TARTU RIIKLIK ÜLIKOOL

K. SOONETS

# TÕENÄOSUSTEOORIA JA MATEMAATILINE STATISTIKA

TARTU 1968

19885

T A R T U   R I I K L I K   Ü L I K O O L
Teoreetilise mehhaanika kateeder

K. Soonets

# T Õ E N Ä O S U S T E O O R I A
# J A   M A T E M A A T I L I N E   S T A T I S T I K A

(Majandusteaduskonna üliõpilastele)

Teine trükk

Tartu 1968

Saateks.

Konspekt sisaldab materjali tõenäosusteooria ja matemaatilise statistika kursuse ulatuses majandusteaduskonnale.

Tõenäosusteooria teoreetilisi küsimusi ja rakendusi matemaatilises statistikas käsitletakse paralleelselt. Keerulisi matemaatilisi tõestusi on välditud. Suurt rõhku on pandud matemaatiliste arutluste ja valemite sõnalisele selgitamisele. See on eeskätt vajalik neile, kelle ettevalmistus matemaatikas ei ole eriti tugev. Paragrahvides 2 ja 3 käsitletud üldisi mõisteid ja arutlusi rakendatakse ning konkretiseeritakse järgmistes paragrahvides. Sellepärast on väga oluline 2. ja 3. paragrahvi sisu põhjalik mõistmine.

Konspekt on varustatud ülesannetega iseseisvaks lahendamiseks. Teoreetilise materjali kinnistamiseks on konspektis mitmel pool jäetud üksikküsimusi iseseisvaks lahendamiseks. Nendele küsimustele vastuse leidmise käigus saab ise kontrollida, kuidas on aru saadud eelnevatest punktidest.

Konspekt on mõeldud kasutamiseks esmajoones majandusteaduskonna üliõpilastele, kuid ta sobib tõenäosusteooria ja matemaatilise statistika põhimõistetega tutvumiseks ka teiste teaduskondade üliõpilastele ( mittematemaatikutele ).

Autor.

## Tõenäosusteooria ja matemaatilise statistika aine.

Igal elualal tuleb massiliselt kokku puutuda nähtustega, mis ka samades tingimustes on teatud määral erinevad. Näiteks ühel katselapil kasvavates viljapeades on terade arv erinev; võrdse liikmete arvuga perekondades ei tarbita piima samal hulgal; tekstiilikäitises töötavate samatüübiliste automaattelgede seisakute arv ühe tunni jooksul on erinev; täringuga igal viskel saadav silmade arv on muutuv. Nähtuste esinemist mõjutavad peamised tegurid on samad, kuid teisejärgulised (kõrvalised) tegurid pole iga samatüübilise nähtuse korral täpselt samad ning need tingivadki erinevusi nähtuste esinemises.

Harilikult võetakse arvesse ainult nähtust oluliselt mõjutavad tegurid, ignoreerides täiesti kõrvalisi tegureid. Selliselt toimides asendame reaalse nähtuse uurimise tema lihtsustatud mudeli uurimisega. Kui aga paljude teisejärguliste tegurite koosmõju oluliselt mõjutab nähtust, siis ei tohi neid tegureid arvestamata jätta. Sageli aga ei ole võimalik täpselt hinnata kõiki nähtust mõjutavaid kõrvalisi tegureid ning iga üksikteguri mõju. Neil juhtudel räägitakse juhusest sõltuvatest nähtustest.

Juhusest sõltuvaks ehk juhuslikuks nimetatakse nähtust, mis esineb paljude kõrvaliste põhjuste koosmõjul ühe ja sama katse kordamisel teatud erinevustega.

Tõenäosusteooria uurimisobjektiks on juhuslikud nähtused. Tõenäosusteooria selgitab ja kirjeldab juhuslike nähtuste massilisel toimumisel kehtivaid seaduspärasusi matemaatiliste meetoditega. Tõenäosusteooria abil saame teha õigeid järeldusi massiliste katsete tulemustest, sest iga üksikkatse tulemus võib sisaldada palju juhuslikke, vaadeldavale nähtuste liigile üldiselt mitteomaseid jooni.

Inimkond on jõudnud oma arengus sellisesse staadiumi, kus reaalses maailmas valitsevate seaduspärasuste uurimiseks ja kirjeldamiseks tuleb uuritavale objektile vastavusse seatud mudelit täpsustada, s.t. veel rohkem lähendada tõelisele objektile. Sellisel juhul tuleb aga seaduspärasuste selgitamiseks kasutada ka tõenäosusteooria meetodeid. Tõenäosusteooriat rakendataksegi erinevates teadusharudes järjest rohkem.

Tõenäosusteoorias loodud mõisted ja nähtuste uurimise meetodid võimaldavad kokku hoida aega ja vahendeid katsete sooritamiseks, kuid ei võimalda neist täiesti loobuda. Juhuslike nähtuste uurimine tugineb lõppkokkuvõttes siiski katsele ja vaatlusele.

Massiliste juhuslike nähtuste vaatluse tulemuste registreerimise, kirjeldamise ja analüüsi üldiste meetodite loomine ning arendamine on matemaatilise statistika aineks. Matemaatiline statistika tegeleb eri liiki massilistele nähtustele ühiste tunnuste ja omaduste uurimisega. Ta jätab kõrvale uuritavate objektide spetsiifilised iseärasused ja käsitleb statistiliste uurimismeetodite formaalset, matemaatilist külge. Matemaatilise statistika meetodeid saab kergesti kohandada väga erinevate konkreetsete nähtuste uurimiseks.

Matemaatiline statistika on lahutamatult seotud tõenäosusteooriaga. Matemaatilises statistikas vaadeldavad spetsiifilistest tunnustest "puhastatud" nähtused alluvad tõenäosusteooria seadustele. Tõenäosusteooria seadused on massiliste juhuslike nähtuste seaduspärasuste matemaatiliseks väljenduseks. Tõenäosusteoorial on matemaatilises statistikas umbes sama osa mis geomeetrial mõõdistamises või joonestamises.

## §1. TÕENÄOSUSTEOORIA ALGMÕISTED.

### 1. S ü n d m u s e d.

Esimeseks põhiliseks mõisteks tõenäosusteoorias on sündmus. Sündmuseks peetakse kõike seda, mille kohta saab rääkida toimumisest ja mittetoimumisest.

Sündmus saab toimuda teatud tingimustes. Kui jälgime mingi sündmuse toimumist või mittetoimumist, räägime katse sooritamisest.

Näiteid:

1. Laskur tulistab märklauda. Katseks on lasu sooritamine, märklaua teatud tsooni tabamine on sündmuseks.
2. Urnis on mitut värvi kuulid. Kuuli võtmine urnist on katse, teatud värvi kuuli saamine on sündmus.
3. Jalatsivabrikus konveieril valmistatud jalanõud liigitame kolme kategooriasse: esimene ja teine sort ning praak. Võtame kontrollimiseks juhuslikult ühe paari - see on katse. Katse tulemuseks on üks sündmustest: esimese sordi, teise sordi või praakpaari saamine.

Sama katse korduval sooritamisel räägime katseseeriast. Iga katseseeriasse kuuluvat katset nimetame üksikkatseks.

Sündmusi liigitame nende toimumise võimalikkuse seisukohalt kindlateks, võimatuteks ja juhuslikeks.

Kui katsel teatud sündmuse toimumine on vältimatu, nimetame teda kindlaks sündmuseks. Kui katsel sündmus toimuda ei saa, nimetame teda võimatuks sündmuseks.

Sündmust nimetame juhuslikuks, kui ta antud katsel võib toimuda või mitte toimuda. Juhusliku sündmuse toimumist või mittetoimumist ei saa garanteerida, teda iseloomustab toimumise võimalikkus.

Näiteid:

1. Olgu urnis kõik kuulid rohelised. Kuuli võtmisel urnist

on rohelise kuuli saamine kindel sündmus, punase kuuli saamine on võimatu sündmus.

2. Viskame kaht täringut korraga. Ühe või teise silmade kombinatsiooni esiletulek on juhuslik sündmus.
3. Kauplus müüb erinevatel päevadel kaupa erinevalt. Päevase läbimüügi suurus on juhuslik sündmus.
4. Automaattööpingil valmistatud detaili kvaliteetseks või praagiks osutumine on juhuslikud sündmused, kuigi esimene sündmus toimub võrratult sagedamini kui teine.
5. Päikese tõusmine läänest on võimatu sündmus.
6. Perekonnas poisslapse sündimine on juhuslik sündmus.
7. Loteriil võitmine on juhuslik sündmus.

Sündmuse juhuslikkus ei ole vastuolus dialektilise materialismi seisukohtadega. Sündmust loeme juhuslikuks selles mõttes, et me kas ei tea kõiki sündmuse toimumise põhjusi või ei suuda nende mõju täpselt arvestada.

Sündmuste tähistamiseks kasutame edaspidi suuri tähti. Nii võime konveierilt tulevate kingapaaride kontrollimisel esimese sordi paari saamisel rääkida sündmusest A, teise sordi paari saamisel sündmusest B ja praakpaari saamisel sündmusest C.

## 2. Sündmuste liigitamisest.

A. Kaht sündmust nimetame teineteist välistavateks, kui ühe sündmuse toimumine antud katsel välistab teise sündmuse toimumise samal katsel.

Näiteid:

1. Vapi ja kirja esiletulek ühel raha viskel on teineteist välistavad sündmused. Visates raha kaks korda järjest võib pärast vapi esiletulekut esimesel viskel tulla teisel viskel kiri. Sellise katseskeemi korral on vapi ja kirja esiletulek mittevälistavad sündmused.
2. Ühe loteriipiletiga samaaegselt 40 rbl. ja 100 rbl. võitmine on välistavad sündmused.

3. Liigitame tööpingil valmistatud detailid I sordi, II sordi ja praakdetailideks. Sama detaili kuulumine I sordi ja praagi hulka on välistavad sündmused. Kui loeme I ja II sordi detailid standardseteks ja praakdetailid mittestandardseteks, siis detaili kuulumine I sorti ja standardsete hulka ei ole enam välistavad. Detaili kuulumine standardsete hulka ei välista võimalust kuuluda samaaegselt ka I sorti.

B. Sündmusi nimetame võrdvõimalikeks, kui võib eeldada, et antud katsel on iga sündmuse toimumise objektiivsed võimalused samad.

Näiteid:

1. Lugedes täringu viskel erineva silmade arvu saamist omaette sündmusteks, võib toimuda üks kuuest sündmusest. 1 kuni 6 silma saamine on võrdvõimalikud sündmused, sest korrapärase täringu korral on loomulik eeldada, et iga tahu pealejäämiseks on võimalused samad.
2. Olgu urnis 6 valget ja 4 musta kuuli. Kirjutame kuulidele järjekorranumbrid ühest kümneni. Kuuli võtmisel ei ole valge ja musta kuuli saamine võrdvõimalikud sündmused. Järjekorranubritega 1, 2, 3 jne. kuuli saamine on võrdvõimalikud sündmused.
3. Televiisori ostmisel kvaliteetse aparaadi ja praakaparaadi saamine ei ole võrdvõimalikud sündmused - esimese sündmuse toimumine on palju oodatavam.
4. Iga loteriipiletiga võitmise võimalused on samad - sündmused on võrdvõimalikud.

C. Sündmusi nimetame ainuvõimalikeks, kui antud katsel neist kindlasti üks toimub.

Näiteid:

1. Täringu viskel 1 - 6 silma saamine on ainuvõimalikud sündmused.
2. Valgeid ja musti kuule sisaldavast urnist kas valge või musta kuuli saamine on ainuvõimalikud sündmused.

3. Tööpingil valmistatud detail saab kuuluda kas I sorti, II sorti või praagi hulka - need kolm sündmust on ainuvõimalikud.
4. Märklaua tulistamisel 1 - 10 silma saamine pole ainuvõimalikud sündmused. Kuul võib märklauast hoopis mööda minna.

D. Ainuvõimalikud üksteist välistavad sündmused moodustavad täieliku sündmuste süsteemi. Üks täielikku sündmuste süsteemi kuuluvatest sündmustest toimub kindlasti.

Näiteid:

1. Raha viskel vapi või kirja tulek moodustavad täieliku sündmuste süsteemi.
2. Täringu viskel kas 1, 2 jne. kuni 6 silma saamine moodustavad täieliku sündmuste süsteemi.
3. Alajaotuses C toodud näites 4 pole tegemist täieliku sündmuste süsteemiga. Loetletud 10 sündmust on küll üksteist välistavad, kuid pole ainuvõimalikud - üks täieliku sündmuste süsteemi moodustamiseks vajalikest nõuetest pole täidetud.

Täieliku sündmuste süsteemi moodustavaid üksiksündmusi nimetame ka võimalusteks ehk juhtudeks. Täringu viskel esineb kuus võimalust; kui 3 silma saamine on sündmuseks A, siis A toimumiseks on üks soodne juht.

E. Katsel võib sündmus A toimuda või mitte toimuda. A mittetoimumist loeme uueks sündmuseks B. Pole oluline, missugune konkreetne sündmus toimub - kui ta pole sündmus A, siis loeme teda sündmuseks B. Üks neist kahest sündmusest toimub kindlasti: nad on ainuvõimalikud ja teineteist välistavad.

Kaht teineteist välistavat ainuvõimalikku sündmust nimetame vastandsündmusteks.

Sündmuse A vastandsündmuse tähistamiseks kasutame sümbolit $\bar{A}$. Sündmused A ja $\bar{A}$ moodustavad täieliku sündmuste süsteemi. Katsel toimub kindlasti üks neist, kolmandat võimalust ei ole.

Näiteid:

1. Vapi ja kirja esiletulek raha viskel on vastandsündmused.
2. Olgu urnis mustad, valged ja punased kuulid. Valge kuuli saamine ja valge kuuli mittesaamine, s.t. kas musta või punase kuuli saamine on vastandsündmused. Kuid valge kuuli saamine pole vastandsündmuseks musta kuuli saamisele, sest on veel kolmanda sündmuse toimumise võimalus - punase kuuli saamine.

## 3. Sündmuse tõenäosuse mõiste.

<u>Tõenäosuse mõiste.</u> Tõenäosusteoorias vaadeldakse sündmusi nende toimumise või mittetoimumise võimalikkuse seisukohalt. Antud katsel võib ühe sündmuse toimumine olla rohkem või vähem võimalik kui mõne teise sündmuse toimumine. Sündmuse toimumise objektiivse võimalikkuse näitajaks valitakse teatud arvuline suurus, mida nimetatakse sündmuse tõenäosuseks. Selgitame tõenäosuse mõistet kahe lihtsa näite varal.

Olgu urnis 7 musta ja 13 valget kuuli, mis on hoolikalt segatud. On selge, et valge kuuli saamiseks on rohkem lootusi kui musta kuuli saamiseks. Vaatame, kuidas oleks mõistlik valida arvulist näitajat, mis iseloomustaks seda, et valge kuuli saamine on rohkem võimalik kui musta kuuli saamine. Urnist saame kuuli võtta 20 erineval viisil, kusjuures valge kuul võib sattuda meie kätte 13 juhul, must 7 juhul. Mõistlik on valida valge kuuli saamise võimalikkuse näitajaks $\frac{13}{20}$. Kui valgete kuulide arv suureneks mustade arvel (kuulide arv urnis ei muutu), siis suurenevad ka võimalused valge kuuli saamiseks. Murrus, mille lugejaks on urnis olevate valgete kuulide arv ja nimetajaks kuulide koguarv, kasvaks lugeja (nimetaja muutumatuks jäädes). Murru väärtus suureneks. Kui lisada urni musti kuule, siis valge kuuli saamise objektiivsed võimalused vähenevad. Meie poolt koostatud murru nimetaja oleks ka suurem ja murru väärtus väheneks. Ütleme, et valge kuuli saamise tõenäosus on $\frac{13}{20}$ ja analoogiliselt musta kuuli saamise tõenäosus on $\frac{7}{20}$. Juhime tähelepanu asjaolule, et kui kuulid oleksid varustatud järjekorranumbritega, siis võimalused mistahes järjekorranumbriga kuuli saamiseks (värv pole oluline) oleksid võrdsed.

Viskame korraga 2 täringut. Kui suur on tõenäosus 4 silma saamiseks? Püüame jälle selgitada kõigi täringute viskel esineda võivate võimaluste arvu ning 4 silma saamise võimaluste arvu. Ühe täringu mingil tahul olev silmade arv võib esile tulla koos iga silmade arvuga teisel täringul - see oleks 6 erinevat võimalust. Kombineerides selliselt ühe täringu kõiki tahke teise täringu kõigi tahkudega, saaksime kokku 6.6 = 36 võimalikku silmade kombinatsiooni. Neli silma saame siis, kui kummalgi täringul on 2 silma või kui ühel täringul on 1 ja teisel 3 silma. Tundub, et analoogiliselt eelmise näitega võiks nüüd 4 silma saamise tõenäosuseks võtta murru $\frac{2}{36}$. See oleks aga väär. Soodsaid juhte 4 silma saamiseks pole ainult kaks. Loetletud soodsad juhud pole võrdvõimalikud: 2 + 2 silma võib esineda ühel viisil, 1 + 3 silma kahel viisil. Seega on 4 silma saamiseks soodsaid võrdvõimalikke juhte 3 ja otsitavaks tõenäosuseks oleks õige valida $\frac{3}{36} \approx 0{,}083$.

Nägime, et sündmuse toimumiseks soodsate juhtude arvu määramisel tuleb arvesse võtta kõik võrdvõimalikud juhud.

Olgu antud n võrdvõimalikust üksteist välistavast sündmusest koosnev täielik sündmuste süsteem. Olgu n võimaliku katsetulemuse hulgas teatud sündmuse A toimumiseks soodsate juhtude arv m. Meid huvitab sündmuse A toimumise võimalikkuse mõõt ehk sündmuse A tõenäosus sooritataval katsel, mille määrame järgmiselt.

Sündmuse matemaatiline tõenäosus on võrdne murruga, mille lugejaks on sündmuse toimumiseks soodsate juhtude arv ja nimetajaks kõigi võimalike juhtude arv.

Kui tähistada sündmuse matemaatilist tõenäosust tähega p ja sulgudes märkida sündmus, mille tõenäosusest on juttu, siis

$$p(A) = \frac{m}{n} .$$

Seega tuleb sündmuse matemaatilise tõenäosuse leidmiseks määrata

a) üksikkatsel kõigi mõeldavate tulemuste ehk võimalike juhtude arv eeldusel, et nad moodustavad täieliku sündmuste süsteemi,

b) meid huvitava sündmuse toimumiseks soodsate juhtude arv,

c) soodsate juhtude ja kõigi võimalike juhtude arvu jagatis.

Antud tõenäosuse leidmise eeskirjast järelduvad järgmised tõenäosuse omadused.

1. Juhusliku sündmuse tõenäosus on positiivne lihtmurd. Tõepoolest, juhusliku sündmuse toimumiseks soodsate juhtude arv m moodustab ainult osa kõigi võimaluste arvust n ja seega on $m < n$ ning

$$0 < \frac{m}{n} < 1 .$$

2. Kindla sündmuse tõenäosus on 1 .
Kindla sündmuse korral on katse iga mõeldav üksiktulemus sündmuse toimumise seisukohalt soodne juht, s.t. $m = n$ ja

$$p = \frac{n}{n} = 1 .$$

3. Võimatu sündmuse tõenäosus on 0 .

Võimatu sündmuse toimumiseks soodsate juhtude arv on 0 ja

$$p = \frac{0}{m} = 0 .$$

Kokkuvõte:

Mis tahes sündmuse matemaatiline tõenäosus p rahuldab tingimusi

$$0 \leq p \leq 1 .$$

Relatiivne sagedus. Igal katsel pole võimalik määrata sündmuse tõenäosust eespool antud skeemi järgi. Alati pole võimalik määrata katsel esineda võivate võrdvõimalike juhtude arvu ja samuti sündmuse toimumiseks soodsate juhtude arvu. Näiteks ebasümmeetrilise täringu korral pole täringu viskel esineda võivad kuus võimalust võrdvõimalikud ja ühe silma saamise tõenäosus ei tarvitse olla $\frac{1}{6}$. Samuti ei saa antud tõenäosuse leidmise eeskirja põhjal leida, kui tõenäone on televiisori valmistamisel praagi esinemine jne. Ometi on nende sündmuste toimumine võimalik ja võimalikkuse astet saab iseloomustada teatud arvulise näitajaga. Olgu näiteks 1000 kontrollitud televiisori hulgas defektidega 50. Kui tootmistingimused oluliselt ei muutu, siis on alust arvata, et ka teistes 1000 aparaadist koosnevates partiides kõigub defektidega aparaatide arv 50 ümber ehk moodustab umbes $\frac{50}{1000} = 0{,}05$ kogu partiist. Arv 0,05 iseloomustab vaadeldud tüüpi televiisorite korral praagi esinemise võimalust. Antud juhul pole võimalik hinnata defektidega televiisorite osa kogutoodangu hulgas teisiti kui katsel saadud andmete põhjal.

Kui sündmuse tõenäosuse leidmine pole teostatav juhtude loendamise skeemi ehk nn. tõenäosuse klassikalise definitsiooni järgi, kasutatakse teisi, katsele tuginevaid meetodeid. Katsel puutume kokku sündmuse relatiivse sageduse mõistega.

Olgu n katsest koosnevas katseseerias teatud sündmus A toimunud m katsel. Sündmuse relatiivne sagedus vaadeldavas katseseerias on võrdne murruga, mille lugejaks on sündmuse toimumiste arv ja nimetajaks sooritatud katsete üldarv.

Kui tähistada relatiivne sagedus tähega w, siis $w = \frac{m}{n}$, kusjuures $0 \leqslant w \leqslant 1$.

Sündmuse toimumiste arvu nimetame edaspidi sündmuse sageduseks.

Lühikeste katseseeriate puhul võib relatiivne sagedus üksikutes katseseeriates olla oluliselt erinev. Näiteks raha viskamisel 10 korda võib vapp ühes katseseerias esineda 2 korda, teises seerias 8 korda - esimesel juhul on vapi esinemise relatiivne sagedus 0,2 , teisel 0,8 . Katsete arvu suurenemisel on relatiivsel sagedusel tendents stabiliseeruda, ta kõigub väikeses ulatuses teatud kindla arvu ümber. Raha viskamisel läheneb vapi esinemise relatiivne sagedus arvule 0,5. Relatiivse sageduse seda omadust on inimkond oma praktilises tegevuses väga palju kordi kontrollinud. J. Bernoulli andis sellele omadusele matemaatilise tõestuse nn. suurte arvude seaduse kujul: sündmuse relatiivne sagedus läheneb katsete arvu suurenemisel sündmuse matemaatilisele tõenäosusele.

Sündmuse tõenäosus on loendatavate juhtude korral leitav katset sooritamata, relatiivne sagedus leitakse alati pärast katsete tegemist.

Tõenäosus ja relatiivne sagedus on lahutamatud mõisted. Näiteks, kui laskuri poolt märklaua tabamise tõenäosus on 0,9 , siis mõistame seda nii, et laskur tabab märklauda keskmiselt 90 lasuga 100-st ehk lasketabavus on 90% . Sündmuse tõenäosus seostub praktikas relatiivse sagedusega suure katsete arvu korral. Suurema tõenäosusega sündmused toimuvad üldiselt sagedamini kui väiksema tõenäosusega sündmused.

Kui sündmuse tõenäosus ei ole leitav juhtude loendamise teel või spetsiaalsete matemaatiliste võtetega, siis võib sündmuse tõenäosuse ligikaudseks väärtuseks võtta tema relatiivse sageduse küllalt pikas katseseerias. Muidugi säilib võimalus ka paljude katsete korral relatiivse sageduse ja tõenäosuse oluliseks erinemiseks, kuid selline võimalus realiseerub väga harva ehk see on vähe tõenäone.

Kui sündmuse tõenäosus on väga väike, siis toimub sündmus harva, s.t. teda võib lugeda praktiliselt võimatuks. Ühele lähedase tõenäosusega sündmust võib pidada praktiliselt kindlaks. Piir, millest väiksema tõenäosuse korral võib sündmust pidada praktiliselt võimatuks, on suhteline ja oleneb konkreetsetest ülesannetest. Analoogiline olukord on sündmuse praktilise kindluse määramisega. Kui langevarju mitteavanemise tõenäosus on 0,01 , siis ei saa seda lugeda praktiliselt võimatuks sündmuseks. Kui aga automaattööpingil on praagi valmistamise tõenäosus 0,01 , siis võime praktiliselt kogu toodangu lugeda kvaliteetseks.

Paljude majanduse valdkonda kuuluvate probleemide lahendamisel tuleb kasutada tõenäosuse mõistet.

Näiteid:

1. Kauplused peavad riietusesemeid ja jalanõusid tellima üksikute numbrite järgi erinevates kogustes, et rahuldada kõigi ostjate soove. Tuleks arvestada, kui tõenäone on, et tellitavad kogused on piisavad.
2. Töö häireteta kulgemiseks tehase tsehhis on tarvis hoida varuseadmeid, millega saaks kohe asendada rikete tõttu rivist väljalangenud seadmeid. Tagavaraseadmete hulk ei tohi olla liiga väike ega liiga suur - mõlemad variandid on ebamajanduslikud. Õigete proportsioonide määramiseks on tarvis teada seadmete riknemise tõenäosusi. Analoogiline olukord on mitmesuguste üksikdetailide tagavaradega.
3. Telefonikeskjaamade võimsuse planeerimisel tuleb arvestada, kui tõenäone on teatud arvu väljakutsete esinemine ajaühikus.
4. Ühe töötaja poolt teenindatav kudumisautomaatide arv sõltub niitide katkemise tõenäosusest. Mida vähem võimalusi

on niitide katkemiseks, seda rohkem automaate saab ühe töötaja järelevalve alla anda.

## 4. N ä i t e i d .

1. Leida metallraha viskamisel vapi esiletuleku tõenäosus.

   Raha viskel võib esineda kaks juhtu: võib tulla vapp või kiri. Vapi tulekuks soodsaid juhte on üks. Seega on vapi esiletuleku tõenäosus $p = \frac{1}{2}$ .

2. Abonent on unustanud vajaliku telefoninumbri kaks viimast numbrit (need on teineteisest erinevad) ja valib need juhuslikult. Kui tõenäone on, et ta valib õiged numbrid?

   Sündmuseks A olgu mõlema numbri õige valik. Telefoniketta 10 numbrit annavad erinevaid võimalusi nii palju, kui palju saab moodustada variatsioone 10 elemendist 2-kaupa:

$$V_{10}^{2} = 10.9 = 90 .$$

Sündmuse A toimumiseks on soodsaid juhte 1. Otsitav tõenäosus

$$p(A) = \frac{1}{90} \approx 0,011 .$$

3. Kaupluses töötab 7 nais- ja 3 meesmüüjat. Ühes vahetuses töötab 3 müüjat. Kui tõenäone on, et ühes juhuslikult valitud vahetuses on 3 meesmüüjat ?

   Sündmuseks olgu 3 meesmüüja töötamine ühes vahetuses. Erinevaid võimalusi müüjate jaotamiseks vahetuste kaupa on nii palju, kui saab moodustada kombinatsioone 10 elemendist 3-kaupa. Kõigi juhtude arv $n = C_{10}^{3} = 120$. Soodsaid juhte sündmuse A toimumiseks on ainult üks. Otsitav tõenäosus $p(A) = \frac{1}{120} \approx 0,008$.

4. Tehase TKO avastas 100 detaili hulgas 5 praakdetaili. Kui suur on praakdetailide relatiivne sagedus?

Iga detaili kontrollimine on katse. Praakdetaili esinemine on meid huvitav sündmus. Vaadeldav sündmus toimus 100 katsel 5 korda. Praagi esinemise relatiivne sagedus vaadeldud detailide partiis $w = \frac{5}{100} = 0{,}05 = 5\%$.

5. Automaattööpingil valmistatud detailidest on kvaliteetseid keskmiselt 97%. Kui suur on arvatav praakdetailide hulk 2000 detailist koosnevas partiis?

Praaki on kogutoodangust keskmiselt 3% ehk praagi keskmine relatiivne sagedus on 0,03. 2000 detaili korral on praakdetailide arv $m = w \cdot n = 0{,}03 \cdot 2000 = 60$.

## 5. Tõenäosuste liitmislause.

Tihti osutub sündmuse tõenäosuse arvutamine juhtude loendamise teel väga keerukaks. Kasutatakse mitmesuguseid kaudseid meetodeid, kus ühtede sündmuste tõenäosuste järgi leitakse teiste sündmuste tõenäosusi.

Kõigepealt loome sündmuste summa mõiste.

Kahe sündmuse A ja B summaks nimetame sündmust C, mis seisneb ükskõik kumma või mõlema sündmuse toimumises.

Kui $C = A + B$, siis sündmust C loeme toimunuks, kui toimub kas sündmus A või sündmus B või mõlemad järjest. Sündmusi A ja B nimetame liitsündmuse C osasündmusteks.

Rohkem kui kahe sündmuse summaks on sündmus, mis seisneb ükskõik millise antud osasündmuse toimumises.

Näiteid:

1. Loeme sündmuseks C jalatsivabriku konveierilt standardse kingapaari saamist. Kingapaar võib kuuluda kas I või II sorti. Standardse jalanõu saamist võime vaadelda koosnevana kahe osasündmuse summast, mida võime sümboolselt märkida järgmiselt

$$C = (I) + (II).$$

2. Täringu viskel paaritu silmade arvu saamine koosneb kolme osasündmuse summast: kas 1 või 3 või 5 silma saamine.
3. Hundi üheaegsel tulistamisel 3 jahimehe poolt võib hundi tapjaks osutuda ükskõik kes kolmest laskjast. Hundi tapmine (sündmus D) on 3 osasündmuse (hundi tapmine kas I, II või III jahimehe poolt) summa.

$$D = A_1 + A_2 + A_3 .$$

Kahes esimeses näites on osasündmused üksteist välistavad. Viimases näites on tegemist üksteist mittevälistavate osasündmustega - ühe jahimehe poolt sooritatud surmav lask ei takista ka teistel jahimeestel sama täpse lasu sooritamist.

Tõenäosuste liitmislause tuletame ainult üksteist välistavate osasündmuste jaoks.

Tõenäosuste liitmislause:

Kahe üksteist välistava sündmuse summa tõenäosus on võrdne osasündmuste tõenäosuste summaga.

Kui $C = A + B$, siis
$$P(C) = p(A) + p(B).$$

Liitmislause õigsuses võib veenduda järgmise arutlusega. Olgu katsel esineda võivate juhtude arv n, mille hulgas olgu osasündmuse A toimumiseks soodsaid juhte a ja osasündmuse B toimumiseks soodsaid juhte b . Üksteist välistavate osasündmuste A ja B korral pole A jaoks soodsate juhtude hulgas soodsaid juhte B jaoks ja vastupidi. Sündmuse $C = A + B$ toimumiseks on soodsaid juhte $a + b$ ja

$$p(C) = \frac{a+b}{n} = \frac{a}{n} + \frac{b}{n} .$$

Kuid $\frac{a}{n} = p(A)$, $\frac{b}{n} = p(B)$, ning seega ongi liitmislause õigsus näidatud.

Liitmislauset kasutame siis, kui osasündmuse kohta saab küsida "kas...või"; sündmus $C = A + B$ tähendab kas $A$ või $B$ toimumist.

Liitmislauset kasutame ka enam kui kahe üksteist välistava osasündmuse korral. Näiteks kolmest osasündmusest koosneva liitsündmuse puhul

$$p(A + B + C) = p(A) + p(B) + p(C).$$

Olgu tegemist n sündmusest koosneva täieliku sündmuste süsteemiga, kus sündmused on tähistatud $A_1, A_2, \ldots, A_n$. Kui tõenäone on, et toimub ükskõik missugune täielikku sündmuste süsteemi kuuluvatest sündmustest? Teisiti väljendades: kui suur on täieliku sündmuste süsteemi moodustavate sündmuste summa tõenäosus? Ükskõik missuguse täielikku sündmuste süsteemi kuuluv sündmus on osasündmuseks liitsündmusele $B = A_1 + A_2 + \ldots + A_n$. Üks osasündmustest $A_i$ ($i = 1, 2, \ldots, n$) toimub kindlasti (seleta, miks!). Seega on liitsündmus B kindel sündmus ja tema tõenäosus

$$p(B) = p(A_1 + \ldots + A_n) = 1.$$

Täieliku sündmuste süsteemi moodustavad sündmused on üksteist välistavad ja võime rakendada liitmislauset. Seega

$$p(A_1 + \ldots + A_n) = p(A_1) + p(A_2) + \ldots + p(A_n) = 1$$

– täieliku sündmuste süsteemi moodustavate sündmuste tõenäosuste summa võrdub ühega.

Järeldus. Sündmuse ja tema vastandsündmuse tõenäosuste summa on üks (seleta, miks!).

$$p(A) + p(\bar{A}) = 1 .$$

Näiteid:

1. Urnis on 30 kuuli: 5 valget, 7 rohelist ja 18 punast. Kui tõenäone on kas valge või rohelise kuuli saamine ühe kuuli juhuslikul võtmisel?

Esimene lahendusviis. Valge kuuli saamiseks (sündmus

V) on soodsaid juhte 5, rohelise kuuli saamiseks (sündmus R) - 7. Kas valge või rohelise kuuli saamine on üksteist välistavate osasündmuste summa (sündmus A). Seega

$$p(\text{kas } V \text{ või } R) = p(V + R) = \frac{5}{30} + \frac{7}{30} = \frac{12}{30} .$$

Teine lahendusviis. Sündmuseks A on kas valge või rohelise kuuli saamine - selle sündmuse toimumiseks on 30 võimaluse hulgas 12 soodsat võimalust. Otsitav tõenäosus

$$p(\text{kas } V \text{ või } R) = p(A) = \frac{12}{30} = \frac{2}{5} .$$

2. Raha ja asjade loterii iga 10 000 pileti kohta loositakse 150 rahalist ja 50 esemelist võitu. Kui tõenäone on ühe piletiga võitmine?

Sündmus "võitmine" (V) on kahe üksteist välistava "rahaline võit" (R) ja "esemeline võit" (E) summa.

$$p(V) = p(R) + p(E) = \frac{150}{10\,000} + \frac{50}{10\,000} = 0{,}02$$

3. Kui tõenäone on kahe täringu viskel saada 7 või 8 silma?

Kahe täringu viskel on erinevaid võrdvõimalikke võimalusi $6 \cdot 6 = 36$. Soodsaid võimalusi 7 silma saamiseks on 6: ühel täringul 1, teisel 6 silma; 2 ja 5 silma; 3 ja 4 silma ning vastupidi. Soodsaid võimalusi 8 silma saamiseks on 5. Kas 7 või 8 silma saamise (sündmus A) tõenäosus

$$p(A) = \frac{6}{36} + \frac{5}{36} = \frac{11}{36} .$$

4. Pikaajaliste vaatluste tulemusena selgus, et remonditöökojas on treipingi 20 peatuse põhjusteks keskmiselt 10 juhul lõiketera vahetamine, 3 juhul transmissiooni rikked, 2 juhul toorikute puudumine. Ülejäänud peatused on tingitud muudest põhjustest. Leida muudest põhjustest tingitud peatuste tõenäosus.

Lõiketera vahetamine (sündmus A), transmissiooni rikked sündmus (B), toorikute puudumine (C) ja muudest põhjustest tingitud peatused (D) moodustavad täieliku sündmuste süsteemi. Kuid

$$p(A) + p(B) + p(C) + p(D) = 1,$$

kust

$$p(D) = 1 - p(A) - p(B) - p(B) - p(C) = 1 - \frac{10}{20} - \frac{3}{20} - \frac{2}{20} = 0{,}25.$$

Lahendage sama ülesanne liitmislauset kasutamata.

5. Statistiliste andmete põhjal võib tütarlapse sündimise tõenäosuseks võtta 0,482. Leida poisslapse sündimise tõenäosus p.

Tütarlapse ja poisslapse sündimine on vastandsündmused ning

$$0{,}482 + p = 1.$$

Poisslapse sündimise tõenäosus $p = 0{,}518$.

Mitmete maade statistilised andmed kinnitavad, et 1000 vastsündinust on poisslapsi keskmiselt 518 ja tütarlapsi keskmiselt 482.

6. Loteriipiletiga võitmise tõenäosus on $\frac{1}{3}$. Kui tõenäone on, et 4 pileti omanik võidab ühega neljast piletist?

Ükskõik missuguse piletiga võitmine (A) on nelja osasündmuse ($A_1$, $A_2$, $A_3$, $A_4$ - 1., 2. jne. piletiga võitmine) summa. Liitmislause põhjal

$$p(A) = p(A_1) + p(A_2) + p(A_3) + p(A_4) = 4 \cdot \frac{1}{3} = 1\frac{1}{3} .$$

Sündmuse tõenäosus on ühest suurem!? Kus on viga? Tõepoolest osasündmused $A_1$, $A_2$, $A_3$, $A_4$ ei ole üksteist välistavad - võidu langemine ühele piletile ei takista samaaegselt võitu langemast ka mõnele teisele piletile. Tõenäosuste liitmislause eespool antud kujul ei ole üksteist mittevälistavate osasündmuste korral rakendatav.

Ühe mooduse üksteist mittevälistavate sündmuste summa tõenäosuse leidmiseks osasündmuste tõenäosuste kaudu anname hiljem.

## 6. Sõltumatud ja sõltuvad sündmused.

Alustame näidetega.

1. näide. Viskame täringut 2 korda. Täringu teisel viskel saadav silmade arv ei olene täringu esimesel viskel saadud silde arvust.
2. näide. Urnis on 10 sinist ja 16 rohelist kuuli. Võtame urnist ühe kuuli, registreerime tema värvi ja paneme kuuli urni tagasi. Teisel katsel sinise kuuli saamine ei olene esimesel katsel saadud kuulist. Esimese katse tulemus ei mõjuta ka teisel katsel sinise kuuli saamise tõenäosust, $p(s) = \frac{10}{26}$. Toimime nüüd aga nii, et esimesel katsel võetud kuuli urni tagasi ei pane. Teisel katsel sinise kuuli saamise tõenäosus sõltub sellest, mis värvi kuul tuli esimesel katsel. Kui esimesel katsel saime sinise kuuli, siis teisel katsel sinise kuuli saamise tõenäosus on $\frac{9}{25}$, kui esimesel katsel saime rohelise kuuli, siis – $\frac{10}{25}$.

Kaht sündmust nimetatakse sõltumatuteks, kui ühe sündmuse tõenäosus ei sõltu teise sündmuse toimumisest või mittetoimumisest.

Kaht sündmust nimetatakse sõltuvateks, kui ühe sündmuse tõenäosus sõltub teise sündmuse toimumisest või mittetoimumisest.

Olgu tegemist sõltuvate sündmustega A ja B. Sündmuse B tõenäosust, arvutatuna eeldusel, et sündmus A on toimunud, nimetame sündmuse B tinglikuks tõenäosuseks. Tingliku tõenäosuse tähistame järgmiselt: p(B/A) - kaldjoone ette märgime sündmuse, mille tinglikust tõenäosusest on juttu, kaldjoone taha sündmuse, mille toimumist eeldame.

3. näide. Kaubandusvõrku varustavad elektripirnidega kaks tehast, kusjuures I tehas annab 70%, II tehas 30% müügile minevatest pirnidest. I tehase poolt toodetud igast 100-st pirnist on keskmiselt 83 standardsed. II tehase toodangus on vastav näitaja 63. Leiame, kui tõenäone on pirni ostmisel standardse pirni saamine.

Kaupluses on keskmiselt igast 100 pirnist 70 toodetud I tehase poolt ja 30 II tehase poolt. I tehase 70 pirnist on keskmiselt $70 \cdot \frac{83}{100}$ standardseid, II tehase 30 pirnist on standardseid $30 \cdot \frac{63}{100}$ - kokku $70 \cdot 0{,}83 + 30 \cdot 0{,}63 = 77$ pirni. Seega on standardse pirni saamise tõenäosus $\frac{77}{100} = 0{,}77$.

Leiame veel standardse pirni saamise tõenäosuse eeldusel, et kaupluses on kõik I tehases toodetud pirnid. I tehases toodetud pirni saamine olgu sündmus A, standardse pirni saamine sündmus B. Seega tuleb leida sündmuse B tinglik tõenäosus. I tehase toodangu hulgast standardse pirni saamise tõenäosus $\frac{83}{100}$ ehk $p(B/A) = 0{,}83$.

4. näide. Käesoleva punkti 2. näite andmetel on teisel katsel sinise kuuli saamise tinglikud tõenäosused

$$p(S/R) = \frac{10}{25}, \quad p(S/S) = \frac{9}{25}.$$

Märkus: Sõltumatute sündmuste A ja B korral ühe sündmuse toimumine või mittetoimumine ei mõjuta teise sündmuse tõenäosust ja

$$p(A/B) = p(A) \qquad p(B/A) = p(B).$$

## 7. T õ e n ä o s u s t e   k o r r u t a m i s l a u s e.

Kahe sündmuse A ja B korrutiseks nimetame liitsündmust, mis seisneb nii sündmuse A kui sündmuse B toimumises. Kahe või enama sündmuse korrutiseks on sündmus, mis seisneb kõigi osasündmuste toimumises.

1. näide. Viskame raha kaks korda. Olgu kirja tulek ühel viskel sündmus A, vapi tulek sündmus B. Kui esimesel viskel tuleb kiri ja teisel vapp, siis räägime, et toimus sündmus AB, mida peamegi osasündmuste A ja B korrutiseks. Mõlemal viskel kirja tulek oleks sündmuseks, mida loeme osasündmuste A ja A korrutiseks AA. Mõlemal viskel vapi saamine on sündmus BB.

2. näide. Tööline teenindab kolme tööpinki. Häireteta töö ühe tunni jooksul I tööpingil olgu sündmus A, II tööpingil - sündmus B, III tööpingil - sündmus C. Sündmuste A ja B korrutiseks AB loeme sündmust, mis seisneb nii I kui II pingi häireteta töös ühe tunni jooksul. Kolme sündmuse A, B ja C korrutiseks on ABC - sündmus, mis seisneb kõigi 3 tööpingi häireteta töös ühe tunni jooksul.

Sündmust, mis seisneb nii ühe kui teise osasündmuse toimumises, võime tähistada ka ainult ühe tähega. Kui soovime aga eraldi rõhutada, et on tegemist liitsündmusega "nii ühe kui teise osasündmuse toimumine", on otstarbekas kasutada osasündmuste korrutist.

Kuidas leida sündmuste korrutise tõenäosust?

Tõenäosuste korrutamislause:

Tõenäosus nii ühe kui teise sõltuva sündmuse toimumiseks on võrdne ühe sündmuse tõenäosuse ja teise sündmuse tingliku tõenäosuse korrutisega, kus tinglik tõenäosus on leitud eeldusel, et üks osasündmustest on juba toimunud;

tõenäosus nii ühe kui teise sõltumatu sündmuse toimumiseks on võrdne mõlema osasündmuse tõenäosuste korrutisega.

Korrutamislauset võime esitada valemi abil:

1) sõltuvate osasündmuste juhul:

$$p(AB) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B) ;$$

2) sõltumatute osasündmuste juhul

$$p(AB) = p(A) \cdot p(B) .$$

Korrutamislause esimese poole õigsuses veendume järgmise arutlusega. Olgu sündmuse A toimumiseks ja B mittetoimumiseks a soodsat võimalust; B toimumiseks ja A mittetoimumiseks b soodsat võimalust. Nii A kui ka B mittetoimumiseks c soodsat võimalust; nii A kui B toimumiseks d soodsat võimalust. Seega oleksid kõik mõeldavad võimalused loendatud -

$$n = a + b + c + d .$$

Sündmuse "nii A kui B toimumine" (AB) tõenäosus $p(AB) = \frac{d}{n}$.
Sama tulemuse saame, kui korrutame sündmuse A tõenäosuse $p(A) = \frac{a + d}{n}$ sündmuse B tingliku tõenäosusega $p(B/A) = \frac{d}{a + d}$.

Sündmuse B tingliku tõenäosuse arvutamisel eeldame, et A on juba toimunud ning kõigi võimaluste arvuks tuleb võtta a + d.

Nüüd näeme, et

$$p(A) \cdot p(B/A) = \frac{a + d}{n} \cdot \frac{d}{a + d} = \frac{d}{n}$$

ehk

$$p(AB) = p(A) \cdot p(B/A) .$$

Näita, et $p(AB) = p(B) \cdot p(A/B)$!

Sõltumatute osasündmuste puhul on $p(B/A) = p(B)$ ja tõepoolest $p(AB) = p(A) \cdot p(B)$.

Korrutamislause kolmest osasündmusest moodustatud sündmuste korrutise juhul:

1) sõltuvate sündmuste korral

$$p(ABC) = p(A) \cdot p(B/A) \cdot p(C/AB) ,$$

kus sündmuste järjekorra valik on vaba;

2) sõltumatute sündmuste korral

$$p(ABC) = p(A) \cdot p(B) \cdot p(C) .$$

3. näide. Olgu 2. näites I tööpingi häireteta töötamise tõenäosus 0,9, II tööpingil 0,8 ja III tööpingil 0,85. Kui tõenäone on, et 1) nii I kui II tööpink töötavad vaadeldava tunni jooksul häireteta; 2) kõik kolm tööpinki töötavad häireteta?

Ühe tööpingi normaalne töö või ka rike ei mõjuta teiste pinkide tööd - sündmused A, B ja C on sõltumatud. Kuid siis

1) $p(AB) = 0,9 \cdot 0,8 = 0,72$; 2) $p(ABC) = 0,9 \cdot 0,8 \cdot$

$\cdot\, 0,85 = 0,612.$

4. näide. Ettevõtte toodangust on 95% standardne, millest 86% kuulub I sorti. Kui tõenäone on, et juhuslikult valitud toode kuulub I sorti?

