TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
MATEMAATILISE STATISTIKA INSTITUUT

Astrid Haas

Üldistatud lineaarne segamudel ESM-uuringu andmetele

Magistritöö (30 EAP)
Finants- ja kindlustusmatemaatika

Juhendaja: Märt Möls

TARTU 2015

Üldistatud lineaarne segamudel ESM-uuringu andmetele

Sotsiaalmeedias ilmuvad tihti artiklid, mis kurdavad õpilaste kodutööde liigse koormuse üle Eesti koolides. Ka haridus- ja teadusministeerium on probleemiga tuttav ning on aastate jooksul kasutusele võtnud erinevaid lahendusi laste koormatuse vähendamiseks.

Käesoleva magistritöö eesmärk on uurida kui palju Eesti koolilapsed reaalselt kodus õpivad. Eesmärgi saavutamiseks hinnatakse üldistatud lineaarne segamudel, mis kirjeldab kodutöödele kulutatud aja sõltuvust teistest tunnustest. Mudeli hindamiseks kasutatakse ESM-meetodil (*experience sampling method*) kogutud andmeid erinevatest Eesti koolidest.

Märksõnad: *üldistatud lineaarsed mudelid*, *uuringud*, *psühholoogilised uuringud.*

Generalized linear mixed model for ESM survey data

It has become a recurrent theme on social media to highlight the perceived excessive amount of coursework load in the Estonian Primary School curriculum. Estonian Ministry of Education and Science has noticed this theme and has suggested possible solutions to reduce this coursework load.

This motivates the goal of this work which was to study the time spent on coursework by Estonian pupils. We accomplish our goal by investigating and assessing the generalized linear mixed model that describes the changes in the time spent on coursework. Data was collected by ESM method (experience sampling method) from various Estonian schools.

Keywords: *generalized linear models*, *research*, *psychological research.*

Sisukord

Sissejuhatus

Sotsiaalmeedias ilmuvad tihti artiklid, mis kurdavad kodutööde liigse koormuse üle Eesti koolides. Näiteks artiklid „Murelik ema: lapsed põlevad koolikoormuse tõttu läbi juba enne täiskasvanuks saamist" (artikkel [1]) ja „Miks lapsed paksud on? Pidin poja trennist ära võtma, sest õppimise kõrvalt ei jäänud tal selleks enam aega!" (artikkel [2]). Õpetajate ja haridusasutuste tööga lähedalt seotud ametnike poolt on ka avaldatud arvamuslugusid, mis väidavad, et lastel on koolis tööd liiga palju (näiteks artiklid [3] ja [4]).

Juba mitu aastat on haridus- ja teadusministeerium püüdnud vähendada põhikooli õpilaste õppekoormust. Näiteks 2010. aastal vastu võetud põhikooli ja gümnaasiumi seadusega vähendati üheksanda klassi nädalast tunniplaani kahe tunni võrra. Aastal 2014 puhastati riiklike õppekavasid, et vältida teemade kordumisi erinevate ainete õppekavades. Õppekavade puhastamist jätkatakse siiani.

Seega on tarvis uurida, kas lapsed tegelikult ka õpivad nii palju, et neil pole aega millekski muuks?

Tallinna Ülikooli psühholoogia instituudi teadurid viisid 2014. aasta kevadel läbi uuringu, mille eesmärgiks oli selgitada, kui palju kulub Eesti koolilastel aega õppimiseks ja missuguseid emotsioone nad kodus õppides tunnevad. Käesolev magistritöö keskendub uuringu sellele osale, mis otsib vastust küsimusele, kui palju lapsed tegelikult kulutavad aega koduste tööde tegemisele.

Magistritöö katsub leida vastust uuritavale küsimusele läbi üldistatud lineaarse segamudeli. Töö esimene osa on referatiivne, et tutvustada üldistatud lineaarse segamudeli teoreetilist poolt. Ühe lihtsa näite abil tutvustatakse võimalusi hinnata üldistatud lineaarset segamudelit statistika rakendustarkvara R abil. Töö teises osas on antud ülevaade läbi viidud uuringust ning viimases osas kirjeldatatakse leitud mudeleid. Leitud mudelite abil antakse ka ülevaade sellest, kui palju aega lapsed koduste tööde tegemiseks kulutavad ja millest ajakulu sõltub. Näiteks uuritakse kas on erinevusi poiste ja tüdrukute poolt õppimisele pühendatud ajal või kas mõnes koolis õpitakse võrreldes teiste koolidega rohkem.

Autor tänab Tallinna Ülikooli psühholoogia instituudi teadureid, kes viisid läbi ESM-uuringu ning lubasid oma uuringu tulemusi kasutada. Lisaks autor tänab ka juhendajat Märt Mölsi, kes andis palju täiendavaid mõtteid ning aitas loendamatul hulgal parandusi teha.

Peatükk 1

Üldistatud lineaarne mudel

Esimene peatükk annab referatiivse ülevaate üldistatud lineaarsest mudelist. Peatüki sissejuhatuse kirjutamiseks on aluseks allikaid [5] ja [6].

Klassikaliste mudelite üks eeldus on, et uuritav tunnus on normaaljaotusega. Kui andmed ei ole tegelikult normaaljaotusega võivad mudeli abil teostatavad testid kaotada võimsuses (võrreldes korrektset jaotust kasutavate testidega), saadud hinnangud võivad osutuda ebatäpseteks ning testid võivad liiga tihti teha esimest liiki viga. Üheks võimaluseks, kuidas seda lahendada on kasutada üldistatud lineaarset mudelit, kus kasutame andmetele sobivat eksponentsiaalsete jaotuste perest pärit jaotust ega proovigi seda teisendada normaaljaotusele lähedaseks.

Üldistatud lineaarse mudeli tähtsamad eeldused on, et uuritava juhusliku suuruse keskväärtus sõltub kirjeldavatest tunnustest ning et uuritava juhusliku suuruse jaotus oleks eksponentsiaalsete jaotuste perest. Otsitava juhusliku suuruse dispersioon võib sõltuda tema keskväärtusest. Üldistatud lineaarne mudel kasutab seosefunktsiooni, sidumaks jaotuse keskväärtust argumentide lineaarkombinatsiooniga.

Juhusliku suuruse y üldistatud lineaarne mudel, kui on teada tunnuse x väärtused, on kujul:

$$\begin{aligned} E(y) &= \mu \\ &= g^{-1}(\beta_0 + \beta_1 x), \end{aligned}$$

kus $E(y)$ on juhusliku suuruse y keskväärtus, x on teadaoleva kirjeldava tunnuse väärtus, β_0 on vabaliige, β_1 on parameeter, mis iseloomustab seost uuritava tunnuse y ja tunnuse x vahel ja $g()$ on seosefunktsioon.

Enamasti nii lihtsaid mudeleid aga ei otsita. Kui uuritavaks tunnuseks on juhuslike suuruste vektor $\boldsymbol{y}$, siis on üldistatud lineaarne mudel kujul:

$$\begin{aligned} E(\boldsymbol{y}) &= \boldsymbol{\mu} \\ &= \boldsymbol{g}^{-1}(\mathbf{X}\boldsymbol{\beta}), \end{aligned}$$

kus $E(\boldsymbol{y})$ on juhuslike suuruste vektori $\boldsymbol{y}$ keskväärtuste vektor, $\mathbf{X}$ on kirjeldavatest tunnustest moodustatud maatriks, $\boldsymbol{\beta}$ on tundmatute parameetrite vektor ja vektorfunktsioon $\boldsymbol{g}()$ on seosefunktsioon, mis rakendab igale argumendi elemendile funktsiooni $g()$.

Kokkuvõtvalt, et kasutada üldistatud lineaarset mudelit peame teadma:

1. eksponentsiaalsete jaotuste perest pärit uuritava tunnuse täpsemat jaotuste peret (näiteks normaaljaotus, Poissoni jaotus, binoomjaotus, ...),

2. seosefunktsiooni $\boldsymbol{g}()$,

3. kirjeldavatest tunnustest moodustatud mudelimaatriksit $\mathbf{X}$, mis on seotud tundmatute parameetrite vektoriga $\boldsymbol{\beta}$ nii, et kehtiks $E(\boldsymbol{y}) = \boldsymbol{g}^{-1}(\mathbf{X}\boldsymbol{\beta})$.

1.1 Üldistatud lineaarne segamudel

Antud alampeatükk on refereeritud allikast [7].

Üldistatud lineaarsetes mudelites on mudelimaatriks $\mathbf{X}$ moodustatud, kas kirjeldavate tunnuste (fikseeritud mõjud) või juhuslike mõjude poolt. Missugune on erinevus fikseeritud ja juhuslikel mõjudel? Näiteks, kui tehakse küsitlus erinevates Eesti koolides ning soovitakse hinnata koolide mõju laste õpiharjumustele. Kui mudelisse panna kool fikseeritud kujul, siis me hindame selles uuringus osalenud koolide mõjusid ning võime kirjeldada uuringus osalenud koolides käivate laste õpikogemusi. Kuid uuringus mitteosalenud õpilaste õpiharjumuste kohta ei saa sellise mudeli abil mingeid järeldusi teha. Kui aga koolid on valitud juhuslikult kõigi Eesti koolide seast, siis võime käsitleda kooli mõju juhuslikuna. Sellisel juhul tekib võimalus üldistada mudeli tulemusi kõigile Eesti koolidele. Kooli juhuslik mõju tähendab seda, et kõik antud kooli lapsed õpivad kodus keskmiselt rohkem või vähem. Kui üldistatud lineaarses mudelis on kasutatud nii fikseeritud mõjuga tunnuseid kui ka juhusliku mõjuga tunnuseid, siis on tegemist üldistatud lineaarse segamudeliga.

Algselt on üldistatud lineaarne mudel kujul $E(\boldsymbol{y}) = \boldsymbol{g}^{-1}(\mathbf{X}\boldsymbol{\beta})$, kus $\boldsymbol{\beta}$ on fikseeritud mõjude parameetrite vektor. Selleks, et saada segamudelit, lisame sinna juurde juhuslike mõjude $\mathbf{u}$ mõju.

Saame mudeli, kus me hindame vektori $\boldsymbol{y}$ tingliku keskväärtust, tingimusel, et on fikseeritud juhuslike mõjude vektor $\mathbf{u}$. Mudeli kuju on

$$E(\boldsymbol{y}|\mathbf{u}) = \boldsymbol{g}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}),$$

kus $\mathbf{Z}$ on teadaolev juhuslike mõjude mudelimaatriks ja $\mathbf{u}$ on juhuslike mõjude parameetrite vektor ning $\mathbf{X}$ on sõltumatute tunnuste poolt moodustatud mudelimaatriks ja $\boldsymbol{\beta}$ on fikseeritud mõjude parameetrite vektor ning $\boldsymbol{g}()$ seosefunktsioon.

Vaatame ühte näidet, et paremini mõista, missugusel kujul on maatriksid $\mathbf{X}$ ja $\mathbf{Z}$. Olgu meil kahest juhuslikult valitud koolist küsitletud kokku nelja last. Laste käest küsiti, kui kaugel oli nende kodu koolist, mis koolis käivad ja aasta jooksul kogutud viite arv. Soovime hinnata kooli ja õpilase kodu vahelise kauguse ning kooli mõju viite saamisele. Fikseeritud mõjuna käsitleme kaugust kodu ja kooli vahel ning olgu kooli mõju juhuslik (tegemist on juhuslikult valitud koolidega). Seosefunktsiooniks kasutame log-seosefunktsiooni.

Mudel koolis j käiva õpilase i korral on kujul:

$$E[(\mathit{Viite\ arv})_i|u_j] = \exp\{\beta_0 + \beta_1 \cdot \mathit{kaugus}_i + u_j + \epsilon_i\},$$

kus u_j on kooli j juhuslik mõju, β_0 on vabaliige ning β_1 tunnuse *kauguse* mõju kirjeldav parameeter. Praegusel juhul on kaks kooli, järelikult $j = \{1, 2\}$ ning kuna küsitleti nelja last, siis $i = \{1, 2, 3, 4\}$. Sama mudeli esitamiseks maatrikskujul defineerime esmalt vektorid $\boldsymbol{\beta}$ ja $\mathbf{u}$:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Mudelimaatriksite $\mathbf{X}$ ja $\mathbf{Z}$ kujud on aga vastavalt:

$$\mathbf{X} = \begin{pmatrix} 1 & Kaugus_1 \\ 1 & Kaugus_2 \\ 1 & Kaugus_3 \\ 1 & Kaugus_4 \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Soovi korral saame kombineerida mudeli maatriksid $\mathbf{X}$ ja $\mathbf{Z}$ üheks maatriksiks: $\mathbf{X}^*$=[$\mathbf{X}$ $\mathbf{Z}$] ning sama moodi tundmatute parameetrite vektorid: $\boldsymbol{\beta}^* = [\boldsymbol{\beta}^T\ \mathbf{u}^T]^T$, kus T tähistab transponeerimist.

$$\boldsymbol{\beta}^* = \begin{pmatrix} \beta_0 \\ \beta_1 \\ u_1 \\ u_2 \end{pmatrix}, \mathbf{X}^* = \begin{pmatrix} 1 & Kaugus_1 & 1 & 0 \\ 1 & Kaugus_2 & 1 & 0 \\ 1 & Kaugus_3 & 0 & 1 \\ 1 & Kaugus_4 & 0 & 1 \end{pmatrix}.$$

See tähendab, et me saame kirjutada segamudeli kujule: '

$$E(\boldsymbol{y}|\mathbf{u}) = \boldsymbol{g}^{-1}(\mathbf{X}^*\boldsymbol{\beta}^*)$$

. Seega ka üldistatud lineaarse segamudeli poolt otsitav tinglik keskväärtus on läbi seosefunktsiooni lineaarne argumenttunnuste suhtes. Üldistatud lineaarset segamudelit kasutades on võimalik hinnata mudel, mille tulemusi saab üldistada laiemale populatsioonile, vaid fikseeritud mõjudega mudeli korral oleksime saanud teha järeldusi vaid uuringus osalenud koolide kohta.

Segamudeli struktuur

Segamudeli otsitava tunnuse vektor $\boldsymbol{y}$ koosneb tinglikult sõltumatutest elementidest, mis on pärit eksponentsiaalsete jaotuste pere jaotusest (mille kirjutame välja kanoonilisel kujul):

$$y_i|\mathbf{u} \sim f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}),\ (y_i \perp y_j)|\mathbf{u} \text{ kui } i \neq j,$$

$$f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}) = \exp\left\{[y_i\gamma_i - b(\gamma_i)]/\tau^2 - c(y_i,\tau)\right\}. \tag{1.1}$$

Tahame leida $f_{Y_i}(y_i)$ marginaalset jaotust. Tinglik keskväärtus avaldub elemendiviisiliselt kujul:

$$\begin{aligned} E(y_i|\mathbf{u}) &= \mu_i \\ g(\mu_i) &= \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}, \end{aligned}$$

kus $g()$ on seosefunktsioon, $\mathbf{x}_i^T$ on mudelimaatriksi $\mathbf{X}$ rida i ning $\boldsymbol{\beta}$ on fikseeritud mõjude parameetrite vektor, $\mathbf{z}_i^T$ on mudelimaatriksi $\mathbf{Z}$ rida i ning $\mathbf{u}$ on juhuslike mõjude parameetrite vektor. Pöörame tähelepanu sellele, et me kasutame μ_i praegusel juhul y_i tingliku keskväärtuse märkimiseks, mitte marginaalse jaotuse keskväärtusena. Tinglik dispersioon on seotud tingliku keskväärtusega μ_i läbi seose

$$D(y_i|\mathbf{u}) = \tau^2 \upsilon(\mu_i), \tag{1.2}$$

kus $\upsilon(\mu_i) := \dfrac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2}$. See seos tuleneb eksponentsiaalse pere jaotusest (1.1) ning skoorifunktsiooni omadustest

$$E\Big[\frac{\partial \log f_{Y_i}(y_i)}{\partial \gamma_i}\Big] = 0 \text{ ja } D\Big(\frac{\partial \log f_{Y_i}(y_i)}{\partial \gamma_i}\Big) = -E\Big[\frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \gamma_i^2}\Big].$$

Määrame juhuslike mõjude jaotuse:

$$\mathbf{u} \sim f_{\mathbf{U}}(\mathbf{u}).$$

Nüüd tuletame $f_{Y_i}(y_i)$ marginaaljaotuse.

Keskväärtus

Me märgime praegu tingliku keskväärtust kui μ_i. Kasutades keskväärtuse omadust $EA = E[E(A|B)]$ võime leida y_i keskväärtuse:

$$\begin{aligned} E[y_i] &= E[E[y_i|\mathbf{u}]] \\ &= E[\mu_i] \\ &= E[g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u})] \end{aligned} \tag{1.3}$$

Üldiselt tulemuseks saadud avaldist ei ole võimalik lihtsustada, kuna $g^{-1}()$ on mittelineaarne funktsioon.

Dispersioon

Keskväärtuse omadus $EA = E[E(A|B)]$ ja dispersiooni omadus $D(A|B) = E(A^2|B) - [E(A|B)]^2$ annavad meile $E[D(A|B)] = D(A) - D[E(A|B)]$. Viimasest võrdusest avaldame D(A) ja kasutame seda leidmaks y_i dispersiooni:

$$\begin{aligned} D(y_i) &= D(E[y_i|\mathbf{u}])) + E[D(y_i|\mathbf{u})] \\ &= D(\mu_i) + E[\tau^2 \upsilon(\mu_i)] \\ &= D(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u})) + E[\tau^2 \upsilon(g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}))], \end{aligned} \tag{1.4}$$

Teine võrdus kehtib, kuna $\mathbf{u}$ korral on $\tau^2 \upsilon(\mu_i)$ seose (1.2) põhjal tinglik dispersioon, kui tegemist on eksponentsiaalse pere jaotusega.

Nii nagu y_i keskväärtusele vastavat avaldist ei saa lihtsustada, ei ole võimalik ka dispersiooni valemit lihtsustada, kui ei tee eeldusi $g^{-1}()$ kohta.

Kovariatsioon

Juhusliku mõjude kasutamine toob endaga kaasa ka korrelatsiooni vaatluste vahel, millel on mõni ühine juhuslik mõju. Korrelatsiooni leidmiseks on vajalik kovariatsiooni teadmine $\mathrm{cor}(y_i, y_j) = \dfrac{\mathrm{cov}(y_i, y_j)}{\sqrt{D(y_i)D(y_j)}}$. Eeldades $\boldsymbol{y}$ elementide tingliku sõltumatust saame kovariatsiooniks:

$$\begin{aligned}\mathrm{cov}(y_i, y_j) &= \mathrm{cov}(E[y_i|\mathbf{u}], E[y_j|\mathbf{u}]) + E[\mathrm{cov}(y_i, y_j|\mathbf{u})] \\ &= \mathrm{cov}(\mu_i, \mu_j) + E[0] \\ &= \mathrm{cov}\big(g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}), g^{-1}(\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{z}_j^T\mathbf{u})\big).\end{aligned} \tag{1.5}$$

Keskväärtuse kui ka dispersiooni leidmine on raskendatud ilma eeldusteta $g^{-1}()$ kohta. Vaatame täpsemalt olukorda, kus me teame, et seosefunktsioon $g()$ on log-seosefunktsioon.

1.2 Logaritmiline seos

See osa on kirjutatud kasutades kirjandust [7] ja [8].

Mis juhtub $\boldsymbol{y}$ marginaalse jaotusega, kui võtta seosefunktsiooniks log-seosefunktsiooni ehk $g(x) = \log x$ ja $g^{-1}(x) = \exp x$ ning millal ja miks on seda seosefunktsiooni kasulik rakendada?

Log-seosefunktsiooni üks tähtsamaid omadusi on see, et keskväärtus on alati positiivne. Seda seosefunktsiooni kasutatakse selliste andmete puhul, kus on eelnevalt teada, et uuritava tunnus väärtus ei saa olla negatiivne. Kui uuritavaks tunnuseks on näiteks mingit teelõiku ajaühikus läbinud autode arv või õunapuu otsas kasvavate õunte arv. Nende tunnuste väärtused ei saa olla negatiivsed, kuna ei ole võimalik, et teelõiku läbiks -3 autot või puu otsas kasvab -20 õuna. Mittenegatiivse juhusliku suuruse keskväärtus ei saa aga olla negatiivne.

