



TARTU RIIKLIK ÜLIKOOL

Teoreetilise mehaanika kateeder

K. Soonets

TÕENÄOSUSTEORIA JA
MATEMAATILINE STATISTIKA

(Majandusteaduskonna üliõpilastele)

Kolmas, täiendatud trükk

Tartu 1974

Kinnitatud Matemaatikateaduskonna nõukogus 29. detsem-
ril 1972. .

S a a t e k s

Konspekt sisaldab materjali tõenäosusteooria ja matemaatilise statistika kursuse ulatuses TRÜ Majandusteaduskonnale.

Tõenäosusteooria teoreetilisi küsimusi ja nende rakedust 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.

Konspektis leidub ülesandeid iseseisvaks lahendamiseks. Teoreetilise materjali kinnistamiseks on mitmel pool jäetud üksikküsimusi iseseisvaks lahendamiseks. Nendele küsimustele vastuse leidmise käigus saab üliõpilane kontrollida, kuidas on aru saadud eelnevatest punktidest.

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

Käesolev, kolmas trükk kujutab eelmiste täiendatud ja parandatud väljaannet. Oluliselt on ümber töötatud valimiteooria ja korrelatsiooniteooria paragrahvid, väiksemaid muudatusi on ka ülejäänud paragrahvides. Harjutusülesannete valik on laiem. Olulisemad valemid on toodud konspekti lõpus.

Avaldan tänu kolleegidele E. Tiidule, R. Tammestele ning M. Kullile ja E. Paabutile suure abi eest konspekti vormistamisel.

Autor

1. Tõenäosusteooria ja matemaatilise statistika aine

Paljudel elualadel tuleb kokku puutada nähtustega, mis ka sarnastes tingimustes on teatud määral erinevad. Näiteks ühel katselapil kasvavates viljapeades on terade arv erinev; võrdse suurusega perekondades ei tarbita piima samal hulgal; tekstiilikäitises töötavate samatüübiliste automaattelgede seisakute arv ühe tunni jooksul on erinev; täringuga igal viskel saadav silmade arv võib olla erinev. 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. Sellisel toimides uurime reaalse nähtuse asemel selle lihtsustatud mudelit. Kui aga paljude teisejärguliste tegurite koosmõju nähtust oluliselt mõjutab, 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 nimetame nähtust, mis esineb paljude kõrvaliste põhjuste koosmõjul ühe ja sama katse kordamisel teatud erinevustega.

✓ Tõenäosusteooria uurimisobjektiks on juhuslikud nähtused: ta selgitab ja kirjeldab nende nähtuste massilisel toimimisel kehtivaid seaduspärasusi matemaatiliste meetoditega. Tõenäosusteooria abil saame teha õigeid järeldusi massiliste katsete tulemustest, sest iga üksikkatse tulemus võib sisaldada palju juhuslikke, vaadeldavale nähtus-

te 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 võib seaduspärasuste selgitamiseks kasutada ka tõenäosusteooria meetodeid ja neid 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 lahutamatu seotud tõenäosusteooriaga. Matemaatilises statistikas vaadeldavad spetsiifilistest tunnustest „puhastatud“ nähtused alluvad tõenäosusteooria seadustele, mis on massiliste juhuslike nähtuste seaduspärasuste matemaatiliseks väljenduseks. Tõenäosusteoorial on matemaatilises statistikas umbes sama osa, mis geomeetrial mõõdistamises või joonestamises.

§ 1. TÕENÄOSUSTEORIA ALGMÕISTED

1. Sündmused

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

Sündmus saab toimuda teatud tingimustes. Teatud tingimuste kompleksi realiseerimist nimetame katseks. Katsel võib sündmus toimuda või mitte toimuda.

Näiteid

1. Laskur tulistab märklauda. Katseks on lasu sooritamine, märklaua teatud tsooni tabamine on sündmuseks.
2. Urnis on mitut värvi kuuliid. 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 räägitakse katseseeriast. Iga katseseeriasse kuuluvat katsset nimetatakse üksikkatseks.

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

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

Sündmust nimetatakse juhuslikuks, kui see antud katsel võib toimuda ja võib ka mitte toimuda. Juhusliku sündmuse toimumist või mittetoimumist ei saa garanteerida, seda iseloomustab toimumise võimalikkus.

N ä i t e i d :

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. Perekonnas poisilapse sündimine on juhuslik sündmus.
6. 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. S ü n d m u s t e l i i g i t a m i s e s t

A. Kaht sündmust nimetatakse teineteist välistavateks, kui ühe sündmuse toimumine antud katsel välistab teise sündmuse toimumise samal katsel ehk, kui nad koos toimuda ei saa.

N ä i t e i d :

1. Vapi ja kirja esiletulek ühel raha viskel on teineteist välistavad sündmused. Kui visata 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. Ildigitame 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 mitte-standardseteks, siis detaili kuulumine I sordi ja standardsete hulka ei ole enam välistavad. Detaili kuulumine standardsete hulka ei välista võimalust kuuluda samaaegselt ka I sorti.

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

N ä i t e i d :

1. Kui lugeda täringu viskel erineva silmade arvu saamist omaette sündmuseks, võib toimuda üks kuuest sündmusest. Ühe kuni kuue 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, mis on ka nummerdatud ühest kümneni. Kuuli võtmisel ei ole valge ja musta kuuli saamine võrdvõimalikud sündmused. Järjekorranumbritega 1, 2, 3 jne. kuuli saamine on võrdvõimalikud sündmused.
3. Televisori ostmisel kvaliteetse aparaadi ja praakaparaadi saamine ei ole võrdvõimalikud sündmused - esimene sündmus toimub sagedamini, ta on palju tõenäosem.
4. Iga loteriipiletiga võitmise võimalused on võrdsed - sündmused on võrdvõimalikud.

C. Olgu antud mingi hulk üksteist välistavaid sündmusi. Kui katsel üks neist kindlasti toimub (ükskõik milline), siis öeldakse, et need sündmused moodustavad täieliku sündmuste süsteemi. Täielikku sündmuste süsteemi moodustavaid sündmusi nimetatakse sageli ka ainuvõimalikeks sündmusteks.

N ä i t e i d :

1. Raha viskel moodustab vapi või kirja tulek täieliku sündmuste süsteemi.
2. Täringu viskel kas 1, 2 jne. kuni 6 silma saamine moodustab täieliku sündmuste süsteemi.

Kui täielikku sündmuste süsteemi kuuluvad üksiksündmused on võrdvõimalikud, siis nimetatakse neid elementaarsündmusteks ehk juhtudeks. Võrdvõimalike sündmuste täieliku süsteemi nimetatakse elementaarsündmuste süsteemiks.

N ä i t e i d :

1. Sümmeetrilise täringu viskel on tegemist 6 juhust koosneva elementaarsündmuste süsteemiga.
2. Mängu alustamiseks on vaja saada täringu viskel 1 silm. Olgu 1 silma saamine sündmus A, mistahes muu arvu silmade saamine sündmus B. Siis A ja B moodustavad küll täieliku sündmuste süsteemi, kuid mitte enam elementaarsündmuste süsteemi, sest A ja B pole ilmselt võrdvõimalikud.

D. Katsel võib sündmus A toimuda või mitte toimuda. A mittetoimumist loeme uueks sündmuseks - „mitte A“. Üks neist kahest sündmusest toimub kindlasti, kolmandat võimalust ei ole. Sündmused A ja „mitte A“ moodustavad täieliku sündmuste süsteemi. Kaht täielikku sündmuste süsteemi moodustavat sündmust nimetatakse vastandsündmusteks.

Sündmuse A vastandsündmuse tähistamiseks kasutatakse harilikult sümbolit \bar{A} . Sündmused A ja \bar{A} moodustavad täieliku sündmuste süsteemi.

N ä i t e i d :

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 tõenäosus

Tõenäosusteoorias vaadeldakse sündmusi nende toimumise või mittetoimumise võimalikkuse seisukohalt. Ühete sündmuste toimumine võib olla rohkem või vähem oodatav kui teistel.

Sündmuse toimumise võimalikkuse astet iseloomustatakse tõenäosuse abil.

Klassikaline tõenäosus. Alustame näitest. Urnis olevast 20 kuulist on 7 mustad ja 13 valged. Kuulid on lisaks nummerdatud (1-7 mustad, 8-20 valged). Juhusliku kuuli võtmisel on valge kuuli saamiseks rohkem lootusi kui musta kuuli saamiseks. Ütleme, et valge kuuli tõenäosus on suurem musta kuuli tõenäosusest ning musta kuuli saamine ja valge kuuli saamine ei ole võrdvõimalikud sündmused. Järjekorranumbrite seisukohalt saab kuuli võtmisel toimuda üks 20-st sündmusest, mis moodustavad täieliku elementaarsündmuste süsteemi: mistahes järjekorranumbriga kuuli saamine on omaette juht. Nende 20 võimaliku juhuse hulgas on 7 sellist, millega kaasneb ka sündmus „kuul on must“ - M ja 13 juhust, millega kaasneb „kuul on valge“ - V. Ütleme, et 20 võimaliku juhu hulgas on 7 soodsat juhtu sündmusele M ja 13 soodsat juhtu sündmusele V. Sündmuse V tõenäosuseks loeme jagatise $13 : 20$. Kui valgete kuulide arv suureneks mustade arvel, siis suureneks ka analoogiline jagatis ja nii kajastuks õigesti asjaolu, et valge kuuli saamise võimalused on kasvanud.

Käsitleme edasi asja üldiselt. Olgu teada n juhust koosnev elementaarsündmuste süsteem. Kõigi juhtude hulgas olgu m juhtu, millega kaasneb sündmuse A toimumine - neid nimetatakse sündmuse A jaoks soodsateks juhtudeks.

Sündmuse A klassikaliseks tõenäosuseks nimetatakse sündmuse jaoks soodsate juhtude arvu m ja kõigi juhtude arvu n jagatist

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

Tõenäosuse leidmiseks toodud definitsiooni järgi tuleb määrata võrdvõimalikest üksiksündmustest koosnev täielik sündmuste süsteem, s.t. kõik võimalikud juhud ja nende hulgast soodsate juhtude arv sündmuse jaoks. Sündmuse toimumise võimalikkuse astet saab sel viisil määrata katseid sooritamata.

Klassikalise tõenäosuse definitsioon ei ole rakendatav, kui võimalike juhtude arv on lõpmatu või pole võimalik välja selgitada võrdvõimalikke üksiksündmusi. Sellistel juhtudel aitab meid geomeetrilise või kõige sagedamini statistilise tõenäosuse mõiste.

Geomeetriline tõenäosus. On antud märklaud pindalaga S . Elementaarsündmusteks on mistahes punkti tabamine. Tuleb leida märklaua osa suurusega s tabamise tõenäosus. Võimalike juhtude ja samuti soodsate juhtude arv on lõpmata suur ning klassikaline tõenäosuse definitsioon ei ole rakendatav.

Geomeetriliseks tõenäosuseks tasapinnal nimetatakse pindalade suhet $s : S$.

Statistiline tõenäosus. Sümmeetrilise täringu viskel on 6 võrdvõimalikku võimalust. Ühe silma saamise tõenäosus on klassikalise tõenäosuse definitsiooni kohaselt $1 : 6$. Mittesümmeetrilise täringu korral pole aga võimalused enam võrdvõimalikud ja mingi silmade arvu saamise tõenäosus pole enam leitav teoreetilistest kaalutlustest lähtudes. Sellisel juhul tuleb toetuda katsele.

Antud tingimustes sooritatakse n katset. Sündmuse toimumiste arv ehk sündmuse sagedus katseseerias olgu m . Sündmuse suhteliseks ehk relatiivseks sageduseks n katsest koosnevas katseseerias nimetatakse jagatist

$$w = \frac{m}{n}.$$

Lühikestes katseseeriates võib suhteline sagedus üksikutes seeriates olla oluliselt erinev. Katsete arvu suurenemisel on suhtelisel sagedusel tendents stabiliseeruda, enamasti kõigub see järjest väiksemas ulatuses teatud kindla arvu ümber.

Statistiliseks tõenäosuseks nimetatakse suhtelist sagedust küllalt suure katsete arvu korral.

Statistilist tõenäosust ei ole põhimõtteliselt võimalik täpselt määrata, sest ei saa korraldada lõpmata pikka katseseeriat. Ka paljude katsete korral säilib võimalus suhtelise sageduse ja tõenäosuse oluliseks erinemiseks, kuid selline võimalus realiseerub väga harva ehk see on vähe tõenäone (sellest tuleb lähemalt juttu suurte arvude seaduse käsitlemisel).

Edaspidi räägime sündmuse tõenäosusest, pööramata tähelepanu sellele, mil viisil see tõenäosus on määratud.

Paljude majanduse valdkonda kuuluvate probleemide lahendamisel tuleb kasutada tõenäosuse mõistet.

N ä i t e i d :

1. Kauplused peavad riietusesemeid ja jalanõusid tellima üksikute numbrite järgi erinevates kogustes, et rahuldada kõigi ostjate soove. Tuleks hinnata tellitavate koguste piisavust.
2. Töö häireteta kulgemiseks tehase tsehhis on tarvis hoolda 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 tõenäosust. 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. Ü l e s a n d e i d

1. Leida mündi viskamisel vapi esiletuleku tõenäosus.

Mündi viskel saab olla kaks juhtu: võib tulla vapp või kiri. Vapi tulekuks soodsaid juhte on üks. Seega on vapi esiletuleku tõenäosus $P(V) = 0,5$.

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 numbriga õige valik. Telefoniketta 10 numbrit annavad erinevaid võimalusi nii palju, kui palju saab moodustada variatsioonide 10 elemendist 2-kaupa:

$$V_{10}^2 = 10 \cdot 9 = 90.$$

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 vahetuste moodustamiseks on nii palju, kui palju saab moodustada kombinatsioonide 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 suhteline sagedus?

Iga detaili kontrollimine on katse. Praakdetaili esinemine on meid huvitav sündmus. Vaadeldav sündmus toimus 100 katsel 5 korda. Praagi esinemise suhteline sagedus

vaadeldud detailide partiis $w = 0,05 = 5\%$.

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 suhteline sagedus on 0,03. Vaadeldavas partiis praakdetailide oodatav arv $m = w \cdot n = 0,03 \cdot 2000 = 60$.

5. Tõenäosuse omadusi

Klassikalise tõenäosuse definitsioonist järeldub rida omadusi, mis on ülekantavad ka geomeetrilisele ja statistilisele tõenäosusele.

1. Võimatu sündmuse tõenäosus on null.

Tõepoolest, kui sündmus on võimatu, siis soodsate juhusete arv $m = 0$ ning $P = 0$.

2. Kindla sündmuse tõenäosus on üks.

Kindla sündmuse puhul $m = n$ ning $P = n : n = 1$.

3. Juhusliku sündmuse tõenäosus asub 0 ja 1 vahel.

Juhusliku sündmuse korral üldiselt $m < n$ ning seega $m : n < 1$.

Kokkuvõtte. Mistahes sündmuse tõenäosus P rahuldab tingimust

$$0 \leq P \leq 1.$$

Juhime tähelepanu järgmisele asjaolule. Klassikalise tõenäosuse definitsiooni järgi on sündmus võimatu, kui selle tõenäosus on null, ja kindel, kui tõenäosus on üks. Geomeetrilise ja statistilise tõenäosuse korral ei tarvitse see enam nii olla. Näiteks tasapinna ühe punkti tabamine on võimalik, kuigi tabamise tõenäosus on null (punkti pindala on null). Katseseerias võib sündmus toimuda igal n katsel ning seega $w = 1$, kuid pole välistatud sündmuse mittetoimumine $n + 1$ katsel - sündmust ei või veel pidada kindlaks. Analoogiliselt sündmuse mittetoimumine n katsel ei

välisest võimalusest selle toimumiseks veelgi pikemas katseseria. Kui sündmuse tõenäosus on küllalt lähedane nullile või null, siis öeldakse, et sündmus on praktiliselt võimatu - see toimub väga harva. Ühele lähedase tõenäosusega sündmusi loetakse praktiliselt kindlateks - need toimuvad peaaegu alati.

6. Sündmuste summa ja korrutis

Kahe sündmuse abil saab defineerida ka mõningaid uusi sündmusi.

Kahe sündmuse A ja B summaks nimetame sündmust, mille toimumine seisneb A või B (või mõlema) toimumises.

Sündmuste A ja B summa märkimiseks kasutame sümbolit $A \cup B$.

Analoogiliselt võib rääkida ka kolme või enama sündmuse summast.

Näide 1. Teatud mängu alustamise tingimuseks on täringu viskel kas 1 (sündmus A) või 6 silma (sündmus B) saamine. Sündmus „mängu alustamine“ (sündmus C) on seega sündmuste A ja B summa $C = A \cup B$.

Näide 2. Toodet loetakse standardseks (sündmus S), kui see kuulub kas kõrgemasse (K) või esimesse (E) sorti. Seega $S = K \cup E$.

Kahe sündmuse A ja B korrutiseks $A \cap B$ (või AB) nimetatakse sündmust, mille toimumine seisneb nii A kui B toimumises.

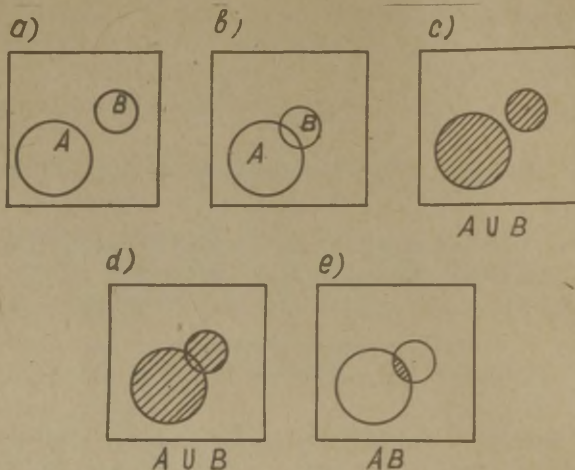
Näide 1. Järgnevale kursusele saamiseks (sündmus K) on vajalik nii arvestuste (A) kui eksamite (E) edukas sooritamine. Sündmust K võib vaadelda sündmuste A ja E korrutisena $S = A \cap E$ ehk $S = AE$.

Näide 2. Preemiat makstakse (sündmus P) töötajale, kes ei riku töödistsipliini (D) ning annab kogu plaani ulatuses kvaliteetset toodangut (K). Seega $P = D \cap K$.

Sündmuste summat ja korrutist (vastavalt $A \cup B$ või $A \cap B$) nimetatakse liitsündmusteks, mis koosnevad osasünd-

mustest A ja B.

Sündmuste summa ja korrutise mõistet on hea illustreerida graafiliselt (joonis 1).



Joonis 1.

Ristküliku osa A mistahes punkti tabamine juhuslikul torkel olgu sündmus A, osa B tabamine sündmus B. Sündmused A ja B on skeemil a) teineteist välistavad, b) mittevälistavad. Skeemil c) on antud juhule a) vastavate sündmuste summa $A \cup B$. Sündmus $A \cap B$ on aga võimatu, sest samaaegselt ei saa torkega tabada piirkondi A ja B. Skeemidel d) ja e) on toodud juhule b) vastavate sündmuste summa ja korrutis. Siit on näha, et sündmuste korrutise all mõistetakse sündmuste ühisosa.

Järgmistes punktides käsitleme seda, kuidas sündmuste A ja B teadaolevate tõenäosuste järgi leida nende sündmuste summa ja korrutise tõenäosusi.

7. Tõenäosuste liitmislause

On teada sündmuste A ja B tõenäosused $P(A)$, $P(B)$.

Leida liitsündmuse $A \cup B$ (kas A või B) tõenäosus.

Tõenäosuste liitmislause tuletame joonise 1 abil. Kogu risküliku punktid moodustavad elementaarsündmuste süsteemi ja risküliku pindala olgu 1. Siis on sündmuste A ja B tõenäosused vastavalt geomeetrilisele tõenäosusele võrdsed osade A ja B pindaladega (näit. $P(A) = s_A : 1 = s_A$). Peab eristama teineteist välistavate ja mittevälistavate sündmuste juhte.

Välistavate sündmuste juhul (skeem 1c) on sündmuseks $A \cup B$ mõlemad viirutatud ringid ning $A \cup B$ tõenäosuseks viirutatud pindalade summa. Seega

$$P(A \cup B) = P(A) + P(B)$$

- teineteist välistavate sündmuste summa tõenäosus võrdub liidetavate tõenäosuste summaga.

Mittevälistavate sündmuste juhul (skeem 1d) on A ja B summa tõenäosus jälle võrdne viirutatud osa pindalaga. Selle saamiseks tuleb mõlema ringi pindalade summast lahutada ringide ühisosa pindala. Lahutatav on aga parajasti sündmuste A ja B korrutise (koosinemise) tõenäosus. Seega

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

- teineteist mittevälistavate sündmuste summa tõenäosus võrdub liidetavate tõenäosuste summaga, millest on lahutatud nende sündmuste koosinemise (korrutise) tõenäosus.

Kolme sündmuse summa korral esitub tõenäosuste liitmislause järgmisel kujul:

välistavate osasündmuste juhul

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) ;$$

mittevälitavate osasündmuste juhul

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

M ä r k u s. Eeskirja sündmuste korrutise tõenäosuse leidmiseks saame järgmises punktis.

Liitmislauset kasutame siis, kui osasündmuste kohta saab küsida „kas... või...“.

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

$$P(B) = P(A_1 \cup A_2 \dots \cup A_n) = 1.$$

Teiselt poolt on täielikku sündmuste süsteemi moodustavad sündmused üksteist välistavad ja nende summa tõenäosuse leidmiseks võib rakendada liitmislauset. Siis

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) = 1$$

- täielikku sündmuste süsteemi moodustavate sündmuste tõenäosuste summa võrdub ühega.

J ä r e l d u s. Sündmuse ja selle vastandsündmuse tõenäosuste summa on üks (seleta, miks!):

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

Näiteid liitmislausel rakendamise kohta on punktis 10.

8. Tõenäosuste korrutamislause

Leiame eeskirja sündmuste korrutise $A \cap B$ tõenäosuse leidmiseks A ja B tõenäosuste kaudu. Olgu n võimaliku katse hulgas m_A juhtu, millega kaasneb sündmuse A toimumine, ja m_B juhtu, millega kaasneb B. Seejuures osal juhtudest võivad sündmused A ja B toimuda koos. Olgu sündmuse A jaoks soodsate juhtude m_A hulgas sündmuse B jaoks soodsate juhtude arv m_{AB} .

Jagatis $m_{AB} : m_A$ määrab sündmuse B tõenäosuse täiendaval tingimusel, et sündmus A on juba toimunud (juhtude üldarvuks on nende juhtude arv, mille korral sündmus A toimub).

Sündmuse B tõenäosust, mis on leitud eeldusel, et sündmus A toimus (toimub), nimetatakse sündmuse B tinglikuks tõenäosuseks tingimusel A $P(B/A)$. Sündmus, mille eelnevat toimumist eeldatakse, märgitakse kaldjoone taha. Loenduvate juhtude arvu abil saab tinglikku tõenäosust avaldada järgmiselt:

$$P(B/A) = \frac{m_{AB}}{m_A} .$$

Teisendame võrduse paremat poolt järgmiselt:

$$\frac{m_{AB}}{m_A} = \frac{m_{AB} : n}{m_A : n} .$$

Teiselt poolt $m_{AB} : n = P(A \cap B)$ ja $m_A : n = P(A)$. Seega

$$\frac{\text{korvatis}}{P(B/A)} = \frac{P(A \cap B)}{P(A)}$$

ehk

$$P(A \cap B) = P(A) \cdot P(B/A) .$$

Tulemus esitab tõenäosuste korrutamislause: kahe sündmuse

korrutise tõenäosus võrdub ühe sündmuse tõenäosuse (ükskõik kumma) ja teise sündmuse tingliku tõenäosuse korrutisega.

Kui kehtib seos

$$p(B/A) = P(A),$$

s.t. et ühe sündmuse tõenäosus ei sõltu teise sündmuse toimumisest või mittetoimumisest, siis nimetatakse sündmusi A ja B teineteisest sõltumatuteks. Vastasel korral öeldakse, et sündmused on sõltuvad.

N ä i d e. Urnis on 10 sinist ja 16 rohelist kuuli. Võtame urnist ühe kuuli, registreerime selle värvi ja pane-
me kuuli urni tagasi. Teisel katsel sinise kuuli saamine (sündmus S) ei olene esimesel katsel saadud kuulist (kas toimus S või R). Esimese katse tulemus ei mõjuta ka teisel katsel sinise kuuli saamise tõenäosust, $P(S) = \frac{10}{26}$. Toimi-
me nüüd aga nii, et esimesel katsel võetud kuuli urni taga-
si 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 si-
nise kuuli saamise tõenäosus on $\frac{9}{25}$, kui esimesel katsel saime rohelise kuuli, siis - $\frac{10}{25}$.

Teise katseskeemi puhul on sündmuse S tõenäosus tei-
sel katsel tinglik tõenäosus vastavalt tingimustel S või R:

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

Esimese katseskeemi puhul seevastu

$$P(S/S) = P(S), \quad P(S/R) = P(S) = \frac{10}{26}.$$

Tõenäosuste korrutamislause sõltumatute sündmuste kor-
ral - kahe sõltumatu sündmuse korrutise tõenäosus võrdub osasündmuste tõenäosuste korrutisega:

$$P(A \cap B) = P(A) \cdot P(B).$$

M ä r k u s. Tulemus on ülekantav ka kolme ja enama sündmuse korrrutise juhule. Näiteks kolme sõltuva osasündmuse juhul

$$P(ABC) = P(A) \cdot P(B/A) \cdot P(C/AB) ;$$

sõltumatute osasündmuste juhul

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

Juhime veel kord tähelepanu järgnevale. Liitmislauset kasutame sellise liitsündmuse puhul, mille kohta saab ütelda „kas see või teine“, korrrutamislause t aga „nii see kui teine“ puhul.

Näited on toodud punktis 10.

9. Tä i s t õ e n ä o s u s e v a l e m .

B a y e s i v a l e m

Alustame näitest. Kauplusse saabus 500 komplekti õmb-lustooteid kolmest vabrikust: 100 komplekti vabrikust K , 150 vabrikust L ja 250 vabrikust M. Vabriku K toodangust kuulub keskmiselt 75 % I sorti. Vabrikute L ja M jaoks on see näitaja vastavalt 90 % ja 80 %. Leida tõenäosus, et huupi võetud komplekt on esimest sorti.

Tähistame sündmuse, et komplekt on esimest sorti, tähega A. Sündmust, et komplekt pärineb vabrikust K (pole tähtis mis sorti) tähistame tähega K ja ülejäänud vastavalt L, M. I sordi komplekti võib saada kas vabriku K, L või M toodangu hulgast. Tõenäosus selleks, et komplekt on pärit teatud vabrikust, on

$$P(K) = \frac{100}{500} , \quad P(L) = \frac{150}{500} , \quad P(M) = \frac{250}{500} .$$

Tõenäosus I sordi komplekti saamiseks vabrikus K valmistatud toodangu hulgast (seega tinglik tõenäosus) on

$$P(A/K) = \frac{75}{100} = 0,75 .$$

Vabrikute L ja M jaoks on tinglikud tõenäosused $P(A/L) = 0,9$, $P(A/M) = 0,8$.

Nüüd leiame, kui tõenäone on, et juhuslik komplekt pärineb vabrikust K ja kuulub I sorti (analoogiliselt L ja M jaoks). See tähendab sündmuse KA tõenäosuse leidmist. Tõenäosuste korrutamislausest saame

$$P(K A) = P(K) \cdot P(A/K) = 0,2 \cdot 0,75 ;$$

$$P(L A) = 0,3 \cdot 0,9 ; \quad P(M A) = 0,5 \cdot 0,8 .$$

Sündmus A on vaadeldav 3 üksteist välistava sündmuse summana: I sordi komplekt kas K või L või M toodangust. Lühemalt: A saab toimuda koos ühega sündmustest K, L või M. Selle tõenäosuse leidmiseks rakendame liitmislauset, sest

$$A = KA \cup LA \cup MA$$

ning

$$\begin{aligned} P(A) &= P(KA) + P(LA) + P(MA) = \\ &= 0,2 \cdot 0,75 + 0,3 \cdot 0,9 + 0,5 \cdot 0,8 = 0,82. \end{aligned}$$

Esimest sorti komplekti saamise tõenäosus on 0,82 (82 %).

Vaatleme ülesannet üldises seades. Saagu sündmus A kaasneda ühega sündmustest H_1, H_2, \dots, H_n , mis moodustavad täieliku sündmuste süsteemi. Sündmuste H_i tõenäosused olgu teada: $P(H_i)$ ($i = 1, 2, \dots, n$). Sündmuse A sündmusega H_i koostoimumise tõenäosus ehk tinglik tõenäosus tingimusel H_i olgu $P(A/H_i)$. Sündmuse A tõenäosuse $P(A)$ leidmiseks kehtib täistõenäosuse valem

$$\begin{aligned} P(A) &= P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + \dots + \\ &+ P(H_n) \cdot P(A/H_n) \end{aligned}$$

ehk lühemalt

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A/H_i) .$$

Valemi tõestuskäik on analoogiline lahendatud näites kasutatud mõttekäigule.

Esineb olukordi, kus sündmuse A toimumise põhjuste kohta tuleb teha mitmesuguseid oletusi ehk nn. hüpoteese. Sündmus A saab kaasneda mõne hüpoteesi realiseerumisega. Hüpoteese võib vaadelda juhuslike sündmustena H_1, H_2, \dots, H_n , millel on teatud tõenäosused $P(H_i)$. Sündmuse A tõenäosus eeldusel, et mingi hüpotees H_i on realiseerunud, on tinglik tõenäosus $P(A/H_i)$. Täistõenäosuse valemi abil leitakse A tõenäosus, arvestades sündmuste H_i toimumise võimalusi. Sooritame katse ja selle käigus toimub sündmus A . See sunnib tavaliselt ümber hindama varem püstitatud hüpoteese ja nende tõenäosusi. Tuleb leida teatud hüpoteesi H_i tõenäosus pärast seda, kui sündmus A on toimunud. See oleks hüpoteesi H_i tinglik tõenäosus tingimusel $A - P(H_i/A)$.

Tingliku tõenäosuse mõiste järgi (vt. 8.p.)

$$P(H_i/A) = \frac{P(H_i A)}{P(A)} ,$$

kus $P(A)$ leitakse täistõenäosuse valemist ning sündmuse $H_i A$ tõenäosus tõenäosuste korrutamislausest. $P(H_i A) = P(H_i) \cdot P(A/H_i)$. Mõlemaid seoseid arvestades saame Bayesi valemi

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{\sum_1 P(H_i) \cdot P(A/H_i)} .$$

Võrduse paremal poolel esinevad tõenäosused on kõik määratud enne katset. Vasak pool tähendab hüpoteesi tõenäosust, võttes arvesse tegelikku katsetulemust.

10. Ü l e s a n d e i d

Alustame ülesannetega liitmis- ja korrutamislauseste kasutamise kohta.

1. Raha ja asjade loterii iga 10000 pileti kohta loositakse 150 rahalist ja 50 esemelist võitu. Kui tõenäone on ühe piletiga võitmine?

Sündmus „võitmine“ (V) on kahe teineteist välistava sündmuse „rahaline võit“ (R) ja „esemeline võit“ (E) summa.

$$P(V) = P(R \cup E) = P(R) + P(E) = \frac{150}{10000} + \frac{50}{10000} = 0,02 .$$

Teine lahendusviis. Sündmusele V on soodsaid juhte $m = 150 + 50 = 200$ ja $P(V) = 200 : 10000 = 0,02$.

2. 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

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

3. Pikaajaliste vaatluste tulemusena selgus, et remonditöökojas on treipingi 20 peatuse põhjusteks keskmiselt 10 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

$$P(A) + P(B) + P(C) + P(D) = 1, \text{ kust}$$

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

Lahendage sama ülesanne liitmislauset kasutamata.

4. 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.

5. Tõenäosus, et nädala jooksul tuleb punaka tooniga valgusti asendada uuega, on 0,7 ja sinaka tooniga valgusti puhul - 0,5. Kui tõenäone on, et asendada tuleb mõlemad valgustid?

Tähistagu A sündmust „asendada punase tooniga valgusti” ja B - „asendada sinine valgusti”. Meid huvitab sündmus „nii A kui B” ehk $C = AB$, kus osasündmused on teineteisest sõltumatud:

$$P(AB) = P(A) \cdot P(B) = 0,7 \cdot 0,5 = 0,35.$$

6. 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 $P(AB) = P(A) \cdot P(B/A) = 0,95 \cdot 0,86 = 0,817$.

7. Tsehhis töötab 3 tööpinki. I tööpingi häireteta töötamise tõenäosus on 0,9, II tööpingil 0,8 ja III töö-

pingil on 0,85. Kui tõenäone on, et 1) nii I kui II tööpinki töötavad vaadeldava tunni jooksul häireteta; 2) kõik kolm tööpinki töötavad häireteta?

Ühe tööpinki 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) $P(ABC) = 0,9 \cdot 0,8 \cdot 0,85 = 0,612$.

8. Teate toimetamiseks ühest punktist teise saadeti teineteisest sõltumatult teele 2 käskjalga. Tõenäosus, et esimene käskjalg jõuab sihtpunkti, on 0,9 ja teisel 0,8. Kui tõenäone on, et teade jõuab sihtpunkti?

Tähistame sündmused „I käskjalg jõuab sihtpunkti“ ja „II käskjalg jõuab sihtpunkti“ vastavalt S_1 ja S_2 . Sündmused S_1 ja S_2 on teineteist mittevälistavad. Tähega S tähistame sündmuse „teade jõuab sihtkohta“. Siis $S = S_1 \cup S_2$ ja

$$P(S) = P(S_1) + P(S_2) - P(S_1 S_2).$$

Sündmused S_1 ja S_2 on teineteisest sõltumatud ning vastavalt sõltumatute sündmuste tõenäosuste korrutamislausele

$$P(S_1 S_2) = P(S_1) \cdot P(S_2).$$

Seega

$$P(S) = 0,9 + 0,8 - 0,9 \cdot 0,8 = 0,98,$$

mis näitab, et teate jõudmine sihtkohta on praktiliselt garanteeritud.

9. Umbrohutõrjet tehes suudab üks lennuk preparaati pritsides umbrohu hävitada tõenäosusega 0,7, teine lennuk - 0,8. Kui suur on umbrohu hävitamise tõenäosus mõlema lennuki kasutamisel?

Umbrohu hävitamine I lennuki abil olgu sündmus H_1 , II lennuki abil - H_2 , umbrohu hävitamine üldse - H .

Esimene lahendusviis. Sündmus H on mittevälstavate sündmuste H_1 ja H_2 summa, kusjuures H_1 ja H_2 on teineteisest sõltumatud. Siis

$$P(H) = P(H_1 \cup H_2) = P(H_1) + P(H_2) - P(H_1) \cdot P(H_2) = 0,94.$$

Teine lahendusviis. Sündmuse H vastandsündmuseks on \bar{H} - kumbki lennuk ei suuda umbrohtu hävitada. Sündmus \bar{H} on omakorda sõltumatute sündmuste - \bar{H}_1 ja \bar{H}_2 - korrutis ning $P(\bar{H}_1 \bar{H}_2) = P(\bar{H}_1) \cdot P(\bar{H}_2)$. Vastandsündmuste jaoks kehtib seos $P(H) + P(\bar{H}) = 1$. Siit saame leida $P(H) = 1 - P(\bar{H})$. Kuid $P(\bar{H}_1) = 0,3$ ning $P(\bar{H}_2) = 0,2$. Seega $P(H) = 1 - 0,3 \cdot 0,2 = 0,94$.

10. Tulistatakse 3 lasku. Märgi tabamise tõenäosused on I lasuga 0,4, II - 0,5, III - 0,7. Kui tõenäone on, et märki tabab vähemalt üks lask?

Tähistame sündmused järgmiselt: I lask tabab - T_1 , II - T_2 , III - T_3 , märgi tabamine üldse (ühe või enama lasuga) - T . Märgi tabamine vähemalt üks kord on vastandsündmuseks märgi mittetabamisele kõigi laskudega. Märgi mittetabamise tõenäosus mingi lasuga on $P(\bar{T}_1) = 1 - 0,4 = 0,6$, $P(\bar{T}_2) = 0,5$, $P(\bar{T}_3) = 0,3$. Märgi mittetabamine kõigi laskudega $\bar{T} = \bar{T}_1 \bar{T}_2 \bar{T}_3$ ning $P(\bar{T}_1 \bar{T}_2 \bar{T}_3) = 0,6 \cdot 0,5 \cdot 0,3 = 0,09$. Nüüd leiame seosest

$$P(T) + P(\bar{T}) = 1;$$

$$P(T) = 1 - P(\bar{T}) = 1 - 0,09 = 0,91.$$

Esitatud mõttekäiku on sobiv kasutada ülesannetes, kus tuleb leida mittevälstavate sündmuste puhul vähemalt ühe (ükskõik, millise) sündmuse toimumise tõenäosus.

Järgnevalt vaatleme täistõenäosuse valemi rakendamist.

11. Aparaaadi monteerimisel kasutatakse neljas tsehhis valmistatud detaile. Praagi tõenäosus tsehhide kaupa on 0,04; 0,03; 0,06; 0,02. Tsehhidest saabus detaile järgmistes kogustes: 30, 20, 50 ja 25 tükki. Kui tõenäone on, et juhuslikult võttes saadakse praakdetail?

