

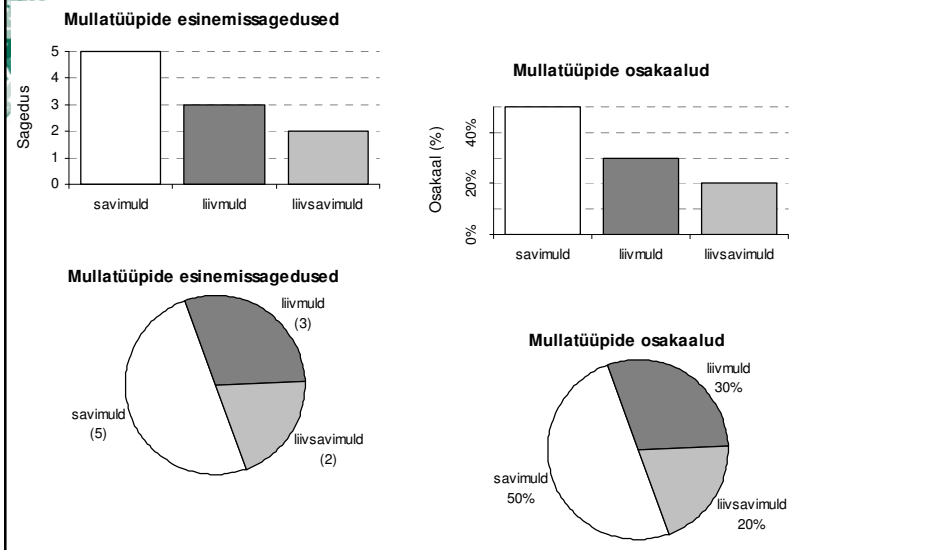
Sagedused ja osakaalud – diskreetne tunnus

Mittearvuliste või diskreetsete tunnuste (erinevate väärtuste arv suhteliselt väike) ülevaatlikuks kirjeldamiseks on lihtne lugeda kokku, mitu korda iga erinevat väärtust esineb ja kirjutada saadud arvud tabeli kujul. Väärtuse esinemiste arvu nimetatakse tema **sageduseks**.

Tihti leitakse lisaks iga väärtuse (protsentuaalne) **osakaal** valimis, mida nimetatakse ka **suhteliseks sageduseks**.

| Mullatüüp | Niiskus | Suvinisu viljakus (kg/ha) | Mullatüüp | Sagedus | Osakaal | Osakaal (%) |
|--------------|---------|------------------------------|--------------|---------|---------|-------------|
| savimuld | niiske | 3624 | savimuld | 5 | 0,5 | 50% |
| liivsavimuld | paras | 4782 | liivmuld | 3 | 0,3 | 30% |
| savimuld | niiske | 4274 | liivsavimuld | 2 | 0,2 | 20% |
| liivmuld | kuiv | 3927 | | | | |
| savimuld | paras | 4630 | | | | |
| liivmuld | paras | 4920 | | | | |
| savimuld | niiske | 4260 | | | | |
| savimuld | paras | 4935 | | | | |
| liivsavimuld | paras | 5035 | | | | |
| liivmuld | kuiv | 4500 | | | | |

Sagedustabeli asemel võib kokkuvõtliku info väärtuste esinemissagedustest esitada ka kas **tulp-** või **ringdiagrammina** (sektordiagrammina).

