

## Andmed

Arvulised ja mittearvulised tunnused, mõõtmine, ankeetküsitlus, järjestustunnused, kodeerimine, andmestik

## Põhikool

Statistikas nimetatakse **andmeteks** süstemaatiliselt kogutud ja analüüsimiseks sobivat teavet, kõige sagedamini on need katse- või mõõtmistulemused, mis iseloomustavad teatavaid **tunnuseid**. Mõõdetavad tunnused võivad olla nii **arvulised** kui ka **mittearvulised**. Inimese kasv ja kaal on arvtunnused (kvantitatiivsed tunnused), samuti on matemaatikaülesande lahendamiseks või suuskadel 5 km läbimiseks kulunud aeg arvtunnus. Arvtunnuste väärtused on arvud.

Silmade värvus ja sugu on mittearvulised (kvalitatiivsed) tunnused.

Statistiline uurimus koosneb üldjuhul andmete kogumisest, nende korrastamisest ning analüüsimisest, järelduste tegemisest ja tõlgendamisest.

Andmete kogumisel tuleb täpselt jälgida kokku lepitud reegleid ja hoolitseda selle eest, et andmed registreeritakse täpselt. Igasugune andmete muutmine või mõõtmistulemuste kõrvalejätmine on ausa statistika seisukohast lubamatu, sest muudab uurimise lõpptulemust ning võib põhjustada ekslike otsuste tegemist.

Sageli kogutakse andmeid **ankeetküsitluse** teel, sel juhul tuleb jälgida, et ankeedi struktuur, sõnastus ja vastusevariantide valik (sõnastus) vastaksid reeglitele, mis tagavad tulemuse korrektsuse.

Uuringutes on võimalik kasutada ka varem teistel eesmärkidel kogutud andmeid. Neid võib saada mitmesugustest andmekogudest sh internetist. Niisuguste andmete kasutamisel tuleb täpselt selgitada nende olemus ja kasutatavad eeldused.

Andmete kogumit nimetatakse **andmestikuks**.

## Gümnaasium

Andmete analüüsimisel tuleb jälgida seda, mis tüüpi on analüüsitavad tunnused. Ankeetide ja testide tulemusena saadakse enamasti mittearvulised tunnused, nende väärtused on väljendatud sõnadega. Näiteks mõõdetakse rahulolu: „Kuidas olete rahul selle tunniga?“ ja vastusevariandid on: väga rahul/ pigem rahul/ ei rahul ega rahulolematu/ pigem rahulolematu/ väga rahulolematu. Sellist tunnust nimetatakse **järjestustunnuseks**.

Järjestustunnuse väärtused sageli **kodeeritakse**, st asendatakse arvudega, jälgides, et säiliks vastusevariantide sisuline järjestus. Käesolevas näites on vastusevariandid järjestatud rahulolu kahanemise järgi ja kodeerida võib need nii: väga rahul: 1, pigem rahul: 2, ei rahul ega rahulolematu: 3, pigem rahulolematu: 4, väga rahulolematu: 5. Kodeeritud tunnused on analüüsimiseks sobivamad, kuigi neil pole kõiki arvtunnuste omadusi.

On ka selliseid mittearvulisi tunnuseid, mille vastusevariandid ei ole sisuliselt järjestatavad, näiteks õpilase sünnilinn, ema elukutse või lemmikraamat. Nende analüüsimiseks on vähe häid võimalusi.

### Millele pöörata tähelepanu

Suuremahulised andmed (ingl. *Big Data*) on termin, millega tähistatakse väga suuri ja keerukaid, omavahel seotud andmestikke ja nende süsteeme. Tänapäeval on suurte andmetega töötamine omandanud väga olulise koha uurimistöös.

### Seotud mõisted

valim ja üldkogum, jaotus, empiiriline jaotus.