Praakdetaili saamine olgu sündmus A; detaili kuulmine tsehhi i toodangusse - B_i ($i = 1, 2, 3, 4$). Siis

$$P(B_1) = \frac{30}{30 + 20 + 50 + 25} = \frac{30}{125}, \quad P(B_2) = \frac{20}{125};$$

$$P(B_3) = \frac{50}{125}, \quad P(B_4) = \frac{25}{125}.$$

Praagi tinglikud tõenäosused eeldusel, et detail pärineb teatud tsehhist, on

$$P(A/B_1) = 0,04; \quad P(A/B_2) = 0,03; \quad P(A/B_3) = 0,06; \quad P(A/B_4) = 0,02.$$

Täistõenäosuse valemi järgi praagi saamise tõenäosus montaažitsehhis on

$$P(A) = \frac{30}{125} \cdot 0,04 + \frac{20}{125} \cdot 0,03 + \frac{50}{125} \cdot 0,06 + \frac{25}{125} \cdot 0,02 = 0,066.$$

12. Viiest urnist 2 sisaldavad kumbki 4 valget ja 3 musta kuuli, üks - 3 valget ja 4 musta ning kaks urni kumbki 5 valget ja 2 musta kuuli. Ühest urnist võetakse üks kuul. Kui tõenäone on, et kuul osutub valgeks?

Olgu kuuli saamine esimesest grupist sündmus B_1 , teisest - B_2 , kolmandast - B_3 . Valge kuuli saamine olgu sündmus V. Sündmuste B_1, B_2, B_3 tõenäosused on

$$P(B_1) = \frac{2}{5}, \quad P(B_2) = \frac{1}{5}, \quad P(B_3) = \frac{2}{5}.$$

Sündmuse V tinglikud tõenäosused on

$$P(V/B_1) = \frac{4}{7}, \quad P(V/B_2) = \frac{3}{7}, \quad P(V/B_3) = \frac{5}{7}.$$

Valge kuuli saamise tõenäosus

$$P(V) = \frac{2}{5} \cdot \frac{4}{7} + \frac{1}{5} \cdot \frac{3}{7} + \frac{2}{5} \cdot \frac{5}{7} = \frac{21}{35} = \frac{3}{5}.$$

Lõpuks käsitleme Bayesi valemi rakendamist.

13. Lähteandmed on 12 näites. Võetud kuul osutus valgeks. Kui tõenäone on, et ta pärineb esimesest urnide gruppist?

Praegu huvitab meid B_1 tõenäosus tingimusel, et V on toimunud. Bayesi valemi kohaselt

$$P(B_1/V) = \frac{P(B_1) \cdot P(V/B_1)}{P(V)}$$

Vajalikud tõenäosused on leitud eelmises näites ja

$$P(B_1/V) = \frac{\frac{2}{5} \cdot \frac{4}{7}}{\frac{3}{5}} = \frac{8 \cdot 5}{5 \cdot 7 \cdot 3} = \frac{8}{21}.$$

14. Aparaaate monteeritakse kõrgema või I sordi detailidest. Keskmiselt 40 % aparaatidest monteeritakse kõrgema sordi detailidest. Kõrgema sordi detailidest aparaadil on tõenäosus tõrgeteta töötamiseks aasta jooksul 0,95, esimese sordi detailidest monteeritud aparaatidel - 0,7. Aparaat läbis katsetused tõrgeteta. Kui tõenäone on, et aparaat koosnes kõrgema sordi detailidest?

Aparaadi tõrgeteta töötamine olgu sündmus A ; aparaat on monteeritud kõrgema sordi detailidest - sündmus H_1 , esimese sordi detailidest - sündmus H_2 . Tunneme huvi, kui oodatav on sündmus H_1 , kui osutus, et toimus sündmus A . Hüpoteeside H_1 ja H_2 tõenäosused on antud: $P(H_1) = 0,4$, $P(H_2) = 0,6$. Sündmuse A tinglikud tõenäosused on samuti antud: $P(A/H_1) = 0,95$, $P(A/H_2) = 0,7$. Bayesi valemis esineva murru nimetajaks on A tõenäosus enne katset:

$$P(A) = 0,4 \cdot 0,95 + 0,6 \cdot 0,7 = 0,8.$$

Otsitav tõenäosus

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{P(A)} = \frac{0,4 \cdot 0,95}{0,8} = 0,475.$$

15. Garaažis seisab 3 autot. Pole teada, kas nad on tangitud. Ühe auto kontrollimisel osutus see tangituks. Kui tõenäone on, et ka ülejäänud autod on tangitud?

Eelnevalt võib püstitada järgmised oletused (hüpoteesid): H_1 - 3 autot on tangitud; H_2 - 2 on tangitud, üks tühja paagiga; H_3 - 1 on tangitud, 2 tühjad. Hüpoteesidest ei ole alust ühtegi eelistada - kõigi tõenäosused loeme võrdseteks ehk $P(H_i) = 0,33$ ($i = 1, 2, 3$).

„Auto on tangitud“ olgu sündmus A. Sündmuse A tinglikud tõenäosused - hüpoteesidel H_i vastavalt

$$P(A/H_1) = 1; \quad P(A/H_2) = \frac{2}{3}; \quad P(A/H_3) = \frac{1}{3}.$$

Leida tuleb hüpoteesi H_1 tõenäosus pärast seda, kui selgus, et vähemalt 1 auto on tangitud:

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{P(A)} = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot 0} = \frac{1}{2}.$$

§ 2. JUHUSLIKE SUURUSTE JAOTUSSEADUSED

1. J u h u s l i k u d s u u r u s e d

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

1. n ä i d e. 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 - mingi silmade arvu saamine sõltub juhusest.

2. n ä i d e. Ühest viljapeast võetud terade kaal on üldiselt muutuv. Üksikute terade erinev kaal on tingitud rea faktorite koosmõjust, mida me ei oska määrata. Räägime, et tera kaal sõltub juhusest.

3. n ä i d e. 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, teinekord 48 tütarlast. 100 vastsündinu kohta tulev tütarlaste arv on juhusest sõltuv ehk juhuslik suurus, mis saab omandada ühe võimalikest väärtustest.

4. n ä i d e. 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 (muutuajat) nimetatakse juhuslikuks, kui see 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 kui juhusliku sündmuse esinemise võimalikkust saab põhimõtteliselt iseloomustada vastava väärtuse esinemise tõenäosuse abil. Niisiis, täieliku ülevaate saamiseks juhusliku suuruse väärtuste esinemisest peab määrama suuruse kõik võimalikud väärtused ja nende väärtuste tõenäosused.

Seadust, mis määrab seose juhusliku suuruse kõigi võimalike väärtuste ja nende tõenäosuste vahel, nimetatakse juhusliku suuruse tõenäosuste jaotusseaduseks ehk juhusliku suuruse jaotusseaduseks.

Juhuslikke suurusi liigitatakse diskreetseteks ja pidevateks.

Juhuslikku suurust nimetatakse diskreetseks, kui selle võimalike väärtuste hulk on loenduv. Diskreetse suuruse võimalikud väärtused erinevad üksteisest mingi lõpliku arvu võrra.

Käesoleva punkti 1., 3., ja 4. näites on tegemist diskreetsete juhuslike suurustega.

Juhuslikku suurust nimetatakse pidevaks, kui see võib omandada mistahes väärtusi teatud lõplikust või lõpmatust reaalarvude vahemikust (vt. ka p. 5).

Näites 2 võib terade kaal omandada mistahes reaalarvulise väärtuse kõige kergema ja kõige raskema kaalu vahel. Muidugi kaalude piiratud täpsuse tõttu ei ole võimalik kuitahes lähedasi kaale eristada.

5. n ä i d e. Inimese jalalaba pikkust saab vaadelda pideva juhusliku suurusena. Selle võimalike väärtuste hulk ei ole loenduv ega saa 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 saab anda ka iga jala numbri esinemise tõenäosuse. Selle põhjal saavad jalatsivabrikud oma toodangut planeerida jalanõude suuruse järgi.

Õmblusvabrikud peavad valmisriiete õmblemisel arvestama inimeste pikkuse ja rinnaümberrõõdu 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, \dots). Väärtuste tõenäosusi tähistame tähega p ; väärtuse x_1 tõenäosus on p_1 ($i = 1, 2, 3 \dots$).

2. J u h u s l i k u s u u r u s e j a o t u s - t a b e l

Üheks lihtsamaks juhusliku suuruse jaotuse esitamise viisiks ehk jaotusseaduseks on jaotustabel.

1. n ä i d e. 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. n ä i d e. Viskame kaht täringut korraga. Saadav silmade arv on juhuslik suurus, mis võib omandada naturaalarvulisi väärtusi 2 kuni 12. Leiame üksikute väärtuste tõenäosused p_i ($i = 2, 3, \dots, 12$). Kahe täringu viskel on kõigi võrdvõimalike juhtude arv 36 (ühe täringu kindla tahuga saab teisel täringul tulla 6 erinevat tahku; erinevate tahkude kombinatsioonid saab olla 6 · 6).

Soodsate juhtude arvu määramise 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 silmale - 4; 6 silmale - 5; 7 silmale - 6; 8 silmale
 - 5; 9 silmale - 4; 10 silmale - 3; 11 silmale - 2; 12
 silmale - 1.

Vastavate väärtuste tõenäosused saame soodsate juh-
 tude ja kõigi juhtude arvu jagamisel. Kui kanda tabeli üh-
 te ritta silmade võimalikud väärtused ja teise ritta tõe-
 näosused, saame tabeli:

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 jaotustabelis esita-
 takse suuruse võimalikud väärtused x_1 ja vastavad tõenäo-
 sused p_1 ($i = 1, 2, \dots, n$):

X	x_1	x_2	\dots	x_n
p	p_1	p_2	\dots	p_n

Pideva juhusliku suuruse jaotuse esitamiseks kasuta-
 takse ka jaotustabelit, kus ühes reas on võimalike väärtus-
 te muutumisvahemiku intervallid ja teises reas väärtuste
 vastavasse intervalli kuulumise tõenäosused. Kui võimalike
 väärtuste muutumisvahemik ehk väärtuste varu asub reaalar-
 vude x_0 ja x_n vahel, siis jaotatakse see vahemik kindlaks
 arvuks n intervalliks (intervallid võivad olla võrdse või
 ka mittevõrdse pikkusega) $x_0 - x_1, x_1 - x_2, \dots, x_{n-1} -$
 $- x_n$. Tähistame juhusliku suuruse väärtuse teatud inter-
 valli langemise tõenäosuse p_1 -ga. Siis saab pideva juhusli-
 ku suuruse intervallitud jaotustabeli

X	$x_0 - x_1$	$x_1 - x_2$	\dots	$x_{n-1} - x_n$
p	p_1	p_2	\dots	p_n

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

M ä r k u s. Kahe intervalli piiril asuv väärtus loetakse kõrgemasse intervalli kuuluvaks.

Juhusliku suuruse jaotustabelis on antud kõik võimalikud väärtused, millest katsel üks kindlasti realiseerub. Suuruse võimalikud väärtused moodustavad seega täieliku sündmuste süsteemi. Täielikku sündmuste süsteemi moodustavate sündmuste tõenäosuste summa on aga 1. Järelikult suuruse jaotustabeli korral peab alati kehtima võrdus

$$P_1 + P_2 + \dots + P_n = \sum_{i=1}^n P_i = 1.$$

3. V a r i a t s i o o n r i d a. S t a t i s t i - l i n e j a o t u s t a b e l

A. Statistika algmõisted.

Statistilise uurimise objektiks olevat nähtuste, esemete või indiviidide kogu nimetatakse 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 kogumi uurimine toimub, nimetatakse tunnuseks.

1. n ä i d e. Uurime Eesti NSV tööstusettevõtteid kvartaliplaani täitmise järgi. Tööstusettevõtted moodustavad kogumi, iga ettevõtte on kogumi element, plaani täitmise protsent on uuritavaks tunnuseks. Antud tunnust saab arvuliselt avaldada.

2. n ä i d e. Rajoonikeskuse keskkooli õpilased jaotame vanemate tööala järgi kolhoosnike, sovhoositöölise tööstusettevõtete töölise ja teenistujate lasteks. Tunnus-

seks on õpilaste vanemate tööala. Antud juhul me ei saa va-
litud tunnust arvuliselt kirjeldada.

Tunnust, mille väärtusteks on arvud, nimetatakse
kvantitatiivseks; tunnust, mida ei saa arvuliselt kirjel-
dada, kvalitatiivseks.

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.

Tunnuse X väärtuse x_1 saame katse või vaatluse teel.
Teeme N katset, mille tulemusena saame tunnuse väärtuste
jada $x_1, x_2, x_3, \dots, x_N$, kus võib olla ka võrdseid liik-
meid.

3. n ä i d e. 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 nimetatakse
statistiliseks reaks. Statistiline rida saadakse katseand-
mete registreerimisel.

Võib juhtuda, et statistilises reas esineb tunnuse
mõni väärtus korduvalt. Nii esineb 3. näite andmetel hinne
„5” statistilises reas 5 korda.

Arvu, mis näitab, mitu korda tunnuse mingi väärtus
esineb statistilises reas, nimetatakse selle väärtuse sage-
duseks. Sagedust märgime tähega f .

Otstarbekas on statistilises reas olevad väärtused
esitada kindlas järjestuses ja ära näidata üksikute väärtu-
tuste sagedused.

Tabelit, kus tunnuse väärtused on esitatud kas kasva-
vas või kahanevas järjekorras koos üksikväärtuste sagedus-
tega, nimetatakse variatsioonreaks. Kui erinevaid väärtu-
si on n tükki, saab variatsioonrea üldiselt kujul

X	x_1	x_2	\dots	x_n
f	f_1	f_2	\dots	f_n

4. n ä i d e. Belmise näite andmetel saame variatsioonreaks

Hinne	5	4	3
Sagedus	5	3	2

Katsete üldarvu tähistame tähega N ja nimetame kogumi mahuks:

$$N = f_1 + f_2 + \dots + f_n = \sum_{i=1}^n f_i .$$

Iga üksikväärtuse x_i sageduse f_i ja kogumi mahu N jagatist nimetatakse väärtuse suhteliseks sageduseks w_i :

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

Väärtuse x_i suhteline sagedus tähendab sedasama mis sündmuse suhteline sagedus. 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	\dots	x_n
P	p_1	p_2	\dots	p_n

Üksikväärtuse x_i tõenäosus p_i näitab vastava väärtuse esinemise võimalikkust katsel. Katsetel realiseeruvad suuruse X väärtused x_i teatud arv f_i korda. Katsetulemused võib esitada kas variatsioonreana või tabelina, kus ühes reas on suuruse väärtused x_i ja teises reas nende suhtelised sagedused w_i . Tabelit

X	x_1	x_2	\dots	x_n
w	w_1	w_2	\dots	w_n

nimetatakse 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_1 ja p_1 vahel ühe ja sama suuruse jaoks.

Statistilise jaotustabeli korral kehtib suhteliste sageduste jaoks omadus

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

Tõepoolest

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

5. n ä i d e. Koostada 4. näite andmetel statistiline jaotustabel.

Hinnete koguarv $N = 10$. Jaotustabeli saame kujul

Hinne	5	4	3
Relatiivne sagedus	0,5	0,3	0,2

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

6. n ä i d e. 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	110.			

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

Võtame plaani täitmise protsendi vahemiku algväärtu-

seks 75 ja lõppväärtuseks 175 ning määrame igasse intervalli kuuluvate ettevõtete arvu loendamise teel (statistika üldkursuses käsitletakse rühmitamise võtteid üksikasjalikumalt).

Plaani täitmise % (\bar{x})	75-85	85-95	95-105	105-115	115-125	125-135	135-145	145-155	155-165	165-175
Ettevõtete arv f	2	2	13	8	6	8	4	2	1	1
Suhteline sagedus w	$\frac{4}{47}$	$\frac{2}{47}$	$\frac{13}{47}$	$\frac{8}{47}$	$\frac{6}{47}$	$\frac{8}{47}$	$\frac{4}{47}$	$\frac{2}{47}$	$\frac{1}{47}$	$\frac{1}{47}$

Kogumi maht on 47. Statistilise jaotustabeli saame intervalli kuuluvate ettevõtete arvu 47-ga jagades. Tabeli 1. ja 3. rida koos vaadelduna on statistiline jaotustabel, 1. ja 2. rida - variatsioonrida.

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 suhteliste 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 tähtsamaks graafikuks juhusliku suuruse tõenäosuste jaotushulknurk ehk polügoon ja histogramm. Mitmesugused teised graafikud leiavad lähema käsitluse statistika üldteoorias.

Selgitame tõenäosuste jaotushulknurga ehk polügooni konstrueerimist diskreetse suuruse korral.

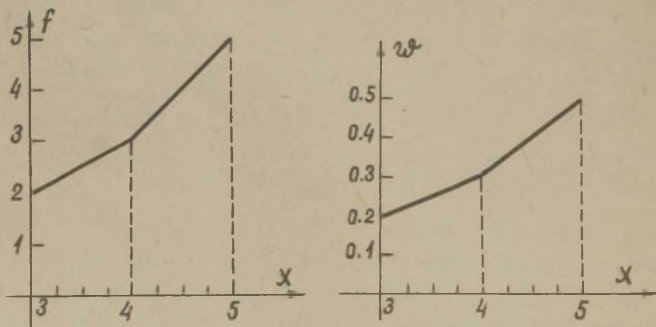
On antud juhusliku suuruse jaotustabel

X	-2	-1	0	1	2	3
p	0,1	0,2	0,25	0,2	0,2	0,05

Kanname koordinaatteljestikus abstsissiteljele valitud mastaabis suuruse väärtused ja ordinaatteljele väärtuste tõenäosused. Kokkukuuluvad suuruse väärtuse x_1 ja tõenäosuse p_1 paarid kujutuvad tasapinna punktidenä (x_1, p_1). Ühendame saadud punktid sirglõikudega ja saame murdjoone, mis kannabki polügooni nime.

Variatsioonrea või statistilise jaotustabeli korral kantakse ordinaatteljele vastavalt sagedused f või suhtelised sagedused w , saades sageduste või suhteliste sageduste polügoonid.

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



Joonis 2.

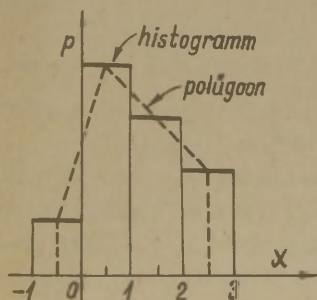
Polügooni abil saab kohe ülevaate sellest, milliste väärtuste esinemise tõenäosused üldse esineda võivate väärtuste hulgas on kõige suuremad või, empiiriliste andmete järgi koostatud polügooni korral, millised väärtused kõige sagedamini esinesid.

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 kantakse intervallid pikkusega Δx ja ehitatakse neile lõikudele ristkülikud, mille kõrgused on võrdelised vastavate tõenäosustega.

1. n ä i d e. Jaotustabelile

X	-1 - 0	0 - 1	1 - 2	2 - 3
p	0,1	0,4	0,3	0,2

vastav histogramm on joonisel 3.



Joonis 3.

histogrammilt võib üle minna polügoonile. Selleks ühendame lõikudega ristkülikute ülemiste aluste keskpunktid. Joonisel 3 on polügoon joonistatud punktiiriga.

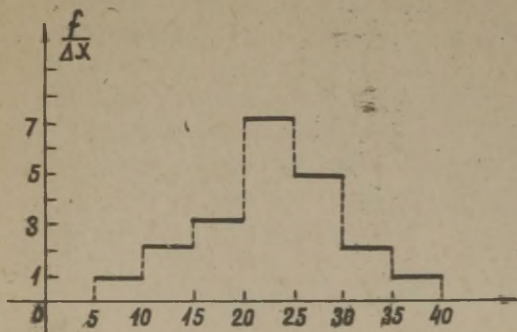
M ä r k u s. Kui tunnuse väärtuste muutumisvahemik pole jagatud võrdseteks intervallideks, siis on otstarbekas kanda ordinaatteljele nn. sageduse keskmine tihedus. Sageduse kesk-

2. n ä i d e. 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 sageduse keskmine

tihedus intervallide järjekorras on 0,8; 1,2; 3,2; 7,2; 4,8; 2; 0,8. Histogramm on joonisel 4.



Joonis 4.

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

5. Juhusliku suuruse jaotusfunktsioon

Diskreetse juhusliku suuruse iseloomustamiseks piisab tema jaotustabeli teadmisest. Pideva juhusliku suuruse korral on intervallitud jaotustabel ligikaudne jaotus. See annab ainult väärtuste antud pikkusega intervalli sattumise tõenäosuse, täpsemalt määramata antud pikkusest lühemasse intervalli sattumise või 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.

Tõenäosust selleks, et juhuslik suurus X omandab antud arvust x väiksema väärtuse, nimetame juhusliku suuruse jaotusfunktsiooniks: $F(x) = P(X < x)$.

Jaotusfunktsioon on reaalarvulise argumendi x funktsioon. Kui juhusliku suuruse jaotusfunktsioon on pidev, siis

nimetatakse juhuslikku suurust pidevaks. Sisuliselt tähendab see, et juhuslikul suurusel on lõpmata palju erinevaid võimalikke väärtusi (vt. p. 1).

Geomeetriliselt võib 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 pikka horisontaalset renni. Kuul võib jääda (juhuslikest teguritest tingituna) peatuma kas ühes või teises punktis. Funktsioon $F(x)$ näitab, kui tõenäone on, et kuul peatub mingis punktis enne fikseeritud kangust x .

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 puhul leida, kui tõenäone on, et juhusliku suuruse väärtused on tõkkest x väiksemad. $F(x)$ kasutatakse nii pideva kui ka diskreetse suuruse korral.

Vaatleme jaotusfunktsiooni omadusi.

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

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

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

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

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

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

3. Jaotusfunktsioon on mittekahanev, s.t.

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

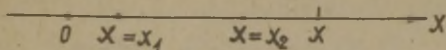
4. Kui juhusliku suuruse kõik võimalikud väärtused asuvad vahemikus (a, b) , siis

$$F(x) = 0, \text{ kui } x \leq a;$$

$$F(x) = 1, \text{ kui } x \geq b;$$

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

Nende omaduste õigsuses veendume geomeetrilise tõlgenduse abil. Juhuslikku suurust X vaatleme arvtelje liikuva juhusliku punktina, mis võib katsel sattuda ühte või teise



Joonis 5.

kohta (joonis 5).

$F(x)$ määrab siis liikuva punkti X fikseeritud punktist x vasakule jäämise tõenäosuse. Kui

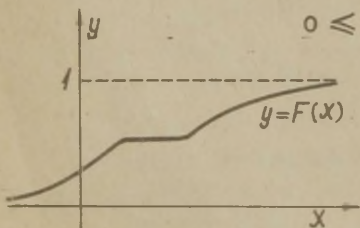
$x \rightarrow -\infty$, siis on sündmuse $X < -\infty$ toimumine võimatu ja seega $F(-\infty) = 0$.

Kui $x \rightarrow +\infty$, siis $X \rightarrow +\infty$ 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 pikeneb arvtelje see osa, kus punkt X võib asuda. Sellega on üldiselt tõenäosus sündmuse $(X < x_2)$ toimumiseks suurem kui sündmuse $(X < x_1)$ toimumiseks. See ütlebki, et $F(x)$ on mittekahanev.

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

K o k k u v õ t e. Jaotusfunktsioon $F(x)$ on mittekahanev ja selle väärtused asuvad 0 ning 1 vahel:



Joonis 6.

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

Jaotusfunktsiooni omadustest järeldub, et selle graafik asub sirgete $y = 0$ ja $y = 1$ vahel ning üldiselt kulgeb tõusvalt (joonis 6). Jaotusfunktsiooni graafikut

nimetatakse ka kumulaadiks.

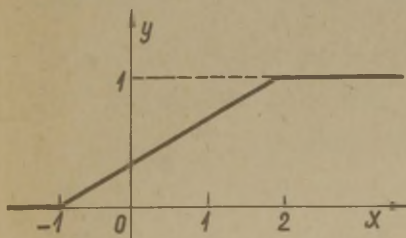
1. n ä i d e. Jaotusfunktsioon on määratud järgmiselt:

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

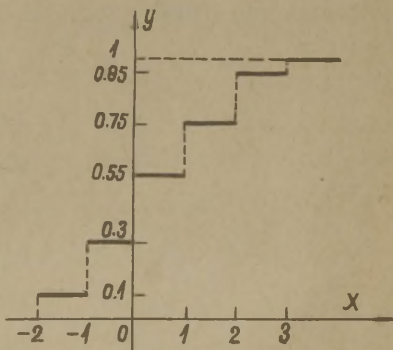
Konstrueerida kumulaat.

Kui $x \leq -1$, ühtib graafik abstsissiteljega, kui $-1 \leq x \leq 2$, on graafikuks sirge tõusuga $\frac{1}{3}$ ja $x \geq 2$ korral saame abstsissiteljega paralleelse sirge $y = 1$ (joonis 7).

Diskreetse juhusliku suuruse kumulaadiks on treppjoon (näit. joonis 8).



Joonis 7.



Joonis 8.

2. n ä i d e. 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 $X < 0$ koosneb üksteist välistavatest osasündmustest: $X = -2$ ja $X = -1$.

Tõenäosuste liitmislause põhjal

$$P(X < 0) = P(0) = 0,1 + 0,2 = 0,3.$$

Analoogiliselt

$$P(1) = 0,1 + 0,2 + 0,25 = 0,55;$$

$$P(2) = 0,1 + 0,2 + 0,25 + 0,2 = 0,75;$$

$$P(3) = 0,75 + 0,2 = 0,95;$$

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

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

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

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

Funktsiooni graafikul on hüpped suuruse võimalikele väärtustele vastavates punktides (joonis 8).

6. Juhusliku suuruse antud vahemikku sattumise tõenäosuse leidmine jaotusfunktsiooni kaudu

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 sattumine antud vahemikku“ ja selle sündmuse tõenäosust tähistame

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

Sündmus $X < x_2$ on kahe teineteist välistava sündmuse $X < x_1$ ja $x_1 < X < x_2$ summa ($X < x_2$) = ($X < x_1$) \cup ($x_1 < X < x_2$). 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.

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \Delta F;$$

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

Käsitleme nüüd pidevaid juhuslikke suursi. Tähistame $x_2 = x_1 + \Delta x$, kus Δx on argumendi juurdekasv ja laseme $\Delta x \rightarrow 0$.

Sündmus $x_1 < X < x_1 + \Delta x$ taandub sündmuseks $X = x_1$, kuid siis ka

$$\begin{aligned} P(X = x_1) &= \lim_{\Delta x \rightarrow 0} P(x_1 < X < x_1 + \Delta x) = \\ &= \lim_{\Delta x \rightarrow 0} [F(x_1 + \Delta x) - F(x_1)] = 0, \end{aligned}$$

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

Seega pideva juhusliku suuruse mistahes üksikväärtuse esinemise tõenäosus on null. Tulemusest $P(X = x_1) = 0$ tuleb nii aru saada, et katsete korduval sooritamisel toimub sündmus $X = x_1$ väga harva, s.t. selle esinemine on väga ebatõenäone. Pideva juhusliku suuruse korral on mõtet rääkida ainult selle mingisse vahemikku (ka kuitahes väikesse) langemise tõenäosusest. Suuruse võimalike väärtuste lõpmata 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 väiksele maa-alale võib langeda palju mürske.

Praktika seisukohalt pakubki huvi esmajoones see, milistes piirides kõigub reisijate arv linnaliinide bussides tippkoormuste ajal.

7. Tõenäosuse tihedus

Olgu antud pideva juhusliku suuruse X jaotusfunktsioon, millel on tuletis.

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 väärtuse esinemise keskmise tõenäosuse pikkusühiku kohta. Seda nimetatakse tõenäosuse keskmiseks tiheduseks lõigul Δx .

Tõenäosuse keskmine tihedus $p_X(x)$ lõigul Δx on leitav jaotusfunktsiooni kaudu valemist

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

olles jaotusfunktsiooni muutumise keskmiseks kiiruseks.

Tõenäosuse keskmise tiheduse piirväärtust vahemiku pikkuse Δx tõkestamatul kahanemisel nimetatakse tõenäosuse tiheduseks $p(x)$ punktis x ehk tihedusfunktsiooniks. Tõenäosuse tiheduse graafikut nimetatakse tõenäosuste jaotuskõveraks.

Definitsiooni järgi
$$p(x) = \lim_{\Delta x \rightarrow 0} p_k(x)$$

Teiselt poolt

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

funktsiooni tuletise definitsiooni järgi.

Seega tõenäosuse tihedus punktis x on võrdne jaotusfunktsiooni tuletisega samas punktis:

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

Tõenäosuse tihedus on juhusliku suuruse jaotuse üheks esitusviisiks.

Tõenäosuse tiheduse omadusi:

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

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

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

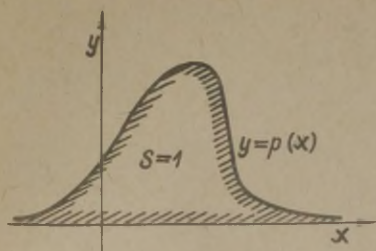
2. Integraal lõpmatutes rajades tõenäosuse tihedusest on võrdne ühega:

$$\int_{-\infty}^{+\infty} p(x) dx = 1, \text{ sest } \int_{-\infty}^{+\infty} p(x) dx = F(+\infty) - F(-\infty) = 1 - 0 = 1$$

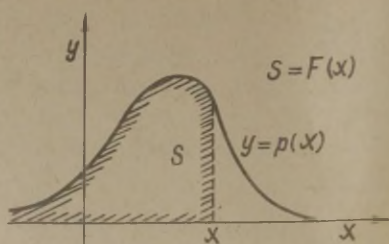
ehk teisiti: sündmus $-\infty < X < +\infty$ on kindel sündmus ning selle tõenäosus on 1.

Geomeetriliselt väljenduvad need omadused järgmiselt: tõenäosuse tiheduse kõver ei asu allpool abstsissstelge ning kogu kõvera ja abstsissstelje vahele jääva kujundi pindala on

võrdne ühega (joonis 9).



Joonis 9.



Joonis 10.

Tihedusfunktsiooniks saab olla neid kaht nõuet rahuldav funktsioon.

Teadaoleva tõenäosuse tiheduse $p(x)$ järgi saab leida jaotusfunktsiooni järgmiselt:

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

Geomeetriliselt vastab jaotusfunktsioonile jaotuskõvera alune pindala punktini x (joonis 10).

1. n ä i d e. Leida tõenäosuse tiheduse

$$p(x) = \frac{1}{\pi(1+x^2)}$$

järgi jaotusfunktsioon.

Saame

$$F(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{dx}{1+x^2} = \frac{1}{\pi} \arctan x \Big|_{-\infty}^x = \frac{1}{\pi} \arctan x + \frac{1}{2}.$$

2. n ä i d e. Funktsioon

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

kannab normaaljaotuse tõenäosuse tiheduse nime. Normaaljaotusel jaotusfunktsioon

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

8. J u h u s l i k u s u u r u s e a n t u d
v a h e m i k k u s a t t u m i s e t ô e -
n ä o s u s e l e i d m i n e t ô e n ä o -
s u s e t i h e d u s e k a u d u

Olgu teada suuruse tõenäosuse tihedus $p(x)$. Leida selle kaudu suuruse vahemikku (x_1, x_2) sattumise tõenäosus.

Võtame arvesse vahemikku sattumise tõenäosuse leidmise eeskirja jaotusfunktsiooni kaudu ning jaotus- ja tihedusfunktsiooni seose. Siis

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} F'(x) dx = \int_{x_1}^{x_2} p(x) dx.$$

Seega

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

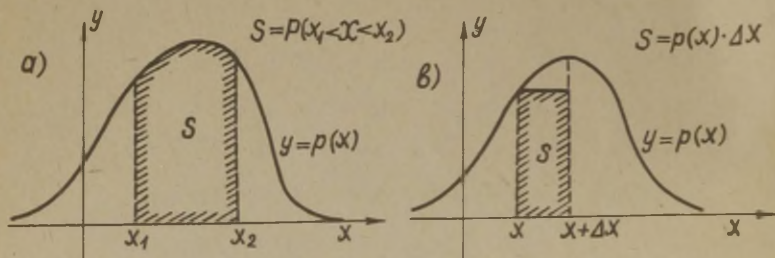
- juhusliku suuruse antud vahemikku sattumise tõenäosus võrdub määratud integraaliga tõenäosuse tihedusest vahemiku otspunktilele vastavates rajades.

Geomeetriliselt vastab suuruse vahemikku sattumise tõenäosusele jaotuskõvera alune pindala vastavas vahemikus (joonis 11a).

Vahemikku $(x, x + \Delta x)$ sattumise tõenäosus ΔF on küllalt väikese Δx korral ligikaudu asendatav diferentsiaaliga dF . Kuid $dF = F'(x) \cdot \Delta x = p(x) \cdot \Delta x$. Seega

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

- suuruse väiksesse vahemikku sattumise tõenäosus võrdub liigikaudu tõenäosuse tiheduse ja vahemiku pikkuse korrutisega. Viga on seda väiksem, mida lühem on vahemik (joonis 11b).



Joonis 11.

Lõpmata väiksesse vahemikku sattumine pideva suuruse korral vastab üksikväärtuse realiseerumisele diskreetse suuruse korral. Seega on vastavuses ka tõenäosused $p(x) \cdot \Delta x$ pideva ja p diskreetse suuruse juhul.

9. Ühtlane jaotus

Illustreerime eespool käsitletud mõisteid ühe konkreetse jaotuse korral.

Kui juhusliku suuruse võimalikud väärtused asuvad kõik vahemikus (a, b) ning tõenäosuse tihedus on selles vahemikus konstantne ning väljaspool nulli, nimetatakse jaotust ühtlaseks. Suurust ennast nimetatakse ühtlase jaotusega suuruseks.

Tõenäosuse tihedus on määratud valemiga

$$p(x) = \begin{cases} 0 & , \text{ kui } x \leq a; \\ \text{const} = c, & \text{ kui } a < x < b; \\ 0 & , \text{ kui } x \geq b. \end{cases}$$

Leiame c väärtuse, arvestades tõenäosuse tiheduse omadusi. Selle kohaselt

$$\int_{-\infty}^{+\infty} p(x) dx = 1.$$

Kuna $p(x)$ on määratud erinevate valemitega arvtelje üksikutes osades, siis

$$\int_{-\infty}^{+\infty} p(x) dx = \int_{-\infty}^a 0 dx + \int_a^b c dx + \int_b^{+\infty} 0 dx = c x \Big|_a^b = c(b-a).$$

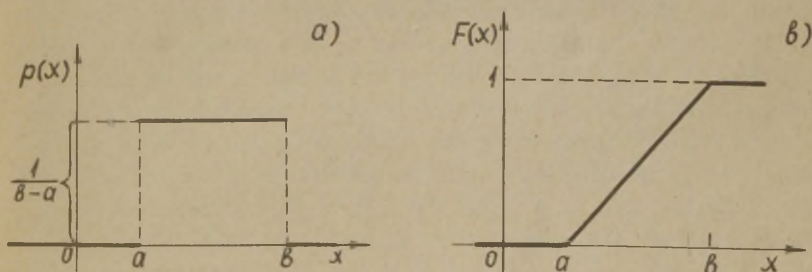
Seega $c(b-a) = 1$ ja $c = 1 : (b-a)$.

Saiame

$$p(x) = \begin{cases} 0 & , x \leq a; \\ \frac{1}{b-a} & , a < x < b; \\ 0 & , x \geq b. \end{cases}$$

Tõenäosuse tiheduse graafik on joonisel 12a. Siit on näha, et tihedusfunktsiooni ja abstsissitelje vahelise kujundi pindala

$$S = \frac{1}{b-a} \cdot (b-a) = 1.$$



Joonis 12.

Jaotusfunktsiooni $F(x)$ leiame tihedusfunktsiooni kaudu. Kui $x \leq a$, siis

$$F(x) = \int_{-\infty}^x 0 \cdot dx = 0;$$

kui $a < x < b$, siis

$$F(x) = 0 + \int_a^x \frac{1}{b-a} dx = \frac{x}{b-a} \Big|_a^x = \frac{x-a}{b-a};$$

kui $x \geq b$, siis

$$F(x) = 0 + \int_a^b \frac{1}{b-a} dx + 0 = \frac{b-a}{b-a} = 1.$$

Saiame

$$F(x) = \begin{cases} 0 & , x \leq a; \\ \frac{x-a}{b-a} & , a < x < b; \\ 1 & , x \geq b. \end{cases}$$

Jaotusfunktsiooni graafik on joonisel 12b.

N ä i d e. Teatud liinil liiguvad bussid täpselt 5-minutiliste intervallidega. Bussipeatusse tuleva liiklusgraafikut mitte teadva reisiija ooteaeg T on juhuslik suurus, mille võimalikud väärtused asuvad 0 ja 5 minuti vahel. Tõenäosus, et bussi tuleb oodata 1 või 2 või mingi muu arv minuteid, on teadmata. Kuna pole mingit alust teatud ooteaega lugeda rohkem võimalikuks kui mõnda teist, siis võib mistahes ooteaja tõenäosust lugeda konstantseks ja tõenäosuse tihedus oleks $p(T) = \frac{1}{5}$, kui $0 < T < 5$. Tõenäosus selleks, et ooteaeg ei ole pikem kui T minutit, avaldub jaotusfunktsiooni kaudu:

$$F(T) = \frac{T-0}{5-0} = \frac{T}{5}.$$

Kahest, viiest minutist lühema ooteaja tõenäosus

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

$$P(T < 5) = F(5) = 1.$$

Üle 5 min. aga tõepoolest 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 nimetatakse juhusliku suuruse arvulisteks karakteristikuteks. Suuruse eri külgede kirjeldamiseks võime luua mitmesuguseid karakteristikuid. Nende hulka ei maksa aga liiga suureks ajada ning need tuleb valida nii, et suuruse meid huvitavad omadused väljenduksid võimalikult hästi.

