

TARTU RIIKLIK ÜLIKOO

U. MERESTE

**KESKMISED JA
VARIATSIOONI-
NÄITARVUD**

TARTU  1965



1970298

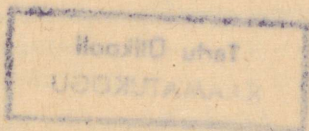
A-26405₁₁

TARTU RIIKLIK ÜLIKOO

Rahanduse ja krediidi kateeder

U. MERESTE

KESKMISED JA VARIATSIOONI- NÄITARVUD



TARTU 1965

Тартуский государственный университет
ЭССР, г. Тарту, ул. Оликооли, 18

У. Мересте
СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАНТНОСТИ
На эстонском языке

2

Tartu Ülikooli
RAAMATUKOGU

190298

S i s u k o r d .

1. Keskmete kasutamise üldaluseid	8
1.1. Keskmete mõiste	8
1.1.1. Keskmine kvantitatiivse üldistusena . .	8
1.1.2. Suurte arvude seadus ja keskmised . . .	9
1.2. A. Quetelet' keskmete teooria ja selle kriitika	12
1.2.1. A. Quetelet' keskmete klassifikatsioon.	12
1.2.2. "Keskmete inimese" teooria.	14
1.3. Keskmete tunnetuslike omadusi.	15
1.3.2. Abstraktsus ja konkreetus.	15
1.3.2. Tüüpilisus.	16
1.3.3. Objektiivsus.	18
1.4. Keskmetele esitatavad nõuded.	19
1.5. Peamised keskmiseliigid.	20
1.5.1. Mahu- ja asendikeskmised.	20
1.5.2. Mahukeskmete üldvalem.	21
1.5.3. Mahukeskmete majorantsus	23
1.6. Keskmete kasutusala piirid.	23
2. Aritmeetiline keskmine	25
2.1. Lihtne ja kaalutud aritmeetiline keskmine. . .	25
2.1.1. Keskmete valem.	25
2.1.2. Variatsioonireak keskmine	26
2.1.3. Variatsioonireak koondamine	28

2.2. Aritmeetilise keskmise omadused	30
2.2.1. Esimene omadus	30
2.2.2. Teine omadus	30
2.2.3. Kolmas omadus.	31
2.2.4. Neljas omadus.	32
2.2.5. Viies omadus	36
2.2.6. Kuues omadus	39
2.3. Keskmise leidmine liitmisemenetlusega.	41
2.4. Keskmise arvutamise tehnika sõltuvus töödeldava rea iseloomust	43
2.4.1. Aritmeetiline keskmine aegrea andmetest.	43
2.4.2. Aritmeetilise keskmise arvutamine inter- vallrea andmetel	45
2.4.3. Lahtised äärerühmad.	46
2.4.4. Nn. isoleeritud keskmised.	49
2.5. Kaalude valik	50
2.6. Tsentrograafiaprobleemide lahendamine	53
3. Kronoloogiline keskmine.	55
3.1. Aritmeetiline ja kronoloogiline keskmine.	55
3.2. Kronoloogilise keskmise üldvalem.	57
3.3. Kronoloogilise keskmise praktiline kasutamine.	58
4. Harmooniline keskmine	60
4.1. Harmoonilise keskmise valem	60
4.2. Millal kasutada harmoonilist, millal aritmeetilist keskmist?	62
4.3. Harmoonilise keskmise praktiline kasutamine	63
4.4. Diskussioon harmoonilise keskmise olemusest	65
5. Geomeetriline keskmine	66
5.1. Aritmeetiline ja geomeetriline proportsioon	66
5.2. Geomeetrilise keskmise valem. Logaritm- keskmine	67
5.3. Geomeetrilise keskmise praktilisest kasutamisest	68

5.4.	Geomeetrilise keskmise kasutamise lihtsam erijuhtum	71
5.5.	Diskussioon geomeetrilise keskmise kasutatavuse üle.	73
6.	Ruutkeskmine ja muud astmekeskmsed.	77
6.1.	Ruutkeskmise valem.	77
6.2.	Ruutkeskmise kasutamine majanduspraktikas . . .	77
6.3.	Muud astmekeskmsed	80
7.	Mediaan ja kvartiilid	80
7.1.	Mediaani mõiste ja leidmine lihtsas statistilises reas	81
7.2.	Mediaani leidmine sõredates moment-variatsioonridades	83
7.3.	Mediaani leidmine intervallitud variatsioonreaga andmetel	84
7.3.1.	Graafiline meetod.	84
7.3.2.	Mediaani arvutamine.	86
7.4.	Kvartiilid	89
7.5.	Detsiilid. Tsentiiilid	91
7.6.	Mediaani praktiline kasutamine	93
8.	Mood	94
8.1.	Moodi mõiste ja leidmine lihtsas statistilises reas	94
8.2.	Moodi leidmine momentreaga andmetel.	95
8.3.	Moodi leidmine intervallitud variatsioonreaga.	95
8.3.1.	Graafiline menetlus.	95
8.3.2.	Moodi arvutamine intervallitud variatsioonreaga andmetel	96
8.4.	Moodi leidmine aritmeetilise keskmise ja mediaani kaudu	99
8.5.	Moodi praktiline kasutamine	101
9.	Variatsiooninäitajad	102

9.1. Varieerumise mõiste ja olemus	102
9.1.1. Vajadus täiendada keskmisi variatsiooninäitavudega	103
9.1.2. Mõningaid näiteid variatsiooninäitavude praktilisest kasutamisest	104
9.2. Elementaarseid variatsiooninäitavate	106
9.2.1. Variatsiooniamplituud	106
9.2.2. Keskmine lineaarhälve	107
9.2.3. Variatsioonikoefitsient - suhteline variatsiooninäitav	108
9.3. Dispersioon ja standardhälve	111
9.3.1. Dispersioon	111
9.3.2. Standardhälve	112
9.3.3. Variatsioonikoefitsient	113
9.3.4. Alternatiivse tunnuse variatsiooninäitavud	114
9.3.5. Dispersioonide liitmise lause	117
9.4. Dispersiooni lihtsustatud arvutamine	118
9.4.1. Dispersiooni omadused ja nende praktiline kasutamine	118
9.4.1.1. Esimene omadus	118
9.4.1.2. Teine omadus	119
9.4.2. Momentide mõiste ja kasutamine dispersiooni arvutamisel	120
9.4.2.1. Momendi mõiste	120
9.4.2.2. Dispersiooni ja aritmeetilise keskmise leidmine momentide meetodil	122
9.4.3. Dispersiooni leidmine liitmisemenetlusega	126
9.5. Muud variatsiooninäitavate	128
9.5.1. G. Gini keskmine absoluuthälve (media differenza).	128
9.5.2. F. Galtoni kvantiilhälve	132
9.5.3. A. Bowli diferentseerumiskoefitsient	135

9.5.4. K. Pearsoni ja R. Fisheri variatsioonikoefitsiendid	136
10. Keskmised ja variatsiooniridade kuju	136
10.1. Sümmeetrilised ja ebasümmeetrilised read . .	136
10.2. Asümmeetria	137
10.2.1. K. Pearsoni asümmeetriakordaja . . .	140
10.2.2. W. Johannseni asümmeetriakordaja . .	141
10.3. Ekstess	142
10.4. Teoreetilise variatsioonikõvera konstrueerimine empiirilistel andmetel	143
11. Lõppmärkusi keskmiste ja variatsiooninäitavade kasutamise kohta	148

1. KESKMISTE KASUTAMISE ÜLDALUSEID.

1.1. Keskmise mõiste.

1.1.1. Keskmine kvantitatiivse üldistusena.

Keskmine on levinumaid kvantitatiivset laadi üldistusi. Rahva üldises kõnepruugis on laialt kasutusel mõisted, nagu "keskmine", "keskmist kasvu", "keskmine sissetulek" jne. Nende mõistetega opereeritakse enamasti kui küllaltki avarate vahemikega, mõtlemata nende täpsele arvulisele kindlaksmääramisele. "Keskealine" ei ole mitte mingi kindla vanusega inimene, vaid see, kelle kohta ei saa öelda, et ta oleks vana, ega ka, et ta oleks noor. Hoolimata niiviisi piiratud vanusevahemiku suurest ulatusest ja piiride ähmasusest, on mõiste "keskealine" siiski väga otstarbekas. Kui suurem täpsus pole nõutav, võimaldab ta mõnigi kord inimesi küllalt tabavalt iseloomustada. Sama võib öelda ka paljude teiste keskmiste kasutamisest.

See tähelepanek viitab ühele keskmise mõistega seotud eriti tähtsale seigale: keskmise mõiste kasutamine võib osutuda ratsionaalseks ka ilma, et seda püütakski numbriliselt avaldada. Seega ei ole keskmine eeskätt mitte mingi kindla meetodi kohaselt leitud arv, vaid loogilises arutlusprotsessis kasutatav üldistav mõtlemiskategooria. Püüdes teda täpselt kindlaks määrata, võib leida muidugi ka tema numbrilise väärtuse.

Nagu teisedki kvantitatiivsed suurused, allub ka keskmise mõiste arvulisele "mõõtmisele". Sellisel "mõõtmisel"

võidakse lähtuda mitmesugustest eri vaatekohtadest, mida tingivad ühelt poolt uuritavate nähtuste konkreetseid iseärasused ja teiselt poolt uurimise ülesanded. See tähendab, et keskmisi võib olla palju, et ühel kogumil võib olla mitu keskmist, mis formaalselt võttes on kõik üheväärselt õiged. Missugust neist kunagi vaja on, seda saab otsustada ainult uuritavate nähtuste iseloomu arvestades.

Keskmete arvukus on näiliselt vastuolus mehhanistliku keskmisekäsitusega: igal kangil saab olla ainult üks keskpunkt, milles ta tasakaalustub! Nii sõnastatud arusaam keskmisest ei ole vale - teame seda hästi oma praktilistest kogemustest - , seda ei saa aga kanda üle keskmise mõistele üldse. Seepärast ongi siin tegemist ainult näilise vastuoluga; tegelikult on mehhanistlik keskmisekäsitus ainult keskmise mõiste rakendamise üks erijuht, mis seega tõestab paljude keskmiste võimalikkust ega räägi sellele hoopiski vastu.

1.1.2. Suurte arvude seadus ja keskmised.

Teatud tingimustel võib üksiknähtuse individuaalväärtust vaadelda juhusuurusena, s.t. niisuguse suurusena, mille väärtused kujunevad mitte ainult mingi üldise seaduspärasuse automaatsete resultaattena, vaid ühtlasi paljude juhuslike, üldisele seaduspärasusele mittealluvate tegurite toimel. Inimeste eluea pikkus allub näiteks objektiivsetele bioloogilistele ja ühiskondlikele seaduspärasustele ja sõltub üldistest majanduslikest, sanitaarsetest jms. elutingimustest. Üksiku isiku elu kestus sõltub peale selle aga ka äärmiselt paljudest juhuslikest asjaoludest: tema elulaadist, tervisest, korterioludest, tootumisrežiimist, iseloomust, õnnetustest, mis teda tabavad, epideemiade levikust tema elukohas jne. Just viimased tegurid võivad osutada sageli otsustavaks ja lühendada või pikendada tunduvalt üksikisiku eluea kestust. Seepärast ei saa mingi riigi rahvastiku või mingi elukutse esindajate elu pikkuse kohta teha mingisuguseid

pädevaid järeldusi ainult ühe või ka mõne üksiku inimese eluea alusel. Pealtnäha juhuslike surmapõhjuste taha peidetud eluea kestust reguleeriv seaduspärasus muutub mõisteta-vaks alles siis, kui võtta vaatluse alla piisavalt suur hulk inimesi ja jälgida nende keskmise eluea muutumist.

Seaduspärasus avaldub keskmiste kaudu, mis on suhteliselt püsivad, sellal kui üksiknähtuste individuaalväärtused on õige muutlikud. Üsikut inimesed elavad näiteks väga erineva arvu aastaid - mõned surevad väga noorelt, teised elavad keskeani, kolmandad jõuavad õige kõrgesse ikka. Pealtnäha pole selles midagi korrapärast. Ometi on näiteks rahvastiku keskmine eluiga (eeldusel, et vaadeldakse piisavalt suurt kogumit) õige püsiv suurus, mis muutub aastast aastasse ainult suhteliselt vähesel määral.

Toodud näitest selgub keskmiste ja suurte arvude seaduse vaheliste seoste kohta mõndagi. Kõigepealt see, et keskmiste tähtsus tunnetusprotsessis põhineb suurte arvude seadusele, et nad kujunevad usaldatavaks informatsiooniallikeks üksnes siis, kui nende arvutamisel on seda seadust arvestatud.

Keskmiste lahutamatu sisuline seos suurte arvude seadusega on leidnud statistikakirjanduses suurt tähelepanu. Seda seadust on nimetatud ka "keskmiste arvude seaduseks". Viimast terminit kasutas näiteks K. Marx oma "Kapitali" I köites.

Suurte arvude seaduse olemuse üle on avaldatud palju lahkuminevaid arvamusi. Eriti suurt huvi on äratanud selle seaduse vahekord majandus- ja looduseadustega. Pidades silmas ainult majandus- ja looduseaduste erisusi ja nägemata nende seaduste omavahelisi seoseid, on nõukogude statistika-kirjanduses minevikus statistilisi seaduspärasusi ja muuhulgas ka suurte arvude seadust eitatud. Tänapäeval oleme jõudnud veendumusele, et statistiline seaduspärasus ei ole mingi kolmas seaduseliik loodus- ja majanduseaduste vahel, vaid viimaste avaldumisvorm. Statistilise seaduspärasuse kujul võib avalduda nii looduseaduste kui ka ühiskondliku seaduse toime.¹

¹ Prof. I.J. Pissarevi järgi võivad nii loodus- kui ühiskonnaseadused avalduda kas automaatsete või statistilis-

Mitmesuguste majandusseaduste statistiliste seaduspärasuste kujul avaldumist käsitledes on F.D. Livšits teinud ettepaneku anda suurte arvude seadusele üldine sõnastus, mille võib võtta kokku nelja järgmisse punkti:

"a) ühtlase koostisega ühiskondlike nähtuste ja protsesside seaduspärasused avalduvad ja neid võib järelikult ka uurida ainult piisavalt suure arvu andmete (üksuste, juhtude) alusel;

b) neid seaduspärasusi võib kvantitatiivselt väljendada ainult keskmiste (rühmkeskmiste; üldkeskmiste) abil;

c) keskmised väljendavad ühtlase koostisega ühiskondlike nähtuste ja protsesside arengu seaduspärasusi sega täpselt, mida suurem arv uuritava nähtuse või protsessi üksikosi uurimistööga hõlmatakse;

d) üksiknähtuste keskmistest ühele või teisele poole hällimine, mida põhjustavad antud nähtuse suhtes ebasoolised, juhuslikud asjaolud, on jaotumuse põhilisi omadusi ning ühikute suure arvu korral nende vastastikune mõju kattub."¹

Suurte arvude seaduse ja keskmiste vahetõlke õigeks mõistmiseks peetagu silmas, et mitte suurte arvude seadus ei määrata kui suureks peab kujunema mingi konkreetne keskmine. Nii- samuti ei ole see seadus võimeline muutma mingit keskmist või tagama selle püsivust pikema aja vältel. Suurte arvude seadus räägib üksnes massnähtustes toimuva seaduspärasuse avaldumise keskmise vormis. Teisiti võiks öelda, et suurte arvude seadus on seadus vormist, milles avalduvad massnähtuste valdkonnas toimivad seadused, sealhulgas ka majandusseadused.

te seaduspärasuste kujul. Loodusseadustest realiseeruh nn. automaatse seaduspärasuse vormis näit. gravitatsiooniseadus, statistilise seaduspärasusena Browni seadus. Ühiskondlike massnähtuste kujunemist mõjustavate seaduste avaldumine toimub pea eranditult statistilise seaduspärasuse kujul. Vt. lähemalt И.Ю. Писарев, Статистическая закономерность и ее характеристики в форме средних, НИИТ, Москва 1959, lk. 8 jj.

¹ Ф.Д. Лившиц, Закон больших (средних) чисел в общественных явлениях, АН СССР, Ученые записки по статистике, том I, Москва 1955, lk. 172.

1.2. Adolphe Quetelet' keskmiste teooria ja selle kriitika.

Ehkki praktiliselt kasutusel juba väga ammustest aegadest, hakati keskmiste tunnetusteoreetilisele küljele tähelepanu pöörama alles möödunud sajandil.

Üks esimesi keskmiste tunnetuslike funktsioonide teaduslikke käsitlejaid oli kuulus belgia statistik Adolphe Quetelet (1796 - 1874, hääld. ketlee), keda peetakse keskmiste teooria rajajaks. Hariduselt astronoom ja füüsik, tegeles Quetelet kõigepealt põhjalikult matemaatikaga, huvitus siis aga filosoofiast ja leidis lõppeks oma pärisala statistikas. Juba võrdlemisi noorelt sai ta mitme akadeemia liikmeks ja teda on peetud moodsa ühiskonnastatistika isaks. Quetelet mõistis statistikat teadusena, mis tegeleb ühiskonna füsioloogia arvulise tunnetamisega.

Oma peateose pealkirjastas ta "Physique sociale" ("Sotsiaalne füüsika"). Ta kasutas ühiskondlike nähtuste vaatlisel loodusteadusest tuntud analoogiaid ning oma aja lapsena püüdis ühiskondlikke seaduspärasusi taandada mehhaanikaseadustele.

1.2.1. Quetelet' keskmiste klassifikatsioon.

Keskmiste teoorias eristas Quetelet a) tüüpilisi (teisiti väljendatult tõelisi ehk reaalseid keskmisi) ja b) mitte-tüüpilisi (teisiti subjektiivseid ehk ideaalseid keskmisi).

Tüüpilised keskmised saadakse mingi ühe eseme korduva mõõtmise tulemustest. Näiteks ühe ja sama maja-kõrguse mitmekordsel mõõtmisel saadakse iga kord mõne sentimeetri või millimeetri võrra erinev tulemus. Selle põhjuseks on võib-olla mõõdulindi erinev pingutatus eri mõõtmistel, tulemuse lugemise ebatäpsused jne. Loomulikult on majal ainult mingi üks kindel kõrgus. Missugune saadud konkreetsetest mõõtmistulemustest sellele kõige enam läheneb, pole teada. Täielik alus on aga oletada, et kui leida kõigist saadud mõõtmistulemustest keskmine, siis on tõenäoline viga kõige väiksem. Mida rohkem

mõõtmisi teha, seda väiksemaks viga kujuneb. Füüsikalises mõõtmispraktikas ja geodeesias on sellised mõõtmistulemuste keskmised olnud ammuigi kasutusel.

Nii leitud keskmist nimetabki Quetelet tõeliseks (objektiivseks, tüüpiliseks, reaalseks) keskmiseks. Antud juhul on tegemist tõelise keskmisega just selles mõttes, et see väljendab suurima mõeldava täpsusega mingi tegelikult olemasoleva eseme objektiivset väärtust.

Mittetüüpilise ehk subjektiivse keskmisega on tegu siis, kui see arvutatakse paljude ühelaadiliste objektide mõõtmetest, näiteks paljude eri majade keskmine kõrgus. Sel juhul võidakse saada niisugune keskmine, mis ei vasta ainsagi konkreetse maja kõrgusele, annab aga üldistava iseloomustuse kogu vaadeldud majade kollektiivile. Quetelet' järgi on selline keskmine ebatüüpiline seepärast, et mingit kindlat majatüüpi, mida saadud kõrgus iseloomustaks, pole olemas. Subjektiivseks on seda keskmist hiljem nimetatud seetõttu, et tema väärtus sõltub muuhulgas ka sellest, missugused majad mõõtmise alla võtta (ilmselt subjektiivselt otsustatav).

Vaatlusteooria terminoloogiat kasutades¹ võib lühidalt öelda, et kordusvaatluse andmetest saadakse Quetelet' järgi objektiivseid keskmisi, kogumivaatluse andmetest subjektiivseid.

Kahe kirjeldatud keskmiserühma erinevuste tõlgendamist on käsitlenud paljud nimekad statistikateoreetikud (Mayr, Lexis, Jahnsen jt.). Peaprobleemiks on olnud seejuures keskmiste tüüpilisus. Vene statistikud on näidanud, et ka Quetelet' poolt mittetüüpiliseks tituleeritud kogumivaatluse andmetest tuletatud keskmine võib olla kõigiti tüüpiline, kui vaatlusega hõlmatakse küllalt suur arv üksikobjekte (Jahnsen). (Vt. ka 1.3.2.)

Milliseid täpsustusi Quetelet' keskmiste käsitluses ka tuleb teha, on tema poolt alustatud keskmiserühmade eristamine ning nende ühiskonnateaduste valdkonnas kasutamise propageerimine kahtlemata Quetelet' suuri teeneid.

¹ Vt. U. Mereste, Statistilise vaatluse teooria, TRÜ, Tartu 1963, lk. 15.

1.22. "Keskmise inimese" teooria.

Quetelet' peamised eksimused, mis kutsusid esile terava kriitika ja diskrimineerisid teatud aja vältel isegi keskmiste kasutamist üldse, tulenesid tema ebaõigest ettekujutusest ühiskonnas toimivatest seaduspärasustest ja nende avaldumisvormidest.

Quetelet' järgi areneb inimühiskond kindlate, loodusesaadustele sarnaste seaduste alusel. Quetelet nägi esimesena statistika peamist ülesannet nende seaduste uurimises. Ta tahtis selgitada välja ühiskondlike nähtuste arengut mõjustavad põhjused ega rahuldunud üksnes andmete kogumise, refereerimise ja intuitsioonile põhineva tõlgendamisega, mis oli temani statistikas valdav. Ühiskondlikke seaduspärasusi mõistis ta aga lihtsalt omamoodi füüsikaseadustena.

Quetelet' järgi tuleb keskmisi vaadelda vahenditena, mis võimaldavad tuua esile nähtustes peituvad tüüpilised jooned. Igasuguseid kogumiliikmete individuaalseid erisusi tuleb aga vaadelda hälvetena tollest tüübist.

Ühiskonnaliikmete mitmesuguste omaduste kohta kogutud andmetest leitud keskmised võimaldavad Quetelet' järgi töötada välja ettekujutuse keskmisest inimesest, mis on vastava rahvuse või inimeste kollektiivi jaoks tüüpiline. Keskmine inimene on iga ühiskondliku organismi raskuskese, mille ümber võnguvad (ostsilleerivad) kõik üksikisendid.

Muidugi on selline keskmiste kombinatsioon ülimalt abstraktne ja tinglik. Quetelet' rõhutabki korduvalt, et tema keskmine inimene on ainult kujutletav olend, paljude ideaalsete keskmiste kombinatsioon.

Ent siin komistab ta iseoma võrdlustele. Ideaalseid keskmisi kui üksnes ideedes olelevaid mõttelisi keskmisi hakkab ta samastama ideaalidega kui ülima täiuslikkuse ja otstarbekuse kehastistega. Arendades õpetust keskmisest inimesest, jõuab ta väideteni, et keskmine kui tüüp ongi igasugune arengu lõplik eesmärk, et keskmine inimene ongi inimliku ilu ideaal jne.

Need Quetelet' tõestamata arvamused sisaldavad vasturääkivusi, mida on hiljem kritiseerinud paljud autorid. A. Quete-

let' põhjendamata seisukohtade kriitika kujunes peagi ketletismi kriitikaks üldse, kusjuures seati kahtluse alla ka keskmiste ühiskondlike ja majanduslike nähtuste uurimisel kasutamise võimalikkus üldse, ühiskondliku seaduspärasuse mõiste jne. Quetelet' teostest paisunud huvipuhang statistiliste probleemide teoreetilise uurimise vastu vaibus peagi. Quetelet'-järgset perioodi iseloomustab statistika ajaloos pikeamat aega kestnud empirism ja teooriast vöördumine. See kestis kuni XIX sajandi lõpuveerandini, millal mitmed matemaatilise statistika valdkonnas tehtud edusammud äratasid taas huvi ka ühiskonna- ja majandusstatistiliste probleemide teoreetilise läbitöötamise vastu.

1.3. Keskmiſte tunnetuslikke omadusi.

1.3.1. Abſtraktsus ja konkreetſus.

Statistika teoorias on vaieldud palju selle üle, kas keskmine on loomult abſtraktsne või konkreetne.

Keskmiſi arvutatakse teatud kogumite kohta hangitud andmetest. Nad iseloomustavad mingi tunnuse väärtust vastavas kogumis. Isikut A iseloomustavad tema individuaalsed omadused: vanus 20 aastat, pikkus 168 sm, kaal 59 kg, jne. Samasugused, ehkki teistsuguse numbrilise väärtusega andmed iseloomustavad ka isikut B, C jne. Kui tahame iseloomustada samade tunnustega teatud kollektiivi, siis on selleks kaks võimalust:

a) loetleda kõigi kollektiivi kuuluvate inimeste vanused, pikkused jne., mis oleks küllaltki tülikas, või

b) tuletada vastavad keskmised: keskmine vanus, keskmine pikkus jne. ning tugineda neile. Viimasel juhul on selge, et leitud keskmised ei käi ühegi konkreetse inimese, vaid kogu kollektiivi kohta.

Võime seetõttu kirjeldada keskmiſi kui kogumi ühe arvuga väljendatud koondiseloomustusi. Võime ütelda, et nagu uuritavate tunnuste konkreetse mõõtmisega kindlaks tehtud väärtused iseloomustavad kogumi üksikliikmeid, nii iseloomustavad vastavate tunnuste keskmised väärtused kogumit tervikuna. Seega on

keskmine kahepalgeline suurus. Kogumi kui terviku iseloomusena on ta täiesti konkreetne. Iga selle üksikliikme suhtes on aga keskmine abstraktsioon.

Keskmine on abstraktne selles mõttes, et üldistades endas paljude üksiknähtuste andmeid on viimaste individuaalsed erisused keskmistest abstraheeritud (elimineeritud).

Keskmete abstraktsus ei takista muidugi nende kasutamist konkreetsete ülesannete lahendamisel.

1.3.2. Tüüpilisus.

Keskmise ja kogumisse kuuluvate üksiknähtuste vahekor-rast kasvab välja keskmiste tüüpilisuse probleem, mis sisal-dub praktiliselt küsimuses: missugusel määral keskmine esin-dab kogumisse kuuluvate üksikliikmete vastavate tunnuste väär-tusi?

Keskmised on seda tüüpilisemad, mida homogeensemad on kogumid.

Kogumi homogeensust tuleb seejuures mõista kaheti. Kva-litatiivse homogeensuse ehk ühtluse all mõistetakse kogumi üksikliikmete põhimõttelist üheliigilisust, mis avaldub tea-tud omaduste komplekside kaudu. Kui näiteks inimese keskmise sissetuleku arvutamisel võetakse arvesse nii tehaseomanikud ja nende sissetulekud kui ka töölised ja nende palgad, siis on saadud keskmine nii töösturite kui tööliste sissetulekute suhtes ebatüüpiline. Põhjuseks on kogumi ebaühtlus, sest oma tulude suuruse ja iseloomu poolest ei saa tööstureid ja töö-lisi, kasumeid ja palku liita üheks kogumiks. Mingi teise tunnuse uurimise seisukohast võib aga sama inimkollektiiv osu-tuda ka homogeenseks, näiteks inimeste antropoloogiliste mõõt-mete uurimisel jms.

Kvantitatiivse homogeensuse all mõistetakse hangitud üksikandmete häälbimise teatud normaalset iseloomu. Homogeenne on siisugune kogum, mille üksikliikmed ei häälbi üksteisest rohkem kui seda vastavate juhuslike suuruste jaotumise sea-duste alusel võib oletada. Kogumite kvantitatiivse ühtluse uurimiseks kasutatakse matemaatilises statistikas nn. homo-geensusteste (näit. χ^2 -testi).

Ka kvantitatiivne ebahomogeensus võib põhjustada keskmiste ebatüüpilisust. Seda võib juhtuda ühiskondlike ja majanduslike nähtuste uurimisel suhteliselt sageli, sagedamini kui näiteks bioloogiliste protsesside puhul. Mitteküllaldasest ühtlusest tekkivate komplikatsioonide kõrvaldamiseks on töötatud välja mitmesuguseid spetsiaalmenetlusi (näit. nn. asendikeskmised, G. Gini variatsioonimõõtarv jne.), mis võimaldavad saada rahuldavaid üldistusi ka kvantitatiivselt ebahomogeensest materjalist; eeldusel et kogumi kvalitatiivne homogeensus on tagatud.

Üldistusena võib öelda, et kui kvalitatiivse homogeensususe puudumine rõõvib keskmistelt ja nende alusel arvutatavalt muudelt näitarvudelt tihti igasuguse sisulise (resp. majandusliku) mõtte, siis defektid kvantitatiivses homogeensususes enamasti ainult komplitseerivad materjali analüüsimist või seavad tõkkeid tuletatavate üldistuste rakendatavusele.

Keskmiisi on kasutatud ka marksismi-leninismi klassikute paljudes ühiskonnateaduslikes töödes. Tähtsaid metodoloogilisi vihjeid keskmiste rakendamises esinenud puuduste kohta on teinud V.I. Lenin. Erilist tähelepanu on ta osutanud seejuures kogumite kvalitatiivse ühtluse tagamise vajadusele, mida omaaegne vene semstvostatistika küllaldaselt silmas ei pidanud. V.I. Lenini töödes keskmiste kasutamise kohta sisalduvad meetoodilised näpunäited on lühidalt kokku võtnud prof. G.I. Baklanov. Tema sõnastatud juhiste kohaselt peab keskmiste arvutamisel:

1) hoolikalt analüüsima kogumi koosseisu, tagama kogumi elementide kvalitatiivse homogeensususe (ühtluse);

2) mitte piirduma keskmiste arvutamisega terve kogumi kohta, vaid laialdaselt kasutama rühmkeskmisi, diferentseeritud keskmisi;

3) uurima kogumi kõige paremaid ja kõige halvemaid elemente, milleks tuleb tunnuse keskmise väärtuse kõrval kindlaks teha ka ta suurim ja väikseim väärtus;

4) peegeldama uue kvaliteedi sügenemise protsessi, kvantitatiivsete muutuste üleminekut kvalitatiivseteks muutusteks, sest parimad näitajad, mis täna on veel üksiknähtused, kujuta-

vad endist uue kvaliteedi tekkimist ja saavad homme tüüpilisteks, massilisteks."¹

Peale kogumite homogeensuse sõltub mittekõikse vaatluse andmetel leiitud keskmiste tüüpilisus veel kasutatud algandmete hulgast. Tüüpilisemad, laiemalt kehtivad ja teoreetiliselt huvitavamad on niisugused keskmised, mis on arvutatud piisavalt suurest algandmete hulgast. Vastavate üksikprobleemidega (kui suur on piisav arv jms.) tegeleb lähemalt väljavõttelise vaatluse teooria, mille raames rakendatakse ulatuslikult töönaosusarvutusi.

1.3.3. Objektiivsus.

Teadusliku käsitluse olulisimad tunnused on objektiivsus ja täpsus. See eeldab ka keskmise arvutamise meetodika täpset kindlaksmääramist. Ainult sellisel juhul on saadavad keskmised objektiivsed.

Siiski tuleb pidada silmas, et teatud subjektiivsuse elemente võib keskmiste väärtustes avalduda ka siis, kui neid arvutatakse õigesti.

Subjektiivseid mõjustusi võib keskmistesse sügeneda eelkõige kasutatavate lähteandmete sihiliku rühmitamisega nii, et saadud rühmakeeskmiised kujuneksid uurijale soodsamateks. Seda laadi vigu võidakse teha tahtlikult, kindlatel majanduslikel või poliitilistel kaalutlustel. (Kuidas andmeid õigesti rühmitada ja missuguseid vigu seejuures võib tekkida, seda käsitletakse üksikasjalisemalt rühmitamisteooria raames.) Ent keskmiste mehhanismi mitte küllalt põhjalikult tundev uurija võib siin eksida ka tahtmatult. Märgates, et lähtematerjali mitmesuguselt rühmitades saadakse erinevaid keskmisi, võib ta valida alateadlikult niisuguse rühmitamisviisi, mis annab tema väidetega kõige paremini sobivaid tulemusi, kuigi see rühmitamisviis võib olla kunstlik ja sisuliselt sobimatu. Loomulikult ei ole sellistel keskmistel mingit teaduslikku tõestusjõudu, nad ei ole objektiivsed. Et niisuguseid komistamisi välti-

¹ G.I. Baklanov, Tõestusstatistika, Tallinn 1951, lk. 54.

da, peab andmete eelrühmitamisel silmas pidama kõiki rühmitamisteooria nõudeid ja püüdma kasutada ainult täiesti objektiivseid rühmitamisaluseid. Eriti oluline on see andmete subjektiivsete tunnuste järgi rühmitamisel; näit. hea, keskmise ja halva põllupinnaga põllumajandusettevõtete rühmade loomisel tekib küsimus: kus on objektiivne piir halva ja keskmise kvaliteediga põllumaa vahel?

