Matemaatiline statistika ja modelleerimine

Käsitlemata teemasid …

EMÜ doktorikool
DK.0007

Tanel Kaart

Valimi suurusest ja testi võimsusest

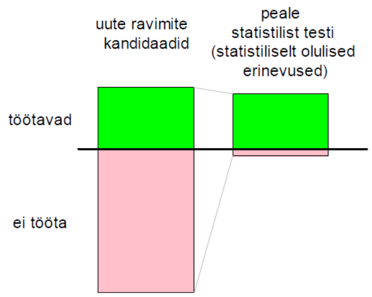
(kasutades Märt Mölsi materjale ja kirjutist aadressilt

http://www.indiana.edu/~statmath/stat/all/power/power.pdf)

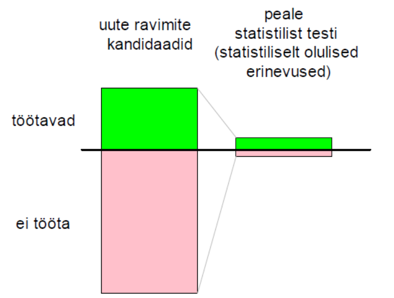
Valimi suurusest ja testi võimsusest

- Statistilise testi tõenäosust antud katseplaani korral tõestada alternatiivne hüpotees nimetatakse testi **võimsuseks**.

Statistiline test kui filter – suur valim.
Testi võimsus ehk *power* on suur.



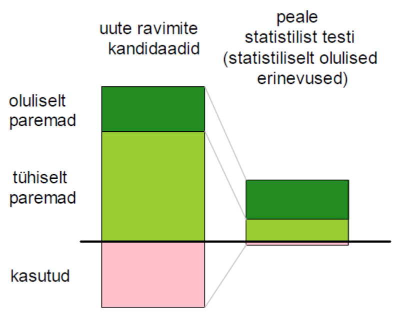
Statistiline test kui filter – väike valim.
Testi võimsus ehk *power* on väike.



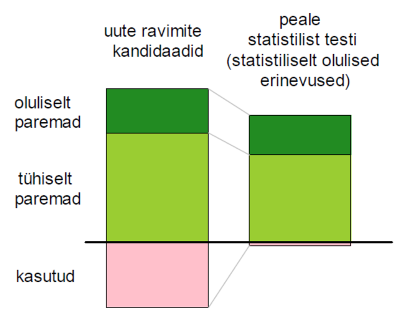
Joonised Märt Mölsi loengukonspektist

Valimi suurusest ja testi võimsusest

Statistiline test kui filter – "parajalt" suur valim.



Statistiline test kui filter – "liiga" suur valim.

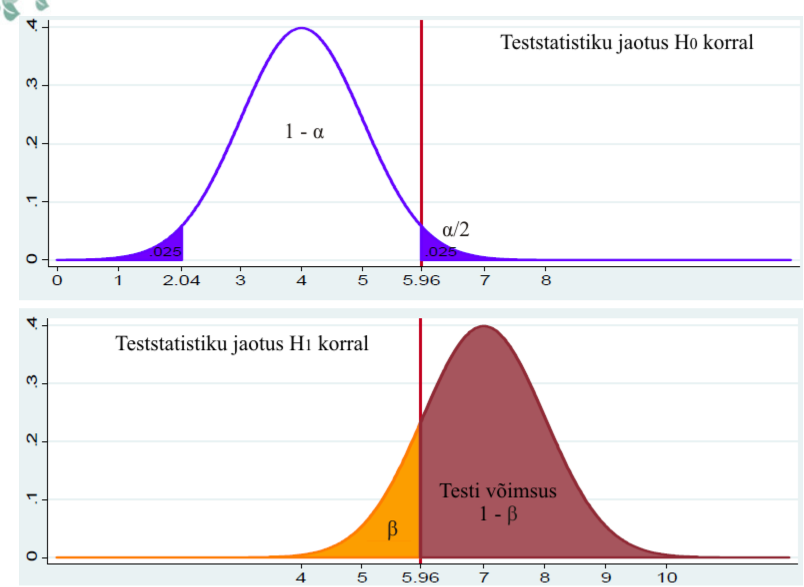


Joonised Märt Mölsi loengukonspektist

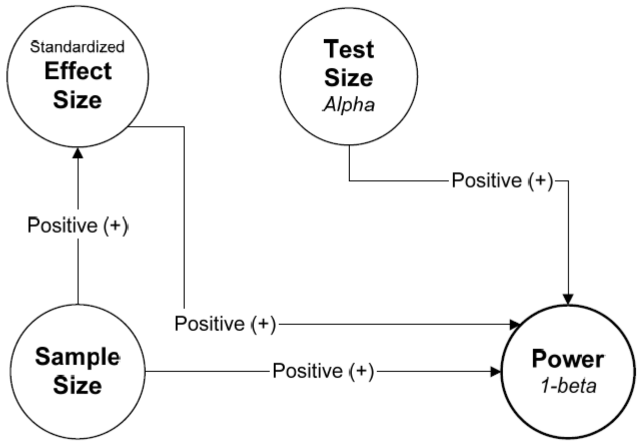
Valimi suurusest ja testi võimsusest

- Testi võimsuse arvutamiseks on enamasti vaja teada:
  - ¤) kavandatava valimi suurust;
  - ¤) analüüsimiseks kasutatavat testi;
  - ¤) uuritava tunnuse hajuvust;
  - ¤) milline alternatiivne hüpotees kehtib.
- Võimsusanalüüsi tööpõhimõte:
  - ¤) leitakse teststatistiku jaotus nullhüpoteesi H_0 korral;
  - ¤) sellele jaotusele vastavalt leitakse teststatistiku kriitiline väärtus, millest suurema teststatistiku väärtuse tulekul me H_0 enam ei usu;
  - ¤) järgnevalt leitakse (mingile konkreetsele) alternatiivsele hüpoteesile vastav teststatistiku jaotus;
  - ¤) viimast nullhüpoteesile vastava teststatistiku jaotusega kõrvutades saame aga leida, kui tõenäoline on, et saame valimi, mille puhul saame nullhüpoteesi kummutada – see ongi testi võimsus antud valimi suuruse korral.