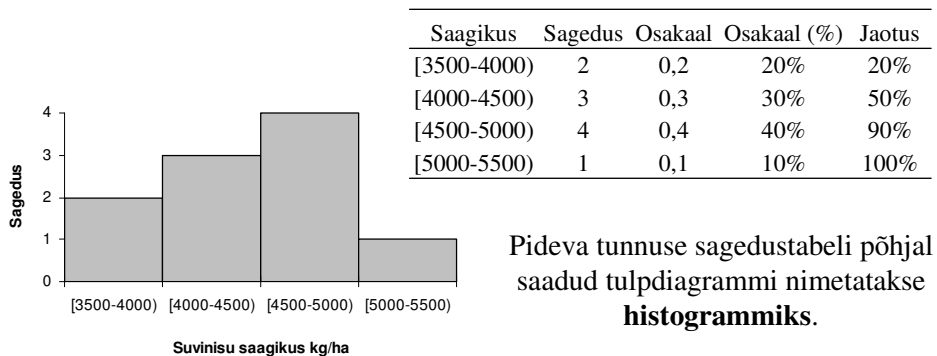


Sagedused ja osakaalud – pidev tunnus

Pidevate tunnuste puhul on tunnuse võimalike väärtuste arv (teoreetiliselt) lõpmatu – seega kui sagedustabelis vastaks igale väärtusele üks rida, siis kaoks praktiliselt erinevus sagedustabeli ja originaalandmete vahel.

Seetõttu jagatakse tunnuse võimalikud väärtused intervallidesse ja sagedustabel näitab, mitu väärtust langeb ühte või teise intervalli.

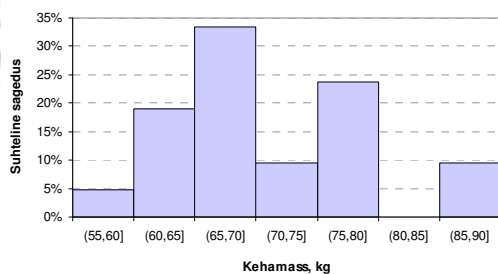
Intervallide arv ei tohiks olla liiga suur ja see oleneb valimi suurusest ($\approx \sqrt{n}$).



Pideva tunnuse sagedustabeli põhjal saadud tulpdiagrammi nimetatakse **histogrammiks**.

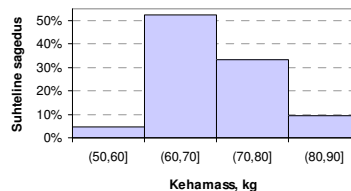
Märkusi ja soovitusi

Erinevalt tulpdiagrammist, mis on antud andmete korral üheselt määratud, võime samade andmete põhjal saada üsna erineva kujuga histograme.



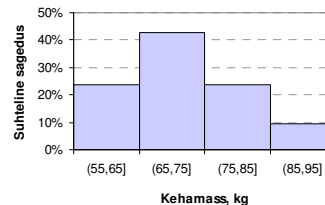
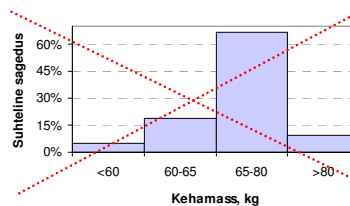
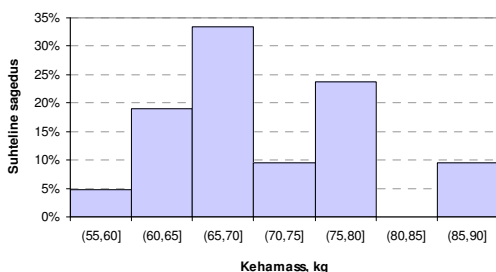
| Lammaste kehamass, kg | Sagedus | Suhteline sagedus |
|-----------------------|---------|-------------------|
| (55,60] | 1 | 4,76% |
| (60,65] | 4 | 19,05% |
| (65,70] | 7 | 33,33% |
| (70,75] | 2 | 9,52% |
| (75,80] | 5 | 23,81% |
| (80,85] | 0 | 0,00% |
| (85,90] | 2 | 9,52% |

