

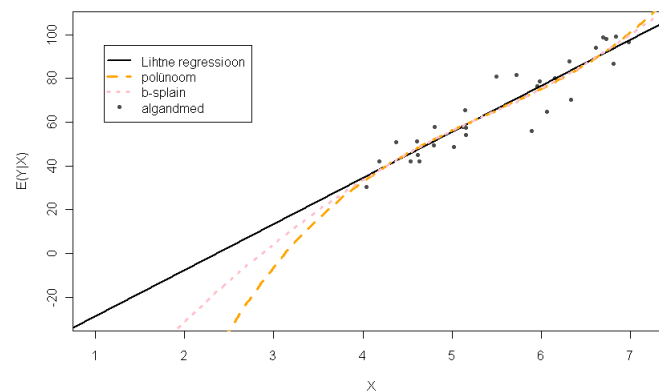
Biomeetria bioloogidele

loeng 6

Üldistatud lineaarne mudel I (*Generalized Linear Model – GLM*)

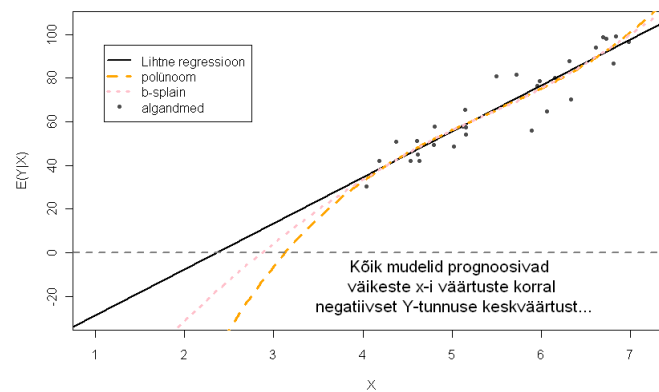
Poissoni regressioon

Automaatse modelleerimise piirid



2

Automaatse modelleerimise piirid



3

Kuidas öelda mudelile, et keskvaartus ei tohi olla negatiivne?

Mudeli

$$E(Y|X) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots$$

asemel kasutame mudelit

$$\ln(E(Y|X)) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots$$

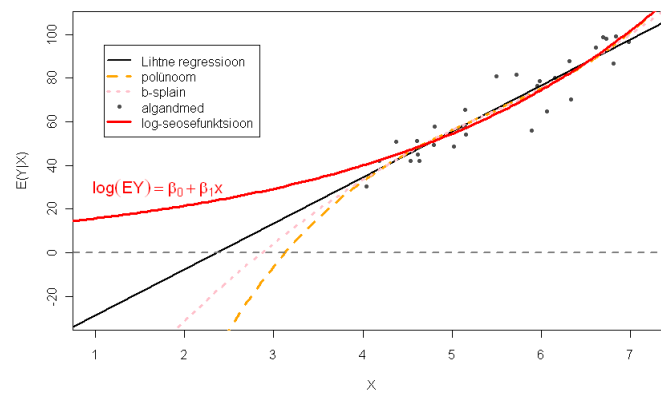
ehk

$$E(Y|X) = \exp(\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots).$$

Siin pole enam tegemist lineaarse mudeliga (küll aga millegi sarnasega – seega üldistatud lineaarne mudel ehk *generalized linear model*)

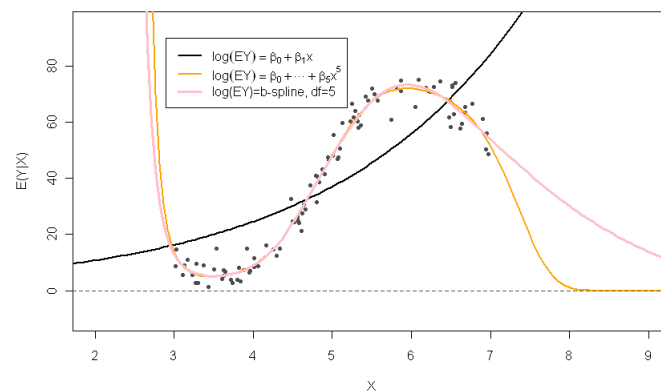
4

Modelleerides $\log(EY)$ -t saab olla kindel, et keskvärtuse hinnang (regressioonisirge) püsib igal juhul mittenegatiivne.



5

Log-seosefunktsiooni kasutamine ei tähenda, et polünoomide/splainide/... kasutamisest pääseksime...



6

Seosefunktsioon

Kui kasutame mudelit

$$g(EY) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$$

Siis funktsiooni $g()$ kutsutakse seosefunktsiooniks (link-function).

Seosefunktsiooni üheks peamiseks rolliks on edasi anda mudelile/arvutile meile teadaolevat lisainformatsiooni keskvärtuse kohta (näiteks seda, et keskvärtus ei saa olla negatiivne).

7

Seosefunktsioonide näiteid

- Keskvärtus ei tohi olla negatiivne:

$$g(EY) = \log(EY)$$

$$\rightarrow EY = \exp(\beta_0 + \beta_1 x_1 + \dots)$$

- Keskvärtus peab olema vahemikus 0..1:

$$g(EY) = \log(EY/(1-EY))$$

$$\rightarrow EY = \exp(\beta_0 + \beta_1 x_1 + \dots) / (1 + \exp(\beta_0 + \beta_1 x_1 + \dots))$$

$$g(EY) = \Phi(EY)$$

$\Phi()$ - normaaljaotuse jaotusfunktsioon

$$\rightarrow EY = \Phi^{-1}(\beta_0 + \beta_1 x_1 + \dots)$$

8

Üldistatud lineaarne mudel – generalized linear model (glm)

- Lineaarne mudel:

$$E(Y|X) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots$$

- Üldistatud lineaarne mudel:

$$g(E(Y|X)) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots$$

Lineaarse mudeli parameetreid suudab arvuti alati hinnata (kuigi hinnang ei pruugi alati olla üheselt määratud).

Üldistatud lineaarse mudeli parameetrite hindamine on tehniliselt raskem ülesanne. Arvuti võib vahel parameetrite hindamisel jääda hätta ja paluda kasutaja abi. Enamasti (kui algandmetega pole probleeme) see siiski õnnestub.

9

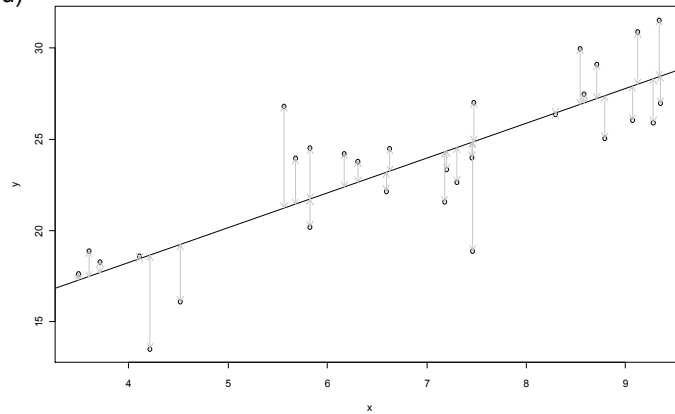
Meeldetuletuseks

- Tundmatut seost tunnuste vahel on võimalik modelleerida näiteks polünoomide või splineide abil (+palju muid võimalusi);
- Kui uuritava tunnuse kohta on midagi teada, siis vastavat lisateavet saab kasutada valides mudeli jaoks sobiva seosefunktsiooni (*link-function*);

