Peptiid NFMMPYV testimise tulemus**

Mbox	F	df1	df2	P
18.8142	0.9004	18	5653	0.5782
Covariance matrices are not significantly different.				

**Tabel 3. Peptiid PGDTTSTF testimise tulemus**

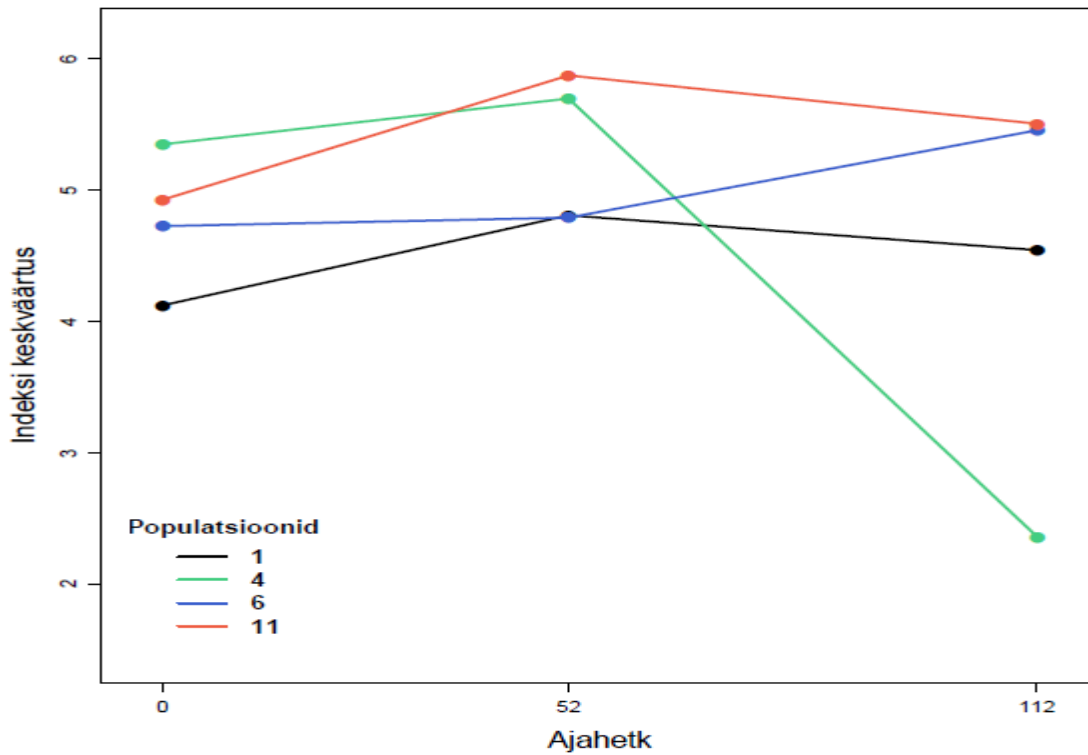
Mbox	F	df1	df2	P
23.0234	1.1018	18	5653	0.3427
Covariance matrices are not significantly different.				

**Tabel 4. Peptiid RTRSKGTL testimise tulemus**

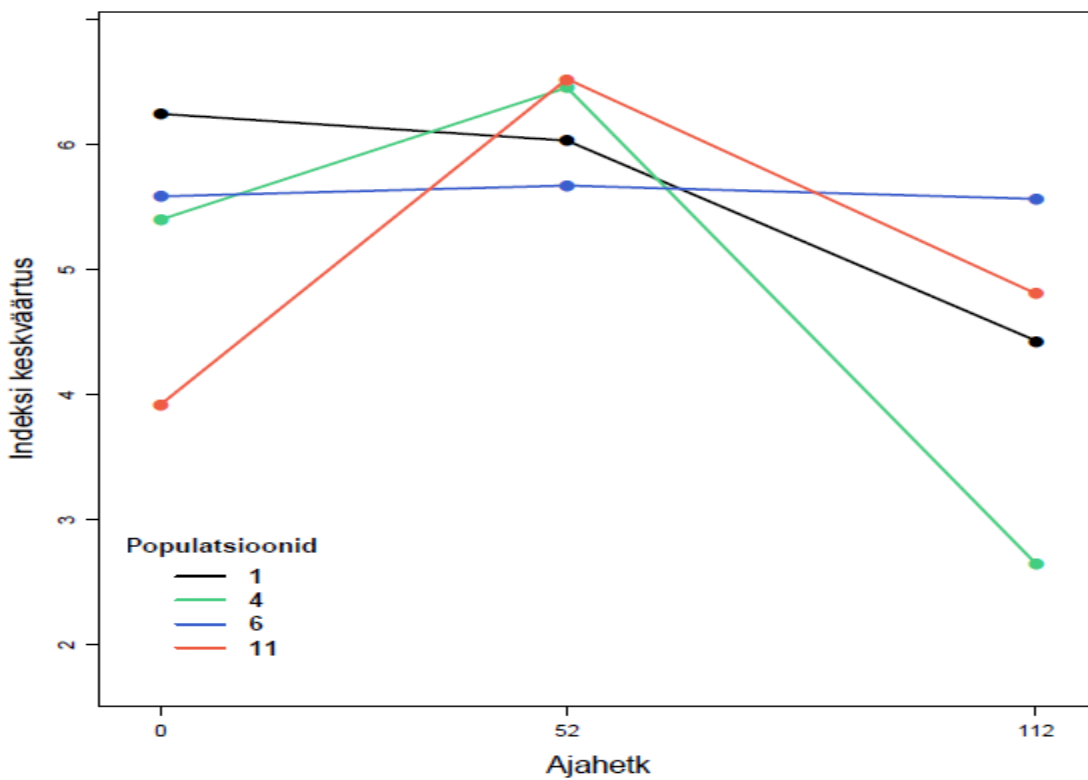
Mbox	F	df1	df2	P
23.0322	1.1022	18	5653	0.3423
Covariance matrices are not significantly different.				

**Tabel 5. Peptiid SFMMPYV testimise tulemus**

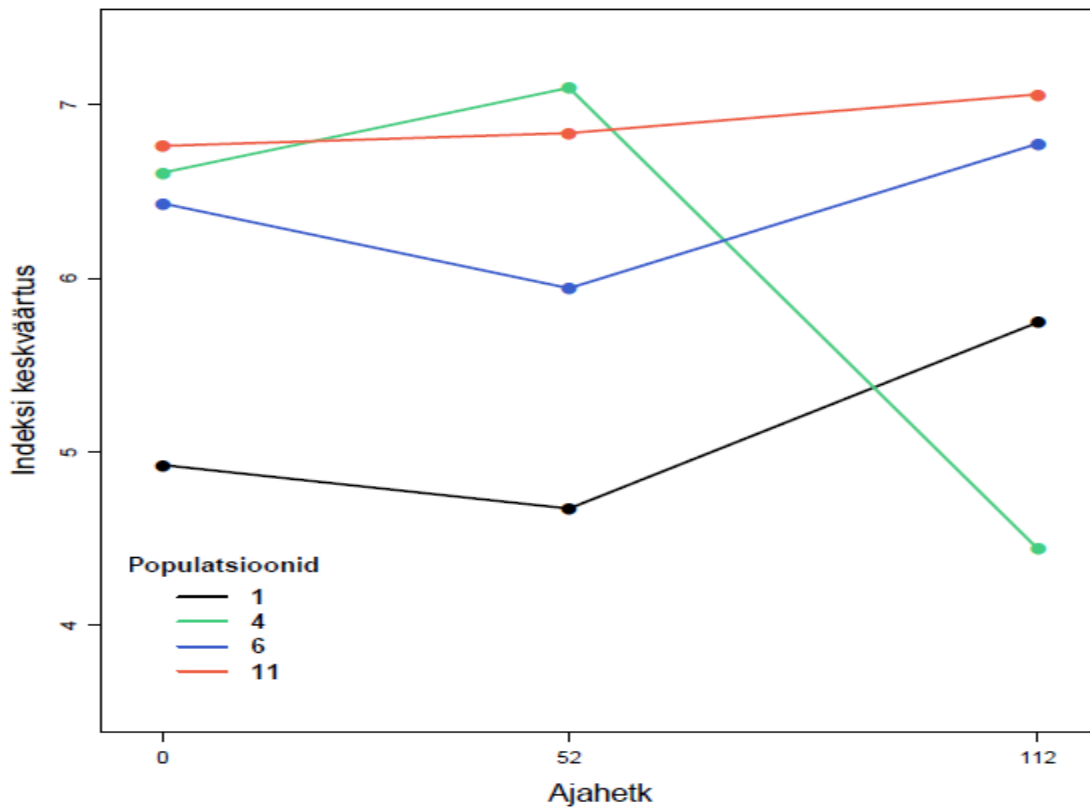
### Lisa 3. Mitmemõõtmelise normaaljaotusega peptiidide profiilide joonised