Kasutades nüüd $g^{-1}(x) = \exp x$ arendame edasi juba leitud marginaalse jaotuse keskväärtuse valemit (1.3):

$$\begin{aligned}E[y_i] &= E[g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u})] \\ &= E[\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}] \\ &= \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\} E[\exp\left\{\mathbf{z}_i^T\mathbf{u}\right\}] \\ &= \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\} M_u(\mathbf{z}_i),\end{aligned}$$

kus $M_u(\mathbf{z}_i) = E[e^{\mathbf{z}_i^T\mathbf{u}}]$ on $\mathbf{u}$ momente genereeriv funktsioon. Eeldame, et $u_i \sim \mathcal{N}(0,\Sigma)$, siis $M_u(\mathbf{z}_i) = \exp\left\{\frac{\Sigma}{2}\right\}$. Järelikult

$$E[y_i] = \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\}\exp\left\{\frac{\Sigma}{2}\right\} = \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \frac{\Sigma}{2}\right\}.$$

Kui $y_i|\mathbf{u}$ on Poissoni jaotusega, siis Poissoni jaotuse omaduste tõttu on y_i tinglik dispersioon $D(y_i|\mathbf{u}) = \mu_i$. Järelikult seose (1.2) põhjal $\tau^2\nu(\mu_i) = \mu_i$. Eeldame nüüd, et meil on $y_i|\mathbf{u}$ Poissoni jaotusest, teame et $\boldsymbol{y}$ elemendid on sõltumatud. Seega kasutades (1.4) saame dispersiooniks:

$$\begin{aligned}
D(y_i) &= D(g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u})) + E[\tau^2\upsilon(g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}))] \\
&= D(\mu_i) + E[\mu_i] \\
&= D(\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}) + E[\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}] \\
&= E[\exp\left\{2(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u})\right\}] - [E[\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}]]^2 + \\
&\quad + E[\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}] \\
&= \exp\left\{2\mathbf{x}_i^T\boldsymbol{\beta}\right\}\Big(M_u(2\mathbf{z}_i) - [M_u(\mathbf{z}_i)]^2 + \exp\left\{-\mathbf{x}_i^T\boldsymbol{\beta}\right\}M_u(\mathbf{z}_i)\Big),
\end{aligned}$$

Eeldame, et $u_i \sim \mathcal{N}(0,\Sigma)$, siis $M_u(\mathbf{z}_i) = \exp\left\{\frac{\Sigma}{2}\right\}$. Järelikult

$$\begin{aligned}
D[y_i] &= \exp\left\{2\mathbf{x}_i^T\boldsymbol{\beta}\right\}\Big(\exp\left\{2\Sigma\right\} - \exp\left\{\Sigma\right\}\Big) + \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\}\exp\left\{\frac{\Sigma}{2}\right\} \\
&= \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \frac{\Sigma}{2}\right\}\Big(\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\}\Big[\exp\left\{\frac{3\Sigma}{2}\right\} - \exp\left\{\frac{\Sigma}{2}\right\}\Big] + 1\Big) \\
&= E[y_i]\Big(\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta}\right\}\Big[\exp\left\{\frac{3\Sigma}{2}\right\} - \exp\left\{\frac{\Sigma}{2}\right\}\Big] + 1\Big).
\end{aligned}$$

Eelmises alampeatükis leidsime ka üldkuju kovariatsioonile ning seega kui kasutame log-seosefunktsiooni, siis kovariatsioon saab kuju:

$$\begin{aligned}
\mathrm{cov}(y_i, y_j) &= \mathrm{cov}(g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}), g^{-1}(\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{z}_j^T\mathbf{u})) \\
&= \mathrm{cov}(\exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u}\right\}, \exp\left\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{z}_j^T\mathbf{u}\right\}) \\
&= \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{x}_j^T\boldsymbol{\beta}\right\}\mathrm{cov}(\exp\left\{\mathbf{z}_i^T\mathbf{u}\right\}, \exp\left\{\mathbf{z}_j^T\mathbf{u}\right\}) \\
&= \exp\left\{\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{x}_j^T\boldsymbol{\beta}\right\}[M_u(\mathbf{z}_i + \mathbf{z}_j) - M_u(\mathbf{z}_i)M_u(\mathbf{z}_j)],
\end{aligned}$$

kus $M_u(\mathbf{z}_i) = E[e^{\mathbf{z}_i\mathbf{u}}]$ ja $M_u(\mathbf{z}_i + \mathbf{z}_j) = E[e^{(\mathbf{z}_i+\mathbf{z}_j)\mathbf{u}}]$.

Ülehajuvus

Ülehajuvuse teema on referatiivne ning on kasutatud allikaid [5] ja [8].

Poissoni jaotusel on omadus, et keskväärtus ja dispersioon on võrdsed ehk $D(y) = E(y)$. Kui see võrdus aga andmetes ei kehti on tegemist, kas alahajuvuse $(D(y) < E(y))$ või ülehajuvusega $(D(y) > E(y))$.

Ülehajuvus võib näiteks tekkida järgnevatel põhjustel:

- Andmestikus on erindid.
- Andmetes, kus on mudeliga kirjeldatud hajuvus, on lisaks mingi muu juhuslikkus (näiteks mõõteviga) ehk teatud osa hajuvusest on kirjeldatud tunnuste kaudu ja mingi osa kirjeldamata.
- Mudeli eeldused ei ole täidetud.
- Ühes grupis asuvate objektide keskväärtused ja dispersioonid ei pruugi olla homogeensed. Gruppidel on küll samad keskväärtused ja dispersioonid, aga grupis olevatel liikmetel on omavahel erinevad keskväärtused ja dispersioonid.
- Kui on tegemist korduvmõõtmistega, siis iga objekt võib alustada küll sama keskväärtuse parameetriga, kuid see võib mõõtmisaja jooksul muutuda näiteks eelnevate sündmuste mõjul.
- Kui uuritav tunnus on mõne muu jaotusega, mitte Poissoni jaotusega.

Ülehajuvuse korral kasutatakse kvaasitõepära meetodit. Kvaasitõepära arvestab dispersiooni kuju kasutades eksponentsiaalse pere jaotuse skaalaparameetrit. Parameetrite vektori $\boldsymbol{\beta}$ hinnang $\hat{\boldsymbol{\beta}}$ ei sõltu skaalaparameetrist φ, seega saame kvaasitõepära meetodil samad hinnangud parameetervektorile β nagu ülehajuvust arvestamata, aga $\mathrm{cov}(\hat{\boldsymbol{\beta}})$ on proportsionaalne skaalaparameetriga φ ja seega muutub kordajate hinnangute hajuvuse hinnang ning võib muutuda kordajate olulisus mudelis.

Selleks, et selgitada kas andmetes on ülehajuvus hinnatakse skaalaparameeter φ seosest $D(y) = \varphi E(y)$. Selleks, et ei oleks mingisugust üle- või alahajuvust andmetes peab kehtima $\hat{\varphi} \approx 1$. Ülehajuvuse kontrollimiseks saab hinnata skaalaparameetri φ, kas:

$$\text{hälbimuse } \hat{\varphi} = \frac{D}{df}, \text{ või Pearsoni statistiku abil: } \hat{\varphi} = \frac{\chi^2}{df}.$$

1.3 Parameetrite hindamine suurima tõepära meetodiga

See alampeatükk on refereeritud kirjandusest [7].

Parameetrite hindamiseks on mitmeid erinevaid meetodeid. Üks levinum parameetrite hindamiseks on suurima tõepära meetod. Kuidas on võimalik anda hinnang vektorile $\boldsymbol{\beta}$ üldistatud lineaarses segamudelis kasutades suurima tõepära meetodit?

Alustame sellest, et kirjutame välja üldistatud lineaarses segamudelis uuritava tunnuse y jaotuse tihedusfunktsiooni:

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \int f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}.$$

Ülaltoodud valemis integreeritakse üle iga vektori $\mathbf{u}$ elemendi. Logaritmiline tõepärafunktsioon on kujul:

$$l = \log f_{\mathbf{Y}}(\boldsymbol{y}) = \log \int \Big(f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) \Big) d\mathbf{u}.$$

Leidmaks hinnanguid fikseeritud mõjudele, võtame tuletise logaritmilisest tõepärafunktsioonist β järgi:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\partial \log f_{\mathbf{Y}}(\boldsymbol{y})}{\partial \boldsymbol{\beta}} \\ &= \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \int \Big(f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) \Big) d\mathbf{u}}{f_{\mathbf{Y}}(\boldsymbol{y})} \\ &= \frac{\int \Big[\frac{\partial}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) \Big] f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}}{f_{\mathbf{Y}}(\boldsymbol{y})}. \end{aligned} \tag{1.6}$$

Viimane võrdus kehtib kuna $f_{\mathbf{U}}(\mathbf{u})$ ei sõltu parameetrist $\boldsymbol{\beta}$. Seetõttu saame:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) &= \Big(\frac{1}{f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u})} \frac{\partial f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} \Big) f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) \\ &= \frac{\partial \log f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}). \end{aligned}$$

Nüüd saame (1.6) ümber kirjutada kujule:

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\int \frac{\partial \log f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}}{f_{\mathbf{Y}}(\boldsymbol{y})} \\
&= \int \frac{\partial \log f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u})}{\partial \boldsymbol{\beta}} f_{\mathbf{Y}|\mathbf{u}}(\boldsymbol{y}|\mathbf{u}) f_{\mathbf{U}|\boldsymbol{y}}(\mathbf{u}|\boldsymbol{y}) d\mathbf{u}.
\end{aligned}
$$

Võrdsustades tuletised nullidega saame võrrandsüsteemi, mille lahendid on parameetrite suurima tõepära hinnangud. Praktikas lahendatakse seda võrrandsüsteemi erinevate numbriliste meetoditega.

1.4 Näide

Olgu meil uuritavaks tunnuseks Y, mis on mõõdetud kahel lapsel kaks korda. Lapse indikaator olgu tähistatud tunnuses Z ning lisaks on teada tunnuse X väärtuse nelja mõõtmise jaoks. Eeldame, et tunnuste Y ja X vahelist seost kirjeldab mudel:

$$E(y_{ij}|laps_i) = \exp\left\{\beta_0 + (\beta_1 + u_i)x_{ij}\right\},$$

kus on tegemist i-nda lapse j-inda vaatluse keskväärtusega, kus u_i kirjeldab kui palju erineb selle lapse tunnuse X ees olev kordaja kogu populatsiooni tunnuse X ees olevast kordajast.

Ülalpool toodud mudelile vastav mudelimaatriks $\mathbf{X}$ on kujul:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{12} \end{pmatrix}, \text{ kus } x_{11}, x_{21}, x_{22}, x_{12} \text{ on tunnuse } X \text{ väärtused.}$$

Kui tavaliselt on mudelimaatriks $\mathbf{Z}$ nullide ja ühtede maatriks, siis meie ülesande püsituse kohaselt on mudelimaatriksis $\mathbf{Z}$ ühtede asemel tunnuse X väärtused. Mudelimaatriks $\mathbf{Z}$ on kujul:

$$\mathbf{Z} = \begin{pmatrix} x_{11} & 0 \\ 0 & x_{21} \\ 0 & x_{22} \\ x_{12} & 0 \end{pmatrix}, \text{ kus } x_{11}, x_{21}, x_{22}, x_{12} \text{ on tunnuse } X \text{ väärtused.}$$

Eeldame, et tunnuse Y tinglik jaotus (tingimusel, et on fikseeritud $\mathbf{u}$) on Poissoni jaotus. Mudeli maatrikskuju on:

$$E(\boldsymbol{y}|\mathbf{u}) = \exp\{\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}\}.$$

Me tahame hinnata vektorit $\boldsymbol{\beta}$. Selleks kasutame suurima tõepära meetodit, mida eelnevas alampeatükis tutvustasime. Eeldame, et $\mathbf{u}$ on normaaljaotusest keskväärtusega null ja dispersiooniga σ^2.

Nüüd kirjutame välja tõepärafunktsiooni. Edasi maksimiseerime logaritmilise tõepärafunktsiooni, et saada parameetrite hinnangud. Tõepärafunktsioon kirjeldatud neljale vaatlusele on:

$$\begin{aligned}
L &= \int \Big(f(\boldsymbol{y}|\mathbf{u}) \cdot f(\mathbf{u}) \Big) d\mathbf{u} \\
&= \int \Big(f(y_{11}|u_1, u_2) \cdot f(y_{12}|u_1, u_2) \cdot f(y_{21}|u_1, u_2) \cdot f(y_{22}|u_1, u_2) \cdot \\
&\quad \cdot f(u_1) \cdot f(u_2) \Big) d\mathbf{u} \\
&= \int \Big(f(y_{11}|u_1) \cdot f(y_{12}|u_1) \cdot f(y_{21}|u_2) \cdot f(y_{22}|u_2) \cdot f(u_1) \cdot f(u_2) \Big) d\mathbf{u} \\
&= \int\int \Big(f(y_{11}|u_1) \cdot f(y_{12}|u_1) \cdot f(y_{21}|u_2) \cdot f(y_{22}|u_2) \cdot f(u_1) \cdot f(u_2) \Big) du_1 du_2.
\end{aligned}$$

Eeldasime, et uuritava tunnuse Y tinglik jaotus fikseeritud $\mathbf{u}$ korral on Poissoni jaotusest ning, et $\mathbf{u} \sim \mathcal{N}(0, \sigma^2)$, seega tõepärafunktsioon konkreetselt nendele andmetele on kujul:

$$\begin{aligned}
L &= \int\int \Big(\frac{\lambda_{11}^{y_{11}}}{y_{11}!} e^{-\lambda_{11}} \cdot \frac{\lambda_{12}^{y_{12}}}{y_{12}!} e^{-\lambda_{12}} \cdot \frac{\lambda_{21}^{y_{21}}}{y_{21}!} e^{-\lambda_{21}} \cdot \frac{\lambda_{22}^{y_{22}}}{y_{22}!} e^{-\lambda_{22}} \cdot \\
&\quad \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-u_1^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-u_2^2}{2\sigma^2}} \Big) du_1 du_2 \\
&= \int\int \Big(\prod_{i=1}^{2} \prod_{j=1}^{2} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-u_j^2}{2\sigma^2}} \Big) du_1 du_2,
\end{aligned}$$

kus $\lambda_{ij} = \exp\{\beta_0 + \beta_1 x_{ij} + u_j x_{ij}\}$. Log-tõepära, mida me maksimiseerime $\hat{\boldsymbol{\beta}}$ ja $\hat{\sigma}$ leidmiseks, on:

$$\begin{aligned}
l &= \log L \\
&= \log\Big(\int\int \Big(f(y_{11}|u_1) f(y_{12}|u_1) f(y_{21}|u_2) f(y_{22}|u_2) f(u_1) f(u_2) \Big) du_1 du_2 \Big) \\
&= \log\Big[\int\int \Big(\prod_{i=1}^{2} \prod_{j=1}^{2} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-u_j^2}{2\sigma^2}} \Big) du_1 du_2 \Big].
\end{aligned}$$

Mudeli parameetrite hinnangute saamiseks võtame kasutusele genereeritud andmed tunnuste Y ja X kohta ning need on välja toodud tabelis 1.1.

Tabel 1.1: Näide

Y	X	Z
24	1,47	1
51	1,66	2
13	0,38	2
16	0,71	1

Väljakirjutatud logaritmilise tõepärafunktsiooni maksimiseerimiseks kasutame R-i sisseehitatud käsku *optim()*, mis annab meie otsitavatele parameetritele hinnangud (vt. lisa A.1):

$$\hat{\beta}_0 = 2,17, \hat{\beta}_1 = 0,91, \hat{\sigma} = 0,14.$$

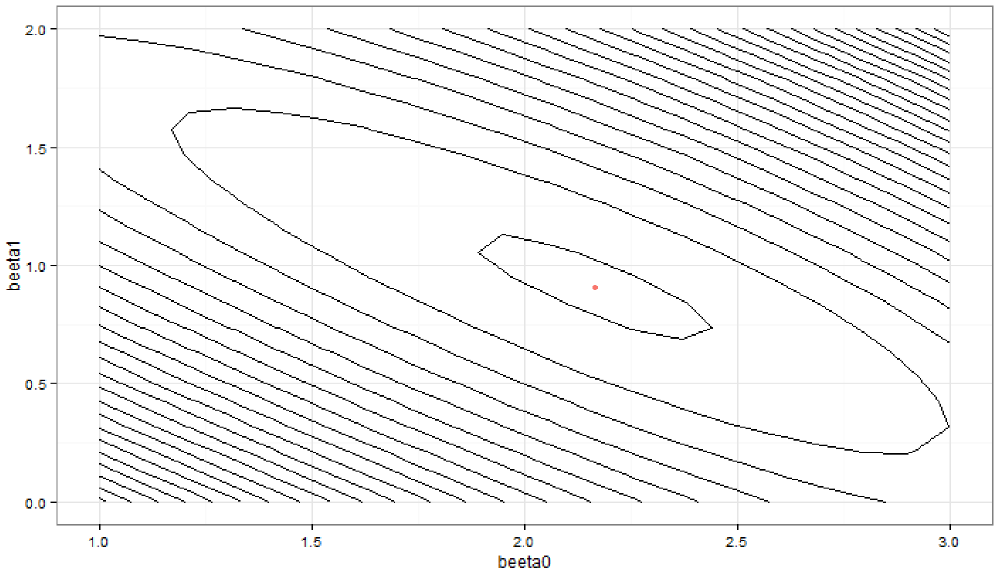
Iga lapse tingliku keskväärtuse hinnangud avalduksid kujul:

$$\widehat{E[y_{ij}|u_j]} = \exp\{2,17 + (0,91 + u_j) \cdot x_{ij}\}.$$

Lapse spetsiifilistest juhuslikest mõjudest on pooled suuremad ja pooled väiksemad nullist (eeldasime, et mõjude keskväärtus on null $Eu_i = 0$). Seetõttu tunnuste Y ja X vahelist seost kirjeldavate joonte mediaani hinnnang kogu populatsioonile on kirja pandud valemiga:

$$\widehat{\text{median}}(E[y_{ij}|x_{ij}]) = \exp\{2,17 + 0,91 \cdot x_{ij}\}.$$

Joonisel (1.1) on fikseeritud σ korral, kuid erinevate $\hat{\boldsymbol{\beta}}$ väärtuste puhul arvutatud logaritmilise tõepärafunktsiooni väärtus kohal $\hat{\boldsymbol{\beta}}$. Praegu on valitud $\sigma = 0,14$. Graafikule kantud samakõrgusjooned (ovaalid) tähistavad $\hat{\beta}$ väärtuseid, mille puhul logaritmilise tõepärafunktsiooni väärtus on sama. Jooniselt on näha, et eksisteerib logaritmiliste tõepärafunktsioonide seas maksimumkoht. Punase täpiga on tähistatud R-iga leitud maksimum.



Joonis 1.1: Maksimiseerimisülesanne $\sigma = 0,14$ korral

Suurima tõepära meetodil üldistatud lineaarse segamudeli parameetrite hinnangute leidmise aitab lihtsamaks muuta R-i lisamoodul glmmADMB.

GlmmADMB pakett

Üldistatud lineaarse segamudeli hindamiseks on mitmeid erinevaid pakette rakendustarkvaras R, kuid praegusel hetkel peaks neist kõige kiiremini ja kõige paremini oma töö ära tegema pakett glmmADMB. Lühend ADMB on tulnud sõnadest *Automatic Differentiation Model Builder*. Pakett glmmADMB hindab mudeli parameetreid kasutades automaatset tuletiste (AD) võtmise meetodit. Meetod AD hindab funktsiooni tuletist ning on täpsem kui numbriliste meetoditega tuletiste leidmine. Jaotuse marginaalse tihedusfunktsiooni leidmisel on vajalik leida integraal ning pakett glmmADMB kasutab selleks Laplace'i lähendust. (Allikas [9])

Kuigi pakett glmmADMB kasutab parameetrite hindamiseks teistsugustel numbrilistel meetoditel põhinevat lähenemist, kui muudes lisamoodulites olevad funktsioonid (nt nlme või lme4), on tema kasutamine sarnane üldistatud lineaarsete segamudelite hindamiseks mõeldud funktsioonidele R-s.

Hindame nüüd tabelis 1.1 toodud andmete pealt näites 1.4 tutvustatud mudel kasutades glmmADMB funktsiooni. Mudeli hindamiseks tuleb R-s anda käsk:

```
glmmadmb(formula=y~x+(x-1|z),data=naidis, family="poisson")
```

Nii juhuslikud kui ka fikseeritud mõjud määratakse ära mudeli kujus, lisaparameetri „formula=" abil. Vasakul pool „∼" märki on sõltuv tunnus ning paremal pool on tunnused, millest sõltuv tunnus sõltub. Kõik fikseeritud mõjud kirjutatakse välja ning eraldatakse plussiga. Praeguse juhul on ainult üks fikseeritud mõju, tunnus X. Juhuslikud mõjud lisatakse samuti plussi abil mudelile. Juhuslikud mõjud kirjutatakse sulgude sisse, kus vasakul pool püstkriipsu on need tunnused, mille eesolevad kordajad on juhuslikult sõltuvad valimisse sattunud lastest ning paremal pool püstkriipsu on lapse identifikaator. Näites me ei soovinud, et juhuslikus mõjud muudaksid vabaliiget, vaid et nad mõjutaksid ainult tunnuse X ees olevat kordajat. Seetõttu oli vaja eemaldada juhuslik vabaliige, mis muidu vaikimisi mudelisse lisatakse. Juhusliku vabaliikme eeldamiseks on juhusliku mõju juures $x-1$. Selleks, et mudel töötaks, tuleb meil määrata ära ka sõltuva tunnuse tinglik jaotus. Kuna me eeldasime, et Y tinglik jaotus on Poissoni jaotus, siis „family=poisson". Ülaltoodud käsu poolt genereeritud väljund asub lisas A.2.

Meil olid andmed kahe lapse kohta ning konkreetselt nende tinglike keskväärtuste hinnangud avalduvad järgmisel:

$$E[\widehat{y_{i1}|laps}_1] = \exp\{2,17 + (0,91 - 0,11) \cdot x_{i1}\} = \exp\{2,17 + 0,81 \cdot x_{i1}\}$$

ja

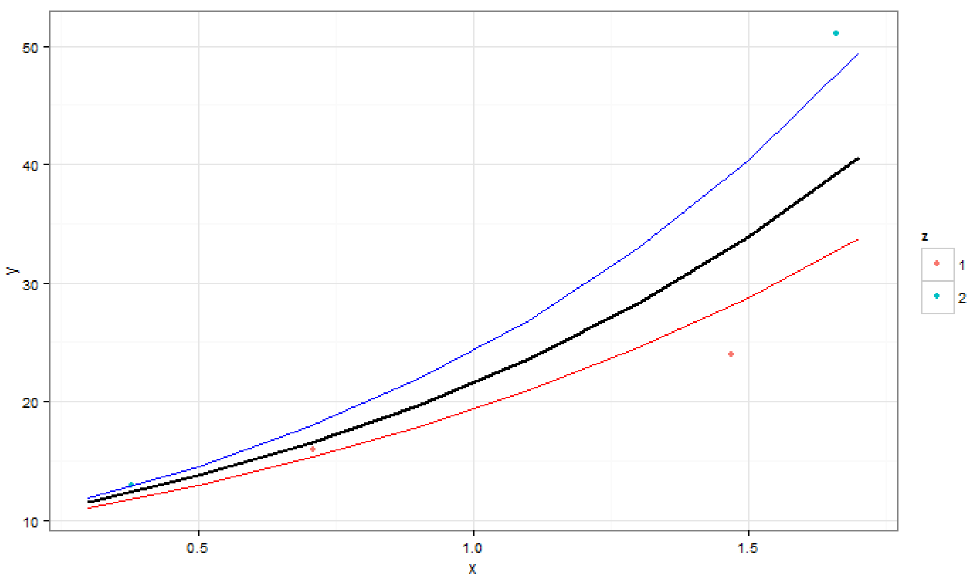
$$E[\widehat{y_{i2}|laps}_2] = \exp\{2,17 + (0,91 + 0,12) \cdot x_{i2}\} = \exp\{2,17 + 1,03 \cdot x_{i2}\}.$$

Juhuslikud mõjud olid lapse spetsiifilised ning eeldus on, et juhuslike mõjude keskväärtus on null ($Eu_i = 0$). Seetõttu laste spetsiifiliste mõjude seas on pooled suuremad ja pooled väiksemad nullist. GlmmADMB paketi tulemusena saadud tunnuste Y ja X vahelist seost kirjeldavate joonte mediaani hinnnang kogu populatsioonile avaldub kujul:

$$\widehat{\text{median}}(E[y_{ij}|x_{ij}]) = \exp\{2,17 + 0,91 \cdot x_{ij}\}.$$

Ülaltoodud mediaan ja keskväärtuste hinnangud kattuvad täpselt eelmises alampeatükis saadud tulemustega.