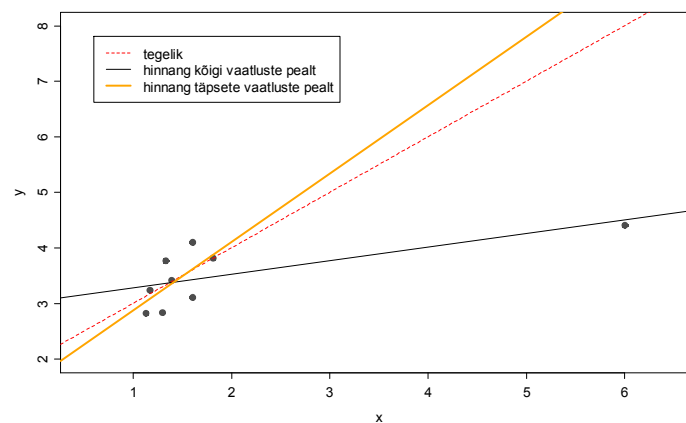
10

Mudeli parameetrite hindamisest

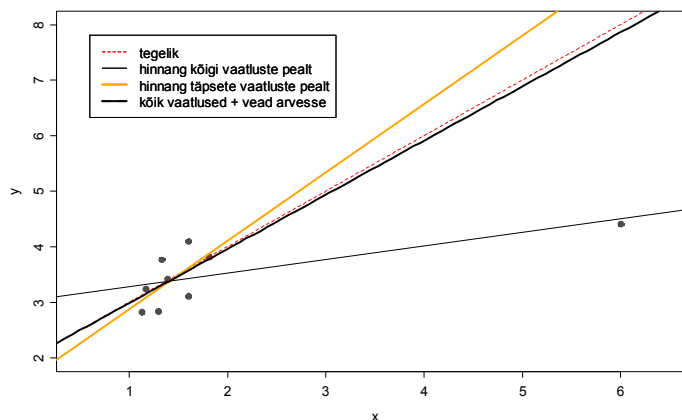
Mudeli tundmatute parameetrite väärtused valitakse selliselt, et tekkivate prognoosivigade (jääkide) ruutude summa oleks minimaalne (vaadatakse ajalooliste andmete prognoosimisel tekkivaid vigu)



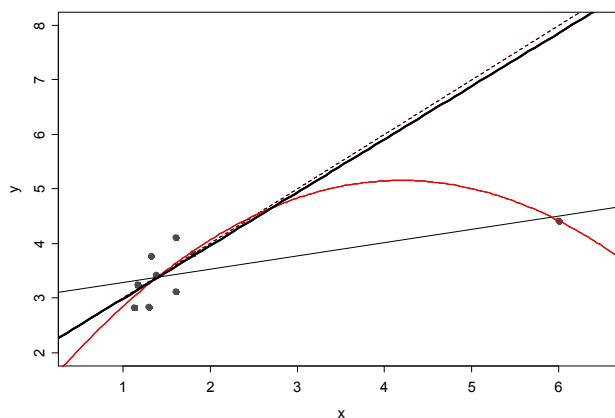
Ebatäpsete vaatluste probleem



Ebatäpsete vaatluste probleem



Ebatäpsete vaatluste probleem – ka statistilised testid võivad hakata eksitama...



Probleem

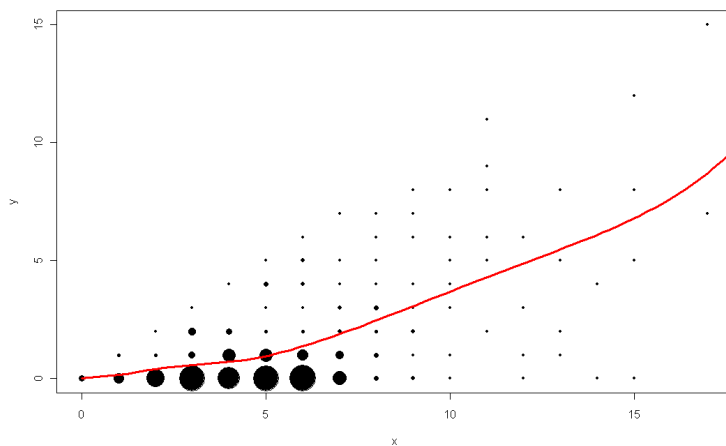
Selleks, et saaks kasutada informatsiooni vaatluste täpsuse kohta, on vaja mudelit vaatluste hajuvuse (dispersiooni) jaoks.

Dispersiooni modelleerimine on aga enamasti üsna raske ja ebatäpne töö...

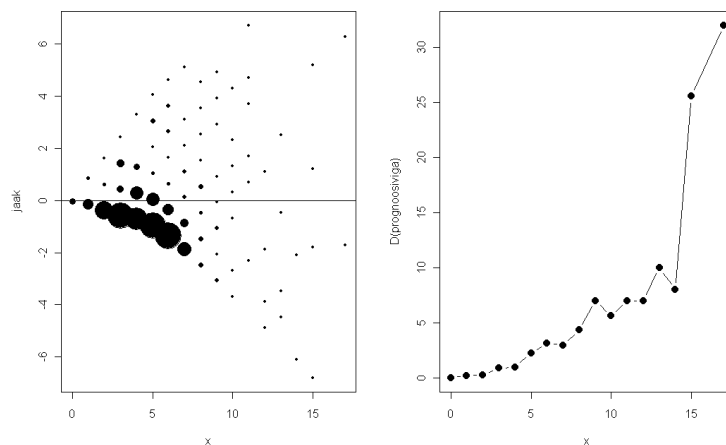
Lihtsad mudelid dispersioonile

- Lihtne regressioon/dispersioonanalüüs:
Kõik vaatlused sama täpsed
- Näide (ühest) alternatiivisest võimalusest:
Vaatluste dispersioon ("täpsus") on seda väiksem, mida suurem on uuritava tunnuse keskväärts...

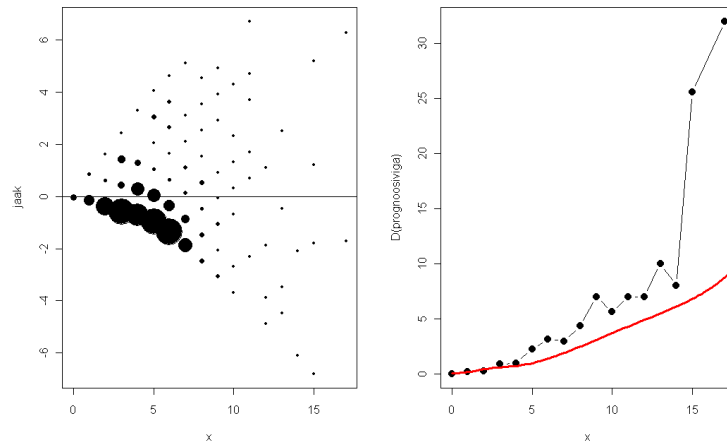
Tüüpiline näide



Jääkide hajuvus suureneb...



Jääkide hajuvus suureneb...



Üldistatud lineaarne mudel (Generalized Linear Model, GLM)

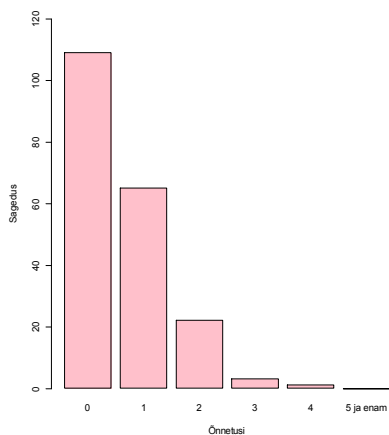
Üldistatud lineaarne mudel \approx

seosefunktsiooni kasutamine
(lisainformatsioon keskvärtuse kohta)

