SISEKAITSEAKADEEMIA

STATISTIKA

Kolmas trükk

HELMO KÄERDI

Õppematerjali 3. trükk on mõeldud statistika õpetamiseks Sisekaitseakadeemia õppekavade järgi, kuid mõnedel erialadel on kogu ainemahu sisu omandamiseks vajalik lisaks kasutada ka teisi Sisekaitseakadeemia väljaandeid: H. Käerdi „Nähtustevaheliste seoste uurimine" (2001 ja 2005) ning H. Käerdi „Statistika ja tõenäosusteooria alused" (1997, 1999).

Õppematerjali 2000. aastal ilmunud esmatrükki täiendati 2002. aastal näitülesannetega ja iseseisvaks lahendamiseks mõeldud ülesannetega. On kasutatav ka teistes kõrgkoolides, kus statistikat ei õpetata põhiainena.

ISBN 9985-67-128-7

Sisekaitseakadeemia
Kase 61 12012 Tallinn
november 2005

Sissejuhatus

Kvalifitseeritud spetsialist peab valdama statistika ja tõenäosusteooria põhimõisteid ning analüüsimeetodeid ja oskama neid vabalt ning loominguliselt kasutada. Senitehtu adekvaatseks hindamiseks ja perspektiivsete töösuundade kavandamiseks peab olema võimalikult palju informatsiooni oma tegevusala iseloomustavate karakteristikute kohta. Enamik informatsioonist on kvantitatiivne, arvudes väljendatud ja töötlemata kujul on sellest raske ülevaadet saada. Statistika ülesandeks ei ole ainult katse- või vaatlusandmete kogumine ja klassifitseerimine, nagu seda tavakäsitluses mõnikord mõistetakse. Niisugune kirjeldav statistika on vaid põhjalikuma andmeanalüüsi esimeseks etapiks. Sellele järgnevates uuringutes eraldatakse nähtust iseloomustavad olulised tunnusjooned, kontrollitakse nähtusele antud hinnangute ning püstitatud hüpoteeside täpsust ja usaldusväärsust ning prognoositakse selle nähtuse võimalikku edasist käitumist.

Suurt osa andmetest mõjutavad juhuslikud faktorid. Juhuslikkuse astme iseloomustamiseks kasutatakse tõenäosuse mõistet. Tõenäosusteooria uurib massilistes või küllalt sagedasti korduvates juhuslikes nähtustes esinevaid seaduspärasusi, seob nende nähtuste teoreetilisi ja eksperimentaalseid karakteristikuid ning on seega statistika meetodite aluseks. Ametlik statistika põhineb enamasti kõiksel uuringul, kuid suure töömahukuse tõttu piirdutakse sageli üldkogumist moodustatud juhusliku valimi vaatlemisega. Valimi kohta teadaolevast informatsioonist lähtudes on võimalik anda hinnanguid üldkogumi parameetrite kohta ning vastata korrektse praktilise kasutamise seisukohalt määrava tähtsusega küsimustele, mis on seotud nende hinnangute täpsuse ja usaldatavusega. Statistika võimaldab kontrollida juhuslike suuruste vahelise mittefunktsionaalse (statistilise) sõltuvuse olemasolu, uurida selle sõltuvuse iseloomu ja tugevust ning prognoosida teatava tõenäosusega juhusliku suuruse võimalikku käitumist tulevikus.

Statistilised meetodid on kasutatavad kuritegevuse, õnnetusjuhtumite, rahvastikuprotsesside ja majanduse valdkonnas tekkivate probleemide analüüsil, lahendamisel ning prognoosimisel, sotsioloogilistes uurimustes, infotöötluses, juhtimis- ja haldusalaste otsuste vastuvõtmisel.

Käesoleva väljaande 1. peatükis tutvustatakse kirjeldaval tasemel statistika põhimõisteid, aga samuti statistilise materjali graafilist esitamist ning indekseid. 2. peatükis käsitletakse tõenäosuse mõistet ning arvutamise põhivõtteid. 3. peatükis defineeritakse juhuslik suurus ning antakse teda iseloomustavaid karakteristikuid. 4. peatükis pööratakse klassikalistest jaotustest põhjalikumat tähelepanu binoomjaotusele ja normaaljaotusele. 5. peatükis vaadeldakse detailsemalt jaotuse parameetrite punkti- ja vahemikhinnanguid. 6. peatükis on esitatud mõned statistiliste hüpoteeside kontrollimise algoritmid.

Korrelatsioon- ja regressioonanalüüsi meetodite tundmaõppimiseks nõutava iseseisva töö tegemiseks on üliõpilastel tarvis lisaks käesolevale väljaandele kasutada autori poolt koostatud õppevahendit [9] või [10].

Sisukord

1. Kirjeldav statistika

1.1. Statistiliste andmete liigitus

Uurimisobjektide vaadeldavaid omadusi nimetatakse statistikas **tunnusteks**. **Tunnus varieerub**, kui ta esineb objektidel erineval määral. Tunnused jagunevad **kvantitatiivseteks ja kvalitatiivseteks**. Kvantitatiivseid tunnuseid (näiteks kaalu, pikkust) on võimalik mõõta mingis meetrilises skaalas, milles on ühik ja nullpunkt. Kui iga vaatlustulemuse saab liigitada teatavatest klassidest ühte konkreetsesse klassi, on tegemist kvalitatiivse tunnusega (näiteks professionaalne klassifikatsioon; jaotus naisteks ja meesteks).

Tunnuste varieeruvuse mõõtmiseks kasutatakse nelja skaalat, mille hierarhia (madalamast kõrgemale) on järgmine: nominaal- ehk nimiskaala, ordinaal- ehk järjestusskaala, intervallskaala ja meetriline skaala. Mida kõrgemal paikneb vaadeldav tunnus selles hierarhias, seda rohkem informatsiooni ta sisaldab.

Nominaalskaala on hierarhias madalaim ja koosneb üksteisest sõltumatutest klassidest või rühmadest, mida ei saa loogiliselt järjestada, kusjuures iga objekt saab kuuluda ainult ühte klassi. Küll aga on võimalik leida sagedused, millega teatud tunnuse variatsioonid esinevad. Nominaalskaalat kasutatakse põhiliselt kvalitatiivsete tunnuste mõõtmiseks. Kui näiteks objektid on sinised, rohelised ja punased, võib neid mõõta nominaalskaalas.

Ordinaalskaala koosneb erinevatest klassidest, mida on võimalik järjestada. Näiteks:

halb	poolt	esimene sort
keskmine	pigem poolt kui vastu	teine sort
hea	neutraalne	kolmas sort
väga hea	pigem vastu kui poolt	
	vastu	

Ordinaalskaalas jäävad klassidevahelised erinevused määratlemata ning seda skaalat kasutatakse (nagu nominaalskaalatki) põhiliselt kvalitatiivsete tunnuste mõõtmiseks.

Kahte järgmist skaalat kasutatakse kvantitatiivsete tunnuste mõõtmiseks.

Intervallskaala määrab üksikute mõõtmistulemuste vahelise kauguse, kusjuures fikseeritud nullpunkti ei eksisteeri ning suhtarve ei saa moodustada (näiteks temperatuur Celsiuse või Fahrenheiti kraadides, kusjuures ei saa öelda, et 20 ^{0}C on kaks korda soojem kui 10 ^{0}C).

Meetriline skaala määrab üksikute mõõtmistulemuste vahelise kauguse, kusjuures eksisteerib fikseeritud nullpunkt. Meetriliste andmete jaoks saab moodustada suhtarve. Näiteks 20-aastane inimene on kaks korda vanem kui 10-aastane.

Skaalade hierarhias on võimalik (mõnikord ka soovitatav) üle minna kõrgemalt skaalalt madalamale (vastupidi pole võimalik). Näiteks võib mingi piirkonna elanike vanused (meetrilises skaalas) rühmitada klassidesse:

0–20 aastat,

20–40 aastat,

40–60 aastat,

60–80 aastat,

80–100 aastat.

Kirjeldatud viisil koostatud ja järjestatud sagedustabel esitab elanike vanused ordinaalskaalas. Kui andmed olidki kogutud eeltoodud klassidesse, konkreetseid vanuseid fikseerimata, siis pole neid võimalik tagasi teisendada hierarhias kõrgemal seisvasse meetrilisse skaalasse.

1.2. Statistika mõiste. Üldkogum ja valim. Rühmitatud andmed

Üldkogumiks nimetatakse antud tunnustega elementide hulka. Turu-uuringute üldkogum on näiteks kõikide potentsiaalsete ostjate hulk. Valimiseelsetes küsitlustes on üldkogumiks kõikide valimisõiguslike elanike hulk. Tüüpiline kõikne uuring on rahvaloendus. Ka enamus ametlikust statistikast on kõikne. Suure töömahukuse, kulukuse või objektide kasutamiskõlbmatuks muutumise tõttu ei ole kõikne vaatlus alati otstarbekas. Näiteks igast 100 000 toodetud turvavööst võidakse kontrollida 10 vöö tõmbetugevust (kõikide kontrollimisel ei annaks tehas üldse toodangut). Kui üldkogumis olevate elementide arv (ehk üldkogumi maht N) on suur, siis piirdutakse sageli neist mingi juhusliku alamhulga, s.o **valimi** vaatlemisega. Tihti on valimi maht n palju väiksem üldkogumi mahust N. **Statistika üheks oluliseks ülesandeks on teha valimi põhjal otsustusi üldkogumi kohta.**

Üldkogumi parameetritele antavate hinnangute korrektsuse huvides peab valim sarnanema struktuurilt üldkogumiga, mis tagatakse valimi elementide võrdtõenäosusliku juhusliku valikuga. Valimi vajalik maht (valimi esinduslikkus ehk representatiivsus) määratakse sõltuvalt üldkogumi parameetritele antavate hinnangute täpsusest ja usaldusväärsusest.

Valimi vaatlemise teel saadud andmed moodustavad korrastamata **statistilise rea**. Kui valimi elemendid korrastada mingi arvtunnuse (näiteks turvavööde puhul tõmbetugevuse) alusel järjestatud hulka, siis saadakse **variatsioonrida**. Juhul kui valimis mahuga n on võrdseid elemente, siis esitatakse variatsioonrida kujul

x_i	x_1	x_2	...	x_m
p_i^* $\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_m}{n}$

kus $p_i^* = \frac{n_i}{n}$ on elemendi x_i sagedus ja $\sum_{i=1}^{m} n_i = n$ ning $\sum_{i=1}^{m} p_i^* = 1$.

Variatsioonrea lühendamiseks rühmitatakse valimi elemendid sageli klassidesse:

	$[a_0;\ a_1[$	$[a_1;\ a_2[$	...	$[a_{m-1};\ a_m]$
$p_i^* = \frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_m}{n}$

Võimaluse korral võetakse kõik klassid ühesuguse ulatusega

$$\Delta x = \frac{x_m - x_1}{m}.$$

ning siis tuleb $a_0 = x_1$, $a_i = a_0 + i \cdot \Delta x$, kus $i = 1, 2, \dots, m$.

Klasside arvu m valikuks võib kasutada järgnevas tabelis toodud ligikaudseid soovitusi.

Valimi maht n	Klasside arv m
alla 30	5
[30; 60[	6
[60; 130[	7
[130; 250[	8
[250; 500[	9
[500; 1000[	10
[1000; 2000[	11
[2000; 4000[	12
[4000; 8000[	13
üle 8000	14

Otstarbeka klasside arvu võib leida ka valemiga

$$m = 1 + \frac{\log n}{\log 2}.$$

1.3. Keskmised

Keskmised jagunevad mahukeskmisteks ja asendikeskmisteks. Tähistame valimi elemendid $x_1, x_2 \dots x_n$.

Mahukeskmised sõltuvad rea mahust $\sum_{i=1}^{n} x_i$ ning olulisemad neist on:

1. **Aritmeetiline keskmine**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{1}$$

Täisarvulise tunnuse aritmeetiline keskmine ei tarvitse olla täisarvuline. Selles mõttes on aritmeetilise küllalt tugev abstraktsioon.

Näide 1. Järgnevas tabelis on toodud esimese ettevõtte seitsme juhuslikult valitud töötaja kuupalgad kroonides (edaspidises analüüsis näites 1a võrreldakse neid tulemusi teise ettevõtte töötajate palkadega). Aritmeetilise keskmisena arvutatud töötaja

	1	2	3	4	5	6	7
x_i	6100	5100	5700	6500	6000	5900	5600

keskmine palk on $\bar{x} = \frac{1}{7} \sum_{i=1}^{7} x_i \approx 5842{,}9$ krooni.

Üldisemalt räägitakse **kaalutud aritmeetilisest keskmisest**

$$\bar{x}_w = \frac{\sum_{i=1}^{m} w_i x_i}{\sum_{i=1}^{m} w_i}, \tag{2}$$

kus variandid $x_1,\ x_2, \ldots, x_m$ korduvad vastavalt kaaludega $w_1,\ w_2, \ldots, w_m$.

Näide 2. Leida järgnevatel andmetel kaalutud aritmeetiline keskmine.

Piirkond	Töötute protsent	Töövõimelise elanikkonna hulk
1	5,1	21 300
2	4,4	18 500
3	1,9	210 400

Valemiga (1) arvutatud aritmeetiline keskmine tuleb $\bar{x} = (5,1 + 4,4 + 1,9)/3 = 3,8$. Kolmas piirkond on kahest esimesest tunduvalt suurem ja seetõttu iseloomustab saadud tulemus keskmist töötute protsenti vääralt. Õige on kasutada valemit (2), millega saadakse

$$\bar{x}_w = \frac{0,051 \cdot 21\ 300 + 0,044 \cdot 18\ 500 + 0,019 \cdot 210\ 400}{21\ 300 + 18\ 500 + 210\ 400} \approx 0,0236.$$

Seega annab kaalutud aritmeetiline keskmine töötute protsendiks 2,36.

Kui valimi elemendid on rühmitatud võrdse ulatusega klassidesse, siis on $x_1,\ x_2, \ldots, x_m$ intervallide keskmised ning $w_1 = n_1,\ w_2 = n_2, \ldots, w_m = n_m$ (vt punkti 1.5 näidet).

2. Harmooniline keskmine $\bar{x}_h$

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{3}$$

ja **kaalutud harmooniline keskmine** $\bar{x}_h$

$$\bar{x}_h = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}}. \tag{4}$$

Kui variandid x_i sõltuvad (kui üldse sõltuvad!) kaaludest w_i pöördvõrdeliselt, siis on õige kasutada aritmeetilist keskmist, kui aga võrdeliselt, siis harmoonilist keskmist.

Näide 3. Dispetšer arvutas liinil sõitnud 14 autobussi keskmise kiiruse (km/h): 68,2; 66,1; 61,1; 68,6; 69,4; 69,9; 65,3; 54,7; 64,2; 65,9; 64,8; 66,7; 63,7; 63,4. Kuna kiirused sõltuvad teepikkusest (mis on siin kõikidel bussidel ühesugune) võrdeliselt, siis kasutatakse keskmise kiiruse leidmiseks harmoonilist keskmist (valemit (3)): $\bar{x}_h = 14/0,2157 \approx 64,9$.

3. Geomeetriline keskmine

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}. \tag{5}$$

Geomeetriline keskmine sobib positiivsete väärtustega tunnuste jaoks (nullväärtusi ei tohi olla), kui tunnusel on üksikuid eriti suuri väärtusi. Sel juhul nihkub aritmeetiline keskmine tugevasti paremale võrreldes geomeetrilise keskmisega, mida üksikud hälbinud väärtused mõjutavad vähem.

Näide 4. Investeeringutelt saadud tulu oli esimesel aastal 12%, teisel 10% ja kolmandal 15%. Tuhandekrooonine algkapital kasvab kolme aasta jooksul
$1000(1+0,12)(1-0,10)(1+0,15)=1416,8$ kroonini. Geomeetriline keskmine

$$\overline{x}_g = \sqrt[3]{1.12 \cdot 1.10 \cdot 1,15} \approx 1.12315$$

annab keskmiseks aastaseks tuluprotsendiks 12,315, mille alusel tuhandekroonine algkapital kasvab kolme aastaga $1000(1+0,12315)^3 = 1416,8$ kroonini, mis langeb kokku eespool arvutatuga. Aritmeetiline keskmine $\overline{x} = (1,12+1.10+1,15)/3 \approx 1.12333$ annab väära tulemuse (keskmiselt 12,333% aastas, mitte aga õigeks osutuva 12,315%).

4. Ruutkeskmine $\qquad \overline{x}_r = \sqrt{\dfrac{\sum\limits_{i=1}^{n} x_i^2}{n}}. \qquad (6)$

Rakendustes leitakse mõnikord hälbed teatavast normist või keskmisest. Osa nendest hälvetest on positiivsed, osa negatiivsed. Hälvete keskmise iseloomustamiseks on otstarbekas kasutada ruutkeskmist, mis väldib hälvete vastastikust kustutamist. Ruutkeskmine võimendab tunnuse üksiku eriti suure või väikese väärtuse osatähtsust, harmooniline keskmine aga tasandab.

Mitme keskmist iseloomustava karakteristiku üheaegne kasutamine annab valimi kohta rohkem teavet, eriti siis, kui nad üksteisest oluliselt erinevad.

Asendi- ehk struktuurikeskmistest on olulisemad mood ja mediaan. Nende täpsemad definitsioonid antakse 3. peatükis. Siin piirdutakse näidete varal toodud kirjeldustega.

Esitame näites 1 antud kuupalgad x_i kasvavas järjekorras, st koostame variatsioonrea:

5100; 5600; 5700; 5900; 6000; 6100; 6500.

Variatsioonrea *keskele* jääv element on **mediaan** $M_eX = 5900$. **Mediaan** on variatsioonrea keskmine element paarituarvulise valimi korral või kahe keskmise elemendi poolsumma paarisarvulise valimi puhul.

Näide 5. Tehase nädalatoodangust võeti 50 raadiot ja uuriti neid põhjalikult enne lõppkontrolli saatmist. Tulemused on esitatud järgmises tabelis.

Defektide arv	0	1	2	3
Raadiote arv	12	15	17	6

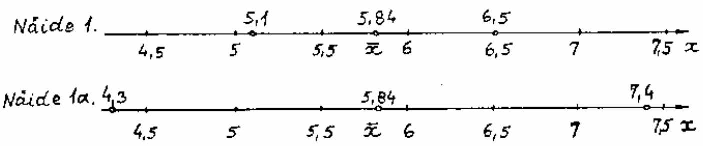
Mood on see tunnuse väärtus, millele vastab suurim sagedus. Kuna suurimale defektsete raadiote arvule 17 vastav defektide arv 2, siis on mood $M_oX = 2$.

1.4. Variatsiooninäitarvud

Näide 1a. Järgnevas tabelis on toodud teise ettevõtte seitsme juhuslikult valitud töötaja kuupalgad kroonides (võrdle näitega 1).

	1	2	3	4	5	6	7
x_i	6200	4300	5300	7400	6300	5900	5500

Näites 1a on aritmeetiline keskmine ja mediaan täpselt samasugused kui näites 1: $\overline{x} \approx 5842,9$ ja $M_eX = 5900$. Ilmselt on aga näites 1a palkade hajuvus suurem kui näites 1 .Seda asjaolu illustreerib järgnev skeem, millele on kantud palk tuhandetes kroonides:



Leiame hälbed aritmeetilise keskmise suhtes $x_i - \overline{x}$ ning liidame need kokku. Osa liidetavaist on positiivsed, osa negatiivsed ja summaks saame nulli. Nagu järgnevast näha, on tulemus kergesti põhjendatav:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} = n \cdot \overline{x} - n \cdot \overline{x} = 0.$$

Et nulliga võrduv tulemus ei iseloomusta palkade hajuvust, siis leitakse aritmeetilise keskmise suhtes võetud hälvete ruudud ja arvutatakse nende aritmeetiline keskmine

$$D^*X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x} \cdot \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}(\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2. \quad (1)$$

D^*X on **statistiline dispersioon**, mida kasutatakse **üldkogumi** hajuvuse kirjeldamiseks. Täpsem analüüs (vt punkti 5.2) näitab, et valemi (1) kasutamisel valimi kohta tehakse süstemaatiline viga ning alahinnatakse parameetri tegelikku väärtust (nagu öeldakse, hinnang on nihutatud). **Valimi** hajuvust on õige iseloomustada **dispersiooni nihutamata hinnanguga**, mis saadakse statistilise dispersiooni D^*X korrutamisel suurusega $n/(n-1)$:

$$s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2\right). \quad (2)$$

Kuna näidetes 1 ja 2 vaadeldakse valimeid, siis kasutatakse dispersiooni hindamiseks valemit (2). Näites 1 tuleb $s^2 \approx 192\ 897[kr^2]$. Näites 1a saadakse $s^2 \approx 926\ 214[kr^2]$, mis on *suurem* kui näites 1.

Valemiga (2) arvutatud dispersiooni nihutamata hinnang on ühikutes [*krooni ruudus*]. Esialgse mõõtühiku [*kroon*] juurde tagasiminekuks arvutatakse **standardhälbe** nihutamata hinnang s:

$$s = \sqrt{s^2}. \qquad (3)$$

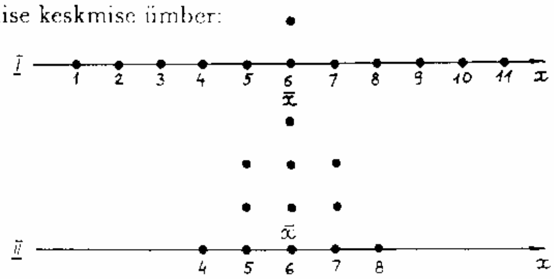
Näites 1 on $s \approx 439,2[kr]$, näites 1a on aga $s \approx 962,4[kr]$, mis on *suurem* ning kajastab seega suuremat palkade hajuvust teises ettevõttes.

Näide 6. Võrdleme järgnevas tabelis toodud kahe valimi elemente.

I valim	1	2	3	4	5	6	6	7	8	9	10	11
II valim	4	5	5	5	6	6	6	6	7	7	7	8

Võrdsed on nii valimite mahud $n_I = n_{II} = 12$ kui ka aritmeetilised keskmised $\bar{x}_I = \bar{x}_{II} = 6$, mediaanid $M_eX_I = M_eX_{II} = 6$ ja moodid $M_oX_I = M_oX_{II} = 6$. Oluliselt erinevad on aga hajuvust iseloomustavad dispersiooni nihutamata hinnangud $s_I^2 = 10$ ja $s_{II}^2 = 1,27$ ning standardhälbe nihutamata hinnangud $s_I = 3,16$ ja $s_{II} = 1,13$.

Kujutades nende valimite elementide paiknemist graafiliselt arvteljel, on selgelt näha I valimi elementide suurem hajuvus ning II valimi elementide suhteliselt palju tihedam koondumine aritmeetilise keskmise ümber:



Kahe või enama jaotuse hajuvuse võrdlemiseks kasutatakse **variatsioonikordajat**

$$V = \frac{s}{\bar{x}} \cdot 100\%, \qquad (3)$$

kus $\bar{x}$ on aritmeetiline keskmine.

Standardhälve iseloomustab hajuvuse absoluutsuurust ega arvesta valimi elementide väärtusi. Variatsioonikordaja on relatiivne mõõt ning võimaldab ülevaatlikult võrrelda ka neid jaotusi, mille elementide väärtused erinevad oluliselt.

Näide 7. Järgnevas tabelis on esitatud kahes hotellis nädala jooksul peatunud ärimeeste arv.

Hotell A	987	988	990	967	973	945	993	965	944	930	948	958
Hotell B	124	138	110	147	198	136	140	162	187	119	155	149

Hotell A: $\bar{x}_A = 965,67$, $s_A = 21,08$, $V_A = 2,18\%$.
Hotell B: $\bar{x}_B = 147,08$, $s_B = 25,98$, $V_B = 17,7\%$.

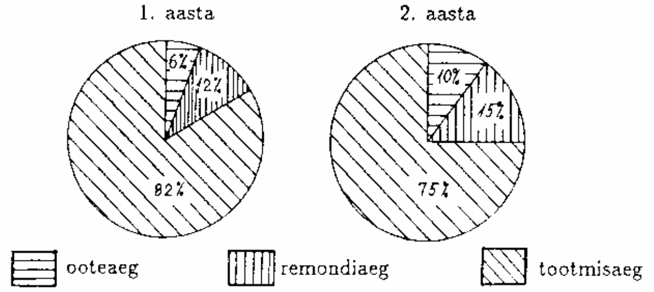
Standardhälbed $s_A = 21,08$ ja $s_B = 25,98$ on mõlemal hotellil lähedased. Variatsioonikordajad $V_A = 2,18\%$ ja $V_B = 17,7\%$ aga erinevad oluliselt. See ütleb, et hotelli A külastajate arv varieerub suhteliselt palju vähem.

1.5. Statistilise materjali graafiline esitamine

Näide 8. Järgnevas tabelis on toodud tehaseseadmete koormamisajad kahe aasta kohta. Esitada tulemused sektordiagrammidena.

	1. aasta		2. aasta	
	päevad	%	päevad	%
Koormamisaeg	270	100	320	100
sealhulgas: ooteaeg	16	6	32	10
remondiaeg	32	12	48	15
tootmisaeg	222	82	240	75

Sektordiagrammi joonestamisel vastab ringi täispindalale 100% vaatlustest (antud ülesandes 100% seadmete koormamisajast). Üksikute sektorite pindalad kajastavad nähtuse tunnuste protsentides väljendatud sagedusi $p_i^* \cdot 100\% = (n_i/n) \cdot 100\%$ (selles ülesandes ooteaja, remondiaja ning tootmisaja protsendid).



Näide 9. 50 juhuslikult valitud autoomaniku keskmine kuuläbisõit oli viimase aasta jooksul järgmine:

821	1157	1341	1090	1218	913	1023	1106	1183	859
591	1293	514	802	1076	1134	831	1177	947	415
342	927	728	554	1451	916	1384	462	618	1415
709	853	1052	798	994	1083	639	733	1131	766
1215	570	942	683	1008	976	1036	886	986	847

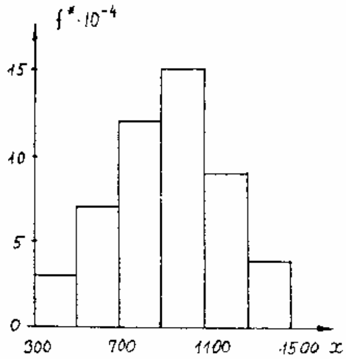
Eeltoodud korrastamata statistilise rea elemendid moodustavad valimi mahuga $n = 50$. Koostame variatsioonrea, rühmitades valimi elemendid klassidesse. Vastavalt punktis 1.2 toodud soovitustele valime 6 võrdse ulatusega $\Delta x = 200$ klassi. Esimese intervalli algus ja viimase intervalli lõpp ei pea kokku langema vastavalt tunnuse vähima väärtusega $x_{min} = 342$ ja tunnuse suurima väärtusega $x_{max} = 1451$. Võtame esimese intervalli alguseks 300. Koondame lähteandmed ja arvutustulemused järgnevasse tabelisse.

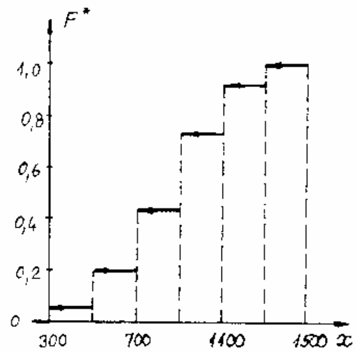
Intervalli nr	Intervall	n_i	p_i^*	f_i^*	F^*
1	[300;500[	3	0,06	0,0003	0,06
2	[500;700[	7	0,14	0,0007	0,2
3	[700;900[	12	0,24	0,0012	0,44
4	[900;1100[	15	0,3	0,0015	0,74
5	[1100;1300]	9	0,18	0,0009	0,92
6	[1300;1500]	4	0,08	0,0004	1

n_i on i-ndasse intervalli jäävate vaatluste arv, kusjuures $\sum n_i = 50$. $p_i^* = n_i/n$ on sagedus, kusjuures $\sum p_i^* = 1$. Sageduse suhteline tihedus on $f_i^* = p_i^*/\Delta x_i$. Tabeli viimases veerus on kumulatiivne sagedus F^* (empiiriline jaotusfunktsioon), mille saamiseks liidetakse eelnevas reas olevale F^* väärtusele juurde käesolevas reas olev p_i^* väärtus. Viimases reas olev F^* tuleb alati võrdne ühega.

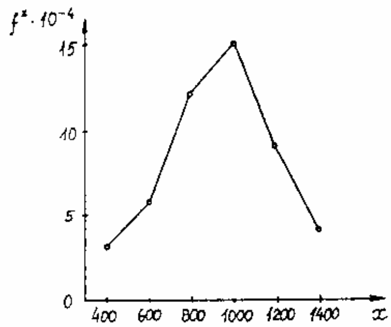
Valimi elementide jaotuse graafiliseks iseloomustamiseks kasutatakse histogrammi. Sageduste **histogrammiks** nimetatakse astmelist kujundit, mis koosneb ristkülikutest, mille alused on võrdsed klassi ulatusega Δx_i, kõrgused aga on võrdsed sageduse suhtelise tihedusega f_i^*. Sellisel viisil defineeritud histogrammi pindala on alati võrdne **ühega**. Viimasel asjaolul on oma sisuline tähendus, mida selgitatakse hiljem. Histogrammis olevate üksikute ristkülikute pindalad on võrdelised vastavasse intervalli sattunud vaatluste arvuga n_i. Seetõttu piirdutakse andmete jaotusest esialgse *lihtsustatud* ülevaate saamiseks sageli sellega, et histogrammis olevate ristkülikute kõrgused võrdsustatakse sellesse intervalli jäävate vaatluste arvuga n_i.

Käesolevale näitele vastav histogramm on toodud järgneva joonise vasakpoolsel graafikul. Selle joonise parempoolsel graafikul on kumulatiivne sagedus (ehk empiiriline jaotusfunktsioon) F^*, mis näitab (juhul $F^* \cdot 100\%$), mitu protsenti vaatluspunktidest on iga klassi ülemisest piirist väiksema väärtusega. Näiteks antud ülesandes on alla 1100 kilomeetri läbi sõitnud 74% autoomanikest (neljanda intervalli lõpus kohal 1100 on $F^* = 0,74$).





Peale histogrammi võib valimi elementide jaotust graafiliselt iseloomustada ka **sageduspolügooniga**. Järgneval joonisel on vertikaalteljele märgitud sageduse suhteline tihedus f_i^* ning horisontaalteljele intervallide keskmised. Intervallide keskmiste kohal on graafikule kantud igale intervallile vastavad f_i^* väärtused. Neid punkte sirglõikudega ühendades saadaksegi sageduspolügoon.



Kui valimi elemendid on rühmitatud võrdse ulatusega klassidesse, siis lähtutakse aritmeetilise keskmise arvutamisel intervallide keskmistest (vt punkti 1.3 valemit (2) ning sama punkti näite 2 järel toodud selgitust):

$$\overline{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = \frac{1}{50}(3 \cdot 400 + 7 \cdot 600 + 12 \cdot 800 + 15 \cdot 1000 + 9 \cdot 1200 + 4 \cdot 1400) = 928.$$

Dispersiooni nihutamata hinnangu s^2 leidmiseks võetakse punkti 1.4 valemis (2) oleva summa

$$\sum_{i=1}^{50} x_i^2 = 3 \cdot 400^2 + 7 \cdot 600^2 + 12 \cdot 800^2 + 15 \cdot 1000^2 + 9 \cdot 1200^2 + 4 \cdot 1400^2$$

arvutamisel samuti aluseks intervallide keskmised.

Standardhälbe hinnanguks tuleb $s = \sqrt{s^2} \approx 264,22$.

Kui lähtuda statistilises reas toodud individuaalvaatlustest, siis saadakse $\overline{x} = 923,9$ ja $s \approx 265,32$. Erinevus rühmitatud andmete alusel saadud tulemustest on küllalt väike.

1.6. Indeksid

Indeksiks nimetatakse statistikas suhtarvu, millega iseloomustatakse mingi nähtuse (näiteks omahinna, toodangu mahu, tööviljakuse) muutumist ajas.

Baasindeks ehk alusindeks I_{bi} arvutatakse, kui vaadeldaval perioodil olemasoleva tunnuse väärtuse p_i ja mingil baasiks valitud ajaperioodil omandatud tunnuse väärtuse p_b suhe:

$$I_{bi} = \frac{p_i}{p_b}. \qquad (1)$$

Sageli võetakse baasperioodiks rea esimene liige ($i = 0$), millele vastab tunnuse väärtus p_0.

Näide 11. Tabelis 1 on toodud ühepereelamute keskmised hinnad (dollarites) aastail 1992–1998 ($i = 0, 1, \ldots, 6$).

Tabel 1

Aasta i	1992	1993	1994	1995	1996	1997	1998
Hind p_i	110 700	116 200	122 300	133 400	132 000	142 900	156 100

Valime baasperioodiks 1992. aasta. Baasindeksi saamiseks jagatakse iga aasta hind baasperioodil kehtinud hinnaga. 1992. aasta hinnaindeks tuleb seega võrdne ühega (protsentides väljendatuna 100%). Tulemused on tabelis 2.

Tabel 2

Aasta i	1992	1993	1994	1995	1996	1997	1998
Hinnaindeks $I_{0,i}$	1	1,050	1,105	1,205	1,192	1,291	1,410

Ahelindeks leitakse kahe järjestikuse perioodi tunnuse väärtuste suhtena. Arvutamisel alustatakse väärtusest 1 ja edasi arvutatakse vaadeldava perioodi ja eelmise perioodi tunnuse väärtuste suhe

$$I_i = \frac{p_i}{p_{i-1}}. \tag{2}$$

Näite 11 andmetel saadavad tulemused

1993: 116 200/110 700=1,050

1994: 122 300/116 200=1,052 jne

on koondatud tabelisse 3.