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

2. Keskväärtus

p. 10.

Asendi karakteristikutest on kõige tähtsam juhusliku suuruse keskväärtus.

1) Diskreetse juhusliku suuruse keskväärtus. Olgu suurus X määratud jaotustabeliga

X	x_1	x_2	...	x_n
P	P_1	P_2	...	P_n

Diskreetse juhusliku suuruse keskväärtuseks nimetatakse suuruse võimalike väärtuste ja nende tõenäosuste korrutiste summat.

Tähistame suuruse X keskväärtuse sümboliga $E(X)$ või m_0 . Jaotustabeli andmetel näeb $E(X)$ arvutamise eeskiri välja järgmiselt:

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

1. n ä i d e. On antud juhusliku suuruse jaotustabel

X	-3	-1	1	2	3
p	0,1	0,15	0,35	0,3	0,1

Leida keskväärtus

$$E(X) = -3 \cdot 0,1 - 1 \cdot 0,15 + 1 \cdot 0,35 + 2 \cdot 0,3 + 3 \cdot 0,1 = \\ = 0,8.$$

2. n ä i d e. Kahe täringu viskamisel saadava võimaliku silmade arvu jaotustabeli andmetel (§ 2, punkt 2) leida silmade arvu keskväärtus.

X	36·p	X·36p
2	1	2
3	2	6
4	3	12
5	4	20
6	5	30
7	6	42
8	5	40
9	4	36
10	3	30
11	2	22
12	1	12
		252

$$E(X) = \frac{1}{36} \cdot 252 = 7.$$

Pikemate jaotustabelite korral on otstarbekas kasutada arvutus-skéemi tabeli vormis.

Käesolevas näites on kõik tõenäosused 1 : 36 kordsed ning seepärast on tabelis tõenäosuste asemel kirjutusviisi hõlbustamiseks võetud 36 korda suuremad väärtused. Lõpuks jagame saadud tulemuse 36-ga.

Juhime tähelepanu asjaolule, et keskväärtus asub alati juhusliku suuruse võimalike väärtuste muutmisvahemikus.

Tõmbame paralleele juhusliku suuruse keskväärtuse ja kehade raskuskeskme vahel. Kujutame ette kaaluta varrast, mille külge riputatakse

se raskused. Valime vardal mingi punkti nullpunktiks ning kinnitame varda külge vihid nendel kaugustel nullpunktist, mis vastavad juhusliku suuruse X võimalikele väärtustele. Vihid valime nii rasked (mingites kaaluühikutes), kui suured on X väärtuste tõenäosused. Vihtide raskuskeskme asukoht vastab siis X keskvaertuse asukohale arvteljel.

3. n ä i d e. Kahepunktiline jaotus. Olgu sündmuse A tõenäosus p ja vastandsündmusel \bar{A} $q = 1 - p$. Sündmuse toimumisele - mittetoimumisele saame vastavusse juhusliku suuruse X järgmiselt: kui katsel A ei toimu (toimub \bar{A}), siis omistame X -le väärtuse $X = 0$ ja kui A toimub, siis $X = 1$. Nende väärtuste tõenäosused on vastavalt q ja p . Saame X jaotustabeli

X	0	1
P	q	p

Saadud jaotust nimetatakse kahepunktiliseks jaotuseks. Leia me kahepunktilise jaotuse keskvaertuse

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

2) Pideva juhusliku suuruse keskvaertus. Olgu teada pideva suuruse tõenäosuse tihedus $p(x)$.

Pideva suuruse keskvaertus defineeritakse järgmiselt:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot p(x) dx .$$

4. n ä i d e. Leida ühtlase jaotuse keskvaertus. Ühtlasel jaotusel saime

$$p(x) = \begin{cases} 0 & , x \leq a; \\ \frac{1}{b-a} & , a < x < b; \\ 0 & , x \geq b. \end{cases}$$

Kui suuruse võimalike väärtuste muutumisvahemik on teada, siis taandub integreerimine lõpmatutes rajades integreerimisele muutumisvahemikule vastavates rajades. Nüüd

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

- ühtlasel jaotusel on keskvaärtuseks suuruse äärmiste väärtuste poolsumma.

3. Aritmeetiline keskmine

Olgu statistilise kogumi tunnuse X uurimiseks sooritatud N katsed ning katseandmed esitatud statistilise reana

$$x_1 \quad x_2 \quad x_3 \quad \dots \quad x_N.$$

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

Keskmise leidmiseks saame eeskirja

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \cdot \sum_{i=1}^N x_i.$$

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

(I)

X	x_1	\dots	x_n
f	f_1	\dots	f_n

(II)

X	x_1	\dots	x_n
w	w_1	\dots	w_n

Variatsioonrea andmetel tuleks \bar{x} leida valemist

$$\bar{x} = \frac{x_1 f_1 + \dots + x_n f_n}{N} = \frac{1}{N} \cdot \sum_{i=1}^n x_i f_i.$$

Viimasel juhul on leitud nn. kaalutud keskmine, esimesel juhul lihtne keskmine. Sagedus f_1 on tunnuse väärtuse x_1 „kaaluks“.

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

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

Kuna

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

siis leitakse aritmeetiline keskmine statistilise jaotustabeli korral valemist

$$\bar{x} = x_1 w_1 + x_2 w_2 + \dots + x_n w_n = \sum_{i=1}^n x_i w_i .$$

Intervallitud ridade juhul tuleb arvutuslikeks väärtusteks x_i võtta intervallide keskkohad!

Kuidas on seotud keskväärtaus ja aritmeetiline keskmine?

Katsete arvu suurendamisel on suuruse X iga väärtuse suhtelisel sagedusel tendents läheneda vastava väärtuse tõenäosusele, s.t. võib võtta

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

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 keskväärtaus.

Seega

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

- juhusliku suuruse keskväärtaus 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 keskväärtause ümber. Alati ei ole suuruse jaotusseadus teada ja ligikaudseks jaotusseaduseks võetakse

katseseeriale vastav variatsioonrida. Leitud aritmeetilise keskmise järgi saab ligikaudselt hinnata, milline on uuritava suuruse keskväärtus.

Aritmeetilise keskmise ja keskväärtuse dimensioon ühtib suuruse dimensiooniga.

M ä r k u s. Tihti kasutatakse sümboli $E(X)$ asemel \bar{x} , mis on täiesti õigustatud, sest $\bar{x} \approx E(X)$.

1. n ä i d e. 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 koristatud hektarite arv päevas. Aritmeetiline keskmine

$$\bar{x} = \frac{22 \cdot 6 + 24 \cdot 5 + 25 \cdot 8 + 26 \cdot 7 + 27 \cdot 8}{34} = \frac{850}{34} = 25 \text{ ha.}$$

2. n ä i d e. 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 (\bar{X})	0	1	2	3	4	5	ja rohkem
p	0,24	0,38	0,26	0,10	0,02	0	

Kontrolliti 50 künnest detailist koosnevat partiid ja 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 arvud $E(X)$ ja \bar{x}

10 detailist koosnevas partiis.

Saame

$$E(X) = 0 \cdot 0,24 + 1 \cdot 0,38 + 2 \cdot 0,26 + 3 \cdot 0,10 + 4 \cdot 0,02 = 1,28 ,$$

$$\bar{x} = \frac{0 \cdot 11 + 1 \cdot 20 + 2 \cdot 12 + 3 \cdot 5 + 4 \cdot 2}{50} = 1,34.$$

Võib arvata, et partiide arvu suurenemisel muutuks $E(X)$ ja \bar{x} erinevus veelgi väiksemaks.

3. n ä i d e. Kontrolltöö hinnete jaotus oli õpperühmas järgmine:

Hinne (X)	2	3	4	5
Suht.sagedus (w) %	10	25	50	15

Leida keskmine hinne.

Kasutame keskmise arvutuseeskirja statistilise jaotustabeli juhul (%-des antud suhtelisi sagedusi tuleb jagada sajaga).

$$\bar{x} = 2 \cdot 0,1 + 3 \cdot 0,25 + 4 \cdot 0,50 + 5 \cdot 0,15 = 3,7.$$

4. Keskväärtuse omadusi

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

<table border="1"><thead><tr><th>X</th><th>x_1</th><th>...</th><th>x_m</th></tr></thead><tbody><tr><td>P</td><td>P_1</td><td>...</td><td>P_m</td></tr></tbody></table>	X	x_1	...	x_m	P	P_1	...	P_m	ja	<table border="1"><thead><tr><th>Y</th><th>y_1</th><th>...</th><th>y_n</th></tr></thead><tbody><tr><td>r</td><td>r_1</td><td>...</td><td>r_n</td></tr></tbody></table>	Y	y_1	...	y_n	r	r_1	...	r_n
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 väärtuspaaride 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 võib vaadelda liitsündmusena, milleks on nii x_k kui y_1 samaaegne esinemine. Siis on ju $x_k + y_1$ tõenäosus võrdne osasündmuste tõenäosuste korrutisega $p_k \cdot r_1$. Summa jaotustabel esitub 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	...	$p_m r_n$

1. n ä i d e. Koostada sõltumatute suuruste summa jaotustabel, kui liidetavate jaotustabelid on

X	0	1	2	ja	Y	-1	0	1
p	0,3	0,5	0,2		r	0,2	0,5	0,3

Summa $X + Y$ väärtuste saamiseks liidame kõik võimalikud X ja Y väärtuste paardid. 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 väärtusega 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 väärtuspaaride 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	...	$x_m y_n$
p	p_1r_1	p_1r_2	...	p_1r_n	p_2r_1	...	$p_m r_n$

2. n ä i d e. 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 cX . Konstanti võib vaadelda suurusena, millel on üks väärtus c tõenäosusega 1. Summa erijuhuks on konstandi ja juhusliku suuruse summa $X + c$. Vastavad jaotustabelid saab kujul

cX	cx_1	cx_2	...	cx_n	ja	$X+c$	x_1+c	x_2+c	...	x_n+c
P	P_1	P_2	...	P_n		P	P_1	P_2	...	P_n

(Põhjendage, miks cX või $X + c$ väärtuste tõenäosused on võrdsed X tõenäosusega!).

Asume keskvaartuse omaduste käsitlemisele.

1. o m a d u s. Konstandi keskvaartuseks on konstantise. Tõepoolest,

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

2. o m a d u s. Suuruste summa keskvaartus on võrdne liidetavate keskvaartuste summaga.

Kahe suuruse korral märgime seda matemaatilises sümbolikas

$$E(X + Y) = E(X) + E(Y).$$

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

X	x_1	x_2
P	p_1	p_2

ja

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	$p_1 r_1$	$p_1 r_2$	$p_2 r_1$	$p_2 r_2$

Leiame summa keskvaartuse X + Y jaotustabelist:

$$\begin{aligned}
 E(X+Y) &= (x_1+y_1)p_1r_1 + (x_1+y_2)p_1r_2 + (x_2+y_1)p_2r_1 + \\
 &+ (x_2+y_2)p_2r_2 = x_1p_1(r_1+r_2) + x_2p_2(r_1+r_2) + \\
 &+ y_1r_1(p_1+p_2) + y_2r_2(p_1+p_2) = (x_1p_1+x_2p_2) + \\
 &+ (y_1r_1+y_2r_2) = E(X) + E(Y) .
 \end{aligned}$$

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

3. o m a d u s. Sõltumatute suuruste korrutise keskvaartus on võrdne tegurite keskvaartuse korrutisega.

Kahe suuruse juhul

$$E(XY) = E(X) \cdot E(Y).$$

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

Järeldused: 1. $E(cX) = c \cdot E(X)$;

2. $E(c+X) = E(X) + c$;

3. $E(AX + B) = AE(X) + B$.

Need tulenevad omadustest 1. - 3. (kontrollige!).

Suuruse väärtuse ja mingi jääva arvu c vahet nimetatakse väärtuse hälbeks selle arvu suhtes. Suuruse X hälbe X-c

jaotustabel on järgmine:

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

Keskväärtuse suhtes leitud hälvet nimetatakse tsentreeritud hälbeks $X - m_0$.

4. o m a d u s. Suuruse tsentreeritud hälbe keskväär-
tus on võrdne nulliga:

$$E(X - m_0) = 0.$$

Tõepoolest

$$E(X - m_0) = E(X) - E(m_0) = m_0 - m_0 = 0.$$

See tähendab, et hälbed keskväärtuse suhtes keskmiselt tasa-
kaalustuvad.

N ä i d e. On teada $E(X) = 8$. Leida suuruste $5X$,
 $X - 2$, $3X + 4$ keskväärtused.

Saame

$$E(5X) = 5 \cdot E(X) = 40;$$

$$E(X-2) = E(X) - 2 = 6;$$

$$E(3X + 4) = 3E(X) + 4 = 28.$$

5. A r i t m e e t i l i s e k e s k m i s e a r v u t u s v ö t t e i d

Pikkade variatsioonriidade või suurte arvude korral
võib aritmeetilise keskmise arvutamine kujuneda palju aega
nõudvaks tööks. Kasutame keskväärtuse omadusi keskmise ar-
vutamise lihtsustamiseks.

Tunnuse X jaoks toome sisse abisuuruse

$$X' = \frac{X - a}{k},$$

kus a ja k on põhimõtteliselt meelevaldsed arvud. Avaldame
suuruse X :

$$X = kX' + a.$$

Keskvärtuse omaduste kohaselt

$$E(X) = k \cdot E(X') + a$$

ehk

$$\bar{X} = k \cdot \bar{X}' + a.$$

Arvuks a valitakse harilikult rea keskkohale vastav või suurima sagedusega tunnuse väärtus (NB! intervallitud rea korral vastava intervalli keskoht). Arvu a nimetatakse ka esialgseks keskmiseks. Teguriks k on otstarbekas võttahälbe $X - a$ suurim ühisjagaja. Võrdsete vahedega väärtuste korral on selleks harilikult väärtuste vahe, intervallitud ridade puhul intervalli pikkus.

Kui sagedustel f_1 on olemas ühine jagaja l , siis on otstarbekas kõiki sagedusi jagada arvuga l ja samuti sageduste summat N vähendada l korda.

Võrdusest $\bar{X} = k X' + a$ tulenevad järgmised arvutus-eeskirjad:

variatsioonrea juhul

$$\bar{X} = \frac{k}{N} \sum_{i=1}^n \frac{x_i - a}{k} \cdot \frac{f_i}{l} + a;$$

statistilise jaotustabeli juhul

$$\bar{X} = k \sum_{i=1}^n \frac{x_i - a}{k} w_i + a.$$

Arvutusteks on kasulik koostada sobiv arvutusskeem.

1. n ä i d e. 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: $a = 11,0$, $k = 0,2$ ja $l = 1$.

Siis

$$\bar{X} = \frac{\sum \frac{x_i - 11}{0,2} f_i}{\sum f_i} \cdot 0,2 + 11.$$

Arvutusteks koostame tabeli, kus $x' = \frac{x - 11}{0,2}$.

x	f	x - 11	x'	x' · f
10,2	2	-8	-4	- 8
10,4	3	-6	-3	- 9
10,6	8	-4	-2	-16
10,8	13	-2	-1	-13
11,0	25	0		-46
11,2	20	2	1	20
11,4	12	4	2	24
11,6	10	6	3	30
11,8	6	8	4	24
12,0	1	10	5	5
	100			<u>103</u> 57

Sellesse ritta, mis vastab väärtusele $x = 11,0$ (seal tulevad edaspidi nullväärtused), märgime $\sum x'_i f_i$ negatiivsete väärtuste summa. Positiivsed liikmed liidetakse eraldi ja siis on kerge leida keskmise avaldises murru lugejat.
Nüüd

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

2. n ä i d e. Leida päevanormi täitmise keskmine protsent saja töölise töötulemuste järgi (vaatlusandmed on tabeli kahes esimeses veerus).

Plaani täit- mise %	Tööliste arv f	x	x-137,5	$\frac{x-137,5}{25} \cdot x'$	x'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	0		-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				<u>26</u> - 48

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

3. n ä i d e. Leida statistilise jaotustabeli andmetel keskmine \bar{x} (lähteandmed on tabeli kahes esimeses veerus).

x	w	x'	x'w
0 - 25	0,05	-2	-0,10
25 - 50	0,15	-1	-0,15
50 - 75	0,40	0	-0,25
75 - 100	0,25	1	0,25
100 - 125	0,15	2	0,30
			0,55
			0,30

$$x' = \frac{x - 62,5}{25}$$

Keskmine

$$\bar{x} = 25 \cdot 0,30 + 62,5 = 70.$$

6. M e d i a a n

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 leitakse mediaani (Me) järjekorranumber valemist

$$\frac{N + 1}{2} .$$

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

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

Kui lähteandmed on esitatud grupeeritult momentreana või intervallitud variatsioonreana, siis leitakse mediaan järgmise mõttekäiguga.

Jagame väärtuste koguarvu pooleks ja $\frac{N}{2}$ määrab mediaani järjekorranumbri. Seejärel liidame sagedusi (leiame sa-

geduste järgsumma) seni, kuni jõuame väärtuseni, mille kohal sageduste summa ületab esimesena $\frac{N}{2}$. Seega oleks mediaan-grupp määratud. Momentrea korral ongi mediaaniks vastav tunnuse väärtus. Intervallitud reas täpsustatakse mediaani asukohta mediaanintervalli sees lineaarse interpolatsiooni teel, lugedes väärtuste jaotuse ühtlaseks.

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

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

Mediaani arvutamine kirjeldatud viisil toimub järgmise valemi järgi:

$$Me = x_{Me} + \frac{k(\frac{N}{2} - L)}{f_{Me}},$$

kus L — mediaanintervallile eelnevate intervallide sageduste järgsumma;

f_{Me} — mediaanintervalli sagedus;

k — intervalli pikkus;

x_{Me} — mediaanintervalli algus.

Aritmeetiline keskmine reageerib rea igä üksikväärtuse muutumisele, mediaanile ei avalda mingit mõju väärtuste muutumine mõlemal pool mediaani, kui kummalgi pool jääb liikmete arv võrdseks. Mediaan reageerib liikmete arvu muutumisele ainult siis, kui ühel või teisel pool mediaani muutub liikmete arv erinevaks. Mediaan on vähe tundlik tunnuse üksikväärtuste muutumise suhtes.

7. M o o d

Moodiks (M_o) nimetatakse rea kõige suurema sagedusega liiget. Käesoleva paragrahvi 5. punkti 1. näites on rea moodiks 11,0. Momentrea puhul saame moodi ilma arvutusteta. Intervallitud rea korral määrame esmalt moodintervalli (intervalli, milles mood asub) suurima sageduse järgi. Moodi asukohta täpsustamiseks intervallisiseselt loobutakse eeldusest, et väärtused jaotuvad intervalli ulatuses ühtlaselt (seda eeldatakse \bar{x} ja M_e leidmisel). Moodintervalli naaberintervalli sageduste abil otsustatakse tunnuse jaotumise üle moodintervallis endas. Liikmed loetakse tihedamalt paigutatuks moodintervalli selles osas, mille naaberintervalli sagedus on suurem. Moodi arvutatakse valemist

$$M_o = X_{M_o} + \frac{f_o - f_{-1}}{(f_o - f_{-1}) + (f_o - f_{+1})} \cdot k,$$

kus X_{M_o} - moodintervalli algus;

k - intervalli pikkus;

f_o, f_{+1}, f_{-1} - vastavalt moodintervalli, sellele järgneva ja eelneva intervalli sagedused.

1. n ä i d e. Leida 5. punkti 2. näite andmetel rea mood. Variatsioonreast nähtub, et

$$X_{M_o} = 100, \quad k = 25, \quad f_{+1} = 30, \quad f_{-1} = 18, \quad f_o = 32.$$

Seega

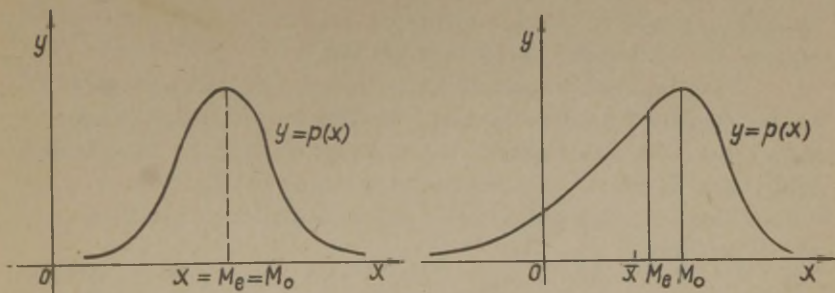
$$M_o = 100 + \frac{(32 - 18) \cdot 25}{(32 - 18) + (32 - 30)} = 100 + \frac{14 \cdot 25}{16} \approx 121,9.$$

Väärtuste intervallisisesest ühtlast jaotumist eeldades tuleks valida moodiks intervalli keskkohaks.

M ä r k u s. Moodi arvutamiseks kasutatakse ka teistsuguseid valemeid, kuid tulemuste erinevused ei ole olulised.

Sümmeetrilise jaotuse korral langevad aritmeetiline keskmine, mood ja mediaan kokku. Ebasümmeetrilise jaotuse

korral on nad erinevad (vt. joonis 13).



Joonis 13.

Mood on võrdne pideva juhusliku suuruse jaotuskõvera kõrgeimale punktile vastava abstsissiga, mediaan - jaotuskõvera alla jäävat pindala poolitava ordinaatlõigu abstsissiga. Mediaan asub \bar{x} ja M_0 vahel aritmeetilisest keskmisest umbes ühe kolmandiku moodi ja keskmise vahemaa kaugusel, s.t.

$$M_e \approx \bar{x} + \frac{1}{3} (M_0 - \bar{x}).$$

8. Keskmise hälve

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

Olgu näiteks juhuslikud suurused antud jaotustabelitega

X	-0,01	0,01
p	0,5	0,5

ja

Y	-100	100
r	0,5	0,5

Mõlema suuruse keskvaartused 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 jaotuse iseloom on väga erinev. Suuruse X väärtused asuvad keskvaartuse ligidal, suuruse Y väärtused hälbibivad kaugemale. Teiste sõnadega: suuruse X hälbed $E(X)$ suhtes on väikesed, suurusel Y suured.

Teades üksnes keskvaartust, moodi või mediaani, ei saa otsustada, kuidas suuruse võimalikud väärtused paigutuvad keskvaartuse suhtes. Tekib vajadus suuruse võimalike väärtuste hajuvust iseloomustava karakteristiku järele.

Kõige lihtsamaks varieerumise näitajaks on variatsiooniamplituud

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

Suurus R ei iseloomusta kuigi hästi väärtuste hajuvust. Võib juhtuda, et suuruse väärtused jaotuvad enamuses tihedalt keskmise ümber, kuid üksikud väärtused jäävad keskmisest kaugemale. Sellisel juhul annab vahe $x_{\max} - x_{\min}$ moonutatud pildi väärtuste jaotumisest. Oleks vaja karakteristikut, mis arvestaks kõigi väärtuste paiknemist. Tsentreeritud hälbe $X - m_0$ keskvaartus selleks ei sobi, sest alati $E(X - m_0) = 0$.

Hajuvuse näitajana kasutatakse mõnikord tsentreeritud hälbe absoluutväärtuse keskvaartust, mida nimetatakse keskmiseks hälbeks ehk absoluutseks hälbeks $d = E(|X - m_0|)$. Keskmise hälbe dimensiooniks on suuruse X dimensioon.

Keskmise hälbe arvutuseeskiri on sõltuvalt lähteandmete esitusviisist järgmine:

teoreetilise jaotustabeli juhul

$$d = \sum |x_1 - m_0| \cdot p_1 ;$$

statistilise jaotustabeli juhul

$$d = \sum |x_1 - \bar{x}| \cdot w_1 ;$$

variatsioonrea juhul

$$d = \frac{\sum |x_1 - \bar{x}| \cdot f_1}{N} .$$

Mida tihedamalt paigutuvad suuruse väärtused keskvaartuse (keskmise) ümber, seda väiksemad on üldiselt hälbed ning väiksem on ka keskmine hälve.

Arvutamisel on otstarbekas kasutada kindlat arvutus-skeemi.

1. n ä i d e. Leida suuruse X keskmine hälve, kui suurus on antud jaotustabeliga

X	1	2	5
p	0,3	0,5	0,2

Belnevalt tuleb leida keskvaartus $E(X)$ ja siis saab leida d . Arvutuskäik on näha tabelist

X	p	xp	$ x - m_0 $	$p x - m_0 $
1	0,3	0,3	1,3	0,39
2	0,5	1,0	0,3	0,15
5	0,2	1,0	2,7	0,54
		2,3		1,08

$$m_0 = \sum x_i p_i = 2,3; \quad d = 1,08.$$

2. n ä i d e. Leida keskmine hälve variatsioonrea andmetel (lähteandmed on tabeli kahes esimeses veerus).

X	f	Xf	$ X - \bar{x} $	$f X - \bar{x} $
4 - 6	2	10	3,76	7,52
6 - 8	5	35	1,76	8,80
8 - 10	12	108	0,24	2,88
10 - 12	6	66	2,24	13,44
	25	219		32,64

$$\bar{x} = \frac{219}{25} = 8,76;$$

$$d = \frac{32,64}{25} \approx 1,31.$$

Intervallitud rea korral võetakse arvutuslikeks väärtusteks ikka intervallide keskkohad.

9. Dispersioon

Hajuvuse karakteristikutena kasutatakse kõige sagedamini dispersiooni ja standardhälvet.

Juhusliku suuruse dispersiooniks nimetatakse suuruse tsentreeritud hälbe ruudu keskvaartust

$$D(X) = E \left[(X - m_0)^2 \right].$$

Dispersiooni tähistatakse sageli ka sümboliga σ^2 . Dispersiooni definitsioonist tuleneb järgmine arvutusvalem:

diskreetse suuruse jaoks

$$D(X) = \sum (x_i - m_0)^2 p_i;$$

pideva suuruse jaoks

$$D(X) = \int_{-\infty}^{+\infty} (x - m_0)^2 p(x) dx.$$

Dispersiooni arvutamisel tuleb eelnevalt leida X keskvaartus m_0 ja seejärel hälvete ruudu $(X - m_0)^2$ keskvaartus.

1. näide. Leida 8. punkti 1. näite andmetel suuruse X dispersioon. Arvutuste käiku saab jälgida tabelist

X	p	Xp	$X - m_0$	$(X - m_0)^2$	$p(X - m_0)^2$
1	0,3	0,3	-1,3	1,69	0,507
2	0,5	1,0	-0,3	0,09	0,045
5	0,2	1,0	2,7	7,29	1,458
		2,3			2,010

$$D(X) = 2,01.$$

Dispersioon on alati mittenegatiivne, sest selle leidmisel tuleb summeerida mittenegatiivseid liidetavaid.

2. näide. Leida kahepunktilise jaotuse dispersioon. Kahepunktilise jaotuse esitasime keskvaartuse käsitlemisel 3. näites ning leidsime juba keskvaartuse $E(X) = p$.

Nüüd

$$D(X) = (0-p)^2 \cdot q + (1-p)^2 \cdot p = p^2q + q^2p = pq(p+q) = pq;$$

$$D(X) = pq.$$

Dispersiooni arvutamine lihtsustub, kui teisendada dispersiooni definitsioonile vastavat valemit. Häälbe ruut $(X - m_0)^2$ on esitatav kujul

$$(X - m_0)^2 = X^2 - 2Xm_0 + m_0^2$$

ning nüüd keskvaertuse omadusi arvestades

$$\begin{aligned} D(X) &= E[(X - m_0)^2] = E(X^2 - 2Xm_0 + m_0^2) = E(X^2) - \\ &- 2m_0E(X) + m_0^2 = E(X^2) - 2m_0^2 = E(X^2) - m_0^2. \end{aligned}$$

Saiame

$$D(X) = E(X^2) - [E(X)]^2$$

- dispersioon on võrdne suuruse ruudu keskvaertuse ja keskvaertuse ruudu vahega.

Eeltoodust tulenevad järgmised arvutuseeskirjad:

diskreetse suuruse jaoks

$$D(X) = \sum_{i=1}^n x_i^2 p_i - \left(\sum_{i=1}^n x_i p_i \right)^2;$$

pideva suuruse jaoks

$$D(X) = \int_{-\infty}^{+\infty} x^2 p(x) dx - \left[\int_{-\infty}^{+\infty} xp(x) dx \right]^2.$$

Dispersiooni arvutamine saadud valemit järgi on oluliselt lihtsam kui definitsioonivalemit järgi.

3. n ä i d e. Leida suuruse X dispersioon jaotustabeli andmetel (lähteandmed on arvutuskeemi kahes esimeses veerus).

x	p	xp	x ²	x ² p
-4	0,1	-0,4	16	1,6
-2	0,25	-0,5	4	1,0
0	0,3	0	0	0
2	0,25	0,5	4	1,0
4	0,1	0,4	16	1,6
		0		5,2

$$E(X) = 0;$$

$$\begin{aligned} D(X) &= 5,2 - 0^2 = \\ &= 5,2. \end{aligned}$$

4. n ä i d e. Leida ühtlase jaotusega suuruse dispersioon.

Ühtlasel jaotusel leidsime keskvaartuse $E(X) = \frac{a+b}{2}$

ja

$$D(X) = \int_{-\infty}^{+\infty} x^2 p(x) dx - \left(\frac{a+b}{2}\right)^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \\ = \frac{1}{3(b-a)} (b^3 - a^3) - \frac{a^2 + 2ab + b^2}{4} = \frac{a^2 - 2ab + b^2}{12}$$

ehk

$$D(X) = \frac{(a-b)^2}{12}.$$

Ruutjuurt dispersioonist nimetatakse keskmiseks ruuthälbeks ehk standardhälbeks

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

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

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

10. D i s p e r s i o o n i o m a d u s i

1. o m a d u s. Konstandi c dispersioon on null.

Tõepoolest

$$D(c) = E(c)^2 - E(c)^2 = c^2 - c^2 = 0 \quad (\text{vt. } E(X) \text{ omadusi}).$$

Konstandil on ainult üks väärtus ja mingit hajumist ei ole.

2. o m a d u s. Kui suuruse võimalikud väärtused omavad ühise teguri c , s.t. $Y = cX$, siis võib selle teguri tuua dispersiooni märgi ette ruutu tõstetuna.

$$D(cX) = c^2 D(X).$$

Omaduse õigsuses veendumiseks lähtume dispersiooni definitsioonist ja võtame arvesse, et $E(cX) = cE(X) = cm_0$.

Siis

$$\begin{aligned} D(cX) &= \sum [cx_1 - E(cX)]^2 p_1 = \sum (cx_1 - cm_0)^2 p_1 = \\ &= \sum c^2(x_1 - m_0)^2 p_1 = c^2 \sum (x_1 - m_0)^2 p_1 = c^2 D(X). \end{aligned}$$

3. o m a d u s. Suuruste summa dispersioon. Piirdume kahe suuruse X ja Y summa dispersiooni juhuga. Tähistame suuruste keskvaärtused vastavalt m_x ja m_y . Siis summa $X + Y$ keskvaärtus $E(X + Y) = m_x + m_y$. Nüüd leiame hälbe ruudu ja teisendame selle avaldist

$$\begin{aligned} [X + Y - (m_x + m_y)]^2 &= [(X - m_x) + (Y - m_y)]^2 = \\ &= (X - m_x)^2 + (Y - m_y)^2 + 2(X - m_x)(Y - m_y). \end{aligned}$$

Summa $X + Y$ dispersiooni avaldise teisendamisel kasutame keskvaärtuse omadusi

$$\begin{aligned} D(X + Y) &= E[(X - m_x)^2 + (Y - m_y)^2 + 2(X - m_x)(Y - m_y)] = \\ &= E(X - m_x)^2 + E(Y - m_y)^2 + 2E(X - m_x)(Y - m_y). \end{aligned}$$

Kaks esimest liidetavat on vastavalt $D(X)$ ja $D(Y)$. Poolt viimasest liidetavast nimetatakse korrelatsioonimomendiks C_{xy} :

$$C_{xy} = E[(X - m_x)(Y - m_y)].$$

Seega

$$D(X + Y) = D(X) + D(Y) + 2C_{xy}$$

- suuruste summa dispersioon võrdub liidetavate dispersioonide ja kahekordse korrelatsioonimomendi summaga.

Kui suurused X ja Y on teineteisest sõltumatud, siis keskvaärtuse 3. ja 4. omaduse kohaselt

$$C_{xy} = E(X - m_x) \cdot E(Y - m_y) = 0.$$

Sõltumatute suuruste summa dispersioon võrdub liidetavate dispersioonide summaga

$$D(X + Y) = D(X) + D(Y).$$

Sõltuvate suuruste korral on korrelatsioonimoment ül-

diselt nullist erinev.

J ä r e l d u s e d

1. Kahe suuruse vahe dispersioon võrdub suuruste dispersioonide summaga

$$D(X - Y) = D(X) + D(Y) \quad (2. \text{ ja } 3. \text{ omaduse põhjal}).$$

2. Kui k ja b on konstandid, siis

$$D(kX + b) = k^2 D(X) \quad (\text{põhjendada!}).$$

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

Viimane järeldus leiab kasutamist dispersiooni arvutamise lihtsustamisel.

11. M o m e n d i d

Juhusliku suuruse arvuliste karakteristikute olulise klassi moodustavad nn. momendid.

Suuruse X s -järku algmomendiks ν_s nimetatakse suuruse s -astme keskvaärtust

$$\nu_s(X) = E(X^s).$$

Suuruse s -järku tsentraalseks momendiks μ_s nimetatakse suuruse tsentreeritud hälbe s -astme keskvaärtust

$$\mu_s(X) = E(X - m_0)^s.$$

Statistilise jaotuse korral tuleb keskvaärtuse all mõista aritmeetilist keskmist.

Momentide arvutusvalemeid vaatleme ainult diskreetse suuruse juhul (sulgudes on toodud arvutuseeskirjad variatsioonrea juhuks).

Esimest järku:

algmoment

$$\nu_1 = E(X^1) = \sum x_i p_i = \left(\frac{1}{N} \sum x_i f_i \right);$$

tsentraalne moment

$$\mu_1 = E(X - m_0) = 0 \quad (\text{alati}).$$

Teist järku:

algmoment

$$\nu_2 = E(X^2) = \sum x_i^2 p_i = \left(\frac{1}{N} \sum x_i^2 f_i \right);$$

tsentraalne moment

$$\mu_2 = E[(X - m_0)^2] = \sum (x_i - m_0)^2 p_i = \left(\frac{1}{N} \sum (x_i - \bar{x})^2 f_i \right).$$

Kolmandat järku:

algmoment

$$\nu_3 = E(X^3) = \sum x_i^3 p_i = \left(\frac{1}{N} \sum x_i^3 f_i \right);$$

tsentraalne moment

$$\mu_3 = E[(X - m_0)^3] = \sum (x_i - m_0)^3 p_i = \left(\frac{1}{N} \sum (x_i - \bar{x})^3 f_i \right).$$

Neljandat järku:

algmoment

$$\nu_4 = E(X^4) = \sum x_i^4 p_i = \left(\frac{1}{N} \sum x_i^4 f_i \right);$$

tsentraalne moment

$$\mu_4 = E[(X - m_0)^4] = \sum (x_i - m_0)^4 p_i = \left(\frac{1}{N} \sum (x_i - \bar{x})^4 f_i \right).$$

Tsentraalsed ja algmomentid on omavahel seotud valemitega

$$\begin{aligned}\mu_2 &= \nu_2 - \nu_1^2; \\ \mu_3 &= \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3; \\ \mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4,\end{aligned}$$

mis võimaldavad tsentraalsete momentide arvutamist lihtsustada.

Eespool toodust on näha, et keskväärtns ja dispersioon avalduvad momentide kaudu järgmiselt:

$$E(X) = \nu_1; \quad D(X) = \mu_2 = \nu_2 - \nu_1^2.$$

Saab näidata, et tsentraalsed momentid ei muutu, kui lähtesuurus X asendada suurusega $X \pm a$. Kui lähtesuuruse X

asemel võtta abisuurus $X' = \frac{X - a}{k}$, siis suuruste $X - a$ ja X' algmomendid on seotud järgmiselt:

$$\nu'_s(X - a) = k^s \cdot \nu'_s(X').$$

See seos võimaldab momentide arvutamist oluliselt lihtsustada, kuna arvude a ja k sobiva valikuga saab X' väärtused muuta küllalt väikesteks. Intervallitud statistiliste jaotuste korral valitakse arvuks a harilikult rea keskel asuva intervalli keskoht ning jagajaks k intervalli pikkus või võrdsete vahemaadega üksikväärtuste puhul rea samm.

N ä i d e. On antud 1000 abonendi jaotus tarbitud elektrienergia koguse järgi (kWh). Leida 2. - 4. järku tsentraalsed momendid. Lähteandmed on arvutusskeemi kahes esimeses veerus.