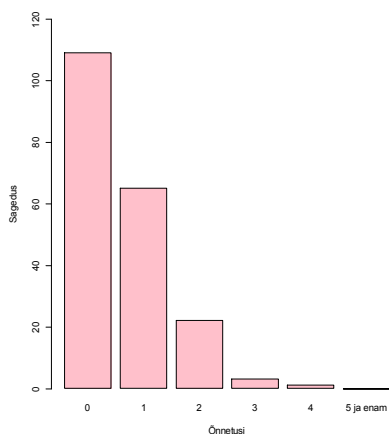
+

mudel vaatluste hajuvuse ehk
dispersiooni kohta

Poissoni jaotus



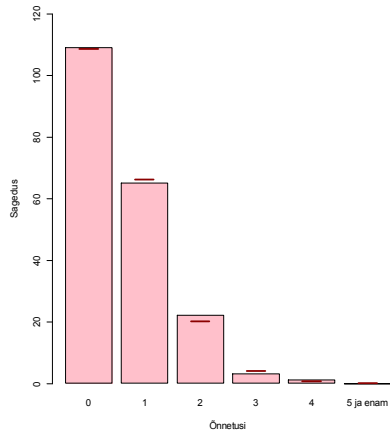
Poissoni jaotus



$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

λ - keskmine õnnetuste arv

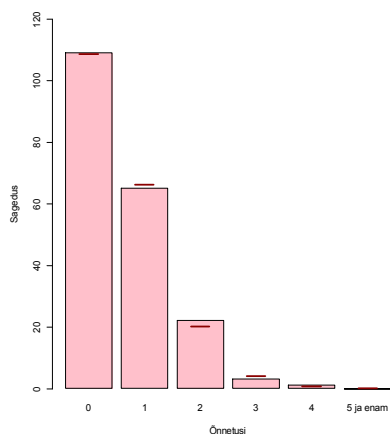
Poissoni jaotus



$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

λ - keskmine õnnetuste arv

Poissoni jaotus



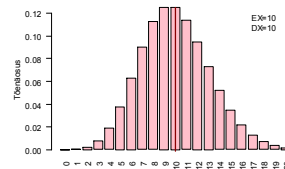
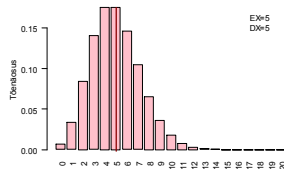
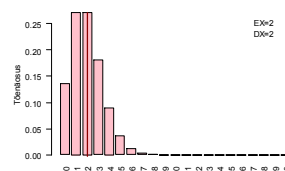
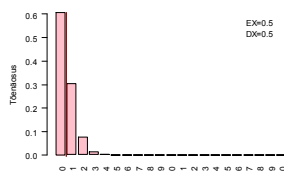
Eeldused:

- Õnnetused toimuvad teineteisest sõltumatult;
- Ohvitsere on igas sõjaväeüksuses ligikaudu sama palju – oodatav õnnetuste arv võiks olla sarnane kõigis üksustes

Poissoni jaotus

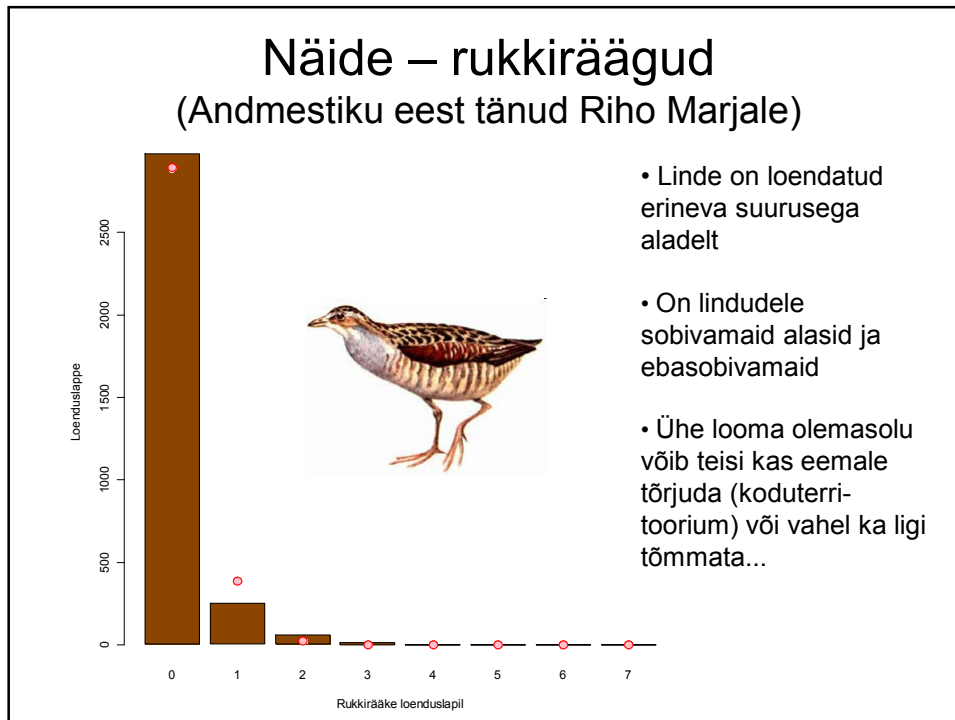
Uuritava tunnuse dispersioon on sama mis keskväärtus.

Kui keskväärtus suureneb, suureneb ka dispersioon



Poissoni regressioon (Log-lineaarne mudel)

- Kui uuritav tunnus on saadud millegi loendamisel (diskreetne tunnus), siis peab uuritava tunnuse keskväärtus olema mittenegatiivne (-> kasutame log-seosefunktsiooni)
- Poissoni jaotuse puhul oli uuritava tunnuse dispersioon võrdne keskväärtusega. Suurema keskväärtusega vaatluse puhul muretsetakse prognoosivea pärast vähem kui väiksema keskväärtusega vaatluse puhul – suure keskväärtuse korral on vaatluse kaal parameetrite hindamisel väiksem.



Konstrueerime regressiooni

$$\log(E(\text{rääke})) = \beta_0 + \beta_1 X$$

$$E(\text{rääke}) = \exp(\beta_0 + \beta_1 X)$$

~~$$E(\text{rääke}) = \exp(\beta_0 + \beta_1 \text{pindala})$$~~

~~$$= \exp(\beta_0) * \exp(\text{pindala})^{\beta_1}$$~~

$$E(\text{rääke}) = \exp(\beta_0 + \beta_1 \log(\text{pindala}))$$

$$= \exp(\beta_0) * \exp(\log(\text{pindala}))^{\beta_1}$$

$$= c_0 * \text{pindala}^{\beta_1}$$

$$= c_0 * \text{pindala}, \text{ kui } \beta_1=1$$

Variante:

```
> summary(glm(raak~log(pindala), family=poisson()))

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.82709     0.12687  -30.16  <2e-16 ***
log(pindala)   0.87901     0.04326   20.32  <2e-16 ***

> summary(glm(raak~log(pindala)+ ..., family=poisson()))

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.251e+00  2.810e-01 -15.129 < 2e-16 ***
log(pindala)   9.763e-01  5.989e-02  16.301 < 2e-16 ***
.....
```

Aga kaks korda suuremal alal elab ju keskmiselt
kaks korda rohkem loomi? Miks ma selles üldse
kahtlen?

```
> summary(glm(raak~1, offset=log(pindala),
              family=poisson()))

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.16475     0.04746  -87.76  <2e-16 ***
```

$$\log(E(\text{rääke})) = \beta_0 + \log(\text{pindala})$$

$$E(\text{rääke}) = c_0 * \text{pindala}$$

$$\exp(-4.16475) = 0.0155336$$

rukkiräaku elab keskmiselt ühel pindalaühikul (hektaril)

Tüüpilisi offset-muutuja näiteid

- Kui suurelt alalt loomi/taimi loendati
offset=log(*pindala*)
- Kui pikk oli vaatlusperiood
offset=log(*vaatlusaeg*)
- Karja suurus, kus õnnetusjuhtumeid või sündmuseid loendati
offset=log(*loomi*)
- Või paljude eeltoodute kombinatsioon:
offset=log(*pindala*)+log(*vaatlusaeg*)

Lisame tunnuseid mudelile...

```
> table(toetustyypp)
toetustyypp
  ei   ksm mahe  ypt
1104  866  256 1082
```

mahe	– mahetoetus (kõige rangemad keskkonnanõuded);
ksm	– keskkonnasõbralik majandamine (keskmised nõuded);
ypt	– ühtne pindalatoetus (keskkonnanõuded otseselt ei ole);
ei	– toetust alale ei maksta või ei ole seda taotletud vastavale alale/põllule.

Lisame tunnuseid mudelile...

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
          family=poisson())
> summary(m2)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.94497    0.11868 -33.241 < 2e-16 ***
factor(toetustyypp) ksm  -0.95392    0.15760  -6.053 1.42e-09 ***
factor(toetustyypp) mahe -0.02797    0.18805  -0.149  0.882
factor(toetustyypp) ypt   0.19152    0.13556   1.413  0.158
factor(toetustyypp) ei     0          .          .          .

