

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
MATEMAATILISE STATISTIKA INSTITUUT

Anu Tensing

Mõjukate vaatluste analüüs lineaarsetes segamudelites

Magistritöö

Juhendaja

Mare Vähi

TARTU

2014

Mõjukate vaatluste analüüs lineaarsetes segamudelites

Lineaarsed segamudelid sisaldavad nii fikseeritud kui ka juhuslikke faktoreid, mistõttu need on väga paindlikud. Lineaarsete segamudelite korral ei eeldata vaatluste sõltumatust. Seega on nimetatud mudelid tihtipeale kasutatavad korduvmõõtmistega uuringutes näiteks meditsiinilistes ja bioloogilistes uuringutes. Kuna üksikute vaatluste välja jätmine mõjutab lineaarsete segamudelite korral ka variatsiooniparameetreid, siis mõjukate vaatluste leidmiseks ei saa kasutada klassikalisi meetodeid, mis ei võta arvesse muutust varieeruvuses. Töö sisaldab ülevaadet lineaarsete segamudelite teooriast, jääkide analüüsist ja mõjukate vaatluste leidmisest. Samuti on rakendatud eelnevat teooriat peptiidide mikrokiipide andmetel, mis sisaldavad korduvmõõtmisi. Märksõnad: *DNA mikrokiibid, longituudanalüüs, mitmetasandilised mudelid, vöörväärtused*

Influence analysis for linear mixed models

Linear mixed models contain both fixed effects and random effects. They are widely used in biological, social and medical studies. Linear mixed models can also be used with correlated data, which make them very useful for data containing repeated measurement or measurement made on clusters. Detecting influential point is complicated because removing the data point affects both fixed and random effects and also covariance parameters. This work contains an overview of influence measures and residual analysis in linear mixed models. In this work, a practical example using peptide microarray data, which contains repeated measurements, is presented.

Key words: *DNA microarrays, longitudinal analysis, multilevel models, outliers*

Sisukord

Infoleht	1
Sissejuhatus	5
1 Lineaarsed segamudelid	6
1.1 Lineaarsed mudelid	6
1.2 Lineaarsed segamudelid	9
1.2.1 Juhuslikele efektide prognoosimine	11
1.2.2 Segamudelite võrrandid	11
1.2.3 Variatsioonikomponentide leidmine	13
1.2.4 Lineaarne segamudel korduvate mõõtmiste korral	14
2 Jääkide analüüs	16
3 Mõjukus	18
3.1 Globaalne mõjukus	19
3.2 Mõju fikseeritud efektide hinnangule	20
3.2.1 Mõju ühe fikseeritud efekti hinnangule	20
3.2.2 Klassi mõju mitmele efektile korraga	21
3.2.3 Mõju fikseeritud efektide lineaarkombinatsioonile	22
3.3 Mõju juhuslike efektide prognoosidele	22
3.3.1 Mõju kõikidele juhuslikele efektidele	22
3.3.2 Erinevust vaadeldud väärtuse ja prognoositud margi- naalse keskmise vahel	23
3.4 Mõju fikseeritud efektide hinnangute ja juhuslike efektide prog- nooside täpsusele	23
3.4.1 Mõju fikseeritud efektide hinnangute täpsusele	24

3.4.2	Mõju juhuslike efektide prognooside täpsusele	25
4	Erindite leidmine peptiidide mikrokiibi andmetel	27
4.1	Peptiidide mikrokiipide tööpõhimõte	27
4.2	Andmed	28
4.3	Peamõjudega mudel	30
4.3.1	Jääkide analüüs	30
4.3.2	Globaalne mõjucus	31
4.3.3	Mõju fikseeritud efektide hinnangule	34
4.3.4	Mõju juhuslikele efektide prognoosidele	35
4.3.5	Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele	37
4.4	Hierarhiline mudel	38
4.4.1	Jääkide analüüs	39
4.4.2	Globaalne mõjucus	40
4.4.3	Mõju fikseeritud efektide hinnangule	42
4.4.4	Mõju juhuslikele efektide prognoosidele	45
4.4.5	Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele	47
4.5	Hierarhiline mudel sõltuvate juhuslike efektidega	48
4.5.1	Jääkide analüüs	50
4.5.2	Globaalne mõjucus	50
4.5.3	Mõju fikseeritud efektide hinnangule	52
4.5.4	Mõju juhuslikele efektide prognoosidele	55
4.5.5	Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele	57
4.6	Tulemused	58

Kokkuvõte	59
Kasutatud kirjandus	61
Lisad	63
Lisa A	63
Lisa B	64
Lisa C	66
Lisa D	67
Lisa E	69
Lisa F	70
Lisa G	72
Lisa H	73
Lisa I	74
Lisa J	75
Lisa K	76
Lisa L	77
Lisa M	79
Lisa N	81
Lisa O	82
Lisa P	83
Lisa Q	84
Lisa R	85
Lisa S	86

Sissejuhatus

Lineaarsed segamudelid on laialdaselt kasutusel uurimaks suuremahulisi ja keerulisi andmestikke. Neid kasutatakse tihti bioloogias, meditsiinis ja muudel taolistel aladel. Lineaarsed segamudelid on klassikaliste lineaarsete mudelite edasiarendus. Lineaarsed mudelid sisaldavad nii fikseeritud kui ka juhuslikke faktoreid, mistõttu on lineaarsed segamudelid väga pädlikud ehk sobivad väga erinevate andmestikega. Erinevalt klassikalistest lineaarsetest mudelitest ei eeldata lineaarsete segamudelitel korral, et vaatlused on omavahel sõltumatud. Seega kasutatakse neid mudeleid tihti uuringutes, kus on tehtud korduvmõõtmisi või andmestik on klasterdatud.

Töö esimeses osas on toodud ülevaade lineaarsete segamudelite teooriast. Töö sisaldab tuletuskäiku lineaarsetest mudelitest lineaarsete segamudeliteni ning suurima tõepära hinnangute leidmist. Teises osas räägitakse jääkide analüüsist. Lineaarsete segamudelite korral saab vaadelda kolme sorti jääke, kusjuures neist kahe abil saab leida erindeid.

Töö kolmas osa keskendub mõjukate punktide leidmisele. Kuna lineaarsed segamudelid sisaldavad nii fikseeritud kui ka juhuslikke faktoreid, siis üksikute elementide välja jätmine mõjutab ka variatsioonikomponente. Klassikalised mõjukuse näitajad ei võta arvesse muutust variatsioonikomponentides. Töös on toodud põhilised näitajad, mille abil hinnata mõjukust erinevate aspektide suhte.

Töö viimases osas on rakendatud eelnev teooria peptiidide mikrokiipide näitel. Tänapäeval on võimalik mikrokiipide abil uurida väga suuremahulisi andmestikke. Töös on kasutatud andmeid meditsiinilisest uuringust, kus igal patsiendil on tehtud mitu mõõtmist, mistõttu vaatlused on sõltuvad.

Töö praktiline osa on läbi viidud kasutades tarkvarapaketti R. Töö on valminud ETF grandis 8294 toel.

1 Lineaarsed segamudelid

Segamudeleid kasutatakse tihti sotsiaalsetes ja meditsiinilistes uuringutes. Nimetatud mudelit on eriti kasulikud, kui igal vaatluse all oleval objektil on tehtud mitu mõõtmist, näiteks meditsiiniline uuring, kus on igalt patsiendilt võetud teatud aja järelt proov, et uurida, kuidas vaadeldav ravim toimib, või vaatlused on jaotatud klassidesse, näiteks testi tulemus erinevates koolides, kus on võetud valim õpilastest.

Lineaarseks segamudeliks nimetatakse lineaarset mudelit, millel on lisaks fikseeritud faktoritele ka juhuslikud faktorid. Seega lineaarsed segamudelid on lineaarsete mudelite edasiarendus. Antud peatükis tuletamegi lineaarsest mudelist lineaarse segamudeli. Antud osa tööst põhineb raamatul West (2007) ja artiklil Harville (1977).

1.1 Lineaarsed mudelid

Klassikaline lineaarne mudel on defineeritud järgmiselt

$$y = X\beta + \epsilon^*, \quad (1.1)$$

kus vektor $y : n \times 1$ sisaldab uuritava tunnuse vaatluseid, maatriks $X : n \times p$ on teadaolev plaanimaatriks ehk disainimaatriks, vektor $\beta : p \times n$ sisaldab fikseeritud faktoritele vastavaid tundmatuid parameetreid ehk fikseeritud efekte ja $\epsilon^* : n \times 1$ on juhuslike vigade vektor. Eeldame, et juhuslike vigade keskväärts on 0, st $E(\epsilon^*) = 0$, ning juhuslikud vead on paarikaupa sõltumatud ja sama dispersiooniga $\sigma_{\epsilon^*}^2$, st $D(\epsilon^*) = \sigma_{\epsilon^*}^2 I_n$. Lisaks eeldame, et juhuslikud vead on normaaljaotusega, siis $\epsilon^* \sim N(0, \sigma_{\epsilon^*}^2 I_n)$. Sellest omakorda saame, et kehtib ka $E(y) = X\beta$ ja $D(y) = \sigma_{\epsilon^*}^2 I_n$. Seega $y \sim N(X\beta, \sigma_{\epsilon^*}^2 I_n)$.

Kuna meil on uuritava tunnuse y jaotus teada, siis üks võimalus hinnata mudeli 1.1 parameetreid β on kasutada suurima tõepära meetodit. Selleks tuleb optimiseerida tõepärafunktsiooni, ent tihtipeale optimiseeritakse selle asemel logaritmilist tõepärafunktsiooni, mille maksimum asub samas kohas. Uuritava tunnuse y tõepärafunktsioon avaldub kujul

$$L(\beta, \sigma_{\epsilon^*}^2) = \frac{1}{(2\pi\sigma_{\epsilon^*}^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma_{\epsilon^*}^2}(y - X\beta)'(y - X\beta)\right\} \quad (1.2)$$

ja sellest omakorda saame, et logaritmiline tõepärafunktsioon on

$$l(\beta, \sigma_{\epsilon^*}^2) = -\frac{n}{2} \ln(2\pi\sigma_{\epsilon^*}^2) - \frac{1}{2\sigma_{\epsilon^*}^2}(y - X\beta)'(y - X\beta). \quad (1.3)$$

Kui me võtame logaritmilisest tõepärafunktsioonist tuletise, siis saame

$$\frac{\partial l(\beta, \sigma_{\epsilon^*}^2)}{\partial \beta} = \frac{1}{2\sigma_{\epsilon^*}^2}(X'y - X'X\beta). \quad (1.4)$$

Võrdsustades saadud logaritmilise tõepärafunktsiooni tuletise nulliga saame võrrandi

$$X'X\beta = X'y. \quad (1.5)$$

Kui maatriks $X'X$ on mittesingulaarne, siis sellel leidub pöördmaatriks $(X'X)^{-1}$. Sellisel juhul saame parameetri β hinnanguks $(X'X)^{-1}X'y$. Üldiselt ei pruugi maatriks $X'X$ olla mittesingulaarse, st ei pruugi leiduda maatriksi $X'X$ pöördmaatriksit $(X'X)^{-1}$.

Tähistame ülaindeksiga "–" üldistatud pöördmaatriksit, st maatriks $(X'X)^{-}$ on maatriksi $X'X$ üldistatud pöördmaatriks. Kusjuures üldistatud pöördmaatriks ei ole üldiselt üheselt määratud. Nüüd defineerime

$$\hat{\beta} = (X'X)^{-}X'y. \quad (1.6)$$

Kuna $\hat{\beta}$ rahuldab üldistatud pöördmaatriksi definitsiooni põhjal võrrandit 1.5, siis $\hat{\beta}$ on suurima tõepära hinnang parameetrite vektorile β . Näeme, et hinnang 1.6 sõltub maatriksi $X'X$ üldistatud pöördmaatriksist, mis ei ole üheselt määratud. Seega $\hat{\beta}$ ei ole üheselt määratud.

Nüüd parim lineaarne nihketa hinnang uuritava tunnuse y keskväärtusele $X\beta$ on

$$X\hat{\beta} = X(X'X)^{-1}X'y, \quad (1.7)$$

kusjuures erinevalt parameetervektori β hinnangust $\hat{\beta}$ on hinnang $X\hat{\beta}$ üheselt määratud vaatamata sellele, et see sisaldab maatriksi $X'X$ üldistatud pöördmaatriksit.

Järgnevas ärme eelda, et mudeli 1.1 vead on paarikaupa sõltumatud ja sama hajuvusega, st ei pruugi kehtida $D(\epsilon^*) = \sigma_{\epsilon^*}^2 I_n$. Tähistame $V = D(\epsilon^*)$. Sellisel juhul saame leida analoogiliselt leida suurima tõepära hinnangud. Nüüd saame uuritava tunnuse y tõepärafunktsiooni kirjutada välja kujul

$$L(\beta, V) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left\{-\frac{1}{2}(y - X\beta)' \Sigma^{-1}(y - X\beta)\right\} \quad (1.8)$$

ja sellest omakorda saame logaritmilise tõepärafunktsiooni kujul

$$l(\beta, V) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma| - \frac{1}{2}(y - X\beta)' \Sigma^{-1}(y - X\beta). \quad (1.9)$$

Võttes logaritmilisest tõepärafunktsioonist tuletise parameetervektori β järgi saame

$$\frac{\partial l(\beta, V)}{\partial \beta} = (X' \Sigma^{-1} y - X' \Sigma^{-1} X \beta)'. \quad (1.10)$$

Võrdsustades saadud logaritmilise tõepärafunktsiooni tuletise nulliga saame võrrandi

$$X'V^{-1}X\beta = X'V^{-1}y. \quad (1.11)$$

Kui maatriks $X'V^{-1}X$ on mittesingulaarne, siis leidub maatriksil $X'V^{-1}X$ pöördmaatriks $(X'V^{-1}X)^{-1}$ ja seega saame parameetervektorile β hinnangu $(X'V^{-1}X)^{-1}X'V^{-1}y$. Üldiselt $X'V^{-1}X$ ei pruugi olla mittesingulaarne. Defineerime nüüd parameetervektorile β hinnangu

$$\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}y. \quad (1.12)$$

Üldistatud pöördmaatriksi definitsiooni põhjal saab näidata, et parameetervektori β hinnang $\hat{\beta}$ rahuldab võrrandit 1.11, kusjuures antud hinnang ei ole üheselt määratud. Nüüd saame uuritava tunnuse y keskvaärtusele $X\beta$ parima lineaarse nihketa hinnangu kujul

$$X\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}y, \quad (1.13)$$

kusjuures antud hinnang on üheselt määratud vaatamata sellele, et see sisaldab maatriksi $X'V^{-1}X$ üldistatud pöördmaatriksit $(X'V^{-1}X)^{-}$, mis ei pruugi olla üheselt määratud.

1.2 Lineaarsed segamudelid

Järgnevas ärme eelda, et mudeli 1.1 vead on paarikaupa sõltumatud ja sama hajuvusega, st ei pruugi kehtida $D(\epsilon^*) = \sigma_{\epsilon^*}^2 I_n$. Olgu mudeli 1.1 vigade dispersioonimaatriks V . Eeldame, et maatriks V on teada. Tähistame

$$\epsilon^* = Zu + \epsilon, \quad (1.14)$$

kus vektor $u : r \times 1$ on juhuslikele faktoritele vastav tundmatute ehk juhuslike efektide vektor, maatriks $Z : n \times r$ vektorile u vastav plaanimaatriks ehk disainimaatriks ja ϵ juhuslike vigade vektor.

Seega saame nii fikseeritud kui juhuslike faktoritega mudeli kujul

$$y = X\beta + Zu + \epsilon, \quad (1.15)$$

kus β on fikseeritud faktoritele vastavate parameetrite ehk fikseeritud efektide vektor, u on juhuslikele faktoritele vastavate parameetrite ehk juhuslike efektide vektor, X ja Z on vastavad plaanimaatriksid ja ϵ on juhuslike vigade vektor. Mudelit 1.15 nimetataksegi lineaarseks segamudeliks.

Kuna vigade ϵ^* keskväärtus on null, siis omakorda eeldame, et juhuslike efektide vektori u ja juhuslike vigade vektori ϵ keskväärtused on samuti nullilised, st $E(u) = 0$ ja $E(\epsilon) = 0$. Nüüd saame uuritava tunnuse y keskväärtuseks $X\beta$, st $E(y) = X\beta$.

Lisaks tähistame $D(u) = G$ ja $D(\epsilon) = R$, kus G ja R on mittesingulaarsed positiivselt määratud maatriksid, ning olgu juhuslike efektide vektor u ja juhuslike vigade vektor ϵ omavahel sõltumatud, st $Cov(u, \epsilon) = 0$. Kuna vektor $X\beta$ on mittejuhuslik, siis uuritava tunnuse y ja vigade ϵ^* dispersioonimaatriksid on võrdsed. Nüüd vastavalt vektorite u ja ϵ dispersioonimaatriksite definitsioonidele ja nende vektorite sõltumatusele saame vigade vektori ϵ^* dispersioonimaatriksiks $ZGZ' + R$, st $V = ZGZ' + R$. Järelikult $D(y) = V = ZGZ' + R$.

Olgu vektorid u ja ϵ mitmemõõtmelise normaaljaotusega, siis saame, et $u \sim N(0, G)$ ja $\epsilon \sim N(0, R)$. Nüüd ka uuritav tunnus y on mitmemõõtmelise normaaljaotusega, st $y \sim N(X\beta, V)$.

Kuna maatriks V on teada, siis saame parameetervektorile β suurima tõepära hinnangu $\hat{\beta} = (X'X)^- X'y$, kusjuures antud hinnang ei pruugi olla üheselt

määratud. Sellest saame omakorda parima lineaarse nihketa hinnangu uuritava tunnuse y keskväärtusele $X\beta$ kujul $X\hat{\beta} = X(X'X)^{-1}X'y$, mis on üheselt määratud.

Kusjuures, kui juhuslikud faktorid puuduvad ja $V = \sigma_{\epsilon}^2 I_n$, siis saame lineaarsest segamudelist 1.15 klassikalise lineaarse mudeli 1.1.

1.2.1 Juhuslikele efektide prognoosimine

Kuna uuritav tunnus y ja juhuslike efektide vektor u on vastavalt jaotustega $N(X\beta, V)$ ja $N(0, G)$ ning juhuslikud faktorid ja juhuslikud vead on omavahel sõltumatud, siis saame vektorite y ja u vaheliseks kovariatsiooniks $Cov(y, u) = Cov(X\beta + Zu + \epsilon, u) = ZD(u) = ZG$. Sellest saame vektorite y ja u ühisjaotuse

$$\begin{pmatrix} y \\ u \end{pmatrix} \sim N\left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V & ZG \\ GZ' & G \end{pmatrix}\right). \quad (1.16)$$

Sellest saame omakorda, et vektori u tinglik jaotus teades vektori y väärtusi on $N(GZ'V^{-1}(y - X\beta), G - GZ'V^{-1}ZG)$. Parima lineaarse nihketa prognoosi \tilde{u} vektori u jaoks saame leides keskväärtuse vektori y suhtes suurusest $u | y$. Seega

$$\tilde{u} = E(u | y) = GZ'V^{-1}(y - X\beta). \quad (1.17)$$

1.2.2 Segamudelite võrrandid

Juhul, kui juhuslikud efektid on fikseeritud, st vektor u on mittejuhuslik, siis saame uuritava tunnuse y keskväärtuseks $X\beta + Zu$ ja dispersioonimaatriksis juhuslike vigade dispersioonimaatriksi R . Seega $y | u \sim N(X\beta + Zu, R)$.

Kuna vektor u on jaotusest $N(0, G)$, siis saame vektorite y ja u ühistiheduseks

$$\begin{aligned} f(y, u) &= f(y | u)f(u) \approx \\ &\approx \exp\left(-\frac{1}{2}(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu)\right) \exp\left(-\frac{1}{2}u'V^{-1}u\right) = \\ &= \exp\left(-\frac{1}{2}[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) + u'V^{-1}u]\right). \end{aligned}$$

Tähistame $\tilde{f}(y, u) = \exp\left(-\frac{1}{2}[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) + u'V^{-1}u]\right)$.

Kuna $f(y, u)$ ja $\tilde{f}(y, u)$ erinevad vaid positiivse konstandi korda, siis funktsiooni $\tilde{f}(y, u)$ naturaalogaritmi maksimiseerimisel saame ühisjaotuse (y, u) suurima tõepära hinnangud. Nüüd

$$\ln \tilde{f}(y, u) = -\frac{1}{2}[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) + u'V^{-1}u].$$

Tähistame $Q(\beta, u) = (y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) + u'V^{-1}u$, siis piisab suurima tõepära hinnangute saamiseks minimiseerida $Q(\beta, u)$. Selleks leiame osatuletised $\frac{\partial Q(\beta, u)}{\partial \beta}$ ja $\frac{\partial Q(\beta, u)}{\partial u}$ ning võrdustame need nulliga. Funktsiooni $Q(\beta, u)$ osatuletised vektorite β ja u järgi on

$$\begin{aligned} \frac{\partial Q(\beta, u)}{\partial \beta} &= -2X'R^{-1}(y - X\beta - Zu), \\ \frac{\partial Q(\beta, u)}{\partial u} &= -2Z'R^{-1}(y - X\beta - Zu) + 2G^{-1}u. \end{aligned}$$

Võrdustades saadud osatuletised nulliga saame võrrandisüsteemi

$$\begin{cases} X'R^{-1}X\beta + X'R^{-1}Zu = X'R^{-1}y, \\ Z'R^{-1}X\beta + (Z'R^{-1}Z + G^{-1})u = Z'R^{-1}y. \end{cases} \quad (1.18)$$

Võrrandeid 1.18 nimetatakse lineaarsete segamudelite võrranditeks, kusjuures hinnang $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ ja prognoos $\tilde{u} = GZ'V^{-1}(Y - X\beta)$ rahuldavad lineaarsete segamudelite võrrandeid. Saadud võrrandisüsteem saa-

me kirjutada ka maatrikskujul

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}. \quad (1.19)$$

1.2.3 Variatsioonikomponentide leidmine

Olgu nüüd dispersioonimaatriks V tundmatu. Kui hinnata dispersioonimaatriks V suurima tõepära meetodiga ja saadud hinnanguga asendada maatriks V hinnangus $\hat{\beta}$ ja prognoosis \tilde{u} , siis võime saada nihkega hinnangu parameetervektori β jaoks ja nihkega prognoosi vektori u jaoks. Samas kasutades selle asemel kitsendatud suurima tõepära meetodit (REML) saame nihketa hinnangud. Asendades dispersioonimaatriksi V kitsendatud suurima tõepära hinnangu parameetervektori β hinnangusse $\hat{\beta}$ ja vektori u prognoosi \tilde{u} saame vastavalt parameetervektori β empiirilise parima lineaarse nihketa hinnangu (EBLUE) ja u empiirilise parima lineaarse nihketa prognoosi (EBLUP).