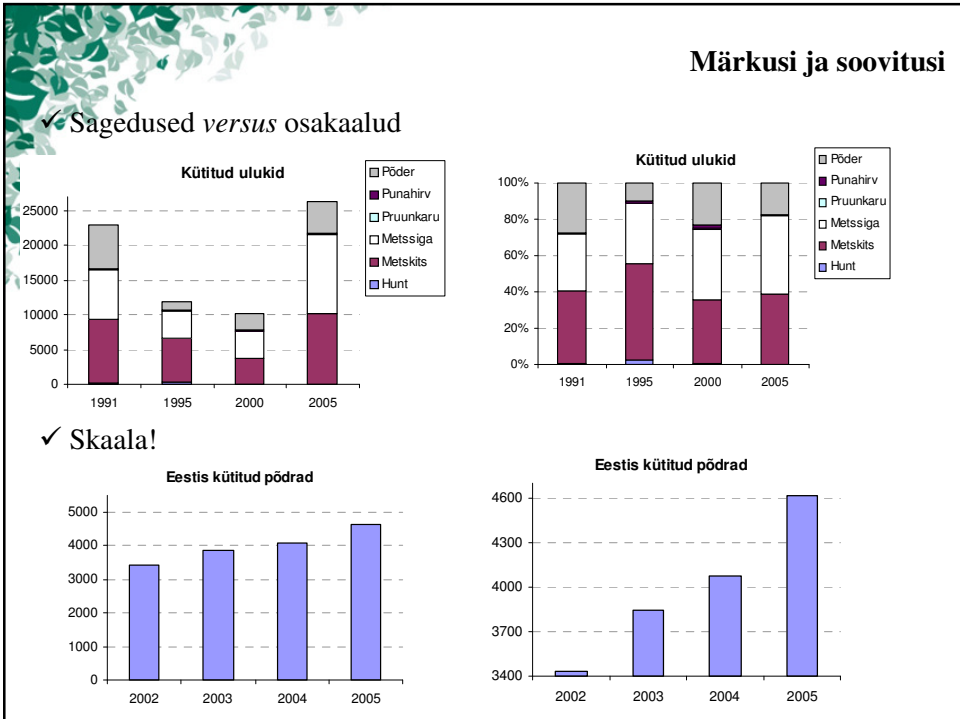
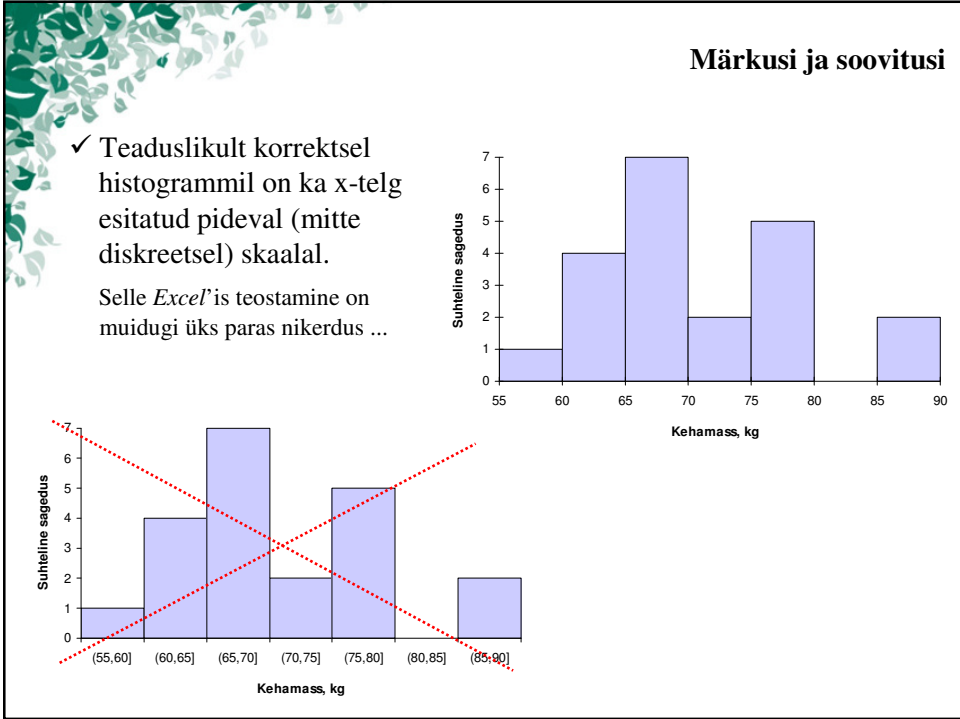
| Lammaste kehamass, kg | Sagedus | Suhteline sagedus |
|-----------------------|---------|-------------------|
| (50,60] | 1 | 4,76% |
| (60,70] | 11 | 52,38% |
| (70,80] | 7 | 33,33% |
| (80,90] | 2 | 9,52% |