Standardse toote saamine olgu sündmus A, I sordi toote saamine - sündmus B. Meid huvitab nii sündmuse A kui B toimumine, s.t. sündmus AB. Sündmuse A tõenäosus on $p(A) = 0,95$; sündmuse B tinglik tõenäosus eeldusel, et A on juba toimunud, on $p(B/A) = 0,86$ ja seega $p(AB) = p(A) \cdot p(B/A) = 0,95 \cdot 0,86 = 0,817$ .

5. näide. Trükikojas on 4 samatüübilist trükimasinat. Iga masina jaoks on antud momendil töötamise tõenäosus 0,9. Kui tõenäone on, et antud momendil vähemalt üks masin töötab?

Ülesannet võib lahendada mitmel viisil.

I. lahendusviis. Vähemalt ühe masina töötamist (sündmus B) võime vaadelda lihtsündmusena; B on järgmiste osasündmuste summa:

kõik 4 masinat töötavad,

3 masinat töötavad, 1 ei tööta,

2 masinat töötavad, 2 ei tööta,

1 masin töötab, 3 ei tööta.

Trükimasina töötamine olgu sündmus A, mittetöötamine - $\bar{A}$ . Sündmuse B osasündmused on omakorda liitsündmused. Näiteks sündmus "kõik 4 masinat töötavad" on sündmuse A korrutis iseendaga 4 korda. Sündmuse B osasündmuste tõenäosused leiame korrutamislause abil ja seejärel B tõenäosuse liitmislause abil.

Kuna $p(A) = 0,9$ , siis masina mittetöötamise tõenäosus

$$p(\bar{A}) = 1 - p(A) = 1 - 0,9 = 0,1 .$$

Kõigi 4 masina üheaegse töötamise tõenäosus $p(AAAA) = 0,9^4$.

Sündmus $AAA\bar{A}$ võib realiseeruda 4 variandis: kas ei tööta I ja töötavad ülejäänud või ei tööta ainult II või ... Seega

$$p(AAA\bar{A}) = (0,9^3 \cdot 0,1) \cdot 4 .$$

Sündmus $AA\bar{A}\bar{A}$ saab realiseeruda nii mitmes variandis kui palju saab moodustada kombinatsioone neljast elemendist kahekaupa, s.t. $C_4^2 = 6$ (pole ju oluline, missugused 2 masinat neljast parajasti töötavad).

$$p(AA\bar{A}\bar{A}) = (0,9^2 \cdot 0,1^2) \cdot 6 .$$

Sündmus $A\bar{A}\bar{A}\bar{A}$ saab realiseeruda jälle 4 variandis ja

$$p(A\bar{A}\bar{A}\bar{A}) = (0,9 \cdot 0,1^3) \cdot 4 .$$

Vähemalt ühe masina töötamise tõenäosus

$$p(B) = 0,9^4 + 0,9^3 \cdot 0,1 \cdot 4 + 0,9^2 \cdot 0,1^2 \cdot 6 + 0,9 \cdot 0,1^3 = 0,999 .$$

2. lahendusviis. Sündmus "vähemalt üks masin töötab" on vastandsündmuseks sündmusele "kõik masinad seisavad". Kuid siis

$$p(B) = 1 - p(\bar{B}) ,$$

kus $p(\bar{B}) = p(\bar{A}\bar{A}\bar{A}\bar{A}) = 0,1^4$ ja $p(B) = 1 - 0,1^4 = 0,999$ .

Analoogiliselt 2. lahendusviisis kasutatud arutlusele saab leida vähemalt ühe osasündmuse tõenäosust katsel, kus võib toimuda mitu sõltumatut osasündmust.

6. näide. Kolm jahimeest tulistavad samaaegselt hunti. Kui

tõenäone on, et hunt tapetakse, teades, et harilikult kulutab igaüks nendest jahimeestest hundi tapmiseks 2 lasku?

Tähistame sündmused "I jne. jahimees tapab hundi ühe lasuga" sümbolitega $A_1$, $A_2$, $A_3$. Antud juhul

$$p(A_1) = p(A_2) = p(A_3) = \frac{1}{2} \text{ ja } p(\bar{A}_1) = p(\bar{A}_2) = p(\bar{A}_3) = 1 - \frac{1}{2}.$$

Hundi tapmine (pole oluline, kas surmavalt tabas 1 või rohkem lasku) on vastandsündmus hundi mittetapmisele $\bar{A}_1\bar{A}_2\bar{A}_3$.

Nüüd saame kohe, et $p(\bar{A}_1\bar{A}_2\bar{A}_3) = \left(\frac{1}{2}\right)^3$ ja tõenäosus hundi tapmiseks vähemalt ühe jahimehe poolt oleks $1 - 0{,}125 = 0{,}875$.

Sündmused $A_1$, $A_2$ ja $A_3$ on üksteist mitte välistavad ja seetõttu ei saanud kasutada tõenäosuste liitmislauset eespool käsitletud kujul.

Ülesanne. Lahendada punktis 5 käsitletud 6. näide õigesti.

## 8. Tõenäosuste liitmislause üksteist mittevälistavate sündmuste puhul.

Üksteist mittevälistavate sündmuste summa tõenäosus on võrdne nende sündmuste tõenäosuste summaga, millest on lahutatud mõlema osasündmuse ühiselt esinemise tõenäosus.

Vastav valem esitub kujul

$$p(A + B) = p(A) + p(B) - p(AB).$$

Lause õigsuses veendume järgmiselt. Sündmused $A\bar{B}$ ja $AB$ on teineteist välistavad: $A$ ei saa samaaegselt toimuda koos $B$ ja $\bar{B}$-ga. Kas $A\bar{B}$ või $AB$ toimumine tähendavad ka $A$ toimumist ja seega

$$p(A) = p(A\bar{B}) + p(AB).$$

Analoogiliselt

$$p(B) = p(\bar{A}B) + p(AB).$$

Sündmus $A + B$ toimub, kui toimub ükskõik missugune üksteist välistavatest sündmustest $A\bar{B}$, $\bar{A}B$, $AB$. Kuid siis üksteist välistavate sündmuste liitmislause põhjal

$$p(A + B) = p(A\bar{B}) + p(\bar{A}B) + p(AB) .$$

Kui viimasesse võrdusse asendada $p(A\bar{B})$ ja $p(\bar{A}B)$ kahest eelmisest, siis saame

$$p(A + B) = p(A) + p(B) - p(AB), \text{ m.o.t.t.}$$

Märkus 1. Kui $A$ ja $B$ on sõltumatud sündmused, siis

$$p(A + B) = p(A) + p(B) - p(A) \cdot p(B) ;$$

sõltuvate sündmuste korral

$$p(A + B) = p(A) + p(B) - p(A) \cdot p(B/A).$$

Märkus 2. Üksteist välistavate sündmuste korral on nende korrutis $AB$ võimatu sündmus ja $p(AB) = 0$ ning jõuame tuttava tulemuseni

$$p(A + B) = p(A) + p(B) .$$

1. näide: Märki tulistatakse samaaegselt kahest kahurist. Märgi tabamise tõenäosus ühe lasuga on I kahuri korral $p_1 = 0,7$ ja II kahuri korral $p_2 = 0,8$. Kui tõenäone on märgi tabamine vähemalt ühest kahurist (sündmus $A_1 + A_2$) ? Sündmused $A_1$ ja $A_2$ on üksteist mittevälistavad, kuid teineteisest sõltumatud.
Vastavalt käesoleva punkti 1. märkusele

$$p(A_1 + A_2) = p_1 + p_2 - p_1 p_2 = 0,7 + 0,8 - 0,7 \cdot 0,8 = \\ = 0,94.$$

Märkus 3. Üksteist mittevälistavate sündmuste liitmislauset saab laiendada ka enam kui kahe sündmuse juhule. Kolme sündmuse puhul saaksime

$$p(A+B+C) = p(A) + p(B) + p(C) - p(AB) - p(AC) - p(BC) + \\ + P(ABC).$$

**Ülesanne.** Lahenda eelmises punktis antud hundi tulistamise ülesanne viimase valemi järgi.

2. näide. Kolme automaattööpingi ühe tunni jooksul häireteta töötamise tõenäosused on vastavalt 0,9, 0,8 ja 0,75. Kui tõenäone on, et vähemalt üks pink langeb rivist välja ühe tunni jooksul?

Tõenäosus I pingi seismajäämiseks on $p(A_1) = 1 - 0,9 = 0,1$; analoogiliselt $p(A_2) = 0,2$ ja $p(A_3) = 0,25$. Nii I kui II pingi seismajäämise tõenäosus on $p(A_1A_2) = 0,1 \cdot 0,2 = 0,02$; analoogiliselt $p(A_1A_3) = 0,1 \cdot 0,25 = 0,025$ ja $p(A_2A_3) = 0,20 \cdot 0,25 = 0,050$. Kõigi kolme pingi seismajäämise tõenäosus $p(A_1A_2A_3) = 0,1 \cdot 0,2 \cdot 0,25 = 0,0050$.

Vähemalt ühe pingi seismajäämise tõenäosus

$$p(A_1 + A_2 + A_3) = 0,1 + 0,2 + 0,25 - 0,02 - 0,025 - 0,05 + 0,005 = 0,460.$$

Mõlemaid ülesandeid saab lahendada ka eelmises punktis kasutatud mõttekäikudega.

## § 2. JUHUSLIKE SUURUSTE JAOTUSSEADUSED.

### 1. Juhuslikud suurused.

Tihti puutume kokku suurustega, mis muutumatute katsetingimuste korral võivad omandada erinevaid väärtusi.

1. näide. Täringu viskamisel võime saada kas 1, 2, 3, 4, 5 või 6 silma, s.t. et on tegemist 6 võimaliku väärtusega.

Ühegi silmade arvu saamist me ei saa ette ennustada – teatud silmade arvu saamine sõltub juhusest.

2. näide. Ühest viljapeast võetud terade kaal on üldiselt muutuv. Üksikute terade erinev kaal on tingi-

tud rea faktorite koosmõjust, mida me ei oska määrata. Räägime, et tera kaal sõltub juhusest.

3. näide. Sünnitusmajas sündinud tütarlaste arv 100 vastsündinu kohta on suurus, mis võib omada väärtusi nullist kuni sajani. Me ei saa ette määrata, milline tuleb tütarlaste arv 100 sündiva lapse kohta. Samuti ei saa kindlaks teha, millistel põhjustel on 100 vastsündinu hulgas kord 46, teine kord 48 tütarlast. 100 vastsündinu kohta tulev tütarlaste arv on juhusest sõltuv ehk juhuslik suurus, mis saab omandada ühe võimalikest väärtustest.

4. näide. Telefonikeskjaamas ühe tunni jooksul tellitud kaugekõnede arv on üksikutel tundidel erinev. Kaugekõnede arv 1 tunni jooksul võib olla 0, 1, 2 ... jne., kusjuures pole võimalik määrata, mis põhjustest on tingitud erinevused. Pikemaajaliste vaatluste põhjal võime ligikaudu otsustada, kui palju kaugekõnesid tellitakse päeva erinevatel tundidel.

Suurust nimetatakse juhuslikuks, kui ta omandab antud tingimustes ühe oma võimalikest väärtustest, mis sõltub juhuslikest põhjustest.

Juhusliku suuruse iga võimaliku väärtuse esinemine antud tingimustes on juhuslik sündmus. Katse kordamisel ilmneb, et võimalike väärtuste esinemise sagedus on erinev. Juhusliku suuruse iga väärtuse esinemise võimalikkust saab põhimõtteliselt iseloomustada vastava väärtuse esinemise tõenäosuse abil. Niisiis, täieliku ülevaate saamiseks juhusliku suuruse väärtuste esinemisest peame määrama suuruse kõik võimalikud väärtused ja nende väärtuste tõenäosused.

Juhuslikke suurusi liigitame diskreetseteks ja pidevateks.

Juhuslikku suurust nimetame diskreetseks, kui tema võimalikud väärtused erinevad üksteisest mingi lõpliku arvu võrra.

Käesoleva punkti 1., 3. ja 4. näites on tegemist diskreetsete juhuslike suurustega. Suuruste väärtused erinevad üksteisest vähemalt 1 võrra.

Juhuslikku suurust nimetame pidevaks, kui ta võib omandada kõiki väärtusi teatud lõplikust või lõpmatust arvude vahemikust.

Näites 2 võib terade kaal omandada mis tahes reaalarvulise väärtuse kõige kergema ja kõige raskema kaalu vahel. Muidugi kaalude piiratud täpsuse tõttu ei ole võimalik kui tahes väikesi kaalude erinevusi ilmestada.

5. näide. Inimese jalalaba pikkust saame vaadelda pideva juhusliku suurusena. Tema võimalike väärtuste hulk ei ole loendatav ja ei ole võimalik anda ka iga võimaliku väärtuse tõenäosust. Jaotame jalalaba pikkuse väiksemateks intervallideks. Saadud intervallid nummerdame ja jala suuruseks loeme sellele intervallile vastavat arvu, millesse langeb jala pikkus. Nüüd on võimalik anda ka iga jalanumbri esinemise tõenäosus. Selle põhjal saavad jalatsivabrikud oma toodangut planeerida jalanõude suuruse järgi. Õmblusvabrikud peavad valmisriiete õmblemisel arvestama inimeste pikkuse ja rinnaümbermõõdu võimalikke väärtusi ning nende esinemissagedusi, et tagada kõigi tarbijate soovide rahuldamist kaubandusvõrgus.

Tähistame edaspidi juhuslikke suurusi tähestiku suurte tähtedega, võimalikke väärtusi väikeste tähtedega. Juhuslik suurus X võib omandada väärtusi $x_1$, $x_2$, $x_3$, ... Täht X on suuruse sümboliks ja x tähistab suuruse võimalikku väärtust. Väärtuste tõenäosusi tähistame tähega $p_i$; väärtuse $x_i$ tõenäosus on $p_i$ ($i$ = 1, 2, 3 ...).

## 2. Juhusliku suuruse jaotustabel.

Iga eeskirja, mis määrab seose juhusliku suuruse kõigi võimalike väärtuste ja nende tõenäosuste vahel, nimetame juhusliku suuruse tõenäosuste ehk lihtsalt juhusliku suuruse jaotuseks. Juhusliku suuruse enda kohta ütleme, et ta allub antud jaotusele.

Üheks lihtsamaks juhusliku suuruse jaotuse esitamise viisiks on jaotustabel.

1. näide. Raha viskel on nii vapi kui kirja saamise tõenäosus $\frac{1}{2}$. Raha viskel esinevaid võimalusi võime vaadelda juhusliku suuruse X väärtustena järgmiselt: vapi tulekul ütleme, et suuruse väärtus on 0, kirja tulekul 1. Seega väärtuse 0 tõenäosus on $\frac{1}{2}$ ja väärtuse 1 tõenäosus $\frac{1}{2}$. Sobiv on juhuslikku suurust esitada tabeli kujul, kus ülemisse ritta kanname suuruse võimalikud väärtused ja alumisse ritta nende väärtuste tõenäosused. Saame tabeli

| x | 0 | 1 |
|---|---|---|
| p | 0,5 | 0,5 |

, mida nimetatakse jaotustabeliks.

2. näide. Viskame kaht täringut korraga. Saadav silmade arv on juhuslik suurus, mis võib omandada naturaalarvulisi väärtusi 2 kuni 12. Leiame üksikute väärtuste tõenäosused $p_i$ $(i = 2, 3, \ldots 12)$. Kahe täringu viskel on kõigi võrdvõimalike juhtude arv 36 (vt. näide § 1, punkt 3, lk. 11).

Soodsate juhtude arvu määrame loendamise teel:

2 silma saamiseks on soodsaid juhte 1;
3 " " " " " 2 (1 ja 2 või 2 ja 1);
4 " " " " " 3 (1 ja 3 või 3 ja 1
2 ja 2);
5 silma - 4; 6 silma - 5; 7 silma - 6; 8 silma - 5; 9 silma - 4; 10 silma - 3; 11 silma - 2; 12 silma - 1.

Vastavate väärtuste tõenäosused saame soodsate juhtude ja kõigi juhtude arvu jagamisel. Kui kanda tabeli ühte ritta silmade võimalikud väärtused ja teise ritta vastavad tõenäosused, saame

| silmade arv | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| p | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

Diskreetse juhusliku suuruse X jaotustabeli saame, kui esitame tabelis suuruse võimalikud väärtused $x_i$ ja vastavad tõenäosused $p_i$ $(i = 1, 2, \ldots, n)$

| x | $x_1$ | $x_2$ | ... | $x_n$ |
|---|---|---|---|---|
| p | $p_1$ | $p_2$ | ... | $p_n$ |

Pideva juhusliku suuruse jaotuse esitamiseks kasutatakse ka jaotustabelit, kus ühes reas on võimalike väärtuste muutumisvahemiku intervallid ja teises reas väärtuste vastavasse intervalli kuulumise tõenäosused. Kui võimalike väärtuste muutumisvahemik ehk väärtuste varu asub reaalarvude $x_0$ ja $x_n$ vahel, siis jaotame selle vahemiku kindlaks arvuks $n$ intervalliks (intervallid võivad olla võrdse või ka mittevõrdse pikkusega) $x_0 - x_1$, $x_1 - x_2$, ... , $x_{n-1} - x_n$. Tähistame juhusliku suuruse väärtuste teatud intervalli langemise tõenäosuse $p_i$-ga. Siis saame pideva juhusliku suuruse intervallitud jaotustabeli

| x | $x_0 - x_1$ | $x_1 - x_2$ | ..... | $x_{n-1} - x_n$ |
|---|---|---|---|---|
| p | $p_1$ | $p_2$ | ..... | $p_n$ |

Intervallitud jaotustabelit kasutame ka diskreetse suuruse korral, kui võimalike väärtuste arv on väga suur ja iga üksikväärtuse võimalikkuse mõõt polegi oluline.

Märkus: Kahe intervalli piiril asuva väärtuse loeme kõrgemasse intervalli kuuluvaks. Juhusliku suuruse jaotustabelis on antud suuruse kõik võimalikud väärtused, millest katsel üks kindlasti realiseerub. Suuruse võimalikud väärtused moodustavad seega täieliku sündmuste süsteemi. Täieliku sündmuste süsteemi moodustavate sündmuste tõenäosuste summa on aga 1. Järelikult suuruse jaotustabeli korral peab alati olema

$$p_1 + p_2 + \ldots p_n = \sum_{i=1}^{n} p_i = 1 .$$

## 3. Variatsioonrida. Statistiline jaotustabel.

A. Statistika algmõisted.

Statistilise uurimise objektiks olevat nähtuste, esemete või indiviidide kogu nimetame kogumiks. Iga kogumisse kuuluv nähtus, ese või indiviid kannab kogumi elemendi nime. Kogumi moodustamine toimub kõigile elementidele mingi ühise omaduse (või omaduste) järgi. Seda ühist omadust, mille järgi kollektiivi uurimine toimub, nimetame tunnuseks.

1. näide. Uurime Eesti NSV tööstusettevõtteid kvartaliplaani täitmise järgi. Tööstusettevõtted moodustavad kogumi, iga ettevõte on kogumi element, plaani täitmise % on uuritavaks tunnuseks. Antud tunnust saab arvuliselt avaldada.

2. näide. Rajoonikeskuse keskkooli õpilased jaotame vanemate tööala järgi kolhoosnikute, sovhoositööliste, tööstusettevõtete tööliste ja teenistujate lasteks. Tunnuseks on õpilaste vanemate tööala. Antud juhul me ei saa valitud tunnust arvuliselt kirjeldada. Tunnust, mida saab arvuliselt kirjeldada, nimetame kvantitatiivseks; tunnust, mida ei saa arvuliselt kirjeldada, nimetame kvalitatiivseks tunnuseks.

Tunnust kirjeldav arv kannab tunnuse väärtuse nime. Kogumi elementide tunnuste väärtused on üldiselt üksteisest erinevad, väärtused varieeruvad. Kui me ei suuda põhjendada tunnuse erinevate väärtuste esinemist, siis räägime jälle juhuslikust suurusest.

Vaatleme kvantitatiivse tunnusega kogumit. Tunnuse tähistame X-ga. Tunnuse X väärtuse $x_1$ saame katse või vaatluse teel. Teeme k katset, mille tulemusena saame tunnuse väärtuste jada

$x_1, x_2, x_3, \ldots, x_k$ , kus võib olla ka võrdseid liikmeid.

3. näide. Üliõpilaste rühma eksamihinded kõrgema matemaatika eksamil olid vastamise järjekorras järgmised:

5, 5, 4, 5, 3, 4, 5, 3, 4, 5 .

Kogumi elementide tunnuse väärtuste jada nimetame statistiliseks reaks. Statistilise rea saame katseandmete registreerimisel.

Võib juhtuda, et statistilises reas esineb tunnuse mõni väärtus korduvalt. Nii esineb 3. näite andmetel hinne "väga hea" statistilises reas 5 korda.

Arvu, mis näitab, mitu korda tunnuse mingi väärtus esineb statistilises reas, nimetame selle väärtuse sageduseks. Sagedust märgime tähega f .

Otstarbekas on statistilises reas väärtused esitada kindlas järjestuses ja ära näidata üksikute väärtuste sagedused.

Tabelit, kus tunnuse väärtused on esitatud kas kasvavas või kahanevas järjekorras koos üksikväärtuste sagedustega, nimetame variatsioonreaks. Kui erinevaid väärtusi on n tükki, saame variatsioonrea üldiselt kujul

| x | $x_1$ | $x_2$ | ..... | $x_n$ |
|---|---|---|---|---|
| f | $f_1$ | $f_2$ | ..... | $f_n$ |

4. näide. Eelmise näite andmetel saame variatsioonreaks

| Hinne | 5 | 4 | 3 |
|---|---|---|---|
| Sagedus | 5 | 3 | 2 |

Katsete üldarvu tähistame N ja nimetame kogumi mahuks.

$$N = f_1 \quad f_2 + _3 \ldots + f_n = \sum_{i=1}^{n} f_i$$

Iga üksikväärtuse $x_i$ sageduse $f_i$ ja kogumi mahu N jagatist nimetame väärtuse relatiivseks sageduseks $w_i$ :

$$w_i = \frac{f_i}{N}$$

Väärtuse $x_i$ relatiivne sagedus tähendab sama, mis sündmuse relatiivne sagedus ( § 1. p. 3.). Väärtuse $x_i$ esinemine on sündmuseks, $x_i$ sagedus vastab sündmuse toimumiste arvule ja kogumi maht $N$ katsete üldarvule.

B. Olgu antud juhusliku suuruse jaotustabel

| $x$ | $x_1$ | $x_2$ | ..... | $x_n$ |
|---|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | | $p_n$ |

Üksikväärtuse $x_i$ tõenäosus $p_i$ näitab vastava väärtuse esinemise võimalikkust katsel. $N$ katsel realiseeruvad suuruse $X$ väärtused $x_i$ teatud arv $f_i$ korda. Esitame katsetulemused kas variatsioonreana või tabelina, kus ühes reas on suuruse väärtused $x_i$ ja teises reas nende relatiivsed sagedused $w_i$.

Tabelit

| $x$ | $x_1$ | $x_2$ | ..... | $x_n$ |
|---|---|---|---|---|
| $w$ | $w_1$ | $w_2$ | ..... | $w_n$ |

nimetame juhusliku suuruse statistiliseks ehk empiiriliseks jaotustabeliks.

Teoreetiline jaotustabel näitab üksikväärtuste esinemise võimalikkust (tõenäosust), statistiline jaotustabel näitab, kuidas üksikväärtused katseseerias realiseerusid. Mida suurem on katsete arv, seda paremas kooskõlas on mõlemad jaotustabelid, s.t. seda väiksemad on erinevused $w_i$ ja $p_i$ vahel ühe ja sama suuruse jaoks.

Statistilise jaotustabeli korral kehtib relatiivsete sageduste jaoks omadus

$$\sum_{i=1}^{n} w_i = 1.$$

Tõepoolest

$$\sum_{i=1}^{n} w_i = \frac{f_1}{N} + \frac{f_2}{N} + \ldots\ldots + \frac{f_n}{N} = \frac{f_1 + \ldots + f_n}{N} = \frac{N}{N}.$$

5. näide. Koostada 4. näite andmetel statistiline jaotustabel.

Hinnete koguarv N = 10. Jaotustabeli saame kujul

| Hinne | 5 | 4 | 3 |
|---|---|---|---|
| Relatiivne sagedus w | 0,5 | 0,3 | 0,2 |

Relatiivsete sageduste summa on 0,5 + 0,3 + 0,2 = 1.

Kui statistilises reas on liikmeid niivõrd palju, et on tülikas anda tunnuse iga väärtuse esinemise sagedust, siis kasutame intervallitud jaotustabelit, mille alumises reas on vastavasse intervalli kuuluvate väärtuste sagedused. Üksikväärtuste kaupa esitatud variatsioonrida nimetatakse ka momentreaks.

6. näide. Uuritavate ettevõtete tootmisplaani täitmise protsendid on järgmised:

100 100 114 102 114 136 131 167 95 102 103 111
104 103 83 78 125 117 127 134 109 118 112 115
123 132 144 147 100 99 89 101 100 161 138 121
107 119 130 100 140 110 90 130 130 150.

Koostada intervallitud variatsioonrida ja statistiline jaotustabel, võttes intervalli pikkuseks 10%.

Võtame plaani täitmise % vahemiku algväärtuseks 75 ja lõppväärtuseks 175 ning määrame igasse intervalli kuuluvate ettevõtete arvu loendamise teel (statistika üldkursuses käsitletakse rühmitamise võtteid üksikasjalikumalt).

| Plaani täitmise % (x) | 75-85 | 85-95 | 95-105 | 105-115 | 115-125 | 125-135 |
|---|---|---|---|---|---|---|
| Ettevõtete arv f (x) | 2 | 2 | 13 | 8 | 6 | 8 |
| Relatiivne sagedus % (w) | $\frac{2}{47}$ | $\frac{2}{47}$ | $\frac{13}{47}$ | $\frac{8}{47}$ | $\frac{6}{47}$ | $\frac{8}{47}$ |

| Plaani täitmise % (x) | 135-145 | 145-155 | 155-165 | 165-175 |
|---|---|---|---|---|
| Ettevõtete arv (x) | 4 | 2 | 1 | 1 |
| Relatiivne sagedus (w) | $\frac{4}{47}$ | $\frac{2}{47}$ | $\frac{1}{47}$ | $\frac{1}{47}$ |

Kogumi maht on 47. Statistilise jaotustabeli saame intervalli kuuluvate ettevõtete arvu 47-ga jagades. Tabeli I ja III rida koos vaadelduna ongi statistiliseks jaotustabeliks.

Näeme, et juhusliku suuruse teoreetilise ja statistilise jaotustabeli koostamise printsiip on sama. Edaspidi räägime üldiselt juhusliku suuruse jaotustabelist, eraldi rõhutamata, kas on tegemist matemaatiliste tõenäosustega või relatiivsete sagedustega.

## 4. P o l ü g o o n  j a  h i s t o g r a m m.

Kõigis teadusharudes kasutatakse mitmesuguseid diagramme ja graafikuid suurustevahelisest sõltuvusest näitlikuma, selgema ülevaate saamiseks. Graafikud võimaldavad matemaatiliste tabelite ja valemite taga peituvaid seoseid konkreetsemalt ja ilmekamalt tunnetada. Tõenäosusteoorias ja matemaatilises statistikas on tähtsamateks graafikuteks juhuslike suuruste tõenäosuste jaotushulknurk ehk polügoon ja histogramm. Mitmesugused teised graafikud leiavad lähema käsitluse statistika üldteoorias.

Selgitame tõenäosuste jaotushulknurga ehk polügooni konstrueerimist diskreetse suuruse korral.

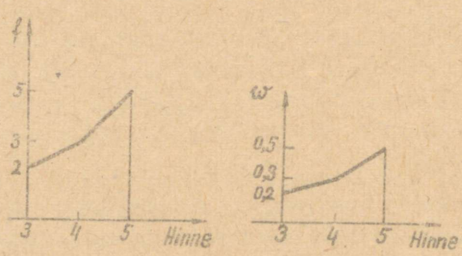
On antud juhusliku suuruse jaotustabel

| x | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| p | 0,1 | 0,2 | 0,25 | 0,2 | 0,2 | 0,05 |

Kanname koordinaatteljestikus abstsissteljele valitud mastaabis suuruse väärtused ja ordinaatteljele väärtuste tõenäosused. Kokkukuuluvad suuruse väärtuse $x_i$ ja tõenäosuse $p_i$ paarid kujutuvad tasapinna punktidena $(x_i, p_i)$. Ühendame saadud punktid sirglõikudega ja saame murdjoone, mis kannabki polügooni nime.

Variatsioonrea või statistilise jaotustabeli korral kanname ordinaatteljele vastavalt sagedused f või relatiivsed sagedused w, saades sageduste või relatiivsete sageduste polügoonid.

Eelmise punkti 4. näite andmetel saame järgmised sageduste ja relatiivsete sageduste polügoonid.




Joonis 1.

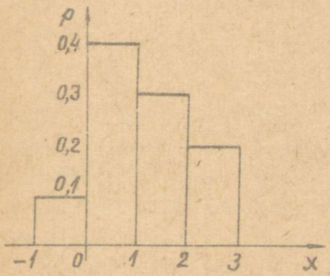
Polügooni abil saame kohe ülevaate sellest, milliste väärtuste esinemise tõenäosused üldse esineda võivate väärtuste hulgas on kõige suuremad või, empiiriliste andmete järgi koostatud polügooni korral, millised väärtused kõige sagedamini esinesid.

Pideva juhusliku suuruse korral on otstarbekas graafiliseks kujutamiseks kasutada histogrammi. Pideva suuruse jaotustabelis on antud suuruse väärtuste teatud intervalli kuulumise tõenäosused, kusjuures harilikult on intervallid võrdse pikkusega. Abstsissteljele kanname intervallid pikkusega $\Delta x$ ja ehitame neile lõikudele ristkülikud, millede kõrgused on võrdelised vastavate tõenäosustega.

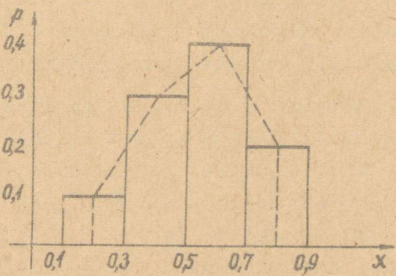
1. näide. Jaotustabelile

| x | -1 - 0 | 0 - 1 | 1 - 2 | 2 - 3 |
|---|---|---|---|---|
| p | 0,1 | 0,4 | 0,3 | 0,2 |

vastav histogramm on joonisel 2.




Joonis 2.




Joonis 3.

Antud juhul on intervalli pikkus $\Delta x = 1$ ja saadud ristkülikute pindalade summa $1.(0,1 + 0,4 + 0,3 + 0,2) = 1$.

Iga ristküliku pindala on võrdne suuruse väärtuste vastavasse intervalli kuulumise tõenäosusega.

2. näide. Juhusliku suuruse jaotustabel on järgmine

| x | 0,1 - 0,3 | 0,3 - 0,5 | 0,5 - 0,7 | 0,7 - 0,9 |
|---|---|---|---|---|
| p | 0,1 | 0,3 | 0,4 | 0,2 |

Konstrueerime histogrammi, võttes ristkülikute kõrgused võrdseks vastavate tõenäosustega. Nüüd on ristkülikute pindalad võrdelised intervallide tõenäosustega, kus võrdeteguriks on intervalli pikkus $\Delta x = 0,2$ . Kogu histogrammi pindala võrdub $0,2( 0,1 + 0,3 + 0,5 + 0,2 ) = 0,2 . 1 = 0,2$.

Kui aga valida abstsissteljel mastaap nii, et intervalli pikkus $\Delta x = 1$ , siis oleks histogrammi pindala 1. Sellisel juhul vastaksid abstsissteljel endistele jaotustele



6.

uued jaotused järgmiselt $x' = \frac{x}{0,2}$.

Sulgudes on antud punktide koordinaadid endises mastaabis.

| | (0) | (0,1) | (0,3) | (0,5) | (0,7) | (0,9) |
|---|---|---|---|---|---|---|
| $x'$ | 0 | 0,5 | 1,5 | 2,5 | 3,5 | 4,5 |

Histogrammi pindala oleks 1.(0,1 + 0,3 + 0,4 + 0,2) = 1. Histogrammilt võime üle minna polügoonile. Selleks ühendame lõikudega ristkülikute ülemiste aluste keskpunktid. Joonisel 3 on polügoon joonistatud punktiiriga.

Märkus. Kui tunnuse väärtuste muutumisvahemik pole jagatud võrdseteks intervallideks, siis on otstarbekas kanda ordinaatteljele nn. sageduse või relatiivse sageduse keskmine tihedus. Relatiivse sageduse keskmise tiheduse all mõistame intervalli kuuluvate väärtuste relatiivse sageduse ja intervalli pikkuse suhet. Sama võtet võib kasutada ka võrdsete intervallide korral.

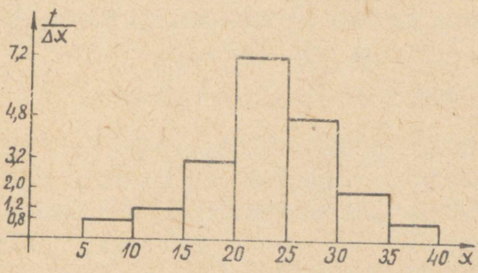
3. näide. Konstrueerida antud variatsioonreale vastav histogramm.

| Tunnuse väärtused | 5–10 | 10–15 | 15–20 | 20–25 | 25–30 | 30–35 | 35–40 |
|---|---|---|---|---|---|---|---|
| Sagedus f | 4 | 6 | 16 | 36 | 24 | 10 | 4 |

Kuna intervalli pikkus $\Delta x = 5$, siis sageduste tihedus intervallide järjekorras on 0,8; 1,2; 3,2; 7,2; 4,8; 2; 0,8.

Polügoon ja histogramm on jaotustabelile vastavaks graafikuks. Diskreetse suuruse korral kasutame ainult polügooni. Pideva suuruse jaoks võib kasutada nii histogrammi kui ka polügooni.

Histogrammi saame kujul




Joonis 4.

## 5. Juhusliku suuruse jaotusfunktsioon.

Diskreetse juhusliku suuruse iseloomustamiseks piisab tema jaotustabeli teadmisest. Pideva juhusliku suuruse korral on intervallitud jaotustabel ligikaudne jaotus. Ta annab ainult väärtuste antud pikkusega intervalli sattumise tõenäosuse, täpsemalt määramata antud pikkusest lühemasse intervalli sattumise või mõne üksikväärtuse esinemise tõenäosust. Suuruse väärtuste loendamatu hulga tõttu ei ole võimalik anda jaotustabelit üksikväärtuste kaupa.

Otstarbekas on anda seos pideva suuruse võimalike väärtuste ja nende tõenäosuste vahel matemaatilise avaldise kujul. Seda võib teha ka diskreetse juhusliku suuruse korral.

Tähistame juhusliku suuruse tähega X ja selle suuruse võimalikke väärtusi x. Oletame, et suuruse X võimalikud väärtused x moodustavad pideva vahemiku ehk teisiti väljendades kujutuvad arvteljel arvtelje katkematu osana.

Juhusliku suuruse jaotusfunktsiooniks nimetatakse funktsiooni $F(x)$, mille väärtused esitavad iga $x$ jaoks tõenäosust selleks, et juhuslik suurus omandab väiksema väärtuse kui $x$.

Kui tähistame tõenäosuse selleks, et $X < x$ lühidalt $P(X < x)$, siis definitsiooni kohaselt

$$F(x) = P(X < x) .$$

Geomeetriliselt võime seda tõlgendada nii: $F(x)$ määrab tõenäosuse selleks, et juhusliku suuruse väärtusele vastav punkt asub arvteljel punktist abstsissiga $x$ vasakul.

Võime näitlikult arutleda ka järgmiselt. Veeretame kuuli mööda lõpmata pikka horisontaalset renni. Kuul võib jääda (juhuslikest teguritest tingituna) peatuma kas ühes või teises punktis. Funktsioon $F(x)$ määrab, kui tõenäone on, et kuul ei veere kaugemale etteantud punktist.

Jaotusfunktsioon on üks juhusliku suuruse jaotuse esitusviise. Ta iseloomustab täielikult juhusliku suuruse väärtuste jaotumist nende esinemise tõenäosuse järgi. Teadaoleva funktsiooni $F(x)$ korral on võimalik iga $x$ korral leida, kui tõenäone on, et juhusliku suuruse väärtused on tõkkest $x$ väiksemad. $F(x)$ kasutame nii pideva kui ka diskreetse suuruse korral.

Vaatleme jaotusfunktsiooni omadusi.

1. Jaotusfunktsiooni piirväärtuseks on null, kui $x \to -\infty$

$$\lim_{x \to -\infty} F(x) = 0$$

ehk kokkuleppeliselt kirjutades $F(-\infty) = 0$.

2. Jaotusfunktsiooni piirväärtuseks on üks, kui $x \to +\infty$.

$$\lim_{x \to +\infty} F(x) = 1$$

ehk $F(+\infty) = 1$.

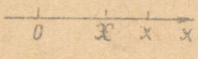
3. Jaotusfunktsioon on mittekahanev, s.t.

$$F(x_2) > F(x_1)\,, \text{ kui } x_2 > x_1 .$$

4. Kui juhusliku suuruse kõik võimalikud väärtused asuvad vahemikus $(a,b)$ siis

$$F(x) = 0, \text{ kui } x < a$$
$$F(x) = 1, \text{ kui } x > b .$$

Nende omaduste õigsuses veendume geomeetrilise tõlgenduse abil. Juhuslikku suurust $X$ vaatleme kui arvtelje liikuvat juhuslikku punkti, mis võib katsel sattuda ühte või teise kohta (joon. 5). $F(x)$ määrab siis liikuva punkti $X$ fikseeritud punktist $x$ vasakule jäämise tõenäosuse. Kui $x \to -\infty$, siis on sündmuse $X < x$ toimumine võimatu ja seega $F(-\infty) = 0$.




Joonis 5.

Kui $x \to +\infty$, siis $X < x$ on kindel sündmus, sest punkt $X$ ei saa asuda enam punktist $x$ paremal. Seega $F(+\infty) = 1$.

Kui $x_2 > x_1$, siis punkti $x$ liikumisel asendist $x = x_1$ paremale poole arvtelje see osa, kus punkt $X$ võib asuda, pikeneb. Sellega on üldiselt tõenäosus sündmuse $(X < x_2)$ toimumiseks suurem kui sündmuse $(X < x_1)$ toimumiseks, kuid see ütlebki, et $F(x)$ on mittekahanev.

Kui punkt $X$ saab asuda ainult vahemikus $(a,b)$, siis sündmus $(X < a)$ on võimatu ja järelikult $F(x) = 0$, kui $x < a$. Sündmus $(X < b)$ on kindel, sest punkt $x$ ei saa eelduse põhjal sattuda paremale punktist $b$. Ammugi on kindel sündmus $(X < x)$, kus $x > b$. Seega $F(x) = 1$, kui $x > b$.

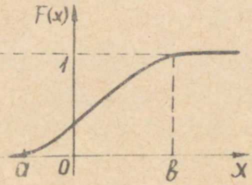
Kokkuvõte:

Jaotusfunktsioon $F(x)$ on mittekahanev ja tema väärtused asuvad 0 ning 1 vahel

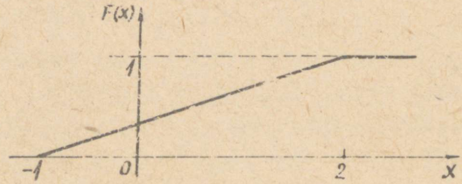
$$0 \leq F(x) \leq 1 .$$

Jaotusfunktsiooni omadustest järeldub, et tema graafik asub sirgete $y = 0$ ja $y = 1$ vahel ning kulgeb tõusvas joones. Jaotusfunktsiooni graafikut nimetatakse ka kumulaadiks.

Olgu teada, et pideva juhusliku suuruse võimalikud väärtused asuvad vahemikus (a , b). Siis kumulaat kulgeb üldjoontes nii, nagu on näha joonisel 6.




Joonis 6.




Joonis 7.

1. näide. Jaotusfunktsioon on määratud järgmiselt

$$F(x) = \begin{cases} 0 & , \text{kui } x < -1 \\ \frac{1}{3}x + \frac{1}{3} & , \text{kui } -1 \leq x \leq 2 \\ 1 & , \text{kui } x > 2 \end{cases}$$

Konstrueerida kumulaat.

Kui $x < -1$, ühtib graafik abstsissteljega; kui $-1 \leq x \leq 2$, on graafikuks sirge tõusuga $\frac{1}{3}$ ja $x > 2$ korral saame abstsissteljega paralleelse sirge $y = 1$ (joon. 7).

Diskreetse juhusliku suuruse kumulaadiks on treppjoon (joon. 8).

2. näide. Koostame juhusliku suuruse jaotustabeli järgi jaotusfunktsiooni ja kumulaadi.

| x | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| p | 0,1 | 0,2 | 0,25 | 0,2 | 0,2 | 0,05 |

Sündmus $X < -2$ on võimatu ja seega $F(-2) = 0$. Juhuslik suurus võib väärtusest $x = -1$ väiksema väärtuse omandada tõenäosusega 0,1 , s.t.

$$P(X < -1) = F(-1) = 0,1.$$

Sündmus $x \leq 0$ koosneb üksteist välistavatest osasündmustest $X = -2$ ja $X = -1$.

Tõenäosuste liitmislause põhjal

$$P(X \leq 0) = F(0) = 0,1 + 0,2 = 0,3$$

Analoogiliselt

$$F(1) = 0,1 + 0,2 + 0,25 = 0,55$$

$$F(2) = 0,1 + 0,2 + 0,25 + 0,2 = 0,75$$

$$F(3) = 0,75 + 0,2 = 0,95$$

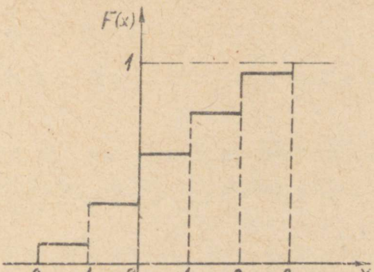
$$F(x) = 0,95 + 0,05 = 1 \text{ , kui } x > 3.$$

Kuna suuruse antud väärtuste vahel pole võimalikke väärtusi, siis jääb F(x) antud väärtuste vahemikus konstantseks.

$$F(x) = \begin{cases} 0 \text{ , kui } & x \quad -2 \\ 0,1 \text{ , kui } & - \quad x \quad -1 \\ 0,3 \text{ , kui } & -1 \quad x \quad 0 \\ 0,55 \text{ , kui } & 0 \quad x \quad 1 \\ 0,75 \text{ , kui } & 1 \quad x \quad 2 \\ 0,95 \text{ , kui } & 2 \quad x \quad 3 \\ 1 \text{ , kui } & x \quad 3 \end{cases}$$

Veendusime, et diskreetse juhusliku suuruse jaotusfunktsiooni väärtuste leidmine toimub järjest üksikväärtuste tõenäosuste juurdeliitmise teel.

Funktsiooni graafikul on hüpped suuruse võimalikele väärtustele vastavates punktides.




Joonis 8.

## 6. Juhusliku suuruse antud vahemikku langemise tõenäosus.

Paljude ülesannete lahendamisel on tarvis leida, kui tõenäone on, et juhuslik suurus omandab väärtuse antud väärtuste $x_1$ ja $x_2$ vahel. Räägime sündmusest "juhusliku suuruse langemine antud vahemikku" ja selle sündmuse tõenäosust tähistame

$$P(x_1 < X < x_2) .$$

Sündmus $X < x_2$ on kahe üksteist välistava sündmuse $X < x_1$ ja $x_1 < X < x_2$ summa ning tõenäosuste liitmislause põhjal

$$P(X < x_2) = P(X < x_1) + P(x_1 < X < x_2)$$

Siit leiame, et

$$P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1) ,$$

s.t.

$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

– juhusliku suuruse antud vahemikku langemise tõenäosus on võrdne jaotusfunktsiooni väärtuste vahega vahemiku otspunktides, ehk juhusliku suuruse antud vahemikku langemise tõenäosus on võrdne jaotusfunktsiooni juurdekasvuga selles vahemikus.

Käsitleme ainult selliseid pidevaid juhuslikke suurusi, mille jaotusfunktsioon on pidev.

Tähistame $x_2 = x_1 + \Delta x$, kus $\Delta x$ on argumendi juurdekasv ja laseme $\Delta x \to 0$.

Siis sündmus $x_1 < X < x_1 + \Delta x$ taandub sündmuseks $X = x_1$.

Kuid siis ka

$$P(X = x_1) = \lim_{\Delta x \to 0} P(x_1 < X < x_1 + \Delta x) =$$
$$= \lim_{\Delta x \to 0} \left[F(x_1 + \Delta x) - F(x_1)\right] = 0,$$

sest pideva funktsiooni juurdekasv kahaneb tõkestamatult koos argumendi juurdekasvu tõkestamatu kahanemisega.

Seega pideva juhusliku suuruse mis tahes üksikväärtuse esinemise tõenäosus on null.

