E-kursuse "Tõenäosusteooria ja statistika" materjalid

Aine maht 3 EAP

Erko Jakobson (Tartu Ülikool), 2013

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

erko.jakobson@ut.ee

2013 kevad, Tartu

Sisukord

Sissejuhatus 4

Aine sissejuhatus 4

1 Juhuslikud sündmused 6

1.1. Sündmuste liigitamine 6

1.2. Venni diagrammid 7

1.3. Tehted sündmustega 8

1.4. Sündmuse sagedus ja tõenäosus 11

1.4.1. Sündmuste tõenäosuse võimalikud väärtused 11

1.4.2. Tõenäosuste liitmisreegel 12

1.4.3. Tõenäosuste korrutamisreegel 12

1.4.4. Tinglik tõenäosus 12

1.5. Tõenäosus sündmuste ruumis Ω 14

1.6. Bayesi valem 16

2 Kombinatoorika 21

2.1 Kombinatoorika ilma korduvate elementideta 21

2.2 Kombinatoorika koos korduvate elementidega 25

3 Juhusliku suuruse jaotused 29

3.1 Valik diskreetse juhusliku suuruse jaotusseaduseid 30

3.1.1 Binoomjaotus ehk Bernoulli jaotus 31

3.1.2 Geomeetriline jaotus 34

3.1.3 Poissoni jaotus 36

3.2 Jaotusseaduste esitamine 38

3.2.1 Jaotustihedus 40

3.2.2 Jaotusfunktsioon 42

3.3 Juhusliku suuruse arvkarakteristikud 46

3.3.1 Keskväärtus 47

3.3.2 Keskmiste kasutamisest 48

3.3.3 Dispersioon, ruuthälve ja standardhälve 49
3.4 Valik pideva juhusliku suuruse jaotusseaduseid 54
3.4.1 Ühtlane jaotus 55
3.4.2 Eksponentjaotus 55
3.4.3 Normaaljaotus 57
3.4.4 Studenti jaotus 61
1.5. Lisa 1. Studenti t-kordaja väärtused 63

Sissejuhatus

Aine sissejuhatus

Looduses ja ühiskonnas – sealhulgas teaduslikus uurimistöös, mitmesugustel tehnikaaladel ning tööstuslikus masstootmises – on sageli tegemist korduvalt esinevate nähtustega, mis toimuvad või kulgevad iga kord veidi erinevalt. Niisuguseid nähtusi nimetatakse juhuslikeks.

Näiteid juhuslikest nähtustest:

- Tänaval vastutulevate inimeste pikkus (vanus, kehakaal, ...);
- Märklaua tabamine märkilaskmisel püssi või vibuga;
- Radioaktiivse elemendi aatomi eluiga;
- Tuule suund ja suurus;
- Viking-loto loosimise kuus võidunumbrit;
- Masstootmises konveierilt tulevate seadmete või detailide pikkus;
- Ühe ja sama eseme pikkus (kaal, elektrijuhtivus, ...) erinevatel (korduvatel) mõõtmistel;

Üldiselt saab kõiki nähtusi ja sündmusi liigitada kindlateks, võimatuteks ja juhuslikeks.

Nähtust nimetatakse kindlaks, kui ta seatud tingimustel alati toimub. Nähtust nimetatakse võimatuks, kui see seatud tingimustel ei toimu iialgi. Nähtus, mis seisneb selles, et käestpillatud ese kukub maha (ja ei lähe selle asemel hoopis lendu), on kindel sündmus Maal, kuid ei ole seda mitte ümber Maa tiirlevas orbitaaljaamas. Samal ajal sündmus „käestpillatud pliiats jääb õhku hõljuma” on Maal võimatu.

Nähtust nimetatakse juhuslikuks, kui see seatud tingimustel võib toimuda või mitte toimuda. Võime öelda ka nii, et seatud tingimustel juhuslik nähtus kulgeb korduvatel katsetel iga kord erinevalt.

Täringu viskamisel silmade 1, 2, 3, 4, 5 või 6 saamine on juhuslik nähtus (eeldusel, et täring on korrapärane ja valmistatud homogeensest materjalist).

Paljudel nähtustel on korduv iseloom, mis on korduvalt realiseeritavate põhjuste (tingimuste) tagajärg. Niisuguste korduvate nähtuste kohta võib kujuneda teatud “objektiivne” ettekujutus nende juhuslikkuse või võimalikkuse kohta. Osutub, korduvaid juhuslikke nähtusi iseloomustavad seaduspärasused, mis ei ole juhuslikud. Selliseid seaduspärasusi nimetatakse tõenäosuslikeks.

Juhuslikes suurustes või nähtustes ilmnevaid mittejuhuslikke seaduspärasusi uurib, põhjendab ja selgitab **tõenäosusteooria**. Juhuslike suuruste tõenäosuslike seaduspärasuste eksperimentaalse uurimisega ja põhjendamisega tegeleb **matemaatiline statistika**. Kui tõenäosusteooria tegeleb peaasjalikult juhuslike sündmuste ja juhuslike suuruste analüüsiga, keskendudes tõenäosustele, jaotusfunktsioonidele ja juhulike suuruste arvkarakteristikute uurimisele, siis matemaatilise statistika eesmärgiks on korduvatel mõõtmistel või vaatlustel saadud andmete statistiliste seaduspärasuste väljaselgitamine ja põhjendamine.

Õpiväljundid, ehk kursuse positiivsele hindele läbinud üliõpilane:

1. mõistab tõenäosusteooria ja statistika põhimõisteid
2. oskab lahendada tavalisemaid tõenäosusteooria ülesandeid
3. oskab luua seoseid tõenäosusteooria ja statistika vahel

4. oskab leida andmestikke kirjeldavaid karakteristikuid (aritmeetiline keskmine, mediaan, standardhälve)
5. oskab kontrollida statistilisi hüpoteese kahe andmestiku võrdlemiseks (t-test, z-test)
6. on võimeline teostama korrelatsioon- ja regressioonanalüüsi.

Hindamismeetodid:

Hindamismeetoditeks on kodused tööd, kirjalik test, grupitöö ning kirjalik eksam. Kodused tööd (õpiväljundid 1 – 6) annavad 10%, test (õpiväljundid 1, 2) annab 20%, grupitöö (õpiväljundid 1, 4 – 6) 10% ning eksam (õpiväljundid 1, 3 – 6) 60% koondhindest. Testis ning eksamil on ülesande juures ära toodud, palju punkte mingi ülesanne annab (kui ülesanne on jagatud mitmesse ossa, siis palju punkte mingi osa annab). Grupitööl hinnatakse töö püstitust, meetodi sobivust, tulemuse õigsust ning töö vormistust. Kõik hindamismeetodid tuleb sooritada positiivsele hindele, s.t. peab saama üle 50% punktidest.

Koondhinne arvutatakse vastavalt praegu kehtivale %-süsteemile ("A" = >90 % jne).

Grupitöö on mingi andmebaasi statistiliste seoste ja hüpoteeside otsimine-kontrollimine vabalt valitud arvutusprogrammiga ning saadud tulemuste vormistamine. Grupi suurus – kuni 5 tudengit. Analüüsitavad andmed leiavad tudengid soovitatavalt ise, aga vajadusel võib andmebaase saada ka õppejõu käest.

Grupitöö eesmärk:

- Praktilise kogemuse saamine andmete statistilise analüüsiga. Väga tõenäoliselt sisaldab ka lõputöö statistilist andmetöötlust.
- Statistikaprogrammidega tutvumine, sest 3 EAP raames arvutipraktikumi ei jõua teha.
- Õpitu rakendamine praktikas.

Test on kirjalik, tõenäosusteooria ülesannete lahendamise peale

Eksam on kirjalik, 50% ülesannete lahendamine, 50% teooria.

Testi ja eksami ülesannete osas on spikerdamine on limiteeritud. Kaastudengeid, raamatuid, konspekte, arvuteid, telefone jne pole lubatud kasutada. Laual võib olla kalkulaator ning üks A4 formaadis lehekülg vabalt valitud käsitsi kirjutatud teksti, nagu valemid, definitsioonid, jne.

1 Juhuslikud sündmused

1.1. Sündmuste liigitamine

Iga teadusharu põhineb teatud arvul põhimõistetel, mille abil antakse teooria loogiline ülesehitus,
Iga teadusharu põhineb teatud arvul põhimõistetel, mille abil antakse teooria loogiline ülesehitus, s.t määratakse keerukamad mõisted. Tõenäosusteoorias on esimesteks niisugusteks mõisteteks katse (eksperiment) ja sündmus. Neid ei defineerita, vaid selgitatakse näidetega.

Tõenäosusteooria baaskontseptsioonideks on **katse** ja **sündmus**. Katse all mõistame teatud tingimuste kompleksi realiseerumist, mille tulemusena võivad toimuda mingid sündmused. Tingimused võivad olla loodud kunstlikult (tahtlikult), või nad eksisteerivad sõltumatult eksperimentaatorist. Võime öelda ka, et katse on niisugune olukord või seisund, mille tulemusena võivad toimuda mingid sündmused.

Katse ~ eksperiment ~ vaatlus

Sõltuvalt sündmuse sisust kasutame nimetuse katse asemel veel nimetusi eksperiment või vaatlus.

Sündmus – fakt või ilming, mis leiab aset katses.

Niisiis katse tulemusena toimuvad teatavad (aimatavad) sündmused. Katse on määratletud, kui on teada katse tingimused ja oodatavate sündmuste hulk. Katse võib olla korraldatud kindla eesmärgi suunitlusega, planeerimisega, aga ka passiivse vaatlusena, lihtsalt andmete registreerimisega. Rõhutame veel, et katse ei tarvitse olla korraldatud inimese poolt; katse võib kulgeda sõltumatult inimesest. Siinjuures on inimene ainult vaatleja osas või toimuva registreerija.

Katse tulemusel toimuv sündmus võib olla deterministlik (ette määratud) või juhulik.

Deterministlik sündmus: korduval katsel on tulemus alati üks ja seesama sündmus.

Juhuslik sündmus: korduval (ühesugusel) katsel on sündmus iga kord erinev.

Tähistame sündmusi ladina tähestiku suurtähtedega, varustades neid vajaduse korral indeksitega: A, B, ... $A1$, $B1$ jne.

Juhuslikuks sündmuseks nimetatakse sündmust, mis katsel võib toimuda või mitte, kusjuures varem ei ole teada, kas sündmus toimub. Juhusliku sündmuse toimumine on küll võimalik, kuid selle reaalne toimumine sõltub mitmetest omavahel seotud või sõltumatutest põhjustest.

Näide 1.1. Lihtsamateks juhuslikeks sündmusteks on täringu viskamisel silmade tulek, kaardi võtmine kaardipakist, mündi viskamine, võit loteriil, inimese eluiga jne.

Näide 1.2. Valime Tartu elanikest juhuslikult suvalise esindaja. Juhuslikuks sündmuseks on meesterahva/naisterahva saamine, aga samuti vanema/noorema kui 35 aastat inimese saamine.

Näide 1.3. Koosnegu toodang n seadmest. Seadme kvaliteedi kontrollimine olgu aga niisugune, et selle tulemusel seade hävib. Et mitte hävitada kogu toodangut, valitakse m seadet ($m < n$) ja kontrollitakse neid. Katse tulemuseks saame defektsete seadmete arvu. Siit võib juba teha järeldusi toodangu kohta.

Näide 1.4. Vaatleme Browni liikumist – vedelikus heljuvate väikeste aineosakeste kaootilist liikumist. Osake muudab oma asendit juhuslikult, põrkudes kokku vedeliku molekulidega. Osakese asend fikseeritud ajamomendil t on aga raskesti prognoositav ja seega juhuslik sündmus. Kui aga aeg t muutub pidevalt, jõuame üldisema mõiste – juhusliku protsessi juurde.

Vaadeldud näidetest tuleneb, et tõenäosusteoorias sündmus on väga üldine ja universaalne mõiste, mille alla mahuvad kõik meid igapäevaelus ümbritsevad nähtused.

Tõenäosusteoorias on eriline koht kahel sündmusel: **kindel sündmus Ω** ja **võimatu sündmus Φ**.

Kindlaks sündmuseks Ω nimetatakse sündmust, mis katse tulemusena alati toimub, näiteks visatud kivi kukub alati maa peale tagasi (ei kehti kosmoses); iga elusolend hukub temperatuuril 1000 ºC.

Võimatu sündmus Φ on sündmus, mis vaadeldava katse tulemusena kunagi ei toimu, näiteks 7 silma saamine täringuviskel või varahommikune temperatuur akna taga –275 ºC.

Kui sündmuse A toimumisest järeldub sündmuse B toimumine, siis tähistame

$$A \subset B$$

s.t. A kuulub B-sse. Sel korral nimetatakse sündmust A ka sündmuse B alamsündmuseks või osasündmuseks ja sündmust B sündmuse A ülemsündmuseks. Näiteks, kui sündmus A tähendab kõiki TT kursusele registreerinuid ja B – kõiki TT kursusele registreerinud tütarlapsi, siis ilmselt $B \subset A$.

Alati kehtivad seosed

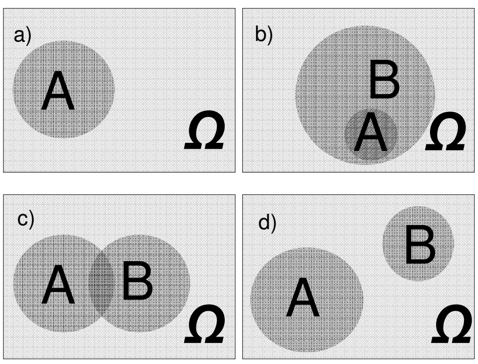
$$\Phi \subset A; \ A \subset \Omega.$$

Need seosed tähendavad, et igasse sündmuste hulka kuulub ka võimatu sündmus Φ ja et iga sündmus on kindla sündmuse Ω alamsündmus.

Sündmusi A ja B nimetatakse võrdseteks või samasteks või identseteks (ekvivalentseteks), kui kehtivad seosed $A \subset B$ ja $B \subset A$, sel korral kirjutatakse $A = B$

1.2. *Venni diagrammid*

Sündmus A olgu ekvivalentne juhusliku punkti sattumisega piirkonda A (ühe-, kahe- või kolmemõõtmelises ruumis), mis on piirkonna ***Ω*** alamhulk. Vaatleme siin kahemõõtmelisi piirkondi. Sel korral kindlat sündmust Ω kujutame ristkülikuna. Joonisel 1.1 on sündmus $A \subset B$ on toodud alajaotusel b.



Joonis 1.1. Venni diagrammid.

1.3. Tehted sündmustega

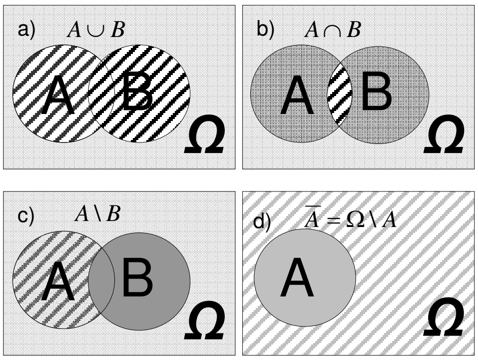
Sündmuste vahelised elementaartehted on järgmised:

a) summa (või): $A \cup B$ – toimub kas A, B või nii A kui ka B.

b) korrutis (ja): $A \cap B$ – toimuvad nii A kui ka B.

c) vahe: $A \setminus B$ – toibub A, aga B ei toimu.

d) vastandsündmus: $\overline{A} = \Omega \setminus A$ – ei toimu sündmust A.



Joonis 1.2. Sündmustevahelised elementaartehted.

Definitsioonidest järeldub, et sündmuste summa ja korrutis on kommutatiivsed, s.o

$$A \bigcup B = B \bigcup A,$$
$$A \bigcap B = B \bigcap A.$$

Kehtivad sündmuste liitmise ja korrutamise assotsiatiivsuse seadus

$$(A \bigcup B) \bigcup C = A \bigcup (B \bigcup C),$$
$$(A \bigcap B) \bigcap C = A \bigcap (B \bigcap C)$$

ja distributiivsuse seadus

$$(A \bigcup B) \bigcap C = (A \bigcap C) \bigcup (B \bigcap C).$$

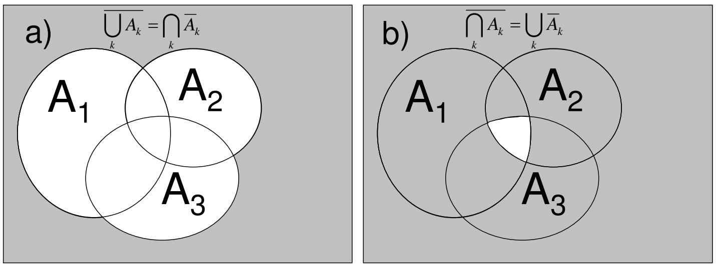
Kasutades vaadeldud tehteid, saab keerukamaid sündmusi (liitsündmusi) avaldada antud sündmuste (lihtsündmuste) kaudu. Ülesannete lahendamisel on sageli otstarbekas lihtsustada valemeid, kus mingid sündmused avalduvad teiste sündmuste kaudu. Seda võib teha vastavalt summa või korrutise kommutatiivsuse, assotsiatiivsuse või distributiivsuse omaduste põhjal. Kasulikud on siinjuures järgmised **duaalsusseosed** ehk **Morgani seadused**:

$$\overline{A \bigcup B} = \overline{A} \bigcap \overline{B}, \ \overline{A \bigcap B} = \overline{A} \bigcup \overline{B},$$

mis formaalselt tähendavad, et korrutamise ja summa tehtemärgid tuleb ära vahetada, kui sündmused asendada vastandsündmustega.

Duaalsusseosed võib üldistada sündmuste hulga jaoks:

$$\overline{\bigcup_k A_k} = \bigcap_k \overline{A_k}, \quad \overline{\bigcap_k A_k} = \bigcup_k \overline{A_k}.$$



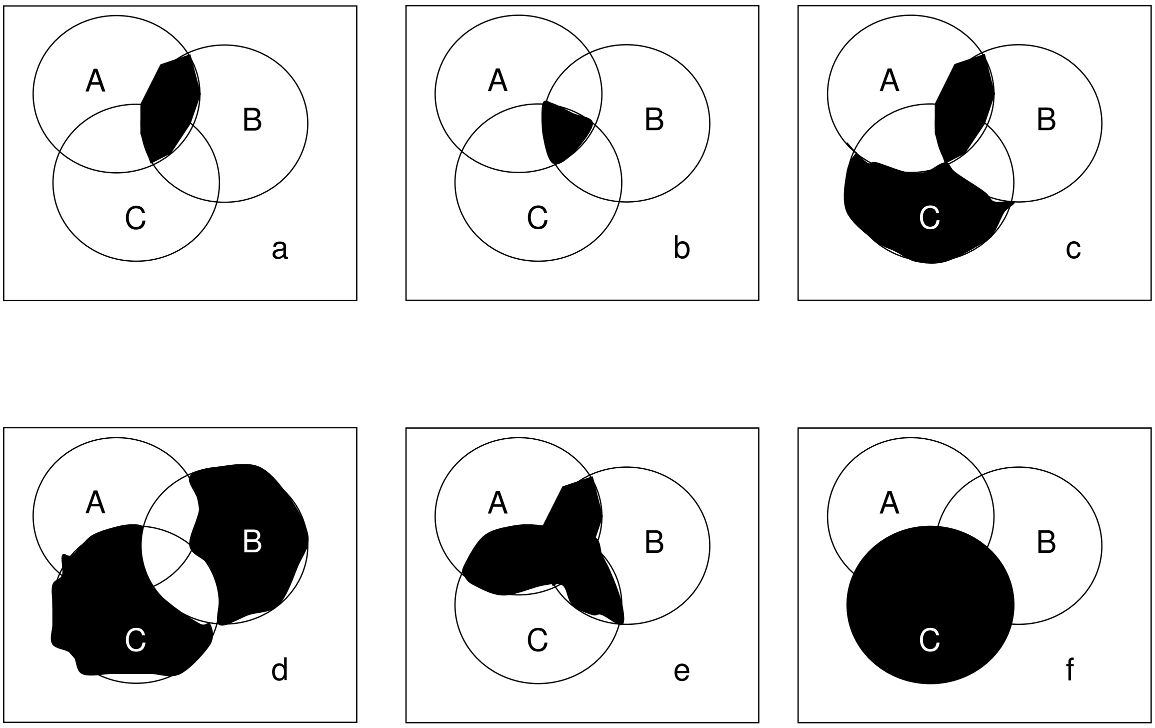
Joonis 1.3. Morgani seosed kolme hulga korral.

<u>ÜLESANDED</u>

<u>**Ülesanne 1.1.**</u> Esitada Venni diagrammil mustaks värvitud ala HT (hulgateooria) tehete abil

1. kasutades märke $(\cap\,;\ \cup\,;\ \backslash)$

2. kasutades märke $(\cap\,;\ \cup\,;\overline{X})$, siin X võib olla A, B, C või HT tehe nende vahel.



Lahendus:

a. $A \cap B$

b. $A \cap B \cap C$

c. $(A \cap B) \cup (C \setminus (A \cup B))$ $\qquad$ $(A \cap B) \cup (\overline{(A \cup B)} \cap C)$

d. $(C \setminus B) \cup (B \setminus (A \cup C))$ $\qquad$ $(\overline{B} \cap C) \cup (\overline{(A \cup C)} \cap B)$

e. $(A \cap B) \cup (A \cap C) \cup (B \cap C)$

f. C

<u>Ülesanne 1.2.</u> Uurida Venni diagrammi järgi, et kui $A \cap B \subset \overline{C}$ ja $A \cup C \subset B$, järelikult $A \cap C = ?$

Lahendus:

Vaatame alustuseks teist tingimust: $A \cup C \subset B$, see tähendab, et nii *A* kui ka *C* peavad kuuluma täielikult *B*-sse. Esimene tingimus, et $A \cap B \subset \overline{C}$ lihtsustub, sest kuna *A* kuulub *B*-sse, siis nende ühisosa on võrdne *A*-ga. Kuna aga *A* peab seega kuuluma *C* vastandsündmusesse, siis, järelikult $A \cap C = \Phi$, s.t. *A*-l ja *C*-l ühisosa puudub.

<u>Ülesanne 1.3.</u> Münti visatakse 3 korda. Kirjuta välja variandid, mille puhul:

A. kulli visatakse täpselt 2 korda
B. kulli visatakse vähemalt 2 korda
C. kull ei tulnud enne, kui tuli kiri
D. esimesena visati kull

leia järgmised sündmused: $\overline{A}$; $A \cup (C \cap D)$; $A \cap \overline{D}$

Lahendus:

Võimalike variantide arv on $2^3 = 8$

Kirjutame alustuseks välja kõik võimalikud variandid, tähistame: kull = 0, kiri = 1:

(000); (001); (010); (011); (100); (101); (110); (111)

A. kulli visatakse täpselt 2 korda:	(001); (010); (100)
B. kulli visatakse vähemalt 2 korda:	(000); (001); (010); (100)
C. kull ei tulnud enne, kui tuli kiri:	(100); (110); (111)
D. esimesena visati kull;	(000); (001); (010); (011)

$\overline{A} = (000) \cup (011) \cup (101) \cup (110) \cup (111)$

$A \cup (C \cap D) = A \cup \Phi = A = (001) \cup (010) \cup (100)$

$A \cap \overline{D} = (100)$

1.4. Sündmuse sagedus ja tõenäosus

Sündmuste tõenäosuse klassikaline definitsioon – Sündmuse *A* tõenäosuseks *P*(*A*), nimetatakse sündmuse toimumiseks soodsate juhuste arvu *m*(*A*) suhet kõigi võimalike juhuste arvusse *n*, kus juhused moodustavad elementaarsündmuste süsteemi.

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

Näide 1.5. Kausis on 4 kollast ja 5 punast ploomi. Kausist võetakse juhuslikult 1 ploom, kui suur on tõenäosus, et see ploom on kollane? Kausis on kokku 4 + 5 = 9 ploomi, seega ploomivalikus on 9 erinevat võimalust, kollaste ploomide ehk soodsate juhuste arv on 4, seega kausist juhuslikult kollase ploomi võtmise tõenäosus on *P*(*A*) = 4 / 9 = 0,44.

Sageli pole siiski tõenäosuse klassikalist definitsiooni võimalik rakendada ja tuleb leida sündmuse suhteline sagedus. Vaatleme katset, mille tulemusena võib toimuda sündmus *A*. Eeldame, et katset võib korrata, kusjuures katsetingimused ei muutu ja igal katsel ei sõltu katsetulemus teiste katsete tulemustest. Nimetame niisuguseid katseid sõltumatuteks katseteks sündmuse *A* suhtes. Olgu selles katseseerias katsete arv *n* ja *m*(*a*) – selles seerias sündmuse *A* toimumiste (esinemiste) arv. Vaadeldavas katseseerias nimetatakse sündmuse *A* sageduseks suurust

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

Kui katsete arv *n* ei ole suur, siis sündmuse sagedus on oluliselt juhusliku iseloomuga. Katsete suure arvu n korral on aga sagedused p_n erinevates katseseeriates, nagu näitab pikaajaline praktika, ligikaudu võrdsed. S.t sagedusel on tendents stabiliseeruda (on stabiilsuse omadus). Siin peab märkima, et tõenäosusteooria uurib ainult niisuguseid sündmusi, millel on stabiilsuse omadus. Teisisõnu, eksisteerib niisugune konstant *P*(*A*), mida nimetatakse sündmuse *A* tõenäosuseks, mille ümber sagedused võnguvad, s.o erinevad vähe sellest konstandist *P*(*A*). Lühidalt, iga lõpliku kuid üha suurema *n* korral kehtib ligikaudne võrdus

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

Ning piirjuhul läheb sündmuse sagedus sündmuse tõenäosuseks:

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

Edaspidi tegelemegi peamiselt sündmuste tõenäosustega.

1.4.1. Sündmuste tõenäosuse võimalikud väärtused

Kuna sündmuse *A* toimumiste arv saab olla $0 \le m(A) \le n$, siis järelikult $0 \le P^*(A) \le 1$, s.t. sündmuse *A* sagedus saab olla vahemikus 0-st 1-ni. Kui *P*(*A*) = 0, on tegemist võimatu sündmusega, kui *P*(*A*) = 1, on tegemist kindla sündmusega.

1.4.2. Tõenäosuste liitmisreegel

Kahe sündmuse *A* ja *B* korral kehtib võrdus

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Tõestus:

Lähtudes joonisest 1.4,

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

ja

$$m(A \cup B) = m(A) + m(B) - l,$$

(liige „–*l*“ seetõttu, et *A* ja *B* ühisosa loeti 2 korda), siis

$$P(A \cup B) = \frac{m(A) + m(B) - l}{n} = \frac{m(A)}{n} + \frac{m(B)}{n} - \frac{l}{n} = P(A) + P(B) - P(A \cap B).$$

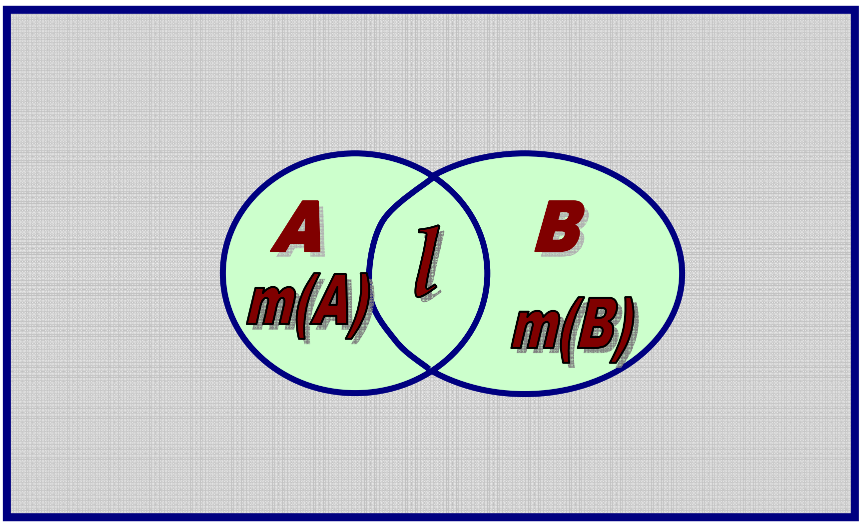
1.4.3. Tõenäosuste korrutamisreegel

Kahe juhusliku sündmuse *A* ja *B* korral kehtivad võrdused

$$P(A \cap B) = P(A)P(B|A),$$
$$P(A \cap B) = P(B)P(A|B),$$

kus $P(A|B)$ on sündmuse *A* tinglik tõenäosus sündmuse *B* suhtes. Ühe sündmuse *A* tõenäosust, mis on leitud tingimusel, et teine sündmus *B* toimus, nimetatakse sündmuse *A* tinglikuks tõenäosuseks sündmuse *B* suhtes ja tähistatakse $P(A|B)$.