Tabel 3

Aasta i	1992	1993	1994	1995	1996	1997	1998
Hinna ahelindeks I_i	1	1,050	1,052	1,091	0,990	1,083	1,092

Näide 12. Kauba hind tõusis 5% ja seejärel langes 5%. Kuidas muutus hind kokkuvõttes?

Antud on ahelindeksi väärtused (vt tabelit 4).

Tabel 4

	Alghind	Vahepealne hind	Lõpphind
Hinna ahelindeks I_i	1	1,05	0,95

Arvutame baasindeksi alghinna suhtes (vt tabelit 5).

Tabel 5

	Alghind	Vahepealne hind	Lõpphind
Hinna baasindeks $I_{0,i}$	1	$1 \cdot 1,05 = 1,05$	$1 \cdot 1,05 \cdot 0,95 = 0,9975$

Kuna lõpphinna ja alghinna baasindeksite vahe 0,9975-1=-0,0025 on negatiivne, siis kokkuvõttes hind langes $0,0025 \cdot 100\% = 0,25\%$.

Näide 13. Investeeringutelt saadud tulu oli esimesel aastal 12%, teisel 10% ja kolmandal 15% (vt punkti 1.3 näidet 4). Leida keskmine aastane tuluprotsent.

Antud on ahelindeksi väärtused (vt tabelit 6).

Tabel 6

	Alg-kapital	Kapital 1. aasta lõpul	Kapital 2. aasta lõpul	Kapital 3. aasta lõpul
Tulu ahelindeks I_i	1	1,12	1,10	1,15

Keskmine aastane tulu on ahelindeksi niisugune väärtus, mis annab kolmanda aasta lõpuks sellesama tulu, mis tabelis 6 toodud ahelindeksid 1,12; 1,10 ja 1,15. Otsitav ahelindeksi väärtus on arvude 1,12; 1,10 ja 1,15 geomeetriline keskmine, nagu seda on selgitatud varem lahendatud näites 4 (vt punkti 1.3):

$$\bar{x}_g = \sqrt[3]{1,12 \cdot 1,10 \cdot 1,15} \approx 1,12315.$$

Seega keskmine aastane tulu on 12,315%.

Eespool käsitletud baas- ning ahelindeksid on **lihtindeksid**, mis iseloomustavad *ühe* tunnuse suhtelist muutumist. Lihtindeks avadub kahe näitarvu suhtena.

Liitindeks mõõdab mitme omavahel seotud tunnuse suhtelist muutumist. Kaalutud liitindeksis sisaldub *erinevate kaaludega* tunnuste kombinatsioon. Kui hindade kaaluks on toodangu mahud, siis räägitakse hinnaindeksist. Kui toodangu mahtude kaaluks on hinnad, siis on tegemist toodangu mahu indeksiga.

Näide 14. Väikeettevõttes valmistatakse kolme toodet. Tabelis 7 on nende toodete hinnad p_i (kroonides), kogused q_i (tükkides) ja maksumused $p_i \cdot q_i$ (kroonides) kahel järjestikusel aastal.

Tabel 7

	Esimene aasta (baasperiood)			Teine aasta (aruandeperiood)		
Toode	hind p_0	kogus q_0	maksumus $p_0 \cdot q_0$	hind p_1	kogus q_1	maksumus $p_1 \cdot q_1$
A	120	560	67 200	140	580	81 200
B	80	240	19 200	88	250	22 000
C	30	1380	41 400	42	1520	63 840

Koondindeks, mis iseloomustab maksumuse muutumist nii hindade kui koguste järgi, on

$$I_Q = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0} \cdot 100\%, \tag{3}$$

kus $\sum p_0 \cdot q_0$ on toodangu maksumus baasperioodil ja $\sum p_1 \cdot q_1$ on toodangu maksumus aruandeperioodil. Näites 14 tuleb

$$I_Q = \frac{81\ 200 + 22\ 000 + 63\ 840}{67\ 200 + 19\ 200 + 41\ 400} \cdot 100\% = 130,70\%,$$

mis ütleb, et maksumus on kasvanud 30,7%. Selle indeksi dünaamikas pole selge hindade ja koguste panus eraldi võetuna.

Laspeyrese hinnaindeksi arvutamisel võetakse kaaludeks baasperioodi toodangu mahud q_0:

$$I_{L_h} = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} \cdot 100\%, \tag{4}$$

kus $\sum p_1 \cdot q_0$ on toodangu maksumus aruandeperioodil baasperioodi hindades. Näites 14 tuleb

$$I_{L_h} = \frac{140 \cdot 560 + 88 \cdot 240 + 42 \cdot 1380}{120 \cdot 560 + 80 \cdot 240 + 30 \cdot 1380} \cdot 100\% = 123,22\%.$$

Tingimusel, et toodangu kogused aruandeperioodil oleksid samad, mis baasperioodil, tõusid hinnad 23,22%.

Tarbijahinnaindeks on Laspeyrese tüüpi hinnaindeks, mis kajastab tarbekaupade ja teenuste hinna muutumist mingis ajavahemikus. Tarbijahinnaindeksit arvutatakse iga kaubarühma (toiduained, rõivad ja jalanõud, alkohol, tubakas jm) ning teenuse (eluaseme-, majapidamis-, tervishoiu-, transpordi- ja side- ning vaba aja veetmise kulutused) osatähtsuse järgi tarbimises.

Eestis kasutatava tarbijahinnaindeksi arvutamisel võetakse alates 1994. aastast kaubarühmi ja teenuseid arvesse järgmiste kaaludega:

	%
toiduained	37,85
alkohoolsed joogid ja tubakas	4,33
rõivad ja jalanõud	7,61
eluase	19,73
majapidamine	3,20
tervishoid	2,01
transport ja side	13,04
vaba aeg	6,56
muud kaubad ja teenused	5,67

Kulutused eluasemele sisaldavad lisaks üürile, kommunaalteenustele ja küttele ka eluaseme remondikulusid.

Majapidamiskulutuste alla kuuluvad näiteks mööbli ja kodumasinate ost ning hooldus, puhastusvahendite, nõude ost jms.

Vabaajakulutuste hulka kuuluvad väljaminekud trükistele, raadiokaupadele, spordi- ja fototarvetele, lilledele, kultuuri- ja spordiüritustele, eneseharimisele ja lasteaiale.

Muude kaupade ja teenuste hulgas on kulutused elutarbelistele teenustele, mitmesugustele tarbeesemetele, isikliku hügieeni vahenditele, söömisele väljaspool kodu ja muudes rühmades arvestamata teenustele.

Alates 1994. aastast on tarbijahinnaindeksi arvutamiseks kasutatavaid esinduskaupu kokku 410.

Tööstustoodangu tootjahinnaindeks iseloomustab Eestis valmistatud tööstustoodete hinnamuutusi. Indeks sisaldab nii sise- kui välisturule valmistatud tööstustoodete hinna ilma käibe- ja aktsiisimaksuta. Lisaks kogu tööstust iseloomustavale tootjahinnaindeksile tuuakse eraldi välja indeksid energeetika, mäetööstuse ja töötleva tööstuse kohta.

Näide 15. Tabelis 8 on antud USA ühe tarbijahinnaindeksi CPI-W (*consumer prise index for wage eaners and clerical workers*) väärtused protsentides aastal 1988 baasperioodiks võetavate aastate 1982–1984 suhtes.

Tabel 8

Kuu (1988)	CPI-W	100/CPI-W
Jaanuar	114,5	0,873
Veebruar	114,7	0,872
Märts	115,1	0,869
Aprill	115,7	0,864
Mai	116,2	0,861
Juuni	116,7	0,857

Nagu tabelist 8 näha, oli CPI-W 1988. aasta juunis 116,7%. Seega oli baasperioodi ja selle kuu vaheline inflatsioon 16,7%. Tabeli 8 viimases veerus on arvutatud suhe 100/CPI-W, mis näitab dollari ostujõu muutust baasperioodiga võrreldes. Kui juunis 1988 oli 100/CPI-W=0,857, siis see tähendab, et võrreldes aastatega 1982–1984 oli üks dollar väärt ainult 85,7 senti.

Paasche hinnaindeksi leidmisel on kaaludeks aruandeperioodi toodangu mahud q_1:

$$I_{P_h} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot 100\%. \tag{5}$$

Näite 14 andmetel saadakse

$$I_{P_h} = \frac{140 \cdot 580 + 88 \cdot 250 + 42 \cdot 1520}{120 \cdot 580 + 80 \cdot 250 + 30 \cdot 1520} \cdot 100\% = 123,55\%.$$

Laspeyrese toodangu mahu indeksi arvutamisel võetakse kaaludeks baasperioodi hinnad p_0:

$$I_{L_m} = \frac{\sum p_0 \cdot q_1}{\sum p_0 \cdot q_0} \cdot 100\%, \tag{6}$$

Tabelis 7 toodud andmetel arvutatakse

$$I_{L_m} = \frac{120 \cdot 580 + 80 \cdot 250 + 30 \cdot 1520}{120 \cdot 560 + 80 \cdot 240 + 30 \cdot 1380} \cdot 100\% = 105,79\%.$$

Tingimusel, et hinnad oleksid aruandeperioodil samad kui baasperioodil, suurenesid toodangu mahud 5,79%.

Paasche toodangu mahu indeksi leidmisel on kaaludeks aruandeperioodi hinnad p_1:

$$I_{P_m} = \frac{\sum p_1 \cdot q_1}{\sum p_1 \cdot q_0} \cdot 100\%. \tag{7}$$

Näite 14 andmete kohaselt tuleb

$$I_{P_m} = \frac{140 \cdot 580 + 88 \cdot 250 + 42 \cdot 1520}{140 \cdot 560 + 88 \cdot 240 + 42 \cdot 1380} \cdot 100\% = 106,07\%.$$

Laspeyrese indeksit sobib rohkem kasutada siis, kui aruandeperioode soovitakse võrrelda ühe ja sama baasperioodiga. Pealegi on aruandeperioodi toodangu mahtude kohta sageli andmeid raskem saada kui baasperioodi kohta. Kui aruadeperiood on baasperioodist ajaliselt kaugel, siis peegeldab Paasche indeks tegelikkust adekvaatsemalt. Vaadeldavas näites 14 erinevad valemitega (4) ja (5) ning samuti valemitega (6) ja (7) arvutatud Laspeyrese indeks ja Paasche indeks teineteisest vähe. See on tingitud asjaolust, et tabelis 7 olevatest algandmetest saadavad kaalud on eespool nimetatud valemites nii baas- kui ka aruandeperioodil lähedased. Sellisel juhul eelistatakse Laspeyrese indeksit.

Börsiindeksid jagunevad oma arvutuseeskirjalt kas aritmeetilisele keskmisele, geomeetrilisele keskmisele, Laspeyrese indeksile või Paasche indeksile baseeruvateks.

Aritmeetilise keskmisena arvutatakse Dow Jones Industrial Average (tugineb 30 New-Yorgi börsil kaubeldaval suuraktsial) ja Nikkei Sock Average (põhineb 225 Tokio börsil noteeritud aktsial).

Äripäeva indeks arvutatakse kui kaalutud aritmeetiline keskmine.

Geomeetrilisel keskmisel baseeruvad *Financial Times* Ordinary Index (Suurbritannia) ja Value Line Composite Average (USA).

Laspeyrese indeksina leitakse FT-SE 100 Index (Suurbritannia) ja Standard and Poors 500 Index (USA).

Paasche hinnaindeksina arvutatakse Helsingi börsiindeks HEX ning sellega lähedasel viisil ka börsiindeks TALSE.

1.7. Ülesanded

1.1. Suure kaubamaja ühe osakonna 20 müüja poolt sooritatud müügitehingute arv päevas, koondatuna variatsioonritta, oli järgmine: 6, 9, 10, 12, 13, 14, 14, 15, 16, 16, 16, 17, 17, 18, 18, 19, 20, 21, 22, 24. Leida keskmine müügitehingute arv päevas, mood, mediaan ja standardhälbe nihutamata hinnang. Kuidas muutuvad need arvkarakteristikud, kui vaatlustulemus 24 asendada suurusega 100?

1.2. Statistikas toimub 3 kontrolltööd, millest igaüks annab 15% lõpphindest. Kodutöö ja eksamitöö annavad lõpphindest vastavalt 25% ja 30%. Üliõpilane sai maksimaalselt võimalikust sajast punktist kontrolltööde eest 75, 82 ja 84 punkti, kodutöö eest 79 punkti ning eksamitöö eest 91 punkti. Missugune on üliõpilase lõpphinne?

1.3. Esimene auto läbis 100 *km* kiirusega 60 km/h, teine auto läbis sama vahemaa kiirusega 40 km/h. Leida nende kahe auto keskmine kiirus.

1.4. Investeeringutelt saadav tulu oli viiel järjestikusel aastal 10%, 8%, 9%, 12% ja 7%. Leida keskmine aastane tuluprotsent.

1.5. Panga peakontoris on jooksevkonto keskmine suurus 2497,00 krooni standardhälbega 1075,86 krooni, harukontoris aga 1487,68 krooni standardhälbega 650,20 krooni. Leida variatsioonikordajad ning võrrelda neid.

1.6. Statistika esimesel kontrolltööl said 55 üliõpilast järgmised tulemused (12-st maksimumpunktist):

10,8	4,2	8,2	8,4	10,6	11,1	9,6	7,1
1,3	11,3	6	10,8	10,6	2,4	8,2	8,2
8,6	7,6	6,2	9,2	4,8	6,4	5	8,3
5,6	8,8	5,8	9,7	9	12	9,4	6,8
8,7	2,6	4,8	7	12	3,2	3,7	8,1
6,6	7	7,7	9,6	4,2	7	4,8	11,4
6,1	9,7	12	5,1	10	6,3	7,5	

Koostada variatsioonrida, rühmitades andmed kuude klassi ulatusega $\Delta x = 2$. Põhjendada klasside arvu ning esimese intervalli alguse ja viimase intervalli lõpu valikut. Joonestada histogramm ning kumulatiivse sageduse (empiirilise jaotusfunktsiooni) graafik. Joonestada sageduspolügoon. Arvutada keskmine hinne ning standardhälve, lähtudes nii individuaalvaatlustest kui ka rühmitatud andmetest.

Kahel eelneval aastal osales samalaadses kontrolltöös vastavalt 52 ja 48 üliõpilast, kusjuures keskmine hinne oli 7,28 standardhälbega 2,93 ja 7,52 standardhälbega 2,69. Hinnata tulemuste dünaamikat kolme aasta jooksul.

1.7. Aktsiakapitali aastane tootlus protsentides 30-aastasel perioodil, millest on inflatsiooni mõju kõrvaldatud, on järgmine:

-3,2	17,4	-13,4	-9,9	20,4	15,1
2,7	-1,6	41,0	20,8	6,1	-21,8
20,9	53,4	10,3	15,1	-13,8	-34,8
24,6	31,1	-1,0	10,3	-1,5	28,3
17,2	3,6	26,0	-13,0	10,6	18,2

Koostada variatsioonrida, rühmitades andmed viide klassi ulatusega $\Delta x = 20$. Põhjendada klasside arvu ning esimese intervalli alguse ja viimase intervalli lõpu valikut. Joonestada histogramm ning kumulatiivse sageduse (empiirilise jaotusfunktsiooni) graafik. Joonestada sageduspolügoon. Arvutada aktsiakapitali keskmine tootlus ning standardhälve, lähtudes nii individuaalvaatlustest kui ka rühmitatud andmetest.

1.8. Tabelis 9 on toodud ettevõttes 6 aasta jooksul valmistatud kaubakogus.

Tabel 9

Aasta i	1	2	3	4	5	6
Toodang p_i tonnides	2400	2800	3000	3600	4200	4400

Leida baasindeksid esimese ja kuuenda aasta suhtes. Leida ahelindeksid. Kuidas iseloomustab ahelindeks toodangu mahu muutusi?

1.9. Teatud hooajalise kauba hind oli kuude lõikes järgmine.

Tabel 10

Kuu i	Märts	Aprill	Mai	Juuni	Juuli	August	September
Hind p_i	5,50	5,45	5,40	5,30	5,10	5,10	5,20

Leida baasindeksid märtsi ja septembri suhtes. Leida ahelindeksid. Kuidas iseloomustab ahelindeks hinnamuutusi?

1.10. Ettevõttes tõsteti esimesel aastal palka 10%, teisel 15% ja kolmandal 5%. Leida keskmine aastane palgatõus.

1.11. Eettevõttes valmistatakse nelja toodet. Tabelis 11 on nende toodete hinnad p_i (kroonides), kogused q_i (tootel A kilogrammides, tootel B liitrites, tootel C meetrites ja tootel D tükkides) ning maksumused $p_i \cdot q_i$ (ühiku kohta kroonides) kahel järjestikusel aastal.

Tabel 11

Toode	Esimene aasta (baasperiood) hind p_0	kogus q_0	maksumus $p_0 \cdot q_0$	Teine aasta (aruandeperiood) hind p_1	kogus q_1	maksumus $p_1 \cdot q_1$
A	2,20	560	1232	2,80	600	1680
B	14	112	1568	15	150	2250
C	40	18	720	50	22	1100
D	9,60	680	6528	9,80	730	7154

Leida koondindeks, mis iseloomustab maksumuse muutumist nii hindade kui ka koguste järgi. Arvutada Laspeyresi ja Paasche hinnaindeksid ning Laspeyresi ja Paasche toodangu mahu indeksid.

1.12. Eettevõte tellis kahel järjestikusel aastal kolme toodet, mille hinnad p_i (kroonides) ja kogused q_i (tükkides) on toodud tabelis 12.

Tabel 12

Toode	Esimene aasta (baasperiood) hind p_0	kogus q_0	maksumus $p_0 \cdot q_0$	Teine aasta (aruandeperiood) hind p_1	kogus q_1	maksumus $p_1 \cdot q_1$
A	21	65	1365	43	46	1978
B	63	124	7812	108	105	11340
C	56	62	3472	57	94	5358

Leida koondindeks, mis iseloomustab maksumuse muutumist nii hindade kui ka koguste järgi. Arvutada Laspeyresi ja Paasche hinna- ning toodangu mahu indeksid.

2. Tõenäosuse mõiste, omadusi ja arvutamise põhivõtteid

2.1. Juhuslikud sündmused

Mingi katse või vaatluse tulemusena võib toimuda teatav **sündmus**. Sündmuseks on näiteks kontrollitud toote praagiks osutumine. Toote kontrollimine on katse, praagiks osutumine aga sündmus. Sündmus on näiteks märklaua tabamine püssilaskmisel, kusjuures laskmine on katse, tabamine aga sündmus. Sündmusi märgitakse tähtedega $A, B, \ldots$.

Sündmused võib jaotada kolme liiki:

1. Kindel sündmus Ω, mis alati toimub. Näiteks vesi külmub normaaltingimustel $0\,^{\circ}$C juures.

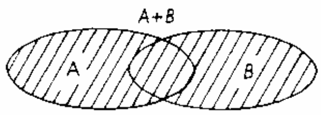
2. Võimatu sündmus $\emptyset$, mis kunagi ei toimu. Näiteks mündi viskamisel on võimatu saada üheaegselt vappi ja kirja.

3. Juhuslik sündmus, mis võib toimuda või mitte toimuda. Näiteks täringu viskamisel võib saada 2, aga võib ka mitte saada.

Tõenäosusteooria uurib massiliselt toimuvates juhuslikes sündmustes esinevaid seaduspärasusi. Näiteks täringu paljukordsel viskamisel tuleb 2 (ligikaudu) 1/6 juhtudest, mis pole enam juhuslik, vaid seaduspärane.

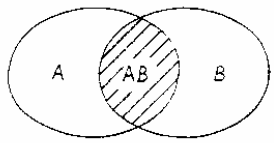
2.2. Tehted sündmustega

Sündmuste A ja B **summaks** $A + B$ (ehk ühendiks $A \cup B$) nimetatakse sündmust, mille toimumine seisneb neist vähemalt ühe, kas A või B toimumises ("A või B või mõlemad"). Kahe sündmuse summale vastav piirkond on järgneval (hulgateooriast tuntud) Venni diagrammil viirutatud:



Näiteks kui sündmus A on märkilaskmisel piirkonna A tabamine ja sündmus B on piirkonna B tabamine, siis sündmus $A+B$ on joonisel oleva viirutatud piirkonna tabamine.

Sündmuste A ja B **korrutiseks** AB (ehk ühisosaks $A \cap B$) nimetatakse sündmust, mille toimumine seisneb nii A kui B toimumises ("A ja B"). Kahe sündmuse korrutisele vastav piirkond on järgneval Venni diagrammil viirutatud:



Kahte sündmust nimetatakse **teineteist välistavaks**, kui nende korrutis (ühisosa) on võimatu sündmus: $AB = \emptyset$ (ehk $A \cap B = \emptyset$).

Kui sündmusi on üle kahe, jäävad nende summa ja korrutise definitsioonid analoogseteks eeltoodutega.

Sündmuse A **vastandsündmuseks** $\bar{A}$ nimetatakse sündmust, mille toimumine seisneb sündmuse A mittetoimumises.

2.3. Sündmuse sagedus

Olgu sooritatud n katset ja mingi sündmus A toimugu selles katseseerias m korda. **Sündmuse** A **sageduseks** selles katseseerias nimetatakse arvu

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

Näide. 5-lasulises seerias tabati märki 2 korda, 50-lasulises 27 korda ja 100-lasulises 51 korda. Tabamise sagedused on

$$P^*(A_5) = \frac{2}{5} = 0,4; \quad P^*(A_{50}) = \frac{27}{50} = 0,54; \quad P^*(A_{100}) = \frac{51}{100} = 0,51.$$

Sageduse omadusi

1. $0 \leq P^*(A) \leq 1$ (sündmus A võib n katse puhul toimuda 0 kuni n korda, st $0 \leq m \leq n$).
2. $P^*(\Omega) = 1$ (kindla sündmuse sagedus on 1).
3. $P^*(\emptyset)=0$ (võimatu sündmuse sagedus on 0).
4. Kui sündmused A ja B on teineteist *välistavad*, siis on võimalik näidata, et

$$P^*(A + B) = P^*(A) + P^*(B).$$

5. Sündmuse A sagedust tingimusel, et sündmus B toimus, nimetatakse sündmuse A **tinglikuks sageduseks** $P^*(A|B)$. Saab tõestada, et sündmuse A tinglik sagedus avaldub kujul

$$P^*(A|B) = \frac{P^*(AB)}{P^*(B)}$$ ja korrutise sageduse võib esitada kahel viisil

$$P^*(AB) = P^*(B) \cdot P^*(A|B) = P^*(A) \cdot P^*(B|A). \tag{1}$$

2.4. Sündmuse tõenäosus

Sündmuse sagedus on määratav vaid pärast katseid. Kuna eri katseseeriate puhul saadakse erinevaid tulemusi (vt eelmise punkti näidet), siis ei ole sagedus sündmuse kõige paremaks iseloomustajaks. Kuid võib tähele panna, et pikas katseseerias sagedus stabiliseerub, kõikudes mingi konstantse väärtuse p ümber. Nii jõuame **tõenäosuse statistilise definitsioonini**: *juhusliku sündmuse tõenäosuseks nimetatakse konstanti, mille ümber grupeeruvad selle sündmuse sagedused katsete arvu suurenedes.*

Eelnev definitsioon iseloomustab tõenäosuse mõistet küllalt sisukalt, kuid ei anna võtteid tõenäosuse arvutamiseks. Seda võimaldab teha **tõenäosuse klassikaline definitsioon** (e Laplace'i mudel): *sündmuse A tõenäosuseks nimetatakse suurust*

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

kus n on kõikide võrdvõimalike elementaarsündmuste arv ja m on sündmuses A sisalduvate elementaarsündmuste arv (m on n-ö soodsate elementaarsündmuste arv).

Elementaarsündmus on niisugune sündmus, mis osasündmusteks enam ei lahutu. Elementaarsündmused on võrdvõimalikud, kui sisulistel kaalutlustel ei saa lugeda ühe elementaarsündmuse toimumise võimalikkust erinevaks teise elementaarsündmuse toimumise võimalikkusest. Klassikaline definitsioon võimaldab tõenäosust arvutada ilma katseid tegemata. Selle definitsiooni puuduseks on elementaarsündmuste võrdvõimalikkuse nõue.

Näide 1. 200-detailise partii hulgast leidis tehnilise kontrolli osakond 8 praakdetaili. Leida praakdetailide esinemise sagedus.

Katsete arv $n=200$.
Sündmus A (praakdetaili esinemine) toimus $m=8$ korda.
Sündmuse A sagedus selles katseseerias $P^*(A) = m/n = 8/200 = 0,04$.

Näide 2. Tehasesse toodi 1000 laagrit. Nende hulka oli sattunud 30 praaklaagrit. Leida tõenäosus, et juhuslikult võetud laager osutub standardseks.

Kõikide võrdvõimalike elementaarsündmuste arv $n=1000$ (s.o võimaluste arv võtta ühte laagrit). Standardse laagri saamise seisukohalt soodsate elementaarsündmuste arv $m = 1000 - 30 = 970$. Sündmuse A, s.o standardse laagri saamise tõenäosus on $P(A) =$ $= m/n = 970/1000 = 0,97$.

2.5. Ühendid

Ühenditeks nimetatakse lõpliku hulga $A_n = \{a_1, a_2, \dots, a_n\}$ (nn põhihulga) elementidest $a_1, a_2, \dots, a_n$ moodustatud alamhulki, mis erinevad üksteisest kas elementide endi, nende järjestuse, arvu või kordsuse poolest.

Järgnevas käsitletakse ühenditest korduvaid elemente mitteomavaid variatsioone, permutatsioone ja kombinatsioone, mille abil lahenduvad kõik käesolevas õppevahendis olevad ülesanded.

Variatsioonideks *nimetatakse põhihulga m-elemendilisi järjestatud alamhulki.* Ehk teisiti: variatsioonideks n erinevast elemendist m elemendi kaupa nimetatakse ühendeid, mis *erinevad üksteisest kas elementide endi või nende järjestuse poolest.*

Leiame variatsioone endid moodustamata kõigi variatsioonide arvu, mis on võimalik moodustada n elemendist m-kaupa. Tähistame otsitava suuruse V_n^m. Ilmselt $V_n^1 = n$. Edasi leiame V_n^2 järgmise mõttekäigu alusel.

Esimese elemendi valikuks leidub n võimalust.

Teise elemendi valikuks leidub $n-1$ võimalust.

Mõlema, nii esimese kui teise elemendi valikuks leidub $n(n-1)$ erinevat võimalust, st $V_n^2 = n(n-1)$.

Sama mõttekäiku jätkates saame $V_n^3 = n(n-1)(n-2)$ ning üldiselt

$$V_n^m = n(n-1)(n-2)\ldots[n-(m-1)]. \tag{1}$$

Näide 1. $V_7^3 = 7\cdot 6\cdot 5 = 210$.

Näide 2. Üliõpilased õpivad üheksat erinevat ainet. Teisel septembril on ette nähtud 4 loengut. Mitu erinevat võimalust on tunniplaani koostamiseks, kui sel päeval peab olema 4 erinevat õppeainet?

$$V_9^4 = 9\cdot 8\cdot 7\cdot 6 = 3024.$$

Näide 3. Mitmel viisil on võimalik valmistada kolmevärvilist kolmest horisontaalvöödist koosnevat lippu, kui üks vööt peab olema sinine ja kasutada on 5 erinevat värvi kangast?

$$V_5^3 - V_4^3 = 36 \text{ ehk } 3\cdot V_4^2 = 36.$$

Permutatsioonideks *nimetatakse põhihulga n-elemendilisi järjestatud alamhulki.* Permutatsioonid on seega ühendid, mis erinevad üksteisest ainult elementide järjestuse poolest. Tähistame n elemendist moodustatud permutatsioone kujul P_n. Ilmselt $P_n = V_n^n = n(n-1)(n-2)\ldots 3\cdot 2\cdot 1$ ehk

$$P_n = 1\cdot 2\cdot\ldots\cdot n = n!. \tag{2}$$

$n! = 1\cdot 2\cdot\ldots\cdot n$ loetakse n-faktoriaal. Näiteks 12!=479001600.

Näide 4. Kuus võrkpallurit on võimalik paigutada väljakule $P_6 = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720$ erineval viisil.

Kui valemit (1) korrutada ja jagada suurusega $(n-m)!$, siis saab selle valemi esitada kujul

$$V_n^m = \frac{n!}{(n-m)!}. \tag{3}$$

Kombinatsioonideks *nimetatakse põhihulga m-elemendilisi alamhulki.* Ehk teisiti, kombinatsioonideks n erinevast elemendist m elemendi kaupa nimetatakse ühendeid, mis *erinevad üksteisest vähemalt ühe elemendi poolest* (kuid elementide *järjekord* pole oluline, vastupidiselt variatsioonidele). Tähistame kombinatsioone n elemendist m kaupa kujul $C_n^m \equiv \binom{n}{m}$. Igast kombinatsioonist C_n^m saab moodustada $m!$ erinevat permutatsiooni. Järelikult $m!\cdot C_n^m = V_n^m$ ja

$$C_n^m = \frac{V_n^m}{P_m}. \tag{4}$$

Valemile (4) võib valemi (3) abil anda teise kuju:

$$C_n^m = \frac{n!}{m!\cdot(n-m)!}. \tag{5}$$

Asendades valemis (5) tähe m vahega $n-m$, saab

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

ja seega

$$C_n^m = C_n^{n-m}. \tag{6}$$

Näide 5.

$$C_5^2 = \frac{5 \cdot 4}{1 \cdot 2} = 10.$$

Näide 6.

$$C_{100}^{97} = C_{100}^3 = \frac{100 \cdot 99 \cdot 98}{1 \cdot 2 \cdot 3} = 161700.$$

Näide 7. 25-liikmelisest kollektiivist on vaja saata 3 delegaati ametiühingukonverentsile. Kuna delegaatide (ühendite keeles elementide) järjekord ei ole oluline, siis on tegemist kombinatsioonidega ning valiku võimalusi on

$$C_{25}^3 = \frac{25 \cdot 24 \cdot 23}{1 \cdot 2 \cdot 3} = 2300.$$

Näide 8. 25-liikmelisest õpperühmast on vaja valida rühmavanem, tema asetäitja ning spordiorganisaator. Kuna pole sugugi ükskõik, missuguseks funktsionääriks konkreetne üliõpilane valitakse, siis on siin (ühendite keeles rääkides) elementide järjekord oluline ning tegemist on variatsioonidega. Valiku võimalusi on seega

$$V_{25}^3 = 25 \cdot 24 \cdot 23 = 13800.$$

Näide 9. Lapsel nimega Katrin on käes 27 ladina tähestiku tähte. Ta võtab nende hulgast juhuslikult 6. Tõenäosus, et ta saab oma nimes olevad tähed kätte õiges järjekorras (sündmus A), on

$$P(A) = \frac{1}{V_{27}^6}.$$

Tõenäosus, et ta saab oma nimes olevad tähed kätte järjekorda arvestamata (sündmus B), on

$$P(B) = \frac{1}{C_{27}^6}.$$

Mõlemal juhul on soodsate elementaarsündmuste arv 1.

Näide 10. Ühele riiulile paigutatakse 10 raamatut juhuslikus järjekorras. Tõenäosus, et 3 mingit kindlat raamatut seejuures kõrvuti satuvad, on

$$P(A) = \frac{8! \cdot 3!}{10!} = \frac{1}{15}.$$

Kõiki võrdvõimalikke elementaarsündmusi on siin 10!, kuna kõik elemendid on korraga vaatluse all ning nende järjekord on oluline. Soodsate sündmuste arvu leidmiseks kujutame ette, et kolm raamatut on kokku seotud ning moodustavad ühe elemendi. Siis on meil 8 elemendi riiulile paigutamiseks 8! võimalust. Neist igas asendis on aga 3! võimalust kolme kõrvuti oleva raamatu ümberpaigutamiseks.

Näide 11. Kaubasaadetises on 4 kõrgema sordi toodet ja 6 esimese sordi toodet. Tõenäosus, et viie juhuslikult võetud toote hulgas on kõik esimese sordi tooted (sündmus A), arvutatakse järgmiselt:

$$P(A) = \frac{C_6^5}{C_{10}^5} = \frac{1}{42}.$$

Tõenäosus, et viie juhuslikult võetud toote hulgas on (täpselt) 2 kõrgema sordi toodet (sündmus B), avaldub aga kujul

$$P(B) = \frac{C_4^2 \cdot C_6^3}{C_{10}^5} = \frac{10}{21}.$$

Soodsate sündmuste arvu leidmiseks võib kasutada näiteks järgmist mõttekäiku. Nelja kõrgema sordi toote hulgast kahe juhuslikuks võtmiseks on C_4^2 võimalust (elementide järjekord pole oluline). $5 - 2 = 3$ esimese sordi toote võtmiseks on aga C_6^3 võimalust. Iga kõrgema sordi toodete varianti võib ühendada kõikide esimese sordi toodete variantidega. Niisiis, soodsate sündmuste arvu saamiseks tuleb kaks eelkirjeldatud kombinatsiooni omavahel korrutada: $C_4^2 \cdot C_6^3$.

Näide 12. Viieliikmelisse uurimisrühma kandideerib 12 kandidaati, neist 4 on naised. Tõenäosus, et uurimisrühma moodustamisel loosimise teel satub sinna 2 naist, arvutatakse lähtudes samast ideest kui näites 11:

$$P(B) = \frac{C_4^2 \cdot C_8^3}{C_{12}^5} = \frac{14}{33}.$$

2.6. Üksteist välistavate sündmuste liitmisteoreem

Tõenäosuse statistilise definitsiooni alusel (vt punkti 2.4) ja lähtudes sündmuse sageduse omadustest (vt punkti 2.3) tuuakse sisse viis tõenäosusteooria aksioomi (neist viimane järgmises punktis). Niisuguse idee põhjal üles ehitatud **tõenäosusteooria aksiomaatika** tagab teooria kooskõla katsetega.