Võib tunduda, et jõudsime vastuolulise tulemuseni. Seni rääkisime, et võimatu sündmuse tõenäosus on null. Sündmus $X = x_1$ on ometi võimalik, kuigi tema tõenäosus on null. Vastuolu siiski ei ole. Olukord - sündmus on võimalik, kuigi tema tõenäosus on null - esineb siis, kui tõenäosuse leidmine ei taandu juhtude loendamisele. Pideva juhusliku suuruse korral see tõepoolest nii ongi. Katsel omandab juhuslik suurus X ühe oma võimalikest väärtustest, kuigi iga üksikväärtuse tõenäosus on null. Sündmuse relatiivne sagedus ei ole isegi suure arvu katsete korral täpselt võrdne sündmuse tõenäosusega, vaid ainult läheneb sellele. Järelikult tulemusest $P(X = x_1) = 0$ tuleb nii aru saada, et katsete korduval sooritamisel tulmub sündmus $X = x_1$ väga harva, s.t. tema esinemine on väga ebatõenäone.

Pideva juhusliku suuruse korral on mõtet rääkida ainult tema mingisse vahemikku (ka kui tahes väikesse) langemise tõenäosusest. Suuruse võimalike väärtuste lõpmatu suure arvu korral on täiesti loomulik, et teatud kindel üksikväärtus saab realiseeruda väga harva. Nii on sõdade kogemused näidanud, et praktiliselt kaks mürsku ühes lahingus täpselt samasse kohta ei lange, kuid väikesele maa-alale võib langeda palju mürske.

Praktika seisukohalt pakubki huvi näiteks see, millistes piirides kõigub reisijate arv linnaliinide bussides tippkoormuste ajal, mitte aga kindel reisijate arv.

## 7. Tõenäosuse tihedus.

Olgu antud pideva juhusliku suuruse X jaotusfunktsioon, mille kohta eeldame, et ta on pidev ning omab tuletise iga x korral.

Suuruse X vahemikku $(x, x + \Delta x)$ langemise tõenäosus avaldub kujul

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x),$$

Jagatis

$$\frac{P(x < X < x + \Delta x)}{\Delta x}$$

määrab juhusliku suuruse X väärtuste esinemise keskmise tõenäosuse pikkusühiku kohta lõigu $\Delta x$ ulatuses, mida nimetame tõenäosuse keskmiseks tiheduseks lõigul $\Delta x$.

Tõenäosuse keskmise tiheduse mõiste jaotusfunktsiooni abil defineerime analoogiliselt sellele, kuidas jõudsime varem keskmise tootmiskiiruse mõisteni tootmisseaduse kaudu või üldiselt funktsiooni muutumise keskmise kiiruseni funktsiooni enda kaudu.

Tõenäosuse keskmine tihedus $p_k(x)$ lõigul $\Delta x$ on leitav valemist

$$p_k(x) = \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Kui $\Delta x \to 0$, siis

$$\lim_{\Delta x \to 0} p_k(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x}$$

Tõenäosuse keskmise tiheduse piirväärtust vahemiku pikkuse $\Delta x$ tõkestamatul kahanemisel nimetatakse tõenäosuse tiheduseks punktis x. Tõenäosuse tihedus oleneb punktist x, s.t. on suuruse võimalike väärtuste x funktsioon, mille tähistamiseks kasutame sümbolit $p(x)$. Tõenäosuse tiheduse graafikut nimetame tõenäosuste jaotuskõveraks.

Teiselt poolt

$$p(x) = \lim_{\Delta x \to 0} p_k(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x)$$

funktsiooni tuletise definitsiooni järgi.

Tõenäosuse tihedus punktis x on võrdne jaotusfunktsiooni tuletisega samas punktis

$$p(x) = F'(x) .$$

Väiksesse vahemikku pikkusega $\Delta x$ langemise tõenäosus

$$F(x+\Delta x) - F(x) \approx dF(x); \quad [\Delta y \approx dy!]$$

Teame, et

$$dF(x) = F'(x) \cdot dx = p(x)\,dx .$$

Tõenäosus juhusliku suuruse langemiseks väiksesse vahemikku $(x,\ x+\Delta x)$ on ligikaudu võrdne punktis $x$ arvutatud tõenäosuse tiheduse ja vahemiku pikkuse $\Delta x$ korrutisega, s.t.

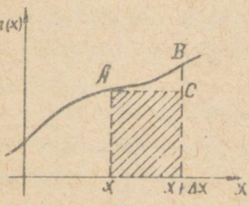
$$p(x) \cdot \Delta x$$

Tõenäosuste jaotuskõvera abil saame seda tulemust tõlgendada kui punktis $x$ alusega $\Delta x$ ja kõrgusega $p(x)$ konstrueeritud ristküliku pindala.

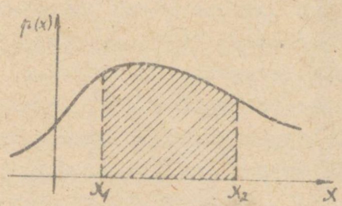
Võrduse

$$P(x < X < x+\Delta x) \approx p(x) \cdot \Delta x$$

viga on seda väiksem, mida väiksem on $\Delta x$.




Joonis 9.




Joonis 10.

Joniselt 9 näeme, et $\Delta x$ vähenemisel väheneb ka kõverjoonse kolmnurga ABC pindala. Pideva suuruse korral on korrutis $p(x) \cdot \Delta x$ sama tähendusega, mida diskreetse suuruse korral üksikväärtuse tõenäosus $p$.

Tõenäosuse tihedus määrab samuti suuruse $X$ võimalike väärtuste realiseerumise võimalused ja ta on juhusliku suuruse jaotuse üheks esitusviisiks.

Teades tõenäosuse tihedust, saab leida juhusliku suuruse vahemikku $(x_1,\ x_2)$ langemise tõenäosuse. Selleks jaotame vahemiku osavahemikeks pikkusega $\Delta x$. Sündmus $x_1 < X < x_2$ on siis suuruse $X$ osavahemikesse pikkusega $\Delta x$ langemiste summa. Vahemikku $\Delta x$ langemise tõenäosus oli aga $p(x) \cdot \Delta x$.

Tõenäosuste liitmislause põhjal

$$P(x_1 < X < x_2) \approx \sum_{x_1}^{x_2} p(x) \cdot \Delta x$$

Täpse võrduseni jõuame, kui suurendame osavahemike arvu piiramatult ja $\Delta x \to 0$. Kuid

$$\lim_{\Delta x \to 0} \sum_{x_1}^{x_2} p(x) \cdot \Delta x = \int_{x_1}^{x_2} p(x)\, dx$$

vastavalt määratud integraali definitsioonile.

Jõudsime tulemusele

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

- juhusliku suuruse antud vahemikku langemise tõenäosus on võrdne tõenäosuse tiheduse määratud integraaliga vahemiku otspunktidele vastavates rajades.

Geomeetriliselt vastab juhusliku suuruse vahemikku $(x_1, x_2)$ langemise tõenäosusele tõenäosuste jaotuskõvera alla jääv kõverjooneline trapetsi pindala vastavas vahemikus (joonis 10). Jaotusfunktsiooni saab leida tõenäosuse tiheduse kaudu järgmiselt:

$$F(x) = \int_{-\infty}^{x} p(x)\, dx$$

Tõepoolest $P(X < x) = P(-\infty < X < x) = \int_{-\infty}^{x} p(x)dx$

Tõenäosuse tiheduse omadusi:

1. Tõenäosuse tihedus on mittenegatiivne funktsioon, s.t.

$$p(x) \geqslant 0 .$$

Tõepoolest, $F(x)$ on mittekahanev funktsioon ja mittekahaneva funktsiooni tuletis on mittenegatiivne.

2. Integraal lõpmatutes rajades tõenäosuse tihedusest on võrdne ühega.

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1, \quad \text{sest} \quad \int_{-\infty}^{+\infty} p(x)\,dx = F(+\infty) - F(-\infty) =$$

$$= 1 - 0 = 1,$$

ehk teisiti: sündmus $-\infty < X < +\infty$ on kindel sündmus ning tema tõenäosus on 1.

Geomeetriliselt väljenduvad need omadused järgmiselt: tõenäosuse tiheduse kõver ei asu allpool abstsisstelge ning kogu kõvera ja abstsisstelje vahele jääva kujundi pindala on võrdne ühega.

1. näide. Jaotusseadus kujul

$$p(x) = \frac{1}{2}\, e^{\frac{-(x-a)^2}{2\sigma^2}} \qquad a, \sigma \text{ - const.}$$

kannab normaaljaotuse tiheduse nime.

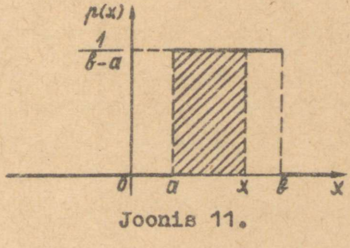
Normaaljaotust käsitleme üksikasjalisemalt eraldi paragrahvis.

2. näide. Ühtlane jaotus.

Jaotust nimetatakse ühtlaseks, kui tõenäosuse tihedus on vahemikus (a, b) konstantne ja väljaspool seda kõikjal 0. Tõenäosuse tihedus on määratud valemiga

$$p(x) = \begin{cases} 0 & \text{, kui } x < a \\ \dfrac{1}{b-a} & \text{, kui } x > a \\ 0 & \text{, kui } x > b \end{cases}$$

Tema graafik näeb välja järgmiselt




Joonis 11.

Vahemikus $(a,b)$ peab olema $p(x) = \dfrac{1}{b-a}$, sest muidu ei oleks täidetud tingimus, et tõenäosuse tiheduse kõvera ja abstsisstelje vaheline pindala võrdub ühega.

Jaotusfunktsiooni $F(x)$ määrame $p(x)$ graafiku abil (joonis 11). Kui $x < a$, siis on $p(x)$ graafiku ja x-telje vaheline pindala null; vahemikus $a < x < b$ on $F(x)$ võrdne viirutatud ristküliku pindalaga; kui $x > b$, siis on $F(x)$ võrdne ühega. Seega

$$F(x) = \begin{cases} 0 & , \text{ kui } x < a \\ \dfrac{x-a}{b-a} & , \text{ kui } a < x < b \\ 1 & , \text{ kui } x > b \end{cases}$$

Ühtlase jaotusega puutume kokku selliste ülesannete lahendamisel, kus on teada, et suuruse võimalikud väärtused asuvad kindlas vahemikus ja kõigi võimalike väärtuste tõenäosused on võrdsed.

Liikugu bussid teatud liinil 5-minutiliste vaheaegadega. Bussipeatusse tuleva reisija ooteaeg T on juhuslik suurus, mille võimalikud väärtused asuvad 0 ja 5 minuti vahel. Tõenäosus selleks, et tuleb oodata kas 1 min. või 2 min. jne. on konstantne ning võrdub $\frac{1}{5}$. Tõenäosuse selleks, et ei tule oodata kauem kui teatud arv minuteid, leiame jaotusfunktsiooni $F(T)$ abil. Kahest minutist lühema ooteaja tõenäosus

$$P(T < 2) = F(2) = \frac{2}{5} ,$$

$$P(T < 3) = F(3) = \frac{3}{5} ; \quad P(T < 5) = F(5) = 1.$$

Üle 5 min. aga tööpoolest ei tule oodata.

## § 3. JUHUSLIKU SUURUSE ARVULISED KARAKTERISTIKUD.

### 1. Arvuliste karakteristikute eesmärk.

Eelmises paragrahvis rääkisime juhusliku suuruse jaotusest. Jaotusseadus mistahes kujul kirjeldab juhuslikku suurust täielikult suuruse väärtuste esinemise võimalikkuse seisukohalt.

Paljudes ülesannetes ei ole eesmärgiks suuruse täielik kirjeldamine, vaid on oluline suuruse teatud külgede esiletoomine. Selleks kasutatakse mitmesuguseid arvulisi parameetreid. Harilikult kirjeldavad need parameetrid suuruse teatud külgi ilmekamalt kui jaotusseadus tervikuna. Neid parameetreid nimetame juhusliku suuruse arvulisteks karakteristikuteks. Suuruse eri külgede kirjeldamiseks võime luua mitmesuguseid karakteristikuid. Nende arvu ei maksa aga liiga suureks ajada ning nad tuleksid valida nii, et suuruse meid huvitavad omadused väljenduksid võimalikult hästi.

Juhusliku suuruse puhul tunneme ikka huvi, kas tema väärtused grupeeruvad mõne väärtuse ümber tihedamalt, kui laias piirkonnas väärtused paigutuvad jne. Käesolevas paragrahvis tutvume asendit ja varieerumist kirjeldavate karakteristikutega diskreetse suuruse jaoks.

### 2. Keskväärtus.

Asendi karakteristikutest on kõige tähtsam juhusliku suuruse keskväärtus.

Olgu suurus x määratud jaotustabeliga

| $x$ | $x_1$ | $x_2$ ...... | $x_n$ |
|---|---|---|---|
| $p$ | $p_1$ | $p_2$ ...... | $p_n$ |

Juhusliku suuruse keskväärtuseks nimetatakse suuruse võimalike väärtuste ja nende tõenäosuste korrutiste summat.

Tähistame suuruse X keskväärtuse sümboliga E(x) või a. Jaotustabeli andmetel naeb E(x) arvutamise eeskiri välja järgmiselt

$$E(x) = a = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^{n} x_i p_i \,.$$[1]

1. näide. Kahe täringu viskamisel saadava võimaliku silmade arvu jaotustabeli andmetel ( § 2 punkt 2) leida silmade arvu keskväärtus.

$$E(x) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} +$$

$$+ 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} +$$

$$+ 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = \frac{252}{36} = 7 \,.$$

2. näide. Sündmuse A toimumise võimalikkust üksikkatsel tõlgendame juhusliku suurusena X , mis omandab väärtuse 1, kui A toimub ja väärtuse 0, kui A ei toimu. Kui A toimumise tõenäosus on p ja mittetoimumise tõenäosus q = 1 - p , siis suuruse X jaotustabel omab kuju

| x | 0 | 1 |
|---|---|---|
| p | q | p |

Leiame E(x) .

$$E(x) = 0 \cdot q + 1 \cdot p = p$$

---

[1] Mõnikord jätame summamärgi juures indeksi i muutumise piirid märkimata.

- üksikkatsel on sündmuse toimumise keskväärtus võrdne sündmuse tõenäosusega.

## 3. Aritmeetiline keskmine.

Olgu statistilise kogumi tunnuse x uurimiseks sooritatud N katset ning katseandmed esitatud statistilise reana

$$x_1 \quad x_2 \quad x_3 \; \ldots\ldots \; x_N .$$

Statistilise rea aritmeetiliseks keskmiseks $\bar{x}$ nimetatakse tunnuse kõigi väärtuste summa ja väärtuste üldarvu jagatist.

Keskmise leidmiseks saame eeskirja

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum x_i}{N} .$$

Tunnuse eri väärtused võivad esineda korduvalt. Harilikult kasutame variatsioonrida (tabel I) või statistilist jaotustabelit (tabel II)

(I)

| x | $x_1 \; \ldots\ldots \; x_n$ |
|---|---|
| f | $f_1 \; \ldots\ldots \; f_n$ |

(II)

| x | $x_1 \; \ldots\ldots \; x_n$ |
|---|---|
| w | $w_1 \; \ldots\ldots \; w_n$ |

Variatsioonrea andmetel tuleks $\bar{x}$ leida valemist

$$\bar{x} = \frac{x_1 f_1 + \ldots + x_n f_n}{N} = \frac{\sum x_i f_i}{\sum f_i} .$$

Mõnikord räägitakse, et viimasel juhul on leitud kaalutud keskmine, esimesel juhul lihtne keskmine. Samade andmete korral saame arvulise tulemuse ühesuguse, erinevus on ainult matemaatilise eeskirja välises kujus. Näiteks korrutis $x_1 . f_1$ tähendab sama, mida $f_1$ korda esinenud väärtuse $x_1$ iseendaga liitmine.

Kui jagada $\bar{x}$ lugejas iga liidetavat eraldi nimetajaga N, jõuame seoseni

$$\bar{x} = x_1 \frac{f_1}{N} + x_2 \cdot \frac{f_2}{N} + \dots + x_n \cdot \frac{f_n}{N} \, .$$

Me teame, et

$$\frac{f_i}{N} = w_i \quad (i = 1, 2, \dots, n) \, .$$

Aritmeetilise keskmise leiame statistilise jaotustabeli korral valemist

$$\bar{X} = x_1 w_1 + x_2 w_2 + \dots + x_n w_n = \sum x_i w_i \, .$$

Kuidas on seotud keskväärtus ja aritmeetiline keskmine? Katsete arvu suurendamisel läheneb suuruse X iga väärtuse relatiivne sagedus vastava väärtuse tõenäosusele, s.t.

$$w_i \approx p_i \quad ( i = 1, 2, \dots , n ) \, .$$

Asendades $\bar{x}$ avaldises $w_i$ tõenäosustega $p_i$ jõuame tulemuseni

$$\bar{x} \approx x_1 p_1 + \dots + x_n p_n \, .$$

Võrduse parem pool on aga suuruse X keskväärtus. Seega

$$\bar{x} \approx E(x)$$

– juhusliku suuruse keskväärtus ja katsel saadud väärtuste aritmeetiline keskmine erinevad teineteisest järjest vähem katsete arvu suurendamisel.

Sooritades suurusega X mitu katseseeriat, saame igas seerias erineva väärtuste jaotustabeli (erinevused võivad olla nii väärtuste osas kui ka võrdsete väärtuste sageduste osas). Katseseeriatele vastavad erinevad keskmised, mis kõiguvad suuruse keskväärtuse ümber. Alati ei ole suuruse jaotusseadus teada ja ligikaudseks jaotusseaduseks võetakse katseseeriale vastav variatsioonrida. Leitud aritmeetilise keskmise järgi saame ligikaudselt hinnata, milline on uuritava suuruse keskväärtus.

Aritmeetilise keskmise ja keskväärtuse dimensioon ühtib suuruse dimensiooniga.

Märkus: Tihti kasutatakse sümboli E(x) asemel $\bar{x}$, mis on täiesti õigustatud, sest $\bar{x} \approx E(x)$.

1. näide. Sovhoosi 34 kombaini koristasid vilja ühel päeval järgmiselt

| x (ha päevas) | 22 | 24 | 25 | 26 | 27 |
|---|---|---|---|---|---|
| f (komb. arv) | 6 | 5 | 8 | 7 | 8 |

Leida ühe kombaini kohta tulev keskmine hektarite arv päevas.

Aritmeetiline keskmine

$$\bar{x} = \frac{22.6 + 24.5 + 25.8 + 26.7 + 27.8}{34} = \frac{850}{34} =$$

$$= 25 \text{ ha} .$$

2. näide. Automaatpingi tehnoloogilise protsessi analüüs on näidanud, et tööpingi õige reguleerimise korral on 10 toodetud detailist koosnevas partiis standardile mittevastavate detailide esinemise jaotustabel järgmine

| det. arv (x) | 0 | 1 | 2 | 3 | 4 | 5 ja rohkem |
|---|---|---|---|---|---|---|
| p | 0,24 | 0,38 | 0,26 | 0,10 | 0,02 | 0 |

Teostati 50 kümnest detailist koosneva partii kontrollmõõtmine, kust selgus, et partiid jaotusid mittestandardsete detailide arvu järgi järgmiselt

| x | 0 | 1 | 2 | 3 | 4 | 5. |
|---|---|---|---|---|---|---|
| f | 11 | 20 | 12 | 5 | 2 | 0 |

Leida mittestandardsete detailide arvu E(x) ja $\bar{x}$ 10 detailist koosnevas partiis.

Saame

$$E(x) = 0.0,24 + 1.0,38 + 2.0,26 + 3.0,10 + 4.0,02 = 1,28 ,$$

$$\bar{x} = \frac{0.11 + 1.20 + 2.12 + 3.5 + 4.2}{50} = 1,34 .$$

Partiide arvu suurendamisel muutuks $E(x)$ ja $\bar{x}$ erinevus veelgi väiksemaks.

## 4. K e s k v ä ä r t u s e  o m a d u s i .

Käsitletavad keskväärtuse omadused kehtivad ka aritmeetilise keskmise jaoks.

Esmalt loome juhuslike suuruste summa ja korrutise mõiste kahe suuruse x ning y jaoks (suurema arvu suuruste jaoks oleks kõik analoogiline). Olgu antud jaotustabelid

| $x$ | $x_1$ ..... $x_m$ |
|---|---|
| $p$ | $p_1$ ..... $p_m$ |

| $y$ | $y_1$ ..... $y_n$ |
|---|---|
| $r$ | $r_1$ ..... $r_n$ |

Juhuslike suuruste summa on juhuslik suurus, mille võimalikeks väärtusteks on liidetavate kõigi võimalike väärtuste summad. Summa $x + y$ võimalikud väärtused oleksid näiteks $x_1 + y_1$, $x_1 + y_2$, $x_2 + y_1$ jne. Mingi väärtuse $x_k + y_l$ tõenäosuse leidmiseks paneme tähele, et $x_k + y_l$ esinemist võime vaadelda liitsündmusena, milloks on nii $x_k$ kui $y_l$ samaaegne esinemine. Siis on ju $x_k + y_l$ tõenäosus võrdne osasündmuste tõenäosuste korrutisega $p_k . r_l$ . Summa jaotustabeli esitame kujul

| $x+y$ | $x_1+y_1$ | $x_1+x_2$ | ... | $x_1+y_n$ | $x_2+y_1$ | ... | $x_m+y_n$ |
|---|---|---|---|---|---|---|---|
| $p$ | $p_1r_1$ | $p_1r_2$ | | $p_1r_n$ | $p_2r_1$ | | $p_mr_n$ |

1. näide: Koostada sõltumatute suuruste summa jaotus-

tabel, kui liidetavate jaotustabelid on

| x | 0 | 1 | 2 |
|---|---|---|---|
| p | 0,3 | 0,5 | 0,2 |

| y | -1 | 0 | 1 |
|---|---|---|---|
| r | 0,2 | 0,5 | 0,3 |

Summa x + y väärtuste saamiseks liidame kõik võimalikud x ja y väärtused. Tõenäosused saame vastavate p ja r korrutamisel. Jõuame tulemusele

| x+y | -1 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|
| p | 0,6 | 0,15 | 0,09 | 0,1 | 0,25 | 0,15 | 0,04 | 0,1 | 0,06 |

Väärtused 0, 1 ja 2 esinevad korduvalt. Väärtuse 0 tõenäosused on 0,15 ja 0,1. Tõenäosuste liitmislause põhjal on 0 tõenäosus 0,15 + 0,1 = 0,25. Toimides analoogiliselt väärtustega 1 ja 2 saame lõplikult

| x + y | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| p | 0,06 | 0,25 | 0,38 | 0,25 | 0,06 |

Sõltumatute juhuslike suuruste korrutis on juhuslik suurus, mille väärtusteks on tegurite kõigi võimalike väärtuste korrutised. Korrutise mingi väärtuse tõenäosuse leiame tegurite tõenäosuste korrutamisel (jälle korrutamislause!). Jaotustabeli koostame järgmisel põhimõttel

| xy | $x_1y_1$ | $x_1y_2$ | ... | $x_1y_n$ | $x_2y_1$ | ..... | $x_my_n$ |
|---|---|---|---|---|---|---|---|
| p | $p_1r_1$ | $p_1r_2$ | ... | $p_1r_n$ | $p_2r_1$ | ..... | $p_mr_n$ |

2. näide. Koostada eelmise näite andmetel suuruse xy jaotustabel.

Leiame tegurite kõik võimalikud korrutised ja vastavad tõenäosused. Tulemused esitame tabelis

| xy | 0 | 0 | 0 | -1 | 0 | 1 | -2 | 0 | 2 |
|---|---|---|---|---|---|---|---|---|---|
| p | 0,06 | 0,15 | 0,09 | 0,1 | 0,25 | 0,15 | 0,04 | 0,1 | 0,06 |

Mõne väärtuse kordumisel võime tabelit koondada. Väärtuse 0 tõenäosus on 0,06 + 0,15 + 0,09 + 0,25 + 0,1 = 0,65. Jaotustabel esitub lõpuks kujul

| xy | - 2 | - 1 | 0 | 1 | 2 |
|---|---|---|---|---|---|
| p | 0,04 | 0,1 | 0,65 | 0,15 | 0,06 |

Juhuslike suuruste korrutise erijuhuks on konstandi ja juhusliku suuruse korrutis C.x. Konstanti võime vaadelda suurusena, millel on üks väärtus tõenäosusega 1. Summa erijuhuks on konstandi ja juhusliku suuruse summa x + C. Vastavad jaotustabelid saame kujul

| cx | $cx_1$ | $cx_2$ ... $cx_n$ |
|---|---|---|
| p | $p_1$ | $p_2$ ... $p_n$ |

;

| x+c | $x_1$+c | $x_2$+ c ... | $x_n$+ c |
|---|---|---|---|
| | $p_1$ | $p_2$ | $p_n$ |

(Põhjendage, miks cx või x + c väärtuste tõenäsused on võrdsed x tõenäosustega!).

Asume keskväärtuse omaduste käsitlemisele.

I omadus. Konstandi keskväärtuseks on konstant ise. Tõepoolest,

$$E(c) = c.1 = c .$$

II omadus. Suuruste summa keskväärtus on võrdne liidetavate keskväärtuste summaga.

Kahe suuruse korral märgime seda matemaatilises sümboolikas

$$E(x+y) = E(x) + E(y) .$$

Näitame selle võrduse kehtivust lihtsal erijuhul (üldjuhul on mõttekäik analoogiline). Omagu suuruste jaotustabelid kuju

| x | $x_1$ | $x_2$ |
|---|---|---|
| p | $p_1$ | $p_2$ |

,

| y | $y_1$ | $y_2$ |
|---|---|---|
| r | $r_1$ | $r_2$ |

.

Summa $x + y$ jaotustabel esitub kujul

| $x + y$ | $x_1 + y_1$ | $x_1 + y_2$ | $x_2 + y_1$ | $x_2 + y_2$ |
|---|---|---|---|---|
| $p$ | $p_1r_1$ | $p_1r_1$ | $p_2r_1$ | $p_2r_2$ |

Leiame summa keskväärtuse $x + y$ jaotustabelist

$$E(x+y) = (x_1+y_1)\,p_1r_1 + (x_1+y_2)p_1r_2 + (x_2+y_1)p_2r_1 +$$

$$+ (x_2+y_2)p_2r_2 = x_1p_1(r_1+r_2) + x_2p_2(r_1+r_2) + y_1r_1(p_1+p_2) +$$

$$+ y_2r_2(p_1+p_2) = (x_1p_1+x_2p_2) + (y_1r_1+y_2r_2) = E(x) + E(y)\,.$$

(Miks $p_1 + p_2 = 1$ ja $r_1 + r_2 = 1$ ?).

III omadus. Sõltumatute suuruste korrutise keskväärtus on võrdne tegurite keskväärtuste korrutisega.

Kahe suuruse juhul

$$E(x.y) = E(x)\,.\,E(y)\,.$$

(Kontrollige omaduse kehtivust eelmise omaduse juures toodud erijuhu jaoks!)

Järeldused: 1. $E(c.x) = c.E(x)$

2. $E(c+x) = E(x) + c$

3. $E(Ax + B) = AE(x) + B$ .

Need tulenevad omadustest I - III (kontrollige!).

Suuruse väärtuse ja mingi jääva arvu $c$ vahet nimetatakse väärtuse hälbeks selle arvu suhtes. Suuruse $x$ hälvete $x - c$ jaotustabel on järgmine

| $x - c$ | $x_1 - c$ | $x_2 - c$ ..... | $x_n - c$ |
|---|---|---|---|
| $p$ | $p_1$ | $p_2$ | $p_n$ |

IV omadus. Suuruse keskväärtuse $a$ suhtes leitud hälvete keskväärtus on võrdne nulliga.

$$E(x-a)=0.$$

Tõepoolest,

$$E(x-a)=E(x)-E(a)=a-a=0.$$

See tähendab, et hälbed keskväärtuse suhtes keskmiselt tasakaalustuvad.

V omadus. Suuruse x ruudu keskväärtus on suurem sama suuruse keskväärtuse ruudust

$$E(x^2)>\left[E(x)\right]^2.$$

Kasutame E(x) asemel sümbolit a ja leiame hälvete ruudu keskväärtuse $E\left[(x-a)^2\right]$. See on positiivne, sest $(x-a)^2$ väärtused on kõik positiivsed. Teisendame seda avaldist ja kasutame lihtsustamiseks eelmisi omadusi.

$$E\left[(x-a)^2\right]=E(x^2-2ax+a^2)=E(x^2)-E(2ax)+E(a^2)=$$

$$=E(x^2)-2a\,E(x)+a^2=E(x^2)-2a^2+a^2=E(x^2)-a^2.$$

Kuid

$$E\left[(x-a)^2\right]=E(x^2)-a^2>0 \quad \text{ehk} \quad E(x^2)>\left[E(x)\right]^2.$$

## 5. Aritmeetilise keskmise arvutusvõtteid.

Pikkade variatsioonridade või suurte arvude korral võib aritmeetilise keskmise arvutamine kujuneda palju aega nõudvaks tööks. Kasutame keskväärtuse omadusi keskmise arvutamise lihtsustamiseks.

Tunnuse x väärtused võime esitada kujul

$$x=(x-b)+b,$$

kus b on meelevaldne arv. Otstarbekas on valida arvuks b variatsioonrea kõige suurema sagedusega väärtus või rea algusest ja lõpust võrdsel kaugusel seisev väärtus. Kui kõik vahed x - b omavad ühisjagaja k, siis võime x väärtused esitada veel kujul

$$x = \frac{x - b}{k} k + b .$$

Tähistame

$$\frac{x - b}{k} = z \; ; \text{ siis } x = kz + b .$$

Eelmise punkti 3. järelduse põhjal

$$E(x) = E(kz + b) = k\,E(z) + b$$

ehk

$$\bar{x} = k\,\bar{z} + b .$$

Oleme taandanud $\bar{x}$ leidmise $z = \frac{x - b}{k}$ keskmise leidmisele.

Aritmeetilise keskmise leidmise eeskirja kasutades saame valemi

$$\bar{x} = \frac{\sum \frac{x_i - b}{k} f_i}{N} k + b .$$

Kui sagedused $f_i$ omavad ühisjagajat l , siis võime kõiki sagedusi vähendada l korda. Keskmise väärtust see ei mõjuta, sest murru väärtus ei muutu tema lugeja ja nimetaja jagamisel sama arvuga.

Valem aritmeetilise keskmise arvutamiseks omandab kuju

$$\bar{x} = \frac{\sum \frac{x_i - b}{k} \cdot \frac{f_i}{l}}{\frac{N}{l}} k + b$$

Kui väärtusi ei soovita vähendada mingi arvu võrra, siis tuleb antud valemis võtta b = 0 ; väärtustel või sagedustel ühisjagaja puudumisel võtame vastavalt k = 1 või l = 1 .

Intervallitud variatsioonrea puhul võtame keskmise arvutamisel väärtusteks intervallide keskkohad. See tähendab, et intervalli piirides loeme tunnuse väärtuste jaotumise ühtlaseks. Seda tingib asjaolu, et harilikult pole olemas andmeid selle kohta, kuidas väärtused intervallisiseselt jaotuvad. Keskmise arvutuseeskirjas valitakse arvuks k harilikult intervalli pikkus või selle kordne.

Arvutusteks on kasulik koostada sobiv arvutusskeem.

1. näide. Leida tunnuse keskmine väärtus, kui variatsioonrida on järgmine:

| x | 10,2 | 10,4 | 10,6 | 10,8 | 11,0 | 11,2 | 11,4 | 11,6 | 11,8 | 12,0 |
|---|---|---|---|---|---|---|---|---|---|---|
| f | 2 | 3 | 8 | 13 | 25 | 20 | 12 | 10 | 6 | 1 |

Võtame keskmise valemis b = 11,0, k= 2 ja l = 1. Siis

$$\bar{x} = \frac{\sum \frac{x_i - 11}{2} f_i}{\sum f_i} \cdot 2 + 11 .$$

Arvutusteks koostame skeemi

| x | f | x - 11 | $\frac{x-11}{2}$ | $\frac{x-11}{2} f$ |
|---|---|---|---|---|
| 10,2 | 2 | - 0,8 | - 0,4 | - 0,8 |
| 10,4 | 3 | - 0,6 | - 0,3 | - 0,9 |
| 10,6 | 8 | - 0,4 | - 0,2 | - 1,6 |
| 10,8 | 13 | - 0,2 | - 0,1 | - 1,3 |
| 11,0 | 25 | 0 | | - 4,6 |
| 11,2 | 20 | 0,2 | 0,1 | 2,0 |
| 11,4 | 12 | 0,4 | 0,2 | 2,4 |
| 11,6 | 10 | 0,6 | 0,3 | 3,0 |
| 11,8 | 6 | 0,8 | 0,4 | 2,4 |
| 12,0 | 1 | 1,0 | 0,5 | 0,5 |
| | 100 | | | 10,3 |
| | | | | 5,7 |

Sellesse ritta, mis vastab X väärtusele 11,0, leiame negatiivsete arvude summa. Positiivsed arvud liidetakse eraldi ja siis on kerge leida keskmise avaldises murru lugejat

$$\bar{x} = \frac{5,7}{100} \cdot 2 + 11 = 11,1 .$$

2. näide. Leida päevanormi täitmise keskmine % saja töölise töötulemuste järgi (vaatlusandmed on tabeli kahes esimeses veerus).

| Plaani täitmise % | Tööliste arv f | x | x - 137,5 | $\frac{x-137,5}{25}$ | $\frac{x-137,5}{25} f$ |
|---|---|---|---|---|---|
| 50 - 75 | 2 | 62,5 | - 75 | - 3 | - 6 |
| 75 - 100 | 18 | 87,5 | - 50 | - 2 | - 36 |
| 100 - 125 | 32 | 112,5 | - 25 | - 1 | - 32 |
| 125 - 150 | 30 | 137,5 | | | - 74 |
| 150 - 175 | 11 | 162,5 | 25 | 1 | 11 |
| 175 - 200 | 6 | 187,5 | 50 | 2 | 12 |
| 200 - 225 | 1 | 212,5 | 75 | 3 | 3 |
| | 100 | | | | 26 |
| | | | | | - 48 |

$$\bar{x} = \frac{-48}{100} \cdot 25 + 137,5 = 125,5 \%.$$

## 6. Mediaan.

Korrastatud statistilise rea mediaaniks nimetatakse seda liiget, millest mõlemale poole jääb võrdne arv liikmeid.

Nii on reas 3 5 6 9 12 13 17 mediaaniks 9, sest temast jääb mõlemale poole 3 liiget.

Kui reas on liikmeid paaritu arv N, siis leiame mediaani (Me) järjekorranumbri valemist

$$Me = \frac{N+1}{2} ,$$

Reas 10 12 15 19 21 22 on liikmeid paarisarv ja sellist liiget polegi, millest mõlemale poole jääb täpselt võrdne arv liikmeid. Sellisel puhul võetakse mediaaniks rea kahe liikme aritmeetiline keskmine. Vaadeldavas näites on

$$Me = \frac{15 + 19}{2} = 17 .$$

Kui lähteandmed on esitatud grupeeritult momentreana või intervallitud variatsioonreana, siis leiame mediaani järgmise mõttekäiguga.

Jagame väärtuste koguarvu pooleks, mis määrab mediaani järjekorranumbri. Seejärel liidame sagedusi seni, kuni jõuame väärtuseni, mille kohal sageduste summa ületab esimesena $\frac{N}{2}$. Seega oleks mediaangrupp määratud. Momentrea korral ongi mediaaniks saadud väärtus. Intervallitud reas täpsustame mediaani asukohta intervalli sees lineaarse interpolatsiooni teel, lugedes väärtuste jaotuse ühtlaseks.

1. näide. Leida eelmise punkti 2.näite andmetel rea mediaan. Variantide sageduste summa on 100, seega võtame mediaaniks rea 50. liikme. Sageduste liitmisel leiame, et mediaan asub intervallis 100 - 125, sest seal 2 + 18 + + 32 > 50. Saadud intervallist tuleb võtta 30. liige (intervallist allapoole jääb 20 liiget). Intervalli 32 liikmele vastab intervalli pikkus 25, 30. liige asub algusest kaugusel $\frac{25}{32} \cdot 30 = 23,4.$

$$Me = 100 + 23,4 = 123,4.$$

Mediaani arvutamine kirjeldatud viisil toimub järgmise valemi järgi

$$Me = X_{Me} + \frac{k\left(\frac{N}{2} - L\right)}{f_{Me}}$$

$L$ - mediaanintervallile eelnevate intervallide sageduste summa,

$f_{Me}$ - mediaanintervalli sagedus,

$k$ - intervalli pikkus,

$X_{Me}$ - mediaanintervalli algus.

Momentrea korral võib aritmeetiliseks keskmiseks tulla väärtus, mis üldse ei ühti rea liikmetega, mediaaniks on ikka üks rea liikmetest (või mähemalt sellele väga lähedane). Aritmeetiline keskmine reageerib iga üksikväärtuse muutumisele, mediaanile ei avalda mingit mõju väärtuste muutumine mõlemal pool mediaani. Mediaan reageerib ainult liikmete arvu muutumisele.

## 7. M o o d .

Moodiks nimetatakse rea kõige suurema sagedusega liiget. Käesoleva paragrahvi 5. punkti 1. näites on rea moodiks 11,0. Momentrea puhul leiame moodi ilma arvutusteta. Intervallitud rea korral määrame esmalt moodintervalli (intervalli, milles mood asub) suurima sageduse järgi. Moodi asukoha täpsustamiseks intervallisiseselt loobume eeldusest, et väärtused jaotuvad intervalli ulatuses ühtlaselt (seda eeldasime $\bar{x}$ ja Me leidmisel). Moodintervalli naaberintervalli sageduste abil otsustame leitud jaotumise üle moodintervallis endas. Liikmed loeme tihedamalt paigutatuks moodintervalli selles osas, mille naaberintervalli sagedus on suurem. Moodi arvutame siis valemist

$$Mo = X_{Mo} + \frac{k \cdot f_{+1}}{f_{+1} + f_{-1}}, \text{ kus}$$

$X_{Mo}$ - moodintervalli alumine piir,

$k$ - intervalli pikkus,

$f_{+1}$, $f_{-1}$ - vastavalt moodintervalli järgneva ja eelneva intervalli sagedus.

1. näide. Leida 5. punkti 2. näite andmetel rea mood. Variatsioonreast näeme, et

$$x_{Mo} = 100, \quad k = 25, \quad f_{+1} = 30, \quad f_{-1} = 18.$$

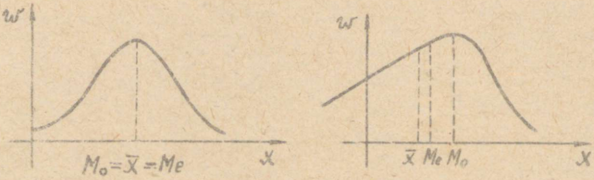
Seega

$$Mo = 100 + \frac{25 \cdot 30}{30 + 18} = 100 + \frac{750}{48} = 115,6 .$$

Väärtuste intervallisisest ühtlast jaotumist eeldades tuleks valida moodiks intervalli keskkoht.

Märkus: Moodi arvutamiseks kasutatakse ka teistsuguseid valemeid, kuid tulemuste erinevused ei ole olulised.

Sümmeetriliste jaotuste korral langevad aritmeetiline keskmine, mood ja mediaan kokku. Ebasümmeetrilise jaotuse korral on nad erinevad.




Joonis 12 .

Mood on võrdne pideva juhusliku suuruse jaotuskõvera kõrgeimale punktile vastava abstsissiga, mediaan - jaotuskõvera alla jäävat pindala poolitava ordinaatlõigu abstsissiga. Mediaan asub aritmeetilisest keskmisest umbes ühe kolmandiku moodi ja keskmise vahemaa kaugusel.

## 8. Dispersioon.

Juhusliku suuruse asendi karakteristikud peegeldavad suuruse väärtuste üldist nivood, kuid nendes ei kajastu väärtuste omavahelised erinevused, väärtuste jaotus üksteise suhtes.

Olgu näiteks juhuslikud suurused antud jaotustabelitega

| $x$ | $-0{,}01$ | $0{,}01$ |
|---|---|---|
| $p$ | $0{,}5$ | $0{,}5$ |

| $y$ | $-100$ | $100$ |
|---|---|---|
| $r$ | $0{,}5$ | $0{,}5$ |

Mõlema suuruse matemaatilised ootused on võrdsed

$$E(x) = -0{,}01 \cdot 0{,}5 + 0{,}01 \cdot 0{,}5 = 0;$$

$$E(y) = -100 \cdot 0{,}5 + 100 \cdot 0{,}5 = 0 ,$$

kuid väärtuste jaotuste iseloom on väga erinev. Suuruse $x$ väärtused asuvad keskväärtuse ligidal, suuruse $y$ väärtused kaugel. Teiste sõnadega: suuruse $x$ hälbed $E(x)$ suhtes on väikesed, suurusel $y$ suured.

Teades üksnes keskväärtust või ka moodi ja mediaani ei saa otsustada, missugused on suuruse võimalikud väärtused ja kuidas nad paigutuvad keskväärtuse suhtes. Tekib vajadus suuruse võimalike väärtuste hajuvust iseloomustava karakteristiku järgi.

Suuruse suurima ja väikseima väärtuse vahe ei iseloomusta kuigi hästi väärtuste hajuvust (kuigi seda mõnikord kasutatakse). Võib juhtuda, et suuruse väärtused jaotuvad enamuses tihedalt keskmise ümber, kuid üksikud väärtused jäävad keskmisest kaugele. Sellisel juhul annab vahe $|x_{max} - x_{min}|$ moonutatud pildi väärtuste jaotumisest. Oleks vaja karakteristikut, mis arvestaks kõigi väärtuste paiknemist. Keskmise suhtes leitud hälvete keskmine ei sobi sellepärast, et $E(x - a) = 0$, kuigi väärtused võivad asuda keskmisest kaugel. Hajuvuse mõõduna kasutatakse mõnikord keskväärtuse suhtes leitud hälvete absoluutväärtuste keskväärtust, mida nimetatakse keskmiseks lineaarhälbeks $d$. Eeskiri $d$ leidmiseks omab kuju

$$d = \sum p_i \cdot |x_i - a| .$$

Mida tihedamalt asuvad suuruse väärtused keskmise ümber, seda väiksem on ka keskmine lineaarhälve. Üldistes matemaatilistes arutlustes on absoluutväärtusi sisaldavate avaldistega tülikas tegelda. Hoopis mugavam on kasutada absoluutväärtuste asemel arvude ruute. Hajumise karakteristikutena kasutatakse kõige sagedamini dispersiooni ja keskmist ruuthälvet. Selgitame nende mõistete tähendust.

Olgu antud juhusliku suuruse jaotustabel. Leiame suuruse keskväärtuse $E(x) = a$ ning väärtuste hälbed $a$ suhtes. Seejärel leiame hälvete ruudud, mille jaotustabel omab kuju

| $(x-a)^2$ | $(x_1-a)^2$ | ..... | $(x_n-a)^2$ |
|---|---|---|---|
| $p$ | $p_1$ | ..... | $p_n$ |

.

Juhusliku suuruse dispersiooniks nimetatakse tema keskväärtuse suhtes arvutatud hälvete ruutude keskväärtust.

Dispersiooni sümbolina kasutame kas $\sigma^2$ või $D(x)$. Vastavalt definitsioonile näeb eeskiri $D(x)$ leidmiseks välja järgmiselt

$$D(x) = E\left[(x-a)^2\right] = \sum_{i=1}^{N}(x_i-a)^2 p_i \ .$$

1. näide. Leida suuruse $x$ keskmine lineaarhälve ja dispersioon, kui on antud jaotustabel

| $x$ | 1 | 2 | 5 |
|---|---|---|---|
| $p$ | 0,3 | 0,5 | 0,2 |

Esmalt leiame

$$a = 1 \cdot 0{,}3 + 2 \cdot 0{,}5 + 5 \cdot 0{,}2 = 2{,}3 \ .$$

Seejärel leiame hälvete absoluutväärtused ja hälvete ruudud. Nende jaotustabeliteks on

| $\lvert x-a \rvert$ | 1,3 | 0,3 | 2,7 |
|---|---|---|---|
| $p$ | 0,3 | 0,5 | 0,2 |

| $(x-a)^2$ | 1,69 | 0,09 | 7,29 |
|---|---|---|---|
| $p$ | 0,3 | 0,5 | 0,2 |

.

Keskmine lineaarhälve

$$d = 1{,}3 \cdot 0{,}3 + 0{,}3 \cdot 0{,}5 + 2{,}7 \cdot 0{,}2 = 1{,}08 ,$$

dispersioon

$$D(x) = 1{,}69 \cdot 0{,}3 + 0{,}09 \cdot 0{,}5 + 7{,}29 \cdot 0{,}2 = 2{,}01 .$$

Dispersiooni arvutamine lihtsustub, kui teisendada definitsioonile vastavat valemit. Matemaatilise ootuse 5. omaduse käsitlemisel saime seose

$$E\left[(x-a)^2\right] = E(x^2) - a^2 .$$

Seega oleme jõudnud tulemuseni

$$D(x) = \sigma^2 = E(x^2) - \left[E(x)\right]^2$$

- dispersioon on võrdne suuruse võimalike väärtuste ruutude keskväärtuse ja suuruse keskväärtuse ruudu vahega.

Kuna $E(x^2) > a^2$, siis on dispersioon alati positiivne.

2. näide. Arvutame eelmise näite andmetel dispersiooni teisendatud arvutusvalemit kasutades.

| $x$ | $p$ | $xp$ | $x^2$ | $x^2p$ |
|---|---|---|---|---|
| 1 | 0,3 | 0,3 | 1 | 0,3 |
| 2 | 0,5 | 1,0 | 4 | 2,0 |
| 5 | 0,2 | 1,0 | 25 | 5,0 |
| | | 2,3 | | 7,3 |

$$D(x) = 7{,}3 - 2{,}3^2 = 2{,}01 .$$

3. näide. Leida sündmuse toimumiste arvu dispersioon üksikkatsel.

