VERBIS AUT RE

EESTI RIIGIKAITSE AKADEEMIA

Ants Kõverjalg

TEADUSTÖÖ METOODIKA ALUSED

II

1994

EESTI RIIGIKAITSE AKADEEMIA

Ants Kõverjalg

TEADUSTÖÖ METOODIKA ALUSED

II

Tallinn 1994

SISUKORD

4. ANDMETÖÖTLUS

4.1. STATISTILISED MEETODID UURIMISTÖÖS

Kaheldakse, kas humanitaarteadused on üldse teadused, ja seda kolmel põhjusel:

1. Puuduvad kriteeriumid nähtuste objektiivseks mõõtmiseks.
2. Nähtused olenevad väga paljudest faktoritest ning seetõttu on nende põhjusi võimatu objektiivselt kindlaks teha.
3. Nähtustevahelisi seoseid ei ole võimalik täppisteaduses kasutatavate täpsete meetoditega kindlaks määrata. Igas teaduses oleneb aga tulemuste objektiivsus sellest, kui täpselt osatakse nähtusi mõõta, neid analüüsida ja uurimistulemusi läbi töötada.

Peab märkima, et humanitaarteadustes domineeris kuni viimase ajani nähtuste kirjeldamine, mis tihtipeale oli kaunis subjektiivne, raskesti mõõdetav ja kontrollitav. Kvantitatiivseid külgi tundmata ei saa aga küllalt põhjalikult tundma õppida nähtuste kvalitatiivseid külgi.

Ilma täpsete kriteeriumideta, mõõtmisteta ja nähtuste kvantitatiivseid ning kvalitatiivseid külgi igakülgselt tundmata ei ole aga mõeldav küllaldane objektiivsus teaduses. Viimast asjaolu arvestades on ka humanitaarteadustes hakatud otsima teid nähtuste kvantitatiivsete külgede senisest täpsemaks mõõtmiseks ning uurimistulemuste usaldatavuse tagamiseks.

Ekslik on ka arvata, et statistiliste meetoditega lahendatakse korraga kõik teaduste kitsaskohad ning et ainuüksi statistiliste meetodite kasutamine teeb teadusest teaduse. Statistika ei ava veel nähtuste olemust. Ta võib fikseerida küll statistiliselt usaldatavaid erinevusi uuritavate nähtuste vahel, kuid põhjusi, miks üks nähtus teisest erineb, statistika selgitada ei suuda. Seda peab tegema teadlase mõistus, tema eruditsioon ning võime uuritavate nähtuste olemust ja nähtusi tekitavaid põhjusi analüüsida.

Nii näiteks võib eksperimentaator kindlaks teha, et tema kasutatud uus õppemeetod andis teise meetodiga võrreldes usaldatavalt erinevaid tulemusi. Selliselt saadud statistilise arvutuse alusel ei saa ta aga öelda, miks just uus meetod andis paremaid tulemusi. Seda võib öelda ainuüksi õppeprotsessi põhjaliku analüüsi, õpetajate ja õpilaste tegevuse igakülgse jälgimise alusel. Teaduses ei piisa ainult faktide fikseerimisest; nende olemust, omavahelisi seoseid ja muutusi nendes on vaja ka teoreetiliselt põhjendada.

Statistilised meetodid on kvantitatiivsete näitajatega nähtuste uurimiseks küll rakendatavad, kuid ainuüksi nende alusel ei saa teha järeldusi, otsustusi ega hüpoteese.

Statistilisi meetodeid tuleb lugeda uurimistöö spetsiaalseteks meetoditeks. Neid tuleb rakendada nii uurimistöö planeerimisel kui kokkuvõtete tegemisel ning neile tuleb vaadata kui vahenditele, mis võimaldavad uurimistulemusi teaduslikult läbi töötada.

Ehitise kvaliteet ei sõltu üksnes sellest, kui hästi on segatud masinas tsemendimört, vaid ka sellest, missuguse kvaliteediga tsemendist, liivast ja lubjast on mört tehtud ning kui oskuslikult on see asetatud kivide vahele. Ka teadusliku uurimistöö tulemus ei olene üksnes statistika kasutamisest, vaid kogutud lähtematerjalide objektiivsusest ning saadud materjalide teoreetilisest analüüsist.

Seega võib teaduslikku tööd piltlikult võrrelda ehitise rajamisega, kus on vähe kasu, et hulk materjali on ehitusplatsile kokku veetud, kui puuduvad seadmed ja töölistel oskused sellest ehitis püstitada.

Nagu kogemused näitavad, tehakse uurimistöös sageli suuri vigu juba vajaliku lähtematerjali kogumisel. Nähtuste hindamise kriteeriumid on tihti subjektiivsed ning võivad oluliselt mõjustada uurimistöö objektiivsust.

Mis kasu on uurimistöö suurepärastest statistilistest meetoditest ja nähtuste heast teoreetilisest analüüsist, kui kogutud lähteandmeid ei saa lugeda objektiivseteks?

Uurimistöös nähtusi hinnates ei tohiks piirduda vaid ühtede kriteeriumidega. Näiteks tuleks mitme kriteeriumi alusel ja mitme meetodiga kontrollitud õpilaste teadmiste taset omavahel korreleerida ning alles siis, kui tulemuste vahel valitseb usaldatav seos, võib uurimistöö tulemustest teha küllaldase usaldatavusega järeldusi.

Praktika näitab, et kui ei ole küllaldast ettevalmistust, erialast teaduslikku eruditsiooni, oskust leida ja analüüsida nähtuste põhjuslikke seoseid, kindlaks määrata uurimistöös kasutatavate mõistete täpset sisu ja mahtu, viivad statistilised arvutused tihti valedele järeldustele. Sageli juhtub seda humanitaarala inimestega, kes ei süvene küllaldaselt matemaatilis-statistiliste meetodite olemusse ega nende kasutusvõimalustesse. Võime kohata uurijat, kes oma töös püüab iga hinna eest rakendada statistilisi meetodeid, ilma aru andmata, kas neid meetodeid kasutada on otstarbekohane ja üldse vajalik. Ainuüksi asjaolu, et statistika on n-ö moodne ning seda kasutavad paljud uurijad, ei õigusta veel selle ükskõik kus ja millal kasutamist.

Uurimistöös ei tohi muidugi minna ka teise äärmusse - hakata tundma hirmu arvude ja matemaatiliste valemite ees ja loobuda nende kasutamisest seal, kus statistika on uurija hädavajalik tööriist. Tuleb arvestada, et statistilised meetodid võimaldavad uurimistöö tulemusi süstematiseerida, töödelda ja nende teaduslikku usaldatavust kontrollida.

Statistika käsitleb nähtuste kogumeid, mis koosnevad hulgast üksiknähtustest, ning statistiliste meetodite kasutamisel taandatakse terve hulk üksikandmeid üheleainsale, uurija poolt representatiivseks peetud väärtusele.

Statistilisi meetodeid kasutatakse eelkõige selleks, et kokkuvõtvalt esitada või võrrelda teatud nähtuste kogumite tulemusi. Üksiknähtust ennast aga vaadeldakse statistikas ainult kogumi liikmena, mille tunnused (iseloomulikud jooned) on kogumit iseloomustada võimaldavad vahendid.

Üksikuid kogumisse kuuluvaid nähtusi nimetatakse ***kogumi elementideks*** (tähistatakse tavaliselt x_i või y_i). Üksikelementide ***summat*** ($\sum x_i$) ***nimetatakse kogumi mahuks ja tähistatakse tähega N.***

Kogumi elementidel on ***vähemalt üks tunnus***, millel on tavaliselt teatud arvväärtus, mille alusel on võimalik üht kogumi elementi teisest eristada. Statistiliste meetodite peamine ülesanne uurimistöös ongi ***leida, mõõta ja registreerida üksiknähtuste (elementide) tunnused ning saadud tulemuste alusel kogumit kui tervikut kvantitatiivselt iseloomustada.***

Nii näiteks tehakse kontrolltöödega teatud hindamissüsteemi alusel kindlaks klassi iga õpilase (kogumi elemendid) teadmiste tase. Selle alusel leitakse kogu klassi (kogumi) teadmiste taset iseloomustav üks hinne (tavaliselt keskmine hinne, mõnikord ka kõige sagedamini esinev hinne). Viimane iseloomustab kogumit tervikuna, mitte selle üksikuid elemente.

Kogumite kvantitatiivseks iseloomustamiseks kasutatakse peamiselt mitmesuguseid keskmisi näitajaid. Nii nagu teisedki kvantitatiivsed suurused, saadakse ka mõõtmiste ja arvutamiste keskmised tulemused. Keskmiste leidmisel lähtutakse uurimise ülesannetest ja uuritavate nähtuste konkreetsetest iseärasustest. Statistiliste meetodite kasutamisel tuleb tähelepanu juhtida ühele olulisele asjaolule: kogumi elementide väärtused kujutavad tavaliselt juhusuurusi, mille väärtused ei kujune mingi üldise seaduspärasuse resultaadina, vaid paljude juhuslike, üldistele seaduspärasustele mittealluvate tegurite toimel.

Näiteks tahetakse õpilaste teadmiste üle otsustada kontrolltööde hinnete põhjal. Hinded olenevad suurel määral muidugi õpilase teadmistest, kuid küllaltki suur tähtsus on üksikhinnete kujunemisel ka paljudel juhuslikel asjaoludel: õpilase psüühiline seisund, väsimus, mahakirjutamise võimalus jms. Seetõttu ei saa kooli (klassi) õpilase teadmiste taseme üle otsustada mõne üksiku õpilase teadmiste taseme (hinde) alusel. Järeldusi kogu kooli (klassi) õpilaste teadmiste taseme üle võime teha alles siis, kui vaatluse alla on võetud uuritavast koolist piisavalt ***suur hulk õpilasi*** ning analüüsitakse nende ***keskmist hinnet***.

Mingite kogumitega seotud seaduspärasused avalduvad keskmiste näitajate kaudu, kusjuures viimased on suhteliselt püsivamad, võrreldes kogumi elementide individuaalväärtustega. Need võivad olla kaunis muutlikud: ühed õpilased teevad kontrolltöö väga halvasti, teised väga hästi. Kooli keskmine hinne on aga enam-vähem püsiv suurus, mis muutub aastate lõikes suhteliselt vähe.

Keskmised näitajad kujunevad aga usaldatavaks teaduslikuks informatsiooniallikaks ainult siis, kui nende arvutamisel on arvestatud suurte arvude seadust. Selle olemust võib lühidalt kokku võtta järgmiselt:

a) ühtlase koostisega nähtuste kogumi seaduspärasused avalduvad ja neid võib uurida ***küllalt suure arvu andmete*** alusel;

b) neid seaduspärasusi võib kvantitatiivselt väljendada ainult ***keskmiste näitajatega;***

c) keskmistega väljendatud kogumi seaduspärasused on seda täpsemad, mida suurem arv uuritava kogumi elemente uurimisobjektiks võetakse;

d) üksiknähtuste hälbed (kõrvalekaldumised) keskmisest ühele või teisele poole, mida põhjustavad nähtuste suhtes ebaolulised, juhuslikud asjaolud, uurimisobjektide suure arvu korral vastastikku kattuvad.

Eespool toodust selgub, et kogumi üksikute elementide näitajad, üksikute tegevusviiside iseloom jääb statistiliste meetodite kasutamisel uurimisorbiidist välja. Nii näiteks võib statistika vahendusel küll kindlaks teha, kui tihedasti koonduvad õpilaste hinded klassi (kogumit) iseloomustava keskmise hinde ümber, kuid õpilaste (elementide) hinnete olemust, üksikute hinnete individuaalseid iseärasusi statistika ei vaatle.

Kogumi maht võib uurimuste ajal kõikuda mõnest elemendist tuhandeni. Nagu varem märgitud, ei ole statistilise materjali suure mahu puhul võimalik ega ka otstarbekohane uurida kogumi kõikide elementide väärtusi. Statistiliselt usaldusväärseid andmeid saadakse siis, kui uuritakse nn ***väljavõtukogumit mahuga n.*** Selline kogum (populatsioon) moodustatakse ***teatud viisil*** valitud osast kogumi elementidest. Sel juhul otsustatakse kogumiga seotud nähtuste kvantitatiivse külje üle kogumi ***teatud osa*** elementide tunnuste põhjal (vt I osa ptk 2.4).

Küllalt suur väljavõtu maht ***n*** ja teaduslikult õigete printsiipidega toimetatud elementide valik üldkogumist väljavõtukogumisse tagab, et väljavõtu uurimisel saadakse enam-vähem samad tulemused, mis kogumi uurimisel.

Statistilised meetodid võimaldavad ka selgitada, kui suure ***väljavõtukogumi*** alusel võib ***usaldatavalt hinnata kogumit kui tervikut.***

Peale selle võib statistiliste meetoditega kindlaks määrata ka uurimistulemuste usaldatavust, kogumite kvantitatiivsete näitajate erinevuse usaldatavust ja eri kogumite kvantitatiivsete tunnuste vahelist seost.

Statistilise kogumi analüüsimisel on üks olulisi näitajaid peale keskmise ka *kogumi elementide tunnuste homogeensus või heterogeensus, st kogumi elementide tunnuste paiknemine keskväärtuse ümber.*

Varieerumine ongi nähtus, kus kogumi elemente iseloomustava sama tunnuse väärtus on kogumi eri liikmetel erinev, st kogumi eri liikmete väärtused varieeruvad.

Sama tunnuse varieeruvus eri kogumites võib olla küllaltki suurte erinevustega, seejuures kogumite aritmeetilised keskmised võivad olla võrdsed. Võttes aga aluseks ainult keskmised, võib mõnikord teha kogumite kohta ebaõigeid järeldusi. Seetõttu on vaja keskmiste abil uuritavate kogumite paremaks iseloomustamiseks kasutada veel ***variatsiooninäitarve*** *(variatsiooni* ***amplituud****,* keskmine lineaarhälve, dispersioon ja standardhälve), st arve, mis väljendavad kogumi üksikliikmete vahelisi erisusi.

Olulisemat

1. **Teaduslike probleemide statistilisel uurimisel on tavaliselt tegemist kolme tööetapiga:**
   - **Vaatlus ja eksperiment mingite nähtuste kohta andmete hankimiseks.**
   - **Saadud andmete kokkuvõtt - üksikandmetest üldistavate arvandmete saamine ja nende töötlemine.**
   - **Andmete analüüs ja interpreteerimine.**
   - **Nendel tööetappidel kasutatavate uurimismeetodite ja kriteeriumide otstarbekohasus ning kasutatavate meetodite oskuslik seostamine uuritavate nähtuste teoreetilise analüüsiga tagavad uuringute maksimaalse objektiivsuse.**
2. **Statistika ülesandeks on kogumi kui terviku kvantitatiivne iseloomustamine.**
3. **Statistikas uuritakse üldkogumi (populatsiooni) asemel sageli väljavõtukogumit ja selle alusel hinnatakse kogumit kui tervikut.**
5. **Keskmiste abil uuritavate kogumite paremaks iseloomustamiseks kasutatakse variatsiooninäitarve.**
6. **Statistilised meetodid on uurimistöö spetsiaalsed meetodid, mida kasutatakse uurimistöö planeerimisel ja kokkuvõtul.**

4.2. MÕÕTMINE TEADUSES

Igal mõõtmisel võrreldakse tulemust mingi skaala väärtusega (mõõtühikuga). Kui reaalteadustes on mõõdetavat suurust võimalik küllaltki täpselt fikseerida mitmesuguste aparaatide ja seadmetega, siis humanitaarteadustes võib kehtivas hindamissüsteemis tulemust vaid pingeritta seada (edasijõudvad ja mahajäävad õpilased; nõrgad, rahuldavad ja head õpilased). Küll aga ei saa sel juhul väita, mitme skaalajaotuse võrra on üks õpilane teisest nõrgem või tugevam.

Teaduses puututakse kokku peamiselt nelja hindeskaalaga.

Nimi- ehk ***nominaalskaalaga*** on tegemist, kui näiteks igale korvpallimeeskonna liikmele antakse mingi number. Nende numbritega on mõeldamatu teha mingeid statistilisi arvutusi. Ei ole ju sel juhul näiteks mängijal 5 eelistusi mängija 1 ees. Ei saa ka öelda, et üks arvnäitaja on sel juhul teistest suurem, väiksem, parem või halvem. Nimiskaalaga on tegemist ka õpilaste soo (poisid, tüdrukud), rahvuse (venelane, eestlane, ukrainlane), huvide (muusika, kujutav kunst, sport), ka õppeedukuse (edasijõudjate %) puhul. Viimasel juhul liigitatakse õpilased edasijõudvateks ja mitteedasijõudvateks. Kui arvutusi teha korrektselt, võib (muidugi siis, kui lähtehinded on objektiivsed) saada küllaltki usutavat informatsiooni. Tuleb aga arvestada, et sel juhul on mõõtühikuks õpilane, mistõttu arvutamisel tuleb lähtuda õpilast esindavast protsentarvust. Kui klassis on näiteks 33 õpilast, siis üks õpilane moodustab õpilaste üldarvust ligikaudu 3%. Kui edasi ei jõua kolm õpilast, siis saame õppeedukuse protsendiks

$$\frac{3 \cdot 100}{33} \approx 91\%,$$

mitte aga 90,90909...% või 90,9%. Tuleb ka silmas pidada, et arvutuste täpsus ei tohi ületada lähteandmete täpsust. Kui aga koolis on 860 õpilast, siis võib teha arvutusi täpsusega kuni 0,1%, sest üks õpilane moodustab 0,12% õpilaste üldarvust.

Järjestusskaala (astmelise skaala) puhul ei ole intervallid skaalajaotuste vahel ühesugused. Õppetöö tulemused võib reastada kas kasvavas või kahanevas järjestuses, kuid nende intervallid ei ole võrdsed. Koolihinded on astmelised suurused, mis ei ole saadud kvantitatiivsel mõõtmisel, vaid subjektiivsel hindamisel. Seejuures märgitakse kõrgema hindega küll mingite omaduste (teadmiste, oskuste, vilumuste) kõrgemat taset võrreldes madalamaga, kuid ei ole võimalik täpselt näidata, kui palju kõrgem tase madalamast erineb. Seetõttu on statistilised arvutused koolihinnetega (s.o aritmeetiliste keskmiste, varieeruvuse, keskmiste erinevuste usaldatavuse jm leidmine) ka ebakorrektne.

Pikkuse puhul võib kindlalt väita, et 50 cm on täpselt kümme korda pikem kui 5 cm. Ei saa aga väita, et hinde "4" saanud õpilane on oma teadmistelt täpselt kaks korda tugevam hinde "2" saanust. Esimesel juhul on mõõteskaala ***intervallid täpselt ühesugused*** kõikides skaalavahemikes: 1 cm on sama pikk vahemikus 10-11 cm ja 100-101 cm. Hinnetevahemik 1-2, nagu näitab koolipraktika, on hoopis väiksem, kui näiteks 2-3 või 3-4. Hinne "3" on teistest tunduvalt rohkem "välja venitatud". Pealegi võivad koolihinded olla äärmiselt subjektiivsed. Näiteks ühe õpetaja pandud hinne "5" on teisel tihti "3". Seetõttu on mõttetu võrrelda ühe regiooni kooli õppetöö tulemusi teise regiooni omadega hinnete alusel. Kõrgkoolide sisseastumiseksamite tulemustest on selgunud, et mõne kooli medaliga lõpetanu teadmised on madalamad medalita õpilaste omadest.

Tuleb juhtida tähelepanu, et objektiivset väärtust ei ole näiteks klassi keskmisel hindel, mis on väljendatud hinde kümnendike või sajandikena (nt keskmine hinne 3,85). Nii arvutatud keskmise täpsus ületab lähteandmete täpsuse ning sel puudub reaalne mõte, sest vastavat skaalajaotust tegelikult ei esine. Saadud keskmine hinne näitab küll seda, et subjektiivselt väljapandud hinded lähenevad rohkem hindele "4" kui "3", kuid ei saa väita, et keskmine hinne oleks 0,15 skaalajaotust väiksem. Küll näitab klassi õpilaste keskmine pikkus 166 cm seda, et 170 cm pikkune õpilane on keskmisest õpilasest (keda ei pruugi tegelikult klassis üldse olla) tinglikult 4 cm võrra pikem (viimane suurus on aga täpselt mõõdetav).

Intervallskaalal (kvantitatiivsel skaalal) on skaalajaotuse intervallid täpselt ühesugused ning on täpselt teada, mitu korda on üks mõõdetav suurus teisest suurem või väiksem (üks meeter on täpselt sada korda pikem kui üks sentimeeter). Intervallskaalale vastavad näiteks temperatuur Celsiuse ja Fahrenheiti järgi, pikkus, õpilaste teadmiste testimisel saadud õigete vastuste arv.

Intervallskaala erijuht on nn ***suhteskaala***, millel on täpselt fikseeritud skaala nullpunkt (näiteks kasv, kaal, aeg, temperatuur Kelvini järgi).