Valime

$$x' = \frac{x - 27,5}{5}.$$

Väärtusele $x' = 0$ vastavasse ritta on märgitud negatiivsete arvude summa veergudes, kus see on vajalik

x	f	x'	fx'	x'^2	fx'^2	x'^3	fx'^3	x'^4	fx'^4
5 - 10	3	-4	- 12	16	48	-64	-192	256	768
10 - 15	13	-3	- 39	9	117	-27	-351	81	1053
15 - 20	70	-2	-140	4	280	- 8	-560	16	1120
20 - 25	190	-1	-190	1	190	- 1	-190	1	190
25 - 30	290	0	-391				-1293		
30 - 35	230	1	230	1	230	1	230	1	230
35 - 40	130	2	260	4	520	8	1040	16	2080
40 - 45	62	3	186	9	558	27	1674	81	5022
45 - 50	12	4	48	16	192	64	768	256	3072
	1000		724		2135		3712		13535
			343				2419		

Nüüd leiame suuruse $X - a$ algmomendid:

$$\nu_1 = \frac{5 \cdot 343}{1000} = 1,715 ; \quad \nu_2 = \frac{5^2 \cdot 2135}{1000} = 53,375 ;$$

$$\nu_3 = \frac{5^3 \cdot 2419}{1000} = 302,375; \quad \nu_4 = \frac{5^4 \cdot 13535}{1000} = 8459,375.$$

Tsentraalsed momendid:

$$\mu_2 = 53,3 - 1,72^2 = 50,42;$$

$$\mu_3 = 302,3 - 3 \cdot 1,72 \cdot 53,38 + 2 \cdot 1,72^2 = 33,91;$$

$$\mu_4 = 8459,38 - 4 \cdot 1,72 \cdot 302,38 + 6 \cdot 53,38 \cdot 1,72^2 - 3 \cdot 1,72^4 = 7299,73.$$

12. Dispersiooni leidmine statistiliste jaotuste korral

Dispersiooni definitsiooni kohaselt saab arvutusees-
kirjaks
variatsioonrea korral

$$\sigma^2 = \frac{\sum (x_1 - \bar{x})^2 f_1}{\sum f_1} = \frac{1}{N} \sum x_1^2 f_1 - \bar{x}^2;$$

statistilise jaotustabeli korral

$$\sigma^2 = \sum (x_1 - \bar{x})^2 w_1 = \sum x_1^2 w_1 - \bar{x}^2.$$

Dispersiooni teisendatud arvutuseeskiri on sümboolselt esi-
tatud kujul

$$\sigma^2 = \overline{x^2} - \bar{x}^2$$

- dispersioon võrdub väärtuste ruutude keskmise ja keskmise ruudu vahega.

Intervallitud variatsioonreas valitakse tunnuse väärtusteks intervallide keskkohad.

Dispersiooni σ^2 arvutamine võib valmistada suurte arvude puhul tehnilisi raskusi. Sellepärast antakse σ^2 leidmise eeskirjale tihti teisendatud kuju. Valemit tei-

sendatakse nii, et saab ära kasutada ka \bar{x} leidmisel tehtud arvutusi. Tuginedes \bar{x} leidmise eeskirjale, saab

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

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

Konkreetse arvutuseeskirja valik oleneb lähteandmete hulgast ja üksikväärtuste suurusest.

1. n ä i d e. Leida standardhälve 3. punkti 1. näite andmetel.

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

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

Standardhälve

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

2. n ä i d e. Leida dispersioon ja standardhälve 3. punkti 2. näite andmetel (automaatpingil toodetud detailide standardsuse kontroll).

Esitame arvutused ühes tabelis, kusjuures puuduvad lahtrid $\sum p_i x_i$ ja $\sum f_i x_i$ leidmiseks, sest keskmine on eespool saadud $E(X) = 1,28$; $\bar{x} = 1,34$.

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

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

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

3. n ä i d e. Aastatel 1889-1890 mõõdeti Moskva 1000 meestööliste pikkus (mõõtmisandmed on tabeli kahes esimeses veerus). Leida tööliste keskmine pikkus ja standardhälve.

$$\text{Valime } k = 3, \quad b = 165,5, \quad l = 1 \text{ ja } x' = \frac{x - 165,5}{3}.$$

x	f	x'	x' ²	x'·f	x' 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	2	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	

Saame

$$\bar{x} = \frac{k}{N} \sum_1 x'_i f_i + a = 165,5 + \frac{10 \cdot 3}{1000} = 165,53 \text{ cm};$$

$$\begin{aligned} \sigma^2 &= \frac{\sum_1 x'_i{}^2 \cdot f_i}{N} \cdot k^2 - (\bar{x} - a)^2 = \frac{4064}{1000} \cdot 9 - \\ &- (165,53 - 165,5)^2 \approx 36,58; \end{aligned}$$

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

Mida väiksem on standardhälve σ , seda tihedamalt paigutuvad suuruse väärtused aritmeetilise keskmise ümber,

kusjuures suuruse väärtuste hälbed keskmise suhtes enamikul juhtudel ei ületa kolmekordset standardhälvet (põhjendatakse hiljem). Hinnata seni lahendatud näidete puhul, kas ja kui palju väärtusi jääb väljapoole tõkkeid $\bar{x} \pm 3\sigma$.

13. Normeeritud hälve

Juhusliku suuruse X tsentreeritud hälbeks nimetame suurust $X - m_0$.

Normeeritud hälbeks t nimetatakse tsentreeritud hälbe ja standardhälbe jagatist

$$t = \frac{X - m_0}{\sigma} .$$

Leiame normeeritud hälbe keskvaartuse ja dispersiooni. Selleks toetume nendele omadustele:

$$E(t) = E\left(\frac{X - m_0}{\sigma}\right) = \frac{1}{\sigma} E(X - m_0) = 0;$$

$$D(t) = D\left(\frac{X - m_0}{\sigma}\right) = \frac{1}{\sigma^2} D(X) = \frac{\sigma^2}{\sigma^2} = 1 .$$

Normeeritud hälbe keskvaartus on alati null ja dispersioon ning standardhälve üks.

§ 4. JUHUSLIKU SUURUSE JAOTUSI

1. Korduvate katsete skeem

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

Oletame, et ühegi lasu tulemus ei sõltu eelmiste lasude tulemustest. Iga lasu puhul kas tabatakse südamikku (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 selle värvus ja pannakse tagasi. Kuuli võtmine on üksikkatse, mida võib korrata vajalik arv kordi. Pärast kuuli tagasipanekut segatakse 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.

Muutumatus katsetingimustes sooritataavaid ühetüübilisi katseid, kus ühegi katse tulemus ei sõltu eelmistest katsetest, nimetatakse sõltumatus 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 räägitakse korduvate katsete skeemist.

2. Bernoulli valem

Tihti tuleb otsida vastust sellistele küsimustele: kui tõenäone (kui usutav, põhjendatud) on, et

- piimakombinaadis vooluliinilt tulevast 1000 korgitud pudelist on 950 vigastamata?

- kauplusesse toodavast 10000 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 ja \bar{A} tõenäosus üksikkatsel on vastavalt p ja $q = 1 - p$?

Tähistame sündmuse „ A toimub n katsel m korda“ sümboliga B_m . Sündmuse A toimumine üksikkatsel järjekorranumbriga i olgu sündmus A_i ja mittetoimumine \bar{A}_i . Sündmus B_m toimub, kui realiseeruvad sündmused A_i ja \bar{A}_i vahekorras m ja $n - m$ korda, kuid ükskõik millises järjestuses. Seega on B_m järgmiste sündmuste summa

$$B_m = (A_1 A_2 \dots A_m \bar{A}_{m+1} \dots \bar{A}_n) \cup (\bar{A}_1 A_2 \dots A_m \bar{A}_{m+1} \dots \bar{A}_{n-1} A_n) \cup \dots \\ \dots \cup (A_1 \bar{A}_2 A_3 \dots A_m \bar{A}_{m+1} \dots \bar{A}_{n-1} A_n) \cup (\bar{A}_1 \dots \bar{A}_{n-m} A_{n-m+1} \dots A_n),$$

kus igat osasündmust võib omakorda vaadelda sündmuste A_i ja \bar{A}_j korrutisena (m korda sündmus A ja $n-m$ korda \bar{A}). Iga summasse kuuluva osasündmuse tõenäosuse saame tõenäosuste korrutamislausest - $p^m q^{n-m}$. Summas on osasündmusi nii palju, kui saab moodustada kombinatsioone n elemendist $A_1, A_2, \dots, \dots, A_m, \bar{A}_{m+1}, \dots, \bar{A}_n$ m -kaupa (igas n elemendist koosnevas kombinatsioonis peab A esinema m korda, kusjuures vähemalt ühes A järjekorranumbris i peab olema erinevus). Kom-

binatsioonide arv C_n^m n elemendist m-kaupa arvutatakse valemist

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

Järelikult

$$P(B_m) = \underbrace{p^m q^{n-m} + \dots + p^m q^{n-m}}_{C_n^m \text{ korda}} .$$

Sündmuse A n katsel m-kordse toimumise ehk sündmuse sageduse m tõenäosust tähistame edaspidi $P_{n,m}$.

Oleme jõudnud Bernoulli valemini

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

On kokku lepitud, et $0! = 1$.

Sageduse väärtuste $m = 0$ ja $m = n$ tõenäosused saab leida ka vahetult tõenäosuste korrutamislausest

$$P_{n,0} = q^n , \quad P_{n,n} = p^n .$$

1. n ä i d e. 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:

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

2. n ä i d e. Lahendame eelmise punkti 1. näites tõstatatud probleemi. Laskude arv $n = 10$, märklaua südamikku tabamuste (sündmus A) arv $m = 8$ ja $p = 0,9$. Siis $q = 1 - 0,9 = 0,1$ ning

$$P_{10,8} = \frac{10!}{8! \cdot 2!} 0,9^8 \cdot 0,1^2 = \frac{10 \cdot 9}{2} \cdot 0,9^8 \cdot 0,01 \approx 0,194 .$$

Juba käesolevast näitest ilmneb, et suuremate arvude

puhul põrkame kokku arvutuslike raskustega.

3. n ä i d e. Tõenäosus ööpäevaks ettenähtud elektrienergia normi mitteületamiseks $p = 0,75$. Kui tõenäone on, et 6 päevast 4 päeval ei ületata normi?

Ülesande andmetest näeme, et $p = 0,75$, $n = 6$, $m = 4$ ja $q = 1 - p = 0,25$.

$$P_{6,4} = \frac{6!}{4! 2!} \cdot 0,75^4 \cdot 0,25^2 = \frac{5 \cdot 6}{2} \frac{3^4}{4^4 4^2} = 0,30.$$

4. n ä i d e. 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 $q = 1 - p = 0,5$.

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

$$P_{4,3} = \frac{4!}{3! \cdot 1!} \left(\frac{1}{2}\right)^3 \cdot \frac{1}{2} = \frac{1}{4} > \frac{7}{32}.$$

Tõenäosem on võita 3 mängu neljast.

3. B i n o o m j a o t u s

Sündmuse sagedus korduvatel katsetel on juhuslik suurus, mis võib omandada täisarvulisi väärtusi 0-st katsete arvuni n . Sageduse mistahes väärtuse m tõenäosus $P_{n,m}$ on leitav Bernoulli valemist

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

Juhuslikku suurust, mille väärtuste m ($m = 0, 1, \dots, n$) tõenäosused on määratavad Bernoulli valemist, nimetatakse binoomjaotusega juhuslikuks suuruseks.

Binoomjaotuse jaotustabel on selline:

m	0	1	2	...	m	...	n
$P_{n,m}$	$P_{n,0}$	$P_{n,1}$	$P_{n,2}$...	$P_{n,m}$...	$P_{n,n}$

Sündmuse sagedus korduvatel katsetel on binoomjaotusega suurus.

1. n ä i d e. 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 valemit

$$P_{5,m} = \frac{5!}{m!(5-m)!} 0,4^m \cdot 0,6^{5-m}.$$

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. n ä i d e. Koostada sageduse m jaotustabel, kui $p = 0,5$ ja $n = 5$. Käesoleval juhul on ka $q = 0,5$ ja avaldises

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

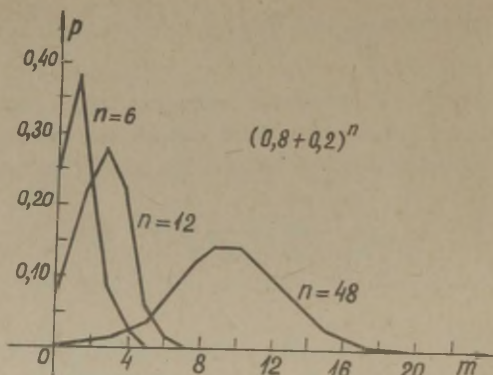
on iga m korral $0,5^m \cdot 0,5^{5-m} = 0,5^5 = 0,03125$.

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 binoomjaotus ebasümmeetriline. Kuid isegi teineteisest erinevate p ja q korral katsete arvu n suurenemisel läheneb jaotus 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

14). Katsete arvu kasvamisel läheneb polügoon sümmeetrilisele kõverale.



Joonis 14.

Käsitleme mõningaid binoomjaotuse omadusi, tuginedes eelmistele näidetele ja joonisele 14.

1) Sageduse m kõigi võimalike väärtuste tõenäosuste summa on üks

$$\sum_{m=0}^n P_{n,m} = 1.$$

Tõepoolest, sageduse m kõik võimalikud väärtused $m = 0, 1, \dots, n$ kui juhuslikud sündmused moodustavad täieliku sündmuste süsteemi ja nende sündmuste tõenäosuste summa on alati üks.

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

3) Katsete arvu n kasvades sama p korral tõenäosuse $P_{n,m}$ väärtused vähenevad.

Kui n kasvab, siis suureneb sageduse m võimalike väärtuste arv. Kuna alati $\sum P_{n,m} = 1$ ja n kasvades summas liidetavate $P_{n,m}$ arv kasvab, siis iga sageduse väärtuse esinemise tõenäosused ise üldiselt vähenevad.

4. Binoomjaotuse karakteristikud

Kelnevalt veendume, et sündmuse sagedus m korduvatel katsetel on esitatav n ühesuguse jaotusega kahepunktilise jaotuse summana. Kahepunktiline jaotus oli eespool defineeritud jaotustabeliga

X	0	1
P	q	p

ja karakteristikud olid $E(X) = p$, $D(X) = pq$. Katsel järjekorranumbriga i võib sündmus A toimuda või mitte toimuda, seega võib igale üksikkatsel seada vastavusse kahepunktilise jaotuse X_i karakteristikutega $E(X_i) = p$, $D(X_i) = pq$ ($i = 1, 2, \dots, n$). Kui üksikkatsel toimub sündmus A , siis katsel vastav $X_i = 1$ ja kui n katsel sündmus A toimub m korda, siis on nullist erinevad ka m suurust X_i . Seega

$$m = X_1 + \dots + X_n = \sum_{i=1}^n X_i.$$

Sõltumatutel katsetel on ka suurused X_i üksteisest sõltumatud. Keskvärtuse ja dispersiooni omaduste kohaselt

$$E(m) = E\left(\sum X_i\right) = \sum E(X_i) = np;$$

$$D(m) = D\left(\sum X_i\right) = \sum D(X_i) = npq$$

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

Jagades sageduse m võimalikud väärtused katsete üldarvuga n , saame suhtelise sageduse $\frac{m}{n} = w$. Ka suhteline sagedus on juhuslik suurus; tema keskvärtus

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

dispersioon

$$D(w) = \frac{1}{n^2} D(m) = \frac{1}{n^2} npq = \frac{pq}{n}; \quad \sigma(w) = \sqrt{\frac{pq}{n}}$$

- binoomjaotusel suhtelise sageduse keskväärtus võrdub sündmuse tõenäosusega üksikkatsel ega sõltu katsete arvust: hajuvus väheneb katsete arvu kasvamisel (dispersiooni avaldises on n murru nimetajas).

Sageduse m seda väärtust, mille tõenäosus on suurim, nimetatakse tõenäoiseimaks sageduseks. Tõenäoiseim sagedus μ on binoomjaotuse mood. Kuna binoomjaotus läheneb sümmeetrilisele jaotusele, siis läheneb ka mood keskväärtusele. Tõenäoiseim sagedus on võrdne keskväärtusega

$$\mu = np,$$

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

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

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

1. n ä i d e. Leida tõenäoiseim 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. n ä i d e. Leida tõenäoiseim 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

$$np + p = 7,2 + 0,8 = 8;$$

$$np - q = 7,2 - 0,2 = 7.$$

Tõenäoiseimad sagedused on 7 ja 8 võrdse tõenäosusega.

3. n ä i d e. Praakdetalli tootmise tõenäosus on 0,035. Leida tõenäoiseim praagi hulk 500 detaili tootmisel.

Nüüd on $p = 0,035$: $n = 500$ ja $np = 500 \cdot 0,035 = 17,5$. Tõenäoiseim praagi hulk asub tõkete $np + p = 17,5 + 0,035 = 17,535$ ja $np - q = 17,5 - 0,966 = 16,535$ vahel. Järeli-

kult $\mu = 17$. Leiame, kui tõenäone on $\mu = 17$ praakdetaili saamine. Tuleb arvutada

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

mis valmistab tõsisel tehnilisi raskusi.

4. n ä i d e. Sündmuse tõenäosus üksikkatsel $p = 0,7$. Leida sageduse ja suhtelise sageduse karakteristikud katseseerias pikkusega 1) $n = 100$; 2) $n = 1600$:

$$1) E(m) = np = 100 \cdot 0,7 = 70;$$

$$D(m) = npq = 70 \cdot 0,3 = 21; \quad \sigma(m) = \sqrt{21} \approx 4,58;$$

$$E(w) = p = 0,7; \quad D(w) = \frac{pq}{n} = 0,0021;$$

$$\sigma(w) \approx 0,0458.$$

$$2) E(m) = 1600 \cdot 0,7 = 1120; \quad D(m) = 1120 \cdot 0,3 = 336;$$

$$\sigma(m) = 18,3; \quad E(w) = 0,7; \quad D(w) = \frac{0,21}{1600} = 0,000131;$$

$$\sigma(w) = 0,0114.$$

5. A s ü m p t o o t i l i n e v a l e m s a g e d u s e t õ e n ä o s u s e l e i d m i s e k s

Suure katsete arvu n korral on tõenäosuste $P_{n,m}$ arvutamine Bernoulli valemi abil väga tülikas. Püüame leida lihtsama mooduse vastava tõenäosuse leidmiseks. Toetume Laplace'i (laplass) lokaalsele piirteoreemile: katsete arvu n tõkestamatul kasvamisel läheneb korrutis $\sqrt{npq} \cdot P_{n,m}$ funktsioonile

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad t = \frac{m - np}{\sqrt{npq}}.$$

Matemaatilises sümboolikas

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

Funktsiooni $\varphi(t)$ nimetatakse normaaljaotuse tihedu-

seks. Seega: katsete arvu kasvamisel läheneb binoomjaotus normaaljaotusele. (Normaaljaotust käsitleme järgnevates punktides).

Laplace'i lokaalteoreemist saamegi nn. asümptootilise valemi sageduse tõenäosuse leidmiseks. Küllalt suure katsete arvu korral (praktiliselt $n \geq 20$) võib juba kasutada seost

$$\sqrt{npq} \cdot P_{n,m} \approx \varphi(t).$$

Saame asümptootilise valemi

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

kus t on sageduse m normeeritud hälve

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

ja funktsioon $\varphi(t)$ on normaaljaotuse tihedus

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

Funktsioon $\varphi(t)$ on tabuleeritud (vt. lisa I).

K o k k u v ô t e. Tõenäosus $P_{n,m}$ arvutatakse küllalt suure katsete arvu korral järgmiselt:

1) leitakse sagedusele m vastav normeeritud hälve

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

2) leitakse saadud t väärtusele vastav $\varphi(t)$ väärtus tabelist, kusjuures $\varphi(t) = \varphi(-t)$;

3) leitakse otsitav tõenäosus

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

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

1. n ä i d e. Praakdetailide tootmise tõenäosus on

0,005. Kui tõenäone on, et 10000-st detailist on praakdetaille 40?

Lähteandmetest $p = 0,005$, $n = 10000$, $m = 40$ leiame

$$q = 1 - p = 0,995; \sqrt{npq} = \sqrt{10000 \cdot 0,005 \cdot 0,995} = 7,05;$$

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

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

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

Bernoulli valemiga arvutades saaksime

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

Ligikaudse väärtuse viga $0,0206 - 0,0197 = 0,0009$ moodustab 4,5 % täpsest väärtusest, mis praktika seisukohalt ei ole olulise tähtsusega. Kui p erineb nullist rohkem, siis on viga veelgi väiksem.

6. Normaaljaotuse tihedus

Normaaljaotuse tõenäosuse tiheduseks nimetatakse üldjuhul funktsiooni

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - m_0)^2}{2\sigma^2}}.$$

Suurust X , mille tõenäosuse tihedus on antud kujuga, nimetatakse normaaljaotusega suuruseks. Võib veenduda, et parameetrite m_0 ja σ tähendus on järgmine:

$$m_0 = E(X); \quad \sigma^2 = D(X).$$

Normaaljaotusega suuruse määramiseks on vaja teada kaht parameetrit: keskväertust ja standardhälvet. Asjaolu, et juhuslik suurus X on normaaljaotusega, mille parameetrid on m_0 ja σ , märgitakse lühidalt kujul $X \sim N(m_0, \sigma)$.

Tähtsaks erijuhuks on normaaljaotus $X \sim N(0,1)$, s.t. kui $m_0 = 0$ ja $\sigma = 1$. Seda nimetatakse normaliseeritud nor-

maaljaotuseks. Vastav tõenäosuse tihedus

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

on tabuleeritud. Tihedusfunktsiooni $\varphi(t)$ graafikut nimetatakse normaalkõveraks ehk Gaussi kõveraks. Gaussi kõvera võrrand on

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

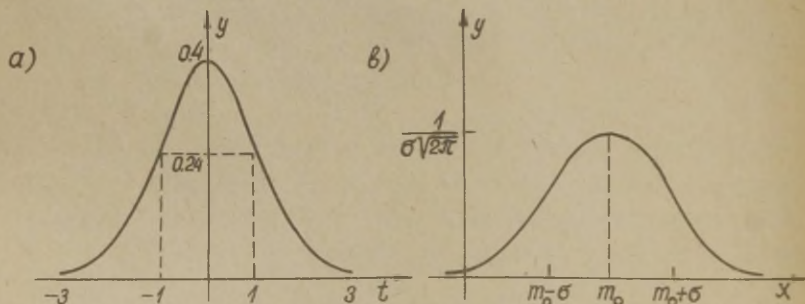
Tihedusfunktsiooni $\varphi(t)$ uurimisel kõrgema matemaatika kursuses selgus, et

1) funktsiooni maksimum on punktis $t = 0$; $y_{\max} = \frac{1}{\sqrt{2\pi}} = 0,3989$;

2) käänupunktide koordinaadid on $(\pm 1; 0,2420)$; kõver on kumer käänupunktide vahel;

3) kõver on sümmeetriline y -telje suhtes, s.t. $\varphi(-t) = \varphi(t)$ (sellepärast ongi $\varphi(t)$ tabelis ainult positiivsed argumendid);

4) abstsissstelg on kõvera asümptoodiks, kusjuures $\varphi(\pm 3)$ on praktiliselt null (täpsemalt $\varphi(\pm 3) = 0,0044$). Gaussi kõver on joonisel 15a.



Joonis 15.

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

1) funktsiooni maksimum asub keskväärtusele vastavas punktis $x = m_0$ ja $y_{\max} = \frac{1}{\sqrt{2\pi}}$;

2) käänupunktid asuvad keskväärtusest kahel pool kau-

gusel σ , s.t. et käänpunktide abstsissid on $x = m_0 \pm \sigma$.
 Käänpunktide vahel on kõver kumer, väljaspool nõgus;

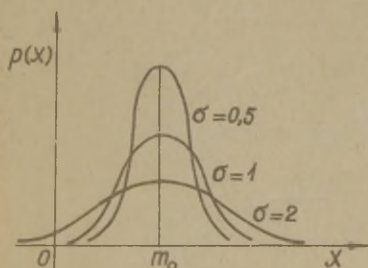
3) kõver on sümmeetriline sirge $x = m_0$ suhtes;

4) kõvera asümptoodiks on abstsissitelg.

Üldise normaaljaotuse tiheduse graafik on joonisel 15b.

Keskväertuse m_0 muutumine jääva σ korral nihutab normaalkõverat abstsissstelje sihis vasakule või paremale. Kui muutub standardhälve σ keskväertuse m_0 muutumatuks jäädes, siis muutub normaalkõvera kuju, kuid nii, et kõvera ja abstsissstelje vaheline pindala on ikka 1. Joonisel 16 on esitatud normaalkõverad kolme erineva standardhälbe korral; arvulised andmed on tabelis

x	m_0	$m_0 \pm 0,5$	$m_0 \pm 1$	$m_0 \pm 2$
$y(\sigma = 1)$	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 σ , seda suurem on y_{\max} ja järsem on kõver ning seda rohkem kõvera all olevast pindalast on keskväertuse m_0 ümbruses. Kui σ kasvab, siis kõver muutub järjest lamemaks ning pindala jaotub abstsissstelje sihis laiemalt.

Üldise normaaljaotuse tiheduse $p(x)$ väärtusi saab arvutada normaliseeritud jaotuse

tiheduse kaudu järgmiselt:

1) leida väärtusele x vastav normaliseeritud hälve

$$t = \frac{x - m_0}{\sigma};$$

2) tabelist leida $\varphi(t)$;

$$3) \quad p(x) = \frac{1}{\sigma} \varphi(t).$$

N ä i d e. Leida normaaljaotusega suuruse $X \sim N(50,4)$ tõenäosuse tihedus kohal $x = 45$.

Normeeritud hälve

$$t = \frac{45 - 50}{4} = -1,25;$$

$$\varphi(-1,25) = \varphi(1,25) = 0,1826.$$

Seega

$$p(45) = \frac{1}{4} \cdot 0,1826 = 0,04565.$$

7. Normaalne jaotusfunktsioon.

Laplace'i funktsioon

Normaaljaotusega suuruse jaotusfunktsiooni nimetatakse ka normaalseks jaotusfunktsiooniks.

Pideva juhusliku suuruse jaotusfunktsioon avaldub tõenäosuse tiheduse kaudu (§ 2, p. 7):

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

Üldise normaaljaotuse jaotusfunktsioon esitub kujul

$$F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x - m_0)^2}{2\sigma^2}} dx;$$

normaliseeritud normaaljaotuse jaotusfunktsioon

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

Viimane funktsioon on tabuleeritud (vt. lisa 4).

Geomeetriliselt vastab jaotusfunktsioonile $F_x(x)$ tiheduse kõvera alla jääv pindala abstsissini x (joonisel 17a ühekordselt viirutatud ala).

Kuna normaalne jaotusfunktsioon ei avaldu elementaar-

funktsioonides, siis toimub tema väärtuste arvutamine tabuleeritud jaotusfunktsiooni väärtuste kaudu. Lähtesuuruselt x läheme üle normeeritud hälbele t , mis on normaliseeritud normaaljaotusega suurus. Tehes $F_x(x)$ avaldises muutuja vahetuse

$$t = \frac{x - m_0}{\sigma}; \quad dx = \sigma dt,$$

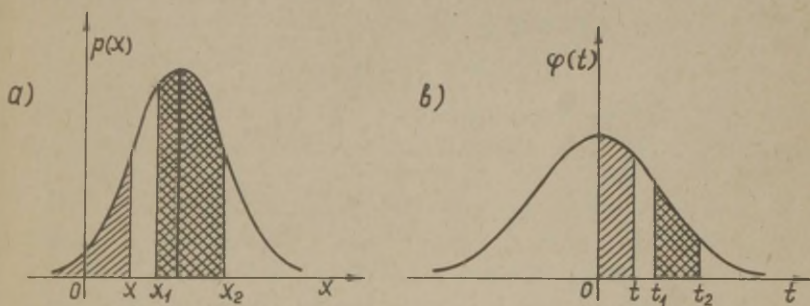
saame

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

Jõudsime tulemuseni

$$F_x(x) = F(t), \quad \text{kus } t = \frac{x - m_0}{\sigma}.$$

Lisas 4 on jaotusfunktsiooni väärtused antud muutuja t vahemikus $-3,9 \leq t \leq 3,9$. Praktiliselt võib võtta väärtustel $t < -3$ $F(t) \approx 0$ ja $t > 3$ korral $F(t) \approx 1$.



Joonis 17.

Normaaljaotuse $X \sim N(0,1)$ jaotusfunktsiooni asemel kasutatakse sageli Laplace'i funktsiooni

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt = \int_0^t \varphi(t) dt$$

või selle kahekordset väärtust

$$\bar{\varphi}(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt = 2 \varphi(t) .$$

Geomeetriliselt vastab Laplace'i funktsiooni absoluutväärtusele Gaussi kõvera alune pindala ordinaatteljest abstsissini t (joonis 17b).

Jaotusfunktsiooni ja Laplace'i funktsiooni vahel kehtib seos

$$F(t) = 0,5 + \varphi(t) .$$

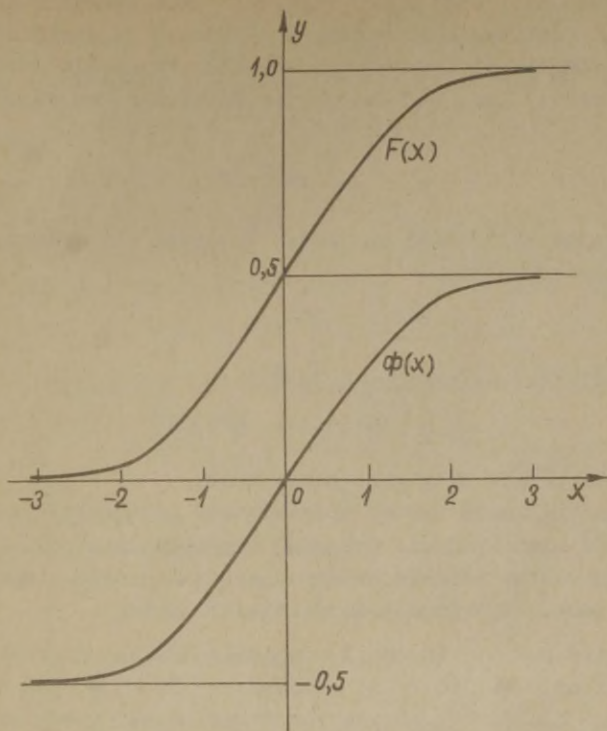
Tõepoolest, funktsiooni $\varphi(t)$ graafiku alune pindala ordinaatteljeni on 0,5 (kogu pindala on 1!), ordinaatteljest edasi lisanduv pindala on võrdne $\varphi(t)$ -ga.

Laplace'i funktsioon on tabuleeritud. Lisades 2 ja 3 on esitatud vastavalt $\varphi(t)$ ja $\bar{\varphi}(t)$ tabelid. Funktsiooni väärtused on toodud ainult argumentide positiivsete väärtuste jaoks. Laplace'i funktsioon on paaritu ja negatiivsete argumentide korral kasutatakse seost

$$\varphi(-t) = -\varphi(t) .$$

Jaotuskõvera $y = \varphi(t)$ sümmeetria tõttu jaotab ordinaattelg jaotuskõvera aluse pindala pooleks ja $F(0) = 0,5$ ning siis $\varphi(0) = 0$. Kui $t < 0$, siis $F(t) < 0,5$ ning $\varphi(t) < 0$. Kui $t \rightarrow -\infty$, siis $F(t) \rightarrow 0$ ja $\varphi(t) \rightarrow -0,5$; kui $t \rightarrow +\infty$, siis $F(t) \rightarrow 1$ ja $\varphi(t) \rightarrow 0,5$. Praktiliselt juba $t = 3$ korral $\varphi(3) \approx 0,5$. Sellepärast polegi sageli Laplace'i funktsiooni tabelis toodud $t = 3$ ületavaid väärtusi.

Normaliseeritud normaaljaotuse jaotusfunktsiooni ja Laplace'i funktsiooni graafikud on joonisel 18.



Joonis 18.

8. Normaaljaotusega suuruse
antud vahemikku sattumise
tõenäosus

Juhusliku suuruse antud vahemikku (x_1, x_2) sattumise tõenäosus on võrdne jaotusfunktsiooni väärtuste vahega vahemiku otspunktides.

Normaaljaotusega suuruse $X \sim N(m_0, \sigma)$ korral ei saa vastavat eeskirja

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

vahetult kasutada, sest $F_x(x)$ väärtused pole otseselt leitavad. Appi võetakse kas suuruse $X \sim N(0,1)$ jaotusfunktsiooni või Laplace'i funktsiooni tabelid. Eelnevalt tuleb leida vahemikule (x_1, x_2) vastav normeeritud hälbe vahemik (t_1, t_2) :

$$t_{1,2} = \frac{x_{1,2} - m_0}{\sigma} .$$

Eelmise punkti valemitele tuginedes kasutame ühte järgmistest arvutuseeskirjadest

$$P(x_1 < X < x_2) = \begin{cases} F(t_2) - F(t_1), \\ \phi(t_2) - \phi(t_1), \\ \frac{1}{2} [\bar{\Phi}(t_2) - \bar{\Phi}(t_1)] \end{cases}$$

sõltuvalt käepärast olevatest tabelitest.

Geomeetriliselt vastab tõenäosusele $(P(x_1 < X < x_2))$ jaotuskõvera alune pindala vahemiku otspunktide vahel. Joonisel 17 on vastav pindala kahekordselt viirutatud - 17a muutuja x jaoks, 17b normeeritud hälbe t jaoks.

1. n ä i d e. Normaaljaotusega juhusliku suuruse X keskväertus on $m_0 = 168$ ja standardhälve $\sigma = 5,9$. Kui tõenäone on, et juhusliku suuruse väärtused asuvad vahemikus $(160, 180)$?

Leiame normeeritud hälbed

$$t_1 = \frac{160 - 168}{5,9} = -1,536; \quad t_2 = \frac{180 - 168}{5,9} = 2,034.$$

Tabelist leiame (lisa 2) lineaarset interpolatsiooni kasutades

$$\begin{aligned} \phi(2,034) &= 0,4790; \quad \phi(-1,356) = -\phi(1,356) = \\ &= -0,4125. \end{aligned}$$

Otsitav tõenäosus

$$P(160 < X < 180) = 0,4790 - (-0,4125) = 0,8915.$$

Keskmiselt 89 % suuruse X väärtustest asub 160 ja 180 vahel.

9. Keskväärtuse suhtes sümmeetriliste hälvete tõenäosus. „Kolme sigma” reegel

Praktika seisukohalt on tähtis ülesanne: kui tõenäone on, et normaaljaotusega suuruse hälbed keskväärtusest mõlemale poole ei ületa antud arvu Δ . Tuleb leida tõenäosus

$$P(m_0 - \Delta < X < m_0 + \Delta) = P(|X - m_0| < \Delta).$$

Tõketele $m_0 \pm \Delta$ vastavad normeeritud hälbed on

$$t_{1,2} = \frac{(m_0 \pm \Delta) - m_0}{\sigma} = \pm \frac{\Delta}{\sigma}.$$

Otsitav tõenäosus avaldub jaotusfunktsiooni kaudu

$$P(|X - m_0| < \Delta) = F\left(\frac{\Delta}{\sigma}\right) - F\left(-\frac{\Delta}{\sigma}\right);$$

Laplace'i funktsiooni kaudu

$$P(|X - m_0| < \Delta) = \Phi\left(\frac{\Delta}{\sigma}\right) - \Phi\left(-\frac{\Delta}{\sigma}\right) = \bar{\Phi}\left(\frac{\Delta}{\sigma}\right)$$

ehk

$$P(|X - m_0| < \Delta) = 2 \cdot \Phi(t_\Delta) = \bar{\Phi}(t_\Delta),$$

kus

$$t_\Delta = \frac{\Delta}{\sigma}.$$

N ä i d e. Kui tõenäone on, et normaaljaotusega suuruse hälbe absoluutväärtus ei ületa standardhälbe kordset $\Delta = k\sigma$?

Normeeritud hälve

$$t_\Delta = \frac{\Delta}{\sigma} = \frac{k \cdot \sigma}{\sigma} = k.$$

Omandagu k väärtused 1, 2, 3:

- $k = 1 \quad P(|X - m_0| < \sigma) = \bar{\Phi}(1) = 0,6826 \text{ (68,26 \%)};$
 $k = 2 \quad P(|X - m_0| < 2\sigma) = \bar{\Phi}(2) = 0,9545 \text{ (95,45 \%)};$
 $k = 3 \quad P(|X - m_0| < 3\sigma) = \bar{\Phi}(3) = 0,9993 \text{ (99,93 \%)}.$

Teiste sõnadega: normaalkõvera alusest pindalast asut väärtuste $m_0 \pm \sigma$ vahel 68,26 %, $m_0 \pm 2\sigma$ vahel 95,45 % ja $m_0 \pm 3\sigma$ vahel 99,93 % ehk 68,26 % juhusliku suuruse väärtustest satuvad keskvaärtusele lähemale kui ühekordne standardhälve jne. Ainult tühine osa (0,07 %) võimalikest väärtustest hälbib keskvaärtuse suhtes rohkem kui kolmekordse standardhälbe võrra kummalegi poole.

Tulemus ongi tuntud „3 σ ” reegli nime all: praktiliselt võib olla kindel, et normaaljaotusega suurus ei hälbi keskvaärtuse suhtes rohkem kui kolmekordse standardhälbe võrra kummalegi poole.

10. S ü n d m u s e s a g e d u s e a n t u d v a h e m i k k u s a t t u m i s e t õ e - n ä o s u s

Käesoleva paragrahvi alguses lahendasime ülesanda 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

- 10 000 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” - $a \leq m \leq b$ ja selle sündmuse tõenäosuse tähistame $P(a \leq m \leq b)$.