Aksioom 1. $0 \leq P(A) \leq 1$.

Aksioom 2. $P(\Omega) = 1$.

Aksioom 3. $P(\emptyset)=0$.

Aksioom 4. Kui sündmused A ja B on teineteist **välistavad**, siis nende summa tõenäosus avaldub kujul

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

Viimase üldistuseks on **üksteist välistavate sündmuste liitmisteoreem**

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots P(A_n). \tag{2}$$

Üksteist välistavad sündmused $A_1, A_2, \ldots, A_n$ moodustavad **täieliku süsteemi**, kui katse tulemusena toimub neist tingimata üks ja ainult üks. Seega $P(A_1+A_2+\ldots+A_n) = 1$.

Järeldus 1. Kui üksteist välistavad sündmused $A_1, A_2, \ldots, A_n$ moodustavad täieliku sündmuste süsteemi, siis vastavalt eelnevale valemile ning liitmisteoreemile

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1. \tag{3}$$

Järeldus 2. Kuna vastandsündmused moodustavad täieliku süsteemi, siis on nende tõenäosuste summa võrdne ühega:

$$P(A) + P(\overline{A}) = 1 \text{ ja } P(A) = 1 - P(\overline{A}). \tag{4}$$

Näide. Urnis on 2 rohelist, 7 punast, 5 kollast ja 10 valget kuuli. Urnist võetakse üks kuul. Tõenäosuse, et see kuul on värviline, võib arvutada kas valemiga (2)

$$P(Värviline) = P(Kollane) + P(Punane) + P(Roheline) =$$

$$= \frac{5}{24} + \frac{7}{24} + \frac{2}{24} = \frac{7}{12}$$

või valemiga (4), toetudes vastandsündmuse mõistele

$$P(Värviline) = 1 - P(Valge) = 1 - \frac{10}{24} = \frac{7}{12}.$$

2.7. Korrutamisteoreem

Toetudes punkti 2.3 valemile (1), tuuakse sisse tõenäosusteooria **viies aksioom**:

$$P(AB) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A). \tag{1}$$

$P(A|B)$ on sündmuse A tõenäosus tingimusel, et sündmus B toimus.
Viienda aksioomi üldistuseks n sündmusele on **korrutamisteoreem**:

$$P(A_1A_2\ldots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1A_2\ldots A_{n-1}). \tag{2}$$

Kui $P(A|B) = P(A)$ ja $P(B|A) = P(B)$, siis sündmused A ja B on teineteisest sõltumatud ning korrutamisteoreem lihtsustub:

$$P(AB) = P(A) \cdot P(B). \tag{3}$$

Näide 1. Tõenäosus, et kõik vahetuse jooksul valmistatud detailid on standardsed, on 0,9. Leida tõenäosus, et kolme vahetuse jooksul ei toodeta ühtegi praakdetaili (sündmus A).

Märkigu A_1 sündmust, et esimeses vahetuses valmistatud detailid on standardsed. A_2 ja A_3 olgu analoogsed, vastavalt teise ja kolmanda vahetusega seotud sündmused. Siis $A = A_1 \cdot A_2 \cdot A_3$. Sündmused A_1, A_2, A_3 on sõltumatud ja seega

$$P(A) = P(A_1A_2A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) = 0,9^3 = 0,729.$$

Näide 2. Õpperühmas on 4 tütarlast ja 16 noormeest. Eksamile ilmujate järjekord koostatakse juhuslikult. Eeldusel, et kõik üliõpilased on lubatud eksamile, leida tõenäosus, et 1) kolm esimest eksamile ilmujat on noormehed (sündmus N);

2) vähemalt üks kolmest esimesest on tütarlaps (sündmus T).

Esimesele küsimusele vastamine nõuab korrutamisteoreemi (2) kasutamist:

$$P(N) = P(N_1N_2N_3) = P(N_1) \cdot P(N_2|N_1) \cdot P(N_3|N_1N_2) = \frac{16}{20} \cdot \frac{15}{19} \cdot \frac{14}{18} = \frac{28}{57}.$$

Teise küsimuse saab lahendada vastandsündmuse tõenäosuse abil:

$$P(T) = 1 - P(N) = 1 - 28/57 = 29/57.$$

Näide 3. Kastis on 25 detaili. Neist 6 on defektiga. Kui suur on tõenäosus, et kahe järgemööda pimesi võetud detaili hulgas on *teine* defektiga?

P(I defektiga ja II defektiga või I defektita ja II defektiga) $=$

$$= \frac{6}{25} \cdot \frac{5}{24} + \frac{19}{25} \cdot \frac{6}{24} = \frac{6}{25}.$$

Näide 4. Vaatlused on näidanud, et iga-aastasel maratonijooksul läbib 90% osavõtjatest 3/4 distantsist ja 80% kogu distantsi. Leida tõenäosus, et 3/4 distantsist läbinud sportlane lõpetab jooksu.

Sündmus A: sportlane on läbinud 3/4 distantsist; $P(A) = 0,9$.
Sündmus B: sportlane lõpetab jooksu.
Sündmus AB: sportlane on läbinud 3/4 distantsist ja lõpetab jooksu; $P(AB) = 0,8$.
Sündmus $B|A$: sportlane lõpetab jooksu tingimusel, et ta on läbinud 3/4 distantsist.

Valemist (1) saab $P(B|A) = \dfrac{P(AB)}{P(A)} = \dfrac{0,8}{0,9} \approx 0,89$.

2.8. Üksteist mittevälistavate sündmuste liitmisteoreem

Näide 1. Esimene üliõpilane lahendab ülesande (sündmus A) tõenäosusega 0,75, teine aga (sündmus B) tõenäosusega 0,8. Kui tõenäone on, et *vähemalt* üks üliõpilane lahendab ülesande, kui nad töötavad teineteisest sõltumatult?

Antud tõenäosusi liites 0,75+0,8=1,55 saame tulemuse üle ühe, mis on ilmselt väär.

Kahe teineteist **mittevälistava** *sündmuse summa tõenäosuse kohta kehtib järgmine* **teoreem**:

$$P(A + B) = P(A) + P(B) - P(AB). \tag{1}$$

Teoreemi geomeetrilist tähendust võib illustreerida järgneva Venni diagrammiga:

Diagrammil olevat ristviirutusega osa tuleb arvesse võtta *üks* kord. Tõlgendades hulkasid A ja B piltlikult märklaudadena, oleks ühisosa kohal muidu kahekordne märklaud.

Vastandsündmusele üle minnes võib valemi (1) asemel kasutada järgmist valemit:

$$P(A+B) = 1 - P(\overline{A}\,\overline{B}). \tag{2}$$

Kolme üksteist mittevälistava sündmuse puhul

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC). \tag{3}$$

Vastandsündmusele üleminek nõuab aga ilmselt väiksemaid arvutusi $P(A+B+C) = 1 - P(\overline{A}\,\overline{B}\,\overline{C})$. Kolme ja enama sündmuse puhul tulekski seega soovitada üleminekut vastandsündmusele.

Lahendame nüüd ülaltoodud **näite 1**. Arvestades, et sündmused A ja B on teineteist *mittevälistavad*, kuid *sõltumatud* (vt lisaks punkti 2.7 valemit (3)), saame vastavalt antud punkti valemile (1)

$$P(A+B) = 0,75 + 0,8 - 0,75 \cdot 0,8 = 0,95$$

ehk vastandsündmusele üleminekul vastavalt valemile (2)

$$P(A+B) = 1 - 0,25 \cdot 0,2 = 0,95.$$

Näide 2. Teatud marsruudil reisimiseks on võimalik kasutada kas rongi (sündmus A), bussi (sündmus B) või lennukit (sündmus C). Tõenäosus, et mingil päeval nimetatud liiklusvahendid ei välju, on vastavalt 0,1; 0,2 ja 0,5. Leida tõenäosus, et vähemalt ühega neist saab sõita sellel marsruudil.

Sündmused A, B ja C on üksteist *mittevälistavad*, kuid *sõltumatud*. Seega

$$P(A+B+C) = 1 - 0,1 \cdot 0,2 \cdot 0,5 = 0,99 \quad \text{ehk} \quad P(A+B+C) =$$

$$= 0,9 + 0,8 + 0,5 - 0,9 \cdot 0,8 - 0,9 \cdot 0,5 - 0,8 \cdot 0,5 + 0,9 \cdot 0,8 \cdot 0,5 = 0,99.$$

2.9. Täistõenäosuse valem

Moodustugu üksteist välistavatest elementaarsündmustest $H_1, H_2, \ldots, H_n$ täielik süsteem. Nimetame neid sündmusi **hüpoteesideks**. Toimugu meid huvitav sündmus A vaid koos ühega hüpoteesidest:

$$A = AH_1 + AH_2 + \ldots + AH_n.$$

Selles seoses olevad liidetavad on üksteist välistavad sündmused, kuna hüpoteesid olid üksteist välistavad. Seega võib kasutada üksteist välistavate sündmuste tõenäosuste liitmisteoreemi:

$$P(A) = \sum_{i=1}^{n} P(AH_i).$$

Korrutamisteoreemi abil

$$P(AH_i) = P(H_i) \cdot P(A|H_i) \quad \text{ning}$$

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

Viimane ongi **täistõenäosuse valem**.

Näide 1. Kauplusse saabusid ühetüübilised elektripirnid, mis olid valmistatud neljas tehases: esimeses 250 pirni, teises 525, kolmandas 275 ning neljandas 950. Tõenäosus, et pirn põleb üle 1500 tunni, on esimeses tehases 0,15, teises 0,30, kolmandas 0,20 ning neljandas 0,10. Riiulitele ladumisel läksid pirnid segamini. Kui suur on tõenäosus, et ostetud pirn põleb üle 1500 tunni?

Sündmus A: pirn põleb üle 1500 tunni. H_1, H_2, H_3, H_4 on hüpoteesid, et pirn on valmistatud vastavalt esimeses, teises, kolmandas või neljandas tehases.

Kuna pirne on kokku 2000, siis hüpoteeside tõenäosused

$$P(H_1) = \frac{250}{2000} = 0,125, \qquad P(H_2) = \frac{525}{2000} = 0,2625,$$

$$P(H_3) = \frac{275}{2000} = 0,1375, \qquad P(H_4) = \frac{950}{2000} = 0,475.$$

Hüpoteesid moodustavad täieliku süsteemi ja nende tõenäosuste summa peab olema üks. Sündmuse A tinglikud tõenäosused on

$$P(A|H_1) = 0,15, \qquad P(A|H_2) = 0,30,$$

$$P(A|H_3) = 0,20, \qquad P(A|H_4) = 0,10.$$

Täistõenäosuse valemi (1) alusel

$$P(A) = \sum_{i=1}^{4} P(H_i)P(A|H_i) =$$

$$= 0,125 \cdot 0,15 + 0,2625 \cdot 0,30 + 0,1375 \cdot 0,20 + 0,475 \cdot 0,10 = 0,1725.$$

Näide 2. Ühes turismigrupis on 3 võõrkeeleoskajat ja 2 mitteoskajat. Teises grupis on need arvud vastavalt 4 ja 4. Esimesest grupist saadeti valikuta teise üks turist. Leida tõenäosus, et nüüd teisest grupist juhuslikult välja kutsutud turist on võõrkeeleoskaja (sündmus A).

Hüpotees H_1: esimesest grupist saadi võõrkeele mitteoskaja.

Hüpotees H_2: esimesest grupist saadi võõrkeeleoskaja.

$$P(H_1) = \frac{2}{5}, \qquad P(H_2) = \frac{3}{5},$$

$$P(A|H_1) = \frac{4}{9}, \qquad P(A|H_2) = \frac{5}{9}.$$

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{2}{5}\cdot\frac{4}{9} + \frac{3}{5}\cdot\frac{5}{9} = \frac{23}{45}.$$

2.10. Bayesi valem (hüpoteesi tinglik tõenäosus)

Olgu antud hüpoteeside täielik süsteem $H_1, H_2, \ldots, H_n$ ning olgu teada nende hüpoteeside tõenäosused $P(H_1), P(H_2), \ldots, P(H_n)$. Tehakse katse, mille tulemuseks on mingi sündmus A, mille tinglikud tõenäosused $P(A|H_1), P(A|H_2), \ldots, P(A|H_n)$ olgu teada. Siis avaldub hüpoteesi H_i tinglik tõenäosus järgneva **Bayesi valemi** kujul:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum\limits_{i=1}^{n} P(H_i)P(A|H_i)}. \tag{1}$$

Valemi (1) kehtivust saab põhjendada täistõenäosuse valemi ning korrutamisteoreemiga.

Bayesi valemi kasutamist nõudvates ülesannetes on iseloomulik, et kõigepealt püstitatakse hüpoteesid. Seejärel tehakse katse, mille tulemus on teada. Sõltuvalt katse tulemusest hinnatakse hüpoteeside esialgsed tõenäosused ümber.

Näide 1. On teada, et 96% toodangust vastab standardile. Lihtsustatud kontrollimise süsteem tunnistab kõlblikuks standardse toote tõenäosusega 0,98, mittestandardse aga tõenäosusega 0,05. Kui suur on tõenäosus, et toode, mis tunnistati kõlblikuks, tõepoolest vastab ka standardile?

Sündmus A: toode tunnistati kõlblikuks.
Hüpotees H_1: toode vastab standardile.
Hüpotees H_2: toode ei vasta standardile.

$$P(H_1) = 0,96, \qquad P(A|H_1) = 0,98,$$

$$P(H_2) = 0,04, \qquad P(A|H_2) = 0,05.$$

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{\sum\limits_{i=1}^{2} P(H_i)P(A|H_i)} = \frac{0,96\cdot 0,98}{0,96\cdot 0,98 + 0,04\cdot 0,05} = 0,9979.$$

Näide 2. Esimene õhutõrjepatarei tegi 10 lasku, teine patarei 8 lasku ja kolmas patarei 12 lasku. Tabamise tõenäosused olid vastavalt 0,3, 0,4 ja 0,2. Üks lask tabas lennukit. Missugusele patareile see lask kõige tõenäosemalt kuulus?

Sündmus A: Lennuki tabamine.
Hüpotees H_1: laskis I patarei, $P(H_1) = 10/30$, $P(A|H_1) = 0,3$.
Hüpotees H_2: laskis II patarei, $P(H_2) = 8/30$, $P(A|H_2) = 0,4$.
Hüpotees H_3: laskis III patarei, $P(H_3) = 12/30$, $P(A|H_3) = 0,2$.

$$P(H_1|A) = 15/43, \quad P(H_2|A) = 16/43, \quad P(H_3|A) = 12/43.$$

Vastuseks on $P(H_2|A) = 16/43$, mis tähendab, et kõige tõenäosemalt kuulus lennukit tabanud lask teisele patareile.

Kuna hüpoteesid moodustavad täieliku süsteemi, siis nii

$$\sum_{i=1}^{3} P(H_i) = 1 \quad \text{kui ka} \quad \sum_{i=1}^{3} P(H_i|A) = 1.$$

2.11. Bernoulli valem. Tõenäoseim sündmuse toimumiste arv

Püstitame järgmise ülesande. Leida tõenäosus, et n ühesuguse ja sõltumatu katse tulemusel toimuks sündmus A täpselt m korda. Olgu igal üksikkatsel sündmuse A toimumise tõenäosus konstantne $P(A) = p$. Vastandsündmuse $\overline{A}$ toimumise tõenäosus on siis $P(\overline{A}) = 1 - p = q$. Tähistame otsitava tõenäosuse $P_{m,n}$. Tõenäosuse $P_{m,n}$ arvutamine oleks liitmis- ja korrutamisteoreemi abil võimalik, kuid katsete arvu suurenedes nõuaks see mahukaid arvutusi. Nimetatud valemitest lähtudes saab tuletada arvutusi lihtsustava **Bernoulli valemi**:

$$P_{m,n} = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m}. \tag{1}$$

Näide 1. Telestuudios on 5 kaamerat. Tõenäosus, et kaamera on vaadeldaval hetkel sisse lülitatud, on 0,65. Kui tõenäone on, et antud hetkel ei ole sisse lülitatud rohkem kui kolm kaamerat?

Sündmus A: antud hetkel ei ole sisse lülitatud rohkem kui kolm kaamerat.

Sündmused $A_0, A_1, \ldots, A_5$: sisse on lülitatud 0, 1, ..., 5 kaamerat.

Seega $A = A_0 + A_1 + A_2 + A_3$ ehk vastandsündmuse kaudu veidi lihtsamalt $A = 1 - A_4 - A_5$ ja

$$P(A) = 1 - P(A_4) - P(A_5) = 1 - P_{4,5} - P_{5,5} = C_5^4 0,65^4 0,35 + C_5^5 0,65^5 = 0,66304.$$

Sündmused, mis seisnevad sündmuse A toimumises n katsel $0, 1, \ldots, n$ korda, on üksteist välistavad ning moodustavad sündmuste täieliku süsteemi. Eelmise näite puhul moodustavad täieliku süsteemi sündmused, et sisse on lülitatud 0, 1, 2, 3, 4 ja 5 kaamerat. Seega $\sum_{m=0}^{n} P_{m,n} = 1$.

Kui arvutada kõik kuus eelmainitud tõenäosust, siis on võimalik teada saada, mitu kaamerat on kõige tõenäosemalt sisse lülitatud:

m	0	1	2	3	4	5
$P_{m,5}$	0,00525	0,04877	0,18115	0,33642	0,31239	0,11603

Kuna $P_{3,5} = 0,33642$ on maksimaalne, siis on kõige tõenäosemalt sisse lülitatud 3 kaamerat.

Tõenäoseimat sündmuse toimumiste arvu m_0, mille puhul $P_{m_0,n}$ on maksimaalne, saab aga lihtsamalt leida. Ilmselt

$$P_{m_0,n} \geq P_{m_0+1,n} \quad \text{ja} \quad P_{m_0,n} \geq P_{m_0-1,n}.$$

Neist võrratustest lähtudes saab

$$np - q \leq m_0 \leq np + p. \tag{2}$$

Näite 1 andmetele vastavalt

$$5 \cdot 0,65 - 0,35 \leq m_0 \leq 5 \cdot 0,65 + 0,65,$$

millest $2,9 \leq m_0 \leq 3,9$ ja seega tõenäoseimalt on sisse lülitatud $m_0 = 3$ kaamerat.

Näide 2. Märgi tabamise tõenäosus on 0,25. Tulistati 21 lasku. Leida tõenäoseim tabamuste arv ning vastav tõenäosus.

$n = 21$, $p = 0,25$, $q = 0,75$. Võrratuste (2) kohaselt

$$21 \cdot 0,25 - 0,75 \leq m_0 \leq 21 \cdot 0,25 + 0,25,$$

millest $4,5 \leq m_0 \leq 5,5$ ja tõenäoseim tabamuste arv $m_0 = 5$.

Vastav tõenäosus arvutatakse Bernoulli valemiga:

$$P_{5,21} = C_{21}^5 \cdot 0,25^5 \cdot 0,75^{16} \approx 0,199.$$

Märkus 1. Kui Bernoulli valemis olev katsete arv n on suur, siis võib Bernoulli valemiga arvutamine olla raskendatud või isegi võimatu. Sellisel juhul kasutatakse ligikaudseid asümptootilisi valemeid, millega tutvutakse järgnevates peatükkides.

Märkus 2. Bernoulli valemi kasutamine eeldab, et sündmuse toimumise tõenäosus igal üksikkatsel p on konstantne. Kui aga *erinevates* tingimustes (mitte ühesugustes, nagu see Bernoulli valemi rakendamise puhul vajalik oli) tehakse n sõltumatut katset, kusjuures sündmuse A toimumise tõenäosus i-ndal katsel on p_i, siis tõenäosus $P_{m,n}$ (et sündmus A toimuks n katsel täpselt m korda) on leitav genereeriva funktsiooni $\psi_n(x)$ abil:

$$\psi_n(x) = (q_1 + p_1 x)(q_2 + p_2 x) \ldots (q_n + p_n x), \tag{3}$$

kus $q_i = 1 - p_i$. $P_{m,n}$ on liikme x^m kordaja.

Näide 3. Üksteise järel saadetakse välja neli raadiosignaali. Iga signaali vastuvõtt on sõltumatu ülejäänud signaalide vastuvõtust. Üksikute signaalide vastuvõtu tõenäosused on 0,2, 0,3, 0,4, 0,5. Leida kolme signaali vastuvõtu tõenäosus.

Kasutades valemit (3), saame:

$$\psi_4(x) = (0,8 + 0,2x)(0,7 + 0,3x)(0,6 + 0,4x)(0,5 + 0,5x) =$$

$$= 0,168 + 0,394x + 0,32x^2 + 0,106x^3 + 0,012x^4.$$

Otsitav $P_{3,4}$ on x^3 kordajaks: $P_{3,4} = 0,106$.

2.12. Ülesanded

2.1. Üliõpilasel tuleb eksamisessioonil sooritada 4 eksamit. Tähistame A_1-ga sündmuse, et üliõpilane sooritas esimese eksami vähemalt rahuldavalt (A_2, A_3, A_4: sooritas vastavalt II, III, IV eksami vähemalt rahuldavalt). Avaldada sündmuste A_1, A_2, A_3, A_4 ning nende vastandsündmuste $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$, $\overline{A}_4$ kaudu järgmised sündmused:

1) üliõpilane sooritab kõik eksamid;

2) üliõpilane sooritab *vähemalt* ühe eksami;

3) üliõpilane saab *vähemalt* ühe mitterahuldava;

4) üliõpilane saab *täpselt* ühe mitterahuldava (suvalisel eksamil).

2.2. 36-st kaardist koosnevast pakist võetakse juhuslikult üks. Kui suur on tõenäosus, et see on poti mastist?

2.3. Kui tõenäone on, et uue passi number lõpeb

1) paaritu numbriga,

2) nulliga?

2.4. Üliõpilane läks eksamile, olles 25 kordamisküsimusest 20 selgeks õppinud. Talle esitati 3 küsimust. Leida tõenäosus, et ta:

1) oskas vastata kõikidele küsimustele;

2) ei osanud vastata ühelegi küsimusele.

2.5. Kui palju võib moodustada kolmekohalisi positiivseid täisarve, millest igaüks koosneks kolmest erinevast numbrist?

2.6. Mitmel viisil on võimalik paigutada neli inimest neljale kaetud söögilauda?

2.7. Lapsel nimega KATRIN on käes 27 ladina tähestiku tähte. Ta võtab nende hulgast juhuslikult 6. Kui suur on tõenäosus, et ta saab kätte oma nime tähed

1) õiges järjekorras;

2) järjekorda arvestamata.

2.8. 36-st kaardist koosnev pakk jagatakse pooleks. Leida tõenäosus, et kummaski pakis oleks kaks ässa.

2.9. Õpperühmas on 8 mees- ja 12 naisüliõpilast. Neist 6 kutsutakse juhusliku valiku teel eksamiruumi. Leida tõenäosus selleks, et

1) kõik sisenejad on tüdrukud,

2) kõik sisenejad on poisid,

3) sisenejate hulgas on 4 tüdrukut.

2.10. Mängude üldise arvu vähendamiseks jaotati 20 võistkonda kahte alagruppi. Kui suur on tõenäosus, et kaks kõige tugevamat võistkonda satuvad

1) erinevatesse alagruppidesse,

2) samasse alagruppi?

2.11. Kui palju on võimalusi tähtede A, I, L, L, N, N, T rittapaigutamiseks? Mitmel juhul saab tulemuseks sõna TALLINN ja kui suur on vastav tõenäosus?

2.12. Üheksal kaardil on tähed E, N, O, S, S, T, U, Õ, Ä. Leida tõenäosus, et neid kaarte juhuslikult ritta ladudes saadakse sõna TÕENÄOSUS.

2.13. Rühmavanem ostis 12 kinopiletit ühte ritta ja jaotas need huupi 12-le üliõpilasele. Kui suur on tõenäosus, et kaks sõpra satuvad kõrvuti?

2.14. Raha ja asjade loterii iga 1000 pileti kohta tuleb 5 rahalist ja 25 esemelist võitu. Kui tõenäone on võita ühe piletiga?

2.15. Tõenäosus, et esimesel tööpingil valmistatud detail on esimest sorti, on 0,7. Samasuguse esimese sordi detaili valmistamise tõenäosus teisel pingil on 0,8. Esimesel tööpingil valmistati kaks detaili ja teisel üks detail. Kui suur on tõenäosus, et kõik valmistatud detailid on esimest sorti?

2.16. Seitsmest kaardist on moodustatud sõna TALLINN. Kaardid segatakse ja neist võetakse juhuslikult neli. Kui suur on tõenäosus, et saadakse nimi INNA?

2.17. Merehädas oleva laeva meeskonnale heidetakse kahelt lennukilt langevarju abil varustust. Tõenäosused selleks, et varustus langeb laevale, on vastavalt 0,75 ja 0,65. Kui suur on tõenäosus selleks, et vähemalt ühelt lennukilt heidetud varustus langeb laevale?

2.18. Tõenäosus saada raamat esimesest raamatukogust on 0,5, teisest 0,7 ja kolmandast 0,4. Kui suur on tõenäosus, et see raamat on olemas vähemalt ühes raamatukogus?

2.19. Kahel laskuril on märklaua tabamise tõenäosused vastavalt 0,7 ja 0,8. Leida tõenäosus, et

1) mõlemad tabavad esimese lasuga;
2) mõlemad lasevad mööda;
3) tabab üks laskudest;
4) tabab vähemalt üks laskudest.

2.20. Visatakse korraga kahte täringut. Kui suur on tõenäosus, et vähemalt ühel täringul tuleb 6 silma?

2.21. Tööline teenindab kolme tööpinki, mis töötavad üksteisest sõltumatult. Tõenäosus, et tunni aja jooksul ei nõua esimene tööpink töölise tähelepanu, on 0,95. Teise tööpingi korral on see tõenäosus 0,9 ning kolmandal 0,8. Leida tõenäosus, et tunni aja jooksul

1) ükski tööpink ei nõua töölise tähelepanu,
2) üks tööpink nõuab tähelepanu,
3) *vähemalt* üks tööpink nõuab töölise tähelepanu.

2.22. Tehas valmistab teatud detaile. Iga detail võib olla defektne tõenäosusega 0,2. Detaile kontrollib kontrolör, kes avastab esineva defekti tõenäosusega 0,9. Kui defekte ei avastata, siis pannakse detail valmistoodangusse. Peale selle võib kontrolör eksida, praakides tõenäosusega 0,05 välja korras detaili. Leida tõenäosus, et kontrollimiseks juhuslikult võetud detail

1) praagitakse välja ekslikult;
2) läheb valmistoodangu hulka defektsena;
3) praagitakse välja.

2.23. Lattu saabunud detailidest on valmistatud esimeses vabrikus 20%, teises 46% ja kolmandas 34%. On teada, et esimese vabriku toodangus leidub mittestandardseid detaile 3%, teises 2% ja kolmandas 1%. Leida tõenäosus, et juhuslikult võetud

1) detail on mittestandardne;

2) mittestandardne detail on valmistatud esimeses vabrikus.

2.24. Tsehhis töötab 25 ühesuguse tootlikkusega tööpinki. Neist 15 pinki on A-tüüpi, 4 B-tüüpi ja 6 C-tüüpi. Tõenäosus, et toode on esimest sorti, on A-tüüpi pingil 0,9, B-tüüpi pingil 0,8 ja C-tüüpi pingil 0,7. Leida tõenäosus, et tsehhi kogutoodangust juhuslikult võetud toode on esimest sorti.

2.25. 18 laskuri hulgast viis tabas märki tõenäosusega 0,8, seitse tõenäosusega 0,7, neli tõenäosusega 0,6 ja kaks tõenäosusega 0,5. Juhuslikult valitud laskur tulistas, kuid ei tabanud märki. Leida tõenäosus tema kuulumiseks igasse vaadeldavasse rühma. Missugusesse rühma see laskur kõige tõenäosemalt kuulub?

2.26. USA kiire majanduskasvu perioodil tugevneb dollar tõenäosusega 0,7, keskmise kasvu perioodil tugevneb dollar tõenäosusega 0,4 ja aeglase kasvu perioodil tõenäosusega 0,2. Kiire majanduskasvu tõenäosus on 0,3, keskmise kasvu tõenäosus on 0,5 ja aeglase kasvu tõenäosus 0,2. Vaadeldaval perioodil dollar tugevnes. Missuguse tõenäosusega on tegemist kiire majanduskasvu perioodiga?

2.27. Haige uurimisel on kahtlus kahe haiguse H_1 ja H_2 suhtes, kusjuures $P(H_1) = 0,6$ ja $P(H_2) = 0,4$. Diagnoosi saab täpsustada analüüsiga, mille tulemuseks on kas positiivne või negatiivne reaktsioon. Haiguse H_1 puhul on positiivse reaktsiooni tõenäosus 0,9, haiguse H_2 puhul 0,5. Analüüsi tehti kaks korda ning mõlemal korral oli tulemus negatiivne. Leida haiguste H_1 ja H_2 põdemise tõenäosused.

2.28. Ants koos kahe sõbraga tulistasid lõbustuspargi tiirus õhupalle. Iga poiss tegi ühe lasu ja tabati kahte palli. Palli tabamise tõenäosus iga lasuga on mõlemal sõbral võrdselt 0,3 ja Antsul 0,6. Leida tõenäosus, et möödalaskja oli Ants.

2.29. Tõenäosus, et iga üliõpilane jõuab hommikul õigeaegselt loengule, on 0,9. Leida tõenäosus, et neljast üliõpilasest vähemalt kolm jõuab õigeaegselt loengule.

Ülesanne 2.30. Nelja üliõpilase õigeaegselt loengule jõudmise tõenäosused on vastavalt 0,8, 0,85, 0,9 ja 0,95. Leida tõenäosus, et neljast üliõpilasest vähemalt kolm jõuab õigeaegselt loengule.

2.31. Tõenäosus autobussi õigeaegseks lõpp-peatusesse saabumiseks on 0,85. Leida tõenäosus, et seitsmest liinile väljunud autobussist jõuab lõpp-peatusesse õigeaegselt 1) viis, 2) vähemalt viis bussi.

Mitu bussi jõuab kõige tõenäosemalt lõpp-peatusesse õigeaegselt?

2.32. Mis on tõenäosem, kas võita võrdse vastasega mängides (viik ei ole võimalik)

1) kolm partiid neljast või neli seitsmest;

2) vähemalt kolm partiid neljast või vähemalt viis partiid kaheksast?

2.33. Tõenäosus, et teatud korvpallur tabab ühe viskega korvi, on 0,4. Korvpallur teeb 11 viset. Kui suur on tõenäoseim korvide arv ja sellele vastav tõenäosus?

2.34. Tõenäosus, et vabriku toode on kõrgemat sorti, on 4/7. Kui suur on tõenäoseim kõrgema sordi toodete arv 13 juhuslikult valitud toote hulgas?

2.35. Leida tõenäoseim sademeteta päevade arv septembri esimeses dekaadis, kui mitmeaastaste vaatluste põhjal esineb septembris sademeid keskmiselt 17 päeval.

3. Juhuslikud suurused

3.1. Juhusliku suuruse mõiste

Juhuslikuks nimetatakse **suurust**, mis katse tulemusena sõltuvalt juhusest omandab ühe ja ainult ühe oma võimalikest väärtustest. Juhuslikke suurusi tähistatakse suurtähtedega $X, Y, \ldots$ ja nende võimalikke väärtusi väiketähtedega x_1, $x_2, \ldots$, y_1, $y_2, \ldots$ Juhuslikud suurused liigitatakse diskreetseteks ja pidevateks.

Diskreetne juhuslik suurus võib katse või vaatluse tulemusena omandada lõpliku või loenduva hulga isoleeritud väärtusi, mis erinevad üksteisest mingi lõpliku arvu võrra. Diskreetseks juhuslikuks suuruseks on näiteks
1) televiisori riknemiste arv aasta jooksul,
2) laskude arv kuni esimese märkitabamiseni,
3) defektiga detailide arv n detailist koosnevas partiis.

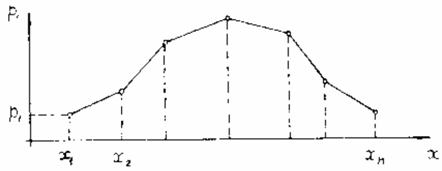
Pidev juhuslik suurus võib omandada kõiki väärtusi mingist lõplikust või lõpmatust vahemikust. Pidevaks juhuslikuks suuruseks on näiteks
1) auto keskmine kütusekulu $[l/100\ km]$,
2) elektrilüliti tööiga,
3) mürsu lennukaugus, mis kuulub mingisse vahemikku $]a, b[$.

3.2. Diskreetse juhusliku suuruse jaotusseadus

Diskreetse juhusliku suuruse jaotusseaduseks nimetatakse vastavust tema kõikide võimalike väärtuste $x_1, x_2, \ldots, x_n$ ja nende tõenäosuste $p_1, p_2, \ldots, p_n$ vahel. Jaotusseadust saab esitada tabeli kujul **jaotusreana**

X	x_1	x_2	$\ldots$	x_n
p	p_1	p_2	$\ldots$	p_n

või graafiliselt **jaotuspolügoonina**



Jaotuspolügoonil olevad punktid ühendatakse vaid piltlikkuse mõttes. Vahemikest $]x_1, x_2[$, $]x_2, x_3[$, $\ldots$, $]x_{n-1}, x_n[$ ei või juhuslik suurus väärtusi omandada.

Sündmused $x_1, x_2, \ldots, x_n$ moodustavad täieliku süsteemi ja seega

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

Jaotusseadus iseloomustab diskreetset juhuslikku suurust tõenäosuslikust seisukohast täielikult.