Valimi suurusest ja testi võimsusest

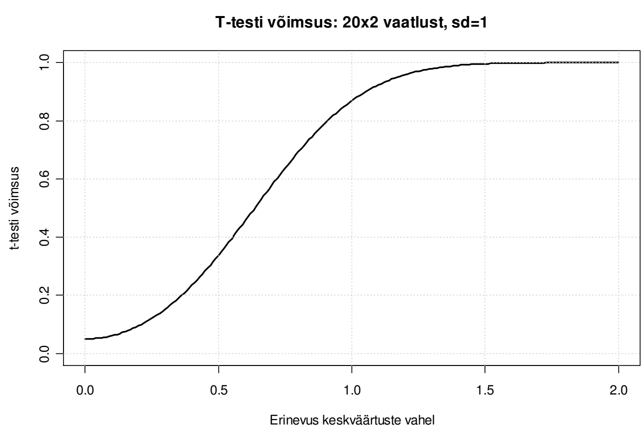


Valimi suurusest ja testi võimsusest



Valimi suurusest ja testi võimsusest

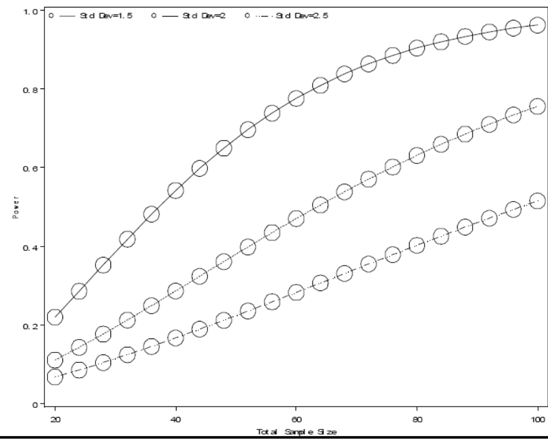
```
erinevus=seq(0,2,0.01)
voimsus=power.t.test(n=20, erinevus, sd=1, sig.level=0.05, strict=TRUE)
plot(erinevus, voimsus$power, type="l", xlab="Erinevus keskväärtuste vahel",
        ylab="t-testi võimsus", main="T-testi võimsus: 20x2 vaatlust, sd=1",
        ylim=c(0,1), lwd=2)
grid()
```



Valimi suurusest ja testi võimsusest

```
PROC POWER;
      TWOSAMPLEMEANS
      ALPHA=.01 SIDES=2
      GROUPMEANS= (2.79 4.12)
      STDDEV= 1.5 2.0 2.5
      NTOTAL=44
      POWER=.;
      PLOT X=N MIN=20 MAX=100 KEY=BYCURVE(NUMBERS=OFF POS=INSET);
RUN;
```

```
            The POWER Procedure
   Two-sample t Test for Mean Difference

         Fixed Scenario Elements

     Distribution                Normal
     Method                       Exact
     Number of Sides                  2
     Alpha                         0.01
     Group 1 Mean                  2.79
     Group 2 Mean                  4.12
     Total Sample Size               44
     Null Difference                  0
     Group 1 Weight                   1
     Group 2 Weight                   1

              Computed Power

                   Std
          Index    Dev     Power

            1      1.5     0.598
            2      2.0     0.324
            3      2.5     0.189
```



Valimi suurusest ja testi võimsusest

- Küsimus teistpidi: kui suurt valimit oleks vaja, et oleks piisavalt suur võimalus tõestada alternatiivse hüpoteesi kehtivust?
- Vt näiteks t-testi (kahe rühma keskväärtuste võrdlus). Siis on valimi suuruse arvutamiseks vaja teada järgmisi näitajaid:
  - ¤) $\delta = \mu_1 - \mu_2$, kahe grupi keskväärtuste vähim erinevus, mille tõestamist peetakse oluliseks;
  - ¤) σ, uuritava tunnuse eeldatav standardhälve;
  - ¤) α, olulisuse nivoo (lubatud tõenäosus eksida, kui tegelikult gruppide keskmised pole erinevad);
  - ¤) $1 - \beta$, soovitav võimsus.

$$n = \frac{2\sigma^2}{\delta^2} f(\alpha, \beta),$$

(mõlemast grupist peab olema mõõdetud n isendit)

$f(\alpha, \beta)$		$1-\beta$			
		0,95	0,9	0,8	0,5
α	0,1	10,8	8,6	6,2	2,7
	0,05	13,0	10,5	7,9	3,8
	0,02	15,8	13,0	10,0	5,4
	0,01	17,8	14,9	11,7	6,6

Valimi suurusest ja testi võimsusest

Kas porganditega nuumatud küülikutel kasvavad pikemad kõrvad kui kaalikatega toidetutel? Kui suurt valimit vajame?

- Valime $\delta = 0{,}5$ cm (see erinevus võiks olla piisav praktilise huvi äratamiseks);
- kirjandust uurides leiame, et $\sigma = 2{,}5$ cm, millest $\sigma^2 = 6{,}25$;
- võtame olulisuse nivooks standardse $\alpha = 0{,}05$ ja testi võimsuseks $1 - \beta = 0{,}8$ (kui tegelikud kõrva pikkuste keskväärtused erinevad 0,5 cm või enam, siis tahame erinevust tõestada vähemalt tõenäosusega 0,8).

Asendades toodud arvud valemisse, saame:

$$n = \frac{2\sigma^2}{\delta^2} f(\alpha, \beta) = \frac{2 \times 6{,}25}{0{,}25} \times 7{,}9 = 395.$$

Sama asi *R*-s:

```
> power.t.test(delta=0.5, sd=2.5, sig.level=0.05, power=0.8, type="two.sample", alternative="two.sided")

     Two-sample t test power calculation

              n = 393.4067
          delta = 0.5
             sd = 2.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

 NOTE: n is number in *each* group
```

Valimi suurusest ja testi võimsusest

Võimalused võimsuse (valimi mahu) arvutamiseks:

• *R*:

`power.t.test` – t-testi võimsus (sõltuvad ja sõltumatud valimid)
`power.anova.test` – ühefaktoriline ANOVA
`power.prop.test` – tõenäosuste (protsentide) võrdlemine

• *SAS*:
Solutions -> Analysis -> Analyst -> Statistics -> Sample size
protseduurid POWER ja GLMPOWER