1.4.4. Tinglik tõenäosus



Joonis 1.4. Tingliku tõenäosuse selgitus.

Lähtume joonisest 1.4:

$$P(A)=\frac{m(A)}{n},\ P(B)=\frac{m(B)}{n},\ P(A\bigcap B)=\frac{l}{n}.$$

Tingliku tõenäosuse mõiste järgi

$$P(B\,|\,A)=\frac{l}{m(A)},\ P(A\,|\,B)=\frac{l}{m(B)}$$

järelikult

$$P(A\bigcap B)=\frac{l}{n}=\frac{m(A)}{n}\cdot\frac{l}{m(A)}=P(A)P(B\,|\,A)$$

Analoogiliselt

$$P(A\bigcap B)=\frac{m(B)}{n}\cdot\frac{l}{m(B)}=P(B)P(A\,|\,B)$$

Sõltumatute sündmuste hulkade korral ei sõltu sündmuste hulga *A* tõenäosus sellest, kas on toimunud sündmuste hulk *C*, ning vastupidi. Seega on $P(A\,|\,C)=P(A)$.
See on ka tingimuseks sündmuste sõltumatuse kohta.

Näide 1.6. Olgu Ω kõik TT kursusele registreerunud, *A* on noormehed ja *B* on füüsikud.

17.01.13 seisuga oli registreerunuid kokku 95, noormehi neist 71. Füüsikuid oli 28. Füüsikutest noormehi oli 23.

$P(A)=\frac{m(A)}{n}=\frac{71}{95}=0{,}747$ on tõenäosus, et juhuslikult valitud tudeng on noormees.

$P(B)=\frac{m(B)}{n}=\frac{28}{95}=0{,}295$ on tõenäosus, et juhuslikult valitud tudeng on füüsik.

$P(A\bigcap B)=\frac{l}{n}=\frac{23}{95}=0{,}242$ on tõenäosus, et juhuslikult valitud tudeng on korraga nii noormees kui ka füüsik.

$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{71}{95}+\frac{28}{95}-\frac{23}{95}=\frac{76}{95}=0{,}8$ on tõenäosus, et juhuslikult valitud tudeng on füüsik või vähemalt noormees.

$P(B\,|\,A)=\frac{l}{m(A)}=\frac{23}{71}=0{,}324$ on tõenäosus, et juhusikult valitud noormees on füüsik, s.t. eelnevalt on paigas, et tegemist on noormehega.

$P(A\,|\,B)=\frac{l}{m(B)}=\frac{23}{28}=0{,}821$ on tõenäosus, et juhusikult valitud füüsik on noormees, s.t. eelnevalt on paigas, et tegemist on füüsikuga.

$$P(A \cap B) = P(A)P(B \mid A) = \frac{71}{95} \cdot \frac{23}{71} = \frac{23}{95} = 0{,}242$$

$$P(A \cap B) = P(B)P(A \mid B) = \frac{28}{95} \cdot \frac{23}{28} = \frac{23}{95} = 0{,}242$$

$P(A \mid B) = 0{,}821 \neq P(A) = 0{,}747$, järelikult A ja B ei ole omavahel sõltumatud.

1.5. *Tõenäosus sündmuste ruumis Ω*

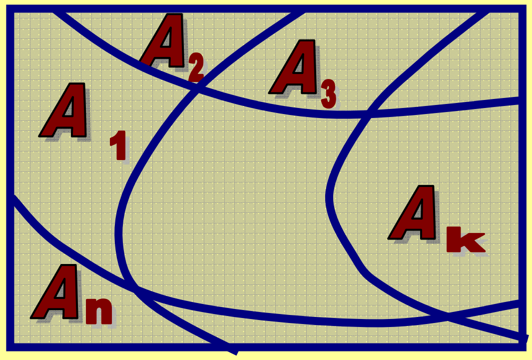
Defineerime esmalt sündmuste täissüsteemi $\{A_1,\ A_2, \ldots, A_n\}$.

(1) sündmused välistavad vastastikku üksteist

$$A_i \cap A_k = \Phi \quad (i \neq k).$$

(2) Sündmused moodustavad täissüsteemi

$$\bigcup_{k=1}^{n} A_k = \Omega$$



Kui üksteist välistavad sündmused $\{A_1,\ A_2, \ldots, A_n\}$ moodustavad täissüsteemi ruumis Ω, siis kehtib

$$\sum_{i=1}^{n} P(A_i) = 1$$

Täissüsteem tähendab, et igal katsel leiab aset mingi sündmus (üks ja parajasti üks sündmus). Välistatud on olukord, kus katses üldse midagi ei juhtu.

Näide 1.7. Olgu sündmuseks A_i mündiviskel kulli tulemine kõigil i-l viskel. Et üksikviskel on kullisaamise tõenäosus (1/2), on

$$P(A_i) = \left(\frac{1}{2}\right)^i$$

Kuna kehtib valem (tõestada iseseisvalt)

$$\sum_{i=1}^{n} r^i = r\frac{1-r^n}{1-r} \quad \rightarrow \quad \sum_{i=1}^{\infty} r^i = \frac{r}{1-r} \ ,$$

(teine summa on koonduv vaid siis, kui $|r| < 1$), saame

$$\sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1.$$

See on näide lõpmatust kuid loenduvast sündmuste täissüsteemist. Saadud tulemus on tõestus sellele, et vaatamata lõpmatusele on see sündmuste süsteem täielik.

ÜLESANDED

Ülesanne 1.4. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 2/3$ ja $P(B) = 1/6$ ning $P(A \cap B) = 1/9$. Leida tõenäosus $P(A \cup B)$?

Lahendus:

Joonistame Venni diagrammi. Sellelt on näha, et $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, seega on

$P(A \cup B) = 2/3 + 1/6 - 1/9 = 13/18$

Ülesanne 1.5. Olgu E ja F kaks sündmuste hulka, mille kohta on teada, et neist vähemalt ühe toimumise tõenäosus on ¾. Mis on tõenäosus, et ei E ega ka F ei toimu?

Lahendus:

Joonistame Venni diagrammi ning kirjutame ülesande lahti valemite keeles. Teame, et $P(E \cup F)$ = ¾, meid huvitab tõenäosus, et $P(\overline{E \cup F}) = P(\Omega) - P(E \cup F) = 1 - 3/4 = 1/4$

Ülesanne 1.6. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 0,3$ ja $P(B) = 0,4$ ning $P(A \cap B) = 0,2$. Leida tõenäosus $P(\overline{A} \cap B)$?

Lahendus: Lahenda ise (vastus on 0,2).

Ülesanne 1.7. Vaatame sündmuseid A, B ja C, mis võivad olla mingi eksperimendi tulemused. Kontrolli, kas tõenäosus, et ainult sündmus A toimub (B ja C ei toimu) on arvutatav valemiga $P(A \setminus B \setminus C) = P(A \cup B \cup C) - P(B) - P(C) + P(B \cap C)$.

Lahendus: Joonista ise Venni diagramm. Sellelt näed, et antud valem kehtib.

Ülesanne 1.8. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 0,4$ ja $P(B) = 0,5$ ning $P(A \cap B) = 0,1$. Leida tõenäosus, et juhuslik sündmus kuuluks hulka A või B, kuid mitte mõlemasse korraga.

Lahendus: Lahenda ise (vastus on 0,7).

Ülesanne 1.9. Ülesanne 1.8. algandmete põhjal kontrolli, kas sündmused A ja B on omavahel sõltumatud.

Lahendus: Sündmuste sõltumatuse korral peab kehtima tingimus $P(A \mid B) = P(A)$.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.5} = 0.2$$

Kuna $P(A \mid B) = 0.2 \neq P(A) = 0.4$, siis sündmused A ja B on omavahel sõltuvad.

1.6. Bayesi valem

Tingtõenäosuse definitsioon oli

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Siin $P(A|B)$ on Tõenäosus A saamiseks tingimusel, et sündmus B on toimunud.
Selles kontekstis nimetatakse sündmust B ka hüpoteesiks.

Kasutame korrutamisvalemit:

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

Siit saab

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Nüüd edasi saab esitada täistõenäosuse B jaoks

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) = P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})$$

Eelneva valemi esimene võrdus on soovituslik tõestada Venni diagramme kasutades. Tundub mõttetu lihtsa asja keeruliseks ajamisena, aga see trikk on Bayesi valemi saamise eelduseks ning on võtmeks mitmete ülesannete lahendamiseks.

Näide 1.8. Olgu sündmuseks B eksamist läbi saamine, sündmuseks A aga eksamiks õppimine. Eelnev valem tähendab seda, et tõenäosust eksamist läbi saada võib vaadelda kahe variandi summana – läbi saamine, kui õpitakse ning läbi saamine, kui ei õpita. Kuna sündmus ja tema vastandsündmus moodustavad täissüsteemi, siis võib sündmuseks A valida tegelikult suvalise sündmuse, tulemust see ei mõjuta. Praktikas on mõistlik valida siiski sündmuseks A midagi, mille kohta on mingisugust lisainfot.

Eelnevate valemite põhjal saabki kokku panna **Bayesi valemi**:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}$$

ÜLESANDED

Ülesanne 1.10. Leida tõenäosus, et klassis, kus on n tudengit, ei esine korduvaid sünnipäevi? Enne arvutamist teha kõhutunde pealt pakkumine, et kui suur peaks olema n, et see tõenäosus oleks suurem kui 0,5, s.t. et tõenäosus, et klassis on korduv sünnipäev oleks suurem, kui et ei ole korduvat sünnipäeva.

Lahendus:

Oletame, et tudengeid on kaks. Esimese tudengi sünnipäeval pole millegagi korduda. Tõenäosus $P(A_1A_2)$, et teise tudengi sünnipäev ei kattuks esimese tudengi sünnipäevaga on:

$$P(A_1A_2) = 1 - \frac{1}{365}.$$

Liige 1/365 viitab, et korduvate sünnipäevade saavutamiseks soodsate kuupäevade arv on 1, kõigi kuupäevade arv aga 365 (jätame liigaastad arvestusest välja). Siin A_1A_2-ga tähistame sündmust, et tudengite A_1 ja A_2 sünnipäevad ei kattu, hiljem tähistame $A_1A_2A_3$-ga sündmust, et tudengite A_1, A_2 ja A_3 sünnipäevad ei kattu.

Oletame, et tudengeid on kolm. Nüüd saame kasutada juba tingimuslikku tõenäosust. Tingimuseks on, et kahel esimesel tudengil on sünnipäevad erinevatel päevadel, s.t. $P(A_1A_2)$. Tõenäosus, et kolmandal tudengil ei ole sünnipäev samal ajal esimese ja teise tudengiga, on $P(A_3| A_1A_2)$:

$$P(\mathrm{A}_3 \mid A_1A_2) = 1 - \frac{2}{365}$$

Kasutades korrutamise reeglit, saame et

$$P(A_1A_2A_3) = P(A_3 \cap A_1A_2) = P(A_3 \mid A_1A_2) \cdot P(A_1A_2) = (1 - \frac{2}{365}) \cdot (1 - \frac{1}{365}) \approx 0{,}9917$$

Oletame nüüd, et me oleme välja arvutanud juba n – 1 tudengi jaoks tõenäosuse. n-nda tudengi puhul oleks tõenäosus, et n-nda tudengi sünnipäev ei kattuks eelnevate tudengite sünnipäevadega:

$$P(A_n \mid A_{n-1}) = 1 - \frac{n-1}{365}$$

Kasutades nüüd jälle korrutamise reeglit, saame, et

$$P(A_n) = P(A_n \cap A_{n-1}) = P(A_n \mid A_{n-1}) \cdot P(A_{n-1}) = (1 - \frac{n-1}{365}) \cdot P(A_{n-1})$$

Rakendades saadud valemit järjest alates n = 4, saame kergesti üldvalemi:

$$P(A_n) = P(A_n \cap A_{n-1}) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) \cdot ... \cdot P(A_n) =$$
$$= (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdot (1 - \frac{3}{365}) \cdot (1 - \frac{4}{365}) \cdot ... \cdot (1 - \frac{n-1}{365})$$

Vastav joonis demos „korduvad_synnip2evad.mcd“

Alates 23-st tudengist on tõenäosus, et on kordseid sünnipäevi, suurem, kui et ei ole.

<u>Ülesanne 1.11.</u> Täringut visatakse 2 korda. *A* on „silmade summa on 4“, *B* on „vähemalt ühe täringu silmade arv on 3“. Kas *A* ja *B* on sõltumatud sündmused?

Lahendus:

A tõenäosus on 3/36, *B* tõenäosus on 11/36, tõenäosus, et kehtivad nii *A* kui ka *B* on 2/36 (variandid {(1; 3); (3; 1)}).

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}$$

Kuna $P(A) = \frac{1}{12} \neq P(A \mid B) = \frac{2}{11}$, siis järelikult pole sündmused *A* ja *B* sõltumatud.

<u>Ülesanne 1.12.</u> Kaardipakist (52 kaarti) võetakse kaks kaarti. Olgu S1, et esimene kaart on poti mastist. Olgu S2, et teine kaart on potimastist. Arvuta a) $P(S1)$; $P(S2 \mid S1)$; $P(S2 \mid \overline{S1})$; b) Mis on tõenäosus, et sõltumata esimese kaardi mastist on teine kaart potimast; c) Kas *S*1 ja *S*2 on sõltumatud sündmused?

Lahendus: Lahenda ise (osa vastused on antud)

a) $P(S2 \mid \overline{S1}) = 13/51$.

c) *S*1 ja *S*2 on sõltuvad sündmused, s.t. *S*2 tõenäosus sõltub sündmuse *S*1 esinemisest.

<u>Ülesanne 1.13.</u> Sinul diagnoositakse üks vähelevinud haigus. Sa tead, et selle haiguse saamise tõenäosus on ainult 1%. Tähistame, et *D* tähendab, et „sul on see haigus“ ja *T* tähendab, et „test ütleb, et sul on see haigus“. On teada, et see test pole täiuslik, ehk $P(T \mid D) = 0{,}98$ ning $P(\overline{T} \mid \overline{D}) = 0{,}95$. Testi tulemus oli positiivne, kuid mis on tõenäosus, et sul tõesti on see haigus?

Lahendus:

Enne lahendama asumist, vaata neid numbreid ning paku kõhutunde järgi, kui suur on tõenäosus, et positiivne testi tulemus tähendab, et sul tõesti on see haigus?

Paneme HT valemitena kirja olemasoleva info:

$P(D) = 0{,}01$, järelikult $P(\overline{D}) = 0{,}99$

$P(T \mid D) = 0{,}98$

$P(\overline{T} \mid \overline{D}) = 0{,}95$, järelikult $P(T \mid \overline{D}) = 0{,}05$

Küsitakse $P(D \mid T) = ?$

Leiame kõigepealt tõenäosuse, et test annab positiivse tulemuse:

$$P(T) = P(T \cap D) + P(T \cap \overline{D}) = P(T \mid D) \cdot P(D) + P(T \mid \overline{D}) \cdot P(\overline{D}) =$$
$$= 0{,}98 \cdot 0{,}01 + 0{,}05 \cdot 0{,}99 = 0{,}0098 + 0{,}0495 = 0{,}0593$$

Kasutame nüüd tingliku tõenäosuse arvutamise reeglit:

$$P(D \mid T) = \frac{P(D \cap T)}{P(T)} = \frac{0{,}0098}{0{,}0593} = 0{,}165$$

Seega, kui testi tulemus on positiivne, on tõenäosus 16,5%, et sa oled haige. Saadud tulemus tekitab kahtlusi, kuna kõik testi kohta käivad tõenäosused on ju üsna head. Samas analüüsime nüüd testi kohta antud tingimust $P(\overline{T} \mid \overline{D}) = 0{,}95$. Oletame, et testitakse 101 inimest, kes 100 on terved. See tingimus aga annab, et test tunnistab terveks 95 inimest, seega on testi järgi haigeid 6 inimest, kuigi tegelikult on ainult 1. Seega on tõenäosus, et haigeks testitu ongi haige ainult 1/6 ≈ 16,7%. Saadud hinnang langeb hästi kokku arvutatud tõenäosusega.

Ülesanne 1.14. Arendame nüüd eelmist ülesannet edasi. Testi järgi oled sa haige. Et kontrollida, et tõesti oled sa haige, teed sa veel ühe testi. See osutus ka positiivseks. Mis on nüüd tõenäosus, et sa tõesti oledki haige? Huvi pärast, proovi uuesti kõhu tunde järgi arvata, kui suur see tõenäosus on.

Lahendus:

Kordame uuesti eelnevat arvutuskäiku, kuid nüüd on haige olemise tõenäosus juba 0,165, seega

$P(D) = 0{,}165$, järelikult $P(\overline{D}) = 0{,}835$

Leiame kõigepealt tõenäosuse, et ka teine test annab positiivse tulemuse:

$$P(T) = P(T \cap D) + P(T \cap \overline{D}) = P(T \mid D) \cdot P(D) + P(T \mid \overline{D}) \cdot P(\overline{D}) =$$
$$= 0{,}98 \cdot 0{,}165 + 0{,}05 \cdot 0{,}835 = 0{,}1617 + 0{,}0418 = 0{,}2035$$

Kasutame nüüd tingliku tõenäosuse arvutamise reeglit:

$$P(D \mid T) = \frac{P(D \cap T)}{P(T)} = \frac{0{,}1617}{0{,}2035} = 0{,}795$$

Seega, kui ka teise testi tulemus on positiivne, on tõenäosus 79,5%, et sa oled tõesti haige. Kahe testi usaldusväärsus osutus märgatavalt paremaks kui ühe testi usaldusväärsus.

OLULINE: see kehtib ainult eeldusel, et teine test on esimesest sõltumatu, s.t. kasutab teistsugust metoodikat! Kui kehva enesetunde korral kraadimine annab kehatemperatuuriks 37.6 ºC, siis uuesti kraadimisel sama tulemuse saamine ei muuda tõenäosust, et sul on gripp, mitte sa pole lihtsalt külmetanud!

<u>Ülesanne 1.15.</u> Seadmel on ülekuumenemise eest hoiatav lamp, mis peaks süttima, kui seade kuumeneb üle. Tähistame *W*-ga sündmust, et „hoiatustuli süttib" ning *F*-ga, et „seade on ülekuumenenud". Seadme ülekuumenemise tõenäosus on 0,1; tõenäosus, et kui seade on tõesti üle kuumenenud, siis süttib hoiatuslamp, on 0,99. Tõenäosus, et lamp süttib, kui seade pole ülekuumenenud, on 0,02. a) Leia tõenäosus, et hoiatustuli on süttinud; b) Leia tingimuslik tõenäosus, et kui lamp süttib, et siis on seade üle kuumenenud.

Lahendus: Lahenda ise (osa vastused on antud)

Lambi süttimise tõenäosus on 11,7%

Lambi süttides on seade ülekuumenenud 84,6% juhtumitest

2 Kombinatoorika

Kombinatoorika valemeid me ei tõesta, eesmärgiks on meenutada/õppida nende valemite rakendamist praktiliste ülesannete lahendamisel.

2.1 Kombinatoorika ilma korduvate elementideta

Permutatsioonid on n elemendilise hulga elementidest moodustatud n-elemendilised järjestatud osahulgad.

Permutatsioonide arv leitakse valemiga $P_n = n!$

Kirjutist $n!$ loetakse – "n faktoriaalis" ja arvutatakse järgmise reegli järgi:

$\boldsymbol{n! = 1 \cdot 2 \cdot 3 \ldots (n - 1) \cdot n}$**.**

Jätke meelde, et $0! = 1$ ja $1! = 1$.

Näited:
1) $1! = 1$, $3! = 1 \cdot 2 \cdot 3 = 6$ ja $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.

2) Neljast tähest (k, a, r, u) on võimalik moodustada tähtede ümberpaigutamise teel $4! = 24$ erinevat sõna.

3) 13 õpilasega klassis on võimalik teha $13! = 6227020800$ erineva järjestusega õpilaste nimekirja.

Kombinatsioonid *n*-elemendist *k*-kaupa on n-elemendilise hulga k-elemendilised osahulgad.

Arvutusvalem on: $C_n^k = \dfrac{n!}{k!(n-k)!}$.

Kombinatsioonide arvu leidmisel elementide järjestus pole oluline, s.t. kui kombinatsioon {Jüri, Mari} on olemas, siis {Mari, Jüri} eraldi kombinatsioonina arvesse ei lähe.

Näide 2.1. kümnest inimesest on võimalik moodustada erinevaid kolmeliikmelisi rühmi $C_{10}^3 = \dfrac{10!}{3!(10-3)!} = \dfrac{10!}{3!7!} = 120$.

Näide 2.2. 30 õpilasega klassis on võimalik kaks korrapidajat ametisse määrata $C_{30}^2 = \dfrac{30!}{2!(30-2)!} = \dfrac{30!}{2!28!} = 435$ erineval viisil.

Variatsioonid n elemendist k kaupa on n-elemendilise hulga k-elemendilised järjestatud osahulgad.

Arvutusvalem on: $V_n^m = \dfrac{n!}{(n-m)!}$

Näide 2.3. 30 lehekandja hulgast on võimalik ametisse määrata lehekandja ja vanemlehekandja $V_{30}^2 = \frac{30!}{(30-2)!} = \frac{30 \cdot 29 \cdot 28!}{28!} = 870$ erineval viisil;

Näide 2.4. Kuueliikmelisest võistkonnast saab neli teatesuusatajat (koos järjestusega) välja valida $V_6^4 = \frac{6!}{(6-4)!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2!}{2!} = 360$ erineval viisil. Kui järjestus pole oluline, siis on tegemist kombinatsioonidega, ehk siis $C_6^4 = \frac{6!}{(6-4)! \cdot 4!} = \frac{6 \cdot 5 \cdot 4!}{2! \cdot 4!} = \frac{30}{2} = 15$

Liitmise reegel - kui mingi elemendi A saab valida *n* erineval viisil ja elemendi B saab valida *m* erineval viisil, siis elemendi "kas A või B" saab valida

***m* + *n* erineval viisil.**

Korrutamise reegel - kui mingi elemendi A saab valida *n* erineval viisl ja elemendi B saab valida *m* erineval viisil, siis elementide paari "A ja B" saab valida

***m* · *n* erineval viisil.**

NB! A ja B valikud peavad olema sõltumatud!!!

Näide 2.5. Martil on kuus "Jaguari", üheksa "Volvot" ja kolm päevinäinud "Moskvitši". Töölesõiduks on tal valida

6 + 9 + 3 = 18 auto vahel;

Näide 2.6. Jürkal on kooliminekuks võimalik valida nelja ülikonna, kolme jope ja viie paari kingade vahel. Jürkal on valikuvõimalusi kokku

$4 \cdot 3 \cdot 5 = 60$.

Näide 2.7. Maril on kooliminekuks võimalik valida nelja kleidi, kolme mütsi ja viie paari kingade vahel. Samas iga müts ei sobi iga kleidiga ning iga kleit iga kingaga. Seetõttu on valikuvõimalusi oluliselt vähem kui 60.

ÜLESANDED

Ülesanne 2.1. Klassis on 6 poissi. Kui mitmel viisil saab neid kehalise kasvatuse õpetaja tunni algul rivvi panna? Kas ta jõuab kõik järjestused 45 minutiga läbi proovida, kui iga ümberjärjestuse peale kulub 5 sekundit?

Lahendus: Lahenda ise (vastus on 720, aega kuluks 60 minutit).

Ülesanne 2.2. Klassi kokkutulekule saabus 32 vana klassikaaslast ning iga tulija kallistas varem kohalejõudnut. Mitu kallistust tehti kokku?

Lahendus: Lahenda ise (vastus on 496).

Ülesanne 2.3. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on 2 õpiku võtmiseks?

Lahendus: Lahenda ise (vastus on 66).

Ülesanne 2.4. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on 2 matemaatikaõpiku võtmiseks?

Lahendus: Lahenda ise (vastus on 10).

Ülesanne 2.5. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on ühe matemaatikaõpiku ja ühe füüsikaõpiku võtmiseks?

Lahendus: Lahenda ise (vastus on 35).

Ülesanne 2.6. Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on mõlemad viiekroonised?

Lahendus: Lahenda ise (vastus on ligikaudu 0,28).

Ülesanne 2.7. Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on need erineva nominaalväärtusega?

Lahendus: Lahenda ise (vastus on ligikaudu 0,56).

<u>Ülesanne 2.8.</u> Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on need sama nominaalväärtusega?

Lahendus: Lahenda ise (vastus on ligikaudu 0,44).

<u>Ülesanne 2.9.</u> Lumivalgekesel on 7 sõpra. Mitu erinevat seltskonda saab Lumivalgeke endale külla kutsuda?

Lahendus: Lahenda ise (vastus on 127).

<u>Ülesanne 2.10.</u> Viking Lotol on 48 numbrit (kuuli) ja võidud on 3, 4, 5 ja 6 õiget numbrit (6 on ‚Jackpot'). Tuletada vastavad võidutõenäosused:

Lahendus:

Arvutame, mitu erinevat võimalust on masinal valida 6 kuuli. Tegemist on kombinatsioonidega, sest järjestus pole oluline:

$$n = C_n^m = \frac{n!}{(n-m)!\cdot m!} = \frac{48!}{(48-6)!\cdot 6!} =$$

$$\frac{48\cdot 47\cdot 46\cdot 45\cdot 44\cdot 43\cdot 42!}{42!\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2} = \frac{8835488640}{720} = 12271512$$

Seega võib tulla ca 12,3 miljonit erinevat võimalust.