Sündmust $a \leq m \leq b$ võib 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

$$P(a \leq m \leq b) = \sum_{m=a}^b P_{n,m} .$$

N ä i d e. Telestudios on 5 kaamerat, mis töötavad üksteisest sõltumatult. 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

$$P_{5,0} = 0,4^5 = 0,01024; \quad P_{5,1} = \frac{5!}{1! 4!} 0,6^1 \cdot 0,4^4 = 0,0768 ;$$

$$P_{5,2} = \frac{5!}{2! 3!} 0,6^2 \cdot 0,4^3 = 0,2304 ;$$

$$P_{5,3} = \frac{5!}{3! 2!} 0,6^3 \cdot 0,4^2 = 0,3456 \text{ ja}$$

$$P(0 \leq m \leq 3) = 0,01024 + 0,0768 + 0,2304 + 0,3456 = \\ = 0,6630,$$

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 võtte väga aeganõudvaks. Kasutame ligikaudset valemit otsitava tõenäosuse leidmiseks. Käesoleva paragrahvi 5. punktis selgitasime, et katsete arvu n kasvades läheneb binoomjaotus normaaljaotusele. Normaaljaotuse puhul aga leiame juhusliku suuruse antud vahemikku sattumise tõenäosuse Laplace'i või jaotusfunktsiooni abil. Sageduselt m läheme üle normeeritud hälbele

$$t = \frac{m - E(m)}{\sigma(m)} = \frac{m - np}{\sqrt{npq}},$$

seejärel leiame m tõketele a ja b vastavad suuruse t tõkkes t_1 ja t_2 ning

$$P(a \leq m \leq b) \approx \phi(t_2) - \phi(t_1) = F(t_2) - F(t_1),$$

mida nimetatakse Laplace'i integraalvalemiks.

Küllalt suure katsete arvu n (iseegi kui $n \geq 20$) korral on viga väga väike.

K o k k u v ô t e. 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

$$P(a < m < b) = \begin{cases} \phi(t_2) - \phi(t_1); \\ F(t_2) - F(t_1). \end{cases}$$

11. T ü ü p ü l e s a n d e i d

Käesolevas punktis selgitame, kuidas binoom- ja normaaljaotuse jaoks tuletatud juhusliku suuruse antud vahemikku sattumise tõenäosuse arvutamise eeskirju kohandada mitmesuguste praktikas esinevate ülesannete lahendamiseks.

1. t ü ü p. 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

$$P = \phi(t_2) - \phi(t_1).$$

N ä i d e. Poisslapse sündimise tõenäosus on 0,515. Kui tõenäone on, et iga 1000 vastsündinu hulgas on poisslas-

te arv 455 ja 555 vahel?

Lähteandmed: $n = 1000$, $p = 0,515$, $a = 455$, $b = 555$;

$q = 1 - p = 0,485$.

Leida $P = P(455 \leq m \leq 555)$.

Selleks arvutame normeeritud hälbe väärtused

$$t_1 = \frac{a - np}{\sqrt{npq}} = \frac{455 - 0,515 \cdot 1000}{\sqrt{1000 \cdot 0,515 \cdot 0,485}} = -3,80;$$

$$t_2 = \frac{b - np}{\sqrt{npq}} = \frac{555 - 0,515 \cdot 1000}{\sqrt{1000 \cdot 0,515 \cdot 0,485}} = 2,53.$$

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.

2. t ü ü p. Kui tõenäone on, et sagedus m ei hälbi oma keskväärtusest $E(m)$ rohkem kui antud arvu Δ võrra?

Antud on Δ , n , p , $q = 1 - p$;

leida tuleb $P(|m - np| < \Delta)$.

Kuna sagedus m allub binoomjaotusele, mis on lähedane normaaljaotusele, siis saab kasutada eeskirja

$$P(|m - np| < \Delta) = \bar{\Phi}(t_{\Delta}), \text{ kus } t_{\Delta} = \frac{\Delta}{\sigma(m)} = \frac{\Delta}{\sqrt{npq}}.$$

N ä i d e. 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 %), $\Delta = 10$; $q = 1 - 0,2 = 0,8$.

Leida $P = P(|m - np| < 10)$.

Templita toodete keskmine 400 toote hulgas on keskväärtuseks

$$E(m) = np = 0,2 \cdot 400 = 80.$$

Nüüd saame

$$t = \frac{10}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = \frac{10}{10 \cdot 0,8} = 1,25;$$

$$P(|m - 80| < 10) = \bar{\Phi}(1,25) = 0,7888.$$

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.

3. t ü ü p. Sooritatakse n sõltumatut katset ja igal katsel on uuritava sündmuse tõenäosus p. Kui tõenäone on, et sündmuse suhteline sagedus w ei erine tõenäosusest p rohkem kui arvu Δ võrra?

Antud on n, p, Δ , $q = 1 - p$.

Võrdse pikkusega katseseeriates on sündmuse sagedus m harilikult erinev ja siis on ka w juhuslik suurus. Oleme huvitatud, kui tõenäone on, et $|w - p| < \Delta$. Binoomjaotuse käsitlemisel selgus, et $E(w) = p$ ja $\sigma(w) = \sqrt{\frac{pq}{n}}$. Analoogiliselt 2. ülesannete tüübile

$$P(|w - p| < \Delta) = \bar{\Phi}(t_{\Delta}), \text{ kus } t_{\Delta} = \Delta \sqrt{\frac{n}{pq}}.$$

N ä i d e. Tõenäosus mittestandardse detaili tootmiseks on 0,1. Kui tõenäone on, et 400 detaili korral mittestandardsete detailide suhteline sagedus ei erine 0,1-st rohkem kui 0,03 võrra?

Lähteandmed: $n = 400$, $p = 0,1$, $\Delta = 0,03$; $q = 1 - 0,1 = 0,9$.

Leida tuleb $P(|\frac{m}{400} - 0,1| < 0,03)$.

Selleks määrame argumendi

$$t_{\Delta} = 0,03 \cdot \sqrt{\frac{400}{0,1 \cdot 0,9}} = 2$$

ja

$$P(|\frac{m}{400} - 0,1| < 0,03) = \bar{\Phi}(2) = 0,9544.$$

Kui kontrollida 400 detailist koosnevaid partiisid kor-

duvalt, siis umbes 95 % partiides ei erine mittestandardsete detailide suhteline sagedus 0,1-st rohkem kui 0,03 võrra ehk asub piirides 0,07 kuni 0,13. Tulemust võib tõlgendada ka nii: 95 % 400 detailist koosnevatest partiidest on sellised, kus mittestandardsete detailide arv asub $0,07 \cdot 400 = 28$ ja $0,13 \cdot 400 = 52$ vahel.

4. t ü ü p. Teades sündmuse tõenäosust üksikkatsel, leida suhtelise sageduse w maksimaalne hälve tema keskvaartuse p suhtes antud tõenäosusega.

Suuruse X väärtuste maksimaalset hälvet oma keskvaartuse $E(X)$ suhtes nimetatakse sageli piirveaks antud tõenäosusega P . Piirviga Δ määrab suuruse X usalduspiirid $E(X) \pm \Delta$ ja nendes piirides asumise tõenäosust nimetatakse usaldusnivooks.

Antud on $n, p, P(|w - p| < \Delta) = P; q = 1 - p$.

Leida tuleb suhtelise sageduse piirviga Δ .

Antud usaldusnivoo P järgi määrame Laplace'i funktsiooni tabelist argumendi t väärtuse, kasutades võrdust $P = \bar{\Phi}(t) = 2\phi(t)$.

Piirvea Δ ja t vaheline sõltuvus esitab kujul

$$t_{\Delta} = \Delta \sqrt{\frac{n}{pq}},$$

kust saab otsitava

$$\Delta = t_{\Delta} \cdot \sqrt{\frac{pq}{n}}.$$

N ä i d e. Automaattööpingil valmistatud toodangust on keskmiselt 10 % praaki. Leida praakdetailide suhtelise sageduse piirviga keskmise praagiprotsendi suhtes usaldusnivoo 0,992 10 000 detailist koosnevas partiis.

Lähteandmed: $n = 10\ 000, p = 0,1$ (10 %),

$$P(|\frac{m}{10000} - 0,1| < \Delta) = 0,992.$$

Laplace'i funktsiooni tabelist leiame argumendi väärtuse $t = 2,65$, mis vastab funktsiooni väärtusele $\bar{\Phi}(t) = 0,992$ (või $2\phi(t) = 0,992$). Nüüd arvutame piirvea

$$\Delta = 2,65 \sqrt{\frac{0,1 \cdot 0,9}{10000}} = 0,00795 \approx 0,008.$$

Keskmiselt 99 % 10 000 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.

5. t ü ü p. Mitu katset tuleks sooritada, et antud tõenäosusega (usaldusnivooga) garanteerida: sündmuse suhteline sagedus ei erine sündmuse tõenäosusest üksikkatsel rohkem kui antud arvu Δ võrra?

Antud on p, q, Δ , $P(|w - p| < \Delta) = P$, leida tuleb katsete arv n .

Ülesanne on analoogiline eelmisega, ainult otsitavaks on Δ asemel n . Leiame jälle võrduse $P = \bar{\Phi}(t)$ abil argumendi t . Kuna t ja n on seotud võrdusega

$$t = \Delta \sqrt{\frac{n}{pq}},$$

siis saame

$$n = \frac{t^2 pq}{\Delta^2}.$$

N ä i d e. Mitu korda on vaja visata münti, et vapi suhteline sagedus ei erineks selle tõenäosusest üksikviskel rohkem kui 0,01 võrra usaldusnivooga 1) $P = 0,6$, 2) $P = 0,954$?

Lähteandmed:

1) $\Delta = 0,01$, $P = 0,6$; varasemast teame, et $p = 0,5$.

Leida n .

Kuna $P = 0,6$, siis leiame tabelist, et $t = 0,84$.

Minimaalne visete arv

$$n = \frac{0,84^2 \cdot 0,5^2}{0,01^2} = 1764.$$

2) $P = 0,954$. Siis $t = 2$ ja $n = 10\ 000$.

Usaldusnivoo suurendamine sama täpsuse korral nõuab katsete arvu suurendamist.

12. Poissoni jaotus

Korduvatel katsetel leiame sündmuse m korda toimumise tõenäosuse $P_{n,m}$ Bernoulli valemi abil või suure katsete arvu n puhul ligikaudsest valemist, mis tähendab binoomjaotuse 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 binoomjaotuse lähendamine normaaljaotusega anda üpris ebatäpseid tulemusi. Saab näidata, et väikese p ja küllalt suure n korral läheneb

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

suurusele

$$P_m = \frac{\mu^m}{m!} e^{-\mu}, \quad \text{kus } \mu = np.$$

Viimase valemiga määratud juhusliku suuruse m tõenäosuste jaotust nimetatakse Poissoni (puassoon) jaotuseks. Suurus μ on sageduse m keskväärtnus, olles ainukeseks jaotuse parameetriks.

Poissoni jaotust rakendatakse 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 üksikatsel (väga väikeses ajavahemikus) on väike ja ajavahemiku lühendamisel väheneb, kuid nii, et $\mu = np = \text{const}$.

1. n ä i d e. Kangur teenindab 800 värtnat. Lõnga katkemise tõenäosus igal värtnal 5 min. jooksul on $0,005$. Leida tõenäosus lõngade katkemise arv ja selle tõenäosus.

Tõenäosus katkemiste arv $\mu = np = 800 \cdot 0,005 = 4$. Leiame tõenäosuse täpsest Bernoulli valemist, samuti ligikaudsete valemitega, kasutades normaaljaotust ja Poissoni jaotust.

Bernoulli valemi järgi

$$P_{800,4} = \frac{800!}{4! 796!} \cdot 0,005^4 \cdot 0,995^{796} = 0,1945;$$

Poissoni valemist

$$P_4 = \frac{4^4}{4!} \cdot e^{-4} = \frac{256}{24} \cdot 0,0183 = 0,1954.$$

Laplace'i lokaalteoreemile tugineva valemi järgi

$$P_{800,4} = \frac{1}{\sqrt{800 \cdot 0,005 \cdot 0,995}} \cdot \varphi(0) = 0,2.$$

Mäeme, et Poissoni jaotuse järgi arvutatud tulemuse viga (0,0009) on väiksem normaaljaotuse järgi arvutatud tulemuse veast (0,0055).

2. n ä i d e. Telefonikeskjaam saab tunnis keskmiselt k väljakutset. Leida tõenäosus m väljakutse saamiseks 1 minuti jooksul.

Ühes minutis saadavate väljakutsete arvu keskvaartuseks loeme väärtuse $\mu = \frac{k}{60}$. Tõenäosus m väljakutse saamiseks minutis oleks Poissoni valemi järgi

$$P_m = \frac{\left(\frac{k}{60}\right)^m}{m!} \cdot e^{-\frac{k}{60}}.$$

3. n ä i d e. Suures raamatuhoidlas on 100 elektripirni, nendest igaihe põlemise tõenäosus on 0,02. Koostada sisselülitatud pirnide sageduse m jaotustabel Poissoni valemi järgi.

Leiame $\mu = np = 100 \cdot 0,02 = 2$ ja $e^{-\mu} = e^{-2} = 0,135335$. Sageduse m väärtustele vastavad tõenäosused arvutame valemist

$$P_m = \frac{2^m}{m!} \cdot e^{-2} = 0,135335 \cdot \frac{2^m}{m!}.$$

Nii saaksime näiteks

$$P_3 = 0,135335 \cdot \frac{2^3}{3!} = 0,18044.$$

Tulemused on antud tabelis, mille alumises reas on

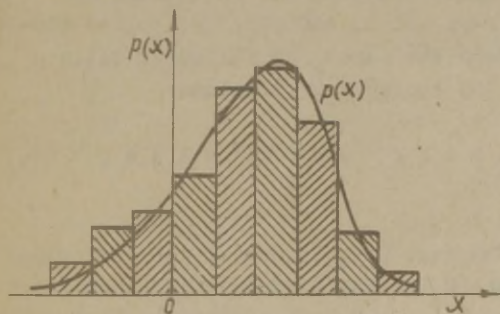
võrdluseks toodud ka Bernoulli valemiga leitud täpsed tulemused. Vastavate väärtuste võrdlemisel näeme, et kooskõla on väga hea. Tabelis puuduvate sageduste (9 jne.) tõenäosused on antud täpsuse piires nullid.

m	0	1	2	3	4	5	6	7	8
$P_{n,m}$	0,135	0,271	0,271	0,180	0,090	0,036	0,012	0,003	0,001
	0,133	0,271	0,273	0,182	0,090	0,035	0,011	0,003	0,001

13. Statistiliste jaotuste tasandamisest

Matemaatilise statistika üheks tähtsaks ülesandeks on juhusliku suuruse uurimisel saadud statistiliste katseandmete järgi suuruse jaotusseaduse määramine. Katsetulemused ei kirjelda uuritavat suurust kunagi täpselt, katseandmetes sisaldub juhuslikkusest ning katseandmete piiratud mahust tingituna ebatäpsusi ja hälbeid suuruse tõelisest jaotusseadusest. Uurija ülesandeks on suurusele omaste seaduspärasuste väljaselgitamine, suuruse empiirilise jaotuse „puhasdamine“ kõrvalisest, ebaolulisest, mittetüüpilisest.

Statistilise jaotuse tasandamise all mõistetakse selle jaotuse lähendamist teatud mõttes hea teoreetilise jaotusseadusega. Antud empiirilise jaotuse põhjal tuleb määrata suurust kirjeldav teoreetiline jaotusseadus kas tõenäosuse tiheduse $p(x)$ või jaotusfunktsiooni $F(x)$ kujul (joonis 19). Püstitatud ülesanne on analoogiline kõrgema matemaatika



Joonis 19.

kursuses käsitletud 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).

Teisel etapil tuleb määrata valitud teoreetilise jaotuse karakteristikud nii, et statistiline ja teoreetiline jaotus oleksid võimalikult heas kooskõlas. Tihti toimub see nii, et osa teoreetilise jaotuse karakteristikuid (esmajoones momendid) valitakse võrdsed statistilise jaotuse vastavate karakteristikutega.

Kolmas etapp seisneb järgnevas. Olgu teoreetiline jaotus koostatud. Alati esineb lahkuminekuid empiirilise ja teoreetilise jaotuse vahel. Tuleb hinnata, kas need lahkuminekuid on tingitud juhuslikest asjaoludest, mis tulenevad katseandmete piiratud mahust, või on nad olulised, olles tingitud tasandava funktsiooni ebaõnnestunud valikust. Selle hindamiseks kasutatakse nn. kooskõla kriteeriume.

14. T a s a n d a m i n e n o r m a a l j a o t u s e s e g a

Suurte arvude seadus (vt. § 5, Ljapunovi teoreem) kindlatab, et üheks sagedamini esinevaks jaotuseks on normaaljaotus. Sellepärast käsitleme natuke lähemalt statistilise jaotuse tasandamist normaaljaotusega.

Sageli saab statistilise jaotuse lähedust normaaljao-

tusele otsustada suhtelise sageduse w polügooni (histogrammi) kuju järgi. Kui see on küllalt sümmeetriline keskmise suhtes ega ole keskosas eriti terava tipuga või jälle liiga lame, siis võib seda lugeda küllalt lähedaseks normaalkõverale ja statistilist jaotust saab lähendada normaaljaotusega.

Sümmeetrilisest jaotusest kõrvalekaldumise astet mõõdetakse nn. asümmeetria kordajaga

$$\alpha = \frac{\mu_3}{\sigma^3}$$

Joonisel 20a on näha nn. vasak- ja parempoolne asümmeetria. Sümmeetrilise jaotuse puhul on 3. järku moment $\mu_3 = 0$ ja $\alpha = 0$. Vasakpoolse asümmeetria puhul on $\alpha < 0$ ja parempoolisel $\alpha > 0$.

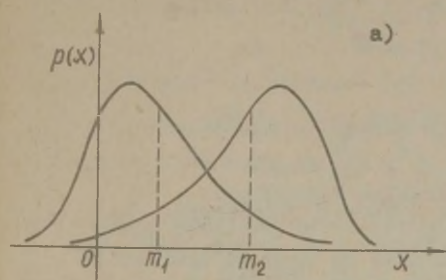
Joonisel 20b on esitatud normaalkõver (1), sellest teravamaga tipuga jaotuskõver (2) ja lamedam kõver (3). Jaotuskõvera tipu teravuse astet normaalkõveraga võrreldes mõõdetakse nn. ekstsessiga

$$\varepsilon = \frac{\mu_4}{\sigma^4} - 3.$$

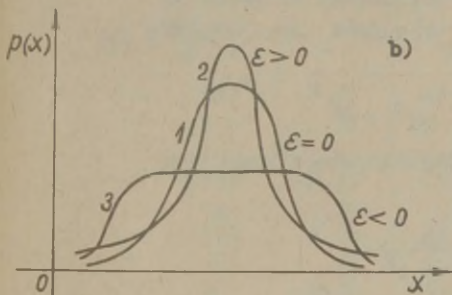
Normaalkõvera korral on 4. järku moment $\mu_4 = 3\sigma^4$ ja $\varepsilon = 0$. Normaalkõveraga

võrreldes on teravamaga tipuga kõvera jaoks $\varepsilon > 0$, lamedama tipu puhul $\varepsilon < 0$.

Eelöeldust järeldub, et kui statistilise jaotuse korral on asümmeetria α ja ekstsess ε küllalt lähedased nul-



a)



b)

Joonis 20.

lile, siis sobib tasandamiseks normaalkõver.

1. n ä i d e. Leida α ja ε eelmise paragrahvi 11. p. näite andmetel. Sain $\mu_3 = 33,91$, $\mu_4 = 7299,7$ ja $\sigma = 50,42$. Seega

$$\alpha = \frac{33,91}{356} \approx 0,095; \quad \varepsilon = \frac{7299,7}{2542} = -0,09.$$

Olgu tasandavaks jaotuseks normaaljaotus, s.t. teoreetilise jaotuse tõenäosuse tihedus valitakse kujul

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - m_0)^2}{2\sigma^2}}.$$

Normaaljaotuse tihedus sõltub kahest parameetrist: keskvaartusest m_0 ja dispersioonist σ^2 , mis on vastavalt võrdsed suuruse 1. ja 2. järku tsentraalse momendiga.

Nn. momentide meetodi kohaselt on parimaks tasandavaks jaotuseks selline, mille momendid on võrdsed statistilise jaotuse vastavate momentidega. Statistilise jaotuse 1. ja 2. järku momendid on vastavalt aritmeetiline keskmine \bar{x} ning dispersioon s^2 . Seega tuleb tasandava normaalkõvera parameetrid m_0 ja σ^2 valida

$$m_0 = \bar{x}, \quad \sigma^2 = s^2$$

ning tasandava jaotuse tihedusfunktsioon omab kuju

$$p(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x - \bar{x})^2}{2s^2}}.$$

2. n ä i d e. Leida § 3, p. 12 3. näite andmetel tasandava normaaljaotuse tõenäosuse tihedus ning võrrelda suuruse X väärtuste antud intervallidesse sattumise relatiivseid sagedusi ning tõenäosuse tiheduse abil leitud intervallidesse sattumise tõenäosusi.

Eespool leidsime $\bar{x} = 165,5$; $s^2 = 36,58$. Teoreetilise

jaotuse (tasandava jaotuse) tihedusfunktsioon on seega

$$p(x) = \frac{1}{6,05 \cdot \sqrt{2\pi}} e^{-\frac{(x - 165,5)^2}{2 \cdot 36,58}}$$

Momentrea korral saaksime üksikväärtuste tõenäosused valemist

$$P(X = x_1) = \frac{1}{s} \varphi(t_1), \text{ kus } t_1 = \frac{x_1 - \bar{x}}{s}.$$

Intervallitud rea korral saame intervalli sattumise tõenäosuse

$$P = P(x_1 < X < x_1 + \Delta x) = \varphi(t_1 + \Delta t) - \varphi(t_1),$$

kus x_1 ja $x_1 + \Delta x$ on intervallide otspunktid ning t_1 , $t_1 + \Delta t$ vastavad normeeritud hälbed ($i = 1, 2, \dots, n$).

Empiirilised suhtelised sagedused leitakse valemist

$$w_1 = \frac{f_1}{N} \quad (i = 1, 2, \dots, n).$$

Arvutuste käiku saab üksikasjalikumalt jälgida järgnevast tabelist. Esialgne w_1 ja p_1 võrdlemine vihjab empiiriliste ja teoreetiliste andmete heale kooskõlale. Sügavam analüüs oleks teostatav nn. „hi ruut“ kooskõla kriteeriumi abil (vt. näiteks [3]).

Intervallid	x	t	$\varphi(t)$	P	w
143-146	143	-3,72	-0,4999	0,0006	0,001
146-149	146	-3,33	-0,4993	0,003	0,002
149-152	149	-2,73	-0,4968	0,009	0,008
152-155	152	-2,23	-0,4871	0,028	0,026
155-158	155	-1,74	-0,4591	0,067	0,065
158-161	158	-1,24	-0,3925	0,122	0,120
161-164	161	-0,74	-0,2703	0,172	0,181
164-167	164	-0,25	-0,0987	0,197	0,201
167-170	167	0,25	0,0987	0,197	0,170
170-173	170	0,74	0,2703	0,122	0,120
173-176	173	1,24	0,3925	0,067	0,064
176-179	176	1,74	0,4591	0,028	0,028
179-182	179	2,23	0,4871	0,009	0,010
182-185	182	2,73	0,4968	0,003	0,003
185-188	185	3,22	0,4993	0,0006	0,001
	188	3,72	0,4999		

7

§ 5. SUURTE ARVUDE SEADUS

1. Keskmise stabiilsusest

Toimigu teatava suuruse väärtuse määramine katseliselt, kusjuures sooritatakse N katset samades tingimustes. Katsetulemused x_1, x_2, \dots, x_N on üldiselt ebatäpsed, sisaldades juhuslikest põhjustest tingituna nn. juhuslikke vigu. Seega on katsetulemused juhuslikud suurused, mida võime lugeda võrdsetel määral usaldatavateks. Määratava suuruse tõeline väärtus m_0 jääb teadmata. Katsetulemused kõiguvad väärtuse m_0 ümber ja viimane on juhuslike suuruste x_i ($i = 1, 2, \dots, N$) keskvärtus, s.t. $E(x_i) = m_0$. Võrdsetes tingimustes sooritatud katsete puhul võib ka kõigi üksikkatsete tulemuste hajuvuse (dispersiooni) lugeda samaks. Tähistame üksiktulemuse dispersiooni $D(x_i) = \sigma^2$ ($i = 1, 2, \dots, N$). Kerkib küsimus: missugust arvu tuleb katsetulemuste põhjal lugeda kõige paremaks suuruse väärtuse m_0 lähendiks? Belistame sellist lähendit, mille hajuvus m_0 suhtes on võimalikult väike.

Katsetulemuste aritmeetiline keskmine

$$\bar{x} = \frac{1}{N} \sum x_i,$$

mis on leitud juhuslike suuruste baasil, on samuti juhuslik suurus. Leiame keskmise keskvärtuse ja dispersiooni, kasutades selleks nende omadusi. Saame

$$E(\bar{x}) = E\left(\frac{1}{N} \sum x_i\right) = \frac{1}{N} \sum E(x_i) = \frac{N \cdot m_0}{N} = m_0;$$

$$D(\bar{x}) = D\left(\frac{1}{N} \sum x_i\right) = \frac{1}{N^2} \sum D(x_i) = \frac{N \cdot \sigma^2}{N^2} = \frac{\sigma^2}{N}.$$

Jõudsime tulemusteni

$$E(\bar{x}) = m_0; \quad D(\bar{x}) = \frac{\sigma^2}{N}$$

— võrdse keskväärtusega juhuslike suuruste aritmeetilise keskmise keskväärtus võrdub üksiksuuruse keskväärtusega ja keskmise dispersioon on N korda väiksem üksiksuuruse dispersioonist. Teiste sõnadega: aritmeetiline keskmine on usaldatavam üksikkatse tulemusest ja seda rohkem, mida suurema hulga juhuslike suuruste baasil on ta arvutatud.

Veelgi olulisemaid järeldusi saab teha rea teoreemide põhjal, mis tänapäeval on tuntud nn. suurte arvude seaduse nime all. Nendel teoreemidel on tähtis koht matemaatilises statistikas, kuna nad seovad tõenäosusteooria abstraktseid tulemusi katseandmetega, andes võimaluse ette näha arvata-vaid tulemusi. Suurte arvude seaduse mõnede erikujudega tutvumegi järgmistes punktides. Kõigi nende puhul on tegemist piirprotsessiga ja neid nimetatakse sageli tõenäosusteooria piirteoreemideks. Suurte arvude seaduse ühe külje esindajateks on Bernoulli ja Tšebôšovi teoreemid. Teise külje moodustab Ljapunovi teoreem, mida selle erakordse üldisuse tõttu nimetatakse ka tsentraalseks piirteoreemiks.

2. B e r n o u l l i j a T š e b ô š o v i t e o r e e m i d

Suurte arvude seaduse alla kuuluvatest teoreemidest on ajaliselt esimene Bernoulli teoreem (1713.a.). Tänapäeval nimetatakse seda ka Bernoulli suurte arvude seaduseks:

Kui korduvatel sõltumatutel katsetel on sündmuse A toimumise tõenäosus p konstantne, siis võib küllalt suure katsete arvu N korral ühele kuitahes lähedase tõenäosusega väita, et sündmuse suhtelise sageduse hälve sündmuse tõenäosuse suhtes saab absoluutväärtuse poolest väiksemaks kuitahes väikesest positiivsest arvust Δ .

Tõestamiseks lähtume eelmises paragrahvis (p. 11) tuletatud seosest

$$P(|w - p| < \Delta) \approx 2 \phi(t_{\Delta}), \quad t_{\Delta} = \Delta \cdot \sqrt{\frac{N}{pq}}.$$

Võrdus on seda täpsem, mida suurem on N (binoomjaotus lähe-

neb normaaljaotusele katsete arvu N kasvamisel). Täpse võrduse saame üleminekul piirile, lastes $N \rightarrow \infty$. Kuid $\lim_{N \rightarrow \infty} t_{\Delta} = \infty$ ja Laplace'i funktsioon argumendi tõkestamatul kasvamisel saab piirväärtuseks 0,5. Seega tõepoolest

$$\lim_{N \rightarrow \infty} P(|w - p| < \Delta) = 1,$$

mis ongi Bernoulli teoreemi matemaatiliseks väljenduseks.

Bernoulli teoreemi ei saa tõlgendada nii, et katsete arvu N piiramatul kasvamisel $w \rightarrow p$. Põhimõtteliselt võib ka suure katsete arvu korral w ja p erinevus olla suur, kuid tõenäosus selleks on väga väike. Seega võib praktiliselt olla kindel, et w ja p erinevus küllalt suure katsete arvu korral ei ole suur. Seda asjaolu väljendatakse ka kujul: N kasvades $w \rightarrow p$ tõenäosuslikult.

Bernoulli teoreemi üldistuseks on 1867.a. Tšebõšovi poolt tõestatud teoreem ehk Tšebõšovi suurte arvude seadus. Selles teoreemis toimub eelmises punktis saadud tulemuste edasine täpsustamine.

Tšebõšovi suurte arvude seadus:

küllalt suure arvu võrdsete keskväertuste ja dispersioonidega sõltumatute juhuslike suuruste puhul võib ühele kuitahes lähedase tõenäosusega väita, et nende suuruste aritmeetiline keskmine erineb ühisest keskväertusest kuitahes vähe.

Kui tähistame juhuslike suuruste x_1, x_2, \dots, x_N keskväertuse tähega m_0 ja nende aritmeetilise keskmise sümboliga \bar{x} , mis on määratud 1. punktis, siis saame Tšebõšovi suurte arvude seaduse matemaatilises sümboolikas esitada kujul

$$\lim_{N \rightarrow \infty} P(|\bar{x} - m_0| < \Delta) = 1.$$

Ka suure N korral võib \bar{x} oluliselt hälbida m_0 suhtes, kuid selle tõenäosus on väga väike. Seda juhtub väga harva ehk see on praktiliselt võimatu. Teiste sõnadega: $\bar{x} \rightarrow m_0$ tõenäosuslikult.

Tšebõšovi suurte arvude seadus on üldisem Bernoulli suurte arvude seadusest. Tšebõšovi teoreem esitatud kujul

kehtib kõigi juhuslike suuruste jaoks, mis rahuldavad eespool nimetatud tingimusi, Bernoulli teoreem käib ainult sündmuse suhtelise sageduse kohta korduvatel katsetel (w on samuti juhuslik suurus).

Bernoulli suurte arvude seaduse kehtivust on püütud korduvalt katseliselt kontrollida. Toome mõnede kontrollimiste tulemused. Buffon (büffoon) viskas 1777.a. münti 4040 korda. Vapi sagedus ja suhteline sagedus olid vastavalt $m = 2048$, $w = 0,507$ (vapi tõenäosus üksikviskel $p = 0,5$). Quetelet (ketlee) registreeris 1846.a. 20 musta ja 20 valget kuuli sisaldavast urnist kuulide võtmise tulemusi. Kuuli võeti 4096 korda, sealjuures oli valge kuuli sagedus 2066 ja suhteline sagedus 0,504. Romanovski viskas 1919. a. 20160 korda üheaegselt 4 münti ja registreeris vappide ja kirjade kõigi kombinatsioonide esinemise sagedused. Bernoulli valemist saab leida ka iga kombinatsiooni tõenäosuse (leidke need iseseisvalt!) Suhtelised sagedused ja tõenäosused on võrdluseks toodud tabelis

	4 vappi	3 v., 1 k.	2 v., 2 k.	1 v., 3 k.	4 kirja
p	0,0625	0,2500	0,3750	0,2500	0,0625
w	0,0586	0,2435	0,3761	0,2522	0,0695

3. L j a p u n o v i t e o r e e m i s i s u s t

Normaaljaotusega suurusi kasutatakse rakendustes kõige sagedamini. Vastuse küsimusele, miks see nii on, annab vene akadeemiku A. Ljapunovi poolt tõestatud teoreem. Kirjeldame selle sisu.

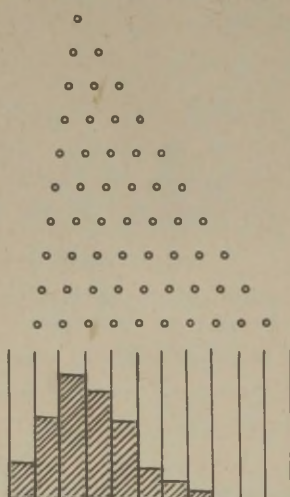
Juhuslik suurus X olgu N sõltumatu juhusliku suuruse X_i ($i = 1, 2, \dots, N$) summa $X = X_1 + X_2 + \dots + X_N$. Nende suuruste jaotusseadused võivad olla väga erinevad, kuid iga liidetava osatähtsus summaarses suuruses X olgu ühtlaselt väike. Siis läheneb summaarse suuruse X jaotus normaaljaotusele liidetavate arvu N kasvamisel.

Kui aga osutub, et mõni komponentidest X_i avaldab olu-

list mõju summaarsele suurusele X , siis ka summaarse suuruse jaotusseadus on lähedane komponendi X_i jaotusseadusele.

Ljapunovi teoreemist järeldub, et kui aritmeetiline keskmine on leitud N ühesuguse jaotusega juhusliku suuruse baasil (N üksikkatse tulemuste keskmine), siis keskmise kui juhusliku suuruse jaotuseks on küllalt suure N korral ligikaudu normaaljaotus. Praktiliselt võib keskmise jaotusseaduse lugeda juba normaalseks $N \geq 10$ korral.

Normaaljaotuse kujunemise illustratsiooniks on nn. Galtoni võre (joonis 21). Plaadi sisse on löödud naelad



malelaua süsteemis. Plaadi alumises servas on lahtrid. Kui seadme ülemises otsas olevast avast kallata sisse kuulikesi, siis jaotuvad nad lahtrites ligikaudu normaalkõvera järgi. Iga kuulike põrkub oma teel vastu teisi kuulikesi ja naelu, mis kujutavad endast juhuslikke mõjustusi. Naelte puudumisel langeksid kuulid valdavas enamuses keskmistesse lahtritesse.

Joonis 21.

4. Suurte arvude seaduse praktiline tähtsus

Praktilises tegevuses toetutakse sageli suurte arvude seadusele. Toome selle kohta mõningaid näiteid.

1. Sündmuse tõenäosuseks loetakse tema suhteline sagedus küllalt pikas katseseerias. Bernoulli suurte arvude sea-

duse järgi on harva oodata tõenäosuse ja suhtelise sageduse olulist erinevust.

2. Katseliselt määratava suuruse väärtuseks võetakse katsetulemuste aritmeetiline keskmine. Tšebõšovi suurte arvude seadus kinnitab, et praktiliselt võib küllaldase arvu katsete korral suuruse tõelise väärtuse ja aritmeetilise keskmise erinevuse lugeda tühiseks.

3. Juhusliku suuruse väärtused hälbivad keskväärtusest mõlemale poole ja aritmeetilises keskmises need hälbed tasakaalustuvad. Keskmises peaaegu ei kajastu suuruse üksikväärtuste konkreetseid iseärasused. Keskmine on vaadeldavale suurusele omase seaduspärasuse peegeldaja. Näiteks ilmastiku prognoosimisel on paljude aastate kohta leitud keskmised näitajad (temperatuur, sademete hulk, õhurõhk jne.) olulisel kohal. Tarbekaupade ja toiduainete koguste määramisel elanikkonna üksikute gruppide jaoks toetutakse ka varemate perioodide põhjal leitud keskmistele näitajatele.

4. Valimimeetod statistikas baseerub Ljapunovi teoreemile. Osutub, et näiteks toodangu kvaliteedi valikuline kontroll võimaldab teha praktiliselt niisama täpseid järeldusi kui kogu toodangu kontroll. Majanduslikult on aga tunduvalt odavam toodete väiksema koguse kontrollimine. Taimede seemnete idanevuse kontroll toimub piiratud kogusega seemnete korral ning tulemused on ülekantavad suurtele kogustele. Kogu viljaseemne idanevuse kontrollimisel polekski enam midagi külvata.

Näeme, et nn. suurte arvude seadus ei ole tähtis ainult matemaatilise teooria seisukohalt, vaid leiab ka laialdast praktilist kasutamist (kas teadlikult või ebateadlikult inimkonna kogemustele toetudes).

§ 6. VALIMITEORIA MATEMAATILISED ALUSED

1. Valimimeetod

1. n ä i d e. Ettevõtte ametiühingukomitee tahab välja selgitada töötajate korteriolusid. Selleks küsitletakse kõiki töötajaid. Küsitlusel saadud andmete põhjal saab teha vajalikke järeldusi ja leida arvulisi näitajaid.

2. n ä i d e. 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. Ütleme, et kasutati valimimeetodit (väljavõttelist meetodit), mille idee on järgmine.

Olgu tegemist mingi objektide hulgaga, milles püütakse uurida tunnust X. Harilikult ei ole võimalik uurida kõiki objekte, vaid nende hulgast võetakse teatud osa, mida uuritakse vajaliku tunnuse seisukohalt. Saadud andmete põhjal tehtud järeldusi üldistatakse kõigile samatüübilistele objektidele.

Vaatluse alla võetud üksikobjektid ehk indiviidid moodustavad valimi (väljavõtte); indiviidide hulk, mille kohta soovitakse järeldusi teha, on üldkogum. Kogumi elementide arvu nimetatakse kogumi mahuks. Üldkogumi mahu tähistame N_0 , valimil - N . Üldkogumi maht N_0 võib olla lõplik või lõpmatu.

Valimimeetodiga uuritakse üldkogumit kaudselt - valimi kaudu. Valimis peab uuritav tunnus jaotuma ligikaudu samuti nagu üldkogumis. Valimi ja üldkogumi kooskõla nimetatakse representatiivsuseks ja öeldakse, et valim esindab üldkogumit (valim on küllalt esindav).

Valimimeetodit kasutatakse laialdaselt tehnoloogiliste protsesside kontrollimisel, toodangu kvaliteedi määramisel, keemiliste ja füsioloogiliste eksperimentide korraldamisel, elanikkonna teenindamise efektiivsuse hindamisel jne.

