

Jaotuse parameetrid (arvkarakteristikud)

Asendiparameetrid, keskmine, aritmeetiline keskmine, mediaan, mood, jaotuse sümmeetrilisus, kvartiilid, karpdiagramm

Põhikool

Asendiparameetrid

Katse tulemusena saadud andmestik võib olla väga mahukas. Ka sagedustabel ega empiirilist jaotust esitav histogramm pole mõnikord piisavalt kompaktne, kui andmestikku soovitakse esitada üheainsa või väheste arvude abil.

Niisuguseid jaotust iseloomustavaid arve nimetatakse **jaotuse parameetriteks** (arvkarakteristikuteks).

Parameetrite seas on kõige tuntumad **asendiparameetrid** (asendikarakteristikud), mis iseloomustavad katsetulemuste üldist suurust ehk empiirilise jaotuse paiknemist arvteljel.

Empiirilise jaotuse keskmine

Kõige tavalisem asendikarakteristik on **keskmine**.

Empiirilise jaotuse keskmine on andmestiku üksikväärtuste aritmeetiline keskmine, mis leitakse, liites kõik väärtused kokku ja jagades summa liidetavate arvuga. Keskmise lähedased väärtused on vaadeldava andmestiku seisukohast tüüpilised, kuid keskmine ei tarvitse ühtida andmestiku ühegi väärtusega.

Näide: Olgu mõõdetud 7A klassi õpilaste pikkused, neid oli mõõtmispäeval koolis 17.

160 148 172 165 163 158 177 181 175 171 155 159 175 166 174 180 173
Arvutamise tulemusena leiame, et õpilaste keskmine pikkus on 167,8. Seega käesoleva klassi kõige tüüpilisema pikkusega on õpilane, kelle pikkus on 168 cm.

Kui tunnus on esitatud jaotustabelina, on lihtsam keskmine arvutada tabeli abil. Selleks tuleb tunnuse väärtused korrutada nende suhteliste sagedustega ja tulemused liita.

Tabel 1. Klassi õpilaste õdede ja vendade arvu jaotus

Õdede ja vendade arv	0	1	2	3	Kokku
Õpilaste arv, kellel on nii palju õdesid-vendi	7	12	4	2	25
Osakaalud	28%	48%	16%	8%	100%

Nii saame klassi õpilaste keskmiseks õdede-vendade keskmise arvu

$$0 \cdot 0,28 + 1 \cdot 0,48 + 2 \cdot 0,16 + 3 \cdot 0,08 = 1,04.$$

Siit järeldub, et kõige tüüpilisemad on need õpilased, kellel on üks vend või õde.

Mediaan

Teine sageli kasutatav asendiparameeter on **mediaan**, mille leidmine ilma arvuti abita on eriti lihtne.

Vaatame taas A klassi õpilaste pikkusi.

Järjestame need suuruse järgi, alates kõige väiksemast, saame uue jada:

148 155 158 159 160 163 165 166 **171** 172 173 174 175 175 177 180 181

Nüüd otsime sellest jadast välja niisuguse pikkusega õpilase, kellest pikemaid ja lühemaid õpilasi on võrdne arv. Ilmselt peab see olema järjestuses üheksas – sellest on 8 väiksemat ja 8 suuremat arvu. Näeme, et pikkuste mediaan antud klassis on 171. Ka sellise pikkusega õpilast võib pidada klassi kõige tüüpilisema pikkusega õpilaseks.

Kui me tahame leida klassi B õpilaste pikkuse mediaani, tekib väike probleem: järjestatud pikkuste jadas ei ole ühte keskmist arvu, sest õpilaste arv on paarisarv.

155 157 159 160 162 163 166 **166 171** 172 173 174 174 174 179 180

Sel juhul loetakse mediaaniks kahe keskmise väärtuse aritmeetiline keskmine. Praegusel juhul on need järjestatud jada kaheksas ja üheksas arv (166 ja 171) ja mediaan on 168,5 cm.

B-klassi õpilaste keskmine pikkus on 167,81, seega samuti ligikaudu 168 cm.