Kui oleme mingi konkreetse 6-se valiku teinud, mis sisaldab kolme soovitud kuuli, siis kõikide erinevate võimaluste arv sellise tulemuse saamiseks on:

$$m_3 = C_6^3 \cdot C_{42}^3 = \frac{6!}{(6-3)!\cdot 3!}\cdot\frac{42!}{(42-3)!\cdot 3!} = \frac{6\cdot 5\cdot 4\cdot 3!}{3\cdot 2\cdot 3!}\cdot\frac{42\cdot 41\cdot 40\cdot 39!}{39!\cdot 3\cdot 2} = 20\cdot 11480 = 229600$$

$$m_4 = C_6^4 \cdot C_{42}^2 = \frac{6!}{(6-4)!\cdot 4!}\cdot\frac{42!}{(42-2)!\cdot 2!} = \frac{6\cdot 5\cdot 4!}{2\cdot 4!}\cdot\frac{42\cdot 41\cdot 40!}{40!\cdot 2} = 15\cdot 861 = 12915$$

$$m_5 = C_6^5 \cdot C_{42}^1 = \frac{6!}{(6-5)!\cdot 5!}\cdot\frac{42!}{(42-1)!\cdot 1!} = \frac{6\cdot 5!}{1\cdot 5!}\cdot\frac{42\cdot 41!}{41!\cdot 1} = 6\cdot 42 = 252$$

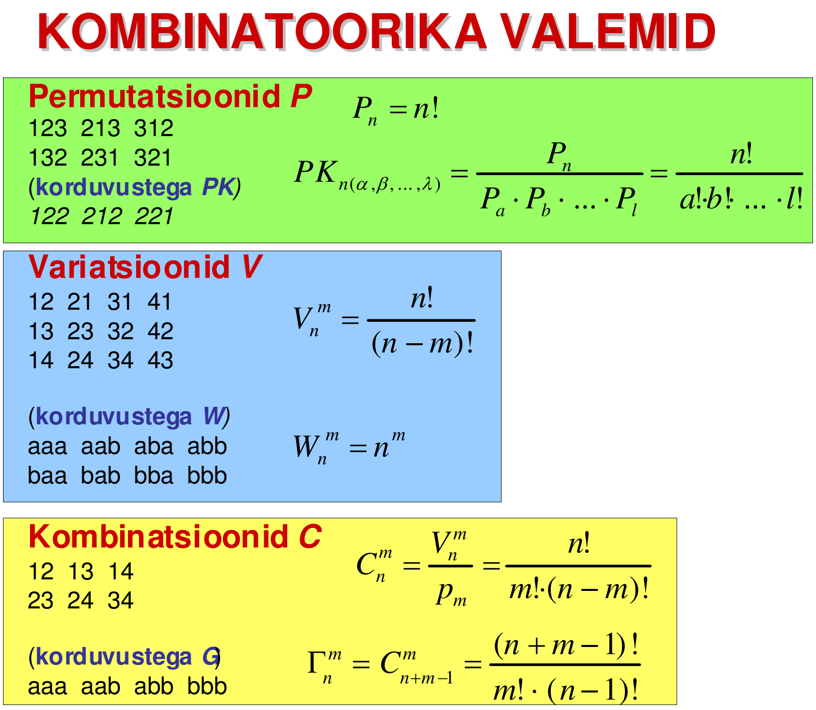
$$p_3 = \frac{m_3}{n} = \frac{229600}{12271512} \approx \frac{1}{53,4}$$

$$p_4 = \frac{m_4}{n} = \frac{12915}{12271512} \approx \frac{1}{950,2}$$

$$p_5 = \frac{m_5}{n} = \frac{252}{12271512} \approx \frac{1}{48696}$$

$p_6 = 1/(12{,}3 \cdot 10^6)$.

2.2 Kombinatoorika koos korduvate elementidega



Joonis 2.1. Kombinatoorika valemid koos lihtsa näitega.

<u>ÜLESANDED</u>

<u>Ülesanne 2.11.</u> Kui mitu erinevat kolmest saiast koosnevat lõunaoodet saab kohvikus tellida, kui menüü sisaldab 11 erinevat nimetust saiasid?

Lahendus:

Selgitame kõigepealt, millal lugeda kahte lõunaoodet (ühendit) teineteisest erinevateks. Antud juhul on erinevad näiteks tellimused ABC ja ABD (erinevad elementide eneste poolest), aga samuti AAB ja ABB (erinevad tellimuste proportsioonide poolest, kuid sisaldavad endas korduvaid elemente). Kokkulangevateks tuleb aga lugeda näiteks tellimused ABC ja BCA, sest tellimuse esitamise järjekord pole oluline. Seega selleks, et leida erinevate võimalike lõunaoodete arvu, tuleb leida kõigi võimalike kordumistega kombinatsioonide arv 11 elemendist 3 kaupa, s.t.

$$\Gamma_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m! \cdot (n-1)!} = \frac{(11+3-1)!}{3! \cdot (11-1)!} = \frac{13!}{3! \cdot 10!} = \frac{13 \cdot 12 \cdot 11}{3 \cdot 2} = 286$$

Ülesanne 2.12. Kui mitmel erineval viisil saab nimes TEELE tähti selliselt ümber paigutada, et kolm tähte E ei satuks kõrvuti?

Lahendus:

Leiame kõigepealt mitu võimalust on nime TEELE tähti erinevalt järjestada. Oleksid kõik tähed erinevad, oleks lahenduseks permutatsioonid, aga kuna esineb korduvusi, siis tuleb kasutada üldisemat valemit – kordumistega permutatsioonide valemit. Tähti T ja L esineb üks kord, tähte E aga kolm korda. Pannes selle sisse valemisse, saame, et:

$$P_{n(1,1,3)} = \frac{5!}{1!\cdot 1!\cdot 3!} = \frac{5\cdot 4\cdot 3!}{3!} = 20$$

Leiame nüüd järjestuste arvu, milles on kolm E-tähte kõrvuti. Selleks tähistame variandi EEE tähega F. Nüüd on vaja leida permutatsioonide arv tähtedest T, L ja F, seega permutatsioonide arv kolmest, mis on 3! = 6.

Seega on kokkuvõtteks võimalik nime TEELE tähti selliselt ümber paigutada, et kolm tähte E ei satuks kõrvuti 20 – 6 = 14.

Ülesanne 2.13. Auto registreerimisnumber koosneb kolmekohalisest arvust (ka arv 031 loetakse kolmekohaliseks) ja kolmetähelisest sõnast. Mitu autot on võimalik registreerida, kui kasutatakse 25-tähelist tähestikku?

Lahendus:

Numbrite erinevate kombinatsioonide arv on korduvustega variatsioonide arv:

$$W_n^m = n^m = 10^3 = 1000$$

samuti on tähtede erinevate kombinatsioonide arv:

$$W_n^m = n^m = 25^3 = 15625$$

Seega on võimalik registreerida

$$N = 15625\cdot 1000 = 15625000,$$

seega on võimalik registreerida ca 15 miljonit autot.

Ülesanne 2.14. Ühiselamu toas elab 3 üliõpilast. Neil on kokku 4 erinevat tassi, 5 erinevat alustaldrikut ning 6 erinevat teelusikat. Kui mitmel viisil saavad nad katta laua hommikuseks kohvijoomiseks, kui nõude paigutus laual pole oluline?

Lahendus:

Tassid, alustaldrikud ning teelusikad on omavahel sõltumatud, seega on tegemist kombinatsioonidega:

$$N_{tassid} = C_n^m = \frac{n!}{m!\cdot(n-m)!} = \frac{4!}{3!\cdot(4-3)!} = \frac{4\cdot 3!}{3!\cdot 1!} = 4$$

$$N_{alustaldrikud} = C_n^m = \frac{n!}{m!\cdot(n-m)!} = \frac{5!}{3!\cdot(5-3)!} = \frac{5\cdot 4\cdot 3!}{3!\cdot 2!} = 10$$

$$N_{teelusikad} = C_n^m = \frac{n!}{m!\cdot(n-m)!} = \frac{6!}{3!\cdot(6-3)!} = \frac{6\cdot 5\cdot 4\cdot 3!}{3!\cdot 3!} = 20$$

Kokku on erinevate võimaluste arv $N = 4\cdot 10\cdot 20 = 800$

<u>Ülesanne 2.15.</u> Kui mitmel viisil saab jagada 36-kaardilise kaardipaki kaheks osaks nii, et mõlemad osad sisaldaks 2 ässa?

Lahendus:

Kuna mõlemas kaardipakis peab olema 2 ässa, siis ülejäänud kaardid võivad jaguneda suvaliselt. Põhimõtteliselt on tegemist kombinatsioonide arvuga, kus elementide arv *m* muutub nullist 32-ni (sest ässasid ei varieerita):

$$N1 = \sum_{i=0}^{32} C_n^i = \sum_{i=0}^{32} \frac{32!}{i!\cdot(32-i)!} = 32!\cdot\sum_{i=0}^{32} \frac{1}{i!\cdot(32-i)!} = 4{,}295\cdot 10^9$$

Nüüd võtame veel arvesse, et 4 marki ässasid saab ka varieerida:

$$N2 = C_n^i = \frac{4!}{2!\cdot(4-2)!} = \frac{4\cdot 3\cdot 2}{2\cdot 2} = 6$$

Seega saab jagada 36-kaardilise kaardipaki kaheks osaks nii, et mõlemad osad sisaldaksid 2 ässa kokku $N = N1\cdot N2 = 2{,}577\cdot 10^{10}$.

Arvutame huvi pärast välja võimaluste arvu erijuhul, kui mõlemal mängijal on võrdne arv kaarte:

$$N = C_{32}^{16}\cdot C_4^2 = \frac{32!}{16!\cdot 16!}\cdot\frac{4!}{2!\cdot 2!} = 601080390\cdot 6 = 3{,}6\cdot 10^9.$$

<u>Ülesanne 2.16.</u> Viis sõpra otsustavad teha lauajalgpalli turniiri, nii et mängitakse paarides ja kõik võimalikud paarid mängiksid kõigi võimalike paaridega korra läbi. Kui oleks neli mängijat, siis oleks kolm mängu (1, 2 – 3, 4), (1, 3 – 2, 4) ja (1, 4 – 2, 3). Mitmest mängust koosneb turniir 5-kesi mängides? Mitu mängu tuleks mängida sama süsteemi järgi, kui oleks 8 sõpra?

Lahendus:

Lahenda ise. Vastused on $n(5) = 15$; $n(8) = 210$.

3 Juhusliku suuruse jaotused

Juhuslik suurus on suurus, mis katse tulemusel omandab ühe või teise, varem mitteteadaoleva väärtuse mingist väärtuste hulgast. Viimast hulka nimetatakse juhusliku suuruse võimalike väärtuste hulgaks.

Juhuslikuks suuruseks on näiteks:

- täringu viskamisel saadud silmade arv,
- telefonikeskjaama saabuvate kutsungite arv,
- spordivõistlustel kaugushüppe võitja tulemus,
- elektripirni põlemiskestus,
- gaasimolekuli kiirus,
- välisõhu temperatuur (mingil kindlal ajahetkel) Tartus Raekoja Platsil.

Siin kahel esimesel suurusel on naturaalarvulised, kolmel järgneval – reaalarvulised positiivsed võimalikud väärtused, viimasel – nii positiivsed kui negatiivsed (kui mõõdame Celsiuse kraadides).

Tähistame juhuslikke suurusi ladina tähestiku suurtähtedega X, Y, ...; nende võimalikke väärtusi aga vastavate väiketähtedega x, y, ... Kui katse tulemusel juhuslik suurus X omandab väärtuse $X = x$, siis öeldakse ka, et x on juhusliku suuruse X realisatsioon. Üldiselt nimetatakse juhuslikuks suuruseks suurust X, kui iga reaalväärtuselise x korral vahemikus

$$-\infty < x < \infty$$

on määratud või antud tõenäosus leida juhuslik suurus punktist x vasemal $P(X < x)$.

Liigitame juhuslikke suurusi diskreetseteks ja pidevateks suurusteks (võimalikud on ka segatüüpi juhuslikud suurused, aga neid selles kursuses ei vaadelda).

Juhuslikku suurust nimetatakse **diskreetseks**, kui selle võimalikeks väärtusteks on üksikud, diskreetsed või isoleeritud väärtused. Võimalike väärtuste hulk on siin lõplik või loenduv hulk. Näiteks täringuviske tulemused saavad olla 1, 2, 3, 4, 5 või 6, teisi väärtuseid ei saa täringuviskel tulla.

Pideva juhusliku suuruse all mõistame niisugust suurust, mille võimalike väärtuste hulk on reaaltelg $-\infty < x < \infty$ või selle mingi alamhulk, näiteks lõik [a, b]. Näiteks tudengi esimene ärkamisaeg on pidev juhuslik suurus alumise piiriga kell 0 ja ülemise piiriga kell 24. Oluline on tähele panna, et erinevaid võimalikke ärkamisaegasid on lõpmatult, sest aega saab (teoreetiliselt) jagada lõpmatult paljudeks ajahetkedeks.

Juhuslikke suurusi iseloomustatakse jaotusseadustega. Öeldakse, et jaotusseadus ehk jaotus ehk tõenäosusjaotus on antud, kui on määratud:

1) juhusliku suuruse võimalike väärtuste hulk,

2) eeskiri, mis võimaldab leida tõenäosust, et juhusliku suuruse väärtus on mingis piirkonnas. Jaotusseaduse enamlevinumateks ja praktilisteks vormideks on jaotustabel, jaotustihedus ja integraalne jaotusfunktsioon.

Kui juhuslikul suurusel on antud jaotusseadus, siis öeldakse, et juhuslik suurus on antud jaotusega või juhuslik suurus allub antud jaotusele.

3.1 Valik diskreetse juhusliku suuruse jaotusseaduseid

Vaatleme diskreetset juhuslikku suurust X võimalike väärtustega x_1, x_2, ..., x_n. See, et juhuslik suurus X omandab ühe neist väärtustest, näiteks $X = x_2$, on tüüpiline juhuslik sündmus mingi tõenäosusega. Tähistame tõenäosused

$$P(X = x_k) = p_k; \qquad (k = 1, 2, \ldots n),$$

mida nimetame diskreetse juhusliku suuruse X **tõenäosusfunktsiooniks**. Ilmselt kehtib

$$\sum_{k=1}^{n} p_k = 1, \tag{3.1}$$

kuna sündmused $X = x_k$ on ainuvõimalikud, üksteist välistavad ja moodustavad täissüsteemi.

Hulka X: x_1, x_2, ..., x_n võib vaadelda ka elementaarsündmuste ruumina, ainult et klassikalist võrdtõenäosuse eeldust ei tehta.

Võrduse (3.1) paremal poolel olev arv 1 on seega jaotatud juhusliku suuruse võimalike väärtuste vahel. Siit tuleneb ka nimetus **jaotus**.

Diskreetse juhusliku suuruse jaotusseaduse lihtsaimaks esitamisvormiks on tabel, kus on loetletud võimalikud väärtused ja neile vastavad tõenäosused.

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Niisugust tabelit nimetatakse juhusliku suuruse **jaotustabeliks.**

Praktikas sobib kasutada jaotustabelit, kui võimalike väärtuste arv ei ole suur. Mõnedel juhtudel saab jaotusseaduse anda ka valemi kujul

$$p_k = P(X = x_k) = g(x_k).$$

Jaotustabeli võib esitada ka graafiliselt, märkides juhusliku suuruse võimalikud väärtused x-teljel ja tõenäosused y-teljel.

Näide 3.1. Teelõigul on 4 valgusfoori. Igaüks neist fooridest lubab liiklusvahendi liikumise tõenäosusega p. Leida liiklusvahendi poolt peatuseta (kuni esimese peatuseni) läbitud fooride arvu jaotustabel. Oluline lisaeeldus on, et foorid vilkugu üksteisest sõltumatult!

Olgu juhuslik suurus X liiklusvahendi poolt peatuseta läbitud fooride arv. Ilmselt X võimalikud väärtused on 0 (ei läbi ühtki foori), 1, 2, 3 ja 4. Tähistame tõenäosuse, et liiklusvahend läbib konkreetse foori p-ga, ja tõenäosuse et ei läbi konkreetset foori q-ga: $q = 1 - p$. Siis tõenäosused avalduvad järgmiselt:

Tõenäosus, et ei läbita ühtki foori:

$$P(X = 0) = q.$$

Tõenäosus, et läbitakse esimene foor, kuid peatutakse teise ees:

$$P(X = 1) = p\,q.$$

Tõenäosus, et läbitakse esimene ja teine foor, kuid peatutakse kolmanda ees:

$$P(X = 2) = p^2\,q.$$

Tõenäosus, et läbitakse esimene, teine ja kolmas foor, kuid peatutakse neljanda ees:

$$P(X = 3) = p^3\,q.$$

Tõenäosus, et peatuseta läbitakse kõik neli foori:

$$P(X = 4) = p^4.$$

Niisiis on jaotustabel kujul

X	0	1	2	3	4
P	q	pq	p^2q	p^3q	p^4

Lihtne on kontrollida, et kehtib võrdus (3.1):

$$\sum_{k=0}^{4} p_k = q(1 + p + p^2 + p^3) + p^4 = (1 - p)(1 + p + p^2 + p^3) + p^4 = 1\,.$$

Konkreetsetel juhtudel $p = 0.1$; $p = 0.5$ ja $p = 0.9$ saame

X	0	1	2	3	4
$P = 0.1$	0,9	0,09	0,009	0,0009	0,0001
$P = 0.5$	0,5	0,25	0,125	0,0625	0,0625
$P = 0.9$	0,1	0,09	0,081	0,0729	0,6561

3.1.1 Binoomjaotus ehk Bernoulli jaotus

Vaatleme n sõltumatut katset, kui igal katsel toimub sündmus A konstantse tõenäosusega p. Sündmuse A toimumise tõenäosus üksikkatsel on p, mittetoimumise tõenäosus aga on $q = 1 - p$. Kui teeme n katset ja saame mingis konkreetses järjestuses sündmuse A toimumise k korda, ning järelikult, vastandsündmuse $\overline{A}$ toimumise $n - k$ korda, siis on sellise tulemuse saamise tõenäosus $p^k q^{n-k}$. Kokku on erinevaid võimalusi saada erinevas järjestuses k soodsat katset võrdne kombinatsioonide arvuga n liikmest k kaupa, s.o C_n^k -ga. Seega

$$P(X = k) = P_n(k) \equiv C_n^k p^k q^{n-k}\,. \qquad (3.2)$$

<u>ÜLESANDED</u>

<u>Ülesanne 3.1.</u> Testis on 10 küsimust, igal küsimusel on 4 valikvastust, millest üks on õige. Kui vastata huupi, siis tulemus allub Bernoulli jaotusele, kus õige vastuse tõenäosus on $P(1) = ¼$. Mis on tõenäosus, et testis saame k õiget vastust? Mis on tõenäosus, et testis saame huupi vastates 5 või 6 õiget vastust?

Lahendus:

Kui $p = ¼$, siis $q = 1 - p = ¾$.

$$P(k) = C_n^k p^k q^{n-k} = \frac{10!}{k! \cdot (10-k)!} \cdot \left(\frac{1}{4}\right)^k \cdot \left(\frac{3}{4}\right)^{10-k} = \frac{10!}{4^{10}} \frac{3^{10-k}}{k! \cdot (10-k)!} = 3{,}46 \cdot \frac{3^{10-k}}{k! \cdot (10-k)!}$$

$$P(5) = 3{,}46 \cdot \frac{3^{10-k}}{k! \cdot (10-k)!} = 3{,}46 \cdot \frac{3^5}{5! \cdot 5!} = 0{,}058$$

$$P(6) = 3{,}46 \cdot \frac{3^{10-k}}{k! \cdot (10-k)!} = 3{,}46 \cdot \frac{3^4}{6! \cdot 4!} = 0{,}016$$

Seega on tõenäosus, et testis saadakse huupi vastates 5 või 6 õiget vastus, P = 0,058 + 0,016 = 0,074 = 7,4 %.

<u>Ülesanne 3.2.</u> Estonian Air teostab lende Tartu-Stockholm 33-kohalise lennukiga Saab-340. Pileteid müüakse välja maksimaalselt 35, sest eeldatakse, et kõik pileti ostnud reisijad lennukile ei jõua. Kui juhtub olukord, kus kohale tuleb rohkem inimesi kui on lennukis kohti, peab lennufirma mahajääjale kompensatsiooni maksma, lisaks ka lennufirma maine langus, seega oletame, et lennufirma arvestab, et 95 % ülemüüdud lendude juhtumitest õigustavad ennast. Mis on lennufirma arvutustes tõenäosus, et pileti ostnud inimene ei ilmu lennukile?

Lahendus:

Teeme eelduse, et alati müüakse ülemüüdud lendudele 35 piletit, sest lennufirma ei saa eeldada, et näiteks pooltel ülemüügi kordadel müüakse 34 piletit.

Lennule jõudmise tõenäosust tähistame p-ga, mittejõudmist q-ga. Leiame tõenäosuse, et lennule jõuab 34 või 35 inimest, s.t. lennufirmal on probleemid, võrdsustame selle viie protsendiga, et ülemüük ei õigusta ennast:

$$P(34) = C_n^k p^k q^{n-k} = \frac{35!}{34! \cdot (35-34)!} \cdot p^{34} \cdot q^1 = 35 \cdot p^{34} \cdot q^1$$

$$P(35) = C_n^k p^k q^{n-k} = \frac{35!}{35! \cdot (35-35)!} \cdot p^{35} \cdot q^0 = p^{35}$$

$$P(34;35) = 35 \cdot p^{34} \cdot q^{1} + p^{35} = 0{,}05$$

$$35 \cdot p^{34} \cdot (1-p)^{1} + p^{35} = 0{,}05$$

$$35 \cdot p^{34} - 34 \cdot p^{35} = 0{,}05$$

$$p^{34}(35 - 34p) = 0{,}05$$

Sellest võrrandist p täpselt leidmine ei ole lihtne, samas on ülesanne üsna kergesti lihtsalt proovimise teel lahendatav, näiteks Mathcadis, sealt sain, et $p = 0{,}871498$. Seega eeldab lennufirma, et keskmiselt 87 % pileti ostnutest jõuab lennukile.

<u>Ülesanne 3.3.</u> Arendame edasi ülesannet 3.2. Estonian Air'i Boeing 737-500 võtab pardale 120 reisijat. Kui müüakse üle samuti 2 piletit, siis mis on tõenäosus, et kõik lennule jõudnud ei mahu pardale?

Lahendus:

Kasutame eelmises osas saadud tõenäosust ning saadud valemit:

$$p^{121}(122 - 121p) = 0{,}87^{121}(122 - 121 \cdot 0{,}87) = 8 \cdot 10^{-7}$$

Seega selle lennuki puhul ülemüügi juhtumitest praktiliselt ei tohiks lennufirmale probleeme tekkida.

<u>Ülesanne 3.4.</u> Pood võtab müüki 1000 lampi. Tõenäosus, et juhuslikult valitud lamp on katki, on $p = 0{,}1$ %. Tähistame X-ga katkiste lampide arvu. Leida tõenäosus, et

a) kõik lambid on terved;

b) üks lamp on katki;

c) katki on rohkem kui 2 lampi.

Lahendus:

$$P(X = k) = C_n^k p^k q^{n-k} = \frac{1000!}{k! \cdot (1000-k)!} \cdot 0{,}001^{k} \cdot 0{,}999^{1000-k}$$

$$P(X = 0) = C_n^k p^k q^{n-k} = \frac{1000!}{0! \cdot (1000-0)!} \cdot 0{,}001^{0} \cdot 0{,}999^{1000} = 0{,}368$$

$$P(X = 1) = C_n^k p^k q^{n-k} = \frac{1000!}{1! \cdot (1000-1)!} \cdot 0{,}001^{1} \cdot 0{,}999^{999} = 0{,}368$$

$$P(X=2)=C_n^k p^k q^{n-k}=\frac{1000!}{2!\cdot(1000-2)!}\cdot 0{,}001^2\cdot 0{,}999^{998}=0{,}184$$

$$P(X>2)=1-P(X=0)-P(X=1)-P(X=2)=$$
$$=1-0{,}368-0{,}368-0{,}184=0{,}080$$

Huvi pärast arvutame välja ka variandid, et katki läheb 3 või 4 lampi:

$$P(X=3)=C_n^k p^k q^{n-k}=\frac{1000!}{3!\cdot(1000-3)!}\cdot 0{,}001^3\cdot 0{,}999^{997}=0{,}061$$

$$P(X=4)=C_n^k p^k q^{n-k}=\frac{1000!}{4!\cdot(1000-4)!}\cdot 0{,}001^4\cdot 0{,}999^{996}=0{,}015$$

On näha, et iga järgnev liige on eelmisest ligikaudu k korda väiksem, tingituna faktoriaali liikmest.

3.1.2 Geomeetriline jaotus

Olukord nagu eelmisel Bernoulli katsete juhul, et sündmuse A toimumise tõenäosus üksikkatsel on p, mittetoimumise tõenäosus aga on $q=1-p$. Nüüd on katseseeria lõpmatu, ja küsitakse, milline on tõenäosus, et esimesed $x-1$ korda tulevad $\overline{A}$, s.o. sündmust ei toimu, ja soodne sündmus A leiab aset x-ndal katsel.

Geomeetrilise jaotuse tõenäosusjaotus on:

$$P(X=k)=q^{k-1}p \qquad (3.3)$$

Ülesanded

Ülesanne 3.5. Langevarju hüpetel on alati on tõenäosus, et langevari ei avane. Oletame, et see tõenäosus on $p = 0{,}1$ %. Arvutada tõenäosus, et a) langevari ei avane 5-ndal hüppel; b) langevari ei avane 1000-ndal hüppel.

Lahendus:

Sündmuseks nimetame, et langevari ei avane. esimesed $k-1$ korda toimub vastandsündmus, s.t. langevari avaneb, selle tõenäosus on q^{k-1}, k-s kord toimub sündmus, s.t. langevari ei avane, selle tõenäosus on p:

$$P(X=1)=q^{k-1}p=0{,}999^0\cdot 0{,}001=0{,}001$$
$$P(X=2)=q^{k-1}p=0{,}999^1\cdot 0{,}001=0{,}000999$$
$$P(X=5)=q^{k-1}p=0{,}999^5\cdot 0{,}001=0{,}000995$$
$$P(X=1000)=q^{k-1}p=0{,}999^{999}\cdot 0{,}001=0{,}000367$$

Näeme, et 1000-ndal hüppel on langevarju mitteavanemine juba palju ebatõenäolisem kui 5-ndal hüppel. See on aga tingitud lihtsast faktist, et tõenäosus, et langevari pole avanenud juba varem, järjest kasvab. Näiteks tõenäosus, et langevari ei avane teisel hüppel on väiksem tõenäosusest, et langevari ei avane esimesel hüppel seetõttu, et on olemas väike tõenäosus, et langevari ei avane juba esimesel hüppel.

Ülesanne 3.6. Oletame, et metsas lõkke süütamisel on tõenäosus, et tule saab süüdatud ühe tikuga p = 25%. Arvutada tõenäosus, et a) lõke sütib esimesest tikust; b) lõke sütib viiendast tikust; c) lõke sütib kahekümnendast tikust.

Lahendus:

Sündmuseks nimetame, et lõke süttib. esimesed $k-1$ korda toimub vastandsündmus, s.t. lõke ei sütti, selle tõenäosus on q^{k-1}, k-s kord toimub sündmus, s.t. lõke süttib, selle tõenäosus on p:

$$P(X=1)=q^{k-1}p=0{,}75^0\cdot 0{,}25=0{,}25=25\%$$
$$P(X=5)=q^{k-1}p=0{,}75^4\cdot 0{,}25=0{,}079=7.9\%$$
$$P(X=20)=q^{k-1}p=0{,}75^{20}\cdot 0{,}25=0{,}0011=0.11\%$$

Näeme, et 20. tikuga lõkke süütamise tõenäosus on üle 200 korra väiksem kui esimesel korral. Seda seetõttu, et tõenäosus, et lõke süttis juba esimese 19 tiku abil, on väga suur.

Ülesanne 3.7. Petsi juures on iga nädal raju pidu, mille käigus käib ta iga kord ka autoga tanklas varusid täiendamas. Oletame, et on 5 % tõenäosus, et jäädakse politseile vahele. Arvutada tõenäosus, et a) ta jääb politseile vahele esimesel peol; b) ta jääb vahele kuu aja pärast 5. peol; c) ta jääb vahele täpselt aasta pärast 52. peol? Teeme eelduse, et pärast politseile vahele jäämist tabab Petsi meelemuutus ning rajud peod lõpevad.

Lahendus:

Sündmuseks nimetame, Pets jääb politseile vahele. esimesed $k-1$ korda toimub vastandsündmus, s.t. ei jää politseile vahele, selle tõenäosus on q^{k-1}, k-s kord toimub sündmus, s.t. jääb politseile vahele, selle tõenäosus on p:

$$P(X=1)=q^{k-1}p=0{,}95^0\cdot 0{,}05=0{,}05$$
$$P(X=5)=q^{k-1}p=0{,}95^4\cdot 0{,}05=0{,}0407$$
$$P(X=52)=q^{k-1}p=0{,}95^{51}\cdot 0{,}05=0{,}0037.$$

3.1.3 Poissoni jaotus

Vaatleme veel kord katseseeriat, milles on n katset ja sündmuse A tõenäosus igal katsel on p. Tõenäosus, et katseseerias sündmus A toimub k korda, avaldub Bernoulli valemiga. Kui n on küllalt suur ja p väga väike, siis valem praktiliseks arvutamiseks ei sobi, kuna on tegemist väga suurte ja väga väikeste arvude korrutamisega. Leiame sobivama valemi.