Segavat mõju andmete objektiivsusele võib avaldada ka keskmise tüübi juhuslik (s.t. ebateadlik) valik. Nagu teada, ei ole samadest lähteandmetest arvatatud erinevad keskmised võrdsed; ühed neist on suuremad, teised väiksemad. Kui iseloomustatakse nähtust, mille väärtust tahetakse näidata võimalikult suuremana ja kasutatakse selleks meelevaldselt (s.o. ilma et nähtuse sisu seda dikteeriks) just suuremat arvulist väärtust tagavat keskmist, siis on tegemist muidugi subjektiivse moonutusega. Nagu edasises näeme, võib iga keskmist kasutada ainult teatud kindlatel piiravatel tingimustel.

1.4. Keskmistele esitatavad nõuded.

Inglise statistik G.U. Yule on formuleerinud mõned üldised formaalsed nõuded, millele keskmine peab vastama.¹ Neid nõudeid arvestades saab selgeks, miks mõned keskmised on eriti populaarsed. Ühtlasi aitavad nad põhjendada mitmete erinevate keskmiste kasutamise vajadust. Keskmistele esitatavad nõuded on lühidalt järgmised:

a) Rea üldistavaks iseloomustamiseks kasutatav keskmine peab olema arvutuste teel objektiivselt määratav ega tohi sõltuda mingisugustest subjektiivsetest kaalutlustest.

b) Keskmine peab sõltuma kõigist statistilisse ritta kuuluvatest andmetest. Kui see nõue pole rahuldatud, ei iseloomusta ta variantide tegelikku jaotumist.

c) On soovitatav, et keskmisel oleks konkreetne tähendus ja et tema omadused oleksid lihtsad ja kergesti taibatavad.

¹ Д.Э. Юл и М.Д. Кендал, Теория статистики, Москва 1960, lk. 136-37.

Eriti tähtis on see seetõttu, et keskmistega opereerivad igapäevases töös ka sellised inimesed, kellelt ei saa nõuda põhjalikke teadmisi statistika teooriast.

d) Keskmine peab olema lihtsalt ja kergesti arvutatav.

e) keskmine ei tohi eriti palju muutuda, kui ta leitakse mitte kõikse, vaid väljavõttelise vaatluse andmete alusel. (See omadus on eriti tähtis väljavõttelise vaatluse teooria seisukohalt, mida käesolevas teemas lähemalt ei käsitleta.)

f) Keskmistega peab olema hõlpus sooritada vajalikke algebralisi tehteid. Üldkogumi keskmine peaks olema kergesti leitav osakogumite keskmiste kaudu.

Nendele omadustele vastab kõige paremini aritmeetiline keskmine. Pole sugugi juhuslik, et just see on olnud sajandeid kõige kasutatavamaks keskmiseligiiks.

1.5. Peamised keskmiseligid.

1.5.1. Mahu- ja asendikeskmised.

Praktikas kasutatavad keskmised jaotatakse kahte rühma: mahukeskmised ja asendikeskmised.

Mahukeskmised on sellised, mille arvuline väärtus reageerib eranditult igale muutusele statistilise rea mistahes liikme väärtuses, s.o. rea liikmete individuaalväärtuste summas ehk rea mahus.

Asendi- ehk struktuurikeskmised seevastu reageerivad ainult niisugustele muutustele rea üksikliikmete väärtuses, millega kaasneb olulisi nihkeid ka rea struktuuris (näiteks rea liikmete asendis üksteise suhtes ehk järjekorras, eri väärtusega rea liikmete osatähtsustes jne.). Paljud rea üksikliikmete väärtustes toimuvad väikesed muutused struktuurikeskmetele mõju ei avalda. Struktuurikeskmiste suurem "tõhusus" reas toimuvate muutuste suhtes on nende peamisi tunnuslikke eeliseid, mis võimaldavad neid kasutada paljudel niisugustelgi puhkudel, kui mahukeskmiste leidmine on liiga tülikas või ei võimalda püstitatud probleemidele vastata.

Peamised praktikas kasutatavad keskmised on järgmised.

A. Mahukeskmised:

- 1) aritmeetiline keskmine;
- 2) harmooniline keskmine;
- 3) geomeetriline keskmine;
- 4) kronoloogiline keskmine;
- 5) ruutkeskmine.

B. Asendi- ehk struktuurikeskmised:

- 1) mood;
- 2) mediaan;
- 3) kvartiilid;
- 4) detsiilid;
- 5) tsentiilid.

Mahukeskmistest kolm esimest - aritmeetiline, harmooniline ja geomeetriline - on tuntud juba väga ammu. Nende kasutuselevõttu on omistatud Pythagorasele, ent ilmselt tunti neid juba varem. Muud keskmised on suhteliselt nooremad. Mood ja mediaan on tulnud kasutusele alles viimasel sajandil, ehkki ebateadlikult on neidki kasutatud juba iidsetest aegadest.

1.5.2. Mahukeskmise üldvalem.

Peamisi mahukeskmisi võib esitada üldistatult valemiga

$$\bar{x} = \sqrt[m]{\frac{\sum_{i=1}^n x_i^m}{n}},$$

$$\bar{x} = \left(\frac{\sum_{i=1}^n x_i^m}{n} \right)^{\frac{1}{m}}. \quad (1)$$

kus $\bar{x}$ on variantide x_i keskmine.

Olenevalt sellest, missugune väärtus antakse suurusele m , saadakse üks neljast peamisest keskmisest. Kui $m = 1$, saadakse aritmeetiline keskmine, kui $m = 2$ - ruutkeskmine, kui $m = -1$ - harmooniline ja kui $m = 0$, siis geomeetriline keskmine.

Vaatleme mõnede keskmiste valemite tuletamist lähemalt.

Aritmeetiline keskmine: $m = 1$

$$\bar{x} = \left(\frac{\sum x^1}{n} \right)^{\frac{1}{1}} = \frac{\sum x}{n}.$$

Ruutkeskmine: $m = 2$

$$\bar{x} = \left(\frac{\sum x^2}{n} \right)^{\frac{1}{2}} = \sqrt{\frac{\sum x^2}{n}}.$$

Harmooniline keskmine: $m = -1$

$$\bar{x} = \left(\frac{\sum x^{-1}}{n} \right)^{-1} = \left(\frac{\sum \frac{1}{x}}{n} \right)^{-1} = \frac{n}{\sum \frac{1}{x}}.$$

Geomeetrilise keskmise tuletamine valemist (1), kus $m = 0$, on mõnevõrra keerulisem. Otsese asendamise puhul saaksime määramatuse

$$\bar{x} = \sqrt[0]{\frac{\sum x^0}{n}} = \sqrt[0]{\frac{n}{n}} = \sqrt[0]{1} = 1^{\frac{1}{0}} = 1^{\infty}.$$

Mõtestatud tulemuseni jõutakse, kui logaritmid võrrandi mõlemaid pooli

$$\log \bar{x} = \frac{\log \left(\frac{\sum x^m}{n} \right)}{m}$$

ning rakendada diferentsiaalarvutusest tuntud L'Hôpitali reeglit. Selle reegli kohaselt saab kahe nullile läheneva suuruse jagatise piirväärtust leida nende tuletiste jagatise piirväärtusena. Saame

$$\frac{1}{m'} \left(\log \frac{\sum x_1^m}{n} \right)' = \frac{\sum x_1^m \log x}{\sum x_1^m},$$

kust selgub, et

$$\lim_{m \rightarrow 0} \frac{\sum x^m \log x}{\sum x^m} = \log \sqrt[n]{\prod x}.$$

$\prod$ on korrutise märk (analoogiliselt summa märgiga $\sum$). Kirjutist $\prod x_i$ tuleb mõista kui kõikide variantide korrutist, s.t. $\prod x_i = x_1 x_2 x_3 \dots x_n$.

Lõpuks selgub, et

$$\lim_{m \rightarrow 0} \log \bar{x} = \lim_{m \rightarrow 0} \log \sqrt[m]{\frac{\sum x^m}{n}} = \log \sqrt[n]{\prod x},$$

kust

$$\bar{x} = \sqrt[n]{\prod x}.$$

Variatsioonrea andmetel on geomeetriline keskmine

$$\bar{x} = \sqrt[n]{\prod x^m}.$$

1.5.3. Mahukeskmiste majorantsus.

Samadest andmetest leitud eri keskmised on erineva numbrilise väärtusega. Ühed neist on suuremad, teised väiksemad. Sellises mittevõrdsuses avalduvat keskmiste omadust nimetatakse nende majorantsuseks (ladinakeelsest sõnast "major" - suurem). Nii selgub, et

$$\bar{x}_{\text{harm.}} < \bar{x}_{\text{geom.}} < \bar{x} < \bar{x}^2 \text{ jne.}$$

Seda keskmiste omadust on lihtsa arvnäitega hõlpus demonstreerida. Arvurea

$$1, 2, 5, 7, 10, 13$$

andmetel on näiteks

$$\bar{x}_{\text{harm.}} = 2,97;$$

$$\bar{x}_{\text{geom.}} = 4,58;$$

$$\bar{x} = 6,33;$$

$$\bar{x}^2 = 7,61.$$

1.6. Keskmiste kasutusala piirid.

Keskmised on ühiskondlike ja majandusalaste probleemide uurimisel asendamatud. Siiski on neile tuginevate järelduste tegemisel tarvis silmas pidada keskmiste tunnetusvõimaluste suhtelist piiratust.

Ehkki keskmised võimaldavad vältida kogumite kirjeldami-

sel pikkade arvuridade esitamist ja iseloomustada mistahes tunnust väga ökonoomselt, ainult ühe arvu kaudu, ei iseloomusta nad seda siiski igakülgselt. Keskmine iseloomustab ainult üldist nivood, mille ümber tunnuse väärtus kõigub, mitte aga variantidevahelisi individuaalseid erisusi. Viimaseks otstarbeks tuleb kasutada spetsiaalseid variatsiooninäitarve (vt. ptk. 9).

Teiseks on tähtis silmas pidada, et keskmine ei sõltu mitte ainult variantide individuaalväärtustest, vaid ka nende suhtelisest esinemissagedusest. Nii on võimalik, et müüdavate kaupade keskmine hind muutub, ehkki kõikide kaupade individuaalsed hinnad jäävad püsivaks. Viimast laadi nihe toimub siis, kui suureneb suhteliselt odavamate või kallimate kaupade osatähtsus (keskmine hind vastavalt langeb või tõuseb).

Struktuuri mõju on vaja arvestada kõigi keskmistena leitatavate töötulemuste (tehase töövõljakuse, toodangu omahinna, rentaabluse, käibekulude taseme, masinate jõudluse, materjali kulunormide jne.) hindamisel. Struktuurimuutuste mõjul võib esineda esimesel pilgul kurioossetena tunduvaid olukordi. Võib juhtuda näiteks, et trusti üheski tehases pole omahind tõusnud, trusti toodangu keskmine omahind aga on tõusnud, või et peavalitsuse kõigis ettevõtteis on töövõljakus langenud, kogu peavalitsuse süsteemis on see aga tõusnud.

Et sellised juhtumid on võimalikud mitte ainult teoorias, vaid esinevad ka tegelikult, demonstreerib hästi 1962. aasta jaanuaris ayastatud nähtus Tartu Keemiakombinaadis. Ainult kahest tehasest koosnevas kombinaadis ületasid mõlemad tehased töövõljakusplaani, kombinaadi plaan aga jäi täitmata.

Seda laadi näidetele tuginedes on katsutud väita, et keskmised ei näitavat midagi, et neid ei saavat üldse usaldada, et nad olevad head ainult prügi silma ajamiseks jne. Loomulikult pole sellised väited õiged. Asi ei seisa sellistel juhtudel mitte keskmistes, vaid oskamatuses neid kasutada ja neid oma arutlustesse lülitada. Enne keskmistele tuginevate hinnangute langetamist tuleb nad struktuurinihete mõjust vabastada. Selleks rakendatavaid meetodilisi võtteid käsitleb üksikasjaliselt indeksiteooria.

2. ARITMEETILINE KESKMINE.

2.1. Lihtne ja kaalutud aritmeetiline keskmine.

Aritmeetiline keskmine on kõige levinum keskmine. Elementaarsed võtted aritmeetilise keskmise leidmiseks on üldtuntud. Tema arvutamisel eksitakse harva. Peamisi vigu on aritmeetilise keskmise arvutamine ka seal, kus õigupoolest tuleks kasutada mõnda teist mahu- või asendikeskmist.

Millal missugust keskmist kasutada, sellele ongi edasises püütud omistada suurimat tähelepanu. Teine tähtsam probleem on aritmeetilise keskmise arvutamise lihtsustamine, et võita sellega arvutustele kuluvat aega ja vältida arvutusvigu.

2.11. Keskmise valem.

Aritmeetilise keskmise üldvalem on

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}, \quad (2)$$

kus x_i on variandid, f_i nende nn. kaalud (ehk teisiti: osatähtsused, sagedused, erikaalud) ja n variantide arv. Iga variandi väärtus võtab siin keskmise väärtuse kujunemisest osa seda suuremal määral, mida suurem on tema kaal (f), seepärast nimetatakse seda kaalutud aritmeetiliseks keskmiseks.

Erijuhul, kui iga variant esineb ainult üks kord, s.o.

$$f_1 = f_2 = \dots = f_n = 1,$$

siis omandab aritmeetiline keskmine eespool juba vaadeldud kuju

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

mida nimetatakse lihtsaks aritmeetiliseks keskmiseks.

Lihtsa aritmeetilise keskmise valemit saab kasutada järgmistel juhtudel:

- a) kui on teada otse vaatlusega hangitud andmete rida, milles on vaatlusüksuste arvuga võrdne arv liikmeid. Enamasti on sellised read väga pikad ja tülikad töödelda;
- b) kui on tegemist variatsioonreaga, milles kõikide variantide sagedused on võrdsed. Selliseid ridu kohtab praktikas väga harva;
- c) kui on tegemist aegreaga, mille liikmete väärtused käivad võrdse pikkusega perioodide või momentide kohta.

2.1.2. Variatsioonrea keskmine.

Kaalutud aritmeetilise keskmise arvutamise tehnikat demonstreerime tabeli 1 andmetel.

Seame ülesande arvutada kõigi vaadeldavate ehitusorganisatsioonide keskmine plaani täitmise protsent. Ülesande lahendamiseks vajalikud lähteandmed sisalduvad tabeli kahes esimeses veerus. Veerus (3) on toodud organisatsioonide rühmitamisel kasutatud intervallirajade keskmised. Et pole teada, kuidas igasse rühma kuuluvad organisatsioonid oma plaani täitmise protsendi järgi rühmasiseselt jaotuvad, võetakse intervalli rajade keskmine rühma tinglikuks keskmiseks plaani täitmise protsendiks. Sellega on intervallrida, mis koosneb veergudest (1) ja (2), muudetud momentreaks, mis koosneb veergudest (2) ja (3). Edasistes arvutustes tuginemegi üksnes neile veergudele.

Variantide individuaalväärtuste asendamine vastava intervallirajade keskmisega pole rangelt võttes muidugi päris õige. Võtta intervalli 10 - 20 keskmiseks väärtuseks 15 oleks täiesti õige ainult sel juhul, kui kõigi vastavas intervallis olevate liikmete keskmine olekski 15. Kuidas aga variandid intervallisiseselt jaotuvad, selle kohta pole enamasti mingisugust informatsiooni. Tõenäoliselt hälbib nende keskmine mõnevõrra intervallirajade keskmisest. Järelikult tekib igas korutises $\pm$ teatud viga, mis avaldab lõpptulemusena mõju ka

Tabel 1.

NSV Liidu tähtsamate ehitusorganisatsioonide jaotumine
1959. aasta 9 kuu plaani täitmise järgi.¹

Plaani täitmise protsent	Organisatsioonide arv	Rühma tinglik keskmine plaani täitmise protsent	
x_1	f_1	x_1	$x_1 f_1$
1	2	3	4
... - 10	20	5	100
10 - 20	23	15	345
20 - 30	66	25	1650
30 - 40	95	35	3325
40 - 50	153	45	6885
50 - 60	217	55	11935
60 - 70	254	65	16510
70 - 80	273	75	20475
80 - 90	282	85	23970
90 - 100	232	95	22040
100 - 110	178	105	18690
110 - 120	100	115	11500
120 - 130	74	125	9250
130 - 140	32	135	4320
140 - 150	27	145	3915
150 - 160	20	155	3100
160 - 170	7	165	1155
170 - 180	3	175	525
180 - 190	5	185	925
190 - 200	8	195	1560
Σ	2069	Σ	162175

¹ Andmed B. И. Ежов, Выравнивание и вычисление рядов распределения, Москва 1961, lk. 19. Tabeli kahes esimeses veerus, mis on kaksikjoonest vasakul, on lähteandmed. Veergudes 3 ja 4 on arvutusteks vajalikud abianndmed.

aritmeetilisele keskmisele. Et aga rea algul tekivad tavaliselt miinus- ja rea lõpul plussmärgiga vead (kui rea maht on küllalt suur ja jaotumus läheneb normaalsele), siis osaliselt need vead kattuvad vastastikku, mistõttu lõplik viga pole eriti suur.

Korrutanud kõik variandid nende sagedusega ning saadud korrutised liitnud, selgub et

$$\sum x_i f_i = 162175$$

ning et otsitav keskmine on

$$\bar{x} = \frac{\sum x f}{\sum f} = \frac{162175}{2069} = 78,38 \approx 78,4 \%$$

Nagu selgub, kujuneb arvutustööde maht keskmise leidmisel küllaltki suureks. On kerge mõista, et kui variandid oleksid mitte kahe- ja kolmekohalised, nagu käesolevas näites, vaid nelja-, viie-, või enamakohalised, muutuks arvutamine nii suurte arvudega väga tülikaks. Ka sagedused võivad olla mõnede ülesannete lahendamisel väga suured, mis raskendab omakorda arvutustöid. (Eesti NSV rahvastiku keskmist vanust arvutades võivad üksikute vanuserühmade sagedused, s.o. inimeste arvud üksikutes rühmades, olla kuni kuuekohalised!). See tõstab üles tähtsa probleemi arvutuste lihtsustamisest.

2.1.3. Variatsioonirea koondamine.

Mõnesuguseid võimalusi töö hõlbustamiseks pakub rea lühendamine tema koondamise teel. Kui muuta kõik intervallid kaks korda pikemaks, siis muutub rida kaks korda lühemaks, millega sooritataivate korrutamistehete arv väheneb 2 korda.

Intervallide liitmise teel variatsioonrida lühendades tuleb ühendatavate intervallide sagedused liita. Saadud rida on toodud tabelis 2.

Tabeli 2 andmetel on keskmist selles mõttes mõnevõrra hõlpsam leida, et arvutusi on vähem. Arvud ise aga ei muutu-

T a b e l 2 .

NSV Liidu tähtsamate ehitusorganisatsioonide jaotumine
1959. aasta 9 kuu plaani täitmise järgi.

Organisatsioonide rühmad plaani täit- mise protsendi järgi	Organisatsioo- nide arv	Rühma ting- lik keskmis- ne plaani täitmise protsent	$x_i f_i$
x_i	f_i	x_i	
1	2	3	4
... 20	43	10	430
20 - 40	161	30	4830
40 - 60	370	50	18500
60 - 80	527	70	36890
80 - 100	514	90	46260
100 - 120	278	110	30580
120 - 140	106	130	13780
140 - 160	47	150	7050
160 - 180	10	170	1700
180 - 200	13	190	2470
Σ	2069	Σ	162490

nud väiksemateks, vaid koguni suuremateks. Selgub, et

$$\bar{x} = \frac{162490}{2069} = 78,53 \approx 78,5 \% .$$

Eespool saime

$$\bar{x} = 78,4 \% .$$

Kuigi tulemuste vahe on suhteliselt väike (ainult 1/100)
tekib küsimus: kumb neist on täpsem?

Kui pidada silmas, et kõige täpsema keskmise saaksime

lihtsa aritmeetilise keskmisena otseste vaatlusandmete alusel, mille kujundamisest võtab iga variant osa oma tõelise individuaalväärtusega ja et viga tuleneb ebatäpsustest, mis sugenevad arvutustesse andmete rühmitamisel, siis on selge, et üldjuhul on tekkinud viga seda suurem, mida pikemad on intervallid. Kuigi otseseid vaatlusandmeid pole käesoleval juhul teada ja seepärast ei saa leida ka kõnealuste ehitusorganisatsioonide "tõelist" keskmist plaani täitmise protsenti, võib siiski olla kindel, et 1. tabeli andmetel leitud keskmise viga on väiksem.

2.2. Aritmeetilise keskmise omadused.

Aritmeetilise keskmise omadustel on suur tähtsus tema tunnetuslike funktsioonide, samuti ka lihtsustatud arvutamise võtete sisu mõistmisel. Vaatleme neist mõningaid peamisi.

2.2.1. Esimene omadus.

Kui suuruse kõik väärtused on omavahel võrdsed, siis on aritmeetiline keskmine üks neist väärtustest, s.t. kui

$$x_1 = x_2 = \dots = x_n = a,$$

siis

$$\bar{x} = \frac{\sum x}{n} = \frac{a + a + \dots + a}{n} = \frac{na}{n} = a.$$

Sellel näiliselt õigegi triviaalselt sõnastatud omadusel on keskmise olemuse mõistmisel suur tunnetuslik väärtus. Eriti tuleb seda meenutada variatsiooniteooria põhimõistete käsitlemisel.

2.2.2. Teine omadus.

Mistahes suuruste summa (vahe) aritmeetiline keskmine on võrdne nende suuruste aritmeetiliste keskmiste summaga (vahega).

Vaadelgu näiteks suuruse z aritmeetilist keskmist $\bar{z}$, kusjuures $z_i = x_i + y_i$ ($i = 1, 2, \dots, n$). Sel juhul võib eespool sõnastatud omadust väljendada sümbolitega

$$\bar{z} = \overline{x + y} = \bar{x} + \bar{y}.$$

Et

$$\sum (x_i + y_i) = \sum x_i + \sum y_i,$$

siis

$$\begin{aligned}\bar{z} &= \frac{\sum z_i}{n} = \frac{\sum (x_i + y_i)}{n} = \frac{\sum x_i + \sum y_i}{n} = \frac{\sum x_i}{n} + \frac{\sum y_i}{n} = \\ &= \bar{x} + \bar{y},\end{aligned}$$

millega eespool sõnastatud omadus on tõestatud. Sama mõttekõik kehtib arusaadavalt ka kahe suuruse vahe puhul.

Seda omadust saab tihtipeale kasutada arvutuste hõlbustamiseks. Kui näiteks on teada, et kasum võrdub realiseeritud hinnatäiendi ja käibekulude absoluutsumma vahega, siis järeldub sellest, et samasugune seos valitseb ka nende keskmiste vahel. Praktikas leitaksegi keskmine rentaablustase hinnatäiendi ja käibekulude keskmiste tasemete vahena.

Keskmiste otsene liitmine või lahutamine tuleb arvesse muidugi ainult siis, kui n on mõlema keskmise (resp. kõigi keskmiste) puhul võrdne.

2.2.3. Kolmas omadus.

Variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste hälvete summa on null, s.o.

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

Ridade omaduse kohaselt

$$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x}.$$

Et keskmise II omaduse alusel

$$x_1 + x_2 + \dots + x_n = \bar{x} + \bar{x} + \dots + \bar{x},$$

s.t.

$$\sum x_i = n\bar{x},$$

siis

$$\sum x_i - n\bar{x} = 0,$$

millest selgubki, et

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

Aritmeetilise keskmise seda omadust on nimetatud tema põhiomaduseks. See iseloomustab ammendavalt keskmise osa üldistava mõtlemiskategooriana, mis tasandab kõik individuaalnähtuste kvantitatiivsed erisused.

Sellega on sisuliselt tõestatud, et aritmeetiline keskmine vastab mehhaanikast tuntud raskuskeskme mõistele. Variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste hälvete summa on null seepärast, et positiivsete ja negatiivsete hälvete summad on võrdsed, s.t. omavahel tasakaalus.

Et aritmeetiline keskmine on identne raskuskeskmega, seda võib demonstreerida vastava katsega. Oletagem, et vaatleme mingit mehhaanilist süsteemi, mis koosneb kangist ja selle külge riputatavatest eri suurusega raskustest. Olgu raskused võrdelised eri teraviljakultuuride külvipinnaga mingis põllumajandusettevõttes ja kangi pikkus võrdeline vastavate kultuuride saagikuste muutumise vahemikuga. Kultuuride saagikused olid järgmised: nisu 30 ts/ha ja rukis 14 ts/ha, ja nende külvipinnad 500 ja 300 hektarit (vt. joon. 1). Neil andmeil on keskmine saagikus 24 ts/ha ja samas punktis tasakaalustub ka ülalpool kirjeldatud süsteem.

2.2.4. Neljas omadus.

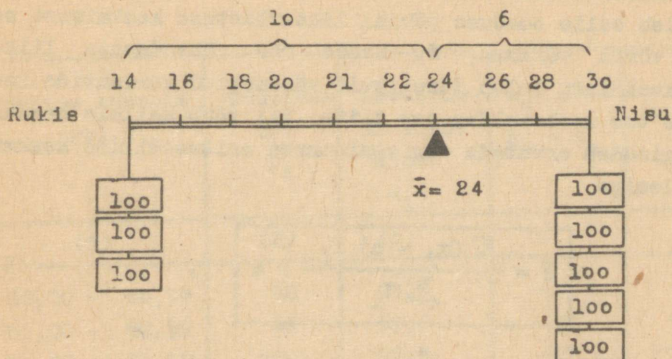
Kui vähendada (suurendada) variantide väärtusi mingi meelevaalse arvu a võrra, siis väheneb (suureneb) nende aritmeetiline keskmine sama arvu võrra.

$$x_1 = 14$$

$$f_1 = 300$$

$$x_2 = 30$$

$$f_2 = 500$$



$$10 \times 300 = 3000$$

$$6 \times 5000 = 3000$$

Joon. 1. Aritmeetiline keskmine on võrreldav raskuskeskmega. Joonisel on kujutatud kahest liikmest koosneva variatsioonrea aritmeetilise keskmise leidmist kaalumise teel (vt. lk.32).

Oletagem, et variantide x_1 vähendamisel saadakse variantid z_1 , kusjuures

$$z_1 = x_1 - a,$$

sel juhul

$$\begin{aligned} \bar{z} &= \frac{\sum z_1}{n} = \frac{\sum (x_1 - a)}{n} = \frac{(x_1 - a) + (x_2 - a) + \dots + (x_n - a)}{n} = \\ &= \frac{(x_1 + x_2 + \dots + x_n)}{n} - \frac{na}{n} = \bar{x} - a. \end{aligned}$$

See on põhilisi omadusi, millele tuginevad keskmiste ja variatsiooninäitavate arvutamise lihtsustamise viisid ja eriti nn. momentide meetod.

Kui variantide väärtused on ebamugavalt suured, siis võib valida mingi meelevaldse arvu a , mille võrra neid vähendada. Saadud vähendatud variantidest arvutatud keskmine tuleb selle omaduse põhjal lähteväärtuse keskmisest sama arvu võrra väiksem. Et saada rea vähendamata liikmete keskmist, tuleb järelikult vähendatud variantide keskmisele too meelevaldne arv liita. Nii võib sellele omadusele tuginedes arvutada variatsioonrea aritmeetilise keskmise ka valemiga

$$\bar{x} = \frac{\sum (x_i - a) f_i}{\sum f_i} + a, \quad (3)$$

ehk lihtsa statistilise rea keskmise vastavalt valemiga

$$\bar{x} = \frac{\sum (x_i - a)}{n} + a,$$

sest

$$\sum f_i = n.$$

Arvutused lihtsustuvad kõige rohkem siis, kui võtta meelevaldseks suuruseks mingi arv rea keskelt, näiteks kõige suurema sagedusega arv. Sel juhul nimetatakse seda meelevaldset suurust esialgseks keskmiseks ja tähistatakse x_0 .

Oletagem, et on teada tabelis 3 toodud andmed, mille alusel tuleb arvutada töölise keskmine kuupalk.

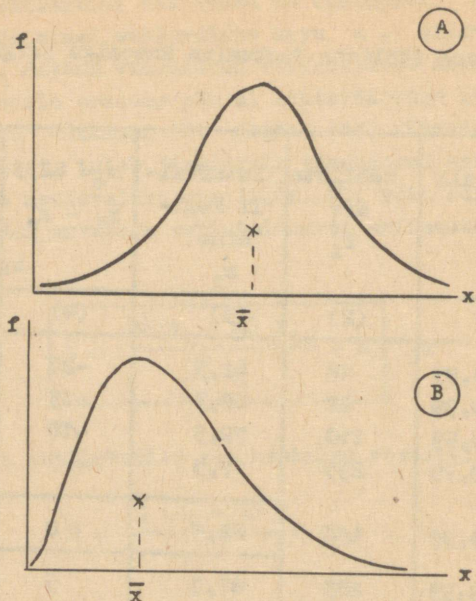
Tabeli 3 andmetel on töölise keskmine palk

$$\bar{x} = \frac{4515}{1553} + 82,5 = 2,9 + 82,5 = 85,4.$$

Tabel 3.

N tehase tööliste jaotumine kuupalga järgi.

Töölise kuupalk (rubl.)	Tööliste arv f_1	Interval- li kesk- mine x_1	$x_e = 82,5$ $x_1 - x_e$	$(x_1 - x_e)f_1$
(1)	(2)	(3)	(4)	(5)
60,00 - 64,99	18	62,5	-20	-360
65,00 - 69,99	42	67,5	-15	-630
70,00 - 74,99	110	72,5	-10	-1100
75,00 - 79,99	295	77,5	- 5	-1475
80,00 - 84,99	438	82,5	0	-3565
85,00 - 89,99	283	87,5	5	1415
90,00 - 94,99	108	92,6	10	1080
95,00 - 99,99	72	97,5	15	1080
100,00 - 104,99	80	102,5	20	1600
105,00 - 109,99	65	107,5	25	1625
110,00 - 114,99	38	112,5	30	1140
115,00 - 119,99	4	117,5	35	140
Kokku	1553	Σ	Σ	8080 - 3565 4515



Joon. 2. Aritmeetilise keskmise koht A - sümmeetrilises ja B - ebasümmeetrilises variatsioonreas. Aritmeetiline keskmine on võrdne variatsioonkõvera aluse pindala raskuskeskme abstsissiga.

2.2.5. Viies omadus.

Kui vähendada (suurendada) variantide väärtusi mingi meelevaldne arv k korda, siis väheneb (suureneb) nende aritmeetiline keskmine sama arv korda.

Kui variantide x_1 vähendamisel k korda saadakse uued variandid z_1 , siis

$$z_1 = \frac{x_1}{k} .$$

Sel juhul

$$\begin{aligned}
\bar{z} &= \frac{\sum z_i}{n} = \frac{\sum \frac{x_i}{k}}{n} = \frac{\frac{x_1}{k} + \frac{x_2}{k} + \dots + \frac{x_n}{k}}{n} = \\
&= \frac{x_1 \frac{1}{k} + x_2 \frac{1}{k} + \dots + x_n \frac{1}{k}}{n} = \\
&= \frac{(x_1 + x_2 + \dots + x_n) \cdot \frac{1}{k}}{n} = \frac{\sum x_i}{n} \cdot \frac{1}{k} = \\
&= \bar{x} \cdot \frac{1}{k} = \frac{\bar{x}}{k} .
\end{aligned}$$

Analoogiliselt eelmisele pakub seegi omadus suurepäraseid võimalusi aritmeetilise keskmise arvutamise lihtsustamiseks. Eriti sobiv on ta siis, kui variandid on mingi arvu kordsed. Valides konstandiks k , tuleb aritmeetiline keskmine leida lihtsustatud korras valemiga

$$\boxed{\bar{x} = \frac{\sum \frac{x}{k} f}{\sum f} \cdot k} \quad (4)$$

See tähendab, et kui k korda vähendatud variantidest arvutada aritmeetiline keskmine on lähteväärtustest sama palju kordi väiksem, siis tuleb $\bar{x}$ (s.o. lähteväärtuste keskmise) leidmiseks saadud keskmist x korda suurendada.

Seda omadust saab kasutada väga sageli koos aritmeetilise keskmise IV omadusega, mispuhul saadakse valem

$$\bar{x} = \frac{\sum \left(\frac{x_i - x_e}{k} \right) f_i}{\sum f_i} k + x_e \quad (4a)$$

Tabel 4.

Palk (inter- vallikesakmine) (rbl.) x_i	Tööliste arv f_i	$\frac{x_i - x_e}{k}$	$\frac{x_i - x_e}{k} f$
1	2	3	4
62,5	18	-4	- 72
67,5	42	-3	-126
72,5	110	-2	-220
77,5	295	-1	-295
82,5	438	0	-713
87,5	283	1	283
92,5	108	2	216
97,5	72	3	216
102,5	80	4	320
107,5	65	5	325
112,5	38	6	228
117,5	4	7	28
Σ	1553	Σ	+1616 - 713

+ 903

Tabelist 3 paistab hästi silma (veerust 4), et kõik varian- tide ja nende aritmeetilise keskmise vahelised hälbed jagu- vad viiega. See tähendab, et meelevaldseks suurusks k tu- leks võtta 5. Andmed ja vahetulemused on toodud tabelis 4). Nii saame aritmeetilise keskmise väärtuseks

$$\bar{x} = \frac{903}{1553} \cdot 5 + 82,5 = 0,58 \cdot 5 + 82,5 =$$

$$= 2,9 + 82,5 = 85,4 \text{ rbl.}$$

Viimane vastus langeb täpselt kokku varem (lk. 34) leitud keskmise väärtusega.