Joonisel 1.2 on paketi glmmADMB tulemused ning algsed väärtused on peale kantud punktikestena. Joonisel tähistab must joon keskväärtuste mediaani hinnangut kogu populatsioonile ning punane ja sinine aga vastavalt esimese lapse ja teise lapse tinglike keskväärtuste hinnanguid. Jooniselt näeme, et kõik kolm joont algavad enam-vähem ühest punktist aga kolme joone tõusunurk on erinev. See on sellepärast, et me lubasime juhuslikul mõjul mõjutada ainult sirge tõusu, mitte vabaliiget.



Joonis 1.2: Mudel glmmADMB kasutades

Peatükk 2

ESM uuring

ESM (*Experience sampling methodology*) on uuringu meetod, mis lubab uurida kogemusi ja emotsioone, mis leiavad aset inimeste elus. Selleks tuleb osalejatel teatud aja tagant täita päevikut, kus nad jäädvustavad oma tundmused ning tegevused. (Allikas [10])

Tallinna Ülikooli psühholoogia instituudi teadurid viisid läbi 2014. aasta kevadel ESM uuringu, mille põhieesmärgiks oli hinnata, kui palju lapsed õpivad, millal nad õpivad ja missuguseid emotsioone nad õppides tunnevad.

Uuring viidi läbi erinevates Eesti koolides ning uuritavateks vanuseklassideks valiti teine ja seitsmes klass. Uuringusse kaasatud koolid ja klassid valiti juhuslikud erinevate Eesti koolide seast. Valimisse sattus nii maakoole kui ka niinimetatud eliitkoole Tallinnast ja Tartust.

Andmete kogumiseks kasutati küsitlusrakendust nutitelefonis või tahvelarvutis, mida õpilased pidid täitma kolm korda igal tööpäeval ning viis korda päevas laupäeval ja pühapäeval. Iga laps pidi küsitlusele vastama ühe nädala jooksul ning küsitluse täitmiseks saadeti meeldetuletuseks lühisõnum. Lastele oli öeldud, et küsitlust tuleb täita peale sõnumi saamist niipea kui võimalik. Tööpäeviti hakati meeldetuletusi saatma alates pärastlõunast iga paarikolme tunni tagant, st ajal, millal lapsed võiksid koduste töödega tegeleda. Loodeti tabada lapsi õppimast, et saaks aimu nende emotsioonidest õppimise ajal. Nädalavahetustel saadeti meeldetuletusi alates hommikust iga nelja-viie tunni tagant. Mõned õpilased võtsid uuringus osalemist tõsiselt ning täitsid küsitlusrakendust ka meeldetuletust saamata.

Küsitluses oli ligikaudu 100 küsimust, kuid iga kord kõigile küsimustele ei olnud vaja vastata, sest osad küsimused sõltusid vastustest. Näiteks kui laps parasjagu õppis, siis küsiti, mis ainet ta õpib ja mida ta tunneb õppimise ajal või kui ta ei õppinud, siis küsiti, millega ta tegeleb ja kas ta peaks õppima. Küsimused, mis esitati sõltumata vastustest olid: „Kaua Sa oled täna juba õppinud?", „Kaua Sa plaanid täna veel õppida?", „Kas Sa vastasid küsimustele ausalt?" ja identifitseerivad küsimused (nimi, kool, klass). Küsitlusrakendus fikseeris iga vastamise korra ajal, sekundi täpsusega, millal alustati küsitluse täitmist ning kas küsitlus lõpetati või mitte. Küsitlus loeti lõpetatuks siis, kui oli vastatud kõigile esitatud küsimustele.

Käesolevas magistritöös kasutame eelkõige küsitluse toimumise aega, vastust küsimusele „Kaua Sa oled täna juba õppinud?", lapse sugu, klassi ja kooli.

Küsimusele „Kaua Sa oled täna juba õppinud?" oli võimalik vasta skaalal: Üldse pole - Veerand tundi - Pool tundi - Tund - 2 tundi - ... - 7 tundi - 7 ja rohkem tundi. Joonistel on kasutatud vastuseid jagatuna kolme kategooriasse: Üldse pole - Alla tunni - Üle tunni, kus esimene kategooria koosneb neist vastustest, mis esialgsel skaalal anti variandile „Üldse pole", teine kategooria koosneb esialgse skaala variantidest „Veerand tundi" ja „Pool tundi" ning kolmas kategooria koondab endas ülejäänud vastused.

Andmete kogumine ei kulgenud plaanipäraselt. Palju kordi lapsed alustasid küsitluse täitmist, kuid ei lõpetanud; mõned lapsed täitsid küsitlust ainult ühel päeval ja siis mitu korda järjest ning teised lapsed vastasid mitmel erineval päeval, kuid igal päeval vaid ühe korra. Kokkuvõttes saadi palju vähem küsitluse täitmisi, kui loodeti, kuigi mõned lapsed vastasid ka, siis kui polnud meeldetuletust saanud. Näiteks saadeti täidetud ankeete ka ajal, mil lapsed oleksid pidanud viibima koolis tundides või ka küsitluseks planeeritud nädalale järgnenud päevadel.

Lisaks loodeti uuringut alustades, et välja valitud klasside kõik lapsed osalevad uuringus. Paraku selgus, et paljud lapsed ei soovinudki uuringus osaleda. Küsitlusandmed on olemas ainult nende laste kohta, kes vähemalt korra alustasid küsitluse täitmist. Järgnevates alampeatükkides välja toodud keskmiste arvutamisel on arvesse läinud ainult need lapsed, kes on vähemalt korra alustanud küsitlusele vastamist.

2.1 Teine klass

Teise klassi õpilasi oli uuringusse kaasatud viiest koolist 11 klassist. Kokku alustas vähemalt korra küsitluse täitmist nutirakenduses 139 teise klassi õpilast. Need 139 last alustasid küsitluse täitmist kokku 1228 korda, millest lõpetati küsitlus 823 korda ning pooleli jäi 405 korda.

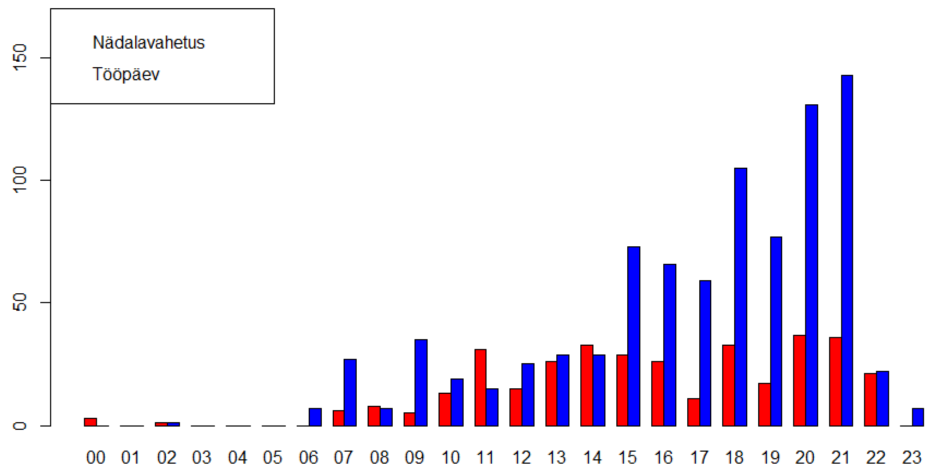
Kõige agaram teise klassi laps alustas küsimustiku täitmist 48 korda. Keskmiselt alustas küsitluses osalema nõustunud laps küsitluse täitmist 8,8 korda. Küsitlust alustas vähemalt üks teise klassi laps, kes ei lõpetanud kordagi küsitluse täitmist (ehk mingile küsimusele jättis vastamata). Vähemalt üks laps lõpetas küsitluse täitmise 30 korral ning keskmiselt lõpetas laps küsitluse täitmise 5,9 korral nende laste seast, kes vähemalt korra alustasid küsitlusele vastamist. Uuringus osales vähemalt üks teise klassi laps, kes ei jätnud küsitluse täitmist mitte kordagi pooleli ja üks laps jättis küsitluse täitmise pooleli 23 korda. Küsitluse täitmist vähemalt korra alustanud teise klassi laste seas jäeti keskmiselt küsitluse täitmine pooleli 2,9 korda. (vt. tabel 2.1)

Tabel 2.1: Teise klassi õpilaste küsitluse täitmine

Mitu korda uuringu jooksul?	Minimaalselt	Keskmiselt	Maksimaalselt
Küsitlust alustati	1	8,83	48
Küsitlus lõpetati	0	5,92	30
Küsitlus jäi pooleli	0	2,91	23

Küsitluse täitmist alustas vähemalt korra 139 last, neist vähemalt korra lõpetas küsitluse täitmise 127. Kui nende 127 lapse küsitluse täitmist vaadata, siis nende seas oli laps, kes lõpetas küsitluse täitmise 30 korda ning keskmiselt need lapsed lõpetasid küsitluse täitmise 6,5 korda. Ühe päeva jooksul keskmiselt on need lapsed lõpetanud küsitluse täitmise 2,1 korda ning maksimaalselt on päev jooksul lõpetatud küsitluse täitmine kuus korda.

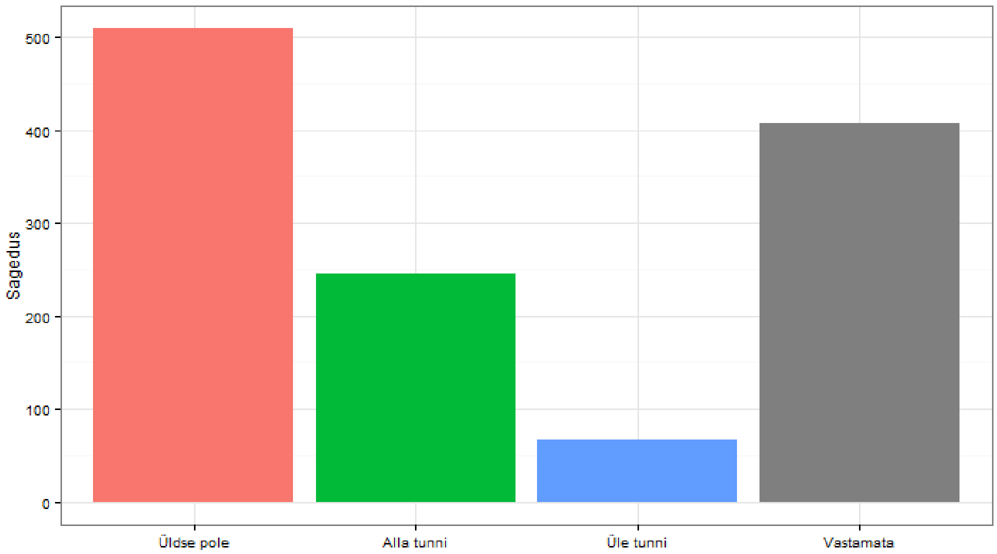
Joonisel 2.1 on need ajad, millal on küsitlust alustatud. Joonisel tähendab näiteks tund 10 ajavahemikku [10:00,11:00). Eraldi on välja toodud vastamine nädala sees ning nädalavahetusel, kuna nädala sees pidi ühes päevas täitma küsimustiku kolm korda ning nädalavahetusel viis korda. Joonist vaadates võib arvata, et meeldetuletuse saatmine toimus tööpäevadel kellaaegadel: 15:00, 18:00 ning 20:00 (või 21:00). Kuna nädalavahetusel oli vähem vastamisi, siis on vastuste laekumise sageduse põhjal keerulisem aimata täpseid meeldetuletuste saatmise kellaaegu.



Joonis 2.1: Teise klassi küsitluse täitmise alustamise ajaline jaotus

Kõige rohkem on küsitluse täitmisi alustatud vahemikus [20:00,22:00) nii tööpäeviti kui ka nädalavahetusel. Tööpäeva hommikuti on ka vahemikes [7:00,8:00) ning [9:00,10:00) palju vastatud (rohkem kui koolitundide ajal kell 10:00-15:00), see võib olla põhjustatud sellest, et kaks uuringusse kaasatud klassi käisid õhtuses vahetustes ehk nende koolipäev hakkaski pärastlõunal ning neile saadeti üks meeldetuletus ka enne kooli.

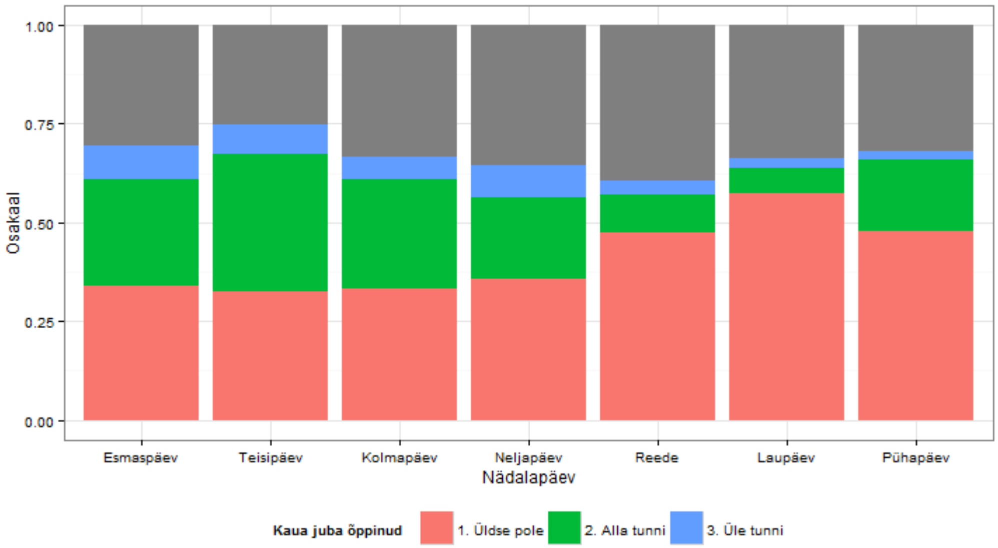
Joonisel 2.2 on näidatud kõik vastused, mis on antud küsimusele „Kaua Sa oled täna juba õppinud?”. Joonise koostamisel pole arvestatud, et ühelt lapselt võis olla mitu vastust. Sellelt jooniselt näeme, et enamikel vastamiskordade ajal ei ole lapsed veel üldse jõudnud õppida. Väga vähesed teise klassi lapsed on ankeedi täitmise ajaks jõudnud õppida enam kui tund aega.



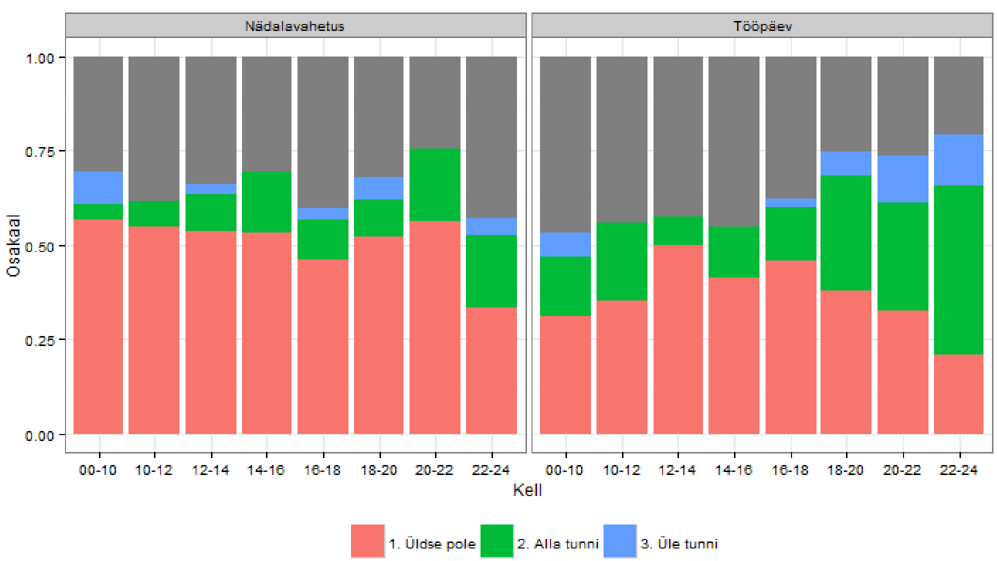
Joonis 2.2: Teise klassi vastamine küsimusele „Kaua Sa oled juba täna õppinud?”

Jooniselt 2.3 näeme, et reedel, laupäeval ja pühapäeval on vastamisel rohkem kasutatud varianti „Üldse pole”. Nädala sees kohtab sagedamini ka teisi vastusevariante.

Joonisel 2.4 on osakaalud kellaaegade lõikes. Kuna nädalavahetustel on väga harva ankeedi täitmise hetkeks oldud mingi aeg õppimas, siis on raske sellelt jooniselt midagi välja lugeda. Tööpäeviti on aga näha, kuidas hilisematel kellaaegadel õppimise hulk järjest suureneb. Kui vahemikus 16-18 oli vähesed õpilased jõudnud õppida, siis kell 22-24 oli rohkem kui pooltel kordadel vastatud, et midagi ollakse õppinud.



Joonis 2.3: Teise klassi vastamine küsimusele „Kaua Sa oled täna juba õppinud?" nädalapäevade kaupa



Joonis 2.4: Teise klassi vastamine küsimusele „Kaua Sa oled täna juba õppinud?" kellaaegade lõikes

2.2 Seitsmes klass

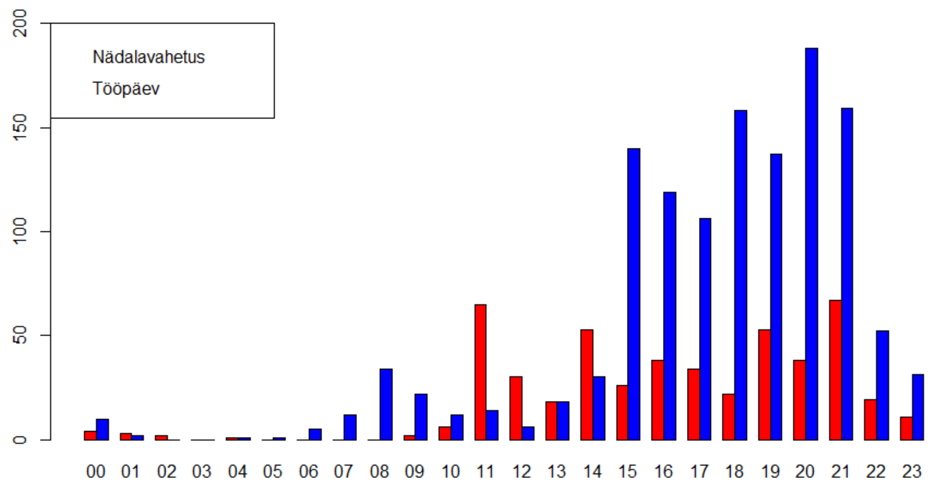
Uuringus osales seitsmest erinevast koolist seitsmenda klassi õpilasi 13 klassist. Kokku alustas vähemalt korra küsitluse täitmist nutirakenduses 183 seitsmenda klassi õpilast. Need 183 last alustasid küsitluse täitmist 1749 korda, millest täielikult lõpetati 1419 korda ning pooleli jäi 330 korda.

Seitsmenda klassi kõige tublim laps alustas küsitluse täitmist 34 korda. Need lapsed, kes olid nõus uuringus osalema, alustasid keskmiselt ankeedi täitmist 9,6 korda. Küsitlust alustas vähemalt üks laps, kes ei lõpetanud kordagi küsitluse täitmist. Uuringus osalema nõustunud seitsmenda klassi laste seas oli vähemalt üks laps, kes lõpetas küsitluse täitmise 32 korral ning keskmiselt lapsed lõpetasid küsitluse täitmise 7,8 korral. Uuringus osales vähemalt üks laps, kes ei jätnud küsitluse täitmist mitte kordagi pooleli ja üks laps, kes jättis 9 korda küsitluse täitmise pooleli. Keskmiselt jäi igal seitsmenda klassi lapsel, kes olid alustanud küsitluse täitmist vähemalt korra, küsitluse täitmine pooleli 1,8 korda. (vt. tabel 2.2)

Tabel 2.2: Seitsmenda klassi õpilaste küsitluse täitmine

Mitu korda uuringu jooksul?	Minimaalselt	Keskmiselt	Maksimaalselt
Küsitlust alustati	1	9,56	34
Küsitlus lõpetati	0	7,75	32
Küsitlus jäi pooleli	0	1,80	9

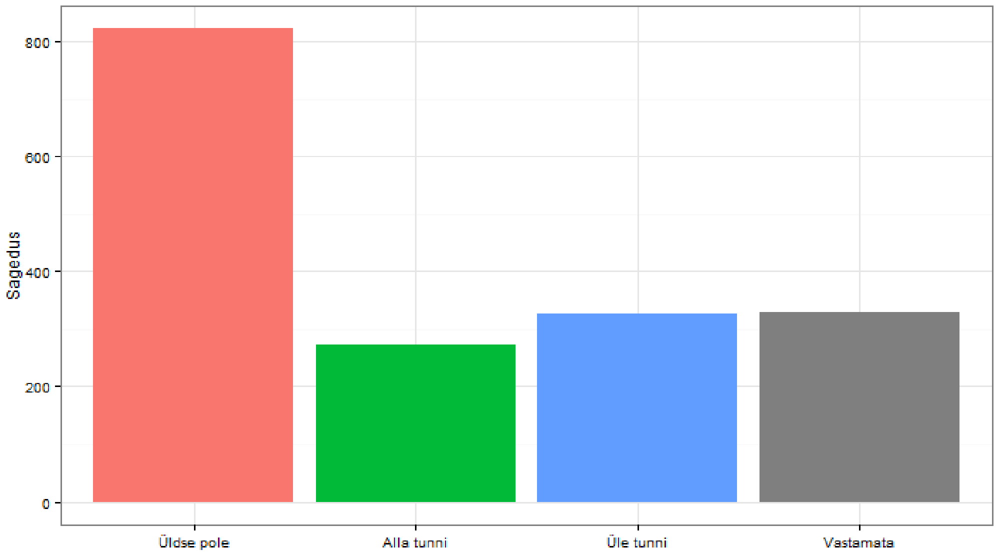
Küsitluse täitmist alustas vähemalt korra 183 seitsmenda klassi last, aga neist vähemalt korra lõpetas küsitluse täitmise 175. Kui nende 175 lapse küsitluse täitmist vaadata, siis nende seas oli laps, kes alustas ankeedi täitmist 34 korda ning keskmiselt alustasid need lapsed küsitluse täitmist 8,1 korda. Ühe päeva jooksul on need lapsed kõige rohkem lõpetanud kaheksa korda küsitluse täitmise ning keskmiselt 2,4 korda.