Intervall- ja ***suhteskaala*** järgi saadud arvnäitajate vahendusel võib leida aritmeetilisi keskmisi, variatsiooni näitarve, keskmiste erinevuse statistilist usaldatavust *F (Fischeri) või t (Studenti) kriteeriumiga ainult sel juhul, kui on tegemist uuritava kogumi elementide väärtuste nn normaaljaotusega.* Sel juhul jaotuvad kogumi elementide väärtused iseloomulikul viisil aritmeetilise keskmise ümber. Normaaljaotusele vastab graafiliselt normaaljaotuskõver (Gaussi-Laplace'i kõver, vt jn 1). Selleks et selgitada, kas saadud empiirilised tulemused vastavad või on ligilähedased normaaljaotusele, tuleb uurijal kasutada alljärgnevaid meetodeid.

1) Arvutada valemiga uuritava kogumi keskmine ja standardhälve (vt lähemalt ptk 4.3.1)

$$\sigma = \frac{(x_i - \bar{x})^2}{N},$$

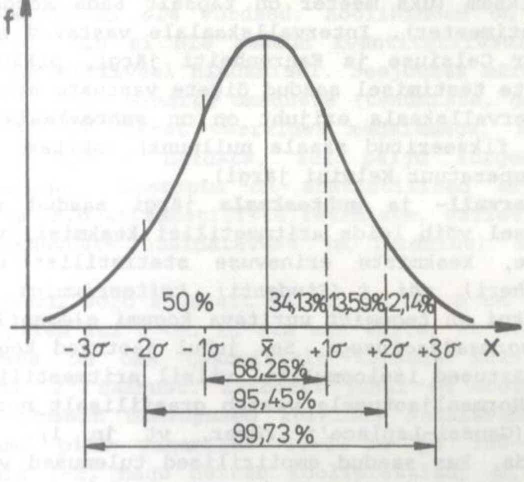
kus σ - standardhälve;

x_i - kogumi elementide individuaalväärtused;

$\bar{x}$ - kogumi aritmeetiline keskmine;

N - kogumi elementide arv (vt ka ptk 4.3 ja 4.4).

Normaaljaotuse puhul on standardhälbe *(σ)* väärtus määratud normaaljaotuskõvera käänupunktidega. Vahemik $\bar{x} \pm \sigma$ hõlmab 68,26%, vahemik $\bar{x} \pm 2\sigma$ 95,45% ning vahemik $\bar{x} \pm 3\sigma$ 99,73% jaotuskõvera alusest pindalast. Graafiliselt võib standardhälvet kujutada normaaljaotuskõvera käänupunkti abstsissina (vt jn 1). Võrreldes samasse teljestikku joonestatud normaaljaotuskõverat ja uurimistöös saadud empiirilist sageduspolügooni, tuleb jõuda selgusele, kas viimane erineb oluliselt normaaljaotuskõverast. Võib võrrelda ka empiirilise ja normaaljaotuskõvera vastavaid pindalasid.



Joonis 1. Normaaljaotuskõver

2) Leida empiirilise sagedusjaotuse ***mood*** (kõige sagedamini esinev väärtus) ja ***mediaan*** (jaotuse keskmine`liige, millest mõlemale poole jääb võrdne arv liikmeid). Kui moodi mediaani ja keskmise väärtused üksteisest palju ei erine, on tegemist normaaljaotusega (vt ptk 4.3).

3) Empiirilise jaotuse vastavust normaaljaotusele võib kontrollida ka χ^2-kriteeriumiga (hiiruutkriteeriumiga).

Alles pärast seda, kui on kontrollitud intervallskaala alusel saadud empiirilisi tulemusi ning on selgitatud jaotuse vastavust normaaljaotusele, on korrektne kasutada nn parameetrilisi hindamiskriteeriume (kogumit iseloomustavaid karakteristikuid: keskmist, variatsiooninäitarve, statistilise olulisuse kriteeriume).

Mitteparameetrilise hindamise kriteeriumi puhul kasutatakse eespool, nimetatud karakteristikute asemel variantide jaotust iseloomustavaid kriteeriume.

Hoolimata sellest, et kirjanduses on püütud näidata ka koolihinnete intervallskaalale üleviimise võimalusi statistiliste arvutuste tegemiseks, ei saa neid seisukohti lugeda põhjendatuiks ega soovitada selliseid võtteid kooli- ja teaduspraktikas.

Põhjendamatult vähe kasutatakse pedagoogikas (teaduslikes uurimistöödes aga peaaegu üldse mitte) intervallskaala erijuhtu - ***alternatiivset (dihhotoomset)*** skaalat ning selle alusel saadud empiiriliste tulemuste töötlemist mitteparameetriliste hindamiskriteeriumide vahendusel (vt ptk 7.2 ja 8).

Olulisemat

1. **Mõõtmine kujutab mingi tulemuse võrdlemist mõõtühikuga (skaala väärtusega).**
2. **Nimiskaalaga mõõdetud tulemustega ei saa teha statistilisi arvutusi.**
3. **Järjestusskaala jaotuste vahed ei ole ühesugused.**
4. **Intervallskaala puhul on intervallid ühesugused ning selle skaala järgi saadud arvnäitajate vahendusel võib teha statistilisi arvutusi.**
5. **Tulemuste erinevuse usaldatavust võib Fischeri ja Studenti kriteeriumiga määrata vaid sel juhul, kui on tegemist uuritava kogumi elementide väärtuste normaaljaotusega.**

4.3. STATISTILISE KOGUMI KESKMISED

Statistiliste kogumite kõige üldistavamaks kirjeldamiseks kasutatakse mitmesuguseid keskmisi näitajaid. Tihtipeale ei ole keskmise mõiste seotud mingi kindla arvuga, vaid kujutab endast suhtlemisel kasutatavat üldistavat mõtlemiskategooriat (nt keskmine õpilane, keskmine õppeedukus jne). Püüdes eespool toodud mõisteid täpsemalt kindlaks määrata, neid võimalikult objektiivsete meetoditega mõõta ja hinnata, võib neile anda ka kindla numbrilise väärtuse.

Tänapäeva teaduse üks olulisi ülesandeid ongi teaduslike nähtuste kvantifitseerimine, st nähtuste kvantitatiivne mõõtmine. Ka teaduses tuleks juhinduda Galileo Galilei kuulsast ütlusest: "Mõõda kõike, mis on mõõdetav, ja tee mõõdetavaks, mis pole veel mõõdetav."

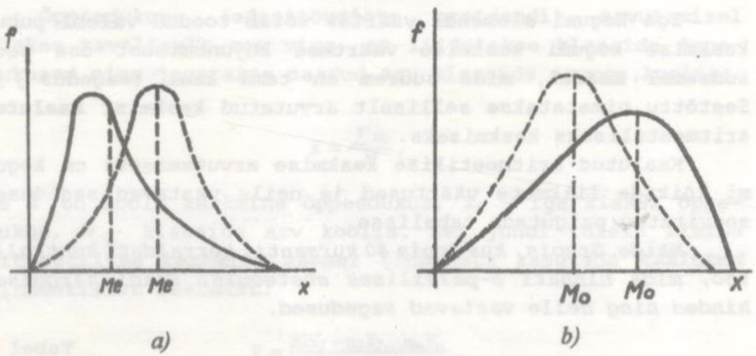
Mõnigi kord kaheldakse, kas humanitaarteadustes üldse ongi võimalik nähtusi mõõta ja neid arvuliselt kindlaks määrata, kas neid ongi võimalik kvantifitseerida. Nagu eespool kirjeldatust selgub, on paljude teaduslike nähtuste puhul võimalik loendamise teel kindlaks määrata sündmuste sagedust. See aga ongi kvantifitseerimise üks primitiivsemaid vorme. Kvantifitseerimise järgmine aste on kogumi mitmesuguste keskmiste (keskväärtuste) kindlaksmääramine. Kuna keskmised kujutavad endast kvantitatiivseid, st mõõtmise ja arvutamisega saadud suurusi ning nende saamisel võib lähtuda eri vaatekohtadest, võib kogumil olla ka ***palju keskmisi***. Praktikas kasutatavad keskmised jaotatakse kahte rühma: ***mahukeskmised*** ja ***asendikeskmised***.

Mahukeskmised on sellised keskmised, mille arvuline väärtus muutub siis, kui muutub kogumi ükskõik millise liikme väärtus. Kui näiteks läheb klassist ära üks lühikesekasvuline õpilane ja tema asemele tuleb üks pikakasvuline, muutub kohe klassi õpilaste keskmine pikkus.

Mahukeskmisena kasutatakse pedagoogilistes uurimistöödes peamiselt aritmeetilist keskmist (kogumi üksikute elementide arvuliste väärtuste summa $\sum x_i$ ja kogumi mahu N jagatis), vähem harmoonilist, geomeetrilist, ruut- ja kronoloogilist keskmist.

Asendi- ehk struktuurikeskmised muutuvad siis, kui toimuvad nihked kogumi struktuuris (muutub nende arv, järjekord jne).

Asendikeskmistena kasutatakse peamiselt **mediaani** (M_e) (korrastatud sagedusjaotuse keskmine liige, millest mõlemale poole jääb võrdne arv liikmeid) ja **moodi** (M_o) (statistilises sagedusjaotuses kõige sagedamini korduv väärtus). Vähem kasutatakse ***kvartiile*** (jagavad sagedusjaotuse neljaks osaks, milles igas on võrdne arv rea liikmeid) ja ***detsiile*** (jagavad statistilise rea kümneks võrdseks osaks). Mediaani ja moodi graafiline tõlgendus on toodud joonisel 2 (*a* ja *b*).



Joonis 2. Moodi ja mediaani graafiline tõlgendus

Kvartiile on kolm: Q_1, Q_2, Q_3.

Mediaani ja kvartiili on lihtne kindlaks määrata ka kumulatiivse sageduse protsendikõvera abil (vt jn 8, ptk 10). Seal lõikab ordinaattelje 50% tähistava jaotuse kohalt tõmmatud horisontaalsirge (50% liin) kõverat. Kõvera ja liini lõikepunktist abstsissteljele tõmmatud vertikaalsirge lõikab viimast teatud jaotuse kohal, mille arvväärtus ongi mediaan.

Kolme kvartiili arvväärtus saadakse 25%, 50% ja 75% liini lõikumisel kõveraga ja lõikepunktidest *x*-teljele vertikaalsirgete tõmbamisega. Nagu näha, langeb keskmise kvartiili väärtus kokku mediaani väärtusega.

Keskmiste tähtsus mingi kvantitatiivse nähtuse tunnetamisel põhineb suurte arvude seadusel (vt ptk 4.2) ja keskmised kujunevad usaldatavaks informatsiooniallikaks vaid siis, kui nende arvutamisel on seda seadust arvestatud.

4.3.1. Aritmeetilise keskmise arvutamine

Aritmeetiline keskmine $\bar{x}$ arvutatakse valemiga

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N},$$

kus x_i - kogumi üksikute elementide väärtus;
f_i - üksiku väärtuse esinemise sagedus;
N - kogumi liikmete arv (kogumi maht);
$\sum f_i = N$.

Iga kogumi elemendi väärtus võtab toodud valemi puhul keskmise kogumi keskmise väärtuse kujunemisest osa seda suuremal määral, mida suurem on tema kaal (sagedus f). Seetõttu nimetatakse selliselt arvutatud keskmist ***kaalutud*** aritmeetiliseks keskmiseks.

Kaalutud aritmeetilise keskmise arvutamiseks on kogumi kõikide liikmete väärtused ja neile vastavad sagedused soovitatav paigutada tabelisse.

Näide. *Grupis, kus õppis 40 kursanti, korraldati kontrolltöö, mida hinnati 5-pallilises süsteemis. Saadi järgmised hinded ning neile vastavad sagedused.*

Tabel 1

Keskmise arvutamine

Hinne x_i	Hinde esinemise sagedus f_i	$f_i x_i$
1	5	5
2	10	20
3	7	21
4	8	32
5	10	50
	$\sum f_i = 40$	$\sum f_i x_i = 128$

Aritmeetilise keskmise arvutamiseks tuleb kõigepealt reastada hinded kasvavas järjekorras. Siis märgitakse tabelisse iga hinde esinemise sagedus. Sageduste summa annab kogumi liikmete üldmahu N. Seejärel korrutatakse iga hinde väärtus selle esinemissagedusega. Lõpuks leitakse tabelis saadud korrutiste summa($\sum f_i x_i$). Selle jagamisel kogumi mahuga N saadaksegi aritmeetiline keskmine.

Meie näite puhul

$$\bar{x} = \frac{128}{40} = 3,2.$$

Juhul kui sagedusjaotuses esineb iga variant ainult üks kord, st $f_1 = f_2 = ... f_n = 1$, omandab valem kuju

$$\bar{x} = \frac{\sum x_i}{N}.$$

Seda nimetatakse ***lihtsaks*** aritmeetiliseks keskmiseks.

Õppeedukuse (edasijõudjate protsendi) arvutamisel tehakse tavaliselt see viga, et liidetakse klasside õppeedukused ning jagatakse saadud arv klasside arvuga koolis:

$$\bar{x} = \frac{\sum x_i}{N},$$

kus $\bar{x}$ on kooli keskmine õppeedukus, x_i - iga klassi õppeedukus, N - klasside arv koolis. Sel juhul tuleks ***lihtsa aritmeetilise keskmise*** asemel tingimata kasutada ***kaalutud aritmeetilist keskmist:***

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 ... n_m\bar{x}_m}{n_1 + n_2 + ... n_m},$$

kus n - on õpilaste arv klassis ja $\bar{x}_i$ iga klassi õppeedukus.

Näide. *Oletame, et tahetakse arvutada 6.-9. grupi õpilaste keskmine õppeedukus protsentides järgmiste näitajate alusel.*

Tabel 2

Õppeedukuse arvutamine

Grupp	Õpilaste arv	Mitteedasijõudjad	Õppeedukus
6	40	5	88
7	40	4	90
8	15	5	67
9	10	4	60

Keskmine õppeedukus arvutatakse tavaliselt järgmiselt:

$$\bar{x} = \frac{88+90+67+60}{4} = \frac{305}{4} = 76\%.$$

Selle arvutuse järgi peaks 6.-9. grupis 76%, s.o 80 õpilast 105-st edukalt õppima. Tegelikult aga õpib edukalt 87 õpilast (105-18=87). Nagu näha, võib väära arvutusega kirjutada koolile juurde tervelt 7 mitteedasijõudvat õpilast. Õige arvutus peaks olema järgmine:

$$\bar{x} = \frac{40\cdot 88+40\cdot 90+15\cdot 67+10\cdot 60}{40+40+15+10} = 83\%.$$

Sama tulemuse saab ka siis, kui 6.-9. grupis edasijõudvate õpilaste arv jagada kogu õpilaste arvuga:

$$\bar{x} = \frac{87\ 100}{105} = 83\%.$$

Vale arvutuse puhul eksiti õppeasutuse kahjuks 7%. Eksimus on eriti suur, kui keskmise õppeedukuse või mõne muu õppetööd iseloomustava näitaja keskmise arvutamisel on gruppides (koolides) õpilaste arv väga erinev.

4.3.2. Keskmiste praktilisest kasutamisest

Keskmisi rakendatakse ulatuslikult uurimistöö andmete iseloomustamiseks. Kuid tuleb ka arvestada, et keskmiste tunnetusvõimalused on suhteliselt piiratud. ***Aritmeetiline keskmine*** võimaldab:

a) iseloomustada uuritavat kogumit ainult ühe arvu kaudu;
b) võrrelda üksikväärtuste suurust aritmeetilise keskmisega;
c) määrata mingi nähtuse arengutendentsi;
d) võrrelda eri kogumeid;
e) arvutada teisi statistilisi näitajaid (paljud statistilised arvutused tuginevad aritmeetilisele keskmisele).

Tuleb aga kohe rõhutada, et aritmeetiline keskmine ei iseloomusta kogumit igakülgselt. Ta iseloomustab ainult mingi kogumi (nähtuse) üldist nivood, mille ümber elementide väärtused kõiguvad. Elementide individuaalseid erinevusi, kogumi homogeensust või heterogeensust aritmeetiline keskmine ei iseloomusta.

Tähelepanu tuleb juhtida ka sellele, et aritmeetiline keskmine ei sõltu mitte ainult elementide individuaalväärtustest, vaid ka nende suhtelisest esinemissagedusest. Nii näiteks muutub grupi keskmine hinne kohe, kui suureneb või väheneb puudulike hinnete arv.

Aritmeetiline keskmine sobib kogumi kirjeldamiseks siis, kui mõõtarvud on jaotunud sümmeetriliselt keskpunkti ümber (vt jn 1). Kui on tegemist kogumi mitmetipulise jaotumisega, siis aritmeetiline keskmine kogumi kirjeldamiseks ei sobi. Sellisel juhul on parem kasutada moodi.

Mediaani kasutatakse siis, kui soovitakse määrata kindlaks jaotuse täpset keskpunkti. Mediaani tunnetuslikuks eeliseks on see, et mõned eriti suure sagedusega intervallid võivad aritmeetilist keskmist oluliselt mõjutada. Mediaani sellised ekstreemsed intervallid ei sega.

Aritmeetiliseks keskmiseks võib olla selline tunnuse väärtus, mida reas ei esinegi, mida pole tegelikult üldse olemaski (vt nt tabel 1). Mediaan on aga alati üks jaotusrea liikmeist või sellele väga lähedane suurus (nt paarisarvulistes ridades). Graafiliselt kujutab mediaan sagedusjaotuse kõvera alust pinda poolitava vertikaaljoone abstsissi. Sümmeetrilise sagedusjaotuse puhul on aritmeetilise keskmise ja mediaani väärtused võrdsed (vt jn 1).

Mediaani alusel, veelgi täpsemalt aga kvartiilide alusel, mis põhinevad sagedusjaotuse rea ositamisel (mitte rühmitamisel), st rea jaotamisel teatud osadeks, võib mõnikord anda üsna hea iseloomustuse rea struktuurile. Nendele tuginedes võib väita, kas rea liikmed jaotuvad keskmise ümber ühtlaselt, kas väärtuste kuhjumine toimub madalamatesse või kõrgematesse intervallidesse jne.

Moodi kasutatakse siis, kui soovitakse kiiresti iseloomustada kogumit temas kõige sagedamini esineva nähtuse alusel. Seejuures peetakse silmas kõige üldisemat väärtust, nt tavalistes tingimustes kõige sagedamini esinevat hinnet, teatud vanusega laste kõige sagedamini esinevat riietusesemete numbrit. Lastejalatsite valmistajad võtavad näiteks aluseks moodi (lastel kõige sagedamini esineva jalatsinumbri), mitte laste jalanumbrite aritmeetilise keskmise.

Valemist

$$M_0 = \bar{x} - 3(\bar{x} - M_e)$$

selgub, et mõõdukalt ebasümmeetriliste ridade puhul paikneb mediaan aritmeetilise keskmise ja moodi vahel, olles seejuures aritmeetilisele keskmisele lähemal. Täiesti sümmeetrilise sagedusjaotuse rea korral langevad aga kõik kolm ühte:

$$M_0 = M_e = \bar{x}.$$

Kokkuvõtteks võib märkida, et uurides sagedusjaotuste puhul aritmeetilise keskmise, mediaani ja moodi omavahelisi väärtusi, võib teha huvitavaid järeldusi sagedusjaotuste struktuuri kohta. Ühe sagedusjaotuse keskmiste võrdlemine teise sagedusjaotuse keskmistega võimaldab aga võrrelda nende jaotuste eripärasusi.

Olulisemat

1. **Statistilist kogumit iseloomustavad mahu- ja asendikeskmised.**
2. **Arvutustes tuleb rangelt eristada lihtsaid ja kaalutud aritmeetilisi keskmisi.**
3. **Keskmiste kasutamisel tuleb silmas pidada üksiktulemiste normaaljaotust.**

4.4. VARIATSIOONINÄITARVUD

Nagu eelnevast (vt ptk 4.3) selgus, on statistilise kogumi analüüsil üks olulisi näitajaid (peale keskmise) ka kogumi elementide tunnuste homogeensus või heterogeensus, st kogumi elementide tunnuste paiknemine keskväärtuse ümber.

Varieerumine ongi nähtus, kus kogumi elemente iseloomustava sama tunnuse väärtus on kogumi eri liikmetel erinev, st kogumi eri liikmete väärtused varieeruvad.

Eelmises osas toodud näitest selgub veel, et sama tunnuse varieeruvus eri kogumites võib olla üsna suurte erisustega, kogumite aritmeetilised keskmised võivad aga seejuures olla võrdsed. Võttes aga aluseks ainult keskmised, võib mõnikord teha kogumite kohta ebaõigeid järeldusi. Seetõttu on keskmiste abil uuritavate kogumite paremaks iseloomustamiseks vaja kasutada veel variatsiooninäitarve, st arve, mis väljendavad kogumi üksikliikmete vahelisi erisusi.

4.4.1. Variatsiooniamplituud

Kogumi kõige suurema ja kõige väiksema liikme väärtuste vahet nimetatakse variatsiooniamplituudiks R (ka hajuvuseks ehk ulatuseks).

$$R = x_{\max} - x_{\min}.$$

Kui kogum on täiesti homogeenne, st kogumi kõikide elementide väärtused on võrdsed, siis $R = 0$.

Variatsiooniamplituudi on lihtne arvutada, kuid ta ei väljenda mitte kogumi kõikide elementide, vaid kõige äärmiste elementide vahelisi erisusi.