Kui toetustüüp="ksm"
log(E(raak)) = -3.94497 -0.95392 + log(pindala)
E(raak) = exp(-3.94497 -0.95392 + log(pindala))
         = 0.007454.. * pindala
```

```
> predict(m2, data.frame(toetustyypp="ksm", pindala=1), type="response")
1
0.007454874
```

Lisame tunnuseid mudelile...

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
          family=poisson())
> summary(m2)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.94497    0.11868 -33.241 < 2e-16 ***
factor(toetustyypp) ksm  -0.95392    0.15760  -6.053 1.42e-09 ***
factor(toetustyypp) mahe -0.02797    0.18805  -0.149  0.882
factor(toetustyypp) ypt   0.19152    0.13556   1.413  0.158

Kui toetustüüp="mahe"
log(E(raak)) = -3.94497 -0.02797 + log(pindala)
E(raak) = exp(-3.94497 -0.02797 + log(pindala))
         = 0.01881803.. * pindala
```

```
> predict(m2, data.frame(toetustyypp="mahe", pindala=1),
          type="response")
1
0.01881807
```

Lisame tunnuseid mudelile...

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
          family=poisson())
> summary(m2)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.94497    0.11868 -33.241 < 2e-16 ***
factor(toetustyypp) ksm  -0.95392    0.15760  -6.053 1.42e-09 ***
factor(toetustyypp) mahe -0.02797    0.18805  -0.149  0.882
factor(toetustyypp) ypt   0.19152    0.13556   1.413  0.158
```

```
Kui toetustüüp="ypt"
log(E(raak)) = -3.94497 +0.19152 + log(pindala)
E(raak) = exp(-3.94497 +0.19152 + log(pindala))
          = 0.02343675.. * pindala
```

```
> predict(m2, data.frame(toetustyypp="ypt", pindala=1), type="response")
1
0.02343682
```

Lisame tunnuseid mudelile...

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
          family=poisson())
> summary(m2)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.94497    0.11868 -33.241 < 2e-16 ***
factor(toetustyypp) ksm  -0.95392    0.15760  -6.053 1.42e-09 ***
factor(toetustyypp) mahe -0.02797    0.18805  -0.149  0.882
factor(toetustyypp) ypt   0.19152    0.13556   1.413  0.158
factor(toetustyypp) ei    0          .          .          .
```

```
Kui toetustüüp="ei"
log(E(raak)) = -3.94497 +0 + log(pindala)
E(raak) = exp(-3.94497 +0 + log(pindala))
          = 0.0193518.. * pindala
```

```
> predict(m2, data.frame(toetustyypp="ei", pindala=1), type="response")
1
0.01935185
```

Parameetrite tähendus I

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
          family=poisson())
```

```
> summary(m2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.94497	0.11868	-33.241	< 2e-16 ***
factor(toetustyypp) ksm	-0.95392	0.15760	-6.053	1.42e-09 ***
factor(toetustyypp) mahe	-0.02797	0.18805	-0.149	0.882
factor(toetustyypp) ypt	0.19152	0.13556	1.413	0.158

Üldist pindalatoetust (ypt) saavatel aladel on rukkiräake

$\exp(0.1915) = 1.211$ korda rohkem kui samasuurtel ilma toetuseta aladel.

Mahepõllunduse toetust saavatel aladel on rukkiräake

$\exp(-0.02797) = 0.972$ korda "rohkem" kui samasuurtel ilma toetuseta aladel.

Rusikareegel:

Kui x on väike, siis $\exp(x) \approx 1+x$

Võrdlustaseme muutus

```
> m2a=glm(raak~relevel(factor(toetustyypp), ref="ypt"), offset=log(pindala),
          family=poisson())
> summary(m2a)
```

Call:

```
glm(formula = raak ~ relevel(factor(toetustyypp), ref = "ypt"),
     family = poisson(), offset = log(pindala))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7609	-0.4891	-0.2931	-0.1669	4.4184

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.75345	0.06551	-57.294	<2e-16
relevel(factor(toetustyypp), ref = "ypt")ei	-0.19152	0.13556	-1.413	0.158
relevel(factor(toetustyypp), ref = "ypt")ksm	-1.14544	0.12266	-9.339	<2e-16
relevel(factor(toetustyypp), ref = "ypt")mahe	-0.21949	0.15990	-1.373	0.170

Kaks erinevat testi

```
> m2=glm(raak~factor(toetustyypp), offset=log(pindala),
+ family=poisson())
```

```
> summary(m2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.94497	0.11868	-33.241	< 2e-16 ***
factor(toetustyypp) ksm	-0.95392	0.15760	-6.053	1.42e-09 ***
factor(toetustyypp) mahe	-0.02797	0.18805	-0.149	0.882
factor(toetustyypp) ypt	0.19152	0.13556	1.413	0.158

Testib rääkude arvu (keskväärtuse) erinevust mahetoetust saavatel ja toetuseta aladel (võrdlustase)

```
> drop1(m2, test="Chisq")
```

Single term deletions

Model:

```
raak ~ factor(toetustyypp)
```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1527.8	2261.5		
factor(toetustyypp)	3	1632.2	2359.9	104.36	< 2.2e-16 ***

Testib, kas mõne toetustüübi korral on keskmine rukkirääkude asutustihedus suurem kui mõne teise toetustüübi korral.

Täiendame mudelit...

```
> m3=glm(raak~factor(toetustyypp)+factor(hairing), offset=log(pindala),
family=poisson())
```

```
> summary(m3)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.95225	0.25419	-19.482	< 2e-16 ***
factor(toetustyypp) ksm	-0.92263	0.15765	-5.852	4.85e-09 ***
factor(toetustyypp) mahe	0.03998	0.18827	0.212	0.8318
factor(toetustyypp) ypt	0.24906	0.13578	1.834	0.0666 .
factor(hairing)hairingutpole	1.04258	0.22927	4.547	5.43e-06 ***