Kitsendatud suurima tõepära meetodi idee on leida selline maatriks $A = (a_1, \dots, a_{n-\text{rank}(X)}) : (n - \text{rank}(X)) \times n$, mille korral veeruvektorid on lineaarselt sõltumatud, st $\text{rank}(A) = n - \text{rank}(X)$, ja $A'X = 0$, ning leida teisendatud andmestiku $A'y$ peale suurima tõepära hinnangud, kusjuures saadud hinnang ei sõltu konkreetsest A valikust.

Kuna maatriksi $P_X = X(X'X)^{-1}X'$ korral

$$(I - P_X)X = X - X(X'X)^{-1}X'X = X - X = 0$$

ja $\text{rank}(I - P_X) = n - \text{rank}(X)$, siis võime võtta $A = (I - P_X)'$. Seega

$$\begin{aligned} A'y &= (I - P_X)y = (I - P_X)(X\beta + Zu + \epsilon) = \\ &= (I - P_X)X\beta + (I - P_X)Zu + (I - P_X)\epsilon = (I - P_X)Zu + (I - P_X)\epsilon. \end{aligned}$$

Tähistame $w = A'y$, siis $w \sim N(0, (I - P_X)V(I - P_X)')$. Nüüd saame logaritmilise tõepära funktsiooni tingiku vektori $u \mid w$ jaoks

$$\begin{aligned} l(u \mid w) &= -\frac{n-r}{2} \ln(2\pi) - \frac{1}{2} \ln | (I - P_X)V(I - P_X)' | - \\ &- \frac{1}{2} w' [(I - P_X)V(I - P_X)']^{-1} w. \end{aligned}$$

Üldisel kujul saab leitud logaritmilisest tõepärast leida u hinnangu numbriliste meetodite abil.

1.2.4 Lineaarne segamudel korduvate mõõtmiste korral

Lineaarseid mudeleid kasutatakse tihti sõltuvuse tõttu korduvate mõõtmiste korral. Üheks erijuhuks on näiteks meditsiinilised uuringud, kus iga inimese kohta tehakse mõõtmiseid iga mingi aja jooksul. Sellisel juhul on loomulik eeldada, et erinevate patsientide vaatlused ei ole üksteisest sõltuvad, kuid iga patsiendi korral on kõik tema vaatlused üldjuhul sõltuvad. Olgu uuringus k patsienti ja i . patsiendil olgu tehtud n_i mõõtmist, kusjuures $\sum_{i=1}^k n_i = n$. Seega saame i . patsiendi jaoks mudeli

$$y_i = X_i\beta + Z_i u_i + \epsilon_i, \quad (1.20)$$

kus $y_i = (y_{i1}, \dots, y_{in_i})$ on i . patsiendi vaatluste vektori, $\beta : p \times 1$ on fikseeritud efektide vektor, $u_i : q \times n_i$ on i . patsiendile vastav juhuslike efektide vektor, $X_i : n_i \times p$ ja $Z_i n_i \times q_i$ on i . patsiendile vastav plaanimaatriksid ning $\epsilon_i : n_i \times 1$ on i . patsiendile vastavate vigade vektor.

Eeldame, et $\epsilon_i \sim N(0, R_i)$, $u_i \sim N(0, D)$, kus R_i ja D on regulaarsed positiivselt määratud, ning e_i ja u_j on iga i, j korral sõltumatud.

Maatrikskuju saame kirjutada antud mudeli järgmiselt

$$y = X\beta + Zu + \epsilon, \quad (1.21)$$

kus $y = (y'_1, y'_2, \dots, y'_k)' : n \times 1$, $X = (X'_1, X'_2, \dots, X'_k)' : n \times p$, $u = (u'_1, u'_2, \dots, u'_k)' : q \times 1$, $Z = \bigoplus_{i=1}^k Z_i$, kus \bigoplus tähistab maatriksite otseusmmat, st Z on plokk-diagonaalmaatriks, mille i . plokk on Z_i , ja $\epsilon = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_k)' : n \times 1$. Sellisel juhul saame dispersioonimaatriksiteks $G = \bigoplus_{i=1}^k D$ ja $R = \bigoplus_{i=1}^k R_i$. Sellest saame omakorda uuritava tunnuse dispersioonimaatriksi kuju $V = \bigoplus V_i$, kus $V_i = Z_i D Z'_i + R_i$.

2 Jääkide analüüs

Tavalise lineaarse mudeli korral vaadeldakse ainult üht sorti jääke, ent lineaarsete segamudelite korral võib vaadelda kolme erinevat sorti jääke (Singer *et al.* 2013, Hilden-Minton 1995):

- Marginaalseteks jääkides nimetakse vaadeldud väärtuse ja hinnatud marginaalse keskmise vahel, st $\hat{e}_{marg} = y - X\hat{\beta}$. Need vastavad vigadele $e_{marg} = y - E(y) = y - X\beta = Zu + \epsilon$.
- Tinglikeks jääkideks nimetatakse vaadeldud väärtuse ja prognoositud väärtuse vahet, st $\hat{e}_{ting} = y - X\hat{\beta} - Z\tilde{u}$, mis vastavad vigadele $e_{ting} = y - E(y | u) = y - X\beta - Zu = \epsilon$. Erijuhul, kui juhuslikud faktorid puuduvad, langevad marginaalsed ja tinglikud jäägid kokku.
- Juhuslikud jäägid $\hat{e}_j = Z\tilde{u}$, mis vastavad juhuslikele vigadele $e_j = E(y | u) - E(y) = X\beta + Zu - X\beta = Zu$.

Teisendamata kujul jäägid ei sobi võimalike erindlike vaatluste leidmiseks, kuna jäägid sõltuvad ka dispersioonimaatriksitest R ja G . Kui vaatluste varieeruvus on erinev, siis võivad väiksema jäägiga vaatlused tegelikuses olla erindlikumad kui suurema jäägiga vaatlused.

Kuna enamasti ei ole jääkide dispersioon teada, siis üldiselt ei saa antud jääke standardiseerida. Selle asemel võib kasutada studentiseeritud või Pearsoni jääke. Studentiseerimisel jagatakse jäägid jääkide standardhälbe hinnanguga. Kui standardhälbe hindamisel on i . vaatlus välja jäetud, siis leitud jääki i . vaatlusele nimetatakse tema sisemiseks studentiseeritud jäägiks, vastasel juhul välimiseks. Kui kasutame jääkide standardhälbe hinnangu asemel y standardhälvet, saame Pearsoni jäägid.

Erinevate jääkide abil saame kontrollida erinevaid näitajaid. Marginaalsete

jääkide abil saab kontrollide fikseeritud faktorite lineaarsust uuritava tunnuse suhtes. Tinglike ja juhuslike jääkide abil saab tuvastada võimalikke erindeid. Lisaks sellele saab tinglike jääkide abil kontrollida mudeli jääkide ϵ normaalsust ja juhuslike jääkide abil fikseeritud faktorite lineaarsust uuritava tunnuse suhtes.

3 Mõjukus

Mõjukus on defineeritud, kui üksiku vaatluse või vaatluste grupi võime muuta olulisi analüüsi-aspekte läbi nende andmestikus olemise või puudumise. Mõjukuse analüüsi eesmärgiks ei ole vaatluste kustutamise teel saada parem mudeli sobivus vähendatud andmestikule, kuigi see võib olla analüüsi tulemus. Eesmärgiks on leida vaatlused, mis on mõjukad, ja uurida, kuidas nad mõjutavad erinevaid aspekte. Antud osa tööst põhineb artiklidel Nieuwenhuis (2012), Zewotir (2005) ja Schabenberger (2004).

Kui juhuslikud efektid puuduvad, siis saame klassikalise lineaarse mudeli. Selle mõjukuse näitajad on hästi teada. Erinevus klassikalise lineaarse mudeli ja lineaarse segamudeli vahel on see, et lineaarse segamudeli korral hinnang $\hat{\beta}$ ja prognoos \hat{u} sõltuvad dispersioonimaatriksitest R , G ja V . Kui andmestikust välja jätta üks vaatlus, saab leida andmestikult uue hinnangu ja uue prognoosi ilma mudelit uuesti sobitamata ainult siis, kui eeldatakse, et dispersioonimaatriksid on teada.

Lineaarsete segamudelite puhul ollakse tihtipeale huvitatud pigem mingi vaatluste grupi mõjukusest, mitte üksikute vaatluste mõjukusest. Näiteks meditsiiniuuringutes soovitakse üldiselt üles leida erindlik patsient. Seega näidaku järgnevalt alaindeks (A), et vaatluse all olev suurus on leitud ilma vaatlusteta hulgast A . Tähistagu ψ vaadeldavate parameetrite kogumit, siis $\psi_{(A)}$ tähistab vaadeldavate parameetrite kogumit, mis on leidud valimilt, millest on välja jäetud hulk A .

3.1 Globaalne mõjukus

Lineaarsete segamudelite korral saab hinnata üldist mõjukust kõikidele parameetritele tõepära kauguse abil, mis näitab kui palju logaritmiline tõepära kogu andmestiku pealt muutub, kui see hinnata vähendatud andmestiku pealt leitud hinnangute abil. Vähendatud andmestiku all mõistame andmestikku, millest on välja jäetud vaatlus või vaatluste hulk, mille mõjukust hinnatakse. Tõepära kaugus näitab, et vaadeldud objektid on mõjukad, ent ei näita, mida nad mõjutavad. Need võivad olla mõjukad nii fikseeritud efektide hinnangutele, juhuslikele efektide prognoosidele kui ka nende täpsusele. Tähistagu θ kõigi parameetrid. Tähistagu $\hat{\theta}$ kõigi parameetrite hinnanguid, mis on leitud kogu andmestikult, ja $\hat{\theta}_{(A)}$ kõigi parameetrite hinnanguid, mis on leitud vähendatud andmestikult, st andmestikult, mis ei sisalda vaatlusi hulgast A . Logaritmilise tõepärafunktsiooni saame kirjutada kuju

$$l(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|V|) - \frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta). \quad (3.1)$$

Tõepära kaugus arvutatakse vastavalt

$$\begin{aligned} LD(A) &= 2\{l(\hat{\theta}) - l(\hat{\theta}_{(A)})\} = \\ &= \frac{1}{2} \ln \frac{|V_{(A)}|}{|V|} + \frac{1}{2} [(y - X\beta_{(A)})' V_{(A)}^{-1} (y - X\beta_{(A)}) \\ &\quad - (y - X\beta)' V^{-1} (y - X\beta)]. \end{aligned} \quad (3.2)$$

Kusjuures $l(\hat{\theta}_{(A)})$ arvutamisel on kasutatud hinnanguid, mis on leitud andmestikult, mis ei sisalda vaatlusi hulgast A , ent arvutatud üle kogu andmestiku, mis sisaldab ka vaatlusi hulgast A . Suur tõepära kauguse väärtus viitab võimalikule mõjukusele.

Arvutades tõepära kaugust võib tõepära asemel kasutada ka kitsendatud tõepära. Tähistame kitsendatud logaritmilise tõepära funktsiooni $l_R(\hat{\theta})$, siis kit-

sendatud tõepära kaugus on $RLD(A) = 2\{l_R(\hat{\theta}) - l_R(\hat{\theta}_{(A)})\}$.

See, et mingi vaatlus või vaatluste grupp on mingis mõttes mõjukas, ei tähenda, et need tuleks andmestikust eemaldama või tuleks mudelit muuta. Näiteks, kui mingi vaatluste hulk on mõjukas juhuslike efektide prognoosidele või nende täpsusele, ei pruugi need mõjutada fikseeritud efektide hinnanguid.

3.2 Mõju fikseeritud efektide hinnangule

3.2.1 Mõju ühe fikseeritud efekti hinnangule

Üks võimalus hinnata, kuidas mingi vaatluste grupp mõjutab ühe fikseeritud efekti hinnangut, on *DFBETAS*, mis mõõdab vahet kogu andmestikult leitud fikseeritud efektide hinnangute ja vähendatud andmestikult leitud fikseeritud efektide hinnangute vahel, mis on jagatud viimase standardhälbe hinnanguga. Tähistagu $\hat{\beta}_i$ hinnangut i . fikseeritud efektile. Siis *DFBETAS* on i . fikseeritud efekti jaoks defineeritud järgmiselt

$$DFBETAS_{i(A)} = \frac{\hat{\beta}_i - \hat{\beta}_{i(A)}}{\sqrt{\hat{D}(\hat{\beta}_{i(A)})}}, \quad (3.3)$$

kus $\hat{\beta}_i$ on esialgne i . fikseeritud efekti hinnang, $\hat{\beta}_{i(A)}$ on i . fikseeritud efekti hinnang, kui vaatluste grupp A on andmestikust välja jäetud, ja $\hat{D}(\hat{\beta}_{i(A)})$ on hinnangu $\hat{\beta}_{i(A)}$ dispersiooni hinnang. Piiriks, millest absoluutväärtuselt suuremad väärtused loetakse mõjukaks, on $\frac{2}{\sqrt{n'}}$, kus n' on vaadeldavate gruppide arv, mille mõjukust hinnata tahame.

Teine võimalus hinnata mõju ühe fikseeritud efekti hinnangule on protsendiline muutus, mis on defineeritud järgmiselt

$$PCHANGE_i = \frac{\hat{\beta}_i - \hat{\beta}_{i(A)}}{\hat{\beta}_i} \cdot 100\%, \quad (3.4)$$

kus β_i ja $\beta_{i(A)}$ on i . fikseeritud efekti hinnang leitud vastavalt kogu hinnangult ja vähendatud andmestikult. Antud näitaja korral ei ole antud piiri, millest suuremad viitaksid suurele mõjukusele, ent siiski suuremad väärtuses viitavad mõjukusele.

3.2.2 Klassi mõju mitmele efektile korraga

Erinevalt näitajast DFBETAS näitab Cooki kaugus, kuidas vaadeldav klass mõjutab kõiki fikseeritud efektide hinnanguid korraga. Tähistame $r = \text{rank}(X)$.

Cooki kaugus on antud valemiga:

$$CD_{(A)}(\hat{\beta}) = \frac{1}{r}(\hat{\beta} - \hat{\beta}_{(A)})'(\hat{D}(\hat{\beta}))^{-1}(\hat{\beta} - \hat{\beta}_{(A)}), \quad (3.5)$$

kus β on esialgne fikseeritud efektide hinnangute vektor, $\beta_{(A)}$ on fikseeritud efektide hinnangute vektor, kui vaadeldav hulga A on fikseeritud efektide hindamisel andmestikust välja jäetud. Kui vaadeldava hulga A korral Cooki kaugus on suurem kui $\frac{4}{n'}$, kus n' on vaadeldavate gruppide arv, siis see vaatluste grupp on mõjukas.

Lisaks Cooki kaugusele võib kasutada mõjukuse hindamiseks mitmemõõtme- list DFBETAS, mis on defineeritud sarnaselt Cooki kaugusele, ent varieeruvus on hinnatud andmestikult, kus on vaatluste grupp A välja jäetud, nagu DFBETAS korral. Seega

$$MDFBETAS_{(A)}(\hat{\beta}) = \frac{1}{r}(\hat{\beta} - \hat{\beta}_{(A)})'(\hat{D}(\hat{\beta}_{(A)}))^{-1}(\hat{\beta} - \hat{\beta}_{(A)}). \quad (3.6)$$

Näitaja MDFBETAS korral saab kasutada sama piiri nagu Cooki kauguse korral. Juhul, kui MDFBETAS arvutatakse ühe fikseeritud efekti jaoks, siis MDFBETAS on võrdne suuruse DFBETAS ruuduga.

3.2.3 Mõju fikseeritud efektide lineaarkombinatsioonile

Olgu $L : p \times s$ maatriks, mille astak on s . Defineerime $\xi = L'\beta$. Siis ξ annab meile s fikseeritud efektide lineaarkombinatsiooni. Kasutades suurima tõepära meetodit saame ξ hinnangu $\hat{\xi} = L'\hat{\beta}$. Olgu $\hat{\xi}_{(A)}$ hinnang, kui gruppi A kuuluvad vaatlused on välja jäetud. Siis analoogiliselt eelnevale saame suurima tõepära hinnanguks $\hat{\xi}_{(A)} = L'\hat{\beta}_{(A)}$.

Mõjukuse näitajana kasutame eelnevalt defineeritud Cooki kaugust või mitmemõõtmiselist DFBETAS

$$CD_{(A)}(\hat{\xi}) = \frac{1}{s}(\hat{\xi} - \hat{\xi}_{(A)})'[\hat{D}(\hat{\xi})]^{-1}(\hat{\xi} - \hat{\xi}_{(A)}), \quad (3.7)$$

$$MDFBETAS_{(A)}(\hat{\xi}) = \frac{1}{s}(\hat{\xi} - \hat{\xi}_{(A)})'[\hat{D}(\hat{\xi}_{(A)})]^{-1}(\hat{\xi} - \hat{\xi}_{(A)}). \quad (3.8)$$

Erijuhuna võime vaadelda olukorda, kui meil huvitab mõjukus t viimase fikseeritud efekti keskmisele. Sellise juhul $L' = [0_{t \times (p-t)} \mid I_t]$, kus $0_{t \times (p-t)}$ on nullidest koosnev $t \times (p-t)$ maatriks. Siis $\xi = (\beta_{p-t}, \dots, \beta_p)$.

Erijuhul, kui ollakse huvitatud mõjust viimase fikseeritud efekti hinnangule, siis $L' = [0_{1 \times (p-1)} \mid 1]$, kus $0_{1 \times (p-1)}$ on nullidest koosnev vektor.

Juhul, kui $L' = I_p$, siis $\xi = (\beta_1, \dots, \beta_p) = \beta$, mistõttu $CD_{(A)}(\hat{\xi})$ ja $MDFBETAS_{(A)}(\hat{\xi})$ on vastavalt näitajad 3.5 ja 3.6.

3.3 Mõju juhuslike efektide prognoosidele

3.3.1 Mõju kõikidele juhuslikele efektidele

Üheks võimaluseks hinnata vaatluste grupi mõjukust juhuslike efektide prognoosile on Cooki kaugus, mis on defineeritud järgmiselt

$$CD(\tilde{u}) = (\tilde{u}_{(A)} - \tilde{u})'[\hat{D}(\tilde{u})]^{-1}(\tilde{u}_{(A)} - \tilde{u}), \quad (3.9)$$

kus \tilde{u} on esialgne juhuslike efektide prognooside vektor, $\tilde{u}_{(A)}$ juhuslike efektide prognooside vektor leitud vähendatud andmestikult, st andmestikult, kus ei sisalda vaatlusi hulgast A , ja $\hat{D}(\tilde{u})$ on esialgse juhuslike efektide prognooside vektori dispersiooni hinnang. Antud näitaja korral ei ole antud piiri, millest suuremad väärtused viitaks mõjukusele, ent suured Cooki kauguse väärtuse viitavad siiski mõjukusele.

3.3.2 Erinevust vaadeldud väärtuse ja prognoositud marginaalse keskmise vahel

PRESS (*Predicted residual Sum of squares*) vigadeks nimetatakse erinevust vaadeldud väärtuse ja prognoositud marginaalse keskmise vahel, kus viimane on leitud ilma kõne all oleva vaatluseta $i \in A$, st

$$\hat{e}_{i(A)} = y_i - x_i' \hat{\beta}_{(A)}, \quad (3.10)$$

kus x_i on i . vaatlusele vastav plaanimaatriksi X reavektor. Kui leitakse PRESS viga vaadeldava grupi A jaoks, siis see on iga gruppi A kuuluva vaatluse PRESS vea ruutude summa, st

$$PRESS_{(A)} = \sum_{i \in A} \hat{e}_{i(A)}^2. \quad (3.11)$$

3.4 Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognooside täpsusele

Lisaks mõjule fikseeritud efektide hinnangutele ja juhuslike efektide prognoosidele vaadatakse ka mõju hinnangute täpsusele. Kui kõnealune vaatluste hulk ei ole mõjukas hinnangute või prognooside suhtes, siis võib see siiski mõjutada hüpoteesiteste tulemusi ja usaldusintervalle, kui see on mõjukas

hinnangu või prognooside täpsusele.