Punkti 2.10 näites 1 on diskreetseks juhuslikuks suuruseks antud hetkel sisselülitatud telekaamerate arv, mis võib omandada väärtusi $0, 1, 2, \ldots, 5$. Selle näite lahendamisel koostatud tabel, mida siin on korratud, ongi jaotusrida:

X	0	1	2	3	4	5
p	0,00525	0,04877	0,18115	0,33642	0,31239	0,11603

Sellele jaotusreale vastav jaotuspolügoon on järgmine:

3.3. Juhusliku suuruse jaotusfunktsioon

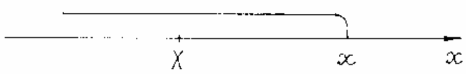
Pideva juhusliku suuruse jaoks pole võimalik jaotusrida välja kirjutada. Üldisemaks juhusliku suuruse (nii diskreetse kui pideva) jaotusseaduse esituseks on jaotusfunktsioon.

Juhusliku suuruse **jaotusfunktsiooniks** *nimetatakse funktsiooni $F(x)$, mis määrab iga reaalarvu x korral tõenäosuse, et juhuslik suurus X omandab väärtuse, mis on väiksem reaalarvust x, s.o*

$$F(x) = P(X < x), \quad \text{kus } x \in]-\infty, \infty[. \tag{1}$$

Pideva juhusliku suuruse definitsioon. *Pidevaks nimetatakse juhuslikku suurust, mille jaotusfunktsioon on pidev.*

Jaotusfunktsioon iseloomustab juhuslikku suurust tõenäosuslikult seisukohalt täielikult. Geomeetriliselt võib jaotusfunktsiooni interpreteerida järgmisel viisil. Olgu juhuslikuks suuruseks X juhuslik punkt, mis katse tulemusena satub x-teljele.



Jaotusfunktsioon määrab iga x puhul tõenäosuse, et juhuslik punkt X asetseb punktist x vasakul.

Jaotusfunktsiooni omadusi

1. $0 \leq F(x) \leq 1$. Jaotusfunktsioon on defineeritud kui tõenäosus; tõenäosus on aga alati mittenegatiivne arv, mis ei ületa ühte.

2. Tõenäosus, et juhuslik suurus omandaks väärtusi lõigul $[\alpha, \beta]$, on võrdne jaotusfunktsiooni muuduga sellel lõigul:

$$P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha). \tag{2}$$

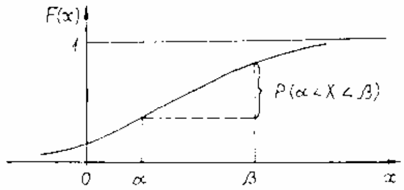
3. Jaotusfunktsioon on mittekahanev funktsioon, st kui

$$\alpha < \beta, \quad \text{siis} \quad F(\alpha) \leq F(\beta).$$

4. Kuna sündmus $A = (X < -\infty)$ on võimatu ja sündmus $B = (X < +\infty)$ on kindel, siis

$$\lim_{x \to -\infty} F(x) = F(-\infty) = 0 \quad \text{ja} \quad \lim_{x \to +\infty} F(x) = F(+\infty) = 1.$$

Eelnevatest omadustest järeldub, et pideva juhusliku suuruse jaotusfunktsiooni graafiku põhimõtteline kuju on järgmine:



3.4. Pideva juhusliku suuruse jaotustihedus

Lõigule $[x, x + \Delta x]$ pikkusega Δx vastav tõenäosus on

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x \quad = \Delta F.$$

Pikkusühikule vastav tõenäosus on $\Delta F / \Delta x$. Kui $\Delta x \to$), siis saadakse pideva **juhusliku suuruse** tõenäosuse tihedus e **jaotustihedus**:

$$\lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} = F'(x) = f(x), \tag{1}$$

mis on niisiis defineeritud kui jaotusfunktsiooni $F(x)$ tuletis.

Jaotustiheduse omadusi

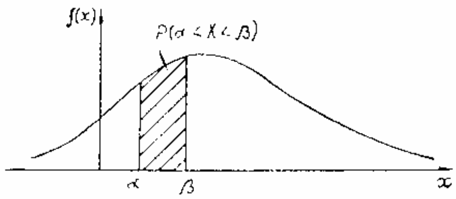
1. $f(x) \geq 0$, mis tuleneb sellest, et $F(x)$ on mittekahanev.
2.
$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

3.
$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

mis järeldub omadusest 2 ning teadmisest, et $F(-\infty) = 1$.

4.
$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx. \tag{2}$$

Jaotustiheduse graafiku põhimõtteline kuju on toodud järgmisel joonisel, kusjuures vastavalt jaotustiheduse kolmandale omadusele on graafiku ning x-telje vaheline pindala võrdne ühega.



Tõenäosus, et pidev juhuslik suurus omandab väärtusi lõigust $[\alpha, \beta]$, on võrdne joonisel viirutatud kõvertrapetsi pindalaga, millega saab antud tõenäosust graafiliselt suhteliselt piltlikumalt iseloomustada, kui jaotusfunktsioon $F(x)$ seda võimaldab (vt eelmise punkti lõpus olevat jaotusfunktsiooni graafikut).

3.5. Juhusliku suuruse keskväärtus

Juhuslikku suurust iseloomustab täielikult tema jaotusfunktsioon (või pideva juhusliku suuruse puhul ka jaotustihedus). Jaotuse oluliste külgede esiletoomiseks kasutatakse aga täiendavalt mitmeid juhusliku suuruse arvkarakteristikuid. Eelkõige on tähtis teada **keskväärtust**, mille ümber grupeeruvad juhusliku suuruse võimalikud väärtused.

Diskreetse juhusliku suuruse keskväärtus e. matemaatiline ootus on

$$EX = M[X] = m_x = \sum_{i=1}^{n} x_i p_i. \tag{1}$$

Vaatleme **diskreetse juhusliku suuruse** keskväärtuse tõenäosuslikku tähendust. Olgu tehtud k katset, milles juhuslik suurus X omandab väärtused $x_1, x_2, \ldots, x_n$ vastavalt $m_1, m_2, \ldots, m_n$ korda, kusjuures $m_1 + m_2 + \ldots + m_n = k$. Juhusliku suuruse X sagedustabel on siis järgmine:

X	x_1	x_2	...	x_n
Sageduse tihedus p_i^*	m_1/k	m_2/k	...	m_n/k

Arvutades juhusliku suuruse vaadeldavate väärtuste aritmeetilise keskmise $\overline{x}$ ning arvestades, et vastavalt tõenäosuse statistilisele definitsioonile (vt punkti 2.4) on sündmuse sagedus $p_i^* = m_i/k$ ligikaudselt võrdne selle sündmuse tõenäosusega p_i, saadakse

$$\overline{x} = (x_1 m_1 + x_2 m_2 + \ldots + x_n m_n)/k =$$

$$= x_1 \frac{m_1}{k} + x_2 \frac{m_2}{k} + \ldots + x_n \frac{m_n}{k} \approx x_1 p_1 + x_2 p_2 + \ldots x_n p_n = EX.$$

Niisiis, diskreetse juhusliku suuruse keskväärtus on ligikaudselt võrdne selle juhusliku suuruse vaadeldavate väärtuste aritmeetilise keskmisega, ja seda täpsemalt, mida suurem on katsete arv.

Pideva juhusliku suuruse $X \in]-\infty, \infty[$ keskväärtuse arvutamisel asendub summeerimine integreerimisega

$$EX = \int_{-\infty}^{\infty} x f(x) dx. \qquad (2)$$

Keskväärtuse omadusi

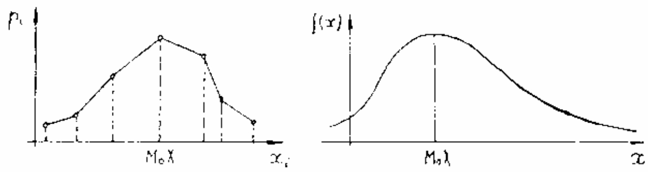
1. $EC = C$, kus $C = \mathrm{const}$.
2. $E(CX) = C \cdot EX$.
3. $E(X + Y) = EX + EY$.
4. Kui X ja Y on sõltumatud, siis $E(XY) = EX \cdot EY$.

3.6. Mood ja mediaan

Diskreetse juhusliku suuruse **mood** M_oX on selle suuruse *kõige tõenäosemalt* esinev väärtus.

Diskreetse juhusliku suuruse **mediaan** M_eX on variatsioonrea *keskmine element* paarituarvulise valimi korral või *kahe keskmise elemendi poolsumma* paarisarvulise valimi puhul.

Pideva juhusliku suuruse **mood** M_oX on selle suuruse niisugune väärtus, mille puhul jaotustihedus on maksimaalne, st $f(M_oX) = max$.

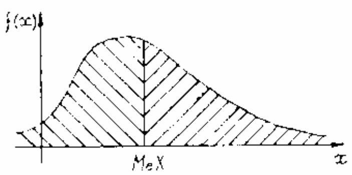


Mõnel juhul võib jaotustihedusel või jaotuspolügoonil olla mitu maksimumi (s.o mitu moodi), kuid maksimum võib ka puududa ning selle asemel võib olla miinimum (st mood puudub).

Pideva juhusliku suuruse **mediaan** M_eX on juhusliku suuruse X selline väärtus, mille puhul

$$P(X < M_eX) = P(X > M_eX) = 0,5.$$

Järgneval joonisel on mõlemat pidi viirutatud pindalad *võrdsed*.



Sümmeetrilisel ühe moodiga jaotusel langevad keskväärtus, mood ja mediaan kokku.

3.7. Dispersioon ja standardhälve

Järgnevas vaatleme arvkarakteristikuid, mis iseloomustavad juhusliku suuruse hajuvuse astet keskväärtuse ümber. Leiame kõigepealt juhusliku suuruse tsentreeritud hälbe keskväärtuse suhtes $X - EX$ ning viimase keskväärtuse $E(X - EX) = EX - E(EX) = EX - EX = 0$. Nulliga võrduv tulemus on selgitatav asjaoluga, et osa hälbeid on positiivsed, osa negatiivsed. Arvutades juhusliku suuruse ja tema keskväärtuse vahe ruudu keskväärtuse, saame juhusliku suuruse **dispersiooni**

$$DX = E(X - EX)^2. \tag{1}$$

Dispersiooni dimensiooniks on juhusliku suuruse dimensiooni ruut.

Sama dimensiooniga, kui juhuslik suurus X, on **standardhälve**

$$\sigma(X) = \sqrt{DX}. \tag{2}$$

Dispersiooni omadusi

1. $DC = 0$, kus C=const.
2. $D(CX) = C^2 DX$.
3. *Sõltumatute* juhuslike suuruste puhul $D(X + Y) = DX + DY$.
4. $$DX = EX^2 - (EX)^2. \tag{3}$$

Tõestus. $DX = E(X - EX)^2 = E[X^2 - 2X \cdot EX + (EX)^2] = EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2$.

Valem (3) sobib dispersiooni praktiliseks arvutamiseks paremini kui definitsioonikohane valem (1). Valemile (3) vastavalt arvutame **diskreetse juhusliku suuruse dispersiooni** kui summa

$$DX = \sum_{i=1}^{n} x_i^2 p_i - (EX)^2 \tag{4}$$

ning **pideva juhusliku suuruse dispersiooni** kui integraali

$$DX = \int_{-\infty}^{\infty} x^2 f(x)dx - (EX)^2. \tag{4}$$

Näide. Juhusliku suuruse X jaotusrida on

x_i	2	3	4	5	6
p_i	0,1	0,2	0,35	0,2	0,15

Leida juhusliku suuruse $3X - 2$ keskväärtus $E(3X - 2)$, dispersioon $D(3X - 2)$ ja standardhälve $\sigma(3X - 2)$.

Ülesande lahendamiseks on kaks võimalust.

1. Arvutame $EX = \sum_{i=1}^{5} x_i p_i = 4,1$ ja keskväärtuse omadusi (vt punkti 3.5) kasutades

saame $E(3X-2) = 3 \cdot EX - 2 = 10,3$.
Märgime, et mood $M_0X = 4$ ja samuti mediaan $M_eX = 4$.
Dispersiooni omadustest lähtuvalt leiame $D(3X-2) = 3^2 \cdot DX = 9 \cdot 1,39 = 12,51$,
milles $DX = EX^2 - (EX)^2 = 18,2 - 4,1^2 = 1,39$ ja $EX^2 = \sum_{i=1}^{5} x_i^2 p_i = 18,2$.
Standardhälve $\sigma(3X-2) = \sqrt{D(3X-2)} = \sqrt{12,51} \approx 3,54$.

2. Teiseks võimaluseks on koostada juhusliku suuruse $3X-2$ jaotusrida:

$3x_i - 2$	4	7	10	13	16
p_i	0,1	0,2	0,35	0,2	0,15

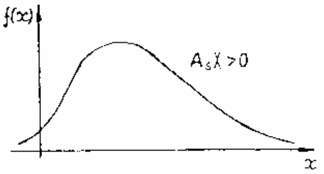
Selle põhjal arvutatakse $E(3X-2) = 10,3$. Jaotusreast on ühtlasi näha, et mood $M_0(3X-2) = 10$ ja mediaan $M_e(3X-2) = 10$. Dispersioon
$D(3X-2) = E(3X-2)^2 - [E(3X-2)]^2 = 118,6 - 10,3^2 = 12,51$.

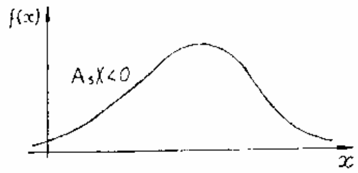
3.8. Asümmeetriategur ja ekstsess

Asümmeetriategur defineeritakse järgmiselt:

$$A_sX = \frac{E(X-EX)^3}{\sigma^3(X)}. \tag{1}$$

Kui $A_sX = 0$, siis on juhuslik suurus jaotatud sümmeetriliselt oma keskväärtuse ümber. Järgnevad jaotustiheduse graafikud iseloomustavad nullist erineva asümmeetriateguri tähendust:

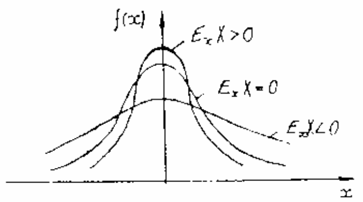




Ekstsess defineeritakse nii:

$$E_xX = \frac{E(X-EX)^4}{\sigma^4(X)} - 3. \tag{2}$$

Edaspidi vaadeldava normaaljaotuse puhul $E_xX = 0$. Kui $E_xX < 0$, siis jaotustiheduse graafik on lamedam kui normaaljaotusel, juhul kui $E_xX > 0$, on ta aga järsema ekstreemumiga:



4. Klassikalised jaotused

4.1. Binoomjaotus

Diskreetne juhuslik suurus X on binoomjaotusega, kui tema võimalike väärtuste hulgaks on täisarvud $0, 1, 2, \ldots, m, \ldots, n$, millele vastavad tõenäosused arvutatakse Bernoulli valemiga (vt punkti 2.11)

$$P(X = m) = P_{m,n} = C_n^m p^m q^{n-m}. \tag{1}$$

Binoomjaotust kirjeldab järgnev jaotusrida:

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

Saab tõestada, et binoomjaotuse keskväärtus

$$EX = np \tag{2}$$

ja dispersioon

$$DX = npq, \quad \text{kus} \quad q = 1 - p. \tag{3}$$

Märkus 1. Kui katsete arv n on suur, siis on Bernoulli valemi kasutamine raskendatud. Sel juhul võib Bernoulli valemi asendada ligikaudsete asümptootiliste valemitega, mida vaadeldakse hiljem.

Näide. Kindlustusagendil on üksikkliendiga lepingu sõlmimise tõenäosus 0,4. Agent kohtus viie kliendiga. Koostada sõlmitud lepingute arvu kui juhusliku suuruse jaotusrida. Leida vaadeldava juhusliku suuruse keskväärtus, dispersioon ja standardhälve.

Järgnevas jaotusreas olevad tõenäosused arvutatakse valemiga (1), kusjuures $n = 5$, $p = 0,4$ ja $q = 0,6$:

x_i	0	1	2	3	4	5
p_i	0,078	0,259	0,346	0,230	0,077	0,010

$EX = np = 5 \cdot 0,4 = 2.$
$DX = npq = 5 \cdot 0,4 \cdot 0,6 = 1,2.$
$\sigma(X) = \sqrt{DX} = \sqrt{1,2} = 1,095.$

Märkus 2. Punkti 2.11 näites 1 vaadeldud juhuslik suurus (antud hetkel sisselülitatud telekaamerate arv) allub binoomjaotusele. Viidatud näites olevas tabelis on selle juhusliku suuruse jaotusrida.

4.2. Poissoni jaotus

Kui Bernoulli valemis olevatest parameetritest n on suur ($n \geq 50$) ja p on väike ($p \leq 0,1$), siis võib Bernoulli valemi asemel kasutada ligikaudset **Poissoni valemit**

$$P(X = m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \tag{1}$$

kus $\lambda = EX = np$.

Valem (1) on seda täpsem, mida suurem on n ja mida väiksem on p.

On võimalik tõestada, et Poissoni jaotuse dispersioon $DX = \lambda$. Seega Poissoni jaotuse puhul

$$EX = DX = \lambda. \tag{2}$$

Näide 1. Ettevõtte telefonijaam teenindab 100 abonenti. Tõenäosus, et iga üksikabonent helistab ühe minuti jooksul, on 0,02. Kas ühe minuti jooksul helistab tõenäosemalt 3 või 4 abonenti?

Antud on $n = 100$, mis on küllalt suur, ning $p = 0,02 < 0,1$.

Arvutame $np = 100 \cdot 0,02 = 2$ ja $npq = 100 \cdot 0,02 \cdot 0,98 = 1,96$.

Võtame $\lambda = np = 2$ ning arvutame valemiga (1)

$$P(X = 3) = \frac{2^3}{3!} e^{-2} = \frac{8}{6 \cdot e^2} \approx 0,1823,$$

$$P(X = 4) = \frac{2^4}{4!} e^{-2} = \frac{16}{24 \cdot e^2} \approx 0,0902.$$

Tulemustest on näha, et kolme abonendi helistamise tõenäosus on suurem kui neljal. Kuna keskväärtus $EX = \lambda = 2$, siis kõige tõenäosemalt helistab 2 abonenti.

Kuna n ei ole väga suur, on küsitud tõenäosuste arvutamine võimalik ka Bernoulli valemiga:

$$P_{3,100} = C_{100}^3 \cdot 0,02^3 \cdot 0,98^{97} \approx 0,1823 \text{ ja } P_{4,100} = C_{100}^4 \cdot 0,02^4 \cdot 0,98^{96} \approx 0,0902.$$

Näide 2. Kaubamaja tegi kindlaks, et nõudmine teatud fotoaparaadi järele allub Poissoni jaotusele. Ühe nädala jooksul ostetakse keskmiselt kaks aparaati. Kui tellida neli aparaati nädalas, kui suur on siis tõenäosus, et nõudmine ületab pakkumise?

Kuna keskväärtus $EX = 2$, siis Poissoni jaotuse parameeter $\lambda = 2$. Vastuseks ülesandes esitatud küsimusele tuleb leida tõenäosus, et nädalas ostetakse üle nelja aparaadi:

$$P(X > 4) = 1 - \sum_{i=0}^{4} P(X = x_i) \approx 0,0527.$$

Jaotusreast

x_i	0	1	2	3	4	5	6	7	8	9
P	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120	0,0034	0,0009	0,0002

on näha, et x_i kasvades kahaneb $P(X = x_i)$ kiiresti. Kuna 9 aparaadi ostmise tõenäosus $P(X = 9) \approx 0,0002$ on juba väga väike, siis saadakse eelnevaga ligilähedaselt võrdne tulemus, kui leitakse tõenäosus, et ostetakse 5 kuni 9 aparaati:

$$\sum_{i=5}^{9} P(X = x_i) \approx 0,0526.$$

Näide 3. Tehases on 10 000 töötajat. Tõenäosus, et aasta jooksul keegi ei hukku tööõnnetusel, on 0,9. Kui suur on ühe töötaja jaoks tõenäosus, et ta ei hukku aasta jooksul tööõnnetusel?

Vastavalt Poissoni valemile

$$P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = 0,9.$$

Kuna $0! \equiv 1$, siis $e^{-\lambda} = 0,9$ ja $\lambda = -\ln 0,9 \approx 0,105$.
Tõenäosus, et üksiktöötaja hukkub, on

$$p = \frac{\lambda}{n} = \frac{-\ln 0,9}{10000}.$$

Üksiktöötaja ellujäämise tõenäosus on

$$q = 1 - p = 1 + \frac{\ln 0,9}{10000} = 0,99998946.$$

4.3. Normaaljaotus

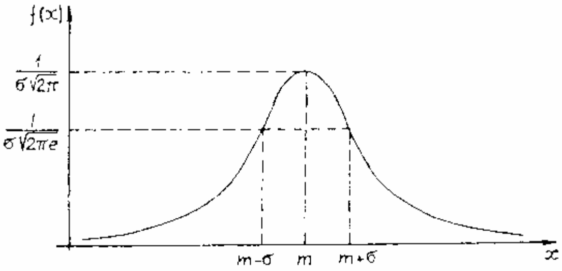
Juhuslik suurus allub normaaljaotusele (e Gaussi jaotusele), kui ta on mõjutatud paljude faktorite poolt, kusjuures iga üksikfaktori mõju juhuslikule suurusele on tühine ning domineerivad faktorid puuduvad.

Normaaljaotuse jaotustihedus defineeritakse järgmiselt ($e^x \equiv \exp[x]$):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{1}$$

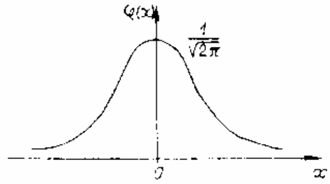
Avaldises (1) olevad parameetrid μ ja σ on vastavalt **keskväärtus** EX ja **standardhälve** $\sigma(X)$.

Normaaljaotuse jaotustiheduse graafiku põhimõtteline kuju on toodud järgneval joonisel (kohtadel $x = \mu - \sigma$ ja $x = \mu + \sigma$ on käänupunktid):



Kui $\mu = 0$ ja $\sigma = 1$, siis räägitakse *normeeritud normaaljaotusest*, mille jaotustiheduse analüütiline avaldis ning vastav graafik on järgmised:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{2}$$



Normaaljaotusega juhusliku suuruse antud lõiku sattumise tõenäosus $P(\alpha \leq X \leq \beta)$ (võrdub vahemikku $P(\alpha < X < \beta)$ sattumise tõenäosusega) leitakse vastavalt punkti 3.4 valemile (2) järgmise integraali abil:

$$P(\alpha \leq X \leq \beta) = \int_\alpha^\beta f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_\alpha^\beta \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx.$$

See integraal ei avaldu elementaarfunktsioonides. Tõenäosuse $P(\alpha \leq X \leq \beta)$ leidmiseks vajalike integraalide arvutamiseks on koostatud funktsiooni

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) dt \tag{3}$$

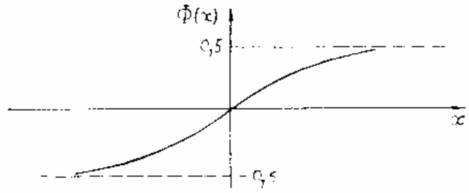
tabelid (vt lisa 1). Funktsiooni $\Phi(x)$ nimetatakse **Laplace'i funktsiooniks** ehk tõenäosuse integraaliks. On võimalik näidata, et selle funktsiooni kaudu

$$P(\alpha \leq X \leq \beta) = \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right). \tag{4}$$

Laplace'i funktsiooni omadusi

1. $\Phi(0) = 0$, sest kui $x = 0$, siis tulevad integraalis (3) võrdsed rajad.
2. $\Phi(\infty) = 0,5$, kusjuures koondumine toimub küllalt kiiresti, näiteks juba $\Phi(3) = 0,49865$.
3. $\Phi(x)$ on paaritu funktsioon, st $\Phi(-x) = -\Phi(x)$ ja tema graafik on sümmeetriline koordinaatide alguspunkti suhtes.

Laplace'i funktsiooni graafiku põhimõtteline kuju on toodud järgmisel joonisel:



Sageli pakub praktilist huvi normaaljaotusega juhusliku suuruse lõiku $[\mu - \varepsilon, \mu + \varepsilon]$ sattumise tõenäosus. Kasutades valemit (4), saab

$$P(|X - \mu| \le \varepsilon) = P(\mu - \varepsilon \le < X \le \mu + \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(\frac{-\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \qquad (5)$$

Valemi (5) alusel leiame

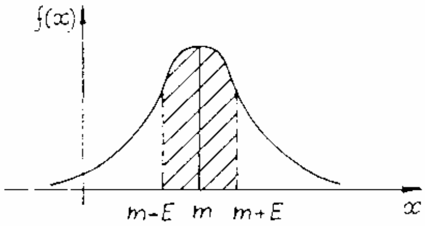
$$P(|X - \mu| < 3\sigma) = 2\Phi(3) \approx 0,9973. \qquad (6)$$

Niisiis *normaaljaotusega juhusliku suuruse hälve keskväärtuse suhtes ületab vaid 0,27% juhtudest* 3σ. Eelnev on nn **kolme sigma reegel**.

Märkus. Normaaljaotuse jaotustiheduse $f(x)$ graafiku laiuse iseloomustamiseks kasutatakse **keskmise vea** mõistet. Keskmine viga E on leitav seosest

$$P(|X - \mu| < E) = 0,5,$$

mis geomeetriliselt tähendab, et jaotustiheduse $f(x)$ graafiku alusest pindalast on 50% määratud vahemikuga $x \in]\mu - E, \mu + E[$. Nimetatud vahemikku jääv pindala on järgneval joonisel viirutatud.



Kuna $P(|X - \mu| < E) = 2\Phi(E/\sigma)$, siis $\Phi(E/\sigma) = 0,25$. Laplace'i funktsiooni $\Phi(x)$ tabelist (vt lisa 1) leiame, et $E/\sigma \approx 0,675$ ja

$$E \approx 0,675\sigma.$$

Näide 1. Tehase toodangu maht allub ligikaudselt normaaljaotusele keskväärtusega 134 786 eset nädalas ja standardhälbega 13 000 eset nädalas.

1. Leida tõenäosus, et nädala toodang ületab 150 000 eset.

2. Leida tõenäosus, et nädala toodang on väiksem kui 100 000 eset.

3. Oletame, et algasid töövaidlused ja toodang langes allapoole 80 000 eset. Administratsioon süüdistab ametiühinguid tootmistempo tahtlikus aeglustamises. Samal ajal väidavad ametiühingud, et vähenenud toodangu maht jäi siiski lubatavate hälvete piiridesse. Kas ametiühingute väited on usutavad?

$$\mu = 134\ 786, \quad \sigma = 13\ 000.$$

1.
$$P(150\ 000 < X < \infty) = \Phi(\infty) - \Phi\left(\frac{150\ 000 - 134\ 786}{13\ 000}\right) =$$
$$= 0,5 - \Phi(1,170) = 0,5 - 0,37900 = 0,121.$$

2.
$$P(0 < X < 100\ 000) = \Phi\left(\frac{100\ 000 - 134\ 786}{13\ 000}\right) - \Phi\left(\frac{0 - 134\ 786}{13\ 000}\right) =$$
$$= \Phi(-2,68) - \Phi(-10,37) = 0,5 - 0,49632 \approx 0,0037.$$

3.
$$P(0 < X < 80\ 000) = \Phi\left(\frac{100\ 000 - 134\ 786}{13\ 000}\right) + 0,5 =$$
$$= 0,5 - \Phi(4,2) = 0,5 - 0,499980 = 0,00002.$$

Kuna viimasena arvutatud tõenäosus on väga väike, siis ei ole ametiühingute väited usutavad. Täiendavalt võib leida hälbe absoluutväärtuse keskväärtuse suhtes

$$|80\ 000 - 134\ 786| = 54\ 786 \approx 4,2\sigma,$$

mis ületab 3σ. Seega ei ole kolme sigma reegli nõuded täidetud, mis samuti kinnitab ametiühingute väidete ebausutavust.

Näide 2. Lennuk lendab 20 km laiuses õhukoridoris. Lennutrajektoori säilitamisel tehtav süstemaatiline viga on 0,5 km, keskmine viga on 2 km. Missuguse tõenäosusega satub lennuk väljapoole õhukoridori?

Juhuslikuks suuruseks X on lennutrajektoori hälbed, mis alluvad normaaljaotusele. Süstemaatiline viga on hälbe keskväärtus $EX = 0,5$ km. Keskmine viga $E = 2$ km ning standardhälve

$$\sigma(X) = \sigma \approx \frac{E}{0,675} = \frac{2}{0,675} \approx 2,96.$$

Tõenäosus, et lennuk satub väljapoole õhukoridori, arvutatakse toetudes vastandtõenäosuse mõistele ning valemile (4):

$$1 - P(-10 \le X \le 10) = 1 - \Phi\left(\frac{10 - 0,5}{2,96}\right) + \Phi\left(\frac{-10 - 0,5}{2,96}\right) =$$
$$= 1 - \Phi(3,21) - \Phi(3,55) = 1 - 0,49933 - 0,49980 \approx 0,0009.$$

Näide 3. Toode kuulub kõrgemasse sorti, kui tema mõõtmete erinevus nominaalmõõtmetest ei ületa absoluutväärtuselt 3,45 mm. Toote mõõtmete juhuslikud erinevused nominaalmõõtmetest alluvad normaaljaotusele, kusjuures standardhälve on 3 mm ja süstemaatilised vead puuduvad. Leida keskmine kõrgemat sorti toodete arv, kui valmistati 40 toodet.

Tõenäosuse, et toode kuulub kõrgemasse sorti, arvutame valemiga (4), milles keskväärtus $\mu = 0$ ja standardhälve $\sigma = 3$:

$$P(-3,45 \leq X \leq 3,45) = 2\Phi\left(\frac{3,45}{3}\right) = 2\Phi(1,15) = 2 \cdot 0,37493 = 0,74986.$$

Eelnev tulemus ütleb, et kõrgemat sorti tooteid on ligikaudu 75%. Seega, keskmine kõrgemat sorti toodete arv 40 toote hulgas on $[40 \cdot 0,74986] = 30$.

Näide 4. Treitakse võlle nimiläbimõõduga 25 mm. Standardhälve on 0,01 mm. Missugustes piirides võib tõenäosusega 0,997 garanteerida võlli läbimõõdu?

Eeltoodud valemis (5) olev ε on selles näites läbimõõdu lubatud hälve keskväärtusest tõenäosusega 0,997. Seega $\Phi\left(\frac{\varepsilon}{\sigma}\right) = \frac{0,997}{2} = 0,4985$. Laplace'i funktsiooni tabelist saame $\varepsilon/\sigma = 2,97$ ning edasi arvutame $\varepsilon = 2,97 \cdot \sigma = 2,97 \cdot 0,01 \approx 0,03$ mm. Niisiis, tõenäosusega 0,997 on võlli läbimõõt piirides $25 \pm 0,03$ mm.

4.4. Binoomjaotuse aproksimeerimine normaaljaotusega

Koostame binoomjaotuse jaotusrea, kui $n = 20$ ja $p = q = 0,5$. Bernoulli valemis $P_{m,n} = C_n^m p^m q^{n-m}$ olevat suurust m võib tõlgendada näiteks kui poiste arvu 20 vastsündinu hulgas, eeldusel et poisi või tüdruku sünni tõenäosus on võrdne (tegelikkuses sünnib poisse veidi rohkem). Täpsusega kolm kohta pärast koma on tõenäosused $P_{0,20} = P_{20,20}$, $P_{1,20} = P_{19,20}$ ja $P_{2,20} = P_{18,20}$ võrdsed nulliga. Teatavasti binoomjaotuse keskväärtus $EX = np = 20 \cdot 0,5 = 10$ ja standardhälve $\sigma(X) = \sqrt{npq} = \sqrt{20 \cdot 0,5 \cdot 0,5} = \sqrt{5} \approx 2,24$.

Arvutame ühtlasi normaaljaotuse jaotustiheduse

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1}$$

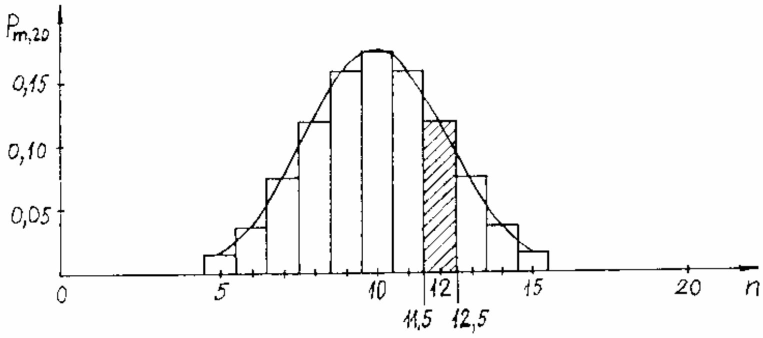
väärtused kohtadel $x = m = 0, 1, \ldots, 20$, võttes normaaljaotuse keskväärtuse ja standardhälbe võrdseks vaadeldava binoomjaotuse vastavate parameetritega: $\mu = 10$ ja $\sigma \approx 2,24$.

Arvutustulemused on järgmises tabelis.

m	0	1	2	3	4	5	6	7	8	9	10
$P_{m,20}$	0,000	0,000	0,000	0,001	0,005	0,015	0,037	0,074	0,120	0,160	0,176
$f(m)$	0,000	0,000	0,000	0,001	0,005	0,015	0,036	0,073	0,120	0,161	0,178

m	11	12	13	14	15	16	17	18	19	20
$P_{m,20}$	0,160	0,120	0,074	0,037	0,015	0,005	0,001	0,000	0,000	0,000
$f(m)$	0,161	0,120	0,073	0,036	0,015	0,005	0,001	0,000	0,000	0,000

Tähele võib panna, et analüüsitavas näites $P_{m,20} \approx f(m)$. Seda ligikaudset võrdust illustreerib binoomjaotusele vastav histogramm ja normaaljaotuse jaotustiheduse graafik:



Näeme, et normaaljaotuse jaotustiheduse graafik on histogrammile heaks lähendiks.