Representatiivse valimi tegemiseks kasutatakse mitmesuguseid võtteid. Nimetagem neist olulisemaid: 1) juhuslik korduv valim, 2) juhuslik kordumatu valim, 3) mehhaaniline valim. Valimi moodustamisel tuleb silmas pidada, et elemendi sattumine üldkogumist valimisse ei tohi sõltuda uuritava tunnusest ja üldkogumi elementidel peab olema valimisse sattumiseks võrdne tõenäosus.

Juhusliku korduva valimi puhul tagastatakse element pärast uurimist üldkogumisse ja ta võib seega uuesti sattuda valimisse.

Juhuslikul kordumatul valimil valitud elemendid üldkogumisse tagastamisele ei kuulu.

Mehhaanilise valimi puhul valitakse valimisse üldkogumi mingil viisil järjestatud elemente võrdsete intervallide tagant, see viis on oma iseloomult seega samuti kordumatu valim.

Valimi moodustamisel tuleb tingimata vältida nn. jämedaid vigu, mis tekivad vaatluste hooletul tegemisel või andmete tahtlikul moonutamisel. Jämedad vead võivad muuta valimi mitteesindavaks. Ebatäpselt reguleeritud mõõteriistad või vaatlusvahendid ja mõnikord vaatleja individuaalsed iseärasused võivad tekitada süstemaatilisi vigu - resultaadid hälbivad õigest tulemusest valdavalt ühele poole. Selline olukord võib esineda näiteks kaalumisel, kui kaalude nullasend ei ole õige. Süstemaatilisi vigu saab harilikult kas katsete hoolikal kordamisel vältida või vähemalt katseandmete läbitöötamisel arvestada ja seega nende mõju kõrvaldada. Edaspidistes arutlustes eeldame, et jämedaid ja süstemaatilisi vigu ei esine.

2. Statistilise hinnangu mõiste

Üldkogumi indiviidide individuaalsed iseärasused, juhuslikud mõõtmisvead ja teised juhuslikud, mittekontrollitavad faktorid on põhjuseks, et uuritava tunnuse valimijaotus ei kopeeri täpselt tunnuse jaotust üldkogumis. Seetõttu esinevad lahkuminekid jaotuste karakteristikute vahel üldkogumis ja valimis ehk nn. representatiivsed vead.

Valimisse sattunud elementide uurimisel registreeritakse uuritava tunnuse väärtused harilikult lihtsa statistilise rea x'_1, x'_2, \dots, x'_n näol. Enne saadud väärtuste analüüsimisele asumist lähteandmed korrastatakse - esitatakse kas variatsioonireana või jaotustabelina intervallitud või momentrea kujul (harilikult väärtuste kasvamise järjekorras). Lühikestes ridades piirdutakse mõnikord lihtsa statistilise reaga.

Punkthinnang. Tunnuse jaotuse järgi valimis tuleb teha otsustusi selle jaotuse kohta ka üldkogumis. Selleks on vaja leida jaotuse karakteristikutele üldkogumis teatud mõttes parimad lähendväärtused valimi põhjal. Valimi põhjal leitud üldkarakteristikute lähendeid nimetatakse statistilisteks hinnanguteks ja need sõltuvad valimisse sattunud individuaalväärtustest ehk valimist ning on seega juhuslikud suurused. Üldkarakteristikute teatud mõttes parimateks lähenditeks peetakse statistilisi hinnanguid, mis rahuldavad järgmisi nõudeid. Soovitav on, et hinnang oleks

- 1) nihketa (ei sisalda nihet);
- 2) efektiivne;
- 3) mõjus ehk konsistentne.

Loetletud nõudeid rahuldavat hinnangut nimetatakse parimaks. Tähistame mingi üldkarakteristiku sümboliga K , sellele vastava hinnangu - k .

Nihketa on selline hinnang k , mille keskväärtus võr-

dub vastava üldkarakteristiku väärtusega, s.t.

$$E(k) = K.$$

Teiste sõnadega: hinnang k hälbib suuruselt K mõlemale poole keskmiselt ühepalju, selline hinnang ei sisalda süstemaatilist viga.

Efektiivne on selline nihketa hinnang, mille hajuvus oma keskväertuse suhtes on vähim. Mida efektiivsem on hinnang, seda väiksemaid kõikumisi on oodata keskväertuse K suhtes.

Mõjus on hinnang, mis valimi mahu suurenedes läheneb tõenäosuslikult hinnatavale üldkarakteristikule (tõenäosusliku lähenemise mõiste esines § 5, p. 2).

Sagedamini hinnatavateks üldkarakteristikuteks on tunnuse keskmine (keskväärtus), dispersioon ja suhteline sagedus üldkogumis (tõenäosus). Tähistame neid edaspidi vastavalt sümbolitega m_0 , σ^2 , p .

Käsitletud statistilisi hinnanguid nimetatakse tavaliselt punkthinnanguteks. Punkthinnang k juhusliku suuruse na jaotub mingi seaduse kohaselt ning pole päris selge, kui suur on tema viga hinnatava üldkarakteristiku K suhtes.

Vahemikhinnang. Sageli määratakse valimi põhjal väärtuste vahemik, mis sisaldab üldkarakteristikut teatud tõenäosusega. Sel juhul räägitakse vahemikhinnangust. Vahemik sõltub valimist ja on seega juhuslik suurus, mis arvteljel võib katta üldkarakteristikut K või asuda seal nii, et K jääb sellest vahemikust välja. Enamasti võetakse vahemikhinnang sümmeetriliselt punkthinnangu k suhtes piirides $k - \Delta$ ja $k + \Delta$ ning vahemikku $(k - \Delta, k + \Delta)$ nimetatakse usalduspiirkonnaks. Tõkkes $k + \Delta$ kannavad usalduspiiride nime. Usalduspiirkond on esitatud joonisel 22. Tõenäosust, millega usalduspiirkond katab üldkarakteristikut K , nimetatakse usaldusnivooks

$$P = P(k - \Delta < K < k + \Delta) = P(|K - k| < \Delta).$$



Joonis 22.

Usaldusnivood P esitatakse ka kujul $P = 1 - \gamma$. Kui näiteks $P = 0,95$ (95 %), siis $\gamma = 0,05$ (5 %). Suurus γ määrab tõenäosuse, et usaldusvahemik ei kata üldkarakteristikut K . Kui $\gamma = 0,05$, siis keskmiselt 5 valimit 100-st annavad usaldusvahemikud, milles ei asu K (mis ei kata üldkarakteristikut K).

Usalduspiirkonna avaldises sisalduvat suurust Δ nimetatakse piirveaks antud usaldusnivool P .

Järgnevatel punktides käsitleme valimiteooria peamisi ülesandeid:

- 1) punkthinnangute leidmine;
- 2) punkthinnangute usaldatavuse hindamine antud usaldusnivooga;
- 3) valimi mahu N selline määramine, et valim esindaks üldkogumit vajaliku täpsusega.

3. Üldkeskmise ja tunnuse tõenäosuse hinnangud

Üldkeskmise hinnang. Olgu moodustatud valim individuaalväärtustega x_i ($i = 1, 2, \dots, N$). Üldkogumi igal elemendil on võrdne võimalus valimisse sattumiseks ning iga x_i keskvärtuseks on üldkeskmine m_0 ja hajuvust mõõdab dispersioon üldkogumis σ^2 . Seega

$$E(x_i) = m_0, \quad D(x_i) = \sigma^2, \quad (i = 1, 2, \dots, N).$$

Eelmise paragrahvi 1. punktis selgus, et aritmeetilise keskmise \bar{x} puhul

$$E(\bar{x}) = m_0, \quad D(\bar{x}) = \frac{\sigma^2}{N}.$$

Võrreldes individuaalväärtuse x_1 ja aritmeetilise keskmise \bar{x} karakteristikuid, nähtub, et nii x_1 kui \bar{x} on üldkeskmise m_0 nihketa hinnangud, kuid \bar{x} on lisaks sellele ka efektiivne ja mõjus - selle hajuvus kahaneb kiiresti valimi mahtu N suurenedes.

Üldkeskmise m_0 parimaks hinnanguks on valimi keskmine \bar{x} , s.t. võetakse $m_0 \approx \bar{x}$.

Suhtelise sageduse hinnang. Olgu valimi abil vaja hinnata, kui suure osa üldkogumist moodustavad teatud tunnust X omavad indiviidid. Teiste sõnadega: milline on tunnust X omavate indiviidide suhteline sagedus p ($p \cdot 100\%$) ehk tunnusega X indiviidide tõenäosus.

Olgu valimi uurimisel selgunud, et tunnusega X indiviidide on valimis m tükki ehk tunnuse suhteline sagedus on $w = \frac{m}{N}$. Binoomjaotuse karakteristikute käsitlemisel nägime, et

$$E(w) = p, \quad D(w) = \frac{pq}{N}.$$

Siit nähtub, et tõenäosuse p nihketa ja efektiivseks hinnanguks on suhteline sagedus w . Bernoulli suurte arvude seadus kinnitab, et $w \rightarrow p$ N kasvades, s.t., et w on ka mõjus hinnang p jaoks.

Tunnuse tõenäosuse p parimaks hinnanguks on tunnuse suhteline sagedus valimis, s.t. võetakse $p \approx w$.

4. Ü l d d i s p e r s i o o n i h i n n a n g

Kui üldkeskmine m_0 on teada, siis on ülddispersiooni hinnanguks suurus

$$\tilde{\sigma}^2 = \frac{1}{N} \cdot \sum_{i=1}^n (x_i - m_0)^2.$$

Aga kuna keskväärus m_0 on enamasti teadmata ja selle hindamine toimub valimi keskmise \bar{x} abil, siis tuleb ka ülddispersiooni hinnang leida \bar{x} abil.

Valimi dispersioon s^2 on leitav valemist

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

(pikemate variatsioonide korral kasutame muidugi nii \bar{x} kui ka s^2 leidmiseks §-s 3 antud lihtsustavaid arvutusvõtteid). Saab näidata, et $E(s^2) = \frac{N-1}{N} \cdot \sigma^2$ - valimi dispersioon on ülddispersiooni nihkega hinnang. See nihe on kõrvaldatav parandusteguri sissetoomisel.

Ülddispersiooni parimaks hinnanguks on korduvas valim

$$\sigma^2 \approx \frac{N}{N-1} s^2,$$

kordumatus valimis

$$\sigma^2 \approx \frac{N}{N-1} \cdot \frac{N_0 - 1}{N_0} \cdot s^2.$$

Kui üldkogumi maht N_0 on lõpmatu, langevad mõlemad valemid ühte:

$$\frac{N_0 - 1}{N_0} = 1 - \frac{1}{N_0} \rightarrow 1.$$

Praktikas on sobiv kasutada esimest (lihtsamat valemit) alati, kui üldkogumi maht on küllalt suur ($N_0 \geq 10N$) sõltumata sellest, kas valim on korduv või kordumatu.

Suurusi σ^2 ja s^2 siduvas valemis esineva kordaja $\frac{1}{N-1}$ mõju on märgatav väikeste valimite puhul. Nii näiteks juba $N = 20$ korral on σ^2 ja s^2 erinevus ainult umbes 5%, $N \geq 50$ korral erinevus praktiliselt kaob.

5. Valimi keskmine standardhälve

Eespool (p. 3) saime punkthinnangud üldkeskmise ja tunnuse tõenäosuse jaoks. Saadud hinnangute kui juhuslike

suuruste täpsus on iseloomustatav nende dispersioonide abil (vastavalt valimi keskmise ja suhtelise sageduse dispersioonid $D(\bar{x})$, $D(w)$).

Edaspidi tuleb samatüübilisi arutlusi läbi viia nii \bar{x} kui w jaoks, kusjuures erinevused on ainult valemite väliskujudes. Seepärast kasutame nende mõlemate kui juhuslike suuruste samaaegseks märkimiseks sümbolit k . Suuruse k dispersiooni tähistame $D(k)$ [$D(\bar{x})$ ja $D(w)$]. Dispersioon $D(k)$ iseloomustab k hajuvust oma keskväertuse ümber (\bar{x} hajuvus m_0 ümber, w hajuvus p ümber). Arvutustes esineb harilikult standardhälve $\sqrt{D(k)}$.

Tähistame standardhälvet $\sqrt{D(k)}$ sümboliga μ ,

$$\mu = \sqrt{D(k)}.$$

Suuruse k dispersioon ja standardhälve μ avalduvad ülddispersiooni kaudu järgmiselt:

suurest üldkogumist ($N_0 \geq 10N$) tehtud valimi puhul

$$D(\bar{x}) = \frac{\sigma^2}{N}, \quad \mu_{\bar{x}} = \frac{\sigma}{\sqrt{N}};$$

$$D(w) = \frac{pq}{N}, \quad \mu_w = \sqrt{\frac{pq}{N}};$$

väikesest üldkogumist ($N_0 < 10N$) tehtud valimi puhul

$$D(\bar{x}) = \frac{\sigma^2}{N} \cdot \frac{N_0 - N}{N_0 - 1} \approx \frac{\sigma^2}{N} \cdot \left(1 - \frac{N}{N_0}\right);$$

$$\mu_{\bar{x}} = \sqrt{\frac{1}{N} \cdot \left(1 - \frac{N}{N_0}\right)} \cdot \sigma;$$

$$D(w) = \frac{pq}{N} \left(1 - \frac{N}{N_0}\right), \quad \mu_w = \sqrt{\frac{pq}{N} \left(1 - \frac{N}{N_0}\right)}.$$

Enamikel juhtudel on σ^2 ja p teadmata ning neid hinnatakse valimi abil (p. 3, 4). Asendades σ^2 ja p nende hinnangutega, saame praktilised arvutuseeskirjad $D(k)$ ja μ leidmiseks:

suure üldkogumi puhul

$$D(\bar{x}) = \frac{s^2}{N-1}, \quad \mu_{\bar{x}} = \frac{s}{\sqrt{N-1}};$$

$$D(w) = \frac{w(1-w)}{N}, \quad \mu_w = \sqrt{\frac{w(1-w)}{N}};$$

väikese üldkogumi puhul

$$D(\bar{x}) = \frac{s^2}{N-1} \left(1 - \frac{N}{N_0}\right), \quad \mu_{\bar{x}} = \sqrt{\frac{s^2}{N-1} \left(1 - \frac{N}{N_0}\right)};$$

$$D(w) = \frac{w(1-w)}{N} \left(1 - \frac{N}{N_0}\right), \quad \mu_w = \sqrt{\frac{w(1-w)}{N} \left(1 - \frac{N}{N_0}\right)}.$$

6. Keskväärtuse ja tõenäosuse usalduspiirkonna leidmine

Leitud punkthinnangute \bar{x} või w (üldiselt hinnang k) ja nende standardhälvete abil saab lahendada kaht tüüpi ülesandeid vahemikhinnangute kohta. Üldisel kujul võib neid ülesandeid formuleerida järgmiselt:

- 1) antud usaldusnivoo järgi leida üldkarakteristiku usalduspiirkond (või usalduspiirid või valimi piirviga);
- 2) antud piirkonna järgi leida üldkarakteristiku selles piirkonnas asumise tõenäosus (ehk usaldusnivoo).

Valimi keskmise ja suhtelise sageduse jaotus on enamasti normaaljaotusele küllalt lähedane, kui valimi maht N on piisavalt suur. Sellepärast saab kasutada arvutuseeskirju, mis on toodud § 4, p. 11. Antud valimi jaoks leitakse esmalt vajalik hinnang k , selle hajuvusmõõdud $D(k)$ ja μ (vt. p. 5).

1. ü l e s a n n e. Antud on usaldusnivoo $P = 1 - \delta$,
s.t.

$$P(|k - K| < \Delta) = P.$$

Leida valimi piirviga Δ ja seejärel usalduspiirkond $k \pm \Delta$. Kasutades Laplace'i funktsiooni, saame seosest $P = \bar{\Phi}(t)$ leida funktsiooni väärtusele vastava argumendi t väärtuse. Teiselt poolt $t = \frac{\Delta}{\mu}$. Seega piirviga $\Delta = t\mu$ ja usalduspiirkonnaks saame $(k - t\mu, k + t\mu)$.

2. ü l e s a n n e. Antud on valimi piirviga Δ . Lei-
da piirveale vastava usalduspiirkonna usaldusnivoo.

Leiame $t = \frac{\Delta}{\mu}$ ja seejärel $\bar{\Phi}(t)$ tabelist usaldus-
nivoo $P = \bar{\Phi}(t)$.

7. N ä i t e i d

1. n ä i d e. Nisu saagikuse määramiseks 10 000 ha
suuruselt pindalalt tehti kontrollmõõtmised 1000 hektaril
korduva valimina. Leida keskmise saagi usalduspiirid usal-
dusnivooga 1) 0,99 ja 2) 0,95. Vaatlusandmed on arvutus-
skeemi kahes esimeses lahtris.

Alustame \bar{x} ja s^2 leidmisest, kasutades abisuurust

$$x' = \frac{x - 26}{2}$$

x Saak ts	f Pindala ha	x'	$x' \cdot \frac{f}{10}$	$x'^2 \cdot \frac{f}{10}$
21-23	150	-2	-30	60
23-25	200	-1	-20	20
25-27	450	0	-50	
27-29	200	1	20	20
	1000		-30	100

Keskmine

$$\bar{x} = \frac{-30}{100} \cdot 2 + 26 = 25,40 \text{ ts},$$

dispersioon

$$s^2 = \frac{100}{100} \cdot 4 - (26 - 25,40)^2 = 3,64 \text{ ts}^2.$$

Leiame valimi keskmise standardhälbe

$$\mu = \sqrt{\frac{s^2}{N-1}} = \sqrt{\frac{3,64}{999}} \approx 0,06 \text{ ts.}$$

Nüüd saame leida valimi piirvea Δ .

1) $P = 0,99$.

Laplace'i funktsiooni tabelist leiame antud tõenäosusele vastava argumenti väärtuse $t \approx 3$ ja

$$\Delta = t \cdot \mu = 3 \cdot 0,06 = 0,18 \text{ ts ;}$$

2) $P = 0,95$.

Analoogiliselt leiame, et $t \approx 2,0$ ja $\Delta = 2 \cdot 0,06 = 0,12$. Keskmise usalduspiirkond on siis

1) $25,40 - 0,18 < m_0 < 25,40 + 0,18$; $25,22 < m_0 < 25,58$;

2) $25,40 - 0,12 < m_0 < 25,40 + 0,12$; $25,28 < m_0 < 25,52$.

(Pöörake tähelepanu sellele, kuidas P muutumine mõjutab piirvea ja usaldusvahemiku muutumist!).

2. n ä i d e. Leida eelmise ülesande andmetel piirveale $\Delta = 0,15$ ts vastav usaldusnivoo.

Arvutame

$$t = \frac{\Delta}{\mu} = \frac{0,15}{0,06} = 2,5$$

ja tabelist leiame usaldusnivoo $P = \bar{\Phi}(2,5) = 0,9876 \approx 0,98$; usaldusnivoo ümardamine toimub allapoole.

Kui moodustada veel valimeid, siis 98 % juhtudest saab üldkeskmisele usalduspiirideks $25,4 \pm 0,15$ ts ja ainult 2 % juhtudest võib anda sellised usalduspiirid, mis ei sisalda üldist keskmist saaki.

3. n ä i d e. Selleks, et selgitada 1000 rbl. mitte-ületavate hoiuste protsenti hoiuste üldarvust, vaadeldi 900 hoiuarvet. Selgus, et 30 % hoiustest ei ületanud 1000 rbl. Kui suure usaldusnivooga võib väita, et selliste hoiuste osatähtsus kõigi hoiustajate hulgas ei erine leitud tulemustest rohkem kui 2 %?

Leiame

$$\mu = \sqrt{\frac{0,3(1-0,3)}{900}} \approx 0,0153 ;$$

piirviga $\Delta = 2\% = 0,02$ ja

$$t = \frac{\Delta}{\mu} = \frac{0,02}{0,0153} = 1,31.$$

Tabelist leiame, et $\bar{\Phi}(1,31) = 0,81$, mis ongi otsitav usaldusnivoo.

8. Valimi mahu määramine

Ülesanne on järgmine. Mitu indiviidi tuleb valimisse võtta, et valimi piirviga ei ületaks antud usalduspiire antud usaldusnivooga?

Vaatleme eraldi suurt (lõpmatut) ja lõplikku üldkogumit.

Antud usaldusnivoo P järgi leiame t ja seosest

$$\Delta = t\mu$$

valimi mahu N , arvestades, et suure üldkogumi puhul

$$\mu = \sqrt{\frac{\sigma^2}{N}}.$$

Saame võrrandi otsitava N suhtes

$$\Delta = t \sqrt{\frac{\sigma^2}{N}},$$

mille lahendamine annab

$$N = \frac{t^2 \sigma^2}{\Delta^2}.$$

Lõpliku üldkogumi puhul on mõttekääk analoogiline eelmisega; ainult, et

$$\mu = \sqrt{\frac{\sigma^2}{N} \left(1 - \frac{N}{N_0}\right)}.$$

Lahendades võrrandi

$$\Delta = t \cdot \sqrt{\frac{\sigma^2}{N} \left(1 - \frac{N}{N_0}\right)}$$

tundmatu N suhtes, saame tulemuseks

$$N = \frac{N_0 t^2 \sigma^2}{N_0 \Delta^2 + t^2 \sigma^2} .$$

Mitte teades ülddispersiooni σ^2 , asendatakse see valimi dispersiooniga s^2 .

N ä i d e. Olgu üldkogumi maht $N_0 = 1000$ ja tunnuse dispersioon üldkogumis $\sigma^2 = 5$. Määrata vajalik valimi maht N nii, et garanteerida valimi piirvea $\Delta = 0,1$ mitteületamist usaldusnivooga $0,997$.

Proovime kõigepealt kasutada esimest valemit mahu N leidmiseks; selleks määrame tabelist $t = 3$ ja

$$N = \frac{3^2 \cdot 5}{0,1^2} = 4500 , \quad N > N_0 .$$

Ilmselt ei sobi eeldus, et üldkogum on valimiga võrreldes suur, ning tuleb kasutada täpsemat valemit.

Seega 1000 individist koosneva üldkogumi puhul tunnuse hajuvusega $\sigma^2 = 5$ ei saagi moodustada sellist korduvat valimit, mis garanteeriks nii väikese piirvea väga kõrge usaldusnivooga.

Teise valemi järgi saame

$$N = \frac{1000 \cdot 3^2 \cdot 5}{1000 \cdot 0,1^2 + 3^2 \cdot 5} = \frac{9000}{11} = 834 ,$$

mis erineb vähe üldkogumi mahust N_0 .

Olgu üldkogumis $N_0 = 100\ 000$ elementi. Siis saaksime vastavateks mahtudeks 4500 ja 4306.

9. Ü l d i s i j ä r e l d u s i

Selgitame, kuidas mõjutavad üksteist suurused N , σ , Δ ja P . Valime kaks suurust konstantseteks (loeme antuteks) ja uurime, kuidas teised 2 suurust teineteist vastastikku mõjutavad. Usaldusnivoolt $P = \bar{\varphi}(t)$ võib üle minna argumendile t , kusjuures suuremale P -le vastab suurem t väärtus. Uuritavad suurused on seotud valemis

$$N = \frac{t^2 \cdot \sigma^2}{\Delta^2} .$$

Antud on P ja σ . Mahu N suurendamine tingib Δ vähenemise (murru lugeja on const). Sisuline tähendus: valimi mahu suurendamisega kaasneb usaldusnivoo muutumatuks jäädes tulemuste täpsuse kasv (piirviga väheneb).

Antud on P ja Δ . Suurendades σ^2 suureneb ka N - suurema hajuvusega tunnuste korral tuleb sama usaldusnivoo-ga täpsuse tagamiseks moodustada suurem valim.

Antud on Δ ja N. Tunnuse hajuvuse σ^2 suurenemine vähendab suurust t ja seega ka vaatlustulemuste usaldusni-wood P (t ja Δ on pöördvõrdelised).

Analüüsige σ ja Δ vahetada antud P ning N puhul, samuti P ja Δ vahetada antud σ ning N puhul!

§ 7. KORRELATSIOONITEOORIA ELEMENTE

1. Statistiline sõltuvus

Kursuse varasemas osas käsitlesime juhuslikke suurusi üksikult, omaette võetuna. See oli vajalik suuruste käitumise uurimise põhimeetodite tundmaõppimiseks. Tegelikuses on suurused omavahel seotud ja suurustevahelise seose väljaselgitamiseks on tarvis samaaegselt uurida mitme suuruse käitumist. Näiteks ettevõtte toodangu omahind sõltub keskmisest toodangu hulgast ühe töölise kohta ja kulutustest tootmise mehhaniseerimisel ning automatiseerimisel. Siin tuleb uurida kolme suuruse samaaegset muutumist. Vastastikusel seoses olevate suuruste arv võib olla ka suurem. Edasises piirdume lihtsama juhuga, kus samaaegselt uuritakse kaht suurust eesmärgiga välja selgitada nende vahelist seost (sõltuvust).

Üheks suurustevahelise sõltuvuse kirjeldamise viisiks on funktsionaalne sõltuvus. Kaks suurust X ja Y on funktsionaalses sõltuvuses, kui ühe suuruse (argumendi X) igale lubatavale väärtusele vastab teise suuruse (funktsiooni Y) kindel väärtus. Funktsionaalse sõltuvuse korral on muutujad X ja Y üldiselt üksüheses vastavuses.

N ä i t e i d

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 toimu-
mugu kapitali suurenemine ainult kasumi näol konstantse
kasuminormiga 100 p % aastas. Kapitali (y) suuruse sõltu-
vus ajast (t) esitub tehtud eeldusel 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.

N ä i t e i d

- Inimese raskus sõltub tema pikkusest.
- Perekonnas tarbitava piima hulk sõltub perekonna suurus-
sest.
- Samatüüblistes ettevõtetes tingib toodangu mahu (X) kasv omahinna (Y) alanemist. Kuid sama toodangu mahuga ettevõ-
tetes on omahind erinev, mis on tingitud sellest, et
tootmistingimused ei ole täpselt samad, omahinda mõjuta-
vad peale toodangu ka teised faktorid.
- Ettevõtte tööliste arv sõltub ettevõtte põhifondide suu-
rusest.
- Omahind on seotud tööviljakusega.
- 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 sõnnikunormidele vastavad erinevad saagid.
Puudub üksühene vastavus suuruste X ja Y vahel.

Toodud näited kuuluvad nn. statistiliste sõltuvuste hulka.

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: 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	jne.
f	1	2	

Omahinna Y väärtustele vastab toodangu mahu X väärtuste jaotus. Väärtusele Y = 150 vastav X jaotus on

X	200	300	400	500	jne.
f	1	1	2	1	

Saab selgitada, kuidas toodangu mahu muutmine mõjutab omahinna muutumist või vastupidi. Sellises käsitluses jäetakse 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. Toodangu mahu X muutumisega kaasneb üldiselt omahinna Y muutumine, kuid suuruse Y muutumine ei ole seletatav ainult suuruse X muutumisega. Mõlema suuruse muutumist mõjutavad paljud muud tegurid, mida käesoleva ülesande seade korral loetakse juhuslikeks, ja sellepärast on ka suurused X ning Y juhuslikud. Suuruse Y muutumine on osaliselt seletatav

suuruse X muutumisega, kuid Y muutumisele avaldavad mõju ka kõrvalised (juhuslikud) tegurid.

Kahe suuruse vahelist sõltuvust (seost) nimetatakse statistiliseks (stohhastiliseks ehk tõenäosuslikuks), kui ühe suuruse igale väärtusele vastab teise suuruse väärtuste jaotus, mis muutub koos esimese suuruse muutumisega.

Statistilise sõltuvuse mõiste on laiem funktsionaalse sõltuvuse mõistest. Funktsionaalne sõltuvus on statistilise sõltuvuse üks piirjuhte: juhuslike lisamõjude puudumisel on teineteisega seotud suurused X ja Y üksüheses vastavuses. Teiseks piirjuhiks on suuruste täielik sõltumatus: ühe suuruse (suuruse jaotuse) muutumine ei sõltu teise suuruse väärtuste muutumisest. Näiteks tehase toodangu maht ei sõltu tehase väravas oleva valvepersonali suurusest.

Kahe juhusliku suuruse vahelise seose uurimisel püütakse selgitada, milline oleks suuruste vastastikune seos kõrvaliste tegurite puudumisel. Selleks väljatöötatud meetodid ongi nn. korrelatsiooniteooria sisuks (correlation - vastastikune suhe, vahekord). Järgnevas tutvume korrelatsiooniteooria kahe põhiülesandega:

1) suurustevahelise seose tugevuse hindamine, s.t. selgitamine, kuivõrd ühe suuruse muutumine oleneb teise vaadeldava suuruse muutumisest;

2) juhuslike suuruste vahelise juhuslikkust sisaldava seose esitamine funktsionaalse (mittejuhusliku) seose abil.

2. L ä h t e a n d m e t e e s i t u s v i i s e

Toimugu mingi kogumi uurimine tunnuste X ja Y järgi N katsel. Igal üksikkatsel registreeritakse uuritavate tunnuste väärtused, saades seega N väärtuspaari (x_1, y_1). Katsetulemusi võib esitada mitmel viisil.

Lihtne korrelatsiooni- ehk loendustabel saadakse sel teel, et paralleelsetesse ridadesse märgitakse kahe suuruse kokkukuuluvad väärtused x_1 ja y_1 :

X	x_1	x_2	...	x_N
Y	y_1	y_2	...	y_N

Eelmise punkti 9. ja 10. näites on vaatlusandmed esitatud lihtsa korrelatsioonitabelina.

Lihtsas korrelatsioonitabelis võib esineda ühe või teise suuruse kordseid väärtusi ja kordseid väärtuspaare.

Pikemate loendustabelite korral on parema ülevaate saamiseks otstarbekas katseandmeid korrastada ühe tunnuse väärtuste kasvamise (kahanemise) järjekorras.

Korrelatsioonitabel saadakse katseandmete korrastamisel samaaegselt mõlema tunnuse järgi. Üks võimalus on eraldi välja kirjutada ühe tunnuse jaotustabelid teise tunnuse üksikväärtuste järgi ja vastupidi (samal põhimõttel nagu eelmise punkti 10. näites alustasime). Seda tehakse harilikult moment- või intervallitud variatsiooniridade kujul. Ülevaatlikum ja kompaktsem on mõlema tunnuse jaotustabelid esitada ühise tabelina. Valitud näite puhul saame nn. korrelatsioonitabeli 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

Korrelatsioonitabelis kantakse esimesse veergu ja ritta kummagi tunnuse erinevad väärtused kas kasvavas või kahanemas järjekorras (võib ka intervallide kaupa) ning ridade ja veergude lõikekohta märgitakse neile vastavate väärtuspaaride esinemise sagedus. Sageduste liitmisel ridasid või veerge mööda saab leida ühe tunnuse mingi väärtuse sageduse, sõltumatult sellest, milliste teise tunnuse väärtustega see koos esines. Esitatud korrelatsioonitabelist nähtub, et omahinnaga $Y = 160$ ja toodangu mahuga $X = 300$ on kaks ettevõtet;

omahinnaga $X = 200$ on üldse 4 ettevõtet; 5 ettevõttes on toodangu omahind $Y = 150$ jne.

Üldisel kujul võib korrelatsioonitabeli esitada järgmiselt:

$X \backslash Y$	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_{22}	...	f_{2j}	...	f_{2l}	m_2
...
x_i	f_{i1}	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

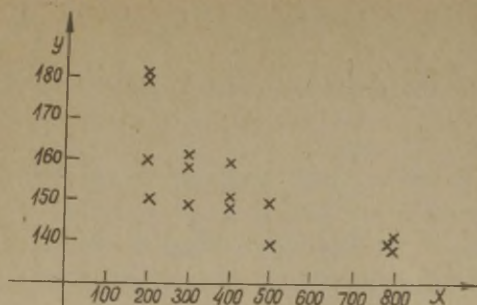
Sageduste liitmine ridade (vastavalt veergude) kaupa annab sageduste summa m_i (n_j). Seega kehtivad järgmised seosed:

$$\sum_{j=1}^l f_{ij} = m_i ; \quad \sum_{i=1}^k f_{ij} = n_j .$$

Sageduse f_{ij} kõigi väärtuste summeerimine annab kogumi mahu N , s.t.

$$\sum_{i,j=1}^{k,l} f_{ij} = \sum_{i=1}^k m_i = \sum_{j=1}^l n_j = N .$$

Korrelatsiooniväli. Sageli on otstarbekas statistilist sõltuvust esitada geomeetriliselt. Iga tunnuste kokku kuuluvate väärtuste paari (x_i, y_j) tõlgendatakse punkti koordinaatidena ristkoordinaadistikus. Igale väärtuste paarile vastab nii palju üksteisega kattuvaid punkte, kui suur on vastav sagedus f_{ij} , ehk teisiti väljendades: igal punktil (x_i, y_j) on kaal f_{ij} . Kaalud võib märkida punktidele juurde või väikese arvu punktide korral märgitakse punktid üksteise ligidusse. Ettevõtete näitele vastab korrelatsiooniväli joonisel 23. Nii korrelatsioonitabeli kui ka korrelatsiooniväl-



Joonis 23.

metele 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. Korrelatsioonikordaja

Asume korrelatsiooniteooria ühe põhiülesande - kahe juhusliku suuruse seose tugevuse käsitlemisele.

Dispersiooni omaduste vaatlemisel selgus, et sõltumata juhuslike suuruste X ja Y summa dispersioon on võrdne liidetavate dispersioonide summaga. Kui aga osutub, et

$$D(X + Y) \neq D(X) + D(Y),$$

siis on suurused X ja Y teineteisest sõltuvad (teineteisega seotud). Sõltuvate suuruste korral kehtib järgmine seos (vt.

3 p. 10):

$$D(X + Y) = D(X) + D(Y) + 2 C_{XY},$$

kus korrelatsioonimoment

$$C_{XY} = E[(X - m_{0x})(Y - m_{0y})],$$

ja 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 lähteand-

Seega korrelatsioonimomendi erinevuse nullist näitab, et juhuslikud suurused on statistilises sõltuvuses. Samaaegselt võrdus $C_{xy} = 0$ ei anna alust vaista, et suurused X ja Y on teineteisest sõltumatud.

Korrelatsioonimomendi dimensiooniks on suuruste X ja Y dimensioonide korrutis. Dimensiooniga suuruste omavaheline võrdlemine on väga tülikas ja seepärast ei kasutata seda vahetult suuruste sõltuvuse astme hindajana. Üheks lihtsamaks ja sagedamini kasutatavaks seose tugevuse hindajaks on korrelatsioonikordaja.

Korrelatsioonikordajaks nimetatakse korrelatsioonimomendi ja juhuslike suuruste standardhälvete korrutise jagatist.

Kui tähistada korrelatsioonikordaja sümboliga r, siis

$$r = \frac{C_{xy}}{\sigma_x \cdot \sigma_y} .$$

Siit nähtub, et r on dimensioonita suurus.

4. Korrelatsioonikordaja omadusi

- 1) Sõltumatute suuruste korrelatsioonikordaja on null.
- 2) Korrelatsioonikordaja ei muutu suuruste x või Y muutmisel mingi konstandi võrra või nende korrutamisel mingi positiivse arvuga. Kui ainult üht suurust arvuga -1 korrutada, siis muudab märki ka korrelatsioonikordaja.

Nende omaduste abil saame edaspidi korrelatsioonikordaja arvutamist lihtsustava eeskirja.

Tõepoolest, kui võtta suuruse X asemel suurus $\xi = X + c$, siis suureneb keskväärtaus m_x ka c võrra ($m_\xi = m_x + c$) ning korrelatsioonimomendi avaldises esinev hälve X - m_x asendub hällbega $\xi - m_\xi$:

$$\xi - m_\xi = X + c - (m_x + c) = X - m_x .$$

Seega r avaldises murru lugeja ei muutu. Muutumatuks jääb ka

murru nimetaja, sest konstandi liitmine juhuslikule suurusele selle dispersioonile (siis ka standardhälvet) ei mõjuta. Suuruse X korrutamisel positiivse arvuga muutub murru lugeja ja samuti nimetaja sama arv kordi (vt. dispersiooni omadusi!) ning murru enda väärtus ei muutu.

3) Korrelatsioonikordaja väärtused ei ole väiksemad kui -1 ja suuremad kui $+1$, s.t. $-1 \leq r \leq 1$ ehk korrelatsioonikordaja absoluutväärtus ei ületa 1 , s.t. $|r| \leq 1$.

Selle omaduse õigsuses veendume järgmise arutlusega. Avaldame $X \pm Y$ dispersiooni suuruste X ja Y dispersioonide σ_x^2 , σ_y^2 ning korrelatsioonikordaja kaudu:

$$D(X \pm Y) = \sigma_x^2 + \sigma_y^2 \pm 2C_{xy}.$$

Korrelatsioonikordaja definitsioonivalemist järgneb

$$C_{xy} = r \cdot \sigma_x \cdot \sigma_y$$

ning seega

$$D(X \pm Y) = \sigma_x^2 + \sigma_y^2 \pm 2r \cdot \sigma_x \cdot \sigma_y.$$

Samasugused võrdused kehtivad ka suuruste X ja Y normeeritud hälvete jaoks, kuna 2. omaduse põhjal r väärtus ei muutu üleminekul normeeritud hälvetele. Lisaks sellele v (vt. § 3, p. 13)

$$D(t_x) = D(t_y) = 1.$$

Eeltoodut arvestades saame

$$D(t_x + t_y) = 2 + 2r, \quad D(t_x - t_y) = 2 - 2r.$$

Kuna dispersioon on alati mittenegatiivne, siis

$$2 + 2r \geq 0 \quad \text{ja} \quad 2 - 2r \geq 0,$$

kust $r \geq -1$, $r \leq 1$.