4.4.2. Keskmine lineaarhälve (keskmine hälve)

Kogumi kõikide liikmete vahelisi erisusi võib elementaarselt kindlaks teha toodud keskmise lineaarhälbe vahendusel. Viimane arvutatakse valemiga

$$\Delta x = \frac{\Sigma f|x_i - \bar{x}|}{N}.$$

Absoluutväärtust kasutatakse valemis seetõttu, et hävitada hälvete märke. Kui kasutada hälvete naturaalväärtusi, oleks Δx mis tahes rea puhul null, sest vastavalt aritmeetilise keskmise põhiomadusele

$$\Sigma f(x_i - \bar{x}) = 0.$$

4.4.3. Dispersioon ja standardhälve

Uurimistöö praktikas kasutatakse kogumi elementide varieeruvuse iseloomustamiseks kõige sagedamini ***dispersiooni*** ja ***standardhälvet***. Need näitarvud iseloomustavad kogumi iga üksiku elemendi hälbeid aritmeetilise keskmise suhtes. Dispersioon leitakse põhimõtteliselt samuti, nagu keskmine lineaarhälvegi, ainult elementide individuaalväärtuste ja kogumi aritmeetilise keskmise vaheliste hälvete $x_i - \bar{x}$ märkide kaotamiseks tõstetakse need väärtused ruutu.

Dispersioon arvutatakse valemiga

$$\sigma^2 = \frac{\Sigma f(x_i - \bar{x})^2}{N}.$$

Dispersiooni ruutjuur annab standardhälbe ehk keskmise ruuthälbe (σ).

Seega:

$$\sigma = \sqrt{\sigma^2} \qquad \text{ehk} \qquad \sigma = \sqrt{\frac{\Sigma f(x_i - \bar{x})^2}{N}}.$$

Märkus. Väljavõtu puhul kasutatakse varieerumise iseloomustamiseks ***standardhälvet***, mis avaldub valemiga

$$s=\sqrt{\frac{\sum f(x_i-\bar{x})^2}{n-1}}.$$

Kui $n>20$, ei tehta vahet keskmise ruuthälbe ja standardhälbe vahel ning need arvutatakse valemiga

$$\sigma=\sqrt{\frac{\sum f(x_i-\bar{x})^2}{N}}.$$

Dispersiooni ja standardhälbe arvutamiseks kantakse andmed jällegi tabelisse, nagu lineaarhälbe puhul.

Tabel 3

Dispersiooni ja standardhälbe arvutamine

Hinne	Hinde sagedus f	fx_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$f(x_i-\bar{x})^2$
1	5	5	-2,2	4,84	24,20
2	10	20	-1,2	1,44	14,40
3	7	21	-0,2	0,04	0,28
4	8	32	+0,8	0,64	5,12
5	10	50	+1,8	3,24	32,40
	N=40	$\sum fx_i$ =128			$\sum$ =76,40

$$\sigma^2=\frac{76,40}{40}=1,91,$$

$$\sigma=\sqrt{1,91}\approx 1,4.$$

Seega tuleb dispersiooni arvutamiseks koostada kõigepealt kogumi elementide sagedusjaotus, arvutada keskmine, leida diferents kogumi üksiku elemendi ja keskmise vahel $(x_i-\bar{x})$, võtta iga üksik vahe ruutu $(x_i-\bar{x})^2$, korrutada see iga elemendi esinemise sagedusega f, leida korrutiste summa ning jagada see kogumi liikmete arvuga N.

Tähelepanu tuleb juhtida sellele, et dispersioon on seotud tõenäosusteooriaga, mistõttu teda saab edukalt kasutada uurimistulemuste statistilise usaldatavuse arvutamisel. Lineaarhälvet ja variatsiooniamplituudi selleks otstarbeks kasutada ei saa. Standardhälve on normaalse sagedusjaotuskõvera (Gaussi-Laplace'i kõvera) käänupunkti abstsissi väärtus.

Väikese kogumi korral, mille liikmed on ritta korrastamata, võib dispersiooni ja standardhälbe arvutada kiiresti valemiga

$$\sigma = \sqrt{\frac{N \sum x_i^2 - (\sum x_i)^2}{N^2}}\,.$$

Olgu x_i väärtused järgmised:

Tabel 4

Erijuht

x_i	x_i^2
3	9
3	9
5	25
6	36
8	64
$\sum = 25$	$\sum = 143$

$$\sigma = \sqrt{\frac{5 \cdot 143 - 25^2}{5^2}} \approx 1{,}8\,.$$

4.4.4. Suhtelised variatsiooninäitarvud

Variatsiooniamplituud, keskmine lineaarhälve ja standardhälve avalduvad samades mõõtühikutes, milles on avaldatud kogumi üksikute elementide väärtused. Kui aga tahetakse võrrelda kahe jaotuse erinevates süsteemides mõõdetud tunnuseid (nt hindamine on ühel juhul toimunud viie, teisel juhul saja palli süsteemis), tuleb kasutada relatiivseid (suhtelisi) näitajaid.

Lähtudes variatsiooniamplituudist saadakse suhteline variatsiooninäitarv V_R, mis avaldub valemiga

$$V_R = \frac{R}{\bar{x}}.$$

Lähtudes keskmisest lineaarhälbest, saadakse suhteline variatsiooninäitarv $V_{\Delta x}$. See arvutatakse valemist

$$V_{\Delta x} = \frac{\Delta x}{\bar{x}}.$$

Praktikas kasutatakse aga kõige rohkem standardhälbest lähtudes saadud suhtelist variatsiooninäitarvu - variatsioonikoefitsienti

$$V = \frac{\sigma}{\bar{x}}.$$

Suhe väljendatakse tavaliselt protsentides, mispuhul saadakse valem

$$V = \frac{\sigma \cdot 100}{\bar{x}}\%.$$

Variatsioonikoefitsient näitab protsentides, missuguse osa moodustab standardhälve σ tulemuste keskmisest.

Eespool toodud valemites dimensioonid taanduvad (samanimeliste suuruste suhe on nimeta suurus), mis lubab võrrelda ka selliseid uurimistulemusi, kus mõõteskaalade ulatus, tunnuste ja väljavõttude maht on erinev.

Näiteks. *Mis varieerub rohkem, kas õpilaste õppeedukus* (x_1) *või nende kontrolltööde* (x_2) *tulemused, mis hinnati kahekümne palli süsteemis?*

Andmed:

$$\bar{x}_1 = 3{,}6 \qquad \bar{x}_2 = 12{,}3$$

$$\sigma_1 = 1{,}4 \qquad \sigma_2 = 3{,}8$$

$$V_1 = \frac{140}{3{,}6} \approx 39\% \qquad V_2 = \frac{380}{12{,}3} \approx 31\%$$

Nagu eespool toodust selgub, on õpilaste õppeedukuse suhteline varieeruvus suurem, võrreldes kontrolltööde tulemuste suhtelise varieeruvusega. Seejuures tuleb aga märkida, et kontrolltööde tulemused olid suhteliselt kesised: õpilased suutsid vastata veidi rohkem kui pooltele küsimustele.

4.4.5. Variatsiooninäitarvude kasutamisest

Praktikas on kujunenud tavaks, et variatsiooninäitarve kasutatakse järgmiselt:

a) *Variatsiooniamplituudi* kasutatakse siis, kui uurijat huvitab varieeruvuse kogu ulatus või varieeruvuse ekstreemväärtused, kui andmeid on napilt teiste näitajate arvutamiseks.

b) *Lineaarhälvet* kasutatakse, kui soovitakse arvutada kõiki hälbeid keskmisest, kui ekstreemsed väärtused liialt mõjutavad dispersiooni.

c) *Dispersiooni* kasutatakse, kui edaspidi arvutatakse ka uurimistulemuste statistilist usaldatavust, kui tahetakse leida kõige stabiilsemaid andmeid.

d) *Variatsioonikoefitsienti* kasutatakse, kui võrreldakse kahe jaotuse erinevates ühikutes mõõdetud tunnuseid.

Suhteliste variatsiooninäitarvudega hindamisel tuleb olla ettevaatlik, sest variatsioonikoefitsient oleneb mitte ainult varieeruvusest, vaid ka keskmise väärtusest. Kui näiteks kontrolltöö raskusastet või mõõtühikuid muudetakse, ei pruugi muutuda variatsioonikoefitsient, küll ka keskmine.

Variatsioonikoefitsient võimaldab hästi võrrelda ühe ja sama testi tulemusi ühes või teises grupis. Kahte erinevat testi võrreldes tuleb aga olla ettevaatlik, sest nende raskusaste võis olla erinev.

Näiteks. *Oletame, et tehti 20 küsimusega kontrolltöö ja saadi järgmised tulemused:*

$$\bar{x} = 12 \quad \text{ja} \quad \sigma = 3{,}6, \qquad V = \frac{360}{12} = 30\%.$$

Oletame, et kontrolltööle lisati viis väga lihtsat küsimust, millele kõik õpilased suutsid õigesti vastata. Järelikult on tulemused järgmised:

$$\bar{x} = 17 \text{ ja } \sigma = 3{,}6, \qquad V = \frac{360}{17} = 21\%.$$

Keskmise muutumisega muutus tunduvalt ka variatsioonikoefitsient.

Olulisemat

1. **Peale keskmise on statistilise kogumi ühed olulisemad näitajad kogumi elementide homogeensus ja heterogeensus.**
2. **Kogumi varieeruvust iseloomustavad: variatsiooniamplituud, keskmine lineaarhälve, dispersioon ja standardhälve.**
3. **Kahe jaotuse eri süsteemides mõõdetud tunnuseid tuleb võrrelda relatiivsete (suhteliste) näitajatega.**
4. **Kõige enam kasutatakse statistikas dispersiooni ja standardhälvet, sest nende baasil on võimalik arvestada kogumite keskmiste erinevuste statistilist usaldatavust.**
5. **Suhtelisi variatsiooninäitarve hinnates peab olema ettevaatlik, sest variatsioonikoefitsient oleneb mitte üksnes varieeruvusest, vaid ka keskmise väärtusest.**

5. UURIMISTÖÖDE TÄPSUSEST

5.1. REPRESENTATIIVNE VÄLJAVÕTUKOGUM

Uurimistöös on mõeldamatu ja ka ebaotstarbekas uurida näiteks kogu uuritavat üldkogumit. Statistiliselt küllaltki usaldusväärseid järeldusi saame teha sobivat väljavõtukogumit (valimit) uurides. Selleks et väljavõtukogum iseloomustaks üldkogumit (kogu uuritavat objekti) igakülgselt, peab ta olema representatiivne ehk esinduslik. Teda peavad iseloomustama väiksemas ulatuses kõik üldkogumile kui tervikule omased tunnused.

Uurimustes tuleb arvestada, et mõõtmise ja arvutamise teel saadud kvantitatiivsed näitajad ei ole kunagi absoluutselt täpsed. Vigade põhjusi on mitmeid.

1) Ebaõigest uurimismetoodikast on tingitud nn ***metoodilised vead*** (nt koolide õppetöö võrdlemine subjektiivsete hinnete järgi arvutatud keskmise õppeedukuse alusel).

2) ***Ebakorrektsused***, mis tulenevad mõõtmisvahendite (mõõteriistade) ja arvutuste ebatäpsusest (nt ebatäpsed mõõteriistad katseisikute reageerimiskiiruse määramiseks ning arvutusvead keskmiste ja variatsiooninäitarvude leidmisel).

3) Väljavõtukogumi põhjal tehtavad ebaõiged järeldused üldkogumi kohta - ***representatiivsusvead***.

Teoreetiliselt on võimalik teha ükskõik kui suurt representatiivsusviga, kuid selle tegemise tõenäosus püütakse muuta võimalikult väikeseks. Pedagoogiliste uurin-gute statistilise veahinnangu puhul leitaksegi piirid, millest viga võib suurem olla vaid teatud tõenäosusega p. Praktikas võetakse tavaliselt $p=5\%$ (0,05) ja $p=1\%$ (0,01). Sel juhul oletatakse, et viga ei ulatu väljapoole neid piire ja räägitakse vea maksimaalsest suurusest 95% või 99% tõenäosustasemel. Esimesel juhul võib viga esineda kuni viiel korral sajast, teisel juhul kuni ühel korral sajast.

Kui on teada mingi väljavõtukogumi aritmeetiline keskmine, dispersioon või standardhälve ning väljavõtukogumi maht, siis võib öelda, et see kogum on statistiliselt kirjeldatud. Kui on tegemist väikese väljavõtukogumiga, mille elementide väärtused on üsna varieeruvad, tekib küsimus, kas saadud keskmine esindab küllalt usaldatavat üldkogumit ja millised võivad olla piirid, mille vahel keskmine väärtus peaks teatava tõenäosusega asuma. Kahjuks tehakse meil sageli väikeste ja mitterepresentatiivsete väljavõtukogumitega uuringute põhjal järeldusi mõne uuritava probleemi kohta tervikuna ning juurutatakse saadud tulemused kohe ka tegelikku praktikasse. Nii näiteks on suurel määral seda teed mindud ka õppekirjandusega eksperimenteerides. Tulemuseks on liiga rasked ja õpilastele ebahuvitavad õpikud.

Kas väljavõtuga saadud andmete põhjal on üldse võimalik arvutada üldkogumi vastavaid väärtusi?

Seda on võimalik teha ainult teatud usalduspiirides. Väljavõtu aritmeetilise keskmise usalduspiirkonna määramiseks kasutatakse tavaliselt kaś aritmeetilise keskmise standardvea ($m_{\bar{x}}$) või selle tõenäose vea (*PE*) mõistet. Valemid on erinevad ning olenevad väljavõtu- ja üldkogumi suhetest, samuti väljavõtukogumi mahust (n).

Aritmeetilise keskmise standardviga ($m_{\bar{x}}$) avaldub väikese väljavõtukogumi korral (kui väljavõtus on vähem kui 30% üldkogumi liikmete üldarvust) valemiga

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

kus σ on väljavõtukogumi standardhälve,

n - väljavõtukogumi maht.

Kui $n<20$, siis võetakse valemisse väärtus (n-1).

Kui väljavõtukogumisse võetakse vähemalt 30% üldkogumi mahust, kasutatakse aritmeetilise keskmise standardvea arvutamiseks valemit

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}},$$

kus N on üldkogumi maht.

Kui standardvea arvutamisel tahetakse arvestada vea tõenäosustaset, tuleb saadud tulemust korrutada konstandiga k, mille mõned väärtused on toodud tabelis 5.

Tabel 5

Tõenäosustaseme koefitsiendid

Tõenäosuse tase	Konstant k
68%	1,00
95%	1,96
99%	2,58

Väljavõtukogumi keskmine esitatakse tavaliselt kujul:

$$\bar{x}_n \pm m_x \cdot k .$$

Praktikas kasutatakse mõnikord ka aritmeetilise keskmise tõenäose vea PE mõistet. $PE = 0{,}6745 \cdot m_{\bar{x}}$ ja väljavõtukogumi keskmine ($\bar{x}_n$) esitatakse kujul $\bar{x}_n \pm PE$.

Statistikas tuleb sageli kasutada ka aritmeetilise keskmise suhtelise (relatiivse) vea mõistet.

See avaldub valemiga

$$\mu = \frac{m_{\bar{x}}}{\bar{x}_n} \cdot 100\% .$$

Näide: *Taheti uurida 2000 õpilase teadmisi füüsikas. Selleks korraldati test 400 juhuslikult valitud õpilasega. Testis esitati 5 küsimust, mis nõudsid üheseid vastuseid. Keskmiseks tulemuseks saadi 3,8 õiget vastust ja standardhälbeks 0,6. Tuleb leida aritmeetilise keskmise võimalik standardviga 95%-lise tõenäosusega.*

Kuna väljavõtukogumi maht on 20% kogumi üldmahust, kasutatakse valemit

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}6}{\sqrt{400}} = \frac{0{,}6}{20} = 0{,}03 ; \quad m_{\bar{x}} = 1{,}96 \cdot 0{,}03 \approx 0{,}06 .$$

(1,96 on konstant k 95%-lise tõenäosustaseme juures.)

Seega 95%-sel tõenäosustasemel

$$\bar{x}_n = 3{,}8 \pm 0{,}06 , \quad \text{st} \quad 3{,}74 < x < 3{,}86 .$$

Selleks et selgitada, kas viga on suur või väike, leitakse suhteline viga:

$$\mu = \frac{0{,}06}{3{,}8} \cdot 100 = 1{,}6\%.$$

Viga on küllaltki väike ja võib väita, et õpilased andsid keskmiselt 3,8 õiget vastust. Õigete vastuste sellist arvu tegelikult ei esine, kuid võib järeldada, et suurem osa õpilasi andsid testi küsimustele 4 või 5 õiget vastust.

Arvutusest selgub, et kui uurida veel samast üldkogumist mingi teise 400-liikmelise väljavõtukogumi samasuguse testi tulemusi, siis 95%-l juhtudest langeb ka selle väljavõtukogumi keskmine tulemus vahemikku $3{,}8 \pm 0{,}06$.

Ülaltoodust selgub, et tehes mingi väljavõtukogumi alusel üldkogumi kohta järeldusi, eriti aga mitme väljavõtukogumi tulemusi võrreldes, tuleb tingimata selgitada, kas tulemuste usalduspiirkonnad (-intervallid) kattuvad või mitte, milline on suhteline viga (viga alla 2% loetakse veel väikeseks veaks) ning kas väljavõtukogumi näitajad erinevad oluliselt üldkogumi vastavatest näitajatest. Üks tüüpilisi ebakorrektsusi nähtuste hindamisel ongi see, et võrreldakse ainult keskmisi näitajaid ega arvestata nende statistilisi usalduspiirkondi.

Kui kehtib normaaljaotus, on väljavõtukogumi põhjal võimalik usaldatavalt kindlaks määrata ka üldkogumi väärtusi, st kogu uuritava populatsiooni parameetreid.

Selleks kehtib valem

$$\overline{X}_N = \bar{x}_n \pm z \cdot m_{\bar{x}},$$

kus $\overline{X}_N$ - üldkogumi keskmise tõenäone esinemispiirkond (usalduspiirkond);

$\bar{x}_n$ - väljavõtukogumi keskmine;

z - väärtus tabelis, mis võetakse usalduspiirkonna piiriks (95%-lise tõenäosuse korral on see 1,96);

$m_{\bar{x}}$ - väljavõtukogumi aritmeetilise keskmise standardviga.

Eeltoodud näite põhjal saame üldkogumi keskmise tõenäoseks esinemispiirkonnaks:

$$\overline{X}_N = 3{,}8 \pm 1{,}96 \cdot 0{,}03 = 3{,}8 \pm 0{,}06.$$

Seega 95%-lise tõenäosuse puhul on üldkogumi keskmine

$$\overline{X}_N = 3{,}8 \pm 0{,}06 \text{, st } 3{,}74 < \overline{X}_N < 3{,}86 .$$

Toodud näidetest selgub, et üldkogumi keskmise usaldus-piirkonnad on samad väljavõtukogumi keskmise usaldus-piirkondadega ning väljavõtukogumi uurimiste tulemuste alusel võib teha ka usaldusväärseid järeldusi üldkogumi kohta.

5.2. VÄLJAVÕTUKOGUMI MAHU KINDLAKSMÄÄRAMINE

Uuritavate nähtuste hindamise üks väga olulisi probleeme on ***väljavõtukogumi mahu kindlaksmääramine***. Ei ole ju mõtet uurida liiga suurt väljavõtukogumit, selleks et teha usaldusväärseid järeldusi uuritava üldkogumi kohta, samuti ei saa seda teha liiga väikese väljavõtukogumi alusel.

Välavõtukogumi kindlaksmääramisel lähtutakse asjaolust, et suhteline viga 95%-lisel tõenäosustasemel ei oleks üle 2%. (Pedagoogilistes uuringutes peetakse 95%-list tõenäosustaset ja viga alla 2% optimaalseks.) Lähtume eespool toodud näitest ja valemitest

$$\mu = \frac{m_{\bar{x}}}{\bar{x}_n} \quad \text{ja} \quad m_{\bar{x}} = 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} ;$$

$$m_{\bar{x}} = \mu \cdot \bar{x}_n ; \quad m_{\bar{x}} = \frac{1{,}96 \cdot \sigma}{\sqrt{n}} ;$$

$$m_{\bar{x}} < 0{,}02 \cdot 3{,}8 ; \quad m_{\bar{x}} = \frac{1{,}96 \cdot 0{,}6}{\sqrt{n}} ;$$

$$0{,}02 \cdot 3{,}8 > \frac{1{,}96 \cdot 0{,}6}{\sqrt{n}} ;$$

$$0{,}076 > \frac{1{,}176}{\sqrt{n}} ;$$

$$\sqrt{n} > \frac{1{,}176}{0{,}076} ; \quad \sqrt{n} > 15 ; \quad n > 225 .$$

Nagu eespool toodud näitest selgub, võib väljavõtukogumi suuruse kindlaks määrata eeleksperimendi tulemuste alusel. Selleks korraldatakse eelekperiment juhuslikult valitud uurimisobjektidega ja selle tulemuste põhjal leitakse väljavõtukogumi keskmine ($\bar{x}$) ning standardhälve (σ). Siis leitakse aritmeetilise keskmise standardviga ($m_{\bar{x}}$) ja suhteline viga (μ) tõenäosuse tasemel 95% või 99% (tähtsate ja suurt usaldusväärsust nõudvates uurimustes). Kui suhteline viga on väike (alla 2%), võib väljavõtu mahtu lugeda normaalseks. Kui viga on suur, tuleb eespool toodud näite alusel leida väljavõtukogumi maht, mille puhul suhteline viga 95%-sel (või 99%-sel) tõenäosustasemel oleks alla 2%.