Vaadeldav näide võimaldab järeldada, et teatud tingimustel võib binoomjaotuse asemel kasutada ligikaudse lähendina (matemaatilise mudelina) normaaljaotust. Niisugune aproksimatsioon on seda täpsem, mida suurem on n. Kui $n \to \infty$, siis binoomjaotus koondub normaaljaotuseks.

Laplace'i lokaalne piirteoreem

Kogemused näitavad, et enamiku praktiliste ülesannete lahendamiseks piisav arvutustäpsus saavutatakse, kui

$$1)\ np > 5 \ \text{juhul}\ p < 0,5 \ \text{või}\ 2)\ nq > 5\ \text{juhul}\ p > 0,5. \tag{2}$$

Niisiis, tingimuse (2) täidetusel võib tõenäosuse $P_{m,n}$ arvutamiseks kasutada Bernoulli valemi asemel ligikaudse lähendina normaaljaotuse jaotustiheduse avaldist (1), milles $\mu = np$ ja $\sigma = \sqrt{npq}$:

$$P_{m,n} \approx \frac{1}{\sqrt{npq}\sqrt{2\pi}} \exp\left[-\frac{(m-np)^2}{2npq}\right]. \tag{3}$$

Valem (3) kehtib täpselt vaid siis, kui $n \to \infty$. Seetõttu öeldakse, et valem (3) on üks tõenäosusteooria piirteoreemidest, täpsemalt **Laplace'i lokaalne piirteoreem** (selles mõttes lokaalne, et tõenäosus $P_{m,n}$ arvutatakse ühes fikseeritud punktis m).

Valemile (3) antakse harilikult alljärgnev kuju:

$$P_{m,n} \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \ \text{kus}\ x = \frac{m-np}{\sqrt{npq}}. \tag{4}$$

Näide 1. Tõenäosus, et lattu saabunud kastides esineb riknenud puuvilja, on 0,1. Leida tõenäosus, et 350-st kastist 25-s on riknenud puuvilja.

Antud on $p = 0,1$, $n = 350$ ja $m = 25$. Arvutame $q = 0,9$.

Kuna $q > 0,5$ ja $np = 350 \cdot 0,1 = 35 > 5$, siis võib binoomjaotuse asemel kasutada ligikaudse lähendina normaaljaotust, mille parameetrid on $\mu = 350 \cdot 0,1 = 35$ ja $\sigma = \sqrt{350 \cdot 0,1 \cdot 0,9} \approx 5,61$. Valemiga (4) arvutame kõigepealt $x \approx -1,78$ ning seejärel küsitud tõenäosuse $P_{25,350} \approx 0,015$.

Märkus. Antud näites on lisaks sellele, et n on suur, ka p väike. Seega võib siin kasutada ka Poissoni valemit (vt punkti 4.2 valemit (1)):

$$P(X = m) \approx \frac{\lambda^m}{m!}e^{-\lambda} = \frac{35^{25}}{25!}e^{-35} = 0,016, \quad kus \quad \lambda = np = 350 \cdot 0,1 = 35.$$

Laplace'i integraalne piirteoreem

Laplace'i lokaalse piirteoreemi (4) põhjendamiseks kasutatud mõttekäik võimaldab eeldusel (2) kasutada binoomjaotuse asemel normaaljaotust tõenäosuse $P(\alpha \leq X \leq \beta)$ arvutamiseks. Võtame punkti 4.3 valemis (4)

$$P(\alpha \leq X \leq \beta) = \Phi\left(\frac{\beta - \mu}{\sigma}\right) - \Phi\left(\frac{\alpha - \mu}{\sigma}\right)$$

$\mu = np$ ja $\sigma = \sqrt{npq}$ ning saame **Laplace'i integraalse piirteoreemi**

$$P(\alpha \leq X \leq \beta) \approx \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right). \tag{5}$$

Integraalsus tähendab siin seda, et valemiga (5) arvutatakse tõenäosust üle kogu lõigu $[\alpha, \beta]$ (mitte lokaalselt fikseeritud punktis m nagu valemis (4)).

Korrektsioonifaktor

Diskreetse binoomjaotuse modelleerimisel **pideva normaaljaotusega** kasutatakse tõenäosuse $P(\alpha \leq X \leq \beta)$ arvutamisel valemiga (5) **korrektsioonifaktorit**, mis tagab tulemuse suurema täpsuse. Analüüsime täiendavalt käesoleva punkti 4.4 sissejuhatavat näidet (vt ka lk 53 oleval joonisel olevat viirutatud piirkonda) ning arvutame valemiga (5) tõenäosuse

$$P(11,5 \leq X \leq 12,5) \approx \Phi\left(\frac{12,5 - 10}{2,24}\right) - \Phi\left(\frac{11,5 - 10}{2,24}\right) = 0,120.$$

Paneme tähele, et nii Bernoulli valemiga leitud teoreetiliselt täpne tõenäosus $P_{12,20}$ kui ka Laplace'i lokaalse piirteoreemiga (4) ligikaudselt arvutatud sama tõenäosus $P_{12,20}$ (need kaks tulemust on lk 53 oleva tabeli selles veerus, kus $m = 12$) ning viimati Laplace'i integraalse piirteoreemiga (5) saadud ligikaudne resultaat $P(11,5 \leq X \leq 12,5)$ langevad kõik kolme kohaga pärast koma kokku ja on 0,120. Niisugune kokkulangevus on selgitatav asjaoluga, et lk 53 joonisel oleva viirutatud ristküliku pindala, mis arvutatakse Bernoulli valemiga, on ligikaudselt võrdne normaaljaotuse jaotustiheduse graafiku aluse pindalaga lõigus [11,5; 12,5], mis leitakse Laplace'i integraalse piirteoreemiga (5). Seega võime tõenäosuse $P_{m,n}$ ligikaudseks arvutamiseks kasutada Laplace'i lokaalse piirteoreemi (4) asemel integraalset piirteoreemi (5) ning leida sellele vastavalt tõenäosuse $P(m - 0,5 \leq X \leq m + 0,5) \approx P_{m,n}$. Sellega ongi seotud **korrektsioonifaktori** olemus. Märgime, et kui tingimuse (2) täidetusel on vaja arvutada tõenäosus $P(m \leq X \leq n)$, siis kasutatakse korrektsioonifaktorit kujul $P(m - 0,5 \leq X < \infty)$, sest normaaljaotuse kõver ei lõpe kohal n, vaid läheb lõpmatusse (tõsi, n asendamine lõpmatusega arvutustulemusi praktiliselt enamasti ei mõjuta, nagu võime veenduda järgnevate näidete põhjal). Tõenäosuse $P(0 \le X \le m)$ arvutamist Laplace'i integraalse piirteoreemiga korrigeeritakse kujul $P(-\infty < X \le m + 0,5)$.

Näide 2. 95% korterivarguste puhul ei saa omanikud oma varandusest midagi tagasi. Leida tõenäosus, et vaadeldud perioodi jooksul toimunud 200 korterivarguse juures sai oma vara tagasi vähemalt: 1) 15 omanikku; 2) 5 omanikku; 3) 10 omanikku.

Antud on $n = 200$ ja $q = 0,95$. Arvutame $p = 0,05$.

Kuna $p < 0,5$ ja $np = 200 \cdot 0,05 = 10 > 5$, siis võib binoomjaotuse asemel kasutada ligikaudse lähendina normaaljaotust, mille parameetrid on $\mu = 200 \cdot 0,05 = 10$ ja $\sigma =$ $= \sqrt{200 \cdot 0,05 \cdot 0,95} \approx 3,08$. Korrektsioonifaktorit kasutades leiame otsitavad tõenäosused

$$1)\quad P(14,5 \le X < \infty) = \Phi(\infty) - \Phi\left(\frac{14,5-10}{3,08}\right) = 0,5 - \Phi(1,46) = 0,5 - 0,42786 \approx 0,072;$$

$$2)\quad P(4,5 \le X < \infty) = \Phi(\infty) - \Phi\left(\frac{4,5-10}{3,08}\right) = 0,5 - \Phi(-1,78) = 0,5 + 0,46246 \approx 0,962;$$

$$3)\quad P(9,5 \le X < \infty) = \Phi(\infty) - \Phi\left(\frac{9,5-10}{3,08}\right) = 0,5 - \Phi(-0,16) = 0,5 + 0,06356 \approx 0,564.$$

Kuna keskväärtus on 10 (keskmiselt saab oma vara tagasi 10 inimest), siis on loogiline, et keskmisest suuremate ootuste tõenäosus (et vähemalt 15 saab vara tagasi) on väike. Keskmisest väiksemate ootuste tõenäosus (et vähemalt 5 saab vara tagasi) on aga küllalt suur.

Näide 3. Tõenäosus saada välgust tabatud on üksikisiku jaoks 1/600 000. Leida tõenäosus, et 6 miljoni juhuslikult valitud inimese hulgast tabab välk vähem kui kuut inimest.

Antud on $n = 6\ 000\ 000$ ja $p = 1/600\ 000$. Arvutame $q = 599\ 999/600\ 000$.

Kuna $p < 0,5$ ja $np = 10 > 5$, siis võib binoomjaotust modelleerida normaaljaotusega. Korrektsioonifaktorit kasutades saame

$$P(-\infty < X \le 5,5) = \Phi(\infty) - \Phi\left(\frac{5,5-10}{\sqrt{npq}}\right) = 0,5 - \Phi(1,42) = 0,5 - 0,42220 = 0,0,0778.$$

4.5. Ülesanded

4.1. Münti visatakse 5 korda. Koostada vappide arvu kui juhusliku suuruse jaotusrida. Leida vaadeldava juhusliku suuruse keskväärtus, dispersioon ja standardhälve.

4.2. Laskur tulistab märklauda 3 korda. Tõenäosus tabada märki igal lasul on 0,4. Koostada tabamuste arvu kui juhusliku suuruse jaotusrida. Leida vaadeldava juhusliku suuruse keskväärtus, mood, dispersioon ja standardhälve.

Matemaatiliselt ekvivalentse lahenduskäiguga on järgmine ülesanne.

Teatud ajalehte loeb 40% elanikkonnast. Küsitleti kolme juhuslikult valitud inimest. Koostada seda ajalehte lugenud inimeste arvu kui juhusliku suuruse jaotusrida.

4.3. Projektijuht määras kindlaks, et allettevõtja täidab 20% tellimustest hilinemisega. Projektijuhil oli allettevõtjale 6 tellimust. Koostada õigeaegselt täitmata tellimuste arvu kui juhusliku suuruse jaotusrida. Leida vaadeldava juhusliku suuruse keskväärtus, mood, dispersioon ja standardhälve.

4.4. Vaatluse all on 4 raadiot, mille igaühe korrasoleku tõenäosus garantiiaja jooksul on 0,8. Leida tõenäosus, et garantiiaja jooksul peab vastu 3 raadiot. Missugune on tõenäoseim garantiiaja jooksul vastupidavate raadiote arv? Koostada garantiiajal vastupidavate raadiote arvu kui juhusliku suuruse jaotusrida. Leida vaadeldava juhusliku suuruse keskväärtus, mood, dispersioon ja standardhälve.

4.5. Tuhandeleheküljelises raamatus on 100 viga. Kui suur on tõenäosus, et juhuslikult valitud leheküljel on vähemalt 2 trükiviga?

4.6. Tehas saadab baasi 600 kõrgekvaliteedilist toodet. Tõenäosus, et toode läheb teel rikki, on 0,002. Kui suur on tõenäosus, et baasi jõudnud toodete hulgas ei ole rikkiläinuid rohkem kui 2?

4.7. Arvuti töös tekivad aeg-ajalt rikked. Rikete arv teatud ajaühikus kui juhuslik suurus allub Poissoni jaotusele. Keskmine rikete arv ööpäevas on 1,5. Leida tõenäosus, et

1) ühe ööpäeva jooksul esineb vähemalt kaks riket;

2) kuue ööpäeva jooksul esineb vähemalt kaks riket.

4.8. 1998. aastal hukkus Eestis liiklusavariide tagajärjel 278 inimest. Leida tõenäosus, et 1) kahe ööpäeva jooksul ei hukku kedagi;

2) ühe ööpäeva jooksul hukkub vähemalt üks inimene.

4.9. Kulla hinna kõikumist lühikesel perioodil modelleerib ligikaudselt normaaljaotus. 1986. aasta septembri keskel oli kulla hinna keskväärtus 409 USA dollarit unts ja standardhälve 12 dollarit unts. Maakleril on kliendi korraldus müüa kuld järgmisel päeval, kui selle hind on piirides 420 kuni 425 dollarit. Kui suur on tõenäosus, et kuld müüakse?

4.10. Mikroprotsessorite valmistamiseks kasutatavas pooljuhis olevate lisandite kontsentratsioon allub normaaljaotusele keskväärtusega 127 miljondikku osa ja standardhälbega 22 miljondikku osa. Pooljuht sobib kasutamiseks vaid siis, kui temas on lisandeid vähem kui 150 miljondikku osa. Mitu protsenti pooljuhtidest on kasutamiskõlblikud?

4.11. Uuriti kinnipeetavate sotsiaalset vastutustunnet vangistuse ajal ning perspektiive nende rehabilitatsiooniks pärast vabastamist. Sotsiaalset vastutustunnet hinnati igale vangile antud testiga, mille eest saadud punktid on normaalselt jaotatud keskväärtusega 100 ja standardhälbega 20. Psühholoogid hindasid iga vangi perspektiivi rehabilitatsiooniks. Need hinnangud on samuti normaalselt jaotatud keskväärtusega 500 ja standardhälbega 100. Konkreetse vangi sotsiaalse vastutustunde testi tulemus oli 146 punkti ja rehabilitatsiooni hinnang oli 335 punkti. Kuidas iseloomustada seda vangi teiste vangide üldisel taustal?

4.12. Kõrgusmõõturi kasutamisel tekivad juhuslikud ja süstemaatilised vead. Süstemaatiline viga on +20 meetrit, juhuslikud vead alluvad normaaljaotusele keskmise veaga 40 meetrit. Kui suur on tõenäosus, et mõõdetud kõrguse väärtus ei erine tegelikust rohkem kui 100 meetrit?

4.13. Normaaljaotusega juhusliku suuruse keskväärtus on 40 ja dispersioon 200. Kui suur on tõenäosus, et

1) juhusliku suuruse mõõdetav väärtus oleks lõigus [30; 80]?

2) juhusliku suuruse kolmest sõltumatust väärtusest ükski ei oleks lõigus [30; 80]?

4.14. Teatud automudeli läbisõit allub normaaljaotusele keskväärtusega 160 000 *km* ja standardhälbega 30 000 *km*. Kui suur on tõenäosus, et ostetud auto läbisõit on piirides 100 000 *km* kuni 180 000 *km*?

4.15. Kahe objekti vahelise kauguse mõõtmisel tekkiv mõõtmisviga allub normaaljaotusele. Keskväärtus on 5 meetrit ja standardhälve 10 meetrit. Leida tõenäosus, et mõõdetud kauguse väärtus erineb tõelisest väärtusest mitte rohkem kui 15 meetrit.

4.16. Lattu toodud tolmuimeja korrasoleku tõenäosus on 0,96. Leida tõenäosus, et kahesajast laos olevast tolmuimejast on korras 190.

4.17. Tehases on kõrgema sordi toodangu osatähtsus 90%. Leida tõenäosus, et juhuslikult valitud 100-st tootest on 96 kõrgemat sorti.

4.18. Praakdetaili tootmise tõenäosus on 0,005. Kui suur on tõenäosus, et 10000 detaili hulgas on 60 praakdetaili?

4.19. Märki tabamise tõenäosus ühel lasul on 0,8. Leida tõenäosus, et sajast lasust tabatakse märki 75 korral.

4.20. Vaatlused näitavad, et 60% üliõpilastest sooritab eksami esimesel katsel. Kasutades binoomjaotuse asemel ligikaudse lähendina normaaljaotust, leida tõenäosus, et teaduskonna 200-st üliõpilasest sooritab vähemalt 65% eksami esimesel katsel.

4.21. Kui suur on tõenäosus, et sajast istutatud puust läheb kasvama 65 kuni 75, kui ühe puu kasvama minemise tõenäosus on 0,7?

4.22. Kui tõenäone on, et münti 100 korda visates tuleb vapp esile 40 kuni 60 korda?

4.23. Vaatlused on näidanud, et 30% kaupluseкülastajatest sooritab ostu. Päevas külastab kauplust keskmiselt 110 inimest. Kui suur on tõenäosus, et konkreetsel päeval on ostjaid 1) vähemalt 30, 2) alla 40?

4.24. Automaatpink annab 90% esimese sordi toodangut. Kui suur on tõenäosus, et 300 detaili hulgas olevate esimese sordi detailide arv ei erine nende tõenäoseimast arvust mitte rohkem kui viie toote võrra?

4.25. Valmistoodangu hulgast juhuslikult valitud aparaat vajab täiendavat reguleerimist tõenäosusega 0,05. Kui toodangu valikulisel kontrollil leitakse, et täiendavat reguleerimist nõuab üle 6% kontrolliks võetud aparaatidest, siis tagastatakse kogu partii. Leida kogu partii tagastamise tõenäosus, kui sellest on kontrolliks võetud 500 aparaati.

4.26. Restorani administraator teab kogemuste põhjal, et laua ettetellinutest tuleb tegelikult kohale 70%. Ühel õhtul võttis administraator vastu 20 tellimust olemasolevale 15 lauale. Kui suur on tõenäosus, et peoseltskondi tuleb kohale rohkem, kui jätkub laudu?

4.27. Kvaliteedi kontrolliga on tuvastatud, et 96% transistoritest peab vastu vähemalt garantiiaja jooksul. Juhuslikult valitakse 15 000 transistori. Leida tõenäosus, et garantiiajal rikki minevate transistoride arv on: 1) 570 kuni 630; 2) vähem kui 615.

5. Jaotuse parameetrite hindamine

5.1. Empiiriline jaotusfunktsioon

Käesolevas peatükis kasutatakse punktides 1.2 ja 1.5 defineeritud statistika põhimõisted **üldkogum, valim, statistiline rida, variatsioonrida ja histogramm.**

Üldkogumi **empiiriliseks jaotusfunktsiooniks** nimetatakse funktsiooni

$$F^*(x) = \sum_{x_i < x} \frac{n_i}{n}. \tag{1}$$

Empiirilise jaotusfunktsiooni graafikuks on treppjoon, kusjuures funktsiooni väärtused kasvavad hüpetena üleminekul rühmitatud andmete järgmisele klassile. Sissejuhatava peatüki punkti 1.5 näites 9 oli kirjeldavas stiilis vaadeldud empiirilist jaotusfunktsiooni ja tema graafikut. Valimi mahu n piiramatul kasvamisel koondub empiiriline jaotusfunktsioon $F^*(x)$ üldkogumi jaotusfunktsiooniks $F(x)$. Koondumine toimub siin spetsiifilisel viisil tõenäosuse järgi. See tähendab, et alates mingist küllalt suurest valimi mahust n muutub $F^*(x)$ ja $F(x)$ vaheline oluline erinevus vähe tõenäoseks.

Empiirilisel jaotusfunktsioonil on samad omadused, kui üldkogumi jaotusfunktsioonilgi:

1) $0 \leq F^*(x) \leq 1$;
2) $F^*(x)$ on mittekahanev;
3) $F^*(x) = 0$, kui $x \leq x_{min}$, $\quad F^*(x) = 1$, kui $x \geq x_{max}$,

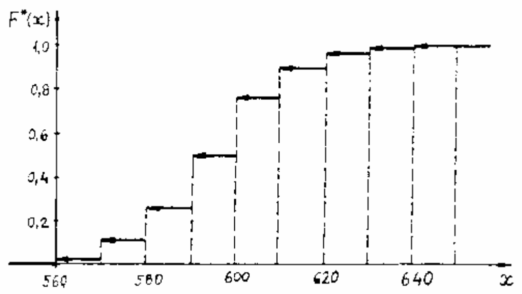
kus x_{min} on tunnuse minimaalne ja x_{max} maksimaalne väärtus.

Näide. Raadiolokaatoriga tehti 300 mõõtmist, selleks et kindlaks teha vea jaotusseadust. Mõõtmistulemused olid diapasoonis 560 kuni 650 m. See diapasoon jagatakse 10 m pikkusteks intervallideks ja saadakse variatsioonrida, mis on toodud järgneva tabeli esimeses, teises ja neljandas veerus. Joonestada empiirilise jaotusfunktsiooni graafik ning histogramm (vt võrdluseks punkti 1.5 näidet 9, kus on lahendatud analoogiline ülesanne).

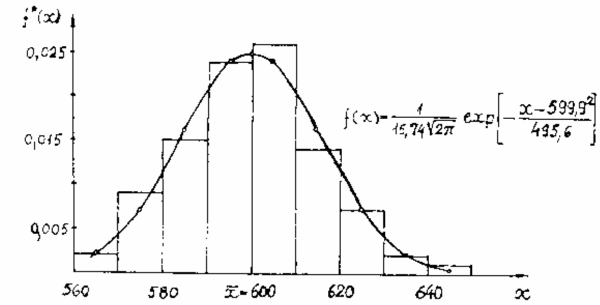
Valimi maht $n = 300$, klassi ulatus $\Delta x = 10$ m. Järgnevas tabelis on x_k intervalli keskmine, n_i antud intervalli sattunud mõõtmiste arv, $p_i^* = n_i/n$ sageduse tihedus ja $f^*(x_k) = p_i^*/\Delta x$ sageduse suhteline tihedus. Tabeli viimases veerus on toodud jaotustiheduse $f(x_k)$ väärtused, mille arvutamist selgitatakse punktis 5.2. Vastavad punktid on kantud histogrammiga ühisele joonisele ning ühendatud pideva joonega.

Intervalli nr	Intervallid	x_k	n_i	$p_i^* = n_i/n$	$F^*(x) = \sum_{x_i<x} n_i/n$	$f^*(x_k) = p_i^*/\Delta x$	$f(x_k)$
1	[560,570[	565	6	0,02	0,02	0,002	0,002
2	[570,580[	575	27	0,09	0,11	0,009	0,007
3	[580,590[	585	45	0,15	0,26	0,015	0,016
4	[590,600[	595	72	0,24	0,50	0,024	0,024
5	[600,610[	605	78	0,26	0,76	0,026	0,024
6	[610,620[	615	42	0,14	0,90	0,014	0,016
7	[620,630[	625	21	0,07	0,97	0,007	0,007
8	[630,640[	635	6	0,02	0,99	0,002	0,002
9	[640,650]	645	3	0,01	1,00	0,001	0,0004

Empiirilise jaotusfunktsiooni $F^*(x)$ graafik:



Histogramm:



5.2. Punktihinnangud ja nendele esitatavad nõuded. Keskväärtuse, dispersiooni ja standardhälbe punktihinnangud

Statistilise analüüsi üheks oluliseks eesmärgiks on teha valimi põhjal järeldusi üldkogumi kohta. Vaatleme siin juhusliku suuruse arvkarakteristikute (keskväärtuse, dispersiooni ja standardhälbe) hindamist valimi alusel. Kuna iga parameetri hindamise tulemusena saadakse üks arv, siis räägitakse arvkarakteristikute **punktihinnangutest**. Olgu juhusliku suuruse X jaotuse tüüp teada, kuid sisaldagu see jaotus tundmatut parameetrit a (milleks on edaspidises käsitluses konkreetselt kas keskväärtus, dispersioon või standardhälve). Tehakse n sõltumatut katset, milles juhuslik suurus X omandab väärtused $x_1,\ x_2, \ldots, x_n$. Ülesandeks on leida parameetri a hinnang

$$\bar{a} = \bar{a}(x_1,\ x_2, \ldots, x_n).$$

Praktilise väärtusega on järgmist kolme nõuet täitvad hinnangud.

1. Üldiselt hinnang $\bar{a}$ erineb parameetri tegelikust väärtusest a. Võimaluse korral püütakse vältida **süstemaatilist viga** parameetri a üle- või alahindamiseks. Süstemaatiline viga puudub, ehk nagu öeldakse, **hinnang on nihutamata**, kui hinnangu keskväärtus võrdub hinnatava parameetri endaga:

$$E\bar{a}^{(n)} = a.$$

2. Hinnang $\tilde{a}$ peab koonduma (tõenäosuse järgi) parameetriks a. See tähendab, et alates mingist küllalt suurest valimi mahust n muutub $\tilde{a}$ ja a vaheline oluline erinevus vähetõenäoseks. Öeldakse, et hinnang peab olema **konsistentne e mõjus**.

3. Üldiselt võib sama mahuga valimi alusel määrata mitu erinevat nihutamata hinnangut. Nende hulgast tuleb eelistada võimalikult väikese dispersiooniga hinnangut, mille puhul on jämeda vea tekkimise tõenäosus väiksem. Parameetri a nihutamata hinnangut $\tilde{a}^{(n)}$ nimetatakse **efektiivseks**, kui tema dispersioon $D\tilde{a}^{(n)}$ on minimaalne kõigi võimalike nihutamata hinnangute hulgast, mis on saadud sama mahuga valimi põhjal. Efektiivse hinnangu kasutamine võimaldab saavutada vajaliku täpsuse minimaalse vaatluste arvuga.

Üldkogumi **keskväärtuse hinnanguks** *on valimi aritmeetiline keskmine:*

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{1}$$

Kuna $$E\overline{x} = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n} EX = EX,$$

siis hinnang (1) on *nihutamata*. On võimalik näidata, et aritmeetiline keskmine on ka konsistentne (mõjus) ja efektiivne hinnang. Seega on aritmeetiline keskmine parimaks võimalikuks keskväärtuse hinnanguks.

Kui valimi dispersiooni hindamiseks kasutada kõikse uuringu andmetel põhinevat üldkogumi dispersiooni arvutuseeskirja (vt punkti 3.7 valemit (1) ja omadust 4), asendades selles keskväärtuse tähise EX üldkogumi andmete alusel arvutatud aritmeetilise keskmise tähisega $\overline{x}$

$$D^*X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x}\cdot\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}(\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2, \tag{2}$$

siis selgub, et **hinnang (2) on nihutatud**:

$$E(D^*X) = \frac{n-1}{n}\cdot DX.$$

Dispersiooni nihutamata hinnanguks on

$$s^2 = \frac{n}{n-1}D^*X = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (x)^2\right). \tag{3}$$

Standardhälbe hinnanguks võetakse

$$s = \sqrt{s^2}, \tag{4}$$

mis on praktilise kasutamise vajadusi rahuldav väga väikese nihkega hinnang (rakenduslikus kirjanduses, nagu siingi edaspidi, on s nimetatud nihutamata hinnanguks).

Märkus 1. Kõikse uuringu puhul **arvutatakse dispersioon** valemiga (2), valikuuringu korral aga **hinnatakse dispersiooni** valemiga (3).

Märkus 2. Paljud taskukalkulaatorid sisaldavad aritmeetilise keskmise ja standardhälbe arvutamise programmi, kusjuures s ja $\sqrt{D^*X}$ tähisteks on klaviatuuril vastavalt s ja σ (või ka σ_{n-1} ja σ_n).

Näide 1. Ettevõtte töölistest valiti juhuslikult viis ja määrati, et nad vajavad ühe detaili valmistamiseks aega vastavalt 21, 26, 28, 31 ja 35 minutit. Leida keskmine aeg, mis kulub ühel töölisel ühe detaili valmistamiseks. Leida dispersiooni ja standardhälbe nihutamata hinnangud.

Keskmine aeg, mis kulub ühel töölisel ühe detaili valmistamiseks, on aritmeetiline keskmine

$$\overline{x} = \frac{1}{5}(21 + 26 + 28 + 31 + 35) = 28,2 \text{ min}$$

Arvutame järgnevalt

$$\frac{1}{5}\sum_{i=1}^{n} x_i^2 = \frac{1}{5}(21^2 + 26^2 + 28^2 + 31^2 + 35^2) = 817,4$$

ning leiame siis dispersiooni nihutamata hinnangu vastavalt valemile (3)

$$s^2 = \frac{5}{5-1}\left(817,4 - 28,2^2\right) = 27,7 \text{ min}^2$$

Valemiga (4) saame standardhälbe nihutamata hinnangu

$$s = \sqrt{27,7} \approx 5,26 \text{ min}$$

Näide 2. Hinnata punkti 5.1 näites toodud jaotuse keskväärtust ja standardhälvet.

Klassidesse rühmitatud andmete puhul lähtutakse aritmeetilise keskmise ja dispersiooni nihutamata hinnangu arvutamisel intervalli keskmistest x_k (vt punkti 1.3 valemit (2) ja punkti 1.5 näidet 9):

$$\overline{x} = \frac{1}{300}(6 \cdot 565 + 27 \cdot 575 + \ldots + 3 \cdot 645) \approx 599,9 \text{ m},$$

$$\frac{1}{300}\sum_{i=1}^{300} x_i^2 = \frac{1}{300}(6 \cdot 565^2 + 27 \cdot 575^2 + \ldots + 3 \cdot 645^2) = 360\ 127$$

$$\text{ja} \quad s^2 = \frac{300}{300-1}(360\ 127 - 599,9^2) \approx 247,8\ m^2 \quad \text{ning} \quad s \approx 15,74\ m.$$

Märkus. s^2 arvutamisel ei tohi ümardada $\overline{x} = 599,9 \approx 600$, sest suurte arvude 360 127 ja $599,9^2 = 359\ 880,01$ või $600^2 = 360\ 000$ väike vahe erineb tunduvalt, olles vastavalt 246,99 või 127 ning põhjustades tulemuses väga suure vea.

Kuna pideva juhusliku suuruse jaotustihedust võib graafiliselt ligikaudselt kirjeldada histogrammiga (vt lk 14), siis võimaldab punktis 5.1 toodud histogrammi kuju, mis meenutab Gaussi kõverat (vt punkti 4.3 esimest joonist), oletada katseandmete alluvust normaaljaotusele. Pannes normaaljaotuse jaotustiheduse avaldisse (punkti 4.3 valem (1)) leitud keskväärtuse ja standardhälbe hinnangud, saadakse

$$f(x) = \frac{1}{15,74\sqrt{2\pi}} \exp\left[-\frac{(x-599,9)^2}{495,6}\right].$$

Iga intervalli keskel arvutatud $f(x_k)$ väärtused on toodud alajaotuses 5.1 oleva tabeli viimases veerus ning nendele väärtustele vastavad punktid on ühtlasi kantud sama alajaotuse juures olevale joonisele. Nende punktide ühendamise teel saadud jaotustiheduse graafik ja histogramm on heas kooskõlas, mis viitab katseandmete küllalt tõenäosele allumisele normaaljaotusele.

5.3. Vahemikhinnangud

Üldkogumi jaotusparameetrite a (konkreetselt EX, DX, $\sigma(X)$) hinnangud $\tilde{a} = \tilde{a}(x_1, x_2, \ldots, x_n)$ (konkreetselt $\bar{x}$, s^2, s), mis on saadud valimi mahuga n vaatlemise teel, on nn **punktihinnangud**, sest need antakse ühe arvuga. Kui üldkogumist moodustada mitu juhuslikku valimit mahuga n ja arvutada neist igaühe põhjal punktihinnang $\tilde{a}$, siis saadakse iga kord erinev tulemus. Järelikult on punktihinnang juhuslik suurus $\tilde{A}$, mis ei anna informatsiooni hinnangu täpsuse ja usaldatavuse kohta. Kui valimi maht on väike, siis võib punktihinnang oluliselt erineda hinnatava parameetri tegelikust väärtusest. Sellisel juhul on korrektsem kasutada **vahemikhinnanguid**, mis võimaldavad määrata punktihinnangute täpsuse teatava usaldatavusega. Olgu $\varepsilon > 0$ ja olgu $|\tilde{a} - a| < \varepsilon$ (ehk $\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon$). Siis iseloomustab ε hinnangu täpsust. Probleem on aga selles, et statistilised meetodid ei võimalda leida niisugust piirkonda $]\tilde{a} - \varepsilon, \tilde{a} + \varepsilon[$, mille puhul saaks kategooriliselt väita, et parameetri tegelik väärtus a asub selles piirkonnas, sest punktihinnang $\tilde{a}$ on *juhuslik* suurus $\tilde{A}$, mis muutub ühelt valimilt teisele üle minnes. Seega on usalduspiirkonnal juhuslikud rajad $]\tilde{A} - \varepsilon, \tilde{A} + \varepsilon[$, mis konkreetse valimi puhul omavad konkreetseid väärtusi $]\tilde{a} - \varepsilon, \tilde{a} + \varepsilon[$.

Usalduspiirkonda määrava ε leidmiseks antakse ette ühele lähedane tõenäosus β (näiteks 0,95 või 0,99), mida nimetatakse **usaldusnivooks**, ning nõutakse, et usalduspiirkond $]\tilde{a} - \varepsilon, \tilde{a} + \varepsilon[$ kataks hinnatava parameetri a tõenäosusega β:

$$P(|\tilde{A} - a| < \varepsilon) = \beta \quad \text{ehk} \quad P(\tilde{A} - \varepsilon < a < \tilde{A} + \varepsilon) = \beta. \tag{1}$$

Punktihinnangu $\tilde{a}$ usalduspiirkonna (vahemikhinnangu) $]\tilde{a} - \varepsilon, \tilde{a} + \varepsilon[$ leidmiseks tuleb:
1) arvutada valimi põhjal punktihinnang $\tilde{a}$;
2) ette anda usaldusnivoo β (näiteks 0,95 või 0,99);
3) leida seosest (1) ε;
4) koostada usalduspiirkond $]\tilde{a} - \varepsilon, \tilde{a} + \varepsilon[$.