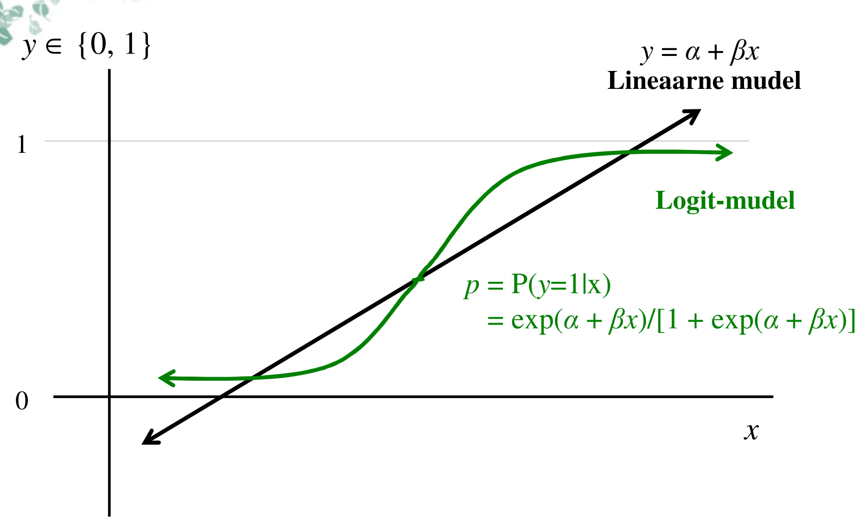
• *STATISTICA*:
Statistics -> Power Analysis

• *G*Power*:
http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/

• http://www.math.uiowa.edu/~rlenth/Power/: Java applets for power and sample size

Logistiline regressioon

Logistiline regressioon



Logistiline regressioon

• Logistilise regressiooni abil leitud tõenäosuste hinnangud jäävad alati 0 ja 1 vahele.

• Logistilise regressiooni mudeli (*logit*-mudeli) kujusid:

$$p = P(y=1|x) = \exp(\alpha + \beta x)/[1 + \exp(\alpha + \beta x)],$$

$$\ln[p/(1-p)] = \alpha + \beta x,$$

$$\text{logit}(p) = \alpha + \beta x,$$

• p on huvi pakkuva sündmuse Y aset leidmise tõenäosus, $P(Y=1)$,

• $p/(1-p)$ on šansside suhe [*odds ratio*],

• $\ln[p/(1-p)]$ on logaritmiline šansside suhe [*log odds ratio*].

Logistiline regressioon

- Tõenäosuse hinnang avaldub kujul:

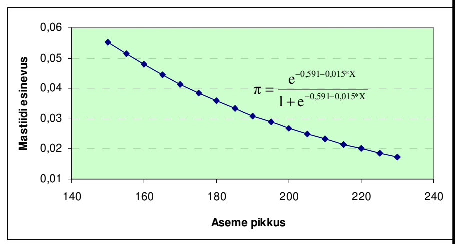
$$p = \exp(\alpha + \beta x)/[1 + \exp(\alpha + \beta x)];$$

  - kui $\alpha + \beta x = 0$, siis $p = 0{,}5$
  - $\alpha + \beta x$ suurenemisel $p \to 1$,
  - $\alpha + \beta x$ vähenemisel $p \to 0$.

- Logistilise regressioonivõrrandi kordaja β eksponent, e^{β}, näitab, kui mitu korda muutub sündmuse aset leidmise šanss (kui palju muutub šansside suhe) argumendi muutumisel ühe ühiku võrra.
  - Näiteks kui $e^{\beta} = 2$, siis kaasneb argumenttunnuse väärtuse suurenemisega 1 võrra sündmuse toimumise šansi kahekordne suurenemine.
  - Negatiivse regressioonikordaja β korral on šansside suhe ühest väiksem, $e^{\beta} < 1$, seega kaasneb argumenttunnuse suurenemisega uuritava sündmuse aset leidmise šansi vähenemine.

Logistiline regressioon

- Mastiidi esinemise tõenäosus (π) on aseme pikkusest (X) prognoositav logistilise regressiooni abil:



- SAS-i protseduur LOGIT:

```
proc logistic data=analyys.andmed1
 descending;
 model mastiit = aseme_pi;
run;
```

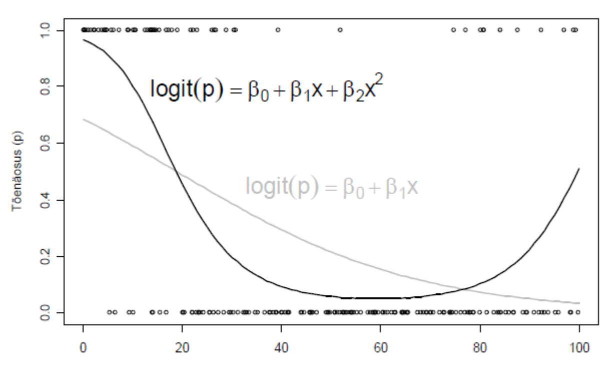
- Programmi väljund

```
The SAS System
The LOGISTIC Procedure
......
Analysis of Maximum Likelihood Estimates
```

Variable	DF	**Parameter Estimate**	Standard Error	Wald Chi-Square	**Pr > Chi-Square**	Standardized Estimate	Odds Ratio
INTERCPT	1	**-0.5809**	0.7074	0.6743	0.4116	.	.
ASEME_PI	1	**-0.0148**	0.00409	13.0188	**0.0003**	-0.156026	**0.985**

Logistiline regressioon

Mis teha, kui pakutud seos ei sobi?



Joonis Märt Mölsi loengukonspektist

Üldistatud lineaarsed mudelid [*generalized linear models*]

• To guarantee the smallest variability of estimates got with least square methods, to test statistical hypothesis and to apply the maximum likelihood methods, the normality of dependent variables is assumed: $\mathbf{y} \sim N(\mathbf{X\beta}, \mathbf{V})$.

• If this assumption is not filled, in most cases (when the probability distribution of dependent variable belongs to the class of exponential distributions) it is possible to fit linear model to transformed dependent variable of the form

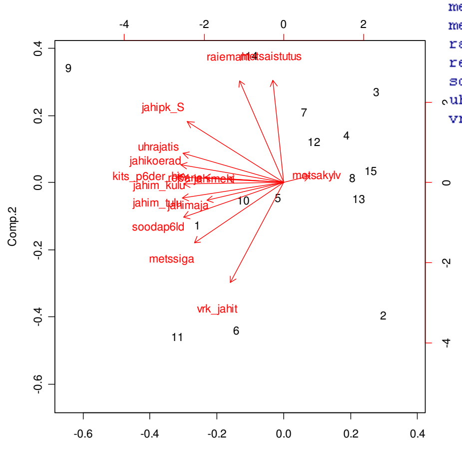
$$g(\mathbf{y}) = \mathbf{X\beta} + \boldsymbol{\varepsilon},$$

where function g is called **link function**.