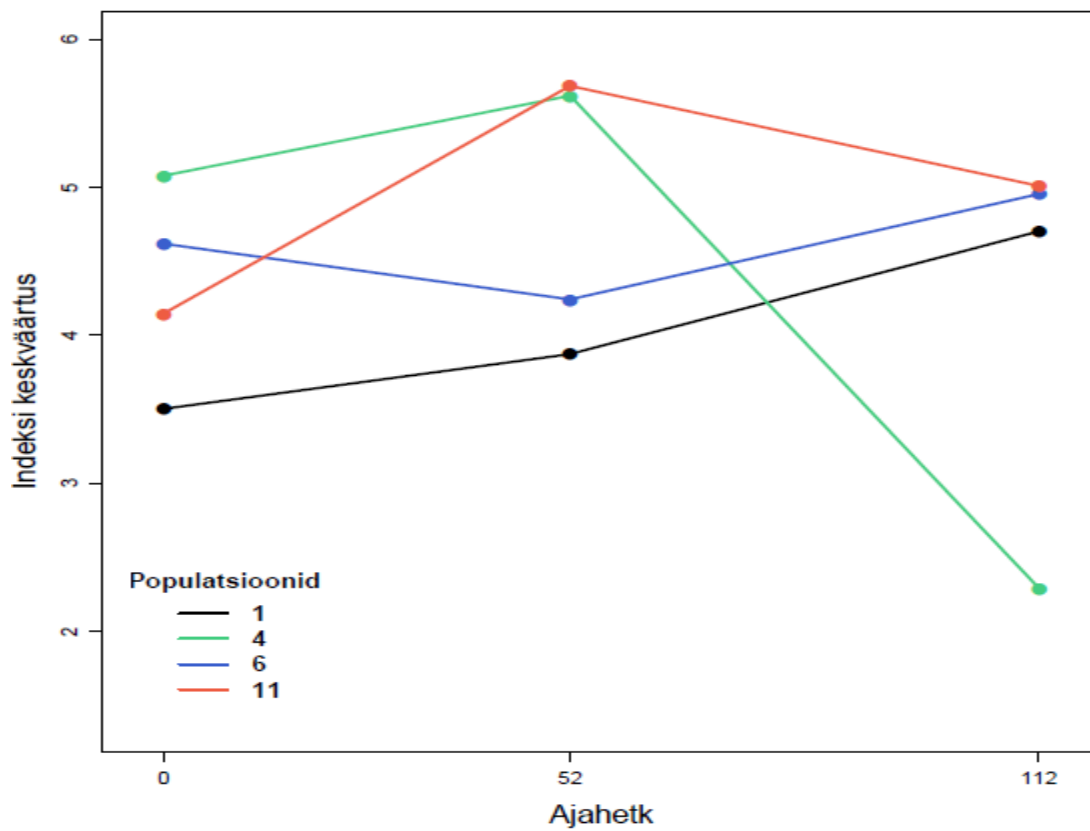
Joonis 8. Peptiidi MLRGIASG keskvaartused 3 ajahetkel populatsioonide kaupa



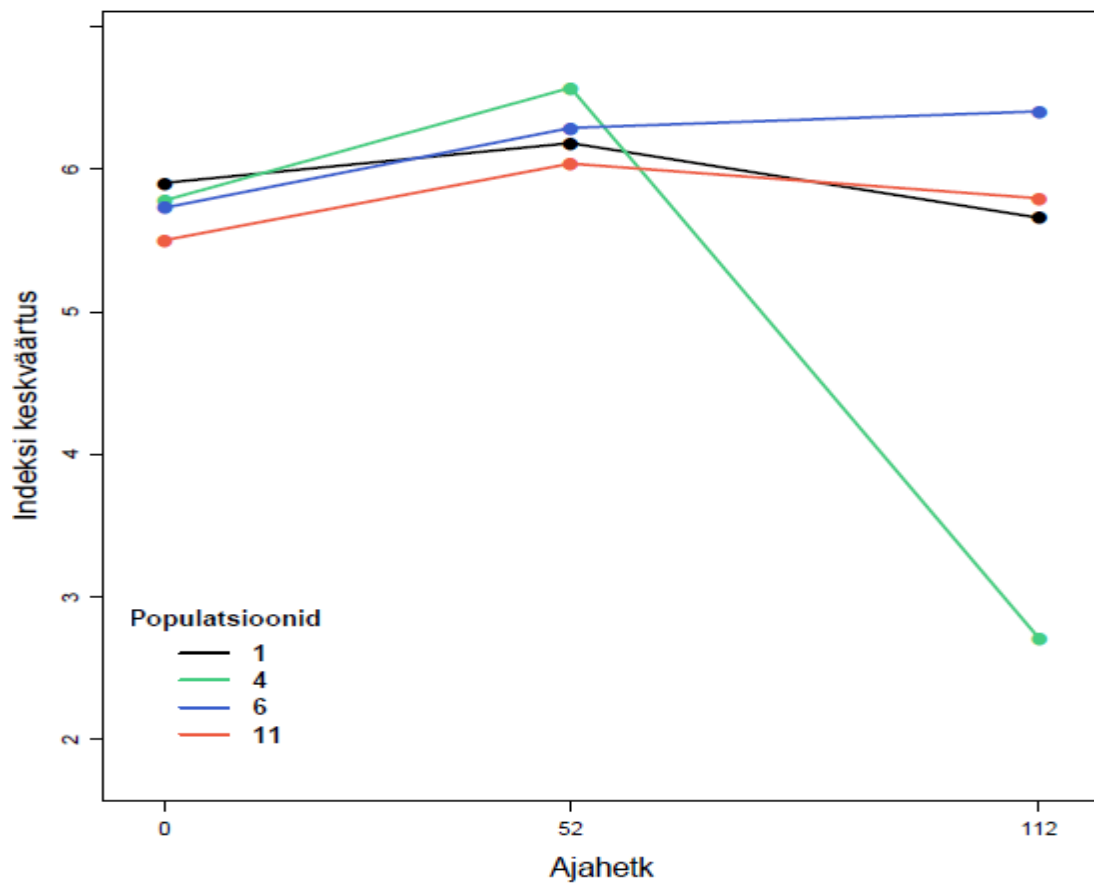
Joonis 9. Peptiidi NFMTPYV keskvaartused 3 ajahetkel populatsioonide kaupa