3.4.1 Mõju fikseeritud efektide hinnangute täpsusele

Hindamiseks mõju fikseeritud efektide β hinnangute täpsusele võib kasutada variatsioonisuhet, mis hindab muutust determinandis, kui vaatluste grupp (A) on andmestikust välja jäetud, st

$$VS(A, \beta) = \frac{\det(\hat{D}(\hat{\beta}_{(A)}))}{\det(\hat{D}(\hat{\beta}))}, \quad (3.12)$$

kus $\hat{\beta}$ on fikseeritud efektide hinnangute vektor kogu andmestikult, $\hat{\beta}_{(A)}$ fikseeritud efektide hinnangute vektor vähendatud andmestikult, mis ei sisalda hulka A kuuluvaid vaatlusi, $\hat{D}(\hat{\beta})$ ja $\hat{D}(\hat{\beta}_{(A)})$ on vastavate dispersioonimaatriksite hinnangud ning $\det(\hat{D}(\hat{\beta}))$ ja $\det(\hat{D}(\hat{\beta}_{(A)}))$ on vastavate dispersioonimaatriksite hinnangute determinandid.

Juhul, kui hulka A kuuluvad vaatlused ei mõjuta fikseeritud efektide hinnangute täpsust, siis vastav dispersioonimaatriks ei muutu, mistõttu ei muutu ka selle determinant. Seega, kui variatsioonisuhe on võrdne ühega, siis grupp A ei ole mõjukas.

Mugavam on kasutada variatsioonisuhte teisendatud kuju

$$VS_2(A, \beta) = | VS(A, \beta) - 1 |, \quad (3.13)$$

mis on võrdne nulliga, kui vaatluste grupp A ei ole mõjukas fikseeritud efektide hinnangute täpsusele. Mida suurem on ?? väärtus, seda mõjukam on hulka A kuuluvad vaatlused fikseeritud efektide hinnangute täpsusele.

Alternatiivselt variatsioonisuhtele võib kasutada ka jälje kaudu defineeritud

suurust

$$VS^{trace}(A, \beta) = | \operatorname{tr}((D(\hat{\beta}))^{-1}D(\hat{\beta}_{(A)})) - p |, \quad (3.14)$$

kus tr tähistab maatriksi jälge. Kui vaatlused hulgast A ei mõjuta fikseeritud efektide hinnangute täpsust, siis maatriksi $(D(\hat{\beta}))^{-1}D(\hat{\beta}_{(A)})$ peadiagonaalil on ühed. Sellisel juhul nende maatriksite jälg on võrdne fikseeritud efektide arvuga p . Seega vaatlused hulgast A ei mõjuta fikseeritud efektide hinnangute täpsust, kui $VR^{trace}(A, \beta)$ on võrdne nulliga, ja mida suuremad on $VR^{trace}(A, \beta)$ väärtus, seda mõjukamad on hulka A kuuluvad vaatlused fikseeritud efektide hinnangute täpsusele.

3.4.2 Mõju juhuslike efektide prognooside täpsusele

Analoogiliselt fikseeritud efektidele saame defineerida variatsioonisuhte hindamiseks mõjukust juhuslike efektide prognooside täpsusele

$$VS(A, u) = \frac{\det(D(\tilde{u}_{(A)}))}{\det(D(\tilde{u}))}, \quad (3.15)$$

kus \tilde{u} on kogu andmestikult leitud juhuslike efektide prognooside vektor ja $\tilde{u}_{(A)}$ on juhuslike efektide prognooside vektor, mis on leitud vähendatud andmestikult, kust on välja jäetud hulka A kuuluvad vaatlused, $D(\tilde{u})$ ja $D(\tilde{u}_{(A)})$ on vastavate prognooside vektorite dispersioonid ning $\det(D(\tilde{u}))$ ja $\det(D(\tilde{u}_{(A)}))$ on vastavate dispersioonimaatriksite determinandid.

Kui vaatlused hulgast A ei ole mõjukad juhuslike efektide prognooside täpsusele, siis variatsioonisuhe $VS(A, u)$ on võrdne ühega. Mugavam on vaadata variatsioonisuhte teisendust kujul

$$VS_2(A, u) = | VS(A, u) - 1 |. \quad (3.16)$$

Kui variatsioonisuhte teisendus 3.16 võrdub nulliga, siis hulka A kuuluvad vaatlused ei ole mõjukad juhuslike efektide prognooside täpsusele, ja mida suurem on variatsioonisuhte teisenduse 3.16 väärtus, seda mõjukam on hulka A kuuluvad vaatlused juhuslike efektide prognooside täpsusele.

Alternatiivselt variatsioonisuhtele võib kasutada ka matriksi jälje abil defineeritud suurust

$$VS^{trace}(A, u) = |tr((D(\tilde{u}))^{-1}D(\tilde{u}_{(A)})) - q|, \quad (3.17)$$

kus tr tähistab jälge. Kui hulka A kuuluvad vaatlused ei ole mõjukad juhuslike efektide prognooside täpsusele, siis matriksi $(D(\tilde{u}))^{-1}D(\tilde{u}_{(A)})$ peadiagonaalil on ühed ja seega nende matriksi jälg on võrdne juhuslike efektide arvuga q . Järelikult hulka A kuuluvad vaatlused ei ole mõjukad juhuslike efektide prognooside täpsusele, kui suurus 3.17 on võrdne nulliga. Mida suurem on 3.17 väärtus, seda mõjukamad on hulka A kuuluvad vaatlused.

4 Erindite leidmine peptiidide mikrokiibi andmetel

4.1 Peptiidide mikrokiipide tööpõhimõte

Peptiidide mikrokiipide tööpõhimõtte ja andmete kirjeldus põhineb artiklil Brus (2009) ja Nahtman (2007).

Valgud on geeniuuringutes olulisel kohal, need sisalduvad DNA-s ja neid saab näiteks kasutada katalüsaatorina. Valgud koosnevad mitmetest aminohapete ahelatest. Lühikesi aminohapete ahelaid nimetatakse peptiidideks. Üldiselt on valgud pikemad ja keerulisemad kui peptiidid ning võivad koosneda mitmest peptiidist.

Antikehad on olulisel kohal organismi immunsussüsteemis. Uuringutes kasutatakse neid erinevate viiruste ja bakterite ära tundmiseks. Antikehad seovad end antigeenidega ehk kindlat sorti peptiididega, mis sisalduvad viirustes ja bakterites. Kontrollimaks, kas proov sisaldab viirust või bakterit, lisatakse proovile antikehad, mis seostuvad antigeenidega. Seejärel puhastatakse proov nii, et alles jäävad vaid need antikehad ja antigeeni, mis on omavahel seostunud. Lõpuks lisatakse proovile fluorestseeruv aine, mis omakorda seostub antigeeni ja antikeha paaridega, üleliigne taaskord eemaldatakse. Saadust uuritakse spetsiaalse skänneri abil. Kui proov sisaldab palju antigeene, peaks töödeldud proov skänneri all helendama. Seevastu proov, mis ei sisalda antigeene ehk on viiruse- või bakterivaba, ei tohi skänneri all helendada.

Mikrokiibid töötavad samal põhimõttel, ent koosnevad tuhandetest süvikutest, milles igaüks võib sisaldada proovi. Seega saab mikrokiipide abil uurida väga suuri andmestikke. Iga mikrokiip koosneb alamkiipidest ja see omakorda plokkidest, mis sisaldavad endas süvikuid. Vastavaid süvikuid, milles asuvad peptiidid, nimetatakse täppideks. Kontrollimaks, kas mikrokiibid toi-

mivad korrektselt, jäetakse osad täpid tühjaks, seega need ei peaks andma mingit signaal, ja osad süvikud sisaldad proovi, mis peaksid andma suurele antigeenide kontsentratsioonile vastava tugeva signaali. Esimesi nimetatakse negatiivseteks kontrolltäppideks ja teisi positiivseteks kontrolltäppideks.

4.2 Andmed

Näites kasutame andmeid viielt patsiendilt, kellel on kliiniline ajalugu teatud sorti tuberkuloosiga. Patsientidele vastav tunnus *patient* on andmestikus identifitseeritu numbritega 1, 4, 5, 6 ja 11. Igale patsiendile anti vaktsiin ja võeti proov neljal erineval ajal. Esimene proov võeti vahetult enne vaktsiini andmist ja ülejäänud proovid võeti vastavalt 56, 112 ja 252 päeva möödudes vaktsiini saamisest. Proovi võtmise ajale vastav tunnus on *Day*, millel on neli võimalikku väärtust: 0, 56, 112 ja 252.

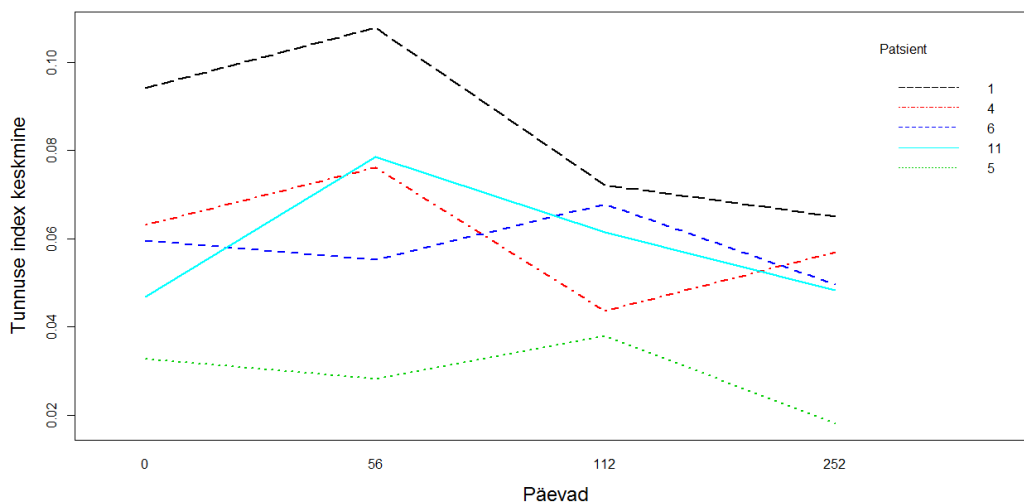
Igalt patsiendilt võetud proov asetati slaididele ja seejärel uuriti mikrokiipide abil. Antud juhul on saadud kaks mõõtmist: täpi helendamise intensiivsus ehk esiplaani mõõtmine $F635_{Median}$ ja täppi ümbritseva ala helendamise intensiivsus ehk tagaplaani mõõtmine $B635_{Median}$. Uuritavaks tunnuse on $indeks = \log_2 \frac{F635_{Median}}{B635_{Median}}$.

Proov on asetatud kolmele erinevale mikrokiibile. Sellele vastav tunnus on *S*, mille võimalikud väärtused on 1,2 ja 3.

Andmestik koosneb positiivsetest kontrollidest. Neid on neli erinevat tüüpi, mis kõik tähistavad erinevaid protsesse. Nende abi üritatakse leida süstemaatilisi vigu. Positiivsete kontrollide puhul peaks tunnuse *Index* väärtus olema suur.

Iga patsiendi korral on tehtud mitu mõõtmist ning kõigil on sama arv mõõtmisi igal päeval ja igal slaidil, siis kõigepealt keskmistame andmed nii, et iga päeva ja iga slaidi kohta jääb igal patsiendil üks mõõtmine. Sellisel juhul on

andmestikus 60 mõõtmist. Eesmärgiks on leida, kas mõni patsientidest on mõjukas mingis aspektis.



Joonis 4.1. Tunnuse *Index* keskmised iga patsiendi korral erinevatel päevadel

Kuna andmestik sisaldab positiivseid kontrolltäppe, siis kõik väärtused peaksid olema suured. Kui tulemused on nullilähedased, siis võib tegemist olla erilise vaatlusega. Joonisel 4.1 on toodud tunnuse *Index* keskmised iga patsiendi iga mõõtmispäeva korral. Näeme, et 5. patsiendi korral on tunnuse *Index* väärtuste keskmine igal päeval kõige väiksem, mistõttu 5. patsient võib osutada erindlikuks. Samas näeme, et 1. patsiendi mõõtmiste keskmised samuti erinevad teiste patsientide mõõtmistest, mistõttu võivad ka 1. patsiendi vaatlused osutada mõjukaks.

4.3 Peamõjudega mudel

Esimeses mudelis on võetud arvesse tunnused *patient* ja *day* juhuslikena. Defineeritud on see järgmiselt

$$y_{ijkl} = \mu + u_{patient(i)} + u_{day(j)} + \epsilon_{ijl}, \quad (4.1)$$

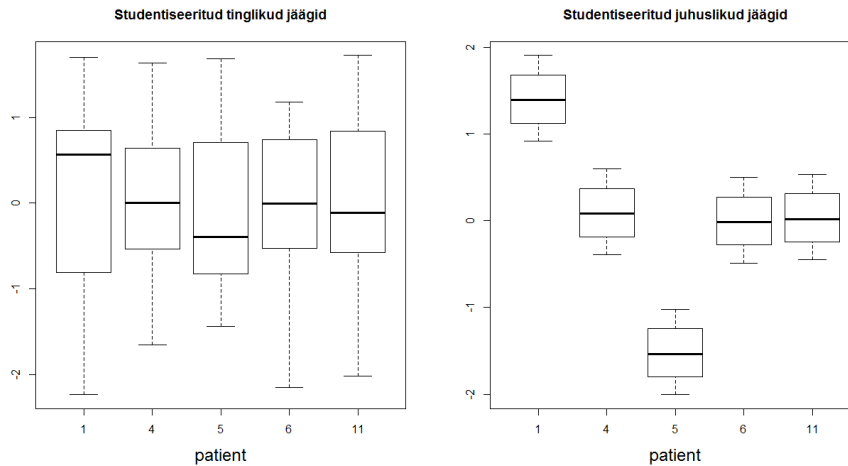
kus y_{ijkl} näitab uuritava tunnuse *index* väärtust i . patsiendi j . päeva korral, $\beta_{slide(k)}$ näitab slaidi mõju, $u_{patient(i)}$ patsiendi mõju, $u_{day(j)}$ päeva mõju ja ϵ_{ijkl} on juhuslik viga. Mudelist on välja jäetud tunnus s , mis on statistiliselt mitteoluline ($p=0.05447$). Kuna vaadeldavaid patsiente on viis ja mõõtmispäevi neli, siis prognoositavaid juhuslikke efekte on üheksa, millest viis vastavad patsientidele ja neli päevadele. Antud mudeli korral on tunnused *patient* ja *day* omavahel sõltumatud ning juhuslikele efektile vastav dispersioonimaatriks G on kahest plokkist koosnev plokk-diagonaalmaatriks, mille esimene plokk on $\sigma_p^2 I_5$ ja teine plokk on $\sigma_d^2 I_4$, kus σ_p^2 ja σ_d^2 on vastavalt tunnustele *patient* ja *day* vastavad varieeruvused. Mudeli vead on sõltumatud. Mudeli saime kujul

$$\begin{aligned} y_{ijkl} = & 0.05819 + 0.02549I_{patient(i=1)} + 0.00167I_{patient(i=4)} - 0.02772I_{patient(i=5)} \\ & - 0.00004I_{patient(i=6)} + 0.00059I_{patient(i=11)} + 0.000922I_{day(j=0)} \\ & + 0.000919I_{day(j=56)} - 0.00131I_{day(j=112)} - 0.00879I_{day(j=252)} + \epsilon_{ijl}. \end{aligned}$$

4.3.1 Jääkide analüüs

Erindite leidmiseks võib kasutada juhuslikke jääke ja tinglikke jääke. Tarkvara pakettis on antud üks käsk jääkide arvutamiseks, so *residuals(m1)*. Kusjuures tuleb tähele panna, et antud käsk arvutab tinglikud jäägid $\hat{\epsilon}_{ting} = y - X\hat{\beta} - Z\tilde{u}$. Erindeid saab tuvastada tinglike ja juhuslike jääkide abil. See-

ga juhuslikud jäägid tuleb arvutada vastavalt valemile $\hat{e}_{juh} = Z\tilde{u}$.

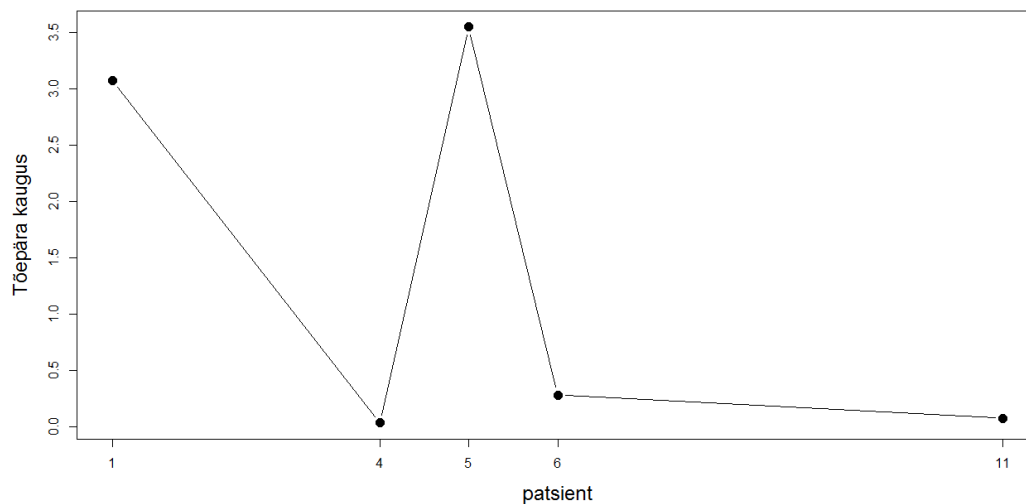


Joonis 4.2. Studentiseeritud tinglike ja juhuslike jääkide jaotus iga patsiendi korral

Joonisel 4.2 on toodud studentiseeritud tinglikud ja juhuslikud jäägid grupeeritud patsientide järgi. Näeme, et studentiseeritud tinglike jääkide põhjal ei ole ükski patsient erindlik, ent studentiseeritud juhuslike jääkide põhjal 1. ja 5. patsiendi korral on jäägid selgelt mittenuollilised.

4.3.2 Globaalne mõjukus

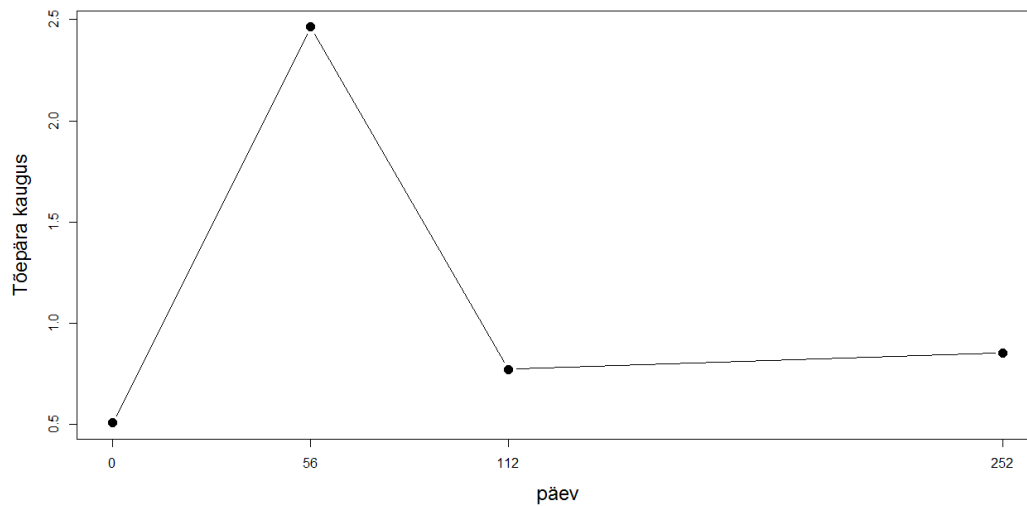
Globaalse mõjukuse hindamiseks saab kasutada tõepära kaugust. Tarkvarapakett R ei paku eraldi käsku tõepära kauguse arvutamiseks, seega tuleb selle leidmiseks kasutada eespool antud valemeid.



Joonis 4.3. Tõepära kaugus iga patsiendi korral

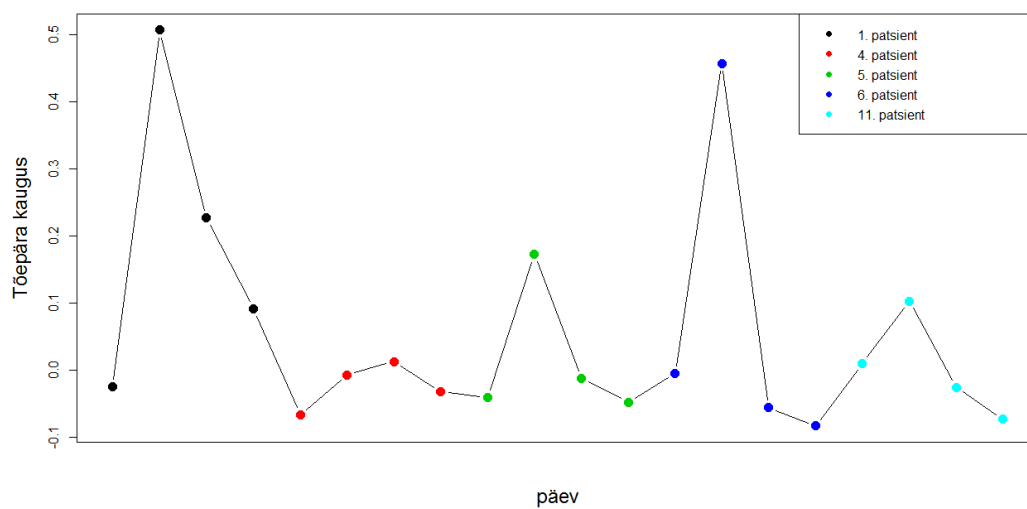
Joonisel 4.3 on toodud valemi 3.2 põhjal arvutatud tõepära kaugus, st on kasutatud suurima tõepära meetodit logaritmilise tõepära arvutamisel. Tõepära kaugus näitab üldist mõjukust. Suurema tõepära kauguse väärtusega patsiendid on mõjukad mingis mõttes. Jooniselt 4.3 näeme, et 1. ja 5. patsient võiksid olla mingis mõttes kõige mõjukamad. Ka jääkide põhjal saime, et need patsiendid on kõige erindlikumad.