Lisaks asendiparameetritele iseloomustavad andmestikku ka hajuvuse parameetrid (vt Hajuvus)

Gümnaasium

Keskmise ja mediaani vahetõrge iseloomustab jaotuse kuju. Kui jaotus on sümmeetriline, st et keskmisest suuremaid ja väiksemaid väärtusi on ühepalju ja nende suhtelised sagedused on ka ühesugused, siis langevad mediaan ja keskmine ühte. Kui keskmine on mediaanist suurem, siis leidub andmete hulgas keskmisest palju suuremaid väärtusi, väiksemad väärtused aga paiknevad keskmise lähedal. Nii on näiteks olukord palkadega: suur osa inimesi saab palka, mis on keskmisest palgast natuke väiksem, aga mõned saavad keskmisest palgast mitmeid kordi suuremat palka. Sellepärast on Eestis (aga ka teistes riikides) mediaanipalk väiksem kui keskmine palk. Vastupidise olukorra puhul on mediaan suurem kui keskmine.

Mittearvulise tunnuse parameeter. Mood

Mittearvuliste väärtustega tunnuse keskmist ei saa arvutada. Kui tunnuse väärtused on järjestatud, siis on võimalik leida mediaan. Kui väärtustel puudub ka järjestus, siis ei saa ka mediaani leida. Näiteks vaatame õpilaste silmade värvi:

Silmade värv	Sinine	Hall	Pruun	Roheline	Kokku
Arv	6	10	5	2	23

Mood on tunnuse kõige sagedamini esinev väärtus, seega silmade värvi puhul on mood „Hall“. Mood on olemas ka arvtunnustel. Näiteks kui vaadata klassi B, siis on kõige rohkem õpilasi, kelle pikkus on 174 cm. See ongi B klassi õpilaste pikkuse mood. A klassis aga on kõik õpilased erineva pikkusega ja moodpikkust ei ole.

Variatsioonrida ja kvartiilid

Väärtuste järgi järjestatud andmestikku nimetatakse **variatsioonreaks**, seda on sobiv kasutada mediaani leidmiseks. Mediaan jagab **variatsioonrea** kaheks võrdseks osaks. Samal viisil võib kaheks jagada ka variatsioonrea kummagi poole, nii saame neli enam-vähem võrdset osa ja lisaks mediaanile kaks uut jaotuspunkti. Variatsioonrea alumist poolt poolitavat punkti nimetatakse alumiseks kvartiiliks: sellest on veerand vaatlustest väiksemad ja kolmveerand suuremad. Variatsioonrea

ülemist poolt poolitab ülemine kvartiil, sellest on kolmveerand vaatlusi väiksemad ja veerand suuremad.

A klassi õpilaste pikkuste variatsioonirida:

148 155 158 **159 160** 163 165 166 **171** 172 173 174 **175 175** 177 180 181

Siin on alumise kvartiili väärtus 159,5 ja ülemise kvartiili väärtus 175.

Kvartiilide vahe (öeldakse ka **kvartiilhaare**) iseloomustab jaotuse hajuvust. A klassi puhul on see 15,5.

Kui andmestikus on vaatlusi palju, ei ole variatsioonrea moodustamine mõttekas. Alumine kvartiil, mediaan ja ülemine kvartiil on vaatlused, mis jagavad suuruse järgi järjestatud andmestiku suhtes 25:75, 50:50 ja 75:25.