Käesoleva paragrahvi 2. punktis koostasime vastava juhusliku suuruse jaotustabeli

| $x$ | 1 | 0 |
|---|---|---|
| $p$ | $p$ | $q$ |

ning leidsime $E(x) = p$.

Dispersioon

$$D(x) = E(x^2) - \left[E(x)\right]^2 = (1^2.p+0^2.q) - p^2 = p-p^2 =$$

$$= p(1-p) = pq .$$

Üksikkatsel võrdub sündmuse toimumiste arvu dispersioon sündmuse ja tema vastandsündmuse tõenäosuste korrutisega.

Positiivset ruutjuurt dispersioonist nimetatakse keskmiseks ruuthälbeks ehk standardhälbeks $\sigma$

$$\sigma = \sqrt{D(x)} .$$

Standardhälbe dimensioon ühtib suuruse dimensiooniga, dispersiooni dimensiooniks on suuruse dimensiooni ruut.

Mida suurem on dispersioon või standardhälve, seda rohkem on suuruse väärtused hajutatud keskväärtuse ümber.

## 9. Dispersiooni omadusi.

I omadus. Konstandi $c$ dispersioon on null.

Tõepoolest

$D(c) = E(c^2) - \left[E(c)\right]^2 = c^2 - c^2 = 0$ (vt. $E(x)$ omadusi).

Konstant omab ainult ühe väärtuse ja mingit hajumist ei ole.

II omadus. Kui suuruse võimalikud väärtused omavad ühise teguri c, st. $y = cx$, siis võib selle teguri tuua dispersiooni märgi ette ruutu tõstetuna

$$D(cx) = c^2D(x) .$$

Omaduse õigsuses veendumiseks lähtume dispersiooni definitsioonist ja võtame arvesse, et $E(cx) = cE(x) = ca$ .

Siis

$$D(cx) = \sum(cx_i-E(cx))^2p_i = \sum(cx_i-ca)^2p_i =$$

$$=\sum c^2(x_i-a)^2p_i = c^2\sum(x_i-a)^2p_i = c^2.D(x) .$$

III omadus. Sõltumatute suuruste summa dispersioon on võrdne liidetavate dispersioonide summaga

$$D(x+y) = D(x) + D(y) .$$

Kui tähistada $E(x) = a_1$ , $E(y) = a_2$ ,

siis $E(x+y) = a_1+a_2$ ning keskväärtuse omaduste põhjal

$$D(x+y) = E\left\{\left[(x+y) - (a_1+a_2)\right]^2\right\} = E\left\{\left[(x-a_1) + (y-a_2)\right]^2\right\} = E\left\{(x-a_1)^2 + 2(x-a_1)(y-a_2) + (y-a_2)^2\right\} =$$

$$= E\left[(x-a_1)^2\right] + E\left[(y-a_2)^2\right] + 2E\left[(x-a_1)(y-a_2)\right] .$$

Kaks esimest liidetavat annavad vastavalt $D(x)$ ja $D(y)$ , viimane liidetav on null (põhjendada!). Seega

$$D(x+y) = D(x) + D(y) .$$

Järeldused:

Kahe suuruse vahe dispersioon võrdub suuruste dispersioonide summaga

$$D(x-y) = D(x) + D(y) \qquad \text{( II ja III omaduse põhjal).}$$

2. Kui $k$ ja $b$ on konstandid, siis

$$D(kx+b) = k^2D(x) \qquad \text{(põhjendada!).}$$

Juhusliku suuruse lineaaravaldise dispersioon võrdub suuruse dispersiooni ja suuruse kordaja ruudu korrutisega.

1. näide. Leida suuruse y dispersioon

| y | 0,2 | 0,4 | 0,8 |
|---|---|---|---|
| p | 0,4 | 0,4 | 0,2 |

Väärtustel on ühine tegur 2. Suuruse y võime esitada kujul $y = 2x$, kus suuruse x jaotustabeliks on

| x | 0,1 | 0,2 | 0,4 |
|---|---|---|---|
| p | 0,4 | 0,4 | 0,2 |

Siis

| x | p | xp | $x^2$ | $x^2p$ |
|---|---|---|---|---|
| 0,1 | 0,4 | 0,04 | 0,01 | 0,004 |
| 0,2 | 0,4 | 0,08 | 0,04 | 0,016 |
| 0,4 | 0,2 | 0,08 | 0,16 | 0,052 |
| | | 0,2 | | 0,072 |

$D(x) = 0{,}072 - 0{,}2^2 = 0{,}032$ ja $D(y) = 2^2 \cdot 0{,}032 = 0{,}128.$

2. näide. Leida suuruste x ja y summa dispersioon, kui suurused on antud jaotustabelitega

| x | 0 | 1 | 2 |
|---|---|---|---|
| p | 0,3 | 0.5 | 0,2 |

| y | -1 | 0 | 1 |
|---|---|---|---|
| r | 0,2 | 0,5 | 0,3 |

Käesoleva paragrahvi 4. punkti 1. näites on x + y jaotustabel koostatud

| x + y | - 1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| p | 0,06 | 0,25 | 0,38 | 0,25 | 0,06 |

Leiame dispersiooni kahel viisil. Arvutuste käik nähtub tabelist.

| $x+y$ | $p$ | $(x+y)\cdot p$ | $(x+y)^2$ | $(x+y)^2\cdot p$ |
|---|---|---|---|---|
| - 1 | 0,06 | - 0,06 | 1 | 0,06 |
| 0 | 0,25 | 0,00 | 0 | 0,00 |
| 1 | 0,38 | 0,38 | 1 | 0,38 |
| 2 | 0,25 | 0,50 | 4 | 1,00 |
| 3 | 0,06 | 0,18 | 9 | 0,54 |
| | | 1,00 | | 1,98 |

$$D(x+y) = 1,98 - 1^2 = 0,98 .$$

| $x$ | $p$ | $x^2$ | $xp$ | $x^2p$ | $y$ | $r$ | $ry$ | $y^2$ | $y^2r$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,3 | 0 | 0 | 0 | -1 | 0,2 | -0,2 | 1 | 0,2 |
| 1 | 0,5 | 1 | 0,5 | 0,5 | 0 | 0,5 | 0,0 | 0 | 0,0 |
| 2 | 0,2 | 4 | 0,4 | 0,8 | 1 | 0,3 | 0,3 | 1 | 0,3 |
| | | | 0,9 | 1,3 | | | 0,1 | | 0,5 |

$$D(x) = 1,3 - 0,9^2 = 0,49 , \quad D(y) = 0,5 - 0,01^2 = 0,49 ;$$

$$D(x) + D(y) = 0,98 = D(x + y) .$$

## 10. Dispersiooni leidmine variatsioonrea andmetel.

Empiiriliste andmete puhul kasutame dispersiooni sümbolina $\sigma^2$ asemel $s^2$.

Variatsioonrea korral saame dispersiooni definitsiooni kohaselt arvutuseeskirjaks

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \sum (x_i - \bar{x})^2 \cdot w_i$$

Standardhälbe $s$ leiame $s^2$ kaudu

$$s = \sqrt{s^2}$$

Dispersiooni teisendatud arvutuseeskiri esitub kujul

$$s^2 = E(x^2) - \left[E(x)\right]^2 = \overline{x^2} - \bar{x}^2$$

- dispersioon võrdub väärtuste ruutude keskmise ja keskmise ruudu vahega.

Intervallitud variatsioonreas valime tunnuse väärtusteks intervallide keskkohad.

Dispersiooni $s^2$ arvutamine võib valmistada suurte arvude puhul tehnilisi raskusi. Sellepärast antakse $s^2$ leidmise eeskirjale tihti teisendatud kuju. Valemit teisendame nii, et saame ära kasutada ka $\bar{x}$ leidmisel teostatud arvutusi. Tuginedes $\bar{x}$ leidmise eeskirjale saame

$$s^2 = \frac{\sum \left(\frac{x_i - b}{k}\right)^2 \frac{f_i}{l}}{\sum \frac{f_i}{l}} \cdot k^2 - (\bar{x} - b)^2$$

Sümbolite $b$, $k$, $l$ tähendus on sama, mis aritmeetilise keskmise leidmisel.

Konkreetse arvutuseeskirja valik oleneb lähteandmete hulgast ja üksikväärtuste suurusest.

1. näide. Leida standardhälve 3. punkti 1. näite andmetel.

Päevas koristati keskmiselt $\bar{x} = 25$ ha. Dispersioon, kui hälvete ruutude keskmine

$$s^2 = \frac{(22-25)^2 \cdot 6 + 1^2 \cdot 5 + 1^2 \cdot 7 + 2^2 \cdot 8}{34} = \frac{98}{34} = 2{,}88(\text{ha}^2)$$

Standardhälve

$$s = \sqrt{2{,}88} = 1{,}7(\text{ha}).$$

2. näide. Leida dispersioon ja standardhälve 3. punkti 2. näite andmetel (automaatpingil toodetud detailide standardsuse kontroll).

Esitame arvutused ühes tabelis, kusjuures puuduvad lahtrid $\sum p.x$ ja $\sum f.x$ leidmiseks, sest keskmine on juba saadud.

| x | p | f | $x^2$ | $px^2$ | $fx^2$ |
|---|---|---|---|---|---|
| 0 | 0,24 | 11 | 0 | 0 | 0 |
| 1 | 0,38 | 20 | 1 | 0,38 | 20 |
| 2 | 0,26 | 12 | 4 | 1,04 | 48 |
| 3 | 0,10 | 5 | 9 | 0,90 | 45 |
| 4 | 0,02 | 2 | 16 | 0,32 | 32 |
| | | 50 | | 2,64 | 145 |

$$\sigma^2 = 2,64 - 1,28^2 = 1 ; \quad \sigma = \sqrt{1} = 1 .$$

$$s^2 = \frac{145}{50} - 1,34^2 = 1,1 ; \quad s = 1,05.$$

3. näide. Aastatel 1889 - 1890 mõõdeti Moskva 1000 meestöölise pikkus (mõõtmisandmed on tabeli kahes esimeses lahtris). Leida tööliste pikkuse keskmine ja standardhälve.

Valime $k = 3$, $b = 165,5$, $l = 1$.

| x | f | $\frac{x-165,5}{3}$ | $\left(\frac{x-165,5}{3}\right)^2$ | $\frac{x-165,5}{3}.f$ | $\left(\frac{x-165,5}{3}\right)^2.f$ |
|---|---|---|---|---|---|
| 143 - 146 | 1 | - 7 | 49 | - 7 | 49 |
| 146 - 149 | 2 | - 6 | 36 | -12 | 72 |
| 149 - 152 | 8 | - 5 | 25 | -40 | 200 |
| 152 - 155 | 26 | - 4 | 16 | -104 | 416 |
| 155 - 158 | 65 | - 3 | 9 | -195 | 585 |
| 158 - 161 | 120 | - 2 | 4 | -240 | 480 |
| 161 - 164 | 181 | - 1 | 1 | -181 | 181 |
| 164 - 167 | 201 | | | -779 | |
| 167 - 170 | 170 | 1 | 1 | 170 | 170 |
| 170 - 173 | 120 | 2 | 4 | 240 | 480 |

| | | | | | |
|---|---|---|---|---|---|
| 173 - 176 | 64 | 3 | 9 | 192 | 576 |
| 176 - 179 | 28 | 4 | 16 | 112 | 448 |
| 179 - 182 | 10 | 5 | 25 | 50 | 250 |
| 182 - 185 | 3 | 6 | 36 | 18 | 108 |
| 185 - 188 | 1 | 7 | 49 | 7 | 49 |
| | 1000 | | | 789 | 4064 |
| | | | | 10 | |

Leiame

$$\bar{x} = b + \frac{\sum \frac{x_i - b}{k} f_i}{N} k = 165,5 + \frac{10.3}{1000} = 165,53 \text{ cm};$$

$$s^2 = \frac{\sum \left(\frac{x_i - b}{k}\right)^2 . f_i}{N} k^2 - (\bar{x} - b)^2 = \frac{4064}{1000} .9-$$

$$-(165,53-165,5)^2 \approx 36,58;$$

$$s = \sqrt{36,58} = 6,05 \text{ cm}.$$

Mida väiksem on standardhälve s, seda tihedamalt jaotuvad suuruse väärtused aritmeetilise keskmise ümber, kusjuures suuruse väärtuste hälbed keskmise suhtes enamikul juhtudel ei ületa kolmekordset standardhälvet (põhjendame hiljem). Hinnake seni lahendatud näidete puhul, kas ja kui palju väärtusi jääb väljapoole tõkkeid $\bar{x} \pm 3s$.

## § 4. JUHUSLIKU SUURUSE JAOTUSI.

### 1. K o r d u v a t e   k a t s e t e   s k e e m.

1. näide. Heal laskuril on märklaua südamiku tabamise tõenäosus iga lasu korral 0,9. Kui tõenäone on tabada südamikku (saada 10 silma) 8 korda 10-st lasust?

Iga lasu tulemus ei sõltu eelmiste laskude tulemustest. Iga lasu puhul kas tabatakse 10 (sündmus A) või ei tabata (sündmus $\bar{A}$), kolmandat võimalust ei ole.

Ülesande lahendame hiljem.

2. näide. Urnis on valged ja mustad kuulid, kusjuures valgete kuulide arvu ja kuulide üldarvu suhe on p. Urnist võetakse ilma valimata kuul, registreeritakse tema värv ja pannakse tagasi. Kuuli võtmine on üksikkatse, mida võime korrata vajalik arv kordi. Pärast kuuli tagasipanekut segame kuulid uuesti. Katsete sellise korralduse juures ei sõltu järgmise katse tulemus eelmistest katsetest ja igal katsel on valge kuuli saamise tõenäosus p.

Urnist kuulide võtmise põhimõttel toimub ka sama liiki toodete kvaliteedi kontroll: toode kas vastab nõutud tingimustele või ei vasta.

Muutumatutes katsetingimustes sooritatavaid ühetüübilisi katseid, kus ühegi katse tulemus ei sõltu eelmistest katsetest, nimetatakse sõltumatuteks katseteks. Kui katsete kordamisel huvitume kogu aeg ainult sellest, kas katsel toimub teatud sündmus A või ei toimu (s.t. toimub sündmus $\bar{A}$ ), siis räägime korduvate katsete skeemist.

### 2. B e r n o u l l i   v a l e m.

Tihti tuleb otsida vastust sellistele küsimustele, nagu näiteks: kui tõenäone (kui usutav, põhjendatud) on, et piimakombinaadis vooluliinilt tulevast 1000 korgitud pudelist on 950 vigastamata?

kauplusse toodavast 10.000 kingapaarist on vähemalt 9900 paari defektideta?

linnaliinidel kurseerivast 50 bussist ei lange päeva jooksul rivist välja rohkem kui 6 bussi?

urnist kuulide võtmisel n korda saame valge kuuli m korral? jne.

Probleemi üldine seade on järgmine: kui suur on tõenäosus sündmuse A m-kordseks toimumiseks n sõltumatul katsel kui A tõenäosus üksikkatsel on p?

Alustame väikesest katsete arvust ja selgitame, missugused võimalused esinevad.

Ühel katsel on kaks võimalust: toimub A või $\bar{A}$ tõenäosustega vastavalt p või q.

Kahel katsel on järgmised võimalused: 1) mõlemal katsel toimub A; 2) esimesel katsel toimub A ja teisel $\bar{A}$; 3) esimesel katsel toimub $\bar{A}$ ja teisel A; 4) mõlemal katsel toimub $\bar{A}$. Loetletud võimalused on liitsündmused. Esimene võimalus kujutab liitsündmust, mis seisneb osasündmuse A kaks korda toimumises. Analoogiliselt tõlgendame ka teisi võimalusi. Iga liitsündmus on osasündmuste korrutis ja me tähistame neid vastavalt varasemale kokkuleppele 1) AA, 2) $A\bar{A}$, 3) $\bar{A}A$, 4) $\bar{A}\bar{A}$. Nende tõenäosused leiame A ja $\bar{A}$ tõenäosuste kaudu korrutamislause abil. Näiteks sündmuse $\bar{A}A$ tõenäosus on qp. Esitame võimalikud liitsündmused ja nende tõenäosused tabelina

| | | | |
|---|---|---|---|
| $AA$ | $A\bar{A}$ | $\bar{A}A$ | $\bar{A}\bar{A}$ |
| $p \cdot p$ | $p \cdot q$ | $q \cdot p$ | $q \cdot q$ |

Kolme katse puhul arutleme samuti kui kahe katse puhul ja jõuame järgmiste võimalike liitsündmusteni

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $AAA$ | $AA\bar{A}$ | $A\bar{A}A$ | $\bar{A}AA$ | $A\bar{A}\bar{A}$ | $\bar{A}A\bar{A}$ | $\bar{A}\bar{A}A$ | $\bar{A}\bar{A}\bar{A}$ |
| $p^3$ | $p^2q$ | $p^2q$ | $p^2q$ | $pq^2$ | $pq^2$ | $pq^2$ | $q^3$ |

Tõenäosuse sündmuse A ühekordseks toimumiseks kahel katsel leiame tõenäosuste liitmislause abil. Sündmus A toimub üks kord, kui ükskõik kumb liitsündmustest $A\bar{A}$ või $\bar{A}A$ realiseerub. Tähistades tõenäosuse sündmuse ühekordseks toimumiseks kahel katsel sümboliga $P_{2,1}$ (esimene indeks näitab katsete, teine sündmuse A toimumiste arvu, saame $P_{2,1} = pq + qp = 2pq$. Analoogiliselt leiame tõenäosuse A toimumiseks 3 katsel vastavalt

3 korda $P_{3,3} = p^3$

2 korda $P_{3,2} = p^2q + p^2q + p^2q = 3\,p^2q$

1 kord $P_{3,1} = pq^2 + pq^2 + pq^2 = 3pq^2$

0 korda $P_{3,0} = q^3$

Kui võrrelda leitud tõenäosusi Newtoni binoomi $(p + q)^3$ arendiga

$$(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3 ,$$

siis näeme, et arendi 1. liige on võrdne tõenäosusega $P_{3,3}$; 2. liige - tõenäosusega $P_{3,2}$ .ne.

Kogu eelnevat mõttekäiku võib üle kanda ka 4,5 ja enama katse juhule, üldiselt n katse juhule.

Tõenäosus sündmuse A m-kordseks toimumiseks n katsel $P_{n,m}$ on võrdne Newtoni binoomi $(p + q)^n$ arendi $(n - m + 1)$-se liikmega

$$P_{n,m} = C_n^m \; p^m \; q^{n-m} .$$

See võrdus kannab Bernoulli valemi nime.

Sümbol $C_n^m$ tähendab kombinatsioonide arvu n elemendist m kaupa ja seda võime leida valemist

$$C_n^m = \frac{n!}{m!\,(n - m)!} ,$$

kusjuures lepime kokku lugeda $0! = 1$ .

Seega

$$P_{n,m} = \frac{n!}{m!\,(n-m)!}\; p^n q^{n-m}$$

1. **näide.** Kui suur on tõenäosus sündmuse 3-kordseks toimumiseks neljal katsel, kui sündmuse tõenäosus üksikkatsel $p = \frac{2}{3}$ ? Antud juhul $n = 4$, $m = 3$, $p = \frac{2}{3}$ ja $q = 1 - p = \frac{1}{3}$ . Paigutame vastavad väärtused Bernoulli valemisse

$$P_{4,2} = \frac{4!}{3!\,(4-3)!} \left(\frac{2}{3}\right)^3 \cdot \frac{1}{3} = \frac{4.3.2}{3.2.1} \cdot \frac{8}{27} \cdot \frac{1}{3} = \frac{32}{81}.$$

2. **näide.** Lahendame eelmise punkti 1. näites tõstatatud probleemi. Laskude arv $n = 10$, märklaua südamiku tabamiste (sündmus A) arv $m = 8$ ja $p = 0,9$ . Siis $q = 1 - 0,9 = 0,1$ ning

$$P_{10,8} = \frac{10!}{8!\,(10-8)!}\, 0,9^8 \cdot 0,1^2 = \frac{10 \cdot 9}{2} \cdot 0,9^8 \cdot \frac{1}{100} =$$

$$= \frac{1}{2} \cdot 0,9^9 \approx 0,194$$

Juba käesolev näide näitab, et suuremate arvude puhul põrkame kokku arvutuslike raskustega.

3. **näide.** Tõenäosus ööpäevaks ettenähtud elektrienergia normi mitteületamiseks $p = 0,75$. Kui tõenäone on, et 6 päevast 4 päeval ei ületata normi?

Ülesande andmetest näeme, et

$$p = 0,75 \;,\; n = 6 \;,\; m = 4 \;\text{ ja }\; q = 1-p = 0,25 \,.$$

Otsitav tõenäosus on

$$P_{6,4} = \frac{6!}{4!(6-4)!}\, 0,75^4 \cdot 0,25^2 = \frac{6.5}{2} \cdot 0,75^4 \cdot 0,25^2 = 0,30$$

4. näide. Mis on tõenäosem: kas võita võrdset vastast 5 korda kaheksast mängust või 3 korda neljast mängust?

Võrdse vastasega mängides on mõlemal poolel võidu tõenäosus $p = 0,5$ ja ka $q = 1-p = 0,5$.

$$P_{8,5} = \frac{8!}{5!\,(8-5)!} \cdot (\tfrac{1}{2})^5 \cdot (\tfrac{1}{2})^3 = \frac{7}{32}$$

$$P_{4,3} = \frac{4!}{3!\,(4-3)!} (\tfrac{1}{2})^3 \cdot \frac{1}{2} = \frac{1}{4} .$$

Tõenäosem on võita 3 mängu neljast.

## 3. Binomiaaljaotus.

Sündmuse A toimumiste arvu m nimetatakse sündmuse sageduseks n katsel. Sagedus m on juhuslik suurus. Bernoulli valemi abil koostame sageduste tõenäosuste $P_{n,m}$ jaotustabeli.

| $m$ | 0 | 1 | 2 ... | $m$ ... | $n$ |
|---|---|---|---|---|---|
| $P_{n,m}$ | $q^n$ | $C_n^1\, pq^{n-1}$ | $C_n^2 p^2 q^{n-2}$ | $C_n^m p^m q^{n-m}$ | $p^n$ |

Tõenäosused jaotustabelis on võrdsed Newtoni binoomi arendi liikmetega ja sellepärast kannabki koostatud jaotustabel binomiaalse jaotustabeli nime.

Korduvatel katsetel allub sündmuse A sagedus m binomiaaljaotusele.

1. näide. Koostada sageduse m jaotustabel, kui $p = 0,4$ ja $n = 5$. Sagedus m võib omandada väärtusi 0,1,2,3,4,5, mille tõenäosused leiame Bernoulli valemist

$$P_{5,m} = C_5^m\ 0,4^m \cdot 0,6^{5-m} .$$

Jaotustabel esitub kujul

| $m$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P_{5,m}$ | 0,078 | 0,259 | 0,346 | 0,230 | 0,077 | 0,010 |

2. näide. Koostada sageduse m jaotustabel, kui $p = 0,5$ ja $n = 5$. Käesoleval juhul on ka $q = 0,5$ ja avaldises

$$P_{5,m} = C_5^m \ 0,5^m \ . \ 0,5^{5-m}$$

on iga m korral $0,5^m \ . \ 0,5^{5-m} = 0,5^5 = 0,03125$.

Tõenäosused leiame korrutades $0,5^5$ kombinatsioonide arvuga 5 elemendist m kaupa. Jõuame jaotustabelini

| $m$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P_{5,m}$ | 0,03125 | 0,15625 | 0,3125 | 0,3125 | 0,15625 | 0,03125 |

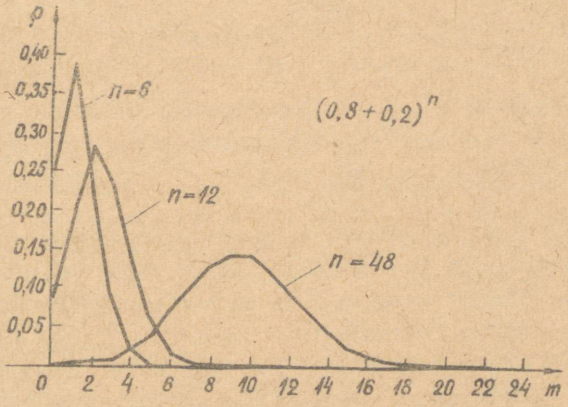
Kahes viimases näites koostatud jaotustabelite võrdlemisel näeme, et teisel juhul on tõenäosuste jaotus sümmeetriline, esimesel juhul ebasümmeetriline. Kui $p = q = 0,5$, säilib sümmeetria iga n korral, $p \neq q$ korral on binomiaaljaotus ebasümmeetriline. Kuid isegi teineteisest erinevate p ja q korral katsete arvu n suurenemisel läheneb jaotustabel sümmeetrilisele. Hästi võib seda näha polügoonidel, mis on koostatud juhuks kui $p = 0,8$, s.t. on leitud binoomi $(0,8 + 0,2)^n$ arendi liikmed kolmel n väärtusel (joonis 13). Katsete arvu kasvamisel läheneb polügoon sujuvale kõverale.

Käsitleme mõningaid binomiaaljaotuse omadusi tuginedes eelmistele näidetele ja joonisele 13.

I. Sageduse tõenäosused $P_{n,m}$ üldiselt kasvavad maksimaalse väärtuseni ja seejärel jälle kahanevad.

II. Katsete arvu n kasvades sama p korral tõenäosuse $P_{n,m}$ väärtused vähenevad. See on hästi näha joonisel 13. Sellele järeldusele võib jõuda ka järgmise arutlusega. Alati on kõigi $P_{n,m}$ summa võrdne ühega (jaotustabeli koostamise printsiip!). Kui katsete arv aga suureneb, siis kasvab ka sageduse m võimalike väärtuste arv ja iga sageduse esinemise võimalused vähenevad.

III. Katsete arvu n kasvades läheneb binomiaaljaotuse polügoon pidevale kõverale, kusjuures polügooni ja abstsisstelje vahelise kujundi pindala on ikka võrdne ühega (seleta, miks!).




Joonis 13.

Leiame binomiaaljaotuse karakteristikud. Sündmuse toimumise sagedust üksikkatsel tõlgendasime 3. paragrahvi 2. punkti 2. näites juhusliku suurusena X, mille karakteristikud olid järgmised

$$E(x) = p; \qquad D(x) = pq.$$

Kui katsete arv on n, siis saame sageduse m, liites n suurust X

$$m = X_1 + X_2 + \dots X_n \quad .$$

Indeks tähe X juures näitab, mitmendale katsele vastavat juhuslikku suurust X vaatleme. Kõigi $X_i$ jaotused on ühesugused ja seega on nende keskväärtused ning dispersioonid võrdsed. Keskväärtuse ja dispersiooni omaduste kohaselt

$$E(m) = np; \quad D(m) = n.pq ; \quad \sigma(m) = \sqrt{npq}$$

- konstantse $p$ korral katsete arvu $n$ kasvades kasvavad ka sageduse keskväärtus, dispersioon ja standardhälve.

Jagades sageduse $m$ võimalikud väärtused katsete üldarvuga $n$ saame relatiivse sageduse $\frac{m}{n}$. Ka relatiivne sagedus on juhuslik suurus; tema keskväärtus

$$E\left(\frac{m}{n}\right) = \frac{1}{n} E(m) = \frac{1}{n} \cdot np = p \,;$$

dispersioon

$$D\left(\frac{m}{n}\right) = \frac{1}{n^2} D(m) = \frac{1}{n^2} npq = \frac{pq}{n} \,; \qquad \sigma\left(\frac{m}{n}\right) = \sqrt{\frac{pq}{n}}$$

- binomiaaljaotusel relatiivse sageduse keskväärtus võrdub sündmuse tõenäosusega üksikkatsel ning ei sõltu katsete arvust; hajuvus väheneb katsete arvu kasvamisel (dispersiooni avaldises on $n$ murru nimetajas).

Sageduse $m$ seda väärtust, mille tõenäosus on suurim, nimetatakse tõenäoseimaks sageduseks. Tõenäoseim sagedus $\mu$ on binomiaaljaotuse mood. Kuna binomiaaljaotus läheneb sümmeetrilisele jaotusele, siis läheneb ka mood keskväärtusele. Tõenäoseim sagedus on võrdne $E(m)$-ga

$$\mu = np \,,$$

kui see korrutis on täisarv. Kui aga keskväärtus ei ole täisarv, siis tuleb tõenäoseimaks sageduseks võtta tõkete $np - q$ ja $np + p$ vahel olev täisarv, s.t.

$$np - q < \mu < np + p.$$

Olles leidnud $\mu$ saame Bernoulli valemist leida ka $P_{n,\mu}$

1. näide. Leida tõenäoseim sagedus, kui $p = 0{,}4$ ja $n = 5$. Leiame $np = 5 \cdot 0{,}4 = 2$ ja kuna $np$ on praegu täisarv, siis $\mu = 2$. Tulemuse võrdlemisel eelmise punkti 1. näite tulemustega näeme, et tõepoolest $\mu = 2$ tõenäosus on suurim.

2. näide. Leida tõenäoseim märgi tabamiste arv üheksast lasust, kui üksiklasuga tabamise tõenäosus on $p = 0,8$.

Katsete arv $n = 9$, $p = 0,8$ ja $q = 1 - 0,8 = 0,2$. Korrutis $np = 7,2$, s.t. ei ole täisarv ja seepärast leiame

$$np + p = 7,2 + 0,8 = 8$$
$$np - q = 7,2 - 0,2 = 7.$$

Tõenäoseimad sagedused on 7 ja 8 võrdse tõenäosusega.

3. näide. Praakdetaili tootmise tõenäosus on 0,035. Leida tõenäoseim praagi hulk 500 detaili tootmisel.

Nüüd on $p = 0,035$; $n = 500$ ja $n \cdot p = 500 \cdot 0,035 = 17,5$. Tõenäoseim praagi hulk asub tõkete $np + p = 17,5 + 0,035 = 17,535$ ja $np - q = 17,5 - 0,966 = 16,535$ vahel. Järelikult $\mu = 17$. Leiame, kui tõenäone on $\mu = 17$ praakdetaili saamine.

Tuleb arvutada

$$P_{500,17} = \frac{500!}{17!\,483!} = 0,035^{17} \cdot 0,965^{483} ,$$

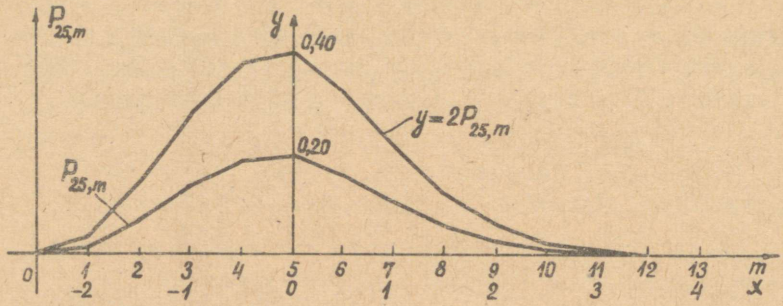
mis valmistab tosiseid tehnilisi raskusi.

## 4. Ligikaudne valem tõenäosuste $P_{n,m}$ leidmiseks.

Suure katsete arvu p korral on tõenäosuste $P_{n,m}$ arvutamine Bernoulli valemi abil väga tülikas. Püüame leida lihtsama mooduse vastava tõenäosuse leidmiseks. Kasutame selleks ka joonisel 13 esitatud polügoone. Näeme, et katsete arvu suurenemisel omandab polügoon järjest korrapärasema kuju. Polügooni haripunkt vastab abstsissile $E(m) = np$ ning polügooni sümmeetriateljeks kujuneb sagedusele $\mu = np$ vastav ordinaatlõik $P_{n,\mu}$. Otstarbekas on valida uus koordinaatide alguspunkt punkti, mille koordinaadid on esialgse teljestiku suhtes $(np, 0)$ (vt. joonis 14).

Tähistame punktide abstsisse uues teljestikus sümboliga $m'$. Üleminekuseos on järgmine

$$m' = m - np .$$




Joonis 14.

Suurte n väärtuste puhul polügooni alus venib järjest pikemaks ja ordinaadid vähenevad. Muudame mastaapi, vähendades kujutusühikut abstsissteljel $\sigma = \sqrt{npq}$ korda ja suurendades seda ordinaatteljel sama arv korda. Siis on punktide m' uued abstsissid

$$x = \frac{m'}{\sqrt{npq}} = \frac{m - np}{\sqrt{npq}}$$

ja sageduste m tõenäosusi kujutavate ordinaatlõikude $P_{n,m}$ asemel lõigud pikkusega

$$y = \sqrt{npq} \cdot P_{n,m} \cdot$$

Seega oleme polügooni ühes suunas kokku surunud ja teises suunas välja venitanud sama arv korda ning polügooni ja abstsisstelje vaheline pindala ei muutunud.

Joonisel 14 on sageduste m tõenäosuste teisendatud polügoon lähteandmetega n = 25; p = 0,2; q = 0,8. Vastavad arvulised väärtused leiame valemitest

$$x = \frac{m - np}{\sqrt{npq}} = \frac{m - 25 \cdot 0,2}{\sqrt{25 \cdot 0,2 \cdot 0,8}} = \frac{m - 5}{2};$$

$$y = \sqrt{npq}\ P_{n,m} = 2P_{25,m}$$

ja on toodud järgmises tabelis

| $m$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $x$ | - 2,5 | - 2,0 | - 1,5 | - 1 | - 0,5 | 0 |
| $P_{25,m}$ | 0,005 | 0,02 | 0,07 | 0,14 | 0,19 | 0,20 |
| $y$ | 0,01 | 0,04 | 0,014 | 0,28 | 0,38 | 0,40 |

| $m$ | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|
| $x$ | 0,5 | 1,0 | 1,5 | 2,0 | 2,5 | 3 |
| $P_{25,m}$ | 0,16 | 0,11 | 0,06 | 0,03 | 0,01 | 0,005 |
| $y$ | 0,32 | 0,22 | 0,12 | 0,06 | 0,02 | 0,01 |

Tabelis ei ole antud sagedusi alates 12-st, sest nende tõenäosused on arvutamise ja samuti joonistamise täpsuse piirides praktiliselt nullid.

Sageduse m asemel on uueks juhuslikuks suuruseks x, mis avaldub binomiaalse jaotuse karakteristikute kaudu

$$x = \frac{m - E(m)}{\sigma(m)} \ .$$

Vahed $m' = m - E(m)$ on sageduse m võimalike väärtuste hälbed nende keskväärtuse suhtes. Hälbed $m'$ on jagatud sageduse standardhälbega. Suurust x nimetatakse sellepärast ka sageduse m normeeritud hälbeks keskväärtuse $E(m)$ suhtes.

Laplace'i lokaalteoreem väidab, et katsete arvu n tõkestamatul kasvamisel lähenevad suuruse $\sqrt{npq} \cdot P_{n,m}$ väärtused piiramatult funktsiooni

$$\varphi(x) = \frac{1}{2} e^{-\frac{x^2}{2}}$$

väärtustele, kus

$$x = \frac{m - np}{\sqrt{npq}}$$

katsete arvu n tõkestamatul kasvamisel läheneb sageduse m tõenäosuste polügoon piiramatult kõverale

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

Matemaatilises sümboolikas võime seda esitada kujul

$$\lim_{n \to \infty} \sqrt{npq} \cdot P_{n,m} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

ehk küllalt suure katsete arvu n korral

$$\sqrt{npq} \cdot P_{n,m} \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

Funktsiooniga $\varphi(x)$ määratud tõenäosuse tihedust nimetatakse normaaljaotuse tiheduseks ja öeldakse, et juhuslik suurus allub normaaljaotusele. Normaaljaotust käsitleme üksikasjaliselt edaspidi.

Laplace'i lokaalteoreemi ja normaaljaotuse mõistet kasutades võime ütelda, et küllalt suure n korral läheneb binomiaaljaotus normaaljaotusele. Bionomiaaljaotusele alluva suuruse võimalike väärtuste tõenäosused saame leida normaaljaotuse tiheduse kaudu. Normaaljaotuse tiheduse $\varphi(x)$ väärtuse antud argumendi x väärtusel leiame harilikult vastavast tabelist (vt. lisa I).

Kokkuvõte. Tõenäosuse $P_{n,m}$ arvutame küllalt suure katsete arvu n korral järgmiselt:

1) leiame sagedusele m vastava normeeritud hälbe

$$x = \frac{m - np}{\sqrt{npq}} ;$$

2) leiame saadud x väärtusele vastava $\varphi(x)$ väärtuse tabelist, kusjuures $\varphi(-x) = \varphi(x)$;

3) leiame otsitava tõenäosuse

$$P_{n,m} = \frac{1}{\sqrt{npq}} \varphi(x) .$$

Saadud $P_{n,m}$ väärtus pole küll päris täpne, kuid viga on juba väga väike $n \geqslant 20$ korral ning kahaneb kiiresti n kasvades. Arvutamine ligikaudse valemi abil on tunduvalt lihtsam ning vähem töömahukas võrreldes Bernoulli täpse valemiga.

1. näide. Praakdetailide tootmise tõenäosus on 0,005. Kui tõenäone on, et 10 000 detailist on praakdetaile 40?

Lähteandmetest $p = 0{,}005$; $n = 10\,000$, $m = 40$ leiame

$$q = 1 - p = 0{,}995 \ ; \quad \sqrt{npq} = \sqrt{1000 \cdot 0{,}005 \cdot 0{,}995} \approx$$

$$\approx 7{,}05 \ ;$$

$$x = \frac{m - np}{\sqrt{npq}} = \frac{40 - 10000 \cdot 0{,}005}{7{,}05} \approx -1{,}42 .$$

Tabelist leiame $\varphi(x)$ väärtuse kohal $x = 1{,}42$, sest $\varphi(1{,}42) = \varphi(-1{,}42) = 0{,}1456$ ja

$$P_{10000,\,40} \approx \frac{0{,}1456}{7{,}05} \approx 0{,}0206 .$$

Bernoulli valemiga arvutades saame

$$P_{10000\,,\,40} = \frac{10000\,!}{40!\;9960!} \, 0{,}005^{40} \cdot 0{,}995^{9960} = 0{,}0197 ,$$

Ligikaudse väärtuse viga 0,0206 - 0,0197 = 0,0009 moodustab 4,5% täpsest väärtusest, mis praktika seisukohalt ei ole olulise tähtsusega.

## 5. Normaaljaotuse tihedus.

A. Eelmises punktis nimetasime funktsiooni

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

normaaljaotuse tiheduseks. Tuleks kontrollida, kas $\varphi(x)$ puhul on täidetud tingimused, mida peab rahuldama tõenäosuse tihedus (vt. § 2, p. 7).

Funktsioon $\varphi(x)$ on iga x korral positiivne, sest positiivse arvu e astmed on mis tahes astendaja puhul positiivsed. Integraalarvutuse kursuses näidatakse, et

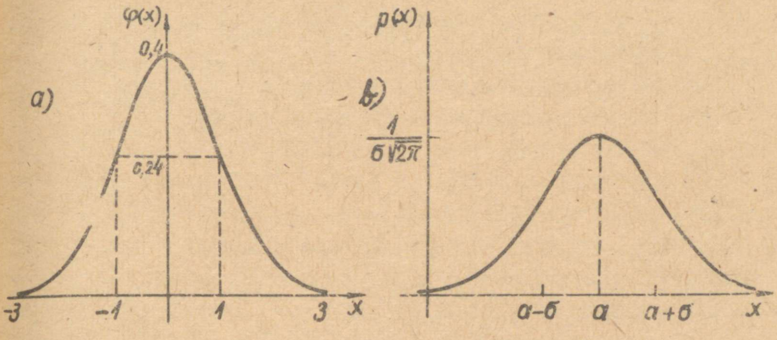
$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1 .$$

Seega täidab $\varphi(x)$ mõlemaid tõenäosuse tihedusele esitatavaid nõudeid.

Normaaljaotuse tiheduse $\varphi(x)$ graafikut nimetatakse normaalkõveraks ehk Gaussi kõveraks. Gaussi kõvera võrrand on

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

Funktsiooni $y = \varphi(x)$ uurimisel selgub (vt. I kursuse materjali), et funktsioonil on maksimum punktis $x = 0$ ja $y_{max} = \frac{1}{\sqrt{2\pi}} \approx 0{,}3989$ ning käänupunktide abstsissid on $x = \pm 1$, millele vastav funktsiooni väärtus on $y(\pm 1) = 0{,}2420$. Kõver on sümmeetriline ordinaattelje suhtes. Gaussi kõver on joonisel 15a.




Joonis 15.

B. Osutub, et juhusliku suuruse tõenäosuse tiheduseks on ka üldisema kujuga funktsioon

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}},$$

kus parameetrid $a$ ja $\sigma$ on reaalarvud. Seda funktsiooni nimetatakse üldise normaaljaotuse tiheduseks.

Eespool defineeritud normaalne tihedus $\varphi(x)$ on funktsiooni $p(x)$ erijuht, mis vastab parameetrite väärtustele $a = 0$, $\sigma = 1$.

Saab näidata, et $p(x)$ avaldises esinev parameeter $a$ on normaaljaotusele alluva juhusliku suuruse keskväärtuseks ja parameeter $\sigma$ on standardhälbeks

$$a = E(x)\,, \qquad \sigma = \sqrt{D(x)}\,.$$

Loetleme funktsiooni $p(x)$ uurimisel saadud tulemusi (vt. I kursuse materjali):

1) funktsiooni maksimum asub keskväärtusele vastavas punktis $x = a$ ja $y_{max} = \frac{1}{\sigma\sqrt{2\pi}}$;

2) käänupunktid asuvad keskväärtusest kahel pool kaugu-

sel $\sigma$, s.t., et käänupunktide abstsissid on $x = a \pm \sigma$. Käänupunktide vahel on kõver kumer, väljaspool nõgus.

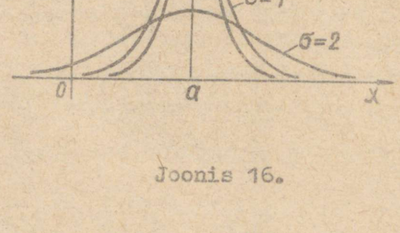
3) kõver on sümmeetriline sirge $x = a$ suhtes,

4) kõvera asümptoodiks on abstsisstelg.

Üldise normaaljaotuse tiheduse graafik on joonisel 15 b.

Keskväärtuse a muutumine jääva $\sigma$ korral nihutab normaalkõverat abstsisstelje sihis vasakule või paremale. Kui muutub standardhälve $\sigma$ keskväärtuse a muutumatuks jäädes, siis muutub normaalkõvera kuju, kuid nii, et kõvera ja abstsisstelje vaheline pindala on ikka 1. Joonisel 16 on esitatud normaalkõverad kolme erineva standardhälbe korral; arvulised andmed on tabelis

| x | a | a ± 0,5 | a ± 1 | a ± 2 |
|---|---|---|---|---|
| y( $\sigma$ = 1) | 0,399 | 0,352 | 0,242 | 0,054 |
| y( $\sigma$ = 0,5) | 0,798 | 0,484 | 0,108 | |
| y( $\sigma$ = 2) | 0,200 | 0,194 | 0,176 | 0,121 |




Joonis 16.

Näeme, et mida väiksem on $\sigma$, seda suurem on $y_{max}$ ja järsem on kõver ning seda rohkem kõvera all olevast pindalast on keskväärtuse a ümbruses. Kui $\sigma$ kasvab, siis kõver muutub järjest lamedamaks ning pindala jaotub abstsisstelje sihis laiemalt.

C. Normaaljaotuse tiheduselt

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

saame üle minna funktsioonile

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

koordinaatide teisendamise teel, asendades x väärtused nn. normeeritud hälvetega

$$t = \frac{x-a}{\sigma}.$$

Üleminek muutujalt x muutujale t tähendab koordinaatide alguspunkti nihutamist punkti x = a, kujutusühiku vähendamist abstsisstelje sihis $\sigma$ korda ja suurendamist ordinaattelje sihis sama arv korda. See on analoogiline punktis 2 kasutatud koordinaatide teisendamisega.

## 6. Normaalne jaotusfunktsioon. Laplace'i funktsioon.

Normaaljaotusele alluva juhusliku suuruse jaotusfunktsioon on kergesti saadav normaaljaotuse tiheduse kaudu. Eespool ( § 1, p. 7) näitasime, et juhusliku suuruse jaotusfunktsioon avaldub tõenäosuse tiheduse kaudu määratud intergraalina

$$P(X < x) = F(x) = \int_{-\infty}^{x} p(x)\,dx.$$

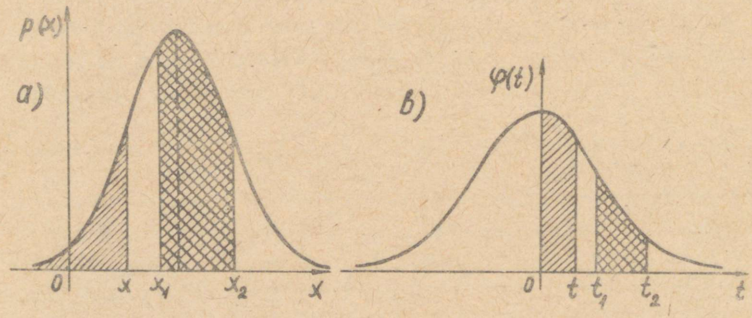


13

Normaaljaotusele vastavat jaotusfunktsiooni nimetame edaspidi normaalseks jaotusfunktsiooniks F(x). Normaalne jaotusfunktsioon on esitatav kujul

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}} dx .$$

Geomeetriliselt vastab jaotusfunktsioonile normaalkõvera alla jääv pindala abstsissini x (joonis 17a).




Joonis 17.

