Mõõtmise võimalused?

Kelle või millega tegemist – **nominaalskaala**

Andmete alamhulkadeks (klassideks) jagamisel kasutatakse nominaalskaalat, mis võimaldab kirjeldada ainult omadusi. Näiteks soo skaalal on kaks väärtust: „mees“ ja „naine“, rahvuse skaalal väärtused „eestlane“, „venelane“, „ukrainlane“, „soomlane“, … „muu“. Andmete nominaalskaalal mõõdetud väärtuste järgi järjestamisel ei ole mõtet. Kuigi kodeerimise abil saab nominaalskaalat esitada ka arvudega, näiteks „mees“ = 1, „naine“ = 2, kuid tehted nende arvudega ei näita nagu järjestaminegi midagi. Nominaalskaalal mõõdetud väärtusi nimetatakse kvalitatiivseteks väärtusteks.

Kes või mis seisab kelle või mille ees või taga – **järjestusskaala**

Andmete järjestamiseks kasutatakse järjestusskaalat. Näiteks inimeste järjestamiseks vanuse järgi väärtusi „noorem“, „vanem“, eksamitulemuste järgi järjestamisel hindeid A, B, C, D, E, F. Paljudes arvamusuuringutes kasutatakse skaalat „väga halb“, „halb“, „keskmine“, „hea“, „väga hea“ või „täiesti nõus“, „peamiselt nõus“, „peamiselt ei ole nõus“, „ei ole üldse nõus“ vms. Järjestusskaalal mõõdetud andmetele võib leida mediaani ehk väärtuse, millest pool andmetest on ühel pool ja pool teisel pool, kuid aritmeetilise keskmise arvutamisel puudub mõte.

Kui suur on kahe mõõdetud väärtuse vahe – vahemik- ehk **intervallskaala**

Temperatuuriskaala on intervallskaala. Eilse ja tänase ilma puhul saab mõõta õhutemperatuuri ja ka nende temperatuuride vahet, kuid mitte seda, mitu korda on eilne õhutemperatuur suurem või väiksem tänasest. Vahemikskaalal võib 0-punkti valida kuhu tahes. Kui mõõdame temperatuuri Celsiuse skaalal, märgib 0 temperatuuri, kus vesi muutub jääks ja 100 temperatuuri, kus vesi aurustub. Reamuri skaalal on 0 samas kohas, kus Celsiuse skaalalgi, kuid vee muutumine auruks toimub temperatuuril 80. Fahrenheiti skaala järgi sulab jää temperatuuril 32 ja vesi keeb temperatuuril 212.

Mitu korda on üks väärtus teisest suurem või väiksem – **suhteskaala**

Suhteskaalal on olemas tõeline 0. Suhteskaalal mõõdetakse lapse pikkust, riidekanga laiust, kauba hinda, ettevõtte kasumit, leibkonna sissetulekut, elamispinna suurust, osaku väärtust jms.

Näide. Olgu mõõdetud 19 põhikoolis käivat õpilast. Mõõtmistulemused on koondatud järgmisesse tabelisse.

Nimi	Sugu	Vanus	Pikkus, m	Kaal, kg
Martin	M	14	1,75	51,03
Kristina	N	13	1,44	38,10
Anna	N	13	1,66	44,45
Maria	N	14	1,60	46,49
Sander	M	14	1,61	46,49
Kristjan	M	12	1,46	37,65
Laura	N	12	1,52	38,33
Sandra	N	15	1,59	51,03
Kevin	M	13	1,59	38,10
Markus	M	12	1,50	45,13
Sofia	N	11	1,30	22,91
Aino	N	14	1,63	40,82
Tiiu	N	12	1,43	34,93
Tiina	N	15	1,69	50,80
Nikita	M	16	1,83	68,04
Rasmus	M	12	1,65	58,06
Artjom	M	15	1,70	60,33
Robin	M	11	1,46	38,56
Jaan	M	15	1,69	50,80

Mida mõõtmistulemused meile ütlevad?

Nimi näitab, millist õpilast on mõõdetud. Nimel korduvaid väärtusi ei ole, see on õpilase indikaator. Sugu on mõõdetud nominaalskaalal ja näitab, kas tegemist on poisi või tüdrukuga. Kuigi soo väärtused võib ümber kodeerida, näiteks M $\rightarrow$ 1 ja N $\rightarrow$ 2, siis arvudega 1 ja 2 ei ole mõtet mingeid tehteid teha ega ka nende põhjal lapsi järjestada. Vanus, pikkus ja kaal on mõõdetud suhteskaalal. Nende puhul võime huvi tunda suurima ja vähima väärtuse vastu, arvutada keskmise ja muid asendinäitajaid, uurida hajuvust keskmise ümber ja tunnustevahelist sõltuvust.