5.3. TULEMUSTE USALDUSPIIRKONNAD

Väikeste väljavõtukogumite korral, kui ei kehti normaaljaotus (üldkogum peab aga olema ligilähedane sellele) ning satandardhälve ei ole täpselt teada, tuleb keskmise tõenäosuslikud hinnangud anda Studenti t-kriteeriumi alusel.

Sel juhul tuleb kõigepealt leida aritmeetilise keskmise standardviga $m_{\bar{x}}$. Edasi leitakse suhe

$$t_{emp} = \frac{\bar{x}_n}{m_{\bar{x}}}.$$

Siis püstitakse hüpotees, et keskmine ei ole usaldatav (nullhüpotees), ja kontrollitakse t-kriteeriumi abil, kui suur on nullhüpoteesi tõenäosus antud väljavõtukogumi elemente (vaatluste arvu) ja suhet t arvestades. Ebatõenäoliseks loetakse hüpoteesi harilikult siis, kui selle tõenäosus on alla 5%.

Kui keskmine on usaldatav ($t_{emp} > t_{kri}$), leitakse keskmise usalduspiirid valemiga

$$\bar{x}_n \pm m_x t_{kri 5\%}$$

Näide. *25 õpilasele korraldati kontrolltöö, mis sisaldas 5 ülesannet; keskmiseks tulemuseks saadi 4 õiget vastust ja standardhälve σ=1,4. Millised on keskmise tulemuse usalduspiirkonnad?*

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1,4}{\sqrt{25}} = \frac{1,4}{5} = 0,28 ,$$

$$t_{emp} = \frac{4,0}{0,28} = 14,2 .$$

t-tabelist leiame, et $t_{kriis\%} 24(n-1)$ vabadusastme korral on 2,06.

Järelikult $t_{emp} > t_{krit} (14,2 > 2,06)$.

Selgub, et nullhüpotees, mis väidab, et keskmine ei ole usaldatav, on väga väikese tõenäosusega. Järelikult tuleb keskmist lugeda usaldatavaks.

Keskmise usalduspiirid

$$4,0 \pm 0,28 \cdot 2,06 ; \; 4,0 \pm 0,58 .$$

Eeltoodust võib järeldada, et kui korraldada veel seeria kontrolltöid samades tingimustes (juhuslikult valitud õpilastega), siis on tõenäoline, et 95%-l juhtudel langeb $\bar{x}_n$ vahemikku 3,42< $\bar{X}$ <4,58.

Näidetest selgub, et mida kitsam on usalduspiirkond uuritava väljavõtukogumi keskmise ümber, seda usaldusväärsem on saadud resultaat. Valemite analüüsist selgub ka, et aritmeetilise keskmise tõenäosus on seda väiksem, mida suurem on standardhälve, ja seda suurem, mida suurem on väljavõtukogumi maht.

Lõpetuseks tuleb rõhutada, et ainuüksi korrektne keskmiste näitajatega opereerimine uuritavate nähtuste hindamisel ei taga veel saadud tulemuste ning nende erinevuse statistilist usaldatavust. Statistilist väljavõtukogumit analüüsides on olulised näitajad ka kogumi elementide tunnuste homogeensus või heterogeensus, st kogumi elementide tunnuste paiknemine keskväärtuse ümber ehk kogumi eri liikmete väärtuste varieeruvus. Sellest ning väljavõtukogumi mahust olenevad väljavõtukogumi aritmeetilise keskmise usalduspiirkonnad. Ilma neid arvestamata ei saa uuritavate nähtuste hindamisel rääkida keskmiste tulemuste korrektsest interpreteerimisest.

Olulisemat

1. **Uuritav väljavõtukogum peab olema representatiivne.**
2. **Mõõtmised ei ole kunagi absoluutselt täpsed.**
3. **Mõõtmistel esinevad metoodilised, mõõtmiste ja arvutuste ebatäpsused ja representatiivsusvead.**
4. **Väljavõtukogumit on võimalik statistiliselt kindlaks määrata.**
5. **Kogumeid võrreldes tuleb kindlaks määrata keskmiste usalduspiirkonnad.**

6. UURIMISTULEMUSTE STATISTILISEST VÕRDLEMISEST

6.1. VÄLJAVÕTTUDE KESKMISTE ERINEVUSTE USALDATAVUS

Uurimistöödes tuleb väga sageli võrrelda kaht väljavõttu, mis kuuluvad ühte ja samasse üldkogumisse. Näiteks tahetakse teada linna- ja maakoolide õpilaste matemaatikaalaste teadmiste taseme erinevust. Ainult keskmiste, nende usalduspiirkondade ning varieeruvuse kõrvutamise alusel ei saa kuigi täpseid andmeid väljavõttude keskmiste erinevuste statistilise usaldatavuse võrdlemiseks.

Praktikas lähtutakse kahe keskmise erinevuse võrdlemisel asjaolust, et mõlema väljavõtu aritmeetiliste keskmiste diferents, kui väljavõtud kuuluvad ühte ja samasse üldkogumisse, on null või juhuslikkuse piires nulli ümber ($\bar{x}_1 - \bar{x}_2 \approx 0$). See ongi nullhüpotees sama üldkogumi kahe väljavõtukogumi diferentside kohta. Enne kui asuda kogumite aritmeetiliste keskmiste erinevuste usaldatavuse uurimisele, tuleb selgitada võrreldavate jaotuste homogeensus või heterogeensus.

Kui ühes väljavõtukogumis on n_1 elementi ja selle kogumi dispersioon on σ_1^2, teises väljavõtukogumis on n_2 elementi ja dispersioon on σ_2^2, kasutatakse dispersioonide võrdlemiseks nende suhet

$$F_{emp} = \frac{\sigma_1^2}{\sigma_2^2},$$

kusjuures ***dispersioonid valitakse selliselt, et nende suhe on alati suurem või võrdne 1-ga.***

Edasi tuleb leida *F-tabelist* (see on igas statistika käsiraamatus; meil vt tabel 6), milline on vabadusastmetega $n_1 - 1$ ja $n_2 - 1$ nullhüpoteesi tõenäosus. Tabeli peast (esimesest reast) tuleb võtta lugeja (suurema dispersiooni) vabadusaste ja tabeli esimesest veerust nimetaja vabadusaste. Tabelis on toodud F väärtused p=5%-le, kuid kasutatakse ka teisi F väärtusi, peamiselt p=1%-le (väga usaldatav tulemus).

Kui selgub, et dispersioonid on oluliselt erinevad, $\sigma_1^2 \gg \sigma_2^2$ ($F_{emp} > F_{krit}$, kui jagamisel saadud suhe F_{emp} erineb oluliselt tabelist saadud väärtusest F_{krit}), st kui nullhüpoteesi tõenäosus on alla 5%, siis on kindel, et kogumid erinevad ja sageli ***ei ole enam vajadustki keskmisi võrrelda***.

Pedagoogikas on dispersioonide erinevused juba suure sisulise tähtsusega. Oletame, et pedagoogilise eksperimendi korraldamise ajal saadi eksperimentaal- ja kontrollgruppides õpilaste teadmiste hindamisel võrdsed hinnete keskmised, kuid dispersioonid olid nendes klassides oluliselt erinevad. See tähendab, et ühes klassis koondusid hinded tihedasti aritmeetilise **keskmise ümber (hinnete jaotumine oli homogeenne)**, teises klassis aga oli palju häid ja väga häid hindeid, aga ka puudulikke ja nõrku hindeid, st **tulemused hajusid (hinnete jaotumine oli heterogeenne)**.

Kui dispersioonid ei erine oluliselt (nullhüpoteesi tõenäosus p>5%), kasutatakse keskmiste erinevuse olulisuse hindamiseks t-kriteeriumi, millest lähemalt järgnevalt.

Selleks et hinnata keskmiste erinevuste olulisust, tehakse kõigepealt kindlaks aritmeetiliste keskmiste diferentside standardviga (m_D).

Selleks kasutatakse praktikas valemit

$$m_D = \sqrt{m_{\bar{x}_1}^{\ 2} + m_{\bar{x}_2}^{\ 2}}\ .$$

Kuna kogumi aritmeetiliste keskmiste standardvead $m_{\bar{x}_1}$ ja $m_{\bar{x}_2}$ arvutatakse valemitega

$$m_{\bar{x}_1} = \frac{\sigma_{n_1}}{\sqrt{n_1}} \quad \text{ja} \quad m_{\bar{x}_2} = \frac{\sigma_{n_2}}{\sqrt{n_2}},$$

siis saadakse valem

$$m_D = \sqrt{\frac{\sigma_{n_1}^{\ 2}}{n_1} + \frac{\sigma_{n_2}^{\ 2}}{n_2}},$$

kus m_D on aritmeetiliste diferentside standardviga, $\sigma_{n_1}^{\ 2}$ ja $\sigma_{n_2}^{\ 2}$ on esimese ja teise väljavõtukogumi dispersioonid ning n_1 ja n_2 on esimese ja teise väljavõtukogumi elementide (vaatluste) arvud.

Tabel 6

Fischeri *F*-kriteerium

(p=5%)

Nime-taja vabadus-aste	Lugeja vabadusaste								
	1	2	3	4	5	6	12	24	∞
1	161,0	199,5	215,7	224,6	230,2	234,0	244,0	249,0	254,3
2	18,5	19,2	19,2	19,3	13,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,5	2,2	2,0	1,7
28	4,2	3,3	3,0	2,7	2,6	2,4	2,1	1,9	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Tabel 7

Studenti *t*-kriteerium

$n-1$	$p=5\%$	$p=1\%$
1	t =12,71	t =63,66
2	4,30	9,93
3	3,19	5,84
4	2,78	4,60
5	2,57	4,03
8	2,30	3,36
10	2,23	3,17
12	2,18	3,06
14	2,15	2,98
16	2,12	2,92
18	2,10	2,88
20	2,09	2,85
22	2,07	2,82
24	2,06	2,80
26	2,05	2,78
28	2,05	2,76
30	2,04	2,75
40	2,02	2,70
60	2,00	2,66
120	1,98	2,62
∞	1,96	2,58

Täpsete arvutuste puhul tuleb erinevuste usaldatavuse kindlaksmääramisel eelkõige selgusele jõuda, kas dispersioonid on homogeensed või heterogeensed, ja vastavalt sellele kasutada ka erinevaid valemeid keskmiste diferentside standardvea arvutamisel.

Kui dispersioonid on oluliselt erinevad (nullhüpoteesi tõenäosus on alla 5%) ja tahetakse võrrelda keskmiste erinevuse statistilist olulisust, kasutatakse järgmisi valemeid:

$$t_{emp} = \frac{D}{m_D} = \frac{|\bar{x}_1 - \bar{x}_2|}{m_D}; \quad t_{emp} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_{n_1}^2}{n_1} + \frac{\sigma_{n_2}^2}{n_2}}};$$

t_{krit} leidmisel Studenti t-kriteeriumi tabelist (vt tabel 7) arvutatakse vabadusastmete arv ligikaudse keskmisena:

$$\text{vabadusastmete arv} = \frac{n_1 + n_2 - 2}{2}.$$

Viimane on peamine oluline erinevus keskmiste diferentside arvutamisel homogeensete ja heterogeensete dispersioonidega väljavõttude puhul.

Kui kogumi elementide arv (katseisikute arv) on mõlemas võrreldavas kogumis enam-vähem ühesugune ($n_1 \approx n_2$) ja mitte väga väike (suurem kui 5), võib kasutada ka homogeensete dispersioonide puhul t_{emp} arvutamiseks eespool toodud valemit.

Veel üks näide. *Õppeaasta lõpul võrreldi õpitulemusi grupis A, kus õppeaasta jooksul rakendati uut meetodit, ja grupis B, kus seda meetodit ei rakendatud.*

Grupis A anti 14, grupis B 10 õiget vastust. Grupis A oli 30, grupis B 20 õpilast; $\sigma_A^2 = 2,2$, $\sigma_B^2 = 6,8$.

Kõigepealt selgitatakse, kas dispersioonid on homogeensed või heterogeensed (oluliselt erinevad):

$$F_{emp} = \frac{6,8}{2,2} \approx 3,1; \quad F_{kriit} = 1,9\,; F_{emp} > F_{kriit}\,(3,1 > 1,9),$$

st nullhüpoteesi tõenäosus on alla 5% ($p<5\%$) ja ka alla 1% ($p<1\%$), mis näitab, et dispersioonid on heterogeensed ja tegemist on täiesti erinevate kogumitega.

Keskmiste diferentside usaldatavuse kontrollimiseks leitakse m_D valemiga, mida rakendatakse heterogeensete dispersioonide puhul. Valemi alusel

$$m_D = \sqrt{\frac{2,2}{30} + \frac{6,8}{20}} \approx \sqrt{0,07 + 0,34} \approx \sqrt{0,41} \approx 0,65\,;$$

$$t_{emp} = \frac{14-10}{0,65} = \frac{4}{0,65} \approx 6,2\,;$$

vabadusastme arv on $\frac{30+20-2}{2} = 24$.

Tabelist (vt tabel 7) leiame t_{kriit} väärtuse 24 vabadusastme puhul:

p=5%, t=2,06,
p=1%, t=2,80.

Eespool toodust järeldub, et nullhüpoteesi tõenäosus p on väiksem kui 5% ja ka väiksem kui 1% ($p<1\%$) ning aritmeetiliste keskmiste diferentside erinevust võib pidada väga usaldusväärseks.

Kokkuvõtteks võib märkida, et

1) nullhüpoteesi saab tagasi lükata (nullhüpotees ei ole võimalik);
2) tegemist on erinevate kogumitega, st uue meetodi rakendamine andis usaldusväärselt paremaid tulemusi, võrreldes seni kasutatavaga.

Tõsist kriitikat on tehtud nn protsendimaaniale (näiteks koole on reastatud õppeedukuse protsendi alusel). Kuid ometi on paratamatu leida klasside, koolide jne töötulemuste võrdlemiseks mingid mõõdetavad ja võrreldavad näitajad. Õppeedukuse protsent ongi olnud üks selline mõõdetav ja võrreldav näitaja. Õppetöö tulemuste hindamisel ei ole seniajani mingeid muid mõõdetavaid näitajaid tavaliselt kasutatavate keskmiste hinnete, keskmiste punktide ja protsentide kõrvale seada olnud. Niikaua kui ei ole leitud muid mõõtmise ja võrdlemise võimalusi, peame olemasolevatega rahulduma.

See ei tähenda hoopiski, et uurijad peaksid õppetöö tulemusi hinnates piirduma ainult keskmiste hinnete ja protsentidega, võttes ainuüksi need aluseks. Kuna meil paraku ei peeta veel kõikides õppeainetes ja kõikides koolides kinni kehtestatud hindamisnõuetest, küllaltki palju esineb hindamisel liberaalsust ja subjektivismi, on ilmne, et ainult hinnete keskmised ja õppeedukuse protsendid ei peegelda objektiivselt õpilaste teadmiste ja oskuste taset. Kindla koha õppetöö kvaliteedi kontrollimisel peavad omandama õppetundide vaatlused, vestlused õpetajate ja õpilastega, õpilaste tööde analüüs ning standardiseeritud kontrolltööd, mis nõuavad igale küsimusele üheseid vastuseid ja mis viiakse läbi neutraalse isiku juuresolekul.

Selleks et otsustada, kas kahe kooli õppeedukuse (ühesuguste kontrolltööde õiged vastused) erinevus protsentides on usaldatav, võib kasutada protsentarvude diferentsi ja protsentarvude difetentsi keskmise vea jagatist t:

$$t = \frac{D}{m_{D\%}}.$$

Viimast nimetatakse ka kriitiliseks jagatiseks. Kontrolltööde õigete vastuste arv protsentides saadakse, kui õpilaste õigete vastuste summa jagada kõigi võimalike õigete vastuste summaga ja saadud arv korrutada 100-ga.

Protsentarvude diferentsi D leiame valemiga $D = p_1 - p_2$.

Protsentarvude diferentsi keskmine viga arvutatakse valemiga

$$m_{D\%} = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}},$$

kus $m_{D\%}$ - protsentarvude diferentsi keskmine viga;

p_1 ja p_2 - uurimisel saadud protsentarvud (esimese ja teise kooli õppeedukus);

q_1 ja q_2 - vastavalt 100-p_1 ja 100-p_2;

n_1 ja n_2 - õpilaste arv esimeses ja teises koolis.

Kui protsentarvude diferents D on 2-3 korda suurem protsentarvude diferentsi keskmisest veast ($t>2$), võib erinevust kahe protsentarvu vahel lugeda usaldatavaks. Kui $t>3$, on erinevus kindlalt usaldatav.

Toome näite. *Olgu uurimisobjektiks kolm õppeasutust, kus õpilaste arv ja õppeedukus on erinev (kontrolltööde tulemused protsentides). Koolis A on õpilasi 100, õppeedukus 90%, koolis B on õpilasi 400, õppeedukus 85% ja koolis C on õpilasi 900 ning õppeedukus 82%. Anda vastus kolmele küsimusele:*

1. Kas kooli A õppeedukus on parem kooli B omast?

$$m_{D\%} = \sqrt{\frac{90 \cdot 10}{100} + \frac{85 \cdot 15}{400}} = \sqrt{9 + 3{,}18} = 3{,}5\%.$$

$$D = 90\% - 85\% = 5\%, \quad t = \frac{5}{3{,}5} = 1{,}4, \quad t<2,$$

erinevus ei ole usaldatav.

2. Kas kooli B õppeedukus on parem kooli C omast?

$$m_{D\%} = \sqrt{\frac{85 \cdot 15}{400} + \frac{82 \cdot 18}{900}} = \sqrt{3{,}18 + 1{,}64} = 2{,}2\%.$$

$$D = 85\% - 82\% = 3\%, \quad t = \frac{3}{2{,}2} = 1{,}4, \quad t<2,$$

erinevus ei ole usaldatav.

3. Kas kooli A õppeedukus on parem kooli C omast?

$$m_{D\%} = \sqrt{\frac{90 \cdot 10}{100} + \frac{82 \cdot 18}{900}} = \sqrt{9 + 1,64} = 3,2\%.$$

$$D = 90\% - 82\% = 8\%, \quad t = \frac{8}{3,2} = 2,5, \quad t > 2,$$

erinevust võib pidada usaldatavaks. (Kui t oleks suurem kui 3, oleks erinevus kindlalt usaldatav.)

Nagu eespool toodust selgub, tuleb õppetöö tulemuste võrdlemisel olla küllaltki ettevaatlik. Ainult keskmiste tulemuste kõrvutamine ei anna veel statistiliselt usaldatavat alust kindlate järelduste tegemiseks. Selle kõrval tuleb tingimata leida ka uuritavate kogumite dispersioonid ning arvutada keskmiste erinevuse statistiline usaldatavus.

Olulisemat

1. **Statistiliste arvutuste puhul on vaja kindlaks teha väljavõttude keskmiste erinevuste usaldatavus.**
2. **Kui dispersioonid erinevad oluliselt, ei ole sageli vaja täiendavaid arvutusi.**
3. **Kui dispersioonid ei erine oluliselt, kasutatakse keskmiste erinevuse olulisuse hindamiseks t-kriteeriumi.**
4. **Täpsete arvutuste puhul tuleb keskmiste erinevuste usaldatavuse kindlaksmääramiseks selgusele jõuda, kas dispersioonid on homogeensed või heterogeensed.**
5. **Õppetöö tulemuste võrdlemisel peab olema ettevaatlik, ainult keskmiste tulemuste võrdlemine ei anna veel statistiliselt usaldatavat alust kindlate järelduste tegemiseks.**

7. KOGUMITE TUNNUSTE VAHELINE SEOS. KORRELATSIOON

7.1. KORRELATSIOONI OLEMUS

Eelmistes osades selgitati meetodeid, mille abil saab kindlaks teha, mil määral kogumi kvantitatiivsete tunnustega väljavõtud üksteisest erinevad.

Uurimistöös on aga sageli vaja kindlaks teha, kas kahe kogumi mõõdetud tunnuste vahel on seos või mitte. Uurijat võib huvitada, kas need õpilased, kes on tugevad näiteks emakeeles, on seda ka võõrkeeles, kas matemaatikas hästi edasijõudvad õpilased on tugevad ka füüsikas ja kas matemaatikas mitterahuldavalt edasijõudjad on mahajääjad ka füüsikas.

Kahe nähtuse vahel võib olla kas funktsionaalne sõltuvus või korrelatiivne seos. Esimesel juhul vastab argumendi mingile kindlale väärtusele funktsiooni mingi kindel väärtus. Näiteks on funktsionaalseks seoseks ringjoone raadiuse ja tema ümbermõõdu vaheline seos ($Ü=2\pi r$), kus ringjoone raadiuse mingile kindlale väärtusele vastab ka ümbermõõdu mingi üks kindel väärtus.