Mida suurem on $|r|$, seda tugevam on suuruste X ja Y vaheline seos ehk korrelatsioon. Osutub, et kui $r = \pm 1$, siis on suurused X ja Y funktsionaalses ja nimelt lineaarses sõltuvuses.

Kui $r > 0$, siis vastab ühe suuruse kasvamisele ka teise suuruse kasvamise tendents ehk suuruste muutumise suund on samapidine. Kui $r < 0$, siis kaasneb ühe suuruse kasvamisega teise kahanemise tendents ehk suuruste muutumise suund on vastupidine.

Rõhutame veel kord, et kui $r = 0$, siis suurused ei tarvitse olla sõltumatud, vaid nende vahel võib isegi eksisteerida lineaarsest sõltuvusest keerukam funktsionaalne sõltuvus.

K o k k u v ô t e. Korrelatsioonikordaja absoluutväärtus näitab suurustevahelise seose tugevust ja just selle seose lähedust lineaarsele funktsionaalsele seosele. Suurematele $|r|$ väärtustele vastab tugevam (tihedam) seos. Korrelatsioonikordaja märk näitab suuruste muutumise suunda.

5. Korrelatsioonikordaja arvutamisest

Korrelatsioonikordaja definitsioonvalemist

$$r = \frac{C_{xy}}{\sigma_x \cdot \sigma_y}$$

lähtudes saab praktilised arvutuseeskirjad r arvutamiseks. Arvutusvalemid kohandame lähteandmete erinevate esitusviiside juhuks.

Lihtsa korrelatsioonitabeli juht. Olgu N katses koosnevas katseseerias registreeritud väärtuspaarid (x_1, y_1) esitatud lihtsa korrelatsioonitabelina.

Korrelatsioonimomenti C_{xy} arvutatakse järgmiselt:

$$C_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) .$$

Viimast valemit teisendades võib saada ka järgmised arvutuseeskirjad:

$$C_{xy} = \frac{1}{N} \sum x_1 y_1 - \frac{1}{N^2} \sum x_1 \sum y_1 = \frac{1}{N} \sum x_1 y_1 - \bar{x} \cdot \bar{y} .$$

Dispersioonide (vastavalt ka standardhälvete) arvutamiseks kasutame kursuse varemast osast tuttavaid valemeid:

$$\sigma_{\alpha}^2 = \frac{1}{N} \sum (\alpha_i - \bar{\alpha})^2 = \frac{1}{N} \sum \alpha_i^2 - \bar{\alpha}^2 \quad (\alpha = x, y).$$

Korrelatsioonitabeli juht. Juhime jälle tähelepanu sellele, et intervallide kaupa antud väärtuste korral tuleb suuruste arvutuslikeks väärtusteks võtta intervallide kesk-kohad. Valemites esinevad sümbolid vastavad 2. punktis too-dud korrelatsioonitabelis kasutatud sümbolitele.

Käesoleval juhul tuleb korrelatsioonimomendi C_{xy} arvu-tamisel arvestada seda, et väärtuspaarid (x_i, y_j) ei esine enam üks kord, vaid korrelatsioonitabelis näidatud f_{ij} kor-da. Saame

$$\begin{aligned} C_{xy} &= \frac{1}{N} \sum_{i,j=1}^{k,l} f_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \\ &= \frac{1}{N} \sum f_{ij}x_i y_j - \frac{1}{N^2} \sum x_i^m \cdot \sum y_j^n = \\ &= \frac{1}{N} \sum f_{ij}x_i y_j - \bar{x} \bar{y}. \end{aligned}$$

Toodud valemis esinev liige (summeerimine toimub mõlema in-deksi i ja j järgi)

$$\sum_{i,j=1}^{k,l} f_{ij}x_i y_j$$

arvutatakse järgmiselt: leitakse korrelatsioonitabelis iga rea ja veeru lõikekohas oleva sageduse f_{ij} , i -rea väärtuse x_i ja j -veeru väärtuse y_j korrutised ning liidetakse need.

Korrelatsioonikordaja avaldise nimetajas olevate standardhälvete leidmine toimub varemast tuttavate valemite järgi.

Uuritavate suuruste X ja Y karakteristikuid, arvutatuna kogu korrelatsioonitabeli andmetel, nimetatakse suuruste üldkarakteristikuteks (üldkeskmine, -dispersioon jne).

Korrelatsioonikordajat on otstarbekas arvutada järgmise arvutusskeemi järgi: leida

- 1) suuruste üldkeskmised,
- 2) ülddispersioonid,
- 3) suuruste kõigi väärtuspaaride korrutiste summa,
- 4) korrelatsioonimoment,
- 5) korrelatsioonikordaja.

Käsitsi arvutamisel saab arvutusi lihtsustada, asendades lähteväärtused sobivalt tinglike väärtustega, nii nagu seda tegime varem keskmise ja dispersiooni arvutamisel. Suurused X ja Y asendame abisuurustega u ja v järgmiselt:

$$u = \frac{X - c_1}{h_1} \quad \text{ja} \quad v = \frac{Y - c_2}{h_2} ,$$

kus c_1 , c_2 , h_1 , h_2 on sobivalt valitud arvud. Korrelatsioonitabelina esitatud andmete korral valitakse harilikult arvudeks c_1 ja c_2 kas X ja Y väärtuste rea keskkohad või suurima sagedusega väärtused (NB! intervallitud ridade korral vastavate intervallide keskkohad). Siis jagunevad hälbed $X - c_1$ ja $Y - c_2$ harilikult võrdlemisi sümmeetriliselt arvu 0 suhtes. Jagajateks h_1 ja h_2 on harilikult sobiv võtta u ja v lugejas olevate hälvete suurimad ühisjagajad. Intervallitud ridade korral on nendeks tavaliselt intervallide pikkused. Edasi toimub korrelatsioonikordaja arvutamine eespool toodud valemitte järgi, kus lähtesuurused X ja Y on asendatud vastavalt abisuurustega u ja v. Korrelatsioonikordaja 2. omaduse kohaselt tema väärtus ei muutu.

6. Ü l e s a n d e i d

Illustreerime eelmises punktis kirjeldatud näidetega.

Ü l e s a n n e 1. 15 ettevõtte põhivahendite maksumuse X (milj. rbl.) ja kogutoodangu mahu Y (milj. rbl.) vaheline sõltuvus on esitatud järgmises tabelis:

X	48	50	48	35	49	40	41	34	45	50	49	38	36	52	51
Y	54	51	46	32	50	39	44	38	52	55	48	35	40	51	45

Leida korrelatsioonikordaja.

Arvutuste lihtsustamiseks võtame abisuurused järgmiselt:

$$u = \frac{X - 45}{1} \quad \text{ja} \quad v = \frac{Y - 44}{1} .$$

Vajalikud arvutused teeme arvutusskeemi abil; tehete iseloom nähtub tabeli peast.

X	Y	u	v	u ²	v ²	u · v
48	54	3	10	9	100	30
50	51	5	7	25	49	35
48	46	3	2	9	4	6
35	32	-10	-12	100	144	120
49	50	4	6	16	36	24
40	39	-5	-5	25	25	25
41	44	-4	0	16	0	0
34	38	-11	-6	121	36	66
45	52	0	8	0	64	0
50	55	5	11	25	121	55
49	48	4	4	16	16	16
38	35	-7	-9	49	81	63
36	40	-9	-4	81	16	36
52	51	7	7	49	49	49
51	45	6	1	36	1	6
Summa		-9	20	577	742	531

Nüüd leiame keskmised ja dispersioonid

$$\bar{u} = - \frac{9}{15} = - 0,600 ; \quad \sigma_u^2 = \frac{577}{15} - 0,36 = 38,3 ;$$

$$\bar{v} = \frac{20}{15} = 1,33 ; \quad \sigma_v^2 = \frac{742}{15} - 1,33^2 = 41,0 .$$

Seega

$$r = \frac{1}{15} \cdot 531 - (-0,4) \cdot 1,33}{38,1 \cdot 47,0} = 0,84 .$$

Viimane tulemus näitab, et seos ettevõtete põhivahendite maksumuse ja kogutoodangu mahu vahel on tugev.

Ü l e s a n n e 2. Grupi Kuibõševi oblasti tööstus-ettevõtete plaanilise tööajafondi (inimpäevades) ehk tööjõu kulu täitmise protsenti (Y) ja tootmisplaani täitmise protsenti (X) iseloomustavad andmed on esitatud korrelatsioonitabelis.

Leida korrelatsioonikoefitsient.

Y \ X	45-55	55-65	65-75	75-85	85-95	95-105	105-115	115-125	125-135	135-145	145-155	155-165	m_x
	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	
45-55	2	1			I						II		3
55-65	1	2		1	1								5
65-75	3	3	1	1									8
75-85			1	1	5	3	2 ⁻²						12
85-95			1	6	5	9	5 ⁻¹	2 ⁻²		1 ⁻⁴		1 ⁻⁶	
95-105	0		1	6	6	20	8	2		1			44
105-115	1	1		3	9	15	6	4	1	1		2	42
115-125				1		8	5	6	2	1	1	2	26
125-135					4	4	3	5	4				20
135-145								4	1		1	2	8
145-155		III				1	IV				2	3	6
155-165					1	1	2	1			1	4	10
m_y	6	7	5	19	31	61	31	24	8	4	5	14	214

Arvutuste lihtsustamiseks on tunnuste X ja Y arvutuslike väärtuste (intervallide keskkohad) asemel võetud abisuurused

$$u = \frac{X - 100}{10} \quad \text{ja} \quad v = \frac{Y - 100}{10}$$

(väärtustele X = 100 ja Y = 100 vastab suurim sagedus).

Ruumi kokkuhoiu mõttes on u ja v väärtused kantud ka lähteandmete tabelisse.

Keskliste ja dispersioonide arvutamise käiku saab jäl-

gida järgmisest tabelist:

u	n_x	um_x	u^2	$u^2 n_x$	v	n_y	vn_y	v^2	$v^2 n_y$
-5	3	-15	25	75	-5	6	-30	25	150
-4	5	20	16	80	-4	7	28	16	112
-3	8	24	9	72	-3	5	15	9	45
-2	12	24	4	48	-2	19	38	4	76
-1	30	30	1	30	-1	31	31	1	31
0	44	-113			0	61	-142	0	
1	42	+42	1	42	1	31	31	1	31
2	26	52	4	104	2	24	48	4	96
3	20	60	9	180	3	8	24	9	72
4	8	32	16	128	4	4	16	16	64
5	6	30	25	150	5	5	25	25	125
6	10	60	36	360	6	14	84	36	504
	214	276		1269		214	228		1306
		163					86		

$$\bar{u} = \frac{163}{214} = 0,76 ;$$

$$\bar{v} = \frac{86}{214} = 0,40 ;$$

$$\sigma_u^2 = \frac{1269}{214} - 0,76^2 = 5,36 ; \quad \sigma_v^2 = \frac{1306}{214} - 0,40^2 = 5,94.$$

Summa $\sum f_{ij} u_i v_j$ leidmiseks saab siin edukalt kasutada nn. „nelja välja“ meetodit. Suuruste u ja v nullväärtustele vastava rea ja veeru „väljalõikamise“ teel jaotub sageduste f_{ij} „väli“ neljaks (lähtetabelis on need väljad märgitud rooma numbritega). „Väljadel“ I ja IV on korrutised $u_i v_j$ positiivsed, II ja III - negatiivsed. Algul leiame $u_i v_j$ „väljade“ kaupa, seejärel $f_{ij} u_i v_j$ ning siis summeerime. Sagedeli märgitakse korrutiste $u_i v_j$ väärtused korrelatsioonitabelisse f_{ij} juurde, et kergendada edasisi arvutusi (näiteks on seda tehtud II välja jaoks).

Nüüd saame väljajade kaupa (otstarbekas on seda teha veergudes):

$$I \quad 5 \cdot 5 \cdot 2 + 5 \cdot 4 \cdot 1 + 4 \cdot 5 \cdot 1 + 4 \cdot 4 \cdot 2 + 4 \cdot 2 \cdot 1 + 4 \cdot 1 \cdot 1 + 3 \cdot 5 \cdot 3 + \\ + 3 \cdot 4 \cdot 3 + 3 \cdot 3 \cdot 1 + 3 \cdot 2 \cdot 1 + 2 \cdot 3 \cdot 1 + 2 \cdot 2 \cdot 1 + 2 \cdot 1 \cdot 5 + 1 \cdot 3 \cdot 1 + \\ + 1 \cdot 2 \cdot 6 + 1 \cdot 1 \cdot 5 = 270 ;$$

$$IV \quad 1 \cdot 1 \cdot 6 + 1 \cdot 2 \cdot 4 + \dots + 6 \cdot 5 \cdot 1 + 6 \cdot 6 \cdot 4 = 646 ;$$

$$II \quad -(2 \cdot 2 + 1 \cdot 5 + 2 \cdot 2 + 4 \cdot 1 + 6 \cdot 1) = -23 ;$$

$$III \quad -(1 \cdot 4 \cdot 1 + 1 \cdot 2 \cdot 3 + 1 \cdot 1 \cdot 9 + 2 \cdot 2 \cdot 1 + 3 \cdot 1 \cdot 4 + 6 \cdot 1 \cdot 1) = -41.$$

Seega

$$\sum f_{1j} u_{1j} v_j = 270 + 646 - (23 + 41) = 852.$$

Nüüd saame

$$r = \frac{\frac{1}{214} \cdot 852 - 0,76 \cdot 0,40}{5,36 \cdot 5,94} = \frac{3,68}{5,65} = 0,65.$$

$r = \frac{\frac{1}{n} \sum_{i,j} x_i y_j - \bar{x} \bar{y}}{s_x s_y} \quad x = u \quad y = v$

Tulemus näitab, et vaadeldud ettevõttes on tööjõu kulu ja tootmisplaanide täitmise protsendi vahel keskmise tugevusega seos. Harilikult loetakse seose tugevust keskmiseks, kui r on 0,5 - 0,7 piires ja tugevaks üle selle.

7. Tinglikud jaotused ja nende karakteristikud

Statistilisele sõltuvusele on iseloomulik, et ühe suuruse igale kindlale väärtusele vastab teise suuruse väärtuste jaotus. Neid jaotusi nimetatakse tinglikeks jaotusteks. Tinglike jaotuste karakteristikud kannavad tinglike karakteristikute nime (näit. tinglikud keskmised, tinglikud dispersioonid jne.) vastukaaluks üldkarakteristikutele.

Selgitame neid mõisteid järgmise näitega.

Korrelatsioonitabelis on esitatud 50 4. kl. õpilase pikkuse X (cm) ja kaalu Y (kg) vaheline seos:

X \ Y	24	25	26	27	28	29	30	n_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
	2	5	9	12	12	7	3	

Kasvule $x_4 = 128$ cm vastav kaalu Y tinglik jaotus esitub variatsioonreana

Y	25	26	27	28
f	1	3	5	1

Kaalule $y_6 = 29$ kg vastav pikkuse tinglik jaotus on

X	129	130	131	132
f	1	2	3	1

Analoogiliselt võib esitada ka ülejäänud X (vastavalt Y) tabelis toodud väärtustele vastavad Y (X) tinglikud jaotused.

Leiame väärtusele $x_4 = 128$ vastava Y tingliku jaotuse tingliku keskmise \bar{y}_4 ja dispersiooni.

y	f	y-f	$y-\bar{y}_4$	$(y-\bar{y}_4)^2$	$f(y-\bar{y}_4)^2$
25	1	25	-1,6	2,56	2,56
26	3	78	-0,6	0,36	1,08
27	5	135	0,4	0,16	0,80
28	1	28	1,4	1,96	1,96
	10	266			6,40

$$\bar{y}_4 = \frac{266}{10} = 26,6 \text{ (kg)} ; \quad \sigma_y^2 = \frac{6,40}{10} = 0,64 \text{ (kg}^2\text{)} .$$

Analoogiliselt saab leida ka teistele X väärtustele vastavad Y tinglikud karakteristikud ja Y väärtustele vastavad X tinglikud karakteristikud. Edaspidi 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
Y	24	25	26	27	28	29	29	30	
\bar{x}	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.

Üldjuhul toimub tinglike keskmiste leidmine järgmiste valemite järgi:

$$\bar{y}_{x_i} = \frac{1}{n_i} \cdot \sum_{j=1}^1 y_j f_{ij} \quad (i = 1, 2, \dots, k);$$

$$\bar{x}_{y_j} = \frac{1}{n_j} \cdot \sum_{i=1}^k x_i f_{ij} \quad (j = 1, 2, \dots, l).$$

Üldkeskmisi saab leida kas vahetult korrelatsioonitabeli andmetel või ka tinglike keskmiste kaudu:

$$\bar{y} = \frac{1}{N} \cdot \sum_{j=1}^1 y_j n_j = \frac{1}{N} \cdot \sum_{i=1}^1 \bar{y}_{x_i} n_i ;$$

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^k x_i n_i = \frac{1}{N} \cdot \sum_{j=1}^1 \bar{x}_{y_j} n_j .$$

Korrelatsioonivälja punkti koordinaatidega (\bar{x}, \bar{y}) nimetatakse korrelatsioonitsentriks.

Käesoleva punkti näite andmetel saame üldkeskmisteks

ja ülddispersioonideks (arvutamisel kasutame lihtsustavaid arvutusvõtteid) järgmised väärtused:

X	x_x	$u = x - 129$	ux	u^2	$u^2 x_x$	Y	y_y	$v = y - 27$	vy	v^2	$v^2 y_y$
125	1	-4	-4	16	16	24	2	-3	-6	9	18
126	3	-3	9	9	27	25	5	-2	10	4	20
127	7	-2	14	4	28	26	9	-1	9	1	9
128	10	-1	10	1	10	27	12	0	-25	0	
129	12	0	-37	0		28	12	1	+12	1	12
130	9	1	+9	1	9	29	7	2	14	4	28
131	5	2	10	4	20	30	3	3	9	9	27
132	2	3	6	9	18						
133	1	4	4	16	16			50	35		114
	50		29		144			10			
			-8								

$$\bar{x} = -\frac{8}{50} + 129 = 128,8 ; \quad \bar{y} = \frac{10}{50} + 27 = 27,2 ;$$

$$\sigma_x^2 = \frac{144}{50} - \left(-\frac{8}{50}\right)^2 = 2,85 ; \quad \sigma_y^2 = \frac{114}{50} - \left(\frac{1}{5}\right)^2 = 2,24 .$$

8. Regressioon

Korrelatsioonikordaja abil hinnatakse juhuslike suuruste vahelise seose tugevust; hinnatakse seda, kui oluline on ühe juhusliku suuruse enda muutumise osa teise juhusliku suuruse muutumises ja kui võrd on suuruste muutumine põhjustatud juhuslikest, kõrvalistest teguritest. Nimetatud probleemi uurimine on korrelatsiooniteooria esimeseks tähtsaks ülesandeks.

Asume teise ülesande käsitlemisele: juhuslikkust sisaldava seose väljendamisele funktsionaalse seose abil. Statistilise sõltuvuse täpsemal kirjeldamisel oleks ideaalne määrata ühe suuruse tinglike jaotuste jaotusseadused

(näit. jaotusfunktsioon või tõenäosuse tihedus) ning nende muutumiseadus sõltuvalt teise suuruse muutumisest. Nn. tinglike jaotusseaduste määramine on aga erakordselt keeruline. Ülesanne lihtsustub, kui tinglike jaotusseaduste asemel leida ühe suuruse tinglike jaotuste olulisemad karakteristikud ning seejärel nende karakteristikute muutumine sõltuvalt teise suuruse muutumisest.

Jaotuse üheks tähtsamaks karakteristikuks on keskaine, mis on peaaegu vaba juhuslikkusest. Seepärast on mõistlik uurida ühe suuruse tinglike keskmiste sõltuvust teise suuruse väärtustest. Seda sõltuvust nimetatakse regressiooniks ehk korrelatiivseks sõltuvuseks. Funktsiooni, mis väljendab ühe suuruse (Y) tinglike keskmiste (\bar{y}_x) sõltuvust teisest suurusest (X), nimetatakse regressioonivõrrandiks (korrelatsioonivõrrandiks)

$$\bar{y}_x = f(x) .$$

Regressioonivõrrandile vastavat graafikut nimetatakse regressioonikõveraks.

Analoogiliselt räägitakse x regressioonist y järgi ning regressioonivõrrandist $\bar{x}_y = f(y)$.

Suurust (tunnust), mis on regressioonivõrrandis argumentiks, nimetatakse faktortunnuseks. Suurust, mille tinglikud keskmised on sõltuvad faktortunnusest, nimetatakse sõltuvaks ehk resultatiivseks tunnuseks.

Kui mõlemad uuritavad tunnused alluvad ühiste faktorite mõjule, siis on otstarbekas uurida kahepoolset regressiooni. Olgu näiteks selgitamisel tarbitud tooraine maksumuse ja valmistoodangu maksumuse vahakord. 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 uuritakse ainult ühepoolset regressiooni, valides faktortunnuseks põhjuse. Näiteks lehmade vanuse ja piimaanni vahelise sõltuvuse uurimisel tuleb faktortunnuseks valida vanus.

Edaspidi kasutame regressioonivõrrandis resultatiivse

tunnuse tingliku keskmise \bar{y}_x (vastavalt \bar{x}_y) asemel sümbelit $y(x)$, s.t. regressioonivõrrandi esitame kujul $y = f(x)$ [$x = g(y)$], mõistes selle all sõltuvust $\bar{y}_x = f(x)$ [$\bar{x}_y = g(y)$].

9. Regressioonivõrrandite koostamisest

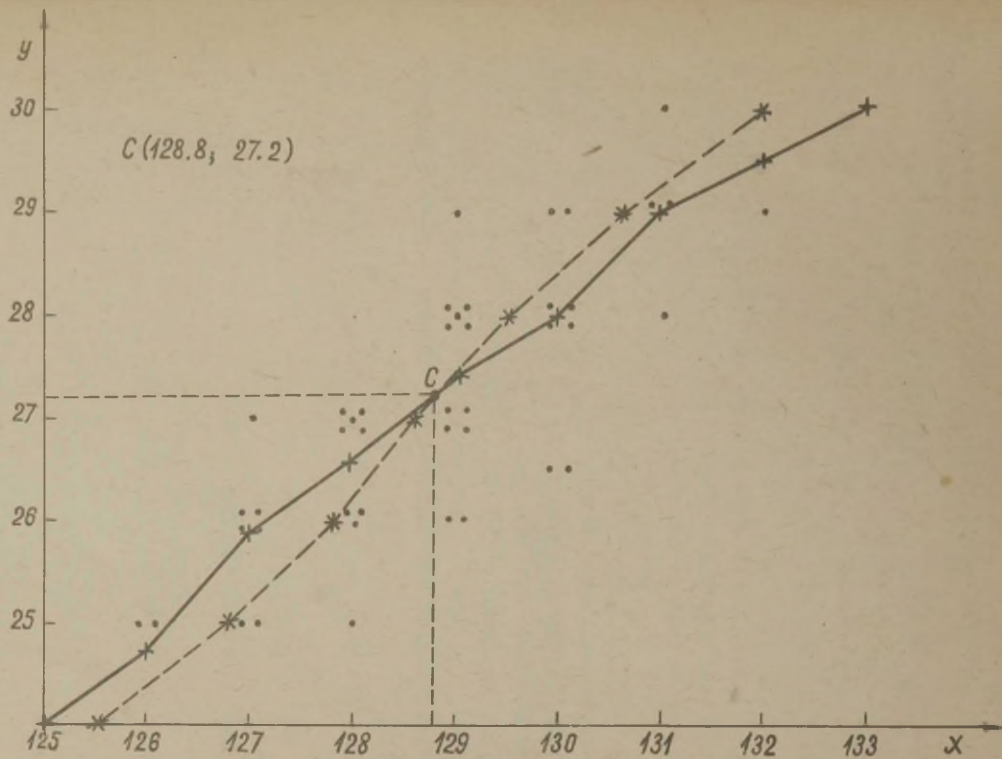
Regressioonivõrrandi koostamine toimub harilikult piiratud mahuga katseandmete põhjal. Meie käsutuses ei ole suuruste kõiki mõeldavaid väärtusi ja pealegi pole need päris täpsed. Leides tinglikud keskmised ja kandes neile vastavad punktid korrelatsiooniväljale, saame nende punktide ühendamisel sirglõikudega eksperimentaalse regressioonikõvera.

N ä i d e. Konstrueerime korrelatsioonivälja ja eksperimentaalsed regressioonikõverad 7. punkti näite andmetel.

Joonisel 24 on ristikestega märgitud \bar{y}_x asukohad ja tärnikestega \bar{x}_y . Ligikaudne regressioonikõver $y \approx f(x)$ on antud pideva, $x \approx g(y)$ katkendliku joonega.

Tinglike keskmiste asukohad pole teada faktortunnuse nendel väärtustel, mis katsetel ei realiseerunud ning regressioonikõver tuleb joonistada katkendlike andmete põhjal. Peale selle pole ka olemasolevad tinglikud keskmised täiesti vabad juhuslikest mõjudest. Seepärast pole mõtet püüda valida regressioonikõverat selliselt, et see läbiks kõiki antud punkte. Regressioonikõver valitakse nii, et see lähendaks katseandmeid teatud mõttes kõige paremini. Harilikult „teatud mõttes parim“ lähend konstrueeritakse vähimruutude meetodil: lähendav funktsioon tuleb valida selliselt, et kõvera suhtes leitud katseliste väärtuste hälvete ruutude summa ehk katseliste väärtuste hajuvus oleks minimaalne. Selle meetodi kasutamine taandub valitud funktsiooni tüübi korral funktsiooni avaldises esinevate parameetrite leidmisele nn. normaalkõrrandite süsteemist.

Praktikas tuleb kõige sagedamini kokku puutuda lineaarse ja ruutfunktsiooniga, s.t. regressioonivõrrandid vali-



Joonis 24.

takse kujul

$y = ax + b$ (sisuliselt $\bar{y}_x = ax + b!$) - lineaarne korrelatsioon;

$y = ax^2 + bx + c$ - paraboolne korrelatsioon.

10. Lineaarsed regressioonivõrrandid

Näitame, kuidas saab lineaarseid regressioonivõrrandeid koostada vahetult lähteandmete tuginedes. Suuruse y regressiooni x järgi esitame kujul

$$y = \rho_{yx} x + b,$$

kus regressioonisirge tõusu ρ_{yx} nimetatakse regressioonikordajaks (esimene indeks tähistab resultatiivset tunnust).

Olgu antud N väärtuspaari (x_1, y_1) (nende hulgas võib olla kordseid väärtusi). Tasandava lineaarfunktsiooni kordajate ρ_{yx} ja b leidmise vähimruutude meetodil toimub normaalvõrrandite süsteemist (vt. kõrgema matemaatika kursusest!):

$$\begin{cases} \rho_{yx} \cdot \sum x_1^2 + b \cdot \sum x_1 = \sum x_1 y_1; \\ \rho_{yx} \cdot \sum x_1 + b \cdot N = \sum y_1. \end{cases}$$

Lahendame süsteemi asendusvõttega:

$$b = \frac{1}{N} \sum y_1 - \frac{1}{N} \sum x_1 \cdot \rho_{yx} = \bar{y} - \bar{x} \cdot \rho_{yx};$$

$$\rho_{yx} \sum x_1^2 + (\bar{y} - \bar{x} \rho_{yx}) \cdot \sum x_1 = \sum x_1 y_1;$$

$$\rho_{yx} = \frac{\sum x_1 y_1 - \bar{y} \cdot \sum x_1}{\sum x_1^2 - \bar{x} \cdot \sum x_1}.$$

Viimase murru lugeja ja nimetaja jagamisel arvuga N saame lugejaks

$$\frac{1}{n} \sum x_1 y_1 - \bar{y} \cdot \bar{x} = c_{xy}$$

ning nimetajaks

$$\frac{1}{n} \sum x_1^2 - \bar{x}^2 = \frac{\sum x_1^2}{n} - \bar{x}^2 = \sigma_x^2.$$

Seega lõplikult on regressioonikordaja

$$\rho_{yx} = \frac{c_{xy}}{\sigma_x^2}.$$

Vabaliikme b saame arvutada ρ_{yx} kaudu. Asendades b avaldise

ρ_{yx} kaudu regressioonivõrrandisse, saame

$$y = \rho_{yx} \cdot x + \bar{y} - \bar{x} \cdot \rho_{yx}$$

ehk regressioonivõrrandi sagedamini kasutatava kuju

$$y - \bar{y} = \rho_{yx} \cdot (x - \bar{x}).$$

Analoogiliselt saab x regressiooni uurimisel y järgi

$$x - \bar{x} = \rho_{xy} \cdot (y - \bar{y}),$$

kus

$$\rho_{xy} = \frac{c_{xy}}{\sigma_y^2}.$$

Pange tähele, et regressioonikordaja avaldise nimetajas seisab faktortunnuse dispersioon.

Kui lähteandmed on esitatud korrelatsioonitabeli kujul, siis leitakse korrelatsioonimoment c_{xy} nii nagu korrelatsioonikordaja arvutamise käigus.

Kasutades arvutuste käigus abisuurusi

$$u = \frac{x - c_1}{h_1} \quad \text{ja} \quad v = \frac{y - c_2}{h_2},$$

saab regressioonikordajad järgmiselt:

$$\rho_{yx} = \frac{h_2}{h_1} \frac{c_{uv}}{\sigma_u^2}; \quad \rho_{xy} = \frac{h_1}{h_2} \frac{c_{uv}}{\sigma_v^2}$$

ning üldkeskmised

$$\bar{x} = h_1 \bar{u} + c_1 \quad \text{ja} \quad \bar{y} = h_2 \bar{v} + c_2 .$$

11. Ü l e s a n d e i d

Selgitame regressioonivõrrandite koostamist arvuliste näidete varal.

Ü l e s a n n e 1. Uurida 6. p. 1. näite andmetel kogutoodangu mahu Y lineaarset regressiooni põhivahendite maksumuse X järgi.

Kasutame r leidmise käigus saadud tulemusi. Saime korrelatsioonimomendi ja dispersiooni

$$C_{uv} = \frac{1}{15} \cdot 531 - (-0,4) \cdot 1,33 = 35,9 ; \quad \sigma_u^2 = 38,3 .$$

Seega regressioonikordaja ρ_{yx} ja üldkeskmised avalduvad järgmiselt:

$$\rho_{yx} = \frac{35,9 \cdot 1}{38,3 \cdot 1} = 0,937 ;$$

$$\bar{x} = \bar{u} + 45 = -0,40 + 45 = 44,6 ; \quad \bar{y} = \bar{v} + 44 = 1,33 + 44 = 45,3 .$$

Regressioonivõrrand omab kuju

$$y - 45,3 = 0,937(x - 44,6) .$$

Omistades x -le kindlaid väärtusi, saaksime leida y teoreetilisi tinglikke keskmisi.

Ü l e s a n n e 2. Koostada regressioonivõrrandid 7. p. näite andmetel.

Kasutame 7. p. lõpus toodud arvutustulemusi vajalike suuruste arvutamiseks:

$$\bar{u} = -\frac{8}{50} = -0,16 ; \quad \sigma_u^2 = \frac{144}{50} = 2,88 ;$$

$$\bar{v} = \frac{10}{50} = 0,20 ; \quad \sigma_v^2 = \frac{114}{50} = 2,28 .$$

Korrelatsioonimomendi avaldises esineva summa

$\sum f_{ij} u_i v_j$ leiame „nelja välja“ meetodil:

x \ y		24	25	26	27	28	29	30
		v						
y	u	-3	-2	-1	0	1	2	3
	125	-4	12		(I)			
126	-3	19	6					
127	-2		24	2	1			
128	-1		2	31	5	1		
129	0			2	4	5	1	
130	1				2	1	2	
131	2					2	4	6
132	3						6	9
133	4			(III)	(IV)			12

$$I - 1 \cdot 12 + 1 \cdot 9 + 2 \cdot 6 + 2 \cdot 4 + 4 \cdot 2 + 1 \cdot 2 + 3 \cdot 1 = 54 ;$$

$$II - -1; \quad III - 0; \quad IV - 56; \quad \sum f_{ij} u_i v_j = 54 + 56 - 1 = 109.$$

Kuna abisuuruste u ja v leidmisel $h_1 = h_2 = 1$, siis saab

$$\rho_{yx} = \frac{\frac{1}{50} \cdot 109 - (-0,16) \cdot 0,20}{2,88} = \frac{2,183}{2,88} = 0,76;$$

$$\rho_{xy} = \frac{\frac{1}{50} \cdot 109 - (-0,16) \cdot 0,20}{2,28} = \frac{2,183}{2,28} = 0,96.$$

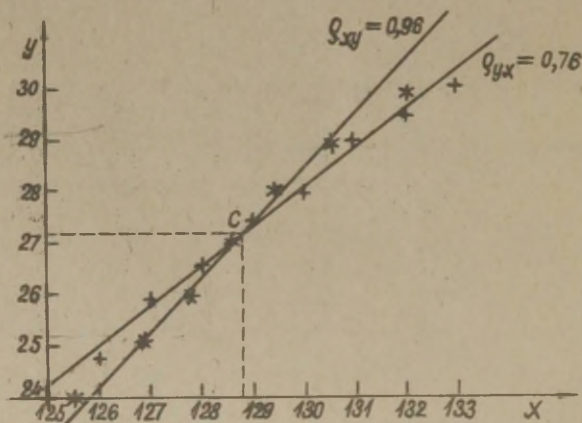
Regressioonivõrrandid esituvad kujul

$$y - 27,2 = 0,76(x - 128,8) ;$$

$$x - 128,8 = 0,96(y - 27,2) .$$

Regressioonisirged on kantud joonisele 25 (juurde on märgitud sirgete tõusud). Regressioonivõrranditest lähtudes saab leida teoreetilised tinglikud keskmised. Järgnevas tabelis on võrdluseks toodud \bar{y}_x mõõtmisandmetest lähtudes (vt. lk. 155) ja regressioonivõrrandist arvatatuna.

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
\bar{y}_{teor}	24,2	25,1	25,8	26,6	27,4	28,1	28,9	29,7	30,4



Joonis 25.

Ülesanne 3. Koostada regressioonivõrrandid
6. p. 2. näite andmetel.

Arvutustes kasutatakse abisuurus

$$u = \frac{x - 100}{10} \quad \text{ja} \quad v = \frac{y - 100}{10} \quad (h_1 = 10, h_2 = 10)$$

ning nende jaoks

$$\bar{u} = 0,76; \quad \sigma_u^2 = 5,36; \quad \bar{v} = 0,40; \quad \sigma_v^2 = 5,94;$$

$$\sum f_{ij} u_i v_j = 852.$$

Nende tulemuste abil leiame

$$r_{yx} = \frac{10}{10} \cdot \frac{\frac{1}{214} \cdot 852 - 0,76 \cdot 0,4}{5,36} = \frac{3,68}{5,36} = 0,69;$$

$$\rho_{xy} = \frac{10}{10} \cdot \frac{3,68}{5,94} = 0,62 ;$$

$$\bar{x} = 10\bar{u} + 100 = 107,6 ; \quad \bar{y} = 10\bar{v} + 100 = 104 .$$

Regressioonivõrrandid on järgmised:

$$y - 104 = 0,69 \cdot (x - 107,6) ; \quad y - 107,6 = 0,62(y - 104) .$$

12. Regressioonikordajate ja korrelatsioonikordajaseostest

Korrelatsiooni ja regressiooni käsitlemisel saime järgmised seosed:

$$r = \frac{c_{xy}}{\sigma_x \sigma_y} ; \quad \rho_{yx} = \frac{c_{xy}}{\sigma_x^2} ; \quad \rho_{xy} = \frac{c_{xy}}{\sigma_y^2} ;$$

$$y - \bar{y} = \rho_{yx}(x - \bar{x}) ; \quad x - \bar{x} = \rho_{xy}(y - \bar{y}) .$$

a) Olgu regressioonikordajad teada. Leida nende kaudu korrelatsioonikordaja.

Leiame korrutise

$$\rho_{yx} \cdot \rho_{xy} = \frac{c_{xy}^2}{\sigma_x^2 \sigma_y^2} = \left(\frac{c_{xy}}{\sigma_x \cdot \sigma_y} \right)^2 = r^2 .$$

Siit

$$r = \pm \sqrt{\rho_{yx} \cdot \rho_{xy}}$$

- korrelatsioonikordaja on võrdne regressioonikordajate geomeetrilise keskmisega; märk ruutjuure ees ühtib regressioonikordajate märgiga.

Regressioonikordajad on alati samamärgilised, olles määratud c_{xy} märgiga (ρ_{yx} nimetajas olev dispersioon on alati positiivne).

b) Olgu korrelatsioonikordaja leitud. Määrata selle

abil regressioonikordajad.

Korrutame r avaldises murru lugejat ja nimetajat ühe standardhälbega (näit. σ_x). Siis

$$r = \frac{c_{xy}}{\sigma_x \cdot \sigma_y} \cdot \frac{\sigma_x}{\sigma_y} = \frac{c_{xy}}{\sigma_x^2} \cdot \frac{\sigma_x}{\sigma_y} = \rho_{yx} \cdot \frac{\sigma_x}{\sigma_y};$$

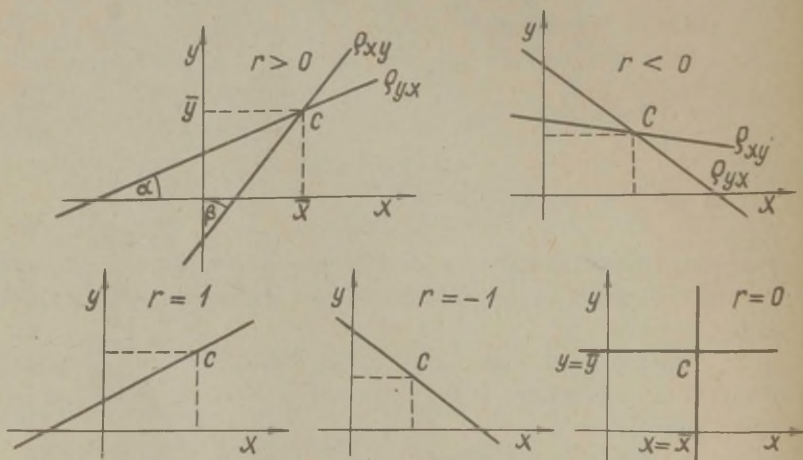
$r = \rho_{yx} \cdot \frac{\sigma_x}{\sigma_y}$, kust avaldame ρ_{yx} (korrumisel suurusega σ_y avaldame ρ_{xy}):

$$\rho_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}, \quad \rho_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}.$$

c) Regressioonisirgete vastastikuse asendi sõltuvus korrelatsioonikordajast.