Märkus. Eelnevas on defineeritud *sümmeetriline* usalduspiirkond, mida praktikas enamasti kasutatakse (sest enamasti kasutatakse parameetri a *nihutamata* hinnangut $\tilde{a}$, mis jaotub oma keskväärtuse $E\tilde{a} = a$ suhtes sümmeetriliselt).

5.4. Normaaljaotuse keskväärtuse usalduspiirkond, kui valimi maht on suur või kui standardhälve on teada

Järgnevat normaaljaotuse keskväärtuse usalduspiirkonna leidmise metoodikat kasutatakse kas juhul, kui valimi maht on suhteliselt suur ($n > 30$), või kui mingi konkreetse ülesande puhul on standardhälve varasemate kogemuste põhjal teada.

Keskväärtuse usalduspiirkonna leidmiseks tehakse n sõltumatut katset, milles juhuslik suurus X omandab väärtused $x_1, x_2, \ldots, x_n$. Selle valimi põhjal leitakse keskväärtuse μ punktihinnang $\bar{x}$ (aritmeetiline keskmine). Kui teha mitu katseseeriat mahuga n, saadakse neist igaühe põhjal erinev aritmeetiline keskmine. Seega võib keskväärtuse μ hinnangut $\bar{x}$ vaadelda kui juhuslikku suurust $\overline{X}$, mis muutub ühelt valimilt teisele üle minnes. Nõuame, et

$$P(|\overline{X} - \mu| < \varepsilon) = \beta, \tag{1}$$

kus β on etteantud usaldusnivoo. Leida tuleb ε, mis määrab usalduspiirkonna $]\bar{x} - \varepsilon, x + \varepsilon[$.

On võimalik näidata, et juhul kui X on normaaljaotusega, siis ka $\overline{X}$ on normaaljaotusega. Normaaljaotusega juhusliku suuruse X vahemikku $]\mu - \varepsilon, \mu + \varepsilon[$ sattumise tõenäosus avaldub Laplace'i funktsiooni kaudu (vt punkti 4.3 valemit (5)) järgmiselt:

$$P(|X - \mu| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \tag{2}$$

Seoste (1) ja (2) vasakud pooled erinevad vaid juhuslike suuruste $\overline{X}$ ja X poolest. Kui valemis (2) asendada X suurusega $\overline{X}$, siis tuleb samas valemis üldkogumi standardhälve $\sigma \equiv \sigma(X)$ asendada valimikeskmise standardhälbega, mis on $\sqrt{n}$ korda väiksem: $\sigma(\overline{X}) = \sigma/\sqrt{n}$. Võrdsustades seejärel seoste (1) ja (2) paremad pooled, saadakse usalduspiirkonda $]\bar{x} - \varepsilon, \bar{x} + \varepsilon[$ määrava ε leidmiseks valem

$$2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = \beta. \tag{3}$$

Kuna standardhälve σ on varasemate kogemuste põhjal harva teada (selle kohta vt näidet 1), siis asendatakse ta küllalt suure valimi puhul ($n > 30$) tema hinnanguga s (vt näidet 2).

Näide 1. Teatavat sorti Prantsuse veinide alkoholisisaldus on jaotatud normaalselt standardhälbega 1,2%. Uuest imporditud partiist juhuslikult valitud 60 veinipudeli keskmine alkoholisisalduse protsent on 9,3. Leida nende veinide keskmise alkoholisisalduse usalduspiirkond usaldusnivooga 0,90.

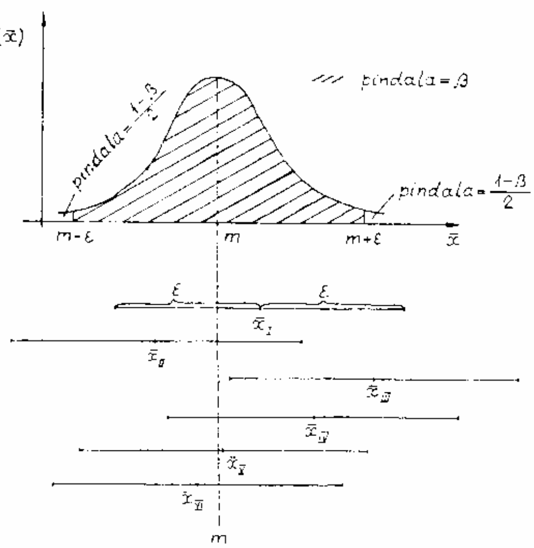
Antud on $\beta = 0,90$, $\quad n = 60$, $\quad \bar{x} = 9,3$, $\quad \sigma = 1,2$. Valemist (3) saame

$$2\Phi\left(\frac{\varepsilon\sqrt{60}}{1,2}\right) = 0,90 \quad \text{ja} \quad \Phi\left(\frac{\varepsilon\sqrt{60}}{1,2}\right) = 0,45.$$

Laplace'i funktsiooni tabelist leiame $\dfrac{\varepsilon\sqrt{60}}{1,2} = 1,645$, millest $\varepsilon \approx 0,25\%$. Usalduspiirkond tuleb]9,3-0,25; 9,3+0,25[ehk]**9,05; 9,55**[.

Leitud usalduspiirkond katab kogu uue veinipartii keskmise alkoholisisalduse tõenäosusega 0,90. See tähendab, et leidub 10% valimeid (mahuga 60), mille usalduspiirkond ei sisaldagi üldkogumi keskväärtust. Ühelt valimilt teisele üle minnes muutub valimi aritmeetiline keskmine ning seoses sellega muutuvad ühtlasi ka usalduspiirkonna rajad. Niisiis, usaldusnivood ei saa siduda üldkogumi keskväärtusega μ, vaid ta on seotud usalduspiirkonna radadega. Seega on kirjutis $P(9,05 < \mu < 9,55) = 0,90$ *vale*, sest $\mu = const$. Kui μ asub usalduspiirkonnas (tegelikult on seda võimatu kinnitada või ümber lükata, sest μ väärtus on tundmatu), siis on sündmus $\mu \in]9,05; 9,55[$ kindel (tõenäosusega 1, mitte 0,90). Kui μ ei asu usalduspiirkonnas, siis on sündmus $\mu \in]9,05; 9,55[$ võimatu (tõenäosusega 0).

Eelnevas näites kirjeldatud konkreetse situatsiooni üldisem analüüs on esitatud järgneval joonisel, millel on kujutatud valimi mahuga n aritmeetilise keskmise $\overline{X}$ (mida on vaadeldud juhusliku suurusena, mis muutub ühelt valimilt teisele üle minnes) jaotustihedust $f(\bar{x})$. Siinjuures $\beta \cdot 100\%$ valimeid annab usalduspiirkonna, mis sisaldab üldkogumi keskväärtust μ (nende hulgast on joonisel I, II, IV, V ja VI usalduspiirkond) ning $(1-\beta) \cdot 100\%$ valimeid annab usalduspiirkonna, mis ei sisalda üldkogumi keskväärtust (joonisel III usalduspiirkond):



Näide 2. Leida punkti 5.1 näite andmetel punkti 5.2 näites 2 arvutatud keskväärtuse usalduspiirkond usaldusnivooga $\beta = 0,99$.

Antud on $\bar{x} = 599,9$, $n = 300$, $s = 15,74$. Valemist (3) leiame ε:

$$2\Phi\left(\frac{\varepsilon\sqrt{300}}{15,74}\right) = 0,99 \text{ ja } \Phi\left(\frac{\varepsilon\sqrt{300}}{15,74}\right) = 0,495, \text{ millest } \frac{\varepsilon\sqrt{300}}{15,74} = 2,58 \text{ ja } \varepsilon \approx 2,4.$$

Keskväärtuse usalduspiirkond tuleb

$]599,9 - 2,4;\ 599,9 + 2,4[$ ehk $]\mathbf{597,5};\ \mathbf{602,3}[$.

5.5. Normaaljaotuse keskväärtuse usalduspiirkond, kui valimi maht on väike. Studenti jaotus

Käesoleva punkti lõpus toodud normaaljaotuse keskväärtuse usalduspiirkonna leidmise algoritmi kasutatakse juhul, kui valimi maht on suhteliselt väike ($n \leq 30$).

Tsentreerime valimikeskmise $\overline{X}$, lahutades temast üldkogumi keskväärtuse: $\overline{X} - \mu$. Normeerides tulemuse valimikeskmise standardhälbe hinnanguga (ehk teisiti nimetatuna standardveaga) $\sigma(\overline{X}) = s/\sqrt{n}$ läbijagamise teel, saame uue juhusliku suuruse

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{s}, \tag{1}$$

mis allub **Studenti jaotusele** (ehk t-jaotusele) **vabadusastmete arvuga** $k = n - 1$, kus n on valimi maht.

Vabadusastmete arv k on valimi elementide arv, mis võivad vabalt muutuda (omandada suvalisi väärtusi). Oletame, et valimis on kaks elementi ning me teame nende aritmeetilist keskmist $\bar{x}$. Siis võib vabalt valida vaid ühte elementi, kuna teine on aritmeetilise keskmise kaudu määratud. Olgu näiteks teada $\bar{x} = 6$. Kui valida $x_1 = 4$, siis $\bar{x} = (x_1 + x_2)/2$ ja x_2 on määratud: $x_2 = 2\bar{x} - x_1 = 2 \cdot 6 - 4 = 8$. Seega on vaadeldavas näites üks vabadusaste: $k = n - 1 = 2 - 1 = 1$.

Valemiga (1) defineeritud juhuslikus suuruses T sisalduva valimi standardhälbe hinnangu s määramiseks peab aritmeetiline keskmine $\bar{x}$ olema teada. Seega üks valimi element on "kinni" ja vabaks jääb neid järelikult $k = n - 1$.

Kui seoses (1) asendada hinnang s üldkogumi standardhälbega σ, siis saadakse normeeritud normaaljaotusega juhuslik suurus (mille keskväärtus on teatavasti null ja standardhälve üks). Seetõttu on mõistetav, et vabadusastmete arvu suurenemisel läheneb Studenti jaotus *kiiresti* normeeritud normaaljaotusele. Kui $k = 30$, siis on Studenti jaotuse erinevus normeeritud normaaljaotusest mõni protsent ja $k \geq 30$ puhul võib keskväärtuse usalduspiirkonna leidmisel lähtuda ka normaaljaotusest (nagu eelmises punktis kirjeldatud). Seega on Studenti jaotus oluline väikese mahuga valimite puhul. Tema eeliseks on sõltumatus tundmatutest parameetritest μ (keskväärtus) ja σ (standardhälve). Märgime, et Studenti jaotuse dispersioon $DT = k/(k-2)$, $(k \geq 3)$ on suurem kui normeeritud normaaljaotuse dispersioon $DX = 1$.

Keskväärtuse usalduspiirkonna leidmiseks läheme võrduses (vt punkti 5.4 seost (1))

$$P(|\overline{X} - \mu| < \varepsilon) = \beta$$

suuruselt $\overline{X}$ seosega (1) üle suurusele T ning saame pärast teisendusi usalduspiirkonda $]\bar{x} - \varepsilon, \bar{x} + \varepsilon[$ määrava ε leidmiseks valemi

$$\varepsilon = \frac{t_{k;\alpha} \cdot s}{\sqrt{n}}, \tag{2}$$

kus $k = n - 1$, $\alpha = (1 + \beta)/2$ ja $t_{k;\alpha}$ on Studenti jaotuse **kvantiil**, s.o niisugune t väärtus, mille korral kehtib võrdus $P(T < t_{k;\alpha}) = \alpha$. Studenti jaotuse kvantiilid $t_{k;\alpha}$ leitakse lisas 2 toodud tabelist.

Keskväärtuse usalduspiirkonna leidmise algoritm

1. Valimi alusel leitakse keskväärtuse ja standardhälbe punktihinnangud $\bar{x}$ ja s.
2. Ette antakse usaldusnivoo β (näiteks 0,95 või 0,99).
3. Arvutatakse $k = n - 1$ ja $\alpha = (1 + \beta)/2$ ning Studenti jaotuse kvantiilide tabelist leitakse $t_{k;\alpha}$.
4. Valemiga (2) arvutatakse $\varepsilon = \dfrac{t_{k;\alpha} \cdot s}{\sqrt{n}}$;
5. Koostatakse usalduspiirkond $]\bar{x} - \varepsilon, \bar{x} + \varepsilon[$.

Näide. Autobaasis kontrolliti teatud automudeli kuue auto kütusekulu, mis osutus järgnevaks (l/100 km): 18,6; 18,4; 19,2; 20,8; 19,4; 20,5. Leida keskmise kütusekulu usalduspiirkond usaldusnivooga $\beta = 0,95$.

Kuna valimi maht $n = 6$ on väike, siis on antud ülesande lahendamiseks õige kasutada Studenti jaotust. Vastavalt eelnevale algoritmile:

1. Arvutame punktihinnangud $\bar{x} = 19,48$ ja $s = 0,98$, kasutades alajaotuses 5.2 toodud valemeid (1), (3) ja (4).
2. Etteantud usaldusnivoo $\beta = 0,95$.
3. Valimi maht $n = 6$, vabadusastmete arv $k = n - 1 = 5$, $\alpha = (1 + \beta)/2 = (1 + 0,95)/2 =$ $=0,975$. Lisas 2 olevast tabelist vaatame kvantiili $t_{k;\alpha} = t(5;\ 0,975) = 2,571$.
4. Valemiga (2) arvutame $\varepsilon = \dfrac{2,571 \cdot 0,98}{\sqrt{6}} = 1,03$.
5. Koostame usalduspiirkonna $]19,48 - 1,03;\ 19,48 - 1,03[$ ehk $]\mathbf{18,45};\ \mathbf{20,51}[$.

5.6. Normaaljaotuse dispersiooni ja standardhälbe usalduspiirkond. χ^2-jaotus

Vaatleme normaaljaotusega üldkogumist saadud valimeid mahuga n, mille dispersiooni nihutamata hinnangut võib käsitleda kui juhuslikku suurust S^2, mis muutub ühelt valimilt teisele üle minnes. Moodustame uue juhusliku suuruse, milles sisaldub S^2:

$$\chi^2 = \frac{S^2 \cdot k}{\sigma^2}. \tag{1}$$

Eelnevas on σ^2 on üldkogumi dispersioon ja $k = n - 1$ vabadusastmete arv. Juhuslik suurus (1) on **χ^2-jaotusega** (loetakse hii-ruut-jaotusega), mis on asümmeetriline ja läheneb vabadusastmete arvu k kasvades normaaljaotusele. Kui $k \geq 30$, siis erineb χ^2-jaotus juba küllalt vähe normaaljaotusest (parameetritega $\mu = k$ ja $\sigma = \sqrt{2k}$).

Lisas 3 on tabuleeritud **χ^2-jaotuse täiendkvantiilid** $\chi^2(\alpha, k)$. Täiendkvantiilid on χ^2 niisugused väärtused, mille puhul kehtib võrdus $P[\chi^2 > \chi^2(\alpha, k)] = \alpha$.

Vastavalt etteantud usaldusnivoole β (näiteks 0,95 või 0,99) avaldub normaaljaotuse **dispersiooni usalduspiirkond** χ^2-jaotuse täiendkvantiilide abil järgnevalt:

$$\left] \frac{s^2 \cdot k}{\chi^2\left(\dfrac{1-\beta}{2}, k\right)}, \frac{s^2 \cdot k}{\chi^2\left(\dfrac{1+\beta}{2}, k\right)} \right[. \tag{2}$$

Usalduspiirkond (2) on dispersiooni punktihinnangu s^2 suhtes asümmeetriline, mis tuleneb χ^2-jaotuse asümmeetriast.

Standardhälbe usalduspiirkonna leidmiseks tuleb piirkonna (2) mõlemast rajast leida ruutjuur (vt järgnevaid näiteid).

Näide 1. Leida punkti 5.5 näite andmetel dispersiooni ja standardhälbe usalduspiirkond usaldusnivooga 0,95.

$s^2 = 0,96; \quad s = 0,98; \quad n = 6; \quad k = n-1 = 5; \quad \beta = 0,95.$

Lisas 3 toodud χ^2-jaotuse täiendkvantiilide tabelist leiame

$$\chi^2\left(\frac{1-\beta}{2}, k\right) = \chi^2(0,025;\ 5) = 12,832,$$

$$\chi^2\left(\frac{1+\beta}{2}, k\right) = \chi^2(0,975;\ 5) = 0,831.$$

Dispersiooni usalduspiirkond on vastavalt seosele (2) järgmine:

$$\left]\frac{0,96\cdot 5}{12,832};\ \frac{0,96\cdot 5}{0,831}\right[\quad \text{ehk} \quad]\mathbf{0,37};\ \mathbf{5,78}[.$$

Standardhälbe usalduspiirkond on

$$\left]\sqrt{0,37};\ \sqrt{5,78}\right[\quad \text{ehk} \quad]\mathbf{0,6};\ \mathbf{2,4}[.$$

Kuna valimi maht on väike, siis on leitud usalduspiirkonnad üsna avarad.

Näide 2. Leida punkti 5.1 näite andmetel punkti 5.2 näites 2 arvutatud dispersiooni ja standardhälbe usalduspiirkond usaldusnivooga 0,99.

$s^2 = 247,8; \quad s = 15,74; \quad n = 300; \quad k = n-1 = 299; \quad \beta = 0,99.$

χ^2-jaotuse täiendkvantiilide

$$\chi^2\left(\frac{1-\beta}{2}, k\right) = \chi^2(0,005;\ 299) = 255 + \frac{1}{3}(585-255) = 365,$$

$$\chi^2\left(\frac{1+\beta}{2}, k\right) = \chi^2(0,995;\ 299) = 152 + \frac{1}{3}(422-152) = 242$$

leidmisel vaatame lisas 3 olevast tabelist

$$\chi^2(0,005;\ 200) = 255, \quad \chi^2(0,005;\ 500) = 585,$$
$$\chi^2(0,995;\ 200) = 152, \quad \chi^2(0,995;\ 500) = 422$$

ja kasutame lineaarset interpolatsiooni.

Dispersiooni usalduspiirkonna leiame vastavalt seosele (2):

$$\left]\frac{247,8\cdot 299}{365};\ \frac{247,8\cdot 299}{242}\right[\quad \text{ehk} \quad]\mathbf{203};\ \mathbf{306}[.$$

Standardhälbe usalduspiirkond on

$$\left]\sqrt{203};\ \sqrt{306}\right[\quad \text{ehk} \quad]\mathbf{14,2};\ \mathbf{17,5}[.$$

5.7. Juhusliku sündmuse tõenäosuse usalduspiirkond

Toimugu sündmus A n ühesugusel ja sõltumatul katsel m korda. Sündmuse A toimumise tõenäosus üksikkatsel p olgu teadmata. Selle tundmatu tõenäosuse *punktihinnanguks* võetakse *sündmuse sagedus* $p^* = m/n$. Märgime $q^* = 1 - p^*$.

Kui (vt ka punkti 4.4 võrratusi (2))

$$1)\ np^* > 5 \text{ juhul } p^* < 0,5 \text{ või } 2)\ nq^* > 5 \text{ juhul } p^* > 0,5, \tag{1}$$

siis võib usalduspiirkonda

$$]p^* - \varepsilon;\ p^* + \varepsilon[\tag{2}$$

määrava ε leidmiseks kasutada ligikaudse lähendina normaaljaotust, asendades punkti 4.3 valemis (5) standardhälbe σ standardveaga $\sigma(p^*) = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{p^*q^*}{n}}$

$$2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{p^* \cdot q^*}}\right) = \beta, \tag{3}$$

kus β on usaldusnivoo.

Näide 1. Suurettevõte korraldas tootmise ümber ja küsis juhuslikult valitud 160-lt töötajalt, kas nad on sellega rahul. 85 vastas jaatavalt. Leida ümberkorraldustega rahul olevate töötajate osakaal usaldusnivooga 0,95.

Antud on valimi maht $n = 160$ ning jaatavalt vastanute arv $m = 85$. Arvutame punktihinnanguks võetava sündmuse sageduse $p^* = m/n = 85/160 \approx 0,531$ ja leiame $q^* = 1 - p^* \approx 0,469$.

Kuna vastavalt tingimusele (1) $nq^* = 160 \cdot 0,469 \approx 75 > 5$, siis võib ε leidmiseks kasutada valemit (3):

$$2\Phi\left(\varepsilon \cdot \sqrt{\frac{160}{0,531 \cdot 0,469}}\right) = 0,95 \quad \text{ehk} \quad \Phi\left(\varepsilon \cdot \sqrt{\frac{160}{0,531 \cdot 0,469}}\right) = 0,475,$$

millest leiame Laplace'i funktsiooni tabeli abil

$$\varepsilon \cdot \sqrt{\frac{160}{0,531 \cdot 0,469}} = 1,96.$$

Siit $\varepsilon \approx 0,077$ ja usalduspiirkond tuleb seose (2) kohaselt $]0,531 - 0,077;\ 0,531 + 0,077[$ ehk $]\mathbf{0,454};\ \mathbf{0,608}[$. Seega usaldusnivooga $\beta = 0,95$ on ümberkorraldustega rahul olevaid töötajaid 45,4% kuni 60,8%.

Näide 2. Viiesajalt juhuslikult võetud valimisõiguslikult elanikult küsiti valimispäeva eelõhtul, kas nad soovivad ühinemist Euroopa Liiduga. Neist 260 vastas jaatavalt. Hinnata valimistel poolt hääletavate protsenti usaldusnivool $\beta = 0,90$.

$n = 500$; $m = 260$; $p^* = m/n = 0,52$; $q^* = 1 - p^* = 0,48$.

Analoogselt näitega 1 leiame $\varepsilon \approx 0,04$ ning usalduspiirkond tuleb seose (2) kohaselt $]\mathbf{0,48};\ \mathbf{0,56}[$. See tähendab, et usaldusnivooga $\beta = 0,90$ on valimistel poolthääletajaid 48% kuni 56%.

Kuna 50% raja jääb usalduspiirkonda, siis püstitame täiendavalt järgmise probleemi: mitut inimest tuleks küsitleda, et (kõrgema) usaldusnivooga $\beta = 0,99$ võiks ennustada valimistulemusi täpsusega $\pm 2\%$, kui eelküsitluste põhjal oli punktihinnang $p^* = 0,52$?

Antud on $\beta = 0,99$ ja $\varepsilon = 0,02$. Leida tuleb representatiivse valimi maht n. Vastavalt valemile (3) saame

$$2\Phi\left(0,02 \cdot \sqrt{\frac{n}{0,52 \cdot 0,48}}\right) = 0,99 \quad \text{ehk} \quad \Phi\left(0,02 \cdot \sqrt{\frac{n}{0,52 \cdot 0,48}}\right) = 0,495,$$

millest leiame Laplace'i funktsiooni tabeli abil

$$0,02 \cdot \sqrt{\frac{n}{0,52 \cdot 0,48}} = 2,575. \quad \text{Siit} \quad n \approx \mathbf{4138}.$$

Niisiis, kui küsitletuid on täpselt 4138, on usaldusnivooga 0,99 poolthääletanuid parajasti 50% kuni 54%. Kuna usaldusnivoo on 0,99, siis võib nüüdki (väikese tõenäosusega) juhtuda, et poolthääletanuid on alla 50%. Kui usaldusnivood tõsta valimi konstantse mahu juures, siis läheb usalduspiirkond punktihinnangu $p^* = 0,52$ ümber avaramaks. Kui suurendada valimi mahtu, läheb usalduspiirkond kitsamaks. Täpse prognoosi saab anda ainult siis, kui *valimi* asemel vaadelda *üldkogumit*, st küsitleda kõiki valimisõiguslikke inimesi (ja sedagi vaid eeldusel, et kõhklejaid ning viimasel hetkel ümbermõtlejaid ei ole).

5.8. Ülesanded

5.1. Mõõdeti 270 detaili. Nende pikkus oli vahemikus 66 kuni 90 *cm*. Jagades selle vahemiku 2 *cm* pikkusteks intervallideks, saadakse järgmine variatsioonrida:

Intervalli nr	Intervallid	n_i
1	[66, 68[	4
2	[68, 70[	12
3	[70, 72[	24
4	[72, 74[	41
5	[74, 76[	50
6	[76, 78[	53

Intervalli nr	Intervallid	n_i
7	[78, 80[	39
8	[80, 82[	26
9	[82, 84[	13
10	[84, 86[	5
11	[86, 88[	2
12	[88, 90]	1

1) Joonestada empiirilise jaotusfunktsiooni graafik ning histogramm.
2) Leida jaotuse keskväärtuse, dispersiooni ja standardhälbe nihutamata hinnangud.

3) Oletades normaaljaotust, panna leitud keskväärtuse ja standardhälbe hinnangud normaaljaotuse jaotustiheduse $f(x)$ avaldisse (punkti 4.3 valemisse (1)). Arvutada jaotustiheduse väärtused iga intervalli keskel $f(x_k)$ ja kanda leitud väärtused histogrammile. Hinnata histogrammi ning jaotustiheduse graafiku kooskõla.

4) Kasutades normaaljaotust, leida keskväärtuse usalduspiirkond usaldusnivooga $\beta =$ $=0{,}99$.

5) Leida dispersiooni ja standardhälbe usalduspiirkond usaldusnivooga $\beta = 0,99$.

5.2. 123 töölist tootsid ühesuguseid detaile. Nende poolt ühe kuu jooksul valmistatud detailide hulk on esitatud järgnevas variatsioonreas:

Intervalli nr	Intervallid	n_i
1	[260; 270[	9
2	[270; 280[	18
3	[280; 290[	55
4	[290; 300[	27
5	[300; 310]	14

Täita ülesande **5.1** punktid 1 kuni 5, võttes usaldusnivoo $\beta = 0,95$.

5.3. Enne uue autorehvimudeli turule laskmist katsetas tehas toodangust juhuslikult valitud 150 rehvi. Katsetuste tulemused on toodud järgnevas variatsioonreas:

Intervalli nr	Intervallid [*tuh. km.*]	n_i
1	[20; 25[	7
2	[25; 30[	14
3	[30; 35[	28
4	[35; 40[	45
5	[40; 45[	30
6	[45; 50[	15
7	[50; 55]	11

Täita ülesande **5.1** punktid 1 kuni 5, võttes usaldusnivoo $\beta = 0,99$.

5.4. Juhuslik suurus X on normaaljaotusega. Standardhälve $\sigma = 3$. Leida keskväärtuse usalduspiirkond usaldusnivooga $\beta = 0,95$, kui aritmeetiline keskmine $\bar{x} = 4,1$ määrati valimist mahuga $n = 36$.

5.5. Suhkrupakkide kaal on jaotatud normaalselt standardhälbega 15 grammi. Juhuslikult valitud 25 paki keskmine kaal on 495 grammi. Leida suhkrupakkide keskmise kaalu usalduspiirkond usaldusnivooga 0,95.

5.6. Terasproovikehade tõmbetugevus oli (kG/mm^2): 34, 53, 48, 48, 46, 38, 45, 42, 47, 50. Leida

1) keskmise tõmbetugevuse usalduspiirkond usaldusnivooga $\beta = 0,95$;

2) dispersiooni ja standardhälbe usalduspiirkond sama usaldusnivooga.

5.7. Kahekümnele juhuslikult valitud inimesele näidati televisioonireklaami ning küsiti nende meeldivushinnangut skaalas $0 \dots 100$. Tulemused olid järgmised: 89, 75, 59, 96, 88, 71, 43, 62, 80, 92, 76, 72, 67, 60, 79, 85, 77, 83, 87, 53. Eeldades normaaljaotust, leida keskmise meeldivushinnangu usalduspiirkond usaldusnivooga $\beta = 0,95$. Leida standardhälbe usalduspiirkond sama usaldusnivooga.

5.8. Valmisriiete kaupluste kett huvitus üliõpilaste kulutustest riietusele õppeaasta esimese kuu jooksul. Kümne üliõpilase küsitlus näitas, et kulutused dollarites olid vastavalt 79.21; 143.62; 95.08; 34.27; 211.87; 98.26; 65.93; 159.11; 127.28; 53.97. Eeldades normaaljaotust, leida ülikooli üliõpilaste keskmiste kulutuste usalduspiirkond usaldusnivooga $\beta = 0,90$. Leida standardhälbe usalduspiirkond sama usaldusnivooga.

5.9. Autobaasi huvitab autode remondiks kuluv aeg. Üheksa auto ekspluatatsiooni analüüs näitas, et autod olid aasta jooksul remondis vastavalt 16, 10, 21, 22, 8, 17, 19, 14, 19 päeva. Eeldades normaaljaotust, leida autobaasi autode keskmise remondiaja usalduspiirkond usaldusnivooga $\beta = 0,90$. Leida standardhälbe usalduspiirkond sama usaldusnivooga.

5.10. Teeninduskombinaadis registreeriti seitsmel juhuslikult valitud päeval kaebusi alljärgnevalt: 10, 12, 8, 5, 11, 9, 14. Eeldades normaaljaotust, leida keskmise kaebuste arvu usalduspiirkond usaldusnivooga $\beta = 0,90$. Leida standardhälbe usalduspiirkond sama usaldusnivooga.

5.11. Lennukompanii küsitles uue lennuliini reisijaid. 347-st juhuslikult valitud küsitletust oli 201 töölähetusel. Leida seda uut lennuliini töölähetuseks kasutavate reisijate osakaal usaldusnivooga $\beta = 0,90$. Mitut reisijat tuleks küsitleda, et usaldusnivooga $\beta = 0,90$ võiks hinnata seda uut lennuliini töölähetuseks kasutavate reisijate osakaalu täpsusega $\pm 3\%$?

5.12. 650 kodus toimunud õnnetusjuhtumi analüüs näitas, et 180 nendest oli tingitud kukkumisest. Leida kukkumisest tingitud õnnetuste osakaal usaldusnivooga $\beta = 0,95$. Mitut õnnetusjuhtumit tuleks analüüsida, et usaldusnivooga $\beta = 0,95$ võiks hinnata kukkumisest tingitud õnnetuste osakaalu täpsusega $\pm 2\%$?

5.13. 676 autojuhi kontroll näitas, et 6 nendest oli alkoholijoobes. Leida alkoholijoobes olevate juhtide osakaal usaldusnivooga $\beta = 0,95$.

6. Statistiliste hüpoteeside kontrollimine

6.1. Statistilised hüpoteesid. Esimest ja teist liiki vead. Kriitilised piirkonnad

Hüpoteesideks võivad olla näiteks oletused, et

1) talverehvi uue mudeli pidurdusteekond on lühem kui vanal mudelil;
2) kahest tööstusharust on ühes töötasu hajuvus väiksem kui teises;
3) ühesugustes tingimustes ühesugust tööd tegevate tööliste tööviljakus allub normaaljaotusele.

Kõikne uuring võimaldaks täpselt vastata küsimusele, kas need hüpoteesid on tõesed või mitte. Üldkogumi analüüs võib aga olla kas väga töömahukas või mõnikord hoopis võimatu. Järgnevas eeldatakse, et kõikset uuringut pole tehtud, vaid on piirdutud valimi vaatlemisega. Sellest hoolimata on vaja otsustada üldkogumi kohta tehtud oletuste kehtivuse või mittekehtivuse üle. Kui hüpoteeside kontrollimisel lähtutakse valimi andmetest, siis räägitakse **statistiliste hüpoteeside** kontrollimisest. Statistilisteks hüpoteesideks on näiteks oletused:

1) kahe jaotuse parameetrite (keskväärtuste või dispersioonide) võrdsusest või olulisest erinevusest;
2) jaotusseaduse tüübi kohta;
3) juhuslike suuruste X ja Y vahelise lineaarse sõltuvuse olemasolust või puudumisest.

Statistiliste hüpoteeside kontrollimise üldised põhimõtted võib süstematiseerida järgmisesse viiesammulisse protseduuri.

1. Nullhüpoteesi ja konkureeriva hüpoteesi püstitamine

Hüpoteesid formuleeritakse teineteist välistavate vastandhüpoteeside paarina H_0 ja H_1, kusjuures nendest hüpoteesidest saab tõene olla vaid üks. Hüpoteesi H_0 nimetatakse nullhüpoteesiks. Hüpoteesi H_1 on erinevate autorite poolt nimetatud:

1) konkureerivaks (alternatiivseks) hüpoteesiks;
2) tõestatavaks (sisukaks) hüpoteesiks.

2. Olulisuse nivoo valik. Esimest ja teist liiki vead

Kuna statistilisi hüpoteese kontrollitakse valimi andmetel, siis tuleb arvestada kahesuguste vigade tekkimise võimalusega:

1) **esimest liiki viga**, mille puhul lükatakse tagasi õige nullhüpotees (ehk teisiti: vastu võetakse tõestatav hüpotees H_1, kuid tegelikult on õige H_0);
2) **teist liiki viga**, mille puhul võetakse vastu vale nullhüpotees, kuid õige on hoopis tõestatav hüpotees H_1.

Statistiliste hüpoteeside kontrollimisel antakse ette esimest liiki vea maksimaalselt lubatav tõenäosus α, mida nimetatakse **olulisuse nivooks**. Sama mahuga valimi puhul mõlemat liiki vigu korraga vähendada ei saa. Kui suurendada olulisuse nivood α, siis kasvab esimest liiki vea oht, kui aga vähendada olulisuse nivood, siis kasvab teist liiki vea oht. Kui võtta olulisuse nivoo $\alpha = 0$, siis võetakse vastu kõik nullhüpoteesid, sealhulgas ka valed. Sagedamini võetakse $\alpha = 0,05$ või $\alpha = 0,01$ (st õige hüpoteesi tagasilükkamise risk on vastavalt 5% või 1%). α valik sõltub konkreetsest ülesandest: kumb on ohtlikum, kas õige nullhüpoteesi tagasilükkamine (esimest liiki viga) või vale vastuvõtmine (teist liiki viga). Võimaluse korral tuleks hüpoteesid sõnastada nii, et esimest liiki viga oleks suurema raskusastmega ja seetõttu ebasoovitavam kui teist liiki viga. Kui näiteks elumaja ehitamisel tekib avariikahtlus, siis peaks nullhüpoteesiks valima soovitule vastupidise väite "mitte jätkata elumaja ehitust". Kui õige nullhüpotees lükatakse tagasi (esimest liiki viga), siis võivad tagajärjeks olla inimohvrid. Kui aga võetakse vastu vale nullhüpotees (teist liiki viga), siis piirdutakse materiaalse kahjuga.