Joonis 2.5: Seitsmenda klassi küsitluse täitmise alustamise ajaline jaotus

Joonisel 2.5 on need ajad, millal on küsitlust alustatud. Joonisel tähendab näiteks tund 10 ajavahemikku [10:00,11:00). Eraldi on välja toodud vastamine nädala sees ning nädalavahetusel, kuna nädala sees pidi ühes päevas täitma küsitlust kolm korda ning nädalavahetusel viis korda. Joonist vaadates võib arvata, et meeletuletuse saatmine toimus tööpäevadel kellaaegadel: 15:00, 18:00 ning 20:00 (või 21:00). Kuna nädalavahetusel on vähem vastamisi, siis on keerulisem vastamissageduse põhjal välja tuua meeldetuletuse saatmise aega. Esile tulnud 4 vahemiku nädalavahetusel on 11:00, 14:00, 19:00, 21:00. Kõige rohkem on küsitluse täitmisi alustatud tööpäeviti vahemikus [20:00,21:00) ning nädalavahetusel vahemikus [21:00,22:00).

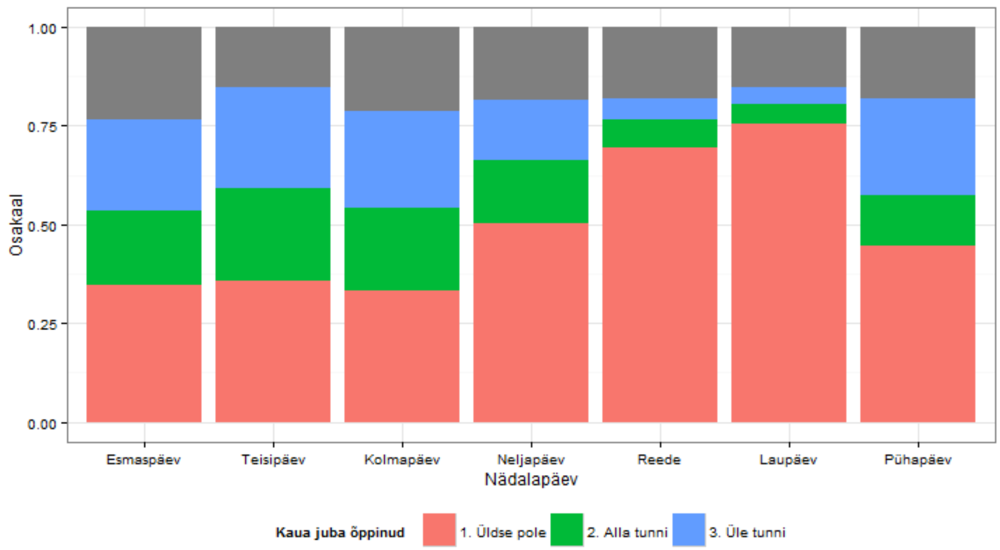
Joonisel 2.6 on näidatud kõik vastused, mis on antud küsimusele „Kaua Sa oled täna juba õppinud?” uuringus osalenud seitsmenda klassi laste poolt. Joonise koostamisel ei arvestatud, et üks laps võis mitu korda vastata. Sellelt joonisel näeme, et enamasti ei oldud vastamise ajaks jõutud veel õppida. Vähesed lapsed on ankeedi täitmise ajaks jõudnud õppida enam kui tund aega, kuid veel vähem on neid lapsi, kes on õppinud alla tunni. Ehk kui ollakse õppinud, siis pigem üle tunni kui alla tunni.



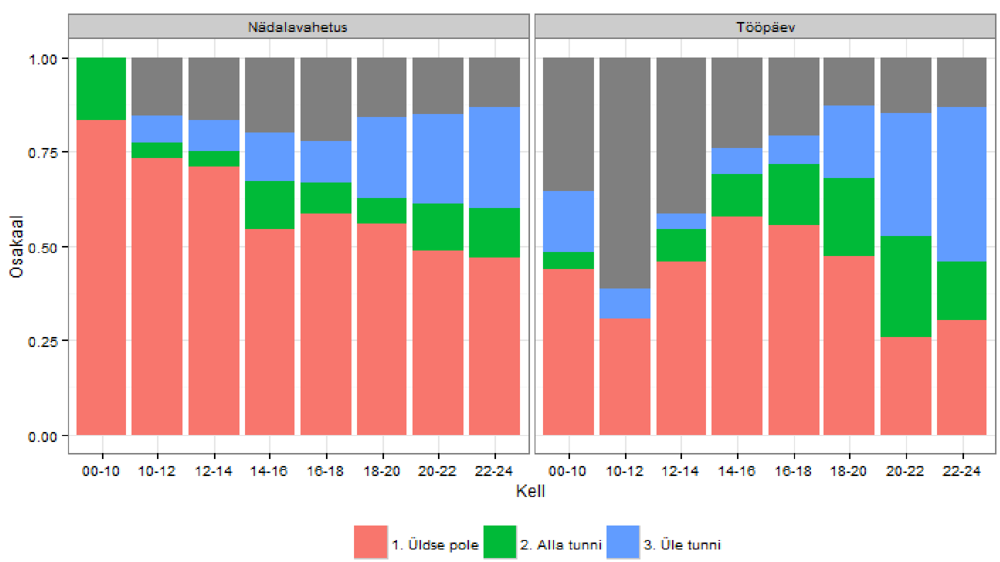
Joonis 2.6: Seitsmenda klassi vastamine küsimusele „Kaua Sa oled täna juba õppinud?”

Jooniselt 2.7 näeme, et reedel ja laupäeval on vastamisel rohkem kasutatud varianti „Üldse pole”. Nädala sees on suuremad osakaalud variantidel, mis ütlevad, et ollakse jõudnud õppida alla tunni või üle tunni. Ehk saab öelda, et tööpäeviti õpitakse rohkem. Osakaalud, alla tunni ja üle tunni õppimistel, on suhteliselt võrdsed.

Jooniselt 2.8 on osakaalud kellaaegade lõikes. Nii nädalavahetusel kui ka tööpäeviti on võimalik näha, et mida hilisem kellaaeg, seda rohkem on ankeedi täitmisel kasutatud vastuse variante, mis näitavad, et lapsed on jõudnud õppida. Kuigi nädalavahetusel on õppimise osakaal ikka väiksem kui mitte õppimise osakaal.



Joonis 2.7: Seitsmenda klassi vastamine küsimusele „Kaua Sa oled täna juba õppinud?” nädalapäevade kaupa



Joonis 2.8: Seitsmenda klassi vastamine küsimusele „Kaua Sa oled täna juba õppinud” kellaaegade lõikes

Peatükk 3

Leitud mudelid

Käesoleva magistritöö üheks eesmärgiks on leida mudel, mis aitaks hinnata, kui palju Eesti koolilapsed õpivad kodus. Mudeli leidmisel on aluseks võetud Tallinna Ülikooli psühholoogia instituudi teadurite poolt koostatud ESM-uuring, millest kasutame eelkõige vastuseid küsimusele „Kaua Sa oled täna juba õppinud?". Lisaks kasutasime ka küsitluse aega, vastaja sugu, kooli ja klassi. Uuritavaks tunnuseks on küsimus „Kaua Sa oled täna juba õppinud?".

Uuringus osales üks kool, kus teise klassi õpilased käisid õhtuses vahetuses. Nende õppimisrütm on teistsugune kui hommikuse vahetuse koolilastel. Seetõttu jäid need klassid analüüsist välja. Lisaks teame, et osad lapsed õpivad hommikuti ning seetõttu on nende vastuste puhul raske hinnata, missuguse päeva koduseid töid teevad ning kas nad õhtul vastates mäletasid, et hommikul juba õppisid. Selliseid vastamisi, kus vastati hommikupoolikul, et ollakse juba õppinud ning sama päeva õhtul vastati, et ei olda üldse veel õppinud, oli vastavalt teises klassis 24 ning seitsmendas klassis 32. Vältimaks vastuolulisi vastamisi päeva jooksul, võtsime analüüsi need ankeeditäitmised, mida alustati ajavahemikus [12:00,00:00]. Analüüsist jäid välja ka need ankeedi täitmised, mis olid lõpetamata ehk küsimusele „Kaua Sa oled täna juba õppinud?" ei olnud vastust antud.

Mudelite hindamiseks kasutame R lisapaketti glmmADMB, millega katsume leida parameetrite hinnangud üldistatud lineaarsele segamudelile. Selleks on meil vaja teada, missugune on uuritava tunnuse tinglik jaotus. Teame, et vastused küsimusele „Kaua Sa oled täna juba õppinud?" on alati positiivsed, seega tinglik keskväärtus nendel andmetel on positiivne. Kuna aga tegemist on loenduvate andmetega (lapsed loendavad palju tunde on nad õppinud), siis võiks olla tegemist Poissoni jaotusega. Kuna jaotuse hindamine on ligilähedane, siis lubame uuritava tunnuse dispersioonil olla kujul $D(y) = \alpha\mu$, kus $D(y)$ on uuritava tunnuse y dispersioon ja μ keskväärtus, ehk kasutame kvaasi-Poissoni mudelit. Kvaasi-Poissoni mudel võimaldab hakkama saada nii üle- kui ka alahajuvuse probleemidega, mis võivad tuleneda näiteks uuritava tunnuse jaotuse mõnevõrra valest valikust. Kui andmed oleksid täpselt Poissoni jaotusega, siis $\alpha = 1$ ja $D(y) = \mu$.

Lisaks uuritava tunnuse jaotusele tuleb valida ka seosefunktsioon. Negatiivne keskväärtus oleks meie uuritava tunnuse suhtes ebaloogiline. Seetõttu valime seosefunktsiooniks log-seosefunktsiooni, mis tagab, et keskväärtuse hinnang ei saa negatiivseks muutuda.

Eelmises peatükis nägime, et õppimisrütm on sarnane nädala alguse poole (esmaspäevast neljapäevani), kuid nädalavahetusel ja reedel õpitakse teistmoodi. Seega hindame kaks mudelit kummalegi klassile. Esimene mudel hindab iga päeva mõju eraldi ning teine hindab esmaspäevast-neljapäevani päeva mõjud ühesuguseks. Kahe mudeli võrdluses tuli välja, et mudel päevatüüpidega kirjeldab paremini teise klassi vastuseid ning mudel nädalapäevadega seitsmenda klassi vastuseid (vt. lisa B).

3.1 Mudel päevatüüpidega

Esimesena tutvustame mudelit, kus on kasutatud päevatüüpe ning see on hinnatud teisele klassile. Seitsmendale klassile hinnatud mudeli leiab lisast C.1.

Uuritavaks tunnuseks on küsimus „Kaua Sa oled täna juba õppinud?”, mille oleme tähistanud vastavalt *kaua*. Mudelis kasutame tunnuseid: *sugu*, *kell*, *kool* ning *päevatüüp*. Erinevad päevatüübid on jagunenud selliselt, et nädala algus on üks päevatüüp ehk hindame, et esmaspäeval, teisipäeval, kolmapäeval ja neljapäeval on ühesugune mõju. Järgmised päevatüübid on reede, laupäev ja pühapäev ehk hindame reedele oma mõju, sellest erineva mõju hindame laupäeval ja ka pühapäeval on oma mõju.

Mudelis on kell esitatud päevades- kellaaeg 12:00 võrdub 0,5 päevaga, ning 1 päev täistab vastavalt kella 00:00). Fikseeritud peamõjuga on tunnused *kell* ja *päevatüüp*. Fikseeritud mõjudena käsitletakse ka *kella* ja *soo*, *kella* ja *päevatüübi* ning *kella*, *soo* ja *päevatüübi* koosmõju. Mudeliga hinnati igale koolile omane tunnuse *kell* kordaja (kooli juhuslik mõju). Hinnati ka igale lapsele spetsiifiline tunnuse *kell* kordaja (lapse juhuslik mõju). Iga lapse igale päevale hinnati spetsiifiline kordaja tunnusele *kell* (konkreetse lapse konkreetse päeva juhuslik mõju).

Peale analüüsiks mittesobivate andmete eemaldamist jäi mudeli hindamisse 97 erinevat teise klassi last, kes käisid neljas erinevas koolis. Need lapsed täitsid ankeeti kokku 560 korda. Allpool oleva mudeli on genereeritud kood, mis asub koos oma väljundiga lisas C.2.

Saadud mudeli kuju:

$$
\begin{aligned}
E[kaua_{ijkl}|u_i, v_j, w_{jk}] &= \exp\Big(-3,271 + 2,191 \cdot kell_{ijkl} \\
&- 2,759 \cdot I_{(Päev_{ijkl} \in \{E,T,K,N\})} \\
&- 1,526 \cdot I_{(Päev_{ijkl}=R)} \\
&+ 3,095 \cdot I_{(Päev_{ijkl}=L)} \\
&- 0,900 \cdot kell_{ijkl} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 3,328 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl} \in \{E,T,K,N\}))} \\
&+ 1,168 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=R)} \\
&- 5,304 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=L)} \\
&+ 0,716 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl} \in \{E,T,K,N\})} \cdot I_{(Sugu_{ijkl}=N)} \\
&- 2,565 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=R)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 1,853 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=L)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ u_i \cdot kell_{ijkl} + v_j \cdot kell_{ijkl} + w_{jk} \cdot kell_{ijkl} \Big),
\end{aligned}
$$

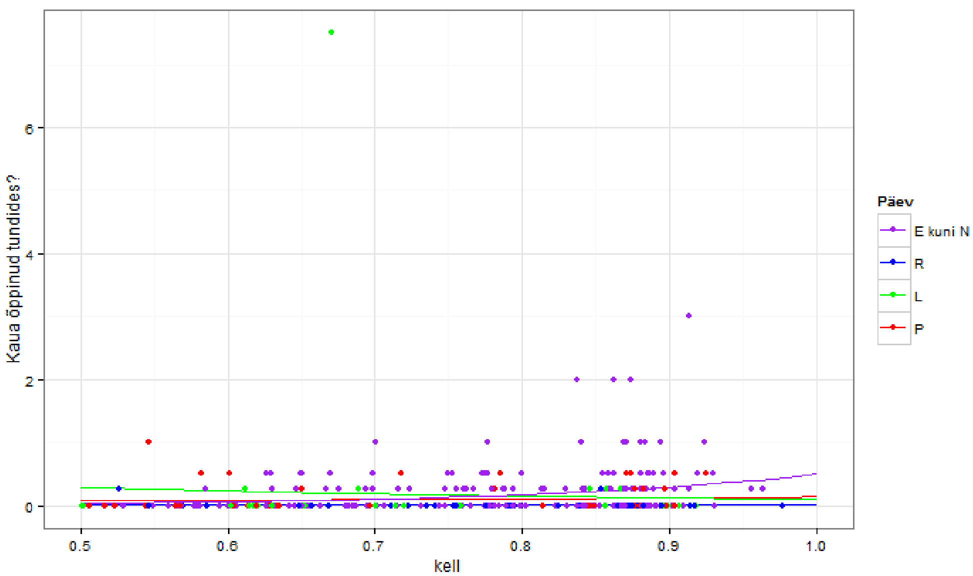
kus u_i on koolide juhuslike mõjude vektor ($i = \{1, 2, 3, 4\}$), v_j on laste juhuslike mõjude vektor ($j = \{1, 2, 3, ..., 97\}$), w_{jk} on konkreetse lapse päevade juhuslike mõjude vektor ($k = \{1, 2, 3, ..., 371\}$), ning $l = \{1, 2, 3, ..., 560\}$ loendab mitmes vaatlus on ja

$$
I_x = \begin{cases} 1, & \text{kui tingimus } x \text{ kehtib} \\ 0, & \text{kui tingimus } x \text{ ei kehti.} \end{cases}
$$

Me kasutasime mudeli hindamiseks kvaasi-Poissoni mudelit, kuid ülehajuvust kirjeldava parameeteri α hinnanguks saime ligikaudu võrdseks ühega ($\hat{\alpha} = 1.001$). Selleks, et õpitud tundide arvu jaotus oleks ligilähedaselt Poissoni jaotusega, peab olema $D(kaua) \approx \mu$. Me teame, et meie uuritav tunnus $kaua$ ei saa kindlasti olla täpselt Poissoni jaotusega, sest õppimisaja hindamiseks oli kasutatud mitte-täisarvulisi vastuseid (nagu 0,25 tundi ja 0,5 tundi). Kuid praegusel hetkel kehtib tingimus $D(kaua) \approx \mu$, mis ütleb meile, et meie uuritava tunnuse dispersiooni ja keskväärtuse vahel on samasugune seos nagu Poissoni jaotusega tunnuse dispersioonil ja keskväärtusel.

Võrdleme mudeli keskmise õppimise hinnanguid üle kõigi juhuslike mõjude kell 21:00 (tunnus $kell = 0,88$), kuna hilisematest aegadest on vähe vastuseid.

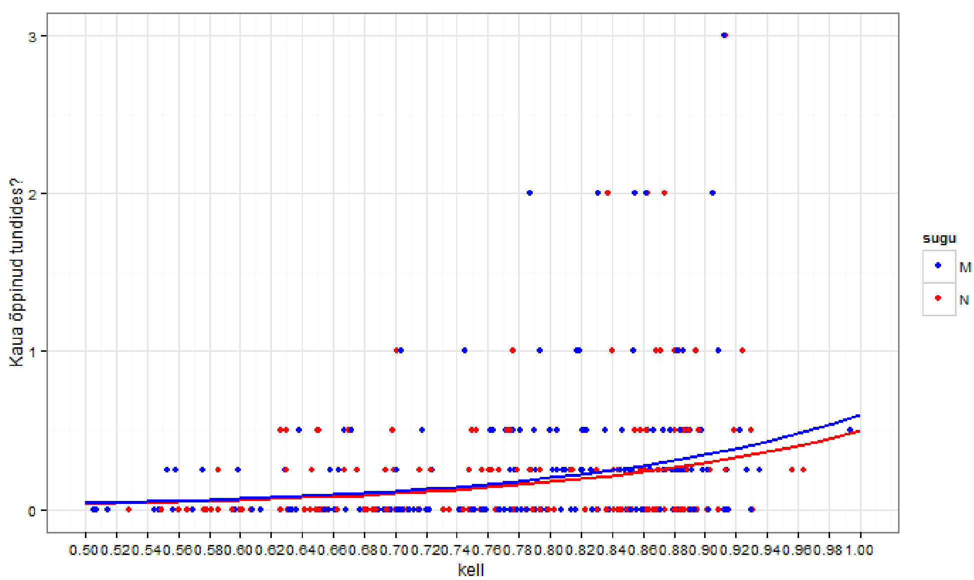
Joonisel 3.1 on punktikeste abil välja toodud kõikide analüüsi jäänud teise klassi tüdrukute vastused päevade kaupa. Joonisel asuvad jooned näitavad, kuidas mudel hindab tüdrukute keskmist õppimist (teise klassi poiste keskmise õppimise kohta tehtud joonis asub lisas C.3). Jooniselt 3.1 näeme, et reede, laupäeva ja pühapäevane tüdrukute keskmine õppimishulk on madalam kui nädala alguses olevate päevade õppimishulk. Nädalalõpu päevadel jääb keskmine õppimine väga nulli lähedale. Esmaspäevast-neljapäevani on teise klassi tüdrukute õppimisaja mediaani hinnang kell üheksa veerand tundi.



Joonis 3.1: Teise klassi tüdrukute õppimismahu hindamine (mudel päevatüüpidega)

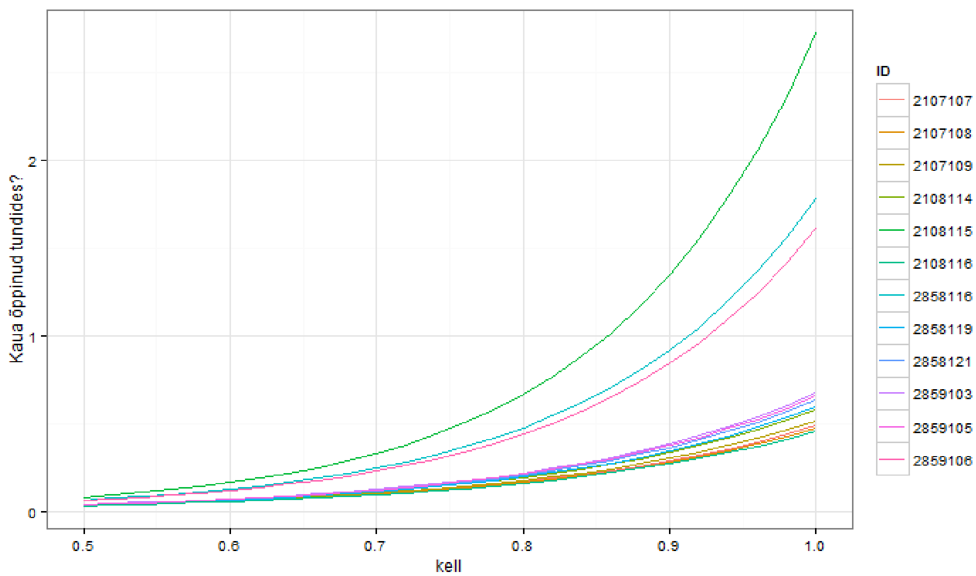
Joonisel 3.1 on näha, et laupäevast keskmist õppimist tähistav joon on kell 12:00 (joonisel 0,5) isegi natuke kõrgemal kui kell 00:00 (joonisel 1). Põhjus, miks varasemal kellaajal on õppimishulk suurem kui hilisemal kellaajal, tuleneb sellest, et mudeli parameetrid hinnatakse. Lisaks on üks teise klassi tüdruk ühel laupäeval väitnud, et on õppinud seitse tundi- mis on väga erandlik vastus. Üldine järeldus laupäeva mudeli kohta oleks, et sõltumata kellaajast pole laupäeviti teise klassi tüdrukud üldse aega õppimisele kulutanud.