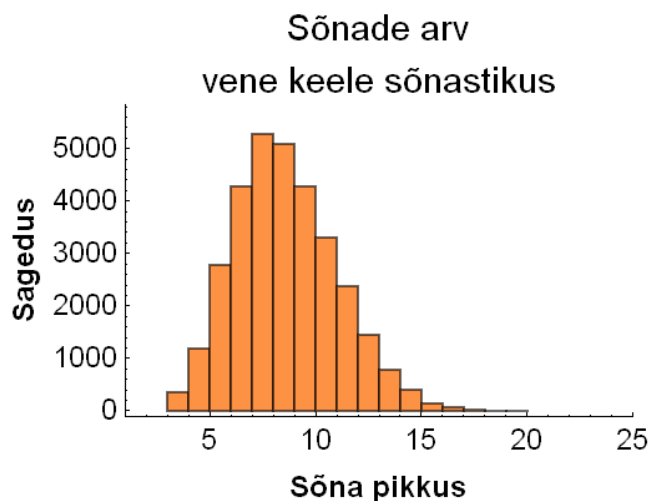
Karpdiagramm

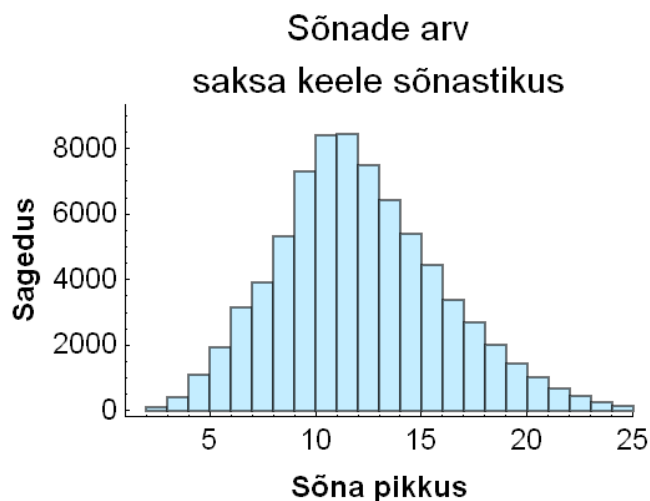
Kvartiilide ja mediaani paiknemine iseloomustab empiirilise jaotuse kuju ja seda näitlikustab **karpdiagramm**. Selle „karbi“ küljed on määratud andmestiku alumise ja ülemise kvartiiliga, keskjoon on mediaani kohal. Märkitud on ka maksimum ja miinimum, suuremate andmestike puhul ka rohkem maksimumi- või miinimumilähedasi väärtusi, mis võivad olla **erindid**.

Erinevate andmestike (osakogumite) karpdiagrammide kõrvutamine on sobiv nende visuaalseks võrdlemiseks.

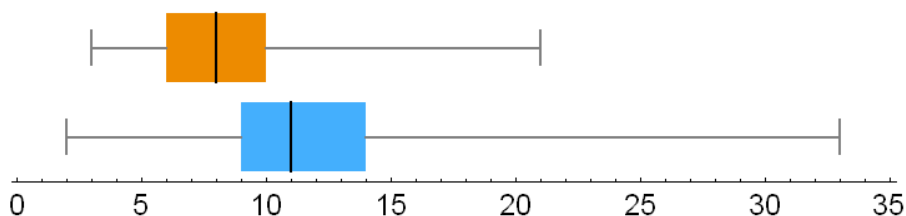
Näide

Olgu uuritavaks tunnuseks vene- ja saksakeelsete sõnade pikkus. Nende tunnuste empiirilist jaotust iseloomustavad allpool esitatud tulpdigrammid.





Täiendava iseloomustuse saamiseks moodustatakse andmestikest ka karpdiagrammid.



Neist järeldeb, et

- Saksa keele kõige lühemad sõnad on lühemad kui vene keele kõige lühemad sõnad;
- Saksa keele kõige pikemad sõnad on palju pikemad kui vene keele kõige pikemad sõnad;
- Saksa keele sõnade mediaanpikkus (millest pooled sõnastiku sõnad on lühemad ja pooled pikemad) on 11 tähte, vene keeles aga 8 tähte.
- Vene keeles on pooled sõnad pikkusega 6–10 tähte, sealjuures on mediaanist lühemaid ja mediaanist pikemaid sõnu praktiliselt võrdselt.
- Saksa keeles on pooled sõnad pikkusega 9–14 tähte, kusjuures veerand sõnadest on pikkusega 9–11 ja veerand pikkusega 11–14 tähte, st et jaotus kindlasti ei ole sümmeetriline.
- Ka venekeelsete sõnade pikkusejaotus ei ole sümmeetriline, vaid on pikkade sõnade suunas „välja venitatud“.