Joonis 10. Peptiidi PGDTTSTF keskväärtused 3 ajahetkel populatsioonide kaupa

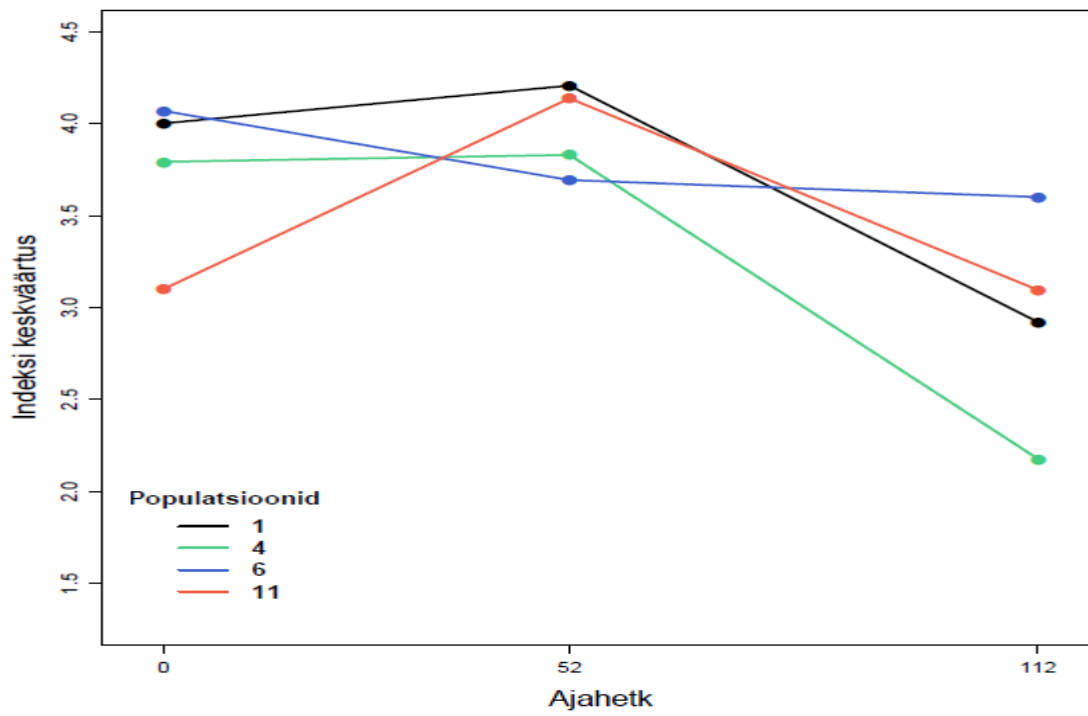


Joonis 11. Peptiidi RTRSKGTL keskväärtused 3 ajahetkel populatsioonide kaupa

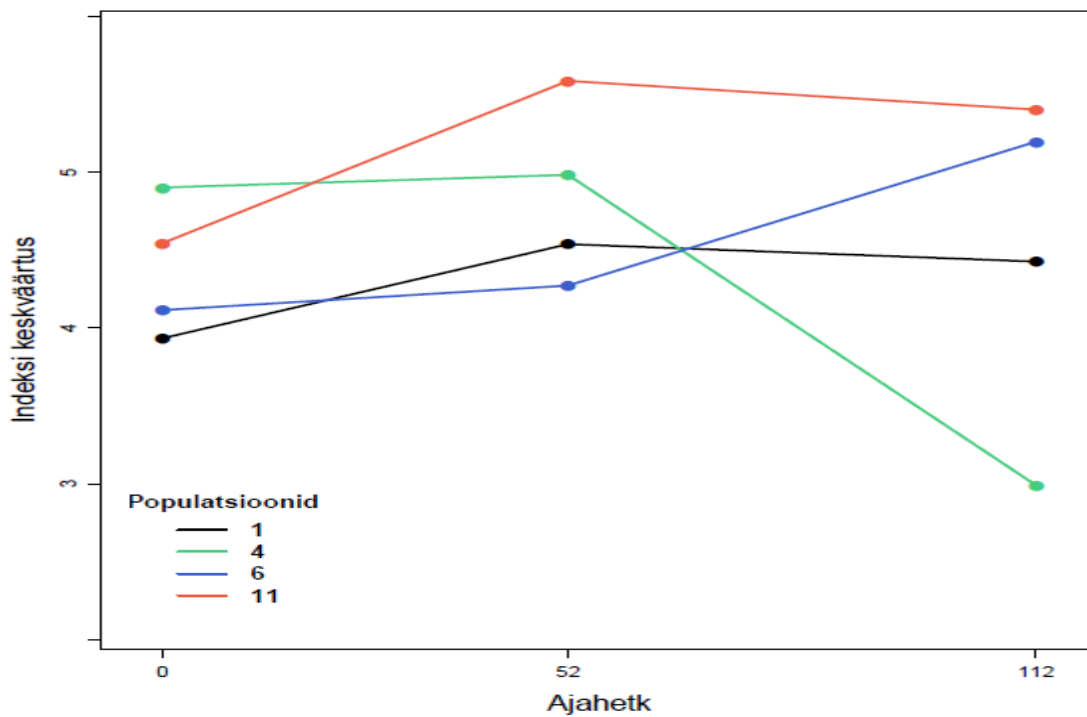


Joonis 12. Peptiidi SFMMTPYV keskväärtsused 3 ajahetkel populatsioonide kaupa

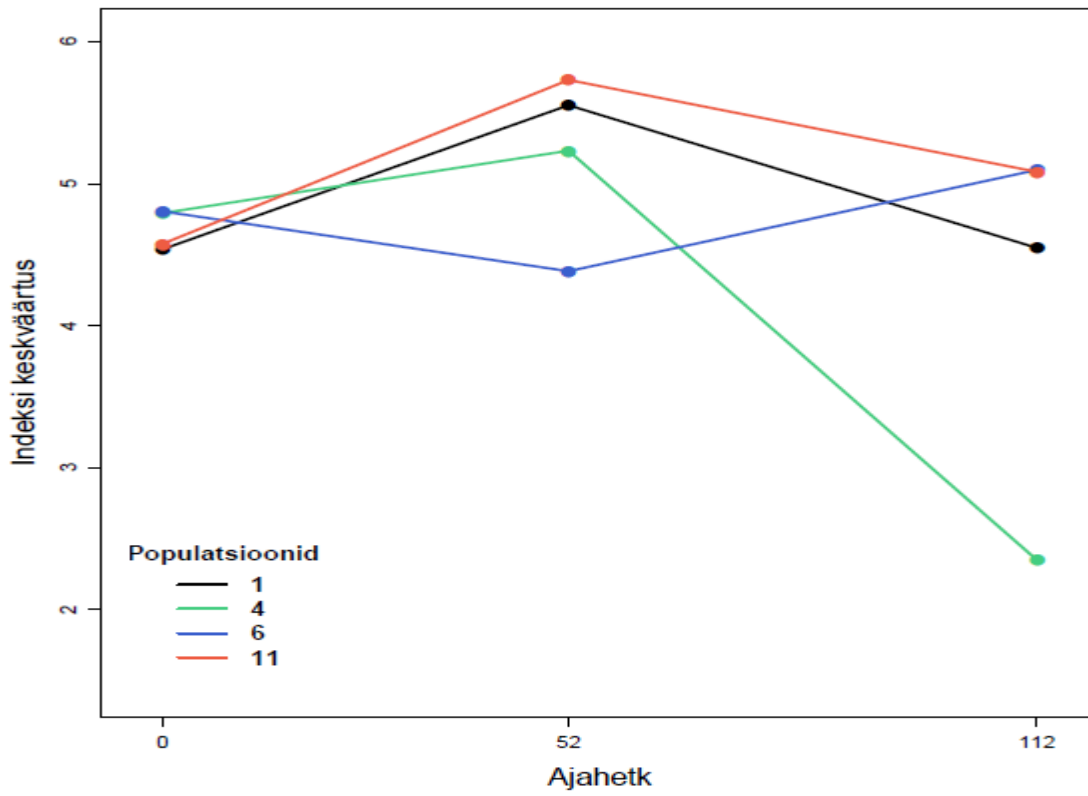
### Lisa 4. Peptiidid, mis ei ole mitmemõõtmelise normaaljaotusega



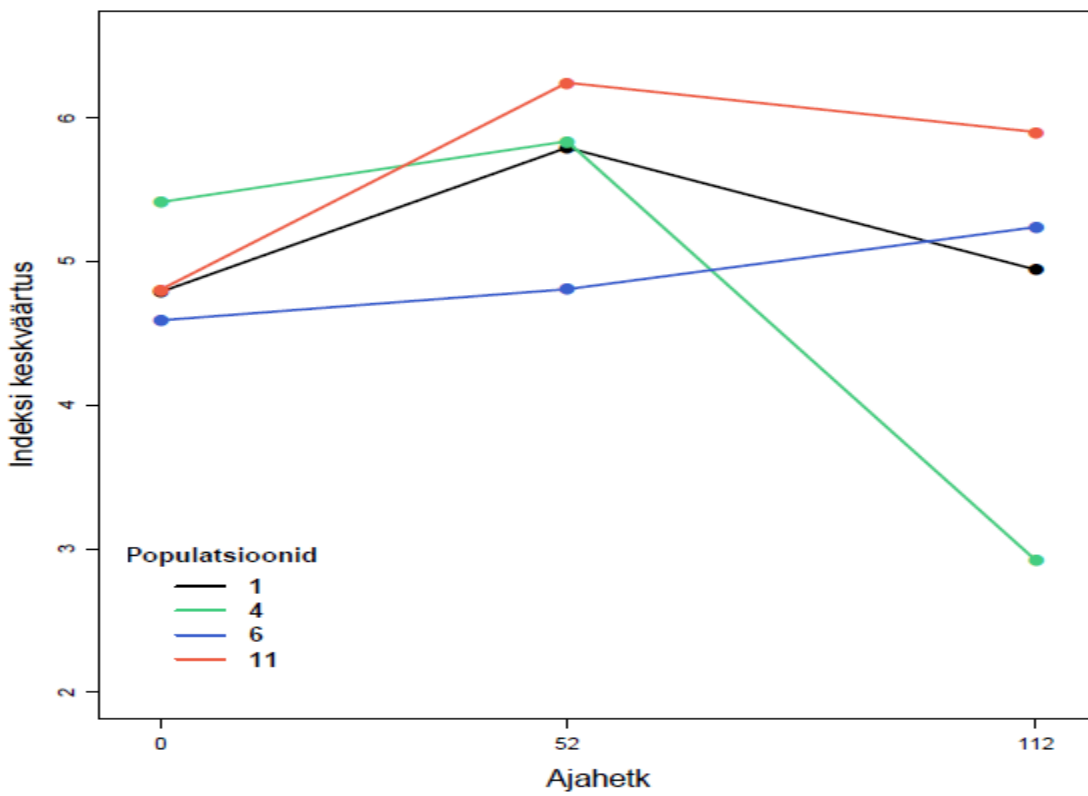
Joonis 13. Peptiidi ASFIMSYW keskväärtsed 3 ajahetkel populatsioonide kaupa