Joonisel 3.2 võrreldakse poiste ja tüdrukute keskmist õppimismahtu nädala alguses. Poiste puhul on rohkem vastatud nädala alguse päevadel, et ollakse õppinud tund või rohkem. Kuid kordagi pole kummastki soost vastatud, et ollakse õppinud juba üle kolme tunni. Seetõttu pole üllatav, et poisid on kella üheksaks õhtul keskmiselt õppinud rohkem kui tüdrukud. Samas erinevus on väga väike, ainult mõned minutid. Mudeli hinnangute järgi otsustades poisid õpivad teise klassi tüdrukutest igal päeval rohkem, välja arvatud laupäeval, kui mudeli hinnangute põhjal ei õpi poisid ega ka tüdrukud üldse (vt lisa C.4).



Joonis 3.2: Teise klassi tüdrukute ja poiste õppimismahtude võrdlemine nädala alguses (mudel päevatüüpidega)

Koolide spetsiifiliste *kella* kordajate standardhälve on $4,548 \cdot 10^{-5}$ ning nende kordajate väärtused jäävad vahemikku $(6,7 \cdot 10^{-10}; 1,8 \cdot 10^{-8})$. Järelikult erinevates koolides õpitakse sama palju. Laste spetsiifiliste *kella* kordajate standardhälve on 0,7279. Lapsele vastav *kella* kordaja ei hinda konkreetse päeva juhuslikku mõju vaid katsub leida terve uuringus osalemise aja tunnuse *kell* keskmist kordajat. Joonisel 3.3 on võetud igast koolist kolm teise klassi last. Valimisse sattus viis poissi (koodidega: 2108114, 2859103, 2858121, 2858116, 2859105) ja seitse tüdrukut (koodidega: 2108115, 2108116, 2107109, 2107108, 2859106, 2107107, 2858119). Kui joonisel 3.2 on, et keskmiselt õpivad poisid rohkem kui tüdrukud, siis joonisel 3.3 on kõige rohkem õppiv teise klassi laps tüdruk (kood 2108116).



Joonis 3.3: Teise klassi 12 juhusliku lapse õppimismahu hindamine nädala alguses (mudel päevatüüpidega)

3.2 Mudel nädalapäevadega

Teise mudelina, mida proovisime sobitada, oli mudel, kus igal päeval oli oma mõju. See kirjeldas paremini seitsmenda klassi vastuseid, kui teine mudel. Seega siin alampeatükis tutvustame seda mudelit seitsmendale klassile. Teisele klassile sai ka samasugune mudel hinnatud, mille tulemused on lisas D.1.

Uuritavaks tunnuseks on küsimus „Kaua Sa oled täna juba õppinud?”, mille tähistame vastavalt *kaua*. Kasutame tunnuseid: *sugu*, *kell*, *kool* ning *nädalapäev*. Mudelis on kell esitatud päevades- kellaaeg 12:00 võrdub 0,5 päevaga, ning 1 päev täistab vastavalt kella 00:00). Fikseeritud peamõjuga on tunnused *kell* ja *nädalapäev*. Fikseeritud mõjudena käsitletakse ka *kella* ja *soo*, *kella* ja *nädalapäev* ning *kella*, *soo* ja *nädalapäev* koosmõju. Mudeliga hinnati igale koolile omane tunnuse *kell* kordaja (kooli juhuslik mõju). Hinnati ka igale lapsele spetsiifiline tunnuse *kell* kordaja (lapse juhuslik mõju). Iga lapse igale päevale hinnati spetsiifiline kordja tunnusele *kell* (konkreetse lapse konkreetse päeva juhuslik mõju).

Peale analüüsiks mittesobivate andmete eemaldamist jäi mudeli hindamisse 167 erinevat seitsmenda klassi last, kes käisid seitsmes erinevas koolis. Need lapsed täitsid ankeeti kokku 1280 korda. Lähemalt saab uurida allpool oleva mudeli R-i koodi ja väljundit lisast D.2.

Saadud mudeli kuju:

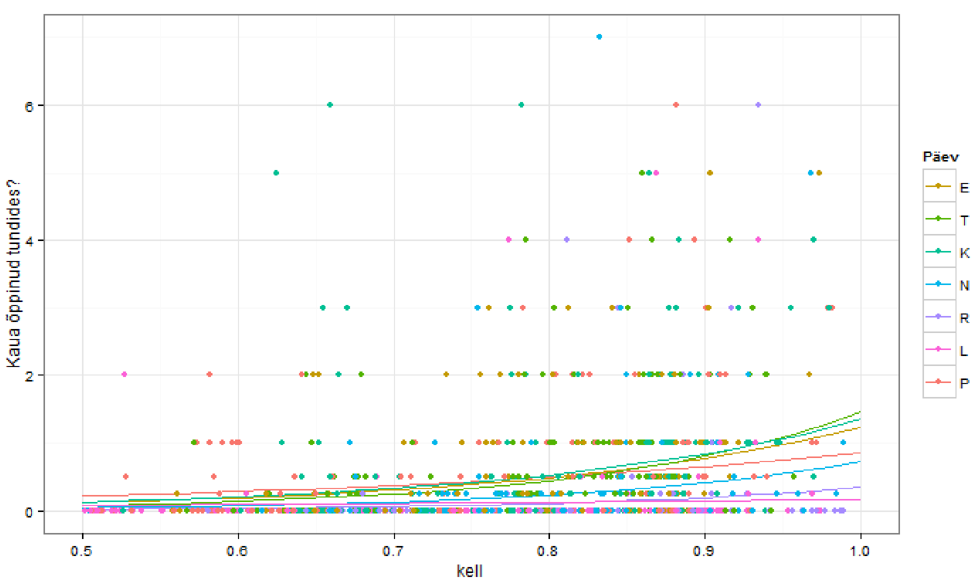
$$
\begin{aligned}
E[y_{ijkl}|u_i, v_j, w_{jk}] &= \exp\Big(-2,951 + 3,571 \cdot kell_{ijkl} \\
&- 1,523 \cdot I_{(Päev_{ijkl}=E)} - 2,647 \cdot I_{(Päev_{ijkl}=T)} \\
&- 1,472 \cdot I_{(Päev_{ijkl}=K)} - 3,225 \cdot I_{(Päev_{ijkl}=N)} \\
&- 3,929 \cdot I_{(Päev_{ijkl}=R)} - 0,350 \cdot I_{(Päev_{ijkl}=L)} \\
&- 0,768 \cdot kell_{ijkl} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 0,961 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=E)} \\
&+ 2,256 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=T)} \\
&+ 1,284 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=K)} \\
&+ 3,054 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=N)} \\
&+ 1,180 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=R)} \\
&- 2,042 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=L)} \\
&+ 0,919 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=E)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 1,082 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=T)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 0,634 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=K)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 0,001 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=N)} \cdot I_{(Sugu_{ijkl}=N)} \\
&- 1,813 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=R)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ 0,722 \cdot kell_{ijkl} \cdot I_{(Päev_{ijkl}=L)} \cdot I_{(Sugu_{ijkl}=N)} \\
&+ u_i \cdot kell_{ijkl} + v_j \cdot kell_{ijkl} + w_{jk} \cdot kell_{ijkl} \Big),
\end{aligned}
$$

kus u_i on koolide juhuslike mõjude vektor ($i = \{1, 2, 3, 4, 5, 6, 7\}$), v_j on laste juhuslike mõjude vektor ($j = \{1, 2, 3, ..., 167\}$), w_{jk} on konkreetsete laste päevade juhuslike mõjude vektor ($k = \{1, 2, 3, ..., 744\}$) ning $l = \{1, 2, 3, ..., 1280\}$ loendab mitmes vaatlus on ja

$$
I_x = \begin{cases} 1, & \text{kui tingimus } x \text{ kehtib} \\ 0, & \text{kui tingimus } x \text{ ei kehti.} \end{cases}
$$

Me kasutasime mudeli hindamiseks kvaasi-Poissoni mudelit, kuid ülehajuvust kirjeldava parameeteri α hinnanguks saime ligikaudu võrdseks ühega ($\hat{\alpha} = 1.0759$). Selleks, et õpitud tundide arvu jaotus oleks ligilähedaselt Poissoni jaotusega, peab olema $D(kaua) \approx \mu$. Me teame, et meie uuritav tunnus $kaua$ ei saa kindlasti olla täpselt Poissoni jaotusega, sest õppimisaja hindamiseks oli kasutatud mitte-täisarvulisi vastuseid (nagu 0,25 tundi ja 0,5 tundi). Kuid praegusel hetkel kehtib tingimus $D(kaua) \approx \mu$, mis ütleb meile, et meie uuritava tunnuse dispersiooni ja keskväärtuse vahel on samasugune seos nagu Poissoni jaotusega tunnuse dispersioonil ja keskväärtusel.

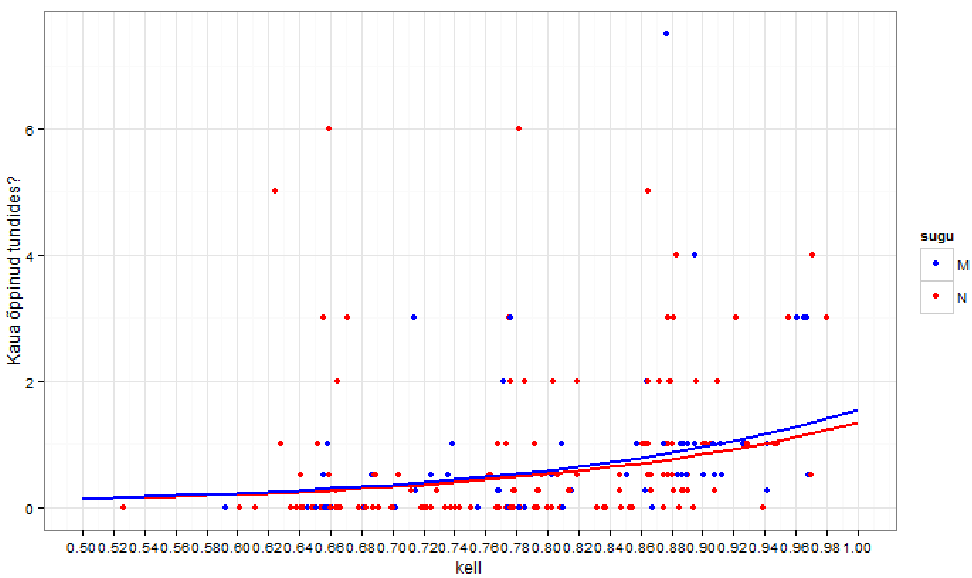
Võrdleme mudeli keskmise õppimise hinnanguid üle kõigi juhuslike mõjude kell 22:00 (tunnus $kell = 0,92$), kuna hilisematest aegadest on vähe vastuseid.



Joonis 3.4: Seitsmenda klassi tüdrukute õppimishulga hindamine (mudel päevadega)

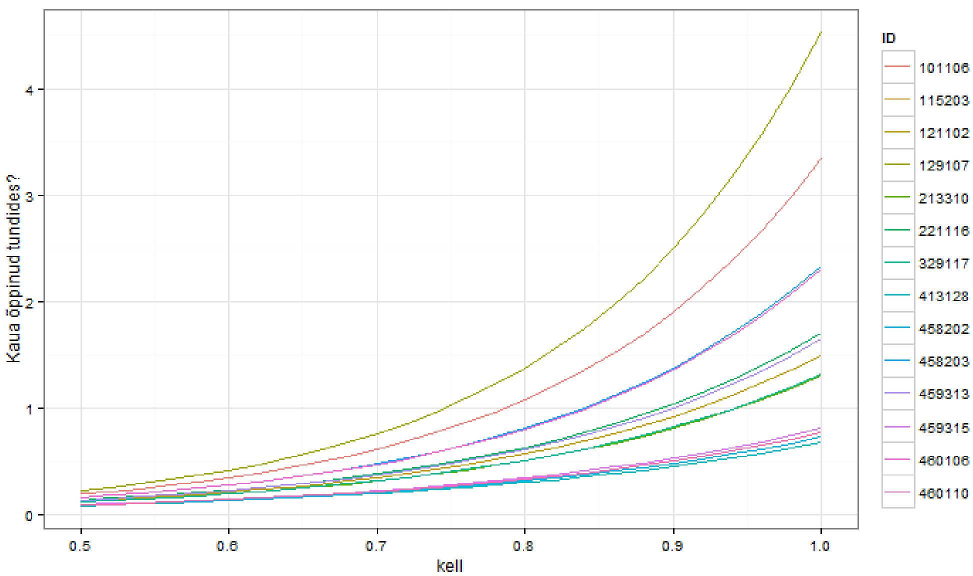
Joonisel 3.4 on punktikeste abil välja toodud kõikide analüüsi jäänud seitsmenda klassi tüdrukute vastused päevade kaupa. Joonisel asuvad jooned näitavad, kuidas mudel hindab tüdrukute keskmist õppimist (seitsmenda klassi poiste keskmise õppimise kohta tehtud joonis asub lisas D.3). Joonised 3.4 näeme, et igal päeval jääb kella kümneks õhtul jõutav keskmine õppimise hulk alla tunni. Mudeli järgi õpivad seitsmenda klassi tüdrukud peale kooli teistest päevadest rohkem teisipäeval, kolmapäeval ja esmaspäeval. Nädala lõpu poole keskmine õppimishulk väheneb. Laupäeval ei õpi mudeli arvates seitsmenda klassi tüdrukud üldse ning ega reedelgi väga palju õppimisse aega ei panustata.

Joonisel 3.5 võrreldakse seitsmenda klassi poiste ja tüdrukute keskmist õppimisemahtu kolmapäeval. Joonisel asuvad punktikesed tähistavad, kuidas on ankeetidele vastatud ning jooned näitavad kolmapäeval hinnatud keskmist õppimisemahtu poiste ja tüdrukute puhul. Jooniselt 3.5 näeme, et analüüsis olevate andmete seas on tundub olevat rohkem tüdrukute ankeedi täitmisi, kui poiste. Mudel hindab, et poiste õppimisaja mediaani hinnang kella kümneks õhtul on üks tund ning tüdrukute mediaani hinnang vastavalt 55 minutit. Mudeli hinnangute järgi otsustades õpivad seitsmenda klassi poisid tüdrukutest keskmiselt rohkem kolmapäeval, neljapäeval ja pühapäeval. Laupäeviti ei õpi mudeli arvates ei poisid ega ka tüdrukud. Esmaspäeval, teisipäeval ja reedel õpivad mudeli hinnangute põhjal tüdrukud keskmiselt rohkem kui seitsmenda klassi poisid. Joonised sugude võrdlemise iga päeva kohta on lisas D.4.



Joonis 3.5: Seitsmenda klassi tüdrukute ja poiste õppimismahu hindamine kolmapäeval (mudel päevadega)

Koolide spetsiifiliste *kella* kordajate standardhälve on $4,548 \cdot 10^{-5}$ ning nende kordajate väärtused jäävad vahemikku $(-5,5 \cdot 10^{-9}; 1,6 \cdot 10^{-8})$. Järelikult erinevates koolides õpitakse sama palju. Laste spetsiifiliste *kella* kordajate standardhälve on 0,7279. Lapsele vastav *kella* kordaja ei hinda konkreetse päeva juhuslikku mõju vaid katsub leida terve uuringus osalemise aja tunnuse *kell* keskmist kordajat. Joonisel 3.6 on võetud igast koolist kaks seitsmenda klassi last. Valimisse sattus neli poissi (koodidega: 121102, 221116, 458202, 459313) ja 10 tüdrukut (koodidega: 213310, 413128, 101106, 115203, 129107, 329117, 458203, 459315, 460106, 460110). Joonisel 3.6 on näha, et laste keskmise õppimisaja hinnangud erinevad üksteisest lausa mitme tunni võrra.



Joonis 3.6: Seitsmes klassi juhusliku 14 lapse õppimismahu hindamine kolmapäeval (mudel päevadega)

Kokkuvõte

Viimastel aastatel on ühiskonnas tihti esinenud arvamust, et koolilaste liialt suur kodutööde koormus ei võimalda neil leida aega hobidele pühendumiseks. Tallinna Ülikooli psühholoogia instituudi teadurite poolt 2014. aasta kevadel läbiviidud uuringus püüti jõuda selgusele, kuivõrd õpilased ise tunnevad end kodutöödega koormatud olevat ja kui kaua nad tegelikult õpivad.

Uuringuga hõlmati koole-klasse üle Eesti, võrdlemaks võimalikke erinevusi nn eliitkoolide ja tavakoolide õpikoormustes. Uurimistöö viidi läbi ESM-meetodil. Küsitlusele vastas siiski loodetust vähem arv õppureid – seega on saadud andmed lünklikud.

Käesolev magistritöö otsis vastust küsimusele, kui palju lapsed reaalselt veedavad kodus aega õppides ning analüüs põhines eelneval uuringul. Leidmaks vastust uuritavale küsimusele, hinnati üldistatud lineaarsed segamudelid, mis näitavad kui palju keskmiselt päeva lõpuks on teise ja seitsmenda klassi lapsed õppinud. Hinnati kaks erinevat segamudelit kummalegi klassile. Mudelite erinevus seisnes selles, et ühes mudelis oli iga nädalapäev oma mõjuga ning teises mudelis oli esmaspäevast neljapäevani päeva mõju ühesugune. Viimast mudelit kutsuti päevatüüpidega mudeliks. Teise klassi puhul osutus paremaks mudeliks päevatüüpidega ning seitsmenda klassi puhul nädalapäevadega mudel.

Mõlema klassi puhul oli sarnane, et klassile parim mudel ei arvanud, et lapsed laupäeval tegeleks õppimisega. Seitsmenda klassi lapsed õppisid keskmiselt rohkem kui teise klassi lapsed. Seitsmenda klassi poiste kõige töökam päev oli pühapäev, kus kella kümneks õhtul oli poiste õppimimahu mediaani hinnnaguks 1,4 tundi. Seitsmenda klassi tüdrukutel aga oli vastavalt kolmapäev ning nende õppimishulga mediaani hinnanguks oli 55 minutit õppimist, kella kümneks õhtul. Teise klassi poiste kõige tublim päevatüüp õppimise suhtes oli nädalaalgus, kus nende õppimismahu mediaani hinnnaguks oli 18 minutit õppimist kella üheksaks õhtul. Teise klassi tüdrukute kõige töökam päevatüüp õppimise suhtes oli samamoodi nädalaalgus, aga nende õppimishulga mediaani hinnanguks oli 16 minutit õppimist, kella üheksaks õhtul.

Teostatud analüüsi üheks eesmärgiks oli välja selgitada, et kas sõltub koolist, kui palju lapsed kodus õpivad. Magistritöös leitud segamudelid ei kinnitanud, et oleks mingi erinevus laste õppimises sõltuvalt sellest, kus koolis nad käivad.

Kokkuvõtteks võib öelda, et läbi viidud ESM-uuringu põhjal tehtud analüüs ütleb, et teise ja seitsmenda klassi lapsed teevad koduseid töid iga päev, välja arvatud laupäeval. Ühtemoodi õpivad niinimetatud eliitkoolide õpilased kui ka maakoolide õpilased. Andmete lünklikkuse tõttu on aga keeruline teha põhjapanevaid järeldusi selle kohta, kui palju Eesti koolilapsed ikkagi kodus õpivad.

Kirjandus

[1] Anonüümne (28.11.2014). Murelik ema: lapsed põlevad koolikoormuse tõttu läbi juba enne täiskasvanuks saamist. [www] `http://rahvahaal.delfi.ee/news/uudised/murelik-ema-lapsed-polevad-koolikoormuse-tottu-labi-juba-enne-taiskasvanuks-saamist?id=70250197` - toimetaja Ragnar Teeveer (viimati vaadatud 12.05.2015)

[2] Maire (20.01.2015). Miks lapsed paksud on? Pidin poja trennist ära võtma, sest õppimise kõrvalt ei jäänud tal selleks enam aega! [www] `http://naistekas.delfi.ee/persoon/lugejakiri/miks-lapsed-paksud-on-pidin-poja-trennist-ara-votma-sest-oppimise-korvalt-ei-jaanud-tal-sellek-enam-aega?id=70602545` (viimati vaadatud 12.05.2015).

[3] Kangur Riina (21.02.2015). Õpetaja: ma pole kusagil näinud nii üleväsinud, tüdinud ja õnnetuid lapsi kui Eestis. [www] `http://rahvahaal.delfi.ee/news/uudised/opetaja-ma-pole-kusagil-nainud-nii-ulevasinud-tudinud-ja-onnetuid-lapsi-kui-eestis?id=70849497` (viimati vaadatud 12.05.2015).

[4] Adamson Andres (25.10.2013) Kuidas laste ja õpetajate ülekoormust vähendada? [www]

`http://opleht.ee/10214-kuidas-laste-ja-opetajate-ulekoormust-vahendada/` (viimati vaadatud 12.05.2015).

[5] Käärik E. (2012). Aine „Andmeanalüüs II” materjalid. Tartu Ülikool.