Märkusi ja soovitusi

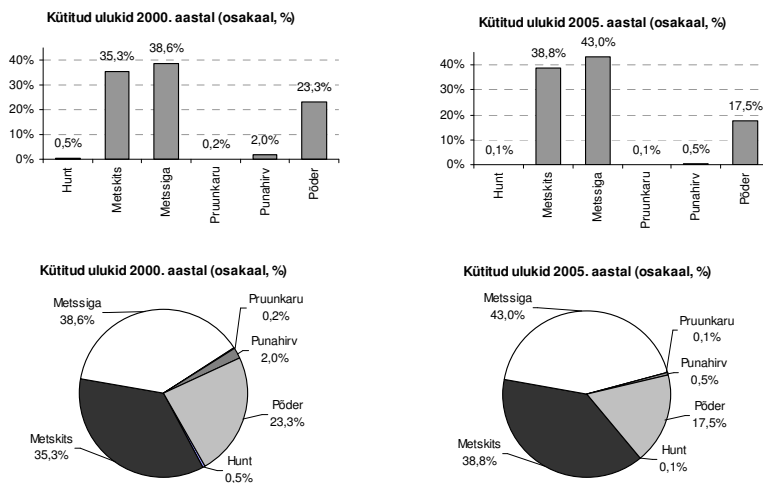
- ✓ On tungivalt soovitav, et kõik kasutatud vahemikud oleksid võrdse pikkusega!
- ✓ Joonisele tuleb kanda ka vahemikud, kuhu ühtki objekti ei sattunud!
- ✓ Avatud vahemikke tuleks võimaluse korral vältida.





Märkusi ja soovitusi

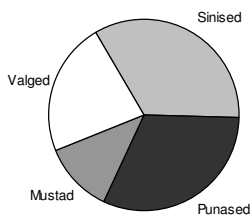
✓ Ringdiagrammide eelistada tulpdigrammi (eriti võrdlemisel).



Märkusi ja soovitusi

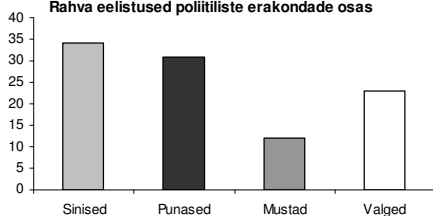
✓ Vältida tuleks 3-mõõtmelisi graafikuid, eriti ringdiagramme.

Rahva eelistused poliitiliste erakondade osas

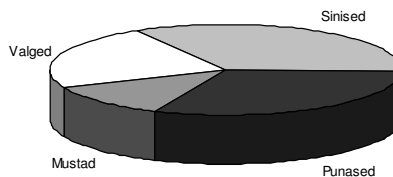


| Erakond | Osakaal (%) |
|---------|-------------|
| Sinised | 34 |
| Punased | 31 |
| Mustad | 12 |
| Valged | 23 |

Rahva eelistused poliitiliste erakondade osas



Rahva eelistused poliitiliste erakondade osas



Arvkarakteristikud

- andmestiku suurus (valimi maht, *sample size*) – n
- (aritmeetiline) keskmine [*average, mean*] – $\bar{x} = \sum_{i=1}^n x_i / n$
- mediaan (nn 50%-punkt) [*median*]
- mood [*mode*] – enim esinev (suurima sagedusega) väärtus

Näide. Uuringu all olnud 5-l haigestunud loomal määrati haiguse peiteajaks vastavalt 8, 16, 12, 60 ja 14 päeva (üks uuritud loomadest oli ilmselt geneetiliselt erinev või siis sai juba mingit muud, haiguse avaldumist pärssivat ravi). Haiguse keskmine peiteaeg on

$$\bar{x} = \frac{8+16+12+60+14}{5} = \frac{110}{5} = 22 \text{ päeva.}$$

Peiteaeg, millest pooltel loomadel avaldus haigus varem ja pooltel hiljem, on leitav kui kasvavalt järjestatud peiteaegade keskmine väärtus e mediaan:

$$8, 12, \boxed{14}, 16, 60 \\ = med$$

Keskmise omadusi

1. $\overline{cx} = c\bar{x}$, kus c on konstant
2. $\overline{x+c} = \bar{x} + c$, kus c on konstant
3. $\overline{x+y} = \bar{x} + \bar{y}$
4. $\sum_{i=1}^n x_i = n\bar{x}$
5. $\overline{f(x)} \neq f(\bar{x})$, kus f on monotoonne teisendus

Mediaani omadusi

1. $med(f(x)) = f(med(x))$, kus f on monotoonne teisendus

Näiteks, kui $med(\log_{10}(x)) = 2$,

siis $\log_{10}(med(x)) = 2 \Rightarrow med(x) = 10^2 = 100$.