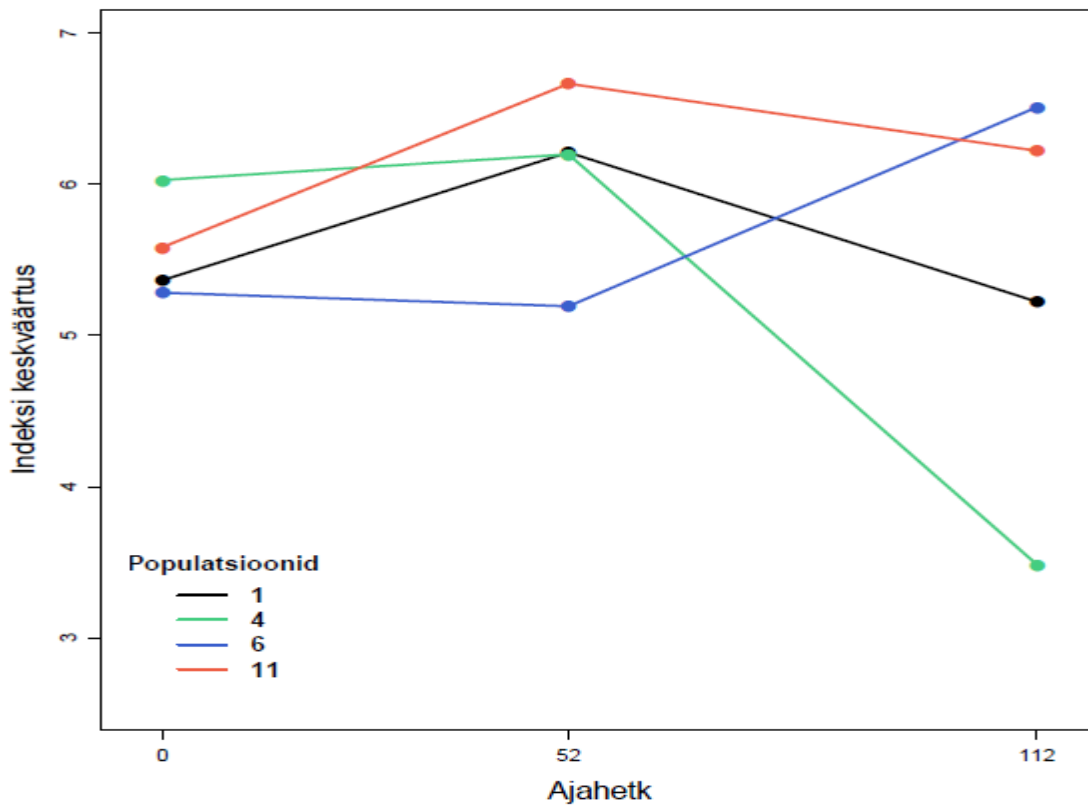
Joonis 14. Peptiidi DSVYKLML keskväärtsed 3 ajahetkel populatsioonide kaupa



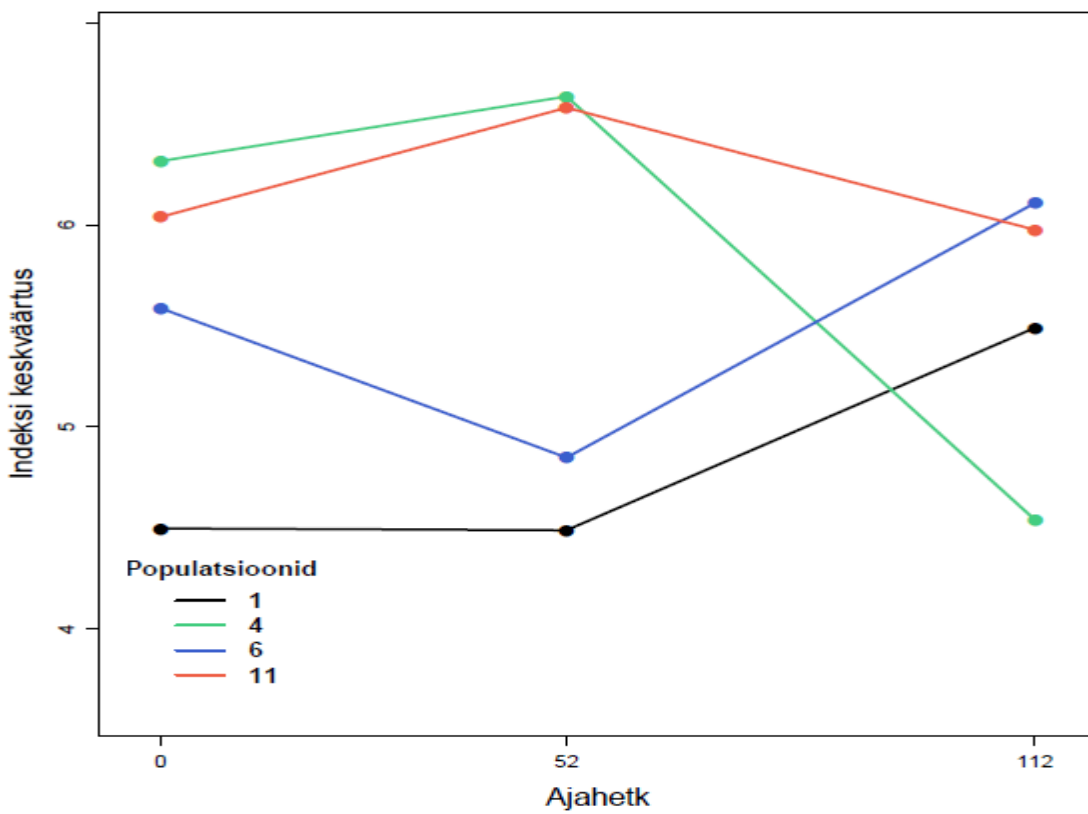
Joonis 15. Peptiidi GLKQKTSF keskvaartused 3 ajahetkel populatsioonide kaupa



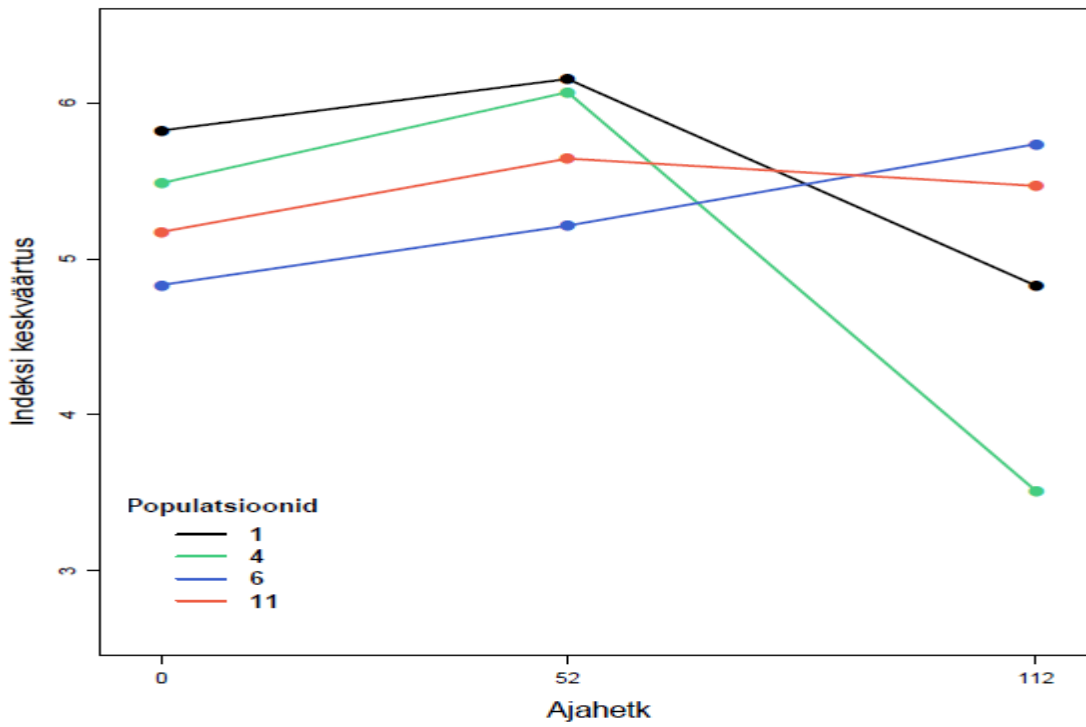
Joonis 16. Peptiidi IRWMPES keskvaartused 3 ajahetkel populatsioonide kaupa



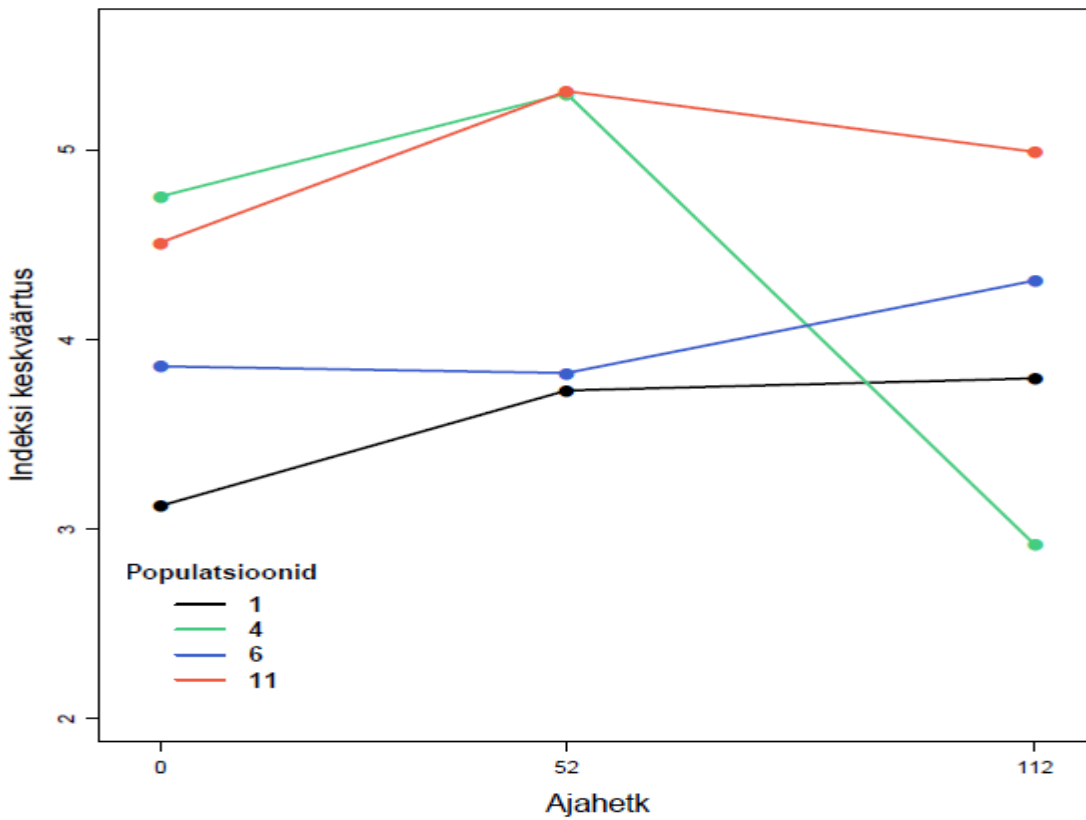
Joonis 17. Peptiidi MNPFIYSL keskvaartused 3 ajahetkel populatsioonide kaupa