Lisaks võime vaadata, kuidas erinevatele päevadele vastavad vaatlused mõjutavad logaritmilist tõepära.



Joonis 4.4. Tõepära kaugus iga päeva korral

Jooniselt 4.4 näeme, et kõige mõjukam teine mõõtmispäev, mis viitab sellele, et mõjukus võib olla tingitud teisest mõõtmispäevast.

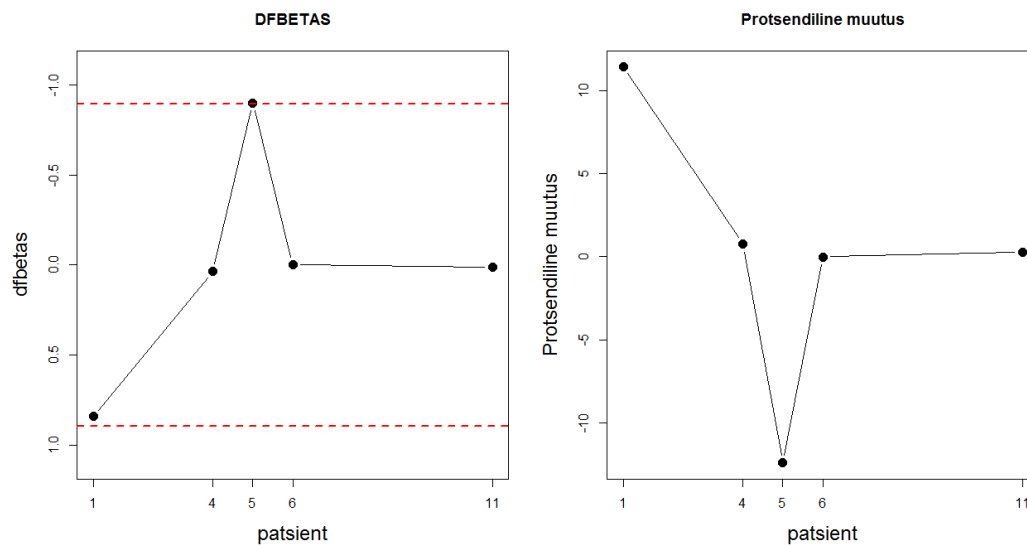


Joonis 4.5. Tõepära kaugus iga patsiendi iga päeva korral

Joonisel 4.5 on toodud tõepära kauguse väärtused, mis on arvutatud iga patsiendi iga päeva korral. Näeme, et nii 1. kui ka 5. patsiendi korral on teisele mõõtmispäevale vastav tõepära kauguse väärtus suurem kui ülejäänud päevade korral. Sama kehtib ka 6. patsiendi kuuenda patsiendi korral, kuigi joonise 4.3 põhjal 6. patsient mõjukas ei näinud olevat.

4.3.3 Mõju fikseeritud efektide hinnangule

Antud andmete korral on meil vaid üks fikseeritud efekt, milleks on vabaliige. Seega leiame antud korral näitaja DFBETAS väärtused ja fikseeritud efektide hinnangute protsendilise muutuse iga patsiendi suhtes. Kuna ühe fikseeritud efekti korral on näitaja DFBETAS väärtus võrdne näitaja MDFBETAS väärtuse ruutjuurega, siis piirdume hetkel vaid näitaja DFBETAS ja protsendilise muutusega, st Cooki kauguse ja näitaja MDFBETAS väärtuseid praegusel juhul ei arvuta.



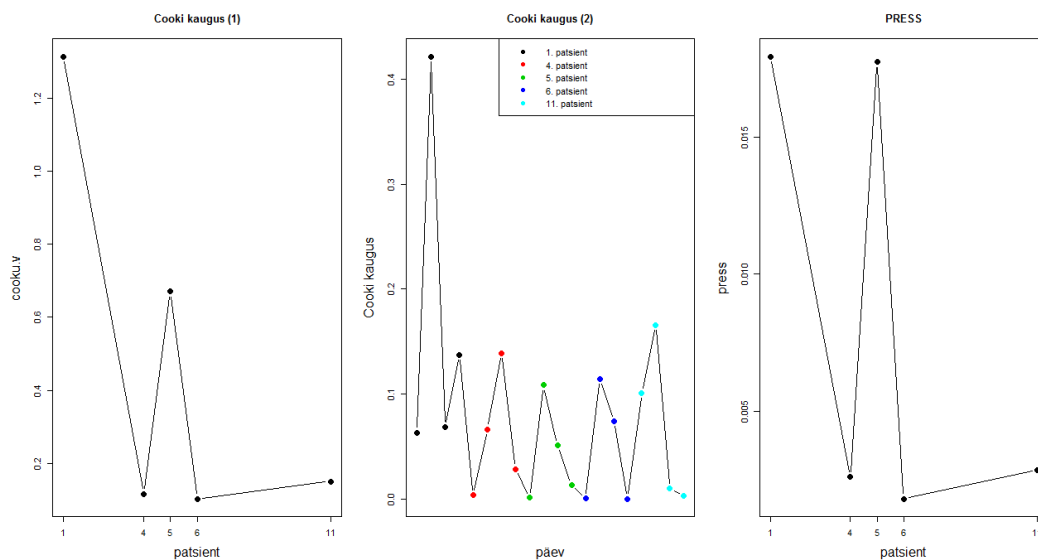
Joonis 4.6. Mõjukusnäitaja DFBETAS ja hinnangu protsendiline muutus, näitavad vastavalt mõju hinnangule ja hinnangu protsendilist muutust

Joonisel 4.6 on toodud antud andmestiku korral leitud DFBETAS, mis näitab, kuidas iga patsiendi vaatlused mõjutavad fikseeritud efektide hinnanguid, ja fikseeritud efekti hinnangu protsendilist muutust. Näitaja DFBETAS korral on toodud piir, millest absoluutväärtuselt suuremad olevad grupid on mõjukad. Selleks on $\frac{2}{\sqrt{n'}}$, kus n' on vaadeldavate gruppide arv (Nieuwenhuis, 2012). Kuna antud juhul on grupeeritud patsientide järgi, siis $n' = 5$, seega piiriks saame $\frac{2}{\sqrt{5}} \approx 0.894$. Antud piir on joonisel märgitud punase katkendjoonega. Näeme, et 5. patsient mõjutab vabaliikme hinnagut. Protsendilise muutuse korral ei ole toodud mõjukuse piiri, ent näeme, et kõige rohkem mõjutavad vabaliikme hinnagut 1. ja 5. patsiendi vaatlused.

4.3.4 Mõju juhuslikele efektide prognoosidele

Mõju juhuslike efektide prognoosidele saab hinnata Cooki kauguse ja näitaja PRESS abil. Eesmärgiks oli leida erindlik patsient. Hindamaks mõju prognoosile Cooki kauguse abil tuleks jätta välja vaadeldava patsiendi vaatlused. Sellisel juhul jääb juhuslike efekte ühe võrra väiksemaks, mis tõttu hindasime Cooki kaugust kahel viisil:

- (1) võrdlesime omavahel hinnanguid, mis jäid alles ka vaadeldava patsiendi vaatluste välja jätmiselt,
- (2) leidsime iga patsiendi iga mõõtmispäeva jaoks Cooki kauguse väärtuse, sellisel juhul jäi juhuslike efektide arv samaks.

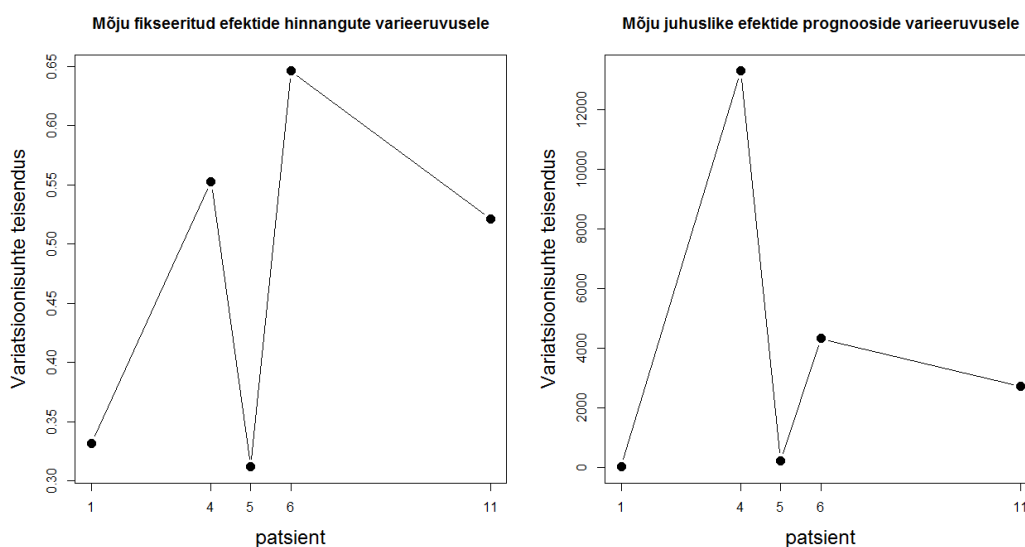


Joonis 4.7. Cooki kaugus kahel viisi ja näitaja PRESS väärtused, näitavad mõju juhuslike efektide prognoosidele

Joonisel 4.7 on toodud iga patsiendi jaoks kaht sorti Cooki kauguse väärtused ja mõjukusnäitaja PRESS väärtus iga patsiendi jaoks. Suurema Cooki kauguse või PRESS väärtusega patsiendid on suurema mõjuga juhuslike efektide prognoosidele. Cooki kauguse põhjal näeme, et juhuslike efektide prognoosidele on kõige mõjukam 1. patsient, kusjuures mõjukus näib tingitud olevat teisest mõõtmiskorrast. Näitaja PRESS põhjal on 1. ja 5. patsient juhuslike efektide prognooside suhtes sama mõjukad.

4.3.5 Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele

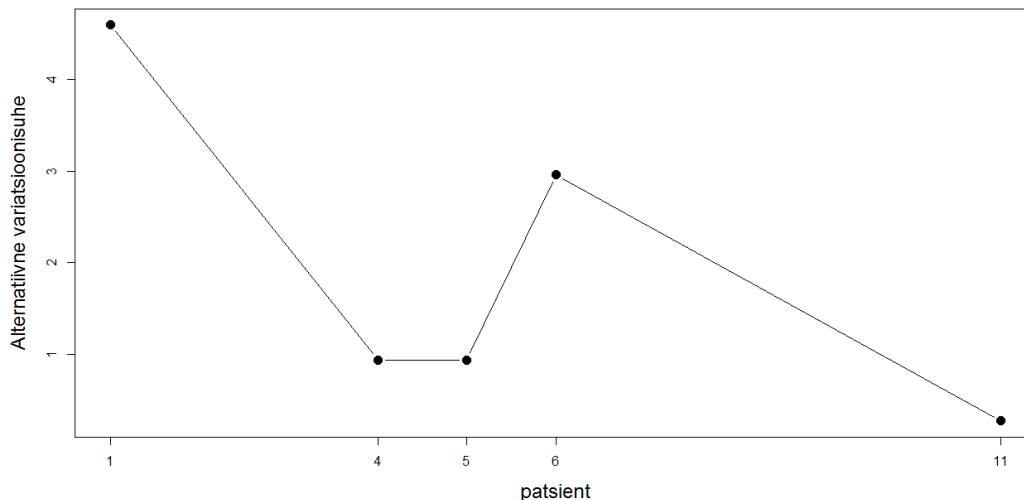
Mõnikord võib juhtuda, et vaatluste grupp ei ole mõjukas fikseeritud efektide hinnangutele ega juhuslike efektide prognoosidele, kui on mõjukas nende täpsusele, mistõttu võib mudel anda kehvad hinnangud.



Joonis 4.8. Variatsioonisuhte teisenduse väärtused iga patsiendi korral

Joonisel 4.8 on toodud nii fikseeritud efektide kui ka juhuslike efektide jaoks leitud variatsioonisuhte teisenduse väärtused, mis on arvatud vastavalt valemitele 3.13 ja 3.16. Antud näitajate korral viitab suur väärtus mõjukusele. Tõepära kauguse põhjal olid globaalselt kõige mõjukamad 1. ja 5. patsient. Jooniselt 4.8 näeme, et vaadeldavad patsiendid ei ole mõjukad fikseeritud efektide hinnangu ega juhuslike efektide prognoosi täpsuse suhtes.

Kuna ainuke fikseeritud efekt on vabaliikmele vastav, siis antud olukorras variatsioonisuhte teisendus 3.13 ja alternatiivne variatsioonisuhe 3.14 langevad kokku.



Joonis 4.9. Alternatiivse variatsioonisuhte väärtused iga patsiendi korral hindamaks mõju juhuslike efektide prognooside täpsust

Joonisel 4.9 on toodud alternatiivse variatsioonisuhte 3.16 väärtused iga patsiendi korral. Suuremad väärtused viitavad suuremale mõjukusele juhuslike efektide prognooside täpsusele. Näeme, et antud juhul on kõige mõjukamad 1. ja 5. patsiendi vaatlused, mis olid ka mõjusad tõepära kauguse põhjal.

4.4 Hierarhiline mudel

Teises mudelis on võetud arvesse kõik tunnused: s , $patient$ ja day . Defineeritud on see järgmiselt

$$y_{ijkl} = \beta_{slide(k)} + u_{patient(i)} + u_{patient(i):day(j)} + \epsilon_{kijl}, \quad (4.2)$$

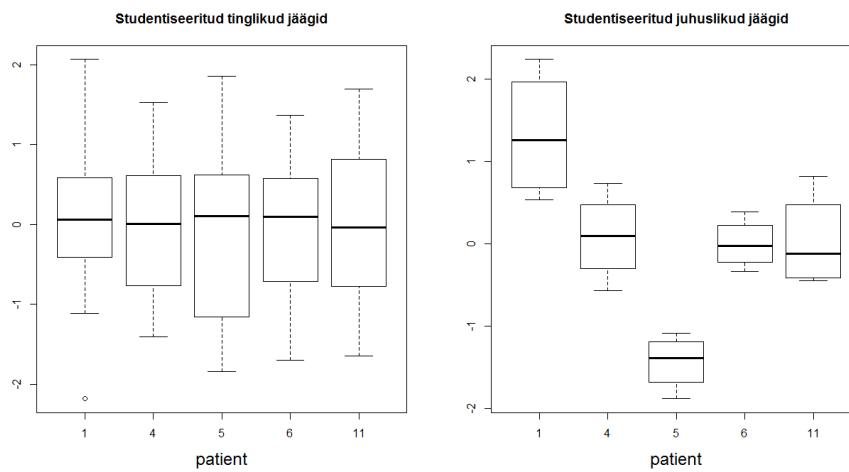
kus y_{ijkl} näitab uuritava tunnuse $index$ väärtust i . patsiendi j . päeva korral, $\beta_{slide(k)}$ näitab k . slaidi mõju, $u_{patient(i)}$ patsiendi mõju, $u_{patient(i),day(j)}$ i . patsiendi päeva mõju ja ϵ_{kijl} on juhuslik viga. Juhuslikke efekte on 25, millest

viis patsientidele vastavat ja neli päevale vastavat efekti iga patsiendi korral. Antud mudeli korral on taaskord tunnused *patient* ja *day* omavahel sõltumatud ning juhuslikele efektidele vastav disperisoonimaatriks on kahest plokkist koosnev plokk-diagonaalmaatriks, mille esimene plokk on $\sigma_d^2 I_2 0$ ja teine plokk on $\sigma_p^2 I_5$, kus σ_d^2 ja σ_p^2 on vastavalt tunnuste *day* ja *patient* varieeruvused. Antud mudeli korral on vead sõltumatud.

Antud juhul saame mudeli

$$\begin{aligned}
y_{ijkl} = & 0.06147348I_{slide(k=1)} + 0.06094877I_{slide(k=2)} + 0.05216196I_{slide(k=3)} \\
& + 0.023480I_{patient(i=1)} + 0.009758I_{patient(i=1),day(j=0)} + 0.020456I_{patient(i=1),day(j=56)} \\
& - 0.007416I_{patient(i=1),day(j=112)} - 0.012942I_{patient(i=1),day(j=252)} \\
& + 0.001542I_{patient(i=4)} + 0.002657I_{patient(i=4),day(j=0)} + 0.012860I_{patient(i=4),day(j=56)} \\
& - 0.012652I_{patient(i=4),day(j=112)} - 0.002218I_{patient(i=4),day(j=252)} \\
& - 0.025531I_{patient(i=5)} + 0.000009I_{patient(i=5),day(j=0)} - 0.003413I_{patient(i=5),day(j=56)} \\
& + 0.004151I_{patient(i=5),day(j=112)} - 0.011465I_{patient(i=5),day(j=252)} \\
& - 0.000032I_{patient(i=6)} + 0.001203I_{patient(i=6),day(j=0)} - 0.002198I_{patient(i=6),day(j=56)} \\
& + 0.007588I_{patient(i=6),day(j=112)} - 0.006607I_{patient(i=6),day(j=252)} \\
& + 0.000542I_{patient(i=11)} - 0.009307I_{patient(i=11),day(j=0)} + 0.015487I_{patient(i=11),day(j=56)} \\
& + 0.002155I_{patient(i=11),day(j=112)} - 0.008107I_{patient(i=11),day(j=252)} + \epsilon_{ijkl}.
\end{aligned}$$

4.4.1 Jääkide analüüs

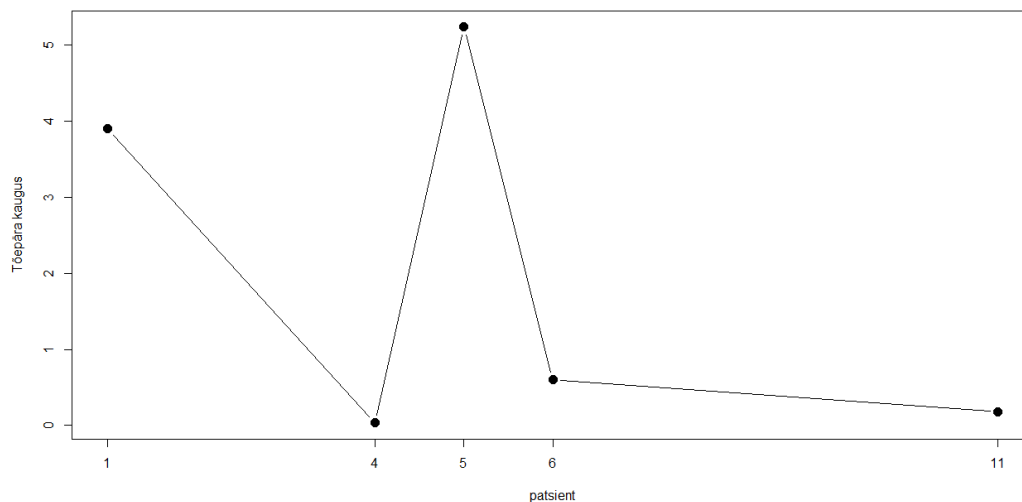


Joonis 4.10. Studentiseeritud tinglike ja juhuslike jääkide jaotus

Joonisel 4.10 on toodud mudeli 4.2 studentiseeritud tinglikud ja juhuslikud jäägid grupeeritud patsientide järgi. Nagu mudeli 4.1 korral on 1. ja 5. patsiendi jäägid selgelt mittenuküllised. Seega on need patsiendid jääkide mõttes erindlikud.

4.4.2 Globaalne mõjusus

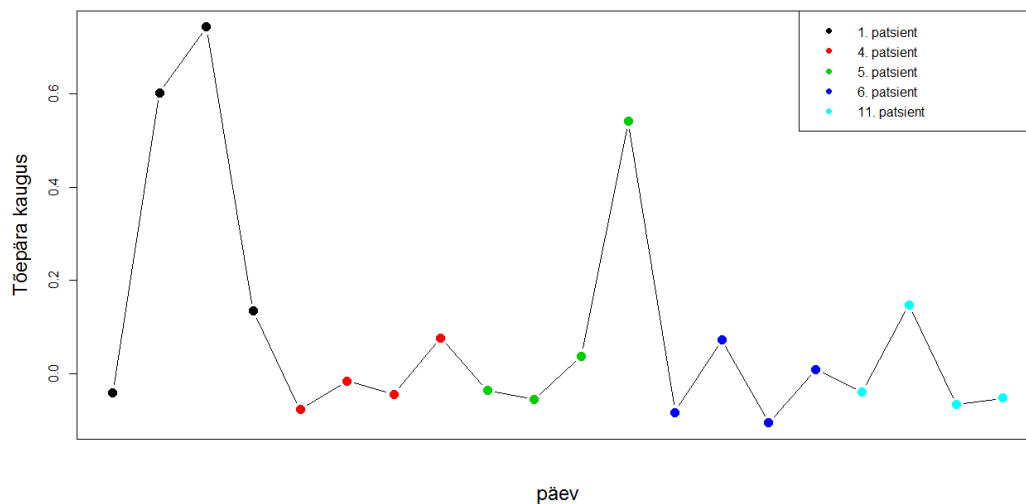
Globaalse mõjukuse hindamiseks saab taaskord kasutada tõepära kaugust.



Joonis 4.11. Tõepära kaugus iga patsiendi korral

Joonisel 4.11 on toodud valemi 3.2 põhjal arvutatud mudeli 4.2 tõepära kaugus. Näeme, et 1. ja 5. patsient võiksid olla mingis mõttes mõjukad nagu ka mudeli 4.1 korral nägime, kusjuures 5. patsient olevat mõjukam. Ka jääkide põhjal saime, et need patsiendid on erindlikud.