Asendinäitajad (aritmeetiline) keskmine, mediaan, kvartiilid

Kui sõna „keskmine" ees mingit täiendit ei ole, mõeldakse tunnuse aritmeetilist keskmist. Aritmeetilise keskmise leidmiseks liidetakse tunnuse väärtused kokku ja jagatakse väärtuste arvuga. Pikkuse aritmeetiline keskmine on

$$\frac{1{,}75+1{,}44+1{,}66+1{,}6+1{,}61+1{,}46+1{,}52+1{,}59+1{,}5+1{,}3+1{,}63+1{,}43+1{,}69+1{,}83+1{,}65+1{,}7+1{,}46+1{,}69}{19}=$$

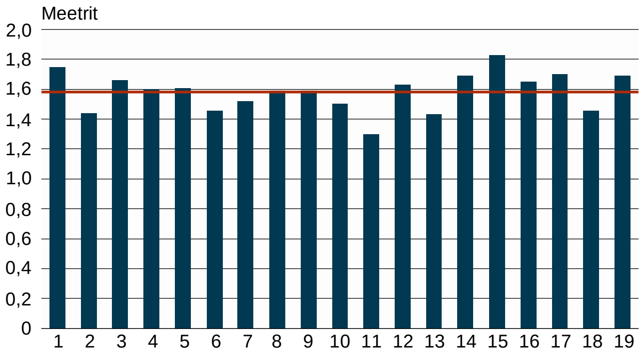
$$=1{,}58$$

Keskmise arvutamiseks on Excelis ja Libre Office'is funktsioon AVERAGE.

Analoogiliselt võib leida kaalu keskmise 45,37.

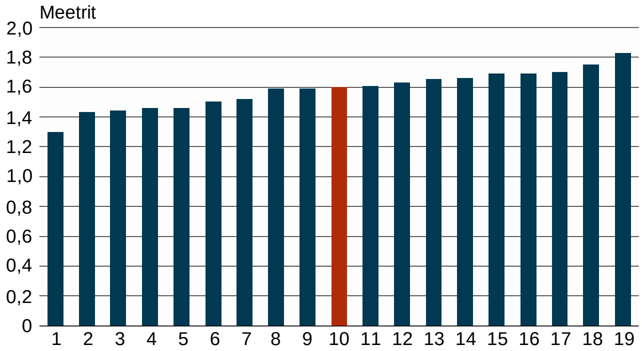
Kui laste pikkused kujutada samas järjekorras nagu tabelis, tekib järgmine pilt, kus punane joon näitab keskmist pikkust.

Õpilased pikkuse järgi



Kui aga järjestada algandmestik pikkuse järgi kasvavalt, saame järgmise pildi.

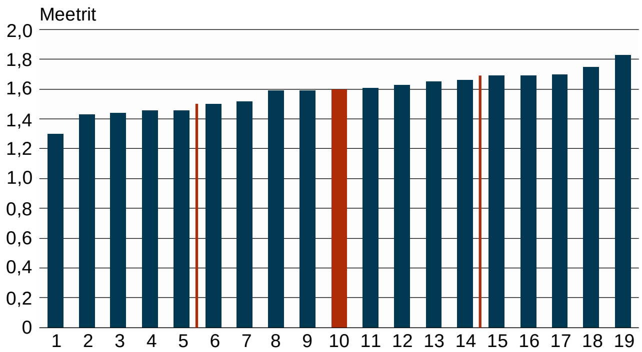
Õpilased pikkuse järgi



Keskel asuvat pikkust 1,6 nimetatakse pikkuse mediaaniks. Pool lastest on lühemad kui 1,6 m ja pool pikemad. Exceli ja Libre Office'i funktsioon MEDIAN.

Laste pikkused võib jagada ka nelja ossa, siis saame peale mediaani ka pikkuse kvartiilid.

Õpilased pikkuse järgi



Hajuvusnäitajad

Asendinäitajad ei iseloomusta andmestikku igast küljest. Arvude 10, 20 ja 30 aritmeetiline keskmine on 20 ning arvude 0, 20 ja 40 aritmeetiline keskmine on niisamuti 20. Esimene kolmik asub väiksemas vahemikus ehk ei ole nii hajuv kui teine. Seega on asendinäitajate kõrvale vaja ka hajuvusnäitajaid.