Korrelatiivse seose puhul aga võib ühe muutuja mingile väärtusele vastata mitu teise muutuja väärtust, mis tavaliselt pole kindlalt määratud. Korrelatsiooni puhul ei saa üldiselt väita, et mingi nähtus sõltub ainult ühest faktorist. Näiteks õpilaste füüsikaülesannete lahendamise oskus võib tõepoolest oluliselt oleneda matemaatikaalastest teadmistest, kuid seda mõjutavad peale selle ka füüsika põhiseaduste tundmine, kontrolltööks ettevalmistumine (õppimine), õpilaste meeleolu jne.

Statistikas kasutatakse kahe tunnuse (x ja y) vahelise seose tugevuse arvulise näitajana korrelatsioonikordajat e korrelatsioonikoefitsienti (tähistatakse kas ϕ, ρ või r).

Korrelatsioonikordaja paikneb alati vahemikus $-1<r<+1$. Kui $r=1$ või $r=-1$, on korrelatsioon kahe tunnuse vahel täielik, st sõltuvus on kirjeldatav lineaarse funktsiooniga

($r=1$ korral x suurenemisel kasvavad ka y väärtused; $r=-1$ korral on tegemist kindlalt negatiivse seosega: x suurenedes y väheneb). Kui $r=0$, siis uuritavate tunnuste vahel seos puudub.

Meie näites tähendaks $r=1$ seda, et iga matemaatikat hästi oskav õpilane oskab hästi ka füüsikat ja ükski õpilane, kes ei oska matemaatikat, ei oska ka füüsikat. Kui $r=-1$, oleks sõltuvus vastupidine: mida paremini õpilane oskab matemaatikat, seda edutum on füüsikas, ja vastupidi. Kui aga sõltuvust matemaatikas ja füüsikas edasijõudmise vahel üldse pole (matemaatikas edukas õpilane võib füüsikas osutuda nii edukaks kui edutuks ja mõlemad võimalused on võrdselt tõenäosed), on $r=0$.

Uuritavate nähtuste puhul korrelatsioonikoefitsiente $r=1$, $r=-1$ ja $r=0$ praktiliselt ei esine. Alati on kas $-1<r<0$ või $0<r<1$. Viimased seosed viitavad alati sõltuvuse olemasolule, ent see võib osutuda kas tugevaks (statistiliselt oluliseks) või nõrgaks (statistiliselt mitteoluliseks). Seose tugevuse määrab koefitsiendi absoluutväärtus ja seose iseloomu koefitsiendi märk.

Tuleb rõhutada, et korrelatsioonikoefitsient ei anna alust seoste põhjuste või tingimuste üle otsustamiseks. Ta küll näitab, et nähtuste vahel eksisteerib vastastikune seos, kuid ei selgita, kas esimene nähtus põhjustab teise, teine esimese või sõltuvad mõlemad nähtused hoopis mingist kolmandast tegurist.

Tunnustevaheliste seoste kindlakstegemiseks ei ole oluline, kas tunnuste väärtused on mõõdetud ühes ja samas mõõtühikute süsteemis või mitte.

Korrelatsioonikordaja arvutamiseks on mitu võimalust. Valik sõltub alljärgnevatest asjaoludest.

- Kõigepealt sellest, kuidas võrreldavad tunnused on esitatud, kas lihtsa statistilise reana või variatsioonireana, kus tunnuste väärtused jaotuvad teatud sagedustega intervallidesse.
- Oluline on, kas tuleb uurida seost kahe või enama tunnuse vahel.
- Lõpuks on tähtis, millisel kujul tunnused on esitatud, kas on tegemist kirjeldavat laadi või täpselt mõõdetavate kvantitatiivsete tunnustega.

7.2. KORRELATSIOONIKOEFITSIENDI (Φ) ARVUTAMINE KVALITATIIVSETE TUNNUSTE ALUSEL

Uurimistöös kasutatavad kvalitatiivsed tunnused on sageli alternatiivsed ja nende puhul tuleb valida üks kahest võimalusest. Näiteks tuleb uurija küsimusele vastata kas "jah" või "ei". Küsimusi, mille puhul tuleb anda ühene valikvastus, kasutatakse eriti sageli ankeetides ja testides.

Alternatiivsete tunnuste puhul arvutatakse korrelatsioonikoefitsient Φ valemiga

$$\Phi=\frac{AD-BC}{\sqrt{(A+B)\cdot(C+D)\cdot(A+C)\cdot(B+D)}},$$

kus A, B, C ja D on üksikud nelja välja skeemis esinevad sagedused. Sellist korrelatsiooni nimetatakse ka **neljaväljaliseks korrelatsiooniks**.

Näide. *Sajast õppeaasta jooksul aktiivselt spordiga tegelenud õpilasest haigestus 20; kolmesajast aktiivselt spordiga mittetegelenust haigestus 180.*

Andmed on esitatud tabelis (vt tabel 8).

Tabel 8

Katseandmed

x Õpilased	y_1 Haigestusid	y_2 Ei haigestunud	Kokku
x_1 Spordiga tegelejad	20 (A)	80 (B)	100 (A+B)
x_2 Spordiga mittetegelejad	180 (C)	120 (D)	300 (C+D)
Kokku	200 (A+C)	200 (B+D)	400 (A+B+C+D)

$$\Phi = \frac{20 \cdot 120 - 80 \cdot 180}{\sqrt{100 \cdot 300 \cdot 200 \cdot 200}} \approx -\frac{12000}{35000} \approx -0,4 .$$

Negatiivne seos -0,4 näitab, et spordiga tegelevad õpilased haigestuvad harvemini kui spordiga mittetegelevad.

Korrelatsioonikoefitsiendi tavaline tõlgendus on toodud tabelis 9.

Tabel 9

Korrelatsioonikoefitsiendi tõlgendus

Koefitsient	Seos
$\lvert\varphi\rvert \le 0,3$	Nõrk
$0,3 \le \lvert\varphi\rvert \le 0,5$	Mõõdukas
$0,5 \le \lvert\varphi\rvert \le 0,7$	Märkimisväärne
$0,7 \le \lvert\varphi\rvert \le 0,9$	Tugev
$0,9 \le \lvert\varphi\rvert \le 1,0$	Väga tugev

Eespool toodud tõlgendused on üsna ebamäärased ja hinnangu andmisel pole arvestatud väljavõtu mahtu. Väljavõttude korral tuleb aga korrelatsioonikoefitsienti tingimata uurida nullhüpoteesi suhtes ja iseloomustada ka selle usaldusväärsust. Selleks on statistika käsiraamatus olemas vajalikud tabelid.

7.3. JÄRJESTUS- EHK ASTAKKORRELATSIOON

Kui väljavõtukogumi tunnuste väärtusi on võimalik reastada kas kasvavas või kahanevas järjekorras ja väljavõtt on väike ($n<30$), on otstarbekas kasutada nn järjestuskorrelatsiooni. Seda meetodit võib kasutada alati, kui on tegemist samaaegselt mõõdetud andmete ridadega.

Selle korrelatsiooni puhul lähtutakse juba vahetutest mõõtarvudest, mitte kvalitatiivsetest tunnustest.

Kui reastada mõõtarvud nende kasvamise või kahanemise järjekorras, vastab igale mõõtarvule kindel kohanumber, seega mõõtarvude endi asemel võib vastavas tabelis reastada kohanumbrid.

Näide. *Oletame, et taheti selgusele jõuda, kas on seos eesti ja võõrkeele õppimise tulemuste vahel. Klassi 15 õpilast reastati õpetajate hinnangute alusel eesti keele ja võõrkeele oskuse järgi.*

Tulemused kanname taas tabelisse (vt tabel 10). Tabeli 2. ja 3. veerus on tunnuste x ja y astakud, mida vaadeldakse kui tunnuste väärtusi ja arvutatakse nendevaheline korrelatsioon.

Tabel 10

Edasijõudmise võrdlus

Õpilased	Kohanumber (eesti keel) x'	Kohanumber (võõrkeel) y'	$D=\lvert x'-y'\rvert$	D^2
1	2	3	4	5
A	14	13	1	1
B	11	14	3	9
C	12	2	10	100
D	4	5	1	1
E	1	3	2	4
F	13	10	3	9
G	15	15	0	0
H	7	8	1	1
I	3	4	1	1
J	5	6	1	1
K	10	9	1	9
L	2	1	1	1
M	8	11	3	9
N	6	7	1	1
O	9	12	3	9
n=15				$\sum D^2$=148

Järjestuskorrelatsioon arvutatakse valemiga

$$\rho = 1-\frac{6\sum D^2}{n(n^2-1)},$$

kus ρ (roo) on järjestuskorrelatsiooni koefitsient, D^2 on iga õpilase mõlema kohanumbri diferentsi ruut ja n on väljavõtu maht.

Meie näites

$$\rho = 1-\frac{6\cdot 148}{15\cdot(15^2-1)} = 1-\frac{888}{3360} = 1-0,26 = 0,74\,.$$

Väikeste väljavõttude puhul tuleb hoolega kontrollida järjestuskorrelatsiooni koefitsiendi usaldatavust nullhüpoteesi suhtes. Selleks kasutatakse eraldi tabelit (vt tabel 12).

Viimase näite puhul tuleb nullhüpotees kõrvale jätta ja lugeda seos väga usaldusväärseks.

7.4. LINEAARNE KORRELATSIOON

Kahe kvantitatiivse normaaljaotusega tunnuse vahelise seose kindlaksmääramiseks on sobivaim lineaarne korrelatsioon.

Lineaarset korrelatsiooni saab arvutada mitmel viisil. Meetodi valik oleneb sellest, millised andmed on varem välja arvutatud. Kui on teada näiteks kummagi tunnuse väärtuste rea aritmeetiline keskmine ja keskmine ruuthälve, on otstarbekas kasutada valemit:

$$r = \frac{\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y},$$

kus $(x_i - \bar{x})$ on iga üksiku x väärtuse hälve aritmeetilise keskmise suhtes, $(y_i - \bar{y})$ on iga üksiku y väärtuse hälve aritmeetilise keskmise suhtes, n on võrreldavate paaride arv ning σ_x ja σ_y on keskmised ruuthälbed.

Näide. *Tuleb leida, kas esineb korrelatsioon kahe tegevuse kiiruse vahel. Kas üht tegevust kiiresti sooritanud õpilased saavad kiiresti hakkama ka teise tegevusega või mitte?*

Mõlema tegevuse sooritamise ajad on kantud tabelisse (vt tabel 11).

Arvutame keskmised:

$$\bar{x} = \frac{60}{10} = 6, \quad \bar{y} = \frac{45}{10} = 4,5 .$$

Keskmised ruuthälbed arvutatakse valemiga

$$\sigma_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}, \quad \sigma_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n}}.$$

Tabel 11

Tegevuskiiruste vaheline seos

Õpilane	Esimese tegevuse aeg x_i	Teise tegevuse aeg y_i	$x_i - x$	$y_i - y$	$(x_i - x)^2$	$(y_i - x)^2$	$(x_i - x)\cdot(y_i - y)$
1	2	3	4	5	6	7	8
A	3	1	-3	-3,5	9	12,25	10,5
B	4	2	-2	-2,5	4	6,25	5,0
C	5	2	-1	-2,5	1	6,25	2,5
D	5	6	-1	+1,5	1	2,25	1,5
E	6	4	0	-0,5	0	0,25	0
F	6	5	0	+0,5	0	0,25	0
G	7	5	+1	+0,5	1	0,25	0,5
H	7	6	+1	+1,5	1	2,25	1,5
I	8	6	+2	+1,5	4	2,25	3,0
J	9	8	+3	+3,5	9	12,25	10,5
N=10	$\sum x_i$=60	$\sum y_i$=45	$\sum$=0	$\sum$=0	$\sum$=30	$\sum$=44,5	$\sum$=35

Antud näites

$$\sigma_x = \sqrt{\frac{30}{10}} = \sqrt{3} \approx 1{,}73 \quad \text{ja} \quad \sigma_y = \sqrt{\frac{44{,}5}{10}} = \sqrt{4{,}45} \approx 2{,}11;$$

korrelatsioonikordaja

$$r = \frac{35}{10 \cdot 1{,}73 \cdot 2{,}11} \approx 0{,}96 .$$

***Märkus**. Arvutuste õigsust saab kontrollida, liites tabeli 4. ja 5. veeru (vt tabel 11).*

Summad $\sum(x_i - \bar{x}) = 0$ ja $\sum(y_i - \bar{y}) = 0$.

Uurimistöö praktikas on tavaliselt tegemist väikeste väljavõttudega ning r usaldatavus määratakse tabelist, kus on antud 5- ja 1-protsendiline jääktõenäosus ning vabadusastmete arv (vt tabel 13). Korrelatsiooni arvutamise puhul on vabadusastmete arv alati n-2, kus n on võrreldavate tunnusepaaride arv.

Nagu statistikakäsiraamatute tabelitest (vt tabel 12 ja 13) selgub, oleneb korrelatsioonikoefitsiendi statistiline usaldatavus suurel määral uuritavate tunnuste arvust. Meie näites, kus $n=10$ ja vabadusaste on 8, osutub $r=0,96$ statistiliselt igati usaldusväärseks (vt tabel 13). Seega võib väita, et need õpilased, kes sooritasid kiiresti esimese tegevuse, on nobedamad ka teise tegevuse puhul.

Tabel 12

Astakkorrelatsiooni usaldatavus

n	Usaldatavus		*n*	Usaldatavus	
	5 %	1 %		5 %	1 %
4	1,000	–	16	0,425	0,601
5	0,900	1,000	18	0,399	0,564
6	0,829	0,943	20	0,377	0,534
7	0,714	0,893	22	0,359	0,508
8	0,643	0,833	24	0,343	0,485
9	0,600	0,783	26	0,329	0,465
10	0,564	0,746	28	0,317	0,448
12	0,506	0,712	30	0,306	0,432
14	0,456	0,645			

Tabel 13

Lineaarkorrelatsiooni usaldatavus

Vabadusaste	Usaldatavus		Vabadusaste	Usaldatavus	
(n-2)	5 %	1 %	(n-2)	5 %	1 %
2	0,95	0,99	10	0,48	0,61
3	0,88	0,96	20	0,42	0,53
4	0,81	0,92	25	0,38	0,49
5	0,75	0,87	35	0,32	0,42
6	0,70	0,83	50	0,27	0,35
7	0,67	0,80	60	0,25	0,33
8	0,63	0,77	80	0,22	0,28
9	0,60	0,74	100	0,19	0,25
10	0,58	0,71	200	0,14	0,18

Olulisemat

1. **Kahe nähtuse vaheline seos võib olla kas funktsionaalne sõltuvus või korrelatiivne seos.**
2. **Kahe tunnuse (*x* ja *y*) vahelise seose tugevuse arvulise näitajana kasutakse korrelatsioonikordajat ehk korrelatsioonikoefitsienti.**
3. **Korrelatsioonikoefitsienti võib arvutada nii kvalitatiivsete kui kvantitatiivsete tunnuste alusel.**
4. **Kui väljavõtukogumi tunnuste väärtusi on võimalik reastada ja väljavõtt on väike, on otstarbekas kasutada nn järjestuskorrelatsiooni.**
5. **Kahe kvantitatiivse normaaljaotusega tunnuse vahelise seose kindlaksmääramiseks on sobivam lineaarne korrelatsioon.**
6. **Korrelatsioonikoefitsiendi statistilist usaldatavust tuleb kontrollida vastavate tabelite abil, sest usaldatavus oleneb suurel määral uuritavate tunnuste arvust.**

8. UURIMISTULEMUSTE VÕRDLEMINE MITTEPARAMEETRILISTE MEETODITEGA

8.1. MITTEPARAMEETRILISTE JA PARAMEETRILISTE MEETODITE KASUTAMINE

Mitteparameetrilisteks kutsutakse selliseid statistilisi meetodeid, mille puhul parameetritele (nähtust iseloomustavatele suurustele) tähelepanu ei osutata ja normaaljaotuse olemasolu ei peeta vajalikuks. Suuremat osa mitteparameetrilisi meetodeid võib kasutada ka siis, kui uurimistöö tulemused on mõõdetud nimi- ja järjestusskaala abil. Seetõttu on mitteparameetriliste meetodite kasutamine sel puhul igati kohane.

Uurijad peavad pöörama tähelepanu sellele, et *F*- ja *t*-kriteeriumi võib rakendada *ainult normaaljaotuse puhul*. Nimelt peavad nende kasutamisel analüüsitavate ja võrreldavate ridade tunnused jaotuma aritmeetilise keskmise ümber normaaljaotusele vastavalt. Sel juhul peab rea diferentsiaalse jaotuskõvera kuju lähenema Gaussi kõverale. Eriti oluline on seejuures sümmeetria ja ühetipulisus.

Tihtipeale aga ei kontrollita, kas on tegemist normaaljaotusega või mitte, ning tulemuste usaldatavuse kontrollimiseks kasutatakse *F*- ja *t*-testi ka siis, kui andmed ei jaotu normaaljaotuse kohaselt. Sageli on aga ka raske kindlaks teha (eriti siis, kui on tegemist väikeste väljavõttudega), kas on tegemist normaaljaotusele lähedase jaotusega või mitte.

Mida suurem on erinevus normaaljaotusest, seda enam on *F*- ja *t*-kriteeriumi kasutamisel eksimisvõimalusi.

Tuleb arvestada, et normaaljaotus on loodus- ja ühiskonnanähtuste puhul kõige sagedasem ning seepärast suure väljavõtu puhul tavaliselt normaaljaotuse seaduse kasutamise vastu ei eksitagi. Kui on tegemist väikeste väljavõttudega, tuleks siiski enne valemite kasutamist kontrollida, kas tunnuste väärtused jaotuvad normaaljaotusele vastavalt või mitte.

Ligikaudselt saab jaotuse üle otsustada jaotuse keskmise, moodi ja mediaani kõrvutamise teel. Kui keskmine, mood ja mediaan langevad kokku, võib jaotuse lugeda normaalseks.

8.2. χ^2-MEETODI KASUTAMINE NELJAVÄLJALISE SKEEMI PUHUL

Uurimistöös võib kasutada ka meetodeid, mis lubavad kahe väljavõtu erinevuse usaldatavust kontrollida mitte ainult normaaljaotuse puhul, vaid ka siis, kui parameetreid ei ole võimalik kvantitatiivselt täpselt hinnata. Ühe sellise võimaluse pakub χ^2-(hii-ruut)meetod neljaväljalise skeemi puhul. Sel juhul leitakse mõlema rea üla- ja allpool mediaani asetsevate juhtumite arv ning uuritakse vastavate katseseeriate sageduste tõenäosust nullhüpoteesi suhtes.

Tähelepanu tuleb aga juhtida sellele, et nimetatud meetodit võib kasutada vaid juhuslike väljavõttude puhul.

Oletame näiteks, et 40 õpilaselt küsiti, kas neile meeldib rohkem iseseisev või tavaline õppetöö. Õpilased pidid esimesele küsimusele vastama kas "jah" või "ei". Kui õpilane vastas esimesele küsimusele "jah", oli teisele vastus "ei". Antud juhul on tegemist nähtuse alternatiivsete tunnustega, st selliste tunnustega, mille puhul tuleb valida üks kahest.

Hinnangute tulemused kantakse tabelisse neljaväljalisse skeemi (vt tabel 14).

Tabel 14

Hinnangu tulemused

Meetod	Hinnang: jah	Hinnang: ei	Kokku
Õppimine iseseisvalt	30 (A)	10 (B)	40 (A+B)
Tavaline õppetöö	10 (C)	30 (D)	40 (C+D)
Kokku	40 (A+C)	40 (B+D)	80 (A+B+C+D)

Selgub, et iseseisvalt õppimist pooldab kolm korda rohkem õpilasi. Selleks et kontrollida, kas selline hinnang on statistiliselt usaldatav või mitte, kasutatakse χ^2-testi. χ^2 arvutatakse valemiga:

$$\chi^2 = \frac{(A-B-1)^2}{A+B},$$

kus A on neljaväljalise skeemi suurem ja B skeemi väiksem arv. Käesolevas näites

$$\chi^2 = \frac{(30-10-1)^2}{30+10} = \frac{361}{40} = 9{,}025 \approx 9{,}03.$$

Vabadusaste antud näite puhul on veergude arv -1. Antud juhul on vabadusastmete arv 2-1=1.

Tabel 15

χ^2**-kriteeriumi tabel**

$n-1$	95 %	99 %	$n-1$	95 %	99 %
1	3,84	6,63	16	26,3	32,0
2	5,99	9,21	17	27,6	33,4
3	7,81	11,3	18	28,9	34,8
4	9,49	13,3	19	30,1	36,2
5	11,1	15,1	20	31,4	37,6
6	12,6	16,8	21	32,7	38,9
7	14,1	18,5	22	33,9	40,3
8	15,5	20,1	23	35,2	41,6
9	16,9	21,7	24	36,4	43,0
10	18,3	23,2	25	37,7	44,3
11	19,7	24,7	26	38,9	45,6
12	21,0	26,2	27	40,1	47,0
13	22,4	27,7	28	41,3	48,3.
14	23,7	29,1	29	42,6	49,6
15	25,0	30,6	30	43,8	50,9

Võrreldes tabelist saadud χ^2 väärtusi, selgub, et $\chi^2_{emp} > \chi^2_{krit99\%} > \chi^2_{krit95\%}$ (9,03>6,63>3,84). Järelikult võib õpilaste juhuslikul valikul saada seda liiki χ^2 väärtused tõenäosusega rohkem kui 99%.