Vaatamaks täpsemalt, millistest mõõtmispäevadest on tingitud 1. ja 5. patsiendi mõjusus, saab leida tõepära kauguse väärtused iga patsiendi iga mõõtmispäeva jaoks.

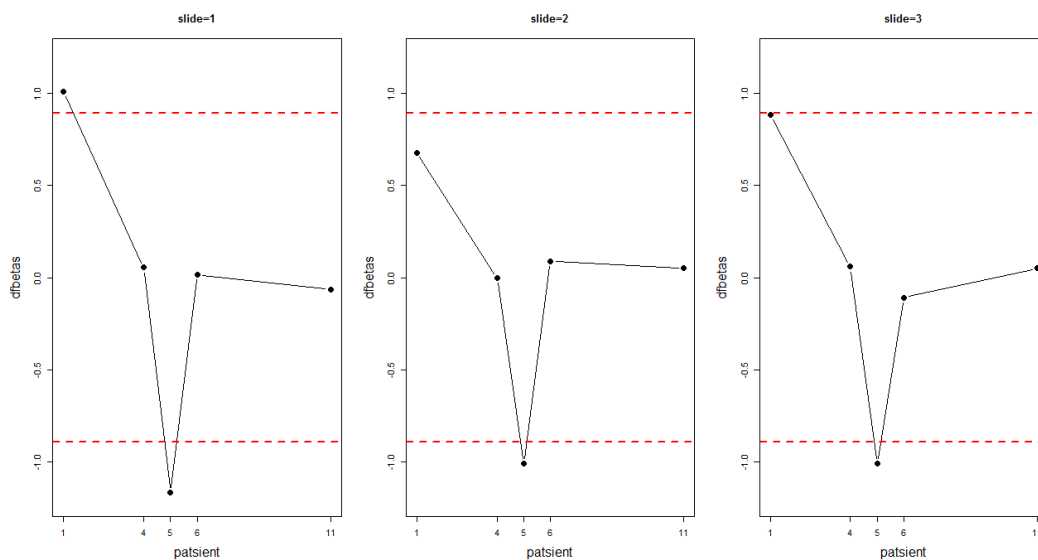


Joonis 4.12. Tõepära kaugus iga patsiendi iga päeva korral

Joonisel 4.12 on toodud tõepära kauguse väärtused iga patsiendi iga päeva korral. Näeme, et 1. patsiendi korral mõjukad teine ja kolmas mõõtmispäev, sama 5. patsiendi korral ei ole teine mõõtmispäev mõjukas nagu oli mudeli 4.1, vaid mõjukas näib olevat viimane mõõtmispäev.

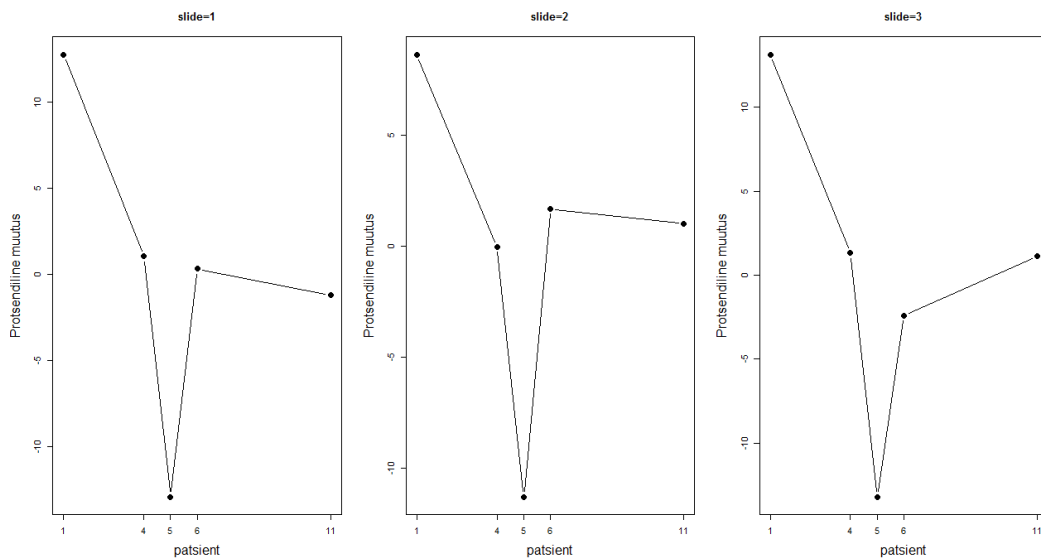
4.4.3 Mõju fikseeritud efektide hinnangule

Kuna antud juhul on fikseeritud faktoreid kolm, siis saame leida mõjukusnäitajad nii iga fikseeritud efekti hinnangu jaoks eraldi kui ka kõigi jaoks korraga.



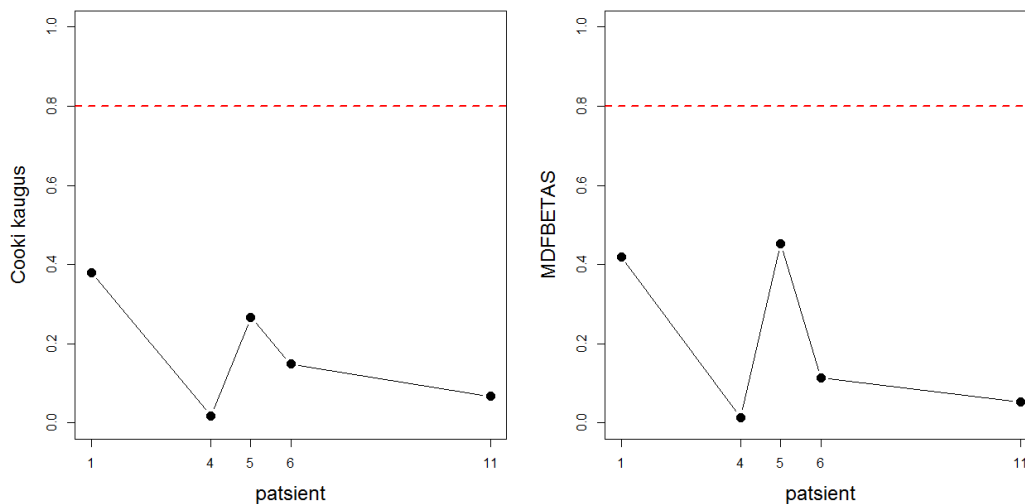
Joonis 4.13. Mõjukusnäitaja DFBETAS iga patsiendi korral, näitab mõjukust iga fikseeritud efekti hinnangu jaoks

Joonisel 4.13 on toodud mudeli 4.2 korral leitud näitaja DFBETAS väärtused. Antud näitaja korral on toodud piir sama mis mudeli 4.1 korral. Seega näeme jooniselt 4.13, et 1. patsiendi vaatlused mõjutavad tunnuse *slide* esimesele tasemele vastava fikseeritud efekti hinnangut ja 5. patsiendi vaatlused mõjutavad tunnuse *slide* igale tasemele vastava fikseeritud efekti hinnangut.



Joonis 4.14. Fikseeritud efektide hinnangute protsendiline muutus iga patsiendi korral

Joonisel 4.14 on näidatud, kui palju fikseeritud efektide hinnangud mudeli 4.2 korral muutuvad, kui välja jätta vastava patsiendi vaatlused. Antud mõjukuse näitaja korral ei ole antud piiri, millest suurema väärtusega patsiendid loetakse mõjukaks, ent sarnaselt näitajale DFBETAS on iga fikseeritud efekti hinnangule kõige mõjukamad 1. ja 5. patsiendi vaatlused.



Joonis 4.15. Cooki kaugus ja näitaja MDFBETAS väärtus iga patsiendi korral

Joonisel 4.15 on toodud leitud Cooki kaugused ja näitaja MDFBETAS väärtused, mis näitavad üldist mõju kõigele fikseeritud faktoritele korraga. Artiklis Nieuwenhuis (2012) on toodud Cooki kauguse korral suure mõjukuse piiriks $\frac{4}{\sqrt{n'}}$, kus n' on vaadeldavate gruppide arv. Saame piiriks $\frac{4}{5} = 0.8$. Näitaja MDFBETAS korral kasutame sama piiri. Jooniselt 4.15 näeme, et vaatamata sellele, et 1. patsiendi vaatlused olid mõjukad tunnuse *slide* esimesele tasemele vastavale hinnangule ja 5. patsiendi vaatlused olid mõjukad tunnuse *slide* igale tasemele vastavale hinnangule, ei ole 1. ega 5. patsiendi vaatlused mõjukad Cooki kauguse ega näitaja MDFBETAS järgi kõigile fikseeritud efektide hinnangutele korraga.

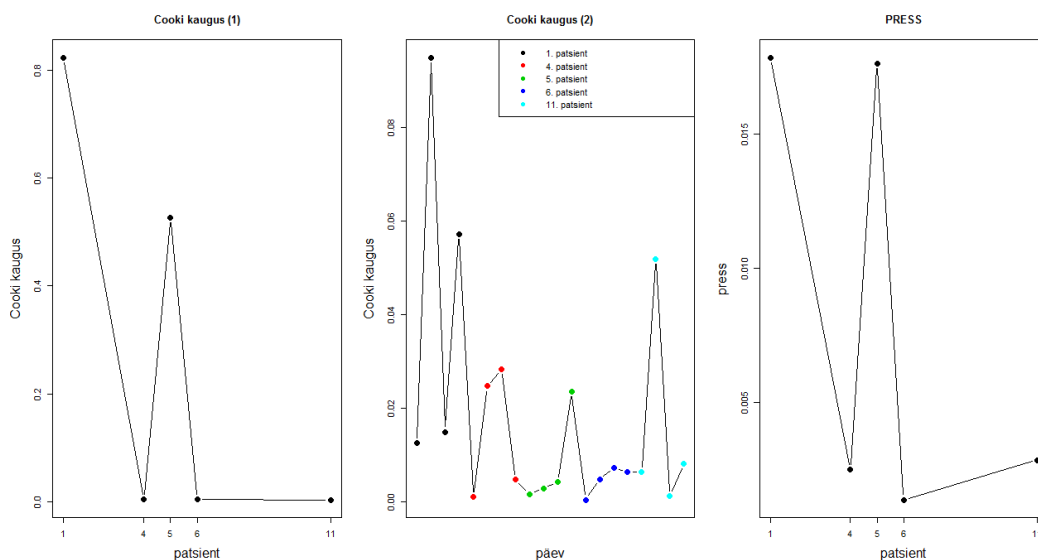
4.4.4 Mõju juhuslikele efektide prognoosidele

Nagu ka mudeli 4.1 korral jättes välja mõningad vaatlused võib juhuslike efektide arv muutuda. Cooki kauguse leidmisel võrdleme vaid juhuslike efekte, mis jäävad alles ka vaatluste välja jätmisel. Antud mudeli korral leiame Cooki

kaugused kahel viisil:

- (1) iga patsiendi jaoks üks Cooki kauguse väärtus, kusjuures juhuslike efektide arv on sellisel juhul viie võrra väiksem.
- (2) iga patsiendi iga päeva jaoks üks Cooki kauguse väärtus, kusjuures juhuslike efektide arv on sellisel juhul ühe võrra väiksem.

Lisaks Cooki kaugusele leiame mõjukusnäitaja PRESS väärtuse iga patsiendi korral.



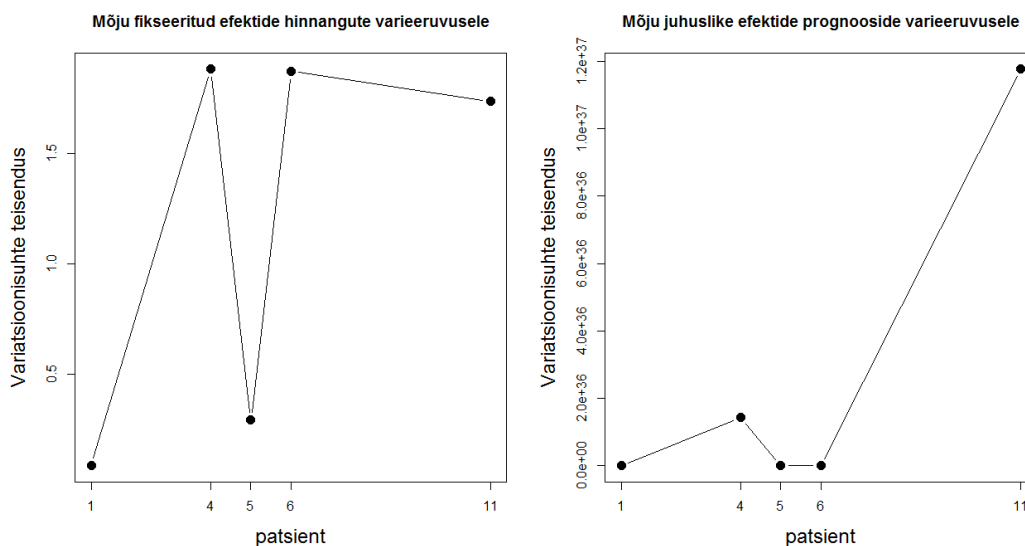
Joonis 4.16. PRESS

Joonisel 4.16 on toodud iga patsiendi jaoks toodud mõjukusnäitaja PRESS väärtus, mis näitavad erinevust vaadeldud väärtuse ja prognoositud marginaalse keskmise vahel, ja kahel viisil arvutatud Cooki kaugused. Nagu ka mudeli 4.1 korral on Cooki kauguse põhjal kõige mõjukam juhuslike efektide prognoosidele 1. patsiendi vaatlused ja mõjukus näib enim tingitud olla teise

mõõtmispäeva poolt. Näitaja PRESS põhjal on 1. ja 2. patsiendi vaatlused sama mõjukad juhuslike efektide prognoosidele.

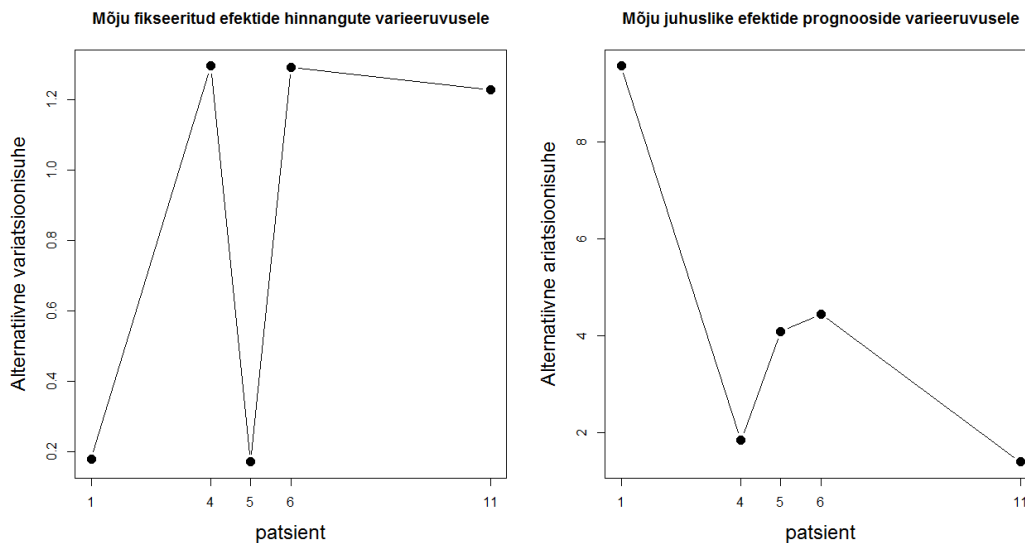
4.4.5 Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele

Vaadeldava mudeli korral saame leida lisaks variatsioonisuhte teisenduste 3.13 ja 3.16 väärtustele ka alternatiivse variatsioonisuhte väärtused, mis on antud võrdustega 3.14 ja 3.17.



Joonis 4.17. Variatsioonisuhte teisenduse väärtused iga patsiendi korral

Joonisel 4.17 on fikseeritud efektide ja juhuslike efektide jaoks toodud vastavalt variatsioonisuhte teisenduse 3.13 ja 3.16 väärtused. Suuremad väärtused viitavad suuremale mõjukusele. Jooniselt 4.17 näeme, et tõepära kauguse põhjal kõige mõjukamad olevad 1. ja 5. patsient ei ole antud näitaja mõjukad fikseeritud efektide hinnangute ega juhuslike efektide prognoosi täpsusele.



Joonis 4.18. Alternatiivse variatsioonisuhte väärtused iga patsiendi korral

Joonisel 4.18 on toodud alternatiivsete variatsioonisuhtete 3.14 ja 3.17 väärtused iga patsiendi korral. Suured alternatiivse variatsioonisuhte väärtused viitavad suurele mõjukusele. Jooniselt 4.18 näeme, et kuigi 1. ja 5. patsiendi vaatlused olid mõjukad tõepära kauguse põhjal, ei ole need mõjukad fikseeritud efektide hinnangute täpsuse suhtes. Samas juhuslike efektide prognooside täpsuse suhtes on kõige mõjukamad 1. patsiendi vaatlused nagu nägime ka mudeli 4.1 korral.

4.5 Hierarhiline mudel sõltuvate juhuslike efektidega

Kolmandas mudelid on samuti võetud arvesse kõik tunnused: s , $patient$ ja day . Defineeritud on see järgmiselt

$$y_{ijkl} = \beta_{slide(k)} + u_{patient(i)} + u_{patient(i):day(j)} + \epsilon_{ijkl}, \quad (4.3)$$

kus y_{ijkl} näitab uuritava tunnuse $index$ väärtust i . patsiendi j . päeva korral, $\beta_{slide(k)}$ näitab k . slaidi mõju, $u_{patient(i)}$ patsiendi mõju, $u_{patient(i),day(j)}$ i .

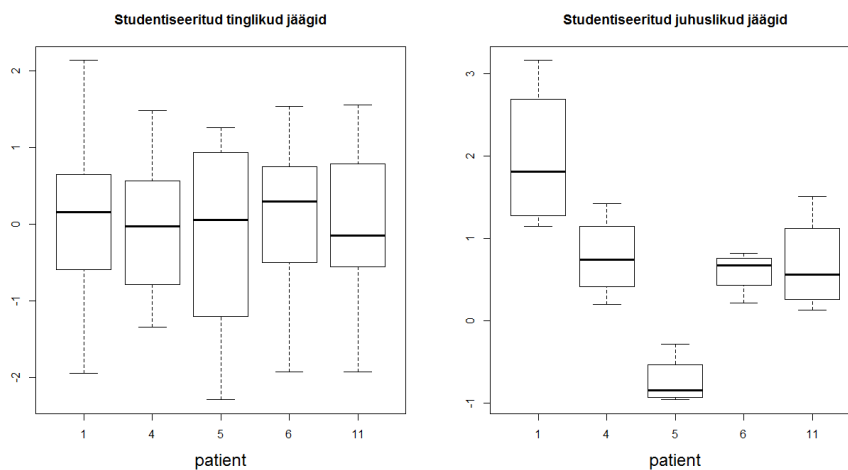
patsiendi j . päeva mõju ja ϵ_{kijl} on juhuslik viga. Juhuslikke efekte on 25, millest viis patsientidele vastavat ja neli päevale vastavat efekti iga patsiendi korral. Vaadeldava mudeli korral on erinevatel päevadel tehtud mõõtmised omavahel korreleeritud, kusjuures iga patsiendi korral on antud korrelatsioonid samad. Seega juhuslikele efektidele vastav dispersioonimaatriks on viiest plokist koosnev plokk-diagonaalmaatriks, mille iga plokk on mõõtmispäevadele vastav dispersioonimaatriks. Vaadeldava mudeli korral on vead taaskord sõltumatud.

Saame mudeli

$$\begin{aligned}
y_{ijkl} = & 0.04835373I_{slide(k=2)} + 0.04782902I_{slide(k=2)} + 0.03904220I_{slide(k=3)} \\
& + 0.043856289I_{patient(i=1),day(j=0)} + 0.06242129I_{patient(i=1),day(j=56)} \\
& + 0.027737701I_{patient(i=1),day(j=112)} + 0.022668428I_{patient(i=1),day(j=252)} \\
& + 0.017164124I_{patient(i=4),day(j=0)} + 0.02808769I_{patient(i=4),day(j=56)} \\
& + 0.003959212I_{patient(i=4),day(j=112)} + 0.012334604I_{patient(i=4),day(j=252)} \\
& - 0.015621182I_{patient(i=5),day(j=0)} - 0.01763876I_{patient(i=5),day(j=56)} \\
& - 0.005491156I_{patient(i=5),day(j=112)} - 0.018904002I_{patient(i=5),day(j=252)} \\
& + 0.012879418I_{patient(i=6),day(j=0)} + 0.01375348I_{patient(i=6),day(j=56)} \\
& + 0.016162201I_{patient(i=6),day(j=112)} + 0.004339895I_{patient(i=6),day(j=252)} \\
& + 0.007663937I_{patient(i=11),day(j=0)} + 0.02984350I_{patient(i=11),day(j=56)} \\
& + 0.014544987I_{patient(i=11),day(j=112)} + 0.002633373I_{patient(i=11),day(j=252)} + \epsilon_{kijl}
\end{aligned}$$

Mudelist näeme, et 5. patsiendi korral on päeva mõju alati negatiivne, teiste patsientide korral on päeva mõju alati vastupidiselt positiivne. See näitab juba näitab, et 5. patsiendi mõõtmised erinevad teiste patsientide omadest.