Lihtsaim hajuvusnäitaja, **haare,** on tunnuse suurima ja vähima väärtuse vahe, st selle vahemiku pikkus, kus kõik selle tunnuse väärtused asuvad. Pikkuse haare on 1,83 – 1,3 = 0,53; kaalu haare on 68,04 – 22,91 = 45,13.

Ajalooliselt on kujunenud, et andmete hajumist keskmise ümber iseloomustatakse parameetritega **dispersioon** ja **standardhälve**. Tunnuse dispersioon leitakse nii, et igast väärtusest lahutatakse aritmeetiline keskmine, vahe tõstetakse ruutu, vahede ruudud liidetakse ja lõpuks jagatakse väärtuste arvuga. Näiteks pikkuse dispersiooni puhul arvutame (lugejas on 19 liidetavat)

$$\frac{(1{,}75-1{,}58)^2+(1{,}44-1{,}58)^2+\ldots+(1{,}46-1{,}58)^2+(1{,}69-1{,}58)^2}{19}=0{,}016$$

Kaalu dispersioon on

$$\frac{(51{,}3-45{,}37)^2+(51{,}3-45{,}37)^2+\ldots+(38{,}56-45{,}37)^2+(50{,}80-45{,}37)^2+}{19}=101{,}094$$

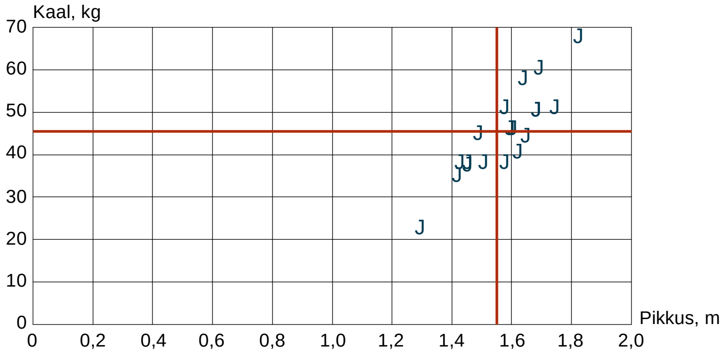
Dispersiooni arvutamiseks on Excelis ja Libre Office'is funktsioon VARP.

Ruutjuurt dispersioonist nimetatakse standardhälbeks (funktsioon STDEVP).

Seose näitajad

Tavaliselt mõõdetakse objekte (indiviide, elemente jms) mitmes mõttes. Näites ülal on iga õpilase kohta andmestikku kantud tunnused sugu, vanus, pikkus ja kaal. Tihti on kaks või enam tunnust mingil määral seotud. Jooniselt on näha, et mida pikem on õpilane, seda suurem kipub olema ka tema kaal. Punane horisontaaljoon näitab kaalu keskmist ja vertikaaljoon pikkuse keskmist.

Õpilaste pikkuse ja kaalu sõltuvus



On välja kujunenud, et tunnustevahelist sõltuvust iseloomustatakse parameetritega **kovariatsioon** ja **korrelatsioonikordaja.** Pikkuse ja kaalu kovariatsioon arvutatakse järgmiselt:

$$\frac{(1{,}75-1{,}58)(51{,}03-45{,}37)+(1{,}44-1{,}58)(38{,}10-45{,}37)+\ldots+(1{,}46-1{,}58)(38{,}56-45{,}37)+(1{,}69-1{,}58)(50{,}80-45{,}37)}{19}=$$

$$=1{,}119$$

(Exceli ja Libre Office'i funktsioon COVAR)

Kui kovariatsiooni väärtus on positiivne, tähendab see seda, et mida suurem on ühe tunnuse väärtus, seda suurem püüab olla ka teise tunnuse väärtus, st sõltuvus on positiivne. Vastupidi, kui kovariatsiooni väärtus on negatiivne, siis on ühe tunnuse suurte väärtuste korral teise tunnuse väärtused pigem väiksemad ehk sõltuvus on negatiivne.

Kovariatsioon näitab küll kahe tunnuse sõltuvuse suunda, kuid mitte suurust. Kui tugev on tunnustevaheline sõltuvus, mõõdab korrelatsioonikordaja:

$$korrelatsioonikordaja = \frac{kovariatsioon}{standardhälve_1 \times standardhälve_2}$$

Korrelatsioonikordaja väärtused asuvad lõigul [–1, 1]. Mida lähemal on korrelatsioonikordaja 1-le, seda tugevam on muutujatevaheline positiivne sõltuvus. Mida lähemal on korrelatsioonikordaja väärtus arvule –1, seda tugevam on muutujate vaheline negatiivne sõltuvus. Nullilähedase korrelatsioonikordaja puhul tunnuste vahel sõltuvus puudub või on väga nõrk.