Täiendava näitena leiame samal viisil ka NSV Liidu tähtsamate ehitusorganisatsioonide 1959. a. 9 kuu keskmise plaani täitmise protsendi (lähteandmeid vaata tabelist 2). Nagu 5. tabelisse koondatud arvutusest selgub, on keskmine plaani täitmise protsent

$$\bar{x} = \frac{\sum \left(\frac{x_1 - a}{k} f \right)}{\sum f} \cdot k + a =$$

$$= \frac{-1183}{2069} \cdot 20 + 90 + -0,573 \cdot 20 + 90 = 78,54 \approx 78,5 \%$$

T a b e l 5 .

x_1	f	x_1	$x_1 - x_0$ $x_0 = 90$	$\frac{x_1 - x_0}{k}$ $k = 20$	$\frac{x_1 - x_0}{k} f$
... - 20	43	10	-80	-4	- 169
20 - 40	161	30	-60	-3	- 483
40 - 60	370	50	-40	-2	- 740
60 - 80	527	70	-20	-1	- 527
80 - 100	514	90	0	0	- 1919
100 - 120	278	110	20	1	278
120 - 140	106	130	40	2	212
140 - 160	47	150	60	3	141
160 - 180	10	170	80	4	40
180 - 200	13	190	100	5	65
	2069				+ 736
					- 1919
					- 1183

2.2.6. Kuues omadus.

Kui vähendada (suurendada) variantide kaalusid mingi meelevaldne arv h korda, siis aritmeetiline keskmine ei muutu, s.t. et jagades kaalusid f_1 konstantse suurusega h

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x \frac{f}{h}}{\sum \frac{f}{h}} = \frac{\frac{\sum xf}{h}}{\frac{\sum f}{h}} = \frac{h \sum xf}{h \sum f} = \frac{\sum xf}{\sum f} .$$

Sellelgi omadusel on suur tähtsus eelkõige arvutuste lihtsustamise seisukohast. On variantide sagedused väga suured, võib neid lihtsustamise eesmärgil vähendada. Kõige lihtsam võte on kaalude vähendamine 10, 100, 1000 jne. korda, mida võib teha kas kaalude väärtustest tarbetute nullide mahatõmbamise (resp. kirjutamata jätmise) või koma vasakule nihutamise ning murruliste kohtade vajaliku täpsuseni ümardamisega. Viimase mooduse kasutamisel tuleb piirduda siiski üksnes niisuguste ümardamisega, millest eri variantide osatähtsused kogumi struktuuris oluliselt ei muutu. Kui variantide suhtelised sagedused muutuvad, avaldab see mõju ka aritmeetilisele keskmisele.

Aritmeetilise keskmise VI omadusega on tihedasti seotud veel üks võimalus tema lihtsustatud arvutamiseks. Selle asemel, et korrutiste summat mingi suurusega jagada, võib sellega üht tegurit igas korrutises ja tulemuses liita. See tähendab, et kaalutud aritmeetilise keskmise valemit võib teisendada järgmiselt

$$\bar{x} = \frac{\sum xf}{\sum f} = \sum \left(x \frac{f}{\sum f} \right) .$$

Selles avaldises on viimasel liikmel iseseisev mõttesisu. See on variandi suhteline sagedus, erikaal ehk osatähtsus antud kogumis. Struktuurisuhtarvuna märgitakse seda tavaliselt tähega ψ (psii)

$$\frac{f}{\sum f} = \psi .$$

Järelikult võib juba eespool toodud iseloomustustele lisaks kirjeldada aritmeetilist keskmist veel ka kui variantide individuaalväärtuste ja nende osatähtsuste korrutiste summat, mida väljendab valem

$$\bar{x} = \sum x_i \psi_i \quad . \quad (5)$$

Selle valemi puhul on eeldatud, et variantide osatähtsused on avaldatud vahetuste suhetena. Kui suuruse ψ_i väärtused on protsendid, tuleb summat $x_i \psi_i$ jagada sajaga, kui promillid, siis tuhandega jne.

Täiendavalt väärib märkimist, et III, IV ja V omadust võib, ja ongi soovitatav rakendada ka korraga, kui lähteandmed seda võimaldavad ja kui selleks tekib praktiline vajadus. Sellisel juhul omandab aritmeetilise keskmise valem kuju

$$\bar{x} = \frac{\sum \frac{x_i - x_0}{k} \cdot \frac{f}{h}}{\sum \frac{f}{h}} \cdot k + x_0 \quad . \quad (6)$$

2.3. Keskmise leidmine liitmismenetlusega.

Seni kirjeldatud võimalused aritmeetilise keskmise arvutamiseks eeldasid variantide väärtuste ja nende sageduste korrutamist. Ehkki arvutuste lihtsustamiseks võis kasutada õige mitmesuguseid mooduseid, on see joon ühine kõigile eespool demonstreeritud arvutuseeskirjadele, mis mahuvad kõik nn. korrutamismenetluse üldnimetuse alla. Liitmismenetlus eeldab seevastu üksnes sageduste liitmist, nende jrgsummade leidmist. Liitmismenetlus on hõlpus ja arvutuskäigu poolest hästi ülevaatlik keskmiste arvutamise võte, mida aga praktikas siiani kahjuks ainult haruharva kasutatakse.

Nagu korrutamismenetluse puhul, nii valitakse ka siin variatsioonrea keskelt meelevaldne arv a . Selle parandus leitakse valemiga

$$b = \frac{\delta_2 - \delta_1}{\sum f} \cdot k, \quad (7)$$

kus k on intervalli pikkus;

δ_1 - sageduste järgsummade järgsumma rea madalamast pooldest;

δ_2 - samasugune summa rea kõrgemast pooldest.

Aritmeetiline keskmine saadakse siis lihtsalt summana

$$\bar{x} = a + b.$$

Järgsummade δ_1 ja δ_2 arvutamise tehnika selgub tabelist 6. Veerus 4 on sageduste järgsummad (s_1), veerus 5 on need omakorda summeeritud.

Tabel 6.

x	x_1	f_1	s_1	s_2
1	2	3	4	5
... - 10	5	20	20	20
10 - 20	15	23	43	63
20 - 30	25	66	109	172
30 - 40	35	95	204	376
40 - 50	45	153	357	733
50 - 60	55	217	574	1307
60 - 70	65	254	828	2135
70 - 80	75	273	1101	3236 = δ_1
80 - 90	85	282		
90 - 100	95	232	686	1867 = δ_2
100 - 110	105	178	454	1181
110 - 120	115	100	276	727
120 - 130	125	74	176	451
130 - 140	135	32	102	275
140 - 150	145	27	70	173
150 - 160	155	20	43	103
160 - 170	165	7	23	60
170 - 180	175	3	16	37
180 - 190	185	5	13	21
190 - 200	195	8	8	8
		2069		

Tabelis esitatud arvutustulemuste kohaselt on

$$\sum_1 = 3236; \quad \sum_2 = 1867.$$

Et $a = 85$ ja $k = 10$, siis

$$b = \frac{(1867 - 3236) 10}{2069} = \frac{-13690}{2069} = -6,62$$

ja

$$x = a + b = 85 - 6,62 = 78,38 \%$$

See tulemus langeb täpselt kokku eespool korrumtamamenetluse abil saadud keskmisega.

2.4. Keskmise arvutamise tehnika sõltuvus töödeldava rea iseloomust.

2.4.1. Aritmeetiline keskmine aegrea andmetest.

Nagu teada, jagunevad aegread periood- ja momentridadeks. Aritmeetilist keskmist saab arvutada ainult perioodridade andmetel. Perioodreas on iga liige seotud mingi ajavahemikuga. Sellise rea näiteks võivad olla tehase aastatoodangute maksumustest ja üksikute dekaadide tootmiskulude summadest koosnevad read. Perioodrea liikmete liitmisel saadakse iseseisva mõttesisuga suurus. Nii on näiteks viie aasta toodangu maksumuste summa võrdne kogu selle viisaastaku jooksul valmistatud toodangu maksumusega. Jagades selle aastate arvuga, saame toodangu aastakeskmise maksumuse.

See on aritmeetilise keskmise arvutamise tüüpiline juhtum aegridade töötlemisel. Nagu arutlustest selgus, tuli seejuures kasutada lihtsa aritmeetilise keskmise valemit, sest üksteisele järgnevad sama suurusjärku perioodid on ühepikkused.

Soovides leida tabeli 7 andmetel näiteks aastakeskmise lubjatoodangu 1955 - 1957, tuleb arvutada

$$\bar{x} = \frac{\sum x}{n} = \frac{104,7 + 92,5 + 111,4}{3} = \frac{308,6}{3} = 102,87$$

Tabel 7.

Eesti NSV lubjatoodangu kasv 1953 - 1957.

Aasta	1953	1955	1956	1957
Toodang (tuh. t.)	99,5	104,7	92,5	111,4

ehk ümardatult: keskmiselt 102 900 tonni aastas.

Kaalutud aritmeetilise keskmise järele tekib vajadus siis, kui perioodid ei ole ühepikkused. Aegrea eri liikmete kaalud peavad sel juhul olema võrdelised vastavate perioodide pikkustega, s.t. et nad peavad suhtuma omavahel samuti nagu vastavate perioodide pikkused

Olgu teada andmed kontrollkarja lehmade aastase piimatoodangu kohta Eestis¹

1924. a.		1954 kg	aastas
1925. a.		2229 kg	—"
1926. a.	} keskmiselt	2553 kg	—"
1927. a.			
1928. a.		2255 kg	—"
1929. a.	} keskmiselt		
1930. a.		2461 kg	—"
1931. a.			
1932. a.		2454 kg	—"
1933. a.		2356 kg	—"

Lehma aastakeskmise piimatoodangu saame sel juhul nii, et omistame kõigile ühe aasta kohta teada olevatele produktiivsustele kaalu 1, kahe aasta keskmisele kaalu 2 ja kolme aasta keskmisele kaalu 3. Vastavad arvutused on tabelis 8, kust selgub, et lehma keskmine aastatoodang oli neil aastail 2371,7 kg ($\frac{23717}{10}$).

Momentreas ei ole aritmeetiline keskmine kasutatav. Sel juhul tuleb arvutada kronoloogiline keskmine (vt. 3.2.).

¹ Andmed statistilisest aastaraamatust "Eesti Põllumajandus" 1933, Tallinn 1934, lk. 38.

Tabel 8.

Aasta	Piimatoodang ühe lehma kohta (kg) (x)	Kaal (f)	xf
1924	1954	1	1954
1925	2229	1	2229
1926-1927	2553	2	5106
1928	2235	1	2235
1929-1931	2461	3	7383
1932	2454	1	2454
1933	2356	1	2356
	x	10	23717

2.4.2. Aritmeetilise keskmise arvutamisest intervallrea andmetel.

Intervallitud variatsioonrea andmetel aritmeetilist keskmist leidma asudes tuleb kõigepealt vahemikena antud variantide väärtused asendada nende intervallikeskmistega. Et enamasti pole olemas informatsioon selle kohta, kuidas variantid intervallisiseselt jaotuvad, siis võetakse intervallikeskmiseks intervallirajade lihtne aritmeetiline keskmine, nagu sellele juba eespool on vihjatud. Sellega ühenduses tõuseb päevakorrale probleem intervallide piiramisel kasutatud töövõtetest, tekib küsimus, kas uuritav tunnus on sõre või pidev jne.

Oletagem, et on teada intervallidena antud variantide väärtused ja nende sagedused

x	f
4,0 - 5,9	9
6,0 - 7,9	20
8,0 - 9,9	32
jne.	

Kui mingisuguseid täiendavaid andmeid pole teada, siis on oletuste tegemiseks kaks võimalust. Võib eeldada, et variant on sõre ja et vahemikud 5,9 - 6,0, 7,9 - 8,0 jne. on tüh-

jad, siis tuleb võtta intervallikeskmisteks

$$\frac{4,0 + 5,9}{2} = 4,95 ; \quad \frac{6,0 + 7,9}{2} = 6,95 \text{ jne.}$$

Võib aga ka eeldada, et variant on pidev ja et rangelt piiratud intervalle on kasutatud ainult määramatuse vältimiseks, mis ligikaudu piiratud intervallide puhul oleks paratamatu. Sel puhul tuleb oletada, et variandil võib olla mis tahes väärtusi ka vahemikus 5,9 - 6,0 ja 7,9 - 8,0 jne. See- ga kuuluvad esimesse intervalli põhimõtteliselt kõik varian- did vahemikus 4,0 - 6,0, teises 6,0 - 8,0 jne. Järelikult tuleks intervallikeskmisteks võtta mitte 4,95, 6,95 jne., vaid

$$\frac{4,0 + 6,0}{2} = 5,0 ; \quad \frac{6,0 + 8,0}{2} = 7,0 \text{ jne.}$$

Täpsemate andmete puudumisel eeldatakse enamasti, et te- gemist on pidevalt varieeruva nähtusega ja talitatakse inter- vallikeskmiste leidmisel nagu viimasel juhul.

2.4.3. Lahtised Äärerühmad.

Lahtisteks Äärerühmadeks nimetatakse variatsioonrea kõige madalamat ja kõige kõrgemat intervalli, mille kohta on antud kas ainult üla- või alapiir. Ridade koostamisel tu- leks selliste rühmade kujundamist võimalikult vältida. Alati pole see aga teostatav.

Reas võib olla kas üks või kaks lahtist Äärerühma. Ühe lahtise Äärerühmaga rea näide leidub tabelis 1. Siin on pii- ramata kõige madalam rühm. Et on tegemist plaani täitmise protsentidega, mis ei saa olla väiksemad kui null, siis on sellegi intervalli alapiir (o) õigupoolest teada. Et inter- valli pole otse niiviisi märgitud, s.o. "o - 10", vaid "... - 10", vihjab sellele, et organisatsioon, kus plaani täitmise protsent oleks null, pole käsitledavas kogumis ilm- selt olemas.

Kahe lahtise Äärerühmaga variatsioonrea näide leidub

tabelis 9. Madalaima intervalli keskmise leidmisel võib oletada, et vähem kui üks päev keegi haiglas ei ole. Seega võib võtta intervalli alapiiriks 1 ja keskmiseks $3 (= \frac{1+5}{2})$ niisamasugustel kaalutlustel, millega leitakse kõigi teistegi intervallide keskmised.

T a b e l 9.

Haiglaravi kestus N linna haiglas.

Haiglas veedetud päevade arv	Inimeste arv	
x_1	f_1	$x_1 f_1$
1	2	3
... - 5	118	
6 - 10	120	8
11 - 15	131	13
16 - 20	140	18
21 - 25	123	23
26 - 30	113	28
31 - ...	129	

874

Raskem probleem tekib rea kõrgeima intervalliga. Pole mingit põhjust oletada, et üle 36 päeva (s.o. 31 + intervalli pikkus 5 päeva) ükski ravialune haiglas ei viibinud. Järelikult ei osutu õigeks ka mõeldav intervallikeskmine ($\frac{31+36}{2} =$) 33 päeva. Missugune see keskmine peaks olema, seda ei saagi täpselt arvutada, kui rea kohta pole võimalik hankida mingisugust täiendavat informatsiooni.

Kui on teada, et lahtine äärerühm sisaldab äärmiselt erineva väärtusega variante (nagu näiteks käesoleval juhul võib oletada), siis tuleb aritmeetilise keskmise arvutamisest enamasti üldse loobuda. Sellistes ridades saadakse tunduvalt tüüpilisem keskmine mediaanina (vt. osa 7).

Aritmeetiline keskmine on seda vähem kasutatav, mida suurem on hajuva koostisega lahtiste äärerühmade osatähtsus kogumis. On lahtise äärerühma osatähtsus suhteliselt väike, siis ei tarvitse sellest tekkiv viga keskmist eriti mõjustada. Kui vaadelda näiteks andmeid perekondade jaotumise kohta Eestis (tabel 10), siis on siingi viimane rühm lah-

tine: 10 ja rohkem liiget. Ent selliseid perekondi on Eestis ainult 218. Nende arv moodustab perekondade üldarvust vähem kui ühe promilli. Seetõttu kujuneb arvutamisel tekkiv viga nii väikeseks, et seda pole praktiliselt tarvis arvestada.

T a b e l 10 .

Perekondade jaotumine liikmete arvu järgi
Eestis ja Armeenias 1959. aastal.¹

Perekonnaliikmete arv	Perekondade arv	
	Eesti NSV-s	Armeenia NSV-s
2	121530	50026
3	92737	58113
4	57097	63367
5	23968	60647
6	8759	49615
7	2872	31908
8	1034	18303
9	389	9265
10 ja rohkem	218	8441
Kokku	308613	349685
Keskmine inimeste arv perekonnas ($\bar{x}$)	3,1	4,8

Tabelis 10 esinevad andmed on huvitavad eriti seetõttu, et nad demonstreerivad lahtiste äärerühmade esinemise võimalust ka momentridades.

Lahtiste äärerühmade käsitlemine ei tekitaks mingisuguseid raskusi, kui oleks täiendavalt teada, mitu haiglapäeva viibisid ravil kõik viimasesse rühma kuuluvad haiged kokku (tabel 9) või kui palju inimesi kuulub üldse perekondadesse, kus on üle 10 liikme (tabel 10) jne. Peetakse soovitavaks, et lahtiste äärerühmadega ridade koostamisel lisataks neile ka sellised andmed.

¹ Andmed Итоги Всесоюзной переписи населения 1959 года, СССР (сводный дом), ЦСУ СССР при Совете Министров СССР, Москва 1962, lk. 242.

2.4.4. Nn. isoleeritud keskmised.

Isoleeritud keskmisteks nimetatakse selliseid aritmeetilisi keskmisi, mida ei arvutata mitte aeg- või variatsioonide andmetel, vaid lihtsalt kahe suuruse suhtena - nn. intensiivsussuhtarvuna. Selle praktikas levinud termini kasutuselevõtmisel on peetud silmas, et keskmise arvutamine on siin statistiliste ridade töötlemisest isoleeritud. Tavaliselt, nagu see eespool ilmses, esineb keskmiste arvutamine statistiliste ridade töötlemise võttena.

Nagu suhtarvude teooriast teada, jagunevad intensiivsussuhtarvud omakorda tihedus- ja sagedussuhtarvudeks. Isoleeritud keskmise kui tihedussuhtarvu näiteks võib olla üldtuntud rahvastiku tiheduse näitarv, mis leitakse murruna

$$\frac{\text{rahvaarv}}{\text{territooriumi pindala (km}^2\text{)}} ,$$

ja mida võiks nimetada teisiti ka ühe ruutkilomeetri kohta tulevaks keskmiseks rahvaarvuks.

Sagedussuhtarvusi kui isoleeritud keskmisi esindavad näiteks suremuskoefitsiendid, sündimuskoeffitsiendid jms.

Õige tihti on niiviisi aritmeetilisi keskmisi hõlpsam arvutada kui vastavate ridade andmeil.

Kui on teada näiteks perekondades elavate inimeste üldarv (956344¹) ja perekondade üldarv (308613), siis võib leida perekonna keskmise suuruse ka lihtsa suhtarvuna

$$\frac{956344}{308613} = 3,1 \text{ inimest.}$$

¹ Perekondades elavate inimeste üldarvu leidmisel on perekonnaliikmete üldarvust 1038308 lahutatud perekondadest lahus elavad perekonnaliikmed 81964. Vt. Итсги Всесоюзной переписи населения 1959 года, Эстонская ССР, Москва 1962, lk. 101

2.5. Kaalude valik.

Tähtsamate ehitusorganisatsioonide plaani täitmise keskmise protsendi leidmisel on eespool (vt. lk. 27) käsitatud variantidena (x_1) üksikute organisatsioonide vastavaid protsente ja kaaludena (f_1) organisatsioonide arvu. See ei ole aga ainus võimalus.

Sama keskmise leidmisel võiks võtta kaaludeks ka veel näiteks ehitustööde üldise maksumuse iga vastava rühma organisatsioonides (kui need andmed oleksid olemas). Muidugi kujuneks nii saadud keskmine plaani täitmise protsent erinevaks sellest, mille leidmisel on kasutatud kaaludena organisatsioonide arvu.

Kumb neist kahest võimalusest siis valida?

Sellele ei saa olla mingit üldkehtivat vastust. Küsimus tuleb lahendada iga kord kooskõlas taotletava eesmärgiga.

Huvitab meid see, kuidas organisatsioonid (resp. ettevõtted, tehased jne.) oma tööülesannetega üldse toime tulevad, siis on loogiline kasutada esimest moodust. Sel juhul vaatleme kõiki organisatsioone ühe kogumi võrdväärsete liikmetena. Asjaolul, et neist mõni on oma töö mahu poolest suurem, mõni väiksem, ei ole olulist tähtsust.

Tahetakse keskmise arvutamisel saada arv, mis näitab, kui suureks kujunes plaani täitmise protsent kõigi tööde seotamisel, mis oli nende 2069 ehitusorganisatsiooni kätte usaldatud, siis on loogilisem võtta kaaluks toodangu suurused.

Viimasel juhul saadav keskmine on võrdne sellega, mis kujuneks nn. isoleeritud keskmisena, kui liidaksime kõigi organisatsioonide tegelikud ja plaanilised toodangud ja jagaksime esimese niiviisi saadud summa teisega.

Analoogilisi juhtumeid kohtab majanduslikus praktikas äärmiselt palju. Näiteks võib plaaniliste töönormide keskmise näitaja arvutamisel võtta kaaluks kas tööliste arvu (1) või plaaniliste normide absoluutsuurused (2).

Demonstreerime neid kahte erinevat kaalumisvõimalust ruumi kokkuhoiu mõttes mõnevõrra lihtsustatud arvnäite abil. Oletagem, et on teada 20 tööstusettevõtte jaotumine tootmisplaani täitmise astme järgi ja et täiendavalt on teada igasse rühma kuuluvate ettevõtete plaanilise toodangu üldsumma (vt. tabel 11). Tabeli esimeses reas olevaid andmeid tuleb mõista järgmiselt: tootmisplaani 80 - 100 %, s.o. keskmiselt 90 % ($= x_1$) täitnud ettevõtteid on arvult 4 ($= f_1$) ja nad pidiid plaani järgi andma kõik kokku 8 milj. rbl. ($= w_1$) eest toodangut.

Kasutades kaaluna ettevõtete arvu, saame plaani täitmise keskmiseks protsendiks

$$\bar{x} = \frac{\sum x_1 f_1}{\sum f_1} = \frac{2260}{20} = 113 \% .$$

Võttes aga kaaluks vastavatesse ettevõtteühmadesse kuuluva te tehaste toodangu plaanilised üldmaksumused, saame

$$\bar{x} = \frac{\sum x_1 w_1}{\sum w_1} = \frac{6100}{52} = 117 \% .$$

T a b e l 11 .

20 tööstusettevõtte jaotumine tootmisplaani täitmise järgi.

Plaani täitmise protsent	Keskmine plaani täitmise protsent rühmas	Ettevõtete arv	Rühma kuuluvate ettevõtete toodangu plaaniline kogumaksumus (milj. rbl.)	Keskmise leidmiseks vajalikud abiarvutused	
				$x_1 f_1$	$x_1 w_1$
x	x_1	f_1	w_1		
(1)	(2)	(3)	(4)	(5)	(6)
80 - 100	90	4	8	360	720
100 - 120	110	10	23	1100	2530
120 - 140	130	5	15	650	1950
140 - 160	150	1	6	150	900
Σ	Σ	20	52	2260	6100

Kummalgi neist keskmistest on oma majanduslik sisu, nagu sellele vihjatud. Seepärast ei saa tekkida küsimust, kumb neist on õigem. Küll aga võib tekkida küsimus, kas neid kasutatakse alati selle tõestamiseks, mida väita tahetakse.

Viimane plaani täitmise keskmine protsent (117 %) langeb täpselt kokku kõigi 20 ettevõtte tegeliku ja plaanilise toodangu kogumaksumuste suhtega ehk plaani täitmise indeksiga. Kõigi ettevõtete plaanilise toodangu kogumaksumus on teada (52 milj. rubl.). Tuletame plaani täitmise protsentide abil tegeliku toodangu maksumuse. See on

$$\frac{\sum x_i w_i}{100} = \frac{6100}{100} = 61.$$

Järelikult on tootmisplaani täitmise indeks

$$I_{\text{tootmispl. täitm.}} = \frac{61}{52} = 1,17 \text{ ehk } 117 \%.$$

Majanduslike ülesannete lahendamisel tekib sageli mitte ainult võimalus, vaid ka vajadus, kasutada samadest lähteandmetest keskmiste arvutamisel kord nii-, kord teistsuguseid kaalusid. Valik mitmesuguste võimaluste vahel tuleb teha vastavalt sellele, missuguse sisuga keskmise leidmist mingi konkreetse ülesande lahendamine eeldab.

Muidugi ei ole variantide kaalumise mitmesuguseid võimalusi valida kõigi ülesannete puhul. On küllalt lihtsa seadega probleeme, kus tulevad kõne alla ainult mingid ühed kaalud.

Lisamärkusena olgu juhitud tähelepanu sellele, et esitatud näidetes on plaani täitmise keskmise protsendi arvutamisel valida, kas kasutada kaaluna ettevõtete arvu või nende toodangu plaanilist maksumust. Ettevõtete toodangu tegeliku maksumuse kaaluna kasutamine ei tule antud juhul kõne alla. Põhimõtteliselt on see küll võimalik, ent siis tuleb loobuda aritmeetilisest keskmisest ja arvutada harmooniline (vt. ptk. 4.3.).

2.6. Tsentrograafiaprobleemide lahendamine.

Tsentrograafilisteks nimetatakse selliseid majanduslike probleeme, mis lahenduvad analoogiliselt raskuskeskme leidmisega. Tüüpiline tsentrograafiaprobleem on näiteks tehase paigutamine kolme linna vahelisele tühjale maa-alale nõnda, et nendest linnadest tehases tööl käima hakkavatel inimestel oleks tarvis iga päev läbida kokku minimaalse pikkusega tee.

Kui kõik muud küsimused kõrvale jätta, siis mõjustavad selle ülesande lahendamist ainult linnade asukohad ja nende elanike arvud. Oletagem, et vastavad andmed on kolme linna A, B, C kohta järgmised:

T a b e l 12 .

Linna koordinaadid	Tähis	A	B	C
Pikkus 12°	λ	15'	30'	35'
Laius 52°	φ	30'	40'	25'
Elanike arv	f	1000	5000	3000

Et linnade geograafilistes koordinaatides on kraadid võrdsed, siis võib need jätta kõrvale ning kogu arvutus sõltuda ainult minutitega.

Tehase asukoha leidmiseks arvutame tema geograafilise pikkuse ja laiuse eraldi, kasutades kummalgi juhul kaalutud aritmeetilise keskmise valemit. Nii saame keskmiseks pikkuseks

$$\bar{\lambda} = \frac{\sum \lambda_i f}{\sum f} = \frac{15 \cdot 1000 + 30 \cdot 5000 + 35 \cdot 3000}{1000 + 5000 + 3000} = \frac{270000}{9000} =$$

$$= 30', \text{ s.o. } 12^{\circ} 30'$$

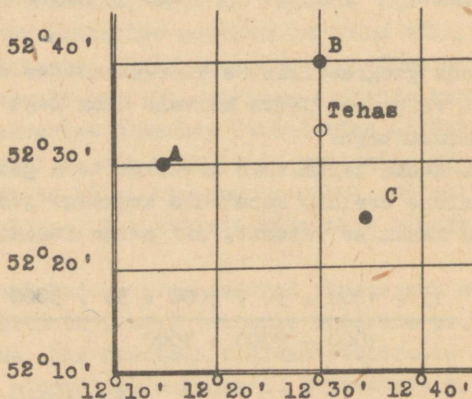
$$\bar{\varphi} = \frac{\sum \varphi_i f}{\sum f} = \frac{30 \cdot 1000 + 40 \cdot 5000 + 25 \cdot 3000}{1000 + 5000 + 3000} =$$

$$= \frac{305000}{9000} = 34'.$$

Seega on tehase asukoha koordinaadid $\lambda = 12^\circ 30'$ ja $\varphi = 52^\circ 34'$. Nagu jooniselt selgub, peaks tehase paiknema kõige suuremale linnale kõige lähemal.

Muidugi ei saa seda viisi tehaste asukohtade määramisel praktikas ilma korrektiivideta kasutada, sest peale linnade suuruse ja asukoha tuleb arvestada veel paljusid muidki asjaolusid - teedevõrku, pinnast, kasutatavat transpordiliiki, veotariife, toodangu äravedu, tootmiseks vajalike veevarude olemasolu jne.

Komplitseeritumad tsentrograafilised meetodid võimaldavad sobivate asukohtade leidmisel arvestada korraga paljude tegurite mõju. Siiski tuleb nendegagi leitud lahendusi tihti peale korrigeerida, sest peale kvantitatiivselt mõõdetavate tegurite avaldavad mõju ka paljud kvalitatiivsed (näit. ma-



Joon. 3. Tehase asukoha leidmine kolme erineva suurusega linna vahealal (vastavalt lk.53 esitatud tingimustele).

janduspoliitilised) tegurid, mille mõju ei saa kvantitatiivselt mõõta ega arvutustesse lülitada.

3. KRONOLOOGILINE KESKMIINE.

3.1. Aritmeetiline ja kronoloogiline keskmine.

Kronoloogiline keskmine on tuletatud aritmeetilisest keskmisest. Seda kasutatakse üksnes aegride (teisiti: "kronoloogiliste ridade") uurimisel, millele viitab ka nimetus.

Nagu teada, jagunevad aegread periood- ja momentridadeks. Perioodridade puhul saab kasutada lihtsat aritmeetilist keskmist (vt. 2.11.).

Momentridades on iga selle rea liige mingi suuruse väärtus just sellel momendil. Momentrea näiteks võib olla andmete rida mingi ettevõtte kassas kuude alguskuupäevadel olnud sularaha summa kohta:

Kuupäev	1.I	1.II	1.III	1.IV	1.V	1.VI	1.VII
Rbl.	50	180	30	670	580	400	1500

Kassas oleva sularaha jääke 1. jaanuaril, 1. veebruaril jne. pole mõtet liita; nii saadav summa oleks mõttetu. Analoogiliselt oleks mõttetu ka aritmeetiline keskmine, mis saadaks, kui nimetatud summa jagataks vastavate kuude arvuga. Et moment- ja perioodsuuruste jagatisel pole mõtet, siis tuleb nad muuta üht laadi suurusteks. Lihtsam on muuta momentsuurused, s.o. kassa bilansilised saldod perioodsuurusteks. Kõige hõlpsam moodus selleks on arvutada üksikute kuude keskmised kassasaldod. Kui mingisuguseid muid täiendavaid andmeid pole käepärast, võib kuu keskmise saldo leida kuu alguse ja lõpu saldo lihtsa aritmeetilise keskmisena. Tähistame kassasaldo tähega $\bar{s}$ ja momentrea iga järjekordse liikme väärtuse s_1 . Sel juhul on jaanuari keskmine saldo

$$\bar{s}_I = \frac{s_{I.I} + s_{I.II}}{2},$$

veebruari keskmine saldo

$$\bar{s}_{II} = \frac{s_{I.II} + s_{I.III}}{2}$$

ja märtsi keskmine saldo

$$\bar{s}_{III} = \frac{s_{I.III} + s_{I.IV}}{2}.$$

Saadud keskmised saldod $\bar{s}_I$, $\bar{s}_{II}$ ja $\bar{s}_{III}$ on perioodsuurused. Nii on esialgne momentrida muundunud perioodreaks ja selle liikmetest võib arvutada lihtsa aritmeetilise keskmise. Leiame nüüd I kvartali keskmise saldo $\bar{s}_I$ kv., mis on

$$\bar{s}_I \text{ kv.} = \frac{\bar{s}_I + \bar{s}_{II} + \bar{s}_{III}}{3}.$$

Esitatud arvnäite andmetel on

$$\bar{s}_I = \frac{50 + 180}{2} = 115;$$

$$\bar{s}_{II} = \frac{180 + 30}{2} = 105;$$

$$\bar{s}_{III} = \frac{30 + 670}{2} = 350$$

ja

$$\bar{s}_I \text{ kv.} = \frac{115 + 105 + 350}{3} = \frac{570}{3} = 190.$$

Viimast arvu (190 rbl.) võibki pidada keskmiseks kassas olnud sularaha summaks esimesel kvartalil. Leidsime selle arvu kahe järgus. Esiteks arvutasime momentrea lähteand-

metest üksikute kuude keskmised. Seejärel leidsime kuukesmistest kvartalikeskmise.

3.2. Kronoloogilise keskmise üldvalem.

Kui käsitleda läbitud arutluskäiku üldistatud kujul, selgub, et niisuguse kaheastmelise arvutusteseeria läbitegemiseks pole mingisugust vajadust. Võime tuletada valemi, mis lubab üldaloodud keskmist arvutada tunduvalt hõlpsamini.

$$\begin{aligned}\bar{s}_I \text{ kv.} &= \frac{\bar{s}_I + \bar{s}_{II} + \bar{s}_{III}}{3} = \\ &= \frac{\frac{s_{1,I} + s_{1,II}}{2} + \frac{s_{1,II} + s_{1,III}}{2} + \frac{s_{1,III} + s_{1,IV}}{2}}{3} = \\ &= \frac{\frac{s_{1,I}}{2} + s_{1,II} + s_{1,III} + \frac{s_{1,IV}}{2}}{3}.\end{aligned}$$

Kui tähistada viimase murru lugejas olevad suurused lihtsalt x_1 , x_2 jne. ning nende arvu (s.o. momentide arvu reas) n , siis võib saadud valemi kujutada järgmisel üldisel kujul

$$\bar{x}_{\text{kon.}} = \frac{\frac{x_1}{2} + x_2 + \dots + x_{n-1} + \frac{x_n}{2}}{n-2}, \quad (8)$$

kus $x_1 = s_1$, s.o. mingi momentsuurus aegreas.

