

Statistikafunktsioonid

Üsna sageli tahetakse teha järeldusi indiviidide või esemete teatud kogu kui terviku kohta või siis uurida looduse ja ühiskonna nähtusi kui juhuslikke protsesse. Näiteks tahetakse teada Eestis elavate 15-aastaste tüdrukute keskmist pikkust või kuidas jaotuvad eestlased silmade värvuse järgi jne. Sellistele küsimustele vastuse leidmine toimub matemaatika ühe haru - **matemaatiline statistika** - meetodeid kasutades.

Uuritavat indiviidide või esemete kogu või uuritavat nähtust, mille kohta tahetakse otsust langetada, nimetatakse **uuritavaks kogumiks**. Kogumit uuritakse tema objektide mingi omaduse, mida nimetatakse **tunnuseks**, seisukohalt. Tunnuseks võib olla näiteks pikkus, kaal, silmade värvus jne.

Uuritava kogumi objektide mõõtmisel saadakse vaadeldava tunnuse väärtuste rida nn. **statistiline rida**.

Näiteks ühe klassi kontrolltööde hinded olid sellised:

2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5.

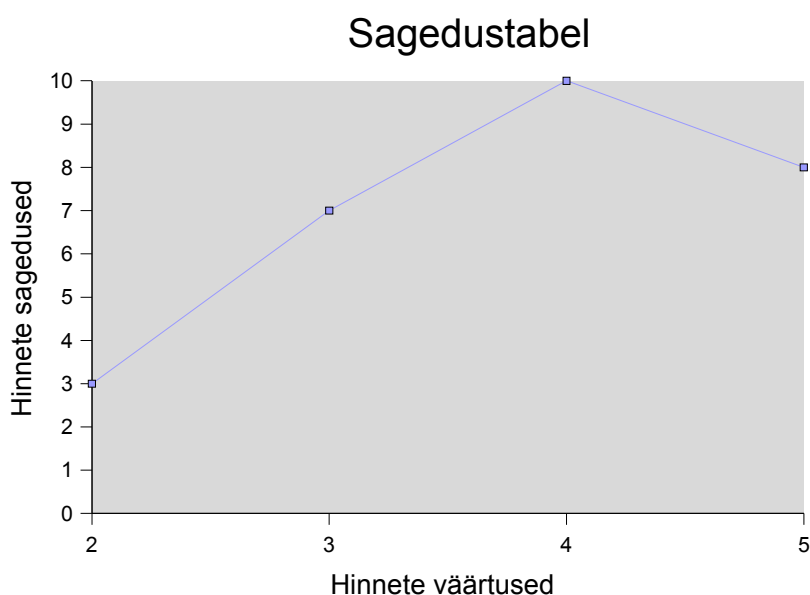
Siit me saame leida hinnete **vähima väärtuse MIN** ja **suurima väärtuse MAX**. Excelis ja Calc'is leitakse sellised väärtused järgnevalt **=MIN(arv1; arv2; ...)** - tagastab miinimumväärtuse argumentide hulgast, kus arv1, arv2 jne. on vastavalt väärtused ning **=MAX(arv1; arv2; ...)** - tagastab maksimumväärtuse argumentide hulgast. Antud näite korral on miinimumväärtuseks 2 ja maksimumväärtuseks 5.

Otstarbekas on esitada andmeid **sagedustabelina**, kus igale väärtusele vastab tema esinemiste arv. Antud näitele vastaks selline tabel:

Hinne (x)	2	3	4	5
Sagedus (f)	3	7	10	8

Kui on antud väärtuste jada, siis on võimalik leida ka Excelis ja Calc'is sagedustabel, kasutades funktsiooni **=FREQUENCY(andmed; klassid)** - tagastab andmemassiivi sagedustabeli. Andmeteks on antud näite korral hinnete rida/vahemik ja klassid on 2, 3, 4, 5. Klassid peavad ka olema ära märgitud vastavasse veergu, mille kõrvale soovitakse leida vastavate väärtuste esinemissagedust.

Saadud andmete põhjal on võimalik joonistada ka **sagedushulknurk** ehk sagedusmurdjoon. Antud näitele vastav diagramm on järgmisel lehel.



Veel on võimalik nüüd kergesti leida **kogumi mahtu N** (tunnuse väärtuste arvu), mis antud näite korral oleks $N = 28$ (sageduste summa).

Ülesanne 1.