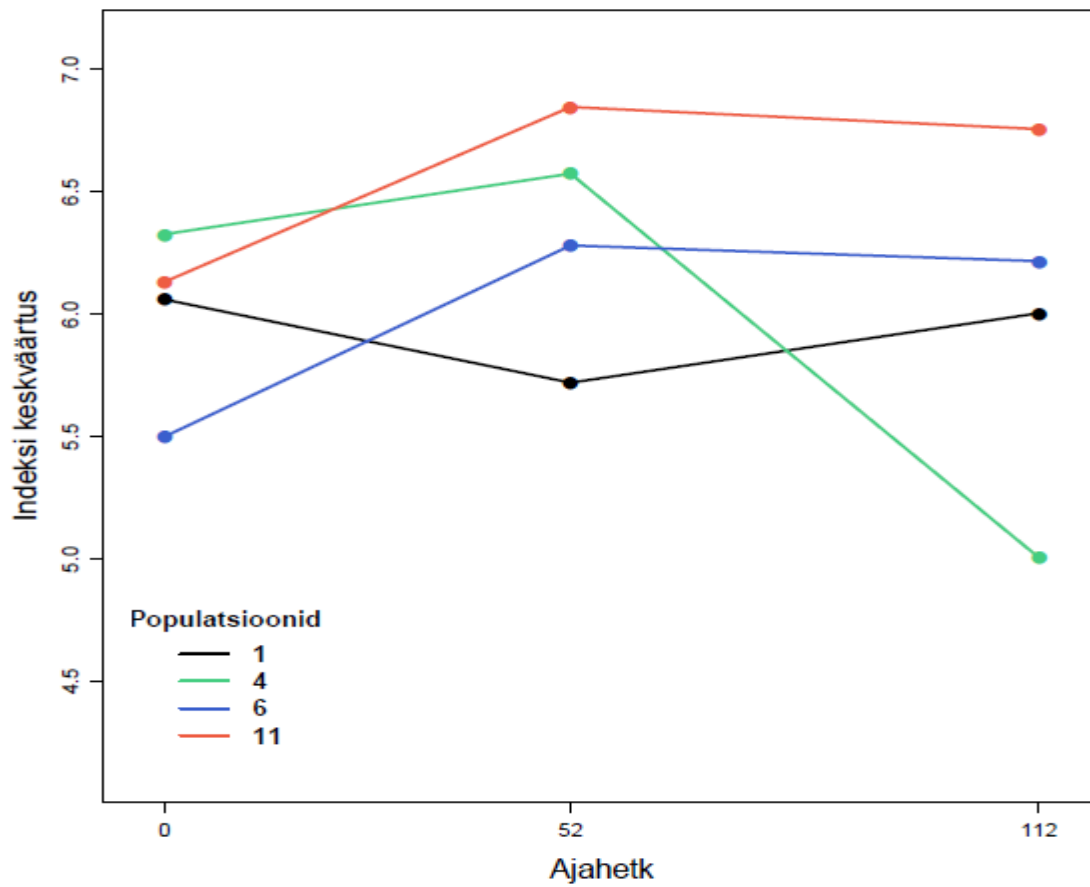
Joonis 18. Peptiidi MYRRSYVF keskvaartused 3 ajahetkel populatsioonide kaupa



Joonis 19. Peptiidi QNSRYQTY keskväärtaus 3 ajahetkel populatsioonide kaupa



Joonis 20. Peptiidi VAMHVAYR keskväärtaus 3 ajahetkel populatsioonide kaupa



Joonis 21. Peptiidi WKYLYRSP keskväärtaus 3 ajahetkel populatsioonide kaupa

## Lisa 5. Peatükk 5 toodud hinnangute tuletuskäigud

1.  $k$ -ndale objekti vaatlusvektori kovaratsiooni avaldamine

$$\begin{aligned} cov(\mathbf{y}_k) &= cov((\mathbf{X} \otimes \mathbf{I}_m)\boldsymbol{\phi}_k + (\mathbf{1} \otimes \mathbf{I}_m)\boldsymbol{\lambda}_k + \mathbf{e}_k) = (\mathbf{1} \otimes \mathbf{I}_m)cov(\boldsymbol{\lambda}_k)(\mathbf{1} \otimes \mathbf{I}_m)^T \\ &+ cov(\mathbf{e}_k) = (\mathbf{1} \otimes \mathbf{I}_m)\boldsymbol{\Sigma}_\lambda(\mathbf{1} \otimes \mathbf{I}_m)^T + (\mathbf{I}_p \otimes \boldsymbol{\Sigma}_e) = (\mathbf{I}_p \otimes \boldsymbol{\Sigma}_e) + (\mathbf{1}\mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda). \end{aligned}$$

Kui vaadata  $cov(\mathbf{y})$ , siis  $cov(\mathbf{y}) = (\mathbf{I}_N \otimes \mathbf{I}_p \otimes \boldsymbol{\Sigma}_e) + (\mathbf{I}_N \otimes \mathbf{1}\mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda)$ .

2. Maatriksi  $\mathbf{B}$  hinnangu  $\hat{\mathbf{B}} = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{Y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}$  tuletamine

Mudeli kuju on antud  $\mathbf{Y}^T = (\mathbf{X} \otimes \mathbf{I}_m)\mathbf{B}\mathbf{A}^T + \mathbf{U}^T$ . Soovitakse kirja panna  $\hat{\mathbf{B}}$ , kasutades vähimruutude hinnangut. Seega viiakse kõigepealt teisendus  $\mathbf{Y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} = (\mathbf{X} \otimes \mathbf{I}_m)\mathbf{B}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} = (\mathbf{X} \otimes \mathbf{I}_m)\mathbf{B}$ . Siis saab  $\mathbf{B}$  hinnangu kirja panna  $\hat{\mathbf{B}} = ((\mathbf{X} \otimes \mathbf{I}_m)^T(\mathbf{X} \otimes \mathbf{I}_m))^{-1}(\mathbf{X} \otimes \mathbf{I}_m)^T\mathbf{Y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{Y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}$ .

3. Parameetervektori  $\boldsymbol{\beta}$  hinnangu  $\hat{\boldsymbol{\beta}} = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{y}$  tuletamine

Et mudeli kuju on  $\mathbf{y}$  jaoks antud kui  $\mathbf{y} = (\mathbf{A} \otimes \mathbf{X} \otimes \mathbf{I}_m)\boldsymbol{\beta} + \mathbf{u}$  ning üldistatud vähimruutude hinnang on hetkel sama, mis vähimruutude hinnang, siis

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= vec(\hat{\mathbf{B}}) = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)vec(\mathbf{Y}^T) \\ &= ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{y}. \end{aligned}$$

4. Parameeter  $\hat{\boldsymbol{\beta}}$  keskvaertuse leidmine

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E(((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{y}) = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)E\mathbf{y} \\ &= ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)(\mathbf{A} \otimes \mathbf{X} \otimes \mathbf{I}_m)\boldsymbol{\beta} = (\mathbf{I}_r \otimes \mathbf{I}_q \otimes \mathbf{I}_m)\boldsymbol{\beta} = \mathbf{I}_{rqm}\boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned}$$

5. Parameeter  $\hat{\boldsymbol{\beta}}$  kovariatsioonimaatriksi tuletamine

$$\begin{aligned} cov(\hat{\boldsymbol{\beta}}) &= ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)cov(\mathbf{y})((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)^T \\ &= ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)\left((\mathbf{I}_N \otimes \mathbf{I}_p \otimes \boldsymbol{\Sigma}_e) + (\mathbf{I}_N \otimes \mathbf{1}\mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda)\right) \\ &((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)^T = ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \boldsymbol{\Sigma}_e \\ &+ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \mathbf{1}\mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda)((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \otimes \mathbf{I}_m)^T \\ &= ((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_e) \end{aligned}$$

$$\begin{aligned}
& +((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{1} \mathbf{1}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_\lambda) \\
& =((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{X}^T \mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_e) + ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{v}_1 \mathbf{p} \mathbf{p}^T \mathbf{v}_1^T (\mathbf{X}^T \mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_\lambda) \\
& =((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{X}^T \mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_e) + ((\mathbf{A}^T \mathbf{A})^{-1} \otimes \mathbf{v}_1 \mathbf{v}_1^T \otimes \boldsymbol{\Sigma}_\lambda),
\end{aligned}$$

sest  $(\mathbf{X}^T \mathbf{X})^{-1}$  on avaldatav kui  $\begin{pmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{1} & \mathbf{Z}^T \mathbf{Z} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{p} & \mathbf{0}^T \\ \mathbf{0} & (\mathbf{Z}^T \mathbf{Z})^{-1} \end{pmatrix}$ , kus  $\mathbf{0}$  tähistab  $(q-1) \times 1$  vektorit. Nüüd  $\mathbf{v}_1 \mathbf{v}_1^T$  on  $q \times q$  maatriks, mille esimese rea esimene element on 1, ja kõik ülejäänud elemendid nullid. Seega

$$p \begin{pmatrix} \frac{1}{p} & \mathbf{0}^T \\ \mathbf{0} & (\mathbf{Z}^T \mathbf{Z})^{-1} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} \mathbf{0}^T \end{pmatrix} \begin{pmatrix} \frac{1}{p} & \mathbf{0}^T \\ \mathbf{0} & (\mathbf{Z}^T \mathbf{Z})^{-1} \end{pmatrix} p = \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} \mathbf{0}^T \end{pmatrix} = \mathbf{v}_1 \mathbf{v}_1^T.$$