4.5.1 Jääkide analüüs

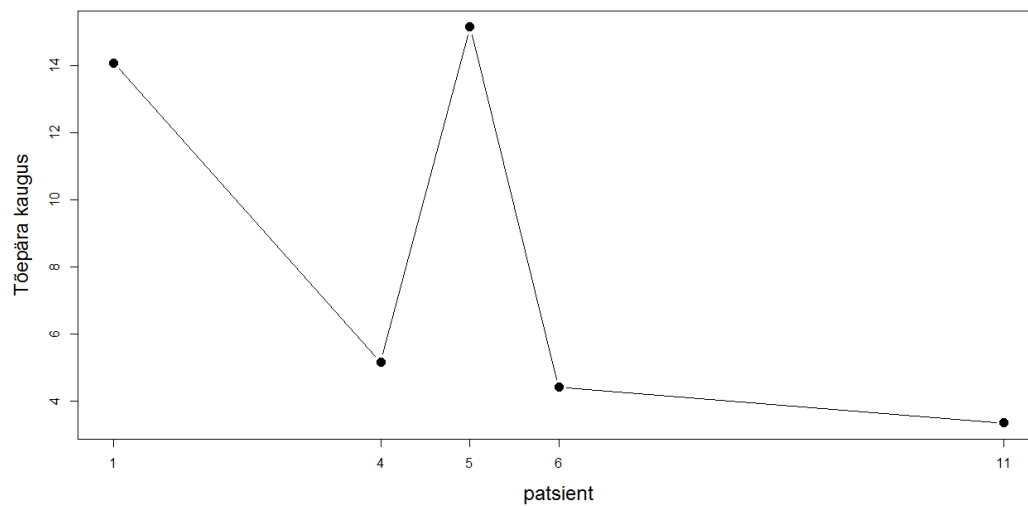


Joonis 4.19. Studentiseeritud tinglike ja juhuslike jääkide jaotus

Joonisel 4.19 on toodud studentiseeritud tinglikud ja juhuslikud jäägid grupeeritud patsientide järgi. Nagu ka mudelite 4.1 ja 4.2 näeme, et studentiseeritud tinglike jääkide põhjal ei ole ükski patsient erindlik, ent studentiseeritud juhuslike jääkide põhjal 1. patsiendi korral on jäägid selgelt mittenullilised. Seega võivad 1. patsiendi vaatlused osutada mõjukaks.

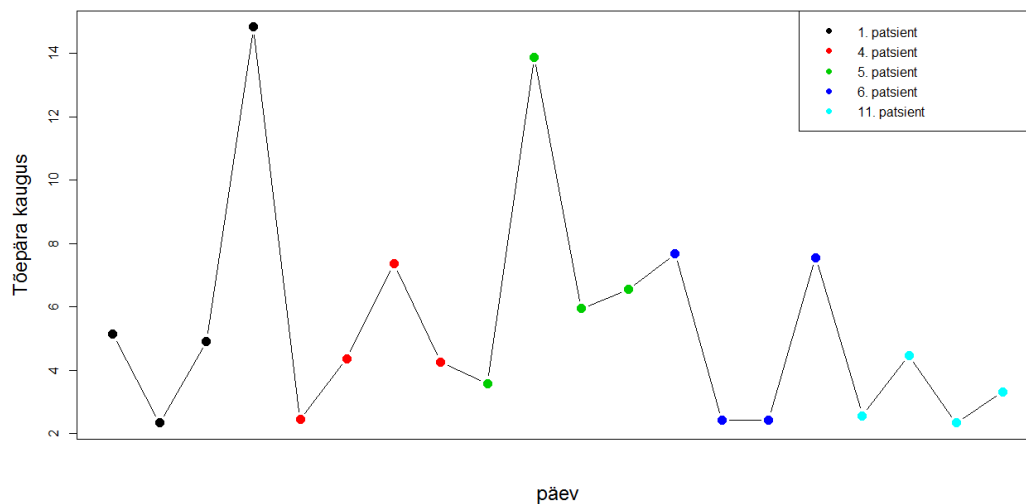
4.5.2 Globaalne mõjusus

Globaalset mõjuskust saab hinnata tõepära kauguse abil.



Joonis 4.20. Tõepära kaugus iga patsiendi korral

Joonisel 4.20 on toodud valemi 3.2 põhjal arvutatud mudeli 4.3 tõepära kauguse väärtused iga patsiendi korral. Näeme, et ka antud mudeli korral on kõige mõjukamad 1. ja 5. patsiendi vaatlused.

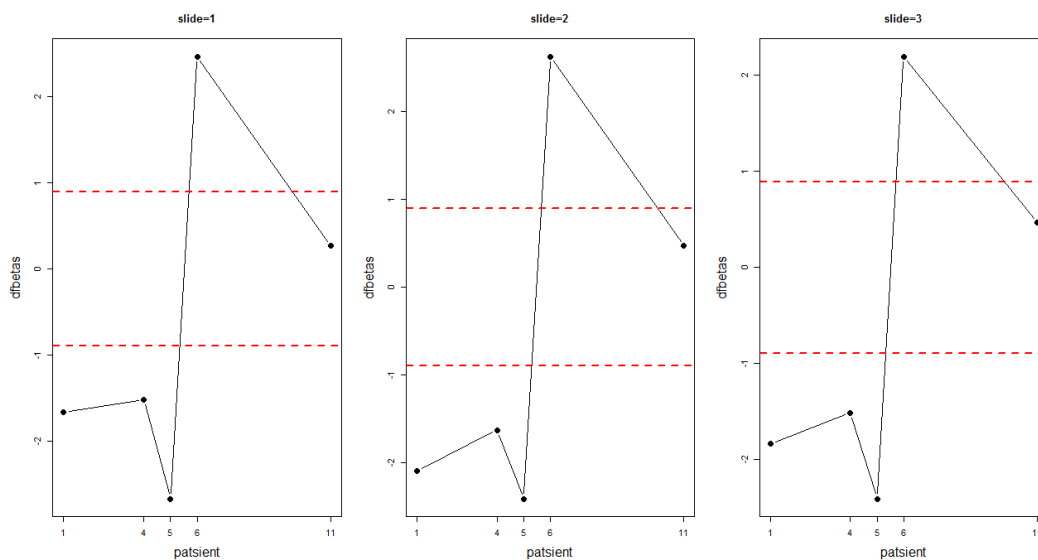


Joonis 4.21. Tõepära kaugus iga patsiendi iga päeva korral korral

Joonisel 4.21 on toodud tõepära kauguse väärtused iga patsiendi iga päeva jaoks. Vaadeldava mudeli korral saame, et 1. patsiendi mõjukus on tingitud suuresti viimase mõõtmispäev vaatluse poolt ja 5. patsiendi mõjukus teise mõõtmispäeva poolt.

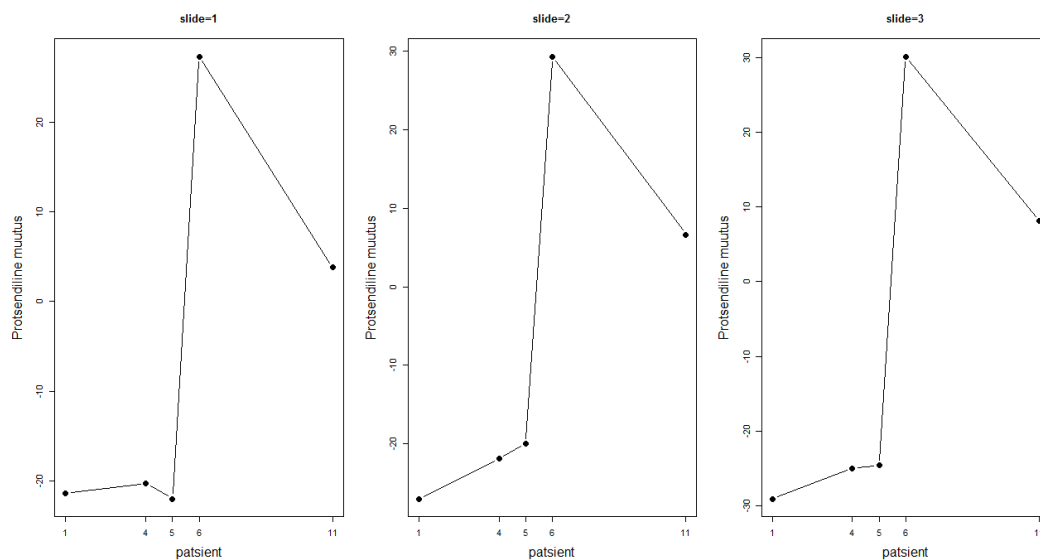
4.5.3 Mõju fikseeritud efektide hinnangule

Mudeli 4.3 korral on kolm fikseeritud efekti, mistõttu arvutame mõjukusnäitajad nii iga fikseeritud efekti jaoks eraldi kui ka kõigi fikseeritud efektide jaoks korraga.



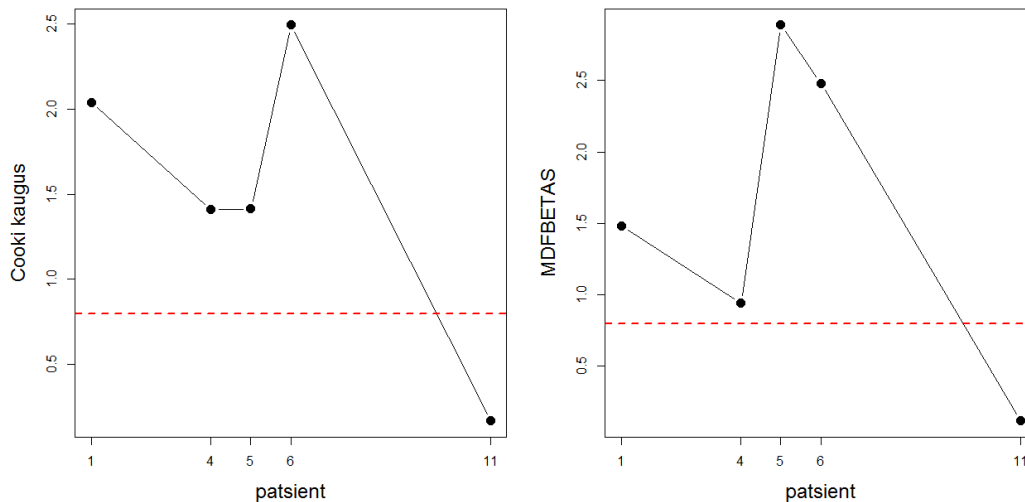
Joonis 4.22. Mõjukusnäitaja DFBETAS väärtused iga patsiendi korral

Joonisel 4.22 on toodud mudeli 4.3 korral leitud DFBETAS, mis näitab, kuidas iga patsiendi vaatlused mõjutavad fikseeritud efektide hinnanguid. Antud näitaja korral on toodud piir sama mis eelnevate mudelite korral korral. See-ga näeme jooniselt 4.22, et kõigi patsientide peale 11. patsiendi vaatlused näivad olevad mõjukad fikseeritud efektide hinnangute suurusele.



Joonis 4.23. Fikseeritud efektide hinnangute protsendiline muutus iga patsiendi korral

Joonisel 4.23 on näidatud, kui palju fikseeritud efektide hinnangud mudeli 4.3 korral muutuvad, kui välja jätta vastava patsiendi vaatlused. Antud mõjukusnäitaja korral ei ole antud piiri, millest suurema väärtusega patsiendid loetakse mõjukaks, ent sarnaselt näitajale DFBETAS on 11. patsiendi vaatlused vähem mõjukad fikseeritud efektide suurusele kui ülejäänud patsientide vaatlused.

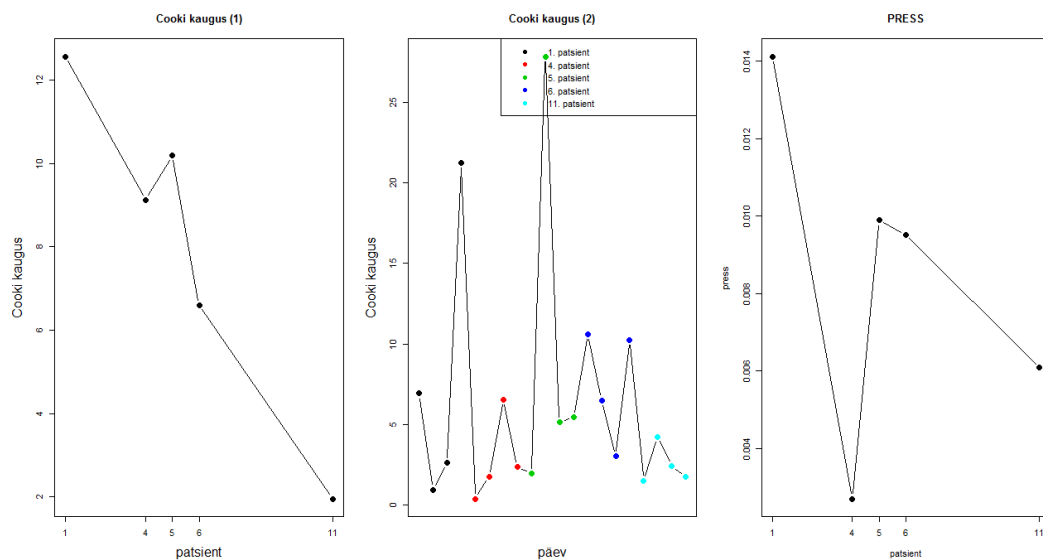


Joonis 4.24. Cooki kaugus ja MDFBETAS väärtus iga patsiendi korral

Joonisel 4.24 on toodud antud andmestiku korral leitud Cooki kaugused ja MDFBETAS väärtused. Need näitavad üldist mõju kõigele fikseeritud faktoritele korraga. Nagu eelnevaltki on Cooki kauguse ja näitaja MDFBETAS mõjukuse piiriks 0.8. Jooniselt 4.24 näeme, et ainult 11. patsiendi vaatlused ei ole mõjukad fikseeritud efektide hinnangute suurusele.

4.5.4 Mõju juhuslikele efektide prognoosidele

Ühe võimalusena võib kasutada näitajat PRESS ja Cooki kaugust, kusjuures taaskord vaatluste eemaldamisel võib juhuslike efektide arv väheneda. Seega arvutame Cooki kauguse väärtused samamoodi nagu mudeli 4.2 korral.

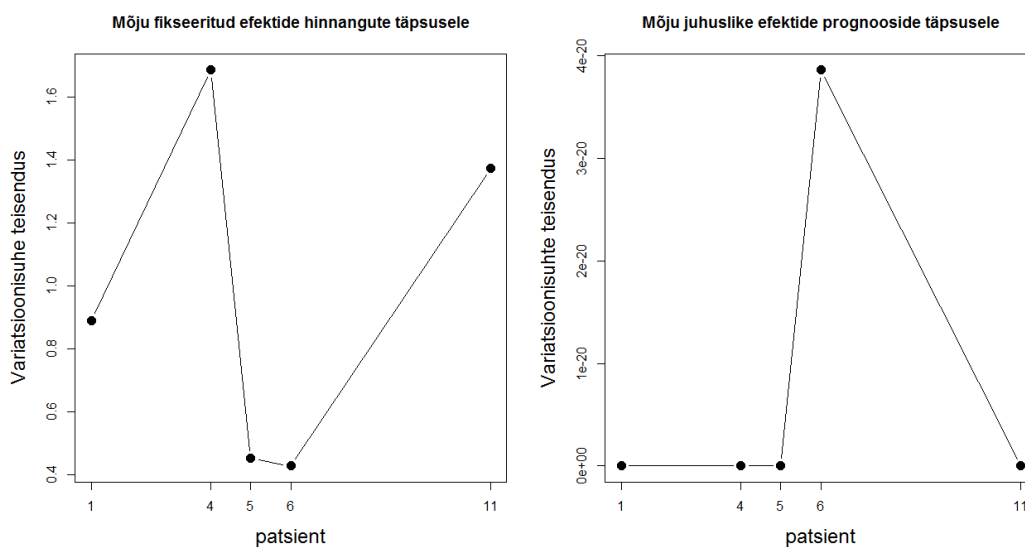


Joonis 4.25. PRESS

Joonisel 4.25 Cooki kauguse väärtused, mis on arvutatud kahel erineval moel, ja mõjukusnäitaja PRESS väärtused mudeli 4.3 korral. Suuremad väärtused viitavad suuremale mõjukusele. Jooniselt 4.25 näeme, et esimesel viisil arvutatud Cooki kauguse väärtuse ja mõjukusnäitaja PRESS väärtuste põhjal on 1. patsiendi vaatlused juhuslike efektide prognooside suurusele kõige mõjukamad. Teisel viisil arvutatud Cooki kauguse väärtuste põhjal näib 1. patsiendi mõjukus olevat tingitud viimasest mõõtmispäevast. Sarnaselt mudelitele 4.1 ja 4.2 on 5. patsiendi vaatlused mõjukamad juhuslike efektide prognooside suurusele kui 4., 6. ja 11. patsiendi vaatlused, ent erinevalt eelnevatest mudelitest näib 5. patsiendi mõjukus olevat tingitud teisest mõõtmispäevast.

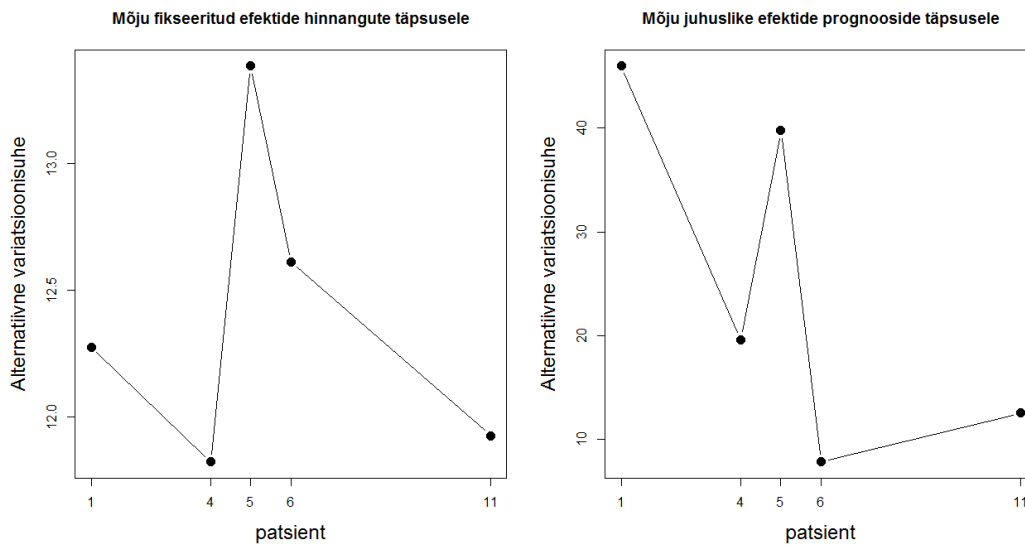
4.5.5 Mõju fikseeritud efektide hinnangute ja juhuslike efektide prognoosi täpsusele

Mudeli 4.3 korral vaatame nii variatsioonisuhte teisendust kui ka alternatiivset variatsioonisuhet.



Joonis 4.26. Variatsioonisuhte teisenduse väärtused iga patsiendi korral

Joonisel 4.26 on toodud variatsioonisuhte teisenduste 3.13 ja 3.16 väärtused iga patsiendi korral. Antud näitaja korral on suurema väärtusega patsiendi väärtused mõjukamad hinnangute või prognooside täpsuse suhtes. Jooniselt 4.26 näeme, et 1. ja 5. patsiendi vaatlused, mis olid mõjukad tõepära kauguse põhjal, ei ole antud kontekstis mõjukad, st need ei ole mõjukad fikseeritud efektide hinnangute ega juhuslike efektide prognooside täpsusele.



Joonis 4.27. Alternatiivse variatsioonisuhte väärtused iga patsiendi korral

Joonisel 4.27 on toodud alternatiivsete variatsioonisuhete 3.14 ja 3.17 väärtused iga patsiendi korral. Näeme, et 5. patsiendi vaatlused mõjutavad nii fikseeritud efektide hinnangute kui ka juhuslike efektide prognooside täpsust. Samas 1. patsiendi vaatlused mõjutavad juhuslike efektide prognoosi täpsust nagu saime ka mudelite 4.1 ja 4.2 korral.

4.6 Tulemused

Kõigi mudelite korral osutusid 1. ja 5. patsiendi vaatlused tõepära kauguse suhtes mõjukaimaks, mis viitab sellele, et nende patsienti puhul tehtud mõõtmised võivad olla vigased. Mudelite 4.1 ja 4.3 korral näis 5. patsiendi vaatluste mõjususe olevat tingitud teisest mõõtmispäevast, mudeli 4.2 korral viimasest mõõtmispäevast. Samas 1. patsiendi korral näib erinevate mudelite korral mõjukus olevat tingitud teisest, kolmandast või viimasest vaatluspäevast.

Mõlema patsiendi vaatlused näisid olevat mingli määral mõjukad nii fikseeri-

tud efektide hinnangute kui ka juhuslike efektide prognooside suuruse suhtes. Alternatiivse variatsioonisuhte põhjal tunduvad 1. patsiendi vaatlused olevalt mõjukad ka juhuslike efektide prognooside täpsuse suhtes, kuid 5. patsiendi vaatlused ei näi olevat mõjukat fikseeritud efektide hinnangute ega juhuslike efektide prognooside täpsuse suhte.

Kokkuvõte

Antud töö eesmärk oli uurida, kuidas leida erindeid ja mõjukaid punkte kasutades lineaarseid segamudeleid. Lineaarsete segamudelite korral on see keeruline, kuna fikseeritud efektide hinnangud ja juhuslike efektide prognoosid sõltuvad nii juhuslike vigade kui ka juhuslike efektide dispersioonimaatriksist, mistõttu klassikalised meetodid antud juhul ei tööta.

Töös on kasutatud peptiidide mikrokiipide andmeid, kus igal patsiendil oli tehtud mitu mõõtmist, mistõttu andmed on omavahel sõltuvad. Seetõttu ongi hea antud andmete korral kasutada lineaarseid segamudeleid.