Peakomponent- ja faktoranalüüs

Peakomponentanalüüs
[*Principal Component Analysis*]

Peakomponentanalüüs

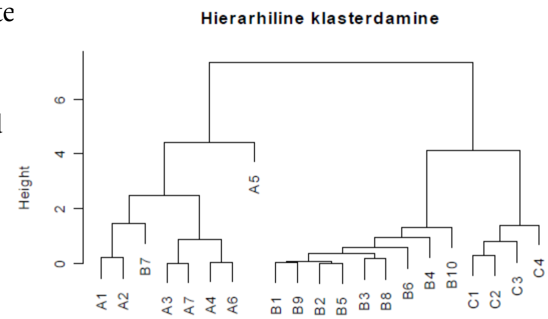


	Comp.1	Comp.2	Comp.3
jahikoerad	-0.32358331	0.086977502	-0.20303098
jahim_kulu	-0.31457085	-0.006760718	0.43823340
jahim_tulu	-0.31954996	-0.077331338	0.41816479
jahimaja	-0.24244314	-0.087138915	0.03381092
jahimehi	-0.17571001	0.016379050	-0.66616633
jahipk_S	-0.30397907	0.300264058	-0.16108449
kits_p6der_hirv	-0.33579007	0.029353228	-0.08126374
metsaistutus	-0.03332323	0.504356000	0.23265124
metsakylv	0.08162153	0.032290689	0.02377239
metssiga	-0.28321337	-0.296317395	0.01682214
raiemaht	-0.13991552	0.501686171	0.12551896
rebane	-0.24964601	0.026143749	-0.17552640
soodap6ld	-0.31674142	-0.169336023	-0.01437659
uhrajatis	-0.31762503	0.142958615	-0.05373451
vrk_jahit	-0.16829529	-0.492769180	0.09880296

Klasteranalüüs

Klasteranalüüs

- Loodusuurija sõitis Asustamata Saarele ja avastas seal suure hulga seni teadusele tundmatuid putukaid.
- Ta mõõtis oma lühikese saarelviibimise jooksul tervel hunnikul seninägematutel putukatel igasuguseid näitajaid (igatsorti pikkuseid ja mustrielementide arvu ja paljut muudki).
- Järgmiseks soovis loodusuurija määratleda, mitmesse alamliiki leitud putukad võiksid kuuluda.
- Et saada esimest ligikaudset lähendit, kust oma uurimistööga pihta hakata, soovis ta leida sarnaste putukate rühmad – kes oleksid siis alamliikide kanditaatideks.
- Selleks söötis ta oma andmed klasteranalüüsi teostavasse programmi, mis joonistas järgmise pildi:



Märt Mölsi loengukonspektist

Klasteranalüüs

Hierarhiline klasterdamine

... on hästi kasutatav siis, kui meil on suhteliselt vähe objekte või kui on oodata, et klastrid suhteliselt selgelt üksteisest eristuvad.

Hierarhiline klasteranalüüs põhineb väga lihtsal algoritmil:

samm-sammult hakatakse omavahel kokku panema kõige sarnasemaid objekte. Näiteks, kui leidub kaks täpselt ühesuguste tulemustega objekti, siis liidetakse nad esimesel sammul üheks klastriks, peale seda võrreldakse kõiki üksikobjekte ja juba tekkinud klastreid ja liidetakse jälle kõige sarnasemad omavahel jne.

Vaatluste omavahelise kauguse määramine:

Eukleidese kaugus: $d\left((x_1, y_1),(x_2, y_2)\right) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

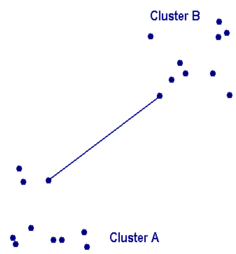
Manhattani kaugus: $d\left((x_1, y_1),(x_2, y_2)\right) = |x_1 - x_2| + |y_1 - y_2|$

…

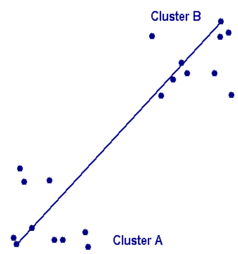
Klasteranalüüs

Klastritevahelise kauguse määramine:

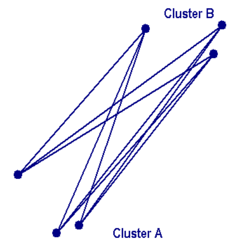
“single”



“complete”



“average”



Klasteranalüüs

k-keskmiste klasterdamine

... sobib grupeerimise meetodiks siis, kui objekte on nii palju, et hierarhilise klasteranalüüsi tulemus muutub ebaülevaatlikuks, kuid ka siis kui me oskame meile sobivat klastrite arvu ligilähedaselt ennustada ning ühtlasi soovime saada ka tekkivate klastrite kirjelduse nende tunnuste osas, mis on grupeerimise aluseks.

Algoritm:

- kõigepealt tuleb määrata klastrite arv, siis
- jagada objektid esialgsetesse klastritesse,
- arvutada välja klastrite keskpunktid ning
- hakata võrdlema igat objekti kõigi klastrite keskpunktidega; kui osutub, et objekti kaugus mõne muu klastri keskpunktist on väiksem kui selle klastri keskpunktist,milles ta parasjagu asub, siis tuleb objekt teise klastrisse ümber tõsta;
- peale objekti ümbertõstmist tuleb pöörduda uuesti sammu 3 juurde ja jätkata protsessi niikaua kui kõik objektid on klastris, mille keskpunktile nad kõige lähemal asuvad.