## 6. Hinnangu $\hat{\boldsymbol{\Gamma}}$ , $E(\hat{\boldsymbol{\gamma}})$ ja $cov(\hat{\boldsymbol{\gamma}})$ tuletamine

Hinnangu  $\hat{\boldsymbol{\Gamma}}$  saab  $\hat{\mathbf{B}}$  hinnangust, sest  $\boldsymbol{\Gamma}$  on  $m(q-1) \times r$  viimast parameetrit  $\mathbf{B}$ -st.  $\hat{\boldsymbol{\gamma}} = \text{vec}(\hat{\boldsymbol{\Gamma}})$  keskvärtus  $\boldsymbol{\gamma}$  tuleneb  $\hat{\boldsymbol{\beta}}$  keskvärtuse valemi abil.

$$E((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \otimes \mathbf{I}_m) \boldsymbol{\gamma} = ((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X} \otimes \mathbf{I}_m) \boldsymbol{\beta}.$$

Pannakse tähele, et  $\mathbf{Z}^T \mathbf{X}$  saab kirjutada kujul  $(\mathbf{Z}^T \mathbf{1} \quad \mathbf{Z}^T \mathbf{Z}) = (\mathbf{0} \quad \mathbf{Z}^T \mathbf{Z})$ , kus  $\mathbf{0}$  tähistab  $(q-1) \times 1$  vektorit. Seega  $(\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{0} \quad \mathbf{Z}^T \mathbf{Z}) = (\mathbf{0} \quad \mathbf{I}_{q-1})$  ja

$$E(\hat{\boldsymbol{\gamma}}) = (\mathbf{I}_r \otimes (\mathbf{0} \quad \mathbf{I}_{q-1}) \otimes \mathbf{I}_m) \boldsymbol{\beta} = \boldsymbol{\gamma}.$$

Kovariatsioonimaatriksi  $cov(\hat{\boldsymbol{\gamma}})$  leidmine on analoogne  $cov(\hat{\boldsymbol{\beta}})$  tuletuskäiguga:

$$cov(\hat{\boldsymbol{\gamma}}) = ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \boldsymbol{\Sigma}_e) + ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{1} \mathbf{1}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \boldsymbol{\Sigma}_\lambda).$$

Kuna  $\mathbf{Z}^T \mathbf{1} = \mathbf{0}$  ja  $\mathbf{1}^T \mathbf{Z} = \mathbf{0}^T$ , kus  $\mathbf{0}$  on  $(q-1) \times 1$  vektor, siis avaldis

$$((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{1} \mathbf{1}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \boldsymbol{\Sigma}_\lambda) = \mathbf{0} \text{ ning seetõttu avaldub } cov(\hat{\boldsymbol{\gamma}}) \text{ kui}$$

$$cov(\hat{\boldsymbol{\gamma}}) = (\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1} \otimes \boldsymbol{\Sigma}_e.$$

## 7. Hinnangu $\hat{\boldsymbol{\mu}}$ ja $E(\text{vec}(\hat{\boldsymbol{\mu}}))$ ja $cov(\text{vec}(\hat{\boldsymbol{\mu}}))$ tuletamine

Hinnang  $\hat{\boldsymbol{\mu}}$  on esimesed  $m \times r$  parameetrit  $\hat{\mathbf{B}}$ -st, seega avaldub  $\hat{\boldsymbol{\mu}}$  kui  $((\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} = \frac{1}{p} (\mathbf{1}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}$ . Keskvärtus  $\text{vec}(\hat{\boldsymbol{\mu}})$  tuleneb  $\hat{\boldsymbol{\beta}}$  hinnangust, ning kuna see on nihketa, siis on ka  $\hat{\boldsymbol{\mu}}$  nihketa:

$$\begin{aligned}
E(\text{vec}(\hat{\boldsymbol{\mu}})) &= E(\text{vec}(\mathbf{I}_m \bar{\mathbf{Y}}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1})) = (\mathbf{I}_m \otimes (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T) E(\text{vec}(\bar{\mathbf{Y}}^T)) \\
&= (\mathbf{I}_m \otimes (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T) \text{vec}(\boldsymbol{\mu} \mathbf{A}^T) = \text{vec}(\mathbf{I}_m \boldsymbol{\mu} \mathbf{A}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1}) = \text{vec}(\boldsymbol{\mu})
\end{aligned}$$

Kovariatsioonimaatriksi  $\text{vec}(\hat{\boldsymbol{\mu}})$ -le saab välja kirjutada, kasutades  $\hat{\boldsymbol{\beta}}$  kovariatsioonimaatriksi valemit:

$$\begin{aligned}
\text{cov}(\hat{\boldsymbol{\mu}}) &= ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{1}^T \mathbf{1})^{-1} \otimes \boldsymbol{\Sigma}_e) + ((\mathbf{A}^T \mathbf{A})^{-1} \otimes (\mathbf{1}^T \mathbf{1})^{-1} p p (\mathbf{1}^T \mathbf{1})^{-1} \otimes \boldsymbol{\Sigma}_\lambda) \\
&= \frac{1}{p} (\mathbf{A}^T \mathbf{A})^{-1} \otimes (\boldsymbol{\Sigma}_e + p \boldsymbol{\Sigma}_\lambda)
\end{aligned}$$

### 8. Maatriksi $\mathbf{S}_e$ jaotuse tuletamine

Maatriksi  $\mathbf{S}_e$  avaldises on omavahel korrutatud jääkide vektorid. Jäägid on aga normaaljaotusega  $N_m(\mathbf{0}, \boldsymbol{\Sigma}_e)$ , seega Wisharti jaotuse definitsiooni kohaselt on  $\mathbf{S}_e$  Wisharti jaotusega, parameetritega  $\boldsymbol{\Sigma}_e$  ja  $N(p-1) - r(q-1)$ . Vabadusastmete arv tuleneb selles, et kokku on  $Np$  vaatlust, kuid avaldises hinnatakse  $N$  erineva inimese keskmist ning igas võrreldavas grupis  $(q-1)$  parameetrit. Seega on vabadusastmete kadu  $N + r(q-1)$ .

### 9. Maatriksi $\mathbf{S}_\lambda$ jaotuse tuletamine

Kovariatsioonimaatriksi  $\mathbf{S}_\lambda$  jaotuse tuletamiseks peab eelnevalt esitama  $\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T$  jaotuse.

$$\begin{aligned}
E(\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T) &= \mathbf{0} \text{ ja } \text{cov}(\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T) = \text{cov}(\bar{\mathbf{Y}}^T) = \text{cov}\left(\text{vec}\left(\frac{1}{p}(\mathbf{1}^T \otimes \mathbf{I}_m) \mathbf{Y}^T \mathbf{I}_N\right)\right) \\
&= \text{cov}\left(\frac{1}{p}(\mathbf{I}_N \otimes (\mathbf{1}^T \otimes \mathbf{I}_m) \text{vec}(\mathbf{Y}^T))\right) = \frac{1}{p^2}(\mathbf{I}_N \otimes (\mathbf{1}^T \otimes \mathbf{I}_m)) \text{vec}(\mathbf{Y}^T) (\mathbf{I}_N \otimes (\mathbf{1} \otimes \mathbf{I}_m)) \\
&= \frac{1}{p^2}(\mathbf{I}_N \otimes \mathbf{1}^T \otimes \mathbf{I}_m) \left( (\mathbf{I}_N \otimes \mathbf{I}_p \otimes \boldsymbol{\Sigma}_e) + (\mathbf{I}_N \otimes \mathbf{1} \mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda) \right) (\mathbf{I}_N \otimes \mathbf{1} \otimes \mathbf{I}_m) \\
&= \frac{1}{p^2} \left( (\mathbf{I}_N \otimes \mathbf{1}^T \otimes \boldsymbol{\Sigma}_e) + (\mathbf{I}_N \otimes p \mathbf{1}^T \otimes \boldsymbol{\Sigma}_\lambda) \right) (\mathbf{I}_N \otimes \mathbf{1} \otimes \mathbf{I}_m) = \frac{1}{p} (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_e) + (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_\lambda).
\end{aligned}$$