Töö praktilise osa eesmärgiks oli uurida, kas leidub mõjukas või erindlik patsient. Vaadatud on jääke ja mõjukust nii fikseeritud efektide hinnangutele kui ka juhuslike efektide prognoosidele ning samas ka mõju nende täpsusele. Mõjukateks osutusid 1. ja 5. patsient, kusjuures mõlema patsiendi vaatlused mõjutavad nii fikseeritud efektide hinnangute kui ka juhuslike efektide prognooside suurust ning 1. patsiendi vaatlused on lisaks mõjukad juhuslike efektide prognoosi täpsusele.

Kasutatud kirjandus

- Brus, Ole. 2009. *A comparison of normalisation methods for peptide microarrays*. Uppsala: U.U.D.M. Project Report 2009:3.
- Hartville, David A. 1977. *Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems*. Journal of the American Statistical Association, Vol 72, No 358 (Jun., 1977), pp. 320-338.
- Hilden-Minton, James Andrew. 1995. *Multilevel Diagnostics for Mixed and Hierarchical Linear Models*. Doktoritöö, California Ülikool.
- Nahtman, Tatjana, Alexander Jernberg, Shahnaz Mahdavifar, Johannes Zerweck, Mike Schutkowski, Markus Maeurer ja Marie Reilly. 2007. *Validation of peptide epitope microarray experiments and extraction of quality data*. Journal of Immunological Methods 328, 1-13.
- Nieuwenhuis, Rense, Manfred te Grotenhuis ja Ben Pelzer. 2012. *influence.ME: Tools for Detecting Influential Data on Mixed Effects Models*. The R Journal 4/2, 38-47.
- Schabenberger, O., *Mixed Model Influence Diagnostics* kogumikus *Proceedings of the 29th Annual SAS Users Group International Conference*. Cary, NC: SAS Institute.
- Singer, M.J., Juvencio S. Nobre ja Francisco M. M. Rocha. 2013. *Diagnostics and treatment for linear mixed models* kogumikus *Proceedings 59th ISI World Statistics Congress* 5486.
- Zewotir, Temesgen, Jacky S. Galpin. 2005. *Influence Diagnostics for Linear Mixed Models*. *Journal of Data Science* 3. 153-177.

- West, Brady T., Kathleen B. Welch ja Andrzej T. Galecki. 2007. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall/CRC.

Lisad

Lisa A. Andmete töötlus

```
>andmed2=andmed[,c(7,9,10,11)]
>head(andmed2)
index s patient day
1 0.108623546 3 5 252
2 -0.040641984 1 5 112
3 -0.108129900 3 5 112
4 0.390789953 1 6 56
5 -0.007593149 3 5 252
6 -0.041519269 3 5 112
>summary(andmed2)
index s patient day
Min. :-3.311333 Min. :1 Min. : 1.0 Min. : 0
1st Qu.: -0.028922 1st Qu.:1 1st Qu.: 4.0 1st Qu.: 42
Median : 0.008824 Median :2 Median : 5.0 Median : 84
Mean : 0.058195 Mean :2 Mean : 5.4 Mean :105
3rd Qu.: 0.056872 3rd Qu.:3 3rd Qu.: 6.0 3rd Qu.:147
Max. : 4.682048 Max. :3 Max. :11.0 Max. :252
>###KESKMISTAMINE S,DAY,PATIENT
>andmed2=andmed[,c(7,9,10,11)]
>head(andmed2)
index s patient day
1 0.108623546 3 5 252
2 -0.040641984 1 5 112
3 -0.108129900 3 5 112
4 0.390789953 1 6 56
5 -0.007593149 3 5 252
6 -0.041519269 3 5 112
>n=5*3*4*4; andm.kesk=matrix(rep(0,n),ncol=4)
>ii=0
>for (i in c(1,4,5,6,11)){
+ p=andmed2[andmed2$patient==i,]
+ for(j in c(1,2,3)){
+ ps=p[ps==j,]
+ for (k in c(0,56,112,252)){
+ psd=ps[ps$day==k,]
+ ii=ii+1
+ andm.kesk[ii,]=c(mean(psd$index),j,i,k)}}}
>colnames(andm.kesk)=c("index","s","patient","day")
>andm.kesk=data.frame(andm.kesk)
>andm.kesk$patient=as.factor(andm.kesk$patient)
>andm.kesk$day=as.factor(andm.kesk$day)
>andm.kesk$s=as.factor(andm.kesk$s)
>#attach(andm.kesk)
>head(andm.kesk)
index s patient day
1 0.09543724 1 1 0
2 0.11774667 1 1 56
3 0.09644635 1 1 112
4 0.06182401 1 1 252
5 0.09797365 2 1 0
6 0.10257727 2 1 56
>str(andm.kesk)
'data.frame': 60 obs. of 4 variables:
 $ index : num 0.0954 0.1177 0.0964 0.0618 0.098 ...
 $ s : Factor w/ 3 levels "1","2","3": 1 1 1 1 2 2 2 2 3 3 ...
 $ patient: Factor w/ 5 levels "1","4","5","6",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ day : Factor w/ 4 levels "0","56","112",...: 1 2 3 4 1 2 3 4 1 2 ...
```

Lisa B. Mudel 4.1 ja selle hinnangud

```
>mm=lmer(index~s+(1|patient)+(1|day),data=andm.kesk)
>mm1=lmer(index~1+(1|patient)+(1|day),data=andm.kesk)
>anova(mm,mm1)
Data: andm.kesk
Models:
mm1: index ~ 1 + (1 | patient) + (1 | day)
mm: index ~ s + (1 | patient) + (1 | day)
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
mm1 4 -312.93 -304.56 160.47 -320.93
mm 6 -314.75 -302.19 163.38 -326.75 5.8202 2 0.05447 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>mm2=lmer(index~(1|patient),data=andm.kesk)
>anova(mm1,mm2)
Data: andm.kesk
Models:
mm2: index ~ (1 | patient)
mm1: index ~ 1 + (1 | patient) + (1 | day)
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
mm2 3 -307.33 -301.05 156.66 -313.33
mm1 4 -312.93 -304.56 160.47 -320.93 7.6049 1 0.005821 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>mm3=lmer(index~(1|day),data=andm.kesk)
>anova(mm1,mm3)
Data: andm.kesk
Models:
mm3: index ~ (1 | day)
mm1: index ~ 1 + (1 | patient) + (1 | day)
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
mm3 3 -275.63 -269.35 140.82 -281.63
mm1 4 -312.93 -304.56 160.47 -320.93 39.303 1 3.629e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>summary(mm1)
Linear mixed model fit by REML ['lmerMod']
Formula: index ~ 1 + (1 | patient) + (1 | day)
Data: andm.kesk

REML criterion at convergence: -313.4133

Random effects:
Groups Name Variance Std.Dev.
patient (Intercept) 3.712e-04 0.01927
day (Intercept) 6.577e-05 0.00811
Residual 1.986e-04 0.01409
Number of obs: 60,groups: patient,5; day,4

Fixed effects:
Estimate Std. Error t value
(Intercept) 0.058195 0.009695 6.003
>### beta hinnagud: vabaliige
>beta=as.matrix(fixef(mm1))
>beta
[,1]
(Intercept) 0.05819474
>
>### u hinnangud: patsient ja päev
>u=as.matrix(rbind(ranef(mm1)$patient,ranef(mm1)$day))
```

```

>u
(Intercept)
1 2.549131e-02
4 1.674237e-03
5 -2.771823e-02
6 -3.526514e-05
11 5.879482e-04
0 9.188431e-04
56 9.185974e-03
112 -1.312876e-03
252 -8.791941e-03
>
>### plaanimaatriksid
>Z = t(as.matrix(mm1@pp$Zt))
>head(Z)
1 4 5 6 11 0 56 112 252
[1,] 1 0 0 0 0 1 0 0 0
[2,] 1 0 0 0 0 0 1 0 0
[3,] 1 0 0 0 0 0 0 1 0
[4,] 1 0 0 0 0 0 0 0 1
[5,] 1 0 0 0 0 1 0 0 0
[6,] 1 0 0 0 0 0 1 0 0
>X = as.matrix(model.matrix(mm1))
>head(X)
(Intercept)
1 1
2 1
3 1
4 1
5 1
6 1
>
>### vaatlused
>y = as.vector(mm1@resp$y)
>head(y)
[1] 0.09543724 0.11774667 0.09644635 0.06182401 0.09797365 0.10257727
>
>q=length(u)
>n=length(y)
>p=1
>
>d=as.matrix(VarCorr(mm1))
>sig2 = attr(VarCorr(mm1),"sc")^2
>sig.p=as.numeric(VarCorr(mm1)$patient)
>sig.d=as.numeric(VarCorr(mm1)$day)
>k.p=length(as.matrix(rbind(ranef(mm1)$patient)))
>k.d=length(as.matrix(rbind(ranef(mm1)$day)))
>Dbeta=as.matrix(vcov(mm1))
>
># kovariatsioonimaatriks
>G=diag(c(rep(sig.p,k.p),rep(sig.d,k.d)))
>R=sig2*diag(n)
>V = Z%*%G%*%t(Z)+R
>V.inv = solve(V)

```

Lisa C. Mudel 4.1: jäägid

```
# tinglikud
r.cond=residuals(mm)

# marginaalsed
r.marg=y-X%*%beta

# juhuslikud
r.juh=Z%*%u

### STUDENTISEERITUD
# tinglikud
r.cond.sv=r.cond/sd(r.cond) # välimised
r.cond.ss=r.cond
for (i in 1:length(r.cond)){
r.cond.ss[i]=r.cond[i]/sd(r.cond[-i])} # sisemised

# marginaalsed
r.marg.sv=r.marg/sd(r.marg) # välimised
r.marg.ss=r.marg
for (i in 1:length(r.marg)){
r.marg.ss[i]=r.marg[i]/sd(r.marg[-i])} # sisemised

# juhuslikud
r.juh.sv=r.juh/sd(r.juh) # välimised
r.juh.ss=r.juh
for (i in 1:length(r.juh)){
r.juh.ss[i]=r.juh[i]/sd(r.juh[-i])} # sisemised
```

Lisa D. Mudel 4.1: tõepära kaugus

```
log.t=-n*0.5*log(2*pi)-0.5*log(det(V))-0.5*t(y-X%*beta)%*%V.inv%*(y-X%*beta)

### iga patsiendi jaoks
LD=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v,REML=F)

  beta.v=as.matrix(fixef(mm.v))
  u.v=as.matrix(rbind(ranef(mm.v)$patient,ranef(mm.v)$day))

  Z.v = t(as.matrix(mm.v@pp$Zt))
  X.v = as.matrix(model.matrix(mm.v))

  q.v=length(u.v)

  sig2.v = attr(VarCorr(mm.v),"sc")^2
  sig.p.v=as.numeric(VarCorr(mm.v)$patient)
  sig.d.v=as.numeric(VarCorr(mm.v)$day)

  G.v=diag(c(rep(sig.p.v,k.p),rep(sig.d.v,k.d)))
  R.v=sig2.v*diag(n)
  V.v = Z%*%G.v%*%t(Z)+R.v
  V.inv.v = solve(V.v)

  log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v))-0.5*t(y-X%*beta.v)%*%V.inv.v%*(y-X%*beta.v)
  LD[j]=2*(log.t-log.t.v)
  j=j+1}

plot(c(1,4,5,6,11),LD,type="b",xlab="patsient",ylab="Tõepära kaugus",xaxt="n",pch=19,cex=1.5,c
axis(1,at=c(1,4,5,6,11),las=0)
# näeme,et tõepära kauguse põhjal on kõige mõjukamad 1. ja 5. patsient

### iga päeva jaoks
LD.day=rep(0,4); j=1
for (i in c(0,56,112,252)){
  andm.v=andm.kesk[andm.kesk$day!=i,]
  mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v,REML=F)

  beta.v=as.matrix(fixef(mm.v))
  u.v=as.matrix(rbind(ranef(mm.v)$patient,ranef(mm.v)$day))

  Z.v = t(as.matrix(mm.v@pp$Zt))
  X.v = as.matrix(model.matrix(mm.v))

  q.v=length(u.v) # juhuslike faktorite arv: 20

  sig2.v = attr(VarCorr(mm.v),"sc")^2
  sig.p.v=as.numeric(VarCorr(mm.v)$patient)
  sig.d.v=as.numeric(VarCorr(mm.v)$day)

  G.v=diag(c(rep(sig.p.v,k.p),rep(sig.d.v,k.d)))
  R.v=sig2.v*diag(n)
  V.v = Z%*%G.v%*%t(Z)+R.v
  V.inv.v = solve(V.v)

  log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v))-0.5*t(y-X%*beta.v)%*%V.inv.v%*(y-X%*beta.v)
  LD.day[j]=2*(log.t-log.t.v)
  j=j+1}
```

```

plot(c(0,56,112,252),LD.day,type="b",xlab="päev",ylab="Tõepära kaugus",xaxt="n",pch=19,cex=1.5)
axis(1,c(0,56,112,252),las=0)

### iga patsiendi iga päeva jaoks
LD.dp=rep(0,20); j=1
for (i in 1:5){
for (k in 1:4){
andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v,REML=F)

beta.v=as.matrix(fixef(mm.v))
u.v=as.matrix(rbind(ranef(mm.v)$patient,ranef(mm.v)$day))

Z.v = t(as.matrix(mm.v@pp$Zt))
X.v = as.matrix(model.matrix(mm.v))

q.v=length(u.v)

sig2.v = attr(VarCorr(mm.v),"sc")^2
sig.p.v=as.numeric(VarCorr(mm.v)$patient)
sig.d.v=as.numeric(VarCorr(mm.v)$day)

G.v=diag(c(rep(sig.p.v,k.p),rep(sig.d.v,k.d)))
R.v=sig2.v*diag(n)
V.v = Z%*%G.v%*%t(Z)+R.v
V.inv.v = solve(V.v)

log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v))-0.5*t(y-X%*%beta.v)%*%V.inv.v%*%(y-X%*%beta.v)
LD.dp[j]=2*(log.t-log.t.v)
j=j+1}}

```

Lisa E. R: Mudel 4.1: mõjukus fikseeritud efektide hinnangutele

```
dfbetas.vabaliige=rep(0,5);uj=1
prots.vabaliige=rep(0,5)
for(i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v)

  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)

  dfbetas.vabaliige[j]=(beta[1]-beta.v[1])/sqrt(Dbeta.v[1,1])
  prots.vabaliige[j]=(beta[1]-beta.v[1])/beta[1]*100
  j=j+1}
```

Lisa F. Mudel 4.1: mõjukus juhuslike efektide prognoosidele

```

### Cooki kaugus (1)
cooku.v=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
andm.v=andm.kesk[andm.kesk$patient!=i,]
mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v)

beta.v=as.matrix(fixef(mm.v))
u.v=as.matrix(rbind(ranef(mm.v)$patient,ranef(mm.v)$day))
u.v2=u[-j]

Z.v = t(as.matrix(mm.v@pp$Zt))
X.v = as.matrix(model.matrix(mm.v))

q.v=length(u.v)
n.v=length(andm.v$index)

sig2.v = attr(VarCorr(mm.v),"sc")^2
k.p.v=length(as.matrix(rbind(ranef(mm.v)$patient)))
k.d.v=length(as.matrix(rbind(ranef(mm.v)$day)))
sig.p.v=as.matrix(VarCorr(mm.v)$patient)
sig.d.v=as.matrix(VarCorr(mm.v)$day)

G.v=diag(c(rep(sig.p.v,k.p.v),rep(sig.d.v,k.d.v)))
R.v=sig2.v*diag(n.v)
V.v = Z.v%*%G.v%*%t(Z.v)+R.v
V.inv.v = solve(V.v)

D.u.v=G.v%*%t(Z.v)%*%V.inv.v%*%Z.v%*%t(G.v)
cooku.v[j]=0.5*t(u.v2-u.v)%*%ginv(D.u.v)%*%(u.v2-u.v)
j=j+1}

### Cooki kaugus (2)
cooku.dp.v=rep(0,20);j=1
for (i in 1:5){
for (k in 1:4){
andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
mm.v=lmer(index~1+(1|patient)+(1|day),data=andm.v)

beta.v=as.matrix(fixef(mm.v))
u.v=as.matrix(rbind(ranef(mm.v)$patient,ranef(mm.v)$day))

Z.v = t(as.matrix(mm.v@pp$Zt))
X.v = as.matrix(model.matrix(mm.v))

q.v=length(u.v)
n.v=length(andm.v$index)

sig2.v = attr(VarCorr(mm.v),"sc")^2
k.p.v=length(as.matrix(rbind(ranef(mm.v)$patient)))
k.d.v=length(as.matrix(rbind(ranef(mm.v)$day)))
sig.p.v=as.matrix(VarCorr(mm.v)$patient)
sig.d.v=as.matrix(VarCorr(mm.v)$day)

G.v=diag(c(rep(sig.p.v,k.p),rep(sig.d.v,k.d)))
R.v=sig2.v*diag(n)
V.v = Z%*%G.v%*%t(Z)+R.v
V.inv.v = solve(V.v)

```

```

D.u.v=G.v%%t(Z)%%V.inv.v%%Z%%t(G.v)
cooku.dp.v[j]=0.5*t(u-u.v)%%ginv(D.u.v)%%(u-u.v)
j=j+1}}

### PRESS
press=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
andm.p=andm.kesk[andm.kesk$patient!=i,]
mm.p=lmer(index~1+(1|patient)+(1|day),data=andm.p)
beta.p=as.matrix(fixef(mm.p))

yi=as.vector(mm@resp$y)[andm.kesk$patient==i]
xi=X[andm.kesk$patient==i,]
ei=yi-xi%%beta.p
press[j]=sum(ei^2)
j=j+1}

```

Lisa G. Mudel 4.1: mõjukus täpsusele

```
### fikseeritud efektid
det1=as.matrix(vcov(mm1));varr=rep(0,5)
VS=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.var=andm.kesk[andm.kesk$patient!=i,]
  mm.var = lmer(index~1+(1|patient)+(1|day),data=andm.var)
  det2=as.matrix(vcov(mm.var))
  varr[j]=as.matrix(vcov(mm.var))
  VS[j]=abs(det2/det1-1)
  j=j+1}

### juhuslikud efektid
detu1=det(G%*%t(Z)%*%V.inv%*%Z%*%G)
VSu=rep(0,5);VSutr=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.var=andm.kesk[andm.kesk$patient!=i,]
  mm.var = lmer(index~1+(1|patient)+(1|day),data=andm.var)

  Z.var = t(as.matrix(mm.var@pp$Zt))

  q.var=length(ranef(mm.var))
  n.var=length(andm.var$index)
  sig2.var = attr(VarCorr(mm.var),"sc")^2
  sigma.var = diag(n.var)
  # kovariatsioonimaatriks
  g.p.var = as.matrix(VarCorr(mm.var)$patient)
  g.d.var = as.matrix(VarCorr(mm.var)$day)
  g.var=c(rep(g.p.var,4),rep(g.d.var,4))
  G.var = diag(g.var)
  V.var = Z.var%*%G.var%*%t(Z.var)+sig2.var*sigma.var
  V.inv.var = solve(V.var) # KOVARIATSIOONIMAATRIKS
  det2=det(G.var%*%t(Z.var)%*%V.inv.var%*%Z.var%*%G.var)
  VSu[j]=abs(det2/detu1-1)
  VSutr[j]=abs(sum(diag(solve(du1)%*%du2))-8)
  j=j+1}
```

Lisa H. Mudel 4.2 ja selle hinnangud

```
mm=lmer(index~s-1+(1|patient/day),data=andm.kesk)

### kas tunnused on olulised?
mm1=lmer(index~s-1+(1|patient),data=andm.kesk)
anova(mm,mm1) # 7.117e-05, seega erinevate patsientide korral on päevad erinevad
mm2=lmer(index~s-1+(1|patient:day),data=andm.kesk)
anova(mm,mm2) # 0.003978, igal patsiendil on erinev vabaliige
mm2=lmer(index~1+(1|patient:day),data=andm.kesk)
anova(mm,mm2) # 0.0007293 slaid on statistiliselt oluline

### Vajalikud hinnangud
### beta hinnangud
beta=as.matrix(fixef(mm))
Dbeta=as.matrix(vcov(mm))

### u hinnangud
u=as.matrix(rbind(ranef(mm)$'day:patient', ranef(mm)$patient))

### plaanimaatriksid
Z = t(as.matrix(mm@pp$Zt))
X = as.matrix(model.matrix(mm))

### vaatlused
y = as.vector(mm@resp$y)

q=length(u)
n=length(y)
p=3

sig2 = attr(VarCorr(mm), "sc")^2
sig.dp=as.matrix(VarCorr(mm)$'day:patient')
sig.p=as.matrix(VarCorr(mm)$patient)
k.dp=length(as.matrix(rbind(ranef(mm)$'day:patient')))
k.p=length(as.matrix(rbind(ranef(mm)$patient)))
# kovariatsioonimaatriks
G=diag(c(rep(sig.dp,k.dp),rep(sig.p,k.p)))
R=sig2*diag(n)
V = Z%*%G%*%t(Z)+R
V.inv = solve(V)
# hinnangu y-le
y.hin=X%*%beta+Z%*%u
```