Normaalset jaotusfunktsiooni esitav integraal ei avaldu elementaarfunktsioonides. Praktiliste ülesannete lahendamiseks võib koostada F(x) väärtuste tabeli, kuid ei ole mõeldav seda teha iga a ja $\sigma$ väärtuse jaoks eraldi. Selle raskuse kõrvaldamiseks tehakse muutuja vahetus, minnes x väärtustelt üle normeeritud hälvetele

$$t = \frac{x - a}{\sigma}$$

Avaldame muutuja x diferentsiaali uue muutuja diferentsiaali kaudu

$$x = \sigma t + a ,$$

$$dx = \sigma dt .$$

Diferentsiaali dx kaudu sisse tulev kordaja $\sigma$ taandub jaotusfunktsiooni avaldise nimetajas oleva teguriga $\sigma$ ja normaalne jaotusfunktsioon omandab kuju

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt .$$

Normaalse jaotusfunktsiooni F(t) asemel kasutatakse harilikult Laplace'i funktsiooni

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{0}^{t} e^{-\frac{t^2}{2}} dt = \int_{0}^{t} \varphi(t)\, dt .$$

Normaalse jaotusfunktsiooni ja Laplace'i funktsiooni seos on järgmine

$$F(t) = \frac{1}{2} + \Phi(t) .$$

Tõepoolest, funktsiooni $\varphi(t)$ graafik on sümmeetriline ordinaattelje suhtes ning seega on mõlemal pool ordinaattelge olev pindala võrdne $\frac{1}{2}$ (kogu pindala on 1 !). Seega

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \int_{0}^{t} e^{-\frac{t^2}{2}} dt.$$

Geomeetriliselt vastab Laplace'i funktsiooni absoluutväärtusele Gaussi kõvera alune pindala ordinaatteljest kuni abstsissile t vastava ordinaatlõiguni (joonis 17b).

Käsitleme mõningaid Laplace'i funktsiooni omadusi.

1. Argumendi väärtusel $t = 0$ on $\Phi(t) = 0$.

Tõepoolest, määratud integraali väärtus võrdse alumise ja ülemise raja puhul on null

$$\Phi(0) = \int_{0}^{0} \varphi(t)\, dt = 0 .$$

2. $$\Phi(+\infty) = \int_{0}^{+\infty} \varphi(t)\, dt = \frac{1}{2} .$$

Funktsiooni $\Phi(t)$ väärtusele lõpmatu ülemise raja puhul vastab Gaussi kõvera alune pindala ordinaatteljest paremal, kuid see on võrdne 1/2 . Praktiliselt on aga juba argumendi väärtusel $t = 5$ $\Phi(t) \approx \frac{1}{2}$ .

3. Funktsioon $\Phi(t)$ on paaritu funktsioon, s.t.

$$\Phi(-t) = - \Phi(t).$$

Gaussi kõvera alused pindalad sümmeetriliselt kahele poole ordinaattelge on võrdsed. Pindala abstsissile $-t$ $(t > 0)$ vastava ordinaatlõigu ja ordinaattelje vahel võrdub

$$\int_{-t}^{0} \varphi(t)\, dt ,$$

pindala ordinaattelje ja abstsissile $t$ vastava ordinaatlõigu vahel võrdub

$$\int_{0}^{t} \varphi(t)\, dt .$$

Seega $\int_{-t}^{0} \varphi(t)\, dt = \int_{0}^{t} \varphi(t)\, dt$ . Määratud integraali omaduste järgi $\int_{-t}^{0} \varphi(t)\, dt = - \int_{0}^{-t} \varphi(t)\, dt$

ning järelikult

$$- \int_{0}^{-t} \varphi(t)\, dt = \int_{0}^{t} \varphi(t)\, dt .$$

ehk $-\Phi(-t) = \Phi(t)$.

4. Funktsioon $\Phi(t)$ on kasvav, s.t. kui $t_2 > t_1$, siis $\Phi(t_2) > \Phi(t_1)$. Suuremale $t$ väärtusele vastav Gaussi kõvera alune pindala on suurem (joonisel 17b pindalad 0 ja $t_2$ ning 0 ja $t_1$ vahel).

Laplace'i funktsiooni väärtuse saame leida tabelist (vt. lisa II), kusjuures argumendi muutumisvahemik on harilikult 0-st 5-ni. Negatiivsete argumendi väärtuste puhul leiame $\Phi(-t) = -\Phi(t)$ ning kui $t > 5$, siis võtame $\Phi(t) = 0{,}5$.

Märkus: Tihti võib kohata Laplace'i funktsiooni defineerituna

$$\frac{2}{\sqrt{2\pi}} \int\limits_0^t e^{-\frac{t^2}{2}} dt,$$

mille väärtused on 2 korda suuremad eespool defineeritud funktsiooni väärtustest. Tabelite kasutamisel tuleb jälgida, missugune on määratud integraali ees olev kordaja. Käesoleva konspekti ulatuses kasutame viimati defineeritud funktsiooni tähisena $\bar{\Phi}(t)$ (lisa III).

Kehtib võrdus

$$\bar{\Phi}(t) = 2\Phi(t).$$

## 7. Juhusliku suuruse väärtuste antud vahemikku sattumise tõenäosus normaaljaotuse puhul.

Sageli tuleb hinnata, kui tõenäone on, et juhusliku suuruse väärtused asuvad teatud vahemikus. Käesolevas punktis vaatleme normaaljaotusele alluva juhusliku suuruse X väärtuste vahemikus $(x_1, x_2)$ asumise tõenäosust. Esimese paragrahvi 6. punktis näitasime, et juhusliku suuruse antud vahemikku langemise tõenäosus on võrdne jaotusfunktsiooni väärtuste vahega vahemiku otspunktides.

Normaalse jaotusfunktsiooni korral

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}} dx - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Minnes muutujalt x üle normeeritud hälvetele (vt. 7.p.), saame

$$P(x_1 < X < x_2) = F(t_2) - F(t_1) ,$$

kus

$$t_1 = \frac{x_1 - a}{\sigma} \; ; \quad t_2 = \frac{x_2 - a}{\sigma} .$$

Kasutades seost

$$F(t) = \frac{1}{2} + \Phi(t)$$

leiame

$$P(x_1 < X < x_2) = \frac{1}{2} + \Phi(t_2) - \left[\frac{1}{2} + \Phi(t_1)\right] =$$

$$= \Phi(t_2) - \Phi(t_1)$$

– tõenäosus normaaljaotusele alluva juhusliku suuruse sattumiseks antud vahemikku on võrdne Laplace'i funktsiooni väärtuste vahega vahemiku otspunktides.

Geomeetriliselt vastab tõenäosusele $P(x_1 < X < x_2)$ Gaussi kõvera alune pindala vahemiku otspunktidele vastavate ordinaatlõikude vahel. Joonisel 17 on vastav pindala kahekordselt viirutatud – 17a esialgse muutuja x jaoks, 17b uue muutuja t jaoks.

Erilist huvi pakub ülesanne: kui tõenäone on, et juhusliku suuruse hälbed keskväärtusest mõlemale poole ei ületa antud arvu $\varepsilon$, s.t. kui tõenäone on, et

$$a - \varepsilon < X < a + \varepsilon$$

ehk

$$|X-a| < \varepsilon .$$

Tõketele $a + \varepsilon$ ja $a - \varepsilon$ vastavad normeeritud hälbed on

$$t_2 = \frac{a + \varepsilon - a}{\sigma} = + \frac{\varepsilon}{\sigma} ; \quad t_1 = \frac{a - \varepsilon - a}{\sigma} = - \frac{\varepsilon}{\sigma}$$

ja

$$P(|X - a| < \varepsilon) = \Phi(\frac{\varepsilon}{\sigma}) - \Phi(-\frac{\varepsilon}{\sigma}) = 2\Phi(\frac{\varepsilon}{\sigma}) = \bar{\Phi}(\frac{\varepsilon}{\sigma}).$$

Kokkuvõte. Normaaljaotusele (keskväärtusega $a$ ja standardhälbega $\sigma$) alluva juhusliku suuruse $X$ antud vahemikku $(x_1, x_2)$ sattumise tõenäosuse leidmiseks tuleb

1) määrata vahemiku otspunktidele $x_1$ ja $x_2$ vastavad normeeritud hälbed

$$t_1 = \frac{x_1 - a}{\sigma}, \quad t_2 = \frac{x_2 - a}{\sigma},$$

2) leida Laplace'i funktsiooni tabelist $\Phi(t_1)$ ja $\Phi(t_2)$;

3) leida $P(x_1 < X < x_2) = \Phi(t_2) - \Phi(t_1)$.

Tõenäosuse $P(|X - a| < \varepsilon)$ leidmiseks tuleb

1) määrata normeeritud hälve $t = \frac{\varepsilon}{\sigma}$;

2) leida tabelist $\Phi(t)$ ja seejärel $2\Phi(t) = \bar{\Phi}(t)$.

1. näide. Normaaljaotusele alluva juhusliku suuruse $X$ keskväärtus on $a = 168$ ja standardhälve $\sigma = 5{,}9$. Kui tõenäone on, et juhusliku suuruse väärtused asuvad vahemikus (160, 180) ?

Leiame normeeritud hälbed

$$t_1 = \frac{160 - 168}{5{,}9} = -1{,}356 ; \quad t_2 = \frac{180 - 168}{5{,}9} = 2{,}034.$$

Tabelist leiame (lisa II ) lineaarset interpolatsiooni kasutades

$$\Phi(2{,}034) = 0{,}4790\ ;\ \Phi(-1{,}356) = -\Phi(1{,}356) = -0{,}4125.$$

Otsitav tõenäosus

$$P(160 < X < 180) = 0{,}4790 - (-0{,}4125) = 0{,}8915.$$

Keskmiselt 89% suuruse X väärtustest asub 160 ja 180 vahel.

2. näide. "3 $\sigma$ reegel".

Leiame, kui tõenäone on, et normaaljaotusele alluva juhusliku suuruse X väärtused ei hälbi keskväärtuse a suhtes rohkem kui standardhälbe $\sigma$ kordsete võrra. Anname tõkkele $\varepsilon$ väärtuse $\sigma$, $2\sigma$, $3\sigma$

$$\varepsilon = \sigma\ ;\ t = \frac{\varepsilon}{\sigma} = 1;\ P(|X-a| < \sigma) = 2\,\Phi(1) = 0{,}6826 \quad (68{,}26\%);$$

$$\varepsilon = 2\sigma\ ;\ t = 2\ ;\ P(|X-a| < 2\sigma) = 2\,\Phi(2) = 0{,}9545 \quad (95{,}45\%);$$

$$\varepsilon = 3\sigma\ ;\ t = 3\ ;\ P(|X-a| < 3\sigma) = 2\,\Phi(3) = 0{,}9993 \quad (99{,}93\%).$$

Selgus, et normaaljaotusele alluva juhusliku suuruse kõigist võimalikest väärtustest satuvad keskmiselt 68,26% keskväärtusele lähemale kui ühekordne standardhälve. Analoogiliselt tõlgendame ka kaht järgmist tulemust. Ainult tühine osa (0,07%) võimalikest väärtustest hälbib keskväärtuse suhtes rohkem kui $3\sigma$ võrra ehk kolmekordset standardhälvet ületavate hälvetega väärtusi esineb väga harva.

"$3\sigma$ reegel": praktiliselt võib olla kindel, et normaaljaotusele alluv juhuslik suurus ei hälbi keskväärtuse suhtes rohkem kui kolmekordse standardhälbe võrra.

Märkus: Funktsiooni $\overline{\Phi}(t)$ tabeleid kasutades on

$$P(x_1 < X < x_2) = \frac{1}{2}\left[\overline{\Phi}(t_2) - \overline{\Phi}(t_1)\right].$$

## 8. Sündmuse sageduse antud vahemikku sattumise tõenäosus.

Käesoleva paragrahvi alguses lahendasime ülesande sündmuse sageduse m tõenäosuse leidmisest korduvatel katsetel. Võrratult sagedamini tuleb praktikas kokku puutuda ülesannetega, kus tuleb selgitada, kui tõenäone (kui usutav) on uuritava sündmuse A sageduse m sattumine antud vahemikku. Näiteks, kui tõenäone on, et

10000 elektripirni hulgas ei ole praakpirne rohkem kui 100 (s.t. 0 kuni 100) ?

märgi tabamuste arv 100 lasust asub 85 ja 95 vahel ?

kauplusse toodavatest naiste suvejalatsitest on üle 80% heledates värvitoonides?

jne.

Ülesande üldine seade on järgmine: Kui suur on tõenäosus n korduval katsel sündmuse A sageduse m sattumiseks vahemikku väärtusest a väärtuseni b, kui sündmuse A tõenäosus üksikkatsel $p(A) = p$ ?

Meid huvitavaks sündmuseks on nüüd "sageduse sattumine a ja b vahele" - $a \leqslant m \leqslant b$ ja selle sündmuse tõenäosuse tähistame $P(a \leqslant m \leqslant b)$.

Sündmust $a \leqslant m \leqslant b$ võime vaadelda osasündmuste "sündmus A toimub $m = a$ korda", "sündmus A toimub $m = a + 1$ korda", ... , "sündmus A toimub $m = b$ korda" summana ja tõenäosuste liitmislause kohaselt

$$P(a \leqslant m \leqslant b) = P_{n,a} + P_{n,a+1} + \dots + P_{n,m} + \dots + P_{n,b} =$$

$$= \sum_{m=a}^{b} P_{n,m} \qquad .$$

Näide. Telestuudios on 5 kaamerat. Tõenäosus selleks, et kaamera on vaadeldaval hetkel sisse lülitatud, on 0,6. Kui tõenäone on, et antud hetkel ei ole üle 3 kaamera sisse lülitatud?

Sündmuseks A on "kaamera on sisse lülitatud". Meid huvitab tõenäosus sündmuse A sageduse m asumiseks 0 ja 3 vahel, ehk $P(0 \leq m \leq 3)$. Selleks leiame

$$P_{5,0} = \frac{5!}{0!\,5!}\,0{,}6^0 \,.\, 0{,}4^5$$ või lihtsamalt, korrutamislause järgi

$$P_{5,0} = 0{,}4^5 = 0{,}01024 \,; \quad P_{5,1} = \frac{5!}{1!\,4!}\,0{,}6^1 .\, 0{,}4^4 = 0{,}0768$$

$$P_{5,2} = \frac{5!}{2!\,3!}\,0{,}6^2 \,.\, 0{,}4^3 = 0{,}2304 \,; \quad P_{5,3} = \frac{5!}{3!\,2!}\,0{,}6^3 \,.$$

$$.0{,}4^2 = 0{,}3456$$

ja

$$P(0 \leq m \leq 3) = 0{,}01024 + 0{,}0768 + 0{,}2304 + 0{,}3456 =$$

$$= 0{,}6630,$$

s.t. keskmiselt 66,3% kogu tööajast ei tööta üle 3 kaamera.

Suurema katsete arvu n ja pika vahemiku (a,b) korral kujuneb eespool kirjeldatud võte väga aeganõudvaks. Kasutame ligikaudset valemit otsitava tõenäosuse leidmiseks. Käesoleva paragrahvi 4. punktis selgitasime, et katsete arvu n kasvades läheneb binomiaaljaotus normaaljaotusele. Normaaljaotuse puhul aga leiame juhusliku suuruse antud vahemikku sattumise tõenäosuse Laplace'i funktsiooni abil. Sageduselt m läheme üle normeeritud hälvetele

$$t = \frac{m - E(m)}{\sigma(m)} = \frac{m - np}{\sqrt{npq}},$$

seejärel leiame m tõketele a ja b vastavad suuruse x tõkked $x_1$ ja $x_2$ ning

$$P(a \leq m \leq b) \cong \Phi(t_2) - \Phi(t_1),$$

mida nimetatakse Laplace'i integraalvalemiks.

Edaspidi jätame ligikaudse võrduse märgi ära, sest küllalt suure katsete arvu n (isegi kui $n \geq 20$) korral on viga väga väike.

Kokkuvõte. Sündmuse sageduse m antud vahemikku (a,b) sattumise tõenäosuse leidmiseks tuleb

1) määrata sageduse m tõketele a ja b vastavad normeeritud hälbed

$$t_1 = \frac{a - np}{\sqrt{npq}}; \quad t_2 = \frac{b - np}{\sqrt{npq}};$$

2) leida Laplace'i funktsiooni tabelist $\Phi(t_2)$ ja $\Phi(t_1)$;

3) leida $P(a \leq m \leq b) = \Phi(t_2) - \Phi(t_1)$.

## 9. Tüüpülesandeid.

Käesolevas punktis selgitame, kuidas binomiaal- ja normaaljaotuse jaoks tuletatud juhusliku suuruse antud vahemikku sattumise tõenäosuse arvutamise eeskirju kohandada mitmesuguste praktikas esinevate tüüpülesannete lahendamiseks.

I tüüp. Leida sageduse m antud vahemikku sattumise tõenäosus

$$P(a \leq m \leq b) = P.$$

Ülesande sellise seade puhul on antud a, b, n, ja p; leida

$$P = \Phi(t_2) - \Phi(t_1).$$

Näide. Poisslapse sündimise tõenäosus on 0,515. Kui tõenäone on, et iga 1000 vastsündinu hulgas on poisslaste arv 455 ja 555 vahel ?

Lähteandmed: $n = 1000$, $p = 0{,}515$, $a = 455$, $b = 555$; kohe saame ka $q = 1 - p = 0{,}485$.

Leida $P = P(455 \leq m \leq 555)$.

Leiame

$$t_1 = \frac{a - np}{\sqrt{npq}} = \frac{455 - 0,515 \cdot 1000}{1000 \cdot 0,515 \cdot 0,485} \approx - 3,80;$$

$$t_2 = \frac{b - np}{\sqrt{npq}} = \frac{555 - 0,515 \cdot 1000}{1000 \cdot 0,515 \cdot 0,485} \approx 2,53.$$

Tabelist määrame $\Phi(2,53) = 0,4943$, $\Phi(- 3,80) =$ $= - \Phi(3,80) = - 0,4999$.

Otsitav tõenäosus

$$P = 0,4943 - (- 0,4999) = 0,9942.$$

Tulemus näitab, et vähemalt 99 juhul 100-st on iga 1000 sünni korral poisslaste arv 455 ja 555 vahel.

II tüüp. Kui tõenäone on, et sagedus m ei hälbi oma keskväärtusest E(m) rohkem kui antud arvu $\varepsilon$ võrra?

Antud on $\varepsilon$, n, p, $q = 1 - p$;

leida tuleb

$$P ( |m - np| < \varepsilon ).$$

Kuna sagedus m allub binomiaaljaotusele, mis on lähedane normaaljaotusele, siis saame kasutada eeskirja

$$P( |X-a| < \varepsilon ) = 2 \Phi(t) , \text{ kus } t = \frac{\varepsilon}{\sigma(x)} .$$

Suurusele X vastab sagedus m, siis ka $E(m) = np = a$ ja $\sigma(x) = \sigma(m) = \sqrt{npq}$. Tehes vastavad asendused jõuame valemini

$$P( |m - np| < \varepsilon ) = 2 \Phi(t) , \text{ kus } t = \frac{\varepsilon}{\sqrt{npq}} .$$

Näide. Valmistoodangust jõuab keskmiselt 20% lattu TKO kontrolltemplita. Kui tõenäone on, et 400 lattu saabuva toote hulgas ei erine templita toodete arv nende keskmisest rohkem kui 10 võrra?

Lähteandmed: $n = 400$, $p = 0,2(20\%)$, $\varepsilon = 10$ ;
$q = 1 - 0,2 = 0,8$.

Leida $P(|m - np| < 10)$.

Templita toodete keskmine 400 toote hulgas on keskväärtuseks

$$E(m) = n.p = 0,2 \cdot 400 = 80 .$$

Nüüd saame

$$t = \frac{10}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = \frac{10}{10 \cdot 0,8} = 1,25 ;$$

$$P(|m - 80| < 10) = 2\,\Phi(1,25) = 2 \cdot 0,3944 = 0,7888.$$

Kui 400 tootest koosnevaid partiisid saadetakse lattu korduvalt, siis umbes 79% partiidest on sellised, kus kontrolltemplita toodete arv on 70 ja 90 vahel, ülejäänud partiides võib vastav näitaja olla 70-st väiksem või 90-st suurem.

III tüüp. Sooritatakse n katset, kusjuures igal katsel on uuritava sündmuse tõenäosus jääv ja võrdne p. Kui tõenäone on, et sündmuse relatiivne sagedus $\frac{m}{n}$ ei erine tõenäosusest p rohkem kui antud arvu $\varepsilon$ võrra?

Antud on $n$, $p$, $\varepsilon$, $q = 1 - p$.

Võrdse pikkusega katseseeriates on sündmuse sagedus m harilikult erinev ja siis on ka $\frac{m}{n}$ juhuslik suurus. Oleme huvitatud, kui tõenäone on, et $\left|\frac{m}{n} - p\right| < \varepsilon$. Vastava tõenäosuse tähistame

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right).$$

Binomiaaljaotuse käsitlemisel näitasime, et $p = E\left(\frac{m}{n}\right)$ ja $\sigma\left(\frac{m}{n}\right) = \sqrt{\frac{pq}{n}}$. Analoogiliselt II tüübis käsitletule asendame X suurusega $\frac{m}{n}$ ja $\sigma(x)$ asemele võtame $\sigma\left(\frac{m}{n}\right)$.

Nüüd

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 2\,\Phi(t), \text{ kus } t = \varepsilon\sqrt{\frac{n}{pq}}$$

Näide. Tõenäosus mittestandardse detaili tootmiseks on 0,1. Kui tõenäone on, et 400 detaili korral mittestandardsete detailide relatiivne sagedus ei erine 0,1-st rohkem kui 0,03 võrra?

Lähteandmed: $n = 400,\ p = 0,1\ ,\ \varepsilon = 0,03\ ;$

$$q = 1 - 0,1 = 0,9\ .$$

Leida $P\left( \left| \frac{m}{400} - 0,1 \right| < 0,03 \right)$.

Määrame argumendi $t = 0,03 \cdot \sqrt{\frac{400}{0,1 \cdot 0,9}} = 2$

ja

$$P\left( \left| \frac{m}{400} - 0,1 \right| < 0,03 \right) = 2\ \Phi(2) = 0,9544.$$

Kui kontrollida 400 detailist koosnevaid partiisid korduvalt, siis umbes 95% partiides ei erine mittestandardsete detailide relatiivne sagedus 0,1-st rohkem kui 0,03 võrra, või siis asub piirides 0,07 kuni 0,13. Tulemust võib tõlgendada ka nii: 95% 400 detailist koosnevatest partiidest on sellised, kus mittestandardsete detailide arv asub 0,07 . 400 = 28 ja 0,13 . 400 = 52 vahel.

IV tüüp. Teades sündmuse tõenäosust üksikkatsel, leida relatiivse sageduse $\frac{m}{n}$ maksimaalne hälve tema keskväärtuse p suhtes antud tõenäosusega.

Suuruse X antud tõenäosusega maksimaalset hälvet oma keskväärtuse E(x) suhtes nimetatakse sageli piirveaks antud tõenäosusega. Piirviga $\varepsilon$ määrab suuruse usalduspiirid $E(x) \pm \varepsilon$ ja nendes piirides asumise tõenäosust nimetatakse usaldusnivooks.

Antud on n, p, $P\left( \left| \frac{m}{n} - p \right| < \varepsilon \right) = P\ ;\ q = 1-p.$

Leida relatiivse sageduse piirviga $\varepsilon$ .

Antud p järgi määrame Laplace'i funktsiooni tabelist argumendi t väärtuse, kasutades võrdust

$$P = 2\ \Phi(t) = \overline{\Phi}(t)\ ;\quad \Phi(t) = \frac{P}{2}\ .$$

Piirvea $\varepsilon$ ja t vaheline sõltuvus esitub kujul

$$t = \varepsilon \sqrt{\frac{n}{p\,q}}\ ,$$

kust leiame otsitava $\varepsilon$

$$\varepsilon = t\sqrt{\frac{pq}{n}} .$$

Näide. Automaattööpingil valmistatud toodangust on keskmiselt 10% praaki. Leida praakdetailide relatiivse sageduse piirviga keskmise praagiprotsendi suhtes usaldusnivooga 0,992 10 000 detailist koosnevas partiis.

Lä' eandmed: $n = 10000$ , $p = 0,1(10\%)$ ,

$$P\left( \left|\frac{m}{10000} - 0,1\right| < \varepsilon \right) = 0,992 .$$

Leida $\varepsilon$ .

Laplace'i funktsiooni tabelist leiame argumendi t väärtuse, mis vastab kahekordsele funktsiooni väärtusele $2\,\Phi(t) =$ $= 0,992$ (või $\bar{\Phi}(t) = 0,992$). Siis $\Phi(t) = 0,496$ ja $t = 2,65$.

Nüüd

$$\varepsilon = 2,65 . \sqrt{\frac{0,1 \cdot 0,9}{10000}} = 0,00795 \approx 0,008$$

Keskmiselt 99% 10000 detailist koosnevatest partiidest on sellised, kus praagi relatiivne sagedus ei erine 0,1-st rohkem kui 0,008 ehk 10%-st rohkem kui 0,8% võrra.

V tüüp. Mitu katset tuleks sooritada, et antud tõenäosusega (usaldusnivooga) garanteerida, et sündmuse relatiivne sagedus ei erine sündmuse tõenäosusest üksikkatsel rohkem kui antud arvu $\varepsilon$ võrra?

Antud on $p$, $q$, $\varepsilon$, $P\left( \left|\frac{m}{n} - p\right| < \varepsilon \right) = P$.

Leida katsete arv n.

Ülesanne on analoogiline eelmisega. Otsitavaks on $\varepsilon$ asemel n. Leiame jälle võrduse $P = 2\,\Phi(t)$ abil argumendi t. Kuna t ja n on seotud võrdusega

$$t = \varepsilon\sqrt{\frac{n}{pq}} ,$$

siis saame

$$n = \frac{t^2 pq}{\varepsilon^2} .$$

Näide. Mitu korda on vaja visata münti, et relatiivne sagedus ei erineks vapi esiletuleku tõenäosusest üksikviskel rohkem kui 0,01 võrra usaldusnivooga 1) $P = 0,6$, 2) $P = 0,954$?

1. Lähteandmed: $\varepsilon = 0,01$, $P = 0,6$; varemast teame, et $p = 0,5$. Leida $n$.

Kuna $P = 0,6$, siis $\Phi(t) = 0,3$ ja tabelist leiame, et $t \approx 0,84$.

Minimaalne visete arv

$$n = \frac{0,84^2 \cdot 0,5^2}{0,01^2} = 1764 .$$

2. $P = 0,954$. Siis $t = 2$ ja $n = 10000$.

Usaldusnivoo suurendamine sama täpsuse korral nõuab katsete arvu suurendamist.

## 10. Poissoni jaotus.

Korduvatel katsetel leiame sündmuse $m$ korda toimumise tõenäosuse $P_{n,m}$ Bernoulli valemi abil või suure katsete arvu $n$ puhul ligikaudsest valemist

$$P_{n,m} = \frac{1}{\sqrt{npq}} \varphi(x) ,$$

mis tähendab binomiaaljaotuse lähendamist normaaljaotusega. Tulemused on seda täpsemad, mida lähemal on $p$ väärtusele 0,5. Kui aga on tegemist harva toimuvate sündmustega, s.t. $p$ on väga väike ($p < 0,1$), võib binomiaaljaotuse lähendamine normaaljaotusega anda üpris ebatäpseid tulemusi. Saab näidata, et väga väikese $p$ ja küllalt suure $n$ korral läheneb

$$P_{n,m} = C_n^m p^m (1-p)^{n-m}$$

suuruselе

$$P_m = \frac{\mu^m}{m!} e^{-\mu}, \quad \text{kus} \quad \mu = np.$$

Viimase valemiga määratud juhusliku suuruse m tõenäosuste jaotust nimetatakse Poissoni jaotuseks. Suurus $\mu$ on sageduse m keskväärtus, olles ainukeseks jaotuse parameetriks.

Poissoni jaotus leiab laialt rakendamist massilise teenindamise küsimuste uurimisel. Poissoni jaotusele allub näiteks telefonikeskjaama väljakutsete sagedus ajaühikus; kiirabiauto väljakutsete sagedus teatud ajavahemikus jne. Harilikult on tegemist olukordadega, kus sündmuse toimumise tõenäosus üksikkatsel (väga väikeses ajavahemikus) on väike ja ajavahemiku lühendamisel väheneb, kuid nii, et $\mu = np = \text{const}$.

1. näide. Kangur teenindab 800 värtnat. Lõnga katkemise tõenäosus igal värtnal 5 min. jooksul on 0,005. Leida tõenäoseim lõngade katkemise arv ja selle tõenäosus.

Tõenäoseim katkemiste arv $\mu = np = 800 \cdot 0{,}005 = 4$. Leiame tõenäosuse täpsest Bernoulli valemist, samuti ligikaudsete valemitega, kasutades normaaljaotust ja Poissoni jaotust.

Bernoulli valem:

$$P_{800,4} = \frac{800!}{4!\,796!} 0{,}005^4 \cdot 0{,}995^{796} = 0{,}1945;$$

normaaljaotus:

$$P_4 = \frac{4^4}{4!} e^{-4} = \frac{256}{24} \cdot 0{,}0183 = 0{,}1954.$$

Näeme, et Poissoni jaotuse järgi arvutatud tulemuse viga (0,0009) on tublisti väiksem normaaljaotuse järgi arvutatud tulemuse veast (0,0055).

2. näide. Telefonikeskjaam saab tunnis keskmiselt k väljakutset. Leida tõenäosus m väljakutse saamiseks 1 minuti jooksul.

Ühes minutis saadavate väljakutsete arvu keskväärtuseks loeme väärtuse $\mu = \frac{k}{60}$. Tõenäosus m väljakutse saamiseks minutis oleks Poissoni valemi järgi

$$P_m = \frac{\left(\frac{k}{60}\right)^m}{m!} e^{-\frac{k}{60}} .$$

3. näide. Suures raamatuhoidlas on 100 pirni, nendest igaühe põlemise tõenäosus on 0,02. Koostame pirnide sisselülitamise sageduse m jaotustabeli (kõiki arvutusi ei ole läbi tehtud).

Poissoni jaotust kasutades leiame $\mu = np = 100 \cdot 0{,}02 = 2$ ja $e^{-\mu} = e^{-2} = 0{,}135335$. Sageduse m väärtustele vastavad tõenäosused arvutame valemist

$$P_m = \frac{2^m}{m!} e^{-2} = \frac{0{,}135335 \cdot 2^m}{m!} .$$

Nii saaksime näiteks

$$P_3 = 0{,}135335 \cdot \frac{2^3}{3!} = 0{,}18044 .$$

Tulemused on antud tabelis, kus on võrdluseks toodud ka Bernoulli valemiga leitud täpsed tulemused. **Arvutused on sooritatud 5-kohase täpsusega selleks,** et üldse näha täpse ja ligikaudse valemi kasutamisel tekkivaid erinevusi. Vastavate väärtuste võrdlemisel näeme, et kooskõla on väga hea. Tabelis puuduvate sageduste (12 jne.) tõenäosused on antud täpsuse piires praktiliselt nullid.

| m | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Poissoni valem | 0,13533 | 0,27067 | 0,27067 | 0,18044 | 0,09022 | 0,036099 |
| Bernoulli valem | 0,13262 | 0,27065 | 0,27342 | 0,18228 | 0,09021 | 0,03535 |

| m | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|
| Poissoni valem | 0,01203 | 0,00343 | 0,00085 | 0,00019 | 0,00003 | 0,00001 |
| Bernoulli valem | 0,01142 | 0,00313 | 0,00073 | 0,00015 | 0,00003 | 0,00001 |

## 11. Statistiliste jaotuste tasandamine.

Käesoleva paragrahvi oleme pühendanud juhusliku suuruse X tähtsamate jaotuste: binomiaaljaotuse, normaaljaotuse ja Poissoni jaotuse käsitlemisele. Praktikas tuleb väga tihti kokku puutuda juhuslike suurustega, mis alluvad nimetatud jaotusele. Sellepärast on neid jaotusi vaja tunda.

Matemaatilise statistika üheks tähtsamaks ülesandeks on juhusliku suuruse uurimisel saadud statistiliste andmete järgi selle suuruse jaotuse määramine. Teatud suuruse X eksperimentaalsel uurimisel saame andmeid selle suuruse kohta ikka piiratud hulgal, kusjuures need andmed pole kunagi absoluutselt täpsed. Uurija ülesandeks on katseandmete läbitöötamisel välja selgitada uuritavale nähtusele omased jooned ning seaduspärasused. Selleks püütakse statistilisi andmeid kirjeldada matemaatiliste meetoditega analüütiliste sõltuvuste abil. Seda nimetatakse statistiliste andmete tasandamiseks. Statistiliste andmete tasandamise ülesande seade on konkreetsemalt järgmine: antud empiirilise jaotuse põhjal määrata suuruse teoreetiline jaotus tõenäose tiheduse p(x) või jaotusfunktsiooni F(x) kujul ehk "määrata tasandav funktsioon". Ülesanne on analoogiline empiiriliste funktsionaalsete sõltuvuste tasandamisega.

Statistiliste andmete tasandamise ülesande võime jagada 3 etapiks:

1) valida sobiv tasandava funktsiooni tüüp (jaotuse tüüp);

2) määrata valitud jaotuse karakteristikud;

3) hinnata statistilise jaotuse ja konstrueeritud teoreetilise jaotuse kooskõla.

Tasandava funktsiooni tüüp määratakse harilikult ülesande sisu ja eelnevaid kogemusi või ka statistilise jaotuse välist kuju arvestades. Sellel etapil on sageli suur osa statistilise jaotuse graafilisel esitusel (polügoon, histogramm). Antud etapi juures me pikemalt ei peatu.

Teisel etapil tulevád määrata valitud teoreetilise jaotuse karakteristikud nii, et statistiline ja teoreetiline jaotus oleksid võimalikult heas kooskõlas. Tihti toimub see nii, et osa teoreetilise jaotuse karakteristikuid valitakse võrdseks statistilise jaotuse vastavate karakteristikutega. Lähemalt käsitleme seda järgmises punktis normaaljaotusega tasandamisel.

Kolmas etapp seisneb järgnevas. Olgu teoreetiline jaotus koostatud. Alati esineb lahkuminekuid empiirilise ja teoreetilise jaotuse vahel. Tuleb hinnata, kas need lahkuminekud on tingitud juhuslikest asjaoludest, mis tulenevad katsete piiratud mahust või on nad olulised, olles tingitud tasandava funktsiooni ebaõnnestunud valikust. Selle hindamiseks kasutatakse nn. kooskõla kriteeriume, mille juures me lähemalt ei peatu.

## 12. Statistiliste andmete tasandamine normaaljaotusega.

A. Normaaljaotus on üks sagedamini esinevaid juhusliku suuruse jaotusi. Vastuse küsimusele, miks see nii on, annab vene akadeemiku A. Ljapunovi poolt tõestatud teoreem. Teoreemi täpne formulatsioon ja tõestus jäävad oma keerukuse tõttu meie kursuse raamidest välja. Selgitame Ljapunovi teoreemist tulenevaid järeldusi.

Kui juhuslik suurus X koosneb suure arvu juhuslike osasuuruste $X_1, X_2, \ldots, X_n$ summast, millest iga liidetav eraldi võetuna mõjutab vähe nende summat, siis suuruse X jaotus on lähedane normaalsele. Praktikas tulebki väga tihti tegelda suurustega, mis on paljude faktorite mõju summaarseteks tulemusteks. Iga üksikfaktori mõju summaarsele tulemusele on tühine. Ljapunovi teoreemi kohaselt on sellise suuruse jaotus lähedane normaalsele ning meil avaneb võimalus selle suuruse muutumist ette näha. Samas on aga iga liidetava jaotus teadmata ning nende suure arvu tõttu ei olegi võimalik iga liidetava mõju summaarsele suurusele täpsemalt arvestada.

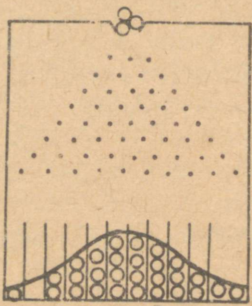
1. näide. Mõõduriista abil toimub katsel teatud füüsikalise suuruse mõõtmine. Iga katse tulemusena saame mõõdetava suuruse ligikaudse väärtuse. Mõõtmistulemust mõjutavad väga paljud kõrvalised tegurid, nagu näiteks väga väikesed temperatuuri, niiskuse jne. muutused, väikesed mõõteriista kõikumised jne. Iga üksiku kõrvalteguri mõju tulemusele on tühiselt väike, kuid nende mõjude summeerimine annab juba märgatava mõõtmisvea. Eespool öeldu põhjal võime lugeda mõõtmisvigu alluvaiks normaaljaotusele.

2. näide. Tehase automaatliinil valmistatakse teatud mõõtmetega võlle. Võllide diameetri vastavus normidele olgu võlli kõlblikkuse määramise aluseks. Juhuslikuks suuruseks X valime toodetud võllide diameetrite hälbed ettenähtud diameetrist. Toorikute materjali väikesed erinevused, temperatuuri kõikumised, vooluliini üksikute osade kulumine ja paljud muud faktorid tingivad igaüks lõpmata väikesi diameetri hälbeid, kuid nende summaarne mõju on märgatav - võllid on erinevate diameetritega. Diameetrite hälbed X alluvad normaaljaotusele.

3. näide. Normaaljaotusele alluvad ka paljud biomeetrilised näitajad, nagu inimese kasv, kaal, jala suurus jne. Riietusesemete tootmisel saab toodangut planeerida suuruse järgi nii, et kaubandusvõrgus ei tule harilikult puudust üksikutest numbritest ja ei teki mittenõutavate numbrite suuri ülejääke.

4. näide. Normaaljaotuse kujunemise illustratsiooniks on järgmine seade (joonis 18). Plaadi sisse on löödud nae-

lad malelaua süsteemis. Plaadi alumises servas on lahtrid. Kui seadme ühes otsas olevast avast kallata sisse kuulikesi, siis jaotuvad nad lahtrites ligikaudu normaalkõvera järgi.



Joonis 18.

Iga kuulike põrkub oma teel vastu teisi kuulikesi ja naelu, mis kujutavad endast juhuslikke mõjustusi. Naelte puudumisel langeksid kuulid ainult keksmistesse lahtritesse.

B. Olgu antud mingi tunnuse statistiline jaotus momentrea või intervallitud rea kujul, kusjuures on selgitatud, et uuritav tunnus allub normaaljaotusele. Antud statistilise jaotuse tasandamisel normaaljaotusega valitakse normaaljaotuse keskväärtuseks a variatsioonrea aritmeetiline keskmine $\bar{x}$ ja standardhälbeks $\sigma$ variatsioonrea andmetel leitud standardhälve s, s.t.

$$a = \bar{x} \quad \text{ja} \quad \sigma = s \, .$$

Seejärel saame normaaljaotuse tiheduse kujul

$$p(x) = \frac{1}{s\sqrt{2\pi}} \, e^{-\frac{(x-\bar{x})^2}{2s^2}} \, .$$

4. näide. Koostada tõenäosuse tiheduse avaldis §3 p. 10 3. näite andmetel, lugedes tööliste pikkuse jaotuse alluvaks normaaljaotusele, valides intervallid nii nagu statistilisel jaotusel.

Lähteandmete põhjal tuleb arvutada aritmeetiline keskmine ja standardhälve. See on meil nimetatud näites juba tehtud.

Saime $\bar{x} = 165{,}5$ cm ; $s = 6{,}05$ cm.

Tasandava normaaljaotuse tihedus esitub kujul

$$p(x) = \frac{1}{6{,}05\sqrt{2\pi}} e^{-\frac{(x-165{,}5)^2}{2 \cdot 36{,}58}} .$$

Intervallitud jaotustabeli koostamiseks tuleb leida tõenä[illegible]sed suuruse X väärtuste asumiseks antud intervallid[illegible]. Arvutuste lihtsustamise huvides asendame X väärtused normeeritud hälvetega

$$t = \frac{x - \bar{x}}{s} = \frac{x - 165{,}5}{6{,}05} ,$$

kus x on intervalli keskkohale vastav väärtus.
Siis saame p(x) asemel kasutada arvutuste käigus $\varphi(t)$ tabeleid. Tõenäosused $p_i$ võime leida ligikaudselt valemist

$$p_i(x) = P(x < X < x + \Delta x) \approx \varphi(t) \cdot \Delta t$$

või täpsemalt

$$p_i(x) = P(\tilde{x}_i < X < \tilde{x}_{i+1}) = \phi(\tilde{t}_{i+1}) - \phi(\tilde{t}_i) .$$

Suurused $\tilde{x}_i$ ja $\tilde{x}_{i+1}$ on intervallide otspunktidele vastavad x väärtused, $\tilde{t}_i$ ja $\tilde{t}_{i+1}$ vastavad normeeritud hälbe t väärtused. Intervallile $\Delta x = 3$ vastav t muut

$$\Delta t = \frac{\Delta x}{s} = \frac{3}{6{,}05} = 0{,}496 .$$

Jaotustabeli moodustavad intervallid ja tõenäosuste p lahter. Võrdluseks on viimases lahtris toodud statistilise rea andmetel relatiivsed sagedused f:1000 = w. Tulemused on heas kooskõlas (võrrelge p ja w vastavaid väärtusi !).

Arvutused on teostatud mõlema valemi järgi. Arvutuste käiku saab jälgida tabelist

| Inter-vallid | $\tilde{x}$ | $\tilde{t}$ | $\Phi(\tilde{t})$ | p | x | t | $\varphi(t)$ | p | w |
|---|---|---|---|---|---|---|---|---|---|
| 143-146 | 143 | -3,72 | -0,4999 | 0,0006 | 144,5 | -3,47 | 0,0010 | 0,0005 | 0,001 |
| 146-149 | 146 | -3,33 | -0,4993 | 0,003 | 147,5 | -2,98 | 0,0047 | 0,002 | 0,002 |
| 149-152 | 149 | -2,73 | -0,4968 | 0,009 | 150,5 | -2,48 | 0,0184 | 0,007 | 0,008 |
| 152-155 | 152 | -2,23 | -0,4871 | 0,028 | 153,5 | -1,98 | 0,0562 | 0,028 | 0,026 |
| 155-158 | 155 | -1,74 | -0,4591 | 0,067 | 156,5 | -1,49 | 0,1315 | 0,065 | 0,065 |
| 158-161 | 158 | -1,24 | -0,3925 | 0,122 | 159,5 | -0,99 | 0,2444 | 0,121 | 0,120 |
| 161-164 | 161 | -0,74 | -0,2703 | 0,172 | 162,5 | -0,50 | 0,3521 | 0,175 | 0,181 |
| 164-167 | 164 | -0,25 | -0,0987 | 0,197 | 165,5 | 0 | 0,3989 | 0,198 | 0,201 |
| 167-170 | 167 | 0,25 | 0,0987 | 0,197 | 168,5 | 0,50 | 0,3521 | 0,175 | 0,170 |
| 170-173 | 170 | 0,74 | 0,2703 | 0,122 | 171,5 | 0,99 | 0,2444 | 0,121 | 0,120 |
| 173-176 | 173 | 1,24 | 0,3925 | 0,067 | 174,5 | 1,49 | 0,1315 | 0,065 | 0,064 |
| 176-179 | 176 | 1,74 | 0,4591 | 0,028 | 177,5 | 1,98 | 0,0562 | 0,028 | 0,028 |
| 179-182 | 179 | 2,23 | 0,4871 | 0,009 | 181,5 | 2,48 | 0,0184 | 0,007 | 0,010 |
| 182-185 | 182 | 2,73 | 0,4968 | 0,003 | 184,5 | 2,98 | 0,0047 | 0,002 | 0,003 |
| 185-188 | 185 | 3,22 | 0,4993 | 0,0006 | 187,5 | 3,47 | 0,0010 | 0,0005 | 0,001 |
| | 188 | 3,72 | 0,4999 | | | | | | |

## §5. SUURTE ARVUDE SEADUS.

### 1. Juhuslike suuruste keskmine ja hajuvus.

Vaatleme n juhuslikku sõltumatut suurust, mida märgime suurte tähtedega $X_1, X_2, \ldots, X_n$. Olgu nende suuruste keskväärtused vastavalt $a_1, a_2, \ldots, a_n$ ja dispersioonid $D_1, D_2, \ldots, D_n$. Leiame antud juhuslike suuruste aritmeetilise keskmise

$$\bar{X} = \frac{X_1 + X_2 + \ldots X_n}{n},$$

mis on samuti juhuslik suurus. Leiame suuruse $\bar{X}$ keskväärtuse ja dispersiooni. Keskväärtuse $E(\bar{X})$ ja dispersiooni $D(\bar{X})$ leidmisel kasutame nende omadusi. Saame

$$E(\bar{X}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \ldots + X_n) =$$

$$= \frac{1}{n}\left[E(X_1) + \ldots + E(X_2)\right] = \frac{a_1 + a_2 + \ldots + a_n}{n};$$

$$D(\bar{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2} \cdot \left[D(X_1) + \ldots + D(X_n)\right] =$$

$$= \frac{D_1 + D_2 \ldots + D_n}{n} \cdot \frac{1}{n}.$$

Tulemused on seega järgmised:

1) juhuslike suuruste aritmeetilise keskmise keskväärtus võrdub nende suuruste keskväärtuste aritmeetilise keskmisega;
2) juhuslike suuruste aritmeetilise keskmise dispersioon on n korda väiksem nende suuruste dispersioonide aritmeetilisest keskmisest.