[6] Vikipeedia (2015) Generalized linear model [www]

`http://en.wikipedia.org/wiki/Generalized_linear_model` (viimati vaadatud 12.05.2015).

[7] McCulloch C. E., Searle S. R. (2001). Generalized, Linear, and Mixed Models. United States of America: A Wiley- Interscience Publication, John Wiley and Sons, Inc.

[8] Lindsey J. K. (2000). Applying Generalized Linear Models. New York: Springer.

[9] GlmmADMB (2015) [www] http://glmmadmb.r-forge.r-project.org/ (viimati vaadatud 12.05.2015).

[10] Vikipeedia (2015) Experience sampling method. [www] http://en.wikipedia.org/wiki/Experience_sampling_method (viimati vaadatud 12.05.2015).

Lisa A

Näide

A.1 Ise kirjutatud funktsioonid

```
> x_norm = c(1.47,1.66,0.38,0.71)
> z = c(1,2,2,1)
> y_pois = round(c(2,1,2,5) + z*15*x_norm)
> naidis = data.frame(y = y_pois,x = x_norm,z = as.factor(z))
> integral_ruumala = function(f,a,b,c,d){
+   samm1 = 100
+   samm2 = 100
+   h1 = (b-a)/samm1 #sammupikkus1
+   h2 = (d-c)/samm2 #sammupikkus2
+   i = c(1:samm1)
+   j = c(1:samm2)
+   x = a + h1*i # X-i väärtused
+   y = c + h2*j # Y-i väärtused
+   pindala = h1*h2 # ruudupindala
+   Sn = 0
+   for (j in 1:(samm1-1)){
+     for (i in 1:(samm2-1)){
+       Sn = Sn + pindala*f((x[i+1]+x[i])/2,(y[j+1]+y[j])/2)
+     }
+   }
+   return(Sn)
+ }
> LogLike = function(par,andmed) {
+   # Teada andmed
+   x = andmed$x
+   y = andmed$y
+   z = andmed$z
+   X=cbind(1,x)
+   Z=cbind(1*(z==1), 1*(z==2))
+   # Anname algväärtused otsitavatele
+   beeta=par[1:2]
+   sigma = par[3]
+   Xb=X%*%beeta
+   # tõepära:
+   toepara = function(u0,u1){
```

```
+     lambda = exp(Xb+Z%*%c(u0, u1)*x)
+     ((prod(dpois(y, lambda =lambda)))*
+     (exp(-u0**2/(2*sigma**2))/(sqrt(2*pi*sigma**2)))*
+     (exp(-u1**2/(2*sigma**2))/(sqrt(2*pi*sigma**2))))
+   }
+   toe = integral_ruumala(toepara, a = -15*sigma, b = 15*sigma,
c = -15*sigma, d = 15*sigma)
+   log = log(toe)
+   return(log)
+ }
> log_like = optim(par = c(2.2,0.9,0.14), fn = LogLike,
andmed = naidis, control=list(fnscale=-1))
> log_like
$par
[1] 2.1666454 0.9045845 0.1396279
$value
[1] -12.08441
$counts
function gradient
      86       NA
$convergence
[1] 0
$message
NULL
```

A.2 Pakett glmmADMB

```
> x_norm = c(1.47,1.66,0.38,0.71)
> z = c(1,2,2,1)
> y_pois = round(c(2,1,2,5) + z*15*x_norm)
> naidis = data.frame(y = y_pois,x = x_norm,z = as.factor(z))
> library(MASS)
> library(glmmADMB)
> naide <- glmmadmb(y~x+(x-1|z),data = naidis, family = "poisson")
> summary(naide)
Call:
glmmadmb(formula = y ~ x + (x - 1 | z), data = naidis, family = "poisson")
AIC: 30.2
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.167      0.304    7.13  9.7e-13 ***
x              0.905      0.249    3.64  0.00027 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Number of observations: total=4, z=2
Random effect variance(s):
Group=z
  Variance StdDev
x  0.01959   0.14
Log-likelihood: -12.0846
> ranef(naide)
$z
           x
1 -0.1091959
2  0.1156203
```

Lisa B

Mudelite võrdlus

Teise klassi mudelite võrdlus

```
> anova(mudel2_2,mudel1_2)
Analysis of Deviance Table

Model 1: kaua_oppinud_pidev ~ kell + paev2 + sugu:kell +
paev2:kell + paev2:kell:sugu
Model 2: kaua_oppinud_pidev ~ kell + nadalapaev + sugu:kell
+ nadalapaev:kell + nadalapaev:kell:sugu
  NoPar  LogLik Df Deviance Pr(>Chi)
1    16 -279.65
2    25 -275.90  9    7.514   0.5838
```

Seitsmenda klassi mudelite võrdlus

```
> anova(mudel2_7,mudel1_7)
Analysis of Deviance Table

Model 1: kaua_oppinud_pidev ~ kell + paev2 + sugu:kell +
paev2:kell + paev2:kell:sugu
Model 2: kaua_oppinud_pidev ~ kell + nadalapaev + sugu:kell
+ nadalapaev:kell + nadalapaev:kell:sugu
  NoPar  LogLik Df Deviance Pr(>Chi)
1    16 -1089.2
2    25 -1078.4  9    21.64  0.01009 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Lisa C

Mudel päevatüüpidega

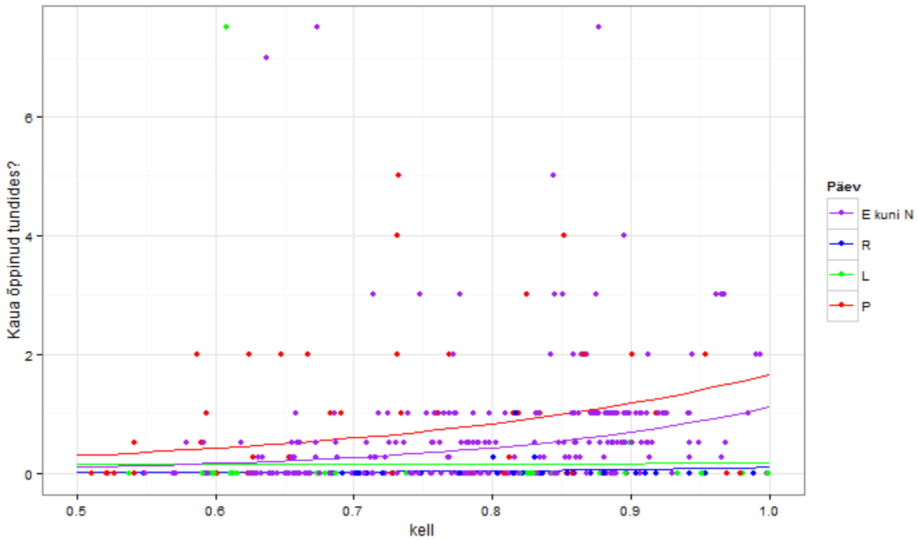
C.1 Seitsmes klass

R-i kood ja väljund

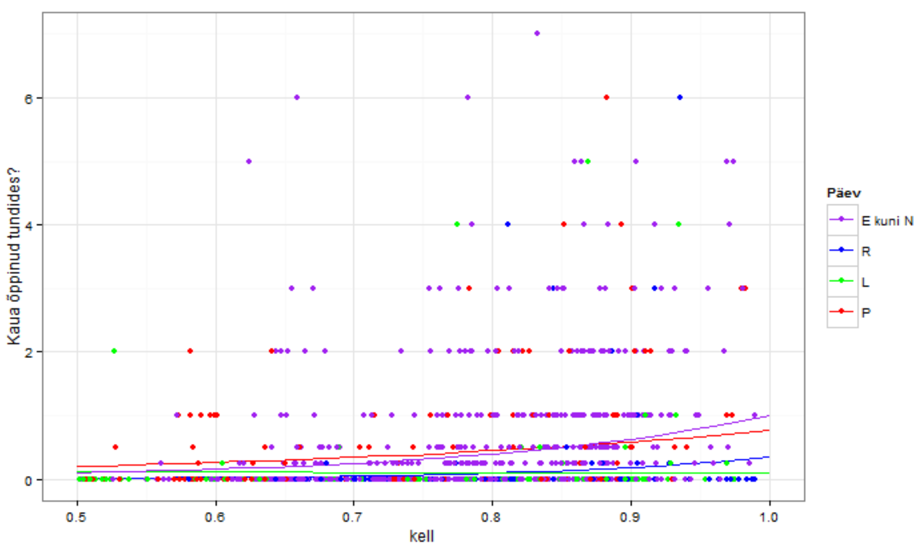
```
> fID_7 = factor(seitsmes3$ID)
> fdate_7 = factor(paste(seitsmes3$ID,seitsmes3$date))
> fKool_7 = factor(seitsmes3$Kool)
> mudel2_7 <- glmmadmb(kaua_oppinud_pidev~kell+paev2+
+sugu:kell+ paev2:kell+paev2:kell:sugu+
+(kell-1|fKool_7/fID_7/fdate_7),
+data=seitsmes3, family="nbinom1")
> summary(mudel2_7)

Call:
glmmadmb(formula = kaua_oppinud_pidev ~ kell + paev2 +
sugu:kell + paev2:kell + paev2:kell:sugu +
    (kell - 1 | fKool_7/fID_7/fdate_7),
    data = seitsmes3, family = "nbinom1")

AIC: 2210.5

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)            -2.956      0.682   -4.34  1.5e-05 ***
kell                    3.465      0.908    3.82  0.00013 ***
paev25                 -4.676      2.583   -1.81  0.07025 .
paev26                  0.865      1.314    0.66  0.51061
paev2algus             -1.715      0.834   -2.06  0.03981 *
kell:suguN             -0.783      0.366   -2.14  0.03255 *
kell:paev25             1.848      3.156    0.59  0.55816
kell:paev26            -3.187      1.780   -1.79  0.07332 .
kell:paev2algus         1.314      1.077    1.22  0.22272
kell:paev25:suguN       2.067      0.972    2.13  0.03347 *
kell:paev26:suguN       0.280      0.698    0.40  0.68815
kell:paev2algus:suguN   0.680      0.368    1.85  0.06457 .
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of observations: total=1280, fKool_7=7,
fKool_7:fID_7=167, fKool_7:fID_7:fdate_7=744
Random effect variance(s):
Group=fKool_7
      Variance   StdDev
kell 2.061e-09 4.54e-05
Group=fKool_7:fID_7
     Variance StdDev
kell    0.758 0.8706
Group=fKool_7:fID_7:fdate_7
     Variance StdDev
kell   0.5706 0.7554

Negative binomial dispersion parameter: 1.001
(std. err.: 0.0001088)

Log-likelihood: -1089.25
> ranef(mudel2_7)
$fKool_7
            kell
13  5.193356e-09
15  5.752236e-09
21 -5.228315e-09
29  1.265401e-08
58  1.842985e-08
59  2.954479e-09
60  1.163250e-08

$‘fKool_7:fID_7‘
                   kell
13:113105  0.0693823122
13:113111 -0.4025782995
13:113113  0.4165689858
13:113120  0.4983017571
13:113121 -0.1361371266
...
$‘fKool_7:fID_7:fdate_7‘
                                     kell
13:113105:113105 2014-04-15  0.1421865548
13:113105:113105 2014-04-16  0.0332891637
13:113105:113105 2014-04-17 -0.1232489948
13:113111:113111 2014-04-15 -0.0437731843
13:113111:113111 2014-04-16 -0.1408344115
...
```

Hinnatud mudelid sugude kaupa

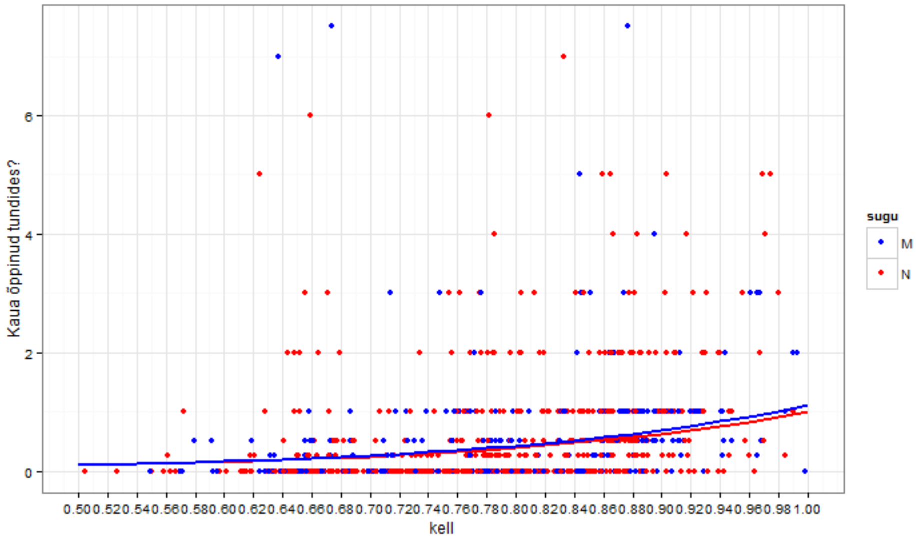


Joonis C.1: Seitsmenda klassi poiste mudelid

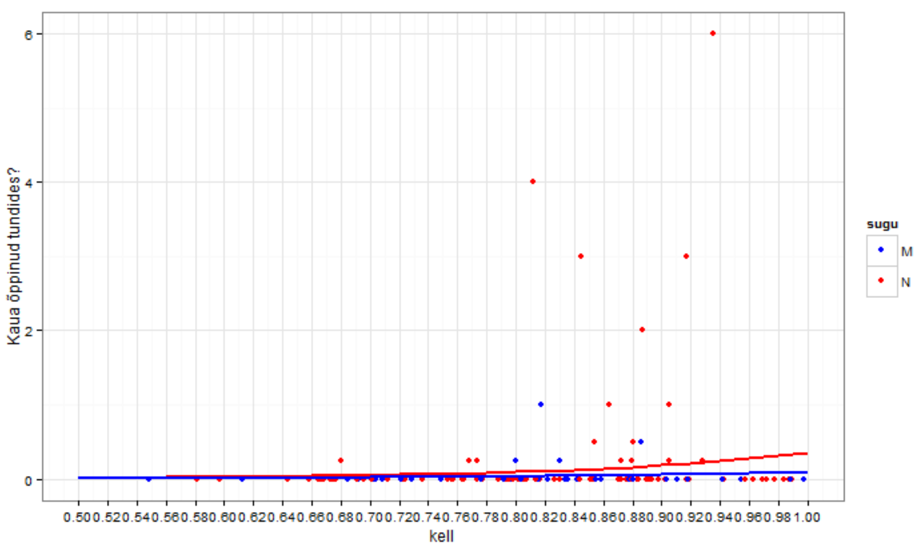


Joonis C.2: Seitsmenda klassi tüdrukute mudelid

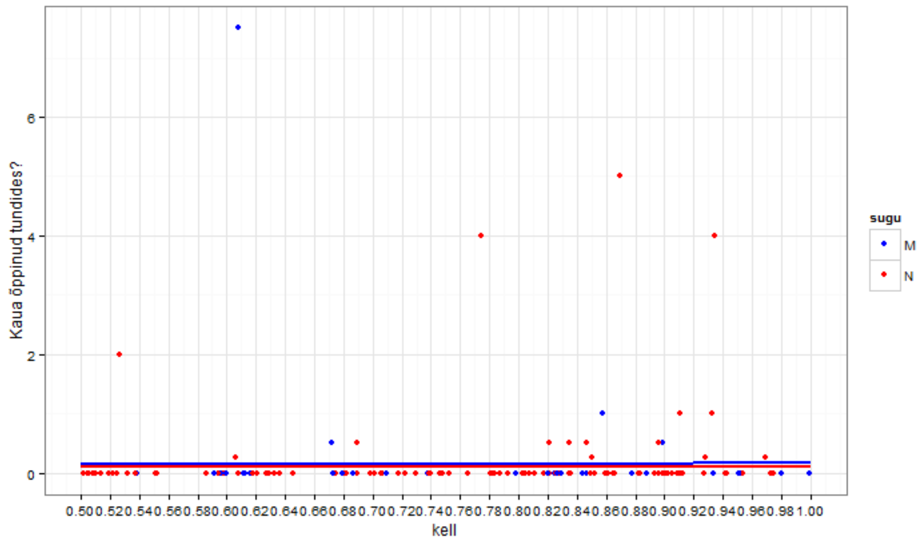
Sugude võrdlemine päevatüüpide kaupa



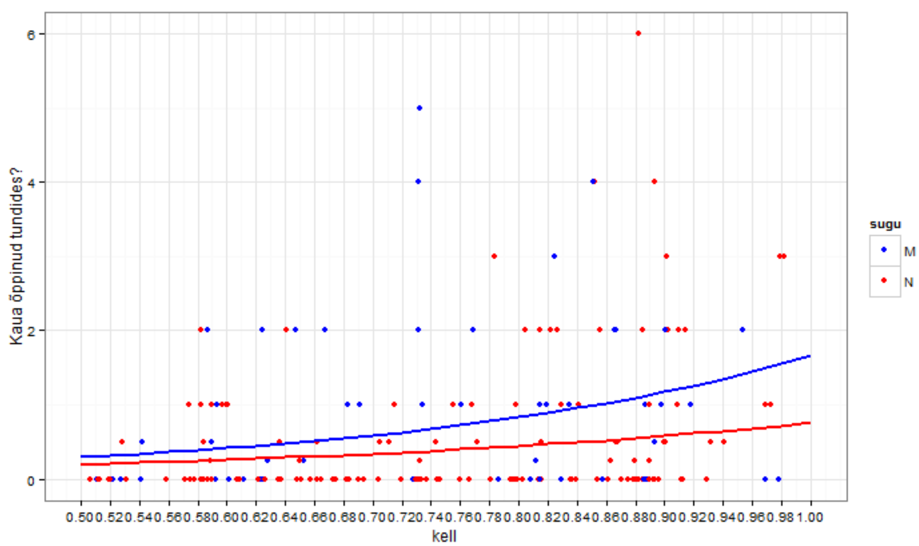
Joonis C.3: Esmaspäevast neljapäevani



Joonis C.4: Reede

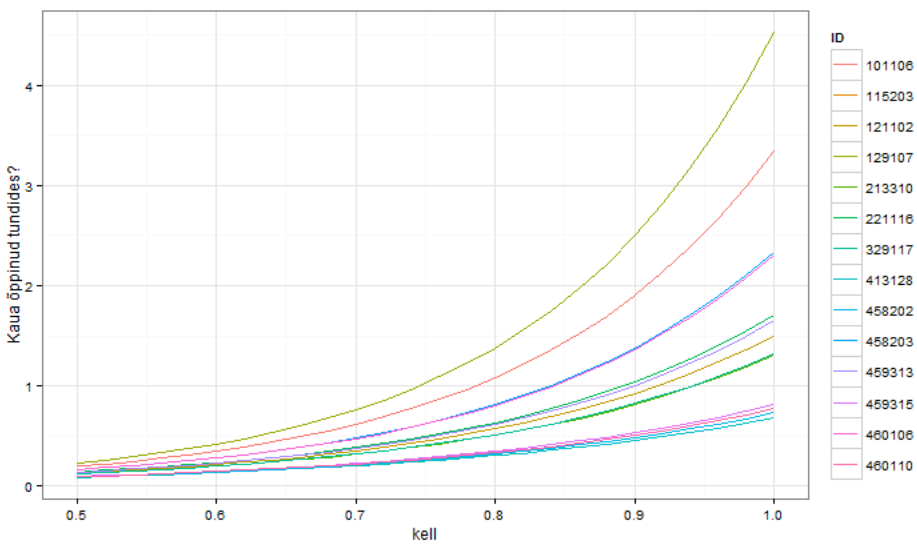


Joonis C.5: Laupäev



Joonis C.6: Pühapäev

Juhuslike laste võrdlemine



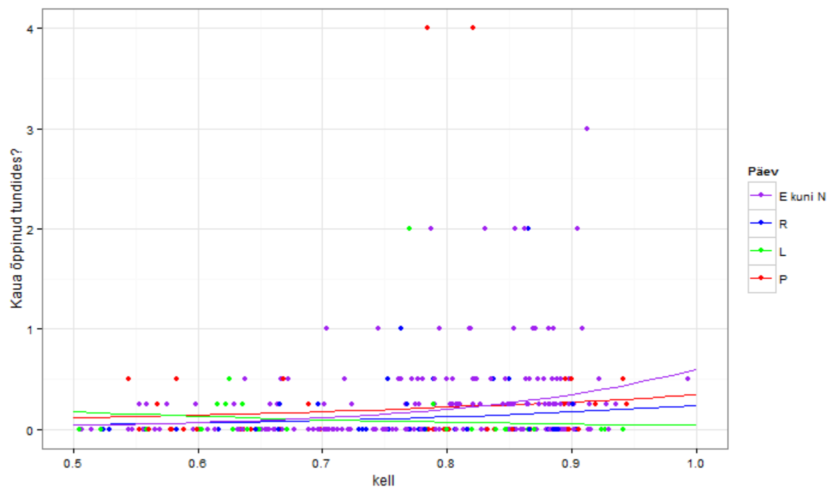
Joonis C.7: Päevatüüp- Esmaspäevast neljapäevani

C.2 Teine klass: R-i kood ja väljund

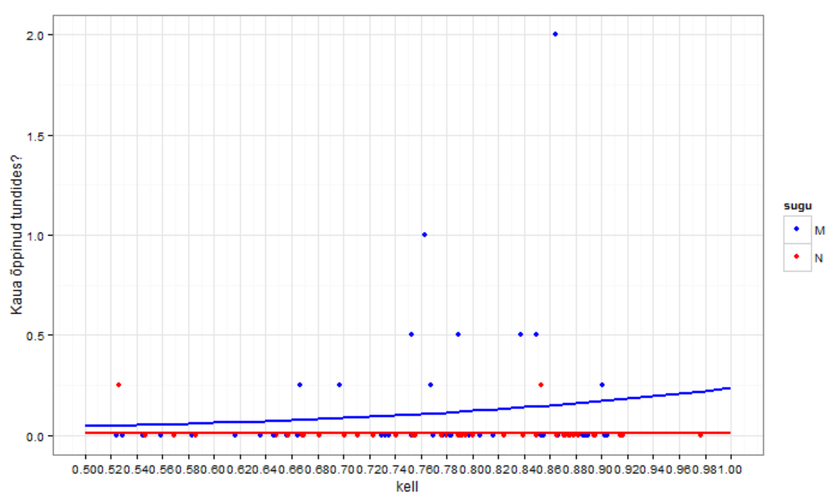
```
> fID_2 = factor(teine3$ID)
> fdate_2 = factor(paste(teine3$ID,teine3$date))
> fKool_2 = factor(teine3$Kool)
> mudel2_2 <- glmmadmb(kaua_oppinud_pidev~kell+paev2+sugu:kell+
+paev2:kell+paev2:kell:sugu+
+(kell-1|fKool_2/fID_2/fdate_2),
+admb.opts=admbControl(shess=FALSE, noinit=FALSE),
+data=teine3, family="nbinom1")
> summary(mudel2_2)

Call:
glmmadmb(formula = kaua_oppinud_pidev ~ kell + paev2 + sugu:kell +
paev2:kell + paev2:kell:sugu +
(kell - 1 | fKool_2/fID_2/fdate_2),
data = teine3, family = "nbinom1",
admb.opts = admbControl(shess = FALSE, noinit = FALSE))

AIC: 591.3

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)             -3.271      1.578   -2.07    0.038 *
kell                     2.191      2.019    1.09    0.278
```

```
paev25                  -1.526      3.780   -0.40    0.686
paev26                   3.095      2.501    1.24    0.216
paev2algus              -2.759      1.947   -1.42    0.157
kell:suguN              -0.900      0.758   -1.19    0.235
kell:paev25              1.168      4.730    0.25    0.805
kell:paev26             -5.304      3.493   -1.52    0.129
kell:paev2algus          3.328      2.444    1.36    0.173
kell:paev25:suguN       -2.565      2.018   -1.27    0.204
kell:paev26:suguN        1.853      1.220    1.52    0.129
kell:paev2algus:suguN    0.716      0.790    0.91    0.365
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of observations: total=560, fKool_2=4,
fKool_2:fID_2=97, fKool_2:fID_2:fdate_2=371
Random effect variance(s):
Group=fKool_2
      Variance    StdDev
kell 2.068e-09 4.548e-05
Group=fKool_2:fID_2
     Variance StdDev
kell   0.5298 0.7279
Group=fKool_2:fID_2:fdate_2
     Variance StdDev
kell   0.5806 0.7619

Negative binomial dispersion parameter: 1.001 (std. err.: 4.1805e-05)

Log-likelihood: -279.653
> ranef(mudel2_2)
$fKool_2
           kell
7  6.859489e-10
8  2.972173e-10
58 3.792274e-09
59 1.824331e-08

$‘fKool_2:fID_2‘
                   kell
7:2107107  -0.019605103
7:2107108  -0.050063109
7:2107109   0.031414459
7:2107112  -0.340770959
7:2107113   0.360336756
...