Lisa I. Mudel 4.2: jäägid

```
#_tinglikud
r.cond=residuals(mm)

#_marginaalsed
r.marg=y-X%*%beta

#_juhuslikud
r.juh=Z%*%u

###_STUDENTISEERITUD
#_tinglikud
r.cond.sv=r.cond/sd(r.cond) #_välimised
r.cond.ss=r.cond
for_(i_in_1:length(r.cond)){
r.cond.ss[i]=r.cond[i]/sd(r.cond[-i])} #_sisemised

#_marginaalsed
r.marg.sv=r.marg/sd(r.marg) #_välimised
r.marg.ss=r.marg
for_(i_in_1:length(r.marg)){
r.marg.ss[i]=r.marg[i]/sd(r.marg[-i])} #_sisemised

#_juhuslikud
r.juh.sv=r.juh/sd(r.juh) #_välimised
r.juh.ss=r.juh
for_(i_in_1:length(r.juh)){
r.juh.ss[i]=r.juh[i]/sd(r.juh[-i])} #_sisemised
```

Lisa J. Mudel 4.2: tõepära kaugus

```
log.t=-n*0.5*log(2*pi)-0.5*log(det(V))-0.5*t(y-X**beta)**V.inv**(y-X**beta)

### iga patsiendi korral
LD=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~s-1+(1|patient/day),data=andm.v,REML=F)
  beta.v=as.matrix(fixef(mm.v))
  sig2.v = attr(VarCorr(mm.v),"sc")^2
  sig.p.v=as.matrix(VarCorr(mm.v)$patient)
  sig.dp.v=as.matrix(VarCorr(mm.v)$'day:patient')
  G.v=diag(c(rep(sig.dp.v,k.dp),rep(sig.p.v,k.p)))
  R.v=sig2.v*diag(n)
  V.v = Z**G.v**t(Z)+R.v
  V.inv.v = solve(V.v)
  log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v))-0.5*t(y-X**beta.v)**V.inv.v**(y-X**beta.v)
  LD[j]=2*(log.t-log.t.v)
  j=j+1}

LDdp=rep(0,20); j=1
for (i in 1:5){
  for (k in 1:4){
    andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
    mm.v=lmer(index~s-1+(1|patient/day),data=andm.v,REML=F)
    beta.v=as.matrix(fixef(mm.v))
    sig2.v = attr(VarCorr(mm.v),"sc")^2
    sig.p.v=as.matrix(VarCorr(mm.v)$patient)
    sig.dp.v=as.matrix(VarCorr(mm.v)$'day:patient')
    G.v=diag(c(rep(sig.dp.v,k.dp),rep(sig.p.v,k.p)))
    R.v=sig2.v*diag(n)
    V.v = Z**G.v**t(Z)+R.v
    V.inv.v = solve(V.v)
    log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v))-0.5*t(y-X**beta.v)**V.inv.v**(y-X**beta.v)
    LDdp[j]=2*(log.t-log.t.v)
    j=j+1}}
```

Lisa K. Mudel 4.2: mõjukus fikseeritud efektide hinnanguite suurusele

```
### DFBETAS
### s=1
dfbetas.s1=rep(0,5); j=1
prots.s1=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)
  dfbetas.s1[j]=(beta[1]-beta.v[1])/sqrt(Dbeta.v[1,1])
  prots.s1[j]=(beta[1]-beta.v[1])/beta[1]*100
  j=j+1}

### s=2
dfbetas.s2=rep(0,5); j=1
prots.s2=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)
  dfbetas.s2[j]=(beta[2]-beta.v[2])/sqrt(Dbeta.v[2,2])
  prots.s2[j]=(beta[2]-beta.v[2])/beta[2]*100
  j=j+1}

### s=3
dfbetas.s3=rep(0,5); j=1
prots.s3=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)
  dfbetas.s3[j]=(beta[3]-beta.v[3])/sqrt(Dbeta.v[3,3])
  prots.s3[j]=(beta[3]-beta.v[3])/beta[3]*100
  j=j+1}

### Cook ja mdfbetas
cook=rep(0,5); j=1
mdfbetas=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=as.matrix(vcov(mm.v))
  cook[j]=t(beta-beta.v)%*%solve(Dbeta.v)%*(beta-beta.v)/p
  mdfbetas[j]=t(beta-beta.v)%*%solve(Dbeta.v)%*(beta-beta.v)/p
  j=j+1}
```

Lisa L. Mudeli 4.2: mõjukus juhuslike efektide prognooside suurusele

```

###_cooki_kaugus_(1)
cooku.v=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
andm.v=andm.kesk[andm.kesk$patient!=i,]
mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
beta.v=as.matrix(fixef(mm.v))
u.v=as.matrix(rbind(ranef(mm.v)$'day:patient',ranef(mm.v)$patient))
u.v2=as.matrix(u[-c(j,j+5,j+10,j+15,j+20)])
Z.v = t(as.matrix(mm.v@pp$Zt))
X.v = as.matrix(model.matrix(mm.v))
q.v=length(u.v)
n.v=length(andm.v$index)
sig2.v = attr(VarCorr(mm.v),"sc")^2
k.p.v=length(as.matrix(rbind(ranef(mm.v)$patient)))
k.d.v=length(as.matrix(rbind(ranef(mm.v)$day)))
sig.p.v=as.matrix(VarCorr(mm.v)$patient)
sig.d.v=as.matrix(VarCorr(mm.v)$day)
G.v=diag(c(rep(sig.p.v,k.p.v),rep(sig.d.v,k.d.v)))
R.v=sig2.v*diag(n.v)
V.v = Z.v%*%G.v%*%t(Z.v)+R.v
V.inv.v = solve(V.v)
D.u.v=G.v%*%t(Z.v)%*%V.inv.v%*%Z.v%*%t(G.v)
cooku.v[j]=0.5*t(u.v2-u.v)%*%ginv(D.u.v)%*(u.v2-u.v)
j=j+1}

###_cooki_kaugus_(2)
cooku.dp.v=rep(0,20);j=1
for (i in 1:5){
for (k in 1:4){
andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
beta.v=as.matrix(fixef(mm.v))
u.v=as.matrix(rbind(ranef(mm.v)$'day:patient',ranef(mm.v)$patient))
u.v2=as.matrix(u[-((k-1)*5+i)])
Z.v = t(as.matrix(mm.v@pp$Zt))
X.v = as.matrix(model.matrix(mm.v))
q.v=length(u.v)
n.v=length(andm.v$index)
sig2.v = attr(VarCorr(mm.v),"sc")^2
k.p.v=length(as.matrix(rbind(ranef(mm.v)$patient)))
k.d.v=length(as.matrix(rbind(ranef(mm.v)$day)))
sig.p.v=as.matrix(VarCorr(mm.v)$patient)
sig.d.v=as.matrix(VarCorr(mm.v)$day)
G.v=diag(c(rep(sig.p.v,k.p.v),rep(sig.d.v,k.d.v)))
R.v=sig2.v*diag(n.v)
V.v = Z.v%*%G.v%*%t(Z.v)+R.v
V.inv.v = solve(V.v)
D.u.v=G.v%*%t(Z.v)%*%V.inv.v%*%Z.v%*%t(G.v)
cooku.dp.v[j]=0.5*t(u.v-u.v2)%*%ginv(D.u.v)%*(u.v-u.v2)
j=j+1}}

### PRESS
press=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
andm.v=andm.kesk[andm.kesk$patient!=i,]
mm.v=lmer(index~s-1+(1|patient/day),data=andm.v)
beta.v=as.matrix(fixef(mm.v))
yi=as.vector(mm@resp$y)[andm.kesk$patient==i]

```

```
xi=X[andm.kesk$patient==i,]  
ei=yi-xi**beta.v  
press[j]=sum(ei^2)  
j=j+1}
```

Lisa M. Mudel 4.1: mõjukus täpsusele

```
### variatsioonisuhte teisendus: fikseeritud efektid
det1=det(vcov(mm))
VS=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.var=andm.kesk[andm.kesk$patient!=i,]
  mm.var=lmer(index~s-1+(1|patient/day),data=andm.var)
  beta.var=as.matrix(fixef(mm.var))
  det2=det(vcov(mm.var))
  VS[j]=abs(det2/det1-1)
  j=j+1}

### variatsioonisuhte teisendus juhuslikud efektid
detu1=det(G%*%t(Z)%*%V.inv%*%Z%*%G)
VSu=rep(0,5);VSu.tr=rep(0,5);VSu.tr2=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.var=andm.kesk[andm.kesk$patient!=i,]
  mm.var=lmer(index~factor(s)+(1|patient/day),data=andm.var)
  Z.var = t(as.matrix(mm.var@pp$Zt))
  q.var=length(ranef(mm.var))
  n.var=length(andm.var$index)
  sig2.var = attr(VarCorr(mm.var),"sc")^2
  sigma.var = diag(n.var)
  # kovariatsioonimaatriks
  k.p.var = 4
  g.p.var = as.matrix(VarCorr(mm.var)$patient)
  k.dp.var = 16
  g.dp.var = as.matrix(VarCorr(mm.var)$'day:patient')
  g.var=c(rep(g.p.var,k.p.var),rep(g.dp.var,k.dp.var))
  G.var = diag(g.var)
  V.var = Z.var%*%G.var%*%t(Z.var)+sig2.var*sigma.var
  V.inv.var = solve(V.var) # KOVARIATSIOONIMAATRIKS
  det2=det(G.var%*%t(Z.var)%*%V.inv.var%*%Z.var%*%G.var)
  VSu[j]=abs(det2/detu1-1)
  j=j+1}

### alternatiivne variatsioonisuhte: juhuslikud efektid
detu1=det(G%*%t(Z)%*%V.inv%*%Z%*%G)
VSu=rep(0,5);VSutr=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.var=andm.kesk[andm.kesk$patient!=i,]
  mm.var=lmer(index~factor(s)+(1|patient/day),data=andm.var)
  Z.var = t(as.matrix(mm.var@pp$Zt))
  q.var=length(ranef(mm.var))
  n.var=length(andm.var$index)
  sig2.var = attr(VarCorr(mm.var),"sc")^2
  sigma.var = diag(n.var)
  # kovariatsioonimaatriks
  k.p.var = 4
  g.p.var = as.matrix(VarCorr(mm.var)$patient)
  k.dp.var = 16
  g.dp.var = as.matrix(VarCorr(mm.var)$'day:patient')
  g.var=c(rep(g.p.var,k.p.var),rep(g.dp.var,k.dp.var))
  G.var = diag(g.var)
  V.var = Z.var%*%G.var%*%t(Z.var)+sig2.var*sigma.var
  V.inv.var = solve(V.var) # KOVARIATSIOONIMAATRIKS
  det2=det(G.var%*%t(Z.var)%*%V.inv.var%*%Z.var%*%G.var)
  VSu[j]=abs(det2/detu1-1)
  du1=(G%*%t(Z)%*%V.inv%*%Z%*%G)[-(c(j,j+5,j+10,j+15,j+20)),-(c(j,j+5,j+10,j+15,j+20))]
  du2=G.var%*%t(Z.var)%*%V.inv.var%*%Z.var%*%G.var
```

```
VSutr[j]=abs(sum(diag( ginv(du1)%*%du2))-20)
j=j+1}
```

Lisa N. Mudel 4.3 ja selle hinnangud

```
mm=lmer(index~s-1+(day-1|patient),data=andm.kesk)

### kas tunnused on olulised?
mm1=lmer(index~1+(day-1|patient),data=andm.kesk)
anova(mm,mm1) # 0.0133 s on oluline
mm1=lmer(index~s-1+(1|patient),data=andm.kesk)
anova(mm,mm1) # 0.003101 ,iga patsiendi korral päevad käituvad erinevalt

n=60
beta=fixef(mm)
u=raneef(mm)$patient;u=t(as.matrix(cbind(u[1,],u[2,],u[3,],u[4,],u[5,])))

Dbeta=vcov(mm) # solve(t(X)%*%V.inv%*%X)-vcov(mm)
p=3 # rank X

X = as.matrix(model.matrix(mm))
Zi=matrix(c(1,0,0,0,1,0,0,0,1,0,0,0,
0,1,0,0,0,1,0,0,0,1,0,0,
0,0,1,0,0,0,1,0,0,0,1,0,
0,0,0,1,0,0,0,1,0,0,0,1),nrow=12)
Z=kronecker(diag(rep(1,5)),Zi)

Gi=matrix(VarCorr(mm)$patient[1:16],nrow=4)
G=kronecker(diag(rep(1,5)),Gi)
Ri=attr(VarCorr(mm),"sc")**2*diag(rep(1,12))
Vi=Zi%*%Gi%*%t(Zi)+Ri
V=kronecker(diag(rep(1,5)),Vi)
V.inv=solve(V)
Du=G%*%t(Z)%*%V.inv%*%Z%*%G
y=andm.kesk$index
```

Lisa O. Mudel 4.3: jäägid

```
### RAW
# tinglikud
r.cond=residuals(mm)

# marginaalsed
r.marg=y-X%%beta

# juhuslikud
r.juh=Z%%uuu

### STUDENTISEERITUD
# tinglikud
r.cond.sv=r.cond/sd(r.cond) # välimised
r.cond.ss=r.cond
for (i in 1:length(r.cond)){
r.cond.ss[i]=r.cond[i]/sd(r.cond[-i])} # sisemised

# marginaalsed
r.marg.sv=r.marg/sd(r.marg) # välimised
r.marg.ss=r.marg
for (i in 1:length(r.marg)){
r.marg.ss[i]=r.marg[i]/sd(r.marg[-i])} # sisemised

# juhuslikud
r.juh.sv=r.juh/sd(r.juh) # välimised
r.juh.ss=r.juh
for (i in 1:length(r.juh)){
r.juh.ss[i]=r.juh[i]/sd(r.juh[-i])} # sisemised
```

Lisa P. Mudel 4.3: tõepära kaugus

```
log.t=-n*0.5*log(2*pi)-0.5*log(det(V))-0.5*t(y-X%*%b)%*%V.inv%*(y-X%*%b)
LD=rep(0,5)

# tõepära kaugus iga patsiendi korral
LD=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  andm.v$patient=as.factor(as.vector(andm.v$patient))
  mm.v=lmer(index~s-1+(day-1|patient),data=andm.v,REML=F)
  beta.v=as.matrix(fixef(mm.v))
  X.v = as.matrix(model.matrix(mm.v))
  Z.v=kroncker(diag(rep(1,4)),Zi)
  Gi.v=matrix(VarCorr(mm.v)$patient[1:16],nrow=4)
  G.v=kroncker(diag(rep(1,4)),Gi.v)
  Ri.v=attr(VarCorr(mm.v),"sc")**2*diag(rep(1,12))
  Vi.v=Zi%*%Gi.v%*%t(Zi)+Ri.v
  V.v=kroncker(diag(rep(1,4)),Vi.v)
  V.inv.v=solve(V.v)
  V.v2=kroncker(diag(rep(1,5)),Vi.v)
  V.inv.v2=solve(V.v2)
  y.v=y[andm.kesk$patient!=i]
  bv=solve(t(X.v)%*%V.inv.v%*%X.v)%*%t(X.v)%*%V.inv.v%*%y.v
  log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v2))-0.5*t(y-X%*%bv)%*%V.inv.v2%*(y-X%*%bv)
  LD[j]=2*(log.t-log.t.v)
  j=j+1}

# tõepära kaugus iga patsiendi iga päeva korral
LDdp=rep(0,20); j=1
for (i in 1:5){
  for (k in 1:4){
    andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
    andm.v$patient=as.factor(as.vector(andm.v$patient))
    mm.v=lmer(index~s-1+(day-1|patient),data=andm.v,REML=F)
    beta.v=as.matrix(fixef(mm.v))
    X.v = as.matrix(model.matrix(mm.v))
    Z.v=Z[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),-j]
    Gi.v=matrix(VarCorr(mm.v)$patient[1:16],nrow=4)
    G.v=kroncker(diag(rep(1,5)),Gi.v)[-j,-j]
    R.v=attr(VarCorr(mm.v),"sc")**2*diag(rep(1,57))
    V.v=Z.v%*%G.v%*%t(Z.v)+R.v
    V.inv.v=solve(V.v)
    V.v2=kroncker(diag(rep(1,5)),Vi.v)
    V.inv.v2=solve(V.v2)
    y.v=y[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8))]
    bv=solve(t(X.v)%*%V.inv.v%*%X.v)%*%t(X.v)%*%V.inv.v%*%y.v
    log.t.v=-n*0.5*log(2*pi)-0.5*log(det(V.v2))-0.5*t(y-X%*%bv)%*%V.inv.v2%*(y-X%*%bv)
    LDdp[j]=2*(log.t-log.t.v)
    j=j+1}}}
```

Lisa Q. Mudel 4.3: mõjukus fikseeritud efektide hinnanguitele

```
### DFBETAS
dfbetas.s1=rep(0,5); j=1
prots.s1=rep(0,5)
dfbetas.s2=rep(0,5)
prots.s2=rep(0,5)
dfbetas.s3=rep(0,5)
prots.s3=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  andm.v$patient=as.factor(as.vector(andm.v$patient))
  mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)
  dfbetas.s1[j]=(beta[1]-beta.v[1])/sqrt(Dbeta.v[1,1])
  prots.s1[j]=(beta[1]-beta.v[1])/beta[1]*100
  dfbetas.s2[j]=(beta[2]-beta.v[2])/sqrt(Dbeta.v[2,2])
  prots.s2[j]=(beta[2]-beta.v[2])/beta[2]*100
  dfbetas.s3[j]=(beta[3]-beta.v[3])/sqrt(Dbeta.v[3,3])
  prots.s3[j]=(beta[3]-beta.v[3])/beta[3]*100
  j=j+1}

### Cook
cook=rep(0,5); j=1
mdfbetas=rep(0,5)
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  andm.v$patient=as.factor(as.vector(andm.v$patient))
  mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)
  beta.v=as.matrix(fixef(mm.v))
  Dbeta.v=vcov(mm.v)
  cook[j]=as.numeric(t(beta-beta.v)%*%solve(Dbeta.v)%*(beta-beta.v)/3)
  mdfbetas[j]=as.numeric(t(beta-beta.v)%*%solve(Dbeta.v)%*(beta-beta.v)/3)
  j=j+1}
```

Lisa R. Mudel 4.3: mõjukus juhuslike efektide prognooside suurusele

```
# Cooki kaugus (1)
cooku.v=rep(0,5); j=1
for (i in c(1,4,5,6,11)){
andm.v=andm.kesk[andm.kesk$patient!=i,]
mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)
beta.v=as.matrix(fixef(mm.v))
u.v=raneef(mm.v)$patient;u.v=t(as.matrix(cbind(u.v[1,],u.v[2,],u.v[3,],u.v[4,])))
u.v2=as.matrix(u[-((j-1)*4+1):(j*4) ])
D.u.v=Du[-((j-1)*4+1):(j*4) ],-((j-1)*4+1):(j*4) ]
cooku.v[j]=0.5*t(u.v2-u.v)%*%solve(D.u.v)%*(u.v2-u.v)
j=j+1}

# Cooki kaugus (2)
cooku.dp.v=rep(0,20);j=1
for (i in 1:5){
for (k in 1:4){
andm.v=andm.kesk[-(c((i-1)*12+k,(i-1)*12+k+4,(i-1)*12+k+8)),]
mm.v=lmer(index~s-1+(day-1|patient),data=andm.v,REML=F)
beta.v=as.matrix(fixef(mm.v))
u.v=raneef(mm.v)$patient;u.v=t(as.matrix(cbind(u.v[1,],u.v[2,],u.v[3,],u.v[4,],u.v[5,])))[-j]
u.v2=as.matrix(u[-j])
D.u.v=Du[-j,-j]
cooku.dp.v[j]=0.5*t(u.v2-u.v)%*%solve(D.u.v)%*(u.v2-u.v)
j=j+1}}

### PRESS
press=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
andm.v=andm.kesk[andm.kesk$patient!=i,]
andm.v$patient=as.factor(as.vector(andm.v$patient))
mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)
beta.v=as.matrix(fixef(mm.v))
yi=andm.kesk$index[andm.kesk$patient==i]
xi=X[andm.kesk$patient==i,]
ei=yi-xi)%*%beta.v
press[j]=sum(ei^2)
j=j+1}
```

Lisa S. Mudel 4.3: täpsusele

```
### fikseeritud efektid
det1=det(vcov(mm))
VS=rep(0,5);VStr=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  andm.v$patient=as.factor(as.vector(andm.v$patient))
  mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)
  det2=det(vcov(mm.v))
  VStr[j]=abs(sum(diag( solve(vcov(mm))%*%vcov(mm.v) ))-16)
  VS[j]=abs(det2/det1-1)
  j=j+1}

### juhuslikud efektid
detu1=(G%*%t(Z)%*%V.inv%*%Z%*%G)[1:16,1:16]
VSu=rep(0,5);VSutr=rep(0,5);j=1
for (i in c(1,4,5,6,11)){
  andm.v=andm.kesk[andm.kesk$patient!=i,]
  andm.v$patient=as.factor(as.vector(andm.v$patient))
  mm.v=lmer(index~s-1+(day-1|patient),data=andm.v)

X.v = as.matrix(model.matrix(mm.v))
Z.v=kronecker(diag(rep(1,4)),Zi)

Gi.v=matrix(VarCorr(mm.v)$patient[1:16],nrow=4)
G.v=kronecker(diag(rep(1,4)),Gi.v)
Ri.v=attr(VarCorr(mm.v),"sc")**2*diag(rep(1,12))
Vi.v=Zi%*%Gi.v%*%t(Zi)+Ri.v
V.v=kronecker(diag(rep(1,4)),Vi.v)
V.v.inv=solve(V.v)
det2=G.v%*%t(Z.v)%*%V.v.inv%*%Z.v%*%G.v
VSu[j]=abs((det(det2)*10**(300))/(det(detu1)*10**(300))*-1)
VSutr[j]=abs(sum(diag( solve(detu1)%*%det2 ))-16)
  j=j+1}
```

Lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Anu Tensing,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

"Mõjukate vaatluste analüüs lineaarsetes segamudelites",

mille juhendaja on Mare Vähi,

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

2.2 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, 02.06.2014.