Analüüsitulemuste esitamisest

(matemaatikute vaatepunktist, mis ei pruugi alati ühtida mittematemaatilistes ajakirjades aegade jooksul välja kujunenud traditsioonidega)

Fondid

- Üldised matemaatilise teksti esitamisel kehtivad reeglid
  - tekst, numbrid, aritmeetilised operaatorid, funktsioonid ja sümbolid esitatakse tavalises püstises kirjas,

  näiteks: $\log_{10}(100) + \sin(90°) = 3$;
  - muutujad (sh statistikud, indeksid, konstandid) esitatakse *kaldkirjas*,

  näiteks: $p < 0.001$; $r = 0.575$ ($p = 0.016$);

  $$\bar{x} \pm s = 45.3 \pm 12.3;$$

  $$y_i = \mu + bx_i + e_i;$$

  $$\mathrm{var}(\bar{P}_s) = \mathrm{var}\left[\bar{P} + \tfrac{1}{2}A_s + \tfrac{1}{n}\sum_{i=1}^{n}(\tfrac{1}{2}A_{d_i} + E_i)\right];$$
  - maatriksid ja vektorid esitatakse **rasvases kirjas**,

  näiteks:

  $$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},\ \text{where } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},\ \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix} \mu \\ b \end{pmatrix},\ \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

Mõningad lisasoovitused

- Kasuta tildet (~) tähistamaks ligikaudset võrdust.
- Arvud (≤ 99) lause alguses kirjutatakse sõnadega. Üheksakümneüheksast suuremate arvude puhul on soovitatav kirjutada lause ümber, vältimaks selle arvuga algamist.
- Teksti see esitatakse arvud kuni üheksani enamasti sõnadega: üks, kaks, kolm, …, üheksa, 10, 11, 12…
  Erand: 2-meetrine mõõdulint; 3 miljonit.
- Arvu ja ühiku vahele käib tühik: näiteks, 75 kg.
  Erand: 75%.
- Märkus: 0.32, mitte .32.
- Tuhandeliste eraldajana kasutatakse inglise keelses tekstis tavaliselt koma (,) enam kui neljakohaliste arvude korral:
  143; 2,461 või 2461; 21,278; 1,409,000…

Kümnendkohtade arv

- Sobiv kümnendkohtade arv:
  - kirjeldavatel statistikutel (keskmine) tavaliselt üks kümnendkoht enam, kui mõõtmistäpsus,
  - standardhälbel (standardveal) tavaliselt üks kümnendkoht enam kui keskmisel,
  - korrelatsioonikordajal kaks kümnendkohta,
  - protsentidel kaks tüvenumbrit;
  - näiteks: 73 ± 5.2; $r = 0.45$; $r = 0.08$; 16%; 1.3%; 0.013%.

		Number of observations	Mean	Standard deviation	T-test (p-value)
Sex	Filly	132	334,4	11,4	0,478
	Colt	131	335,2	9,1	
Foaling time	Spring	182	338,1	9,8	<0,0001
	Autumn	81	327,2	6,9	
Breed	Estonian	23	326,1	11,5	<0,0001
	Tori	240	335,6	9,8	

Kümnendkohtade arv

- Kui olulisuse tõenäosuse (*p*-väärtuse) esitamine on sobilikum usalduspiiride esitamisest, siis tuleks ära tuua täpne *p*-väärtus ühe või kahe informatiivse kümnendkohaga ($p < 0.1$ korral) või kahe kümnendkohaga ($p > 0.10$ korral).
- Kui $p < 0.001$ (mõnikord ka $p < 0.0001$), siis täpset väärtust ei esitata.
- Näiteks:
  - $p = 0.03$ või $p = 0.032$, aga mitte $p = 0.0324786$ või $p < 0.05$;
  - $p = 0.007$ või $p = 0.0072$, aga mitte $p = 0.0072213$;
  - $p < 0.001$ või $p < 0.0001$, aga mitte $p = 0.000043$;
  - $p = 0.09$, $p = 0.74$, aga mitte $p > 0.05$.

Materjal ja metoodika

Materjali ja metoodika kirjutamisel võiks eneselt küsida:

- **Kas suudab keegi teine üksnes selles osas kirjas oleva alusel**
  - **teostada sama eksperimendi sama tulemusega või**
  - **koguda (kasvõi teoreetiliselt) sama struktuuriga andmestiku ja kasutada samu matemaatilisi meetodeid või mudeleid?**

Kirjas peavad olema

- uurimisobjektid (lehm või laut või farm või …; piirkond või põld/katselapp või taim või …),
- protseduur (eksperimendi disain, andmete kogumise kirjeldus),
- statistika (andmetöötlus).

Materjal ja metoodika

- Study area
- Data collection

2. Materials and Methods

2.1. Study Area and Sites

Xiangxi River, a sixth-order river, originating from the Shennongjia Forest Region, is an important tributary of the Yangtze River, China. It has a length of 94 km, with a catchment area of 3,099 km², and a natural fall of 1540 m from the headwaters to its confluence with the Yangtze River at Xiangxi River Mouth. Therefore, many small hydropower stations were built within the watershed.

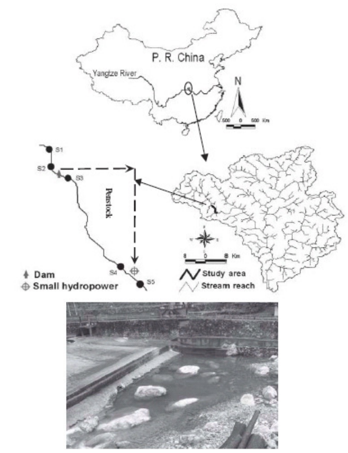


Figure 1. The location of Xiangxi River in China (top), the small hydropower plant and the sampling sites in Xiangxi River (middle) and the picture of S3 (bottom).

Materjal ja metoodika

- Study area
- Data collection

population dynamics.

In the present paper we analyse the role of temperature changes in the population dynamics of lake smelt *Osmerus eperlanus eperlanus* m. *spirinchus* Pallas, a key species in the fish community of Lake Peipsi. Lake smelt is a freshwater dwarf form of European smelt *Osmerus eperlanus* (L.) (Kuderskij & Fedorova, 1977; Pihu, 2003) and is a small-sized, slender, shoaling fish with a characteristic cucumber smell (Quigley et al., 2004). In Lake Peipsi, lake smelt has historically been among the most important commercial fish species in terms of catch size (Pihu & Kangur, 2001). Due to their intermediate trophic position as a consumer of plankton and as a prey for higher predators, lake smelt has the potential to introduce