$‘fKool_2:fID_2:fdate_2‘
                                       kell
7:2107107:2107107 2014-04-17  -0.0240728865
7:2107107:2107107 2014-04-18  -0.0031560519
7:2107107:2107107 2014-04-19  -0.0694162263
7:2107107:2107107 2014-04-22  -0.0267343600
7:2107107:2107107 2014-04-25  -0.0032841348
...
```

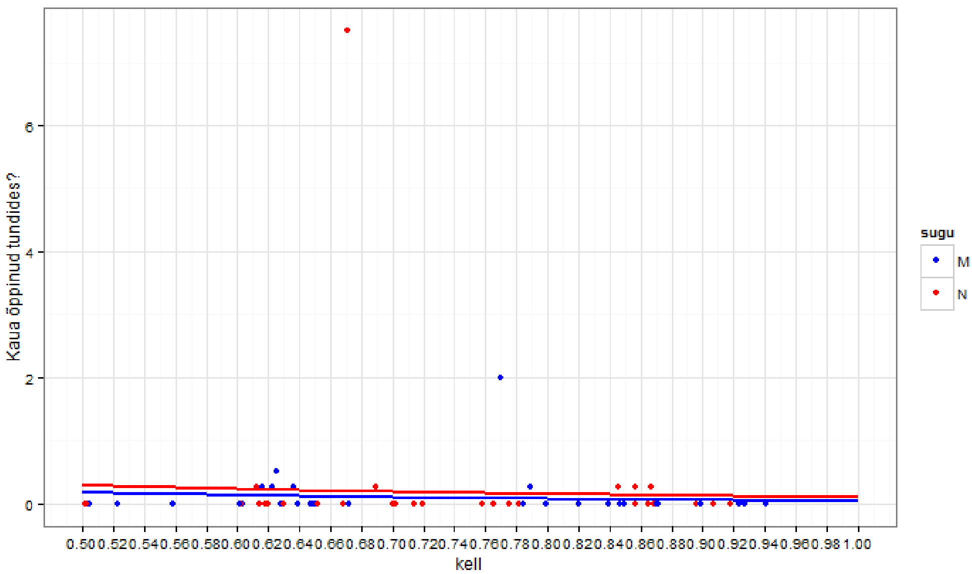
C.3 Teine klass: poistele hinnatud mudelid



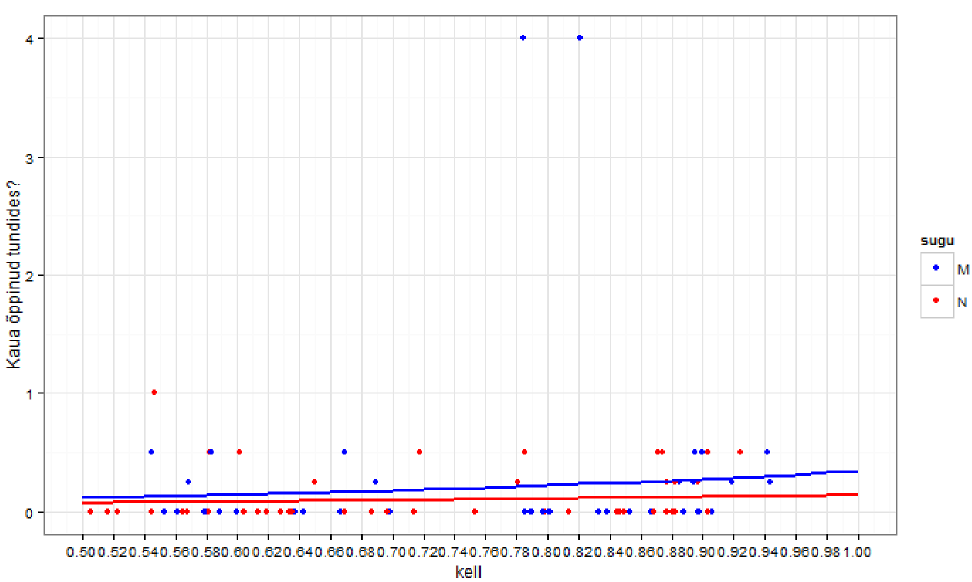
C.4 Teine klass: sugude võrdlemine päevatüüpide kaupa



Joonis C.8: Reede



Joonis C.9: Laupäev



Joonis C.10: Pühapäev

Lisa D

Mudel nädalapäevadega

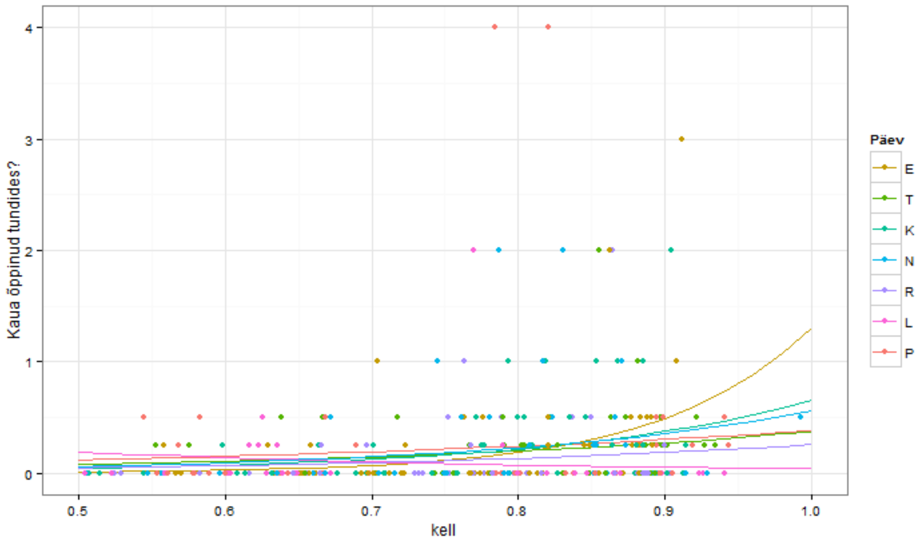
D.1 Teine klass

R-i kood ja väljund

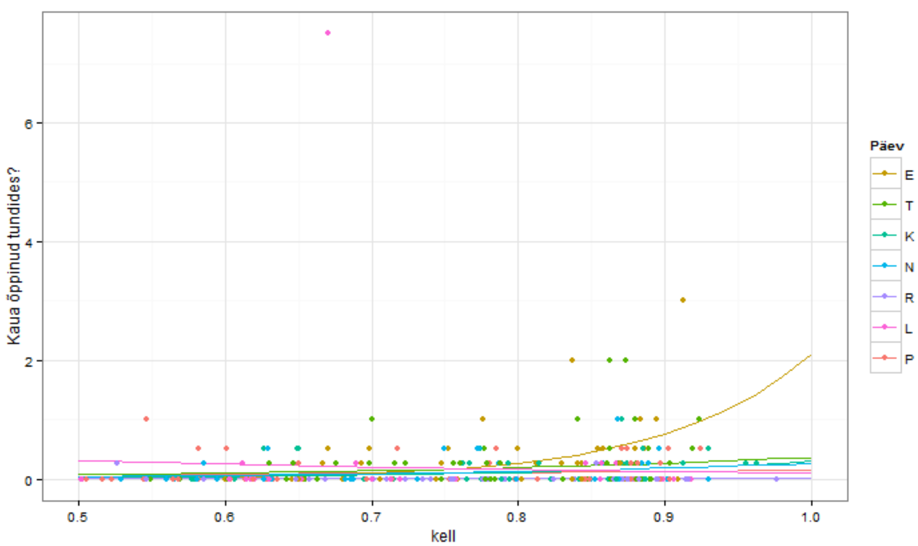
```
> fID_2 = factor(teine3$ID)
> fdate_2 = factor(paste(teine3$ID,teine3$date))
> fKool_2 = factor(teine3$Kool)
> mudel1_2 <- glmmadmb(kaua_oppinud_pidev~kell+nadalapaev
+sugu:kell+nadalapaev:kell+nadalapaev:kell:sugu+
+(kell-1|fKool_2/fID_2/fdate_2),
+ admb.opts=admbControl(shess=FALSE,noinit=FALSE),
+ data=teine3, family="nbinom1")
> summary(mudel1_2)

Call:
glmmadmb(formula = kaua_oppinud_pidev ~ kell + nadalapaev +
sugu:kell + nadalapaev:kell + nadalapaev:kell:sugu +
(kell - 1 | fKool_2/fID_2/fdate_2),
data = teine3, family = "nbinom1",
admb.opts = admbControl(shess = FALSE, noinit = FALSE))

AIC: 601.8

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)              -3.266      1.558   -2.10    0.036 *
kell                      2.296      1.989    1.15    0.248
nadalapaev1              -6.286      2.949   -2.13    0.033 *
nadalapaev2              -0.931      2.555   -0.36    0.715
nadalapaev3              -2.767      2.923   -0.95    0.344
nadalapaev4              -1.807      2.848   -0.63    0.526
nadalapaev5              -1.497      3.756   -0.40    0.690
nadalapaev6               3.219      2.447    1.32    0.188
kell:suguN               -0.967      0.742   -1.30    0.193
kell:nadalapaev1          7.525      3.584    2.10    0.036 *
kell:nadalapaev2          0.913      3.153    0.29    0.772
```

```
kell:nadalapaev3          3.311      3.626    0.91    0.361
kell:nadalapaev4          2.190      3.549    0.62    0.537
kell:nadalapaev5          1.104      4.693    0.24    0.814
kell:nadalapaev6         -5.519      3.425   -1.61    0.107
kell:nadalapaev1:suguN    1.444      0.899    1.61    0.108
kell:nadalapaev2:suguN    0.937      0.908    1.03    0.302
kell:nadalapaev3:suguN    0.172      0.985    0.17    0.862
kell:nadalapaev4:suguN    0.237      1.035    0.23    0.819
kell:nadalapaev5:suguN   -2.502      2.011   -1.24    0.214
kell:nadalapaev6:suguN    1.956      1.198    1.63    0.103
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of observations: total=560, fKool_2=4,
fKool_2:fID_2=97, fKool_2:fID_2:fdate_2=371
Random effect variance(s):
Group=fKool_2
     Variance StdDev
kell  0.03396 0.1843
Group=fKool_2:fID_2
     Variance StdDev
kell   0.5148 0.7175
Group=fKool_2:fID_2:fdate_2
     Variance StdDev
kell   0.4363 0.6605

Negative binomial dispersion parameter: 1.001
(std. err.: 4.2115e-05)

Log-likelihood: -275.896
> ranef(mudel1_2)
$fKool_2
           kell
7   0.002241696
8  -0.032555732
58  0.050008347
59  0.235520922