Meeste kingakaupluses müüdi ühe tunni jooksul 20 paari kingi numbritega 39, 41, 40, 41, 44, 40, 42, 41, 43, 39, 42, 41, 42, 38, 42, 41, 43, 41, 39, 40. Koostage vastav sagedustabel ja joonistage sagedushulknurk. Milline oli kõige väiksem ja kõige suurem kinganumber (kasutage MIN ja MAX)?

Et võrrelda erineva mahuga kogumeid, on otstarbekas esitada sageduste asemel **suhtelised sagedused**. Suhteline sagedus tähistatakse w , mis leitakse valemi

$$w = \frac{f}{N} \quad \text{või} \quad w = \frac{f}{N} \cdot 100 \% \quad \text{kaudu.}$$

Tabelit, kus tunnuse väärtustele on seatud vastavusse nende esinemise suhteline sagedus, nimetatakse **jaotustabeliks**. Hinnete näite korral näeks jaotustabel välja selline:

x	2	3	4	5
w	11%	25%	36%	28%

Jaotustabeli vastavat sirglõikdiagrammi nim. **jaotushulknurgaks**.

Ülesanne 2.

Koostage oma klassi õpilaste jaotustabel silmade värvuse järgi. Koostage vastav sektordiagramm.

Statistiliste andmete kogumisele järgneb andmete töötlemine ehk **andmeanalüüs**. Selle käigus korrastatakse sobivalt andmestikku ning leitakse teatud suurused, nn. **karakteristikud**, mis iseloomustavad tunnuse väärtuse jaotust kui tervikut ühest või teisest seisukohast. Ühtedeks karakteristikuteks on aritmeetiline keskmine, mediaan ja mood.

Aritmeetiliseks keskmiseks nimetatakse kogumi tunnuse kõigi väärtuste summa ja väärtuste (objektide) arvu (märgitud N-ga, ei tähista kogumi mahtuvust) jagatist. Leitakse sellise valemi kaudu

$$\bar{x} = \frac{a_1 + a_2 + a_3 + \dots + a_N}{N}$$

Hinnete näite korral oleks aritmeetiline keskmine võrdne 3,82-ga. Excelis ja Calc'is saadakse sama tulemus, kui kasutada funktsiooni **=AVERAGE(arv1; arv2; ...)** - tagastab valimi keskmise.

Mediaaniks nimetatakse tunnuse väärtust, millest suuremaid (või võrdseid) ja väiksemaid (või võrdseid) liikmeid on variatsioonireas ühepalju. Kui variatsioonireas on paaritu arv liikmeid - N on paaritu arv (N tähistab siin liikmete arvu mitte kogumi mahtu), on mediaaniks variatsioonirea keskmine liige, kui aga variatsioonireas on paarisarv liikmeid, loetakse mediaaniks kahe keskmise liikme aritmeetiline keskmine. Lühemalt:

$$Me = x_i, \text{ kus } i = \frac{1}{2}(N + 1), \text{ kui } N \text{ on paaritu arv,}$$

$$Me = \frac{1}{2}(x_i + x_{i+1}), \text{ kus } i = \frac{N}{2}, \text{ kui } N \text{ on paarisarv.}$$

Excelis ja Calc'is on antud funktsiooni nimi **=MEDIAN(arv1; arv2; ...)** - tagastab antud valimi mediaani. Meie antud näiteülesande korral on mediaan 4, sest N on paaritu arv.

Moodiks nimetatakse tunnuse kõige sagedamini esinevat väärtust. Moodi tähistatakse sümboliga Mo. Excelis ja Calc'is leitakse mood, kasutades funktsiooni **=MODE(arv1; arv2; ...)** - tagastab valimi kõige sagedasema väärtuse. Illustreeritud näite korral oleks selleks 4, sest seda hinnet esineb kõige rohkem.

Ülesanne 3.

Ühe klassi noormeeste kinganumbrite variatsioonirida on 39, 39, 40, 40, 40, 40, 41, 41, 41, tütarlastel aga 35, 35, 35, 35, 36, 39. Leidke vastavad aritmeetilised keskmised, mediaanid ja moodid.

Ülesanne 4.

Leidke oma klassi õpilaste silmade värvuse mood.

Ülesanne 5.

Leidke oma klassi tütarlaste ja noormeeste keskmine pikkus. Koostage vastav sagedus- ja jaotustabel ning vastavad hulknurgad.