Teist liiki vea esinemise tõenäosus ei ole ülesande püstitamisel määratud (ei ole piiratud, nii nagu esimest liiki vea tekkimise tõenäosus on piiratud olulisuse nivooga α). Osutub, et mõnes ülesandes võib teist liiki vea tõenäosus ulatuda isegi väärtuseni $1-\alpha$ (H_0 võetakse vastu, kuid eksimise tõenäosus on $1-\alpha$).

3. Statistilise kriteeriumi valik

Statistiline kriteerium (statistik) püütakse konstrueerida nii, et teist liiki vea tõenäosus oleks minimeeritud. Kasutatakse näiteks z, t, χ^2 *ja* F statistikuid, mis on seotud vastavalt normaal-, Studenti, χ^2- ja Fisheri jaotusega.

4. Kriitiliste punktide leidmine ja kriitilise piirkonna konstrueerimine

Valitud statistilisest kriteeriumist Z ja olulisuse nivoost α lähtudes leitakse **kriitilised punktid** z_{kr} ja statistilise kriteeriumi Z võimalike väärtuste hulk jagatakse kaheks mittelõikuvaks alamhulgaks:

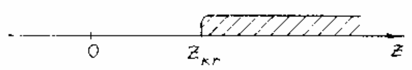
1) kriitiliseks piirkonnaks, millesse sattumisel lükatakse nullhüpotees tagasi ning loetakse sisukas hüpotees tõestatuks;

2) ülejäänud piirkonnaks, millesse sattumisel võetakse nullhüpotees vastu, ehk nagu sageli öeldakse, pole alust nullhüpoteesi tagasi lükata.

Nullhüpoteesi ei ole võimalik tõestada. Kui nullhüpotees tuli vastu võtta, siis peaks võimaluse korral analüüsi jätkama ning tegema vaadeldava probleemiga seoses täiendavaid uuringuid.

Kriitilised piirkonnad liigitatakse parem-, vasak- ja kahepoolseteks. Iga konkreetse statistilise hüpoteesi kontrollimiseks kasutatava kriitilise piirkonna vajalik kuju sõltub konkureeriva hüpoteesi H_1 püstitusest. Kui konkureeriv hüpotees sisaldab võrratust kujul "$>$" (näiteks $H_1: \ EX > EY$), siis kasutatakse parempoolset kriitilist piirkonda, kui võrratust "$<$", siis vasakpoolset. Kui H_1 ei määra suunda, vaid tingimus on formuleeritud mittevõrdusena "$\neq$" (näiteks $H_1: \ EX \neq EY$), siis kasutatakse kahepoolset sümmeetrilist kriitilist piirkonda. Vaatleme lähemalt parempoolset ja kahepoolset sümmeetrilist kriitilist piirkonda, mida järgnevas praktiliselt kasutatakse.

Parempoolne kriitiline piirkond jääb punktist $z_{kr} > 0$ paremale:

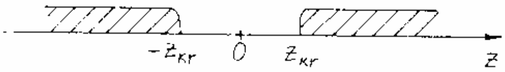


Kriitilise punkti z_{kr} leidmiseks antakse ette olulisuse nivoo α (näiteks $\alpha = 0,05$) ja nõutakse, et nullhüpoteesi kehtivuse korral oleks $Z > z_{kr}$:

$$P(Z > z_{kr}) = \alpha. \tag{1}$$

Sõltuvalt statistilise kriteeriumi valikust omandab tingimus (1) konkreetse kuju. Praktikas kasutatavate statistiliste kriteeriumide kriitiliste punktide leidmiseks on koostatud tabelid, millest olulisemad on toodud käesoleva õppevahendi lisades 1 kuni 4.

Kahepoolne sümmeetriline kriitiline piirkond on sümmeetriline nullpunkti suhtes ning jääb punktist $-z_{kr}$ vasakule ja punktist $+z_{kr}$ paremale:



Kriitiline punkt z_{kr} leitakse tingimusest

$$P(Z > z_{kr}) = \alpha/2. \tag{2}$$

5. Järeldused hüpoteeside kehtivuse kohta

Valimi andmetel arvutatakse statistilise kriteeriumi empiiriline väärtus z_{emp}. Kriteeriumi z_{emp} arvutamiseks kasutatav eeskiri sõltub konkreetse ülesande sisulisest püstitusest. Mõnede praktiliste ülesannete lahendamise algoritme vaadeldakse käesoleva õppevahendi järgnevates punktides 6.2 ja 6.3. Kui z_{emp} satub kriitilisse piirkonda, st parempoolse kriitilise piirkonna puhul on

$$z_{emp} > z_{kr} \tag{3}$$

või kahepoolse sümmeetrilise kriitilise piirkonna puhul on

$$|z_{emp}| > z_{kr}, \tag{4}$$

siis lükatakse nullhüpotees H_0 tagasi. Sellega on sisukas hüpotees tõestatud, sest leidus vähemalt üks valim, mille korral nullhüpotees ei kehti. Kui võrratus (3) või (4) pole täidetud, siis pole alust nullhüpoteesi tagasi lükata, kusjuures ei saa väita, et sellega oleks nullhüpotees tõestatud.

6.2. Kahe normaaljaotuse keskväärtuse võrdlemine

Olgu kahe üldkogumiga seotud juhuslikud suurused X ja Y (ehk teisiti: meid huvitavad tunnused X ja Y) normaaljaotusega ning sõltumatud. Märgime nende juhuslike suuruste keskväärtused vastavalt μ_x ja μ_y. Püstitame nullhüpoteesi üldkogumite keskväärtuste võrdsuse kohta $H_0: \ \mu_x = \mu_y$. Konkureeriva hüpoteesi formuleerimise võimalustest vaatleme kahte:

1) $H_1: \ \mu_x \neq \mu_y$;

1) $H_1: \ \mu_x > \mu_y$.

Esimesel juhul kasutatakse hüpoteeside kontrollimiseks kahepoolset sümmeetrilist kriitilist piirkonda, teisel juhul parempoolset kriitilist piirkonda.

Olgu üldkogumitest võetud valimite $x_1, x_2, \ldots, x_n$ ja $y_1, y_2, \ldots, y_m$ mahud n ja m. Nende valimite andmetel arvutatakse statistilise kriteeriumi empiiriline väärtus

$$z_{emp} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{n} + \dfrac{s_y^2}{m}}}, \tag{1}$$

kus $\overline{x}$ ja $\overline{y}$ on valimikeskmised ning $\dfrac{s_x^2}{n}$ ja $\dfrac{s_y^2}{m}$ on valimikeskmiste dispersioonihinnangud.

Kui valimite mahud n ja m on

1) suured ($n > 30$ ja $m > 30$), siis on z_{emp} ligikaudselt normeeritud normaaljaotusega ning kriitiliste punktide z_{kr} leidmiseks kasutatakse normaaljaotust; kui aga valimite mahud on

2) väikesed ($n \leq 30$ ja $m \leq 30$), siis on z_{emp} ligikaudselt Studenti jaotusega ning kriitiliste punktide z_{kr} leidmiseks kasutatakse Studenti jaotust.

Sõltuvalt konkureeriva hüpoteesi H_1 püstitusest ning valimite mahtudest n ja m formeerub neli statistiliste hüpoteeside kontrollimise algoritmi. Statistilise kriteeriumi empiiriline väärtus z_{emp} leitakse kõigil neljal juhul valemiga (1).

1. $\mathbf{H_0 : \mu_x = \mu_y; \quad H_1 : \mu_x \neq \mu_y; \quad n > 30, \quad m > 30.}$

Ette antakse olulisuse nivoo α ja leitakse kriitiline punkt z_{kr} kahepoolsele sümmeetrilisele kriitilisele piirkonnale seosest

$$P(|Z| < z_{kr}) = 2\Phi(z_{kr}) = 1 - \alpha, \tag{2}$$

kus Z on normeeritud normaaljaotusega juhuslik suurus ja $\Phi(z)$ on Laplace'i funktsioon. Vahet $1 - \alpha$ võib tõlgendada kui usaldusnivood β.

Kui $|z_{emp}| > z_{kr}$, siis lükatakse nullhüpotees H_0 tagasi.

Kui $|z_{emp}| < z_{kr}$, siis pole alust nullhüpoteesi H_0 tagasi lükata.

Näide 1. Suurettevõttes võrreldi ametiühingusse kuuluvate töötajate ja sinna mittekuuluvate töötajate puudumisi aasta jooksul. Juhuslikult valitud 50 ametiühinguliiget puudusid igaüks keskmiselt 9,3 päeva, kusjuures standardhälve oli 3,1 päeva. Juhuslikult valitud 45 ametiühingusse mittekuulujat puudusid igaüks keskmiselt 8,7 päeva standardhälbega 2,3 päeva. Kontrollida olulisuse nivool $\alpha = 0,05$ väidet, et ametiühingusse kuulumine mõjutab kõikide ettevõtte töötajate töötajate puudumisi aasta jooksul.

$n = 50, \quad \overline{x} = 9,3$ päeva, $\quad s_x = 3,1$ päeva.

$m = 45, \quad \overline{y} = 8,7$ päeva, $\quad s_y = 2,3$ päeva.

Ülesande viimases lauses on formuleeritud tõestatav hüpotees $H_1 : \mu_x \neq \mu_y$. Nullhüpotees on $H_0 : \mu_x = \mu_y$.

Valemiga (1) arvutame

$$z_{emp} = \frac{9,3 - 8,7}{\sqrt{\frac{3,1^2}{50} + \frac{2,3^2}{45}}} \approx 1,08.$$

Vastavalt olulisuse nivoole $\alpha = 0,05$ leiame kriitilise punkti seosest (2):

$$2\Phi(z_{kr}) = 1 - 0,05 \text{ ehk } \Phi(z_{kr}) = 0,475, \text{ millest } z_{kr} = 1,96.$$

Kuna $z_{emp} < z_{kr}$ $(1,08 < 1,96)$, siis pole alust nullhüpoteesi tagasi lükata, st puudutud päevade arv ei sõltu ametiühingusse kuulumisest.

2. $H_0: \mu_x = \mu_y$; $H_1: \mu_x > \mu_y$; $n > 30$, $m > 30$.

Ette antakse olulisuse nivoo α ja leitakse kriitiline punkt z_{kr} parempoolsele kriitilisele piirkonnale normaaljaotuse abil seosest

$$2\Phi(z_{kr}) = 1 - 2\alpha. \tag{3}$$

Kui $z_{emp} > z_{kr}$, siis lükatakse nullhüpotees H_0 tagasi.

Kui $z_{emp} < z_{kr}$, siis pole alust nullhüpoteesi H_0 tagasi lükata.

Näide 2. Kontrollida olulisuse nivool $\alpha = 0,05$ väidet, et esimene kiirabiteenistus reageerib väljakutsetele aeglasemalt kui teine. Esimeses teenistuses jälgiti 40 väljakutset, kusjuures keskmine kohalejõudmisaeg oli 14,0 minutit standardhälbega 2,1 minutit. teises teenistuses jälgiti 50 väljakutset, kusjuures keskmine kohalejõudmisaeg oli 12,4 minutit standardhälbega 1,5 minutit.

$n = 40$, $\bar{x} = 14,0$ min, $s_x = 2,1$ min.

$m = 50$, $\bar{y} = 12,4$ min, $s_y = 1,5$ min.

Nullhüpotees $H_0: \mu_x = \mu_y$, konkureeriv hüpotees $H_1: \mu_x > \mu_y$.

Vastavalt valemile (1) arvutame $z_{emp} \approx 4,061$.

Olulisuse nivoo $\alpha = 0,05$ ning kriitilise punkti leiame seosest (3):

$$2\Phi(z_{kr}) = 1 - 2 \cdot 0,05 \text{ ehk } \Phi(z_{kr}) = 0,45, \text{ millest } z_{kr} = 1,645.$$

Kuna $z_{emp} > z_{kr}$ $(4,061 > 1,645)$, siis lükatakse nullhüpotees tagasi, st esimene teenistus reageerib väljakutsetele aeglasemalt kui teine.

3. $H_0: \mu_x = \mu_y$; $H_1: \mu_x \neq \mu_y$; $n \leq 30$, $m \leq 30$.

Valemiga (1) arvutatud z_{emp} ligikaudselt Studenti jaotusega, mille vabadusastmete arv on

$$k = \frac{(s_x^2/n + s_y^2/m)^2}{\dfrac{(s_x^2/n)^2}{n-1} + \dfrac{(s_y^2/m)^2}{m-1}}. \tag{4}$$

Valemiga (4) saadud tulemus ümardatakse lähima *väiksema* täisarvuni.

Ette antakse olulisuse nivoo α ja leitakse kriitiline punkt

$$z_{kr}\left(k;\ 1 - \frac{\alpha}{2}\right) \tag{5}$$

kahepoolsele sümmeetrilisele kriitilisele piirkonnale Studenti jaotuse kvantiilide tabelist.

Kui $|z_{emp}| > z_{kr}$, siis lükatakse nullhüpotees H_0 tagasi.

Kui $|z_{emp}| < z_{kr}$, siis pole alust nullhüpoteesi H_0 tagasi lükata.

Näide 3. Võrrelda kahe firma sigarettide nikotiinisisaldust:

$n = 20$, $\bar{x} = 3,51$ mg, $s_x = 0,42$ mg,

$m = 24$, $\bar{y} = 3,65$ mg, $s_y = 0,33$ mg.

Kontrollida olulisuse nivool $\alpha = 0,05$ väidet, et mõlema firma sigaretid on sama nikotiinisisaldusega.

Nullhüpotees H_0 : $\mu_x = \mu_y$, konkureeriv hüpotees H_1 : $\mu_x \neq \mu_y$. Kuna $n < 30$, siis lähtume Studenti jaotusest ning arvutame valemiga (4)

$$k = \frac{(0,42^2/20 + 0,33^2/24)^2}{\frac{(0,42^2/20)^2}{19} + \frac{(0,33^2/24)^2}{23}} \approx 35,8,$$

mille ümardame lähima *väiksema* täisarvuni $k = 35$.

Vastavalt olulisuse nivoole $\alpha = 0,05$ leiame kriitilise punkti seosest (5):

$$z_{kr}\left(35;\ 1 - \frac{0,05}{2}\right) = z_{kr}(35;\ 0,975) = 2,030.$$

Valemiga (1) arvutame $z_{emp} \approx -1,21$.

Kuna $|z_{emp}| < x_{kr}$ $(1,21 < 2,04)$, siis pole alust nullhüpoteesi tagasi lükata, st mõlema firma sigaretid on ühesuguse nikotiinisisaldusega.

4. $\mathbf{H_0}$: $\mu_x = \mu_y$; $\mathbf{H_1}$: $\mu_x > \mu_y$; $\mathbf{n \leq 30}$, $\mathbf{m \leq 30}$.

Valemiga (4) arvutatakse vabadusastmete arv k, ümardades tulemuse lähima väiksema täisarvuni. Ette antakse olulisuse nivoo α ja leitakse kriitiline punkt z_{kr} parempoolsele kriitilisele piirkonnale

$$z_{kr}(k;\ 1 - \alpha) \tag{6}$$

Studenti jaotuse kvantiilide tabelist.

Kui $z_{emp} > z_{kr}$, siis lükatakse nullhüpotees H_0 tagasi.

Kui $z_{emp} < z_{kr}$, siis pole alust nullhüpoteesi H_0 tagasi lükata.

Näide 4. Laserplaadimängijate tootja alandas vähesel määral hinda. Enne hinnaalandust vaadeldi ühes piirkonnas 15 juhuslikult valitud nädala keskmist läbimüüki, mis oli 6598 dollarit standardhälbega 844 dollarit. Pärast hinnaalandust oli 12 juhuslikult valitud nädala keskmine läbimüük 6870 dollarit standardhälbega 669 dollarit. Kas vähene hinnaalandus oli küllaldane laserplaadimängijate läbimüügi suurendamiseks?

$n = 12$, $\bar{x} = 6870$, $s_x = 669$,

$m = 15$, $\bar{y} = 6598$, $s_y = 844$.

Valime olulisuse nivooks $\alpha = 0,1$. Vastavalt ülesandes esitatud küsimusele püstitame nullhüpoteesi ja konkureeriva hüpoteesi kujul H_0 : $\mu_x = \mu_y$ ja H_1 : $\mu_x > \mu_y$.

Valemiga (1) arvutame statistilise kriteeriumi empiirilise väärtuse $z_{emp} \approx 0,910$.

Kuna valimite mahud on väikesed, siis lähtutakse kriitilise punkti leidmisel Studenti jaotusest. Valemiga (4) arvutatakse vabadusastmete arv $k \approx 25,0$. Vastavalt olulisuse nivoole $\alpha = 0,1$ ja seosele (6) leitakse Studenti jaotuse kvantiilide tabelist kriitiline punkt $z_{kr}(25;\ 1 - 0,1) = z_{kr}(25;\ 0,9) = 1,316$.

Kuna $z_{emp} < x_{kr}$ $(0,910 < 1,316)$, siis pole alust nullhüpoteesi tagasi lükata, st vähene hinnaalandus ei suurendanud laserplaadimängijate läbimüüki.

Keskväärtuste võrdlemine üldkogumite võrdsete dispersioonide korral

Järgmises punktis 6.3 võrreldakse kahe normaaljaotusega juhusliku suuruse X ja Y dispersioone. Kui sellekohase hüpoteesi kontrollimisest järeldub teatud olulisuse nivool üldkogumite dispersioonide võrdsus $DX = DY$, siis võib nende üldkogumite keskväärtuste võrdlemiseks kasutada järgmist lihtsustatud algoritmi. Kõigepealt arvutatakse valimite dispersioonihinnangute s_x^2 ja s_y^2 ühishinnang

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \tag{7}$$

kui hinnangute s_x^2 ja s_y^2 kaalutud keskmine, kus kaaludeks võetakse valimite vabadusastmete arvud $n-1$ ja $m-1$ ning nimetajas on nende vabadusastmete summa $(n-1)+(m-1) = n+m-2$.

Statistilise kriteeriumi empiirilise väärtuse arvutamiseks kasutatakse valemi (1) asemel järgmist valemit:

$$z_{emp} = \frac{\overline{x} - \overline{y}}{\sqrt{s_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}}. \tag{8}$$

See algoritm sobib eriti hästi valimi väikeste mahtude $n \leq 30$, $m \leq 30$ puhul, võimaldades vältida keeruka valemi (4) kasutamist. Kriitiline punkt z_{kr} leitakse sõltuvalt konkureeriva hüpoteesi H_1 püstitusest kas kujul (5) või (6), kusjuures vajalikuks vabadusastmete arvuks võetakse $k = n + m - 2$.

Suurte valimimahtude puhul ($n > 30$, $m > 30$) saab kriitilise punkti z_{kr} sõltuvalt hüpoteesi H_1 formuleeringust kas seost (2) või (3).

Näide 5. Kasutada näite (4) lahendamiseks eelkirjeldatud lihtsustatud algoritmi.

Arvutame esmalt valemiga (7) dispersioonihinnangute ühishinnangu

$$s_p^2 = \frac{(12-1)669^2 + (15-1)844^2}{12+15-2} = 595835$$

ja seejärel valemiga (8) statistilise kriteeriumi empiirilise väärtuse

$$z_{emp} = \frac{6870 - 6579}{\sqrt{595835\left(\frac{1}{12} + \frac{1}{15}\right)}} \approx 0,91.$$

Tulemus langeb täpselt kokku näites 4 valemiga (1) saaduga. Et ka vabadusastmete arv $k = 12 + 15 - 2 = 25$ on sama kui näites 4 valemiga (4) arvutatud vabadusastmete arv, siis on edasine lahenduskäik ja järeldused identsed näites 4 esitatuga.

Näide 6. Kasutada näite 3 lahendamiseks eelkirjeldatud lihtsustatud algoritmi.

Dispersioonihinnangute ühishinnang tuleb vastavalt valemile (7) $s_p^2 = 0,14$ ning valemiga (8) arvutatud statistilise kriteeriumi empiiriline väärtus on $z_{emp} = -1,24$ (valemiga (1) tuli $z_{emp} = -1,21$). Vabadusastmete arv $k = 20 + 24 - 2 = 42$ on küll mõnevõrra erinev valemiga (4) saadust ($k = 31$), kuid kriitiline punkt $z_{kr}(42;\ 0,975) = 2,018$ on siiski väga lähedane näites 3 leitule ($z_{kr} = 2,04$). Järeldus on sama kui näites 3: kuna $|z_{emp}| < x_{kr}$ $(1,24 < 2,018)$, siis pole alust nullhüpoteesi tagasi lükata, st mõlema firma sigaretid on ühesuguse nikotiinisisaldusega.

Esitatud lahenduskäiku peab siiski suhtuma teatud ettevaatusega, sest tekib küsimus üldkogumite dispersioonide võrdsusest, mida siin on teatavasti eeldatud. Praktika näitab, et valemite (7) ja (8) kasutamine tuleb kõne alla, kui s_x^2 ja s_y^2 erinevad vähem kui kaks korda. Siin on see nõue täidetud, sest $s_x^2/s_y^2 = 1,62 < 2$. Õigem oleks siiski kõigepealt kontrollida hüpoteesi üldkogumite dispersioonide võrdsuse kohta (vt järgmist punkti 6.3) ning alles pärast kinnituse saamist selle väite tõesuse kohta võiks asuda keskväärtusi võrdlema lihtsustatud algoritmiga (7), (8).

6.3. Kahe normaaljaotuse dispersiooni võrdlemine. Fisheri jaotus

Olgu juhuslikud suurused X ja Y normaaljaotusega ning sõltumatud. Olgu valimite $x_1, x_2, \ldots, x_n$ ja $y_1, y_2, \ldots, y_m$ mahud n ja m. Märgime indeksiga 1 valimi, mille dispersioonihinnang on suurem. Niisiis järgnevas on $s_1^2 > s_2^2$ ning k_1 on suurema ja k_2 väiksema dispersioonihinnanguga valimi vabadusastmete arv. Moodustame χ^2-jaotusega sõltumatud juhuslikud suurused (vt punkti 5.6 seost (1))

$$\chi_1^2 = \frac{S_1^2}{\sigma_1^2}k_1 \quad \text{ja} \quad \chi_2^2 = \frac{S_2^2}{\sigma_2^2}k_2,$$

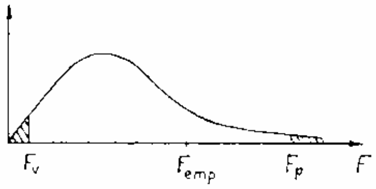
kus $\sigma_1^2 \equiv DX$ ja $\sigma_2^2 \equiv DY$ on üldkogumite dispersioonid ning S_1 ja S_2 on juhuslike suurustena vaadeldud dispersioonide nihutamata hinnangud. Moodustame kahe eelneva juhusliku suuruse abil uue juhusliku suuruse

$$\frac{S_1^2}{S_2^2} = \frac{\chi_1^2\sigma_1^2/k_1}{\chi_2^2\sigma_2^2/k_2}.$$

Kui $\sigma_1^2 = \sigma_2^2$ ehk $DX = DY$ (niisugusel kujul püstitame järgnevas nullhüpoteesi), siis on juhuslik suurus

$$\frac{S_1^2}{S_2^2} = \frac{\chi_1^2/k_1}{\chi_2^2/k_2} = F(k_1, k_2)$$

Fisheri jaotusega e F-jaotusega. Fisheri jaotuse jaotustiheduse graafiku põhimõtteline kuju on järgmine:



Lisas 4 on tabuleeritud Fisheri jaotuse täiendkvantiilid olulisuse nivoodel 0,05 ja 0,01. Kuna Fisheri jaotus sõltub kahest parameetrist k_1 ja k_2, siis on tabelid küllalt mahukad.

Püstitame olulisuse nivool α nullhüpoteesi, et üldkogumite dispersioonid on võrdsed: $H_0 : DX = DY$ (ehk $\sigma_1^2 = \sigma_2^2$). Konkureeriva hüpoteesi erineva formuleerimise võimalustest vaatleme kõigepealt juhust $H_1 : DX \neq DY$ (ehk $\sigma_1^2 \neq \sigma_2^2$).

Nullhüpoteesi kehtivust kontrollitakse järgneva skeemi alusel.

1. Ette antakse olulisuse nivoo α.

2. Fisheri jaotuse täiendkvantiilide tabelist (vt lisa 4) leitakse kriitilised punktid *kahepoolsele kriitilisele piirkonnale* (vt eelnevat joonist, kus Fisheri jaotuse jaotustiheduse graafiku all olevad mõlemad viirutatud pindalad on ühesuurused ning kumbki võrdub $\alpha/2$)):

$$F_p = F(k_1,\ k_2,\ \alpha/2) \text{ ja } F_v = \frac{1}{F(k_2,\ k_1,\ \alpha/2)},$$

kus k_1 on *suurema* dispersioonihinnanguga valimi vabadusastmete arv.

3. Arvutatakse empiirilise kriteeriumi $F_{emp} = \frac{s_1^2}{s_2^2}$ väärtus, kus $s_1^2 > s_2^2$.

4. Kui $F_v < F_{emp} < F_p$, siis pole alust nullhüpoteesi H_0 tagasi lükata. Kui need võrratused ei ole täidetud, lükatakse nullhüpotees tagasi.

Märkus 1. Kuna valimid on indekseeritud nii, et $s_1^2 > s_2^2$, siis on $F_{emp} > 1$ ning alati on täidetud tingimus $F_v < 1 < F_{emp}$. Järelikult võib eelneva algoritmi neljandas punktis asendada võrratused $F_v < F_{emp} < F_p$ võrratusega $F_{emp} < F_p$. Nii on järgnevates näidetes 1 ja 2 ka toimitud.

Näide 1. Kaupluse esimese kassa juures mõõdeti 25 kliendi teenindusaega, teise kassa juures 21 kliendi teenindusaega. Standardhälvete nihutamata hinnangud olid vastavalt 2,5 minutit ja 3,1 minutit. Kontrollida olulisuse nivool 0,02 kahe kassa teenindusaja dispersioonide erinevust.

Nullhüpotees $H_0 : DX = DY$ $(\sigma_1^2 = \sigma_2^2)$.

Konkureeriv hüpotees $H_1 : DX \neq DY$ $(\sigma_1^2 \neq \sigma_2^2)$.

1. Ette antakse olulisuse nivoo $\alpha = 0,02$.

2. Kuna $3,1^2 > 2,5^2$, siis $k_1 = 21 - 1 = 20$ ja $k_2 = 25 - 1 = 24$ ning kriitiline punkt $F_p = F(20;\ 24;\ 0,01) = 2,74$.

3. Empiiriline kriteerium $F_{emp} = \frac{s_1^2}{s_2^2} = \frac{3,1^2}{2,5^2} = 1,54$.

4. Kuna $F_{emp} < F_p$ $(1,54 < 2,74)$, siis pole alust nullhüpoteesi H_0 tagasi lükata, st üldkogumite dispersioonid ei erine teineteisest.

Näide 2. Kontrollida eelmise punkti näite 2 andmetel nullhüpoteesi dispersioonide erinevuse kohta olulisuse nivool 0,10.

Nullhüpotees $H_0 : DX = DY$ $(\sigma_1^2 = \sigma_2^2)$.

Konkureeriv hüpotees $H_1 : DX \neq DY$ $(\sigma_1^2 \neq \sigma_2^2)$.

1. Ette antakse olulisuse nivoo $\alpha = 0,10$.

2. Kuna $2,1^2 > 1,5^2$, siis $k_1 = 40 - 1 = 39$ ja $k_2 = 50 - 1 = 49$ ning kriitiline punkt $F_p = F(39;\ 49;\ 0,05) \approx 1,63$.

3. Empiiriline kriteerium $F_{emp} = \frac{s_1^2}{s_2^2} = \frac{2,1^2}{1,5^2} \approx 1,96$.

4. Kuna $F_{emp} > F_p$ $(1,96 > 1,63)$, siis lükatakse nullhüpotees H_0 tagasi, st üldkogumite dispersioonid erinevad teineteisest oluliselt.

Märkus 2. Olgu konkureeriv hüpotees püstitatud kujul $H_1: DX > DY$ $(\sigma_1^2 > \sigma_2^2)$. Siis kontrollitakse nullhüpoteesi $H_0: DX = DY$ $(\sigma_1^2 = \sigma_2^2)$ kehtivust järgmiselt.

1. Ette antakse olulisuse nivoo α.

2. Kasutatakse *parempoolset kriitilist piirkonda*

$$P[F > F_{kr}(k_1,\ k_2,\ \alpha)] = \alpha,$$

mille alusel on kriitiliseks punktiks $F_{kr}(k_1,\ k_2,\ \alpha)$.

3. Empiiriline kriteerium arvutatakse, nagu eespoolgi, valemiga $F_{emp} = \frac{s_1^2}{s_2^2}$.

4. Kui $F_{emp} > F_{kr}$, siis lükatakse nullhüpotees H_0 tagasi. Kui $F_{emp} < F_{kr}$, siis pole alust nullhüpoteesi H_0 tagasi lükata.

Näide 3. Kontrollida eelmise punkti näite 1 andmetel olulisuse nivool $\alpha = 0,05$ väidet, et ametiühinguliikmete puudumiste dispersioon on suurem kui ametiühingusse mittekuulujate puudumiste dispersioon.

Nullhüpotees $H_0: DX = DY$ $(\sigma_1^2 = \sigma_2^2)$.

Konkureeriv hüpotees $H_1: DX > DY$ $(\sigma_1^2 > \sigma_2^2)$.

1. Olulisuse nivoo $\alpha = 0,05$.

2. Kuna $3,1^2 > 2,3^2$, siis $k_1 = 50 - 1 = 49$ ja $k_2 = 45 - 1 = 44$ ning kriitiline punkt $F_{kr}(49;\ 44;\ 0,05) \approx 1,64$.

3. Empiiriline kriteerium $F_{emp} = \frac{s_1^2}{s_2^2} = \frac{3,1^2}{2,3^2} = 1,817$.

4. Kuna $F_{emp} > F_{kr}$ $(1,817 > 1,64)$, siis lükatakse nullhüpotees H_0 tagasi, st ametiühinguliikmete puudumiste dispersioon on oluliselt suurem kui ametiühingusse mittekuulujatel.

Näide 4. Kontrollida eelmise punkti näite 3 andmetel olulisuse nivool $\alpha = 0,05$ väidet, et esimese firma sigarettide nikotiinisisalduse dispersioon on suurem.

Nullhüpotees $H_0: DX = DY$ $(\sigma_1^2 = \sigma_2^2)$.

Konkureeriv hüpotees $H_1: DX > DY$ $(\sigma_1^2 > \sigma_2^2)$.

1. Olulisuse nivoo $\alpha = 0,05$.

2. Kuna $0,42^2 > 0,33^2$, siis $k_1 = 20 - 1 = 19$ ja $k_2 = 24 - 1 = 23$ ning kriitiline punkt $F_{kr}(19;\ 23;\ 0,05) \approx 2,07$.

3. Empiiriline kriteerium $F_{emp} = \frac{s_1^2}{s_2^2} = \frac{0,42^2}{0,33^2} = 1,62$.

4. Kuna $F_{emp} < F_{kr}$ $(1,62 < 2,07)$, siis pole alust nullhüpoteesi tagasi lükata ning mõlema firma sigarettide nikotiinisisalduse dispersioon tuleb vaadeldud valimite andmetest lähtudes lugeda võrdseks. Kui selles tekib kahtlusi, siis tuleks teha täiendavaid uuringuid, suurendades valimite mahtu. Antud valimite põhjal saadud tulemuste põhjal saab ühtlasi siiski järeldada, eelmise punkti 6.2 näites 6 on keskväärtuste võrdlemine lihtsustatud algoritmiga õigustatud, kuna selleks vajalik tingimus $DX = DY$ on täidetud.

6.4. Ülesanded

6.1. Kahes linnas uuriti sissetulekuid elaniku kohta ning saadi järgmised tulemused:

$$n = 1200, \quad \bar{x} = 2803 \text{ krooni}, \quad s_x = 416 \text{ krooni};$$
$$m = 1000, \quad \bar{y} = 2724 \text{ krooni}, \quad s_y = 364 \text{ krooni}.$$

Kontrollida väidet nende linnade elanike sissetulekute erinevuse kohta olulisuse nivool $\alpha = 0,01$. Kontrollida väidet dispersioonide erinevuse kohta okulisuse nivool $\alpha = 0,02$.

6.2. Katsetati kahe firma elektripirne ning saadi järgmised tulemused:

$$n = 20, \quad \bar{x} = 980 \text{ tundi}, \quad s_x = 85,4 \text{ tundi};$$
$$m = 20, \quad \bar{y} = 1015 \text{ tundi}, \quad s_y = 104,7 \text{ tundi}.$$

Kontrollida väidet nende firmade pirnide keskmise tööea erinevuse kohta olulisuse nivool $\alpha = 0,05$. Kontrollida olulisuse nivool $\alpha = 0,10$ dispersioonide erinevust.