8.3. SIGNATUURTEST

Üks lihtsamaid mitteparameetrilisi meetodeid, mida saab kasutada kahe erineva kvalitatiivse tunnusega juhusliku rea võrdlemiseks, on nn märgi- ehk signatuurtest.

Näiteks huvitab uurijat küsimus, kas esimeses grupis korraldatud kontrolltöö tulemused on usaldatavalt paremad kontrolltöö tulemustest teises grupis. (Eeldame muidugi, et hindamisel kasutati ühesuguseid objektiivseid kriteeriume, näiteks kümnele küsimusele õigesti antud vastuste arvu.) Tulemused kantakse paariviisi tabelisse juhusliku reana. Selle aluseks võib olla näiteks õpilaste järjekorranumber klassis (vt tabel 16).

Tabel 16

Õigete vastuste arv

Õpilaste järjekorranumber grupi nimekirjas	1	2	3	4	5	6	...	n
Õigeid vastuseid esimeses grupis x_i	6	7	5	8	5	6	...	6
Õigeid vastuseid teises grupis y_i	4	8	3	7	5	7	...	6
Vahe $x_i - y_i$ märk	+	-	+	+	0	-	...	0

Nagu tabelist selgub, võib iga arvupaari x_i ja y_i vahel olla ka seos $x_i > y_i(+)$; $x_i < y_i(-)$ või $x_i = y_i(0)$. Tabeli neljandasse ritta ei märgita kontrolltööde tulemuste arvuliste väärtuste vahet, vaid tulemuste suurusjärkude võrdlustulemus märkide "+", "-" või "0" abil. See lihtsustab arvutustööd tunduvalt.

Tulemuste kokkuvõtmisel jäetakse "0"-märgid vaatlusest kõrvale ning vähendatakse vastavalt ka võrreldavate arvupaaride hulka. Edasi leitakse märkide "+" ja "-" sagedused (arvud) ning kontrollitakse nullhüpoteesi abil, kas märkide "+" ja "-" erinev arv esineb juhuse tõttu või mitte. Nullhüpoteesile võib kõige paremini leida vastuse z-testi abil. Selle jaoks kehtib valem

$$z=\frac{\left[\left(f"+"-n\cdot P\{+\}\right)-0,5\right]}{\sqrt{n\cdot\left(P\{+\}\cdot P\{-\}\right)}},$$

kus $f"+"$ ($f"-"$) on sagedamini (harvemini) esinevate märkide arv;

n - võrreldavate arvupaaride arv;

$P\{+\}$ - märgi "+" sageduse oletatav tõenäosus;

$P\{-\}$ - märgi "-" sageduse oletatav tõenäosus;

0,5 - koefitsient.

Märkus: *Märkide "+" ja "-" esinemine on võrdtõenäoline ja võrdne poolega, st P=0,5.*

Oletame, et vaadeldi 40 arvupaari, kusjuures märk "+" esines 25, märk "-" 10 korral ja märk "0" 5 korral. Seega

$$z=\frac{\left[(25-35\cdot 0,5)-0,5\right]}{\sqrt{35\cdot 0,5\cdot 0,5}}=\frac{7}{\sqrt{8,75}}\approx 2,4.$$

Tabel 17

z-jaotus

z	%	z	%	z	%
0,01	0,40	0,70	25,80	2,8	49,74
0,05	1,99	0,80	28,81	3,2	49,93
0,10	3,98	0,90	31,59	3,6	49,98
0,15	5,96	1,0	34,13	4,0	49,99
0,20	7,93	1,2	38,49		
0,25	9,87	1,4	41,92		
0,30	11,79	1,6	44,52		
0,40	15,54	1,8	46,41		
0,50	19,15	2,0	47,72		
0,60	22,57	2,4	49,18		

Tabelist leitud z väärtus 2,4 näitab, et ühel pool aritmeetilist keskmist asub z väärtuse ja aritmeetilise keskmise vahel 49,18% juhte. Teisel pool asub samuti 49,18% juhte, kokku seega 98,36%. Jääktõenäosus (p) on antud näite puhul 100-98,36=1,64%. Seega 1%<p<5%, mille alusel võib väita, et erinevus juhuslike ridade vahel on 5%-lise olulisuse nivoo puhul usaldatav, 1%-lisel olulisuse nivool aga mitteusaldatav.

Märgitesti tulemuste usaldatavust on lihtne kontrollida ka tabeli abil (vt tabel 18).

Tabel 18

Märgitesti tulemuste usaldatavuse kontrollimine

n	$p=1\%$	$p=5\%$	n	$p=1\%$	$p=5\%$
10	0	1	55	17	19
15	2	3	60	19	21
20	3	5	65	21	24
25	5	7	70	23	26
30	7	9	75	25	28
35	9	11	80	28	30
40	11	13	85	30	32
45	13	15	90	32	35
50	15	17			

Tabelist saab leida kriitilise väärtuse $\overline{m}_n$, mis näitab, milline võib olla suurim miinuste (või plusside) arv etteantud usaldatavuse tagamiseks.

Meie näites, kus n=35, võib 1%-lise olulisuse nivoo korral olla miinuseid 9; 5%-lise nivoo korral 11. Kuna meil oli miinuseid 10, on tulemus 5%-lise olulisuse nivoo korral tõepoolest usaldatav, 1%-lise olulisuse nivoo puhul aga mitte.

Tabelis on toodud 90 vaatluspaari kriitilised väärtused. Kui vaatluspaare on üle 90, võib $\overline{m}_n$ väärtuse 5%-lisel olulisuse nivool üsna täpselt arvutada valemiga

$$\overline{m}_n = \frac{n-1}{2} = -0,98\sqrt{n+1},$$

kus n on vaatluspaaride arv.

8.4. Märgitesti kasutamise teine võimalus

Märgitesti võib kasutada ka siis, kui $x_i - y_i$ pole arvude vahe, vaid kontrolltöö mingile küsimusele antud õigete ("+") või valede ("-") vastuste arv.

Vaatame näiteks kontrolltööd eksperimentaal- ja kontrollgrupis, kus iga küsimuse vastus hinnati kas õigeks (Õ) või valeks (V). Viimaste hulka arvati ka vastamata jäänud küsimus.

Tulemused kantakse tabelisse. Kui eksperimentaalgrupi õpilane andis küsimusele õige, kontrollgrupi õpilane aga vale vastuse, kirjutatakse tabeli neljanda rea lahtrisse "+"; kui eksperimentaalgrupi õpilane andis vale vastuse ja kontrollgrupi õpilane õige vastuse, kirjutatakse "-" (vt tabel 19).

Tabel 19

Vastused

Õpilase jrk nr	1.	2.	3.	4.	5.	...	n
Eksperimentaalklass x_i	Õ	Õ	V	V	Õ	...	Õ
Kontrollklass y_i	V	Õ	V	Õ	V	...	V
Vahe $x_i - y_i$ märk	+	0	0	-	+	...	+

Kui vastused olid ühesugused, kas õiged või valed, märgitakse lahtrisse "0" ja see tulemus jäetakse hiljem vaatlusest välja.

Oletame, et 30 õpilaspaari puhul saadi plusse 20, miinuseid 5 ja nulle 5. Kasutades eespool toodud valemit, saame:

$$z = \frac{\left[(20 - 25 \cdot 0,5) - 0,5\right]}{\sqrt{25 \cdot 0,5 \cdot 0,5}} = \frac{7}{\sqrt{6,25}} = \frac{7}{2,5} = 2,8 \approx 3.$$

Tabelist selgub, et $z = 3$ korral on 49,865% sagedustest ühel pool aritmeetilist keskmist. Teisel pool asub sama palju sagedusi, järelikult on tõenäosus 99,79%. Tabeli 18 põhjal on kriitiline väärtus $\overline{m}_n$ 1%-lise olulisuse nivoo puhul 5 miinusmärki ja 5%-lise olulisuse nivoo puhul 7 miinusmärki. Järelikult on meie tulemus usaldatav nii 1%-lise kui ka 5%-lise olulisuse nivoo korral.

Olulisemat

1. **Mitteparameetriliste statistiliste meetodite puhul nähtusi iseloomustavatele suurustele (parameetritele) tähelepanu ei pöörata.**
2. **Nende kasutamisel ei peeta normaaljaotuse olemasolu vajalikuks.**
3. **F- ja t-kriteeriume võib rakendada ainult normaaljaotuse puhul.**
4. **Kui aritmeetiline keskmine, mood ja mediaan langevad ligikaudu kokku, võib jaotust lugeda normaalseks.**
5. **Kui parameetreid ei ole võimalik kvantitatiivselt täpselt hinnata, võib kahe väljavõtu erinevust kontrollida χ^2-meetodi kasutamisega neljaväljalise skeemi abil.**
6. **Kahe erineva kvantitatiivse tunnusega juhusliku rea võrdlemiseks on märgi- ehk signatuurtest.**
7. **Märgitesti võib kasutada ka siis, kui peale arvude vahede on olemas ka mingid kvalitatiivsed näitajad ("+", "-", "Õ", "V", "jah", "ei").**
8. **Tulemuste usaldatavust tuleb kontrollida nii 1%-lise kui ka 5%-lise olulisuse nivoo puhul.**

9. UURIMISTETULEMUSTE ESITAMINE TABELITENA

9.1. UURIMISTULEMUSTE ESITAMISE VIISID

Uurimistöös saadud kvantitatiivseid andmeid esitatakse tavaliselt neljal eri viisil:

a) loetletakse töö tekstis;
b) esitatakse tabelitena;
c) esitatakse arvjoonistena (diagrammid, jaotuspolügoonid ja -kõverad jne);
d) esitatakse valemitena.

Esimest viisi kasutatakse siis, kui esitatavaid arvandmeid on vähe.

Kvantitatiivsed näitajad (arvandmed) esitatakse uurimistöös tavaliselt mitmesuguste tabelitena. Tabelid võimaldavad mingi nähtuse kvantitatiivseid tunnuseid esitada kokkusurutult. Hästi koostatud tabelist võib iga arvu kohta leida tema mõistmiseks vajaliku seletuse, ilma et seda oleks vaja kirjalikus tekstis korrata.

Tuleb arvestada, et tabel ei ole mitte üksnes kogutud andmete esitamise, vaid ka nende tõlgendamise vahend. Ei ole sugugi ükskõik, millises tabelis on arvandmed esitatud. Tabel ei ole esitatavate arvude suhtes kaugeltki ükskõikne. Ta aitab lahti mõtestada temasse kantud arvude olemust ning nendevahelisi seoseid.

Selleks et tabel annaks paremini edasi uurimistööga kogutud informatsiooni, tuleb tunda tabeli koostamise ja kasutamise metoodikat ja teada, missugune tabel sobib ühel või teisel juhul kõige paremini. Igas tabelis on kaks osa: a) tekstiosa ja b) arvuosa.

Tekstiossa kuuluvad tabeli pealkiri, tabeli pea ja arvude selgitamiseks vajalikud sõnalised seletused, täiendavad märkused ja viited.

Arvuossa kuuluvad statistilised arvandmed ja tabeli number. See osa esitatakse tabelis sageli ka arvuväljana.

Tabelid peavad olema lihtsad, et lugeja saaks kiiresti selge ja täpse ülevaate tabelisse märgitud arvandmetest ning nendevahelistest seostest.

Tabelid tuleb võimaluse korral paigutada vahetult pärast tekstis nende kohta käivat viidet. Kui aga tabelid ei ole tekstiga otseselt seotud ja neid on väga palju, võib nad paigutada töö lisasse. Tabel on soovitatav paigutada nii, et seda vaadates ei tuleks lehte pöörata. Kui aga lehe pööramine on möödapääsmatu, tuleb tabel paigutada lehele nii, et selle vaatamiseks tuleb lehte pöörata kellaosuti liikumise suunas.

Kõikidel tabelitel peab olema pealkiri, mis paigutatakse tabeli kohale. Pealkiri peab kokkuvõtlikult iseloomustama tabelisse koondatud andmeid. Pealkiri peaks olema üldine ja hõlmama sisuliselt kõike, mille kohta tabel andmeid sisaldab. Vajaduse korral võib pealkirja all tuua kas väiksema kirjaga või sulgudes täpsustavaid märkusi sisaldava alapealkirja.

Kõik tabelid nummerdatakse alates töö algusest araabia numbritega. Tabeli number kirjutatakse tabeli pealkirjast ühe rea võrra kõrgemale tabeli parema nurga kohale. Kirjutatakse näiteks *Tabel 20.*

Teaduslikus töös kasutatava tabeli näidis on toodud järgmisel leheküljel (vt ptk 9.2. *Tabeli struktuur,* lk 63).

9.2. TABELI STRUKTUUR

Tabeli number Tabel 20

Tabeli pealkiri

Kursuse koosseis

Kursus	I grupp		II grupp	
	tütar-lapsi	poeg-lapsi	tütar-lapsi	poeg-lapsi
1	2	3	4	5
1. A	0	18	11	8
2. B	10	25	15	13
3. C		***Lahter 3/2***		
4. D			***Lahter 4/4***	

Tabeli pea — ***Veergude pealkirjad*** — ***Veergude alapealkirjad*** — ***Veerunumbrid***

Tabeli read ja reanumbrid

Tabeli esiveerg — ***Tabeli veerud (graafid)*** — ***Tabeli lahter 4/5***

Väiksemas tabelis järjekorranumbreid (reanumbreid) tabelisse ei märgita. Vajaduse korral (kui tabelisse mahutatud arvud iseloomustavad paljusid objekte) võib järjekorranumbri märkida tabeli esiveergu iga objekti ette, pannes numbri järele punkti.

Selleks et kergendada uurimistöö tekstis viitamist tabeli veergudele, kasutatakse veergude (graafide) nummerdamist.

Veergude tähendust selgitavad pealkirjad paigutatakse tabeli peasse. Kui tabeli peas on lahtrid kitsad ja sinna ei ole võimalik pealkirju paigutada horisontaalselt, siis kirjutatakse tekst lahtritesse vertikaalselt alt üles. (Tabeli pööramisel kellaosuti liikumise suunas peab tekst jääma horisontaalseks.)

Ridade sisu seletavad pealkirjad paigutatakse tabeli esiveergu.

Kui tabelis toodud parameetritel on ainult üks dimensioon (nt õpilaste vanus aastates), siis viimase lühendatud tähis paigutatakse tabeli üldpealkirja tabeli kohale.

Kui aga arvudel tabeli veergudes on erinevad dimensioonid, märgitakse nende lühendid iga veeru alapealkirjana.

Tabeli ühes veerus toodud parameetrite kümnendkohtade arv peab olema ühesuurune.

Tabeli pead ei ole lubatud diagonaalselt joonida.

Tabeli kohta käiva märkuse võib anda ka tavalise jooncaluse märkusena.

9.3. Tabelite liigitus

Tabelid võib veergude või ridade liigendatuse järgi jagada kolme gruppi: a) liht- ehk loetlustabelid, b) rühmtabelid, c) kombinatsioontabelid.

Lihttabel sisaldab vaadeldavasse kogumisse kuuluvate üksiknähtuste kohta kogutud andmete loetelu.

Näiteks

Tabel 21

Õpilaste arv kursustel

Kursus	Õpilaste arv
1. kursus	132
2. kursus	125

Rühmtabel on selline tabel, kus uuritava kogumi üksikliikmed rühmitatakse mingi ühe tunnuse järgi.

Näiteks

Tabel 22

Kursuste jaotus õpilaste arvu järgi

Õpilaste arv kursusel	Kursuste arv
20 - 100	8
101 - 200	12
201 - 300	6

Kombinatsioontabel sisaldab andmeid rohkem kui ühe tunnuse järgi rühmitatud kogumi kohta.

Näiteks

Tabel 23

Õpilaste koosseis

Grupid	Õpilased								Kokku
	poeglapsed				tütarlapsed				
	kuni 18 a	18-20 a	20-22 a	kokku	kuni 18 a	18-20 a	20-22 a	kokku	
1	2	3	4	5	6	7	8	9	10
1. grupp	42	36	53	131	53	41	48	142	273
2. grupp	...	...	...	...	...	...	...	...	...

Kui tabeli peas anda algul õpilaste vanus, siis, jaotades iga vanusegrupi omakorda poeg- ja tütarlasteks, saame eri vanusega õpilaste soolise jaotumise (eelmises tabelis saadi poeg- ja tütarlaste jaotumine vanuse järgi).

Nagu eespool toodust selgub, pakub kombinatsioontabel liht- ja rühmtabelitega võrreldes tunduvalt paremaid võrdlemis- ja analüüsimisvõimalusi.

Kombinatsioontabelit tuleb aga eristada ***liittabelist***, mis koosneb kahest või enamast ühele sõrestikule kantud liht- või rühmtabelist.

Näiteks

Tabel 24

Õpilaste sugu ja vanus

Grupp	Sugu		Vanus		
	poeglapsed	tütarlapsed	kuni 18 a	18-20 a	21-22 a
1	2	3	4	5	6
1. grupp	131	142	95	77	101
2. grupp	...	...	...	...	...

Nagu näha, moodustavad liittabeli 2. ja 3. veerg iseseisva tabeli, 4., 5. ja 6. veerg aga teise iseseisva tabeli. Nagu selgub, ei saa selle tabeli abil kindlaks teha, kui palju on kuni 18-aastaste õpilaste hulgas poeglapsi, kui palju tütarlapsi. (Kombinatsioontabeli vahendusel on see aga võimalik.)

Kuna eespool toodud tabelis on rühmitamistunnused isoleeritud (andmed kogumi kohta on saadud küll kahe eri rühmitamistunnuse järgi, kuid neid tunnuseid ei ole omavahel kombineeritud), siis ei saa seda tabelit nimetada ka kombinatsioontabeliks.

Kombinatsioontabeli ja liittabeli võrdlus näitab, et esimene on andmete esitamise seisukohalt tunduvalt sisukam. Sellest võib välja lugeda samad andmed, mis liittabelistki, kuid peale selle selgub kombinatsioontabelist ka tütarlaste ja poeglaste jaotumine vanuse järgi või eri vanusega õpilaste sooline jaotumine.

Tabeli valiku otsustab uurija sõltuvalt uurimistöö ja tabeli koostamise eesmärkidest.

Kui tabel on vajalik *esitustabelina*, st see peab andma hea ülevaate uurimistöö tulemustest, on soovitatav valida kombinatsioontabel, sest liittabelis jääb osa uurimistöö tulemusi esitamata. Kombinatsioontabeli vahendusel on ka uurimistöö tulemusi võimalik sügavamalt analüüsida.

Uurimistöö käigus tehtavate vaatlusandmete ülesmärkimiseks, nende kokkuvõtuks ja arvutil töötlemiseks (nn *töötabeliteks*) sobivad sageli paremini liht-, rühm- ja liittabelid.

Sageli on andmete kogumiseks, nende kokkuvõtmiseks, läbitöötamiseks ja esitamiseks väga otstarbekohane kombinatsioonitabeli eriliik, nn ***risttabel***.

Risttabel võimaldab uurida kombineeritavate tunnuste vahelisi seoseid (nt ladina ruut, korrelatsioonitabelid jm). Risttabelid on omal kohal ka matemaatiliste meetodite rakendamise korral. Risttabelisse sobivad ükskõik millise kombinatsioontabeli andmed.

Näiteks kombinatsioontabelis (vt tabel 23) toodud õpilaste jaotumise soo ja vanuse järgi võib risttabelis esitada alljärgneva näite kohaselt.

Tabel 25

Õpilaste koosseis

Sugu	Vanus			Kokku
	kuni 18 a	18-20 a	20-22 a	
Poeglapsed	42	36	53	131
Tütarlapsed	53	41	48	142
Kokku	95	77	101	273

Nagu selgub, annab risttabel vastuse ka sellele, kui palju on igas vanuserühmas poeg- ja tütarlapsi kokku (konbinatsioontabelis nõudis see lisaarvutust).

Risttabelis võib veerge olla muidugi niisama palju kui ridu (vt ladina ja ladina-kreeka ruudud, I osa lk 48).

Kui näiteks uurijat huvitab küsimus, kui palju õpilasi on ühest või teisest maakonnast läinud mingisse õppeasutusse ja kui palju on maakond saanud lõpetanuid sellest õppeasutusest tagasi (eeldame, et kõik sisseastunud kooli ka lõpetasid), võib risttabeli koostada järgmise näite kohaselt.

Tabel 26

Õppima läinute ja lõpetajate jaotuvus maakonniti

Maakond	Maakond sai lõpetajaid				
	A	B	C	Muud maakonnad	Kokku
A	20	10	15	40	85
B	31	6	12	15	64
C	15	2	1	5	23
Muud maakonnad	40	6	7	10	63
Kokku	106	24	35	70	235

Tabeli ridade kokkuvõttest selgub, kui palju on iga maakond õpilasi õppeasutusse andnud, veergude kokkuvõttest aga selgub, kui palju on maakond lõpetanuid tagasi saanud.

Nagu näha, andis õpilasi kõige rohkem maakond A ja sai neid ka kõige rohkem tagasi. Maakond B andis aga 64 õpilast, lõpetajaid sai aga ligi kolm korda vähem. Maakond C andis 23 õpilast, sai aga lõpetajaid rohkem

Millistesse teistesse maakondadesse on õpilased läinud, selgub iga maakonna realt. Millistest maakondadest aga lõpetajaid saadi, seda näeb iga maakonna veerust. Kui palju mingi maakond oma õpilastest tagasi sai, see selgub tabeli vasakust ülanurgast paremasse alanurka kulgeval diagonaalil asuvatest lahtritest.