```
> table(hairing)
hairing
      hairing hairingutpole
      267             3041
```

Exp(1.042..) = 2.836..
Häiringuteta aladel on 2,8 korda enam rukkirääke kui sama toetustüübiga samasuurtel häiringuga aladel...

Koosmõjust I. Peamõjud (*Main effects*)

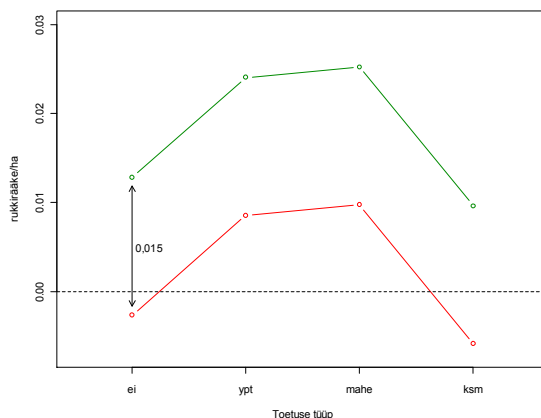
Tavaline lineaarne mudel

$$E \text{ rääk/pindala} = \mu + \alpha_{\text{toetustüüp}} + \beta_{\text{häiring}}$$

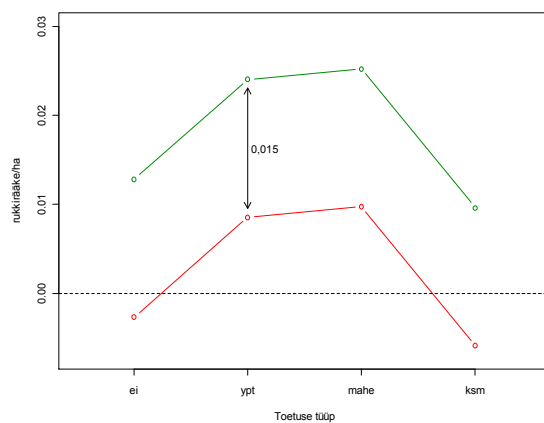
μ	=	-0.0026		
α_{ksm}	=	-0.0032	$\beta_{\text{häiringut pole}}$	= 0.015
α_{mahe}	=	0.0124	$\beta_{\text{häiring}}$	= 0 (võrdlustase)
α_{ypt}	=	0.0112		
α_{ei}	=	0 (võrdlustase)		

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.002640	0.004826	-0.547	0.584415
factor(toetustüüp) ksm	-0.003194	0.003163	-1.010	0.312726
factor(toetustüüp) mahe	0.012404	0.004831	2.567	0.010294 *
factor(toetustüüp) ypt	0.011198	0.002996	3.738	0.000189 ***
factor(häiring) häiringut pole	0.015486	0.004476	3.460	0.000547 ***

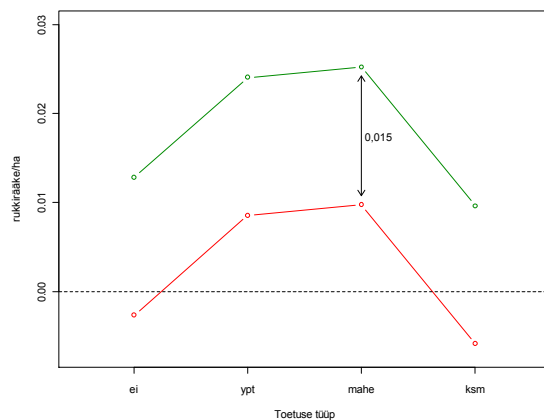
Peamõjudega tavaline lineaarne mudel (ANOVA)



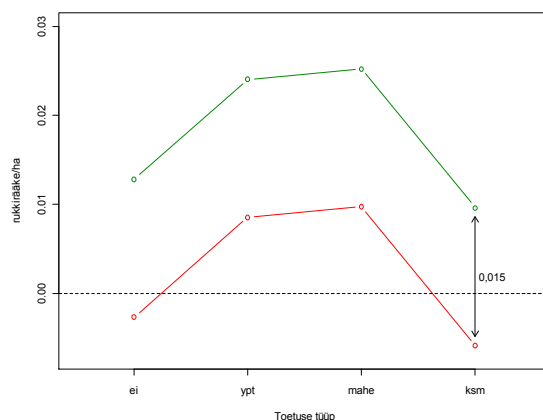
Peamõjudega tavaline lineaarne mudel (ANOVA)



Peamõjudega tavaline lineaarne mudel (ANOVA)



Peamõjudega tavaline lineaarne mudel (ANOVA)



Koosmõjust I. Koosmõjud (*interaction*) Tavaline lineaarne mudel

$$E \text{ rääk/pindala} = \mu + \alpha_{\text{toetustüüp}} + \beta_{\text{häiring}} + (\alpha\beta)_{\text{toetustüüp, häiring}}$$

$$\mu = -0.0026$$

$$\alpha_{ksm} = -0.0032$$

$$\alpha_{mahe} = 0.0124$$

$$\alpha_{ypt} = 0.0112$$

$$\alpha_{ei} = 0 \text{ (võrdlustase)}$$

$$(\alpha\beta)_{ksm, pole} = -0.00317$$

$$(\alpha\beta)_{mahe, pole} = 0.01497$$

$$(\alpha\beta)_{ypt, pole} = 0.00069$$

$$(\alpha\beta)_{ei, pole} = 0$$

$$\beta_{\text{häiringut pole}} = 0.015$$

$$\beta_{\text{häiring}} = 0 \text{ (võrdlustase)}$$

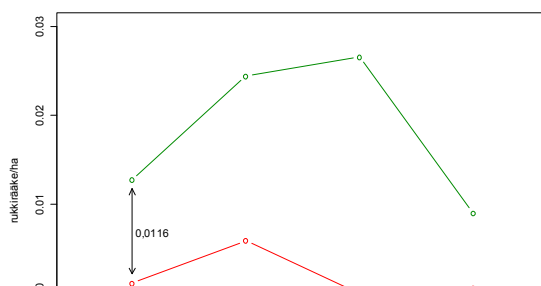
$$(\alpha\beta)_{ksm, häiring} = 0$$

$$(\alpha\beta)_{mahe, häiring} = 0$$

$$(\alpha\beta)_{ypt, häiring} = 0$$

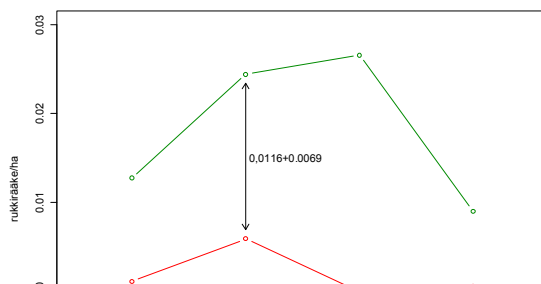
$$(\alpha\beta)_{ei, häiring} = 0$$

Koosmõjudega tavaline lineaarne mudel (ANOVA)



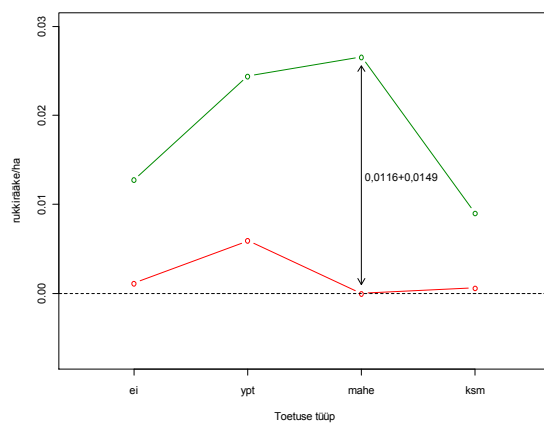
(Intercept)	0.0011620340	
factor (toetustyypp) ksm	-0.0005550818	
factor (toetustyypp) mahe	-0.0011620340	
factor (toetustyypp) ypt	0.0047687493	
factor (hairing) hairingutpole	0.0115739846	
factor (toetustyypp) ypt: factor (hairing) hairingutpole	0.0068889103	
factor (toetustyypp) mahe: factor (hairing) hairingutpole	0.0149714617	
factor (toetustyypp) ksm: factor (hairing) hairingutpole	-0.0031756087	

Koosmõjudega tavaline lineaarne mudel (ANOVA)

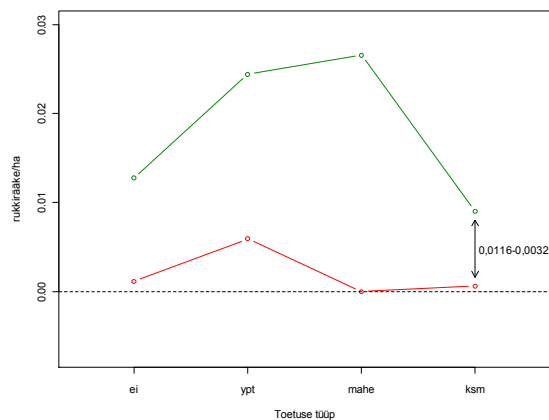


(Intercept)	0.0011620340	
factor (toetustyypp) ksm	-0.0005550818	
factor (toetustyypp) mahe	-0.0011620340	
factor (toetustyypp) ypt	0.0047687493	
factor (hairing) hairingutpole	0.0115739846	
factor (toetustyypp) ypt: factor (hairing) hairingutpole	0.0068889103	
factor (toetustyypp) mahe: factor (hairing) hairingutpole	0.0149714617	
factor (toetustyypp) ksm: factor (hairing) hairingutpole	-0.0031756087	

Koosmõjudega tavaline lineaarne mudel (ANOVA)



Koosmõjudega tavaline lineaarne mudel (ANOVA)



Koosmõjudeta Poissoni regressioon