Sellisel kujul avaldatud keskmist nimetataksegi kronoloogiliseks keskmiseks. Nagu eespool toodud käsitlusest järeldub, on kronoloogiline keskmine oma saamisloolt momentrea andmetel arvutatud aritmeetiliste keskmiste aritmeetiline keskmine.

Kirjanduses esineb kronoloogilise keskmise valem sageli

ka veidi teisendatud kujul, kas näiteks

$$\bar{x}_{\text{kron.}} = \frac{1}{n-1} \left(\frac{x_1 + x_n}{2} + \sum_{i=2}^{n-1} x_i \right)$$

ehk

$$\bar{x}_{\text{kron.}} = \frac{x_1 + 2x_2 + \dots + 2x_{n-1} + x_n}{2(n-1)}.$$

Eespool toodud arvnäite kohaselt on I kvartali keskmine saldo

$$\begin{aligned}\bar{s}_I &= \frac{\frac{50}{2} + 180 + 30 + \frac{670}{2}}{3} = \\ &= \frac{25 + 180 + 30 + 335}{3} = \frac{570}{3} = 190 \text{ rubl.}\end{aligned}$$

ja esimese poolaasta keskmine saldo

$$\begin{aligned}\bar{s} &= \frac{\frac{50}{2} + 180 + 30 + 670 + 580 + 400 + \frac{1500}{2}}{6} = \\ &= \frac{2635}{6} = 439,17 \text{ rubl.}\end{aligned}$$

Kronoloogilise keskmise valemi kasutamisest saadav ajasääst on ilmne. Poole aasta keskmise saldo arvutamisel oleks tulnud teha kokku 14 arvutustehet, s.o. leida 7 summat ja 7 jagatist. Kronoloogilise keskmise arvutamisel piisab kokku 4 tehtest (tuleb leida 3 jagatist ja 1 summa).

3.3. Kronoloogilise keskmise praktiline kasutamine.

Momentridade keskmisi saab leida ainult kronoloogilise keskmise valemi alusel. Et raamatupidamisbilansid koostatak-

se alati mingiks kindlaks momendiks (1. kuupäevaks), siis on kõik bilansiaandmetest koosnevad read momentread ja neist saab arvutada ainult kronoloogilisi keskmisi.

Ülesandeid, mida majandusettevõtete praktikas lahendatakse selle keskmisega, on härmiselt palju. Loetleme neist mõningaid: tööliste, autode, masinate, loomade jne. keskmiste arvude leidmine üksikute kuude esimesteks päevadeks teada olevate arvude alusel; materjalide, kaupade, kokkuostetud põllumajandussaaduste ja käibevahendite keskmiste varude suuruse leidmine jne.

Kuigi kronoloogilise keskmise näol on tegemist suhteliselt lihtsa näitarvuga ja tema kasutusala on väga hõlpus piirata, tuntakse teda praktikas siiani suhteliselt vähe. Esineb ikka veel juhtumeid, kus näiteks keskmiste saldode, keskmiste kaubavarude suuruse jm. selletaoliste näitajate arvutamisel kasutatakse aritmeetilist keskmist. Halb pole seejuures mitte ainult see, et niimoodi saadakse väärad keskmised. Halvem on, et lülitades ebaõigesti leitud keskmised mitmesugustesse muudesse analüütilistesse arvutustesse, saadakse ka nende tulemused tunduvalt moonutatud kujul. Mii võib kogu analüüs mõttetaks osutuda.

Kahjuks ei eksita kronoloogilise keskmise kasutamisel mitte ainult praktikas, vaid jämedaid vigu leidub ka õpikutes. Täiesti desorienteerivaid soovitusi keskmiste varude arvutamise kohta antakse näiteks mõnedes kaubandusökoonoomika õpikutes. Suure autorite kollektiivi koostatud õpiku "Nõukogude kaubanduse ökonoomika" põllumajandussaaduste varumist käsitlevas peatükis soovitab selle autor prof. Dneprovski arvutada laos olevate põllumajandussaaduste kvartalikeskmise varu suurus valemiga

$$\bar{s}_{IV \text{ kv.}} = \frac{s_{1.IX} + s_{1.X} + s_{1.XII} + s_{1.I}}{4},$$

kus $s_{1.IX}$, $s_{1.X}$ jne. on varud vastavate kuude esimestel kuupäevadel.¹ Et siin on tegemist momentreaga, siis on sel-

¹ vt. Экономика советской торговли, Учебник, Госполитиздат, Москва 1962, lk. 125 ja 300.

line arvutus mõttetu. Samasuguseid desorienteeruvaid näpunäiteid leidub kahjuks ka paljudes teistes õpikutes.

4. HARMOONILINE KESKMIINE.

4.1. Harmonilise keskmise valem.

Formaalselt kujutab harmooniline keskmine endast aritmeetilise keskmise pöördsuurust, mis on leitud variantide pöördväärtusest. See selgub nende valemite võrdlemisest. Nagu aritmeetilisel keskmisel, nii on ka harmoonilisel kaks avaldamiskuju:

lihtne harmooniline keskmine

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x}} \quad , \quad (9)$$

mida kasutatakse lihtsate ridade puhul ja

kaalutud harmooniline keskmine, mida arvutatakse variatsiooniridade andmetel

$$\bar{x}_h = \frac{\sum f}{\sum \frac{1}{x} f} \quad . \quad (10)$$

Harmonilist keskmist saab arvutada täpselt samade andmete alusel, mis on vajalikud aritmeetilise keskmise leidmiseks. Järelikult pole mingisugust puhtformaalset takistust arvutada mistahes statistilisest reast ükskõik kumb, kas aritmeetiline või harmooniline keskmine. Ometi selgub praktiliste ülesannete lahendamisel, et tõsioludele vastavaid tulemusi võimaldab neist saada igal üksikul konkreetsel juhul ainult üks, kas aritmeetiline või harmooniline keskmine.

Esitame selle kohta lihtsa näite.

Oletagem, et kaks autot läbisid kumbki 100 km, s.o.

kokku 200 km. Auto A sõitis selle maa läbi kiirusega 60 ja auto B 40 km/tunnis. Kui suur oli keskmine autode sõidukiirus?

Püüame seda ülesannet lahendada kõigepealt aritmeetilise keskmisega. Selgub, et

$$\bar{x} = \frac{40 + 60}{2} = 50 \text{ km/t.}$$

Võime siin kasutada lihtsat keskmist, sest variantide kaalud (kummagi poolt läbitud tee pikkused) on võrdsed, s.o. $f_A = f_B = 100 \text{ km}$. Täpselt samasuguse tulemuse annaks ka kaalutud keskmine

$$\bar{x} = \frac{40 \cdot 100 + 60 \cdot 100}{100 + 100} = 50 \text{ km/t.}$$

Saadud keskmise õigsust saab kontrollida. Kui autode keskmine kiirus oli 50 km/t. ja nad läbisid üldse 200 km, siis pidi neil selleks kuluma kokku 4 tundi ($200 : 50 = 4$). Tegelikult kummagi auto poolt sõiduks kulutatud aegade summa peaks olema sellega võrdne. Leiame, et auto A kulutas sõiduks

$$\frac{100}{60} = 1,67 \text{ tundi}$$

ja auto B

$$\frac{100}{40} = 2,50 \text{ tundi,}$$

s.o. kokku 4,17 tundi. Meie leitud keskmist kiirust arvestades aga võis kulutada ainult 4 tundi. Järelikult pole aritmeetiline keskmine antud juhul sisuliselt õige, see annab liiga suure keskmise kiiruse.

Õige tulemuse saame harmoonilise keskmisena:

$$\bar{x}_{\text{harm.}} = \frac{100 + 100}{\frac{1}{40} \cdot 100 + \frac{1}{60} \cdot 100} = 48 \text{ km/t.}$$

Kontrollime sedagi tulemust. Kiirusega 48 km/t. kulus autodel 200 km läbimiseks kokku

$$\frac{200}{48} = 4,166 \approx 4,17 \text{ tundi,}$$

mis langeb täpselt kokku autode tegelike sõidukestuste sumмага. Järelikult annab harmooniline keskmine siin ainuõige tulemuse.

4.2. Millal kasutada harmoonilist, millal aritmeetilist keskmist?

Eespool kirjeldatud tööviisi kasutamine, mis eeldab mõlema keskmise arvutamist ja tulemuste "proovimist", et neist õige välja valida, on võrdlemisi tülikas. Seepärast taotleatakse anda käepärasemat reeglit, mis lubaks juba enne arvutamisele asumist kindlaks teha, kumb keskmine arvutada. Kirjanduses leidub selleks palju mitmesuguseid reegleid ja soovitusi, ent enamasti on need küllaltki keerulised.

Meie arvates on kõige lihtsam ja "töökindlam" järgmine kolmest sammust koosnev arutlusk skeem, mis hõlmab tegelikult kogu vastava ülesande lahendusakäigu kirjelduse.

1. samm. Määratakse kindlaks, missugune suurus on variant ja missugune kaal. Variantiks on ikka see, mille keskmist otsitakse. Eespool toodud näidisülesandes taotleme leida keskmist kiirust. Järelikult on kiirus variant ja läbitud tee pikkus kaal.

2. samm. Leitakse mingi kolmas suurus, mille kaudu kaal ja variant on omavahel seotud ning avaldatakse need võrrandina nii, et variant on avaldise ühel, kaal koos neid ühendava suurusega teisel poolel. Meie näites on teada, et kiirust ja tee pikkust ühendavaks suurus on aeg, kusjuures

$$\text{kiirus} = \frac{\text{tee}}{\text{aeg}} .$$

Saadud valemit nimetatakse seosvalemiks.

3. samm. Valitakse sobiv keskmine, olenevalt sellest, mida kasutatakse kaaluna. Kui kaal võetakse seosvalemi lugejast, siis tuleb arvutada harmooniline keskmine, kui nimetajast, siis aritmeetiline.

Eespool toodud näidisülesandes olid kaaludena kasutusel tee pikkused. Järelikult tuli kasutada harmoonilise keskmise valemit. Nagu nägime, tagaski just see ainuõigeid tulemusi.

Oleks muidu analoogilise seadega probleemi lahendamisel olnud kaaludena kasutada autode sõiduajad, tulnuks kasutada aritmeetilist keskmist.

Sama reegli võib sõnastada ka lühemalt: kui otsitav keskmine suurus on kaaludest võrdelises olenevuses, tuleb arvutada harmooniline, kui pöördvõrdelises, siis aritmeetiline keskmine.

4.3. Harmoonilise keskmise praktiline kasutamine.

Aritmeetilise ja harmoonilise keskmise kasutamise juhtumite eristamise kriteerium määrab tegelikult ära ka harmoonilise keskmise kasutamise juhtumid majanduslikus praktikas. Võib ütelda, et ülesandeid, kus ainuõigeid tulemusi tagab ainult harmooniline keskmine, esineb praktikas väga tihti. Taotlemata nende ammendavat loetlemist, esitame tabeli vormis lühikese ülevaate harmoonilise keskmise kasutamise peamistest juhtudest (vt. tabel. 12).

Eespool on võrdlemisi üksikasjaliselt käsitletud plaani täitmise keskmise protsendi arvutamist (vt. 2.5). Seejuures sai mõõdamannes märgitud, et aritmeetilise keskmise leidmisel tuleb kõne alla kasutada ainult toodangu plaanilist, mitte tegelikku maksumust. Kui eespool toodud arutluskäiku järgides pidada silmas, et

$$\text{plaani täitmise näitaja} = \frac{\text{aruandeline näitaja}}{\text{plaaniline näitaja}},$$

ja suhtuda sellesse kui nn. seosvalemisse, siis on arusa-

T a b e l 12 .

Keskmise valik mõnede praktikas tihti
esinevate ülesannete lahendamisel.

Otsitav suurus	Seosvalem	Kaaluks kasutatav suurus	Keskmise liik
Keskmine kiirus	$kiirus = \frac{tee}{aeg}$	$\frac{tee\ pikkus}{sõidu\ kestus}$	$\frac{harm.}{aritm.}$
Keskmine sõiduaeg	$aeg = \frac{tee}{kiirus}$	$\frac{tee\ pikkus}{kiirus}$	$\frac{harm.}{aritm.}$
Keskmine hind	$hind = \frac{maksumus}{kogus}$	$\frac{maksumus}{kogus}$	$\frac{harm.}{aritm.}$
Keskmine kogus	$kogus = \frac{maksumus}{hind}$	$\frac{maksumus}{hind}$	$\frac{harm.}{aritm.}$
Keskmine tööviljakus	$tööviljakus = \frac{toodang}{tööliste\ arv}$	$\frac{toodang}{tööliste\ arv}$	$\frac{harm.}{aritm.}$
Keskmine tööliste arv	$tööliste\ arv = \frac{toodang}{töövilj.}$	$\frac{toodang}{tööviljakus}$	$\frac{harm.}{aritm.}$
Keskmine materjali kulunorm	$materjali\ kulunorm = \frac{materjali\ kogus}{toodete\ arv}$	$\frac{materjali\ kogus}{toodete\ arv}$	$\frac{harm.}{aritm.}$
Keskmine plaani täit- mise näitaja	$plaani\ täitmise\ näitaja = \frac{aruanne}{plaan}$	$\frac{aruanne}{plaan}$	$\frac{harm.}{aritm.}$
Toote kesk- mine oma- hind	$omahind = \frac{tootmiskulu\ summa}{toodete\ arv}$	$\frac{tootmiskulude\ summa}{toodete\ arv}$	$\frac{harm.}{aritm.}$
Keskmine saagikus	$saagikus = \frac{saak}{külvipind}$	$\frac{saak}{külvipind}$	$\frac{harm.}{aritm.}$

dav, miks see nii on. Niisamuti on ka selge, et plaani täitmise näitajaid tegelike andmetega kaaludes tuleb kasutada just harmoonilist keskmist. See kehtib loomulikult mitte ainult tootmisplaani, vaid üldse kõigi plaani täitmise keskmiste näitajate kohta.

4.4. Diskussioon harmoonilise keskmise olemusest.

NSV Liidu ja rahvademokraatiamaade statistikakirjanduses on viimase kümne aasta jooksul arutatud korduvalt harmoonilise keskmise kasutamisega seotud küsimusi. Peamisi probleeme on olnud tema "iseseisvus": kas harmooniline keskmine on omaette keskmiseliik või tuleb teda vaadelda aritmeetilise keskmise lihtsa teisendina?

Viimase seisukoha pooldajatena on esinenud prof. D. Oparin ja M. Iljevski. Nad väidavad, et harmoonilise keskmise näol on tegemist üksnes teise viisiga aritmeetilise keskmise arvutamiseks (D. Oparin) või aritmeetilise keskmise arvutamise lihtsustatud võttega (M. Iljevski).¹ Põhimõtteliselt samasuguseid arvamusi leidub ka mõnedes üldkasutatavates õpikutes.

Neile seisukohtadele vaidleb väga veenvalt vastu F.D. Livšits, kes põhjalikus teoreetilises uurimuses tõestab, et harmoonilist keskmist ei saa vaadelda aritmeetilise keskmise "lihtsa tehnilise teisendina".²

¹ Vt. D. Oparini ja M. Iljevski artikleid ajakirjas "Вестник статистики" 1955, 4.

² Б.Д. Лившиц, О дисгармонической средней, АН СССР, Ученые записки по статистике, том 17, Москва 1959, 312-331.

5. GEOMETRILINE KESKINE.

5.1. Aritmeetiline ja geomeetriline proportsioon.

Aritmeetilisest proportsioonist selgub, mitme ühiku võrra mingi väärtus on teisest suurem või väiksem; geomeetriline proportsioon näitab seevastu, mitu korda üks väärtus teisest erineb.

Kui on teada järgmine aritmeetiline proportsioon

$$a - b = b - c ,$$

siis on selle mõlemal poolel esinev liige ehk ühisliige võrdne a ja c aritmeetilise keskmisega. Sest

$$a + c = 2b ,$$

kust

$$b = \frac{a + c}{2}$$

Geomeetrilist proportsiooni

$$\frac{a}{b} = \frac{b}{c}$$

teisendades näeme, et

$$ac = b^2 ,$$

kust

$$b = \sqrt{ac} .$$

Saadud avaldis ongi geomeetrilise keskmise valem. Võime öelda, et geomeetrilise proportsiooni ühisliige on võrdne tema teiste liikmete geomeetrilise keskmisega.

Kahe proportsiooniliigi erisustega ja kaudselt ka geomeetrilise keskmise olemusega on seotud teaduse ajaloost hästi tuntud vaidlus Galilei ja kellegi Nocchioli vahel XVII sajandil. Vaidlus tekkis küsimusest, kumb eksib rohkem, kas see ostja, kes arvab, et tegelikult 100 krooni väärt oleva hobuse hind on 1000 krooni või see, kes arvab, et hobuse hind on ainult 10 krooni.

Absoluutarvudes mõõtes eksib muidugi esimene rohkem, tervelt 900 (= 1000 - 100) krooni võrra, sellal kui teise eksitus on ainult 90 (= 100 - 10) krooni. Kui aga seada küsimus teisiti: mitmekordselt neist üks või teine hobuse odavamaks või kallimaks hindas, siis selgub, et mõlemad eksisid ühepalju. Esimene hindas hobuse 10 korda kallimaks kui see tegelikult oli, teine 10 korda odavamaks.

Viimases käsitluses eksisid mõlemad ühepalju seetõttu, et hobuse tõeline väärtus on kummagi hindaja poolt nimetatud väärtuste keskmine, nimelt geomeetriline keskmine, s.o.

$$\sqrt{10 \cdot 1000} = 100.$$

Sellest näitest selgub, et geomeetriline keskmine on kordsete (multiplikatiivsete) suuruste keskmine. (Aritmeetiline keskmine on seevastu liidetavate suuruste keskmine.)

5.2. Geomeetrilise keskmise valem. Logaritmkeskmine.

Nagu sellele viitas juba eespool toodud mahukeskmiste üldvalem, leitakse geomeetriline keskmine põhimõtteliselt samuti nagu aritmeetiline. Kõik tehted aga asenduvad astme võrra kõrgematega - liitmine asendub korrutamisega, jagamine juurimisega, korrutamine astendamisega.

Geomeetrilise keskmise valem avaldatakse enamasti üldkujul

$$\bar{x}_{\text{geom.}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (11)$$

ehk

$$\bar{x}_{\text{geom.}} = \sqrt[n]{\prod_{i=1}^n x_i}.$$

Et valemi (11) kasutamine on arvutustehniliselt võrdlemisi

tülikas, siis kasutatakse seda enamasti logaritmitud kujul

$$\log \bar{x}_{\text{geom.}} = \frac{1}{n} \sum_{i=1}^n \log x_i \quad . \quad (12)$$

Viimases avaldumisvormis on seda kirjanduses nimetatud ka logaritmeskmiseks.

Kaalutud geomeetrilises keskmises esinevad variantide kaalud mitte teguritena, vaid astendajatena ja juurijaks on vastavalt asendajate summa.

$$\bar{x}_{\text{geom.}} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n}} .$$

Logaritmitud kujul

$$\log \bar{x}_{\text{geom.}} = \frac{\sum_{i=1}^n f_i \log x_i}{\sum_{i=1}^n f_i} .$$

5.3. Geomeetrilise keskmise praktilisest kasutamisest.

Geomeetrilise keskmise abil saab lahendada hulga matemaatilisi probleeme. Leidnud näiteks mingi ristküliku või risttahuka külgede keskmise pikkuse, võib konstrueerida ruudu või kuubi, mille pind- või ruumala on sama suur kui vastaval lähtekujundil. Seda laadi ülesandeid tuleb vahel lahendada majanduslike arvjooniste koostamisel.

Kõige sagedamini kasutatakse aga majandusstatistikas geomeetrilist keskmist aegridade uurimisel, nimelt keskmise kasvutempo arvutamisel.

Kui on teada aegrida, mille absoluutsed väärtused on

$$y_1, y_2, \dots, y_n$$

kus y tähistab näiteks piimatoodangut mingil aastal, siis võib neil andmetel leida kokku $n-1$ ahelindeksit ehk aastast kasvutempot, mis näitab kui intensiivselt on toimunud y kasv aastast aastasse

$$i_{y(2/1)} = \frac{y_2}{y_1} ; \quad i_{y(2/3)} = \frac{y_3}{y_2} ; \quad \dots ;$$

$$i_{y(n/n-1)} = \frac{y_n}{y_{n-1}} .$$

Nagu teada, annab ahelindeksite korrutis viimase aasta alusindeksi,¹

$$\begin{aligned} i_{y(2/1)} \cdot i_{y(3/2)} \cdot \dots \cdot i_{y(n/n-1)} &= \\ &= \cdot \frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \dots \cdot \frac{y_n}{y_{n-1}} = \frac{y_n}{y_1} = i_{y(n/1)} . \end{aligned}$$

Ahelindeksite liitmisel ja seega ka nende aritmeetilisel keskmisel pole praktilist mõtet. Et individuaalindeksitest saab ühes arvus väljenduva mõttega üldistuse ainult nende korrutamise teel, siis tähendab see, et ka nende keskmise arvutamisel tuleb kasutada niisugust keskmist, mis eeldab lähtesuuruste korrutamist omavahel. Teisiti väljendatult tähendab see, et mistahes nähtuse keskmine kasvutempo (keskmine ahelindeks) tuleb arvutada geomeetrilise keskmisena.

Elektrienergia toodangu keskmine kasvutempo oli tabel 13 andmeil näiteks ajavahemikul:

$$\begin{aligned} 1951 - 1955 \\ I_{\text{geom.}} &= \sqrt[4]{1,256 \cdot 1,116 \cdot 1,161 \cdot 1,101} = \\ &= \sqrt[4]{1,792} = 1,157 \text{ s.o. } 15,7 \% \text{ aastas.} \end{aligned}$$

¹ Ahel- ja alusindeksite seoseid käsitletakse üksikasjalisemalt indeksteoorias. Vt. U. Mereste, Indeksiteooria, TRÜ, Tartu 1961, lk. 13.

T a b e l 13 .

Elektrienergia toodang
Eesti NSV-s.

Aasta	Milj. kWt	Indeks eelmise aasta suhtes (ahel- indeks)
1940	190,0	-
1945	123,6	-
1950	435,6	-
1951	525,3	-
1952	659,6	1,256
1953	736,1	1,116
1954	854,3	1,161
1955	940,9	1,101
1956	1024,8	1,089
1957	1093,9	1,097
1958	1159,2	1,060
1959	1267,5	1,094

Tulemust kontrollides võib veenduda, et

$$1,157^4 = 1,7920 .$$

Geomeetrilise keskmise laialdasemat praktilist kasutamist takistab majanduselus see, et algandmete reast, milles on ka negatiivseid ja nulliga võrdseid liikmeid, ei saa teda arvutada. Keskmise tempo arvutamisel pole see omadus aga mingiks takistuseks, sest individuaalsete tempode seas selliseid suurusi praktiliselt ei esine.

5.4. Geomeetrilise keskmise arvutamise lihtsam erijuhtum.

Elkirjeldatud viisi (valemit 11) on geomeetrilise keskmise arvutamiseks otstarbekas kasutada siis, kui indeksid on eelnevalt olemas. Kui aga lähteandmeteks on absoluutarvud, on geomeetrilise keskmise leidmiseks olemas palju otstarbekam viis. Meenutame, et ahelindeksite korrutis annab viimase alusindeksi ja et see omakorda pole midagi muud kui absoluutarvude rea viimase ja esimese liikme jagatis. See tähendab, et geomeetrilise keskmise valemis võib juuremärgi all oleva indeksi korrutise asendada suuruste y_n ja y_1 jagatisega. Nii saame geomeetrilise keskmise lihtsustatud arvutamiseks uue valemi

$$\bar{I}_g = \sqrt[n-1]{\frac{y_n}{y_1}} . \quad (13)$$

Selles valemis väärrib tähelepanu eeskätt juurija $n - 1$. Juurijaks on seega arv, mis on ühe võrra väiksem kui vastava aegrea liikmete arv. Arvestades, et ka indeksi arv on alati 1 võrra väiksem kui liikmete arv (2 liikmest saab arvutada 1 indeksi, 3 saab arvutada 2 indeksit jne.), tähendab see teiste sõnadega, et ka siin on juurija võrdne juuremärgi all mõtteliselt olevate ahelindeksite arvuga, ehkki seal on käesoleval juhul tegelikult absoluutarvud.

Geomeetrilise keskmise juurijaks oleva arvu õige kindlaksmääramine on arvutustulemuste õigsuse tagamiseks väga tähtis. Kõige sagedamini tehakse vigu just selles.

Viimane moodus geomeetrilise keskmise arvutamiseks on eriti soodus sellistel puhkudel, kus saab tugineda ainult lünklikele andmetele. Valemite 11 ja 12 abil, mis eeldavad kõigi ahelindeksite teadmist, ei saaks keskmist tempot sel juhul üldse leida.

Tabelis 12 sisalduvatest andmetest ei piisa näiteks elektrienergia toodangu kasvutempo arvutamiseks aastail 1945 - 1950, kui selleks tahetakse kasutada valemit (11). Valemi (13) järgi on see aga kergesti leitav.

$$I_{\text{geom.}} = \sqrt[50-45]{\frac{525,3}{435,6}} = \sqrt[5]{1,2059} = 1,038.$$

Ent valem (13) on kasutamiseks otstarbekam ka enamasti siis, kui on teada aegrea kõik liikmed. See on mõneti hõlpsam, sest jääb ära ahelindeksite arvutamine ja nende korrutamise.

Nii võib näiteks eespool arvutatud elektrienergia toodangu aastakeskmise kasvutempo ajavahemikul 1951 - 1955 leida tunduvalt väiksema vaevaga

$$I_{\text{geom.}} = \sqrt[55-51]{\frac{940,9}{525,3}} = \sqrt[4]{1,791} \approx 1,157.$$

Mida pikemat ajavahemikku käsitleda, seda otstarbekam on leida geomeetriline keskmine mitte ahelindeksite korrutisest, vaid vastavast alusindeksist.

Ajavahemiku 1945 - 1959 kohta on näiteks elektrienergia toodangu kasvanud keskmiselt

$$\sqrt[59-45]{\frac{1267,5}{123,6}} = \sqrt[14]{10,25} \approx 1,18,$$

s.o. 18 % iga aasta. Sama vastuse saamine valemi (11) abil

oleks eeldanud 14 ahelindeksi leidmist ja nende omavahelist korrutamist, mis oleks tähendanud 17 liigse korrutamise- ja jagamistehte sooritamist.

Üldse on valemit (11) geomeetrilise keskmise arvutamiseks mõtet kasutada ainult siis, kui ainsateks andmeteks on ahelindeksid, rea liikmete absoluutseid väärtusi aga pole millegipärast teada.

5.5. Diskussioon geomeetrilise keskmise kasutatavuse üle.

Geomeetrilise keskmise kasutamise võimalustest on avaldatud õigegi vastuolulisi arvamusi. Sel pinnal arenenud teaduslik mõttevahetus hõlmab peamiselt kolme probleemi:

- 1) ridade iseloom, mille kohta on mõtet leida geomeetrilist keskmist;
- 2) geomeetrilise keskmise kujul leitud indeksite majanduslik põhjendatus;
- 3) geomeetrilise keskmise sobivus keskmise kasvutempo arvutamiseks.

Esimeses küsimuses on võtnud korduvalt sõna prof. I.J. Pissarev, kes väidab, et geomeetrilist keskmist ei saa kasutada ajas toimunud muutuste iseloomustamiseks mitte alati, vaid ainult nendel juhtudel, kui uuritavas protsessis endas kätkeb pikemaajaline ühes ja samas geomeetrilises progressioonis kulgev kasvutendents.¹ Kui sellist püsivat tendentsi pole, siis pole geomeetrilise keskmise kasutamine õigustatud. Mitmesuguseid piiravaid tingimusi on selle keskmise rakendusala seadnud teisedki autorid.

Geomeetrilise keskmisena leitud indeksite kasutatavus on nõukogude statistikas olnud juba väga kaua õige terava vaidluse aineks. Teatava reaktsioonina geomeetriliste in-

¹ И.Ю. Писарев, "Некоторые вопросы теории статистики", "Вопросы экономики" 1948, nr. 7. Samas on toodud ka bibliograafilisi viiteid.

deksite propageerimisel ameerika statistikute poolt käesoleva sajandi alguskümnenditel ja nende ebaõnnestunud rakenduskatsetele NSV Liidus, asusid paljud statistikateoreetikud seda laadi indekse tehes täiesti eitavale seisukohale. Nii sõnastasid B. Jastremski ja V. Hotimski juba 1930. aastal absoluutse keelu: "geomeetrilist indeksit kui tegelikku dünaamikat moonutavat näitarvu ei tohi mitte mingil juhul kasutada."¹ Ligikaudu samasugust seisukohta korratakse ka paljudes hilisemates õpikutes.²

Ometi on see seisukoht ilmselt liiga tõrjuv. Mitmed teised autorid (akad. V. Nemtšinov, L. Kazinets, I. Pissarev jt.) on selles küsimuses eriarvamusel ning on seda ka põhjendanud. Eriti väärivad märkimist prof. I. Pissarevi toodud näited geomeetriliste keskmistena arvatud indekse kohta, mille majanduslik mõtestatus on väljaspool kahtlust.³ (Selle probleemi üksikasjalisem käsitus kuulub indeksteooriasse.)

Suurima tähtsusega on diskuteeritavatest küsimustest kolmas: kas geomeetriline keskmine üldse sobibki nähtuste keskmise kasvutempo leidmiseks. Et see on siiani tolle keskmise peamisi kasutusalasid majanduslike nähtuste uurimisel ja rahvamajanduse plaanimisel, siis seatud küsimuse eitav lahendamine puudutaks õige paljusid majandusliku tegevuse alasid.

Geomeetrilise keskmise kui keskmise kasvutempo näitaja puudused peituvad nn. summaveas, mis on tema puhul paratamatu.

Geomeetrilise keskmise summaviga on üldiselt tuntud. 1958. aastal soovitas hiina statistik Fin Di-Si uut keskmise kasvutempo arvutamise meetodit, millega summaviga ei teki.⁴

¹ Теория математической статистики под редакцией проф. Б.С. Ястремского и В.И. Хотимского, Москва 1930, lk. 125.

² Vt. näit. Теория статистики, Госстатиздат, Москва 1930 lk. 180, Москва 1956, lk. 172.

³ Vt. И.Ю. Писарев, Некоторые вопросы статистического изучения труда и производства, НИИТ, Москва 1961, lk. 21.

⁴ Фин Ди-Си, "К вопросу о методологии исчисления среднего темпа роста" 1958, nr. 12, lk. 42.

Keskmise väärtused leitakse valemiga

$$\bar{q} = R - (\bar{q}^2 + \bar{q}^3 + \dots + \bar{q}^n),$$

kus

q - vaadeldavate aastate kasvutempod (alusindeksid);

R - kõigi aastate alusindeksite summa ja

n - indeksite arv.

Et selle valemi kasutamine nõuab õige komplitseeritud arvutusi, toimub keskmise tempo tegelik määramine vastava tabeli abil. Tabelis on iga R väärtuse jaoks olemas keskmine kasvutempo $\bar{q}$, mis oleneb vastavasse pikemasse ajavahemikku kuuluvate aastate (resp. osaperioodide) arvust n .

Kui on, näiteks, teada mingi ettevõtte kapitaalmahutuste summad üksteisele järgnevatel aastatel (vt. tabel 14), siis toimub arvutus järgmiselt:

1) leitakse kõigi alusindeksite summa R

$$1,05 + 1,79 + 1,44 + 1,71 + 2,64 = 8,23$$

2) leitakse tabelist (vt. tabel 15) R väärtusele vastav $\bar{q}$ (veerg $n = 5$):

$$\bar{q} = 1,171$$

mis ongi otsitav keskmine kasvutempo.

T a b e l 14 .

	Aastad (resp. osaperioodid)					
	baasi-aasta	I	II	III	IV	V
Kapitaalmahutuste summa (tuh. rubl.)	200	210	278	288	342	528
Alusindeksid	-	1,05	1,39	1,44	1,71	2,64

Tabel 15.