Näitest järeldub, et risttabel võimaldab selgitada terve rea huvitavaid probleeme õppima läinute ja lõpetajate jaotuvuse kohta maakonniti.

Kokkuvõtteks võib märkida, et tabelid on teadusliku töö üks oluline koostisosa. Nad võimaldavad arvulisi näitajaid esitada kompaktselt, süstematiseeritult ja ülevaatlikult. Tabelis toodud andmeid on hea omavahel võrrelda ja analüüsida. Uurimistöös ei tohi aga ka tabelitega liialdada ega anda nende vahendusel kogu uurimistöö sisu. Töö tekstis tuleb anda vaid need tabelid, mis tutvustavad teadustöö lõpptulemusi ja on aluseks teaduslike järelduste tegemisel. Tekstiga vähem seotud, tööks vajalikke lähtematerjale pakkuvad ning nende läbitöötamisega seotud tabelid antakse uurimistöö lisana.

Arvandmete analüüsimisel ja esitamisel on teaduslikus uurimistöös tabelitega rööbiti soovitatav kasutada ka sagedusjaotuste graafilist kujutamist.

Olulisemat

1. **Uurimistöö andmed esitatakse kirjaliku tekstina, tabelitena, arvjoonistena või valemitena.**
2. **Tabelites on tekstiosa (pealkiri, veergude ja ridade pealkirjad jne) ja arvosa (statistilised arvandmed).**
3. **Tabelil peab olema number ja pealkiri.**
4. **Tabel paigutatakse tavaliselt kohe pärast tekstisse tabeli kohta tehtud viidet.**
5. **Tabelid liigitatakse liht-, liit-, kombinatsioon- ja risttabeliteks.**
6. **Eristatakse ka esitustabelit (annab ülevaate uurimistöö tulemustest) ja töötabelit (vaatlusandmete ülesmärkimiseks, nende kokkuvõtuks ja arvutil töötlemiseks).**

10. GRAAFILISTE MEETODITE KASUTAMINE UURIMUSTE KOKKUVÕTUL

10.1. ARVJOONISTE KASUTAMISE EESMÄRGID JA NÕUDED

Numbrilisi andmeid on sageli otstarbekas esitada ja analüüsida mitmesuguste arvjoonistena (graafikutena).

Arvjoonised võimaldavad:

1) esitada uuritava nähtuse olemust näitlikult;
2) soodustada uuritavate nähtuste analüüsimist;
3) populariseerida uurimistulemusi.

Graafikud esitavad uurimistöö tulemusi ülevaatlikult ja piltlikult, nad köidavad tavaliselt rohkem tähelepanu kui tabelid. Graafikute vahendusel on võimalik numbriliselt väljendatud fakte ja nendevahelisi seoseid, mis on sageli küllaltki abstraktsed, esitada konkreetsemas ja arusaadavamas vormis. Läbimõeldult esitatud graafikute vahendusel on võimalik kiiresti ja kergesti märgata ka nähtustes valitsevaid seaduspärasusi, mida tabelite abil on sageli raske teha. Tähelepanu tuleb juhtida ka sellele, et graafikud jäävad paremini meelde ning hiljem meenuvad nendes esitatud seosed kiiremini, võrreldes tabelites esitatuga.

Uurija peab hästi tundma graafikute koostamise ja analüüsi metoodikat. Kuigi ta võib-olla ise neid ei joonista, peab ta valima graafilise esitamise viisi (graafiku tüübi) ning andma selle eskiisi.

Graafikute puhul tuleb silmas pidada mõningaid üldnõudeid.

a) Graafiliselt ei ole soovitatav anda mitte kõiki uurimistulemisi, vaid ainult kokkuvõtte või analüüsi põhiresultaadid, millele tahetakse erilist tähelepanu osutada. Graafiku kõige olulisem nõue on selle näitlikkus, mistõttu graafik ei tohi olla üle koormatud liigsete joonte ega kujunditega.

b) Igale graafikule tuleb anda sobivad mõõtmed. Need peavad võimaldama graafikut joonestada ja lugeda. Kui algjoonisest tahetakse teha reproduktsioone (fotosid), soovitatakse algjoonise joonmõõtmed valida 2-8 korda suuremad reproduktsiooni joonmõõtmetest (mõõt 1:2 kuni 1:8).

c) Joonestamisel tuleb silmas pidada graafiku laiuse ja kõrguse sobivaid vahekordi. Soovitatakse, et graafiku lühem külg oleks $\sqrt{2}$ (1,414) korda lühem selle pikemast küljest.

d) Graafikute kujundus ja paigutus peab soodustama nende lugemist. Ühel leheküljel ei tohiks olla üle ühe graafiku, kusjuures selle suurus ei tohiks ületada töö lehekülje formaadi suurust. Graafiku peaks asetama teksti kohe pärast viidet graafikule.

e) Graafiku peab kujundama selliselt, et nähtuse olulisemad faktid ja seosed oleksid vähemtähtsatest selgesti eraldatavad. Olulist osa etendab seejuures eri värvuste, tähistusviiside ja kirja kasutamine. Kunstiliselt hästi kujundatud graafikud soodustavad nende loetavust.

Sagedusjaotuse esitamiseks kasutatakse peamiselt kaht esitusviisi: *sageduspolügooni* ja *histogrammi*, mõnikord ka *kumulatiivset sagedusgraafikut* ja *kumulatiivse sageduse protsendilist kõverat*.

Mõõtarve, mis näitavad mingi nähtuse dünaamikat (muutusi õpilaste arvus, vanuses, töövõimes jm) esitatakse joongraafikutena, joon- või riba- ehk tulpdiagrammidena ja sektordiagrammidena.

Vähem kasutatakse uurimistöödes figuurdiagramme, kartogramme ja kartodiagramme.

10.2. Sageduspolügoon

Tabel 27

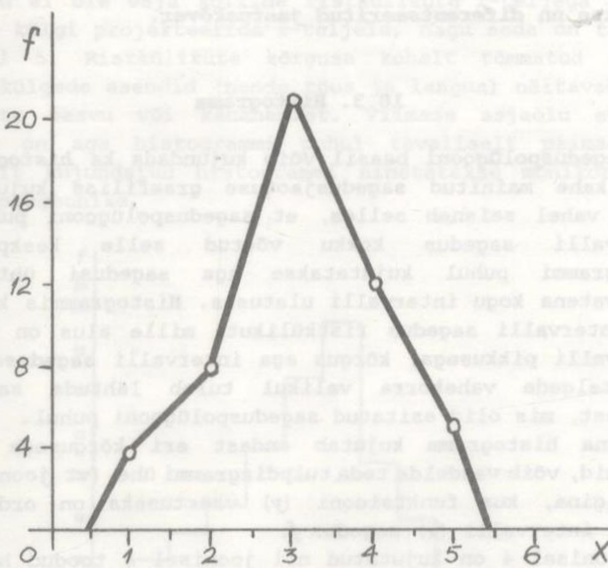
Sagedustabel

Hinne x	Sagedus f
1	4
2	18
3	21
4	12
5	5

Sageduspolügooni ehitamiseks tuleb uurimistulemused kanda eelkõige sagedustabelisse. Nii võib 50 õpilase kontrolltööhinded jaotada järgmiselt (vt tabel 27).

Sageduspolügooni joonestades tuleb anda õige vahekord eelkõige tema abstsissteljele (x-teljele) ja ordinaatteljele (y-teljele). Liiga pikk abstsisstelg venitab polügooni välja, liiga lühike surub selle kokku. Praktikas valitakse abstsisstelje ja ordinaattelje ühikud selliselt, et polügooni kõrgus moodustaks 60-80% selle laiusest. Siis tõmmatakse abstsisstelg ning selle algusest vartikaalselt üles ordinaattelg.

x-telg on antud juhul vaja jaotada vähemalt seitsmeks ühikuks. Kui valida iga ühiku pikkuseks 1 cm, oleks x-telje pikkus 7 cm, y-telje pikkuseks aga 60-80% sellest, s.o 4,2 kuni 5,6 cm. Võttes y-teljeks 5,6 cm ja jaotades selle samuti 1 cm pikkusteks lõikudeks, saame y-teljele 5 lõiku. Iga lõigu vahe tähistaks siis nelja sagedusühikut (sageduspolügooni kuju antud juhul vt jooniselt 3).



Joonis 3. Sageduspolügoon

Kui kõikide intervallide keskkohale on punktid kantud, ühendatakse need omavahel lühikeste joontega, moodustades sellega sageduspolügooni. Selleks et viia sageduspolügooni otsad *x*-teljeni, lisatakse *x*-teljele lisaks sagedustabelis olevatele intervallidele veel kaks täiendavat intervalli. Üks nendest paigutatakse joonise algusse, teine lõppu. Nendes intervallides on sagedus 0. Järelikult alustatakse sageduspolügooni joonestamist vähemalt pool intervalli enne esimest ja lõpetatakse pool intervalli pärast viimast intervalli abstsissteljel.

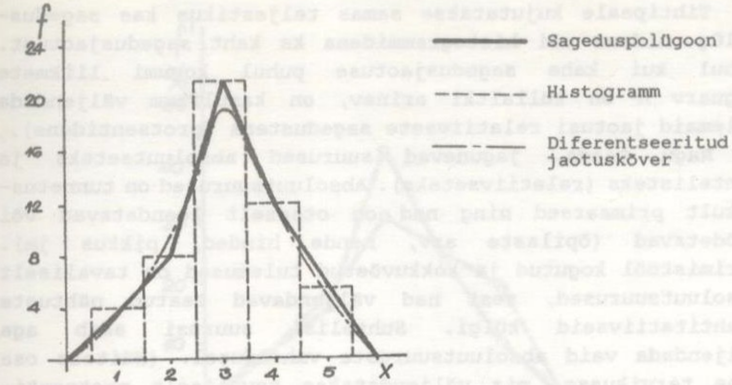
Parema ja täpsema pildi uurimistöö kokkuvõtuandmete seaduspärasusest annab nn tasandatud sageduspolügoon ehk jaotuskõver. Selle täpseks joonestamiseks võetakse kasutusele rida jooksvaid "keskmisi", mille alusel määratakse uued või kohandatud intervallide sagedused. Käesolevas töös seda arvutuskäiku lähemalt ei vaadelda. Uurimistöös on tihti küllaldane, kui tõmmata jaotuskõver silma järgi, siludes polügooni teravaid nurki (vt joonis 4). Sel juhul saadakse nn ***diferentseeritud jaotuskõver***.

10.3. Histogramm

Sageduspolügooni baasil võib kujundada ka ***histogrammi***. Vahe kahe mainitud sagedusjaotuse graafilise kujutamisviisi vahel seisneb selles, et sageduspolügooni puhul on intervalli sagedus kokku võetud selle keskpunkti, histogrammi puhul kujutatakse aga sagedusi ühtlaselt jaotuvatena kogu intervalli ulatuses. Histogrammis kujutab iga intervalli sagedus ristkülikut, mille alus on võrdne intervalli pikkusega, kõrgus aga intervalli sagedusega. *x*- ja *y*-telgede vahekorra valikul tuleb lähtuda samadest nõuetest, mis olid esitatud sageduspolügooni puhul.

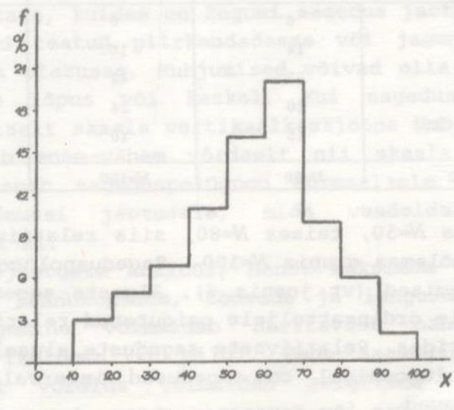
Kuna histogramm kujutab endast eri kõrgusega ristkülikuid, võib vaadelda teda tulpdiagrammi ühe (vt joonis 14) eriliigina, kus funktsiooni (y) väärtuseks on ordinaatteljel intervalli (x) sagedus f.

Joonisel 4 on kujutatud nii joonisel 3 toodud hinnete sageduspolügoon, sellele joonestatud histogramm kui ka diferentseeritud jaotuskõver.



Joonis 4. Sagedusjaotused

Iga intervalli kujutab histogrammis eraldi ristkülik, mistõttu ei ole vaja kõikide ristkülikute *x*-teljega risti olevaid külgi projekteerida *x*-teljele, nagu seda on tehtud joonisel 5. Ristkülikute kõrguse kohalt tõmmatud ristküliku külgede asendid (nende tõus ja langus) näitavad kas sageduste kasvu või kahanemist. Viimase asjaolu esiletoomine on aga histogrammi puhul tavaliselt primaarne. Selliselt kujundatud histogrammi nimetatakse mõnikord ka trepp-polügooniks.



Joonis 5. Histogramm

Tihtipeale kujutatakse samas teljestikus kas sageduspolügoonidena või histogrammidena ka kaht sagedusjaotust. Juhul kui kahe sagedusjaotuse puhul kogumi liikmete koguarv N on küllaltki erinev, on kasulikum väljendada mõlemaid jaotusi relatiivsete sagedustena (protsentidena).

Nagu teada, jagunevad suurused absoluutseteks ja suhtelisteks (relatiivseteks). Absoluutsuurused on tunnetuslikult primaarsed ning nad on otseselt loendatavad või mõõdetavad (õpilaste arv, nende hinded, pikkus jm). Uurimistööl kogutud ja kokkuvõetud tulemused on tavaliselt absoluutsuurused, sest nad väljendavad teatud nähtuste kvantitatiivseid külgi. Suhtelisi suurusi saab aga väljendada vaid absoluutsuuruste vahendusel. (Näiteks osa suhe tervikusse, mis väljendatakse tavaliselt protsentides.) Absoluut- ja suhtarvud väljendavad absoluutsete ja suhteliste suuruste väärtusi.

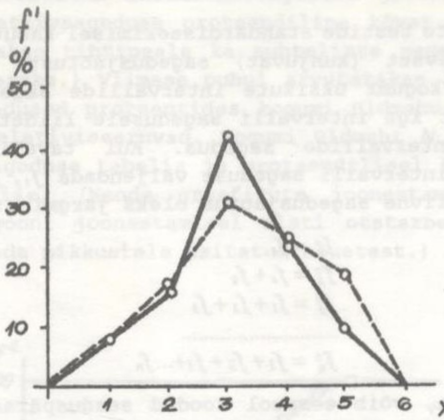
Oletame, et hinnete jaotuvus gruppides A'ja B vastab tabelis 28 toodule:

Tabel 28

Õpilaste hinnete jaotuvus

Hinne x	Grupi A absoluut-sagedus f^1	Grupi B absoluut-sagedus f^2	Grupi A relatiivne sagedus f_1'	Grupi B relatiivne sagedus f_2'
1	4	6	8	8
2	8	14	16	17
3	21	25	42	31
4	12	20	24	25
5	5	15	10	19
	N=50	N=80	N=100	N=100

Kui grupis N=50, teises N=80, siis relatiivsete sageduste puhul mõlemas grupis N=100. Sageduspolügoonid oleksid siis järgmised (vt joonis 6). Algsete sagedusjaotuste asemel on selle ordinaatteljele paigutatud relatiivne sagedus f protsentides. Relatiivsete sageduste alusel joonestatud sageduspolügoonidel on sagedused intervallide kaupa hästi võrreldavad.



Joonis 6. Relatiivsed sagedusjaotused

Sageduspolügooni ja histogrammi kasutamisel tuleb arvestada, et histogramm on sageduspolügoonist täpsem, sest ta peegeldab selgesti piiratud pinna piire. Kui aga uurija soovib samas teljestikus võrrelda kaht või enamat graafikut, siis on seda sobivam teha sageduspolügooni vahendusel. Histogrammi puhul võivad vertikaalsed ja horisontaalsed jooned sageli kattuda.

Mõlemad eespool toodud graafilised meetodid võimaldavad kujutada, kuidas on kogumi sagedus jaotunud: kas see on kuhjunud teatud piirkondadesse või jaguneb ühtlaselt kogu skaala ulatuses. Kuhjumised võivad olla skaala alguses, selle lõpus või keskel. Kui sagedused paiknevad sümmeetriliselt skaala vertikaalkeskjoone ümber ning sagedusi esineb enam-vähem võrdselt nii skaala alguses kui lõpus, läheneb sageduspolügoon normaalsele sagedusjaotusele e Gaussi jaotusele, mida vaadeldakse lähemalt ptk 4.2, lk 11.

Sagedusjaotuste analüüs, nende maksimum- ja miinimumväärtuste, käänukohtade, tõusude ja languse iseloomu jm kindlakstegemine võimaldab uuritavast nähtusest saada küllaltki iseloomuliku pildi. Kahe sageduspolügooni või histogrammi võrdlus võimaldab selgitada eri nähtuste omavahelist seost, nende sarnasust või eripära.

10.4. Kumulatiivsageduse graafik

Igasuguste testide standardiseerimisel kasutatakse sageli kumulatiivset (kuhjuvat) sagedusjaotust. Selle puhul vaadeldakse kogumi üksikute intervallide sagedust kumulatiivselt, st iga intervalli sagedusele liidetakse kõikide eelnevate intervallide sagedus. Kui tavalise sagedusjaotuse iga intervalli sagedust väljendada $f_1, f_2, f_3, \ldots, f_n$, siis kumulatiivne sagedusjaotus oleks järgmine:

$$
\begin{aligned}
f_1^c &= f_1 \\
f_2^c &= f_1 + f_2 \\
f_3^c &= f_1 + f_2 + f_3 \\
&\cdots\cdots\cdots \\
f_n^c &= f_1 + f_2 + f_3 + \ldots f_n
\end{aligned}
$$

Nagu näha, võib eespool toodud seaduspärasuste põhjal koostatud tabeli abil selgitada, mitu juhtumit hõlmati iga intervalli lõpuni.

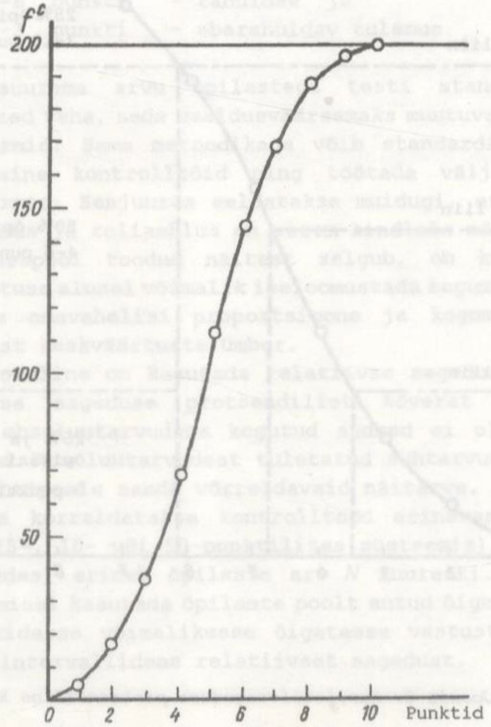
Näiteks korraldati test kahesaja õpilase liigutustäpsuse uurimiseks. Hindamisel kasutati 10-punktilist skaalat (märklauda). Igal õpilasel lasti lüüa vasaraga, mille külge oli kinnitatud teravik, alasil asuva väikese märklaua pihta üks löök. Tulemused on toodud tabelis (vt tabel 29) tavalises ja kumulatiivsageduse jaotuses ning kumulatiivsageduse protsendiliste andmetena.

Tabel 29

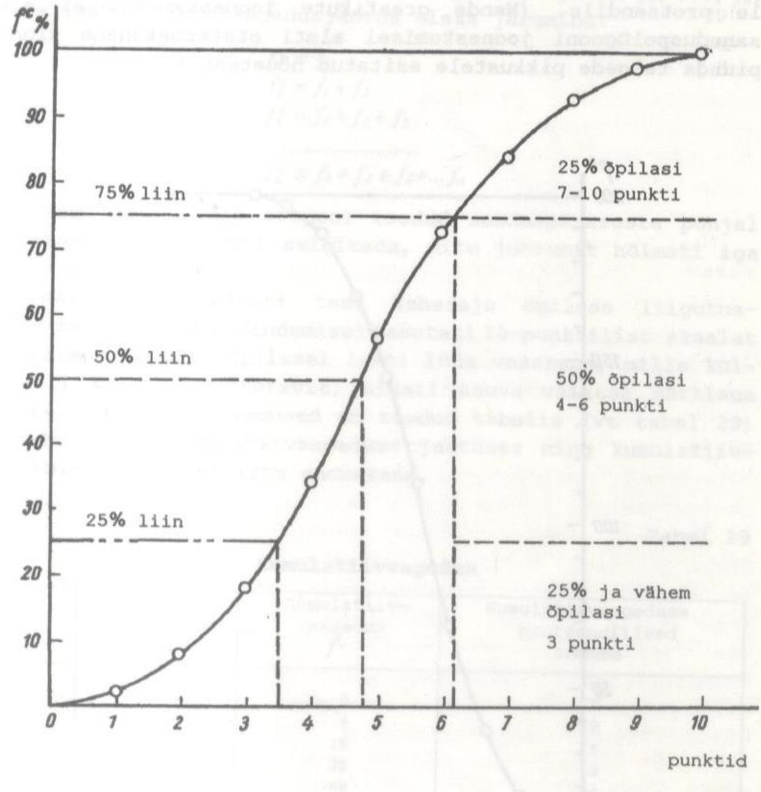
Kumulatiivsagedus

x_i	Sagedus f	Kumulatiivsagedus f^c	Kumulatiivsageduse protsendilised andmed
0	0	0	0
1	4	4	2
2	12	16	8
3	20	36	18
4	32	68	34
5	44	112	56
6	32	144	72
7	24	168	84
8	20	188	94
9	8	196	98
10	4	200	100
		N=200	N=100 %

Graafiliselt on joonisel 7 kujutatud eespool toodud tabeli alusel koostatud kumulatiivsageduse graafik ja joonisel 8 kumulatiivsageduse protsendiline kõver. (Seda kõverat nimetatakse tihtipeale ka suhteliste sageduste järgusummade kõveraks.) Viimase puhul arvutatakse kumulatiivsed esinemissagedused protsentides kogumi üldmahu N suhtes ja seega nad relativiseeruvad. Kogumi üldmaht N vastab seega kumulatiivsageduse tabelis ja protsendilisel kõveral sajale protsendile. (Nende graafikute joonestamisel ei ole sageduspolügooni joonestamisel alati otstarbekohane kinni pidada telgede pikkustele esitatud nõuetest.)