$`fKool_2:fID_2`
                    kell
7:2107107  -0.0434891005
7:2107108  -0.0440695770
7:2107109   0.0097676477
7:2107112  -0.3161408624
7:2107113   0.3414981435
...
$`fKool_2:fID_2:fdate_2`
                                      kell
7:2107107:2107107 2014-04-17  -0.0173620334
7:2107107:2107107 2014-04-18  -0.0024844518
7:2107107:2107107 2014-04-19  -0.0559957779
7:2107107:2107107 2014-04-22  -0.0309564032
7:2107107:2107107 2014-04-25  -0.0025867019
...
```

Sugudele hinnatud mudelid

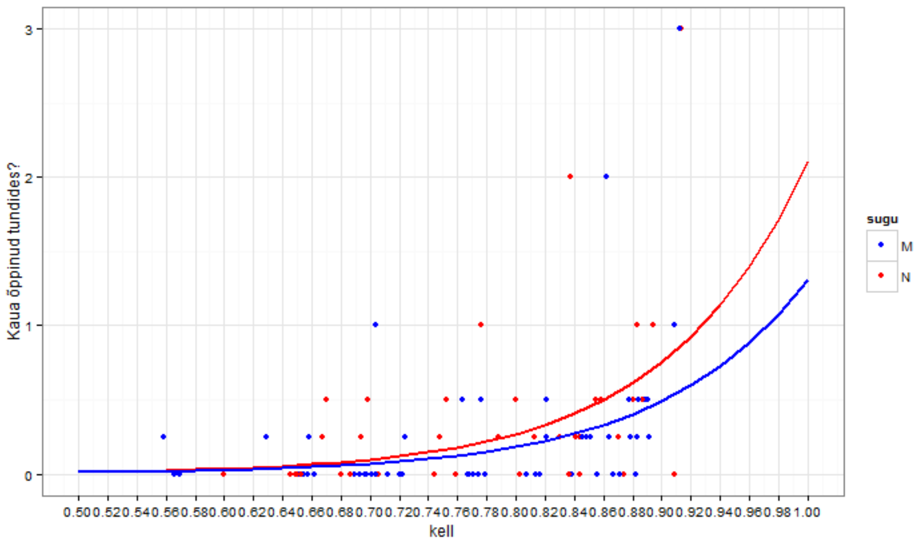


Joonis D.1: Poiste mudelid

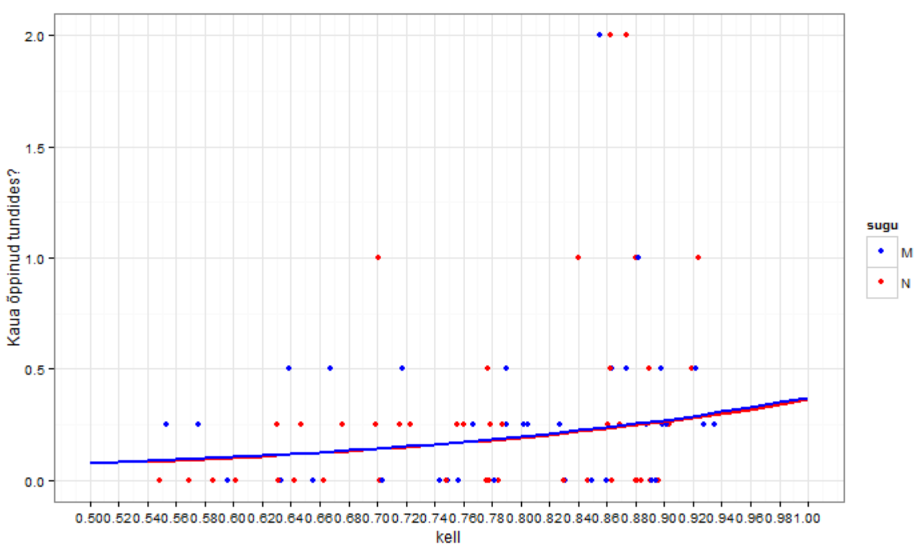


Joonis D.2: Tüdrukute mudelid

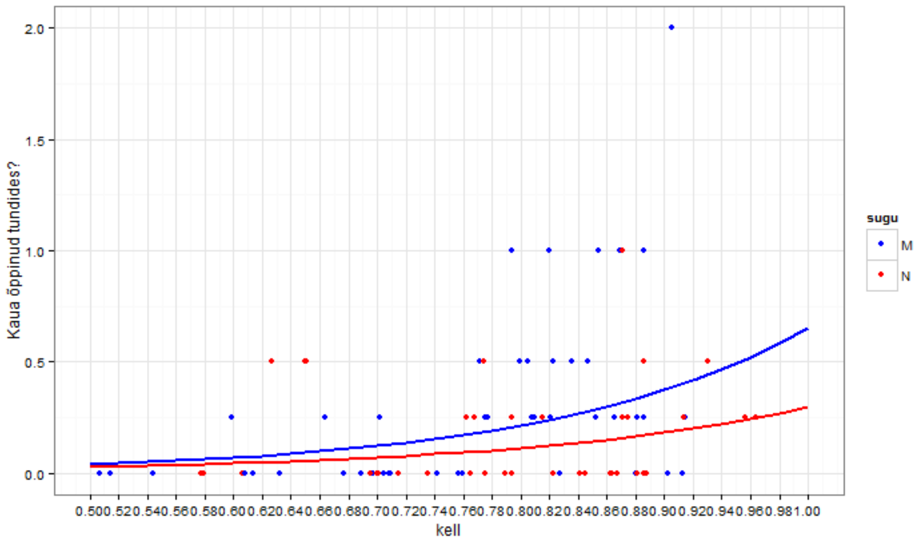
Sugude võrdlemine päevade kaupa



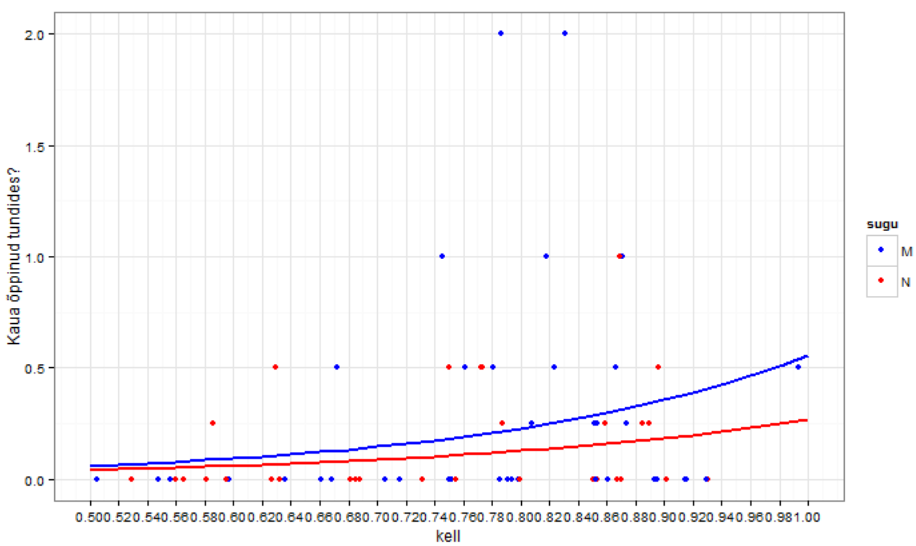
Joonis D.3: Esmaspäev



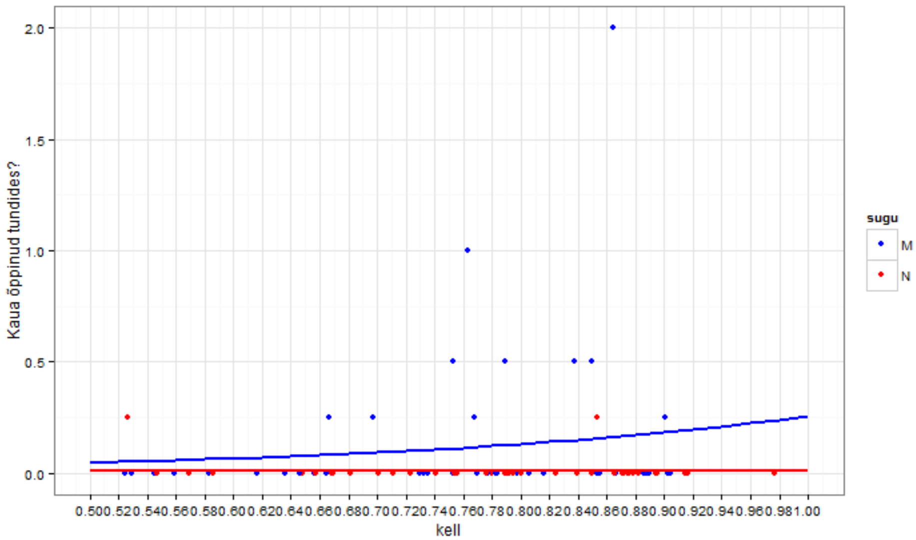
Joonis D.4: Teisipäev



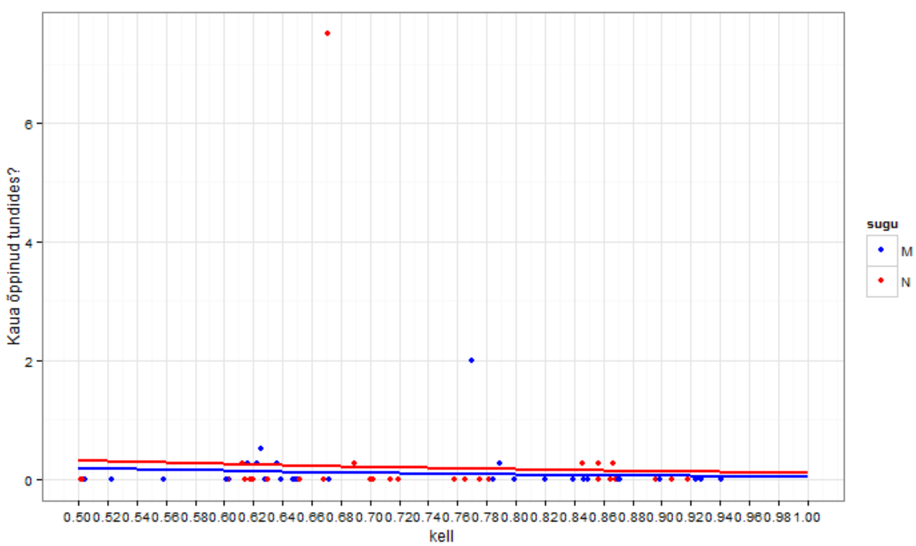
Joonis D.5: Kolmapäev



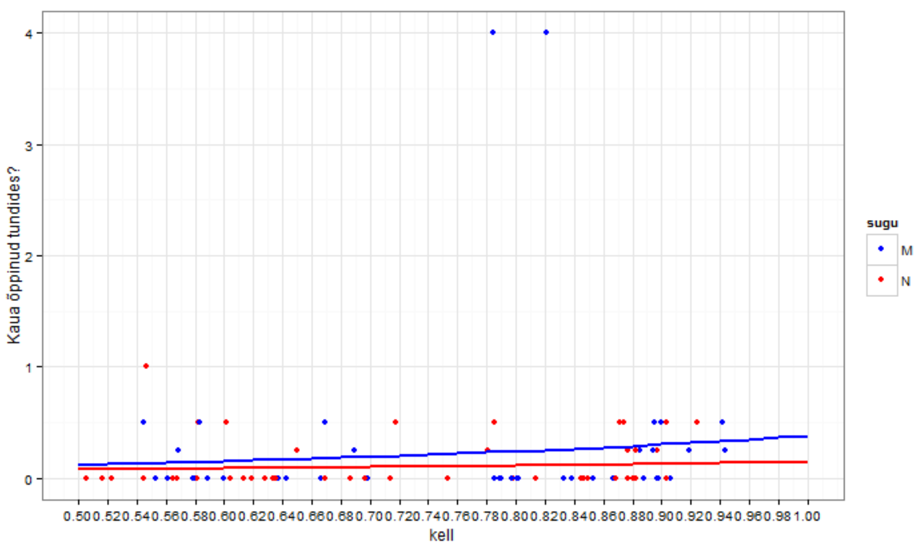
Joonis D.6: Neljapäev



Joonis D.7: Reede

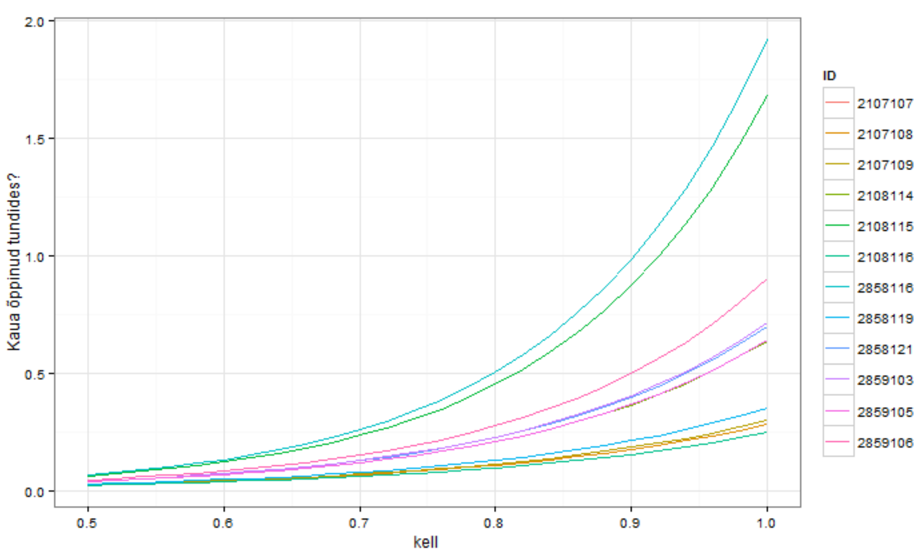


Joonis D.8: Laupäev



Joonis D.9: Pühapäev

Juhuslike laste võrdlemine kolmapäeval

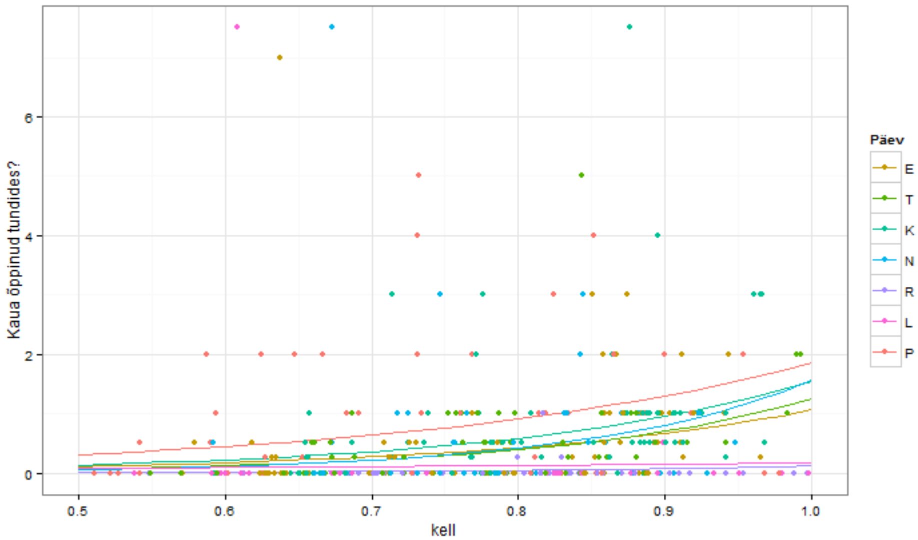


D.2 Seitsmes klass: R-i kood ja väljund

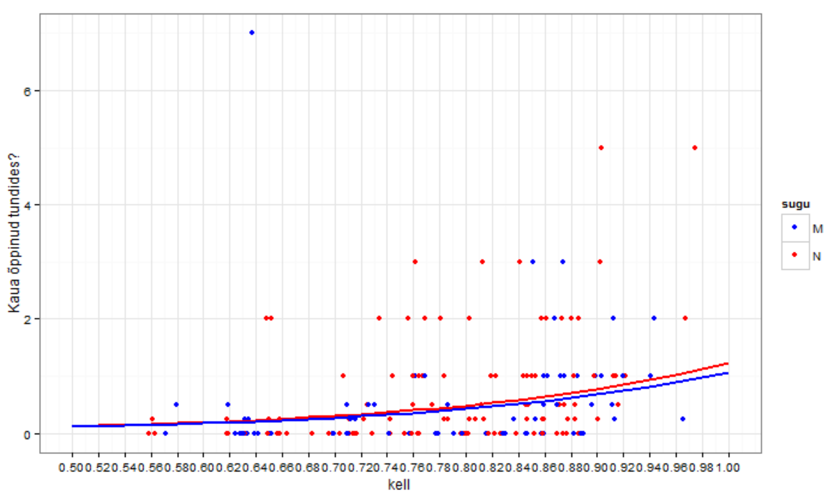
```
> fID_7 = factor(seitsmes3$ID)
> fdate_7 = factor(paste(seitsmes3$ID,seitsmes3$date))
> fKool_7 = factor(seitsmes3$Kool)
> mudel1_7 <- glmmadmb(kaua_oppinud_pidev~kell+nadalapaev+
+sugu:kell+ nadalapaev:kell+nadalapaev:kell:sugu+
+(kell-1|fKool_7/fID_7/fdate_7),
+ data=seitsmes3, family="nbinom1")
> summary(mudel1_7)

Call:
glmmadmb(formula = kaua_oppinud_pidev ~ kell + nadalapaev +
sugu:kell + nadalapaev:kell + nadalapaev:kell:sugu +
    (kell - 1 | fKool_7/fID_7/fdate_7),
    data = seitsmes3, family = "nbinom1")

AIC: 2206.9

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           -2.951264   0.668000   -4.42    1e-05 ***
kell                   3.570869   0.880010    4.06    5e-05 ***
nadalapaev1           -1.522579   1.235200   -1.23    0.218
nadalapaev2           -2.647042   1.210200   -2.19    0.029 *
nadalapaev3           -1.471957   1.114400   -1.32    0.187
nadalapaev4           -3.225191   1.620200   -1.99    0.047 *
nadalapaev5           -3.928525   2.418400   -1.62    0.104
nadalapaev6           -0.350090   2.187600   -0.16    0.873
kell:suguN            -0.768311   0.337820   -2.27    0.023 *
kell:nadalapaev1       0.961474   1.536400    0.63    0.531
kell:nadalapaev2       2.256287   1.518700    1.49    0.137
kell:nadalapaev3       1.284085   1.375600    0.93    0.351
kell:nadalapaev4       3.054325   1.971800    1.55    0.121
kell:nadalapaev5       1.179758   2.943700    0.40    0.689
kell:nadalapaev6      -2.041834   2.377700   -0.86    0.390
kell:nadalapaev1:suguN  0.919219   0.414850    2.22    0.027 *
kell:nadalapaev2:suguN  1.082418   0.423590    2.56    0.011 *
kell:nadalapaev3:suguN  0.633657   0.394950    1.60    0.109
kell:nadalapaev4:suguN  0.000445   0.477700    0.00    0.999
kell:nadalapaev5:suguN  1.812961   0.878130    2.06    0.039 *
kell:nadalapaev6:suguN  0.721885   0.956420    0.75    0.450
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of observations: total=1280, fKool_7=7,
fKool_7:fID_7=167, fKool_7:fID_7:fdate_7=744
Random effect variance(s):
Group=fKool_7
      Variance   StdDev
kell 2.061e-09 4.54e-05
Group=fKool_7:fID_7
     Variance StdDev
kell    0.765 0.8747
Group=fKool_7:fID_7:fdate_7
     Variance StdDev
kell   0.2796 0.5287
```

```
Negative binomial dispersion parameter: 1.0759
(std. err.: 0.073255)

Log-likelihood: -1078.43
> ranef(mudel1_7)
$fKool_7
            kell
13  2.324775e-09
15  3.521077e-09
21 -5.485281e-09
29  1.133377e-08
58  1.634143e-08
59  1.761673e-09
60  1.219047e-08

$`fKool_7:fID_7`
                  kell
13:113105 -0.003320734
13:113111 -0.412218480
13:113113  0.384098163
13:113120  0.394716535
13:113121 -0.118254023
...
$`fKool_7:fID_7:fdate_7`
                                     kell
13:113105:113105 2014-04-15  0.0498030735
13:113105:113105 2014-04-16 -0.0194574236
13:113105:113105 2014-04-17 -0.0315590952
13:113111:113111 2014-04-15 -0.0202743182
13:113111:113111 2014-04-16 -0.0907308124
...
```

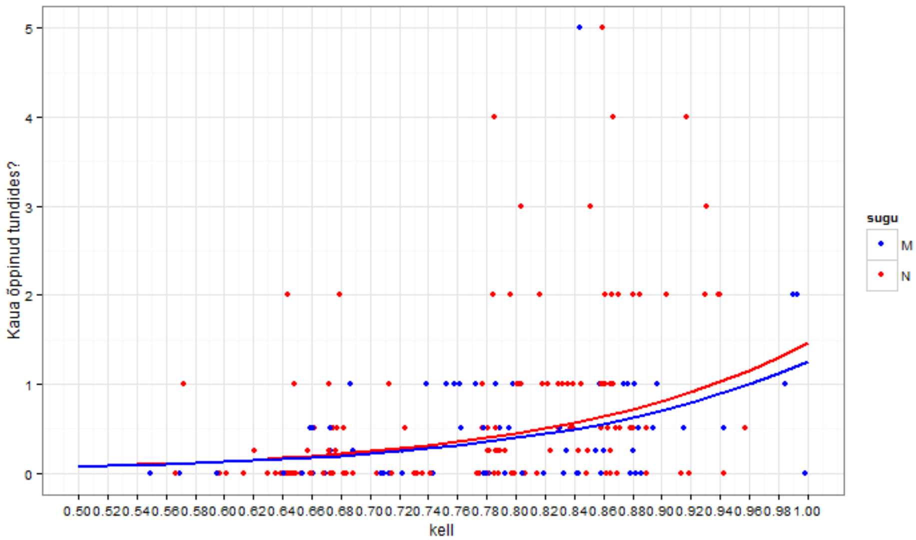
D.3 Seitsmes klass: poistele hinnatud mudelid



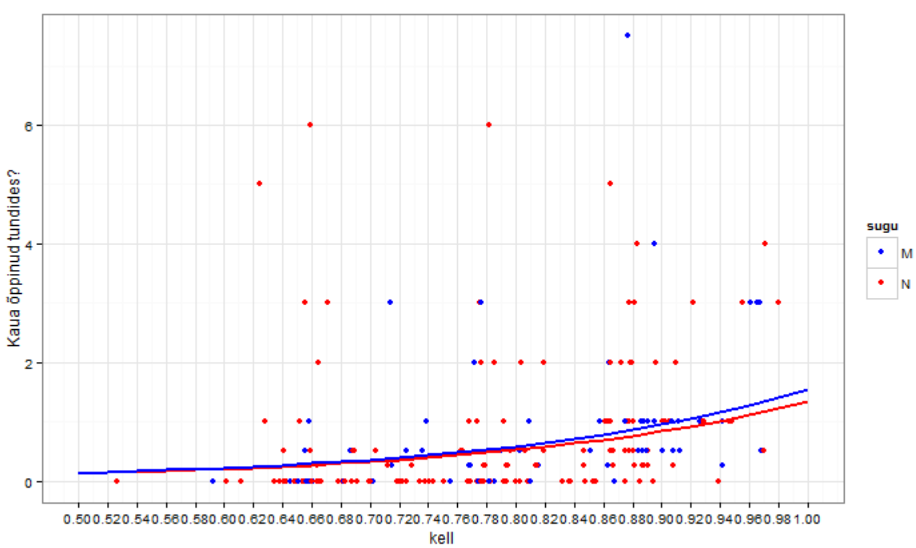
D.4 Seitsmes klass: sugude võrdlemine



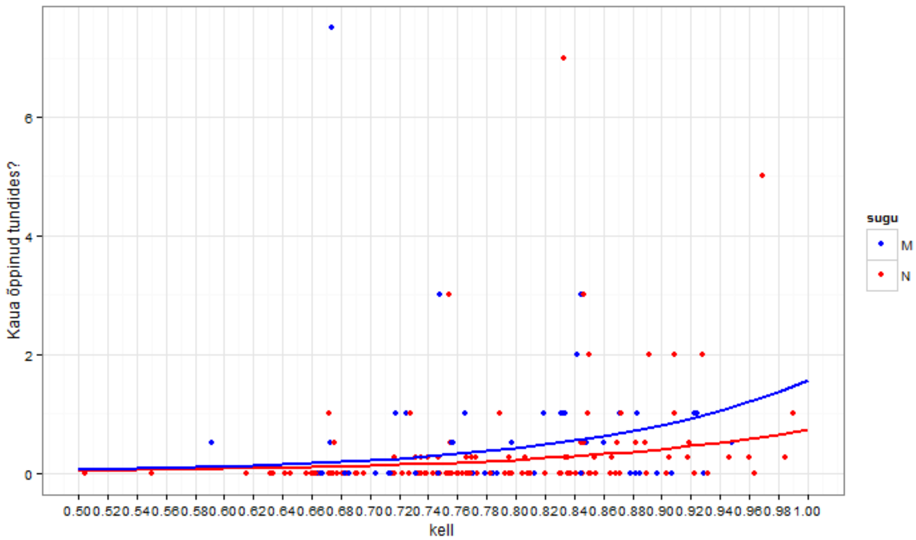
Joonis D.10: Esmaspäev



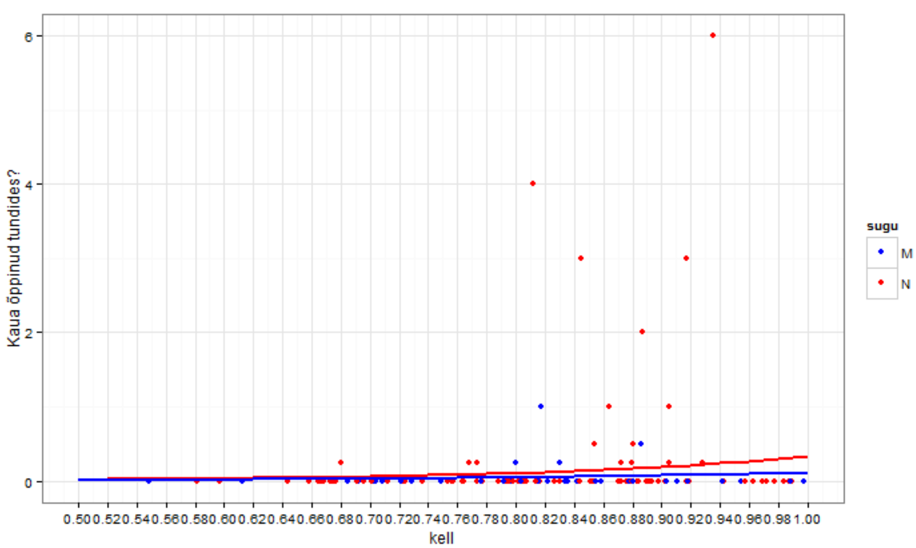
Joonis D.11: Teisipäev



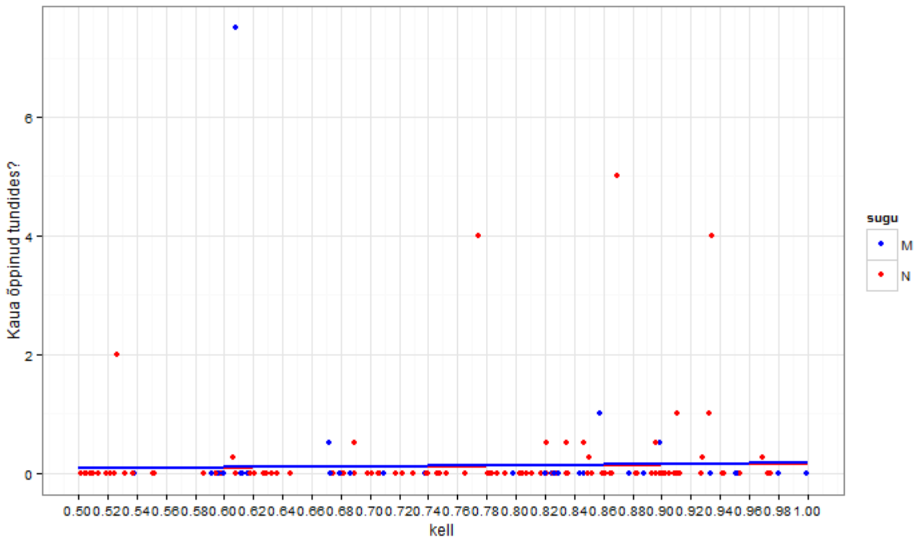
Joonis D.12: Kolmapäev



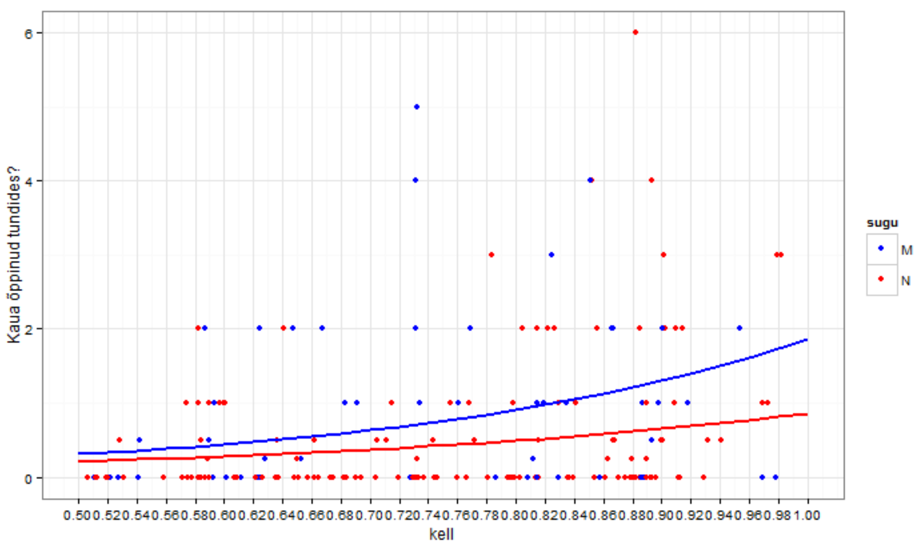
Joonis D.13: Neljapäev



Joonis D.14: Reede



Joonis D.15: Laupäev



Joonis D.16: Pühapäev

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Astrid Haas

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Üldistatud lineaarne segamudel ESM-uuringu andmetele", mille juhendaja on Märt Möls

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

   (b) 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, 13.05.2015