Vaatleme erijuhtu. Suuruse X määramiseks teostame n mõõtmist. Selle suuruse keskväärtus a on teadmata. Kuna katsetulemused sisaldavad juhuslikke vigu, siis võime suuruse X mõõtmistulemusi $X_1, X_2, X_3 \ldots X_n$ pidada juhuslikeks sõltumatuteks suurusteks, millel on ühesugune jaotusseadus. Siis on nende keskväärtused ja dispersioonid võrdsed.

$$E(X_1) = E(X_2) = \ldots = E(X_n) = a$$

$$D(X_1) = D(X_2) = \ldots = D(X_n) = \sigma^2 .$$

Suuruse X väärtuste aritmeetilise keskmise keskväärtus

$$E(\bar{X}) = \frac{n.a}{n} = a$$

ja dispersioon

$$D(\bar{X}) = \frac{n.\sigma^2}{n} \cdot \frac{1}{n} = \frac{\sigma^2}{n} .$$

Võrdse keskväärtusega juhuslike suuruste aritmeetilise keskmise keskväärtus võrdub üksiksuuruse keskväärtusega ja dispersioon on n korda väiksem üksiksuuruse dispersioonist. Kuna dispersioon iseloomustab juhusliku suuruse hajuvust, siis on aritmeetiline keskmine usaldatavam, kui juhuslikud suurused, millest keskmine on leitud. Keskmise hajuvus on seda väiksem, mida suurema hulga juhuslike suuruste baasil ta on arvutatud.

Veelgi olulisemaid järeldusi saab teha mitmetest teoreemidest, mis on tuntud suurte arvude seaduse nime all. Neil teoreemidel on tähtis koht matemaatilises statistikas, kuna nad seovad tõenäosusteooria abstraktseid tulemusi katseandmetega, andes võimaluse ette näha arvatavaid tulemusi. Suurte arvude seadusega tutvumegi järgmistes punktides.

## 2. Bernoulli suurte arvude seadus.

Suurte arvude seaduse alla kuuluvatest teoreemidest on ajaliselt esimene Bernoulli teoreem (1713.a.), mida tänapäeval nimetatakse ka Bernoulli suurte arvude seaduseks.

Bernoulli suurte arvude seadus:

Kui korduvatel sõltumatutel katsetel on sündmuse A toimumise tõenäosus p konstantne, siis võib küllalt suure katsete arvu n korral ühele kui tahes lähedase tõenäosusega väita, et sündmuse relatiivse sageduse hälve sündmuse tõenäosuse suhtes saab absoluutväärtuse poolest väiksemaks kuitahes väikesest positiivsest arvust $\varepsilon$.

Tõestamiseks lähtume eelmises paragrahvis (p. 9) tuletatud seosest

$$\mathcal{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \approx 2\,\Phi(t)\,; \quad t = \varepsilon \cdot \sqrt{\frac{n}{pq}}\,.$$

Võrdus on seda täpsem, mida suurem on n (binomiaaljaotus läheneb normaaljaotusele katsete arvu n kasvamisel). Täpse võrduse saame üleminekul piirile, lastes $n \to \infty$. Kuid

$$\lim_{n\to\infty} t = \lim_{n\to\infty} \varepsilon \cdot \sqrt{\frac{n}{pq}} = \infty$$

ja Laplace'i funktsiooni väärtus lõpmata suurel argumendi väärtusel omab piirväärtusena 0,5. Seega tõepoolest

$$\lim_{n\to\infty} \mathcal{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1\,,$$

mis ongi Bernoulli teoreemi matemaatiliseks väljenduseks.

Bernoulli teoreemi ei saa tõlgendada nii, et katsete arvu kasvamisel $\frac{m}{n} \to p$. Isegi suure katsete arvu korral võib juhtuda, et relatiivne sagedus kaldub tõenäosusest p tunduvalt kõrvale. Teoreem väidab, et tõenäosus relatiivse sageduse $\frac{m}{n}$ tunduvaks kõrvalekaldumiseks tõenäosusest p on väga väike. See tähendab, et relatiivse sageduse tunduvad kõrvalekaldumised tõenäosusest esinevad küllalt suure katsete arvu korral väga harva, s.t. on praktiliselt võimatud.

Bernoulli suurte arvude seaduse kehtivust kinnitavad paljude teadlaste poolt sooritatud katsed. Toome mõnede katsete tulemused. Buffon (büffon) viskas 1777.a. münti 4040 korda. Vapp tuli esile 2048 ja kiri 1992 korda. Vapi esinemise relatiivne sagedus on $\frac{2048}{4040} = 0{,}507$ ja kirja esinemise relatiivne sagedus 0,493. Nii vapi kui kirja esiletuleku tõenäosus ühel viskel on 0,5.

Quetelet (ketlee) registreeris 1846.a. urnist kuulide võtmise tulemusi. Urnist, mis sisaldas 20 musta ja 20 valget kuuli, võeti kuule 4096 korda, kusjuures valge kuuli esiletuleku sagedus oli 2066, mustal kuulil 2030. Valge kuuli relatiivne sagedus 0,504 erineb valge kuuli esiletuleku tõenäosusest p = 0,5 ainult 0,004 võrra.

Romanovski registreeris 1919.a. nelja mündi üheaegset viset 20160 korda. Tähistame vapi ja kirja esinemise kombinatsioone järgmiselt: vapp kolmel mündil ja kiri ühel mündil - 3v. 1k. jne. Tabelis on toodud vastavad tõenäosused üksikkatsel ja katseandmete põhjal leitud relatiivsed sagedused

| | 4v. | 3v. 1k. | 2v. 2k. | 1v. 3k. | 4k. |
|---|---|---|---|---|---|
| p | 0,0625 | 0,2500 | 0,3750 | 0,2500 | 0,0625 |
| $\frac{n}{m}$ | 0,0586 | 0,2435 | 0,3761 | 0,2522 | 0,0695 |

Leidke iseseisvalt võimalike vapi ja kirja kombinatsioonide tõenäosused.

Kõik kirjeldatud katsed on heas kooskõlas Bernoulli suurte arvude seadusega.

## 3. Tšebõševi suurte arvude seadus.

Suurte arvude seaduse üldisemal kujul tõestas kuulus vene matemaatik Tšebõšev 1867. aastal. Tõestuse komplitseerituse tõttu ei saa me seda käesolevas kursuses esitada.

Käesoleva paragrahvi 1. punktis käsitlesime juhuslikke suurusi, millel on sama keskväärtus ja dispersioon. Selgus, et juhuslike suuruste keskmise hajuvus on seda väiksem, mida suuremast arvust suurustest keskmine on leitud. Tšebõševi teoreem (suurte arvude seadus) täpsustab asja veelgi.

Tšebõševi suurte arvude seadus:
Küllalt suure arvu võrdsete keskväärtuste ja dispersioonidega sõltumatute juhuslike suuruste puhul võib ühele kui tahes lähedase tõenäosusega väita, et nende suuruste aritmeetiline keskmine erineb ühisest keskväärtusest kui tahes vähe.

Kui tähistame juhuslike suuruste $X_1, X_2, \dots, X_n$ keskväärtuse tähega a ja nende aritmeetilise keskmise sümboliga $\bar{X}$, mis on määratud 1. punktis, siis saame Tšebõševi suurte arvude seaduse matemaatilises sümboolikas esitada kujul

$$\lim_{n \to \infty} P(|\bar{X} - a| < \varepsilon) = 1.$$

Ka suure n korral võib $\bar{X}$ oluliselt hälbida a suhtes, kuid selle tõenäosus on väga väike. Seda juhtub väga harva ehk see on praktiliselt võimatu.

Tšebõševi suurte arvude seadus on üldisem Bernoulli suurte arvude seadusest. Tšebõševi teoreem esitatud kujul kehtib kõigi juhuslike suuruste jaoks, mis rahuldavad eespool nimetatud tingimusi, Bernoulli teoreem käib ainult sündmuse relatiivse sageduse kohta korduvatel katsetel ( $\frac{m}{n}$ on samuti juhuslik suurus).

foodud suurte arvude seadus on erijuhuks veelgi üldisematel eeldustel kehtivast Tšebõševi teoreemist. Suurte arvude seadus Bernoulli ja Tšebõševi kujul moodustavad selle seaduse ühe külje. Teise külje moodustab Ljapunovi teoreem, mis väidab, et teatud tingimusi rahuldavate juhuslike suuruste summa jaotus on lähedane normaaljaotusele küllalt suure arvu suuruste korral. Üheks oluliseks tingimuseks on see, et ükski summasse kuuluvatest juhuslikest suurustest ei tohi domineerida.

Kõigi nimetatud teoreemide puhul on tegemist piirprotsessiga ja neid nimetataksegi sageli tõenäosusteooria piirteoreemideks. Ljapunovi teoreemi nimetatakse tema erakordse üldisuse tõttu ka tsentraalseks piirteoreemiks.

## 4. Suurte arvude seaduse praktiline tähtsus.

Tšebõševi teoreemist tuleneb rida tähtsaid järeldusi.

1. Juhuslike suuruste väärtused kõiguvad oma keskväärtuste suhtes mõlemale poole ja aritmeetilises keskmises need hälbed kustutavad üksteist. Seega on juhuslike suuruste aritmeetiline keskmine peaaegu vaba juhuslikkusest. Keskmises peaaegu ei väljendugi juhuslike suuruste konkreetsed iseärasused. Keskmine on vaadeldavale protsessile omase seaduspärasuse peegeldaja.

2. Katsel määratava suuruse väärtuseks võetakse mõõtmistulemuste aritmeetiline keskmine. Tšebõševi teoreemi kohaselt on selline talitusviis õigustatud. Olgu mõõtmistulemused $X_1$, $X_2$ ... , $X_n$. Me võime neid vaadelda n juhusliku suurusena. Kuna tavaliselt ühe katse tulemused ei mõjuta teiste katsete tulemusi, siis on ka nõue suuruste sõltumatusest täidetud. Kui mõõtmised on teostatud ilma süstemaatiliste vigadeta, s.t. kui tulemused kõiguvad arvatava tõelise väärtuse ümber mõlemale poole, on suuruste $X_1$, $X_2$ jne. keskväärtused võrdsed määratava suuruse tõelise väärtusega. Kui mõõteriistad garanteerivad teatud täpsuse, siis

on ka mõõtmistulemuste hajuvus piiratud ning võrdsetes mõõtmistingimustes võime lugeda iga üksikmõõtmise dispersiooni samaks. Seega on Tšebõševi teoreemi nõuded täidetud ja teoreemi enda kohaselt võime olla praktiliselt kindlad, et mõõtmistulemuste aritmeetiline keskmine erineb väga vähe suuruse tõelisest väärtusest.

3. Suure hulga samatüübiliste objektide omaduste üle otsustatakse nende hulgast valitud piiratud arvuga objektide järgi. Uuritavate objektide arv on samatüübiliste objektide üldarvuga võrreldes küll väike, kuid siiski küllalt suur üldiste järelduste tegemiseks.

Sel põhimõttel toimub näiteks taimede seemnete idanevuse kontroll, toodete kvaliteedi hindamine jne.

Bernoulli teoreemist järeldub, et küllalt suure katsete arvu puhul relatiivne sagedus erineb harilikult väga vähe sündmuse tõenäosusest üksikkatsel. Katseseeriate kordamisel saame enamikul juhtudel üksteisest vähe erinevad relatiivsed sagedused, mis on lähedased sündmuse tõenäosusele. Suuri lahkuminekuid esineb harva ehk praktiliselt neid ei esine. Siit tuleneb ka põhjendus sellele, et sündmuse tõenäosus loetakse küllalt suure arvu katsete puhul võrdseks katseandmetest leitud sündmuse relatiivse sagedusega.

Suurte arvude seaduse kehtivust on inimkond oma praktilises tegevuses lõpmata palju kordi kontrollinud.

## § 6. VÄLJAVÕTTELISE MEETODI MATEMAATILINE TEOORIA.

### 1. V ä l j a v õ t t e l i n e   m e e t o d.

1. näide. Ettevõtte ametiühingukomitee tahab välja selgitada töötajate korteriolusid. Selleks küsitletakse kõiki töötajaid. Küsitlusel saadud andmete põhjal saab teha vajalikke järeldusi ja leida vajalikke arvulisi näitajaid.

2. näide. Bussipark on huvitatud reisijate sõidumaa pikkusest. Ei ole mõeldav küsitleda kõiki reisijaid ja pealegi veel pikema aja jooksul. Küsitletakse ainult osa reisijaid üksikute päevade teatud tundidel. Saadud andmete põhjal tehtud järeldusi üldistatakse kõigile reisijatele. Ütleme, et kasutati väljavõttelist meetodit. Väljavõttelise meetodi idee on järgmine:

Olgu tegemist mingi objektide hulgaga, milles püüame määrata meid huvitavat tunnust X. Harilikult ei ole võimalik uurida kõiki objekte, vaid valime nende hulgast välja teatud osa, mida uurime vajaliku tunnuse seisukohalt. Saadud andmete põhjal tehtud järeldusi püüame üldistada kõigile samatüübilistele objektidele.

Vaatluse alla võetud objektid moodustavad väljavõtte; objektide hulk, millest valik tehakse, kannab üldkogumi nime. Kogumi elementide arvu nimetame edaspidi kogumi mahuks. Üldkogumi mahtu tähistame N, väljavõttel - n.

Väljavõttelise meetodiga uurime üldkogumit kaudselt - väljavõtte kaudu. Väljavõttes peab uuritav tunnus jaotuma ligikaudu samuti nagu üldkogumis. Väljavõtte ja üldkogumi vaatluse tulemuste kooskõla nimetame representatiivsuseks ja ütleme, et väljavõte representeerib üldkogumit (väljavõte on küllalt esinduslik).

Väljavõttelist meetodit kasutatakse laialdaselt tehnoloogiliste protsesside kontrollimisel, toodangu kvaliteedi määramisel, keemiliste ja füsioloogiliste eksperimentide korraldamisel, elanikkonna teenindamise efektiivsuse hindamisel jne.

Representatiivse väljavõtte tegemiseks kasutatakse mitmesuguseid võtteid. Nimetagem neist olulisemaid: 1) juhuslik korduv väljavõte, 2) juhuslik kordumatu väljavõte, 3) mehhaaniline väljavõte. Väljavõtte tegemisel tuleb silmas pidada, et elemendi valik üldkogumist ei tohi sõltuda uuritavast tunnusest ja üldkogumi elementidel peab olema väljavõttesse sattumiseks võrdne tõenäosus. Nende tingimuste garanteerimise võtteid käsitletakse statistika üldteoorias.

Juhuslik korduv väljavõte on selline, kus element pärast uurimist tagastatakse üldkogumisse ja ta võib seega uuesti sattuda väljavõttesse.

Juhuslikul kordumatul väljavõttel valitud elemendid üldkogumisse tagastamisele ei kuulu.

Mehhaanilisel väljavõttel valitakse väljavõttesse üldkogumi mingil viisil järjestatud elemente võrdsete intervallide tagant, see viis on oma iseloomult seega samuti kordumatu väljavõte.

## 2. Väljavõttelise meetodi kasutamisel esinevatest vigadest.

Tähistame tunnuse X keskmise väärtuse ja dispersiooni üldkogumis sümbolitega a ning $\sigma^2$, väljavõtus vastavalt $\bar{x}$ ja $s^2$.

Väljavõtte uurimisel saadud tulemuste täpsus sõltub kogumi mahust, uuritava tunnuse hajuvusest ja uurimise täpsusest.

Vaatluste hooletu teostamine, valearvestused, ja teiste isikute poolt andmete tahtlik moonutamine tingivad jämedaid vigu. Mõõteriistade ja vaatlusvahendite ebatäpsus ja vaatlejate individuaalsed iseärasused tingivad süstemaatilisi vigu - vaatluse resultaadid on ebatäpsed, kusjuures hälbimine toimub ühes suunas. Jämedaid ja süstemaatilisi vigu saab põhimõtteliselt vältida või vaatlusandmete läbitöötamisel arvestada. Edasistes arutlustes eeldame, et nimetatud vigu ei esine.

Lisanduvad juhuslikud vead, mis on tingitud paljudest juhuslikest, mittekontrollitavatest faktoritest. Juhuslike faktorite mõjul kõiguvad vaatlustulemused kord ühele, kord teisele poole ning nende kõikumiste koguefekt on väljavõttes küllalt suure mahu puhul suurte arvude seaduse järgi praktiliselt null.

Väljavõtte matemaatiline teooria pöörab peatähelepanu nn. representatiivsuse vigade määramisele ja hindamisele. Representatiivsuse vigadeks nimetame lahkuminekuid üldkogumi ja väljavõtte karakteristikute vahel. Nad on tingitud sellest, et väljavõte ei kopeeri täpselt üldkogumit. Üldkogumis on a ja $\sigma^2$ jäävad suurused, kuigi meile harilikult tundmatud. Väljavõtteid saame moodustada erinevalt ja üldiselt on $\bar{x}$ ning $s^2$ varieeruvad, nad sisaldavad juhuslikkuse elementi. Tähistame üldkogumi mingi karakteristiku tähega K, väljavõttes olgu vastav karakteristik $k$, mis on üldkogumi karakteristiku statistiliseks hinnanguks. Väljavõtte korduval moodustamisel saame erinevad k väärtused $k_1$, $k_2$ ...... Karakteristik k on juhuslik suurus ja tema väärtused grupeeruvad keskväärtuse E(k) ümber. Selleks, et karakteristk k oleks üldkogumi karakteristiku K nihutamata statistiliseks hinnanguks, on vaja, et

$$E(k) = K.$$

Karakteristiku k väärtused varieeruvad ja me võime rääkida ligikaudsest võrdusest $k \approx K$. Representatiivsuse viga $|K - k|$ on samuti juhuslik suurus ja me saame hinnata, kui tõenäone on, et $|K - k|$ ei ületa antud tõket $\varepsilon$, s.t.

$$P = P(|K - k| < \varepsilon)$$

Tõenäosust P nimetame väljavõtte usaldusastmeks, tõket $\varepsilon$ väljavõtte maksimaalseks veaks antud tõenäosusega. Tingimust $|K - k| < \varepsilon$ võime vaadelda kujul

$$-\varepsilon < K - k < \varepsilon$$

ehk

$$k - \varepsilon < K < k + \varepsilon \quad .$$

Mida väiksem on $\varepsilon$, seda täpsemalt on määratud üldkogumi karakteristik K ehk seda vähem erinevad teineteisest usalduspiirid $k - \varepsilon$ ja $k + \varepsilon$ antud tõenäosusega P.

Väljavõtte matemaatilise teooria peamisteks ülesanneteks on

1) määrata representatiivsed väljavõtte karakteristikud,
2) hinnata leitud karakteristikute täpsust antud usaldusastmega,
3) määrata väljavõtte maht nii, et ta antud tingimustes representeeriks üldkogumit vajaliku täpsusega.

## 3. Ü l d k e s k m i s e  h i n d a m i n e .

Üldkeskmise nihutamata statistiliseks hinnanguks tuleb valida selline väljavõtte karakteristik k, et kehtiks võrdus

$$E(k) = a.$$

Suurte arvude seaduse käsitlemisel näitasime, et n juhusliku suuruse aritmeetilise keskmise keskväärtus on võrdne nende suuruste keskväärtuste aritmeetilise keskmisega. Üldkogumi elementidel on võrdne võimalus sattuda väljavõttesse ja seega on väljavõtte iga elemendi puhul uuritava tunnuse X väärtus $x_i$ juhuslik suurus ning tema keskväärtuseks on üldkogumi keskmise a, s.t.

$$E(x_i) = a \qquad (i = 1,2, \ldots , n).$$

Tunnuse X väärtuste aritmeetiline keskmine on

$$\bar{x} = \frac{\sum x_i}{n} \quad ;$$

nende väärtuste keskväärtuste keskmine aga

$$\frac{a + a + \ldots + a}{n} = \frac{n \cdot a}{n} = a .$$

Seega eespool nimetatud teoreemi põhjal on $E(\bar{x}) = a$. Võrdleme viimast tulemust lähteseosega $E(k) = a$. Näeme, et üldkeskmise a nihutamata statistiliseks hinnanguks tuleb valida väljavõttekeskmine $\bar{x}$.

Väljavõttekeskmise leiame väljavõtte uurimisel saadud variatsioonrea andmetel.

## 4. Ü l d d i s p e r s i o o n i  h i n d a m i n e .

Väljavõtte andmetel saame leida väljavõtte dispersiooni $s^2$. Saab näidata, et $s^2$ keskväärtus ei lange kokku üld-dispersiooniga, s.t.

$$E(s^2) \neq \sigma^2 ,$$

Üldispersiooni $\sigma^2$ nihutamata statistiliseks hinnanguks võib võtta korduval väljavõttel suuruse

$$\frac{n}{n-1} s^2 ,$$

kordumatul väljavõttel

$$\frac{n}{n-1} \cdot \frac{N-1}{N} s^2 .$$

Kui üldkogumi maht N on küllalt suur, siis

$$\frac{N-1}{N} = 1 - \frac{1}{N} \longrightarrow 1$$

ja võime üldispersiooni kordumatul väljavõttel leida samast seosest nagu korduval väljavõttel.

Väljavõttedispersiooni leiame valemist

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

Pannes $s^2$ avaldise ülddispersiooni ligikaudsesse avaldisse, saame

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1}, \qquad \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}}.$$

Seega leiame ülddispersiooni väljavõtte andmetel samuti nagu väljavõttedispersiooni, ainult elementide arv tuleb võtta ühe võrra väiksem väljavõtte mahust. Kui väljavõte on küllalt suur, siis olulist erinevust $\sigma^2$ ja $s^2$ vahel ei ole.

## 5. Väljavõttekeskmise dispersioon.

Väljavõttekeskmine $\bar{x}$ on juhuslik suurus, mille keskväärtus on a, kuid $\bar{x}$ jaotus on teadmata. Väljavõttekeskmise hajuvust üldkeskmise ümber hindame väljavõttekeskmise dispersiooni abil, mida tähistame sümboliga $\sigma^2_{\bar{x}}$.

Vaatleme eraldi korduvat ja kordumatut väljavõtet.

Korduv väljavõte.

Korduval väljavõttel on üldkogumi elementidel igal katsel võrdne tõenäosus sattuda väljavõttesse ning tunnuse X üksikväärtused on üksteisest sõltumatud juhuslikud suurused. Suurte arvude seaduse käsitlemisel näitasime, et n sõltumatu juhusliku suuruse aritmeetilise keskmise dispersioon on n korda väiksem üksiksuuruste dispersioonist. Tunnuse X iga üksikväärtuse $x_i$ dispersiooniks üldkogumis loeme $\sigma^2$. Aritmeetilise keskmise $\bar{x}$ dispersioon oma keskväärtuse a suhtes on seega

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

- väljavõttekeskmise dispersioon on korduval väljavõttel n korda väiksem tunnuse dispersioonist üldkogumis.

Kordumatu väljavõte.

Kordumatul väljavõttel elemente üldkogumisse ei tagastata ja iga järgmise elemendi sattumine väljavõttesse sõltub eelnevate katsete tulemustest. Ei saa enam rääkida tunnuse X väärtuste sõltumatusest. Siis aga ei saa kasutada ka eelmist mõttekäiku.

Saab näidata, et kordumatul väljavõttel keskmise dispersioon üldkeskmise suhtes avaldub kujul

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} .$$

Võrreldes väljavõttekeskmise dispersioone korduval ja kordumatul väljavõttel, näeme, et kordumatul väljavõttel on viga väiksem kui korduval väljavõttel.

Murd $\frac{N-n}{N-1} < 1$ ja seega

$$\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} < \frac{\sigma^2}{n} .$$

Kui üldkogumi maht on suur ja väljavõtte mahust tunduvalt suurem, siis $N \approx N-1$ ja

$$\frac{N-n}{N-1} \approx \frac{N-n}{N} = 1 - \frac{n}{N} \rightarrow 1 .$$

Väljavõtte keskmise dispersioon sõltub üldkogumi dispersioonist $\sigma^2$ ja väljavõtte mahust $n$.

Mida suurem on väljavõtte maht n, seda vähem hälbib $\bar{x}$ üldkeskmise a suhtes. Mida väiksem on tunnuse dispersioon üldkogumis, seda väiksem on ka $\bar{x}$ hälbimine a suhtes.

Väljavõttekeskmise $\bar{x}$ kooskõla üldkeskmisega a ehk $\bar{x}$ täpsust hindame standardhälbe $\sigma_{\bar{x}}$ abil. Standardhälve

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Suurust $\sigma_{\bar{x}}$ nimetatakse ka keskmiseks väljavõtteveaks. Kuna n väike suurendamine ei muuda oluliselt $\sqrt{n}$ väärtust, siis järeldub sellest, et väljavõtte mahu väike suurendamine ei suurenda oluliselt keskmise $\bar{x}$ täpsust. Soovides näiteks suurendada keskmise täpsust 10 korda, peaksime suurendama väljavõtte mahtu 100 korda ( $\sqrt{100} = 10$ ).

Juhime veel tähelepanu suuruste $\sigma^2$, $s^2$ ja $\sigma^2_{\bar{x}}$ erinevale ühendusele:

$\sigma^2$ iseloomustab tunnuse väärtuste hajuvust üldkogumis üldkeskmise a ümber;

$s^2$ - tunnuse väärtuste hajuvust väljavõttes väljavõttekeskmise $\bar{x}$ ümber;

$\sigma^2_{\bar{x}}$ - väljavõttekeskmiste $\bar{x}$ hajuvust üldkeskmise a ümber.

## 6. Keskmise väljavõttevea leidmine väljavõtte standardhälbe kaudu.

Eelmises punktis näitasime, kuidas leida $^{2}_{\bar{x}}$ ülddispersiooni $^{2}$ kaudu. Kuna $^{2}$ on harilikult teadmata ja teda leiame ligikaudu $s^2$ kaudu, siis on otstarbekas anda eeskiri $^{2}_{\bar{x}}$ ja $_{\bar{x}}$ leidmiseks vahetult $s^2$ kaudu. Selleks kasutame 4. punktis saadud $^{2}$ ja $s^2$ vahelist seost ning 5. punkti valemeid $^{2}_{\bar{x}}$ ja $^{2}$ seose kohta.

Korduval väljavõttel

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{n \cdot s^2}{(n-1)n} = \frac{s^2}{n-1}$$

ja kordumatul väljavõttel

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \left(\frac{n}{n-1} \cdot \frac{N-1}{N} s^2\right) \frac{1}{n} \cdot \frac{N-n}{N-1} =$$

$$= \frac{s^2}{n-1} \cdot \frac{N-n}{N} = \frac{s^2}{n-1}\left(1 - \frac{n}{N}\right).$$

Kokkuvõte. Dispersiooni ja standardhälbe arvutamise eeskirjad on järgmised:

korduval väljavõttel

$$\sigma^2 = \frac{n}{n-1}\, s^2 ;$$

$$\sigma_{\bar{x}}^2 = \frac{s^2}{n-1} ; \qquad \sigma_{\bar{x}} = \sqrt{\frac{s^2}{n-1}} ;$$

kordumatul väljavõttel:

$$\sigma^2 = \frac{n}{n-1}\left(1 - \frac{1}{N}\right) s^2 ;$$

$$\sigma_{\bar{x}}^2 = \frac{s^2}{n-1}\left(1 - \frac{n}{N}\right) ; \qquad \sigma_{\bar{x}} = \sqrt{\frac{s^2}{n-1}\left(1 - \frac{n}{N}\right)} \approx$$

$$\approx \sqrt{\frac{s^2}{n-1}} .$$

Suuruste $\bar{x}$ ja $s^2$ arvutamine toimub tavalise eeskirja järgi väljavõtte jaoks koostatud variatsioonrea andmetel. Eespool (2. punktis) loetletud ülesannetest saame nüüd lahendada esimest.

## 7. Väljavõttekeskmise usaldatavus.

Olles leidnud **väljavõttekeskmise** $\bar{x}$ , peame hindama, millistes piirides võib asuda üldkeskmine teatud tõenäosusega. See tähendab 2. punktis loetletud väljavõtteteooria teise ülesande lahendamist.

Ülesanne on järgmine: Määrata tõkked, mida üldkeskmise ja

väljavõttekeskmise erinevus ei ületa antud tõenäosusega. Tõkkeid, mille vahel üldkeskmine a võib asuda, nimetatakse tema usalduspiirideks ja usalduspiirides asumise tõenäosust keskmise usaldusnivooks. Keskmise a usalduspiirideks on $\bar{x} + \varepsilon$ ja $\bar{x} - \varepsilon$ ; usaldusnivoo tähistame P-ga. Vahe $|\bar{x} - a|$ ei tohi ületada tõket $\varepsilon$ usaldusnivooga P ning arvu $\varepsilon$ nimetatakse väljavõtte piirveaks.
Formuleeritud ülesande võime sõnastada ka järgmiselt: leida väljavõtte piirviga antud usaldusnivooga. Väljavõttekeskmine allub normaaljaotusele või on sellele lähedane ning piirvea $\varepsilon$ leiame nii nagu normaaljaotuse puhul hindame suuruse väärtuste hälbeid keskväärtuse suhtes. Käesoleval juhul on suuruse väärtuste osas võimalikud väljavõttekeskmised $\bar{x}$ ja nende keskväärtuseks üldkeskmine a.

Tõenäosuse selleks, et $|\bar{x} - a| < \varepsilon$ leiame Laplace'i funktsiooni abil:

$$P(|\bar{x} - a| < \varepsilon) = 2\,\Phi(t).$$

Teades usaldusnivood P ja väljavõtteviga $\sigma_{\bar{x}}$, leiame

1) võrduse $P = 2\,\Phi(t)$ abil funktsiooni $\Phi(t)$ tabelist tõenäosuse väärtusele P vastava argumendi väärtuse t;

2) seosest $$t = \frac{\varepsilon}{\sigma_{\bar{x}}}$$

väljavõtte piirvea

$$\varepsilon = t \cdot \sigma_{\bar{x}}$$

(vt. Laplace'i integraalvalemi rakendusi).
Olles leidnud $\varepsilon$ saame a usalduspiirid $\bar{x} \pm \varepsilon$.

Kui on ette antud piirviga $\varepsilon$ või usalduspiirid, leiame üldkeskmise antud usalduspiiridesse kuulumise tõenäosuse (usaldusnivoo) järgmiselt:



18.

1) leiame argumendi

$$t = \frac{\varepsilon}{\sigma_{\bar{x}}};$$

2) leiame Laplace'i funktsiooni tabelist $2\,\Phi(t)$, mis ongi usaldusnivooks P.

Mida suurem on usaldusnivoo, seda laiemad on usalduspiirid (pikemas vahemikus asumist on kergem garanteerida).

## 8. Näiteid.

Lahendame mõned ülesanded eespool käsitletud kahe väljavõtteteooria põhiülesande kohta.

Nisu saagikuse määramiseks 10 000 ha suuruselt pindalalt tehti kontrollmõõtmised 1000 hektaril, mis valiti korduva väljavõtte teel. Leida keskmise saagi usalduspiirid usaldusastmega 1) 0,997 ja 2) 0,9 . Vaatlusandmed on arvutusskeemi kahes esimeses lahtris.

Alustame $\bar{x}$ ja $s^2$ leidmisest.

| $x$ Saak (ts) | $f$ Ha arv | $\frac{x-16}{2}$ | $\frac{x-16}{2}\frac{f}{10}$ | $\left(\frac{x-16}{2}\right)^2\cdot\frac{f}{10}$ |
|---|---|---|---|---|
| 11 - 13 | 150 | - 2 | - 30 | 60 |
| 13 - 15 | 200 | - 1 | - 20 | 20 |
| 15 - 17 | 450 | 0 | - 50 | |
| 17 - 19 | 200 | 1 | 20 | 20 |
| | 1000 | | - 30 | 100 |

Keskmine

$$\bar{x} = \frac{-30}{100}\cdot 2 + 16 = 15{,}4 \text{ ts}$$

dispersioon

$$s^2 = \frac{100}{100}\cdot 4 - (16 - 15{,}4)^2 = 3{,}64 \text{ ts}^2$$

Leiame väljavõttevea

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{3,64}{999}} \approx 0,06 \text{ ts}$$

Nüüd saame leida väljavõtte piirvea $\varepsilon$ :

1) $P = 0,997$ .

Laplace'i funktsiooni tabelist leiame antud tõenäosusele vastava argumendi väärtuse $t = 3$ ja

$$\varepsilon = t \cdot \sigma_{\bar{x}} = 3 \cdot 0,06 = 0,18 \text{ ts};$$

2) $P = 0,91$.

Analoogiliselt leiame, et $t = 1,67$ ja $\varepsilon = 1,67 \cdot 0,06 = 0,1$ ts.

Keskmise saagi usalduspiirid on siis

1) $15,4 - 0,18 < a < 15,4 + 0,18$ ; $15,22 < a < 15,58$

2) $15,4 - 0,1 < a < 15,4 + 0,1$ ; $15,3 < a < 15,5$ .

(Pöörake tähelepanu sellele, kuidas P muutmine mõjutab piirvea ja usaldusvahemiku muutumist!).

Leida eelmise ülesande andmetel piirveale $\varepsilon = 0,15$ ts vastav usaldusnivoo.

Arvutame

$$t = \frac{\varepsilon}{\sigma_{\bar{x}}} = \frac{0,15}{0,06} = 2,5$$

ja tabelist

$$P = 2\Phi(2,5) = 0,9876.$$

Viimast tulemust tõlgendame järgmiselt:

Kui moodustada veel väljavõtte, siis 98,76% juhtudest saame üldkeskmisele usalduspiirideks $15,4 \pm 0,15$ ts ja ainult 1,2% juhtudest võib anda sellised usalduspiirid, mis ei sisalda üldist keskmist saaki.

## 9. Tunnuse relatiivse sageduse hindamine.

Rea probleemide uurimisel me ei huvitu teatud tunnuse aritmeetilisest keskmisest üldkogumis, vaid sellest, kui suure osa moodustavad vastavat tunnust omavad elemendid üldkogumist. Nii ollakse huvitatud praagi protsendist ettevõtte kogutoodangus, eriharidusega töötajate arvust antud kutsealal jne.

Olgu N elemendist koosnevas üldkogumis M elemendil ühine tunnus X. Selle tunnuse relatiivne sagedus on

$$p = \frac{M}{N} .$$

Väljavõtte uurimisel saame sama tunnuse relatiivseks sageduseks

$$w = \frac{m}{n} .$$

Oleme huvitatud tunnuse relatiivse sageduse p hindamisest üldkogumis tunnuse relatiivse sageduse w järgi väljavõttes antud usaldusastmega.

Vahe $|w - p|$ piirvea $\varepsilon$ ja p usalduspiiride määramiseks tuleb leida relatiivse sageduse w viga $\sigma_w$ ( sama tähendus w jaoks nagu $\sigma_{\bar{x}}$ keskmise $\bar{x}$ jaoks). Relatiivse sageduse dispersiooni $s^2$ (w) leidsime binomiaaljaotuse käsitlemisel

$$s^2 (w) = w \cdot (1-w)$$

ning analoogiliselt $\sigma_{\bar{x}}$ leidmisele korduval väljavõttel

$$\sigma_w = \sqrt{\frac{s^2(w)}{n}} = \sqrt{\frac{w(1-w)}{n}}$$

kordumatul väljavõttel

$$\sigma_w = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} .$$

Edasi toimuvad piirvea leidmine ja muud arutlused samal põhimõttel nagu $\bar{x}$ jaoks.

Näide: Selleks, et selgitada 1000 rbl. mitteületavate hoiuste protsenti hoiuste üldarvust, vaadeldi 900 hoiuarvet. Selgus, et 30% hoiustest ei ületanud 1000 rbl. Kui suure usaldusastmega võib väita, et selliste hoiuste osatähtsus kõigi hoiustajate hulgas ei erine leitud väljavõttelisest tulemusest rohkem kui 2% ?

Leiame

$$\sigma_w = \sqrt{\frac{0{,}3(1-0{,}3)}{900}} \approx 0{,}0153$$

piirviga $\varepsilon = 2\% = 0{,}02$ ja

$$t = \frac{\varepsilon}{\sigma_w} = \frac{0{,}02}{0{,}0153} = 1{,}31$$

Tabelist leiame, et $2\,\Phi(1{,}31) = 0{,}8098$, mis ongi otsitav tõenäosus.

## 10. Väljavõtte mahu määramine.

Ülesanne on järgmine. Mitu elementi tuleb väljavõttesse valida, et väljavõtte piirviga ei ületaks antud usalduspiire antud usaldusnivooga?

Vaatleme eraldi korduvat ja kordumatut väljavõtet.

Korduv väljavõte.

Antud usaldusastme P järgi leiame t ja seosest

$$\varepsilon = t\,\sigma_{\bar{x}}$$

väljavõtte mahu n, arvestades, et

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

Saame

$$\varepsilon = t\sqrt{\frac{\sigma^2}{n}},$$

kust

$$n = \frac{t^2\sigma^2}{\varepsilon^2}.$$

Kordumatu väljavõte.

Mõttekäik on analoogiline eelmisega; ainult et

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Lahendades võrrandi

$$\varepsilon = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$$

tundmatu n suhtes saame tulemuseks

$$n = \frac{N \cdot t^2 \sigma^2}{N \varepsilon^2 + t^2 \sigma^2}.$$

Näide: Olgu üldkogumi maht $N = 1000$ ja tunnuse dispersioon üldkogumis $\sigma^2 = 5$. Määrame vajaliku väljavõtte mahu n, et garanteerida väljavõtte piirvea $\varepsilon = 0{,}1$ mitteületamist usaldusastmega 0,997.

Korduv väljavõte.

Mahu n leidmiseks määrame tabelist $t = 3$ ja

$$n = \frac{3^2 \cdot 5}{0{,}1^2} = 4500.$$

Seega 1000 elemendist koosneva üldkogumi puhul ei saagi moodustada sellist väljavõtet korduval väljavõttel, mis garanteeriks nii väikese piirvea väga kõrge usaldusastmega.

Kordumatu väljavõte.

Maht

$$n = \frac{1000 \cdot 3^2 \cdot 5}{1000 \cdot 0{,}1^2 + 3^2 \cdot 5} = \frac{9000}{11} = 834,$$

mis erineb vähe üldkogumi mahust N.

Olgu üldkogumis $N = 100\ 000$ elementi. Siin saaksime vastavateks arvudeks 4500 ja 4306. Seega pole mahtude erinevus korduval ja kordumatul väljavõttel enam suur. Tulemus aga näitab veel kord seda, et kordumatu väljavõtte puhul saavutame sama täpseid tulemusi mis korduval väljavõttel väiksemat arvu elemente uurides.

## 11. Üldisi järeldusi.

Selgitame, kuidas mõjutavad üksteist suurused n, $\sigma$, $\varepsilon$ ja P. Valime kaks suurust konstantseteks (loeme antuteks) ja uurime, kuidas teised 2 suurust teineteist vastastikku mõjutavad. Suuruselt P võime kõikjal üle minna argumendile t, kusjuures suuremale P-le vastab suurem t väärtus. Uuritavad suurused on seotud valemis

$$n = \frac{t^2 \sigma^2}{\varepsilon^2} .$$

Antud on P ja $\sigma$. Mahu n suurendamine tingib $\varepsilon$ vähenemise (murru lugeja on const). Sisuline tähendus: väljavõtte mahu suurendamisega kaasneb usaldusnivoo muutumatuks jäädes tulemuste täpsuse kasv (piirviga väheneb).

Antud on P ja $\varepsilon$. Suurendades $\sigma^2$ suureneb ka n - suurema hajuvusega tunnuste jaoks tuleb sama usaldusnivooga täpsuse tagamiseks moodustada suurem väljavõte.

Antud on $\sigma$ ja $\varepsilon$. Muutumatu täpsuse korral saab usaldusnivood tõsta väljavõtte mahtu suurendades.

Antud on $\varepsilon$ ja n. Tunnuse hajuvuse suurenemine vähendab vaatlustulemuste usaldusnivood.

Analüüsige $\sigma$ ja $\varepsilon$ vahekorda antud P ning n puhul, samuti P ja $\varepsilon$ vahekorda antud $\sigma$ ning n puhul!

## § 7. KORRELATSIOONITEOORIA ELEMENTE.

### 1. Statistiline sõltuvus.

Matemaatika üheks tähtsamaks mõisteks on funktsiooni mõiste. Kahe suuruse vahelist sõltuvust nimetatakse funktsionaalseks, kui ühe suuruse (argumendi) igale väärtusele vastab teise suuruse (funktsiooni) kindel väärtus.

Näiteid: 1. Antud rahasumma (m) korral on ostetava kauba hulk (y) ja kauba hind (x) pöördvõrdelises sõltuvuses, mida saab analüütiliselt esitada kujul

$$y = \frac{m}{x} .$$

2. Kera ruumala (V) ja raadiuse (R) sõltuvus väljendub kujul

$$V = \frac{4}{3} \pi R^3 .$$

3. Olgu kapitali algväärtus teatud ajamomendil $y_0$ ning toimugu kapitali suurenemine ainult kasumi näol konstantse kasuminormiga 100 p% aastas. Kapitali (y) suuruse sõltuvus ajast (t) esitub tehtud eeldustel valemiga

$$y = y_0 (1 + p)^t .$$

Funktsionaalne sõltuvus kahe suuruse vahel on matemaatiline abstraktsioon, mis peegeldab reaalseid seoseid suuruste vahel.

Leidub aga palju suurusi, mis on üksteisest sõltuvad, kuid ei ole täidetud funktsionaalse sõltuvuse definitsioonis rõhutatud tingimus, et ühe suuruse igale väärtusele vastab teise suuruse kindel väärtus.

Näiteid:

4. Inimese raskus sõltub tema pikkusest.

5. Perekonnas tarbitav piima hulk sõltub perekonna suurusest.

6. Samatüübilistes ettevõtetes toodangu mahu (x) kasv tingib omahinna (y) alanemise. Kuid sama toodangu mahuga ettevõtetes on omahind erinev. See erinevus on tingitud sellest, et tootmistingimused ei ole täpselt samad, omahinda mõjutavad peale toodangu mahu ka teised faktorid.

7. Ettevõtte tööliste arv sõltub ettevõtte põhifondide suurusest.

8. Omahind on seotud tööviljakusega.

9. Teravilja saagikus sõltub väetise hulgast. Toome tabeli, mis iseloomustab rukkisaagi (y) sõltuvust maale antud sõnniku hulgast (x)

| x $\frac{t}{ha}$ | 10 | 10 | 12 | 12 | 18 | 20 | 20 | 25 | 30 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| y $\frac{ts}{ha}$ | 8 | 12 | 10 | 15 | 14 | 16 | 18 | 18 | 16 | 20 |

Samadele sõnnikukogustele vastavad erinevad saagid. Puudub üksühene vastavus suuruste x ja y vahel.

Toodud näited kuuluvad nn. statistiliste sõltuvuste hulka. Kahe suuruse vahelist sõltuvust nimetatakse statistiliseks, kui ühe suuruse igale antud väärtusele vastab teise suuruse väärtuste jaotus, mis muutub koos esimese suuruse muutumisega.

10. Toome andmed omahinna (y) ja päevase toodangu mahu (x) kohta 15 samatüübilises ettevõttes.

| Toodangu maht (x) | 200 | 400 | 300 | 200 | 300 | 200 | 400 | 800 |
|---|---|---|---|---|---|---|---|---|
| Omahind (y) | 150 | 160 | 150 | 160 | 160 | 180 | 150 | 140 |

| Toodangu maht (x) | 500 | 800 | 200 | 300 | 400 | 500 | 800 |
|---|---|---|---|---|---|---|---|
| Omahind (y) | 140 | 140 | 180 | 160 | 150 | 150 | 140 |

Samadele x väärtustele vastavad erinevad y väärtused ja vastupidi ehk ühe suuruse kindlatele väärtustele vastavad teise suuruse erinevad jaotused. Nii vastab väärtusele x = 200 suuruse y jaotus

| y | 150 | 160 | 180 |
|---|---|---|---|
| f | 1 | 1 | 2 |

;

väärtusele x = 300

| y | 150 | 160 |
|---|---|---|
| f | 1 | 2 |

jne.

Omahinna y väärtustele vastab toodangu mahu x väärtuste jaotus. Väärtusele y = 150 vastav x jaotus on

| x | 200 | 300 | 400 | 500 |
|---|---|---|---|---|
| f | 1 | 1 | 2 | 1 |

jne.

Sageli tuleb üldkogumit uurida mitme tunnuse järgi. Harilikult tuntakse siis huvi, kuidas ühe tunnuse muutumine mõjutab teise tunnuse muutumist ehk kuidas uuritavad tunnused on teineteisega seotud. Eelmises näites ongi toodud andmed, mis on saadud ettevõtete uurimisel omahinna ja toodangu mahu järgi. Saame selgitada, kuidas toodangu mahu muutumine mõjutab omahinna muutumist või vastupidi. Sellises käsitluses jätame kõrvale teised faktorid, mis mõjutavad nii omahinda kui ka toodangu mahtu. Arvestamata faktorite koguefekt aga tingibki seda, et ühe suuruse igale väärtusele vastab teise suuruse väärtuste jaotus, mitte aga enam kindel väärtus.

Statistilise sõltuvuse üheks piirjuhuks on funktsionaalne sõltuvus: funktsiooni muutumine on tingitud argumendi muutumisest, kui lisafaktoreid mõjumas ei ole. Teiseks piirjuhuks on suuruste täielik sõltumatus: ühe suuruse muutumine ei avalda mingit mõju teise suuruse muutumisele. Näiteks tehase toodangu maht ei sõltu sellest, kui suur on valvepersonal tehase sissepääsu juures.