Joonisel 26 on esitatud regressioonisirgete paar ja on märgitud nende tõusunurgad (nurk faktortunnuse telje ja sirge vahel) ning

$$\rho_{yx} = \tan \alpha, \quad \rho_{xy} = \tan \beta.$$



Joonis 26.

Urime Y regressiooni X järgi, s.t. regressioonisirget tõusunurgaga α :

$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x}) .$$

Sellest võrrandist nähtub, et mida väiksem $|r|$, seda väiksem on sirge kalle faktortunnuse X telje suhtes, s.t. seda vähem reageerivad tunnuse Y tinglikud keskmised tunnuse X muutumisele. Vaadeldes X regressiooni Y järgi, täheldame, et $|r|$ vähenedes väheneb regressioonisirge ja y-telje vaheline nurk ning tunnuse X tinglikud keskmised reageerivad nõrgemini tunnuse Y muutumisele ($r = 0$ korral $y = \bar{y} = \text{const}$, mis ei sõltu faktortunnuse X muutumisest). Tinglike keskmiste sõltuvus faktortunnusest on seda tugevam, mida suurem on $|r|$ ja vastavalt regressioonisirgete kalle faktortunnuste telje suhtes. Kui $r = 0$, on regressioonisirged rööbiti faktortunnuste telgedega ning seega omavahel risti. Regressioonisirgete vaheline nurk hakkab $|r|$ kasvades vähenema ning piirjuhul $|r| = 1$ jõuavad sirged ühtimiseni (X ja Y on täpselt lineaarses sõltuvuses). Väärtused $y = \bar{y}$ ja $x = \bar{x}$ rahuldavad mõlemat regressioonivõrrandit, s.t. regressioonisirged läbivad korrelatsioonitsentri.

K o k k u v ô t e. Suurustevaheline seos on seda tugevam, mida väiksem on regressioonisirgete-vaheline nurk. Regressioonisirged lõikuvad alati korrelatsioonitsentris.

Joonisel 26 on toodud mõeldavad regressioonisirgete paarid sõltuvalt regressioonikordajate märgist (järjestage need skeemid suurustevahelise seose tugevuse järgi!).

Ü l e s a n n e 1. Leida 11. p. 2. näite andmetel korrelatsioonikordaja.

Õpilaste kaalu ja kasvu seose tugevuse näitajaks saab

$$r = \sqrt{\rho_{yx} \cdot \rho_{xy}} = +\sqrt{0,76 \cdot 0,96} = 0,85.$$

Ü l e s a n n e 2. On antud regressioonivõrrandid $y - 21 = -1,5(x + 4)$ ja $x + 4 = -0,6(y - 21)$. Määrata korrelatsioonitsenter ja korrelatsioonikordaja.

Korrelatsioonitsentri koordinaadid on $x = -4$, $y = 21$.
Korrelatsioonikordaja $r = -\sqrt{1,5 \cdot 0,6} = -0,95$.

H a r j u t u s ü l e s a n d e i d

I

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 viskel kolmega jaguva silmade arvu saamine?

3. Õpperühmas on 8 nais- ja 17 meesüliõpilast. Kui suur on naisüliõpilaste suhteline sagedus? Kui tõenäone on, et üliõpilaste nimekirjas on esikohal naisüliõpilane?

4. Tehase TKO avastas 100 valmistootte hulgas 5 mitte-standardset toodet. Leida standardsete toodete suhteline sagedus.

5. Märklaua tabamise suhteline sagedus on 0,85. Leida tabamuste keskmine arv, kui sooritati 120 lasku.

6. Kui tõenäone on, et uue passi saamisel passi number lõpeb paarisarvuga? nulliga?

7. Õppejõud palus kursusevanemal kutsuda konsultatsioonile 3 üliõpilast 6-st kontrollitöö 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 valikuta võtmisel I või II sordi toote saamine?

11. Loteriil võidavad kõik piletid. 200 võidu hulgas on hinnalisematest võitudest 10 täitesulepead, 30 täitepliatsit, 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 raamav 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 kumbki aparaat ei vaja remonti garantiiaja jooksul?

17. TKO ei kontrolli valmistoodangu partiist üle 5 toote. Ühe toote kaupa kontrollides praaktoote avastamisel praagitakse välja kogu partii. Kogutoodangus on keskmiselt 4% praaki. Leida partii praagiks tunnistamise tõenäosus.

18. Autokaupluse saabus 100 autot - 70 ühest ja 30 teisest tehasest. Esimese tehase toodangust on keskmiselt 80 %, teisel 40 % tumeda polstrikattega. Kui tõenäone on juhuslikul kontrollimisel tumeda polstrikattega autole sattumine?

19. Ühes turismigrupis on võõrkeele oskajaid 3 ja mitteoskajaid 2, teises grupis vastavalt 4 ja 4. Esimesest grupist saadeti valikuta teise üks turist. Kui tõenäone on, et nüüd teisest grupist valikuta väljakutsutud turist on võõrkeele oskaja?

20. Samasooliste kaksikute sündimise tõenäosus on 2 korda suurem kui erisoolistel. Poisslapse sündimise tõenäosus on 0,5. Kui tõenäone on, et kohates perekonnas poisslapselast kaksikut ka teine laps on poiss?

21. Kastis on 4 detaili teadmata kvaliteediga. Ühe kontrollimisel selgus, et see on kvaliteetne. Detail pandi kasti tagasi. Kui tõenäone on, et kõik detailid on kvaliteetsed?

II

1. Koostada 3 täringu viskel saadavate silmade summa jaotustabel ja polügoon.

2. Koostada kaupluses päeva jooksul müüdud 45 kingipaari suuruste variatsioonrida, statistiline jaotustabel 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

5. Juhusliku suuruse tõenäosuse tiheduseks on $p(x) = \frac{1}{\pi(1+x^2)}$. Kui suur on tõenäosus suuruse sattumiseks vahemikku $(-1, 1)$?

III

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 standardhälve.

3. Leida § 2 ülesande 2 andmetel kingapaaride suuruse aritmeetiline keskmine, mood, mediaan ja standardhälve.

4. Leida § 2 ülesande 3 andmetel traktoripargi keskmine võimsus ja 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.

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. Kaugõppiija 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 pääseb vabariigi koondvõistkonda?

4. Kogutoodangus on keskmiselt 3 % praaki. Leida tõenäosena 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 suhtelise sageduse hälve sündmuse tõenäosuse $p = 0,40$ suhtes ei ületa 0,04 usaldusnivooga 0,90?

7. Mitu seemet tuleks idanemisproovi tegemiseks võtta, et võiks väita: idanemisvõimeliste seemnete suhteline sagedus ei erine 0,9-st rohkem kui 0,02 võrra usaldusnivooga 0,977?

8. Kontrolliti 3000 elektripirni. Praagi suhteline sagedus on 0,15. Kui tõenäone on, et praagi suhteline sagedus kontrollitud pirnide hulgas ei erine 0,15-st rohkem kui 0,01 võrra?

9. Sündmuse tõenäosus üksikkatsel on 0,3. Leida sündmuse suhtelise sageduse piirviga tõenäosuse 0,3 suhtes, mida võib garanteerida usaldusnivooga 0,95.

10. Tõenäosus lennuki tabamiseks õhutõrjekahurist ühe lasuga on 0,01. Sooritatakse 100 lasku. Leida tõenäosus len-

nuki tabamiseks 2 mürsuga Poissoni, Bernoulli ja Laplace'i valemite abil.

11. Juhuslik suurus X allub normaaljaotusele, mille parameetrid on $m_0 = 5$ ja $\sigma = 2$. Kui tõenäone on, et X väärtused asuvad vahemikus (1, 10)?

12. Normaaljaotusele alluva juhusliku suuruse parameetrid on $m_0 = 3$ ja $\sigma = 0,5$. Kui tõenäone on, et X väärtused ei hälbi keskvaartuse 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.

VI

1. Toodangu kvaliteedi hindamiseks moodustati 10 000 tootest koosneva üldkogumi jaoks valim mahuga $N = 300$. Esimese sordi toodangu suhteline sagedus valimis oli 0,6. Määrata esimese sordi toodangu usalduspiirid üldkogumis usaldusnivooga 0,999.

2. Kartuli tärklisesisalduse selgitamiseks tärklise hulk 560 kartulil. Mõõtmistul riatsioonreana

Tärklise hulk %	(9, 10)	(10, 11)	(11, 12)
Kartulite arv	1	2	6

Tärkl. %	(14, 15)	(15, 16)	(16, 17)	(17, 18)	(18,
Kartul. arv	26	46	94	106	110

76

Tärgl. %	(20, 21)	(21, 22)	(22, 23)	(23, 24)	(24, 25)	(25, 26)
Kartul. arv	39	14	7	1	0	1

Määrata üldkeskmise usalduspiirid usaldusnivooga 0,95.

3. Kvalifitseeritud tööliste keskmise tööstaazi selgitamiseks kontrolliti 10 000 tööliste tööraamatuid korduva valimi teel. Keskmiseks tööstaaziks saadi 15,5 aastat standardhälbega 2,2 aastat. Kui tõenäone on, et üldkeskmise erinevus valimi keskmisest rohkem kui 0,5 aastat?

4. Uuritakse telefonikõnede keskmist pikkust. Mitme kõne pikkus tuleb registreerida, et tõenäosusega 0,997 võiks garanteerida: valimi keskmise ja üldkeskmise erinevus ei ületa 10 sekundit standardhälbega 2,5 minutit?

5. Raamatukogude lugejate keskmise vanuse määramiseks kasutatakse valimimeetodit. Mitu lugejakaarti tuleks 30 000 lugejakaardi hulgast kontrollimiseks võtta, et usaldusnivooga 0,99 võiks garanteerida: üldkeskmise ja valimi keskmise erinevus ei ületa 1 aastat standardhälbega 5 aastat?

VII

1. On antud andmed 1950. a. seisuga 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 korrelatsioonikordajad.

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:

$$X - \text{cm}; \quad Y - \text{m}^3/\text{s}.$$

$y \backslash x$	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 korrelatsioonikordaja ja koostada regressioonivõrrandid.

3. 30 kolhoosi kartulisaagi X ($\frac{tS}{ha}$) ja tsentneri omahinna Y (rbl.) jaotus on antud korrelatsioonitabeliga

$y \backslash x$	160-190	190-220	220-250	250-280
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 korrelatsioonikordaja.

Vastused

I

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. 0,86
16. 0,853
17. 0,18
18. 0,68
19. 0,51

III

1. $E(x)=1,65$; $D(x)=1,15$
2. $E(x)=13,928$; $D(x)=28,345$
4. $\bar{x} = 1890$
5. $\bar{x} = 70$, $Mo = 65$, $Me = 69$.

IV

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. $m_0 = 86,3,$
 $\sigma^2 = 1802,2.$

VI

1. $0,6 \pm 0,094$
2. $\Delta = 0,19$
3. 0,8444
4. 1985
5. 166

VII

1. $\rho_{yx} = 2,5,$
 $\rho_{xy} = 0,384,$
 $\rho_{xz} = 0,76,$
 $\rho_{zx} = 1,32$
2. $r = 0,896$
3. $r = -0,685.$

T ä h t s a m a d v a l e m i d

Valemite numeratsioonid tähistab 1. number konsepti paragrahvi, kust valem on pärit.

Tähistusi

p, P	- sündmuse tõenäosus
m	- sündmuse sagedus
n	- elementaarsündmuste koguarv; binoomjaotusel ka katsete üldarv
w	- suhteline sagedus
x_1, y_1	- juhusliku suuruse üksikväärtused
f_1	- üksikväärtuste sagedused
N	- kogumi maht
m_0	- keskvääratus
D, σ^2	- dispersioon
t	- normeeritud hälve
\bar{x}, \bar{y}	- suuruste x ja y aritmeetilised keskmised
a	- esialgne keskmine
Me	- mediaan
Mo	- mood
F	- jaotusfunktsioon
ϕ	- Laplace'i funktsioon
Δ	- piirviga
C_{xy}	- korrelatsioonimoment
r	- korrelatsioonikordaja

Tõenäosus

Klassikaline tõenäosus

$$P(A) = \frac{m_A}{n},$$

kus m_A on sündmuse jaoks soodsate juhusete arv.

Suhteline sagedus

$$w = \frac{m}{n}.$$

Tõenäosuste liitmise lause

Kui A, B on suvalised sündmused, siis

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.3)$$

Väljavõetavate sündmuste juhul

$$P(A \cup B) = P(A) + P(B). \quad (1.4)$$

Tõenäosuste korrutamise lause

Kui A, B on suvalised sündmused, siis

$$P(A \cap B) = \begin{cases} P(A) \cdot P(B/A); \\ P(B) \cdot P(A/B). \end{cases} \quad (1.5)$$

Sõltumatute sündmuste juhul

$$P(A \cap B) = P(A) \cdot P(B). \quad (1.6)$$

Täistõenäosuse valem

Kui H_1 ($i = 1, 2, \dots, n$) moodustavad täieliku sündmuste süsteemi, siis

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A/H_i). \quad (1.7)$$

Bayesi valem

Hüpoteesi H_1 tõenäosus tingimusel A

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{\sum_{i=1}^n P(H_i) \cdot P(A/H_i)}. \quad (1.8)$$

Juhusliku suuruse X jaotusfunktsioon

$$F(x) = P(X < x), \quad (2.1)$$

tõenäosuse tihedus

$$p(x) = F'(x), \quad (2.2)$$

antud vahemikku (x_1, x_2) sattumise tõenäosus

$$P(x_1 < X < x_2) = \begin{cases} F(x_2) - F(x_1), \\ \int_{x_1}^{x_2} p(x) dx. \end{cases} \quad (2.3)$$

Juhusliku suuruse karakteristikud

Keskvaartuse valem

diskreetse juhusliku suuruse jaoks

$$E(x) = \sum_{i=1}^n x_i p_i ; \quad (3.1)$$

pideva juhusliku suuruse jaoks

$$E(x) = \int_{-\infty}^{+\infty} x p(x) dx . \quad (3.2)$$

Aritmeetilise keskmise valem

statistilise rea x_1, x_2, \dots, x_N korral

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^N x_i ; \quad (3.3)$$

variatsioonrea korral

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^n x_i f_i . \quad (3.4)$$

Valemid arvutuste lihtsustamiseks

$$\bar{x} = \frac{k}{\frac{N}{l}} \cdot \sum_{i=1}^n \frac{x_i - a}{k} \cdot \frac{f_i}{l} + a \quad (3.6)$$

(k ja l on sobivalt valitud arvud; erijuhtudel võib olla kas $a = 0$ või $l = 1$ või $k = 1$).

Keskmise hälbe valem

jaotustabeli korral

$$d = \sum_{i=1}^n |x_i - m_0| p_i ; \quad (3.7)$$

variatsioonrea korral

$$d = \frac{1}{N} \sum_{i=1}^n |x_i - \bar{x}| f_i ; \quad (3.8)$$

statistilise jaotustabeli korral

$$d = \sum_{i=1}^n |x_i - \bar{x}| w_i . \quad (3.9)$$

Dispersiooni valem

diskreetse juhusliku suurus korral

$$D(x) = \sum_{i=1}^n (x_i - m_0)^2 p_i , \quad (3.10)$$

$$D(x) = \sum_{i=1}^n x_i^2 p_i - \left(\sum_{i=1}^n x_i p_i \right)^2 ; \quad (3.11)$$

pideva juhusliku suuruse korral

$$D(x) = \int_{-\infty}^{+\infty} (x - m_0)^2 p(x) dx , \quad (3.12)$$

$$D(x) = \int_{-\infty}^{+\infty} x^2 p(x) dx - \left[\int_{-\infty}^{+\infty} x p(x) dx \right]^2 \quad (3.13)$$

statistilise rea korral

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 , \quad (3.14)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2 ; \quad (3.15)$$

variatsioonrea korral

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 f_i , \quad (3.16)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n x_i^2 f_i - \bar{x}^2 ; \quad (3.17)$$

statistilise jaotustabeli korral

$$\sigma^2 = \sum_{i=1}^n (x_i - \bar{x})^2 w_i , \quad (3.18)$$

$$\sigma^2 = \sum_{i=1}^n x_i^2 w_i - \bar{x}^2 . \quad (3.19)$$

Valem arvutuste lihtsustamiseks

$$\sigma^2 = \frac{k^2}{\frac{N}{l}} \cdot \sum_{i=1}^n \left(\frac{x_i - a}{k} \right)^2 \cdot \frac{f_i}{l} - (\bar{x} - a)^2 \quad (3.20)$$

(erijuhtudel võib olla kas $a = 0$ või $l = 1$ või $k = 1$).

Standardhälve

$$\sigma = \sqrt{D(x)} \quad (3.21)$$

Mediaan

$$Me = x_{Me} + \frac{k \left(\frac{\sum f_i}{2} - f_s \right)}{f_0} , \quad (3.22)$$

kus x_{Me} , f_0 , k - vastavalt mediaanintervalli alampiir, sagedus, pikkus;

f_s - mediaanintervallile eelnevate intervallide sageduste summa.

Mood

$$Mo = x_{Mo} + \frac{k(f_0 - f_{-1})}{(f_0 - f_{-1}) + (f_0 - f_1)} , \quad (3.23)$$

kus x_{Mo} , f_0 , k - vastavalt moodintervalli alampiir, sagedus, pikkus;

f_{-1} , f_1 - moodintervallile eelneva ja järgneva intervalli sagedused.

Binomialjaotus

Sündmuse A m korda toimumise tõenäosus n katsel (sageduse m tõenäosus)

Bernoulli valemi abil

$$P_{n,m} = \frac{n!}{m! (n-m)!} p^m q^{n-m}, \quad (4.1)$$

kus $p = P(A)$, $q = 1 - p$;

asümptootilise valemi abil

$$P_{n,m} = \frac{1}{\sqrt{npq}} \varphi(t), \quad (4.2)$$

kus $t = \frac{m - np}{\sqrt{npq}}$ ja $\varphi(t)$ on normaaljaotuse tõenäosuse tihedus (vt. tabelist).

Sageduse m karakteristikud

$$E(m) = np; \quad (4.3)$$

$$D(m) = npq; \quad (4.4)$$

$$\sigma(m) = \sqrt{npq}; \quad (4.5)$$

$$Mo(m) = np \quad (\text{kui täisarv}); \quad (4.6)$$

$np - q \leq Mo(m) \leq np + p$ (täisarv sellest lõigust).

Suhtelise sageduse $w = \frac{m}{n}$ karakteristikud

$$E(w) = p; \quad (4.7)$$

$$D(w) = \frac{pq}{n}; \quad (4.8)$$

$$\sigma(w) = \sqrt{\frac{pq}{n}}. \quad (4.9)$$

Normaaljaotus

$X \sim N(m_0, \sigma)$, m_0 - keskvääratus, σ - standardhälve

Tõenäosuse tihedus erijuhul $X \sim N(0, 1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; \quad (4.10)$$

üldjuhul $X \sim N(m_0, \sigma)$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - m_0)^2}{2 \cdot \sigma^2}}; \quad (4.11)$$

$$p(x) = \frac{1}{\sigma} \varphi(t), \quad t = \frac{x - m_0}{\sigma} \quad (4.12)$$

ja jaotusfunktsioon vastavatel juhtudel

$$X \sim N(0, 1) \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx; \quad (4.13)$$

$$X \sim N(m_0, \sigma) \quad F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x - m_0)^2}{2\sigma^2}} dx; \quad (4.14)$$

$$F_x(x) = F(t), \quad t = \frac{x - m_0}{\sigma}. \quad (4.15)$$

Laplace'i funktsioon

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt; \quad (4.16)$$

$$\bar{\phi}(t) = 2 \phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt; \quad (4.17)$$

$$F(t) = 0,5 + \phi(t). \quad (4.18)$$

Juhusliku suuruse X antud vahemikku sattumise tõenäosus

$$P(x_1 < X < x_2) = \begin{cases} F(t_2) - F(t_1), \\ \Phi(t_2) - \Phi(t_1), \\ \frac{1}{2}[\bar{\Phi}(t_2) - \bar{\Phi}(t_1)], \end{cases} \quad (4.19)$$

$$t_{1,2} = \frac{x_{1,2} - m_0}{\sigma}$$

$$P(|x - m_0| < \Delta) = \bar{\Phi}(t_\Delta) = 2\Phi(t_\Delta), \quad (4.20)$$

$$t_\Delta = \frac{\Delta}{\sigma}.$$

Sageduse ja suhtelise sageduse

antud vahemikku sattumise tõenäosus n katsel:

täpne valem

$$P(a_1 \leq m \leq a_2) = \sum_{m=a_1}^{a_2} P_{n,m}; \quad (4.21)$$

asümptootiline valem (sobib suurte n väärtuste korral, $n \geq 20$).

$$P(a_1 \leq m \leq a_2) = \Phi(t_2) - \Phi(t_1); \quad (4.22)$$

$$t_i = \frac{a_i - np}{\sqrt{npq}}, \quad i = 1, 2;$$

keskväärtuse suhtes sümmeetriliste hälvete tõenäosus

$$P(|m - np| < \Delta) = \bar{\Phi}(t_\Delta), \quad t_\Delta = \frac{\Delta}{\sqrt{npq}}; \quad (4.23)$$

$$P(|w - p| < \Delta) = \bar{\Phi}(t_\Delta), \quad t_\Delta = \frac{\Delta}{\sqrt{\frac{pq}{n}}}. \quad (4.24)$$

Valimiteooria

Tähistused:

- m_0, D - üldkeskmine ja -dispersioon
- \bar{x}, σ^2 - valimi keskmine ja dispersioon
- p, w - tunnuse suhteline sagedus üldkogumis ja

valimis
 N_0, N - üldkogumi ja valimi maht

Üldkarakteristikute hinnangud

$$m_0 \approx \bar{x}, \quad p \approx w \quad (6.1)$$

lõpmatu üldkogumi korral ($N_0 \geq 10N$)

$$D \approx \frac{N}{N-1} \sigma^2; \quad (6.2)$$

lõpliku üldkogumi korral ($N_0 < 10N$)

$$D \approx \left(1 - \frac{1}{N_0}\right) \frac{N}{N-1} \sigma^2. \quad (6.3)$$

Keskmise hajuvus $D(\bar{x})$

lõpmatu üldkogumi korral

$$D(\bar{x}) = \frac{D}{N}, \quad D(\bar{x}) \approx \frac{\sigma^2}{N-1}, \quad (6.4)$$

lõpliku üldkogumi korral

$$D(\bar{x}) = \left(1 - \frac{N}{N_0}\right) \frac{D}{N}, \quad D(\bar{x}) \approx \left(1 - \frac{N}{N_0}\right) \frac{\sigma^2}{N-1}; \quad (6.5)$$

valimi keskmise standardhälve

$$\mu = \sqrt{D(\bar{x})}. \quad (6.6)$$

Üldkeskmise usalduspiiride $\bar{x} \pm \Delta$ või usaldusnivoo P leidmine

1) antud usaldusnivoo $P = \bar{\Phi}(t)$ järgi leida tabelist t ja piirviga

$$\Delta = t\mu; \quad (6.7)$$

2) antud piirvea Δ korral leida

$$t = \frac{\Delta}{\mu} \quad (6.8)$$

ja tabelist usaldusnivoo $P = \bar{\Phi}(t)$.

Suhtelise sageduse hajuvus $D(w)$

lõpmatu üldkogumi korral

$$D(w) = \frac{w(1-w)}{N}; \quad (6.9)$$

lõpliku üldkogumi korral

$$D(w) = \left(1 - \frac{N}{N_0}\right) \cdot \frac{w(1-w)}{N}; \quad (6.10)$$

valimi suhtelise sageduse standardhälve

$$\mu = \sqrt{D(w)}. \quad (6.11)$$

Suhtelise sageduse p usalduspiirid $w \pm \Delta$

(vt. üldkeskmise usalduspiirid (6. 7.-8.).

Valimi maht

valim tehtud lõpmatust üldkogumist

$$N = \frac{t^2 \cdot D}{\Delta^2}; \quad (6.12)$$

valim tehtud lõplikust üldkogumist

$$N = \frac{t^2 D N_0}{t^2 D + \Delta^2 N_0}. \quad (6.13)$$

Korrelatsiooniteooria

Korrelatsioonikordaja

$$r = \frac{C_{xy}}{\sigma_x \sigma_y} = \pm \sqrt{\rho_{yx} \cdot \rho_{xy}}. \quad (7.1)$$

Korrelatsioonimoment

Lihtsa korrelatsioonitabeli abil

$$C_{xy} = \begin{cases} \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}), \\ \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x} \cdot \bar{y}; \end{cases} \quad (7.2)$$

korrelatsioonitabeli abil

$$\begin{aligned}
 C_{xy} &= \frac{1}{N} \sum_{i,j=1}^{k,l} f_{ij} x_i y_j - \bar{x} \bar{y} = \\
 &= \frac{1}{N} \sum_{i,j=1}^{k,l} (x_i - \bar{x})(y_j - \bar{y}) f_{ij} . \quad (7.3)
 \end{aligned}$$

Regressioonivõrrandid (lineaarsed)

$$\begin{aligned}
 y - \bar{y} &= \rho_{yx}(x - \bar{x}) ; \\
 x - \bar{x} &= \rho_{xy}(y - \bar{y}) . \quad (7.4)
 \end{aligned}$$

Regressioonikordajad

$$\rho_{yx} = \frac{C_{xy}}{\sigma_x^2} ; \quad \rho_{xy} = \frac{C_{xy}}{\sigma_y^2} . \quad (7.5)$$

Abisuuruste $u = \frac{x - c_1}{h_1}$, $v = \frac{y - c_2}{h_2}$ kasutamisel saadakse

regressioonivõrrandis esinevad suurused järgmiselt:

$$\begin{aligned}
 \rho_{yx} &= \frac{h_2}{h_1} \cdot \frac{C_{uv}}{\sigma_u^2} ; \quad \rho_{xy} = \frac{h_1}{h_2} \cdot \frac{C_{uv}}{\sigma_v^2} ; \\
 \bar{x} &= h_1 \bar{u} + c_1 ; \quad \bar{y} = h_2 \bar{v} + c_2 . \quad (7.6)
 \end{aligned}$$

Normaalse juhusliku suuruse $X \sim N(0, 1)$
tõenäosuse tihedus

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

x	0	1	2	3	4	5	6	7	8	9	x
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973	0,0
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918	0,1
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825	0,2
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697	0,3
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538	0,4
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352	0,5
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144	0,6
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920	0,7
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685	0,8
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444	0,9
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203	1,0
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965	1,1
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736	1,2
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518	1,3
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315	1,4
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127	1,5
1,6	1109	1092	1074	1057	1040	1023	1006	989	973	957	1,6
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804	1,7
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669	1,8
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551	1,9
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449	2,0
2,1	0410	0431	0422	0413	0404	0396	0388	0379	0371	0363	2,1
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290	2,2
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229	2,3
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180	2,4
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139	2,5
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107	2,6
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081	2,7
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061	2,8
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046	2,9
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034	3,0
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025	3,1
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018	3,2
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013	3,3
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009	3,4
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006	3,5
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004	3,6
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003	3,7
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002	3,8
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001	3,9
x	0	1	2	3	4	5	6	7	8	9	x

L i s a 4

Normaalse juhusliku suuruse $X \sim N(0,1)$

jaotusfunktsioon

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

t	0	1	2	3	4	5	6	7	8	9
-0,0	,5000	,4960	,4920	,4880	,4840	,4801	,4761	,4721	,4681	,4641
-0,1	,4602	,4562	,4522	,4483	,4443	,4404	,4364	,4325	,4286	,4247
-0,2	,4207	,4168	,4129	,4090	,4052	,4013	,3974	,3936	,3897	,3859
-0,3	,3821	,3783	,3745	,3707	,3669	,3632	,3594	,3557	,3520	,3483
-0,4	,3446	,3409	,3372	,3336	,3300	,3264	,3228	,3192	,3156	,3121
-0,5	,3085	,3050	,3015	,2981	,2946	,2912	,2877	,2843	,2810	,2776
-0,6	,2743	,2709	,2676	,2643	,2611	,2578	,2546	,2514	,2483	,2451
-0,7	,2420	,2389	,2358	,2327	,2297	,2266	,2236	,2206	,2177	,2148
-0,8	,2119	,2090	,2061	,2033	,2005	,1977	,1949	,1922	,1894	,1867
-0,9	,1841	,1814	,1788	,1762	,1736	,1711	,1685	,1660	,1635	,1611
-1,0	,1587	,1562	,1539	,1515	,1492	,1469	,1446	,1423	,1401	,1379
-1,1	,1357	,1335	,1314	,1292	,1271	,1251	,1230	,1210	,1190	,1170
-1,2	,1151	,1131	,1112	,1093	,1075	,1056	,1038	,1020	,1003	,0985
-1,3	,0968	,0951	,0934	,0918	,0901	,0885	,0869	,0853	,0838	,0823
-1,4	,0808	,0793	,0778	,0764	,0749	,0735	,0721	,0708	,0694	,0681
-1,5	,0668	,0655	,0643	,0630	,0618	,0606	,0594	,0582	,0571	,0559
-1,6	,0548	,0537	,0526	,0516	,0505	,0495	,0485	,0475	,0465	,0455
-1,7	,0446	,0436	,0427	,0418	,0409	,0401	,0392	,0384	,0375	,0367
-1,8	,0359	,0351	,0344	,0336	,0329	,0322	,0314	,0307	,0301	,0294
-1,9	,0288	,0281	,0274	,0268	,0262	,0256	,0250	,0244	,0239	,0233
-2,0	,0228	,0222	,0217	,0212	,0207	,0202	,0197	,0192	,0188	,0183
-2,1	,0179	,0174	,0170	,0166	,0162	,0158	,0154	,0150	,0146	,0143
-2,2	,0139	,0136	,0132	,0129	,0125	,0122	,0119	,0116	,0113	,0110
-2,3	,0107	,0104	,0102	,0099	,0096	,0094	,0091	,0089	,0087	,0084
-2,4	,0082	,0080	,0078	,0075	,0073	,0071	,0069	,0068	,0066	,0064
-2,5	,0062	,0060	,0059	,0057	,0055	,0054	,0052	,0051	,0049	,0048
-2,6	,0047	,0045	,0044	,0043	,0041	,0040	,0039	,0038	,0037	,0036
-2,7	,0035	,0034	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026
-2,8	,0026	,0025	,0024	,0023	,0023	,0022	,0021	,0021	,0020	,0019
-2,9	,0019	,0018	,0018	,0017	,0016	,0016	,0015	,0015	,0014	,0014
t = -3,0	-3,1	-3,2	-3,3	-3,4	-3,5	-3,6	-3,7	-3,8	-3,9	
F(t) = ,0013	,0010	,0007	,0005	,0003	,0002	,0002	,0001	,0001	,0000	,0000

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986

$t = 3,0$	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$F(t) = ,9987$,9990	,9993	,9995	,9997	,9998	,9998	,9999	,9999	1,0000

Kirjandust

1. A. Lepamaa. Tõenäosusteooria ja matemaatiline statistika. Tallinn, TPI rotaprint, 1968.
2. A. Lepamaa. Metoodiline juhend tõenäosusteooria ülesannete lahendamiseks. Tallinn, TPI rotaprint, 1971.
3. E. Tiit. Matemaatiline statistika I. Tartu, TRÜ rotaprint, 1971.
4. E. Tiit. Matemaatilise statistika tabelid I. Tartu, TRÜ rotaprint, 1971.
5. E. Tiit. Tõenäosusteooria I. Tartu, TRÜ rotaprint, 1970.
6. В. А. Белинский, И. Л. Калихман и др. Высшая математика основами математической статистики. Москва, 1965.
7. И. Ф. Суворов. Краткий курс высшей математики (для экономических вузов). Москва, 1961.
8. Э. С. Маркович. Курс высшей математики с элементами теории вероятностей и математической статистики. Москва, 1972.

S i s u k o r d

	lk.
Saateks	3
① Tõenäosusteooria ja matemaatilise statistika aine	4
1. TÕENÄOSUSTEORIA ALGMÕISTED	6
② 1. Sündmused	6
2. Sündmuste liigitamisest	7
③ 3. Sündmuse tõenäosus	10
4. Ülesandeid	13
④ 5. Tõenäosuse omadusi	14
⑤ 6. Sündmuste summa ja korrutis	15
④ 7. Tõenäosuste liitmislause	17
⑤ 8. Tõenäosuste korrutamislause	19
⑥ 9. Täistõenäosuse valem. Bayesi valem	21
10. Ülesandeid	24
2. JUHUSLIKE SUURUSTE JAOTUSSEADUSED	31
① 1. Juhuslikud suurused	31
② 2. Juhusliku suuruse jaotustabel	33
3. Variatsioonrida. Statistiline jaotus- tabel	35
4. Polügoon ja histogramm	39
③ 5. Juhusliku suuruse jaotusfunktsioon	42
6. Juhusliku suuruse antud vahemikku sat- tumise tõenäosus ja leidmine jaotus- funktsiooni kaudu	47
④ 7. Tõenäosuse tihedus	48
8. Juhusliku suuruse antud vahemikku sat- tumise tõenäosuse leidmine tõenäosuse tiheduse kaudu	51
9. Ühtlane jaotus	52

3. JUHUSLIKU SUURUSE ARVULISED KARAKTERISTIKUD . . .	55
1. Arvuliste karakteristikute eesmärk . . .	55
2. Keskväärtus	55
3. Aritmeetiline keskmine	58
4. Keskväärtuse omadusi	61
5. Aritmeetilise keskmise arvutusvõtteid . . .	65
6. Mediaan	68
7. Mood	70
8. Keskmine hälve	71
9. Dispersioon	74
10. Dispersiooni omadusi	76
11. Momendid	78
12. Dispersiooni leidmine statistiliste jaotuste korral	81
4. JUHUSLIKU SUURUSE JAOTUSI	85
1. Korduvate katsete skeem	85
2. Bernoulli valem	86
3. Binoomjaotus	88
4. Binoomjaotuse karakteristikud	91
5. Asümptootiline valem sageduse tõenäosuse leidmiseks	93
6. Normaaljaotuse tihedus	95
7. Normaalne jaotusfunktsioon. Laplace'i funktsioon	98
8. Normaaljaotusega suuruse antud vahemikku sattumise tõenäosus	101
9. Keskväärtuse suhtes sümmeetriliste hälvete tõenäosus. „Kolme sigma” reegel . . .	103
10. Sündmuse sageduse antud vahemikku sattumise tõenäosus	104
11. Tüüpülesandeid	106
12. Poissoni jaotus	111
13. Statistiliste jaotuste tasandamine . . .	113
14. Tasandamine normaaljaotusega	114

5. SUURTE ARVUDE SEADUS	118
1. Keskmise stabiilsusest	118
2. Bernoulli ja Tšebôšovi teoreemid	119
3. Ljapunovi teoreemi sisust	121
4. Suurte arvude seaduse praktiline tähtsus	122
6. VALIMITEORIA MATEMAATILISED ALUSED	124
1. Valimimeetod	124
2. Statistilise hinnangu mõiste	126
3. Üldkeskmise ja tunnuse tõenäosuse hinnangud	128
4. Ülddispersiooni hinnang	129
5. Valimi keskmine standardhälve	130
6. Keskväärtuse ja tõenäosuse usaldus- piirkonna leidmine	132
7. Näiteid	133
8. Valimi mahu määramine	135
9. Üldisi järeldusi	136
7. KORRELATSIOONITEORIA ELEMENTE	138
1. Statistiline sõltuvus	138
2. Lähteandmete esitusviise	141
3. Korrelatsioonikordaja	144
4. Korrelatsioonikordaja omadusi	145
5. Korrelatsioonikordaja arvutamisest	147
6. Ülesandeid	149
7. Tinglikud jaotused ja nende karak- teristikud	153
8. Regressioon	156
9. Regressioonivõrrandite koostamisest	158
10. Lineaarsed regressioonivõrrandid	160
11. Ülesandeid	162
12. Regressioonikordajate ja korrelat- sioonikordaja seosest	165

Harjutusülesandeid	168
Vastused	175
Tähtsamad valemid	177
Tõenäosus	177
Juhusliku suuruse karakteristikud	179
Keskmise hälve	179
Binoomjaotus	182
Normaaljaotus	182
Valimiteooria	184
Korrelatsiooniteooria	186
Kirjandust	193

К. Соонегс
ТЕОРИЯ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
(Для студентов экономического факультета)

Третье, дополненное издание
На эстонском языке

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

Vastutav toimetaja Ü. Lepik
Korrektor V. Lang

Paljundamisle antud 19.12.73. Trükipaber nr.2.
30x42. 1/4. Trükipoognaid 12,25. Tingtrükipoog-
naid 11,39. Arvestuspoognaid 7,08. Trükiarv 500.

MB 11167. Tell. nr.111.

TRÜ rotaprint 1974. ENSV, Tartu, Pälsoni tn. 14.

Hind 20 kop.