Materials and methods

Study area

Lake Peipsi is situated on the border of Estonia and Russia (Fig. 1). Its surface area is 3,555 km^2, mean depth 7.1 m and maximum depth 15.3 m (Jaani, 2001a) and its altitude is 30 m above sea level. The volume of the whole lake is 25 km^3 and the residence time of water is about 2 years. The only outflow is through the Narva River into the Gulf of Finland. Natural water level fluctuations have shown an overall range of 3.04 m over the last 80 years, with an average annual range of 1.15 m (Jaani, 1996). L. Peipsi is situated in the northern region of the temperate zone with variable

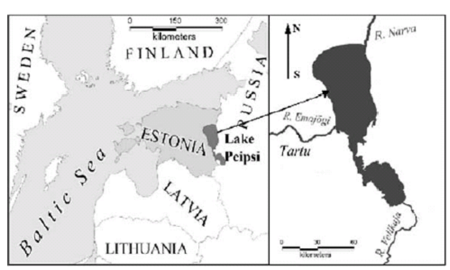


Fig. 1 Location of Lake Peipsi

Materjal ja metoodika

- **Experiment based study** – more attention on the experimental design and less on statistical methods
  - Experimental design
  - Sample size calculations
  - Experiment description
    - Don't assume that the reader knows what you did, even when "the reader" is your instructor.
    - Instead of writing:
      First pour agar into six petri plates. Then inoculate the plates with the bacteria. Then put the plates into the incubator . . .
    - Simply describe how the experiment was done:
      Six petri plates were prepared with agar and inoculated with the bacteria. The plates were incubated for ten hours.
  - Deviations from initial experimental design.

Materjal ja metoodika

- **Experiment based study**
  - Statistical analysis – as a rule a standard statistical methods (t-test, ANOVA, correlation and regression analysis, chi-square test, …)
  - Probability level
  - Error measure (standard deviaton, standard error)
  - Example:
    An unpaired, two-tailed Student's t-test was used to compare the frequency of ……… between males and females. $P < 0.05$ was considered significant and variability was measured with standard error of the mean (*S.E.M.*).

Materjal ja metoodika

- **Model based study** – more attention on the statistical methods
  - Standard statistical methods don’t need additional explanations.
  - More specific methods and complex models need extra explanations and references.

Principal component analysis (PCA) was applied to the landscape elements and the bumblebee species found (Table 4). The results of PCA are interpreted using scatters of the site scores connected with group centroids (star plot), and via correlation circles. The effect of the grouping factor was estimated using the multivariate randomisation test (Manly, 1995) with 1000 permutations. The influence of landscape structure on the bumblebee community was studied with co-inertia analysis (Dolédec and Chessel, 1994). This is a two-table ordination method that facilitates the establishment of connections between tables with data domains that contain the same or even different number of variables. This method combines various standard single-table ordination methods such as principal component analysis and correspondence analysis. Dolédec and Chessel (1994) give a detailed description of the co-inertia analysis in their studies.

Materjal ja metoodika

Data analysis

The analysis of the data comprised three major steps. Firstly, we calculated the community-wide characteristics of clonal growth (Table 2). For each plot, we calculated the weighted average of each clonal growth parameter (M_{pj}) as:

$$M_{pj} = \Sigma\, a_i\, p_{ij} \qquad (2)$$

Here a_i is the cover (ranging from 0 to 1) for the species i in the plot and p_{ij} is the value of the j-th clonal growth parameter for the species i.

We measured the difference in extent of clonal mobility among the coexisting species with the coefficient of variation of the medians of rhizome increment for all species in one plot. The coefficient of variation was chosen as it is an estimate of variance independent of the value of the sample mean. To correct also for a possible bias associated with the differences in sample size, we used the following correction (Sokal and Rohlf, 1995):

$$\mathrm{CV}_{\mathrm{increment}} = (1 + 1/4n)(\mathrm{St.Dev}_{\mathrm{increment}} \times 100/M_{\mathrm{increment}}) \qquad (3)$$

Here $\mathrm{CV}_{\mathrm{increment}}$ stands for the corrected coefficient of variation for median rhizome increment per plot, n is sample size, $\mathrm{St.Dev}_{\mathrm{increment}}$ is the standard deviation of rhizome increment for one plot, and $M_{\mathrm{increment}}$ is the mean of the rhizome increments for all species found in a plot.

Secondly, we estimated the effect of environmental conditions on the averages of clonal growth parameters, species density, and ramet density as well as the general differences between the communities in average clonal growth. To test for the general relationships between different environmental variables and vegetation characters (incl. community-wide parameters of clonal growth), we built a squared correlation matrix with the Pearson’s r and probabilities of error (p-level). To correct for the mass effect we employed the Bonferroni type correction with the Dunn-Šidák method (Sokal and Rohlf, 1995) and obtained the critical p-level (experimentwise error rate) using the following equation:

$$p_{\mathrm{critical}} = 1 - (1 - 0.05)^{1/k} \qquad (4)$$

Materjal ja metoodika

Results were evaluated statistically using a general linear mixed model assuming a first-order autoregressive variance structure of repeated measurements from the individual cows (SAS Inst. Inc. 2006). In order to estimate the effects of different factors on the milk coagulation, compositional parameters, the following models were used:

$y_{ijklmof} = \mu + parity_i + lactmonth_j + \alpha\beta\kappa\_Cn_k + \beta\_Lg_l + a_o + pe_f + \varepsilon_{ijklmof}$

where: $y_{ijklmnop}$ = milk coagulation (log RCT, E_{30}), production (daily milk yield) or compositional trait (milk fat and protein contents), μ = general mean, $parity_i$ = fixed effect of parity class i (i = 1 to 4), $lactmonth_j$ = fixed effect of month of lactation j (j = 1 to 11), $\alpha\beta\kappa\_Cn_k$ = fixed effect of aggregate α_{s1}-, β- and κ-Cn genotypes k (k = 1 to 15), β_Lg_l = fixed effect of β-Lg genotype l, (l = 1 to 3); a_o = random additive genetic effect of animal o, N(0, $A\sigma^2_a$); pe_f = random permanent environmental effect of farm f, N(0, $I\sigma^2_{pe}$); ε = random residual effect with spatial power covariance structure, N(0, R).

mowing frequency once a year. The following model was tested:

$$= \mu + a_1 HABITAT + a_2 L + a_3 L^2 + a_4 B + a_5 B^2 + a_6 LB \qquad (5)$$

tes a dependent variable (species density, ramet density and erages of clonal growth parameters), μ is the intercept, L stands ration coefficient, B is biomass, and a_1–a_6 are the coefficients. *TE* was included in the model as a random factor. The factors L sted in *HABITAT*, while the factors *SITE* and *HABITAT* were munity type. Type 3 test of fixed effects was used with the iter- ed Maximum Likelihood (REML) procedure to estimate the ariance components. To test for the dependence of the variables ght availability and different mowing regimes within one com- s well as for the differences in the average values of the variables hree community types, we used the statement ESTIMATE. In alculated the differences in the least squares means, however, as imilar results they are not presented here.