Joonis 7. Kumulatiivsageduse graafik



Joonis 8. Kumulatiivsageduse protsendiline kõver

Praktikas leiab kasutamist peamiselt kumulatiivsageduse protsendiline kõver. Sellele võib tõmmata 25%, 50%, 75% tähistavad jooned (liinid). Viimased jaotavad kõvera neljaks osaks, millest igaühes on võrdne arv liikmeid. Kahte keskmisse ossa langeb 50%, kummassegi äärmisse 25% juhte. Kui lugeda rahuldavaks neid tulemusi, mis langesid kõvera kahte keskmisse ossa, heaks neid, mis langesid ülemisse ossa, ebarahuldavaks tulemusi, mis langesid alumisse, saamegi teiste samavanuste õpilaste löögitäpsuse mingil määral usaldatavad hindamisnormid. Need oleksid järgmised:

7-10 punkti - hea tulemus,
4-6 punkti - rahuldav ja
0-3 punkti - ebarahuldav tulemus.

Mida suurema arvu õpilastega testi standardiseerimiseks katsed teha, seda usaldusväärsemaks muutuvad tulemuste hindamisnormid. Sama metoodikaga võib standardiseerida ka mõne õppeaine kontrolltöid ning töötada välja vastavad hindamisnormid. Seejuures eeldatakse muidugi, et kontrolltöö valiidsus ja reliaablus on varem kindlaks määratud.

Nagu eespool toodud näitest selgub, on kumulatiivse sagedusjaotuse alusel võimalik iseloomustada kogumi, s.o üksikute osade omavahelisi proportsioone ja kogumi liikmete varieeruvust keskväärtuste ümber.

Eriti oluline on kasutada relatiivse sagedusjaotuse ja kumulatiivse sageduse protsendilist kõverat siis, kui esialgsed absoluutarvudena kogutud andmed ei ole omavahel võrreldavad. Absoluutarvudest tuletatud suhtarvud võimaldavad aga tihtipeale saada võrreldavaid näitarve.

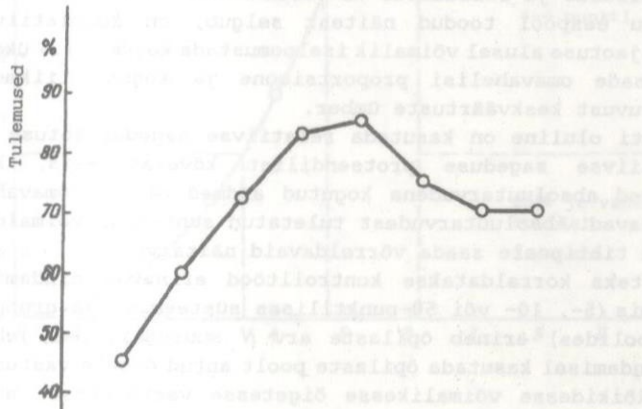
Näiteks korraldatakse kontrolltööd erinevas hindamissüsteemis (5-, 10- või 50-punktilises süsteemis) ja gruppides (koolides) erineb õpilaste arv N suuresti. Sel juhul võib hindamisel kasutada õpilaste poolt antud õigete vastuste suhet kõikidesse võimalikesse õigetesse vastustesse ning sageduste intervallidena relatiivset sagedust.

10.5. JOONGRAAFIK

Joongraafiku vahendusel antakse tavaliselt edasi mõnedes mõõtarvudes toimunud muutusi. Käesolevas töös (vt joonis 9) on kujutatud õppeaasta jooksul ühes grupis korraldatud kontrolltööde keskmiste suhteliste hinnete muutust. (Iga töö hindamisel arvestati, mitu protsenti küsimustest suutis iga õpilane vastata ning nende hinnete alusel leiti grupi keskmine suhteline hinne.)

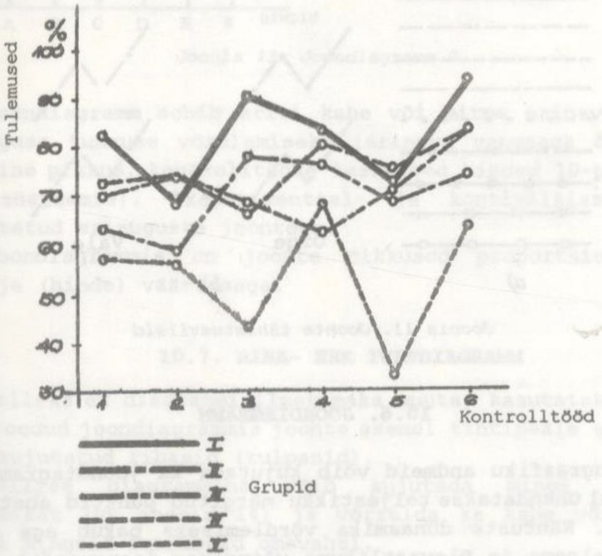
Nagu graafikust selgub, tõusis keskmine suhteline hinne kuni viienda tööni, hiljem tuli uuesti teatud langus.

Joongraafiku puhul toimitakse analoogselt sageduspolügooni joonestamisega. Algul tõmmatakse teljestik, siis paigutatakse abstsissteljele kontrolltööde hinded, ordinaatteljele aga vastavad hinded (suhtelised hinded). Edasi kantakse iga töö kohale punktiga vastava töö keskmine hinne. Hiljem ühendatakse punktid murdjoonega.



Joonis 9. Joongraafik

Kahe või mitme mõõtarvude rea dünaamika võrdlemiseks kantakse need tavaliselt (juhul kui arvridu ei ole üle 5-6) ühte teljestikku. Iga arvrea muutus märgitakse erineva murdjoonega (vt joonis 10).

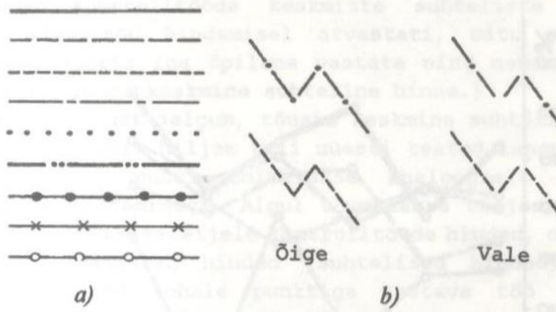


Joonis 10. Joongraafikud

Kui mõõtarvud on igas reas ühesugustes ühikutes, võib väärtuste graafikusse märkimiseks kasutada absoluutarve. Kui mõõtarvud on aga eri ühikutes, tuleb kasutada arvude relatiivset väärtust, mis tavaliselt väljendatakse protsentides, harvem lihtsa suhtarvuna. (Viimane saadakse näiteks iga kontrolltöö puhul saavutatud punktide arvu ja maksimaalselt saada võiva punktide arvu suhtena.)

Selleks et joongraafikutel mõõtarvude ridu paremini üksteisest eristada, kasutatakse nende tähistamiseks eri tähistusviise. Kõige parem oleks neid muidugi eristada siis, kui nad oleksid märgitud erinevate kontrastsete värvidega.

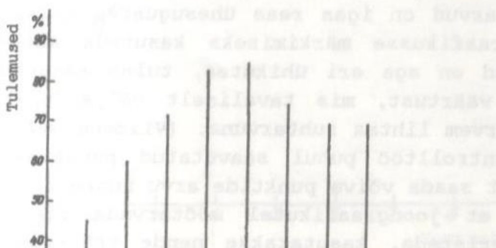
Kuna teadustöö joonistest tuleb sageli teha koopiaid, joonestatakse graafikud enamasti must-valged, kasutades joonte märkimiseks eri tähistusi. Mõningad näited joonte võimalikest tähistusviisidest on toodud joonisel 11a). Samal joonisel on ka näidatud, kuidas jooni on soovitatav murda (b).



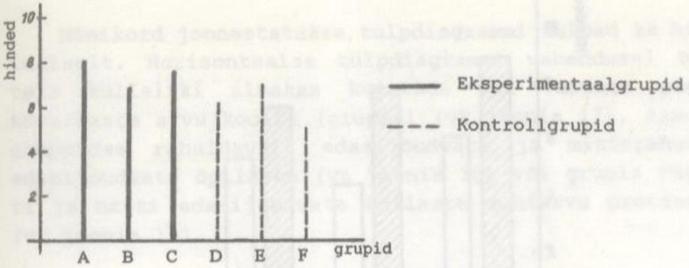
Joonis 11. Joonte tähistusviisid

10.6. JOONDIAGRAMM

Joongraafiku andmeid võib kujutada ka joondiagrammiga. Sel juhul ühendatakse teljestikku märgitud punktid abstsissteljega. Nähtuste dünaamika võrdlemiseks pakub aga mõnevõrra selgema ja ülevaatlikuma võimaluse joongraafik. Mitme nähtuse dünaamika võrdlemine ühel joondiagrammil, eriti siis, kui neid on rohkem kui 2-3, muudab graafiku kaunis kirjuks.



Joonis 12. Joondiagramm 1



Joonis 13. Joondiagramm 2

Joondiagramm sobib eriti kahe või mitme erineva grupi ühesuguse tunnuse võrdlemiseks (erineva vanusega õpilaste keskmine pikkus, kontrolltööde keskmised hinded 10-pallilises süsteemis). Eksperimentaal- ja kontrollklassid on tähistatud erisuguste joontega.

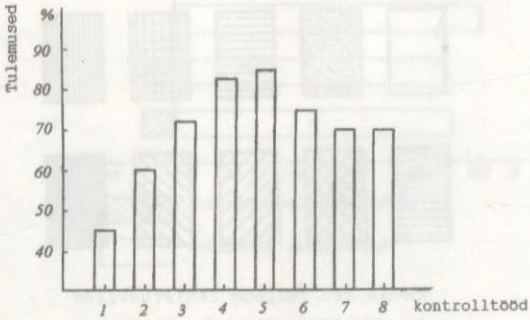
Joondiagrammis on joonte pikkused proportsionaalsed muutuja (hinde) väärtusega.

10.7. RIBA- EHK TULPDIAGRAMM

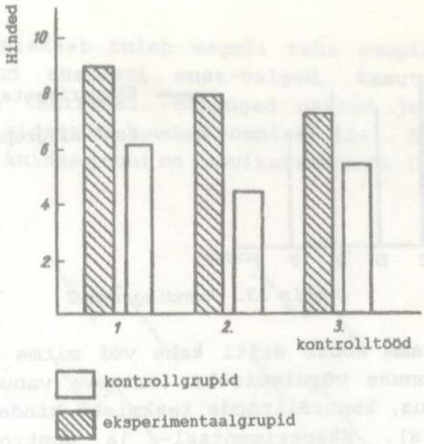
Selleks et diagrammi ilmekamaks muuta, kasutatakse eespool toodud joondiagrammis joonte asemel tihtipeale erisuguselt kujutatud ribasid (tulpasid).

Sellise diagrammina võib kujutada mingi nähtuse dünaamikat (vt joonis 14) ja võrrelda ka kahe või enama kogumi mingeid tunnuseid omavahel.

Joonisel 15 on kujutatud eksperimentaal- ja kontrollgruppides tehtud kolme eri kontrolltöö keskmisi hindeid.



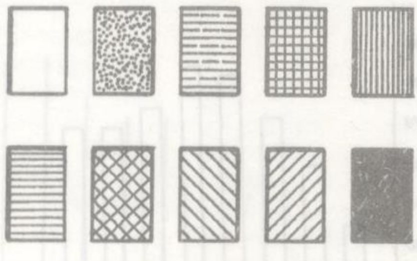
Joonis 14. Tulpdiagramm 1



Joonis 15. Tulpdiagramm 2

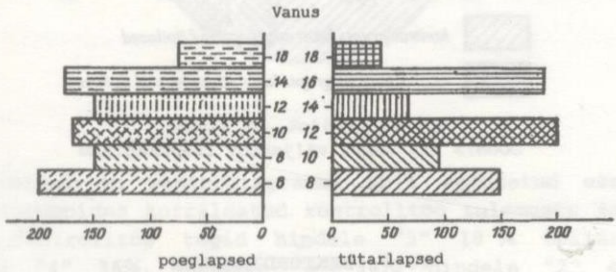
Tulpdiagrammi võib joonestada selliselt, et tulbad külgnevad vahetult (kahe kõrvuti seisva tulba küljed märgitakse ühe joonega) (vt joonis 17), või jäetakse tulpade parema eristamise, peamiselt nende pikkuse parema esiletoomise eesmärgil tulpade vahele ruumi (joonis 18).

Nagu juba varem märgitud, on tulpdiagrammi eriliik histogramm. Sageli kantakse tulpdiagrammi teljestikku 6-7, mõnikord isegi rohkem tulpa. Selleks et neid paremini üksteisest eraldada, kasutatakse tulpade erinevaid tähistusviise. Mõningad tähistusviisid on kujutatud joonisel 16.

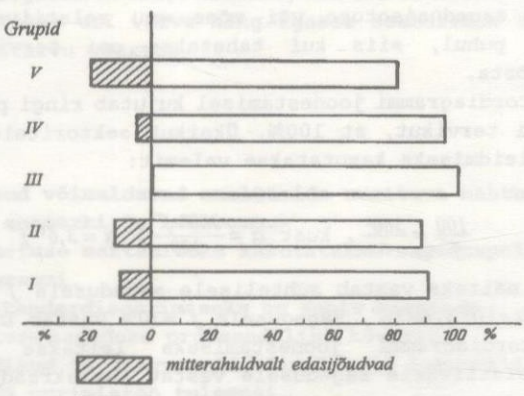


Joonis 16. Tulpade tähistusviise

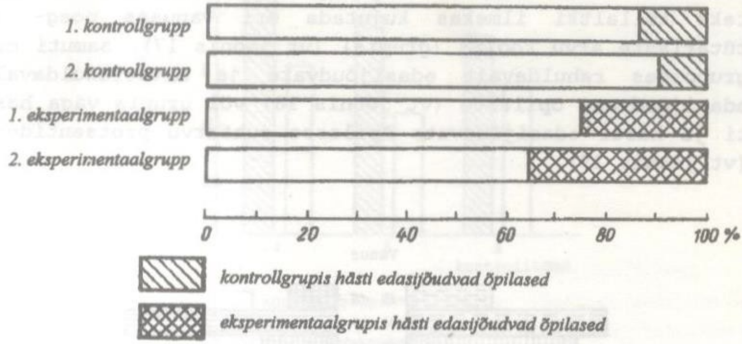
Mõnikord joonestatakse tulpdiagrammi tulbad ka horisontaalselt. Horisontaalse tulpdiagrammi vahendusel on näiteks küllaltki ilmekas kujutada eri vanuses poeg- ja tütarlaste arvu koolis (grupis) (vt joonis 17), samuti eri gruppides rahuldavalt edasijõudvate ja mitterahuldavalt edasijõudvate õpilaste (vt joonis 18) või grupis väga hästi ja hästi edasijõudvate õpilaste suhtarvu protsentides (vt joonis 19).



Joonis 17. Horisontaalne tulpdiagramm



Joonis 18. Õpilaste edasijõudmine



Joonis 19. Suhet väljendav tulpdiagramm

10.8. SEKTORDIAGRAMM

Mingi kogumi jagunemist osadeks on tihtipeale näitlik kujutada ringi kui terviku jagamise teel osadeks, sektoriteks. Sektordiagrammi on eriti sobiv kasutada protsentuaalse sagedusjaotuse või mõne muu relatiivse sagedusjaotuse puhul, siis kui tahetakse osi tervikust eriti esile tõsta.

Sektordiagrammi joonestamisel kujutab ringi pindala nähtust kui tervikut, st 100%. Üksikut sektoritele vastavate kaarte leidmiseks kasutatakse valemit:

$$\frac{100}{f'} = \frac{360^\circ}{\alpha}, \text{ kust } \alpha = \frac{360^\circ \cdot f'}{100}; \quad \alpha = 3{,}6^\circ \cdot f'$$

Nii näiteks vastab suhtelisele sagedusele f'=10%, kaar (nurk) α=10·3,6=36°, sagedusele f'=30% vastab α=108°.

Sektordiagrammi joonestamiseks leitakse kõigepealt igale relatiivsele sagedusele vastav kaarekraadide arv α. Malli vahendusel mõõdetakse vastav kaar ringjoonele või nurk ringi tsentri juures. Edasi joonestatakse välja vastavad sektorid ja vajaduse korral viirutatakse need eri joonestikuga.

Joonis 20. Sektordiagramm

Joonisel on sektordiagrammi abil kujutatud eksperimentaalgruppides korraldatud kontrolltöö tulemuste jagunemist. Kontrolltöö tegid hindele "5" 18 % õpilastest, hindele "4" 36%, hindele "3" 40%, hindele "2" 4% ja hindele "1" 2% õpilastest.

Sektoritele vastavad kaared oleksid siis: 64,8°; 129,6°; 144,0°; 14,4°; 7,2°.

Sektordiagramm on eriti ilmekas, kui eri sektorite märkimiseks kasutatakse eri värve ning igasse sektorisse kirjutatakse ka suhtarvu väärtus.

Olulisemat

1. **Arvjoonised võimaldavad analüüsida uuritava nähtuse olemust, eesmärki ja iseloomu.**
2. **Sagedusjaotuse esitamiseks kasutatakse sageduspolügooni ja histogrammi.**
3. **Testide standardiseerimiseks on sobiv kasutada kumulatiivse sageduse protsendilist kõverat.**
4. **Joongraafikud ja diagrammid võimaldavad lakooniliselt edasi anda uurimistöö tulemusi.**
5. **Riba- ehk tulpdiagrammid võimaldavad graafiliselt esitatavat materjali muuta ilmekamaks.**
6. **Jagunemist osadeks on näitlik kujutada sektordiagrammina.**

KIRJANDUS

1. Kõverjalg, A. (1987). Pedagoogiliste uurimistööde usaldatavus - probleem või mitte? - *Nõukogude Kool*, nr 7, lk 32-37.

2. Kõverjalg, A. (1990). Statistika korrektne kasutamine pedagoogikas. - *Haridus*, nr 5, lk 27-30 ja nr 6, lk 25-27.

3. Kõverjalg, A. (1990). Uurimistulemuste statistilisest võrdlemisest. - *Haridus*, nr 9, lk 29-31.

4. Kõverjalg, A. (1990). Kogumite tunnuste vaheline seos. Korrelatsioon. - *Haridus*, nr 10, lk 35-38.

5. Kõverjalg, A. (1990). Uurimustulemuste võrdlemine mitte-parameetriliste meetoditega. - *Haridus*, nr 12, lk 28-30.

6. Kõverjalg, A. (1992). Tähelepanekuid haridusuuringute hetkeseisust. - *Haridus*, nr 4, lk 18-21.

7. Mereste, U., Saarepera, M. (1983). Arvjoonised. Tln, Valgus, 247 lk.

8. Mereste, U. (1975). Statistika üldteooria. Tln, Valgus, 496 lk.

9. Garrelt, H. (1960). *Statistics in psychology and Education*. New York, London, Toronto.

10. Glauss, G., Epner, H. (1970). *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*. Berlin, 367 S.

11. Stanley, Y.C., Glass G.V. (1970). *Statistical Methods in Education and psychology*. New Yersey. (Tõlge vene keelde, 1986.)

12. Кыверялг А. (1971). Вопросы методики педагогических исследований. Часть 1-2. Тлн, Валгус, 134 и 228 с.

13. Кыверялг А. (1980). Методы исследования в профессиональной педагогике. Тлн, Валгус, 334 с.

14. Суходольский Г. В. (1972). Основы математической статистики для психологов. Л, Изд-во ЛГУ, 429 с.

15. Фресс П., Пиаже Ж. (1966). Экспериментальная психология. М, Прогресс, 420 с.

Eesti Riigikaitse Akadeemia kirjastus. EE0020 Tallinn, Kase 61.
AS *Trükkal*. 12.1994. Tellimus nr 637.

Trükkal