Seega on  $\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T$  maatriksnormaaljaotusega  $N_{m,N}\left(\mathbf{0}, \frac{1}{p} \boldsymbol{\Sigma}_e + \boldsymbol{\Sigma}_\lambda, \mathbf{I}_N\right)$  ning  $\sqrt{p}(\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T)$ , kasutades lisas 5, punktis 12 välja toodud teoreemi 2.2.2, on maatriksnormaaljaotusega  $N_{m,N}(\mathbf{0}, \boldsymbol{\Sigma}_e + p \boldsymbol{\Sigma}_\lambda, \mathbf{I}_N)$ . Kasutades Wisharti jaotuse definitsiooni, saab defineerida suuruse

$p(\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T)(\bar{\mathbf{Y}}^T - \boldsymbol{\mu} \mathbf{A}^T)^T \sim W_m(\boldsymbol{\Sigma}_e + p \boldsymbol{\Sigma}_\lambda, N)$ . Järgnevalt läheb vaja peatükis 1 antud omadust 1, kus maatriksiks  $\mathbf{R}$  võetakse  $\mathbf{I}_N - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ , saades

$$p(\bar{\mathbf{Y}}^T - \boldsymbol{\mu}\mathbf{A}^T)(\mathbf{I}_N - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)(\bar{\mathbf{Y}}^T - \boldsymbol{\mu}\mathbf{A}^T)^T \sim W_m(\boldsymbol{\Sigma}_e + p\boldsymbol{\Sigma}_\lambda, N - r).$$

Avades sulud ja kasutades teadmist, et  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}^T \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}$ , siis

$$\begin{aligned} & p(\bar{\mathbf{Y}}^T - \boldsymbol{\mu}\mathbf{A}^T)(\mathbf{I}_N - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)(\mathbf{I}_N - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)(\bar{\mathbf{Y}}^T - \boldsymbol{\mu}\mathbf{A}^T)^T \\ &= p(\bar{\mathbf{Y}}^T - \bar{\mathbf{Y}}^T \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T - \boldsymbol{\mu}\mathbf{A}^T \\ &\quad + \boldsymbol{\mu}\mathbf{A}^T \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)(\bar{\mathbf{Y}} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \bar{\mathbf{Y}} - \mathbf{A}\boldsymbol{\mu}^T + \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \mathbf{A}\boldsymbol{\mu}^T) \\ &= p(\bar{\mathbf{Y}}^T - \hat{\boldsymbol{\mu}}\mathbf{A}^T)(\bar{\mathbf{Y}}^T - \hat{\boldsymbol{\mu}}\mathbf{A}^T)^T = \mathbf{S}_\lambda \sim W_m(\boldsymbol{\Sigma}_e + p\boldsymbol{\Sigma}_\lambda, N - r). \end{aligned}$$

10. Hinnangu  $\hat{\boldsymbol{\Sigma}}_e$  tuletamine

$\hat{\boldsymbol{\Sigma}}_e$  nihketa hinnang tuleneb järgnevalt:  $E(\mathbf{S}_e) = (N(p-1) - r(q-1))\boldsymbol{\Sigma}_e$ , seega  $\hat{\boldsymbol{\Sigma}}_e = \mathbf{S}_e / (N(p-1) - r(q-1))$ .  $\hat{\boldsymbol{\Sigma}}_\lambda$  nihketa hinnang tuleneb analoogselt:

$$E(\mathbf{S}_\lambda) = (N-r)(p\boldsymbol{\Sigma}_\lambda + \boldsymbol{\Sigma}_e), \text{ seega } \hat{\boldsymbol{\Sigma}}_\lambda = \left( \frac{\mathbf{S}_\lambda}{N-r} - \hat{\boldsymbol{\Sigma}}_e \right) \frac{1}{p}.$$

11.  $E(\text{vec}(\hat{\mathbf{\Delta}}))$  ja  $\text{cov}(\text{vec}(\hat{\mathbf{\Delta}}))$  leidmine

$H_0$  kehtides on  $\mathbf{M}\boldsymbol{\Psi}\mathbf{C}^T = \mathbf{0}$ . Seega  $E(\text{vec}(\hat{\mathbf{\Delta}})) = (\mathbf{C} \otimes \mathbf{M})\text{vec}(\boldsymbol{\Psi}) = \text{vec}(\mathbf{M}\boldsymbol{\Psi}\mathbf{C}^T) = \mathbf{0}$ .

Nüüd saab, et  $\text{cov}(\text{vec}(\hat{\mathbf{\Delta}})) = (\mathbf{C} \otimes \mathbf{M})((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1}) \otimes \boldsymbol{\Sigma}_e)((\mathbf{C} \otimes \mathbf{M}))^T = \mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T \otimes \mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T$ .

12. Maatriksi  $\mathbf{H} = \mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T(\mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T)^{-1}\mathbf{C}\hat{\boldsymbol{\Psi}}^T\mathbf{M}^T$  jaotuse tuletamine

Maatriksi  $\mathbf{H} = \mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T(\mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T)^{-1}\mathbf{C}\hat{\boldsymbol{\Psi}}^T\mathbf{M}^T$  jaotuse saamiseks pannakse kirja kõigepealt  $\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T$  jaotus. On teada, et  $\text{vec}(\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T)$  ja  $\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T$  jaotused on omavahel seotud järgmise eeskirjaga (Rosen & Kollo, 2005): kui

$$\text{vec}(\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T) \sim N_{cg}(\mathbf{0}, \mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T \otimes \mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T),$$

siis

$$\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T \sim N_{c,g}(\mathbf{0}, \mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T, \mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T).$$

Järgnevalt kasutatakse von Roseni & Kollo raamatus toodud teoreemi 2.2.2, mis ütleb, et iga  $\mathbf{R}: w \times c$  ja  $\mathbf{Q}: t \times g$  korral

$$\mathbf{R}\mathbf{M}\hat{\boldsymbol{\Psi}}\mathbf{C}^T\mathbf{Q}^T \sim N_{w,t}(\mathbf{R}\mathbf{O}\mathbf{Q}^T, \mathbf{R}\mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T\mathbf{R}^T, \mathbf{Q}\mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T\mathbf{Q}^T).$$

Maatriksiks  $\mathbf{R}$  võetakse ühikmaatriks  $\mathbf{I}_c$  ja maatriksiks  $\mathbf{Q}$ :

$$\mathbf{Q} = (\mathbf{C}(\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T)^{-\frac{1}{2}}$$

Seega  $\mathbf{I}_c \mathbf{M}\hat{\Psi}\mathbf{C}^T\mathbf{Q}^T \sim N_{c,g}(\mathbf{0}, \mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T, \mathbf{I}_g)$ . Kasutades Wisharti jaotuse definitsiooni, saab defineerida uue muutuja

$$\mathbf{H} = \mathbf{M}\hat{\Psi}\mathbf{C}^T\mathbf{Q}^T\mathbf{Q}\mathbf{C}\hat{\Psi}^T\mathbf{M}^T = \mathbf{M}\hat{\Psi}\mathbf{C}^T(\mathbf{C}((\mathbf{Z}^T\mathbf{Z})^{-1} \otimes (\mathbf{A}^T\mathbf{A})^{-1})\mathbf{C}^T)^{-1}\mathbf{C}\hat{\Psi}^T\mathbf{M}^T,$$

mis on Wisharti jaotusega, parameetritega  $\mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T$  ja  $g$ .

### 13. Maatriksi $\mathbf{G}$ jaotuse tuletamine

$\mathbf{G} = \mathbf{M}\mathbf{S}_e\mathbf{M}^T$  jaotus tuleneb Wisharti jaotuse omadusest, mis ütleb, et kui  $\mathbf{M}: c \times m$  on täisastakuga ja  $\mathbf{S}_e$  on Wisharti jaotusega  $W_m(\boldsymbol{\Sigma}_e, N(p-1) - r(q-1))$ , siis  $\mathbf{M}\mathbf{S}_e\mathbf{M}^T \sim W_c(\mathbf{M}\boldsymbol{\Sigma}_e\mathbf{M}^T, N(p-1) - r(q-1))$ .