2. $\sum_{i=1}^n x_i \neq n \times med(x)$

Vaatluste hajuvus

- miinimum, maksimum, haare [*range*] = $\max - \min$
- standardhälve [*standard deviation*] – $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
- dispersioon [(*sample*) variance] – s^2
- standardviga [*standard error*] – $se = s/\sqrt{n}$

Näide. Uuriti 5 metsiku ja 4 puhtatöulise laborihiire reaktsiooni ärritajale. Tulemuseks saadi järgmised väärtused:

metsikud hiired – 15, 45, 30, 10, 25; labori hiired – 20, 25, 30, 25.

Keskmiised reaktsioonid kummagi grupi jaoks on

$$\bar{x}_m = \frac{15+45+30+10+25}{5} = \frac{125}{5} = 25, \quad \bar{x}_l = \frac{20+25+30+25}{4} = \frac{100}{4} = 25.$$

$$s_m = \sqrt{\frac{(15-25)^2 + (45-25)^2 + (30-25)^2 + (10-25)^2 + (25-25)^2}{5-1}} = \sqrt{\frac{750}{4}} = \sqrt{187,5} \approx 13,69;$$

$$s_l = \sqrt{\frac{(20-25)^2 + (25-25)^2 + (30-25)^2 + (25-25)^2}{4-1}} = \sqrt{\frac{50}{3}} \approx \sqrt{16,67} \approx 4,08.$$

Standardhälbe ja dispersiooni omadusi

1. $s^2(cx) = c^2 s^2(x)$, kus c on konstant
2. $s(cx) = cs(x)$
3. $s^2(x+c) = s^2(x)$
4. $s(x+c) = s(x)$
5. kui x ja y on sõltumatud uuritavad tunnused, siis

$$s^2(x+y) = s^2(x) + s^2(y)$$


Teades vaid uuritava tunnuse keskvaartust (populatsiooni keskmist) ja standardhälvet, võime uuritava tunnuse väärtuste kohta öelda järgmist:

- vähemalt 3/4 uuritava tunnuse väärtustest asuvad keskvaartusele lähemal kui kaks standardhälvet (enamasti asub kahe standardhälbe kaugusel keskvaartusest umbes 95% vaatlustest);
- vähemalt 8/9 uuritava tunnuse väärtustest asuvad keskvaartusele lähemal kui kolm standardhälvet (enamasti asub kolme standardhälbe kaugusel keskvaartusest rohkem kui 99% vaatlustest).

Arvkarakteristikud Näiteid kirjandusest

A comparison of the methods for determination of the rennet coagulation properties of milk

Acta Agriculturae Scand Section A, 2005; 55: 145–148

 Taylor & Francis
Taylor & Francis Group

IVI KÜBARSEPP¹, MERIKE HENNO¹, OLAV KÄRT¹ & TUOMO TUPASELA²

¹Department of Animal Nutrition, Institute of Veterinary Medicine and Animal Science, Estonian University of Life Sciences, Kreutzwaldi 48, 51006 Tartu, Estonia, and ²MTT Food Research, Myllytie 1, 31600, Jokioinen, Finland

Table I. Means, ranges and standard deviations (SD) for milk composition and rennet coagulation parameters.