Keskmise kasvutempo $\bar{q}$ väärtused (Fin Di-Si järgi)
(väljavõtte tabelist).¹

Keskmine kasvu- tempo	Alusindeksite summa (R)			
	kahe aasta kohta	kolme aasta kohta	nelja aasta kohta	viie aasta kohta
$\bar{q}$	n = 2	n = 3	n = 4	n = 5
(1)	(2)	(3)	(4)	(5)
1,1600	2,5056	4,0665	5,8771	7,9774
1,1610	2,5089	4,0738	5,8902	7,9999
1,1620	2,5122	4,0811	5,9042	8,0226
1,1630	2,5156	4,0887	5,9182	8,1459
1,1640	2,5189	4,0964	5,9317	8,0685
1,1650	2,5222	4,1033	5,9453	8,0912
1,1660	2,5256	4,1109	5,9594	8,1148
1,1670	2,5289	4,1182	5,9729	8,1373
1,1680	2,5322	4,1256	5,9867	8,1605
1,1690	2,5356	4,1382	6,0008	8,1840
1,1700	2,5389	4,1405	6,0144	8,2069
1,1710	2,5422	4,1479	6,0282	8,2300
1,1720	2,5456	4,1555	6,0423	8,2536
1,1730	2,5489	4,1628	6,0559	8,2765
1,1740	2,5522	4,1704	6,0700	8,3001
1,1750	2,5556	4,1778	6,0889	8,3236

¹ Täielikum väljavõtte tabelist on toodud ajakirjas
"Вестник статистики" 1958, № 12.

6. RUUTKESKMINE JA MUUD ASTMEKESKMISED.

6.1. Ruutkeskmise valem.

Ruutkeskmise leitakse aritmeetilise keskmisega põhi-
joontes sarnase valemi alusel. Erisuseks on, et variantide
algväärtused tõstetakse eelnevalt ruutu. Et nad seega mitme-
kordistuvad ning ka dimensioonid tõusevad ruutu, siis tuleb
lähteandmete rida üldistavalt iseloomustava ruutkeskmise
leidmiseks teha läbi veel juurimistehe.

Ruutkeskmise arvutamise lähtevalem on variatsioonirida-
de andmete töötlemisel

$$\overline{x^2} = \sqrt{\frac{\sum x^2 f}{\sum f}} \quad (14)$$

ja lihtsa statistilise rea puhul

$$\overline{x^2} = \sqrt{\frac{\sum x^2}{n}} \quad (15)$$

Vastupidi järgnevates peatükkides käsitletavatele asen-
dikeskimestele iseloomustab ruutkeskmist kõige ilmekamalt just
see, et rea ekstreemsed liikmed (väikseimad ja suurimad)
avaldavad tema väärtusele eriti silmapaistvat mõju. Variat-
siooninäitavude tuletamisel, kus on eesmärgiks väljendada
ühes arvus just eeskätt variantidevahelisi erisusi, on see
ruutkeskmise üks peamisi eeliseid.

6.2. Ruutkeskmise kasutamine majanduspraktikas.

Majanduslike probleemide lahendamisel kasutatakse rea
üksikandmetest arvutatud ruutkeskmist suhteliselt harva, mõ-

nede suhteliselt spetsiaalse iseloomuga ülesannete lahendamiseks. Mitmesuguste teiste statistiliste meetodite raames on sellel keskmisetüübil aga väga suur tähtsus, näiteks dispersioonanalüüsis, korrelatsioonikoefitsientide leidmisel, riidade tasandamisel jne.

Ruutkeskmise arvutamist, ühtlasi aga ka tema rakendamist praktiliste probleemide lahendamisel demonstreerib järgmine näide.

Saaveskisse saabus partii 5 m pikkusi palke. On vaja arvutada saadud puitmaterjali üldine hulk tihumeetrites. Selleks on tarvis teada palgi keskmist diameetrit. Huupi väljavõetud ja mõõdetud palkide diameetrid on 20, 30 ja 40 cm. Kui suur on palgi keskmine diameeter saabunud partiis? Bel-dame, et nimetatud kolmes jämeduses palke on kõiki võrdsel arvul.

Kõigepealt selgitagem, missugust keskmist käesoleval juhul tuleks kasutada.

Vaatleme palki tingimisi silindrina. Sel puhul on palgi ruumala v

$$v = \frac{\pi x^2}{4} \cdot h ,$$

kus x on palgi diameeter ja h pikkus. Kolme palgi ruumala on järelikult

$$v_1 + v_2 + v_3 = \frac{\pi x_1^2 h}{4} + \frac{\pi x_2^2 h}{4} + \frac{\pi x_3^2 h}{4} .$$

Asendame mõõdetud palkide tegelikud diameetrid nende keskmisega. Palkide üldine ruumala sellest ei muutu, s.o.

$$\frac{\pi x_1^2 h}{4} + \frac{\pi x_2^2 h}{4} + \frac{\pi x_3^2 h}{4} = \frac{\pi \bar{x}^2 h}{4} + \frac{\pi \bar{x}^2 h}{4} + \frac{\pi \bar{x}^2 h}{4} .$$

Jagame nüüd võrrandi mõlemad pooli avaldisega $\frac{\pi h}{4}$.

Saame

$$x_1^2 + x_2^2 + x_3^2 + 3 \bar{x}^2 ,$$

$$\bar{x} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2}{3}},$$

millest selgub, et palgi keskmine diameeter tuleb antud juhul leida ruutkeskmisena.

Lähteandmeid kasutades saame

$$\bar{x}_{\text{ruutk.}} = \sqrt{\frac{20^2 + 30^2 + 40^2}{3}} = \sqrt{\frac{2900}{3}} \approx \sqrt{966,67} \approx 31,1 \text{ (cm)}.$$

Kontrollime nüüd, kas seatud ülesannet poleks saanud lahendada ka aritmeetilise keskmise abil. Arvutame selleks kolme mõõdetud palgi ruumala

a) lähteandmetel

$$5 \cdot \left(\frac{0,2^2 \pi}{4} + \frac{0,3^2 \pi}{4} + \frac{0,4^2 \pi}{4} \right) = \frac{1,45 \pi}{4};$$

b) ruutkeskmise kaudu

$$3 \cdot 5 \cdot \frac{0,311^2 \pi}{4} = 15 \cdot \frac{0,0967 \pi}{4} = \frac{1,45 \pi}{4};$$

c) aritmeetilise keskmise kaudu

$$3 \cdot 5 \cdot \frac{0,3^2 \pi}{4} = 15 \cdot \frac{0,09 \pi}{4} = \frac{1,35 \pi}{4}.$$

Selgub, et ruutkeskmise kaudu arvutatud palkide üldruumala langeb kokku lähteandmetest leitud ruumalaga, sellal kui aritmeetiline keskmine annab ilmselt vähendatud tulemuse.

Uuritava palgisaadetise koostise kohta ülesande algul püstitatud eeldusi meenutades võib keskmist 31,1 cm vaadelda kõigi saabunud palkide keskmise diameetrina ja selle abil leida nende üldise mahu.

Muudame eelmises punktis lahendatud ülesande andmestikku ning oletame, et keskmise diameetri leidmiseks mõõdeti kokku kümnet palki ja et neist oli läbimõõduga 20 cm - 3, läbimõõduga 30 cm - 2 ja 40 cm jämedusi 5, ning et täpselt samasugus-

tes proportsioonides jaotuvad kõik eri jämedusega palgid kogu saadetises. Et siin eri jämedusega palkide osatähtsus on erinev, tuleb leida kaalutud ruutkeskmine, mis on meie andmetel

$$\bar{x}_{\text{ruutk.}} = \sqrt{\frac{20^2 \cdot 3 + 30^2 \cdot 2 + 40^2 \cdot 5}{3 + 2 + 5}} = \sqrt{1100} \approx 33,2 \text{ (cm).}$$

6.3. Muud astmekeskmised.

Peale ruutkeskmise kasutatakse veel muidki astmekeskmi - kuupkeskmist jt. Kõik nad arvutatakse põhimõtteliselt samasuguse valemi järgi, kusjuures variandid on tõstetud vastavasse astmesse.

Nii saadakse kuupkeskmine

$$\sqrt[3]{\frac{\sum x_i^3 f_i}{\sum f_i}},$$

neljanda astme keskmine

$$\sqrt[4]{\frac{\sum x_i^4 f_i}{\sum f_i}}$$

jne.

7. MEDIAAN JA KVARTIILID.

7.1. Mediaani mõiste ja leidmine lihtsas statistilises reas.

Mediaan on korrastatud rea keskmine liige, millest mõlemale poole jääb võrdne arv liikmeid. Teisiti ongi mediaani nimetatud ka keskliikmeks.

Kui korrastatud reas on liikmeid paaritu arv, siis on mediaani väga hõlpus leida. Näiteks reas

$$S_1 = 5, 8, 9, 13, 16, 18, 20, 25, 39,$$

kus on 9 liiget, on mediaan (tähis Me) 16, s.o. 5. liige ükskõik kummast rea otsast lugedes. Kui tähistada rea liikmete arvu n , siis on mediaani järjekorranumber reas

$$\frac{n + 1}{2}$$

Reas, kus on paarisarv liikmeid, saadakse sama valemit kasutades mediaani järjekorranumbriks selline arv, mis asub küll rea keskel, ent kahe keskmise liikme vahel.

Näiteks kümnest liikmest koosnevas reas

$$S_2 = 5, 8, 9, 13, 16, 18, 20, 25, 39, 40$$

on mediaani järjekorranumber

$$\frac{10 + 1}{2} = 5,5 .$$

Praktiliselt rõhutab see, et reas, kus on paarisarv liikmeid, üht keskmist liiget polegi - rea keskme moodustavad kaks liiget. Et siiski saada üht mediaani väärtust, leitakse mediaan nende keskliikmete väärtuste lihtsa aritmeetilise keskmisena, s.t.

$$Me = \frac{16 + 18}{2} = 17 .$$

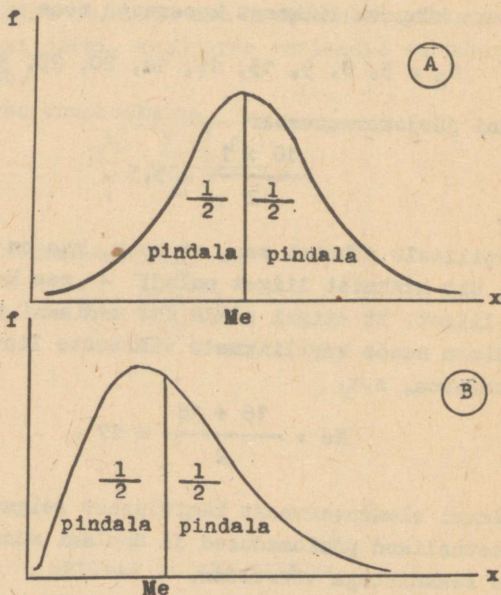
Mediaani elementaarsest käsitlesest selguvad hästi kõik tema tunnetuslikud põhiomadused ja ühtlasi mõnesugused eelised muude keskmistega võrreldes.

Rea alguses ja lõpus olevate variantide suurused ning omavaheline jaotumine temale mõju ei avalda. Seepärast on mediaan lahtiste häärerühmadega variatsiooniridades eriti sobiv asendama aritmeetilist keskmist.

Mediaani tüüpilisus on alati tunduvalt suuremal määral tagatud kui aritmeetilisel keskmisel. Espool nägime, et aritmeetiliseks keskmiseks osutub tihti just selline tunnuse

väärtus, mida reas ei esinegi ja mis pole seega reale üldse omane. Mediaaniga seda juhtuda ei saa. Mediaan on alati üks rea liikmeist (või vähemalt sellele väga lähedane suurus, näit. paarisarvulistes ridades).

Graafilises interpretatsioonis kujutab mediaan endast variatsioonkõvera alust pindala poolitava vertikaaljoone abstsissi (vt. joon. 4). Nagu jooniselt näha (vrd. joon. 2), on täiesti sümmeetrilise rea puhul aritmeetiline keskmine ja mediaan võrdsed.



Joon. 4. Mediaani koht A - sümmeetrilises ja B - ebasümmeetrilises variatsioonreas. Mediaan jagab jaotuskõvera aluse pindala kaheks võrdseks osaks.

7.2. Mediaani leidmine sagedates moment-variatsioonridades.

Sagedates moment-variatsioonridades tuleb mediaani leidmiseks

- 1) leida mediaani järjekorranumber;
- 2) leida mediaanintervall.

Et mediaan on rea keskmine liige, siis on tema järjekorranumber võrdne sageduste poolsummaga, s.o.

$$\frac{\sum f}{2} .$$

Täpsem oleks leida mediaani järjekorranumber analoogiliselt eelmises punktis kasutatud valemiga

$$\frac{\sum f + 1}{2} .$$

Sagedamini kasutatakse siiski esimest moodust, sest mediaani arvutatakse enamasti ikka suurte kogumite kohta, kus tema järjekorranumber ühe liikme liitmisest oluliselt ei muutu.

Tugineme andmetele perekondade suuruse kohta Eesti NSV-s 1959. aastal (vt. tabel 10). Perekondade üldarv oli 308613. Saame mediaani järjekorranumbriks

$$\frac{308613}{2} = 154306,5 \approx 154307 .$$

Et murru lugejale ühe liitmine tõepoolest midagi ei muuda, on käesoleva näite puhul täiesti selge, sest ka

$$\frac{308613 + 1}{2} = 154307 .$$

Teiseks leiame mediaanintervalli. Tavaliselt toimub see üksikute variantide sageduste jargsumma leidmisel. Intervall, mille sageduste jargsumma ületab esimesena mediaani järjekorranumbri, peitubki mediaan. Meie näites on selleks juba

teine intervall, sest

$$121530 + 92732 = 214262 > 154307 .$$

Et selles intervallis on kõikide variantide väärtused võrdsed, siis on mediaani väärtus üks nendest, s.o.

$$Me = 3 .$$

Muidugi saab mediaani niiviisi leida ainult sõreda tunnuse väärtustest koosneva momentrea andmetel, milles variantide ja vastava tunnuse väärtuste loomulikud vahed on võrdsed. Muud momentread tuleb teisendada intervallriksadeks. Intervallriksadest on mediaani leidmine mõnevõrra keerulisem.

7.3. Mediaani leidmine intervallitud variatsioonrea andmetel.

7.3.1. Graafiline meetod.

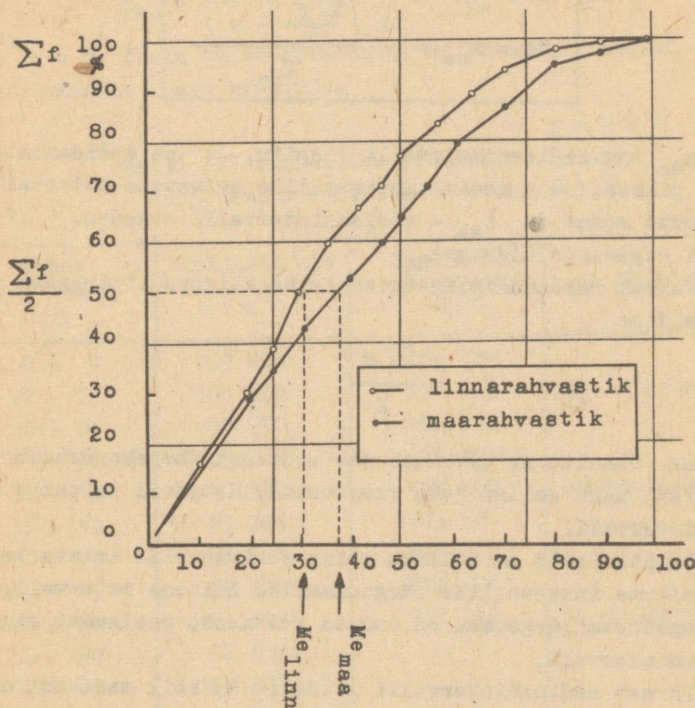
Nagu moodi, nii on ka mediaani leidmine mõnevõrra komplikseeritud sel juhul, kui on tegemist intervallitud variatsioonreaga.

Graafiliselt kujutab mediaan endast variatsioonkõvera alust pindala poolitava ordinaadi ja abstsissitelje lõikumispunkti väärtust (vt. joon. 4). Tavaliselt variatsioonkõveralt on aga mediaani küllaltki tülikas leida. Tunduvalt hõlpsam on seda teha nn. suhteliste sageduste jrgsummade kõvera abil, mida on demonstreeritud joonisel 5.

Variantide väärtusi mõõdetakse sel diagrammil rõhtskaala abil ja variantide suhtelisi sagedusi (protsentides) püstskaala abil. Mediaani väärtuse projekteerib abstsisssteljele kõvera selle punkti ordinaat, mida läbib 50-protsendilise sageduse abstsiss.

Joonisel 5 on kujutatud jrgsummade kõveratena Eesti NSV linna- ja maarahvastiku vanuseline struktuur. Mediaani ehk rea keskmise liikme graafiliseks leidmiseks piisab, kui projekteerida ordinaatteljelt 50 % kohalt rõhtjoon jrgsummade kõverale ja sealt püstjoon abstsisssteljele. Jrgsumma-

Eesti NSV rahvastiku vanuseline jaotumine
1959. aastal.¹



Joon. 5. Mediaani graafiline leidmine suhteliste sageduste järgsummade kõveralt.

de kõvera ordinaadi ja rõhttelje lõikumiskohale saadaksegi mediaani väärtus. Antud juhul saame kaks mediaani, kusjuures selgub, et maarahvastiku keskmine vanus on mõnevõrra suurem kui linna-rahvastikul.

¹ Итоги Всесоюзной переписи населения 1959 года, Москва 1962, lk. 19.

7.3.2. Mediaani arvutamine.

Intervallreas leitakse mediaan järgmise valemiga

$$Me = x_{Me} + \frac{k \left(\frac{\sum f}{2} - w \right)}{f_{Me}}, \quad (16)$$

kus x_{Me} on mediaanintervalli alapiir, k on mediaanintervalli pikkus, w - mediaanintervallile eelnevate intervallide sageduste summa ja f_{Me} - mediaanintervalli sagedus. f on siingi sageduste üldsumma.

Valemi rakendamiseks tuleb leida kõigepealt lugejas oleva avaldise

$$\frac{\sum f}{2}$$

väärtus. Sisuliselt tähendab see mediaani järjekorranumbri leidmist, nagu selles juba veendusime. Seejärel leitakse mediaanintervall.

Praktiliselt on selleks kõige sobivam viis hakata leidma üksikute intervallide järgsummasid. Esimene intervall, kus sageduste järgsumma on suurem kui nende poolsumma, ongi mediaanintervall.

Pärast mediaanintervalli leidmist on kõik mediaani arvutamiseks vajalikud lähtesuurused teada.

Rakendame eespool toodud valemit Eesti NSV rahvastiku keskmise vanuse leidmiseks. Vajalikud andmed on tabelis 16.

Mediaani järjekorranumber on

$$\frac{\sum f}{2} = \frac{1196791}{2} = 598395,5 \approx 598396.$$

Mediaanintervalli leidmiseks arvutame sama tabeli 3. veerus sageduste järgsummasid, kuni esimese mediaani järjekorranumbrit ületava summani, milleks on 657 800. Järelikult on

see vahemik mediaanintervall. Mediaani valemisse vajalikud väärtused on nüüd teada:

$$x_{Me} = 30; w = 559\,371; f_{Me} = 98\,429$$

ja $k = 5$ (siin tuleb arvestada, et intervall hõlmab 5 aastat, viimane kaasa arvatud).

T a b e l 16 .

Eesti NSV rahvastiku vanuseline koostis
1959. aasta rahvaloenduse andmeil.¹

Vanus (aastates) x_i	Inimeste arv f_i	Sageduste järgsummad $\sum f_i$	Märkused
0 - 9	189 220	189 220	Q_1 -intervall
10 - 19	168 303	357 523	
20 - 24	98 770	456 293	
25 - 29	103 078	559 371	
30 - 34	98 429	657 800	Mediaanintervall
35 - 39	75 581	733 381	
40 - 44	62 341	795 722	
45 - 49	77 904	873 626	
50 - 54	73 616	947 242	Q_3 -intervall
55 - 59	68 558		
60 - 69	102 071		
70 - 79	59 942		
80 - 89	17 640		
90 - 99	1 270		
100 ja vanemad	13		
vanus teadmata	55		
	1 196 791		

¹ Andmed Итоги Всесоюзной переписи населения 1959 года, Эстонская ССР, ЦСУ СССР, Москва, 1962, lk. 19.

Seega saame

$$\begin{aligned} Me &= 30 + \frac{5(598396 - 559371)}{98429} = \\ &= 30 + \frac{5 \cdot 39025}{98429} = 30 + \frac{195125}{98429} = \\ &= 30 + 1,98 = 31,98 \approx \underline{32} \text{ aastat.} \end{aligned}$$

Leiame täiendava näitena mediaani 2069 tähtsama ehitus-organisatsiooni plaani täitmise näitajaist (tabel 1) andmeil.

Mediaani järjekorranumber on

$$\frac{\sum f}{2} = \frac{2069}{2} = 1034,5 \approx 1035.$$

Edasi leiame intervallide sageduste jrgsummad, kirjutades selleks vajalikud andmed tabelist välja

f_1	$\sum f_1$
20	20
23	43
66	109
95	204
153	357
217	574
254	828
273	1101

Viimases intervallis ($x_1 = 70 - 80$) on variandid järjekorranumbritega 829 kuni 1101. Järelikult peitub seal ka mediaan, s.o. rea 1035. liige.

Leitud andmetel on

$$\begin{aligned} Me &= 70 + \frac{10(1035 - 828)}{273} = 70 + \frac{2070}{273} = \\ &= 70 + 7,6 = 77,6 \% . \end{aligned}$$

Kasutatud valemi esimeseks liikmeks on x_{Me} - mediaan-intervalli alapiir, mida võib pidada tingimisi ligikaudseks mediaaniks ja

$$\frac{k \left(\frac{\sum f}{2} - w \right)}{f_{Me}},$$

mida võib vaadelda x_{Me} parandusliikmena.

Väärib tähelepanu, et kuna $x_{Me} < Me$, siis peab parandusliige olema alati positiivne. Muutub mediaani valemis olevaid liikmeid ühendav märk miinuseks, peab arvutustes olema viga.

Parandusliiget võib vaadelda omakorda koosnevana kahest osast, intervalli pikkusest k ja võrdetegurist, s.o.

$$k \cdot \frac{\frac{\sum f}{2} - w}{f_{Me}}.$$

Viimane väljendab seda, mitmendiku võrra mediaanintervalli mahust on mediaan suurem tema intervalli alapiirist. Mida suurem on madalamate intervallide osatähtsus rea liikmete üldarvus, seda enam läheneb mediaani väärtus intervalli alapiirile ja vastupidi - mida suurem on kõrgemate intervallide osatähtsus rea liikmete üldarvus, seda enam läheneb mediaani väärtus rea ülapiirile.

7.4. Kvartiilid.

Pikemate ridade puhul võib osutada vajalikuks iseloomustada neid mitte ainult mediaaniga, vaid ka kvartiilidega.

Kvartiilid on mediaaniga analoogiliselt arvutatud struktuurikeskmised; nad jagavad korrastatud statistilise rea neljaks osaks, millest igaüks on võrdne arv liikmeid. Kvartiile on kolm, neid nimetatakse vastavalt esimeseks, teiseks ja kolmandaks kvartiiliks ning tähistatakse

$$q_1, q_2, q_3.$$

Teine kvartiil on võrdne mediaaniga, s.o.

$$Q_2 = Me.$$

Esimene kvartiil on õigupoolest mediaan rea esimesest poolest ja kolmas - mediaan rea teisest poolest. Nii võib arvutada kvartiile juba eespool kirjeldatud mediaani valemi abil.

$$Q_1 = x_{1.Q} + \frac{k \left(\frac{\sum f}{4} - w_{1.Q} \right)}{f_{1.Q}} \quad (17)$$

ja

$$Q_3 = x_{3.Q} + \frac{k \left(\frac{\sum f}{4} - w_{3.Q} \right)}{f_{3.Q}}, \quad (18)$$

kus $x_{1.Q}$ ja $x_{3.Q}$ on vastavate kvartiilide intervallid;

$w_{1.Q}$ on esimesele kvartiilintervallile eelnevate intervallide sageduste summa;

$w_{3.Q}$ - kolmandale kvartiilintervallile eelnevate sageduste summa;

$f_{1.Q}$ ja $f_{3.Q}$ - vastavate kvartiilintervallide sagedused.

Kvartiilid leitakse graafiliselt samuti nagu mediaan. Selleks koostatakse suhteliste sageduste järgsummade kõver (vt. joon. 3). Esimene kvartiil loetakse selle vertikaalskaala 25 % ja kolmas 75 % kohalt.

Käsitades kvartiile struktuurikeskmistena tuleb märkida, et kogu rea keskmine on neist ainult Q_2 . Esimene ja kolmas kvartiil on vastavate reaosade keskmised. See tähendab, et kvartiilide arvutamine rajaneb rea ositamisele (viimast mõistet tuleb eristada rühmitamisest).

Kvartiilid annavad rea struktuurile võrdlemisi hea ja mitmekülgse kirjelduse. Nende tähtsus ulatub aga veelgi kaugemale. Eriti populaarsed on nad lihtsate variatsiooninäitajate arvutamisel, sest kvartiilide erisustele (vahedele) toetudes saab rea liikmete omavaheliste erisuste kohta arvutada õige mitmekesise sisuga üldistavaid näitajaid.

Näide. (Andmed tabel 16, lk. 87). Leida Q_1 ja Q_3 Eesti NSV rahvastiku vanuse kohta.

Nagu mediaani leidmisel, tuleb ka siin määrata kõigepealt kvartiilide järjekorranumbrid. Need on

$$Q_1 \text{ jaoks } \frac{1 \ 196 \ 791}{4} \approx 229 \ 198$$

ja

$$Q_3 \text{ jaoks } \frac{3 \cdot 1 \ 196 \ 791}{4} \approx 897 \ 594 .$$

Selgub, et Q_1 peitub intervallis 10 - 19. Võime seega arvutada

$$\begin{aligned} Q_1 &= 10 + \frac{10(299 \ 198 - 189 \ 220)}{168 \ 303} = \\ &= 10 + \frac{10 \cdot 109 \ 978}{168 \ 303} = 10 + 6,5 = 16,5 \end{aligned}$$

ning

$$\begin{aligned} Q_3 &= 50 + \frac{10(897 \ 594 - 873 \ 626)}{73 \ 616} = \\ &= 50 + \frac{10 \cdot 23 \ 968}{73 \ 616} = 50 + 3,25 \approx 53,3 . \end{aligned}$$

7.5. Detsiilid. Tsentiilid.

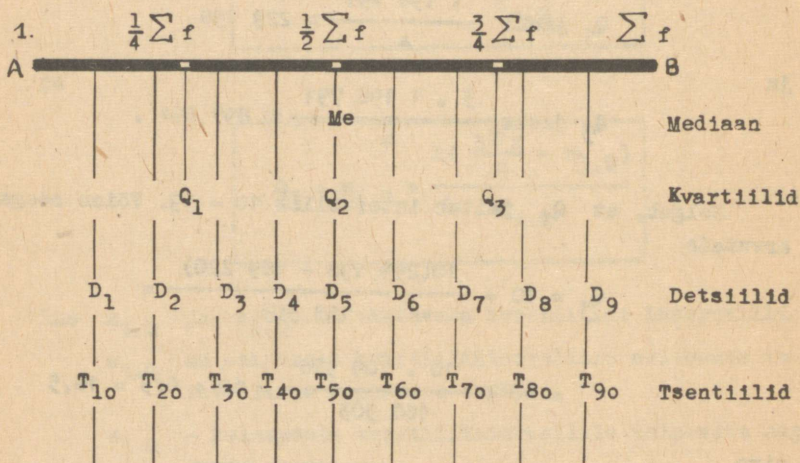
Korrastatud ridade veelgi detailsemal esitamisel, nende jagamisel kümneks või sajaks võrdse liikmete arvuga osaks, saadakse vastavalt kas detsiilid või tsentiilid.

Detsiilid jaotavad statistilise rea kümneks osaks, millest igaühes on võrdne arv liikmeid. Detsiile on üheksa. Neid tähistatakse

$$D_1, D_2, \dots, D_9.$$

Nagu jooniselt 6 selgub, on

$$D_5 = Q_2 = Me.$$



Joon. 6. Tähtsamad asendikeskmised ja nende omavahelised seosed. Lõigu AB pikkus on võrdeline liikmete üldarvuga statistilises reas. Otse üksteise all olevad asendikeskmised on võrdsed, näit. $Me = Q_2 = D_5 = P_{50}$.

Tsentiilid (ehk pertsentiilid) jaotavad statistilise rea sajaks võrdse liikmete arvuga osaks. Neid on 99, nad tähistatakse

$$P_1, P_2, \dots, P_{99}.$$

Nagu jooniselt 6 näitlikult selgub, on täiskümnetega tähistatud tsentiilid võrdsed vastavate ühelistega tähistatud

detsiilidega, s.o.

$$P_{10} = D_1, \quad P_{20} = D_2 \text{ jne.}$$

ning

$$Q_1 = P_{25}, \quad Q_2 = P_{50} \text{ ja } Q_3 = P_{75}.$$

Detsiilide ja tsentiilide arvutuslik leidmine on küllalt tülikas ja vaevanõudev. Töõdes, mis ei pretendeeri suurele täpsusele, leitakse nende väärtused enamasti graafilisel teel.

7.6. Mediaani praktiline kasutamine.

Mediaani kasutatakse enamasti variatsiooniridade andmete töötlemisel. Ta sobib kasutada põhimõtteliselt samadel puhkudel kui aritmeetiline keskmine.

Täiesti asendamatu on mediaan lahtiste äärerühmadega ridade puhul, nagu näit. tabelis 9 esitatud rida, mille aritmeetilise keskmise arvutamine võib anda õige illusoorseid tulemusi.

Ühe patsiendi keskmise haiglapäevade arvu leiame selle tabeli andmeil mediaanina järgmiselt:

$$1) \frac{\sum f}{2} = \frac{874}{2} = 437$$

2)	$\frac{f_1}{\sum f_1}$	
	118	118
	120	238
	131	369
	140	509

interv. 16 - 20 p.

$$3) Me = 16 + \frac{5(437 - 369)}{140} =$$

$$= 16 + \frac{340}{140} = 16 + 2,4 = \underline{\underline{18,4}} \text{ päeva.}$$

8. MOOD.

8.1. Moodi mõiste ja leidmine lihtsas statistilises reas.

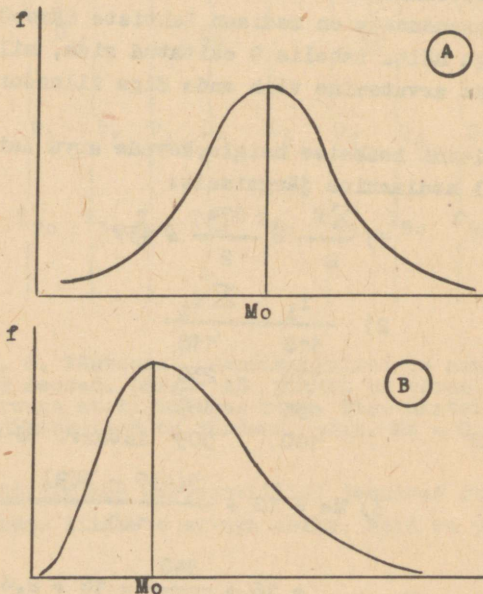
Moodiks nimetatakse statistilises reas kõige sagedamini korduvat väärtust. Lihtsast statistilisest reast (kui see pole just väga pikk) on moodi hõlpus leida silma järgi, ilma et see nõuaks mingisuguseid arvutusi või isegi rea korrastamist.

Statistilises reas

8; 13; 9; 10; 9; 25; 7; 9; 16; 9; 9; 5

on mood 9, sest see esineb kõige sagedamini.

Graafilises interpretatsioonis kujutab mood endast variatsioonrea maksimumi abstsissi (vt. joon. 7).



Joon. 7. Moodi koht variatsioonreas.
Mood vastab jaotuskõvera maksimumile.

8.2. Moodi leidmine momentrea andmeil.

Lihtne on moodi leida ka variatsioonreast, kui see on momentrida. Sel juhul on mood see variant, mille sagedus on kõige suurem.

Kui kasutada eespool toodud näidet perekondade suuruse kohta Eestis ja Armeenias 1959. aastal (vt. tabel 10), siis piisab selle lihtsast vaatamisest, et leida kõige sagedamini esinev perekonna suurus. Eestis on selleks kaheliikmeline perekond, Armeenias neljaliikmeline.

Nii saab moodi leida muidugi ainult sõreda tunnuse väärtsusi sisaldavast momentreast, kus momentide vahe võrdub loomulike vahemikega vastava sõreda tunnuse võimalike väärtsuste vahel (näit. perekonnaliikmete arv 2, 3, 4 jne.). On aga tegemist momentridadega, mis on saadud intervallridadest sel teel, et intervallid on asendatud nende aritmeetiliste keskmistega, näiteks

keskmise plaani tähitmise protsent	90	110	130	150	170
ettevõtete arv	4	10	5	1	20

siis tulevad sellised read (algandmed tabelis 11) muuta kõigepealt taas intervallridadeks ja töödelda seejärel nende tõelisele iseloomule vastavalt.

8.3. Moodi leidmine intervallitud variatsioonreas.

8.3.1. Graafiline menetlus.

Mõnevõrra keerulisem on leida moodi intervallitud variatsioonreas. Selgitame seda kõigepealt lihtsa arvnäite vara graafiliselt.

Oletagem, et on teada järgmine variatsioonrida:

x	f
5 - 10	7
10 - 15	15
15 - 20	40
20 - 25	30
25 - 30	8
	100

Histogrammina kujutatuna esineb see rida joonisel 8.