Eeldame, et korrutis np on konstantne ja tähistame

$$\lambda = np\,, \qquad (3.4)$$

täpsemini

$$\lambda = \lim_{\substack{n\to\infty \\ p\to 0}} np\,.$$

Poissoni jaotuseks nimetatakse jaotust

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k = 0, 1, ...) \qquad (3.5)$$

Jaotust nimetatakse samuti väikeste arvude seaduseks ehk harva esinevate sündmuste seaduseks, sest see kirjeldab harva esinevaid nähtusi.

Poissoni jaotus annab tõenäosuse saada hästi pikas katseseeria pikkusega n soodne sündmus k-l korral, kui üksiksündmuse tõenäosus on $p = \lambda / n << 1$.

Hinnanguliselt võib binoomjaotust lähendada Poissoni jaotusega, kui $p < 0{,}1$. See aga tähendab, et fikseeritud λ korral peab võtma katseseeria pikkuse tingimusest $n \ge \lambda / p = 10\lambda$. Kui $\lambda = 10$, siis näiteks on usaldusväärne valida n = 100.

ÜLESANDED

Ülesanne 3.8. Kõnekeskus saab minutis keskmiselt 20 kõnet. Leia tõenäosus, et veerand minuti jooksul tuleb a) 3 kõnet; b) 5 kõnet; c) 7 kõnet; d) 10 kõnet? Lei a tõenäosus, et ühe sekundi jooksul tuleb 2 kõnet?

Lahendus:

$\lambda = np = 0{,}25 \cdot 20 = 5$

$$P(X = 3) = e^{-5}\frac{5^3}{3!} = 0{,}14$$

$$P(X = 5) = e^{-5}\frac{5^5}{5!} = 0{,}18$$

$$P(X=7)=e^{-5}\frac{5^7}{7!}=0{,}10$$

$$P(X=10)=e^{-5}\frac{5^{10}}{10!}=0{,}02$$

Tõenäosus, et ühe sekundi jooksul tuleb 2 kõnet, on:

$$\lambda=np=\frac{1}{60}\cdot 20=\frac{1}{3}$$

$$P(X=2)=e^{-\frac{1}{3}}\frac{\left(\frac{1}{3}\right)^2}{2!}=0{,}040$$

Ülesanne 3.9. Kaubamaja tegi kindlaks, et nõudmine teatud fotoaparaadi järele allub Poissoni jaotusele. Ühe nädala jooksul ostetakse keskmiselt kaks aparaati. Kui kord nädalas uuendatakse kaubamaja laoseisu nii, et poes oleks 4 aparaati, siis kui suur on tõenäosus, et nõudmine ületab pakkumise?

Lahendus:

Kuna keskväärtus $EX = 2$, siis Poissoni jaotuse parameeter $\lambda = 2$. Vastuseks ülesandes esitatud küsimusele tuleb leida tõenäosus, et nädalas ostetakse üle nelja aparaadi. Selleks tuleb leida tõenäosus, et tahetakse osta 5 aparaati, 6 aparaati, 7 aparaati jne. See aga tähendaks lõputult palju arvutusi. Lihtsam on arvutada vastandsündmuse tõenäosust, st et ostetakse neli või vähem aparaati.

$$\begin{aligned}
P(X\le 4)&=P(0)+P(1)+P(2)+P(3)+P(4)=\sum_{k=0}^{4}e^{-2}\frac{2^k}{k!}=\\
&=e^{-2}\frac{2^0}{0!}+e^{-2}\frac{2^1}{1!}+e^{-2}\frac{2^2}{2!}+e^{-2}\frac{2^3}{3!}+e^{-2}\frac{2^4}{4!}=\\
&=0{,}1353\frac{1}{1}+0{,}1353\frac{2}{1}+0{,}1353\frac{4}{2}+0{,}1353\frac{8}{6}+0{,}1353\frac{16}{24}=\\
&=0{,}1353+0{,}2707+0{,}2707+0{,}1804+0{,}0902=0{,}9473
\end{aligned}$$

Kuna

$$P(X\le 4)+P(X>4)\equiv 1,$$

siis

$$P(X>4)=1-P(X\le 4)=1-0{,}9473=0{,}0527.$$

Seega on tõenäosus, et nõudmine ületab pakkumise 5,27 %.

Ülesanne 3.10. Tehases on 10000 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?

Lahendus:

Kõigepealt on vaja leida suurus λ, leiame selle vastavalt Poissoni valemile:

$$P(X=0)=e^{-\lambda}\frac{\lambda^0}{0!}=e^{-\lambda}=0{,}9 \mid \ln()$$
$$\ln(e^{-\lambda})=-\lambda=\ln(0{,}9)\approx -0{,}105 \Rightarrow \lambda\approx 0{,}105.$$

Tõenäosus, et üks konkreetne töötaja hukkub on arvutatav seosest

$$\lambda=np \Rightarrow p=\frac{\lambda}{n}=\frac{0{,}105}{10000}=0{,}0000105\ .$$

Tõenäosus, et see konkreetne töötaja ei hukku aasta jooksul tööõnnetusel on:

$$q=1-p=1-0{,}0000105=0{,}9999895\ .$$

Seega on ühe töötaja jaoks tõenäosus q = 0,9999895, et ta ei hukku aasta jooksul tööõnnetusel.

3.2 Jaotusseaduste esitamine

Pidevalt muutuva suuruse X korral on tõenäosus, et tuleb etteantud fikseeritud väärtus $X = x$, võrdne nulliga:

$$P(X=x)=0.$$

Et seda demonstreerida, vaatleme näiteks lõiku [0, 1] ehk $0\le x\le 1$.Teeme oletuse, et juhuslik suurus on jaotunud selle lõigul ühtlaselt. Küsime – kui suur on tõenäosus, et sellest intervallist juhuslikult valitud arv on võrdne arvuga x = 0.3181345232 4234902384 0924723898 78979 ?

Küsimus on sisuliselt mõttetu, sest tõenäosuse arvutamise reegli järgi

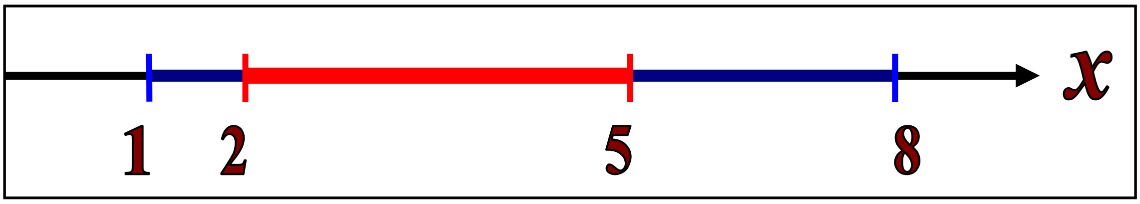
$$P=\frac{\text{soodsad võimalused}}{\text{kõik võimalused}}$$

on tulemuseks

$$P=\frac{1}{\infty}=0\,.$$

Seepärast on mõtet vaid tõenäosusel, et juhusliku suuruse realisatsioon asub mingis intervallis $x_1<X<x_2$, või poollõigul $x_1\le X<x_2$, või rahuldab tingimust $X<x$, kus x on ette antud.

Näide 3.2. Juhuslik suurus x võib omandada võrdse tõenäosusega mistahes väärtusi lõigust [1, 8]. Kui suur on tõenäosus, et juhuslikult valitud x väärtus kuulub alamlõiku [2, 5] (joonis 3.1)



Joonis 3.1.

Otsitav tõenäosus määratakse vastavate lõikude pikkuste suhtena

$$p = \frac{\text{soodsa piirkonna pikkus}}{\text{kogu piirkonna pikkus}} = \frac{5-2}{8-1} = \frac{3}{7} = 0{,}428\,.$$

Siin on ilmselt tegemist geomeetrilise tõenäosuse juhtumiga, kus geomeetriliseks kujundiks on sirglõik ning juhusliku sündmuse realiseerumise tõenäosus on võrdeline selle lõigu pikkusega.

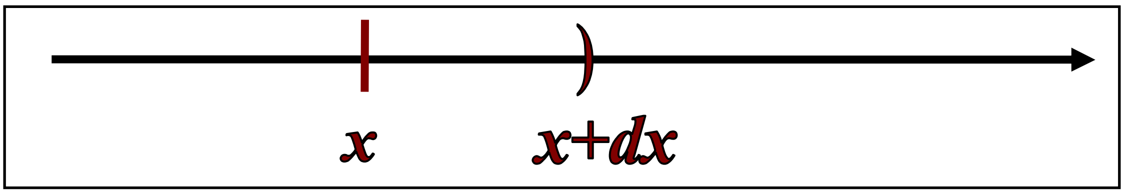
Toodud näites ei olnud soodsas piirkonnas [2, 5] eelistatumaid ega vähemeelistatumaid alasid. Praktilistes ülesannetes aga harilikult on.

Näide 3.3. Kujutagu X endast patsiendi kehatemperatuuri ºC**,** siis praktiliselt kõik võimalikud x väärtused katab piirkond [34, 42]. Selles piirkonnas on aga temperatuuride alampiirkond [36, 38] üldiselt hoopis tõenäosem kui [40, 42].

Näide 3.4. Kujutagu x endast täiskasvanud mehe kehakaalu (korrektsem oleks öelda massi!) kilogrammides. Kogu lubatav x väärtuste piirkond on hinnanguliselt [10, 300]. Samas on ilmne, et alampiirkond [60, 90] on palju tõenäosem kui sama lai piirkond [160, 190].

Juhusliku suuruse võimalike väärtuste erinevate piirkondade tõenäosust väljendab **juhusliku suuruse jaotustihedus.** Kasutatakse ka termineid **tõenäosustihedus, tihedusfunktsioon, jaotusfunktsioon.** NB! Termin **jaotusfunktsioon** on levinud ja kasutusel füüsikute ja keemikute hulgas ning vastavate suundade erialakirjanduses. Matemaatikud tähistavad selle terminiga tavaliselt suurust, mida käesolevas kursuses (hiljem, allpool) tähistame terminiga **integraalne jaotusfunktsioon.** Seepärast peab olema termini **jaotusfunktsioon** kasutamisega ettevaatlik ja alati täpsustama, mida selle termini all konkreetselt silmas peetakse. <u>Käesolevas kursuses on **jaotusfunktsioonil** alati juhusliku suuruse **integraalse jaotusfunktsiooni** tähendus.</u>

Vaatleme reaalteljel (lõpmata) lühikest poolintervalli $[x,\, x + dx)$, mis sisaldab vasema otspunkti x, kuid ei sisalda parempoolset otspunkti $x + dx$ (vt joonis 3.2.).



Joonis 3.2.

Poolintervalli kasutamine on tarvilik, et üks ja sama punkt ei oleks korraga mitmes alamhulgas. Vaadeldes kaht kõrvutiseisvat poolintervalli $[x, x + dx)$ ja $[x + dx, x + dx + dx_1)$ on kindel, et nende kokkupuutepunkt $x + dx$ kuulub parempoolsele intervallile, kuid ei kuulu vasempoolsele, ning samal ajal ei ole ka olukorda, kus see punkt ei kuuluks kummalegi alamhulgale.

Tähistame tõenäosuse, et pideva muutkonnaga juhuslik suurus X satuks sellesse poolintervalli $[x, x + dx)$, sümboliga dP. Ilmselt on tõenäosus dP võrdeline elementaarlõigu pikkusega dx:

$$dP = f(x)dx\,. \tag{3.6}$$

Võrdetegurit $f(x)$ nimetatakse juhusliku suuruse tõenäosustiheduseks e. **jaotustiheduseks.** Vastavalt toodud valemile võib kirjutada

$$f(x) = \frac{dP}{dx},$$

s.o tõenäosustihedus on tõenäosus, et X satub punkti x lähiümbrusesse, arvutatuna ühikulise intervalli kohta.

Jaotustihedus määrab pideva juhusliku suuruse tõenäosuslikus tähenduses täielikult ja üheselt – mingeid muid täiendavaid karakteristikud enam vaja ei ole.

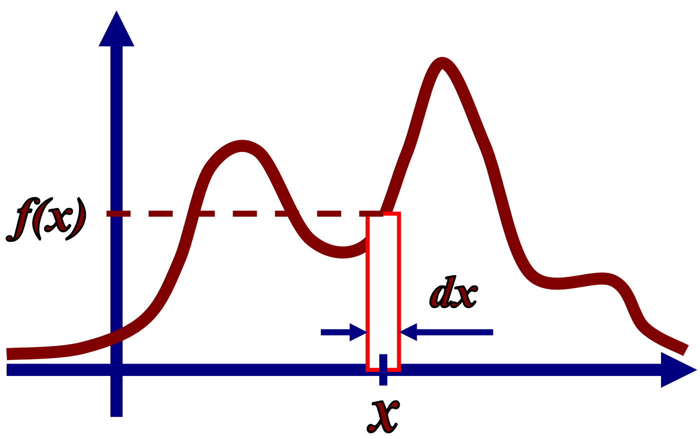
3.2.1 Jaotustihedus

Jaotustiheduse omadused:

1. Mittenegatiivsus:

$$f(x) \geq 0\,.$$

See omadus järeldub määrangust (3.6). Kuna loeme intervalli dx positiivseks ja kehtib $dP \geq 0$, siis järeldub meie väide. Jaotustiheduse põhimõtteline kuju on toodud joonisel 3.3.



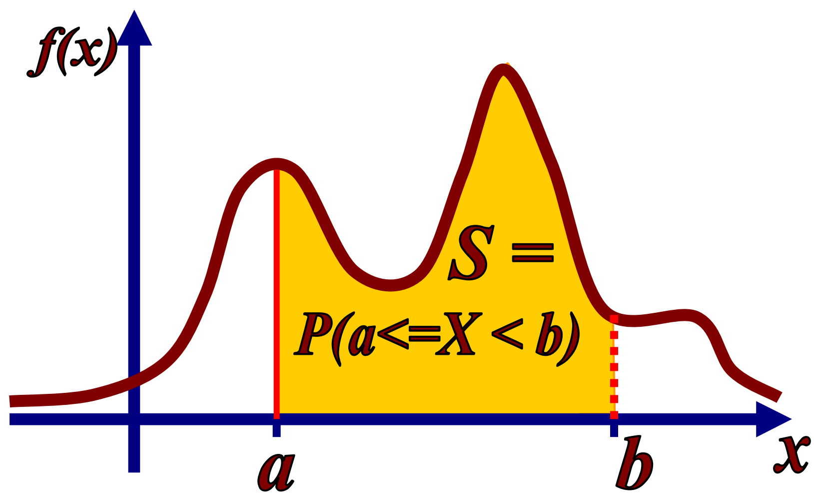
Joonis 3.3.

2. Elementaartõenäosuse geomeetriline tõlgendus. Elementaartõenäosus dP on ristküliku pindala, mille aluse laius on dx ja kõrgus on $f(x)$ (joonis 3.3). Sel põhjusel nimetatakse elementaartõenäosust dP ka ***tõenäosuselemendiks.***

3. Tõenäosus lõplikul lõigul. Summeerides valemi (3.6) üle elementaarintervallide dx lõplikul poolintervallil $[a, b)$, s.o summeerides intervalli elementaartõenäosused dP, saame tõenäosuse, et juhuslik suurus X satub sellesse vahemikku:

$$P(a \le X < b) = \int_a^b dP = \int_a^b f(x)dx. \tag{3.7}$$

Vastavalt määratud integraali definitsioonile on paremal oleva integraali väärtuseks kõvera $f(x)$ alla jääva, vasemalt ja paremalt vertikaaljoontega $x = a$ ja $x = b$ ning alt reaalteljega piiratud kujundi pindala S (joonis 3.4.)

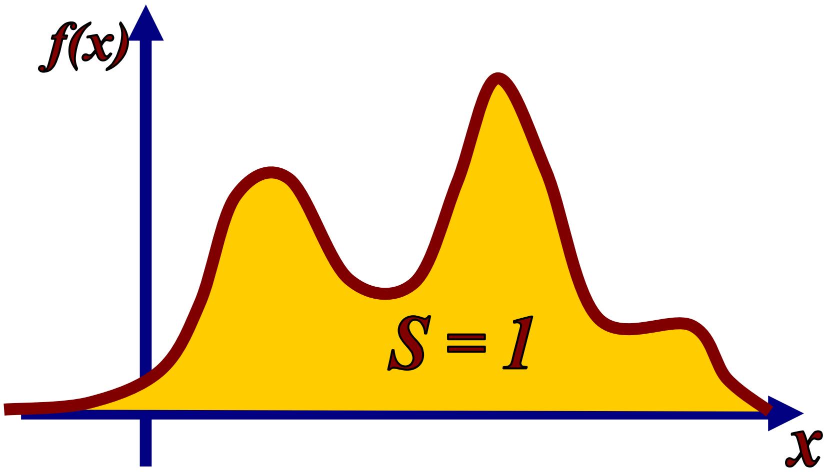


Joonis 3.4.

4. Normeerimistingimus:

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

See tingimus väljendab tõsiasja, et suvalise X väärtuse saamine peab olema tõene sündmus.



Joonis 3.5.

Geomeetriliselt tähendab see, et jaotustiheduse ja x-telje poolt piiratud kujundi kogupindala on võrdne ühega, nagu näidatud joonisel 3.5.

Märkus. Iga funktsioon, mis rahuldab mittenegatiivsuse tingimust ja normeerimistingimust, on tõenäosustihedus (s.t võib olla mingi tõenäosustihedus).

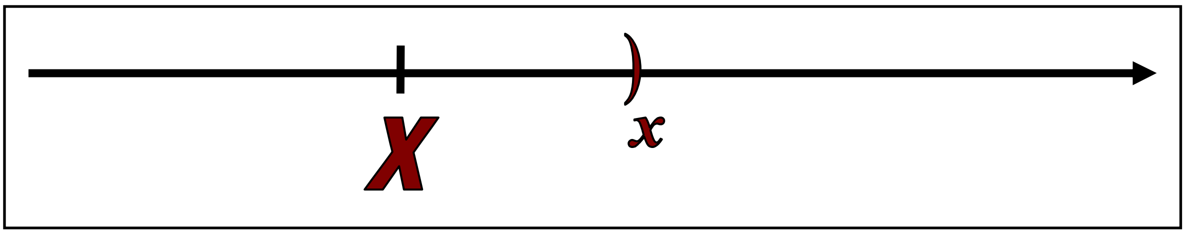
3.2.2 Jaotusfunktsioon

Integraalne jaotusfunktsioon $F(x)$ on kõrvuti jaotustihedusega teine enamlevinud vahend pideva juhusliku suuruse X kirjeldamiseks.

Juhusliku suuruse X (integraalseks) **jaotusfunktsiooniks** $F(x)$ nimetatakse tõenäosust, et kehtib võrratus $X < x$:

$$F(x) = P(X < x)\,. \tag{3.8}$$

Seega, funktsioon $F(x)$ on tõenäosus, et juhuslik punkt X asetseb arvsirgel punktist x vasakul, piirkonnas $(-\infty,\, x)$ (joonis 3.6.).



Joonis 3.6.

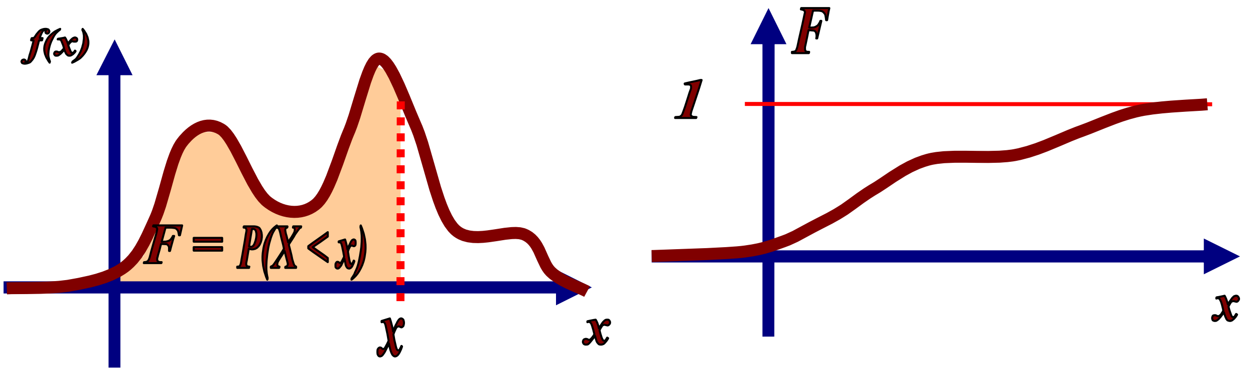
Ümarsulg $\boldsymbol{x}$ väärtuse kohal tähistab asjaolu, et tegemist on paremalt avatud intervalliga, mille otspunkt (käesoleval juhul $\boldsymbol{x}$) ei kuulu selle intervalli koosseisu.

Võttes avaldises (3.7) vasakpoolse raja võrdseks miinus lõpmatusega ja parempoolse raja b võrdseks x-ga, saame

$$P(-\infty < X < x) \quad \equiv \quad P(X < x) \quad = \quad \int_{-\infty}^{x} f(x')dx' .$$

Seega

$$F(x) \quad = \quad \int_{-\infty}^{x} f(y)dy . \qquad (3.9)$$



Joonis 3.7.

Nagu näeme, on jaotusfunktsioon leitav, kui on teada jaotustihedus $f(x)$. Vastupidi, teades jaotusfunktsiooni, on võimalik arvutada jaotustihedus:

$$f(x) \quad = \quad \frac{dF(x)}{dx} \qquad (3.10)$$

Kuigi matemaatiliselt on jaotustihedus ja jaotusfunktsioon üksteisest tuletatavad, on mõlemad siiski tarvilikud, sageli lihtsustab valemite 3.9 ja 3.10 kasutamine oluliselt paljude esmapilgul väga keeruliste ülesannete lahendamist.

ÜLESANDED

Ülesanne 3.11. Juhusliku suuruse X jaotusfunktsioon on

$$F(x) = x^2; \qquad 0 \le x \le 1$$

Arvutage a)vastav jaotustihedus; b) tõenäosus $P(0{,}25 \le X \le 0{,}75)$

Lahendus:

$$f(x) = \frac{dF(x)}{dx} = \frac{d(x^2)}{dx} = 2x \qquad 0 \le x \le 1$$

$$P(0{,}25 \le X \le 0{,}75) = F(0{,}75) - F(0{,}25) = 0{,}75^2 - 0{,}25^2 = 0{,}5625 - 0{,}0625 = 0{,}5.$$

Ülesanne 3.12. Defineeri mingi jaotustihedus $f(x)$, mis oleks vahemikus (–1; 2) mittekonstantne ning mujal võrdne nulliga. Leia selle funktsiooni jaotusfunktsioon ning vahemiku (0; 1) tõenäosus.

Lahendus:

See on ülesanne, millel pole ühest lahendust, s.t. õigeid lahendusi on mitu. Sobib iga funktsioon, millel pole negatiivseid väärtuseid ning mis oleks normeeritud ühele. Valime näiteks ruutfunktsiooni, sest sellel puuduvad negatiivsed väärtused:

$$f(x) = \begin{cases} cx^2; -1 \le x \le 2 \\ 0;\ mujal \end{cases}.$$

Nüüd leiame konstandi c, et ta oleks normeeritud ühele:

$$I = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-1} 0dx + \int_{-1}^{2} c \cdot x^2 dx + \int_{2}^{\infty} 0dx = \frac{cx^3}{3}\Big|_{-1}^{2} = 0 + \frac{c(2^3 - (-1)^3)}{3} + 0 = \frac{9c}{3} = 3c \equiv 1 \quad \Rightarrow \quad c = \frac{1}{3},$$

seega on sobivaks jaotustiheduseks $f(x) = \begin{cases} \frac{1}{3}x^2; -1 \le x \le 2 \\ 0;\ mujal \end{cases}.$

Kuna jaotustiheduse kuju sõltub piirkonnast, siis tuleb ka jaotusfunktsioon leida iga piirkonna jaoks eraldi:

$$-\infty < x < -1 \qquad F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0dy = 0$$

$$-1 < x < 2 \qquad F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{-1} 0dy + \int_{-1}^{x} \frac{1}{3}y^2 dy = 0 + \frac{y^3}{3 \cdot 3}\Big|_{-1}^{x} = \frac{x^3 - (-1)^3}{9} = \frac{x^3 + 1}{9}$$

$$2 < x < \infty \qquad F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{-1} 0dy + \int_{-1}^{2} \frac{1}{3}y^2 dy + \int_{2}^{x} 0dy = 0 + \frac{y^3}{3 \cdot 3}\Big|_{-1}^{2} + 0 = \frac{2^3 - (-1)^3}{9} = \frac{8+1}{9} = 1$$

Ehk kokkuvõtlikult on jaotusfunktsioon:

$$F(x) = \begin{cases} 0;\ x < -1 \\ \frac{x^3 + 1}{9};\ -1 < x < 2 \\ 1;\ x > 2 \end{cases}$$

vahemiku (0; 1) tõenäosus on:

$$P(0 \le X \le 1) = F(1) - F(0) = \frac{1^3 + 1}{9} - \frac{0^3 + 1}{9} = \frac{2}{9} - \frac{1}{9} = \frac{1}{9} \approx 11{,}1\%$$

Erineva jaotustiheduse kuju valikul võib ka tõenäosus tulla erinev!

Ülesanne 3.13. Radioaktiivse aine pooldumist kirjeldab valem $f(t)=\frac{1}{t_0}\exp(-t/t_0)$, kus parameeter t_0 on antud aine poolestusaeg (ajaintervall, mille jooksul jääb ainet $e \approx 2{,}718$ korda vähemaks. Tseesium 137 poolestusaeg on 43 aastat. Leia radioaktiivse aine pooldumist kirjeldav tihedusfunktsioon. Palju kulub aega, et esialgsest Cs ainehulgast jääks järgi ainult 1%?

Lahendus:

$$f(t)=\frac{1}{t_0}\exp(-t/t_0)$$

$$F(t)=\int_0^{t_1} f(t)dt=\int_0^{t_1}\frac{1}{t_0}\exp(-t/t_0)dt=\frac{1}{t_0}\exp(-t/t_0)\cdot\left(-t_0\right)\Big|_0^{t_1}=$$

$$=\exp(-t/t_0)\Big|_{t_1}^{0}=1-\exp(-t_1/t_0)$$

Tingimus, et esialgsest Cs ainehulgast jääks järgi ainult 1% tähendab, et esialgsest ainehulgast (100%) on ära poolestunud 100% – 1% = 99%, seega meid huvitab jaotusfunktsiooni väärtus kohal 99% ehk siis kohal 0.99:

$$1-\exp(-t_1/t_0)=0{,}99 \Rightarrow \exp(-t_1/t_0)=0{,}01 \Rightarrow -t_1/t_0=\ln(0.01)=-4{,}61$$

$$t_1=4{,}61\cdot t_0=4{,}61\cdot 43=198$$

Seega oleks esialgsest ainehulgast 198 aasta pärast alles 1%.

Ülesanne 3.14. Ruumi temperatuurikontrollisüsteem hoiab ruumi temperatuuri vahemikus 20 °C kuni 24 °C. Temperatuuri tsüklilise muutumise tulemuseks on arkussiinustemperatuurijaotus, keskväärtusega 22 °C. Mitu protsenti ajast on ruumi temperatuur vahemikus 21 °C kuni 23 °C? Mitu protsenti ajast on ruumi temperatuur alla 20,5 °C või üle 23,5 °C?

Arkussiinusjaotuse jaotustihedus on $f(x)=\dfrac{1}{\pi\sqrt{x(1-x)}},\quad 0<x<1$

ning tema jaotusfunktsioon on $F(x)=\dfrac{2\cdot\arcsin(\sqrt{x})}{\pi},\quad 0<x<1.$

Tõestame kõigepealt jaotustihedusele vastava jaotusfunktsiooni. Vaatame ainult piirkonda $0 < x < 1$, sest kuna arkussiinusjaotus on nullist erinev ainult piirkonnas (0; 1), on tema jaotusfunktsioon

$F(x) = 0$ piirkonnas $x \le 0$ ja $F(x) = 1$ piirkonnas $x \ge 1$.