## 2. L ä h t e a n d m e t e  e s i t u s v i i s e.

Üldkogumi uurimisel saame harilikult andmed tunnuste vastavate väärtuste registreerimise järjekorras nn. loendustabeli näol (näiteks eelmise punkti 10. näites toodud andmed). Üldiselt võime loendustabeli esitada kujul

| $x$ | $x_1$ | $x_2 \ldots\ldots x_N$ |
|---|---|---|
| $y$ | $y_1$ | $y_2 \ldots\ldots y_N$ |

,

kus N on väärtuste üldarv, mille hulgas võib esineda ka kordseid väärtusi.

Selline tabel sobib suuruste vahelise sõltuvuse täpsemaks uurimiseks siis, kui ühe suuruse väärtusi ei esine kordselt ja selle suuruse väärtused on esitatud järjestatult või siis, kui väärtuste hulk on väike. Sellist tabelit nimetatakse ka lihtsaks korrelatsioonitabeliks.

Korrelatsioonitabel. Lihtsa korrelatsioonitabeli andmed esitatakse korrastatult mõlema tunnuse järgi. Seejuures ei ole oluline, kas tunnuste väärtused on antud intervallitult või üksikväärtustena. Üks võimalus on eraldi välja kirjutada ühe suuruse jaotustabelid teise suuruse üksikväärtuste järgi ja vastupidi (samal põhimõttel, nagu eelmise punkti 10. näites alustasime). Ühe tunnuse üksikutele väärtustele vastavaid teise tunnuse jaotusi nimetatakse tinglikeks jaotusteks. Palju ülevaatlikum ja kompaktsem on esitada mõlema suuruse järgi koostatud tinglikud jaotused ühise tabelina. Valitud näite puhul saame sellise koondtabeli järgmisel kujul:

| x \ y | 140 | 150 | 160 | 180 | |
|---|---|---|---|---|---|
| 200 | - | 1 | 1 | 2 | 4 |
| 300 | - | 1 | 2 | - | 3 |
| 400 | - | 2 | 1 | - | 3 |
| 500 | 1 | 1 | - | - | 2 |
| 800 | 3 | - | - | - | 3 |
| | 4 | 5 | 4 | 2 | 15 |

Selliselt koostatud tabelit nimetatakse korrelatsioonitabeliks. Ridade ja veergude lõikekohas olevad arvud näitavad ettevõtete arvu, millele vastab vaadeldav suuruste x ja y väärtuste paar. Nii on omahinnaga y = 160 ja toodangu mahuga x = 300 kaks ettevõtet jne. Üksikutes veergudes ja ridades olevate sageduste liitmisel saame vastavalt võrdse omahinnaga (y) või võrdse toodangu mahuga (x) ettevõtete arvu. Nii on 5 ettevõttes toodangu omahind y = 150, 3 ettevõttes on toodangu maht x = 400. Sageduste summa alumises reas ja viimases veerus on alati võrdne uuritavate elementide arvuga.

Kui tunnuste väärtused on intervallide kaupa, siis valitakse tunnuste diskreetseteks väärtusteks intervallide keskkohad.

Üldisel kujul võib korrelatsioonitabeli esitada järgmiselt:

| x \ y | $y_1$ | $y_2$ ..... | $y_j$ ..... | $y_l$ | $m_x$ |
|---|---|---|---|---|---|
| $x_1$ | $f_{11}$ | $f_{12}$ | $f_{1j}$ | $f_{1l}$ | $m_1$ |
| $x_2$ | $f_{21}$ | . . . . . . . . . . | | $f_{2l}$ | $m_2$ |
| $x_i$ | | | $f_{ij}$ | $f_{il}$ | $m_i$ |
| $x_k$ | $f_{k1}$ | . . . . . | $f_{kj}$ | $f_{kl}$ | $m_k$ |
| $n_y$ | $n_1$ | $n_2$ | $n_j$ | $n_l$ | N |

Sageduste $f_{xy}$ liitmine tulpade kaupa annab sageduste summad $n_y$; sageduste $f_{xy}$ liitmine ridade kaupa annab summad $m_x$. Seega

$$n_j = f_{1j} + f_{2j} + \dots + f_{ij} \dots f_{kj} = \sum_i f_{ij};$$

$$m_i = f_{i1} + f_{i2} + \dots + f_{ij} + \dots f_{il} = \sum_j f_{ij}$$

Täht summa märgi all näitab, kumma indeksi järgi liitmine on teostatud.

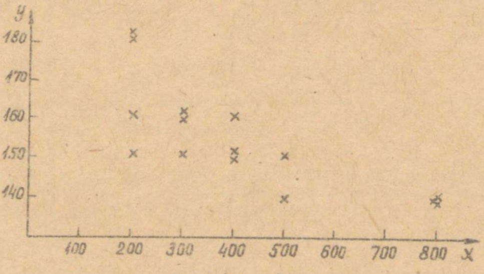
Sageduste $m_i$ summa on võrdne sageduste $n_j$ summaga

$$m_1 + m_2 + \dots m_k = n_1 + n_2 + \dots + n_l = N$$

ehk

$$\sum_i m_i = \sum_j n_j = N.$$

Korrelatsiooniväli. Sageli on otstarbekas statistilist sõltuvust esitada geomeetriliselt. Iga tunnuste kokkukuuluvate väärtuste paari $(x_i, y_j)$ tõlgendame punkti koordinaatidena ristkoordinaadistikus. Igale väärtuste paarile vastab nii palju üksteisega kattuvaid punkte, kui suur on vastav sagedus $f_{xy}$ ehk teisiti väljendades: igal punktil $(x_i, y_j)$ on kaal $f_{xy}$. Kaalud võime märkida punktidele juurde või väikese arvu punktide korral märgime punktid üksteise ligidusse. Ettevõtete näitele vastab korrelatsiooniväli joonisel 19.



Joonis 19.

Nii korrelatsioonitabeli kui ka korrelatsioonivälja abil saab jälgida ühe suuruse üldist muutumise tendentsi teise suuruse muutumise taustal. Toodangu omahinnal on tendents langeda toodangu mahu kasvades.

Statistilise sõltuvuse üksikasjalikum uurimine toimub korrastatud lähteandmetele tuginedes. Sõltuvuse tabeliline esitus jääb lähtepunktiks andmete matemaatilisel läbitöötamisel. Korrelatsioonivälja abil näeme mõnikord suuruste muutumise tendentsi paremini kui korrelatsioonitabelist.

Korrelatsiooniväli aitab seega tihti leida suunda, milles tuleb liikuda sõltuvuse täpsemal uurimisel ja seaduspärasuste väljaselgitamisel.

## 3. Tinglike jaotuste karakteristikute leidmine.

1. näide. Kasvu (x) ja kaalu (y) iseloomustavad andmed 50 4. klassi õpilase kohta on esitatud korrelatsioonitabelis.

| x \ y | 24 | 25 | 26 | 27 | 28 | 29 | 30 | $m_x$ |
|---|---|---|---|---|---|---|---|---|
| 125 | 1 | | | | | | | 1 |
| 126 | 1 | 2 | | | | | | 3 |
| 127 | | 2 | 4 | 1 | | | | 7 |
| 128 | | 1 | 3 | 5 | 1 | | | 10 |
| 129 | | | 2 | 4 | 5 | 1 | | 12 |
| 130 | | | | 2 | 5 | 2 | | 9 |
| 131 | | | | | 1 | 3 | 1 | 5 |
| 132 | | | | | | 1 | 1 | 2 |
| 133 | | | | | | | 1 | 1 |
| $n_y$ | 2 | 5 | 9 | 12 | 12 | 7 | 3 | 50 |

Ühe suuruse väärtustele vastavaid teise suuruse väärtuste jaotusi nimetasime tinglikeks jaotusteks. Tinglike jaotuste karakteristikuid nimetatakse tinglikeks karakteristikuteks. Nii räägitakse tinglikest keskmistest, tinglikest dispersioonidest ning standardhälvetest jne.

Kasvule $x_4 = 128$ cm vastav kaalu y tinglik jaotus omab variatsioonrea

| y | 25 | 26 | 27 | 28 |
|---|---|---|---|---|
| f | 1 | 3 | 5 | 1 |

mille keskmine

$\bar{y}_4 = \frac{1}{10}(1.25 + 3.26 + 5.27 + 1.28) = 26{,}6$ kg on väärtusele $x_4 = 128$ vastavaks tinglikuks keskmiseks.

Samale väärtusele vastav tinglik dispersioon on

$$D_4(y) = \frac{1}{10} \cdot \left[ (2{,}56 \cdot 1 + 0{,}36 + 0{,}16 \cdot 5 + 1{,}96 \cdot 1) \right] = 0{,}64$$

ja tinglik standardhälve

$$\sigma_4(y) = \sqrt{0{,}64} = 0{,}8 \text{ kg.}$$

Analoogiliselt saab leida ka teistele x väärtustele vastavad y tinglikud karakteristikud ja y väärtustele vastavad x tinglikud karakteristikud. Edasipidi puutume peamiselt kokku tinglike keskmistega. Leides tinglikud keskmised mõlema tunnuse jaoks, saame tinglike keskmiste jaotustabelid

| $x$ | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 |
|---|---|---|---|---|---|---|---|---|---|
| $\bar{y}_x$ | 24,0 | 24,7 | 25,9 | 26,6 | 27,4 | 28,0 | 29,0 | 29,5 | 30,0 |

ja

| $y$ | 24,0 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|
| $\bar{x}_y$ | 125,5 | 126,8 | 127,8 | 128,6 | 129,5 | 130,6 | 132,0 |

Näeme, et koos ühe tunnuse muutumisega muutuvad ka teise tunnuse tinglikud keskmised.

2. näide. Koostada tinglike keskmiste jaotustabel järgmise korrelatsioonitabeli andmetel.

| x \ y | 1 | 2 | 3 | 4 | 5 | $m_x$ |
|---|---|---|---|---|---|---|
| 1 | 1 | | | | 1 | 2 |
| 2 | 1 | 1 | | 1 | 1 | 4 |
| 3 | | 2 | 1 | 2 | | 5 |
| 4 | | | 1 | | | 1 |
| $n_y$ | 2 | 3 | 2 | 3 | 2 | 12 |

Lihtne on kontrollida, et tinglike keskmiste jaotustabelid on järgmised

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\bar{y}_x$ | 3 | 3 | 3 | 3 |

| $y$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\bar{x}_y$ | 1,5 | 2,67 | 3,5 | 2,67 | 1,5 |

Tunnuse y muutumine tingib ka $\bar{x}_y$ muutumise. Tunnuse x väärtustele vastavad y tinglikud jaotused on küll erinevad, kuid tinglikud keskmised $\bar{y}_x$ ei muutu.

Üldjuhul toimub korrelatsioonitabeli andmetel tinglike keskmiste leidmine järgmiselt

$$\bar{y}_i = \frac{y_1 f_{i1} + \dots + y_l f_{il}}{m_i} = \frac{1}{m_i} \sum_{j=1}^{l} y_j f_{ij} \quad (i = 1,2 \dots k)$$

$$\bar{x}_j = \frac{x_1 f_{1j} + \dots + x_k f_{kj}}{n_j} = \frac{1}{n_j} \sum_{i=1}^{k} x_i f_{ij} \quad (j = 1,2 \dots l)$$

Kogu korrelatsioonitabeli ulatuses arvutatud aritmeetilisi keskmisi nimetatakse üldkeskmisteks. Nende leidmine toimub aritmeetilise keskmise definitsiooni kohaselt.

$$\bar{x} = \frac{1}{N} (x_1 m_1 + \dots + x_k m_k) = \frac{1}{N} \sum_{i=1}^{k} x_i m_i \;;$$

$$\bar{y} = \frac{1}{N} (y_1 n_1 + \dots + y_l n_l) = \frac{1}{N} \sum_{j=1}^{l} y_j n_j \,.$$

Punkti koordinaatidega $(\bar{x}, \bar{y})$ nimetatakse korrelatsioonitsentriks.

Üldkeskmisi saab leida ka tinglike keskmiste kaudu

$$\bar{y} = (\bar{y}_1 m_1 + \bar{y}_2 m_2 + \dots + \bar{y}_k m_k) \cdot \frac{1}{N} = \frac{1}{N} \sum_{i=1}^{k} m_i \bar{y}_i \;;$$

$$\bar{x} = (\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots \bar{x}_2 n_2) \cdot \frac{1}{N} = \frac{1}{N} \sum_{j=1}^{l} n_j \bar{x}_j \,.$$

Saadud seoseid kasutame edaspidi.

3. näide. Leiame 1. näite andmetel õpilaste keskmise kaalu.

Kasutame keskmise arvutamist lihtsustavat valemit (vt. § 3 p. 5)

$$\bar{y} = \frac{\sum (y - b)\, n_y}{N} + b$$

ja valime b = 27

| $y$ | $n_y$ | $y - 27$ | $(y - 27)\, n_y$ |
|---|---|---|---|
| 24 | 2 | -3 | -6 |
| 25 | 5 | -2 | -10 |
| 26 | 9 | -1 | -9 |
| 27 | 12 | 0 | -25 |
| 28 | 12 | 1 | 12 |
| 29 | 7 | 2 | 14 |
| 30 | 3 | 3 | 9 |
| | | | 10 |

$$\bar{y} = \frac{10}{50} + 27 = 27{,}2 \text{ kg.}$$

## 4. Korrelatiivne sõltuvus.

Ühe suuruse sõltuvuse uurimine teise suuruse varieerumisest on matemaatiliselt tülikas. Otstarbekam on ülesanne taandada lihtsamate sõltuvuste uurimisele. Tinglikke jaotusi saab kirjeldada nende arvuliste karakteristikute abil. Iga tinglikku jaotust saame kirjeldada keskmise, dispersiooni, standardhälbe ja teiste arvuliste karakteristikute abil. Seega ongi statistilise sõltuvuse kirjeldamiseks sobiv määrata ühe tunnuse tinglike karakteristikute muutumine teise tunnuse muutumise taustal. Jaotuse põhiliseks karakteristikuks on aritmeetiline keskmine ja harilikult uuritaksegi ühe suuruse tinglike keskmiste sõltuvust teise suuruse muutumisest.

Tähistame suuruse x väärtustele vastavad suuruse y tinglikud keskmised $\bar{y}_x$ ja y väärtustele vastavad suuruse x tinglikud keskmised $\bar{x}_y$.

Suuruse y sõltuvust suurusest x nimetatakse korrelatiivseks, kui suuruse x igale väärtusele vastab suuruse y väärtuste jaotus ja tema tinglikud keskmised $\bar{y}_x$ ei jää konstantseteks. Analoogiliselt defineeritakse ka suuruse x korrelatiivne sõltuvus suurusest y.

Märkus 1. Korrelatiivse sõltuvuse asemel räägitakse tihti korrelatiivsest seosest uuritavate suuruste vahel.

Märkus 2. Korrelatiivse sõltuvuse definitsioonist järeldub, et ta on kitsam statistilise sõltuvuse mõistest. Üks suurus võib olla statistilise sõltuvuses teisest suurusest, kuid jaotustabelid võivad olla sellised, et tinglikud keskmised on kogu aeg samad(2. näide 3. punkt).

Märkus 3. Võib esineda ka olukordi, kus näiteks suurus y on korrelatiivses sõltuvuses suurusest x, kuid x ei ole korrelatiivses sõltuvuses suurusest y (2. näide 3. punkt).

Eelmise punkti 1. näites on õpilaste kaalud ja kasvud korrelatiivses sõltuvuses, kusjuures see sõltuvus on vastastikune. Järgmises näites on suurused x ja y vastastikuses statistilises sõltuvuses - tinglikud jaotused muutuvad; samuti on suurus x korrelatiivses sõltuvuses suurusest y, kuid suurus y ei ole korrelatiivses sõltuvuses suurusest x, sest $\bar{y}_x$ = const.

Korrelatiivse sõltuvuse definitsioonist järgneb, et tinglik keskmine $\bar{y}_x$ on argumendi x funktsioon, mida võib valemi kujul esitada

$$\bar{y}_x = \varphi(x)$$

Esitatud sõltuvust nimetatakse korrelatsioonivõrrandiks ehk y regressiooniks x järgi. Korrelatsioonivõrrandile vastavat graafikut nimetatakse regressioonikõveraks.

Suuruse y muutumisele avaldavad peale x muutumise mõju ka muud faktorid, mille mõju me ei saa kõrvaldada. See tingib y väärtuste hajumise ja y tinglike jaotuste tekkimise. Küllalt suure katsete arvu puhul on alust arvata, et kõrvalistest teguritest tingitud y hälbed kustutavad üksteist (tugineme suurte arvude seadusele). Järelikult korrelatsioonivõrrand $\bar{y}_x = \wp(x)$ kirjeldab keskmiselt y muutumist x mõjul.

Analoogiliselt räägitakse x regressioonist y järgi ning korrelatsioonivõrrandist $\bar{x}_y = f(y)$.

Kui suurused x ja y oleksid funktsionaalses sõltuvuses, siis oleksid seosed

$$\bar{y}_x = \wp(x) \quad \text{ja} \quad \bar{x}_y = f(y)$$

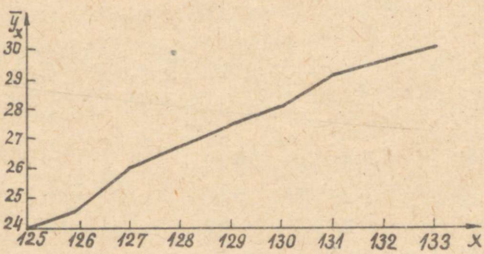
ühe ja sama sõltuvuse kaks esitusviisi. Võrrand $\bar{x}_y = f(x)$ on saadud võrrandist $\bar{y}_x = \wp(x)$ viimase lahendamisel x suhtes.
Regressioonikõverad langevad kokku.

Korrelatiivse sõltuvuse korral on üks kord argumendiks suurus x ja $\bar{y}_x$ tema funktsiooniks, teine kord on argumendiks y ja tema funktsiooniks $\bar{x}_y$. Kuna x ja y ei ole üksüheses vastavuses, siis me ei saa ühest korrelatsioonivõrrandist teist. Regressioonikõverad on samuti erinevad. Argumendi osasse valitud tunnust nimetatakse faktortunnuseks.

Kui mõlemad uuritavad tunnused alluvad ühiste faktorite mõjule, siis on otstarbekas uurida kahepoolset regressiooni. Nii on näiteks tarbitud tooraine maksumuse ja valmistoodangu maksumuse vahekorra selgitamisel. Faktortunnuseks tuleks ühel juhul võtta tooraine maksumus, teisel juhul toodangu maksumus. Kui aga üks tunnus on teise muutumise põhjuseks ja vastupidist mõju ei ole, siis uurime ainult ühepoolset regressiooni, valides faktortunnuseks põhjuse. Näiteks lehmade vanuse ja piimaanni vahelise sõltuvuse uurimisel tuleb faktortunnuseks valida vanus.

## 5. L i n e a a r n e   k o r r e l a t s i o o n .

Korrelatsioonivõrrandi koostamine toimub katseandmete põhjal. Tinglike keskmiste jaotustabeli abil saame koostada eksperimentaalse regressioonikõvera, milleks tuleb harilikult murdjoon. Näiteks 3. punktis käsitletud näites y regressioon x järgi esitub graafiliselt allpool toodud kujul




Joonis 20.

Katseandmete piiratud hulga tõttu sisaldub tinglikes keskmistes juhuslikkuse elemente, mis tingivad punktide kõrvalekaldumise teatud pidevast kõverast. Katsete arvu piiramatul kasvamisel läheneb murdjoon pidevale regressioonikõverale, mille võrrandiks oleks korrelatsioonivõrrand

$$\bar{y}_x = \varphi(x).$$

Meil tuleb korrelatsioonivõrrand koostada piiratud katseandmete abil. See tähendab funktsiooni avaldise konstrueerimist funktsiooni tabeli abil, milles funktsiooni väärtused pole päris täpsed. Teatavasti kasutatakse selleks harilikult vähimruutude meetodit.

Vähimruutude meetodi kohaselt tuleb eelnevalt valida sobiv funktsiooni tüüp katseandmete tasandamiseks ja seejärel määrata funktsiooni avaldises esinevad parameetrid nii, et valitud teoreetilise funktsiooni suhtes arvutatud katseliste funktsiooni väärtuste hälvete ruutude summa oleks minimaalne. Parameetrite leidmine taandub normaalvõrrandite süsteemi lahendamisele.

Praktikas tuleb kõige sagedamini kokku puutuda lineaarsete ja ruutfunktsioonidega, s.t. valitakse

$\bar{y}_x = ax + b$ – lineaarne korrelatsioon,

$\bar{y}_x = ax^2 + bx + c$ – paraboolne korrelatsioon.

Peatume lähemalt lineaarse korrelatsiooni juures. Lineaarse korrelatsiooni võrrandeid nimetame edaspidi lihtsalt regressioonivõrrandeiks ja nende graafikuid regressioonisirgeteks.

Regressioonivõrrandeiks oleksid

$$\bar{y}_x = ax + b$$

$$\bar{x}_y = cy + d.$$

Viime arutlused läbi ainult $\bar{y}_x = ax + b$ jaoks. Tulemused kanname üle ka teisele võrrandile, võttes x asemele y ja $\bar{y}_x$ asemele $\bar{x}_y$.

## 6. Regressioonivõrrandite koostamine.

Olgu antud argumendi u ja funktsiooni v katsel saadud väärtused $u_i$ ja $v_i$ ($i = 1, 2, \ldots, N$). Katseandmete tasandamisel lineaarfunktsiooniga

$$v = \alpha u + \beta$$

määratakse parameetrid $\alpha$ ja $\beta$ normaalvõrrandite süsteemist

$$\alpha \sum_{i=1}^{N} u_i^2 + \beta \sum_{i=1}^{N} u_i = \sum_{i=1}^{N} u_i v_i$$

$$\alpha \sum_{i=1}^{N} u_i + \beta N = \sum_{i=1}^{N} v_i .$$

Regressioonivõrrandi koostamisel on argumendi osas x ja funktsiooni osas tinglik keskmine $\bar{y}_x$. Võrrandisüsteemis esinevate summade leidmisel peame arvestama seda, et iga erinev väärtus esineb $m_i$ korda. Regressioonivõrrandi $\bar{y}_x = ax + b$ kordajad a ja b leiame süsteemist

$$\begin{cases} a \sum\limits_{i=1}^{k} m_i x_i^2 + b \sum\limits_{i=1}^{k} m_i x_i = \sum\limits_{i=1}^{k} m_i x_i \bar{y}_i ; \\ a \sum\limits_{i=1}^{k} m_i x_i + b . N = \sum\limits_{i=1}^{k} m_i \bar{y}_i . \end{cases} \qquad (I)$$

Arvutades korrelatsioonitabeli andmetel tinglikud keskmised $\bar{y}_i$, paigutame nad süsteemi (I) ning viimase lahendamine annab otsitavate a ja b väärtused. Asendades saadud a ja b väärtused regressioonivõrrandisse, olemegi koostanud vajaliku võrrandi.

Püüame normaalvõrrandite süsteemi teisendada niisugusele kujule, et saaksime kõik vajalikud arvutused teostada korrelatsioonitabeliga määratud andmetest lähtudes. Siis langeks ära vajadus tinglike keskmiste leidmiseks.

Kolmanda punkti lõpus näitasime, kuidas saab üldkeskmist $\bar{y}$ avaldada tinglike keskmiste ja üldkeskmist $\bar{x}$ korrelatsioonitabeli andmete $x_i$ kaudu. Kui võrrelda seal saadud seoseid süsteemi (I) esimeses võrrandis tundmatu b kordajaga või teises võrrandis a kordajaga ning teises võrrandis vabaliikmega, näeme, et

$$\sum_{i=1}^{k} m_i x_i = N \bar{x} ; \qquad \sum_{i=1}^{k} m_i \bar{y}_i = N \bar{y} . \qquad (II)$$

Dispersiooni arvutamiseks kasutasime ühe eeskirjana

$$\sigma^2 = \overline{x^2} - \bar{x}^2 .$$

Tähistame tunnuste x ja y ülddispersioonid vastavalt

$s_x^2$ ja $s_y^2$ . Siis

$$s_x^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{N} \sum_{i=1}^{k} m_i x_i^2 - \bar{x}^2 ,$$

kust

$$\sum_{i=1}^{k} m_i x_i^2 = (s_x^2 + \bar{x}^2) \cdot N. \qquad \text{(III)}$$

Teisendame ka esimese võrrandi vabaliiget

$$\sum_{i=1}^{k} m_i x_i \bar{y}_i .$$

Tinglike keskmiste $\bar{y}_i$ leidmine toimub järgneva eeskirja kohaselt (vt. 3. punkt).

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{l} y_j \cdot f_{ij} , \qquad (i=1,2,\dots,k),$$

kust

$$m_i \bar{y}_i = \sum_{j=1}^{l} y_j f_{ij} .$$

Paigutame vabaliikmesse $m_i \bar{y}_i$ asemele saadud summa

$$\sum_{i=1}^{k} m_i \bar{y}_i x_i = \sum_{i=1}^{k} x_i \left( \sum_{j=1}^{l} y_j f_{ij} \right) .$$

Siin on meil tegemist kõigi $x_i$ ja $y_j$ väärtuste ning neile vastavate sageduste $f_{ij}$ korrutiste summaga, mida tähistame järgmiselt

$$\sum_{i,j=1}^{k,l} x_i y_i f_{ij} = \sum_{i=1}^{k} m_i \bar{y}_i x_i \qquad \text{(IV)}$$

Kui asendada süsteemis (I) vastavad liikmed seoste (II), (III), (IV) järgi, saame süsteemi kujul

$$\begin{cases} a(s_x^2 + \bar{x}^2) \cdot N + b\bar{x}N = \sum\limits_{i,j=1}^{k,l} x_i y_i f_{ij} ; \\ a\bar{x}N + bN = N\bar{y} \end{cases}$$

Avaldame teisest võrrandist b ning asendame esimesse

$$N \cdot a\,(s_x^2 + \bar{x}^2) + N(\bar{y} - a\bar{x}) \cdot \bar{x} = \sum_{i,j=1}^{k,l} x_i y_i f_{ij} ,$$

kust leiame kordaja a

$$a = \frac{\frac{1}{N} \sum\limits_{i,j=1}^{k,l} x_i y_j f_{ij} - \bar{x}\bar{y}}{s_x^2} .$$

Ka regressioonivõrrandisse paigutame b avaldise a kaudu. Regressioonivõrrand omandab kuju

$$\bar{y}_x = ax + b = ax + \bar{y} - a\bar{x}$$

ehk

$$\bar{y}_x - \bar{y} = a(x - \bar{x}) .$$

Seega langes ära vajadus eraldi arvutada b väärtus. Regressioonisirge tõusu a nimetatakse regressioonikordajaks ja tähistatakse harilikult $\rho_{yx}$ (y regressioon x järgi). Kui räägitakse x regressioonist y järgi, siis tähistatakse $\rho_{xy}$.

Korrelatsioonitsentri koordinaadid $(\bar{x}, \bar{y})$ rahuldavad regressioonisirge võrrandit, mis näitab, et regressioonisirge läbib korrelatsioonitsentrit.

Kokkuvõte. Regressioonivõrrandite koostamiseks tulevad leida korrelatsioonitsentri kordinaadid $(\bar{x}, \bar{y})$ ja regressioonikoefitsiendid valemitest

$$\rho_{yx} = \frac{C_{xy}}{s_x^2}\ ; \qquad \rho_{xy} = \frac{C_{xy}}{s_y^2},$$

kus

$$C_{xy} = \frac{1}{N} \sum_{i,j=1}^{k,l} x_i y_j f_{ij} - \bar{x}\,\bar{y}$$ kannab kovariatsiooni nime ning seejärel esitada regressioonivõrrandid kujul

$$\bar{y}_x - \bar{y} = \rho_{yx}(x - \bar{x})\ ;$$

$$\bar{x}_y - \bar{x} = \rho_{xy}(y - \bar{y})\ .$$

Arvutuste lihtsustamiseks minnakse sageli lähtemuutujatelt x ja y üle uutele muutujatele u ning v järgmiste seoste abil

$$u = \frac{x - C_1}{h_1}\ ; \qquad v = \frac{y - C_2}{h_2}\ .$$

Konstantideks $C_1$ ja $C_2$ valitakse harilikult kas suurima sagedusega x ja y väärtus või rea algusest ja lõpust võrdsel kaugusel asuv väärtus. Suurused $h_1$ ja $h_2$ valitakse nii, et u ja v oleksid absoluutväärtuselt võimalikult väikesed täisarvud (intervallitud ridade puhul sageli intervalli pikkus).

Regressioonikoefitsiendid $\rho_{yx}$ ja $\rho_{xy}$ leiame siis valemitest

$$\rho_{yx} = \frac{h_2}{h_1}\,\frac{C_{uv}}{s_u^2}\ ; \rho_{xy} = \frac{h_1}{h_2}\,\frac{C_{uv}}{s_v^2},$$

kus

$$C_{uv} = \frac{1}{N} \sum_{i, j=1}^{k, l} u_i v_j f_{ij} - \bar{u}\,\bar{v}$$

ja $s_u^2$ ning $s_v^2$ on vastavalt suuruste u ja v dispersioonid.

## 7. Näide.

Selgitame konkreetse ülesande varal, kuidas regressioonivõrrandite koostamine võiks toimuda.

Tšehhoslovakkia 120 põllumajandusartellis uuriti 1 ha kohta tuleva puhastulu ja tööpäeva eest makstava tasu vahekorda (rahaline arvestus on kroonides). Andmed on esitatud korrelatsioonitabelina (vt. Ф. Эгермайер, В. Грузин, В. Влах, Основы статистики, Москва, 1961).

| Tulu \ Tasu | 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 |
|---|---|---|---|---|---|---|---|---|---|---|
| 00- 400 | 1 | 2 | 1 | | | | | | | |
| 400- 800 | | 1 | 12 | 8 | | 2 | | | | |
| 800-1200 | | | 7 | 5 | 7 | 3 | | | | |
| 1200-1600 | | | | 4 | 6 | 9 | 3 | | | |
| 1600-2000 | | | | | 2 | 6 | 5 | 1 | | |
| 2000-2400 | | | | | 2 | 4 | 8 | 3 | | |
| 2400-2800 | | | | | | 1 | 5 | 1 | 2 | |
| 2800-3200 | | | | | | 2 | | 2 | | |
| 3200-3600 | | | | | | | | | 1 | 2 |
| 3600-4000 | | | | | | | | | 1 | 1 |

Arvutuste lihtsustamiseks asendame kõigepealt intervallid nende keskkohtadele vastavate väärtustega x (tulu) ja y (tasu) ning seejärel läheme üle uutele muutujatele u ning v, mis on seotud tunnustega x ja y järgmiselt

$$u = \frac{x - c_1}{h_1} ; \quad v = \frac{y - c_2}{h_2} .$$

Käesolevas näites on valitud

$c_1 = 1800$ ; $c_2 = 16,5$ ; $h_1 = 400$ ; $h_2 = 3$ .

Siis

$$u = \frac{x - 1800}{400} \qquad v = \frac{y - 16,5}{3}$$

Sooritades vajalikud arvutused, esitame korrelatsioonitabeli kujul (vt. lk. 164).

Alumises reas ja viimases veerus on vastavate ridade (veergude) sageduste summad.

Regressioonikoefitsientide avaldistes esinevad keskmised $\bar{u}$ ja $\bar{v}$ ja dispersioonid $s_u^2$ , $s_v^2$ arvutame järgmiste eeskirjade kohaselt:

| $u$ | $m_x$ | $um_x$ | $u^2$ | $u^2m_x$ |
|---|---|---|---|---|
| - 4 | 4 | - 16 | 16 | 64 |
| - 3 | 23 | - 69 | 9 | 207 |
| - 2 | 22 | - 44 | 4 | 88 |
| - 1 | 22 | - 22 | 1 | 22 |
| 0 | 14 | -151 | | |
| 1 | 17 | 17 | 1 | 17 |
| 2 | 9 | 18 | 4 | 36 |
| 3 | 4 | 12 | 9 | 36 |
| 4 | 3 | 12 | 16 | 48 |
| 5 | 2 | 10 | 25 | 50 |
| | | 69 | | |
| | | - 82 | | 568 |

| $v$ | $n_y$ | $vn_y$ | $v^2$ | $v^2n_y$ |
|---|---|---|---|---|
| - 5 | 1 | - 5 | 25 | 25 |
| - 4 | 3 | -12 | 16 | 48 |
| - 3 | 20 | -60 | 9 | 180 |
| - 2 | 17 | -34 | 4 | 68 |
| - 1 | 17 | -17 | 1 | 17 |
| 0 | 27 | -128 | | |
| 1 | 21 | 21 | 1 | 21 |
| 2 | 7 | 14 | 4 | 28 |
| 3 | 4 | 12 | 9 | 36 |
| 4 | 3 | 12 | 16 | 48 |
| | | + 59 | | |
| | | - 69 | | 471 |

Tabeli andmetest järeldub, et

$$\bar{u} = \frac{-82}{120} = -0,683 \ ; \quad s_u^2 = \frac{568}{120} - 0,683^2 = 4,267 \ ;$$

$$\bar{v} = \frac{-69}{120} = -0,575 \ ; \quad s_v^2 = \frac{471}{120} - 0,575^2 = 3,594 \ .$$

| x \ y | u \ v | 1,5 | 4,5 | 7,5 | 10,5 | 13,5 | 16,5 | 19,5 | 22,5 | 25,5 | 28,5 | $m_x$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | −5 | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 | 4 | |
| 200 | −4 | 1 | 2 | 1 | | | | | | | | 4 |
| 600 | −3 | | 1 | 12 | 8 | | 2 | | | | | 23 |
| 1000 | −2 | | | 7 | 5 | 7 | 3 | | | | | 22 |
| 1400 | −1 | | | | 4 | 6 | 9 | 3 | | | | 22 |
| 1800 | 0 | | | | | 2 | 6 | 5 | 1 | | | 14 |
| 2200 | 1 | | | | | 2 | 4 | 8 | 3 | | | 17 |
| 2600 | 2 | | | | | | 1 | 5 | 1 | 2 | | 9 |
| 3000 | 3 | | | | | | 2 | | 2 | | | 4 |
| 3400 | 4 | | | | | | | | | 1 | 2 | 3 |
| 3800 | 5 | | | | | | | | | 1 | 1 | 2 |
| $n_y$ | | 1 | 3 | 20 | 17 | 17 | 27 | 21 | 7 | 4 | 3 | 120 |

Summa $\sum u_i v_j f_{ij}$ arvutamiseks on otstarbekas kasutada vahetult teisendatud korrelatsioonitabelit. Ridade ja veergude lõikepunktides olevate sageduste kõrvale kirjutame vastavate u ja v väärtuste korrutised. Nullväärtustele vastavad rida ja veerg jagavad tabeli neljaks, kusjuures I ja III välja puhul on kõik korrutised positiivsed, II ja IV välja puhul negatiivsed. Korrutades $u_i v_j$ vastavate sagedustega, on väljade kaupa kerge summeerida saadavaid korrutisi. Nullreale ja veerule vastavad korrutised on nullid

| u \ v | -n5 | - 4 | - 3 | - 2 | - 1 | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| - 4 | 1 $\boxed{20}$ | 2 $\boxed{16}$ | 1 $\boxed{12}$ | | | | | | | |
| - 3 | | 1 $\boxed{12}$ | 12 $\boxed{9}$ | 8 $\boxed{6}$ | | | | | | II |
| - 2 | I | | 7 $\boxed{6}$ | 5 $\boxed{4}$ | 7 $\boxed{2}$ | | | | | |
| - 1 | | | | 4 $\boxed{2}$ | 6 $\boxed{1}$ | | 3 $\boxed{-1}$ | | | |
| 0 | | | | | | | | | | |
| 1 | | | | | 2 $\boxed{-1}$ | | 8 $\boxed{1}$ | 3 $\boxed{2}$ | | III |
| 2 | | | | | | | 5 $\boxed{2}$ | 1 $\boxed{4}$ | 2 $\boxed{6}$ | |
| 3 | | | | | | | | 2 $\boxed{6}$ | | |
| 4 | IV | | | | | | | | 1 $\boxed{12}$ | 2 $\boxed{16}$ |
| 5 | | | | | | | | | 1 $\boxed{15}$ | 1 $\boxed{20}$ |

I väljale vastav summa on 322; III - 131 ; II - - 3; IV - - 2. Kogusumma on 322 + 131 - 3 - 2 = 448.

Järgnevalt leiame

$$C_{uv} = \frac{448}{120} - 0{,}683 \,.\, 0{,}575 = 3{,}340 \,.$$

Regressioonikoefitsiendid

$$\rho_{yx} = \frac{3{,}340.3}{4{,}267.400} = 0{,}0059 \;; \quad \rho_{xy} = \frac{3{,}340.400}{3{,}594.3} = 124$$

Regressioonivõrrandi..e koostamiseks on tarvis teada $\bar{x}$ ja $\bar{y}$. Viimased leiame $\bar{u}$ ja $\bar{v}$ abil. Selleks avaldame x ja y uute muutujate u ning v kaudu ja kuna keskmiste vahel kehtib samasugune seos, siis

$$\bar{x} = 400\,\bar{u} + 1800 = 400.(-0,683) + 1800 \approx 1527,$$

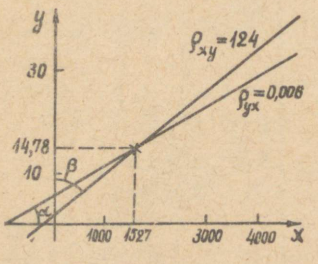
$$\bar{y} = 3\,\bar{v} + 16,5 = 3.(-0,575) + 16,5 \approx 14,78 .$$

Regressioonivõrrandid saame kujul

$$\bar{y}_x - 14,78 = 0,0059\,(x - 1527) ;$$

$$\bar{x}_y - 1527 = 124\,(y - 14,78) .$$

Regressioonisirged on näha joonisel 21. Tõus $\rho_{yx} = \tan\alpha$ ; $\rho_{xy} = \tan\beta$ , sest esimesel juhul on argumendiks x ja abstsisseteljeks x-telg, teisel juhul on argumendiks y ja abstsisseteljeks y-telg. Kui valida ühikud mõlemal teljel võrdsed, siis oleks regressioonisirgete kalle x-telje suhtes väga väike: $\tan\alpha = 0,006$ .




Joonis 21.

8. R e g r e s s i o o n i v õ r r a n d i t e k o o s t a m i n e l i h t s a k o r r e - l a t s i o o n i t a b e l i a n d m e t e l.

Kui lähteandmete hulk on piiratud, siis pole vajadust nende korrastamiseks mõlema tunnuse järgi ja võime piirduda lihtsa korrelatsioonitabeliga

| x | $x_1$ | $x_2$ ... | $x_N$ |
|---|---|---|---|
| y | $y_1$ | $y_2$ ... | $y_N$ |

Sellisel juhul leiame üldkeskmised valemitest

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \ ; \quad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \ ,$$

dispersiooni

$$s_x^2 \ \frac{1}{N} \left[ \sum_{i=1}^{N} x_i^2 - \left( \sum_{i=1}^{N} x_i \right)^2 . \ \frac{1}{N} \right] ;$$

$$s_y^2 = \frac{1}{N} \left[ \sum_{i=1}^{N} y_i^2 - \left( \sum_{i=1}^{N} y_i \right)^2 \frac{1}{N} \right]$$

ja

$$C_{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i - \bar{x} \cdot \bar{y} \ .$$

Kui regressioonikordajate avaldistes murru lugejat ja nimetajat korrutada katsete arvuga N, saame arvutuseeskirjad

$$\rho_{yx} = \frac{N \sum xy - \sum x \cdot \sum y}{N \sum x^2 - \left( \sum x \right)^2}$$

$$\rho_{xy} = \frac{N \sum xy - \sum x \cdot \sum y}{N \sum y^2 - \left( \sum y \right)^2}$$

Näide. Leida lineaarne korrelatsioon Nõukogude Liidu rahvusliku tulu (y) ja tööviljakuse kasvu (x) vahel tööstuses.

Andmed on protsentides 1950 aastaga võrreldes:

| Aastad | 1950 | 1951 | 1952 | 1953 | 1954 | 1955 |
|---|---|---|---|---|---|---|
| x | 100 | 110 | 117 | 125 | 133 | 144 |
| y | 100 | 112 | 125 | 136 | 153 | 168 |

Arvutuste lihtsustamiseks läheme üle uutele muutujatele

$$u = x - 125 \; ; \quad v = y - 136 \; .$$

Vajalike summade arvutamisel kasutame arvutusskeemi

| x | y | u | v | uv | $u^2$ | $v^2$ |
|---|---|---|---|---|---|---|
| 100 | 100 | - 25 | - 36 | 900 | 625 | 1296 |
| 110 | 112 | - 15 | - 24 | 360 | 225 | 576 |
| 117 | 125 | - 8 | - 11 | 88 | 64 | 121 |
| 125 | 136 | 0 | 0 | 0 | 0 | 0 |
| 133 | 153 | 8 | 17 | 136 | 64 | 289 |
| 144 | 168 | 19 | 32 | 608 | 361 | 1024 |
| | | - 21 | - 22 | 2092 | 1339 | 3306 |

Edasi

$$\rho_{yx} = \frac{6 \, . \, 2092 - 21.22}{6.. \, 1339 - 21^2} = \frac{12090}{7593} = 1,5 \; ;$$

$$\rho_{xy} = \frac{12090}{6.3306 - 22^2} = 0,66 \; .$$

Regressioonivõrrandid võime esitada muutujates u ja v või muutujates x ja y . Viimasel juhul leiame

$$\bar{x} = \bar{u} + 125 = - \frac{21}{6} + 125 = 121,5 \; ;$$

$$\bar{y} = \bar{v} + 136 = - \frac{22}{6} + 136 = 132,33 \; .$$

Regressioonivõrrandid on

$$\bar{y}_x - 132,3 = 1,5(x - 121,5) \; ; \; \bar{x}_y - 121,5 = 0,66(y - 132,3).$$

## 9. Korrelatsioonikoefitsient.

Vaatleme tunnuse y sõltuvust tunnusest x . Et paremini aru saada, arutleme eelmise punkti näite andmetel: y – rahvuslik tulu; x – tööviljakuse kasv (mõlemad protsentides 1950. aastaga võrreldes).

Olgu eesmärgiks määrata suuruse y väärtus mingi suuruse x väärtuse puhul. Selleks võime kasutada mitut moodust.

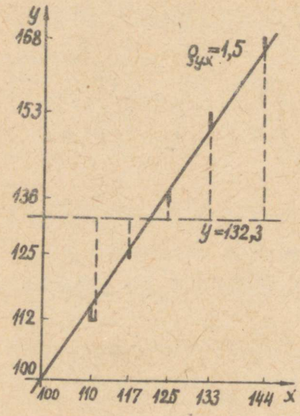
1. Valime suuruse x valitud väärtusele vastavaks y väärtuseks tema keskmise $\bar{y}$ . Keskmise suhtes leitud hälvete ruutude keskmine iseloomustab viga, mille teeme, võttes otsitavaks y väärtuseks tema keskmise $\bar{y}$ . Kuna keskmise suhtes leitud hälvete ruutude keskmine on minimaalne, siis võime lugeda sel juhul keskmise $\bar{y}$ kõige sobivamaks y lähendiks.

See tähendab, et loeme kogu ajavahemikus 1950–1955 rahvusliku tulu kasvuks 132,33%. Graafilisel kujutamisel tähendab selline valik seda, et tööviljakuse kasvule vastavaks rahvusliku tulu muutumist näitavaks jooneks on horisontaalne sirge y = 132,33 (joonisel 22 punktiiriga märgitud sirge) ning punktide hälbed on leitud selle sirge suhtes.

2. Kasutame suuruse y väärtuste määramiseks x järgi regressioonivõrrandit. Näiteks väärtusele x = 110 vastav y väärtus on

$$\bar{y}_{110} = 132,3 + 1,6(110 - 121,5) = 113,9 \; .$$

Tegelik väärtus oli 112. Analoogiliselt saame leida igale x-le vastava y väärtuse. Sellise määramise juures tehtud viga hindame suuruse y tegelike väärtuste hälvete ruutude keskmise abil, kus hälbed on leitud regressioonivõrrandist arvutatud $\bar{y}_x$ suhtes. Graafiliselt vastab sellele punktide hälvete leidmine regressioonisirge suhtes (vt. joonis 22).




Joonis 22.

Keskmise $\bar{y}$ suhtes leitud hälvete ruutude keskmine on dispersioon $s_y^2$. Regressioonisirge suhtes leitud y väärtuste hälvete ruutude keskmise tähistame sümboliga $s_{yx}^2$, kus esimene indeks tähistab sõltuvat muutujat.

Suuruse $s_{yx}^2$ leidmisel arvestame, et $y = v + 136$ ja dispersiooni omaduste põhjal

$$D(y) = D(v + 136) = D(v)$$

ehk

$$s_y^2 = s_v^2 = \overline{v^2} - \bar{v}^2 = \frac{3306}{6} - \left(\frac{22}{6}\right)^2 = 537{,}6 .$$

Keskmise ruutvea $s_{yx}$ leidmiseks kasutame arvutusskeemi, võttes arvesse, et

$$\bar{y}_x = 132{,}3 + 1{,}5(x - 121{,}5) .$$

| x | y | (x-121,5)1,5 | $\bar{y}_x$ | $y-\bar{y}_x$ | $(y-\bar{y}_x)^2$ |
|---|---|---|---|---|---|
| 100 | 100 | - 32,25 | 100,1 | -0,1 | 0,01 |
| 110 | 112 | - 17,25 | 115,0 | -3 | 9,00 |
| 117 | 125 | - 6,75 | 125,6 | -0,6 | 0,36 |
| 125 | 136 | + 5,25 | 137,6 | -1,6 | 2,56 |
| 133 | 153 | 17,25 | 149,6 | +3,4 | 11,56 |
| 144 | 168 | 33,75 | 166,0 | 2,0 | 4,00 |
| | | | | | 27,49 |

$$s^2_{yx} = \frac{27{,}49}{6} = 4{,}58$$

Suuruste $s_y^2$ ja $s_{yx}^2$ väärtustest leiame ruutjuure

$$s_y = \sqrt{537,6} = 23,18 \; ; \; s_{yx} = \sqrt{4,58} = 2,14 \; .$$

Suhe $s_{yx} : s_y$ näitab, mitu korda väiksema vea oleme teinud y leidmisel regressioonivõrrandist lähtudes võrreldes hindamisega keskmise abil

$$\frac{s_{yx}}{s_y} = \frac{2,14}{23,18} = 0,092 = (9,2\%) \; .$$

Tulemus näitab, et regressioonivõrrandist leitud y väärtuste hinnang on tublisti täpsem. Regressioonivõrrandi kasutamisel tehtud viga moodustab 9,2% veast, mis tekkis y määramisel keskmise abil.