## Lisa 6. R-i kood peatükis 5 esitatud ja peatükis 6.4 rakendatud teooria kasutamiseks

```
#m, r, N, q, p, X, Z, B, A ja y defineerimine nii nagu peatükis 5

m=5; r=1; N=44; p=3

X=(matrix(c(1,1,1,-1,0,1,-1,2,-1),nrow=3,ncol=3))

A=t(t(rep(r,N)))

Im=diag(m)

Z=(X[,2:3])

q=dim(Z)[2]+1

#kk_valim oli andmestik, mis sisaldas kõigi 44 indiviidi jaoks 3 ajahetke
#(0,56,112) mõõtmisi 14 peptiidi jaoks

#Andmestiku on juurde tekitatud tunnus "nr", mis näitab, millisest algsest
#patsiendist on vaatlus genereeritud

#tunnus "nimed" sisaldab kõiki algselt analüüsitud 14 peptiidi nime.

#Sorteerin andmestiku analüüsiks sobivale kujule

#ja võtan välja mitmemõõtmelise normaaljaotusega peptiidid, mida oli 5
#tükki, nimede vektoris asukohtadega 5, 8, 9, 11 ja 12,

#tekitades nii uue andmestiku "korduseks"

korduseks=kk_valim[order(kk_valim$ID, kk_valim$nr, kk_valim$day),]
korduseks=korduseks[korduseks$ID %in% nimed[c(5,8,9,11,12)],]
korduseks=korduseks[order(korduseks$patient, korduseks$day, korduseks$ID),]

#Defineerin transponeeritud Y

Yt=matrix(korduseks$index, nrow=p*m, ncol=N)

#Defineerin vec(Y)

y=t(t(korduseks$index))

#beta, matriksi B ja müü hinnangud

beta=(solve(t(A)%*(A))%*%t(A) %x% (solve(t(X)%*(X))%*%t(X)) %x% Im) %*%y

B=((solve(t(X)%*(X))%*%t(X)

%x% Im)%*% matrix(korduseks$index, ncol=44)%*%A%*%solve(t(A)%*(A)))

müü_hinnang=tapply(korduseks$index, list(korduseks$ID, korduseks$patient),
mean)%*%A%*%solve(t(A)%*(A))

#Hindan valimi pealt vigade ja indiviidide efekti kovariatsioonimatriksid
#Kõigepealt Se ja Ee hindamine

Se=0
```

```

sumkokku=0
summa=0
#See on artiklis üle kõigi ajahetkede keskmine kõigile inimestele.
Ykeskmine=tapply(korduseks$index,list(korduseks$ID,korduseks$patient),mean)
#Siin hindan Ee
for (k in 1:N) {
sumkokku=sumkokku+summa
summa=0
zx=0
for (j in 1:p) {
vekt=(t(t(korduseks$index[(1+zx):(5+zx)]))-Ykeskmine[,k]-(t(Z[j,])))
%x%Im)%*%beta[6:15]%*%(t(A[,1])))%*%t((t(korduseks$index[(1+zx):(5+zx)]))-
-Ykeskmine[,k]-(t(Z[j,]))) %x% Im)%*%beta[6:15]%*%(t(A[,1])))
zx=zx+5
summa=summa+vekt
}}
Se=sumkokku
Ee=sumkokku/(N*(p-1)-r*(q-1))
#Siin hindan Sl ja El
summa=0
for( k in 1:N) {
summa=summa+(t(t(Ykeskmine[,k]))-
müü_hinnang)%*%(t(A[,1])))%*%(t(Ykeskmine[,k]-müü_hinnang)%*%(t(A[,1])))
}
Sl=p*summa
El=1/p * (Sl/(N-r) - Ee) # siin leidub negatiivseid dispersiooni hinnanguid
#Negatiivsed dispersiooni hinnangud asendatakse 0-ga.
asendus=ifelse(diag(El)<0,0,diag(El))
diag(El)=asendus
#Defineerin yk kovariatsioonimaatriks Oomega
oomega=(diag(p)%x%Ee)+(t(t(rep(1,p))))%*%t(rep(1,p))%x%El)
#Siin defineerin eeta, millega hiljem M*Oomega*t(C) hüpoteese kontrollida,
#sisaldab gammasid
eeta=cbind(beta[6:10],beta[11:15])

```

```

#S-i hinnang valimi pealt, kontrollimaks, kas liitsümmeetria eeldus võiks
#paika pidada:

S=1/(N-r) * (Yt**(diag(44)-A**solve(t(A)**A)**t(A))**t(Yt))

#Testin, kas hinnang valimi pealt ja liitsümmeetria eeldusega saadud
#kovariatsioonimaatriksid võiksid olla samad:

statistikW =(N-1)*(log(det(oomega))-log(det(S))+tr(S**solve(oomega))-m*p)

konstant = (2*(p*m)**2 +3*(p*m)-1)/(6*(N-1)*(p*m+1))

#hii-ruut jaotusega statistiku arvutamine

statistik =(1-konstant)*statistikW

f = p*m*(p*m+1)/2; kriitiline = qchisq(0.95,f)

#Kuna statistik on suurem kui kriitiline väärtus, võtan vastu H1: Oomega ja
#S ei ole samad

#Järgnevasse analüüsi tuleks suhtuda teatud kahtlusega,
#sest formaalselt ei õnnestunud S-i ja Oomega samasust tõestada
#Uurin, kas kõik peptiidid võiksid käituda samamoodi

M=matrix(c(-1,1,0,0,0,0,-1,1,0,0,0,0,-1,1,0,0,0,0,-1,1),ncol=5,byrow=T)

C=diag(2)

H=M**eeta**t(C)**solve(C**(solve(t(Z)**Z)

%x% solve(t(A)**A))**t(C))**C**t(eeta)**t(M)

G=M**Se**t(M)

lambda=prod(1/(1+eigen(solve(G)**H)$values))

kriitiline=0.83 # vaadatud Wilksi lambda alumine kriitiline väärtus 4
#lineaarse kombinatsiooni korral, p=5

lambda

kriitiline # võtan vastu H1: ei käitu

M=matrix(c(1,0,0,-1,0,1,0,0,0,-1),ncol=5,byrow=T)

C=diag(2)

H=M**eeta**t(C)**solve(C**(solve(t(Z)**Z)

%x% solve(t(A)**A))**t(C))**C**t(eeta)**t(M)

G=M**Se**t(M)

lambda=prod(1/(1+eigen(solve(G)**H)$values))

kriitiline=0.897 # vaadatud Wilksi lambda alumine kriitiline väärtus

lambda

kriitiline

# Võtan vastu H0: vaadeldud 3 peptiidi käituvad sarnaselt

```

## Kasutatud kirjandus

- Bartlett, M. (1938). Further Aspects of the Theory of Multiple Regression. *Proceedings of the Cambridge Philosophical Society*, 34, 33-40.
- Brus, O. (2009). *A comparison of normalisation methods for peptide microarrays*. Uppsala: U.U.D.M. Project Report 2009:3.
- Fujikoshi, Y., Ulyanov, V. V., & Shimizu, R. (2010). *Multivariate Statistics. High-Dimensional and Large-Sample Approximations*. New Jersey: John Wiley & Sons, Inc.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Kollo, T. (2012). *Mitmemõõtmelise statistika loengukonspekt*. Tartu: Tartu Ülikool.
- Kollo, T. (16. 05 2013. a.). suuline konsultatsioon.
- Liaw, A. (2004). *Box's M-test for testing homogeneity of covariance matrices*. Kasutamise kuupäev: 13. 04 2013. a., allikas Homepage of Ranjan Maitra: <http://www.public.iastate.edu/~maitra/stat501/Rcode/BoxMTest.R>
- Macedo, M., & Waterson, T. (04. 06 2008. a.). *Profile Analysis*. Kasutamise kuupäev: 24. 04 2013. a., allikas Homepage of Edward F. Connor: <http://userwww.sfsu.edu/efc/classes/biol710/manova/profileanalysis.htm>
- Maitra, R. (28. 02 2013. a.). *Multivariate Analysis of Variance II*. Kasutamise kuupäev: 14. 04 2013. a., allikas Statistics 501 (Multivariate Statistical methods) Lectures: <http://www.public.iastate.edu/~maitra/stat501/lectures/MANOVA-II.pdf>
- Nahtman, T., Jernberg, A., MahdaviFar, S., Schutkowski, M., Maeurer, M., & Reilly, M. (2007). Validation of peptide epitope microarray experiments and extraction of quality data. *Journal of Immunological Methods*, 382, 1-13.
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Ravishanker, N., & Dey, D. K. (2002). *A First Course in Linear Model Theory*. Boca Raton: Chapman & Hall/CRC.

- Reinsel, G. (1982). Multivariate Repeated-Measurement or Growth Curve Models with Multivariate Random Effects Covariance Structure. *Journal of the American Statistical Association*, Vol. 77, No. 377, 190-195.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. New York: John Wiley & Sons.
- Rosen, D. v., & Kollo, T. (2005). *Advanced Multivariate Statistics with Matrices*. Dordrecht: Springer.
- Siegrist, C.-A. (2008). Vaktsiinide immunoloogia. rmt: S. A. Plotkin, W. A. Orenstein, & P. A. Offit, *Vaktsiinid* (lk 17-36). Elsevier Health Sciences.
- Sikut, R. (01. märts 2001. a.). *Rakubioloogia loengu konspekt*. Kasutamise kuupäev: 02. 04 2013. a., allikas <http://cellbio.ebc.ee/rakubio/immuun.htm>
- Timm, N. H. (2002). *Applied Multivariate Analysis*. New York: Springer-Verlag.
- Valge, M. (2012). *Tunnuste informatiivsete alamhulkade leidmine peptiidide mikrokiibi andmete näitel*. Tartu: Tartu Ülikool.
- Wikipedia. (27. 03 2013. a.). *False Discovery Rate*. Kasutamise kuupäev: 12. 04 2013. a., allikas Wikipedia: [http://en.wikipedia.org/wiki/False\\_discovery\\_rate](http://en.wikipedia.org/wiki/False_discovery_rate)

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Kristi Läll, sünnikuupäevaga 29.01 1989,

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Mitmemõõtmeline analüüs peptiidide käitumise uurimiseks“, mille juhendajad on Mare Vähi ja Anne Selart,

1.1 reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus/Tallinnas/Narvas/Pärnus/Viljandis, \_\_14.05 2013\_\_\_\_\_ (kuupäev)