Statistika tarkvara

- The well-known commercially-available statistics software like SAS, SPSS, R, Minitab, Genstat, Statistica need no further reference, although it is helpful to provide the version number, as these packages are constantly being updated and features change.
- The more specific software needs reference.

241–280, 281–320 and 321–365 days within the year). Regression analyses were carried out with SAS System, Release 8.2, SAS Institute Inc., 1996, using GLM procedure. For studying effect of water temperature and water level on the smelt

Computations and graphical displays of the co-inertia analysis were obtained using the ADE-4 package (Thiolouse et al., 1997).

Kirjeldav statistika

Tabelid

- Tables can replace lengthy descriptions (e.g. to summarize).
  - Tables should not repeat information from the text or from any pictorial display. Ideally, each should provide a complete set of information that is comprehensible if the text is removed.
  - Do not use TAB's to make a table.
  - Do not use vertical lines, but "play" with horizontal lines (and spacing and empty columns).

Table 2. Mean pod and hay yields (t ha^{-1}) of five groundnut cultivars grown at Wad Medani, Sudan in 1984–86.

	Pod yield				Hay yield			
Cultivar	1984	1985	1986	Mean	1984	1985	1986	Mean
Early Bunch	2.47	3.34	2.14	2.65	3.56	2.40	2.61	2.86
Georgia 119–20	1.73	3.38	2.29	2.46	3.50	3.88	2.81	3.39
UF79–1499	1.61	2.05	2.34	2.00	5.23	1.82	3.26	3.44
Apollo	1.47	3.38	1.93	2.26	4.65	3.44	2.81	3.63
MH 383	1.75	3.70	2.14	2.53	4.78	4.58	2.99	4.12
Mean	1.81	3.17	2.17	2.38	4.34	3.22	2.88	3.49
s.e.	0.235	0.223	0.167	0.210	0.571	0.292	0.264	0.373

Tabelid

Line	Number of stallions	Inside lines				Between lines	
		Coancesty coefficient		Inbreeding coefficient		Coancesty coefficient	
		Average	Max	Average	Max	Average	Max
Ahti	7	0,092	0,276	0,012	0,029	0,037	0,161
Eni	4	0,110	0,255	0,004	0,009	0,041	0,163
Raspel	9	0,090	0,287	0,014	0,047	0,040	0,163
Taru	1	0,000	0,000	0,002	0,002	0,026	0,088
Taube	2	0,074	0,074	0,012	0,020	0,028	0,088

Line	Number of stallions	Inside lines				Between lines	
		Coancesty coefficient		Inbreeding coefficient		Coancesty coefficient	
		Average	Max	Average	Max	Average	Max
Ahti	7	0,092	0,276	0,012	0,029	0,037	0,161
Eni	4	0,110	0,255	0,004	0,009	0,041	0,163
Raspel	9	0,090	0,287	0,014	0,047	0,040	0,163
Taru	1	0,000	0,000	0,002	0,002	0,026	0,088
Taube	2	0,074	0,074	0,012	0,020	0,028	0,088

Joonised

- Diagrams should not repeat information in the text or in a table and should be fully comprehensible even if removed from the paper.

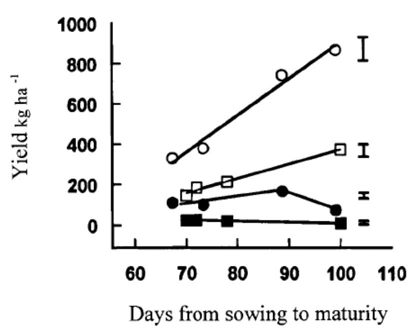
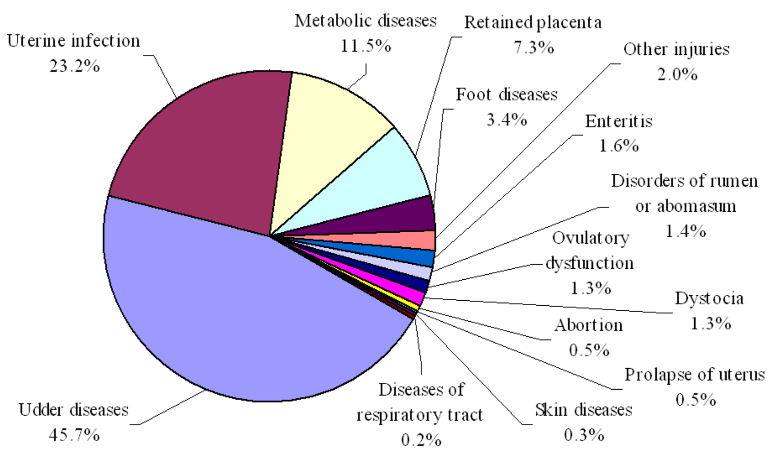


Fig. 2. Relationship between biomass (○●) and seed (□■) yield, and days from sowing to maturity, among four varieties of cowpea grown at Minjibir (open symbols) and Malam Madori (closed symbols) in Nigeria in 1992. Bars are s.e.d. between means. Fitted lines: ○$Y = -919(156) + 18.2(1.88)X$, $r^2 = 0.98$; □$Y = -349(42) + 7.2(0.52)X$, $r^2 = 0.99$; ■$Y = 61(3.5) - 0.5(0.04)X$, $r^2 = 0.99$; ● line fitted by eye.

Joonised

- Use the pie chart only if you really want to describe, how the whole amount is divided into the parts.

Figure 1. The panorama of multifactorial diseases in Estonian dairy cows



Korrelatsioonikordajad

Table 6. Pearson's partial correlation coefficients (r) between studied variables. The effect of 5 partial variables was excluded: light, biomass, their squared effects and combined effect. $CV_{increment}$ is the estimate of the difference in extent of clonal mobility (coefficient of variation of rhizome increments) in coexisting species. All communities are analysed together, $n = 104$.