Moodi leidmisel läbitakse kaks järjestikust töötappi:

1) leitakse moodintervall. Selleks on moodi definitsiooni kohaselt see intervall, mille sagedus on kõige suurem. Antud näites intervall, mille variantide väärtused on 15 - 20;

2) leitakse moodi väärtus erilise graafilise menetlusega. Selleks ühendatakse moodintervallis asuva histogrammitulba tipud B ja C diagonaalsirgete abil eelmise ja järgmise histogrammitulba tippudega D ja A. Diagonaalide lõikumispunkti E abstsiss annabki moodi väärtuse.

Selle võtte puhul väärrib tähelepanu, et moodi väärtus sõltub ainult moodintervalli pikkusest ja kolme intervalli - moodintervalli ning sellele eelneva ja järgneva intervalli sagedusest.

Mida suurem on moodintervallile järgneva intervalli sagedus, seda enam läheneb moodi väärtus intervalli ülapiirile ja vastupidi.

8.3.2. Moodi arvutamine intervallitud variatsioonrea andmeil.

Moodi arvutamisel juhindutakse valemist

$$M_o = x_{mo} + \frac{k(f_{mo} - f_{mo-1})}{(f_{mo} - f_{m-1}) + (f_{mo} - f_{mo+1})} \quad , \quad (19)$$

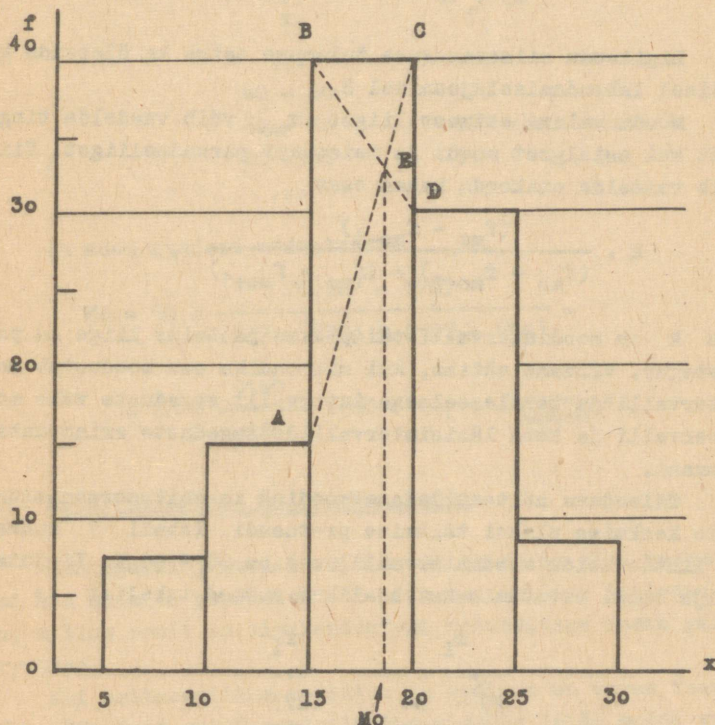
kus x_{mo} on moodintervalli alapiir;

k - moodintervalli pikkus;

f_{mo} - moodintervalli sagedus;

f_{mo-1} - moodintervallile eelneva intervalli sagedus ja

f_{mo+1} - moodintervallile järgneva intervalli sagedus.



Joon. 8. Moodi graafiline leidmine.

See valem baseerub põhimõtteliselt samasugusele lähene-
misviisile, nagu eespool kirjeldatud graafiline menetlus.

Sel juhul tuleb leida eelkõige moodintervall: see on in-

tervall 15 - 20, mille sagedus on 40. Sellega on määratud kõigi moodi valemis sisalduvate suuruste väärtused.

$$\begin{aligned} Mo &= 15 + \frac{5(40 - 15)}{(40 - 15) + (40 - 30)} = \\ &= 15 + \frac{5 \cdot 25}{25 + 10} = 15 + \frac{125}{35} = 15 + 3,57 = \underline{\underline{18,57}}. \end{aligned}$$

Iidikaudu niisamasuguse tulemuse saime ka ülesande graafilisel lahendamisel joonisel 8.

Moodi valemi esimest liiget x_{mo} võib vaadelda tinglikult kui esialgset moodi ja teist kui parandusliiget. Viimast võib vaadelda omakorda kahes osas

$$k \cdot \frac{(f_{mo} - f_{mo-1})}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})},$$

kus k on moodintervalli täispikkus ja teine liige on parandustegur. Viimane näitab, kui mitmendiku osa moodustab moodintervalli ja temale eelneva intervalli sageduste vahe moodintervalli ja tema lähisintervallide sageduste erinevuste summast.

Täiendava näitena leiame moodina ka ehitusorganisatsioonide keskmise plaani täitmise protsendi. Tabeli 1 andmetel on lihtne leida moodintervalli, see on 80 - 90 %. Kirjutame välja moodi arvutamiseks vajalikud andmed tabelist 1:

x_i	f_i	
...	...	
70 - 80	273	
80 - 90	282	moodintervall.
90 - 100	232	
...	...	

Neil andmeil on

$$Mo = 80 + \frac{10(282 - 273)}{(282 - 273) + (282 - 232)} =$$

$$= 80 + \frac{90}{90 + 50} = 80 + \frac{90}{140} = \underline{\underline{80,64}} \% .$$

Missugust mõju avaldab moodile andmete rühmitamisel kasutatud intervallide pikkus, see tuleb ilmsiks, kui arvutame sama kogumi kohta moodi ka kaks korda pikematest intervallidest. Vajalikud andmed on tabelist 2 väljakirjutatuna järgmised:

x_i	f_i
...	...
40 - 60	370
60 - 80	527 moodintervall
80 - 100	514
...	...

Ja mood nendest andmetest

$$\begin{aligned} Mo &= 60 + \frac{20(527 - 370)}{(527 - 370) + (527 - 514)} = \\ &= 60 + \frac{3140}{164} = 60 + 19,14 = \underline{\underline{79,14}} \% . \end{aligned}$$

8.4. Moodi leidmine $\bar{x}$ ja Me kaudu.

Mõningatel juhtudel võib moodi leidmine olla raskendatud kas andmete puundumise või siis rea erilise iseloomu tõttu, milles moodi on ligikaudsetegi meetoditega raske välja arvutada.

Kui aritmeetiline keskmine ja mediaan on varem teada, võib mõõdukalt ebasümmeetrilistes ridades leida moodi valemiga

$$Mo = \bar{x} - 3 (\bar{x} - Me) . \quad (20)$$

Põhimõtteliselt väljendab see valem mõõdukalt ebasümmeetrilistes ridades hästi silmahakkavat tõsiasja, et mediaan paikneb aritmeetilise keskmise ja moodi vahel, olles seejuures aritmeetilisele keskmisele lähemal ($\frac{1}{3}$ vahemaad aritmeetilisest keskmisest, $\frac{2}{3}$ mediaanist). Nende sellise järjestuse meenutamiseks soovitatakse järgmist mnemoonilist reeglit. Teame, et inglise keeles on aritmeetiline keskmine "mean", mediaan "median" ja mood "mode". Kui seada need sõnad tähes-tiku järjekorda, siis asuvad nad just nii: a) mediaan aritmeetilise keskmise ja moodi vahel ning b) aritmeetilisele keskmisele lähemal kui moodile.

Täiesti sümmeetrilistes variatsiooniridades langevad kõik kolm keskmist ühte, s.o.

$$Mo = \bar{x} = 3 \cdot 0 = \bar{x}.$$

Esipool käsitletud näidetest võib pidada tingimisi mõõdukalt ebasümmeetrilise rea näiteks armeenlaste perekondade jaotumise kohta teada olevaid andmeid (vt. tabel 10). Selles reas on $\bar{x} = 4,8$ ja $Me = 5$. Kaudselt arvutatuna on

$$Mo = 4,8 - 3(4,8 - 5) = 4,8 + 0,6 = \underline{5,4} \approx \underline{5}.$$

Otse algandmetest leituna tuleb mood samuti 5.

Andmed Eesti perekondade jaotumise kohta sama tunnuse järgi moodustavad rea, mida tuleb pidada mitte mõõdukalt, vaid täiesti ebasümmeetriliseks. Seetõttu see valem seal ka enam ei kehti.

Esitatud valemist saab tuletada ka analoogilised seosed mediaani ja aritmeetilise keskmise leidmiseks

$$Me = \frac{Mo + 2\bar{x}}{3} \quad \text{ja} \quad \bar{x} = \frac{3Me - Mo}{2}.$$

Viimaseid kasutatakse aga üksnes vastavate seoste hindamiseks, mitte Me või $\bar{x}$ leidmiseks, sest nii üks kui teine on täpsemini arvutatav kui mood. Seetõttu pole mõtet leida neid moodi kui suhteliselt väiksema täpsusega määrata-

va suuruse kaudu. Küll aga on mõtet leida meodi aritmeetilise keskmise ja mediaani kaudu.

8.5. Moodi praktiline kasutamine.

Moodi kasutatakse keskmiste hindade arvutamisel turustatistikas, rõivanumbrite suuruse kindlaksmääramisel ja paljudel muudel puhkudel.

Kaupade hinnad sõltuvad kolhoositurul nõudmise ja pakumise vahekorrast. Seetõttu võib iga müüja nõuda oma kaubalt erinevat hinda. Keskmise hinna arvutamine kaalutud aritmeetilise keskmisena ei tuleks kolhoositurul arusaadavatel põhjustel kõne alla. See nõuaks esiteks liiga palju tööd. Teiseks võiks nii saadud keskmine hind tulla selline, mida päeva jooksul üheltki ostjalt pole küsitud ja mis oleks järelikult ebatüüpiline.

Moodi leidmiseks tarvitseb statistikul vaid turul ringi liikuda ja pärida, mida kauba eest küsitakse. Isegi ilma mingisuguseid märkusi tegemata kujuneb kiiresti kujutus tüüpilisest päevahinnast. Moodi otstarbekust rõhutab keskmiste turuhindade leidmisel eriti see, et tema väärtus langeb täpselt kokku ettekujutusega, mis ostjaskonna valdaval osal on selle päeva keskmisest hinnast.

Ka keskmiste maailmaturuhindade arvutamisel kasutatakse tavaliselt moodi.

Keskmiste rõivanumbrite suuruse väljatöötamisel jaotatakse inimesed eelnevalt soo, vanuse ja kehaehituse järgi rühmadeks. Mõõtmete kombinatsioon, mis rühmas kõige sagedamini esineb - s.o. mood - võetakse normiks. Sellega on tagatud, et igas rühmas on maksimaalselt võimalik arv inimesi, kellele vastavate mõõtmete järgi valmistatud rõivad on parajad.

Oletagem, et seistakse ülesande ees määrata kindlaks kolmeaastaste laste jaoks valmistatavate mütside ümbermõõt. Selleks tuleb korraldada eelkõige statistiline vaatlus - mõõta ära piisavalt suure arvu juhuslikult võetud laste pea ümbermõõdud. Saadud mõõtmetest valitakse välja see, mis kor-

dub kõige sagedamini; see ongi pea übermõõtude mood, mille järgi tuleb valmistada kolmeaastaste laste mütse.

Lihtne on mõista, mis võiks juhtuda, kui mütsid valmistatakse samade vaatlusandmete aritmeetilise keskmise kohaselt. Tõenäosus, et saadakse niisugused mütsid, mis pole parajad ühelegi vastava vanusega lapsele, oleks väga suur, sest aritmeetiliseks keskmiseks saadakse enamasti just niisugune arv, mis ise vaatlusandmete reas ei esine. Moodi kasutamisel on sellised üllatused välditud. Samadel põhjustel tuginetakse moodile ka muude massiliselt valmistatavate rõivastusesemete, jalatsite jt. mõõtmete leidmisel. Põhimõtteliselt samuti arvutatakse ka hobuste jm. veoloomade rakmete mõõtmed.

9. VARIATSIOONINÄITARVUD.

9.1. Varieerumise mõiste ja olemus.

Nähtust, et sama tunnuse väärtused on koguni eri liikmetel erinevad, nimetatakse varieerumiseks.

Kui koguda andmeid näiteks mingi ühe tehase tööliste soo, vanuse, tööstaazi, kvalifikatsiooni, palga, laste arvu ja tööviljakuse kohta, saadakse võrdlemisi kirju andmete rida. Ühtlasi selgub, et iga uuritava tunnuse väärtuse varieerumises avaldub mõnesuguseid iseärasusi.

Suuri erisusi võib täheldada ka sama tunnuse varieerumises eri kogumites. Kui vaadelda näiteks vanuse varieerumist statsionaarsete ja mittestatsionaarsete üliõpilaste ning mingi suure tehase töötajaskonna seas, siis ilmneb arvatavasti, et kõige vähem varieerub statsionaarsete üliõpilaste vanus, kõige rohkem tehase töötajaskonna vanus. Siit selgub, et peale muude karaktersete joonte iseloomustab nähtusi ka veel nende väärtuste omapärane varieerumus.

Niisamuti nagu tunnuste keskmistes väärtustes toimuvad nihked, võimaldavad ka nende varieerumisastmete muutused teha tähtsaid sisulisi järeldusi. Seepärast ei saa piirduda

üksnes keskmise tööviljakuse, loomade keskmise produktiivsuse jms. jälgimisega, vaid tuleb uurida ka vastavate tunnuste varieerumist.

9.1.1. Vajadus täiendada keskmisi variatsiooninäitavudega.

Kogumite kvantitatiivsete üldistustena on keskmised ühekülgsed selles mõttes, et nad kajastavad üksnes variantide väärtuste üldist nivood, mitte aga nende omavahelisi erisusi. Seepärast on võimalik, et täiesti eriilmeliste statistiliste ridade aritmeetilised keskmised tulevad võrdsed. Juhindudes üksnes keskmistest, võib seetõttu teha mõnigi kord ebaõigeid järeldusi.

Oletagem, et on teada andmed tööviljakuse dünaamika kohta paljudes analoogilisi tooteid väljavates ettevõtetes. Nagu teada, peetakse tööviljakuse tõusu või langust tehase töö kvaliteedi üheks kõige tähtsamaks näitajaks. Nende andmete alusel võib seada kõik tehased järjekorda: töö on halvenenud neis tehastes, kus tööviljakus on langenud ja paranenud neis, kus see tõusis. Niisama loogiline näib ka see, kui järeldame, et tehastes, kus tööviljakuse kasv on võrdne, on töötatud ühtviisi hästi, et need tehased on oma töö headuse poolest ühel järjel. See viimane järeldus ei tarvitse aga kaugeltki õige olla ja enamasti polegi seda.

Olgu teada viis tehast, millest igaühes on tööviljakus tõusnud 5 protsenti, s.o. tööviljakuse kasvu koefitsient on 1,05 ehk 105 %. Tehastes on 4 - 5 tootmisjaoskonda. Peale selle on teada andmed tööviljakuse muutumise kohta ka üksikutes tootmisjaoskondades (vt. tabel 17). Näite lihtsustamise mõttes eeldagem, et sama tehase kõikides tootmisjaoskondades on võrdne arv töölisi. Sel juhul on tehase tööviljakuse kasvu koefitsient leitav tsehhide koefitsientide lihtsa aritmeetilise keskmisena. Ehkki kõikide tehaste keskmised kasvud on võrdsed, on andmetest näha, et lood tööviljakusega on kulgenud eri tehastes äärmiselt erinevalt.

Täiesti ühtlaselt on tööviljakus tõusnud ainult I tehases. Kõigi teiste andmed on hoopis ebaühtlasemad. Kolmes vii-

mases tehases on tootmisjaeskondi, kus tööviljakus on isegi tunduvalt langenud (10 - 40 %). Tööviljakuse üldine tõus on neis tehastes saavutatud üksnes selle arvel, et mõnedes teistes tsehhides on tööviljakust eriti silmapaistvalt tõstetud (kuni 80 %). Sellised suured tehasesisesed erisused reedavad enamasti ettevõtte ebarahuldavat tööd. Järelikult on üksnes keskmiste näitajate alusel langetatud otsus nimetatud tehaste võrdselt heast tööst täiesti väär. Ehkki nende töö lõpptulemused on võrdsed, on olukord neis ettevõtetes siiski väga erinev. Keskmistest see aga ei selgu. Siit tekibki vajadus täiendada keskmiste abil uuritavatele kogumitele antavat üldist iseloomustust veel selliste näitajatega, mis väljendaksid kogumi üksikliikmete vahelisi erisusi, ehk nn. variatsiooninäitavudega.

T a b e l 17 .

Tööviljakuse muutumine N peavalitsuse tehastes.

Tehase nr.	Tööviljakuse kasv (%) üksikutes tootmisjaeskondades					Tööviljakuse keskmine kasv tehases (%)
I	105	105	105	105	105	105
II	100	101	101	102	121	105
III	80	90	95	130	130	105
IV	80	100	105	110	130	105
V	60	80	100	180	x	105

9.1.2. Mõningaid näiteid variatsiooninäitavude praktilisest kasutamisest.

Variatsiooninäitavude praktilise kasutamise võimalusi on palju. Eriti tõhusad on nad näiteks dispersioon- ja korrelatsioonanalüüsi lülitatuna. Nende meetodite käsitlemine ei mahu aga käesoleva teema raamidesse. Piirdume seetõttu ainult paari kõige lihtsama näite esitamisega, kus variatsiooninäitavudel on keskmiste kõrval iseseisev tunnetuslik väärtus.

Oletagem, et tuleb langetada otsus, kumb kahest lina-sordist on kvaliteedilt parem, kas A või B. Lina kvaliteedi põhiliseks kriteeriumiks on saadava kiu pikkus. Olgu keskmine kiu pikkus neil sortidel võrdne, s.o.

$$\bar{x}_A = \bar{x}_B .$$

Teiseks tähtsamaks kvaliteeti mõjustavaks asjaoluks on lina-kiudude pikkuse ühtlus. Parema lina, mille kiud on ühtlasema pikkusega. Sellisest sordist valmistatud kedrus tuleb suhteliselt tugevam kui samasuguse keskmise kiupikkusega, ent ebaühtlasemast sordist valmistatud kedrus. Viimases võib olla küll väga pikki kiude, ent seal on ka ilmselt mõningaid väga lühikesi. Kohtadest, kuhu satuvad lühemad kiud, jääb kedrus nõrgaks ja katkeb hõlpsalt.

Küsimus sellest, kas lina kiudude pikkus on ühtlane või ebaühtlane, kattub küsimusega kiupikkuse varieeruvusest. Ühtlasema pikkusega on see lina-sort, mille varieeruvus on väiksem, suurem sellel, mille varieeruvus on suurem.

Seatud probleemi on ilmselt äärmiselt lihtne lahendada, kui kummagi sordi kohta oleks teada ka kiu pikkuse keskmine varieerumisnäitaja V . Kui $V_A > V_B$, siis on sort B parem.

See otsus kehtib aga muidugi ainult lina kui valmistootte kvaliteedi kohta. Kumba neist sortidest on otstarbekam kasvatama hakata, selle küsimuse lahendamisel tuleb arvestada veel paljusid muidki asjaolusid, mispuhul tuginetakse jällegi keskmistele ja neid täiendavatele variatsiooninäitajatele.

Oletagem näiteks, et A, B ja C sort tagavad pike-maaegsete vaatluste andmetel kõik ühesuguse keskmise saagikuse, s.o.

$$\bar{y}_A = \bar{y}_B = \bar{y}_C ,$$

ent saagikuse varieerumisnäitajad V_1 on üksteisest erinevad ja nimelt

$$V_A > V_B > V_C .$$

Siit selgub, et kõige stabiilsemaid saake on taganud aastast-aastasse sort C. Järelikult on saagi suuruse seisukohalt just sort C kõige parem, sest seda kasvatades on hea saak kõige kindlamini tagatud.

Edasi tuleks selgitada, missugune on sortide B ja C kvaliteetide erisus, sest sedagi ei tohi unustada. Mis sorti lõppeks külvata, selle lahendamine eeldab järelikult õige mitmesuguseid võrdlusi, mille juures tuleb arvestada muidugi ka eri sortide nõudeid pinnase, niiskuse, väetise jms. suhtes. Eeltoodust peaks olema selge, et kõigi nende võrdluste tegemisel ei saa aga piirduda üksnes keskmistega, vaid tuleb arvestada ka vastavate tunnuste varieerumist.

9.2. Elementaarseid variatsiooninäitarve.

Lihtsamateks variatsiooninäitavudeks on variatsiooniamplituud, keskmine lineaarhälve ning viimasest arvutatud variatsioonikoefitsient. Neid kõlbab kasutada selliste praktiliste ülesannete lahendamisel, mis ei eelda tõenäosusteooria rakedamist.

9.2.1. Variatsiooniamplituud.

Kõige üldisema ettekujutuse annab tunnuse väärtuse varieerumisest variatsiooniamplituud R , mis leitakse rea kõige suurema ja väiksema liikme väärtuste vahena, s.o.

$$R = x_{\max} - x_{\min} \quad . \quad (21)$$

Ehkki äärmiselt lihtsakoeline, võimaldab ta siiski eristada võrdlemisi näitlikult eespool toodud andmeid tehaste tööviljakuse muutumise kohta. Kui on teada, et

$$\begin{aligned} R_I &= 0 ; & R_{II} &= 21 ; & R_{III} &= 50 ; \\ R_{IV} &= 50 ; & R_{IV} &= 120 , \end{aligned}$$

siis on ilmne, missuguses tehases on tööviljakus muutunud ühtlasemalt, missuguses vähem ühtlaselt. Variatsiooniamplituud on seda suurem, mida suuremad on rea üksikliikmete erisused. Kui kõikide variantide väärtused on võrdsed, siis $R = 0$.

Variatsiooniamplituudi suurimaks eeliseks on tema arvutamise lihtsus. Peamiseks puuduseks on, et ta ei väljenda mitte rea kõigi, vaid üksnes ekstreemsete liikmete vahelisi erisusi. Seetõttu võib juhtuda variatsiooniamplituudidega sama, mis aritmeetilise keskmisega: suhteliselt erinevate riidade variatsiooninäitartvud võivad osutada võrdseteks.

9.2.2. Keskmine lineaarhälve.

Erinevalt variatsiooniamplituudist üldistab keskmine lineaarhälve kogumi kõigi liikmete vahelisi erisusi ja on seega tunduvalt täiuslikum näitartv. Sisuliselt kujutab ta endast variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste absoluutsete hälvete aritmeetilist keskmist.

Lihtsas statistilises reas leitakse ta valemiga

$$\bar{d} = \frac{\sum /x_i - \bar{x}/}{n}$$

ja variatsioonreas

$$\bar{d} = \frac{\sum /x_i - \bar{x}/f}{\sum f} . \quad (22)$$

Lineaarhälbe valemis rakendatakse variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste hälvete absoluutväärtusi selleks, et hävitada hälvete märke. Vastavate hälvete naturaalkväärtuste kasutamisel oleks $\bar{d}$ mistahes rea puhul null, sest vastavalt aritmeetilise keskmise 1. omadusele on

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

Tabeli 17 andmetel arvutame näiteks tehaste keskmised line-

kaarhálbed järgmiselt (vajalikud arvutused on koondatud tabelisse 18):

$$\bar{d}_I = \frac{0}{5} = 0 ; \quad \bar{d}_{II} = \frac{26}{5} = 5,2 ; \quad \bar{d}_3 = \frac{100}{5} = 20 ;$$

$$\bar{d}_{IV} = \frac{60}{5} = 12 ; \quad \bar{d}_V = \frac{150}{4} = 37,5 \% .$$

Keskmise lineaarhálbe tunnetuslikud eelised hakkavad hästi silma, kui võrrelda näiteks andmeid III ja IV tehase kohta, mida variatsiooniamplituud ei võimaldanud eristada, s.o.

$$\bar{d}_{III} = 20 \quad \text{ja} \quad \bar{d}_{IV} = 12 .$$

Neist andmetest selgub, et tööviljakuse muutumine on olnud III tehases tunduvalt ebaühtlasem kui IV tehases. Seda järeldust saab käesolevas lihtsas näites lähteandmete järgi kontrollida, mispuhul selgub, et $\bar{d}$ väärtus peegeldab tegelikke erisusi III ja IV kogumi vahel põhimõtteliselt õigesti.

T a b e l 18 .

Tsehh	I tehas		II tehas		III tehas		IV tehas		V tehas	
	x_1	$/x_1 - \bar{x}/$	x_1	$/x_1 - \bar{x}/$	x_1	$/x_1 - \bar{x}/$	x_1	$/x_1 - \bar{x}/$	x_1	$/x_1 - \bar{x}/$
1	105	0	100	5	80	25	80	25	60	45
2	105	0	101	4	90	15	100	5	80	25
3	105	0	101	4	95	10	105	0	100	5
4	105	0	102	3	130	25	110	5	180	75
5	105	0	121	16	130	25	130	25	x	x
	-	0	-	26	-	100	-	60	-	150

9.2.3. Variatsioonikoefitsient - suhteline variatsiooninäit- arv.

Keskmine lineaarhálve saadakse alati samades mõõtühikut-
tes, milles avaldatakse variante. Seepärast ei saa tema abil

võrrelda erinevates ühikutes mõõdetavate suuruste varieerumise intensiivsust. Selline vajadus esineb praktikas aga õige sageli. Näiteks võib olla vaja lahendada järgmine probleem: mis varieerub rohkem, kas nisu saagikus või keskmine suvetemperatuur? Saagikust mõõdetakse, nagu teada, tsentnerites hektarilt, temperatuure aga kraadides. Kui selgub, et saagikuse keskmine lineaarhälve on 12 ts/ha ja temperatuuril 8°, siis on seatud küsimust võimatu lahendada.

Teiseks keskmise lineaarhälbe puuduseks on, et tema suurus sõltub variantide absoluutväärtusest ehk nivoost, mille ümber variantide väärtused kõiguvad. Oletagem näiteks, et tahetakse lahendada küsimust, mis on viimaste aastate jook-sul suhteliselt rohkem kõikunud, kas kolhoosi või keskmise kolhoosniku sissetulek. Et kolhoosi sissetulekud kõiguvad sadades tuhandetes ja keskmise kolhoosniku sissetulekud sadades rublades, siis on juba ette kindel, et esimeste lineaarhälve on suurem, teistel väiksem, ehkki kolhoosniku sissetulek võib olla suhteliselt rohkem muutunud.

Nii üks kui teine nimetatud puudustest on keskmisele lineaarhälbele omased seetõttu, et ta on absoluutne variatsiooninäitaja ega peagi vastama selliselt seatud küsimustele.

Neid probleeme saab lahendada suhtelise variatsiooni-näitarvu ehk variatsioonikoefitsiendi $V_{\bar{d}}$ arvutamisega, mis leitakse

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \quad (23)$$

Et sellest samanimeliste suuruste suhtest dimensioonid taanduvad, siis on variatsioonikoefitsient nimetu suurus ja järelikult täiesti võrreldav mistahes teise nähtuse kohta arvutatud variatsioonikoefitsiendiga.

Variatsioonikoefitsient avaldatakse enamasti protsentides, mispuhul teda korrutatakse veel sajaga. Keskmise lineaarhälbe ja variatsioonikoefitsiendi arvutamist demonstreerime tabeli 2 andmetel (arvutused vt. tabel 19).

Selgub, et 2069 tähtsama ehitusorganisatsiooni plaani

täitmise protsendi keskmine lineaarhälve on

$$\bar{d} = \frac{51617,2}{2069} = 24,95 \%$$

ehk

$$v_{\bar{d}} = \frac{24,95}{78,38} = 0,318 \text{ ehk } 31,8 \%$$

T a b e l 19 .

x_1	f_1	$ x_1 - \bar{x} $ $\bar{x} = 78,4$	$ x_1 - \bar{x} f_1$
(1)	(2)	(3)	(4)
10	43	68,4	2941,2
30	161	48,4	7792,4
50	370	28,4	10508,0
70	527	8,4	4426,8
90	514	11,6	5962,4
110	278	31,6	8784,8
130	106	51,6	5469,6
150	47	71,6	3365,2
170	10	91,6	916,0
190	13	111,6	1450,8
Σ	2069	Σ	51617,2

Võrreldes keskmise lineaarhõlbega on variatsioonikoeffitsiendi peamine erisus selles, et kui esimene mõõdab varieerumishõlbuse absoluutset, siis viimane mõõdab selle suhtelist intensiivsust.

Analooiliselt võib leida suhtelist variatsiooninäitartvu ka variatsiooniamplituudist, jagades selle aritmeetilise keskmisega, s.o.

$$v_R = \frac{R}{\bar{x}} \quad (24)$$

Variatsioonikoefitsiente on nimetatud kirjanduses ka ost-sillatsiooninäitajaiks.

9.3. Dispersioon ja standardhälve.

9.3.1. Dispersioon.

Dispersioon leitakse üldjoontes samuti nagu keskmine lineaarhälve. Variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste hälvete $x_i - \bar{x}$ märkide kaotamiseks ei lülitata aga arvutustesse mitte nende absoluutväärtusi, vaid tõstetakse nad ruutu. Sellega saavutatakse sama eesmärk nagu lineaarhälbe leidmisel, ühtlasi aga saadakse näitaja, mida on hõlpus seostada tõenäosusteooria rakendustega. See omadus ongi teinud dispersioonist kõige väärtuslikuma variatsiooninäitarvu.

Niisiis on dispersioon variantide individuaalväärtuste ja nende aritmeetilise keskmise vaheliste hälvete ruutkeskmine; teisiti nimetatakse ka ruutkeskmiseks hälbeks.

Dispersiooni valem on lihtsa statistilise rea puhul

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

ja variatsioonrea puhul

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{\sum f} \quad , \quad (25)$$

kus dispersioon on tähistatud σ^2 .

Ruutkeskmisena on dispersiooni mõõtühikuks variandi dimensiooni ruut. On x_i mõõtühikuks meeter, siis on dispersioon ruutmeetrites. Tihti saadakse aga niiviisi kas raskesti lahtimõtestatavad või üldse mõttetud ühikud, kui variantide mõõtühikuks on näiteks: tükk, rubla, tonn, protsent jne. Sellised tõlgendusraskused on aga hõlpsasti ületatavad dispersioonist ruutjuure leidmise teel.

Dispersiooni arvutamise tehnika selgub tabelist 20 (lähteandmed vt. tabel 11). Tabelisse koondatud arvutuste andmeil on

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f}{\sum f} = \frac{5020}{20} = 251 (\%^2) .$$

T a b e l 20 .

x_i	f_i	$x_i - \bar{x}$ $\bar{x} = 113$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f$
90	4	- 23	529	2116
110	10	- 3	9	90
130	5	17	289	1445
150	1	37	1369	1369
	20	Σ		5020

9.3.2. Standardhälve.

Dispersiooni ruutjuur annab standardhälbe ehk keskmise ruuthälbe¹ (tähis σ). Seda võidakse arvutada ühega järgmistest valemitest

$$\sigma = \sqrt{\sigma^2} ;$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} ;$$

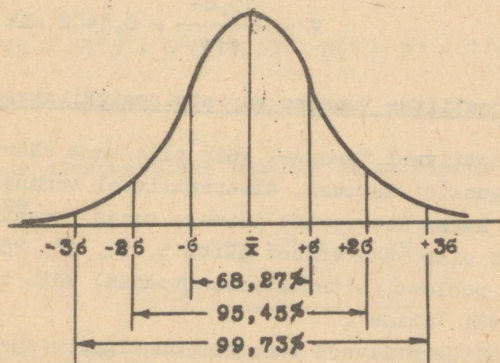
$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f}{\sum f}} . \quad (26)$$

Tabelis 20 tehtud arvutuste kohaselt on

¹'Keskmiist ruuthälvet' (σ) ei tohi muidugi segi ajada osas 9.3.1 käsitletud 'ruutkeskmise hälbega' (σ^2). Et sel-line segimineku oht nende terminite samakõlalisisuse tõttu on olemas, siis on käesoleval juhul nähtavasti otstarbekas kasutada eelistatavalt vastavaid võõrsõnu 'standardhälve' ja 'dispersioon'.

$$\sigma = \sqrt{251} = 15,84 \% .$$

Standardhälve avaldub alati samades mõõtühikutes, milles variandidki. Temaga võib opereerida põhimõtteliselt samuti nagu keskmise lineaarhällbega. Eriliseks eeliseks viimasega võrreldes on aga see, et standardhälve on seotud normaalse jaotumisseadusega. Graafilises interpretatsioonis on standardhälve Gauss-Laplace'i kõvera käänupunkti abstsiss (vt. joon. 9).



Joon. 9. Standardhälve väärtus on määratud normaalse jaotuskõvera käänupunktidega. Vahemik $x \pm \sigma$ hõlmab 68,27 % normaalse variatsioonikõvera alusest pindalast.