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(x)dx =$$

$$= \int_0^x \frac{1}{\pi\sqrt{x(1-x)}}dx = \int_0^x \frac{1}{\pi\sqrt{(1-x)}\cdot\sqrt{x}}dx = \qquad y = \sqrt{x} \Rightarrow x = y^2 \Rightarrow dx = 2y \cdot dy$$

$$= \frac{1}{\pi}\int_0^x \frac{1}{\sqrt{1-\left(\sqrt{x}\right)^2}}\frac{dx}{\sqrt{x}} = \frac{2}{\pi}\int_0^{\sqrt{x}} \frac{1}{\sqrt{1-y^2}}dy = \qquad (0;x) \Rightarrow (\sqrt{0};\sqrt{x}) = (0;\sqrt{x})$$

$$= \frac{2}{\pi}\left[\arcsin(y)\right]_0^{\sqrt{x}} = \frac{2\cdot\arcsin(\sqrt{x})}{\pi} \qquad \int \frac{1}{1-y^2}dy = \arcsin(y)$$

Nüüd tuleb teha ülesande lahendamiseks muutujavahetus, et sobituda arkussiinusjaotuse valemiga:

Vaatame paare: $x_1 = 0 \rightarrow t_1 = 20$

$x_2 = 1 \rightarrow t_2 = 24$

Vastav üleminekuvalem on ilmselt $x = \dfrac{t-20}{4}$.

Teeme tabeli meid huvitavatest temperatuuridest, vastavatest x-i väärtustest ning tihedusfunktsiooni väärtustest:

t	***x***	***F*(*x*)**
20,5	1/8	0,23
21	¼	0,33
23	¾	0,67
23,5	7/8	0,77

Seega on vahemikus 21 °C kuni 23 °C ruumi temperatuur 0,67 – 0,33 = 34% ajast. Alla 20,5 °C või üle 23,5 °C on 0,23 + (1 – 0,77) = 46%. On huvitav märkida, et tsükliliste temperatuurimuutuste korral püsib temperatuur kõige kauem vahemiku ekstreempunktide läheduses ja kõige lühemat aega ettenähtud temperatuuril. See omadus nähtub jaotustiheduse graafikust ja on intuitiivselt mõistetav, vaadeldes siinuskõvera kulgu.

3.3 Juhusliku suuruse arvkarakteristikud

Eespool nägime, et juhusliku suuruse jaotusseadus iseloomustab täielikult juhuslikku suurust tõenäosuslikult vaatekohalt. Jaotusseadus võimaldab leida juhusliku suurusega seotud iga sündmuse tõenäosust. Jaotusseaduse põhikujudeks on teatavasti jaotustabel diskreetse juhusliku suuruse puhul ja jaotusfunktsioon (jaotustihedus) pideva juhusliku suuruse korral.

Pideva suuruse korral on jaotustiheduse eksperimentaalne leidmine sageli väga kulukas ja töömahukas ülesanne. Enamasti aga ei olegi jaotusseadust tarvis teada (ei ole tarvis nii täielikku infot). Piisab, kui kasutada nn juhusliku suuruse arvkarakteristikuid, mis iseloomustavad juhusliku suuruse integraalseid omadusi. Ilma liialdamata võib öelda, et tõenäosusteooria rakendamisel praktiliste ülesannete lahendamiseks on oluline osata kasutada juhusliku suuruse arvkarakteristikuid, jättes kõrvale jaotusseadused.

Olulisemateks arvkarakteristikuteks on keskväärtus ja dispersioon. Peale nende põhikarakteristikute kasutatakse veel suurt hulka teisi arvkarakteristikuid nagu kvantiilid, mediaan, mood, momendid, asümmeetriakordaja, ekstsessikordaja, karakteristlik funktsioon, entroopia jmt.

3.3.1 Keskväärtus

Keskväärtus on juhusliku suuruse tähtsaim arvkarakteristik, mis iseloomustab juhusliku suuruse paiknemist.

Juhusliku suuruse *X* keskväärtust tähistatakse matemaatikas ja füüsikas üsna mitmel erineval viisil. Levinumad tähistused on

m, m_x, $m[x]$, $\bar{x}$, $\langle x \rangle$ – füüsikute hulgas;

EX, $M[X]$ – matemaatikute hulgas.

Siin tähis *m, M* on võetud ingliskeelse sõna *mean* (keskmine) esitähest, *E* aga on tulenenud prantsuskeelse sõna *espérance* (lootus, ootus) esitähest.

Diskreetse juhusliku suuruse $X = \{x_1, x_2, ..., x_n\}$ keskväärtuseks nimetatakse suurust (arvu)

$$m = \sum_{k=1}^{n} x_k p_k \,, \qquad (3.11)$$

kus $p_k = P(X = x_k)$.

Kui võimalike väärtuste hulk on loenduv, siis

$$m = \sum_{k=1}^{\infty} x_k p_k \,, \qquad (3.12)$$

kusjuures eeldatakse, et summa paremal koondub (on lõplik). Kui summa ei koondu, siis harilikult keskväärtust juhuslikule suurusele ei omistata.

Pideva juhusliku suuruse *X*, mille jaotustihedus on $f(x)$ ja võimalike väärtuste hulk on kogu reaaltelg, keskväärtuseks nimetatakse arvu

$$m(x) = \int_{-\infty}^{+\infty} x f(x) dx \,, \qquad (3.13)$$

eeldusel, et integraal eksisteerib (koondub absoluutselt).

Märkus. Valem keskväärtuse arvutamiseks langeb kokku valemiga varda massikeskme arvutamiseks, kui varda mass on võrdne ühega. Teisisõnu, keskväärtus on ühikmassiga varda staatiline moment.

Valemi 3.13 edasiarendus x^n keskväärtuse jaoks on:

$$m(x^n) = \int_{-\infty}^{+\infty} x^n f(x) dx \,, \qquad (3.14)$$

Näide 3.5. Ühtlase jaotuse keskväärtus.

Ühtlase jaotuse jaotustihedus on defineeritud kui

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{mujal} \end{cases} \tag{3.15}$$

Arvutame nüüd ühtlase jaotuse keskväärtuse kasutades valemit (3.13):

$$m = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^{a} x \cdot f(x)dx + \int_{a}^{b} x \cdot f(x)dx + \int_{b}^{\infty} x \cdot f(x)dx =$$

$$= \int_{-\infty}^{a} x \cdot 0dx + \int_{a}^{b} x \cdot \frac{1}{b-a}dx + \int_{b}^{\infty} x \cdot 0dx = 0 + \int_{a}^{b} x \cdot \frac{1}{b-a}dx + 0 =$$

$$= \frac{1}{b-a}\int_{a}^{b} x\,dx = \frac{1}{b-a}\left(\frac{x^2}{2}\right)_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

See on lõigu [*a; b*] keskpunkt.

3.3.2 Keskmiste kasutamisest

Defineerime siinkohal veel mõned keskmisi väärtuseid kirjeldavad juhuslike suuruste arvkarakteristikud:

- Mediaan on arv, millest suuremaid ja väiksemaid väärtuseid on variatsioonireas ühepalju, ehk siis jaotusfunktsiooni kaudu väljendades $F(x_{mediaan}) \equiv 0.5$.
- Mood on tunnuse kõige sagedamini esinev väärtus.
- Kaalutud keskmine on arv, mis saadakse, kui aritmeetilise keskmise arvutamisel antakse erinevatele väärtustele erinevad kaalud. Kaaluks võivad olla näiteks mõõtetulemuste standardhälbed.

Näide 3.6. Suvel märgistatakse raadiomajakaga 9 kuldnokka, et teada saada kuldnokkade keskmine rände kestvus. Sügisel lahkusid kõik 9 kuldnokka Eestist, kuid vaatlusperioodi lõpuks suve alguses oli Eestisse tagasi jõudnud ainult 7 kuldnokka, kelle rände kestvused olid vastavalt 146; 152; 152; 154; 156; 159; 159 päeva. Mitu päeva kestis sellel aastal keskmiselt kuldnokkade ränne?

> Vastus: nihketa hinnang oleks mediaan, ehk 156 päeva. Aritmeetilist keskmist saab arvutada ainult suve alguseks Eestisse tagasi jõudnud lindude rännet arvesse võttes, see oleks 154 päeva, kuid see ei arvesta võimalusega, et kadunud 2 lindu siiski kunagi, näiteks suvel, siiski jõudsid tagasi. Ka annaks aritmeetiline keskmine selgelt nihutatud hinnangu juhul, kui mõni lindudest saabuks juba sügisel miskipärast Eestisse tagasi.

Kui meid huvitab kõige tüüpilisem väärtus, siis seda näitab kõige suurema sagedusega väärtus mood. Mood on sageduse kõrgpunkt, ta ei näita, kas ja kui palju on temast suuremaid ja vähemaid väärtuseid. Nominaaltunnuste korral (näiteks rahvus, elukutse) leitakse keskmisena mood.

Mediaani leidmisel ei arvestata tunnuse väärtusi vaid ainult suurusjärjestust. Mediaani kasutatakse siis, kui on eesmärgiks leida täpne andmete jaotuse keskpunkt, või kui andmete hulgas on ekstremaalseid väärtuseid, mis oluliselt mõjutavad keskväärtust.

Keskväärtus sõltub kõigist tunnuse väärtustest, kuid ta ei pruugi ise olla tunnuse väärtus. Keskväärtus võib sattuda vahemikku, kus tunnusel on vähe väärtuseid või need puuduvad hoopis. Siiski kasutatakse keskväärtust küllalt sageli, sest ta on aluseks teiste statistiliste näitajate (näiteks standardhälve, korrelatsioonikordaja) määramisele.

Näide 3.7. Oletame, et üks firma koosneb juhatajast ja 9-st töötajast. Töötajate kuupalk on 500 EUR, juhata kuupalk on 5500 EUR. Keskmine palk selles firmas on

$$m = \frac{9 \cdot 500 + 1 \cdot 5500}{10} = 1000 .$$

Samas mediaanpalk selles firmas on keskmisest palgast poole väiksem, 500 EUR.

Üks teine firma koosneb neljast insenerist, kuupalgaga 1200 EUR ning koristajast, kuupalgaga 200 EUR. Keskmine palk selles firmas on

$$m = \frac{4 \cdot 1200 + 1 \cdot 200}{5} = 1000 .$$

Samas mediaanpalk selles firmas on keskmisest palgast suurem, 1200 EUR.

Seega, sõltuvalt palkade jaotusest, võib olla keskmine palk mediaanpalgast nii suurem kui ka väiksem. Eesti keskmisena on siiski keskmine palk mediaanpalgast märgatavalt kõrgem, seda tingib ühelt poolt miinimumpalk, millest väiksemat palka ei ole võimalik maksta, samuti tõsiasi, et meil on teatud hulk inimesi, kes saavad ülisuuri palkasid. Näiteks oli statistikaameti andmeil Eestis 2010. aastal täistööajaga töötaja keskmine brutokuutöötasu 819 eurot ja mediaan 688 eurot.

3.3.3 Dispersioon, ruuthälve ja standardhälve

Teades juhusliku suuruse keskväärtust, ei saa veel otsustada, milliseid väärtusi juhuslik suurus võib omandada ja kuidas nad on hajutatud keskväärtuse ümber. Kui jaotustihedus on hästi kitsas, kui juhusliku suuruse hajuvus on väike, siis on keskväärtusest küll, et iseloomustada juhuslikku suurust. Kui aga jaotustihedus on lai ja lame? Meil on tarvis suurust, mis võimaldaks kvantitatiivselt hinnata jaotuse „kitsust“ või „hajuvust“. Selliseks suuruseks on dispersioon.

Dispersiooni defineerimiseks toome esmalt sisse juhusliku suuruse hälbe ehk tsentreeritud juhusliku suuruse:

$$\varepsilon_i = x_i - m[X]. \tag{3.16}$$

Pole raske veenduda, et hälbe keskväärtus on võrdne nulliga. Tõepoolest,

$$\sum_{i=1}^{n} \varepsilon_i = 0 .$$

See on seletatav sellega, et võimalikud hälbed on erinevate märkidega ja summaarselt kompenseeruvad. Järelikult ei iseloomusta hälve hajuvust. Tsentreerimine tähendab geomeetriliselt seda, et koordinaatide alguspunkt viiakse punkti $m[X]$.

Juhusliku suuruse **dispersiooniks** (ingliskeelne nimetus variance) nimetatakse suurust

, (3.17)

mis võetakse üheks hajuvuse karakteristikuks. Seega on dispersioon juhusliku suuruse üheks nn juhuslikkuse määra iseloomustajaks.

Teised levinumad tähised dispersioonile on veel σ^2 (dispersioon on siin tähistatud kui standardhälbe ruut). Käesolevas kursuses kasutame dispersioonile tähist D.

ja kui tahame rõhutada või esile tuua, millise juhuliku suurusega on tegu, siis tähist $D(X)$.

Vastavalt definitsioonile (3.17) on diskreetse juhusliku suuruse dispersiooni arvutusvalem

$$D = \sum_i (x_i - m)^2 p_i . \tag{3.18}$$

Niisiis, tegu on tõenäosustega kaalutud üksikrealisatsioonide hälvete ruutude summaga. Analoogiliselt, pideva juhuliku suuruse korral annab (3.13)

$$D = \int_{-\infty}^{\infty} (x - m)^2 \, f(x)\, dx . \tag{3.19}$$

Dispersiooni praktiliseks arvutamiseks sobib kasutada **Steineri valemit**:

$$D[X] = m[X^2] - (m[X])^2 . \tag{3.20}$$

Tõestus: lähtume valemist (3.19):

$$\begin{aligned}
D[X] &= \int_{-\infty}^{+\infty} (X - m[X])^2 f(X) dX = \int_{-\infty}^{+\infty} \left(X^2 - 2 \cdot X \cdot m[X] + (m[X])^2\right) f(X) dX = \\
&= \int_{-\infty}^{+\infty} (X^2) f(X) dX + \int_{-\infty}^{+\infty} (-2 \cdot X \cdot m[X]) f(X) dX + \int_{-\infty}^{+\infty} (m[X])^2 f(X) dX = \\
&= m[X^2] - 2 \cdot m[X] \cdot \int_{-\infty}^{+\infty} X \cdot f(X) dX + (m[X])^2 \cdot \int_{-\infty}^{+\infty} f(X) dX = \\
&= m[X^2] - 2 \cdot m[X] \cdot m[X] + (m[X])^2 \cdot 1 = m[X^2] - (m[X])^2
\end{aligned} ,$$

Dispersiooni ruutjuurt nimetatakse standardhälbeks, ruutkeskmiseks hälbeks ehk ruuthälbeks. Definitsioonidest on näha, et dispersiooni dimensioon on võrdne juhusliku suuruse dimensiooni (mõõtühiku) ruuduga, standardhälbe dimensiooniks on aga juhusliku suuruse dimensioon. Seetõttu kasutatakse praktikas harilikult standardhälvet.

Mõõtemääramatuste arvutamisel kasutatakse alati standardhälvet!!!

$$\sigma = \sqrt{D} \tag{3.21}$$

Näide 3.8. Täringuvisete keskväärtus ning standardhälve.

Lahendamiseks teeme kõigepealt tabeli:

	x	x^2	$x-m$	$(x-m)^2$	p_i
	1	1	–2,5	6,25	1/6
	2	4	–1,5	2,25	1/6
	3	9	–0,5	0,25	1/6
	4	16	0,5	0,25	1/6
	5	25	1,5	2,25	1/6
	6	36	2,5	6,25	1/6
Σ	21	91	0	17,5	1

Leiame kõigepealt keskväärtuse, lähtudes valemist (2.11):

$$m = \sum_{k=1}^{n} x_k p_k = \frac{1}{6} \cdot 21 = 3{,}5\,.$$

Lähtudes valemist (2.19), saame:

$$D = \sum_i (x_i - m)^2 p_i = 17{,}5 \cdot \frac{1}{6} \approx 2{,}92\,.$$

Lähtudes Steineri valemist (2.21), saame:

$$D = m[X^2] - m^2 = \frac{91}{6} - 3{,}5^2 \approx 15{,}17 - 12{,}25 = 2{,}92\,.$$

Täringuviske standardhälve on

$$\sigma(x) = \sqrt{D(x)} = \sqrt{2{,}92} = 1{,}71\,.$$

Näide 3.9. Ühtlase jaotuse standardhälve.

Ühtlase jaotuse jaotustihedus on:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{mujal} \end{cases}$$

Ühtlase jaotuse keskväärtuse leidsime juba varem näites 3.5:

$$m(x)=\int_a^b x\frac{1}{b-a}dx=\frac{1}{b-a}\int_a^b x\,dx=\frac{1}{b-a}\left(\frac{x^2}{2}\right)_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{(b-a)(b+a)}{2(b-a)}=\frac{a+b}{2}$$

Dispersiooni arvutamisel kasutame Steineri valemit (3.20). Leiame esmalt teist järku algmomendi (ehk x^2 keskväärtuse):

$$m(x^2)=\int_a^b x^2\frac{1}{b-a}dx=\frac{1}{b-a}\int_a^b x^2\,dx=\frac{1}{b-a}\left(\frac{x^3}{3}\right)_a^b=$$

$$=\frac{b^3-a^3}{3(b-a)}=\frac{(b-a)(a^2+ab+b^2)}{3(b-a)}=\frac{a^2+ab+b^2}{3}.$$

Nüüd saame

$$D(x)=m(x^2)-(m(x))^2=\frac{a^2+ab+b^2}{3}-\left(\frac{a+b}{2}\right)^2=$$

$$=\frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}=\frac{a^2-2ab+b^2}{12}=\frac{(b-a)^2}{12}.$$

Standardhälve tuleb siis

$$\sigma(x)=\sqrt{D(x)}=\frac{1}{\sqrt{3}}\frac{b-a}{2}\approx 0{,}577\,\frac{b-a}{2}.$$

ÜLESANDED

Ülesanne 3.15. Jaotusfunktsioon on defineeritud järgmiselt:

$$f(x)=\begin{cases}5x, & 1\le x\le 2\\ 0, & \text{mujal}\end{cases}$$

Tee funktsiooni graafik, hinda silma järgi, kui suur võiks olla keskväärtus. Arvuta antud funktsiooni täpne keskväärtus, kanna see joonisele.

Lahendus:

$$m(x)=\int_{-\infty}^{\infty} f(x)\cdot x\cdot dx=\int_1^2 5x^2\cdot dx=\cdot\frac{5x^3}{3}\Big|_0^b=\frac{5}{3}\cdot(8-1)=\frac{35}{3}=11{,}67$$

Keskväärtus on väljaspool funktsiooni määramispiirkonda, kuid sisuliselt ei saa see nii olla. Kus on viga?

Kontrollime jaotusfunktsiooni vastavust normeerimistingimusele:

$$\int_{-\infty}^{\infty} f(x)\cdot dx = \int_{1}^{2} 5x \cdot dx = \cdot\frac{5x^2}{2}\Big|_1^2 = \frac{5}{2}\cdot(4-1) = \frac{15}{2} = 7{,}5 \neq 1$$

Seega ei olegi tegemist jaotusfunktsiooniga!!!

<u>Ülesanne 3.16.</u> Kolmnurkjaotust võib defineerida ka järgnevalt: piirkondades $x < 0$ ning $x > b$ on tema väärtus 0; vahemikus $[0; b]$ on tema väärtus ühtlaselt kasvav ($f(x) = cx$). Arvuta koefitsient c, jaotusfunktsioon, keskväärtus ja mediaan ning standardhälve. Leia tõenäosus, et sündmus satuks piirkonda $(m \pm \sigma)$; $(m \pm 2\sigma)$. Joonista graafik ning kanna sellele keskväärtus ja mediaan ning viiruta piirkond $(m \pm 2\sigma)$.

Lahendus:

Kuna peab kehtima normeerimistingimus, siis

$$\int_{-\infty}^{\infty} cx\cdot dx = \int_{0}^{b} cx\cdot dx = c\cdot\frac{x^2}{2}\Big|_0^b = c\cdot\frac{(b^2)}{2} \equiv 1 \;\Rightarrow\; c = \frac{2}{b^2}$$

Seega on selle kolmnurkjaotuse tihedusfunktsioon:

$$f(x) = \begin{cases} \dfrac{2x}{b^2}, & 0 \le x \le b \\ 0, & \text{mujal} \end{cases}$$

Leiame nüüd jaotusfunktsiooni:

$$F(x) = \int_{-\infty}^{x} \frac{2x}{b^2}dx = \int_{0}^{x} \frac{2x}{b^2}dx = \frac{x^2}{b^2}\Big|_0^x = \frac{x^2}{b^2}$$

Arvutame nüüd keskväärtuse:

$$m(x) = \int_0^b x\frac{2x}{b^2}dx = \frac{2}{b^2}\int_0^b x^2\,dx = \frac{2}{b^2}\left(\frac{x^3}{3}\right)_0^b = \frac{2b^3}{3b^2} = \frac{2}{3}b \approx 0{,}67b$$

Arvutame mediaani kasutades jaotusfunktsiooni:

$$F(x) = \frac{x^2}{b^2} = 0{,}5 \;\Rightarrow\; x = \sqrt{0{,}5}\cdot b \approx 0{,}71b$$

Arvutame standardhälbe arvutamiseks vajaliku x_2 keskväärtuse:

$$m(x^2) = \int_0^b x^2\frac{2x}{b^2}dx = \frac{2}{b^2}\int_0^b x^3\,dx = \frac{2}{b^2}\left(\frac{x^4}{4}\right)_0^b = \frac{2b^4}{4b^2} = \frac{b^2}{2}.$$

Arvutame standardhälbe, kasutades Steineri valemit:

$$\sigma(x)=\sqrt{m(x^2)-m^2(x)}=\sqrt{\frac{b^2}{2}-\left(\frac{2b}{3}\right)^2}=b\sqrt{\frac{1}{2}-\frac{4}{9}}=\frac{b}{\sqrt{18}}$$

Leiame jaotusfunktsiooni väärtused meid huvitavates punktides:

$$m-2\sigma=\frac{2}{3}b-\frac{2b}{\sqrt{18}}\approx b(0{,}67-0{,}48)=0{,}19b$$
$$m-\sigma=\frac{2}{3}b-\frac{b}{\sqrt{18}}\approx b(0{,}67-0{,}24)=0{,}43b$$
$$m+\sigma=\frac{2}{3}b+\frac{b}{\sqrt{18}}\approx b(0{,}67+0{,}24)=0{,}91b$$
$$m+2\sigma=\frac{2}{3}b+\frac{2b}{\sqrt{18}}\approx b(0{,}67+0{,}48)=1{,}15b$$

$$F(m-2\sigma)\approx F(0{,}19b)=0{,}19^2=0{,}036$$
$$F(m-\sigma)\approx F(0{,}43b)=0{,}43^2=0{,}185$$
$$F(m+\sigma)\approx F(0{,}91b)=0{,}91^2=0{,}828$$
$$F(m+2\sigma)=F(1{,}15b)=1$$

$P(m \pm \sigma) = 0{,}828 – 0{,}185 = 0{,}643 = 64{,}3\%$

$P(m \pm 2\sigma) = 1 – 0{,}036 = 0{,}964 = 96{,}4\%$

Ülesanne 3.17. Kolmnurkjaotuse jaotustihedus on defineeritud järgnevalt: piirkondades $x < 5$ ning $x > 8$ on tema väärtus 0; vahemikus [5; 8] on tema väärtus ühtlaselt kasvav ($f(x) = cx$). Leia koefitsient c; jaotusfunktsioon; keskväärtus, mediaan; standardhälve; tõenäosus, et sündmus satuks piirkonda ($m(x)$, 12)?

Vastused:

$$c=\frac{2}{39};\ F(x)=\begin{cases}0; & x<5\\ \frac{1x^2}{39}-\frac{25}{39}; & 5<x<8\\ 1; & x>8\end{cases};\ m(x)=6{,}62;\ \text{med}=\sqrt{\frac{89}{2}};\ s(x)=0{,}822;\ P=0{,}517$$

3.4 Valik pideva juhusliku suuruse jaotusseaduseid

Erinevate protsesside kirjeldamiseks on maailmas loodud väga palju erinevaid jaotusseaduseid. Antud kursuse raames vaatleme neist ainult mõnda olulisemat.

3.4.1 Ühtlane jaotus

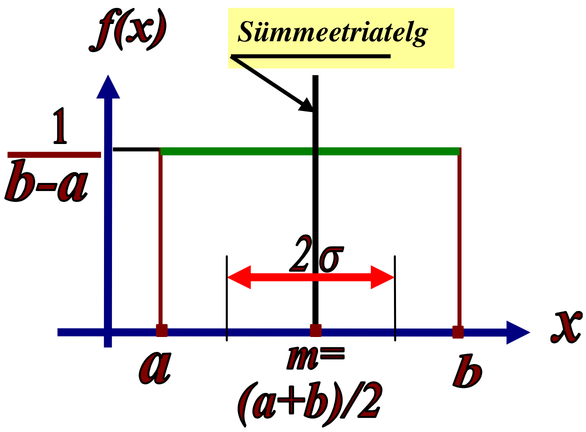
Pideva juhusliku suuruse X ühtlaseks jaotuseks (ka ristkülikjaotuseks) lõigul [a; b] nimetatakse jaotust, mille jaotustihedus sellel lõigul on nullist erinev konstant. Kuna peab kehtima normeerimistingimus, siis

$$\int_{-\infty}^{\infty} c \cdot dx = \int_{a}^{b} c \cdot dx = c \cdot x\Big|_a^b = c \cdot (b-a) \equiv 1 \;\Rightarrow\; c = \frac{1}{b-a}$$

Seega on ühtlase jaotuse jaotustihedus:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{mujal} \end{cases} \tag{3.22}$$

Joonisel 3.8 on toodud ühtlase jaotuse jaotustiheduse graafik.



Joonis 3.8. Ühtlase jaotuse jaotustihedus.

Ühtlase jaotuse keskväärtus ja standardhälve on:

$$m(x) = \frac{a+b}{2}$$

$$\sigma = \frac{1}{\sqrt{3}}\frac{b-a}{2} \approx 0.577\,\frac{b-a}{2}.$$

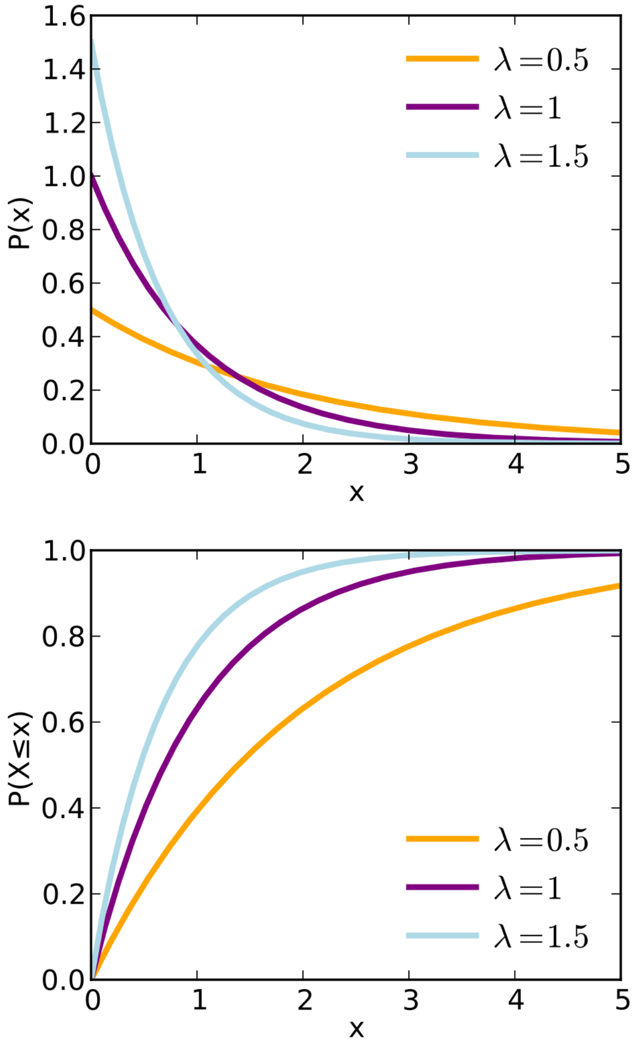
Tuletuskäigud toodud vastavalt näidetes 3.5 ja 3.9.