6.3. Programmeerijate tööd hõlbustava programmi LINC efektiivsuse selgitamiseks võrreldi kahe programmeerijate rühma tööd ühe ja sama ülesande lahendamisel, kusjuures esimene rühm ei kasutanud programmi LINC. Tulemused olid järgmised:

$$n = 45, \quad \bar{x} = 26 \text{ min}, \quad s_x = 8 \text{ min};$$
$$m = 32, \quad \bar{y} = 21 \text{ min}, \quad s_y = 6 \text{ min}.$$

Kontrollida olulisuse nivool $\alpha = 0,01$ väidet, et LINC lühendab programmeerimisaega. Kontrollida olulisuse nivoodel $\alpha = 0,01$ ja $\alpha = 0,05$ väidet, et programmi LINC kasutajate lahendusaja dispersioon on väiksem.

6.4. Statistika eksamil said meesüliõpilased 72, 60, 98, 66, 85, 76, 79, 80 ja 68 punkti ning naisüliõpilased 81, 67, 90, 78, 81, 87 ja 76 punkti sajast võimalikust. Kontrollida olulisuse nivool $\alpha = 0,05$ väiteid, et 1) naiste keskmine hinne on kõrgem; 2) meeste hinnete hajuvus on suurem.

6.5. Firma kogus andmeid kaugekõnede pikkuse kohta müügiosakonnas ja teenindusosakonnas ning sai järgmised tulemused:

$$n = 40, \quad \bar{x} = 10,56 \text{ min}, \quad s_x = 8,15 \text{ min};$$
$$m = 25, \quad \bar{y} = 6,93 \text{ min}, \quad s_y = 5,43 \text{ min}.$$

Kontrollida (kasutades normaaljaotust) olulisuse nivoodel $\alpha = 0,05$ ja $\alpha = 0,02$ väidet, et mõlemas osakonnas on kaugekõne keskmine pikkus võrdne. Kontrollida olulisuse nivool $\alpha = 0,02$ väidet dispersioonide võrdsuse kohta.

6.6. Kontrollida olulisuse nivoodel $\alpha = 0,05$ ja $\alpha = 0,10$ väidet, et kahe automudeli keskmine kütusekulu on erinev.

$$n = 10, \quad \bar{x} = 7,05 \text{ [l/100 km]}, \quad s_x = 1,25;$$
$$m = 12, \quad \bar{y} = 7,90, \quad s_y = 0,85.$$

Kontrollida olulisuse nivool $\alpha = 0,10$ väidet dispersioonide erinevuse kohta.

6.7. Kuritegude arv suurlinna kaheksas juhuslikult valitud piirkonnas oli aasta jooksul enne naabrusvalveprogrammi käivitamist 14, 7, 4, 5, 17, 12, 8, 9 ja pärast naabrusvalveprogrammi käivitamist 2, 7, 3, 6, 8, 13, 3, 5. Kontrollida olulisuse nivoodel $\alpha = 0,01$ ja $\alpha = 0,05$ väidet, et naabrusvalveprogramm aitas kaasa kuritegevuse vähendamisele.

Vastused

1.1. $\overline{x} = 15,85$; $M_0X = 16$; $M_eX = 16$; $s = 4,46$ ja $\overline{x} = 19,65$; $M_0X = 16$; $M_eX = 16$; $s = 19,34$. **1.2.** 83,2. **1.3.** 48. **1.4.** 10,91865. **1.5** 43,1% ja 43,7%. **1.6.** $\overline{x} = 7,58$; $\sqrt{D^*X} = 2,65$ ja $\overline{x} = 7,65$; $\sqrt{D^*X} = 2,53$. **1.7.** $\overline{x} = 9,30$; $s = 18,92$ ja $\overline{x} = 9,33$; $s = 20,67$. **1.10.** 10,99%. **1.11.** I_Q=121,26%, I_{L_h}=107,60%, I_{P_h}=107,75%, I_{L_m}=112,54%, I_{P_m}=112,69%.

2.1. 1) $A = A_1A_2A_3A_4$; 2) $B = A_1 + A_2 + A_3 + A_4$; 3) $C = \overline{A}_1 + +\overline{A}_2 + \overline{A}_3 + \overline{A}_4$; 4) $D = \overline{A}_1A_2A_3A_4 + A_1\overline{A}_2A_3A_4 + A_1A_2\overline{A}_3A_4 + A_1A_2A_3\overline{A}_4$. **2.2.** 0,25. **2.3.** 1) 0,5; 2) 0,1. **2.4.** 54/115; 1/230. **2.5.** 468. **2.6.** 24. **2.7.** 1) $4{,}69\cdot10^{-9}$; 2) $3{,}38\cdot10^{-6}$. **2.8.** 0,3974. **2.9.** 1) 0,024; 2) 0,0007; 3) 0,358. **2.10.** 1) 10/19; 2) 9/19. **2.11.** 5040; 4; 1/1260. **2.12.** 1/181440. **2.13.** 1/6. **2.14.** 0,03. **2.15.** 0,392. **2.16.** 1/420. **2.17.** 0,9125. **2.18** 0,91. **2.19.** 1) 0,56; 2) 0,06; 3) 0,38; 4) 0,94. **2.20.** 11/36. **2.21.** 1) 0,684; 2) 0,283; 3) 0,316. **2.22.** 1) 0,04; 2) 0,02; 3) 0,22. **2.23.** 1) 0,0186; 2) 0,32. **2.24.** 0,836. **2.25.** 10/57; 21/57; 16/57; 10/57. **2.26.** 0,467. **2.27.** 0. 6; 0,94. **2.28.** 1/8. **2.29.** 0,9477. **2.30.** 0,92455. **2.31.** 1) 0,21; 2) 0,93; 3) 6. **2.32.** Neli seitsmest; vähemalt viis kaheksast. **2.33.** 4; 0,2365. **2.34.** 7 ja 8. **2.35.** 5.

4.1. $EX = 2,5$; $DX = 1,25$; $\sigma(X) \approx 1,118$. **4.2.** $EX = 1,2$; $M_0X = 1$; $DX = 0,72$; $\sigma(X) \approx 0,8485$. **4.3.** $EX = 1,2$; $M_0X = 1$; $DX = 0,96$; $\sigma(X) \approx 0,98$. **4.4.** 0,4096; 3 või 4; $EX = 3,2$; $M_0X = 3$ või 4; $DX = 0,64$; $\sigma(X) = 0,8$. **4.5.** 0,0047. **4.6.** 0,88. **4.7.** 1)0,442; 2) 0,9988. **4.8** 1) 0,218; 2) 0,533. **4.9.** 0,087. **4.10.** 85,2%. **4.11.** Sotsiaalse vastutustunde poolest on kõrgema 1% hulgas; rehabilitatsiooni perspektiivi poolest on madalama 5% hulgas. **4.12.** 0,89 **4.13.** 1) 0,758; 2) 0,01163. **4.14.** 0,725. **4.15.** 0,81859. **4.16.** 0,11. **4.17.** 0,18. **4.18.** 0,02. **4.19.** 0,046. **4.20.** 0,085. **4.21.** 0,7242. **4.22.** 0,9545. **4.23.** 1) 0,77; 2) 0,94. **4.24.** 0,831. **4.25.** 0,18. **4.26.** 0,2327. **4.27.** 1)0,7887; 2) 0,7340.

5.1. $\overline{x} \approx 76,21$; $s \approx 4,06$; keskväärtuse usalduspiirkond $]75,6;\ 76,9[$. **5.2.** $\overline{x} \approx 286,545$; $s^2 \approx 109,83$; $s \approx 10,48$; keskväärtuse usalduspiirkond $]284,7;\ 288,4[$; dispersiooni usalduspiirkond]87; 144[; standardhälbe usalduspiirkond $]9,3;\ 12,0[$. **5.3.** $\overline{x} \approx 38,03$; $s^2 \approx 56,085$; $s \approx 7,489$; keskväärtuse usalduspiirkond $]36,44;\ 39,63[$; dispersiooni usalduspiirkond $]42,1;\ 76,6[$; standardhälbe usalduspiirkond $]6,5;\ 8,8[$. **5.4.**]3,12; 5,08[. **5.5.**]489; 501[. **5.6.** 1)]41,0; 49,2[ja]41,6; 48,6[, 2)]15,3; 107,7[ja]3,9; 10,4[. **5.7.**]68,2; 81,2[ja]10,6; 20,4[. **5.8.**]75,6; 138,2[ja]39,4; 86,9[. **5.9.**]13,25; 19,19[,]3,44; 8,19[. **5.10.**]7,7; 12,0[ja]2,0; 5,6[. **5.11.**]0,535; 0,623[; 733. **5.12.**]0,243; 0,311[; 1924. **5.13.**]0,00181; 0,0159[.

6.1. $4,75 > 2,575$ ($z_{emp} > z_{kr}$) ja H_0 lükatakse tagasi. **6.2.** 1) $|-1,158| < 2,028$ ($|z_{emp}| < z_{kr}$) ja pole alust H_0 tagasi lükata; 2) $0,46 < 1,51 < 2,16$ ($F_v < F_{emp} < F_p$) ja pole alust H_0 tagasi lükata. **6.3.** 1) $3,13 > 2,33$ ($z_{emp} > z_{kr}$) ja H_0 lükatakse tagasi (LINC lühendab programmeerimisaega); 2a) $1,78 < 2,26$ ($F_{emp} < F_{kr}$) ja pole alust H_0 tagasi lükata; 2b) ($F_{emp} \approx F_{kr}$) ja ollakse H_0 tagasilükkamise piiril. **6.4.** 1) $0,846 < 1,771$ ($z_{emp} > z_{kr}$) ja keskmised hinded on võrdsed; 2) $2,26 < 4,15$ ($F_{emp} < F_{kr}$) ja hinnete hajuvus on võrdne. **6.5** 1a) $2,154 > 1,96$ ($z_{emp} > z_{kr}$) ja H_0 lükatakse tagasi; 1b) $2,154 < 2,33$ ($z_{emp} < z_{kr}$) ja pole alust H_0 tagasi lükata; 2) $0,437 < 2,253 < 2,49$ ($F_v < F_{emp} < F_p$) ja pole alust H_0 tagasi lükata. **6.6.** 1) $\alpha = 0,05 : |-1,85| < 2,145$ ($|z_{emp}| < z_{kr}$) ja kütusekulu on võrdne; $\alpha = 0,01 : |-1,85| > 1,761$ ($|z_{emp}| > z_{kr}$) ja kütusekulu erineb; 2) $2,38 < 2,90$ ($F_{emp} < F_p$) ja dispersioonid on võrdsed. **6.7.** $\alpha = 0,01 : 1,79 < 2,650$ ($z_{emp} < z_{kr}$) ja programm on ebaefektiivne; $\alpha = 0,05 : 1,79 > 1,771$ ($z_{emp} > z_{kr}$) ja programm on efektiivne.

Kasutatud kirjandus

1. Aarma, A., Vensel, V. Statistika teooria põhikursus. Külim, 1996.
2. Aczel, A.D. Complete Business Statistics. Irwin, 1989.
3. Allik, I., Ruubel, K. Tõenäosusteooria ülesanded. Tallinn, TTÜ, 1992.
4. Allik, I., Roos, J. Tõenäosusteooria ja matemaatilise statistika ülesannete kogu. Tallinn, TPI, 1983.
5. Aruküla, H. Abiks majandusmatemaatika õppijaile VI. Matemaatiline statistika. Tallinn, TPI, 1982.
6. Gurski, J. Tõenäosusteooria ja matemaatilise statistika elemendid. Tallinn, Valgus, 1986.
7. Hanke, J.E. Understanding Business Statistics. Irwin, 1991.
8. Kiviste, A. Matemaatiline statistika MS Exceli keskkonnas. Gensi Tarkvara, 1999.
9. Käerdi, H. Statistika ja tõenäosusteooria alused. Teine, täiendatud trükk. Tallinn, Sisekaitseakadeemia, 1999.
10. Käerdi, H. Statistika ja tõenäosusteooria alused. Tallinn, Eesti Riigikaitse Akadeemia, 1997.
11. Käerdi, H. Statistika ja tõenäosusteooria majanduserialadele. Tallinn, Eesti Kõrgem Kommertskool. Esimene trükk 1994. Teine, täiendatud trükk 1995. Kolmas, täiendatud trükk 1996. Neljas, täiendatud trükk 1997. Viies, täiendatud trükk 1998.
12. Listra, E. Äristatistika I. Tallinn, TTÜ, 1998.
13. Lucey, T. Quantitative Techniques. Third Edition. London, DP publications LTD, 1989.
14. Lõhmus, A., Petersen, I., Roos, H. Kõrgema matemaatika ülesannete kogu. Tallinn. Valgus, 1982.
15. Matre, J.G., Gilbreath, G.H. Statistics for Business and Economics. Third Edition. Homewood, Illinois, BPI Irwin, 1987.
16. Newbold, P. Statistics for Business and Economics. Second Edition. Prentice Hall, 1988.
17. Starnbacher, K. Statistik im Betrieb. 8. Aufl., Wiesbaden, Gabler, 1991.
18. Storm, R. Wahrscheinlichkeisrechnung, mathematische Statistik und statistische Qualitätskontrolle. 9. Aufl., Leipzig, Fachbuchverl., 1988.
19. Sõrmus, I. Abiks majandusmatemaatika õppijaile V. Tõenäosusteooria. Tallinn, TPI, 1980.
20. Tamm, V. Statistika baasmõisted ja -meetodid majanduses. TÜ, 1996.
21. Tiit, E.-M., Möls, M. Rakendusstatistika lühikursus. Eesti Statistikaselts, Tartu, 1997.
22. Tiit, E., Parring, A., Möls, T. Tõenäosusteooria ja matemaatiline statistika. Tallinn, Valgus, 1977.
23. Tooding, L.-M. Andmeanalüüs sotsiaalteadustes. TÜ, 1998.
24. Triola, M.F. Elementary Statistics. Fourth Edition. Redwood City, CA, Benjamin/Cummings, 1989.
24. Täht, T. Statistika. Eesti Kõrgem Kommertskool, 1998.

LAPLACE'I FUNKTSIOONI $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x \exp\left(-\frac{t^2}{2}\right) dt$ TABEL

x	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0.1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0.2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0.3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0.4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0.5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0.6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0.7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0.8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0.9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1.0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1.1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1.2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1.3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1.4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1.5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1.6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1.7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1.8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1.9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2.0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2.1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2.2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2.3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2.4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2.5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2.6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2.8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2.9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3.0	0.49865	3.1	49903	3.2	49931	3.3	49952	3.4	49966	
3.5	49977	3.6	49984	3.7	49989	3.8	49993	3.9	49995	
4.0	499968									
4.5	499997									
5.0	500000									

STUDENTI JAOTUSE KVANTIILIDE $t_{k;\alpha}$ TABEL

	α $\{\ \alpha = (1+\beta)/2,$ kus β on usaldusnivoo $\}$							
k	0.80	0,90	0,95	0,975	0.990	0.995	0.9975	0.999
1	1.376	3.078	6,314	12,71	31.82	63.66	127,3	318.3
2	1.061	1.886	2.920	4.303	6,965	9.925	14.09	22.33
3	0.978	1.638	2.353	3.182	4,541	5.841	7.453	10.22
4	0.941	1.533	2.132	2.776	3.747	4.604	5,598	7.173
5	0.920	1.476	2.015	2.571	3.365	4.032	4,773	5.893
6	0,906	1,440	1.943	2.447	3.143	3.707	4.317	5.208
7	0.896	1.415	1.895	2.365	2,998	3.499	4.029	4.785
8	0,889	1.397	1.859	2.306	2,896	3.355	3.832	4.501
9	0,883	1.383	1.833	2,262	2.821	3.250	3.690	4.297
10	0.879	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	0,876	1.363	1.796	2.201	2,718	3.106	3.497	4.025
12	0,873	1.356	1.782	2,179	2.681	3,054	3.428	3.930
13	0,870	1,350	1.771	2,160	2,650	3.012	3.372	3.852
14	0.868	1.345	1.761	2.145	2,624	2.977	3.326	3.787
15	0,866	1.341	1.753	2.131	2,602	2.947	3.286	3.733
16	0,865	1.337	1.746	2,120	2.583	2.921	3.252	3.686
17	0,863	1,333	1.740	2,110	2,567	2,898	3.222	3.646
18	0,862	1.330	1,734	2,101	2,552	2,878	3.197	3.611
19	0,861	1.328	1.729	2.093	2.539	2,861	3.174	3.579
20	0,860	1.325	1.725	2.086	2,528	2,845	3.153	3.552
21	0.859	1.323	1,721	2,080	2.518	2,831	3.135	3.527
22	0,858	1.321	1.717	2.074	2.508	2,819	3.119	3.505
23	0.858	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	0.857	1.318	1.711	2,064	2.492	2.797	3.090	3.467
25	0.856	1,316	1,708	2,060	2,485	2.787	3.078	3.450
26	0.856	1,315	1,706	2,056	2.479	2.779	3.067	3.435
27	0,855	1.314	1.703	2,052	2.473	2,771	3.056	3.421
28	0,855	1.313	1.701	2.048	2,467	2,763	3.047	3.408
29	0.854	1,311	1,699	2.045	2.462	2,756	3.038	3.396
30	0.854	1.310	1,697	2,042	2,457	2.750	3.030	3.385

STUDENTI JAOTUSE KVANTIILIDE $t_{k;\alpha}$ TABEL

	α $\{ \alpha = (1+\beta)/2$, kus β on usaldusnivoo $\}$							
k	0.80	0.90	0.95	0.975	0.990	0.995	0.9975	0.999
32	0.853	1.309	1.694	2.037	2.449	2.738	3.015	3.365
34	0.853	1.307	1.691	2.032	2.441	2.728	3.002	3.348
36	0.852	1.305	1.688	2.028	2.434	2.719	2.990	3.333
38	0.852	1.304	1.686	2.024	2.429	2.712	2.980	3.319
40	0.851	1.303	1.684	2.021	2.423	2.704	2.971	3.307
42	0.851	1.302	1.682	2.018	2.418	2.698	2.963	3.296
44	0.850	1.301	1.680	2.015	2.414	2.692	2.955	3.286
46	0.850	1.300	1.679	2.013	2.410	2.687	2.949	3.277
48	0.849	1.299	1.677	2.011	2.407	2.682	2.948	3.269
50	0.849	1.298	1.676	2.009	2.403	2.678	2.937	3.261
55	0.848	1.297	1.673	2.004	2.396	2.668	2.925	3.256
60	0.848	1.296	1.671	2.000	2.390	2.660	2.915	3.232
65	0.847	1.295	1.669	1.997	2.385	2.654	2.906	3.220
70	0.847	1.294	1.667	1.994	2.381	2.648	2.899	3.211
80	0.846	1.292	1.664	1.990	2.374	2.639	2.887	3.195
90	0.845	1.291	1.662	1.987	2.368	2.632	2.878	3.183
100	0.845	1.290	1.660	1.984	2.365	2.626	2.871	3.174
120	0.844	1.289	1.658	1.980	2.358	2.617	2.860	3.159
150	0.844	1.287	1.655	1.976	2.351	2.609	2.849	3.145
200	0.843	1.286	1.653	1.972	2.345	2.601	2.838	3.131
250	0.843	1.285	1.651	1.169	2.341	2.596	2.832	3.123
300	0.842	1.284	1.650	1.968	2.339	2.592	2.828	3.118
400	0.842	1.284	1.649	1.966	2.336	2.588	2.823	3.111
500	0.842	1.283	1.648	1.965	2.334	2.586	2.819	3.107

NORMAALJAOTUSE KVANTIILID

∞	0.842	1.282	1.645	1.960	2.326	2.756	2.807	3.090

χ^2-JAOTUSE TÄIENDKVANTIILIDE $\chi^2(\alpha, k)$ TABEL

	α									
k	0.995	0.99	0.975	0,95	0,90	0.10	0,05	0,025	0.01	0,005
1			0,001	0,004	0,016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7,378	9.210	10.597
3	0.072	0.115	0.216	0.352	0,584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0,711	1.064	7,779	9.488	11.143	13,277	14.860
5	0.412	0.554	0.831	1.145	1,610	9.236	11.071	12,833	15.086	16.750
6	0.676	0,872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0,989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18,475	20.278
8	1.344	1.646	2.180	2.733	3,490	13.362	15,507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14,684	16.919	19,023	21,666	23.589
10	2,156	2,558	3.247	3,940	4,865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4,575	5,578	17,275	19,675	21.920	24.725	26.757
12	3.074	3.571	4.404	5,226	6.304	18,549	21,026	23,337	26.217	28.299
13	3.565	4,107	5.009	5.892	7,042	19,812	22.362	24.736	27,688	29.819
14	4.075	4,660	5,629	6.571	7.790	21,064	23.685	26,119	29,141	31.319
15	4.601	5.229	6.262	7.261	8,547	22,307	24,996	27.488	30.578	32.801
16	5.142	5.812	6.908	7,962	9.312	23,542	26,296	28.845	32.000	34.267
17	5.697	6.408	7.564	8,672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9,390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27,204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
22	8,643	9.542	10.982	12.338	14,042	30,813	33.924	36.781	40.289	42.796
24	9.886	10,856	12.401	13,848	15.659	33,196	36.415	39.364	42,980	45.559
26	11.160	12.198	13,844	15,379	17.292	35,563	38,885	41,923	45.642	48.290
28	12.461	13.565	15.308	16.928	18.939	37,916	41.337	44,461	48.278	50.993
30	13.787	14.954	16.791	18.493	20,599	40.256	43.773	46,979	50,892	53.672
40	20.707	22.164	24.433	26.509	29,051	51,805	55.758	59,342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.43	104.22
80	51.172	53.540	57.153	60.391	64.278	96,578	101.88	106.63	112.33	116.32
90	59.196	61.754	65.647	69,126	73,291	107.57	113.15	118.14	124.12	128.30
100	67.328	70.065	74.222	77,929	82.358	118.50	124.34	129.56	135.81	140.17
110	75.550	78.458	82.867	86,792	91,471	129.39	135.48	140.92	147,41	151.95
120	83.852	86.923	91.573	95.705	100.62	140.23	146.57	152.21	158.95	163.65
130	92.222	95.451	100.33	104.66	109,81	151.05	157.61	163.45	170,42	175.28
140	100.66	104.03	109,14	113.66	119.03	161.83	168.61	174.65	181.84	186.85
150	109.14	112.67	117.99	122.69	128.28	172.58	179.58	185.80	193.21	198.36
200	152.24	156.43	162,73	168.28	174.84	226.02	233.99	241.06	249.45	255.26
500	422.30	429.39	439.94	449.15	459.93	540.93	553.13	563.85	576.49	585.21

FISHERI JAOTUSE TÄIENDKVANTIILIDE $F(k_1, k_2, \alpha)$ TABEL

olulisuse nivool $\alpha = 0,05$ (ülemine rida) ja $\alpha = 0,01$ (alumine rida)

	Suurema dispersioonihinnanguga valimi vabadusastmete arv k_1										
k_2	1	2	3	4	5	6	7	8	9	10	12
1	161 4052	200 4999	216 5403	225 5625	230 5764	234 5859	237 5928	239 5981	241 6022	242 6056	244 6106
2	18.51 98.50	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.35 99.36	19,37 99,37	19,38 99,39	19.39 99.40	19.41 99.42
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.89 27.67	8.85 27,49	8.81 27.34	8.79 27.23	8.74 27.05
4	7.71 21.20	6.94 18.00	6.59 16,69	6.39 15.98	6,26 15.52	6.16 15.21	6.09 14.98	6.04 14,80	6.00 14,66	5.96 14.55	5.91 14.37
5	6.61 16.26	5.79 13,27	5.41 12.06	5.19 11.39	5,05 10.97	4.95 10.67	4.88 10.46	4.82 10,29	4.77 10.16	4.74 10.05	4.68 9.89
6	5.99 13.74	5.14 10,92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4,06 7.87	4.00 7,72
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.72	3.64 6.62	3.57 6,47
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.18	3.44 6.03	3.39 5.91	3.35 5.81	3.28 5,67
9	5.12 10.56	4.26 8.02	3,86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.61	3.23 5.47	3.18 5.35	3.14 5.26	3.07 5,11
10	4.96 10.04	4.10 7.56	3,71 6.55	3.48 5,99	3.33 5,64	3.22 5.39	3.14 5.20	3.07 5.06	3.02 4.94	2,98 4.85	2,91 4.71
11	4.84 9.65	3.98 7.21	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.89	2.95 4.74	2.90 4.63	2.85 4.54	2.79 4.40
12	4.75 9.33	3.89 6.93	3.49 5.95	3.26 5,41	3.11 5.06	3.00 4.82	2.91 4.64	2.85 4.50	2.80 4.39	2.75 4.30	2,69 4.16
13	4.67 9.07	3.81 6.70	3.41 5.74	3.18 5.21	3.03 4,86	2.92 4.62	2.83 4.44	2.77 4.30	2.71 4.19	2.67 4.10	2.60 3.96
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.04	2.96 4,70	2.85 4,46	2.76 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.53 3.80
15	4.54 8.68	3.68 6.36	3.29 5,42	3.06 4.89	2,90 4.56	2,79 4.32	2.71 4,14	2.64 4.00	2.59 3.89	2.54 3.80	2.48 3.67
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4,77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.42 3.55
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4,67	2,81 4.34	2.70 4.10	2.61 3.93	2.55 3.79	2.49 3.68	2.45 3.59	2.38 3.46

FISHERI JAOTUSE TÄIENDKVANTIILIDE $F(k_1, k_2, \alpha)$ TABEL

olulisuse nivool $\alpha = 0.05$ (ülemine rida) ja $\alpha = 0.01$ (alumine rida)

	Suurema dispersioonihinnanguga valimi vabadusastmete arv k_1										
k_2	14	16	20	24	30	40	50	75	100	200	∞
1	245 6143	246 6169	248 6209	249 6235	250 6261	251 6287	252 6302	253 6323	253 6334	254 6352	254 6366
2	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.47	19.48 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50
3	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.59 26.41	8.58 26.35	8.57 26.27	8.55 26.23	8.54 26.18	8.53 26.12
4	5.87 14.25	5.84 14.15	5.80 14.02	5.77 13.93	5.75 13.84	5.72 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.63 13.46
5	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.41 9.13	4.39 9.08	4.36 9.02
6	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.93	3.67 6.88
7	3.53 6.36	3.49 6.27	3.44 6.16	3.41 6.07	3.38 5.99	3.34 5.91	3.32 5.86	3.29 5.78	3.27 5.75	3.25 5.70	3.23 5.65
8	3.24 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.12	3.02 5.07	3.00 5.00	2.97 4.96	2.95 4.91	2.93 4.86
9	3.03 5.00	2.99 4.92	2.93 4.81	2.90 4.73	2.86 4.65	2.83 4.57	2.80 4.52	2.77 4.45	2.76 4.42	2.73 4.36	2.71 4.31
10	2.86 4.60	2.83 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.66 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.54 3.91
11	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.51 3.81	2.47 3.74	2.46 3.71	2.43 3.66	2.40 3.60
12	2.64 4.05	2.60 3.97	2.54 3.86	2.51 3.78	2.47 3.70	2.43 3.62	2.40 3.57	2.36 3.49	2.35 3.47	2.32 3.41	2.30 3.36
13	2.55 3.86	2.51 3.78	2.46 3.66	2.42 3.59	2.38 3.51	2.34 3.43	2.31 3.38	2.28 3.30	2.26 3.27	2.23 3.22	2.21 3.17
14	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.35	2.27 3.27	2.24 3.22	2.21 3.14	2.19 3.11	2.16 3.06	2.13 3.00
15	2.42 3.56	2.38 3.49	2.33 3.37	2.29 3.29	2.25 3.21	2.20 3.13	2.18 3.08	2.15 3.00	2.12 2.98	2.10 2.92	2.07 2.87
16	2.37 3.45	2.33 3.37	2.28 3.26	2.24 3.18	2.19 3.10	2.15 3.02	2.12 2.97	2.09 2.89	2.07 2.86	2.04 2.81	2.01 2.75
17	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.10 2.92	2.08 2.87	2.04 2.79	2.02 2.76	1.99 2.71	1.96 2.65

FISHERI JAOTUSE TÄIENDKVANTIILIDE $F(k_1, k_2, \alpha)$ TABEL

olulisuse nivool $\alpha = 0,05$ (ülemine rida) ja $\alpha = 0,01$ (alumine rida)

	Suurema dispersioonihinnanguga valimi vabadusastmete arv k_1										
k_2	1	2	3	4	5	6	7	8	9	10	12
18	4.41	3.55	3.16	2.93	2.77	2.66	2,58	2.51	2.46	2.41	2.34
	8.29	6.01	5,09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28
	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23
21	4.32	3.47	3.07	2,84	2.68	2.57	2.49	2.42	2.37	2.32	2.25
	8.02	5.78	4,87	4.37	4.04	3.81	3,64	3,51	3,40	3,31	3.17
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23
	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2,19	2.12
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3,23	3.12	3.03	2.90
30	4.17	3.32	2,92	2.69	2,53	2,42	2,33	2,27	2.21	2.16	2.09
	7.56	5.39	4.51	4.02	3,70	3.47	3.30	3,17	3.07	2.98	2.84
34	4,13	3,28	2.88	2.65	2.49	2.38	2.29	2,23	2.17	2.12	2.05
	7.44	5.29	4,42	3,93	3.61	3.39	3,22	3.09	2.98	2.89	2.76
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95
	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2,89	2.79	2.70	2.56
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2,07	2,02	1,97	1.89
	7.01	4.92	4.08	3.60	3.29	3.07	2,91	2.78	2.67	2.59	2.45
100	3.94	3.09	2,70	2.46	2,31	2,19	2,10	2.03	1.97	1.93	1.85
	6,90	4.82	3.98	3.51	3,21	2.99	2,82	2,69	2.59	2.50	2,37
200	3,89	3.04	2,65	2,42	2.26	2.14	2.06	1,98	1.93	1,88	1.80
	6.76	4,71	3.88	3,41	3.11	2.89	2.73	2.60	2.50	2.41	2.27
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75
	6.63	4.61	3.78	3,32	3.02	2.80	2.64	2.51	2.41	2.32	2.18

FISHERI JAOTUSE TÄIENDKVANTIILIDE $F(k_1, k_2, \alpha)$ TABEL

olulisuse nivool $\alpha = 0,05$ (ülemine rida) ja $\alpha = 0.01$ (alumine rida)

	Suurema dispersioonihinnanguga valimi vabadusastmete arv k_1										
k_2	14	16	20	24	30	40	50	75	100	200	∞
18	2.29	2.25	2,19	2.15	2.11	2.06	2,04	2.00	1.98	1.95	1,92
	3.27	3.19	3,08	3,00	2.92	2,84	2.78	2.71	2.68	2.62	2.57
19	2.26	2.21	2.15	2.11	2.07	2.03	2.00	1.96	1.94	1.91	1.88
	3.19	3.12	3,00	2.92	2.84	2.76	2.71	2.63	2.60	2.55	2.49
20	2,22	2.18	2,12	2,08	2.04	1,99	1.97	1.92	1,91	1.88	1.84
	3.13	3.05	2,94	2,86	2.78	2,69	2.64	2.56	2.54	2.48	2.42
21	2.20	2.16	2,10	2.05	2.01	1.96	1.94	1.89	1.88	1.84	1.81
	3.07	2.99	2,88	2.80	2.72	2,64	2.58	2.51	2.48	2.42	2.36
22	2,17	2.13	2,07	2,03	1.98	1.94	1,91	1.87	1.85	1.81	1.78
	3.02	2.94	2,83	2,75	2.67	2.58	2.53	2,46	2.42	2.36	2.31
23	2.15	2,11	2,05	2.00	1.96	1.91	1.88	1.84	1,82	1.79	1.76
	2.97	2.89	2,78	2,70	2.62	2,54	2,48	2.41	2.37	2.32	2.26
24	2.13	2,09	2,03	1,98	[illegible]	1.89	1.86	1.82	1.80	1.77	1.73
	2.93	2,85	2,74	2.66	[illegible]	2,49	2.44	2.36	2.33	2.27	2.21
26	2.10	2.05	1.99	1.95	[illegible]	1.85	1.82	1,78	1.76	1.73	1.69
	2.86	2.78	2,66	2.58	2.50	2,42	2.36	2.28	2.25	2.19	2.13
28	2.06	2,02	1,96	1.91	1.87	1.82	1.79	1.75	1.73	1.69	1.65
	2,80	2,71	2,60	2,52	2.44	2,35	2.30	2.22	2.19	2.13	2.06
30	2.04	1.99	1.93	1,89	1.84	1.79	1.76	1.72	1.70	1.66	1.62
	2,74	2,66	2,55	2.47	2.38	2.30	2.25	2.16	2.13	2.07	2.01
34	1.99	1.95	1.89	1.84	1.80	1.75	1.71	1.67	1.65	1.61	1.57
	2.66	2.58	2,46	2.38	2.30	2.21	2.16	2.08	2.04	1.98	1.91
40	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1,61	1.59	1.55	1.51
	2.56	2,48	2.37	2.29	2.20	2.11	2.06	1.97	1.94	1.87	1.80
50	1.89	1.85	1,78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.44
	2.46	2.38	2,26	2,18	2.10	2,00	1.95	1.86	1.82	1.76	1.68
70	1.84	1,79	1.72	1,67	1.62	1.57	1.53	1.47	1.45	1.40	1.35
	2.35	2,27	2.15	2,07	1.98	1,88	1,83	1.74	1.70	1.62	1.53
100	1.79	1,75	1.68	1,63	1.57	1,52	1.48	1,42	1.39	1.34	1.28
	2.26	2.19	2.06	1,98	1.89	1,79	1.73	1,64	1.60	1.52	1.43
200	1.74	1.69	1.62	1.57	1.52	1,46	1.41	1,35	1.32	1.26	1.19
	2.17	2.09	1.97	1.88	1.79	1.69	1.63	1.53	1.48	1.39	1.28
∞	1.69	1.64	1,57	1.52	1.46	1.39	1.35	1.28	1.24	1.17	1.00
	2.08	2.00	1,88	1.79	1.70	1.59	1.52	1.41	1.36	1.25	1.00