	Ramet density	Species density	Ramet life span	Rhizome increment	Branching intensity
Species density	0.44 ***				
Ramet life span	–0.09	–0.04			
Rhizome increment	-0.13	–0.17	0.81 ***		
Branching intensity	0.39 ***	0.23 *	0.34 **	–0.39 ***	
$CV_{increment}$	–0.26 **	–0.07	0.09	–0.02	–0.30 *

*** – p < 0.001; ** – p < 0.01; * – p < 0.05

Table 6. Pearson's partial correlation coefficients (r) between studied variables. The effect of 5 partial variables was excluded: light, biomass, their squared effects and combined effect. $CV_{increment}$ is the estimate of the difference in extent of clonal mobility (coefficient of variation of rhizome increments) in coexisting species. All communities are analysed together, $n = 104$. Statistically significant p-levels (p) are in bold script. The critical experimentwise error rate $p_{critical} = 0.0034$

	Ramet density		Species density		Ramet life span		Rhizome increment		Branching intensity	
	r	p	r	p	r	p	r	p	r	p
Species density	0.44	**< .0001**								
Ramet life span	–0.09	0.360	–0.04	0.690						
Rhizome increment	–0.13	0.210	–0.17	0.084	0.81	**< .0001**				
Branching intensity	0.39	**< .0001**	0.23	0.024	0.34	**0.001**	–0.39	**< .0001**		
$CV_{increment}$	–0.26	0.010	–0.07	0.471	0.09	0.366	–0.02	0.819	–0.30	**0.002**

Mudelid – parameetrite hinnangud jmt

Table 2. Estimates ±SE of effect of parity on studied traits (zero refers to class of comparison) and percentages of non-coagulated (NCM) and poorly (E_{30}<20 mm) coagulated (NK_{20}) milk samples of all samples in respective parity class.

Trait	Parity				P value
	1	2	3–4	≥5	
Number of samples	125	63	106	40	
Daily milk yield, kg	0^{a}	$0.72±1.03^{ab}$	$1.50±0.88^{b}$	$2.20±1.06^{b}$	0.1605
Fat, %	0^{a}	$0.50±0.19^{b}$	$0.31±0.16^{b}$	$0.19±0.19^{ab}$	0.0665
Protein, %	0^{a}	$0.13±0.09^{a}$	$0.07±0.08^{a}$	$0.05±0.09^{a}$	0.5461
log RCT	0^{a}	$-0.02±0.04^{ab}$	$-0.06±0.03^{b}$	$-0.02±0.04^{ab}$	0.2798
[1]E_{30}, mm	0^{ab}	$0.92±2.32^{a}$	$0.98±1.97^{a}$	$-3.47±2.31^{b}$	0.2321
NCM, %	10.4	6.3	1.9	–	
NK_{20}, %	10.4	6.3	5.6	5.0	

[1]Estimates of curd firmness of coagulating (E_{30}>0 mm) milk samples

[a,b]Estimates within row with differing letters in superscript are significantly different (P<0.05)

Table 2. Statistical results for the effect of adding calcium chloride (40% w/w) and water on chemical and functional properties of Mozzarella cheese after 42 d of storage at 4°C.

		Contrast		
Source/variable	Model[1]	Uninjected vs. Calcium	Uninjected vs. Water	Calcium vs. Water
	P<	*P*<	*P*<	*P*<
Calcium	0.001	0.001	NS[2]	0.001
Moisture	0.001	0.001	NS	0.001
pH	0.001	0.001	NS	0.001
Weight	0.001	0.001	NS	0.001
Cohesiveness	0.01	0.001	NS	0.001
Hardness	0.05	0.01	NS	0.001
Melting	0.001	0.001	NS	0.001

[1]$Y_{jkl} = \mu + T_j + B_k + e_{jk} + d_{jkl}$, where Y is the variable of interest, μ is the overall mean, T is the treatment effect, B is the block effect, e is the error term, and d is the subsample effect.

[2]NS: not significant i.e., $P \geq 0.05$.

Mudelid – parameetrite hinnangud jmt

J. Dairy Sci. 86:2308–2319

Relationships Among Milk Yield, Body Condition, Cow Weight, and Reproduction in Spring-Calved Holstein-Friesians

F. Buckley*, K. O'Sullivan†, J. F. Mee*, R. D. Evans*, and P. Dillon*
*Dairy Production Research Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland
†Statistical Laboratory, Dept. of Statistics, University College Cork, Ireland

Table 5. Association between milk production, BCS, and BW variables, and SR21.[1]

Model[2]	OR	95% CI	*P*-value
Model relating milk production variables to likelihood of SR21	n = 2753		R^2 = 0.164
Estimated 200-d milk protein content (g/kg)			
<31.8	1		
31.8 to 33.0	1.20	0.90–1.62	NS
33.1 to 34.4	1.52	1.10–2.09	0.011
>34.4	1.54	1.10–2.14	0.012
Protein-to-fat ratio at herd SBD			
<0.81	1		
0.81 to 0.90	1.34	0.98–1.84	0.066
0.91 to 1.00	1.11	0.80–1.52	NS
>1.00	1.45	1.03–2.05	0.036
Model relating BCS variables to likelihood of SR21	n = 2204		R^2 = 0.131
Average BCS between 60 and 100 d of lactation (BCS units)			
≤2.50	0.59	0.44–0.78	<0.001
2.75 to 3.0	1		
≥3.25	0.90	0.63–1.31	NS
Model relating BW variables to likelihood of SR21	n = 1483		R^2 = 0.194
BW at herd SBD (kg)			
<483	1		
483 to 529	1.33	0.82–2.17	NS
530 to 576	1.20	0.70–2.07	NS
>576	1.90	1.00–3.60	0.048
BW loss from precalving to nadir (kg)			
>131	1		
110 to 131	1.81	1.15–2.86	0.011
88 to 109	1.01	0.65–1.55	NS
<88	1.17	0.72–1.90	NS
BW gain from herd SBD to 90 d thereafter (kg)			
<17	1		
17 to 34	1.08	0.69–1.70	NS
35 to 52	1.64	1.00–2.69	0.052
>32	1.60	0.91–2.82	0.100

[1]SR21 = Submission in the first 3 wk of the breeding season, OR = odds ratio, CI = confidence interval, SBD = herd start of breeding date, n = number of cows included in analysis, NS = P > 0.10.

[2]All models were adjusted for herd, calving period, lactation number, proportion of Holstein-Friesian genes, breeding value for milk yield, and degree of calving assistance.