Selgitame $s_{yx}$ ja $s_y$ vahekorda. Võivad esineda järgmised juhud:

1. Suurused y ja x on funktsionaalses sõltuvuses. Siis $s_{yx} = 0$, sest hälbed puuduvad ja $s_{yx} : s_y = 0$ . Sel juhul saame ühe suuruse täpselt määrata teise kaudu.

2. Suuruste y ja x vahel puudub täiesti lineaarne korrelatiivne seos. Siis suuruse y tinglikud keskmised ei reageeri üldse suuruse x muutumisele, s.t. $\bar{y}_x = \bar{y}$ . Kuid sellisel juhul on ka $s_{yx} = s_y$ ja $s_{yx} : s_y = 1$ .

3. Suurused on korrelatiivses sõltuvuses. Siis $s_{yx} : s_y < 1$ . Kaks esimest juhtu on kolmanda piirjuhud. Mida väiksem on suhe $s_{yx} : s_y$ , seda tugevamalt (tihedamalt) on seotud suurused y ja x. Suurusi loeme seda tihedamalt seotuks, mida rohkem ühe suuruse muutumine mõjutab teise suuruse muutumist. Selles mõttes ongi funktsionaalne sõltuvus kõige tihedam.

Otstarbekas on suurustevahelise seose tiheduse mõõduks võtta suhte $s_{yx} : s_y$ asemel selline suurus, mis suureneks seose tiheduse kasvades. Selliseks suuruseks valitakse nn. korrelatsioonikoefitsient.

Korrelatsioonikoefitsiendiks nimetatakse suurust, mis on defineeritud järgmiselt

$$r = \pm \sqrt{1 - \frac{s_{yx}^2}{s_y^2}} .$$

"+"- märk valitakse siis, kui mõlemate suuruste muutumise tendents on samapidine (positiivne korrelatsioon) ja "-" märk siis, kui muutumise tendents on vastupidine (negatiivne korrelatsioon).

Mida väiksem on $s_{yx} : s_y$, seda väiksem on tema ruut ja seda suurem on ruutjuure märgi all oleva avaldise väärtus. Siis on suurem ka korrelatsioonikoefitsiendi absoluutväärtus r. Kui suhe $s_{yx} : s_y \to 1$, siis ruutjuure alune läheneb nullile ja $|r| \to 0$. Seega tõepoolest korrelatsioonikoefitsient r iseloomustab suurustevahelise seose tugevust.

## 10. Korrelatsioonikoefitsiendi arvutamine.

Korrelatsioonikoefitsienti r võib arvutada tema definitsioonvalemi abil

$$r = \pm \sqrt{1 - \frac{s_{yx}^2}{s_y^2}} .$$

1. näide. Punktis 8 toodud näite andmetel saame punktis 9 tehtud täiendavaid arvutusi kasutades

$$r = + \sqrt{1 - \left(\frac{2,14}{23,18}\right)^2} = \sqrt{0,992} = 0,996 .$$

Otstarbekam on teisendada valem r leidmiseks niisugusele kujule, et saaksime kasutada regressioonivõrrandite koostamisel tehtud arvutusi. Pärast teisendusi saame

$$r = \frac{C_{xy}}{s_x s_y},$$

kus kovariatsioon

$$C_{xy} = \frac{1}{N} \sum_{i,j=1}^{k,l} x_i y_j f_{ij} - \bar{x}\,\bar{y} .$$

Kovariatsiooni märk määrab ka r märgi, sest $s_x > 0$, $s_y > 0$.

Kui arvestada regressioonikoefitsientide avaldisi

$$\rho_{yx} = \frac{C_{xy}}{s_x^2} ; \quad \rho_{xy} = \frac{C_{xy}}{s_y^2} ,$$

siis näeme, et

$$r = \pm \sqrt{\rho_{yx} \cdot \rho_{xy}}$$

– korrelatsioonikoefitsient on võrdne regressioonikoefitsientide geomeetrilise keskmisega.

Arvutades r regressioonikoefitsientide kaudu, valime r märgiks kovariatsiooni $C_{xy}$ märgi.

2. näide. Leiame korrelatsioonikoefitsiendi 8. punkti näite andmetel regressioonikoefitsientide abil.

$$r = + \sqrt{1,6 \cdot 0,66} = \sqrt{0,99} = 0,995.$$

Tulemus ühtib 1. näite tulemusega (erinevus viimases kohas on tingitud ümardamisest arvutuste käigus).

3. näide. Leiame korrelatsioonikoefitsiendi 7. punkti näite andmetel.

Uutelt muutujatelt u ja v ei tarvitse enam tagasi minna esialgsetele muutujatele x ja y , sest

$$r = \pm\sqrt{\rho_{yx} \cdot \rho_{xy}} = \pm \sqrt{\frac{h_2\, C_{uv}}{h_1\, s_u^2} \cdot \frac{h_1\, C_{uv}}{h_2\, s_v^2}} = \frac{C_{uv}}{s_u \cdot s_v} .$$

Eespool saime

$$c_{uv} = 3{,}340 \; ; \; s_u^2 = 4{,}267 \; ; \; s_v^2 = 3{,}594 \,.$$

Seega

$$r = \frac{3{,}340}{4{,}267 \,.\, 3{,}594} = \frac{3{,}340}{2{,}066 \,.\, 1{,}896} = 0{,}85 \,.$$

## 11. Korrelatsioonikoefitsiendi sisust.

Korrelatsioonikoefitsient r on defineeritud valemiga

$$r = \pm \sqrt{1 - \frac{s_{yx}^2}{s_y^2}}$$

ja teda saab leida regressioonikoefitsientide kaudu valemist

$$r = \pm \sqrt{\rho_{yx} \cdot \rho_{xy}} \,.$$

Korrelatsioonikoefitsiendi omadusi selgitame nendele valemitele tuginedes.

1. Korrelatsioonikoefitsiendi absoluutväärtus ei ületa ühte, s.t.

$$|r| \leq 1 \; ;$$

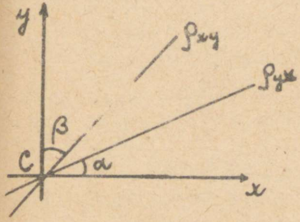
$$-1 \leq r \leq 1 \,.$$

Üheksandas punktis selgitasime, et alati on täidetud tingimus

$$0 \leq \frac{s_{yx}}{s_y} \leq 1$$

ja ruutjuure aluse avaldise väärtus ei ületa ühte. Kuid ruutjuur arvudest, mis ei ületa ühte, ei ületa ka ühte.

Seda võime illustreerida ka geomeetriliselt. Joonisel 23 on esitatud regressioonisirged, kusjuures koordinaatide alguspunkt on valitud korrelatsioonitsentrisse. Suuruse y määramisel x järgi loeme argumendi väärtusi x-teljelt ja $\rho_{yx}$ määrab vastava regressioonisirge tõusu x - telje suhtes, s.t. $\rho_{yx} = \tan\alpha$. Suuruse x määramisel y järgi loeme argumendi väärtusi y - teljelt ning vastavalt



Joonis 23.

$$\rho_{xy} = \tan\beta .$$

Joonistelt näeme, et $\beta < 90° - \alpha$ ja siis ka $\tan\beta < \tan(90° - \alpha) = \cot\alpha$. Kuna $\tan\alpha \cdot \cot\alpha = 1$, siis $\tan\alpha \cdot \tan\beta < \tan\alpha \cdot \cot\alpha = 1$.

2. Funktsionaalse sõltuvuse korral on $r = \pm 1$ ehk $|r| = 1$. Funktsionaalse sõltuvuse korral on suurused x ja y üksüheses vastavuses, olles seotud ühe võrrandiga $y = ax + b$, millele vastab ainult üks sirge (joonis 24 a). Siis $\alpha + \beta = 90°$ ja $\tan\beta = \tan(90° - \alpha) = \cot\alpha$ ning

$$\tan\alpha \cdot \tan\beta = \tan\alpha \cdot \cot\alpha = 1$$

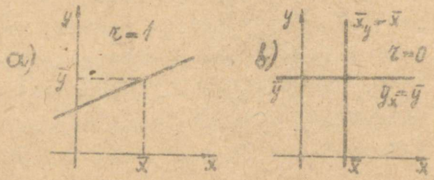
ehk

$$\rho_{yx} \cdot \rho_{xy} = r^2 = 1 .$$

3. Kui korrelatsioonikoefitsient $r = 0$, siis puudub suuruste vahel lineaarne korrelatsioon. Võib esineda kõverjooneline korrelatsioon.

Nüüd on $1 - \dfrac{s_{yx}^2}{s^2} = 0$ ehk $s_{yx} = s_y$. Punktis 9 veendusime, et sellisel juhul lineaarne korrelatsioon puudub - ühe suuruse tinglikud keskmised ei muutu üldse teise suuruse muutumise taustal.

Tingimusele $r = 0$ vastab $\rho_{yx} \cdot \rho_{xy} = 0$ ehk regressioonisirged on teineteisega risti (joonis 24 b).




Joonis 24.

Korrelatsioonikoefitsiendi r absoluutväärtus iseloomustab uuritavate suuruste seose tugevust. Mida lähemal on absoluutväärtus $|r|$ ühele, seda tugevam on suurustevaheline seos ning seda väiksem on regressioonisirgetevaheline nurk.

4. Selgitame suuruste $s_{yx}^2$ ja $s_y^2$ tähendust. Dispersioon $s_y^2$ iseloomustab tegelike y väärtuste varieerumist, $s_{yx}^2$ – tegelike y väärtuste varieerumist regressioonivõrrandist leitud väärtuste $\bar{y}_x$ suhtes, kusjuures

$$s_y^2 = \frac{1}{N}\sum_j (y_j - \bar{y})^2 n_j = D(y) \; ; \; s_{yx}^2 = \frac{1}{N}\sum_j (\bar{y}_{xj} - y)^2 n_j = D_{yx} \; .$$

Leiame veel regressioonivõrrandist leitud väärtuste $\bar{y}_{xj}$ dispersiooni

$$s_{yc}^2 = \frac{1}{N}\sum (\bar{y}_{xj} - \bar{y})^2 n_j = D(\bar{y}_x) \; .$$

Näitame, et

$$s_y^2 = s_{yx}^2 + s_{yc}^2$$

ehk, teistes tähistes,

$$D(y) = D_{yx} + D(\bar{y}_x) \; .$$

Selleks lähtume samasusest $y = \bar{y}_x + (y - \bar{y}_x)$ ning leiame y dispersiooni kui summa dispersiooni

$$D(y) = D\left[\bar{y}_x + (y - \bar{y}_x)\right] = D(\bar{y}_x) + D(y - \bar{y}_x) .$$

(vt. summa dispersioon!)

Esimene liidetav on juba tuttav. Teine liidetav $D(y - \bar{y}_x)$ aga ongi tegelike y väärtuste dispersioon võrrandist leitud $\bar{y}_x$ suhtes, s.t. $D_{yx} = s^2_{yx}$ .

Kui suurused y ja x oleksid lineaarses funktsionaalses sõltuvuses, siis $y = \bar{y}_x$ ja $s^2_{yx} = 0$ ning $s^2_y = s^2_{yc}$ , s.t. et y muutumine on täielikult määratud x muutumisega. Suurus $s^2_{yc}$ iseloomustabki suuruse y varieerumise seda osa, mis on tingitud sõltumatu suuruse x muutumisest. Sellepärast nimetatakse $s^2_{yc}$ mõnikord ka suuruse y "seletatud" variatsiooni mõõduks. Arvestades dispersioonidevahelist seost saame, et

$$s^2_y - s^2_{yx} = s^2_{yc}$$

ja

$$r^2 = 1 - \frac{s^2_{yx}}{s^2_y} = \frac{s^2_y - s^2_{yx}}{s^2_y} = \frac{s^2_{yc}}{s^2_y} .$$

Saadud suhe näitab, kui suur osa suuruse y varieeruvusest on tingitud sõltumatu suuruse x muutumisest. Ülejäänud osa y varieeruvusest tuleb lugeda tingituks muudest mõjutustest.

7. punkti näite andmetel saime r = 0,85. Leiame $r^2 = 0,73$ , mis näitab, et 73% päevapalga varieeruvusest on seletatav puhastulu muutumisega, ülejäänud 27% on tingitud muudest mõjutustest.

Korrelatsioonikoefitsiendi ruutu $r^2$ nimetatakse ka determinismikoefitsiendiks.

Suuremale $r^2$ väärtusele vastab ka suurem r väärtus. Kokkuvõttes saame öelda, et korrelatsioonikoefitsient r iseloomustab suurustevahelise seose tugevust, võimaldades hinnata, kui oluliselt mõjutab sõltumatu suuruse muutumine sõltuva suuruse muutumist.

5. Kui korrelatsioonitabel esitab väljavõtte andmeid, siis suurustevahelise seose kuju ja tugevus ei tarvitse üldkogumis ühtida väljavõtte jaoks saaduga. Selgub, et küllalt suure väljavõtte ($N \geqslant 50$) korral võib üldkogumi korrelatsioonikoefitsiendi $r_ü$ statistiliseks hinnanguks võtta väljavõtte andmetel leitud r keskmise ruutveaga

$$\sigma_r = \frac{1 - r^2}{\sqrt{N}}$$

ja võib olla praktiliselt kindel, et $r_ü$ asub usalduspiirides

$$r - 3\sigma_r \leq r_ü \leq r + 3\sigma_r .$$

Näide: Leiame korrelatsioonikoefitsiendi $r_ü$ usalduspiirid Tšehhoslovakkia kõigi põllumajanduslike kooperatiivide jaoks (näide 7. punktis).

Keskmine ruutviga

$$\sigma_r = \frac{1 - 0{,}853^2}{\sqrt{120}} = \frac{0{,}272}{10{,}95} \approx 0{,}025.$$

Seega

$$0{,}928 \leq r_ü \leq 0{,}778.$$

## Ülesanded.

### 1.

1. Kastis on 50 ühesugust detaili, millest 5 on värvitud siniseks. Kui tõenäone on, et valikuta võetud detail on sinist värvi?

2. Kui tõenäone on täringu viskamisel kolmega jaguva silmade arvu saamine?

3. Õpperühmas on 8 nais- ja 17 meesüliõpilast. Kui suur on naisüliõpilaste relatiivne sagedus? Kui tõenäone on, et üliõpilaste nimekirjas on esikohal naisüliõpilane?

4. Tehase TKO avastas 100 valmistoote hulgas 5 mittestandardset toodet. Leida standardsete toodete relatiivne sagedus.

5. Märklaua tabamise relatiivne sagedus on 0,85. Leida tabamuste arv, kui sooritati 120 lasku.

6. Kui tõenäone on, et uue passi saamisel passi number lõpeb paarisnumbriga? Nulliga?

7. Õppejõud palus kursusevanemal kutsuda konsultatsioonile 3 üliõpilast 6 kontrolltöö mitterahuldavalt kirjutanud üliõpilase hulgast. Kursusevanem unustas kutsutute nimed ja saatis 3 üliõpilast oma äranägemisel. Kui tõenäone on, et ta saatis õppejõu poolt kutsutud üliõpilased?

8. Tõenäosus märklaua tulistamisel 10 silma saamiseks on 0,1, 9 silma saamiseks 0,3, 8 või vähema silma saamiseks 0,6. Kui tõenäone on ühe lasu sooritamisel vähemalt 9 silma saamine?

9. Raha- ja asjadeloterii iga 1000 pileti kohta tuleb 5 rahalist ja 25 esemelist võitu. Kui tõenäone on võita ühe piletiga?

10. Valmistoodangu hulgas on 40 I sordi, 60 II sordi, 45 III sordi ja 5 praaktoodet. Kui tõenäone on vali-

kuta võtmisel I või II sordi toote saamine?

11. Loteriil võidavad kõik piletid. 200 võidu hulgas on hinnalisematest võitudest 10 täitesulepead, 30 täitepliiatsit, 40 keraamilist tuhatoosi, Ülejäänud on väiksemad võidud. Mis on tõenäosem, kas ühe piletiga saada hinnalisem või väiksem võit?

12. Detaili valmistamisel sooritatakse 3 põhioperatsiooni. Esimesel operatsioonil on praagi tõenäosus 0,01, teisel - 0,02, kolmandal 0,03. Leida standardse detaili valmistamise tõenäosus eeldusel, et praagi tekkimine ühel operatsioonil ei mõjuta teiste operatsioonide tulemusi.

13. Tõenäosus raamatu asumiseks esimeses raamatukogus on 0,5, teises- 0,7 , kolmandas - 0,4. Kui tõenäone on, et raamat on olemas vähemalt ühes raamatukogus?

14. Tõenäosus ülesande lahendamiseks üliõpilase A poolt on 0,75 , üliõpilase B poolt 0,80 . Kui tõenäone on, et ülesande lahendab ükskõik kumb üliõpilastest, kui mõlemad üliõpilased töötavad teineteisest eraldi?

15. Kiirrongi õigeaegse saabumise tõenäosus on 0,95. Kui tõenäone on, et 3 kiirrongi saabuvad hilinemata?

16. Kodanik ostis televiisori ja raadiovastuvõtja. Tõenäosus häireteta tööks garantiiaja jooksul on televiisoril 0,85 , raadiovastuvõtjal 0,98. Kui tõenäone on, et mõlemad aparaadid ei vaja remonti garantiiaja jooksul?

## 2.

1. Koostada 3 täringu viskamisel saadavate silmade summa jaotustabel ja polügoon.

2. Koostada kaupluses päeva jooksul müüdud 45 kingapaari suuruste variatsioonrida ja polügoon järgmistel andmetel:

39 41 40 42 41 40 42 44 40 43 42 41 43 39 42
41 42 39 41 37 43 41 38 43 42 41 40 41 38 44
40 39 41 40 42 40 41 42 40 43 38 39 41 41 42 .

3. Koostada histogramm ja kumulaat kolhooside jaotusele

traktoripargi võimsuse järgi.

| Võimsus (HJ) | kuni 1000 | 1000--1500 | 1500--2000 | 2000--2500 | 2500--3000 | Üle 3000 |
|---|---|---|---|---|---|---|
| Kolhooside arv (%) | 5,5 | 27,2 | 34,6 | 18,8 | 7,8 | 6,1 |

4. Mõõtmisvea X jaotustabeli põhjal koostada jaotusfunktsioon ja kumulaat.

| Vea intervall | (0,3) | (3,6) | (6,9) | (9,12) |
|---|---|---|---|---|
| p | 0,035 | 0,050 | 0,166 | 0,212 |

| Vea intervall | (12,15) | (15,18) | (18,21) | (21,24) |
|---|---|---|---|---|
| p | 0,212 | 0,195 | 0,090 | 0,090 |

3.

1. Juhusliku suuruse X jaotustabeli põhjal leida keskväärtus, keskmine lineaarhälve ja dispersioon

| X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| p | 0,2 | 0,25 | 0,35 | 0,10 | 0,10 |

2. Leida § 2 ülesande 4 andmetel mõõtmisvea keskväärtus, dispersioon ja keskmine standardhälve.

3. Leida § 2 ülesande 2 andmetel kingapaaride suuruse aritmeetiline keskmine ja keskmine standardhälve.

4. Leida § 2 ülesande 3 andmetel traktoripargi keskmine võimsus ja keskmine standardhälve.

5. Perekondade jaotus tulu järgi ühe perekonnaliikme kohta on järgmine:

| Tulu(rbl). | Kuni 25 | 25 - 50 | 50 - 75 | 75 - 100 | Üle 100 |
|---|---|---|---|---|---|
| Perek. arv (%) | 5 | 15 | 40 | 25 | 15 |

Leida keskmine tulu, jaotuse mood ja mediaan.

4.

1. Urnis on 25 kuuli: 10 punast ja 15 rohelist. Urnist võetakse kuule 5 korda, kusjuures pärast katset pannakse need urni tagasi. Koostada võimalike katsetulemuste jaotustabel (Bernoulli valem) ja polügoon.

2. Kaugõppija tellib õpikuid posti teel. Tõenäosus selleks, et õpik on kõvas köites, on 0,7. Kui tõenäone on, et tellitud 7 õpikust on 4 kõvas köites?

3. Spordiklubi esindusvõistkonnas on 10 meistersportlast. Tõenäosus meistersportlase arvamiseks vabariigi koondvõistkonda on 0,5. Kui tõenäone on, et 8 sportlast 10-st valitakse vabariigi koondvõistkonda?

4. Kogutoodangus on keskmiselt 3% praaki. Leida tõenäoseim praagi hulk 800 toote hulgas ja vastava sageduse tõenäosus.

5. Münti visatakse 40 korda. Kui tõenäone on vapi esiletulek 15 kuni 25 korda?

6. Mitu katset on vaja sooritada, et usaldusnivooga 0,90 võib väita, et relatiivse sageduse hälve sündmuse tõenäosuse $p = 0,40$ suhtes ei ületa 0,04?

7. Mitu seemet tuleks idanevusproovi tegemiseks võtta,

et tõenäosusega 0,977 võiks väita, et idanemisvõimeliste seemnete relatiivne sagedus ei erine 0,9-st rohkem kui 0,02 võrra?

8. Kontrolliti 3000 elektripirni. Praagi relatiivne sagedus on 0,15. Kui tõenäone on, et praagi relatiivne sagedus kontrollitud pirnide hulgas ei erine 0,15-st rohkem kui 0,01 võrra?

9. Sündmuse tõenäosus üksikkatsel on 0,3. Leida sündmuse relatiivse sageduse piirviga tõenäosuse 0,3 suhtes, mida võib garanteerida tõenäosusega 0,95.

10. Tõenäosus lennuki tabamiseks õhutõrjekahurist ühe lasuga on 0,01. Sooritatakse 100 lasku. Leida tõenäosus lennuki tabamiseks 2 mürsuga Poissoni, Bernoulli ja Laplace'i valemiga.

11. Juhuslik suurus X allub normaaljaotusele, mille parameetrid on $a = 5$, $\sigma = 2$. Kui tõenäone on, et X väärtused asuvad vahemikus (1,10)?

12. Normaaljaotusele alluva juhusliku suuruse parameetrid on $a = 3$ , $\sigma = 0,5$. Kui tõenäone on, et X väärtused ei hälbi keskväärtuse suhtes rohkem kui 1,3 võrra?

13. Traadi tõmbetugevuse jaotus 100 traadi katsetamisel saadi järgmine

| Koormus (kg) | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 | 100-110 | 110-120 |
|---|---|---|---|---|---|---|---|
| Traatide arv | 3 | 9 | 20 | 28 | 24 | 12 | 4 |

Tasandada antud statistilist jaotust normaaljaotusega. Võrrelda normaaljaotuse kaudu leitud sagedusi eksperimentaalsete tulemustega.

## 6.

1. Toodangu kvaliteedi hindamiseks moodustati 10 000 tootest koosneva üldkogumi jaoks väljavõte mahuga n = 300.

Esimese sordi toodangu relatiivne sagedus väljavõttes oli 0,6. Määrata esimese sordi toodangu usalduspiirid üldkogumis usaldusnivooga 0,999 a) korduval, b) kordumatul väljavõttel.

2. Kartuli tärklisesisalduse selgitamiseks mõõdeti tärklise hulk 560 kartulil. Mõõtmistulemused on antud variatsioonreana

| Tärklise hulk % | 9 - | 10 - | 11 - | 12 - | 13 - | 14 - | 15 - | 16 - | 17 - |
|---|---|---|---|---|---|---|---|---|---|
| Kartulite arv | 1 | 2 | 6 | 12 | 19 | 26 | 46 | 94 | |

| Tärklise hulk % | 17 - | 18 - | 19 - | 20 - | 21 - | 22 - | 23 - | 24 - | 25 - | 26 |
|---|---|---|---|---|---|---|---|---|---|---|
| Kartulite arv | 106 | 110 | 76 | 39 | 14 | 7 | 1 | 0 | 1 | |

Määrata üldkeskmise usalduspiirid usaldusnivooga 0,96.

3. Kvalifitseeritud tööliste keskmise tööstaaži selgitamiseks kontrolliti 10 000 töölise tööraamatuid korduva väljavõtte teel. Keskmiseks tööstaažiks saadi 15,5 aastat keskmise standardhälbega 4,8 aastat. Kui tõenäone on, et üldkeskmine ei erine väljavõttekeskmisest rohkem kui 0,5 aastat.

4. Uuritakse telefonikõnede keskmist pikkust. Mitme kõne pikkus tuleb registreerida, et tõenäosusega 0,997 võiks garanteerida, et väljavõttekeskmise ja üldkeskmise erinevus ei ületa 10 sekundit keskmise ruuthälbega 2,5 minutit?

5. Raamatukogude lugejate keskmise vanuse määramiseks kasutatakse väljavõttelist meetodit. Mitu lugejakaarti tu-

leks 30 000 lugejakaardi hulgast kontrollimiseks võtta, et usaldusnivooga 0,99 võiks garanteerida, et üldkeskmise ja väljavõtte keskmise erinevus ei ületa 1 aastat keskmise standardhälbega 5 aastat?

## 7.

1. On antud andmed 1950.a. võrreldes (protsentides) NSVL rahvamajanduses töötavate inimeste arvu (Y), põhifondide (X) ja tööstuse kogutoodangu (Z) kohta. Koostada regressioonivõrrandid, mis väljendavad Y ja X, samuti Z ja X vahelist sõltuvust. Leida vastavad korrelatsioonikoefitsiendid.

| Aastad | 1950 | 1951 | 1952 | 1953 | 1954 | 1955 |
|---|---|---|---|---|---|---|
| X | 100 | 110 | 121 | 133 | 147 | 164 |
| Y | 100 | 105 | 109 | 112 | 122 | 124 |
| Z | 100 | 116 | 130 | 145 | 165 | 185 |

2. Korrelatsioonitabelis on toodud andmed jõe veenivoo (X) ja ettevõtete poolt tarvitatud vee hulga kohta päevade järgi. Kõrgus X on mõõdetud teatud algnivoo suhtes.

$$X - \text{cm}; \quad Y - \frac{m^3}{s}$$

| y \ x | 30 | 60 | 90 | 120 | 150 | 180 | 210 |
|---|---|---|---|---|---|---|---|
| 100 | 3 | 2 | 2 | 1 | | | |
| 250 | 1 | 4 | 3 | 1 | | | |
| 300 | | 1 | 5 | 2 | | | |
| 400 | | 1 | 4 | 4 | 1 | | |
| 500 | | | | 2 | 3 | | |
| 600 | | | | 1 | 4 | 1 | |
| 700 | | | | 1 | 3 | 2 | 1 |
| 800 | | | | | 1 | 4 | 5 |

Leida korrelatsioonikoefitsient ja koostada regressioonivõrrandid.

3. 30 kolhoosi kartulisaagi X ($\frac{ts}{ha}$) ja tsentneri omahinna Y (rbl.) jaotus on antud korrelatsioonitabeliga

| y \ x | 60-90 | 90-120 | 120-150 | 150-180 |
|---|---|---|---|---|
| 2,5-3,5 | - | 1 | 2 | 2 |
| 3,5-4,5 | - | 3 | 5 | 2 |
| 4,5-5,5 | 5 | 4 | 3 | - |
| 5,5-6,5 | 3 | - | - | - |

Koostada regressioonivõrrandid ja leida korrelatsioonikoefitsient.

V a s t u s e d

1.

1. 0,1 , 2. 0,333. 3. 0,32. 4. 0,95. 5. 102
6. 0,5;0,1 7. 0,05. 8. 0,4. 9. 0,03. 10. 0,667
12. 0,94 13. 0,91 14. 0,95 15. 0,59
16. 0,86 17. 0,853.

2.

1. $E(x)=1,65$ $D(x)=4,15$, 2. $E(x)=13,928$, $D(x)=28,345$.
4. $\bar{x} = 1890$ 5. $\bar{x}=70$, $M_o = 65$, $M_e = 69$.

4.

2. 0,177. 4. $\mu=24$, $p=0,083$. 5. 0,886 6. 406

7. 1160 8. 0,8764 9. 0,021. 10. 0,184, 0,185, 0,282.

11. 0,971 12. 0,768 13. $a=86,3$ , $\sigma^2 = 1802,2$.

6.

1. $0,6 \pm 0,094$ , $0,6 \pm 0,093$ . 2. $\varepsilon = 0,19$. 3. 0,8444

4. 1985 5. 166.

7.

1. $\rho_{yx} = 2,5$ , $\rho_{xy} = 0,384$ , $\rho_{xz} = 0,76$ , $\rho_{zx} = 1,32$

2. $r = 0,896$ 3. $r = -0,685$.

881

## Lisa I

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{t^2}{2}}.$$

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,3989 | 3989 | 3989 | 3988 | 3986 | 3984 | 3982 | 3980 | 3977 | 3973 |
| 0,1 | 3970 | 3965 | 3961 | 3956 | 3951 | 3945 | 3939 | 3932 | 3925 | 3918 |
| 0,2 | 3910 | 3902 | 3894 | 3885 | 3876 | 3867 | 3857 | 3847 | 3836 | 3825 |
| 0,3 | 3814 | 3802 | 3790 | 3778 | 3765 | 3752 | 3739 | 3726 | 3712 | 3697 |
| 0,4 | 3683 | 3668 | 3653 | 3637 | 3621 | 3605 | 3589 | 3572 | 3555 | 3538 |
| 0,5 | 3521 | 3503 | 3485 | 3467 | 3448 | 3429 | 3410 | 3391 | 3372 | 3352 |
| 0,6 | 3332 | 3312 | 3292 | 3271 | 3251 | 3230 | 3209 | 3187 | 3166 | 3144 |
| 0,7 | 3123 | 3101 | 3079 | 3056 | 3034 | 3011 | 2989 | 2966 | 2943 | 2920 |
| 0,8 | 2897 | 2874 | 2850 | 2827 | 2803 | 2780 | 2756 | 2732 | 2709 | 2685 |
| 0,9 | 2661 | 2637 | 2613 | 2589 | 2565 | 2541 | 2516 | 2492 | 2468 | 2444 |
| 1,0 | 0,2420 | 2396 | 2371 | 2347 | 2323 | 2299 | 2275 | 2251 | 2227 | 2203 |
| 1,1 | 2179 | 2155 | 2131 | 2107 | 2083 | 2059 | 2036 | 2012 | 1989 | 1965 |
| 1,2 | 1942 | 1919 | 1895 | 1872 | 1849 | 1826 | 1804 | 1781 | 1758 | 1736 |
| 1,3 | 1714 | 1691 | 1669 | 1647 | 1626 | 1604 | 1582 | 1561 | 1539 | 1518 |
| 1,4 | 1497 | 1476 | 1456 | 1435 | 1415 | 1394 | 1374 | 1354 | 1334 | 1315 |
| 1,5 | 1295 | 1276 | 1257 | 1238 | 1219 | 1200 | 1182 | 1163 | 1145 | 1127 |
| 1,6 | 1109 | 1092 | 1074 | 1057 | 1040 | 1023 | 1006 | 0989 | 0973 | 0957 |
| 1,7 | 0940 | 0925 | 0909 | 0893 | 0878 | 0863 | 0848 | 0833 | 0818 | 0804 |
| 1,8 | 0790 | 0775 | 0761 | 0748 | 0734 | 0721 | 0707 | 0694 | 0681 | 0669 |
| 1,9 | 0656 | 0644 | 0632 | 0620 | 0608 | 0596 | 0584 | 0573 | 0562 | 0551 |
| 2,0 | 0,0540 | 0529 | 0519 | 0508 | 0498 | 0488 | 0478 | 0468 | 0459 | 0449 |
| 2,1 | 0440 | 0431 | 0422 | 0413 | 0404 | 0396 | 0387 | 0379 | 0371 | 0363 |
| 2,2 | 0355 | 0347 | 0339 | 0332 | 0325 | 0317 | 0310 | 0303 | 0297 | 0290 |
| 2,3 | 0283 | 0277 | 0270 | 0264 | 0258 | 0252 | 0246 | 0241 | 0235 | 0229 |
| 2,4 | 0224 | 0219 | 0213 | 0208 | 0203 | 0198 | 0194 | 0189 | 0184 | 0180 |
| 2,5 | 0175 | 0171 | 0167 | 0163 | 0158 | 0154 | 0151 | 0147 | 0143 | 0139 |
| 2,6 | 0136 | 0132 | 0129 | 0126 | 0122 | 0119 | 0116 | 0113 | 0110 | 0107 |
| 2,7 | 0104 | 0101 | 0099 | 0096 | 0093 | 0091 | 0088 | 0086 | 0084 | 0081 |
| 2,8 | 0079 | 0077 | 0075 | 0073 | 0071 | 0069 | 0067 | 0065 | 0063 | 0061 |
| 2,9 | 0060 | 0058 | 0056 | 0055 | 0053 | 0051 | 0050 | 0048 | 0047 | 0046 |
| 3,0 | 0,0044 | 0043 | 0042 | 0040 | 0039 | 0038 | 0037 | 0036 | 0035 | 0034 |
| 3,1 | 0033 | 0032 | 0031 | 0030 | 0029 | 0028 | 0027 | 0026 | 0025 | 0025 |
| 3,2 | 0024 | 0023 | 0022 | 0022 | 0021 | 0020 | 0020 | 0019 | 0018 | 0018 |
| 3,3 | 0017 | 0017 | 0016 | 0016 | 0015 | 0015 | 0014 | 0014 | 0013 | 0013 |
| 3,4 | 0012 | 0012 | 0012 | 0011 | 0011 | 0010 | 0010 | 0010 | 0009 | 0009 |
| 3,5 | 0009 | 0008 | 0008 | 0008 | 0008 | 0007 | 0007 | 0007 | 0007 | 0006 |
| 3,6 | 0006 | 0006 | 0006 | 0005 | 0005 | 0005 | 0005 | 0005 | 0005 | 0004 |
| 3,7 | 0004 | 0004 | 0004 | 0004 | 0004 | 0004 | 0003 | 0003 | 0003 | 0003 |
| 3,8 | 0003 | 0003 | 0003 | 0003 | 0003 | 0002 | 0002 | 0002 | 0002 | 0002 |
| 3,9 | 0002 | 0002 | 0002 | 0002 | 0002 | 0002 | 0002 | 0002 | 0001 | 0001 |

# Lisa II

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,00000 | 00399 | 00798 | 01197 | 01595 | 01994 | 02392 | 02790 | 03188 | 03586 |
| 0,1 | 03983 | 04380 | 04776 | 05172 | 05567 | 05962 | 06356 | 06749 | 07142 | 07535 |
| 0,2 | 07926 | 08317 | 08706 | 09095 | 09483 | 09871 | 10257 | 10642 | 11026 | 11409 |
| 0,3 | 11791 | 12172 | 12552 | 12930 | 13307 | 13683 | 14058 | 14431 | 14803 | 15173 |
| 0,4 | 15542 | 15910 | 16276 | 16640 | 17003 | 17364 | 17724 | 18082 | 18439 | 18793 |
| 0,5 | 19146 | 19497 | 19847 | 20194 | 20540 | 20884 | 21226 | 21566 | 21904 | 22240 |
| 0,6 | 22575 | 22907 | 23237 | 23565 | 23891 | 24215 | 24537 | 24857 | 25175 | 25490 |
| 0,7 | 25804 | 26115 | 26424 | 26730 | 27035 | 27337 | 27637 | 27935 | 28230 | 28524 |
| 0,8 | 28814 | 29103 | 29389 | 29673 | 29955 | 30234 | 30511 | 30785 | 31057 | 31327 |
| 0,9 | 31594 | 31859 | 32121 | 32381 | 32639 | 32894 | 33147 | 33398 | 33646 | 33891 |
| 1,0 | 34134 | 34375 | 34614 | 34850 | 35083 | 35314 | 35543 | 35769 | 35993 | 36214 |
| 1,1 | 36433 | 36650 | 36864 | 37076 | 37286 | 37493 | 37698 | 37900 | 38100 | 38298 |
| 1,2 | 38493 | 38686 | 38877 | 39065 | 39251 | 39435 | 39617 | 39796 | 39973 | 40147 |
| 1,3 | 40320 | 40490 | 40658 | 40824 | 40988 | 41149 | 41309 | 41466 | 41621 | 41774 |
| 1,4 | 41924 | 42073 | 42220 | 42364 | 42507 | 42647 | 42786 | 42922 | 43056 | 43189 |
| 1,5 | 43319 | 43448 | 43574 | 43699 | 43822 | 43943 | 44062 | 44179 | 44295 | 44408 |
| 1,6 | 44520 | 44630 | 44738 | 44845 | 44950 | 45053 | 45154 | 45254 | 45352 | 45449 |
| 1,7 | 45543 | 45637 | 45728 | 45818 | 45907 | 45994 | 46080 | 46164 | 46246 | 46327 |
| 1,8 | 46407 | 46485 | 46562 | 46638 | 46712 | 46784 | 46856 | 46926 | 46995 | 47062 |
| 1,9 | 47128 | 47193 | 47257 | 47320 | 47381 | 47441 | 47500 | 47558 | 47615 | 47670 |
| 2,0 | 47725 | 47778 | 47831 | 47882 | 47932 | 47982 | 48030 | 48077 | 48124 | 48169 |
| 2,1 | 48214 | 48257 | 48300 | 48341 | 48382 | 48422 | 48461 | 48500 | 48537 | 48574 |
| 2,2 | 48610 | 48645 | 48679 | 48713 | 48745 | 48778 | 48809 | 48840 | 48870 | 48899 |
| 2,3 | 48928 | 48956 | 48983 | 49010 | 49036 | 49061 | 49086 | 49111 | 49134 | 49158 |
| 2,4 | 49180 | 49202 | 49224 | 49245 | 49266 | 49286 | 49305 | 49324 | 49343 | 49361 |
| 2,5 | 49379 | 49396 | 49413 | 49430 | 49446 | 49461 | 49477 | 49492 | 49506 | 49520 |
| 2,6 | 49534 | 49547 | 49560 | 49573 | 49585 | 49598 | 49609 | 49621 | 49632 | 49643 |
| 2,7 | 49653 | 49664 | 49674 | 49683 | 49693 | 49702 | 49711 | 49720 | 49728 | 49736 |
| 2,8 | 49744 | 49752 | 49760 | 49767 | 49774 | 49781 | 49788 | 49795 | 49801 | 49807 |
| 2,9 | 49813 | 49819 | 49825 | 49831 | 49836 | 49841 | 49846 | 49851 | 49856 | 49861 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3,0 | 0,49865 | 3,1 | 49903 | 3,2 | 49931 | 3,3 | 49952 | 3,4 | 49966 |
| 3,5 | 49977 | 3,6 | 49984 | 3,7 | 49989 | 3,8 | 49993 | 3,9 | 49995 |
| 4,0 | 499968 | | | | | | | | |
| 4,5 | 499997 | | | | | | | | |
| 5,0 | 49999997 | | | | | | | | |

$$\bar{\Phi}(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0080 | 0160 | 0239 | 0319 | 0399 | 0478 | 0558 | 0638 | 0717 |
| 0.1 | 0.0797 | 0876 | 0955 | 1034 | 1113 | 1192 | 1271 | 1359 | 1429 | 1507 |
| 0.2 | 0.1585 | 1662 | 1741 | 1819 | 1897 | 1974 | 2051 | 2128 | 2205 | 2282 |
| 0.3 | 0.2358 | 2434 | 2510 | 2586 | 2661 | 2737 | 2812 | 2886 | 2961 | 3035 |
| 0.4 | 0.3108 | 3182 | 3255 | 3328 | 3401 | 3473 | 3545 | 3616 | 3688 | 3759 |
| 0.5 | 0.3829 | 3900 | 3969 | 4039 | 4108 | 4177 | 4245 | 4313 | 4381 | 4448 |
| 0.6 | 0.4515 | 4581 | 4647 | 4713 | 4778 | 4843 | 4908 | 4971 | 5035 | 5098 |
| 0.7 | 0.5161 | 5223 | 5285 | 5346 | 5407 | 5468 | 5528 | 5587 | 5646 | 5705 |
| 0.8 | 0.5763 | 5821 | 5878 | 5935 | 5991 | 6047 | 6102 | 6157 | 6211 | 6265 |
| 0.9 | 0.6319 | 6372 | 6424 | 6476 | 6528 | 6579 | 6629 | 6680 | 6729 | 6778 |
| 1.0 | 0.6827 | 6875 | 6923 | 6970 | 7017 | 7063 | 7109 | 7154 | 7199 | 7243 |
| 1.1 | 0.7287 | 7330 | 7373 | 7415 | 7457 | 7499 | 7540 | 7580 | 7620 | 7660 |
| 1.2 | 0.7699 | 7737 | 7775 | 7813 | 7850 | 7887 | 7923 | 7959 | 7995 | 8030 |
| 1.3 | 0.8064 | 8098 | 8132 | 8165 | 8198 | 8230 | 8262 | 8293 | 8324 | 8355 |
| 1.4 | 0.8385 | 8415 | 8444 | 8473 | 8501 | 8529 | 8557 | 8584 | 8611 | 8638 |
| 1.5 | 0.8664 | 8690 | 8715 | 8740 | 8764 | 8789 | 8813 | 8836 | 8859 | 8882 |
| 1.6 | 0.8904 | 8926 | 8948 | 8969 | 8990 | 9011 | 9031 | 9051 | 9070 | 9090 |
| 1.7 | 0.9109 | 9127 | 9146 | 9164 | 9181 | 9199 | 9216 | 9233 | 9249 | 9266 |
| 1.8 | 0.9281 | 9297 | 9312 | 9328 | 9342 | 9357 | 9371 | 9385 | 9399 | 9412 |
| 1.9 | 0.9426 | 9439 | 9451 | 9464 | 9476 | 9488 | 9500 | 9512 | 9523 | 9534 |
| 2.0 | 0.9545 | 9556 | 9566 | 9576 | 9587 | 9596 | 9606 | 9616 | 9625 | 9634 |
| 2.1 | 0.9643 | 9651 | 9660 | 9668 | 9677 | 9684 | 9692 | 9700 | 9707 | 9715 |
| 2.2 | 0.9722 | 9729 | 9736 | 9743 | 9749 | 9756 | 9762 | 9768 | 9774 | 9780 |
| 2.3 | 0.9786 | 9791 | 9797 | 9802 | 9807 | 9812 | 9817 | 9822 | 9827 | 9832 |
| 2.4 | 0.9836 | 9841 | 9845 | 9849 | 9853 | 9857 | 9861 | 9865 | 9869 | 9872 |
| 2.5 | 0.9876 | 9879 | 9883 | 9886 | 9889 | 9892 | 9895 | 9898 | 9901 | 9904 |
| 2.6 | 0.9907 | 9910 | 9912 | 9915 | 9917 | 9920 | 9922 | 9924 | 9926 | 9929 |
| 2.7 | 0.9931 | 9933 | 9935 | 9937 | 9939 | 9941 | 9942 | 9944 | 9946 | 9947 |
| 2.8 | 0.9949 | 9951 | 9952 | 9954 | 9955 | 9956 | 9958 | 9959 | 9960 | 9962 |
| 2.9 | 0.9963 | 9964 | 9965 | 9966 | 9967 | 9968 | 9969 | 9970 | 9971 | 9972 |
| 3.0 | 0.9973 | 9974 | 9975 | 9976 | 9976 | 9977 | 9978 | 9979 | 9979 | 9980 |
| 3.1 | 0.9981 | 9981 | 9982 | 9983 | 9983 | 9984 | 9984 | 9985 | 9985 | 9986 |
| 3.2 | 0.9986 | 9987 | 9987 | 9988 | 9988 | 9989 | 9989 | 9989 | 9990 | 9990 |
| 3.3 | 0.9990 | 9991 | 9991 | 9991 | 9992 | 9992 | 9992 | 9993 | 9993 | 9993 |
| 3.4 | 0.9999 | 9994 | 9994 | 9994 | 9994 | 9994 | 9995 | 9995 | 9995 | 9995 |
| 3.5 | 0.9995 | 9996 | 9996 | 9996 | 9996 | 9996 | 9996 | 9996 | 9997 | 9997 |
| 3.6 | 0.9997 | 9997 | 9997 | 9997 | 9997 | 9997 | 9998 | 9998 | 9998 | 9998 |
| 3.7 | 0.9998 | 9998 | 9998 | 9998 | 9998 | 9998 | 9998 | 9998 | 9998 | 9999 |
| 3.8 | 0.9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 |
| 3.9 | 0.9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 | 9999 |

S i s u k o r d

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

К.П. Соонэтс

ТЕОРИЯ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА

На эстонском языке
Издание второе

Vastutav toimetaja E. Jõgi

Korrektor A. Norberg

TRÜ rotaprint 1968. Trükipoognaid 12,13. Tingtrükipoognaid 11,03. Arvestuspoognaid 8,75. Trükiarv 1000. Paber 30x42/4. Paljundamisele antud 3. VII 1968. MB 04393. Tell. nr. 439.

Hind 25 kop.

1.–

Hind ~~25~~ kop.