```
> m3=glm(raak~factor(toetustyypp)+factor(hairing), offset=log(pindala),
family=poisson())
```

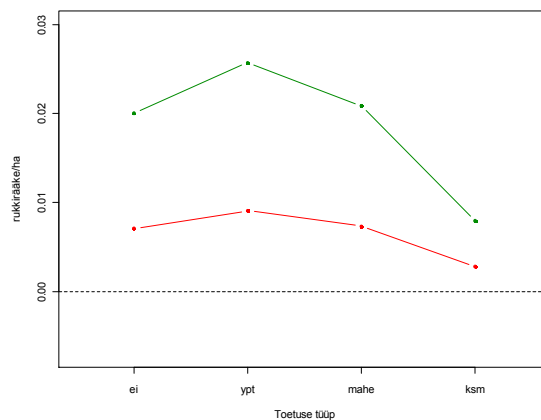
```
> summary(m3)
```

Coefficients:

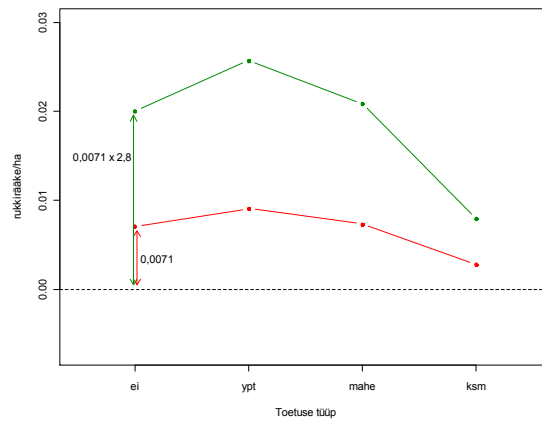
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.95225	0.25419	-19.482	< 2e-16 ***
factor(toetustyypp) ksm	-0.92263	0.15765	-5.852	4.85e-09 ***
factor(toetustyypp) mahe	0.03998	0.18827	0.212	0.8318
factor(toetustyypp) ypt	0.24906	0.13578	1.834	0.0666 .
factor(hairing)hairingutpole	1.04258	0.22927	4.547	5.43e-06 ***

Exp(1.042..) = 2.836..
Häiringuteta aladel on 2,8 korda enam
rukkiräake kui sama toetustüübiga
samasuurtel häiringuga aladel...

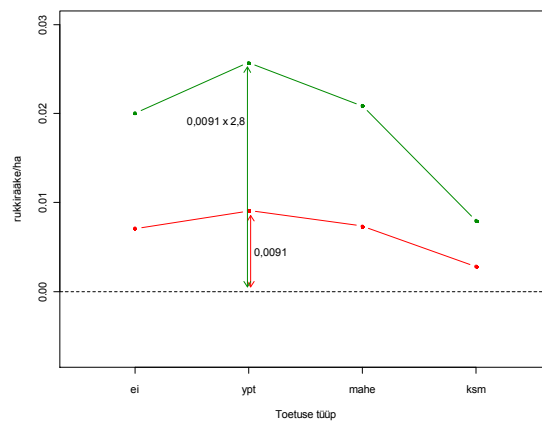
Koosmõjudeta Poissoni regressioon



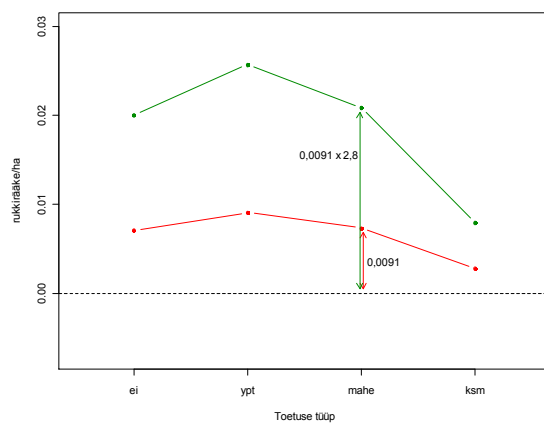
Koosmõjudeta Poissoni regressioon



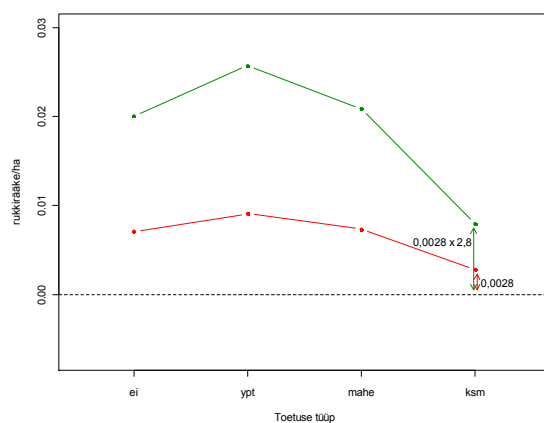
Koosmõjudeta Poissoni regressioon



Koosmõjudeta Poissoni regressioon



Koosmõjudeta Poissoni regressioon



Koosmõjudega Poissoni regressioon

```
> m4=glm(raak~factor(toetustyypp)+factor(hairing)+
          factor(toetustyypp)*factor(hairing),
          offset=log(pindala), family=poisson())
```

```
> summary(m4)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.2807	1.0000	-5.281	1.29e-07
factor(toetustyypp)ksm	-0.7445	1.1547	-0.645	0.519
factor(toetustyypp)mahe	-13.2456	328.7088	-0.040	0.968
factor(toetustyypp)ypt	0.8372	1.0308	0.812	0.417
factor(hairing)hairingutpole	1.3766	1.0071	1.367	0.172
factor(toetustyypp)ksm:factor(hairing)hairingutpole	-0.1783	1.1656	-0.153	0.878
factor(toetustyypp)mahe:factor(hairing)hairingutpole	13.3411	328.7088	0.041	0.968
factor(toetustyypp)ypt:factor(hairing)hairingutpole	-0.6104	1.0399	-0.587	0.557

Toetust ei saa, häiring, 1 hektarine maalapp:

$$E \text{ rääk} = \exp(-5.28 + 0 + 0 + 0 + \log(1)) = \exp(-5.28)$$

Toetust ei saa, häiringut pole, 1 hektarine maalapp:

$$E \text{ rääk} = \exp(-5.28 + 0 + 1.3766 + 0 + \log(1)) \\ = \exp(-5.28) * \exp(1.3766) = \exp(-5.28) * 3,96$$

Koosmõjudega Poissoni regressioon

```
> m4=glm(raak~factor(toetustyypp)+factor(hairing)+
          factor(toetustyypp)*factor(hairing),
          offset=log(pindala), family=poisson())
```

```
> summary(m4)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.2807	1.0000	-5.281	1.29e-07
factor(toetustyypp)ksm	-0.7445	1.1547	-0.645	0.519
factor(toetustyypp)mahe	-13.2456	328.7088	-0.040	0.968
factor(toetustyypp)ypt	0.8372	1.0308	0.812	0.417
factor(hairing)hairingutpole	1.3766	1.0071	1.367	0.172
factor(toetustyypp)ksm:factor(hairing)hairingutpole	-0.1783	1.1656	-0.153	0.878
factor(toetustyypp)mahe:factor(hairing)hairingutpole	13.3411	328.7088	0.041	0.968
factor(toetustyypp)ypt:factor(hairing)hairingutpole	-0.6104	1.0399	-0.587	0.557

Mahetoetus, häiring, 1 hektarine maalapp:

$$E \text{ rääk} = \exp(-5.28 -13.25 + 0 + 0 + \log(1)) = \exp(-5.28-13.25) \\ = 0,00000008...$$

Mahetoetus, häiringut pole, 1 hektarine maalapp:

$$E \text{ rääk} = \exp(-5.28 -13.25 + 1.3766 + 13.34 + \log(1)) \\ = \exp(-5.28-13.25) * \exp(14.72) = \exp(-5.28) * 2470670$$

Kui keskväärtus peaks olema 0...

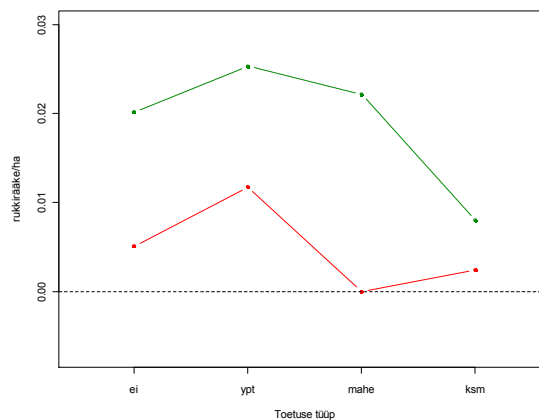
```
> m4=glm(raak~factor(toetustyypp)+factor(hairing)+
  factor(toetustyypp)*factor(hairing), offset=log(pindala),
  family=poisson(), epsilon=1e-14, maxit=100)

> coef(summary(m4))
```

	Estimate	z value	Pr(> z)
(Intercept)	-5.2807133	-5.280713e+00	1.286819e-07
factor(toetustyypp) ksm	-0.7444680	-0.7445	-6.447282e-01
factor(toetustyypp) mahe	-27.2455135	-13.2456	-7.558289e-05
factor(toetustyypp) ypt	0.8371654	0.8372	8.121697e-01
factor(hairing) hairingutpole	1.3766102	1.3766	1.366881e+00
factor(toetustyypp) ksm:factor(hairing) hairingutpole	-0.1783113	-0.1783	-1.529721e-01
factor(toetustyypp) mahe:factor(hairing) hairingutpole	27.3409750	13.3411	7.584771e-05
factor(toetustyypp) ypt:factor(hairing) hairingutpole	-0.6104007	-0.6104	-5.869797e-01

In addition: Warning messages:
1: glm.fit: fitted rates numerically 0 occurred

Koosmõjudega Poissoni regressioon



Koosmõjust võib õnneks antud näite puhul loobuda...