3.4.2 Eksponentjaotus

Eksponentjaotuse jaotustihedus $f(t)$ ja jaotusfunktsioon $F(t)$ on:

$$f(t) = \lambda \exp(-\lambda t)$$
$$F(t) = 1 - \exp(-\lambda t)'$$

Siin $\lambda = 1 / t_0$, eksponentjaotuse keskväärtus ja standardhälve on mõlemad võrdsed parameetriga t_0.



Joonis 3.9. Eksponentjaotuse jaotustihedus ja jaotusfunktsioon.

Näide 3.10. Radioaktiivse aine pooldumist kirjeldab valem $f(t) = \frac{1}{t_0}\exp(-t/t_0)$, kus parameeter t_0 on antud aine poolestusaeg (ajaintervall, mille jooksul jääb ainet $e \approx 2{,}718$ korda vähemaks. Tseesium 137 poolestusaeg on 43 aastat. Leia Cs pooldumise keskväärtus, mediaan ja standardhälve.

Kasutame arvutuses poolestusaja pöördväärtust $\lambda = 1/t_0$:

$$f(t) = \lambda \exp(-\lambda t) = \frac{1}{43}\exp\left(-\frac{t}{43}\right)$$

$$F(t) = 1 - \exp\left(-\frac{t}{43}\right)$$

Seega on eksponentjaotuse keskväärtus ja standardhälve vastavalt

$$m(t) = \sigma(t) = t_0 = 43 \; aastat \, .$$

Leiame nüüd mediaani:

$$F(t) = 1 - \exp(-\lambda t) = 0{,}5 \;\Rightarrow\; \exp(-\lambda t) = 0{,}5$$

$$-\lambda t = \ln(0{,}5) = -0{,}69 \;\Rightarrow\; t = \frac{0{,}69}{\lambda} = 0{,}69 \cdot t_0 = 0{,}69 \cdot 43 \approx 30$$

Seega on tseesiumi pooldumise keskväärtus 43 aastat, aga mediaan 30 aastat. Teisiti võib mediaanväärtust sõnastada nii, et 30 aastaga on pool esialgsest tseesiumi kogusest ära pooldunud.

3.4.3 Normaaljaotus

Normaaljaotus e. ***Gaussi jaotus*** on üks tähtsamaid jaotusseadusi juhuslikele suurustele, mis on jaotunud kogu reaalteljele ja võivad omandada väärtusi vahemikus $(-\infty, \infty)$.

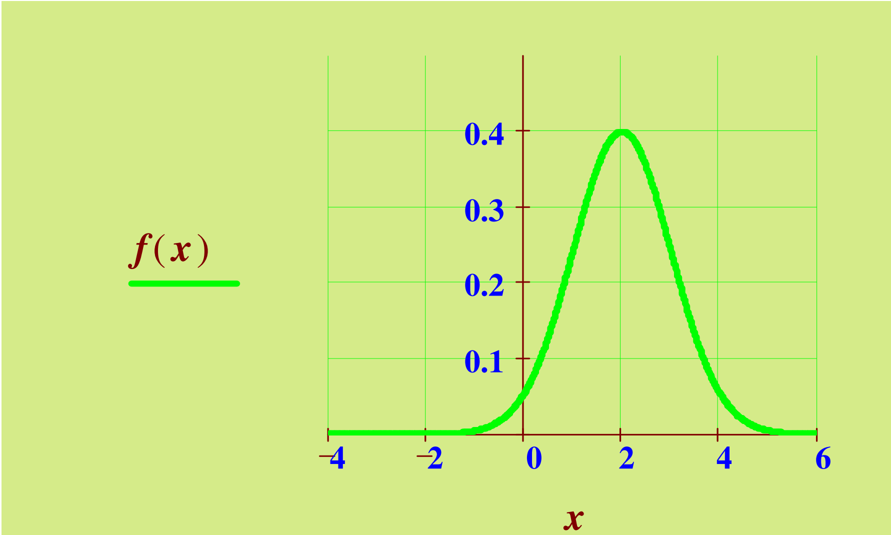
Juhusliku suuruse jaotust nimetatakse **normaaljaotuseks** ehk **Gaussi jaotuseks**, kui jaotustihedus on

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right), \tag{3.23}$$

Kus a on fikseeritud reaalarv (võib olla nii negatiivne kui positiivne, aga võib olla ka null), ja $\sigma > 0$ on fikseeritud positiivne reaalarv. Antud loengukursuse raames me ei tõesta, vaid ainult konstateerime, et parameeter a on **normaaljaotuse keskväärtus** ja σ on **normaaljaotuse standardhälve**.

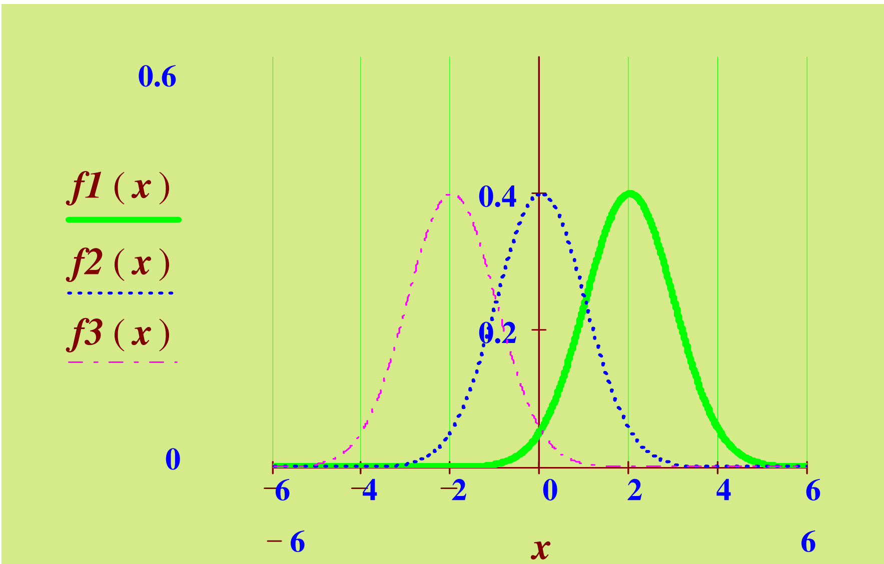
Jaotustiheduse määramispiirkond on kogu reaaltelg ***R*** (s.t argument x võib omandada väärtusi kogu reaalteljel).

Normaaljaotuse tiheduse graafik on parameetrite $a = 2$, $\sigma = 1$ korral toodud joonisel 3.9.



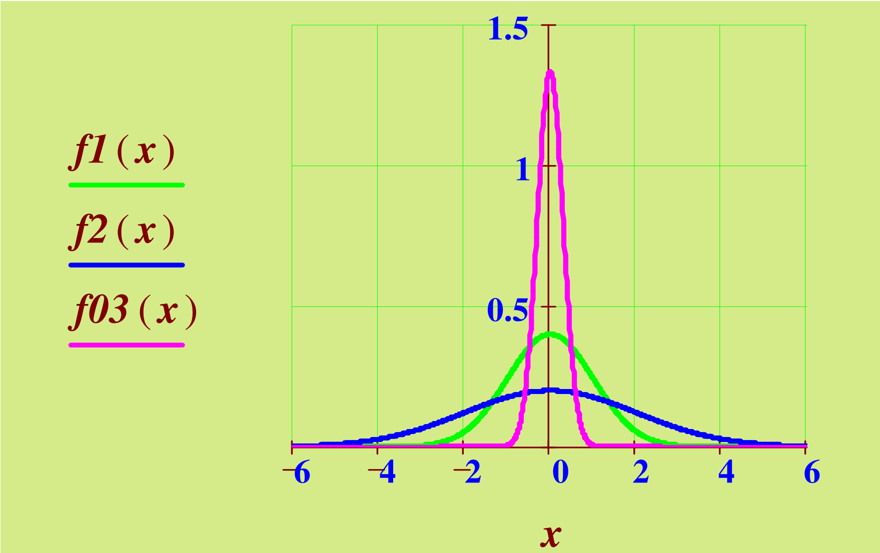
Joonis 3.9. Normaaljaotuse jaotustihedus.

Parameetri a muutumisel joone asend muutub x-telje suhtes: a kasvades jaotus nihkub paremale (toimub kujundi x-telje sihiline paralleellüke). Mida suurem on a, seda paremal paikneb kõver (joonis 3.10).



Joonis 3.10. Erinevad normaaljaotused σ = 1 korral: a = –2 (kõver *f3(x)*), a = 0 (kõver *f2(x)*) ja a = 2 (kõver *f1(x)*).

Kui parameeter σ kasvab, siis kahanevad funktsiooni väärtused ja joon muutub lamedamaks – kõver surutakse kokku y-telje suunas. Kui σ kahaneb, siis muutub joon teravatipulisemaks – kõver venitatakse välja y-telje suunas (joonis 3.11).



Joonis 3.11. Normaaljaotus (a = 0) erinevate σ-de korral: $f1(x)$ – σ = 1, $f2(x)$ – σ = 2 ja $f03(x)$ – σ = 0,3.

Paljud looduses toimuvad protsessid on kirjeldatavad normaaljaotuse abil: loomulik müra, transpordivoo kiirus, teatud vanuserühma meeste, naiste ja laste pikkus, vererõhk jne.

Nagu teistegi jaotusseaduste puhul, on ka normaaljaotuse puhul mingisse intervalli sattumise tõenäosuse leidmiseks vaja teada tema jaotusfunktsiooni. Praktikas ei ole normaaljaotuse (3.23) integreerimine just lihtne ülesanne. Selles loengukursuses me ei tõesta, et normaaljaotuse jaotusfunktsioon on:

$$F(x)=\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-a}{\sigma\sqrt{2}}\right)\right]$$
$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt, \qquad (3.24)$$

Siin erf(x) tähistab veafunktsiooni. Praktiliste ülesannete lahendamisel kasutatakse valmisprogrammide abi. Mathcad'is on normaaljaotuse parameetritega a ja σ jaotustihedus kohal x arvutatav funktsiooniga dnorm(x, a, σ) ning jaotusfunktsioon kohal x funktsiooniga pnorm(x, a, σ).

Excel'is on jaotustihedus arvutatav käsuga normdist(x, a, σ, 0) ning jaotusfunktsioon käsuga normdist(x, a, σ, 1).

Näide 3.11. Arvutame välja tõenäosuse et normaaljaotuse puhul satuks sündmus intervalli (a ± σ), (a ± 2σ), (a ± 3σ). Kuna need tõenäosused ei sõltu jaotuse parameetritest, siis valime võimalikult lihtsad parameetrid, näiteks a = 0 ja σ = 1. Leiame Excel'i abil meid huvitavad jaotusfunktsiooni väärtused:

Normdist(–3, 0, 1, 1) = 0,00135
Normdist(–2, 0, 1, 1) = 0,0228
Normdist(–1, 0, 1, 1) = 0,159

Normdist(1, 0, 1, 1) = 0,841
Normdist(2, 0, 1, 1) = 0,9772
Normdist(3, 0, 1, 1) = 0,99865

Tegelikult piisaks ka neist ainult esimese kolme arvutamisest, sest normaaljaotus on sümmeetriline, ning seega on tõenäosus, et toimub sündmus $x < a - z$ võrdne tõenäosusega, et toimub sündmus $x > a + z$.

$P(a \pm 3\sigma) = 0{,}99865 - 0{,}00135 = 0{,}9973 = 99{,}73$ %

$P(a \pm 3\sigma)$ = 1 – 2(0,00135) = 0,9973 = **99,73 %**
$P(a \pm 2\sigma)$ = 1 – 2(0,0228) = 0,9545 = **95,45 %**
$P(a \pm 1\sigma)$ = 1 – 2(0,159) = 0,6827 = **68,27 %**

Seega satub vahemikku ± 1σ ligikaudu 68 %, vahemikku ± 2σ ligikaudu 95 % ja vahemikku ± 3σ ligikaudu 99,7 % sündmustest.

Normaaljaotust, kus keskväärtus $a = 0$ ja standardhälve $\sigma = 1$ nimetatakse **standardiseeritud normaaljaotuseks N(0; 1)**.

Üleminekuvalem suvaliselt normaaljaotuselt $X \sim N(a;\sigma)$ standardiseerutud normaaljaotuseks on:

$$\frac{X-a}{\sigma} \sim N(0;1)\,.$$

Ülesanne 3.18. Olgu teada, et 20-aastaste noormeeste pikkused alluvad normaaljaotusele keskväärtusega a = 183 cm ja standardhälbega σ = 7 cm. Kui suur osa noormeestest on pikemad kui 2 m; pikemad kui 2,1 m; pikkuse vahemikus 180 cm kuni 190 cm?

Lahendus: (kasutades Exceli funktsiooni NORMDIST):

$$P(X > 200) = 1 - P(X < 200) = 1 - F(200) = 1 - NORMDIST(200;183;7;1) = 1 - 0{,}992421 = 0{,}007579$$
$$P(X > 210) = 1 - P(X < 210) = 1 - F(210) = 1 - NORMDIST(210;183;7;1) = 1 - 0{,}999943 = 0{,}000057$$
$$P(180 < X < 190) = P(X < 190) - P(X < 180) = F(190) - F(180) =$$
$$= 0{,}841345 - 0{,}334118 = 0{,}507227.$$

Ülesanne 3.19. Ühe gümnaasiumi sisseastumiskatseks on komplekstest. Testi tulemused alluvad normaaljaotusele keskväärtusega m = 500 punkti ja standardhälbega σ = 100 punkti. 108-le kohale oli 427 kandidaati. Toomas sai katsetel 598 punkti. Kas sellest peaks piisama gümnaasiumisse pääsemiseks?

Lahendus: (kasutades Exceli funktsiooni NORMDIST):

Gümnaasiumisse võetakse vastu 108 / 427 = 0,25293 = 25,29% kandidaatidest.

$$P(X > 573) = 1 - P(X < 573) = 1 - F(598) = 1 - NORMDIST(573;500;100;1) = 1 - 0{,}7673 = 0{,}2327 \approx 23\%.$$
Kuna Toomas oli parima 23% sisseastujate seas ja vastu võeti 25,29%, siis peaks Toomase tulemusest piisama gümnaasiumisse pääsemiseks.

3.4.4 Studenti jaotus

Inglise teadlane Student, alias W.S. Gosset tõestas, et:

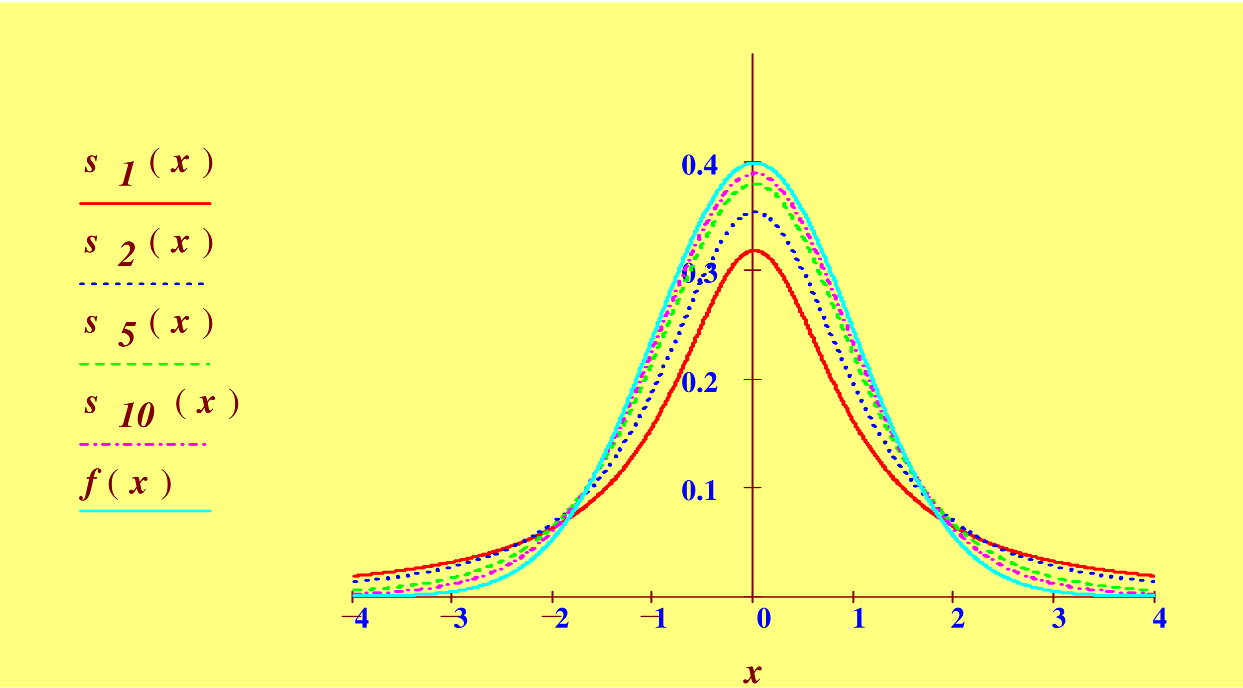
Kui X on normaalne juhuslik suurus, siis juhuslik suurus

$$T = \frac{\bar{x}_N - m[X]}{u(\bar{x}_N)} \tag{3.25}$$

allub jaotusele tihedusega

$$s_{N-1}(t) = C_{N-1}\left(1 + \frac{t^2}{N-1}\right)^{-N/2}. \tag{3.26}$$

Siin C_{N-1} on normeerimistegur. Jaotust (3.26) nimetatakse (N-1 vabadusastmega) ***Studenti jaotuseks***. Studenti jaotus erinevate $N-1$ väärtuste korral on toodud joonisel 3.12, millelt on ka näha, kuidas see funktsioon vabadusastmete arvu kasvades läheneb standardiseeritud normaaljaotusele.



Joonis 3.12. Ühe, kahe, viie ja kümne vabadusastmega Studenti jaotus $s_{N-1}(x)$ ning standardiseeritud normaaljaotus $f(x)$.

Studenti jaotuse oluline omadus on, et juhusliku suuruse T jaotusfunktsioon $s_{N-1}(t)$ sõltub vaid ühest parameetrist (nn vabadusastmete arvust) $N-1$ ja ei sõltu üldse juhusliku suuruse (mõõdetava suuruse) X konkreetsest dispersioonist ega keskväärtusest (küll on aga oluline X normaalsus). Selles mõttes on tegu universaalfunktsiooniga, mis iseloomustab suvaliste ühesuguse normaaljaotusega juhuslike suuruste summa üldisi omadusi.

Otsene järeldus asjaolust, et juhuslik suurus (3.25) allub Studenti jaotusele on, et usaldusnivoole p vastav mõõtmistulemuse paiknemise intervall on esitatav kujul

$$P\left[\bar{x}_N - t_{N-1}(p)u(\bar{x}_N) < x < \bar{x}_N + t_{N-1}(p)u(\bar{x}_N)\right] = p. \qquad (3.27)$$

Siin suurus $t_{N-1}(p)$ on (Studenti) ***t*-kordaja** – parandustegur etteantud usaldusnivoo p ja $N-1$ vabadusastme korral. t-kordaja on samuti ainult ühest parameetrist N sõltuv universaalne usaldusnivoo funktsioon, mis on leitav Studenti jaotusest. Tavaliselt antakse t-kordaja kindlate p väärtuste jaoks tabuleeritult. Studenti t-kordaja tabuleeritud väärtused on toodud **lisas 1.** Mathcadis on t-kordaja leidmiseks funktsioon $qt\left(\frac{1+p}{2}, \nu\right)$; Excelis funktsioon $TINV(1-p, \nu)$.

Valemit (3.27) nimetatakse ***Studenti testiks***.

Suurte N -ide korral läheb Studenti jaotus üle standardiseeritud normaaljaotuseks. Tabelis 1 toodud Studenti t-kordaja tabelist on näha, et näiteks $t_2(95\%) \to 4{,}30$ ja $t_9(99{,}73\%) \to 4{,}09$ jne. Erinevus on oluline alas $N < 30$ ja väga oluline alas $N < 10$.

Ülesanne 3.19. Olgu teada, et 20-aastaste noormeeste pikkused alluvad normaaljaotusele, jaotuse parameetrite määramiseks mõõdeti nelja 20-aastase noormehe pikkused. Keskmiseks noormeeste pikkuseks tuli 181 cm standardhälbega 8 cm. Lähtudes Studenti jaotusest, millisesse vahemikku peaks jääma 95% noormeeste pikkused? Kui suur oleks see vahemik siis, kui mõõdetud noormeeste arv oleks olnud 60?

Lahendus: (kasutades Exceli funktsiooni TINV:

$t_3(95\%) = TINV(0{,}05;3) = 3{,}182$

$95\% = m(x) \pm t_3(95\%) \cdot s(x) = 181 \pm 3{,}182 \cdot 8 = 181 \pm 25 \Rightarrow (156;206).$

$t_{59}(95\%) = TINV(0{,}05;59) = 2{,}000$

$95\% = m(x) \pm t_{59}(95\%) \cdot s(x) = 181 \pm 2{,}000 \cdot 8 = 181 \pm 16 \Rightarrow (165;197).$

Seega samade parameetrite juures annab suurem valim täpsema hinnangu.

Käesoleva konspekti koostamisel on kasutatud prof. Rein Rõõmu loengukonspekti „Tõenäosusteooria ja statistika, I vihik“ materjale.

1.5. Lisa 1. Studenti t-kordaja väärtused

Studenti t-kordaja väärtused $t_\nu(p)$ sõltuvalt vabadusastmete arvust $\nu = N - 1$ ja soovitavast usaldusnivoost p .

Vabadusastmete arv $\nu = N - 1$	Osa p protsentides					
	68,27[I]	90	95	95,45[I]	99	99,73[I]
1	1,84	6,31	12,71	13,97	63,66	235,80
2	1,32	2,92	4,30	4,53	9,92	19,21
3	1,20	2,35	3,18	3,31	5,84	9,22
4	1,14	2,13	2,78	2,87	4,60	6,62
5	1,11	2,02	2,57	2,65	4,03	5,51
6	1,09	1,94	2,45	2,52	3,71	4,90
7	1,08	1,89	2,36	2,43	3,50	4,53
8	1,07	1,86	2,31	2,37	3,36	4,28
9	1,06	1,83	2,26	2,32	3,25	4,09
10	1,05	1,81	2,23	2,28	3,17	3,96
11	1,05	1,80	2,20	2,25	3,11	3,85
12	1,04	1,78	2,18	2,23	3,05	3,76
13	1,04	1,77	2,16	2,21	3,01	3,69
14	1,04	1,76	2,14	2,20	2,98	3,64
15	1,03	1,75	2,13	2,18	2,95	3,59
16	1,03	1,75	2,12	2,17	2,92	3,54
17	1,03	1,74	2,11	2,16	2,90	3,51
18	1,03	1,73	2,10	2,15	2,88	3,48
19	1,03	1,73	2,09	2,14	2,86	3,45
20	1,03	1,72	2,09	2,13	2,85	3,42
25	1,02	1,71	2,06	2,11	2,79	3,33
30	1,02	1,70	2,04	2,09	2,75	3,27
35	1,01	1,70	2,03	2,07	2,72	3,23
40	1,01	1,68	2,02	2,06	2,70	3,20
45	1,01	1,68	2,01	2,06	2,69	3,18
50	1,01	1,68	2,01	2,05	2,68	3,16
100	1,005	1,660	1,984	2,025	2,626	3,077
∞	1,000	1,645	1,960	2,000	2,576	3,000

(I) Suuruse x jaoks, mida kirjeldab normaaljaotus keskväärtusega $\bar{x}$ ja standardhälbega σ_x, sisaldab vahemik $x \pm k\sigma_x$ $p = 68{,}27$; 95,45 ja 99,73 protsenti jaotusest vastavalt $k = 1$; 2 ja 3 korral.

Euroopa Liit
Euroopa Sotsiaalfond

Eesti tuleviku heaks

BeSt
e-õppe programm

Tõenäosusteooria ja statistika

II osa – Statistika

LOFY.03.056

Erko Jakobson, PhD

erko.jakobson@ut.ee

2013 kevad, Tartu

Sisukord

1 Statistilise andmeanalüüsi põhimõisted ... 3

1.1 Üldkogum ja valim ... 3

1.2 Statistilise tunnuse karakteristikud ... 4

2 Statistilised hüpoteesid ja nende tõestamine ... 6

2.1 Statistiliste hinnangute usaldusväärsus ... 6

2.2 Statistiline hüpotees ... 6

2.2.1 Vead hüpoteeside kontrollimisel ... 7

2.2.2 Statistiliste hüpoteeside tõestamine ... 7

3 Testid statistiliste hüpoteeside tõestamiseks ... 8

3.1 F-test normaaljaotusega üldkogumite dispersioonide võrdlemiseks ... 9

3.2 Studenti t-test normaaljaotusele alluva juhusliku suuruse keskväärtuse kohta ... 11

3.2.1 Studenti t-test keskväärtuse võrdlemiseks konstandiga ... 11

3.2.2 Studenti t-test keskväärtuste võrdlemiseks võrdse dispersiooniga üldkogumite korral ... 13

3.2.3 Studenti t-test keskväärtuste võrdlemiseks erineva dispersiooniga üldkogumite korral ... 13

3.2.4 Studenti t-test keskväärtuste võrdlemiseks paarikaupa andmete korral ... 15

Lisa 1. Studenti t-kordaja väärtused ... 23

1 Statistilise andmeanalüüsi põhimõisted

1.1 Üldkogum ja valim

Üldkogumi (populatsiooni) all mõeldakse teatavat nähtust või protsessi looduses või ühiskonnas, mida statistilise andmeanalüüsi abil uuritakse. Üldkogumit kirjeldab lõputu hulk iseloomulikke tunnuseid.

Valimi all mõistetakse üldkogumi kohta kättesaadavat mõõteinformatsiooni, s.o. lõplikku arvu tunnuseid, mida mõõdetakse üldkogumi üksikutel objektidel.

Näide 1.1:

Üldkogum	**Elementaarobjektid**	**Mõõdetavad tunnused**
Maailma järved	Üksikud järved	Pindala, kõrgus merepinnast, sügavus, vee läbipaistvus ...
Peipsi järve vesi	N kohas x ajal võetud veeproovid	Temperatuur, pH, läbipaistvus, värvus, hapnikusisaldus ...
Eesti ilmastik aastal 2012	Iga kolme tunni mõõtmis-andmed n ilmajaamast	Temperatuur, niiskus, õhurõhk, pilvisus, kiirgusbilanss ...
TTS loengukursus	Üksikud üliõpilased	Sugu, vanus, pikkus, kaal, hinded, IQ, eriala ...
Eesti metsad	Üksikud metsatükid	Metsatüki keskmine vanus, keskmine kõrgus, kasvukoha-tüüp, aastane juurdekasv
Eesti metsad	Üksikud puud	Vanus, läbimõõt, kõrgus, liik, aastane juurdekasv

Valim peab olema üldkogumit esindav (representatiivne). Selleks peab olema

1. valim küllalt arvukas;
2. valimi jaotus lähedane üldkogumi jaotusele.

Valimi koostamisel peab üldkogumi objektidel olema võrdne tõenäosus valimisse sattumiseks. Lõpmatu valimi korral peab valimi koostamise eeskiri olema üldkogumi suhtes ühtlase jaotusega.