9.3.3. Variatsioonikoefitsient.

Keskmise lineaarhällbe taoliselt on ka standardhälve absoluutne hällbimis mõõt. Seepärast ei saa ka tema abil võrrelda 1) erinevates ühikutes mõõdetavate tunnuste varieerumist ja 2) väga erineva nivoo ümber toimuvaid kõikumisi. Nagu keskmise lineaarhällbe puhul, nii aitab need puudused ka siin ületada variatsioonikoefitsiendi arvutamine, mis leitakse standardhälve ja aritmeetilise keskmise suhtena, s.o.

$$v = \frac{\sigma}{\bar{x}}$$

(27)

ehk

$$v = \frac{\sqrt{\sigma^2}}{\bar{x}} .$$

Eespool lahendatud lihtsa arvnäite järgi on

$$v = \frac{15,84}{113} = 0,1402 \text{ ehk } 14,02 \% .$$

9.3.4. Alternatiivse tunnuse variatsiooninäitarvud.

Kvantitatiivsel tunnusel võib olla tema väärtusvaru piires mitmesuguseid väärtusi, alternatiivsel tunnusel saab neid olla ainult kaks. Konkreetse inimese vanus (kvantitatiivne tunnus) võib olla väga erinev näit. 5, 16, 18, 25, 61 jne. aastat; soo poolest (alternatiivne tunnus) võib ta olla ainult kas 'mees' või 'naine'.

Ehkki alternatiivsed tunnused on algselt enamasti atribuutiivsed, on neid hõlpus taandada kvantitatiivseteks. Selleks tähistame kahest võimalikust väärtusest ühe arvuga 1 ja teise arvuga 0.

Oletagem, et mingis inimkollektiivis on 400 meest ja 600 naist. Meeste osatähtsus antud kollektiivis, tähistame selle tähega p , on sel juhul

$$p = \frac{400}{1000} = 0,4 \quad \text{ehk } 40 \%$$

ja naiste osatähtsus vastavalt $1 - p = 1,0 - 0,4 = 0,6$ ehk 60 %. Tähistame mehe (x_1) arvuga 1 ja naise (x_2) arvuga 0. Et mehi ja naisi pole võrdselt, tuleb nende keskmine arv kollektiivis leida kaalutud aritmeetilise keskmisena, võttes kaaludeks f nende osatähtsused. Neil tingimustel selgub, et

$$\bar{x} = \frac{\sum x_1 f}{\sum f} = \frac{1 \cdot p + 0(1 - p)}{p + (1 - p)} = p .$$

Selgub, et alternatiivse tunnuse keskmine on võrdne vastavat tunnust omavate kogumiliikmete osatähtsusega (resp. erikaaluga, suhtelise osamääraga, suhtelise sagedusega) antud kogumis.

Kui lülitada need tulemused, s.o. $x_1 = 1$; $x_2 = 0$; $\bar{x} = p$; $f_1 = p$; $f_2 = 1 - p$, tavalisse dispersiooni valemisse, siis saame

$$\begin{aligned}\sigma^2 &= \frac{\sum (x_i - \bar{x})^2 f}{\sum f} = \frac{(1 - p)^2 p + (0 - p)^2 (1 - p)}{p + (1 - p)} = \\ &= (1 - p)^2 p + p^2 (1 - p) = p(1 - p) (1 - p) + p = \\ &= p(1 - p) .\end{aligned}$$

Seega on alternatiivse tunnuse dispersiooni valem üllatavalt lihtne

$$\boxed{\sigma^2 = p(1 - p)} . \quad (28)$$

Meie arvnaites järelikult

$$\sigma^2 = 0,4 \cdot 0,6 = 0,24 .$$

Siit saame standardhälbe

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,24} = 0,49$$

ja variatsioonikoefitsiendi

$$v = \frac{\sigma}{\bar{x}} = \frac{0,49}{0,6} = 0,82 \text{ ehk } 82 \% .$$

Alternatiivse tunnuse varieerumise omapäraga on määratud ka tema dispersiooni maksimaalne väärtus. Teame, et korutis on maksimaalne, kui tegurid on võrdsed. On selge, et alternatiivse tunnuse varieerumus on kõige suurem, kui $p = 0,5$. Antud juhul tähendab siis, kui mehi ja naisi on kollektiivis võrdsel arvul. Sel juhul on

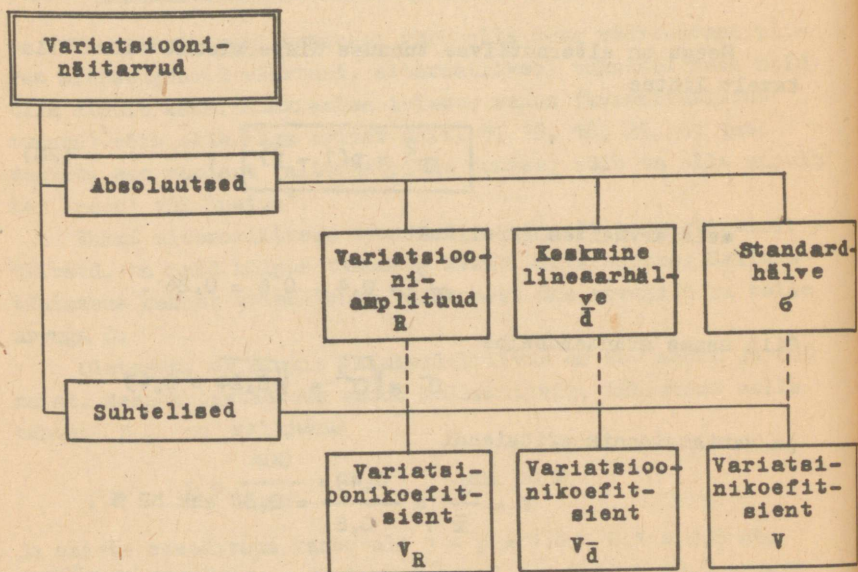
$$\sigma^2 = 0,5 \cdot 0,5 = 0,25$$

$$\sigma = 0,5$$

ja

$$V = \frac{0,5}{0,5} = 1,0 \text{ ehk } 100 \% .$$

Koosneks vaadeldav inimrühm ainult meestest või ainult naistest, siis oleks $\sigma^2 = 1 \cdot 0 = 0$, mis tähendab, et uuritav tunnus ei varieeru või teisiti sõnastatult, et see on oma koostiselt ühtlane (homogeenne).



Joon. 10. Variatsiooni-näitavude liigitelu.

9.3.5. Dispersioonide liitmise lause.

Kui jaotada uuritav kogum teatud tunnuste järgi rühmadeks ning leida iga rühma kohta eraldi aritmeetiline keskmine, siis on teada, et üldkeskmine on rühmkeskmiste keskmine, s.o.

$$\bar{x} = \frac{\sum \bar{x}_i f}{\sum f},$$

kus $\bar{x}_i$ on rühmkeskmised ja $\bar{x}$ üldkeskmine.

Analoogiliselt võib arvutada eraldi iga rühma kohta ka dispersiooni ja on mõistetatav, et teatud side peab valitsema ka üld- ja rühmdispersioonide vahel. Seda sidet väljendab dispersioonide liitmise lause: ülddispersioon on võrdne rühmdispersioonide keskmise ja rühmadevahelise dispersiooni summaga, s.t.

$$\sigma^2 = \sigma_1^2 + \delta^2, \quad (29)$$

kus σ^2 on ülddispersioon;
 σ_1^2 - rühmdispersioonide keskmine ja
 δ^2 - rühmadevaheline dispersioon.

Võrduse (29) parempoolseid liikmeid arvutatakse algandmetest

$$\sigma_1^2 = \frac{\sum \sigma_i^2 f}{\sum f}, \quad (30)$$

millest nähtub, et on tegemist rühmdispersioonide kaalutud aritmeetilise keskmisega ja

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f}{\sum f}, \quad (31)$$

kust selgub, et rühmadevaheline dispersioon on õigupoolest rühmkeskmiste dispersioon üldkeskmistest.

Nagu üld- ja rühmkeskmiste vahelisi seoseid, nii saab ka üld- ja rühmdispersioonide vahelisi seoseid kasutada õige mitmeks otstarbeks. Nende abil saab leitud tulemusi vastastikku kontrollida (1). Kui kasutada olevad andmed on millegi pärast lünklikud, nagu seda praktikas kahjuks tihti peale ette tuleb, siis saab mõningaid puuduvaid rühmdispersioone leida teiste kaudu, mille arvutamiseks on andmed olemas (2). Kõige tähtsam on aga see, et dispersioonide liitmise lausele baseerub olulisel määral üks kaasaja kõige teravmeelsemaid statistilise analüüsi võtteid - dispersioonanalüüs (3), mille kasutuselevõtmine tähendas olulist pööret kõigi nn. eksperimentaalteaduste uurimismetoodikas.

Dispersioonanalüüs võimaldab uurida täpselt ka niisuguste tegurite mõju, mis ei allu otsesele mõõtmisele. Nii võib selle meetodi abil kindlaks teha näiteks sordi, künnisügavuse, künni-, külvi- ja koristustähtaja, väetise hulga ja keskmise suvetemperatuuri mõju põllukultuuride saagikusele; töölise vanuse, soo, kvalifikatsiooni ja kehaehituse mõju tema tööviljakusele jne. Et analoogilise seadega probleeme esineb majanduslikus praktikas tihti, siis on dispersioonanalüüs üks kõige lubavamaid majandusliku analüüsi meetodeid. (Dispersioonanalüüsi üksikasjalik käsitus ei mahu käesoleva teema raamesse).

9.4. Dispersiooni lihtsustatud arvutamine.

Dispersiooni arvutamine kujuneb valem 26 järgi sageli väga vaevanõudvaks, sest hälvete ruutimise tõttu kasvavad arvud väga suurteks. Nagu aritmeetilise keskmise arvutamist, nii saab ka dispersiooni leidmist mitmesuguste võtetega tunduvalt lihtsustada. Selleks sobivad kõigepealt mõned dispersiooni omadused.

9.4.1. Dispersiooni omadused ja nende praktiline kasutamine.

9.4.1.1. Esimene omadus.

Kui vähendada variante mingi meelevaldse arvu võrra, siis dispersioon ei muutu, s.o.

$$\frac{\sum [(x_i - a) - (\bar{x} - a)]^2}{n} = \frac{\sum (x_i - a - \bar{x} + a)^2}{n} =$$

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

See omadus on seotud aritmeetilise keskmise 4. omadusega. Tema praktilist kasutamist demonstreeriv arvnäide (vt. tabel 21) veenab meid saadavate tulemuste täielikus kokkulangemises eespool leitud tulemustega. Võttes meelevaldseks arvuks $a = 9$, saame $\bar{z} = 23$, ning dispersiooni

$$\sigma^2 = \frac{5020}{20} = 251 (\%)^2.$$

Aritmeetilise keskmise 4. omaduse kohaselt on

$$\bar{x} = \bar{z} + a = 23 + 90 = 113 \%.$$

T a b e l 21.

x_i	f_i	z_i $(x_i - a)$	$z_i - \bar{z}$	$(z_i - \bar{z})^2$	$(z_i - \bar{z})^2 f_i$
90	4	0	- 23	529	2116
110	10	20	- 3	9	90
130	5	40	17	289	1445
150	1	60	37	1369	1369
Σ	20	Σ	Σ	Σ	5020

9.4.1.2. Teine omadus.

Kui vähendada variante mingi meelevaldne arv k korda, siis väheneb dispersioon k^2 korda, s.o.

$$\frac{\sum (\frac{x_i}{k} - \frac{\bar{x}}{k})^2}{n} = \frac{\sum (\frac{x_i - \bar{x}}{k})^2}{n} = \frac{1}{k^2} \cdot \frac{\sum (x_i - \bar{x})^2}{n} = \frac{1}{k^2} \sigma^2.$$

Seda omadust saab kasutada järgmise dispersiooni lihtsustatud arvutamise eeskirja formuleerimisel:

$$\sigma^2 = \frac{\sum \left(\frac{x_1 - \bar{x}}{k} \right)^2 f_1}{\sum f_1} \cdot k^2 \quad (32)$$

Valemit (32) eespool toodud arvnäite puhul rakendades saame (vt. tabel 22)

$$\sigma^2 = \frac{50,20}{20} \cdot 10^2 = 2,51 \cdot 100 = 251 ,$$

s.o. täpselt samasuguse tulemuse nagu varem.

T a b e l 22 .

x_1	f_1	$\frac{x_1}{k} = z_1$	$z_1 f_1$	$z_1 - \bar{z}$ $\bar{z} = 11,3$	$(z_1 - \bar{z})^2$	$(z_1 - \bar{z})^2 f$
90	4	9	36	- 2,3	5,29	21,16
110	10	11	110	- 0,3	0,09	0,90
130	5	13	65	1,7	2,89	14,45
150	1	15	15	3,7	13,69	13,69
Σ	20	Σ	226	Σ	Σ	50,20

Dispersiooni seda omadust kasutatakse enamasti koos eelmise omadusega või siis momentide meetodiga.

9.4.2. Momentide mõiste ja kasutamine dispersiooni arvutamisel.

9.4.2.1. Momendi mõiste.

Momentideks (ehk jaotusmomentideks) nimetatakse statistikas variantide individuaalväärtuste ja mistahes arvu vaheliste hälvete mingis astmes arvutatud keskmisi.

Kui variantide hälbed arvutatakse nullist, nimetatakse momente algmomentideks, kui aritmeetilisest keskmisest, siis keskmomentideks ja kui mingist meelevaldselt võetud arvust, siis tingmomentideks.

Olenevalt sellest, mitmendas astmes hälvete keskmisega on tegemist, jaotatakse neid esimest, teist, kolmandat jne. järku momentideks.

Näitlikku ülevaadet mitmesugustest momentidest pakub tabel 23. Tabelis on tähistatud algmomentid kreeka tähega γ (nüü), keskmomentid μ (müü) ja tingmomentid η (eeta). Momenti järku tähistab tema juurde kirjutatud araabia number. Nii on γ_2 teist järku alg- ja μ_4 neljandat järku keskmoment.

T a b e l 23 .

	Algmomentid (γ)	Keskmomentid (μ)	Tingmomentid (η)
1. järku momentid	$\gamma_1 = \frac{\sum (x - 0)f}{\sum f}$	$\mu_1 = \frac{\sum (x - \bar{x})f}{\sum f}$	$\eta_1 = \frac{\sum (x - a)f}{\sum f}$
2. järku momentid	$\gamma_2 = \frac{\sum (x - 0)^2 f}{\sum f}$	$\mu_2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$	$\eta_2 = \frac{\sum (x - a)^2 f}{\sum f}$
3. järku momentid	$\gamma_3 = \frac{\sum (x - 0)^3 f}{\sum f}$	$\mu_3 = \frac{\sum (x - \bar{x})^3 f}{\sum f}$	$\eta_3 = \frac{\sum (x - a)^3 f}{\sum f}$
4. järku momentid	$\gamma_4 = \frac{\sum (x - 0)^4 f}{\sum f}$	$\mu_4 = \frac{\sum (x - \bar{x})^4 f}{\sum f}$	$\eta_4 = \frac{\sum (x - a)^4 f}{\sum f}$

Tabelit 23 vaadeldes selgub, et selles on mitmeid juba varemgi hästi tuntud suurusi. Nii on:

esimest järku algmoment γ_1 = aritmeetiline keskmine,

esimest järku keskmoment $\mu_1 = \text{null}$
 teist järku keskmoment $\mu_2 = \text{dispersioon}$.

Eri liiki ja järku momentide vahel valitseb hulk mitmesuguseid seoseid, mistõttu üht liiki momente võib leida ka mitte otse algandmetest, vaid teiste momentide kaudu. Kõige tülikam on arvutada keskmomente, sest see eeldab kahe arvutusetapi läbimist: 1) aritmeetilise keskmise leidmist ja 2) dispersiooni arvutamist. Mõlemad need tööetapid nõuavad pea ühepalju tööd.

Keskmomente, sealhulgas ka dispersiooni, võib leida aga ka tingmomentide kaudu, millega töö lihtsustub peaaegu poole võrra, sest tingmomentide arvutamine ei eelda aritmeetilise keskmise leidmist.

9.4.2.2. Dispersiooni ja aritmeetilise keskmise leidmine momentide meetodil.

Teist järku keskmoment võrdub teist järku tingmomendi ja esimest järku tingmomendi ruudu vahega, s.o.

$$\mu_2 = \eta_2 - \eta_1^2. \quad (33)$$

Sellele lausele baseeruvat arvutuskäiku nimetataksegi dispersiooni leidmiseks momentide meetodil. Vastav valem on järgmine

$$\sigma^2 = \frac{\sum (x_1 - a)^2 f}{\sum f} - \left(\frac{\sum (x_1 - a) f}{\sum f} \right)^2 \quad (34)$$

ehk lihtsa statistilise rea andmetel

$$\sigma^2 = \frac{\sum (x_1 - a)^2}{n} - \left(\frac{\sum (x_1 - a)}{n} \right)^2$$

Leidnud momentide meetodil dispersiooni, piisab tehtud arvutustest praktiliselt alati, et leida hõlpsal viisil ka

aritmeetiline keskmine. Selleks tuleb esimest järku tingmomen-
dile liita tinglik suurus a , mille võrra variantide
üksikväärtusi vähendati, s.o.

$$\bar{x} = \frac{\sum (x_i - a)f}{\sum f} + a.$$

Kui tähistada variantide individuaalväärtuste ja mee-
valdse suuruse a vahelisi hälbeid tähega e , s.o.

$$x_i - a = e,$$

siis võib kirjutada valemi 34 tunduvalt lihtsamalt

$$\sigma^2 = e^2 - \bar{e}^2$$

ja

$$\bar{x} = \bar{e} + a.$$

Momentide meetodi praktiline kasutamine selgub tabelist
24. Täiendavalt on rakendatud seal ka veel dispersiooni II
omadust, mis eeldab tingmomentide korrutamist vastavate mee-
levaldselt võetud jagatistega. Tabelis arvutatud andmetel on

$$\begin{aligned} \sigma^2 &= \frac{20787}{2069} \cdot 10^2 - \left(\frac{-1369}{2069} \cdot 10 \right)^2 = 1004,688 - 6,6167^2 = \\ &= (1004,688 - 43,781) = 960,907 \approx 960,9 \end{aligned}$$

$$\text{ning } \sigma = \sqrt{960,9} = 30,99.$$

Aritmeetiline keskmine on neil andmeil

$$\bar{x} = 85 - 6,62 = 78,38.$$

Kasutatud arvutustabelis 24 on enamik arve nii lihtsad,
et nende peast leidmine ei tekita raskusi. Momentide meetodi
poolt pakutav tšõ- ja ajasääst saab eriti ilmsiks, kui võr-
velda arvutustabeleid 24 ja 25. Viimases on arvutused, mis on
vajalikud dispersiooni leidmiseks valemiga 26. Seejuures on
eeldatud, et $\bar{x}$ on juba varem teada. Täieliku pildi valemi

T a b e l 24 .

Arvutused dispersiooni leidmiseks momentide
meetodil.

x_i	f_i	x_i	$x_i - a$	$\frac{x_i - a}{k}$	$(\frac{x_i - a}{k})f$	$(\frac{x_i - a}{k})^2$	$(\frac{x_i - a}{2})^2 f$
...- 10	20	5	- 80	- 8	- 160	64	1280
10 - 20	23	15	- 70	- 7	- 161	49	1127
20 - 30	66	25	- 60	- 6	- 396	36	2376
30 - 40	95	35	- 50	- 5	- 475	25	2375
40 - 50	153	45	- 40	- 4	- 612	16	2448
50 - 60	217	55	- 30	- 3	- 651	9	1953
60 - 70	254	65	- 20	- 2	- 508	4	1016
70 - 80	273	75	- 10	- 1	- 273	1	273
80 - 90	282	85			-3236		
90 - 100	232	95	10	1	232	1	232
100 - 110	178	105	20	2	356	4	712
110 - 120	100	115	30	3	300	9	900
120 - 130	74	125	40	4	296	16	1184
130 - 140	32	135	50	5	160	25	800
140 - 150	27	145	60	6	162	36	972
150 - 160	20	155	70	7	140	49	980
160 - 170	7	165	80	8	56	64	448
170 - 180	3	175	90	9	27	81	243
180 - 190	5	185	100	10	50	100	500
190 - 200	8	195	110	11	88	121	968
Σ	2069	Σ	Σ	Σ	1867	Σ	20787
					-3236		
					-1369		

Tabel 25 .

Arvutused dispersiooni leidmiseks valemi
26 järgi.

x_1	f_1	$\bar{x} = 87,4$ $x_1 - \bar{x}$	$(x_1 - \bar{x})^2$	$(x_1 - \bar{x})^2 f$
5	20	- 73,4	5387,56	107751,20
15	23	- 63,4	4019,56	92449,88
25	66	- 53,4	2851,56	188202,96
35	95	- 43,4	1883,56	178938,20
45	153	- 33,4	1115,56	170680,68
55	217	- 23,4	547,56	118820,52
65	254	- 13,4	179,56	45608,24
75	273	- 3,4	11,56	3155,88
85	282	6,6	43,56	12283,92
95	232	16,6	275,56	63929,92
105	178	26,6	707,56	125945,68
115	100	36,6	1339,56	133956,00
125	74	46,6	2171,56	160695,44
135	32	56,6	3203,56	102513,92
145	27	66,6	4435,56	119760,12
155	20	76,6	5867,56	117351,20
165	7	86,6	7499,56	52496,92
175	3	96,6	9331,56	27994,68
185	5	106,6	11363,56	56817,80
195	8	116,6	13595,56	108764,48
Σ	2069	Σ	Σ	1988117,64

26 rakendamisel tehtava töö mahust saaks siis, kui liita tabelid 1 ja 25.

Tähelepanдав on, et rohkem vaevanõudval arvutusviisil pole mingisuguseid eeliseid ka saadavate tulemuste täpsuse mõttes. Ka 25 tabeli andmeil on

$$\sigma^2 = \frac{1988177,64}{2069} = 960,907.$$

Suure tömahukuse tõttu võibki dispersiooni leidmise lihtsustamata arvutusviisi soovitada ainult äärmiselt lihtsate ülesannete lahendamiseks.

9.4.3. Dispersiooni leidmine liitmismenetlusega.

Liitmismenetlusega leitakse dispersioon valemi järgi

$$\sigma^2 = k^2 \left[\frac{2(E_1 + E_2) - (\delta_1 + \delta_2)}{\sum f} - b'^2 \right], \quad (35)$$

kus peale osas 2.3 kirjeldatud suuruste δ_1 ja δ_2 on E_1 - järgsummade s_2 järgsumma rea madalamast pooldest (vt. tabel 26, veerg 5) ja

E_2 - samasugune summa rea kõrgemast pooldest ning

$$b' = \frac{\delta_2 - \delta_1}{\sum f}.$$

Arvutuste lihtsustamiseks on sobiv töötada järgmise skeemi kohaselt:

$$1) \quad 2(E_1 + E_2) = 2(8042 + 4903) = 25890$$

$$(\delta_1 + \delta_2) = -(3236 + 1867) = -5103$$

$$20787$$

$$2) \quad \frac{2(E_1 + E_2) - (\delta_1 + \delta_2)}{\sum f} = \frac{20787}{2069} = 10,04688 \approx 10,047$$

$$3) \quad b' = \frac{1867 - 3236}{2069} = -0,662$$

T a b e l 26

Arvutused dispersiooni leidmiseks
liitmismenetlusega.

x_1	f_1	s_1	s_2	s_B
(1)	(2)	(3)	(4)	(5)
5	20	20	20	20
15	23	43	63	83
25	66	109	172	255
35	95	204	376	631
45	153	357	733	1364
55	217	574	1307	2671
65	254	828	2135	4806
75	273	1101	3236 = δ_1	8042 = E_1
85	282			
95	232	686	1867 = δ_2	4903 = E_2
105	178	454	1181	3036
115	100	276	727	1855
125	74	176	451	1128
135	32	102	275	677
145	27	70	173	402
155	20	43	103	229
165	7	23	60	126
175	3	16	37	66
185	5	13	21	29
195	8	8	8	8
Σ	2069	Σ	Σ	Σ

$$4) G^2 = 10^2 [10,047 - (-0,662)^2] = \\ = 100(10,047 - 0,438) = 100 \cdot 9,609 = \underline{960,9}.$$

Järelikult annab liitmismenetlus siingi täpselt sama tulemuse nagu korrutamismenetlus.

9.5. Muid variatsiooninäitarve.

Peale variatsiooniamplituudi, keskmise lineaarhälbe, dispersiooni ja standardhälbe kasutatakse veel muidki näitarve. Nad ei ole küll nii levinud, ent tihtipeale võimaldavad nad tunnuste väärtuste varieerumist isegi mõnevõrra piltlikumalt väljendada kui näiteks standardhälve. Esitame lühikesse ülevaate veel mõnedest variatsiooninäitervudest, millest mitmed tuginevad asendikeskmistele.

9.5.1. C. Gini keskmine absoluuthälve (media differenza).

Üldiselt arvutatakse variatsiooninäitarve aritmeetilise keskmise suhtes. Väga rangelt võttes on see täiesti mõtestatud ainult siis, kui uuritavad read lähenevad oma struktuurilt normaalsetele variatsioonriidadele. Looduslike, eriti bioloogiliste nähtuste uurimisel on see nõue enamasti rahuldatud. Ühiskondlike ja majanduslike nähtuste uurimisel pole see tingimus aga sugugi mitte alati rahuldatud. Eriti majanduslike nähtuste käsitlemisel saadakse tihti normaalsetest jaotusriidadest tunduvalt hälbivaid andmeridu. Sellistes kogumites on õigem mõõta mitte eri variantide hälbeid aritmeetilisest keskmisest, vaid kõigi võimalike teguripaaride vahelisi erisusi.

C. Gini soovitabki:

- a) leida kõigi teguripaaride vahelised hälbed, vaadeldes selleks kahekaupa kõiki kogumi liikmeid;
- b) saadud hälbed liita ja
- c) nende summa jagada hälvete arvuga.

Gini keskmine hälve tuleks selle metoodika kohaselt leida valemiga

$$\mathcal{E} = \frac{2}{n(n-1)} \cdot \sum (n+1-2i) x_{n-i+1} \quad , \quad (36)$$

kus $\mathcal{E}$ (epsilon) on Gini variatsiooninäitaja;

n - liikmete arv reas;

i - variandi järjekorranumber kahanevas järjestuses kor-
rastatud statistilises reas;

x - rea liige.

Demonstreerime $\mathcal{E}$ arvutamist paari lihtsa arvnäite va-
ral.

1. näide. Kogumis on viis liiget

$$x_1 = 4, 7, 12, 15, 20$$

Valemi järgi saame

$$\begin{aligned} \mathcal{E} &= \frac{2}{5(5-1)} (5+1-2 \cdot 1) 20 + (5+1-2 \cdot 2) 15 + \\ &+ (5+1-2 \cdot 3) 12 + (5+1-2 \cdot 4) 7 + (5+1-2 \cdot 5) 4 = \\ &= \frac{2}{20} (4 \cdot 20 + 2 \cdot 15 + 0 \cdot 12 - 2 \cdot 7 - 4 \cdot 4) = \frac{1}{10} (80 + 30 - 14 - 16 = \\ &= \frac{80}{10} = 8 . \end{aligned}$$

Nagu arvutustest selgub, tulevad kordajad $(n+1-2i)$ va-
riantidele, mille väärtus on keskmisest suurem, positiivsed
ja keskmisest väiksematele negatiivsed.

Sama ülesannet võib lahendada ka teisiti. Teeme kõige-
pealt kindlaks, mitu hälvet antud kogumis üldse saab leida.
Hälvete arv võrdub kombinatsioonidega viiest elemendist kahe-
kaupa, s.o.

$$\frac{n(n-1)}{2} = \frac{5 \cdot 4}{2} = 10 .$$

Seejärel kirjutame vastavad 10 variandipaari välja, leiame nende vahed ja viimaste summa, s.o.

Variandipaarid	Vahe
20 - 15	5
20 - 12	8
20 - 7	13
20 - 4	16
15 - 12	3
15 - 7	8
15 - 4	11
12 - 7	5
12 - 4	8
7 - 4	3
	80

Siit on

$$\mathcal{E} = \frac{80}{10} = 8.$$

Nagu selgub, pole Gini variatsiooninäitaja arvutamine vähemalt lihtsates ridades kuigivõrd tülikam kui dispersiooni leidmine.

Variatsioonirea andmetel on selle leidmine muidugi komp-
litseeritum. Demonstreerime seda praktilise näite varal, ka-
sutatud tööviisi teoreetilist kirjeldust pakkumata.

2. näide. On teada andmed 200 töölise tunnipalga kohta:

palk	tööliste arv
x_1	f_1
1.80	10
1.60	20
1.40	50
1.20	100
1.00	20
	200

Neil andmetel leiame otsitava koefitsiendi järgmiselt:

a) korraldame variatsioonirea kahanevasse järjekorda
(väiksemad variandid ülal, suuremad all);

b) korrutame variandid nende sagedustega;

c) saadud korrutised ($x_i f_i$) korrutame veel läbi kõigi järgmiste variantide sageduste summaga (z_i), millesse variandi enda sagedust ei ole arvatud. Järgmiste variantidega talitame samuti, ent antud liikmest suurema liikme sageduse lülitame sageduste summasse miinusemärgiga;

d) liidame leitud korrutised ja jagame n elemendist kahekaupa võetud kombinatsioonide arvuga, s.o. antud juhul arvuga

$$\frac{200(200 - 1)}{2} = 19900 .$$

Tabel 27 andmeil, kuhu on koondatud vajalikud arvutused, leiame, et

$$\bar{E} = \frac{4040}{19900} = 0,203 .$$

Nagu muudest eespool kirjeldatud absoluutsetest variatsiooninäitavusest, võib ka sellest leida suhtelise. Selleks jagatakse $\bar{E}$ aritmeetilise keskmisega.

Lõpuks olgu märgitud, et Gini variatsiooninäitav osutub vajalikuks ikka ainult siis, kui on tööpoolest tegemist niisuguse kogumiga, mille jaoks pole olemas piisavalt tüüpilist keskmist. Seda võib juhtuda kogumites, kus pole üldse niisuguseid variante, mille sagedus oleks teistest silmapaistvalt suurem või siis sellistes, kus taolisi on mitu. Viimasel juhul on tegemist paljude maksimumide või miinimumide reaga.

T a b e l 27 .

C. Gini variatsiooninäitavuse leidmine.

x_i	f_i	$x_i f_i$	teiste variantide sageduste summa (z_i)	$x_i f_i z_i$
1.80	10	18,0	(20+50+100+20) = 190	3420
1.60	20	32,0	(50+100+20-10) = 160	5120
1.40	50	70,0	(100+20-10-20) = 90	6300
1.20	100	120,0	(20-10-20-50) = -60	-7200
1.00	20	20,0	(-10-20-50-100)=-180	-3600
Σ	200	Σ	Σ	4040

9.5.2. F. Galtoni kvartiihälve.

F. Galtoni kvartiihälve

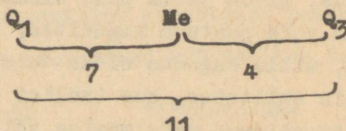
$$K^{\text{Galt.}} = \frac{1}{2} (Q_3 - Q_1) \quad (37)$$

täidab ligikaudu samasugust tunnetuslikku funktsiooni nagu variatsiooniamplituud, iseloomustab aga mitte korrastatud rea otspunktide, vaid kummagi rea poole keskliikmete erisusi.

Demonstreerime Galtoni kvartiihälbe arvutamist paari lihtsa näitega.

1. näide. Oletagem, et on teada andmed tööviljakusplaani täitmise kohta 11 tehases (% plaanist), kusjuures tehased on järjestatud plaani täitmise tulemuste järgi.

Tehase järjek.nr.	1	2	3	4	5	6	7	8	9	10	11
Plaani täitmise %	97	103	104	104	105	111	112	113	114	118	120



Neil andmetel on Galtoni kvartiihälve

$$K^{\text{Galt.}} = \frac{114 - 104}{2} = 5 \%$$

Variatsioonreas arvutame Galtoni kvartiihälbe tabeli 28 andmetel, kus on kõrvutatud perekondade jaotumist suuruse järgi kahes vabariigis - Eestis ja Armeenias. See pakub erilist huvi seetõttu, et Eestis on perekonnad kõige väiksemad ja Armeenias kõige suuremad NSV Liidus.

Arvutame kõigepealt järjekorranumbrid:

Q_1 leidmiseks -

Eestis

Armeenias

$$\frac{308613}{4} \approx 77129;$$

$$\frac{349685}{4} \approx 87696;$$

Me leidmiseks -

$$\frac{308613}{2} \approx 154307;$$

$$\frac{349685}{2} \approx 174842;$$

Q_3 leidmiseks -

$$\frac{3 \cdot 308613}{4} \approx 231386; \quad \frac{3 \cdot 349685}{4} \approx 263089.$$

Et tegemist on momentreaga, siis piisab mediaani leidmiseks vastava järjekorranumbriga variandi ülesotsimisest. Kasutame selleks sageduste järgsummasid $\sum f_i$ (veerud 3 ja 5 tabelis 28).

Selgub, et Eesti NSV-s on

$$Q_1 = 2; \quad Me = 3; \quad Q_3 = 4$$

ja Armeenia NSV-s

$$Q_1 = 2; \quad Me = 5; \quad Q_3 = 6.$$

Neil andmetel on

$$K_{Eesti}^G = \frac{4 - 2}{2} = 1 \text{ inimene}$$

ja

$$K_{Arm.}^G = \frac{6 - 2}{2} = 2 \text{ inimest.}$$

Galtoni kvartiilhälvet on nimetatud teisiti ka veel "töenäoliseks hälbeks". Viimane nimetus on õigustatud selles mõttes, et vahemikku

$$Me \pm K^G,$$

s.o. Eestis 3 ± 1 ja Armeenias 5 ± 2 inimest, mahub tööpoolest üle poole kõigist võimalikest variantidest. Järelikult on töenäosus kohata just selliseid perekondi tunduvalt

suurem kui see, et kohatakse sellest vahemikust väiksemaid või suuremaid perekondi. Öeldut saab lihtsalt tõestada kasutatud arvnäite abil. Tabeli 29 andmetel mahutab nimetatud vahemik Eestis 88 ja Armeenias 78 % kõigist perekondadest.