| | Mean | Min. | Max. | SD |
|----------------------------|-------|------|-------|-------|
| Fat, % | 3.94 | 2.70 | 8.08 | 0.790 |
| Protein, % | 3.41 | 2.56 | 4.62 | 0.456 |
| Lactose, % | 4.81 | 4.38 | 5.18 | 0.174 |
| Formagraph | | | | |
| RCT, min | 9.5 | 3.5 | 35 | 4.95 |
| E ₃₀ , mm | 26.3 | 0 | 52 | 10.34 |
| Optigraph | | | | |
| R _{initial} , min | 6.63 | 3.73 | 19.00 | 2.423 |
| R, min | 9.53 | 4.36 | 31.59 | 4.320 |
| A ₃₀ , V | 13.72 | 0 | 35.98 | 5.855 |

Arvkarakteristikud Näiteid kirjandusest

ISSN 1392-2130. VETERINARIA IR ZOOTECNIKA. T. 36 (38), 2006

EFFECT OF BODY CONDITION SCORE AT PARTURITION ON THE PRODUCTION PERFORMANCE, FERTILITY AND CULLING IN PRIMIPAROUS ESTONIAN HOLSTEIN COWS

Jaak Samuël, Katri Ling, Hanno Jaakson, Tanel Kaart, Olav Kart
Institute of Veterinary Medicine and Animal Science, Estonian University of Life Sciences, 1 Kreutzwaldi St.
51006, Tartu, Estonia. tel. + 372-7-313-476; fax: + 372-7-313-477; e-mail: Jaak.Samuelyel@emu.ee

Table 2. Fertility parameters of the first parity Estonian Holstein cows grouped by BCS at parturition

| Fertility parameters | Body condition score at calving | | |
|--|---------------------------------|-------------------|----------------|
| | ≤3.0 (n = 26) | 3.25–3.5 (n = 39) | ≥3.75 (n = 21) |
| Interval calving to first service (days) | 91 ± 4.1 | 83 ± 3.5 | 88 ± 5.4 |
| First service conception rate (%) | 17 | 23 | 0 |
| Service period (days) | 82 ± 14.4 | 72 ± 13.9 | 77 ± 15.8 |
| Days open (of those pregnant) | 173 ± 13.7 | 155 ± 14.8 | 165 ± 16.6 |
| Services per conception | 3.0 ± 0.36 | 3.0 ± 0.32 | 3.6 ± 0.42 |
| Number of cows not pregnant | 1 | 5 | 5 |

Values are arithmetical means ± S.E.

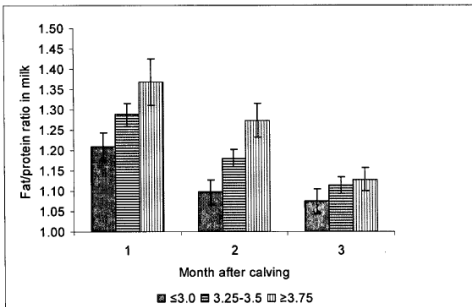


Figure 2 Milk fat/protein ratio of the first lactation Estonian Holstein cows during the first 3 months after calving. Values are means ± S.E. Milk fat/protein ratio was different ($P < 0.05$) between BCS ≤3.0 (*thin*) and ≥3.75 (*fat*) groups during the first and second months of lactation

- Variatsioonikordaja [*coefficient of variation*] – $v = \frac{s}{\bar{x}} \times 100\%$

Aga mis siis, kui keskmine on negatiivne?

Näide.

| | Piim, kg | Rasv, % | Valk, % | SRA, tuh/ml | Energia- bilanss, MJ |
|--------------|-------------|------------|------------|----------------|-------------------------|
| Keskmine | 30,23 | 4,13 | 3,17 | 695,92 | -36,24 |
| St. hälve | 5,32 | 0,74 | 0,24 | 1111,99 | 52,99 |
| Var. kordaja | 17,60 | 17,98 | 7,59 | 159,79 | -146,22 |

Kvantiilid, protsentiilid

- kvantiilid** – alumine kvartiil e 25%-punkt ja ülemine kvartiil e 75%-punkt [*lower, upper quartile*]
- kvantiilide vahe** [*interquartile range, IQR*] – kasutatakse varieeruvuse iseloomustamiseks
- detsiilid, protsentiilid e protsendipunktid/**kvantiilid**
 - α -kvantiiliks [*α -quantile*] nimetatakse sellist uuritava tunnuse väärtust, millest väiksemate väärtuste osakaal mõõtmistulemuste seas on α .
- min, max