Näide 1.2: Olgu üldkogumiks ühe gümnaasiumi õpilased. Valimiks ei sobi siis näiteks kõik klasside 7 – 9 õpilased. Valiku tegemiseks võib:

- nummerdada kõik õpilased ning siis juhuslike arvude generaatori abil leida valimisse minevad õpilased;
- valida valimisse kõik õpilased, kelle isikukoodi viimane number on näiteks 3 või 4;

1.2 Statistilise tunnuse karakteristikud

Tunnuse karakteristikud on statistilist jaotust iseloomustavad parameetrid. Selgub, et väga paljudel juhtudel ei olegi uuritava suuruse täpset jaotusseadust vaja teada (ja sageli see polekski reaalselt võimalik), vaid piisab valimitest mõnede parameetrite leidmisest. Neist kõige tähtsam on **keskväärtus** ehk matemaatiline ootus:

$$m(x) = \bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1.1}$$

Suuruse x hajuvust keskväärtuse $m(x)$ ümber kirjeldab **empiiriline standardhälve**:

$$\sigma(x) = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \tag{1.2}$$

Suuruse x aritmeetilise keskväärtuse hajuvust kirjeldab **aritmeetilise keskmise empiiriline standardhälve**:

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{N}} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N \cdot (N - 1)}} \tag{1.3}$$

Lisaks kasutatakse statistikas veel palju erinevaid karakteristikuid (assümeetria kordaja, ekstsess, geomeetriline ja kaalutud keskmine, kvantiilid ja kvartiilid, mood jne), kuid selle kursuse raamidesse need ei mahu.

Näide 1.3: Olgu üldkogumiks ühe ilmajaama kahe meetri temperatuur ühel päeval. Valimiks võeti iga kolme tunni tagant mõõdetud temperatuur. Mõõtmisandmed olid järgmised: (12; 9; 8; 15; 21; 22; 19; 16) °C. Teeme tabeli:

nr	x	$x - m(x)$	$(x - m(x))^2$
1	12	–3	9
2	8	–7	49
3	7	–8	64
4	15	0	0
5	21	6	36
6	22	7	49
7	19	4	16
8	16	1	1
SUM	120	0	224
AVER	15,0		

Valimi keskväärtuseks on:

$$m(x) = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{120}{8} = 15\ °C\,.$$

Valimi empiiriliseks standardhälbeks on:

$$\sigma(x) = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{224}{7}} = \sqrt{32} = 5{,}66\ °C\,.$$

Valimi aritmeetilise keskmise empiiriline standardhälve on:

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{N}} = \frac{\sqrt{32}}{\sqrt{8}} = \sqrt{4} = 2\ °C\,.$$

Valimi empiiriline standardhälve annab infot üksikute mõõdetud temperatuuride (üksikelementide) hajuvuse kohta; aritmeetilise keskmise standardhälve annab infot valimi keskväärtuse hajuvuse kohta üldkogumi keskväärtusest.

2 Statistilised hüpoteesid ja nende tõestamine

2.1 Statistiliste hinnangute usaldusväärsus

Statistilise andmeanalüüsi põhiülesandeks on tunnuste parameetrite hindamine ning tunnustevaheliste seoste otsimine ja kindlakstegemine. Hinnatakse nii tunnuste väärtusi endid kui ka nende hinnangute usaldusväärsust. Usaldusväärsuse hindamise vajadus tekib seoses sellega, et me tahame meile valimi kujul kättesaadava informatsiooni põhjal otsustada üldkogumis kehtivate seoste üle. Kas see otsus on usaldusväärne ja kui suure vea me teeme üldkogumi kohta otsustades? Usaldusväärsuse hindamise kriteeriumina anname ette **statistilise olulisuse nivoo α**, mis näitab, kui suur on tõenäosus teha viga üldistades valimi põhjal tehtud järeldusi üldkogumi kohta.

2.2 Statistiline hüpotees

Statistiline hüpotees on väide üldkogumi või tema parameetrite kohta. Kontrollimiseks esitatakse alati kaks teineteist välistavat hüpoteesi, millest ainult üks saab olla tõene:

- sisukaks hüpoteesiks H1 nimetatakse väidet, mida soovitakse tõestada;

- nullhüpoteesiks H0 nimetatakse sisuka hüpoteesi vastuväidet.

Sisukas hüpotees tuleb olemasolevate andmete põhjal tõestada või tagasi lükata. Nullhüpoteesi ei ole võimalik tõestada, selle vastuvõtmine võib tähendada lihtsalt uuringute jätkamist.

Näited parameetri kohta käivatest hüpoteesidest:

- H0: $\nu = \nu_0$
  H1: $\nu \neq \nu_0$
  Kõige lihtsam hüpotees parameetri ν väärtuse kohta on väide, et see parameeter on võrdne mingi konstandiga ν_0.

- H0: $\nu \leq \nu_0$
  H1: $\nu > \nu_0$
  Ühepoolne hüpotees, mis väidab, et parameetri ν väärtus on mingist teadaolevast arvust ν_0 suurem.

- H0: $\nu \geq \nu_0$
  H1: $\nu < \nu_0$
  Ühepoolne hüpotees, mis väidab, et parameetri ν väärtus on väiksem kui mingi teadaolev konstant ν_0.

2.2.1 Vead hüpoteeside kontrollimisel

Et otsustusi tehakse valimi põhjal, siis kaasnevad sellega paratamatult vead. Võimalikud on järgmised vead:

	Võetakse vastu H0	Võetakse vastu H1
Kehtib H0	Õige	1. liiki viga
Kehtib H1	2. liiki viga	Õige

Hüpoteesid sõnastatakse nii, et eriti ebasoovitav on esimest liiki viga. Seda arvestades konstrueeritakse otsuse vastuvõtmise kriteerium nii, et esimest liiki vea tegemise suurim lubatav tõenäosus (olulisuse nivoo α) oleks võimalikult väike. Olulisuse nivooks valitakse mingi väike arv, sageli 0,1; 0,05; 0,01 (sõltuvalt selles, kui rasketele tagajärgedele võib 1. liiki vea tegemine viia). Füüsikas kasutatakse tavaliselt olulisuse nivood 0,05. Üldiselt toimub esimest liiki vea tõenäosuse vähendamine teist liiki vea arvelt.

Näide 2.1: Olgu meil püstitatud järgmine hüpoteeside paar:

H0: geenmuundatud toit kujutab ohtu inimeste tervisele;
H1: geenmuundatud toit ei kujuta ohtu inimeste tervisele.

1. liiki viga tehakse siis, kui võetakse vastu sisukas hüpotees H1 ning lubatakse kasutada GMO toiduained ning alles hiljem selgub, et see rikub inimeste tervist.

2. liiki viga tehakse siis, kui jäädakse nullhüpoteesi H0 juurde ja ei lubata kasutada GMO toiduaineid, kuigi need ei kujutaks ohtu inimese tervisele.

2.2.2 Statistiliste hüpoteeside tõestamine

Hüpoteeside kontrollimisel püütakse tõestada sisukas hüpotees nullhüpoteesi kummutamise teel. Selleks arvutatakse valimi andmete põhjal teatud teststatistik, mille teoreetiline jaotus nullhüpoteesi kehtivuse korral on teada. Juhul kui leitud teststatistiku väärtus on ebatõenäoline, võrreldes tema teoreetilise jaotusega, loetakse nullhüpotees kummutatuks ja sisukas hüpotees tõestatuks. Kui sisukat hüpoteesi tõestada ei õnnestu, jäädakse nullhüpoteesi juurde, mis võib tähendada nii seda, et

1) olukord vastas nullhüpoteesile;

kui ka seda, et

2) valimi maht oli liiga väike sisuka hüpoteesi tõestamiseks.

3 Testid statistiliste hüpoteeside tõestamiseks

Statistiliste hüpoteeside tõestamiseks kasutatakse erinevaid teste. Millist valida, sõltub tunnuse tüübist, tunnuse jaotusest, valimi suurusest ja testitavast karakteristikust.

Tunnuse jaotuse kohta tehtavatest eeldustest lähtudes jagatakse testid parameetrilisteks ja mitteparameetrilisteks.

Parameetrilised testid – nende puhul eeldatakse, et uuritav juhuslik suurus allub mingile tuntud tüüpjaotusele. Kõige sagedamini eeldatakse, et tunnus allub normaaljaotusele.

Mitteparameetrilised testid ei tee eeldusi uuritava tunnuse jaotuse kohta. Mitteparameetrilised jaotused on teoreetiliselt paremini põhjendatud, kuid nende võimsus on reeglina väiksem kui parameetrilistel testidel, sest nad raiskavad eelinformatsiooni tunnuse jaotusseaduse kohta.

3.1 Pearsoni X^2-test jaotusseadusele alluvuse testimiseks

Pearsoni χ^2-testi kasutatakse empiirilise jaotuse võrdlemiseks mõne klassikalise jaotusseadusega, enamasti näiteks normaaljaotusega. Esiteks tuleb leida jaotuse parameetrite sobivad väärtused (hinnangud), seejärel arvutatakse empiirilisele sagedustabelile vastavad teoreetilised sagedused ning lõpuks otsustatakse empiiriliste ja teoreetiliste sageduste kooskõla üle Pearsoni χ^2-testi abil.

Valimi jaotuse võrdlemiseks teoreetilise jaotusega tuleb andmed jagada klassidesse. Rusikareegel on, et klasside arv *k* võiks olla ligikaudu võrdne ruutjuurega valimi suurusest. Teoreetilise jaotuse, näiteks normaaljaotuse jaoks, arvutatakse samuti välja erinevatesse klassidesse sattuvate elementide arv, kusjuures teoreetilise jaotuse parameetriteks kasutatakse valimi keskväärtust ning standardhälvet. Valimi ja teoreetilise sageduse kooskõla hindamiseks arvutatakse χ^2-statistik valemiga:

$$\mathrm{X}^2 = \sum_{i=1}^{k} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i},$$

kus n_i on i-ndasse klassi sattunud valimi elementide arv;
$\tilde{n}_i$ on i-ndasse klassi sattunud teoreetilise jaotuse elementide arv;
k on klasside arv.

Kui vaadeldav valim oleks oletatava teoreetilise jaotusega (normaaljaotusega) üldkogumist, siis on χ^2-statistik χ^2-jaotusega vabadusastmete arvuga

$$\nu = k - r - 1,$$

kus *k* on klasside arv;

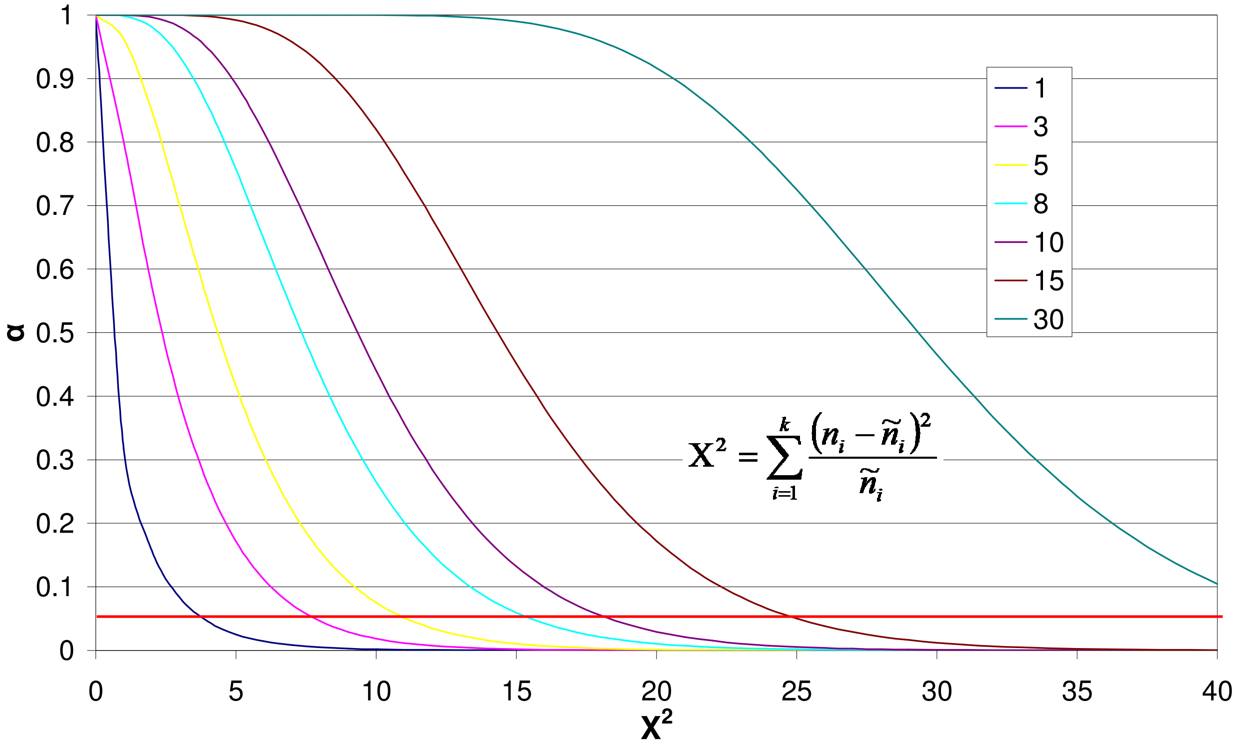
r on teoreetilise jaotuse parameetrite arv, mida on hinnatud valimist. Normaaljaotuse puhul arvutatakse valimist keskväärtus ning standardhälve, seega on $r = 2$.

Excelis arvutatakse χ^2-jaotuse täiendjaotusfunktsiooni väärtuseid funktsiooniga

α = CHITEST(χ^2; ν).

H0 juurde (valim allub võrreldavale teoreetilisele jaotusele) tuleb jääda, kui $\alpha > 0{,}05$;

H1 on tõestatud (valim EI ALLU võrreldavale teoreetilisele jaotusele), kui $\alpha < 0,05$.



Joonis 3.1. χ^2-jaotuse täiendjaotusfunktsioon sõltuvalt vabadusastmete arvust (erinevad värvid). H1 leiab kinnitust 95% usaldusnivool, kui funktsiooni väärtus on väiksem kui 0,05 (ehk punasest joonest all pool).

3.2 F-test normaaljaotusega üldkogumite dispersioonide võrdlemiseks

Olgu s_1^2 ja s_2^2 normaaljaotusega üldkogumist pärit valimite dispersioonid. Järjestame üldkogumid nii, et $s_1^2 > s_2^2$.

Olgu nullhüpoteesiks väide, et valimid on ühesuguse dispersiooniga üldkogumitest:

H0: $s_1^2 = s_2^2$.

Sisukaks hüpoteesiks on väide, et valimid on erinevate dispersioonidega üldkogumitest:

H1: $s_1^2 \neq s_2^2$.

Dispersioonide võrdlemiseks arvutatakse F-statistik:

$$F = \frac{s_1^2}{s_2^2},$$

mis on eelduse $s_1^2 = s_2^2$ korral F-jaotusega juhuslik suurus, vabadusastmete arvuga $N_1 - 1$ ja $N_2 - 1$, kus N_1 ja N_2 on vastavate valimite mahud.

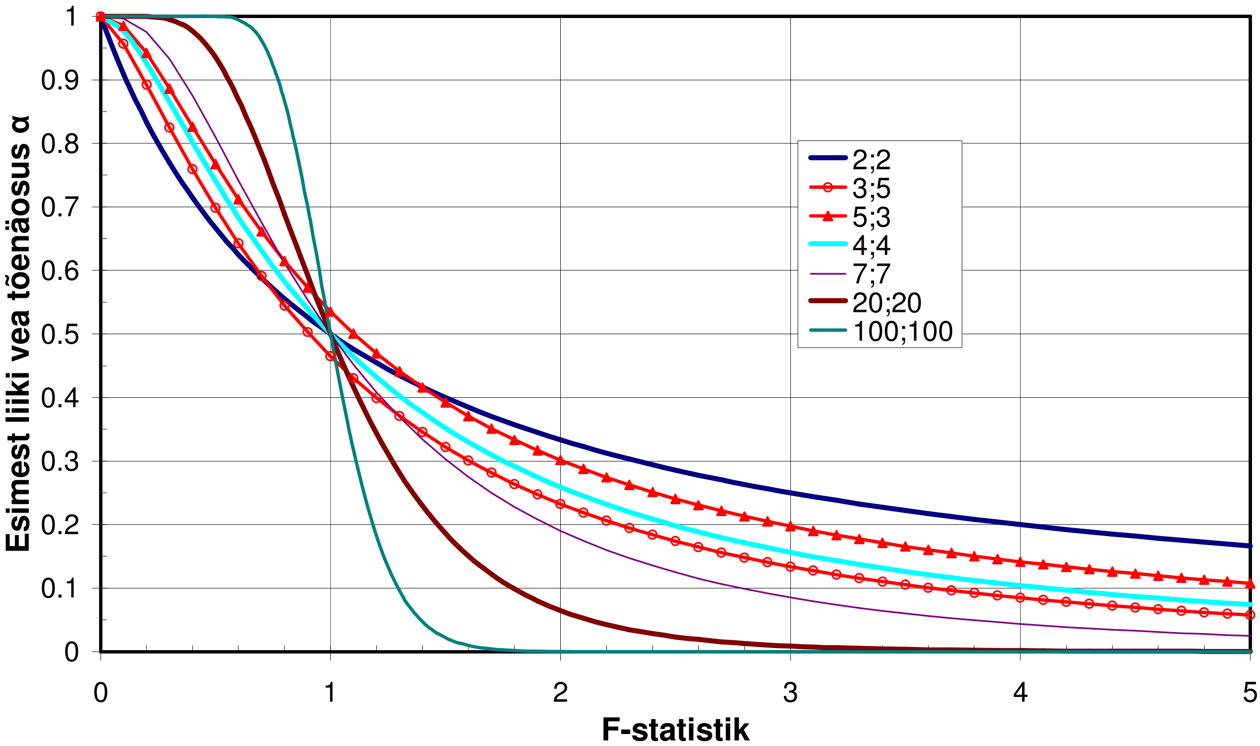
Sisukas hüpotees H1 loetakse tõestatuks, kui

$$F = \frac{s_1^{\ 2}}{s_2^{\ 2}} > F_{\frac{\alpha}{2}, N_1-1; N_2-1},$$

kus $F_{\frac{\alpha}{2}, N_1-1; N_2-1}$ on F-jaotuse α/2 täiendkvantiil, vabadusastmete arvuga $N_1 - 1$ ja $N_2 - 1$. Tõenäosus eksida, võttes vastu otsuse, et valimitele vastavate üldkogumite dispersioonid on erinevad, on maksimaalselt α.

Kui $F = \frac{s_1^{\ 2}}{s_2^{\ 2}} < F_{\frac{\alpha}{2}, N_1-1; N_2-1}$, siis tuleb jääda nullhüpoteesi H0 juurde. F-jaotuse täiendkvantiili arvutamiseks on näiteks Excelis käsk x = FINV(α; $N_1 - 1$; $N_2 - 1$).

Alternatiivne lähenemine on leida tõenäosus, et F-jaotus parameetritega $N_1 - 1$; $N_2 - 1$ on suurem kui arvutatud F-statistik. Selleks on Excelis käsk α = FDIST(F; $N_1 - 1$; $N_2 - 1$). Saadav tõenäosus α näitab, kui suur on tõenäosus teha esimest liiki viga, võttes vastu ühepoolse sisuka hüpoteesi.



Joonis 3.1. Esimest liiki vea tõenäosus α sõltuvalt F-statistikust ning valimite vabadusastmete arvust ühepoolse hüpoteesi korral.

OLULINE: F-jaotuse kuju sõltub vabadusastmete järjekorrast, ehk FDIST(x;3;5) ≠ FDIST(x;5;3).

Näide 3.1: 2012. aastal tegi ühes aines kontrolltöö 82 tudengit ning tulemuste keskväärtus oli m_2012 = 73,3 ning standardhälve oli s_2012 = 18,7.

2013 aastal tegi kontrolltöö 89 tudengit ning tulemuste keskväärtus oli m_2013 = 59,3 ning standardhälve oli s_2013 = 19,1. Kas kontrolltööde tulemuste dispersioonid on ühesugused usaldusnivool 95%?

Lahendus:

F-statistik on $F = \frac{s_1^{\,2}}{s_2^{\,2}} = \frac{19{,}1^2}{18{,}7^2} = 1{,}043.$

F-jaotuse α/2 täiendkvantiil on $F_{\frac{\alpha}{2},N_1-1;N_2-1} = FINV(0{,}025;81;88) = 1{,}533$.

Kuna $F = 1{,}043 < F_{\frac{\alpha}{2},N_1-1;N_2-1} = 1{,}533$, siis tuleb jääda nullhüpoteesi juurde ehk 2012. ja 2013. aasta kontrolltööde dispersioonid olid võrdsed.

Kui tahta teada, kui suur on tõenäosus teha esimest liiki viga, võttes vastu ühepoolse sisuka hüpoteesi, kasutame funktsiooni FDIST:

$$\alpha = FDIST(1{,}043;81;88) = 0{,}422\,.$$

Seega – tuleb jääda nullhüpoteesi juurde. Kui võtta vastu sisukas hüpotees, et dispersioonid on erinevad, siis on esimest liiki vea tegemise tõenäosus 42%.

3.3 Studenti t-test normaaljaotusele alluva juhusliku suuruse keskväärtuse kohta

3.3.1 Studenti t-test keskväärtuse võrdlemiseks konstandiga

Studenti test ehk t-test on normaaljaotusele alluva juhusliku suuruse keskväärtuste võrdlemiseks. T-testi puhul kasutatakse t-statistikut ehk Studenti suhet, mille väärtuse saab arvutada valimist järgnevalt:

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)}$$

siin N on valimi maht ja $s(z)$ on standardhälve. Saab tõestada, et kui üldkogum allub normaaljaotusele, siis t-statistik allub $N - 1$ vabadusastmega Studenti jaotusele. Meenutame, et Studenti jaotus sõltus ainult vabadusastmete arvust ning soovitud usaldusnivoost.

Studenti testi puhul peab sisuka hüpoteesi vastuvõtmiseks olema täidetud tingimus:

- ühepoolse hüpoteesi puhul $|t| > t_{\alpha,\nu}$;
- kahepoolse hüpoteesi puhul $|t| > t_{\frac{\alpha}{2},\nu}$,

siin $t_{\alpha,\nu}$ on Studenti parameeter, mis sõltub soovitud usaldusnivoost α ning vabadusastmete arvust $\nu = N - 1$, näiteks Excelis on Studenti parameeter arvutatav käsuga TINV(α; ν).

Näiteks olulisuse nivoo α = 0,05 puhul tuleb ühepoolse hüpoteesi kontrollimisel kasutada α = 0,05, kahepoolse hüpoteesi kontrollimisel aga α = 0,025, sest viga võib tulla mõlemas suunas.

Alternatiivne lähenemine on leida tõenäosus, et t-jaotus vabadusastmete arvuga ν = N – 1 on suurem kui arvutatud t-statistik. Selleks on Excelis käsk α = TDIST(t; N_1 – 1; *tails*). Siin parameetri *tails* väärtusest sõltub, kas lähtutakse ühepoolsest (*tails* = 1) või kahepoolsest (*tails* = 2) hüpoteesist. Saadav tõenäosus α näitab, kui suur on tõenäosus teha esimest liiki viga, võttes vastu sisuka hüpoteesi.

Näide 3.2: 2013 aastal tegi ühes aines kontrolltöö 89 tudengit ning tulemuste keskväärtus oli m_2013 = 59,3 punkti ning standardhälve oli s_2013 = 19,1 punkti. Varasemate aastate keskmine kontrolltöö tulemus oli 67 punkti. Kas 2013. aasta kontrolltöö tulemused olid oluliselt madalamad varasemate aastate keskmisest usaldusnivool 95%?

Lahendus:

Paneme kirja hüpoteesid:

H0:2013. aasta kontrolltöö tulemused ei olnud oluliselt madalamad kui varasemate aastate keskmine.

H1:2013. aasta kontrolltöö tulemused olid oluliselt madalamad kui varasemate aastate keskmine.

Leiame t-statistiku:

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)} = \sqrt{89}\,\frac{67 - 59{,}3}{19{,}1} = 3{,}80\,.$$

Kuna meil on tegemist ühepoolse hüpoteesiga, siis

$$t_{\alpha,\nu} = t_{0{,}05{,}88} = TINV(0{,}05;88) = 1{,}987$$

Kuna $t = 3{,}80 < t_{0{,}05{,}88} = 1{,}987$, siis saab vastu võtta sisuka hüpoteesi.

Alternatiivne võimalus ülesannet lahendada on leida esimest liiki vea tegemise tõenäosus:

α = TDIST(t; N_1 – 1; *tails*) = TDIST(3,80; 88; 1) = 0,00013 = 0,013%

Kuna esimest liiki vea tegemise tõenäosus α = 0,013% on väiksem kui lubatud 5%, siis võib vastu võtta sisuka hüpoteesi.

Arendame ülesannet edasi – kas samade keskväärtuste ja standardhälvete väärtustega saaks sisuka hüpoteesi vastu võtta ka siis, kui tudengeid oleks olnud ainut 20?

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)} = \sqrt{20}\,\frac{67 - 59{,}3}{19{,}1} = 1{,}803\,.$$

α = TDIST(t; N_1 – 1; *tails*) = TDIST(1,803; 19; 1) = 0,0436 = 4,4%

Kuna esimest liiki vea tegemise tõenäosus α = 4,3% on väiksem kui lubatud 5%, siis võib ka sellel juhul vastu võtta sisuka hüpoteesi.

3.3.2 Studenti t-test keskväärtuste võrdlemiseks võrdse dispersiooniga üldkogumite korral

Olgu meil teada kahe normaaljaotusega üldkogumi valimi aritmeetilised keskmised ning standardhälbed. Oletame, et valimid on **võrdse dispersiooniga üldkogumitest** (saab kontrollida näiteks F-testiga). Keskmiste võrdlemiseks arvutatakse t-statistik:

$$t = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\left[\frac{N_1 + N_2}{N_1 \cdot N_2}\right] \cdot \left[\frac{(N_1 - 1) \cdot s_1^2 + (N_2 - 1) \cdot s_2^2}{N_1 + N_2 - 2}\right]}}.$$

Vabadusastmete arv Studenti parameetri arvutamiseks on:

$$\nu = N_1 + N_2 - 2.$$

3.3.3 Studenti t-test keskväärtuste võrdlemiseks erineva dispersiooniga üldkogumite korral

Olgu meil teada kahe normaaljaotusega üldkogumi valimi aritmeetilised keskmised ning standardhälbed. Oletame, et valimid on **erineva dispersiooniga üldkogumitest** (saab kontrollida näiteks F-testiga). Keskmiste võrdlemiseks arvutatakse t-statistik:

$$t = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}.$$

Vabadusastmete arv Studenti parameetri arvutamiseks on:

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 + 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 + 1}} - 2.$$

Näide 3.3: Olgu meil kaks valimit normaaljaotusega üldkogumitest, nende valimi mahud, keskväärtused ning standardhälbed on järgmised: N_1 = 10; $m(x_1)$ =50; $s(x_1)$ = 15; N_2 = 20; $m(x_2)$ =45; $s(x_2)$ = 10. Kas esimese üldkogu keskväärtus on oluliselt suurem teise üldkogu keskväärtusest usaldusnivool 95%?

Lahendus:

Esmalt on vaja sooritada F-test kontrollimaks, kas üldkogumite dispersioonid on ühesugused või erinevad.

F-statistik on $F = \frac{s_1^{\,2}}{s_2^{\,2}} = \frac{15^2}{10^2} = 2{,}25.$

Kui tahta teada, kui suur on tõenäosus teha esimest liiki viga, võttes vastu ühepoolse sisuka hüpoteesi, kasutame funktsiooni FDIST:

$$\alpha = FDIST(2{,}25;9;19) = 0{,}065\,.$$

Kuna $\alpha = 0{,}065 > 0{,}05$, siis tuleb jääda H0 juurde, ehk üldkogumite dispersioonid on ühesugused.

Leiame nüüd t-statistiku ning vastava vabadusastmete arvu ühesuguste dispersioonide eeldusel:

$$t = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\left[\frac{N_1 + N_2}{N_1 \cdot N_2}\right] \cdot \left[\frac{(N_1 - 1)\cdot s_1^{\,2} + (N_2 - 1)\cdot s_2^{\,2}}{N_1 + N_2 - 2}\right]}} = \frac{|50 - 45|}{\sqrt{\left[\frac{10+20}{10\cdot 20}\right] \cdot \left[\frac{(10-1)\cdot 15^2 + (20-1)\cdot 10^2}{10+20-2}\right]}} = .$$

$$= \frac{5}{\sqrt{0{,}15 \cdot \left[\frac{2025+1900}{28}\right]}} = \frac{5}{4{,}585} = 1{,}090.$$

$$\nu = N_1 + N_2 - 2 = 10 + 20 - 2 = 28\,.$$

Esimest liiki vea tegemise tõenäosus on:

α = TDIST(*t*; *v*; *tails*) = TDIST(1,09; 28; 1) = 0,143 = 14,3% > 5%.