```
> m4=glm(raak~factor(toetustyypp)+factor(hairing)+
  factor(toetustyypp)*factor(hairing), family=poisson(),
  offset=log(pindala))
> drop1(m4, test="Chisq")
Model:
raak ~ factor(toetustyypp) + factor(hairing) + factor(toetustyypp) *
  factor(hairing)

```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1492.2	2233.9		
factor(toetustyypp):factor(hairing)	3	1499.2	2234.8	6.9876	0.07229 .

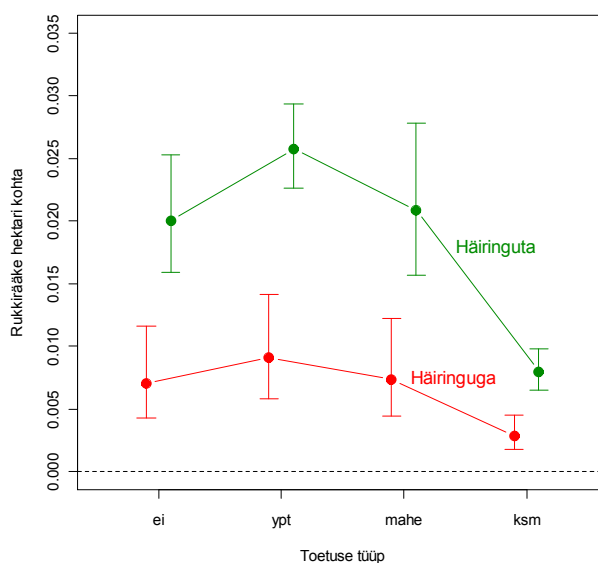
```
> m4v=glm(raak~factor(toetustyypp)+factor(hairing)+
  factor(toetustyypp)*factor(hairing), family=poisson(),
  offset=log(pindala))
> drop1(m4v, test="Chisq")
Model:
raak ~ factor(toetustyypp) + factor(hairing) + factor(toetustyypp) *
  factor(hairing)

```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1527.8	2261.5		
factor(toetustyypp)	3	1507.9	2257.2	19.504	< 2.2e-16 ***
factor(hairing)	1	1527.8	2261.5	28.657	8.642e-08 ***

```
> table(hairing, toetustyypp)
      toetustyypp
hairing      ei      ksm      mahe      ypt
hairing      31      79      30      127
hairingutpole 1073  787      226      955
```

Vahetulemus



Vahetulemuse joonistanud programm

```

m4v=glm(raak~factor(toetustyypp)+factor(hairing), family=poisson(), offset=log(pindala))

abi_hairinguta=predict(m4v, data.frame(toetustyypp=c("ei","ypt","mahe","ksm"),
  hairing=rep("hairingutpole",4), pindala=1), type="link", se.fit=TRUE)
abi_hairinguga=predict(m4v, data.frame(toetustyypp=c("ei","ypt","mahe","ksm"),
  hairing=rep("hairing",4), pindala=1), type="link", se.fit=TRUE)

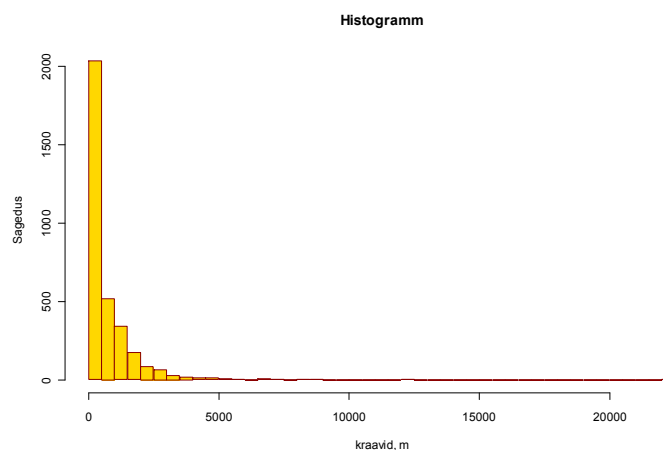
prognos_hairinguta=exp(abi_hairinguta$fit)
prognos_hairinguga=exp(abi_hairinguga$fit)

UI_hairinguta_alumine=exp(abi_hairinguta$fit-1.96*abi_hairinguta$se.fit)
UI_hairinguta_ylemine=exp(abi_hairinguta$fit+1.96*abi_hairinguta$se.fit)
UI_hairinguga_alumine=exp(abi_hairinguga$fit-1.96*abi_hairinguga$se.fit)
UI_hairinguga_ylemine=exp(abi_hairinguga$fit+1.96*abi_hairinguga$se.fit)