T a b e l 28 .

Perekondade jaotumine liikmete arvu järgi Eestis ja Armeenias 1959. aastal.

Liikmete arv perekonnas	Eesti NSV		Armeenia NSV	
	Perekondade arv f_i	Järg- summa $\sum f_i$	Perekondade arv f_i	Järg- summa $\sum f_i$
(1)	(2)	(3)	(4)	(5)
2	121530	121530	50026	50026
3	92732	214262	58113	108139
4	57097	271359	63367	131506
5	23968		60647	232153
6	8759		49615	281768
7	2872		31908	
8	1034		18303	
9	389		9265	
10 ja rohkem	212		8441	
	308613		349685	

Galtoni kvartiilhälve on absoluutseid variatsiooninäitajaid. Suhtelise variatsiooninäitaja ehk variatsioonikoefitsiendi saame sellest, kui jagame kvartiilhälbe mediaaniga, s.o.

$$v^G = \frac{Q_3 - Q_1}{2 Me} \quad (38)$$

Nii selguks, et vastav näitav on

$$v^G_{\text{Eesti}} = \frac{4 - 2}{2 \cdot 3} = \frac{2}{6} = 0,33 \text{ ehk } 33 \%$$

ja

$$V_{Arm.}^G = \frac{6 - 2}{2 \cdot 5} = \frac{4}{10} = 0,40 \text{ ehk } 40 \%,$$

millest selgub, et ka suhteliselt varieerub perekondade suurus Armeenias tunduvalt rohkem kui Eestis. Põhimõtteliselt sama järelduse saab teha algandmete alusel ka n.-ö. silma järgi. Koefitsiendi väärtus seisneb selles, et tema abil saab varieerumises avalduvaid erisusi ka mõõta.

T a b e l 29 .

	Eesti NSV	Armeenia NSV
Perekondade üldarv	308613	349685
Me + K ^G liikmega perekondade arv	271359	273650
Vahemikku Me + K ^G kuuluvate perekondade osatähtsus nende üldarvust	0,88	0,78

9.5.3. A. Bowle'i diferentseerumiskoeffitsient.

A. Boele'i diferentseerumiskoeffitsient leitakse valemi-
ga

$$V^{Bow.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad . \quad (39)$$

Erinevalt Galtoni kvartiilhälbest on Bowle'i diferentseerumiskoeffitsient dimensioonitu suhtarv. Mingisugust variatsioonikoeffitsienti sellest järelikult enam arvutada pole vaja.

Tuginedes andmetele perekondade suurusest Eestis ja Armeenias (vt. tabel 28) saame Bowle'i diferentseerimiskoeffitsientideks

$$v_{\text{Eesti}}^{\text{Bow.}} = \frac{4 - 2}{4 + 2} = \frac{2}{6} = 0,33 \text{ ehk } 33 \%$$

ja

$$v_{\text{Arm.}}^{\text{Bow.}} = \frac{6 - 2}{6 + 2} = \frac{4}{8} = 0,50 \text{ ehk } 50 \%$$

Nagu eespool Galtoni kvartiilhälbest tuletatud v^G , nii ka $v^{\text{Bow.}}$ järgi selgub, et Armeenias varieerub perekonna suurus rohkem kui Eestis.

9.5.4. K. Pearsoni ja R. Fisheri variatsioonikoefitsiendid.

Tunduvalt komplitseerituma sisuga on mitmesuguste keskmomentide suhtena leitud variatsioonikoefitsiendid. Neist on tähtsamad:

a) Pearsoni variatsioonikoefitsient

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \beta_2 = \frac{\mu_3}{\mu_2^2},$$

kus μ_2 ja μ_3 on vastavalt teist või kolmandat järku keskmomendid.

b) Fisheri variatsioonikoefitsiendid

$$\gamma_1 = \sqrt{\beta_1} \quad \text{ja} \quad \gamma_2 = \beta_2 - 3,$$

kus β_1 ja β_2 on vastavad Pearsoni variatsioonikoefitsiendid.

10. KESKMISED JA VARIATSIOONRIDADE KUJU.

10.1. Sümmeetrilised ja ebasümmeetrilised read.

Variatsiooniridade kujus avalduvad nähtuse varieerumist iseloomustavad seaduspärasused. Kui uuritakse küllalt suuri kogumeid ja kui nähtuse väärtus sõltub täiesti juhuslikest

asjaoludest, saadakse nn. normaalne jaotuskõver. See on sümmeetriline kõver, mida tema kuju järgi on nimetatud ka kella-kõveraks või esmakordsete uurijate järgi Gaussi-Laplace'i kõveraks. Normaalsest jaotuskõverast ükskõik kummale poole hõlbivad kõverad on ebasümmeetrilised. Suuna järgi eristatakse parem- ja vasakpoolset ebasümmeetrilisust (vt. joon. 11). Suuruse järgi tehakse vahet nn. mõõduka ja täieliku ebasümmeetrilisuse vahel.

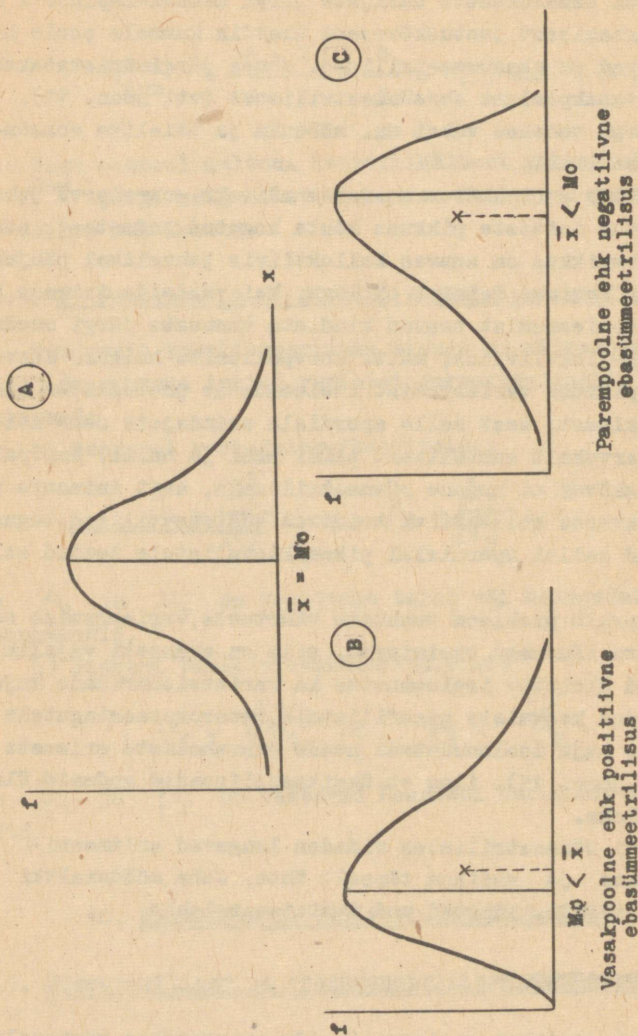
Normaalne jaotuskõver saadakse näiteks suure arvu juhuslikult võetud inimeste pikkuse kohta kogutud andmetest, sest inimese kehapikkus on suures kollektiivis juhuslikel põhjustel kujunev suurus. Teisiti on lugu, kui vaadelda inimese kehapikkuse varieerumist teatud kindlate tunnuste järgi moodustatud inimkollektiivides, näit. korvpallurite hulgas. Korvpallurite pikkuse varieerumist iseloomustab jaotuskõver hõlbiks normaalsest, sest selle spordiala esindajate seas leidub eriti arvukalt suhteliselt pikki mehi ja naisi. Korvpallurite jaotuskõver ei kujune sümmeetriliseks, sest inimeste valikul vastavasse kollektiivi toimivad süsteematilised tegurid, mis annavad sellel spordialal pikemakärvulistele teatud eeliseid.

Kui kerkib probleem tunnuste väärtuste varieerumise seaduspärasuste lähemast uurimisest, siis on enamasti vajalik anda teatud üldistav iseloomustus ka variatsioonridade kujule. Tuginedes keskmiste graafilistele interpreteeringutele võib ridade kuju iseloomustada nende omavaheliste erisuste kaudu (vt. joon. 11), ilma et üksikasjalisemaid andmeid üldse teada oleks.

Täiesti sümmeetrilistes ridades langevad aritmeetiline keskmine ja mediaan täpselt ühte. Juba mõõdukaltki ebasümmeetrilistes lähevad nad üksteisest lahku.

10. 2. Asümmeetria.

Ebasümmeetrilisi variatsioonridu eristatakse vastavalt sellele, kummale poole kaldub nende maksimum. Iseloomustab rida suhteliselt väiksema väärtusega variantide rohkus, mis



Joon. 11. Variatsioonrea kuju kirjeldamine keskmiiste kaudu. Sümeetrilises reas on kõik kolm keskmist võrdsed (A), ebahümeetrilistes kalduvad asendikeskmised aritmeetilisele kas vasakule (B) või paremale (C), mispuhul on tegemist vastavalt nimetatava ebahümeetrilisusega.

avaldub sageduste kiires kasvus ja sellele järgnevas aeglasel kahanemises, siis nimetatakse rida vasakule poole ebasümmeetriliseks ehk vasakkaldeliseks. Paremale poole ebasümmeetrilises ehk paremkaldelises reas on vastupidi. Variantide sagedused tõusevad seal aeglaselt, langevad aga suhteliselt järsult. (Vt. joon. 11)

Vasakkaldelistes ridades on keskmistest kõige väiksem mood, sest maksimum on vasakule "nihutatud", kus rea liikmete väärtused on väiksemad. Sellises reas valitseb keskmiste vahel seos

$$Mo < \bar{x} < Me,$$

mille kõiki üksikvõrratusi võib pidada mistahes rea vasakkaldelisuse tunnusteks.

$$Mo < \bar{x}; \quad \bar{x} < Me; \quad Mo < Me.$$

Paremkaldelisuse puhul on seosed vastupidised, s.o.

$$Mo > \bar{x} > Me$$

ja võrratusi

$$Mo > \bar{x}; \quad \bar{x} > Me; \quad Mo > Me$$

voib pidada rea paremkaldelisuse tunnusteks.

Ridade asümmeetrilisuse mõõduks võib kasutada näiteks aritmeetilise keskmise ja moodi vahet

$$\boxed{A = \bar{x} - Mo} \quad . \quad (46)$$

On see hálve plussmärgiga, siis on rida positiivselt s.o. vasakule poole ebasümmeetriline, kui miinusmärgiga, siis negatiivselt, s.o. paremale poole ebasümmeetriline.

Eespool on näiteks perekondade liikmete arvu järgi jaotumise kohta teada, et Eestis on $\bar{x} = 3,1$ ja $Mo = 2$, Armeenias aga $\bar{x} = 4,8$ ja $Mo = 4$. Nii saame

$$A_{\text{Eesti}} = 3,1 - 2 = 1,1 \text{ inimest}$$

$$A_{\text{Arm}} = 4,8 - 4 = 0,8 \text{ inimest ,}$$

millest selgub, et perekendade jaotumist iseloomustavad variatsioonread on positiivselt asümmeetrilised ja Eestis on vastav rida tunduvalt ebasümmeetrilisem kui Armeenias.

10.2.1. K. Pearsoni asümmeetriakordaja.

Mida ebasümmeetrilisem on rida, seda enam lähuvad aritmeetiline keskmine ja mood teineteisest lahku. Teiselt poolt on ilmne, et mida laugjam on jaotuskõver (s.o. mida suurem on tema standardhälve), seda vähem rea ebasümmeetrilisus mõjule pääseb. Kujult hästi püstakas variatsioonreas, s.o. reas, kus standardhälve on väike, mõjub moodi ja aritmeetilise keskmise teineteisest kõrvalekalduumine seevastu rea sümmeetrilisusele väga silmatorkavalt.

Nende tähelepanekute alusel konstrueeriski K. Pearson järgmise asümmeetria näitartvu

$$A^P = \frac{\bar{x} - Mo}{\sigma} , \quad (41)$$

kus asümmeetrilisus on seatud hälbest $(\bar{x} - Mo)$ võrdelisse ja standardhälbest pöördvõrdelisse olenevasse.

Kui rida on sümmeetriline, s.o.

$$\bar{x} = Mo ,$$

siis on Pearsoni asümmeetriakordaja $A^P = 0$.

On rida vasakule asümmeetriline, siis on $\bar{x} < Mo$ ja $A^P < 0$, s.o. negatiivne arv. Parempoolse asümmeetria korral on $\bar{x} > Mo$ ja ka $A^P > 0$, s.o. positiivne arv. Seepärast nimetataksegi vasakpoolset asümmeetriat ühtlasi negatiivseks ja parempoolset positiivseks asümmeetriaks.

Tabel 11 andmeil, kus $\bar{x} = 113$, $Mo = 115,45$ ning $\sigma = 15,84$ on

$$A^P = \frac{113 - 115,45}{15,84} = \frac{-2,45}{15,84} = -0,155.$$

Järelikult on siin tegemist negatiivse asümmeetriaga.

Tabelis 1 toodud näites on $\bar{x} = 78,38$; $M_0 = 80,6$;

$\sigma = 30,99$. Seega on

$$A^P = \frac{78,4 - 80,6}{30,99} = \frac{-2,2}{30,99} = -0,0715,$$

millest selgub, et siingi on tegemist samasuunalise asümmeetriaga; siiski on selle aste märksa väiksem. See tõestub ka kõverate graafilisel võrdlemisel.

10.22. W. Johannseni asümmeetriakordaja.

W. Johannseni asümmeetriakordaja on kolmandat järku keskmomendi ja kuupi tõstetud standardhälbe suhe, s.o.

$$A^J = \frac{\mu_3}{\sigma^3} \quad (42)$$

ehk teisiti avaldatult

$$A^J = \frac{\sum (x_i - \bar{x})^3 f}{\sigma^3 \sum f}.$$

Tabel 11 andmetel (vt. tabel 30) on

$$\mu_3 = \frac{26280}{20} = 1314;$$

$$\sigma^3 = 15,84^3 = 3975,84$$

ning

$$A^J = \frac{1314}{3976} = 0,3305.$$

T a b e l 30 .

x_1	f_1	$(x_1 - \bar{x})$	$(x_1 - \bar{x})^3$	$(x_1 - \bar{x})^3 f_1$
90	4	-23	-12167	-48668
110	10	- 3	- 27	- 270
130	5	17	4913	24565
150	1	37	50653	50653
	20			26280

Kui rida on täiesti sümmeetriline, siis on $A^J = 1$.

10.3. Ekstsess.

Ekstsessiks nimetatakse variatsioonkõvera suhtelist püstakust, kusjuures võrdlusbaasiks on normaalse jaotuskõvera kuju. Ekstsessiivseteks ehk positiivse ekstsessiga kõveraks nimetatakse selliseid, mis on normaalsest tunduvalt püstakamad. Laugjamaid käsitatakse negatiivse ekstsessiga kõveraina.

Normaalse jaotusrea andmeil on neljandat järku keskmomendi (μ_4) väärtus $3/4$ ja teist järku keskmomendi (μ_2) väärtus $1/2$. Järelikult on neljandat järku keskmomendi ja teist järku keskmomendi ruudu suhe alati konstantne

$$\frac{\mu_4}{\mu_2^2} = \frac{3/4}{1/4} = 3 .$$

Ekstsessi näitavaks E võetakse uuritava konkreetse rea andmeil arvatatud neljandat ja teist järku keskmomentide suhte ja arvu 3 vahe, s.o.

$$E = \frac{\mu_4}{\mu_2^2} - 3 . \quad (43)$$

On uuritav rida oma kujult sarnane normaalse jaotuskõveraga, saadakse

$$E = 3 - 3 = 0.$$

Normaalsest püstakamates ridades on $E > 0$ ja laugjamates vastupidi: $E < 0$.

Tabelis 11 toodud näite andmetel (arvutused vt. tabel 31) on

$$\mu_4 = \frac{3411940}{20} = 170597 ;$$

$$\mu_2^2 = (G^2)^2 = 251^2 = 63001$$

ning

$$E = \frac{170597}{63001} - 3 = 2,708 - 3 = -0,292 .$$

See tähendab, et näiterida on normaalsest jaotuskõverast lamedam.

10.4. Teoreetilise variatsioonikõvera konstrueerimine empiirilistel andmetel.

Mistahes empiirilise variatsioonirea kuju uurimisel, tema asümmeetria ja ekstsessi leidmisel tuginetakse variantide normaalsele jaotumusele kui etaloonile, millega kõiki konkreetseid andmeid võrreldakse. Juba tavalistegi variatsiooninäitarvude leidmisel on normaalne jaotuskõver aluseks, millele ridade iseloomustus rajatakse. Standardhälve vastab oma geomeetrilise mõtte poolest normaalse jaotusrea käänupunktile. Arvutades mingi konkreetse nähtuse varieerumise iseloomustamiseks standardhälbe, vaatleme selle varieerumist sisuliselt niisugusel eeldusel, nagu toimuks see normaalse jaotumisseaduse alusel, ehkki tegelikkuses juhtub selliste nähtustega kokkupuuteid harva.

Nii selgub, et majanduslike nähtuste varieerumises avalduvate seaduspärasuste uurimine tugineb võrdlemisi ulatus-

likult empiiriliste jaotumuste ja vastava mahuga normaalse jaotumuse võrdlemisele. Suurt teoreetilist huvi pakub mitte ainult tegelikke kogumeid iseloomustavate elementide (keskmiste, standardhälvete jne.) võrdlemine vastavate teoreetiliste elementidega, vaid ka nende terviklik võrdlemine.

Kõige lihtsam moodus selleks on võrrelda empiirilisi ja teoreetilisi jaotumusi graafiliselt. Praktiliselt taandub see jaotuskõvera konstrueerimisel mingi tegelikkuses esineva nähtuse kohta kogutud empiiriliste andmete alusel ning selle võrdlemisele teoreetilise variatsioonkõveraga. Empiirilise kogumi asümmeetrilisuse aste ja suund ("+" või "-") ja eks-
tsess ilmnevad sel puhul eriti näitlikult.

Empiiriliste kogumite ja teoreetiliste jaotumuste võrdlusele saab rajada tihtipeale õige kaugele küündivaid järeldusi. Missugusteks need järeldused millalgi kujunevad, see sõltub olulisel määral nähtuste konkreetsest iseloomust, mitmesugustest ühiskondlikest teguritest jms., mida tuleb igal üksikjuhul spetsiaalselt uurida. Enamasti võib aga arutlusel tugineda ligikaudu järgmisele skeemile. Nagu teada, kehtib normaalne jaotusseadus niisuguste nähtuste suhtes, mille konkreetset üksikväärtused sõltuvad laias laastus vaadates peaaesjalikult juhuslikest asjaoludest. Kui mingi nähtuse varieerumine kaldub normaalsest jaotuskõverast oluliselt kõrvale, siis on see järelikult tingitud mingisugustest süstemaatilistelt toimivatest teguritest. Missugused need tegurid on, kuidas neid kvalifitseerida jne., seda tuleb lahendada iga konkreetse uurimuse raames vastavate spetsiaalmeetoditega. Ridade kuju uurimine aga võimaldab selliste süstemaatilistelt toimivate tegurite mõju esile tuua ja seda mõõta: mida enam empiiriline rida normaalsest hälbib, seda suurem on süstemaatilistelt toimivate tegurite mõju.

Ridade kuju uurimisel on keskse tähtsusega põhimõte, et iga tegelikkuses esineva nähtuse kohta normaalses oludes kogutud andmed annavad sellele nähtusele tüüpilise variatsioonkõvera. Viimane võib olla kas sümmeetriline, nagu Gaussi-Laplace'i kõver, või ka ebasümmeetriline. Kui tegelike andme-

te alusel konstrueeritud variatsioonkõver kaldub sellest tüüpilisest kõverast kõrvale, siis tähendab see, et nähtuse kujunemist on hakanud mõjustama mingisugused uued asjaolud. Kontrollides niiviisi näiteks automaattööpingil valmistatud detailide tähtsamate mõõtmete järgi jaotumist, võib teha tehnoloogilise protsessi kulgemise kohta õige tähtsaid järeldusi, saada õigeaegselt andmeid toodangu kvaliteedi muutumisest ja vältida praaktodete valmistamist.

Empiirilistel andmetel teoreetiliste jaotuskõverate konstrueerimisel tuginetakse funktsiooni

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

väärtuste tabelile (vt. lisa 1).

Praktiliselt läbitakse järgmised tööetapid:

1) leitakse variantide individuaalväärtuste x_1 ja nende aritmeetilise keskmise $\bar{x}$ vahelised hälbed. Vastavad arvutused 2069 tähtsama ehitusorganisatsiooni plaani täitmise kohta (lähteandmed on võetud tabelist 1) eeldusel et $\bar{x} = 78,4$, on arvutustabelis 31 veerus (3);

2) leitakse normhälbed t_1 valemiga

$$t_1 = \frac{x_1 - \bar{x}}{\sigma} .$$

Normhälvete väärtused on 31 tabeli veerus (4). Arvestatud on seejuures, et $\sigma = 30,99$. Normhälvete arvutamisel peetakse silmas seda, kui täpsed jaotusfunktsiooni väärtuste tabelleid on kasutada. Käesoleval juhul leitakse t_1 väärtused sajandiku täpsusega, sest sellise täpsusega on ka tabel lisas 2;

3) leitakse normhälvete t_1 väärtustele vastavad normaalse jaotusfunktsiooni väärtused $\psi(t)$ lisas 1 toodud tabelist. Vastavad andmed on 31. tabeli veerus (5);

4) arvutatakse teoreetiliste sageduste leidmiseks vajalik konstant

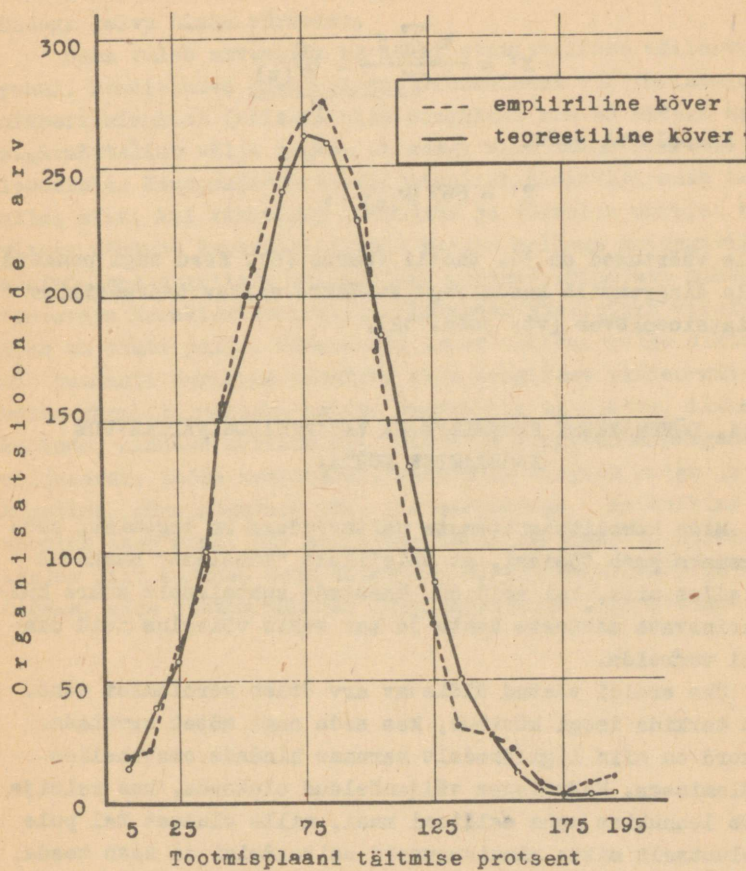
Tabel 31.

x_i	f_i	$x_i - \bar{x}$	$t = \frac{x_i - \bar{x}}{\sigma}$	$\varphi(t)$	f'
(1)	(2)	(3)	(4)	(5)	(6)
5	20	-73,4	-2,37	0,02406	16,1
15	23	-63,4	-2,05	0,04879	32,6
25	66	-53,4	-1,72	0,09089	60,7
35	95	-43,4	-1,40	0,14973	99,96
45	153	-33,4	-1,08	0,22265	148,6
55	217	-23,4	-0,76	0,29887	199,5
65	254	-13,4	-0,43	0,36371	242,8
75	273	- 3,4	-0,11	0,39654	264,7
85	282	6,6	0,21	0,39024	260,5
95	232	16,6	0,54	0,34482	230,2
105	178	26,6	0,86	0,27562	184,0
115	100	36,6	1,18	0,19886	132,8
125	74	46,6	1,50	0,12952	86,5
135	32	56,6	1,83	0,07477	49,9
145	27	66,6	2,15	0,03955	26,4
155	20	76,6	2,47	0,01888	12,6
165	7	86,6	2,79	0,00814	5,4
175	3	96,6	3,12	0,00307	2,0
185	5	106,6	3,44	0,00107	0,7
195	8	116,6	3,74	0,00037	0,2
	2069				

$$\frac{k \sum f}{\sigma}$$

mis on vaadeldavas näites

$$\frac{10 \cdot 2069}{30,99} = 667 ;$$



Joon. 12. Eritusorganisatsioonide jaotumine
1959. aasta 9 kuu plaani täitmise järgi.

5) leitakse vaadeldava nähtuse empiirilistele väärtustele vastavad teoreetilised sagedused f'

$$f' = \frac{k \sum f}{\sigma} \cdot \phi(t)$$

ehk käesoleval juhul

$$f' = 667 \phi(t) ,$$

mille väärtused on 31. tabeli veerus (6). Need ongi punktid, mille diagrammile kandmisega saadakse vastav teoreetiline variatsioonkõver (vt. joon. 12).

11. LÕPPMÄRKUSI KESKMISTE JA VARIATSIOONINÄITARVUDE KASUTAMISE KOHTA.

Mida komplitseeritumate näitarvudega on tegemist, seda selgemaks saab tõsiasi, et täielikult "kõnelema" hakkavad nad alles siis, kui neid on kasutada suhteliselt suure hulga erinevate nähtuste kohta ja kui tekib võimalus neid omavahel võrrelda.

Üks eraldi võetud üldistav arv ütleb võrdlemisi vähe. Võib tekkida isegi küsimus, kas seda ongi mõtet arvutada. Olukord on siin ligilähedalt sarnane hindade omavahelise võrdlemisega. Kujutlegem väljamõeldud olukorda, kus reisija astub lennukist maha sellisel maal, mille oludest tal pole absoluutselt mitte mingisuguseid eelteadmisi ja saab teada, et 1 kg leiva eest nõutakse sellisel maal 15 rahaühikut. On leib siin odav või kallis? Sellele ei suudaks ta vastata, ehkki leiva hind on teada, sest mingi ühe kauba hind jääb majandusliku näitarvuna absoluutselt sisutuks, kui seda ei saa võrrelda teiste kaupade hinnaga. Et sellele küsimusele vastata, tuleks hankida täiendavaid andmeid teiste kaupade - liha, piima, või, rõivastusesemete jne. - hindade ja rahva sissetuleku kohta. Alles seejärel võib ütelda, kas

leib on odav või kallis. Tähelepandav on, et järeldused võivad kujuneda seda sisukamaks, mida suurema arvu muude hindadega leiva hinda võrrelda.

Sama tuleb arvestada ka kõigi statistiliste näitarvude puhul. Ponnistused ühest ainsast keskmisest või variatsioonikoefitsiendist (näiteks nisu saagikuse kohta) midagi eriti sügavmõttelist välja lugeda ei saagi viia kuigi suurtele tulemustele. Kaugemaleküündivaid sisulisi üldistusi saab teha alles siis, kui käsitlust laiendada ja võrrelda uurijat huvitava nähtuse karakteristikuid teiste sellega kuidagiviisi seotud, analoogiliste või millegi poolest erinevate nähtuste vastavate karakteristikutega. Võimalusi taolisteks võrdlusteks on alati palju. Missugused neist valida, selle dikteerib enamasti uurimise eesmärk. Nisu saagikuse varieerumist võiks võrrelda näiteks teiste kultuuride saagikuse, ilmastikuolude, pinnase kvaliteedi, külvi- ja koristustähtaegade, viljasordi, tööde kvaliteedi, kasutatud väetise hulga ja koostise, künnisügavuse jms. varieerumisega. Iga selline võrdlus avab mingi uue külje uuritavas nähtuses ja objektist saadav pilt kujuneb seda mitmekülsemaks, sügavamaks ja tihedamaks, mida rohkem nähtuse külgi käsitluse alla võetakse.

Funktsiooni $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ väärtused

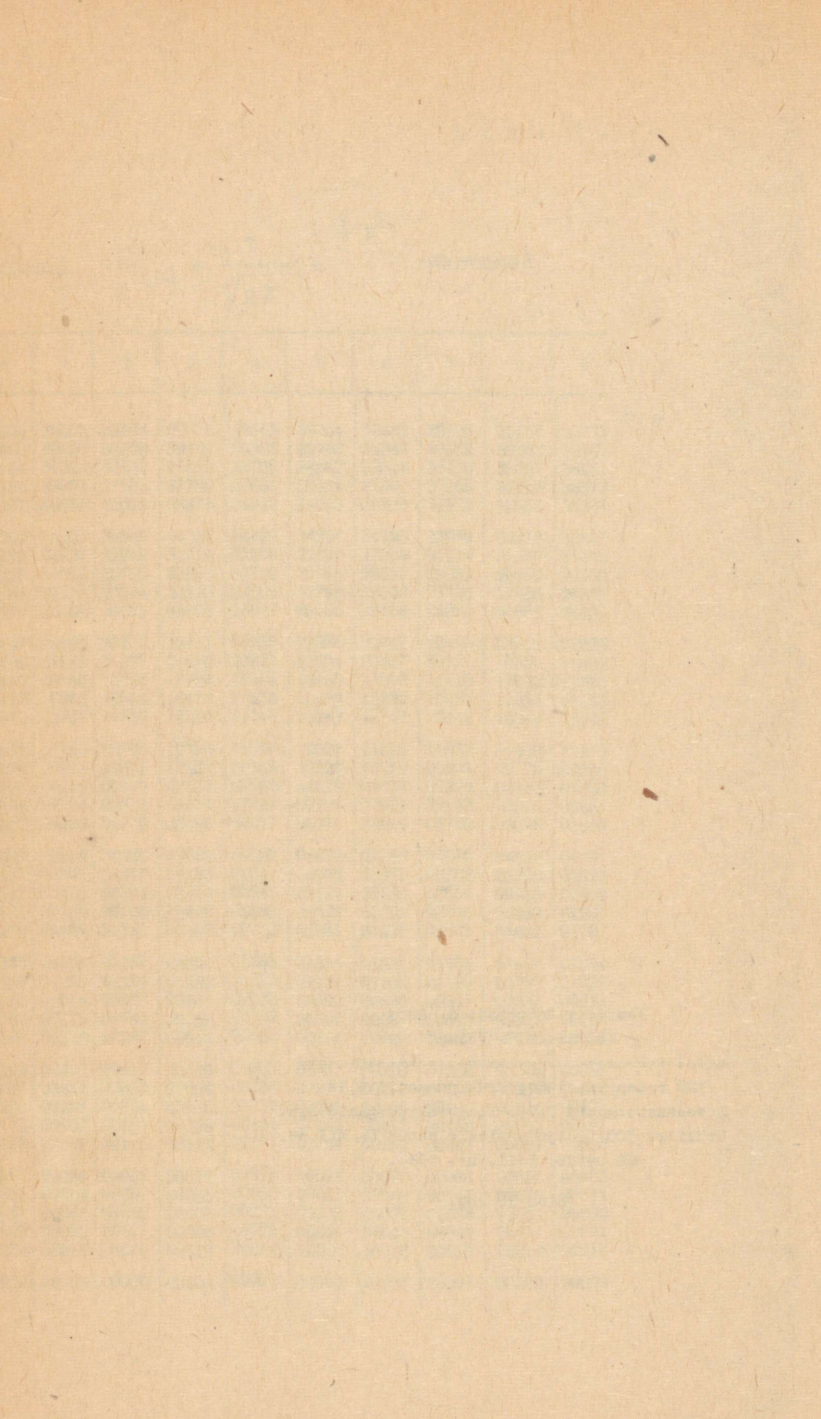
t	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29430	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	9893	9728	9566
1,7	09405	09246	09089	08938	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
4,0	00013	00009	00006	00004	00002	00002	00001	00001	00000	00000

Vastutav toimetaja S. Aron
Korrektor E. Võhandu

=====

TRÜ rotaprint 1964. Trükipoognaid 9,31.
Arvestuspoognaid 7,19. Tingtrükipoognaid 8,5.
Trükiarv 500. Paljundamisele antud 16. XII 64.
MB 09386. Tell. nr. 1634.

Hind 22 kop.



Hind 22 kop.

A

26405

190298

TÜ RAAMATUKOGU



1 0300 00377803 4