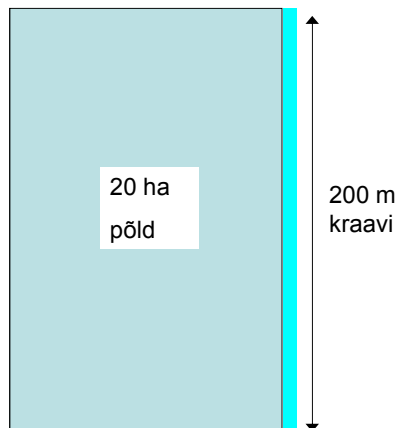
plot(1:4+0.1, prognos_hairinguta, col="green4", pch=20, cex=2, xlab="Toetuse tüüp",
  ylab="Rukkiräake hektari kohta", xaxt="n", ylim=c(0, 0.035), xlim=c(0.5,4.5),
  type="b")
points(1:4-0.1, prognos_hairinguga, col="red", pch=20, cex=2, type="b")
axis(1, at=1:4, c("ei","ypt","mahe","ksm"))
abline(h=0, lty=2)
arrows(1:4+0.1, UI_hairinguta_alumine, 1:4+0.1, UI_hairinguta_ylemine, code=3,
  angle=90, length=0.1, col="green4")
arrows(1:4-0.1, UI_hairinguga_alumine, 1:4-0.1, UI_hairinguga_ylemine, code=3,
  angle=90, length=0.1, col="red")
text(3.75, 0.018, "Häiringuta", col="green4", cex=1.2)
text(3.4, 0.0075, "Häiringuga", col="red", cex=1.2)

```

Pideva tunnuse lisamine

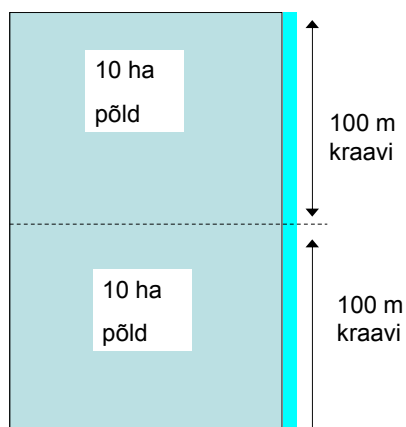


Millises skaalas lisada?



$$\begin{aligned}
 E \text{ rääke} &= \exp(\beta_0 + \beta_1 * \text{kraav} + \log(\text{pindala})) \\
 &= c_0 * \exp(\beta_1 * \text{kraav}) * \text{pindala} \\
 &= c_0 * \exp(\beta_1 * 200) * 20
 \end{aligned}$$

Millises skaalas lisada?



Enne kaheks jagamist:

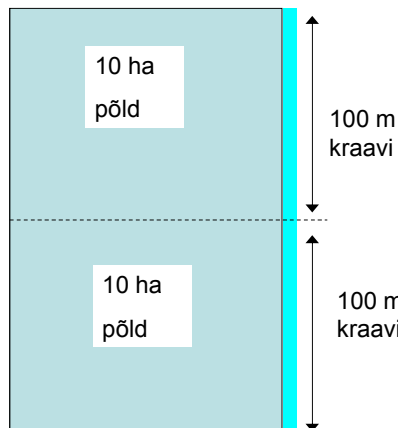
$$\begin{aligned}
 E \text{ rääke} &= \exp(\beta_0 + \beta_1 * \text{kraav} + \log(\text{pindala})) \\
 &= c_0 * \exp(\beta_1 * \text{kraav}) * \text{pindala} \\
 &= c_0 * \exp(\beta_1 * 200) * 20
 \end{aligned}$$

Pärast kaheks jagamist:

$$\begin{aligned}
 E \text{ rääke} &= \exp(\beta_0 + \beta_1 * \text{kraav} + \log(\text{pindala})) \\
 &= c_0 * \exp(\beta_1 * \text{kraav}) * \text{pindala} \\
 &= c_0 * \exp(\beta_1 * 100) * 10
 \end{aligned}$$

$$\begin{aligned}
 &2 * c_0 * \exp(\beta_1 * 100) * 10 \neq \\
 &c_0 * \exp(\beta_1 * 200) * 20
 \end{aligned}$$

Millises skaalas lisada?



Enne kaheks jagamist:

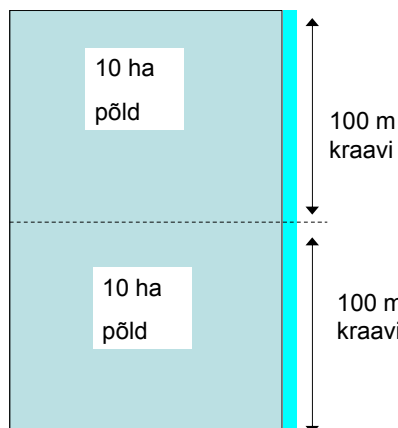
$$\begin{aligned}
 E \text{ rääke} &= \exp(\beta_0 + \beta_1 * (\text{kraav}/\text{pindala}) \\
 &\quad + \log(\text{pindala})) \\
 &= c_0 * \exp(\beta_1 * \text{kraav}/\text{pindala}) \\
 &\quad * \text{pindala} \\
 &= c_0 * \exp(\beta_1 * 200/20) * 20
 \end{aligned}$$

Pärast kaheks jagamist:

$$\begin{aligned}
 E \text{ rääke} &= c_0 * \exp(\beta_1 * \text{kraav}/\text{pindala}) * \text{pindala} \\
 &= c_0 * \exp(\beta_1 * 100/10) * 10
 \end{aligned}$$

$$\begin{aligned}
 2 * c_0 * \exp(\beta_1 * 100/10) * 10 &= \\
 c_0 * \exp(\beta_1 * 200/20) * 20 &
 \end{aligned}$$

Millises skaalas lisada?



Enne kaheks jagamist:

$$\begin{aligned}
 E \text{ rääke} &= \exp(\beta_0 + \beta_1 * \log(\text{kraav}/\text{pindala}) \\
 &\quad + \log(\text{pindala})) \\
 &= c_0 * (\text{kraav}/\text{pindala})^{\beta_1} * \text{pindala} \\
 &= c_0 * (200/20)^{\beta_1} * 20
 \end{aligned}$$

Pärast kaheks jagamist:

$$\begin{aligned}
 E \text{ rääke} &= c_0 * (\text{kraav}/\text{pindala})^{\beta_1} * \text{pindala} \\
 &= c_0 * (100/10)^{\beta_1} * 10
 \end{aligned}$$

$$\begin{aligned}
 2 * c_0 * (100/10)^{\beta_1} * 10 &= \\
 c_0 * (200/20)^{\beta_1} * 20 &
 \end{aligned}$$

Lisame kraavid...

```
> mudell=glm(raak~factor(toetustyypp)+factor(hairing)+I(kraavid2/pindala)+
  offset(log(pindala)), family=poisson()); summary(mudell)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.9283311	0.2603598	-18.929	< 2e-16 ***
factor(toetustyypp) ksm	-0.9357706	0.1606201	-5.826	5.68e-09 ***
factor(toetustyypp) mahe	0.0364291	0.1884282	0.193	0.8467
factor(toetustyypp) ypt	0.2401695	0.1373258	1.749	0.0803 .
factor(hairing) hairingutpole	1.0417357	0.2292775	4.544	5.53e-06 ***
I(kraavid2/pindala)	-0.0001896	0.0004568	-0.415	0.6781

AIC: **2236.7**

```
> drop1(mudell, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1499.0	2236.7		
factor(toetustyypp)	3	1606.7	2338.3	107.665	< 2.2e-16 ***
factor(hairing)	1	1527.6	2263.3	28.601	8.894e-08 ***
I(kraavid2/pindala)	1	1499.2	2234.8	0.180	0.6717

Lisame kraavid...

```
> mudell=glm(raak~factor(toetustyypp)+factor(hairing)+
  I(log(kraavid2/pindala))+offset(log(pindala)), family=poisson())
> summary(mudell)
```

Coefficients:

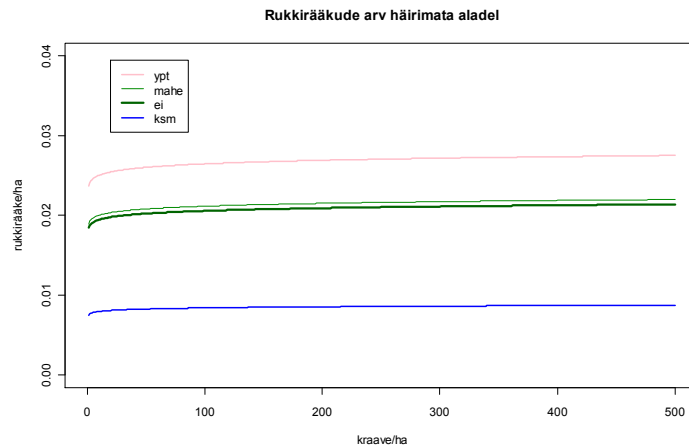
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.04603	0.27111	-18.613	< 2e-16 ***
factor(toetustyypp) ksm	-0.89704	0.15963	-5.620	1.91e-08 ***
factor(toetustyypp) mahe	0.02807	0.18859	0.149	0.8817
factor(toetustyypp) ypt	0.25243	0.13583	1.858	0.0631 .
factor(hairing) hairingutpole	1.05070	0.22940	4.580	4.65e-06 ***
I(log(kraavid2/pindala))	0.02412	0.02396	1.007	0.3139

AIC: **2235.8**

```
> drop1(mudell, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1498.1	2235.8		
factor(toetustyypp)	3	1597.2	2328.9	99.038	< 2.2e-16 ***
factor(hairing)	1	1527.3	2262.9	29.117	6.812e-08 ***
I(log(kraavid2/pindala))	1	1499.2	2234.8	1.032	0.3098

Kraavid ei mõjuta rukkirääkude arvukust?



Või siiski?

```
> m0=glm(raak~factor(toetustyypp)+factor(hairing)+
  I(log(kraavid2/pindala))*factor(toetustyypp)+offset(log(pindala)),
  family=poisson()); summary(m0)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.47287	0.28552	-15.666	< 2e-16
factor(toetustyypp) ksm	-1.66525	0.24428	-6.817	9.29e-12
factor(toetustyypp) mahe	-0.74736	0.50966	-1.466	0.142542
factor(toetustyypp) ypt	-0.49604	0.23186	-2.139	0.032400
factor(hairing) hairingutpole	1.06369	0.22942	4.636	3.55e-06
I(log(kraavid2/pindala))	-0.16301	0.04913	-3.318	0.000906
factor(toetustyypp) ksm:I(log(kraavid2/pindala))	0.25323	0.06739	3.758	0.000172
factor(toetustyypp) mahe:I(log(kraavid2/pindala))	0.23311	0.11973	1.947	0.051540
factor(toetustyypp) ypt:I(log(kraavid2/pindala))	0.23316	0.06216	3.751	0.000176

```
> drop1(m0, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(Chi)
<none>		1480.5	2224.1		
factor(hairing)	1	1510.4	2252.1	29.955	4.422e-08 ***
factor(toetustyypp):I(log(kraavid2/pindala))	3	1498.1	2235.8	17.685	0.0005108 ***

Seos kraavide arvu ja rukkirääkude arvu vahel...