Seega tuleb jääda nullhüpoteesi juurde, ehk esimese üldkogu keskväärtus ei ole oluliselt suurem teise üldkogu keskväärtusest usaldusnivool 95%.

Kontrollime huvi pärast, kui suur tuleks α, kui dispersioonid oleksid olnud erinevad:

$$t = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{\sqrt{\frac{s_1^{\,2}}{N_1} + \frac{s_2^{\,2}}{N_2}}} = \frac{|50-45|}{\sqrt{\frac{15^2}{10} + \frac{10^2}{20}}} = \frac{5}{\sqrt{22{,}5+5}} = \frac{5}{5{,}24} = 0{,}953\,.$$

Vabadusastmete arv Studenti parameetri arvutamiseks on:

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1+1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2+1}} - 2 = \frac{(27{,}5)^2}{\frac{\left(\frac{15^2}{10}\right)^2}{10+1} + \frac{\left(\frac{10^2}{20}\right)^2}{20+1}} - 2 = \frac{(27{,}5)^2}{46{,}02 + 1{,}19} - 2 = 14\,.$$

Esimest liiki vea tegemise tõenäosus on:

α = TDIST(t; ν; *tails*) = TDIST(0,953; 14; 1) = 0,178 = 17,8% > 5%.

Seega ka sellisel juhul tuleks jääda H0 juurde.

3.3.4 Studenti t-test keskväärtuste võrdlemiseks paarikaupa andmete korral

Oletame, et meil on tegemist valimiga normaaljaotusega üldkogumist, kus valimi objekte on mõõdetud kahel korral, näiteks enne teatud toimingut ja pärast seda. Testi tulemusena soovime teada saada, kas läbiviidud toiming muutis üldkogumi keskväärtust või mitte.

Keskmiste võrdlemiseks arvutatakse t-statistik kujul

$$t = \sqrt{N} \cdot \frac{\overline{\Delta}}{s(\Delta)},$$

kus $\overline{\Delta}$ on mõõtmispaaride vahe keskväärtus ning $s(\Delta)$ on mõõtmispaaride vahe standardhälve. Vabadusastmete arv on $\nu = N - 1$.

Näide 3.4: Kahel kinnisvarahindajal lasti üksteisest sõltumatult hinnata kümmet kinnisvaraobjekti. Tulemused olid järgmised (kEUR):

Hindaja 1	60	44	40	92	33	37	48	69	28	71
Hindaja 2	62	47	40	93	31	40	50	68	29	74
Δ	–2	–3	0	–1	2	–3	–2	1	–1	–3

Kas hinnangud olid süstemaatiliselt erinevad, usaldusnivool 95%?

Lahendus:

Leiame statistikud:

$$m(\Delta) = \frac{\sum_{i=1}^{N} \Delta_i}{N} = \frac{-12}{10} = -1{,}2\,;$$

$$s(\Delta) = \sqrt{\frac{\sum_{i=1}^{N} (\Delta_i - m(\Delta))^2}{N-1}} = 1{,}75 .$$

Leiame t-statistiku:

$$t = \sqrt{N} \cdot \frac{\overline{\Delta}}{s(\Delta)} = \sqrt{10} \cdot \frac{1{,}2}{1{,}75} = 2{,}17 ,$$

Esimest liiki vea tegemise tõenäosus on **kahepoolse hüpoteesi** korral:

α = TDIST(t; v; *tails*) = TDIST(2,17; 9; 2) = 0,058 = 5,8% > 5%,

Seega ei leidnud kinnitust sisukas hüpotees, et hinnangud on süstemaatiliselt erinevad.

Püstitame uue sisuka hüpoteesi, et teine hindaja hindas kinnisvaraobjekte esimesest kallimaks.

Esimest liiki vea tegemise tõenäosus on nüüd **ühepoolse hüpoteesi** korral:

α = TDIST(t; v; *tails*) = TDIST(2,17; 9; 1) = 0,029 = 2,9% < 5%,

seega leidis kinnitust sisukas hüpotees, et teine hindaja hindas kinnisvaraobjekte esimesest kallimaks (vale järelduse tõenäosus alla 2,9%).

4 Korrelatsioonanalüüs

4.1 Lineaarne korrelatsioonikordaja

Kahe juhusliku suuruse üheaegsel mõõtmisel saadud andmetest on võimalik leida nendevahelise lineaarse seose olemasolu ning tugevus. Kui suuruste vaheline seos pole lineaarne, on sageli võimalik leida teisendus (ruutu võtmine, logaritmimine vms) mille järel seos juba muutub lineaarseks.

Lineaarse korrelatsioonikordaja parim hinnang saadakse:

$$R = \frac{\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left(\sum_{i=1}^{N} (x_i - \overline{x})^2\right)\left(\sum_{i=1}^{N} (y_i - \overline{y})^2\right)}} .$$

Korrelatsioonikordaja standardhälve on:

$$s(R) = \sqrt{\frac{1-R^2}{N-2}} .$$

Lineaarse korrelatsioonikordaja parima hinnangu saab leida Exceli käsuga CORREL($\{x_{\mathrm{i}}\}$;$\{y_{\mathrm{i}}\}$).

Kui tahta arvutada paljude samaaegselt mõõdetud suuruste omavahelisi korrelatsioone, siis selleks on Excelis paketi *Tools/DataAnalysis* all korrelatsioonimaatriksi arvutamise pakett *Correlation*.

Determinatsioonikordaja

$$D = R^2$$

näitab, milline osa koguhajuvusest on seletatav korrelatsiooniga.

4.2 Lineaarse korrelatsioonikordaja hinnangu usaldatavus

Valimi põhjal arvutatud lineaarse korrelatsioonikordaja punkthinnangu usaldatavuse kontrollimiseks püstitatakse nullhüpotees H0: lineaarset seost juhuslike suuruste *X* ja *Y* vahel ei ole, üldkogumi regressioonisirge tõus on null:

H0: $k = 0$.

Sisukaks hüpoteesiks on väide, et üldkogumi regressioonsirge tõus erineb nullist:

H1: $k \neq 0$.

Püstitatud hüpoteeside kontrolliks arvutatakse F-statistik valemiga

$$F = \frac{(N-2)\cdot R^2}{1-R^2}.$$

Nullhüpoteesi õigsuse korral on antud statistik F-jaotusega, vabadusastmete arvuga $\nu_1 = 1$ ja $\nu_2 = N - 2$. Sisukas hüpotees H1 (tunnuste vahel on usaldusväärne lineaarne seos) loetakse tõestatuks, kui

$$F = \frac{(N-2)\cdot R^2}{1-R^2} > F_{\alpha;1;N-2},$$

kus $F_{\alpha;1;N-2}$ on F-jaotuse α-täiendkvantiil. Tõenäosus eksida, väites, et üldkogumi regressioonsirge tõus erineb nullist , on maksimaalselt α. Juhul kui

$$F = \frac{(N-2)\cdot R^2}{1-R^2} \leq F_{\alpha;1;N-2},$$

tuleb jääda nullhüpoteesi juurde, st me ei saa tõestada olulisuse nivooga α, et üldkogumi regressioonsirge tõus on nullist erinev.

Mõnes allikas pakutakse hüpoteeside kontrolliks F-testi asemel t-testi, kuid kuna $F_{\alpha;1;N-2} \equiv \left(t_{\alpha;N-2}\right)^2$, siis annavad mõlemad testid identseid tulemusi.

Näide 4.1: Kui suur peab olema korrelatsioonikordaja *R*, et regressioonsirge tõus oleks statistiliselt usaldusväärselt nullist erinev usaldusnivool 95%, kui andmepaaride arv on a) $N = 8$; b) $N = 100$; c) $N = 10000$?

Lahendus:

Lähtume põhivalemist:

$$F = \frac{(N-2)\cdot R^2}{1-R^2} > F_{\alpha;1;N-2} = F_{kr}$$
$$(N-2)\cdot R^2 > F_{kr}\cdot(1-R^2)$$
$$N\cdot R^2 - 2\cdot R^2 > F_{kr} - F_{kr}\cdot R^2$$
$$R^2(N-2+F_{kr}) > F_{kr}$$
$$|R| > \sqrt{\frac{F_{kr}}{N-2+F_{kr}}}$$

$$F_{\alpha;1;8-2} = FINV(0{,}05;1;6) = 5{,}987 \;\Rightarrow\; |R| > \sqrt{\frac{F_{kr}}{N-2+F_{kr}}} = \sqrt{\frac{5{,}987}{8-2+5{,}987}} = 0{,}707.$$

$$F_{\alpha;1;100-2} = FINV(0{,}05;1;98) = 3{,}938 \;\Rightarrow\; |R| > \sqrt{\frac{F_{kr}}{N-2+F_{kr}}} = \sqrt{\frac{3{,}938}{100-2+3{,}938}} = 0{,}197.$$

$$F_{\alpha;1;10000-2} = FINV(0{,}05;1;9998) = 3{,}842 \;\Rightarrow\; |R| > \sqrt{\frac{F_{kr}}{N-2+F_{kr}}} = \sqrt{\frac{3{,}842}{10000-2+3{,}938}} = 0{,}0196.$$

Seega peab N = 8 andmepaari vahelise korrelatsiooni absoluutväärtus olema suurem kui 0,71, et trend oleks usaldusväärne, samas kui N = 100 andmepaari vahelise korrelatsiooni absoluutväärtus peab olema suurem kui 0,20 ning N = 10000 andmepaari korral kõigest suurem kui 0,02.

4.3 Punkthinnang regressioonsirge tõusule

Juhusliku suuruse Y regressioonisirge võrrandi

$$Y = k\cdot X + b$$

leidmiseks X suhtes on vaja mõõta kahe tunnuse X ja Y väärtusi.

Kasutades vähimruutude meetodit, minimeeritakse mõõdiste ja lähendussirge vaheliste hälvete ruutude summa:

$$\sum_{i=1}^{N}\left(y_i - kx_i - b\right)^2 = S \to \min$$

Võrdsustades osatuletised sirge tõusu k ja vabaliikme b järgi nulliga ning lahendades saadud võrrandisüsteemi, saab sirge parameetritele järgmised hinnangud, samuti Exceli vastavad funktsioonid:

$$k = \frac{\sum_{i=1}^{N}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2};$$ k = SLOPE(y-plokk; x-plokk);

$b = \bar{y} - k\cdot\bar{x}$; b = INTERCEPT(y-plokk, x-plokk).

Regressioonsirge parameetrite abil saab funktsioontunnuse väärtuseid prognoosida argumenttunnuse suvalise väärtuse korral:

$$\hat{y}_i = k \cdot x_i + b \,.$$

Mõõdetud funktsioontunnuse väärtuse y_i ja selle prognoosi y_i vahet nimetatakse prognoosijäägiks ε_i:

$$\varepsilon_i = y_i - \hat{y}_i \,.$$

Regressioonsirge sobivuse hindamiseks arvutatakse jääkdispersioon:

$$s_e^{\,2} = \frac{1}{N-2} \sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2 \,.$$

Ruutjuurt jääkdispersioonist nimetatakse jääkstandardhälbeks ehk prognoosiveaks ehk mudeli standardveaks:

$$s_e = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2} \,.$$

Jääkstandardhälve iseloomustab funktsioontunnuse keskmist erinevust regressioonijoonest.

Näide 4.2: Tudengid mõõtsid kuue erineva ekraani diagonaali pikkuse d ning pikselite pindala S ning said allolevad tulemused. Leia diagonaali pikkuse ja pikselite pindala vaheline korrelatsioon, kontrolli trendi usaldusväärsus usaldusnivool 95%, leia lineaarse regressioonsirge parameetrid ning jääkstandardhälve, samuti pikselite pindala standardhälve. Sinisega täidetud tabeli osa on juurde arvutatud.

nr	***d* (in)**	***S* (mm²)**	***d – m*(*d*)**	**(*d – m*(*d*))²**	***S – m*(*S*)**	**(*S – m*(*S*))²**	**(*d – m*(*d*)) (*S – m*(*S*))**
1	5	0,014	-11	121	-0,036	0,001296	0,396
2	13	0,047	-3	9	-0,003	0,000009	0,009
3	17	0,032	1	1	-0,018	0,000324	-0,018
4	17	0,063	1	1	0,013	0,000169	0,013
5	20	0,066	4	16	0,016	0,000256	0,064
6	24	0,078	8	64	0,028	0,000784	0,224
SUM	**96**	**0,3**	**0**	**212**	0	0,002838	0,688

Lahendus:

Lineaarse korrelatsioonikordaja parim hinnang saadakse:

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^{N}(x_i - \bar{x})^2\right)\left(\sum_{i=1}^{N}(y_i - \bar{y})^2\right)}} = \frac{0{,}688}{\sqrt{212 \cdot 0{,}00284}} = 0{,}887.$$

Selleks, et trend oleks usaldusväärne, peab korrelatsioonikordaja R olema suurem kui

$$|R| > \sqrt{\frac{FINV(0{,}05;1;4)}{N - 2 + FINV(0{,}05;1;4)}} = \sqrt{\frac{7{,}71}{6 - 2 + 7{,}71}} = 0{,}811.$$

Seega on trend usaldusväärne.

Lineaarse regressioonsirge parameetrid on:

$$k = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{0{,}688}{212} = 0{,}003245;$$

$$b = \bar{y} - k \cdot \bar{x} = 0{,}3/6 - 0{,}003245 \cdot 16 = -0{,}0019.$$

Täiendame jääkstandardhälbe arvutamiseks tabelit:

nr	***d* (in)**	***S* (mm^2)**	$\hat{S}$	$S - \hat{S}$	$(S - \hat{S})^2$
1	5	0,014	0.014301887	-0.0003	0,1E-06
2	13	0,047	0.040264151	0.0067	45,4E-06
3	17	0,032	0.053245283	-0.0212	451,4E-06
4	17	0,063	0.053245283	0.0098	95,2E-06
5	20	0,066	0.062981132	0.0030	9,1E-06
6	24	0,078	0.075962264	0.0020	4,2E-06
SUM	**96**	**0,3**	0.3	4,3E-17	605,2E-06

Regressioonsirge jääkstandardhälve on:

$$s_e(S) = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{6-2}605{,}2 \cdot 10^{-6}} = 0{,}0123.$$

Pikselite pindala standardhälve on:

$$s(S) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(S_i - \overline{S}\right)^2} = \sqrt{\frac{1}{6-1}0{,}002838} = 0{,}0238$$

Nagu näeme, on pikselite pindala lähendamisel keskväärtusega standardhälve praktiliselt kaks korda suurem kui lähendamisel regressioonsirgega.

4.4 Vahemikhinnang regressioonsirge tõusule

Eespool toodud regressioonsirge parameetrite arvutusvalemitega arvutatavad suurused on valimihinnangud. Nende täpsuse hindamiseks eeldatakse, et prognoosijäägid on normaaljaotusega. Sellisel juhul on regressioonsirge kordajate hinnangud normaaljaotusega, mille standardhälbed saab arvutada valemitega:

$$s(k) = s_e \cdot \sqrt{\frac{1}{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2}};$$

$$s(b) = s_e \cdot \sqrt{\frac{1}{N} + \frac{\overline{d}^2}{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2}}.$$

Näide 4.2 jätk: leiame sirge parameetrite standardhälbed:

$$s(k) = s_e \cdot \sqrt{\frac{1}{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2}} = 0{,}0123 \cdot \sqrt{\frac{1}{212}} = 0{,}000845;$$

$$s(b) = s_e \cdot \sqrt{\frac{1}{N} + \frac{\overline{d}^2}{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2}} = 0{,}0123 \cdot \sqrt{\frac{1}{6} + \frac{16^2}{212}} = 0{,}0144.$$

Sirge parameetrite vahemikhinnangud usaldusnivool 95% on leitavad kasutades Studenti kattetegurit vabadusastmete $\nu = N - 2$ jaoks:

$$t_{0{,}05;6-2} = TINV(0{,}05;4) = 2{,}78.$$

Sirge parameetrite vahemikhinnangud on seega:

$$\left[k - t \cdot s(k); k + t \cdot s(k)\right] = \left[0{,}00324 - 2{,}78 \cdot 0{,}000845; 0{,}00324 + 2{,}78 \cdot 0{,}000845\right] = \left[0{,}00089; 0{,}00559\right];$$

$$\left[b - t \cdot s(b); b + t \cdot s(b)\right] = \left[-0{,}0019 - 2{,}78 \cdot 0{,}0144; -0{,}0019 + 2{,}78 \cdot 0{,}0144\right] = \left[-0{,}042; 0{,}038\right].$$

Seega tõesti on selle sirge tõus statistiliselt nullist erinev. Sirge vabaliige aga ei ole statistiliselt nullist erinev.

Regressioonalüüsi tegemiseks on Excelis paketi *Tools/DataAnalysis* all regressioonsirge arvutamise pakett *Regression*.

Kasutades seda näide 4.2 andmetel, saame väljundiks

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.88698099
R Square	0.786735277
Adjusted R Square	0.733419097
Standard Error	0.012300867
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.002232755	0.002232755	14.75603217	0.018438132
Residual	4	0.000605245	0.000151311		
Total	5	0.002838			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.001924528	0.014419921	-0.133463169	0.900272348	-0.041960646	0.03811159
d (in)	0.003245283	0.000844827	3.841358116	0.018438132	0.000899667	0.005590899

Näeme, et statistikapakett andis samad tulemused, ainult oluliselt vähema vaevaga.

Kasutatud kirjandus:

- Andmeanalüüs MS Exceli abil. Tanel Kaart, 2000. http://www.eau.ee/~ktanel/kool_ja_too/stat_excelis/
- Matemaatiline statistika MS Excel keskkonnas. Andres Kiviste, 1999
- Matemaatiline statistika. Ülesannete kogu. Kadrin Keres, Aleksander Levin, 2006
- Mõõtmised ja mõõtemääramatused loengukonspekt, Erko Jakobson, 2012
- Statistika ja andmeanalüüsi kursus. Evely Leetma, http://math.ut.ee/~evelyl/Statistika.php?id=0
- Statistiline andmeanalüüs. Loengukonspekt keskkonnatehnoloogia eriala üliõpilastele. Ülle Kikas, 2002.
- Tõenäosusteooria ja statistika II vihik. Rein Rõõm, 2012

Lisa 1. Studenti t-kordaja väärtused

Studenti t-kordaja väärtused $t_\nu(p)$ sõltuvalt vabadusastmete arvust $\nu = N - 1$ ja soovitavast usaldusnivoost p.

Vabadusastmete arv $\nu = N - 1$	Osa p protsentides					
	68,27(I)	90	95	95,45(I)	99	99,73(I)
1	1,84	6,31	12,71	13,97	63,66	235,80
2	1,32	2,92	4,30	4,53	9,92	19,21
3	1,20	2,35	3,18	3,31	5,84	9,22
4	1,14	2,13	2,78	2,87	4,60	6,62
5	1,11	2,02	2,57	2,65	4,03	5,51
6	1,09	1,94	2,45	2,52	3,71	4,90
7	1,08	1,89	2,36	2,43	3,50	4,53
8	1,07	1,86	2,31	2,37	3,36	4,28
9	1,06	1,83	2,26	2,32	3,25	4,09
10	1,05	1,81	2,23	2,28	3,17	3,96
11	1,05	1,80	2,20	2,25	3,11	3,85
12	1,04	1,78	2,18	2,23	3,05	3,76
13	1,04	1,77	2,16	2,21	3,01	3,69
14	1,04	1,76	2,14	2,20	2,98	3,64
15	1,03	1,75	2,13	2,18	2,95	3,59
16	1,03	1,75	2,12	2,17	2,92	3,54
17	1,03	1,74	2,11	2,16	2,90	3,51
18	1,03	1,73	2,10	2,15	2,88	3,48
19	1,03	1,73	2,09	2,14	2,86	3,45
20	1,03	1,72	2,09	2,13	2,85	3,42
25	1,02	1,71	2,06	2,11	2,79	3,33
30	1,02	1,70	2,04	2,09	2,75	3,27
35	1,01	1,70	2,03	2,07	2,72	3,23
40	1,01	1,68	2,02	2,06	2,70	3,20
45	1,01	1,68	2,01	2,06	2,69	3,18
50	1,01	1,68	2,01	2,05	2,68	3,16
100	1,005	1,660	1,984	2,025	2,626	3,077
∞	1,000	1,645	1,960	2,000	2,576	3,000

(I) Suuruse x jaoks, mida kirjeldab normaaljaotus keskväärtusega $\bar{x}$ ja standardhälbega σ_x, sisaldab vahemik $x \pm k\sigma_x$ $p = 68{,}27$; 95,45 ja 99,73 protsenti jaotusest vastavalt $k = 1$; 2 ja 3 korral.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Sissejuhatus:

Juhuslikes suurustes või nähtustes ilmnevaid mitte-juhuslikke seaduspärasusi uurib, põhjendab ja selgitab <u>tõenäosusteooria</u>. Juhuslike suuruste tõenäosuslike seaduspärasuste eksperimentaalse uurimisega ja põhjendamisega tegeleb <u>matemaatiline statistika</u>.

Kui <u>tõenäosusteooria</u> tegeleb peaasjalikult juhuslike sündmuste ja juhuslike suuruste analüüsiga, keskendudes tõenäosustele, jaotusfunktsioonidele ja juhulike suuruste arvkarakteristikute uurimisele, siis <u>matemaatilise statistika</u> eesmärgiks on korduvatel mõõtmistel või vaatlustel saadud andmete statistiliste seaduspärasuste väljaselgitamine ja põhjendamine.

Sissejuhatus:

On olemas kolme sorti valesid: suured valed, väiksed valed ja statistika.

Mark Twain

Sissejuhatus:

1. **Eestis oli 2010. aastal täistööajaga töötaja keskmine brutokuutöötasu 819 eurot. Võrreldes 700 eurose kuupalgaga inimesega saavad**
   a) **üle poolte täistööajaga töötajatest temast <u>kõrgemat</u> palka.**
   b) **üle poolte täistööajaga töötajatest temast <u>madalamat</u> palka.**

   2010. aasta mediaanpalk oli 688 eurot.

2. **Järjesta tõenäosuse järgi, et juhuslikult valitud isik aasta jooksul (2010 aasta andmetel):**

a)	saab lotomiljonäriks (kroonides)	V	13 tk
b)	hukkub liikluses	II	79 tk
c)	sooritab enesetapu	I	221 tk
d)	kaitseb LOTE-s doktorikraadi	III	47 tk
e)	on Põlva vallavolikogu liige	IV	15 tk

Erko Jakobson

Teenistuskäik

Töökoht ja amet	2013 - ... Tartu Observatoorium; Teadur (0.70) 2012 - ... Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Füüsika Instituut, Atmosfäärifüüsika labor ; Teadur (0.50) 2010 - 2012 Tartu Observatoorium; Järeldoktor (1.00) 2010 - 2011 Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Füüsika Instituut, Atmosfäärifüüsika labor ; Insener (0.50) 2008 - 2009 Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Füüsika Instituut, Atmosfäärifüüsika labor ; insener (0.70) 2008 - 2009 Tartu Ülikool, Bioloogia-geograafiateaduskond; Teadur (0.30) 2006 - 2008 Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Ökoloogia- ja Maateaduste Instituut, Loodusgeograafia ja maastikuökoloogia õppetool; Erakorraline teadur (0.50) 2005 - 2008 Tartu Ülikool, Loodus- ja tehnoloogiateaduskond, Tartu Ülikooli Keemia Instituut, Katsekoda; insener (0.50) 2001 - 2002 AS Alarmnet, tehnik
Haridustee	2004 - 2009 PhD, keskkonnafüüsika, Tartu Ülikool 2002 - 2004 MSc, keskkonnafüüsika, Tartu Ülikool 1997 - 2002 BSc, füüsika,Tartu Ülikool 1994 - 1997 Võru Kreutzwaldi Gümnaasium
Teadusorganisatsiooniline ja -administratiivne tegevus	2009 - ... Magistriõppekava "Rakenduslik mõõteteadus" kaitsmiskomisjoni liige


tara
taraARCTIC 2007-2008

Õpiväljundid:

Õpiväljundid, ehk kursuse positiivsele hindele läbinud üliõpilane:

1. mõistab tõenäosusteooria ja statistika põhimõisteid;
2. oskab lahendada tavalisemaid tõenäosusteooria ülesandeid;
3. oskab luua seoseid tõenäosusteooria ja statistika vahel;
4. oskab leida andmestikke kirjeldavaid karakteristikuid (aritmeetiline keskmine, mediaan, standardhälve);
5. oskab kontrollida statistilisi hüpoteese kahe andmestiku võrdlemiseks (t-test, z-test);
6. on võimeline teostama korrelatsioon- ja regressioonanalüüsi.

Hindamismeetodid:

Hindamismeetoditeks on:

1. Moodle testid (10%);
2. kirjalik test (20%) tõenäosusteooria ülesannete peale;
3. grupitöö (10%);
4. kirjalik eksam (60%).

Koondhinne arvutatakse vastavalt praegu kehtivale %-süsteemile ("A" = >90% jne).

Eksamile pääsemise eelduseks on Moodle testidest, kirja-likust testist ja grupitööst vähemalt 50% punktide saamine.

Eksam on kirjalik, 50% ülesannete lahendamine, 50% teooria.

Hindamismeetodid:

Grupitöö on mingi andmebaasi statistiliste seoste ja hüpoteeside otsimine-kontrollimine vabalt valitud arvutusprogrammiga ning saadud tulemuste vormistamine. Grupi suurus – kuni 5 tudengit.

Grupitöö eesmärk:

- **Praktilise kogemuse saamine andmete statistilise analüüsiga. Väga tõenäoliselt sisaldab ka lõputöö statistilist andmetöötlust.**
- **Statistikaprogrammidega tutvumine, sest 3 EAP raames arvutipraktikumi ei jõua teha.**
- **Õpitu rakendamine praktikas.**

Hindamismeetodid:

Testi ja eksami ülesannete osas on spikerdamine limiteeritud. Kaastudengeid, raamatuid, konspekte, arvuteid, telefone jne pole lubatud kasutada. Laual võib olla kalkulaator ning üks A4 formaadis lehekülg vabalt valitud käsitsi kirjutatud teksti, nagu valemid, definitsioonid, jne.

Ajakava:

1. – 7. nädal – tõenäosusteooria;

8. Nädal – Kirjalik test (3. aprill 2013);

9. nädal – testi analüüs, sissejuhatus statistikasse, grupitööde planeerimine;

10. – 14. nädal – statistika.

15. – 16. nädal – statistika, grupitööde analüüs.

Osa materjalist kordab aines “Mõõtmised ja mõõtemääramatused” õpitut (jaotustihedus, pidevad jaotused jms).

Moodle

**Paralleelselt aine lugemisega on tegemisel ka Moode e-tugi.
kohustuslik on osaleda ainult kontrolltööl, grupitööl ning eksamil.
Kõik loengus esitatavad materjalid panen Moodle keskkonda üles.**

https://moodle.ut.ee

Tõenäosusteooria ja statistika (LOFY.03.056)

Palun lisage Moodlesse enda pilt ja õpitav eriala!

Juhuslikud sündmused:

Tõenäosusteooria baaskontseptsioonideks on <u>katse</u> ja <u>sündmus</u>. Katse all mõistame teatud tingimuste kompleksi realiseerumist, mille tulemusena võivad toimuda mingid sündmused. Tingimused võivad olla loodud kunstlikult (tahtlikult), või nad eksisteerivad sõltumatult eksperimentaatorist. Võime öelda ka, et katse on niisugune olukord või seisund, mille tulemusena võivad toimuda mingid sündmused.

Sündmus – fakt või ilming, mis leiab aset katses.

Juhuslikud sündmused:

- SÜNDMUS
  - JUHUSLIK SÜNDMUS
  - DETERMINISTLIK SÜNDMUS
    - KINDEL SÜNDMUS
    - VÕIMATU SÜNDMUS

Lihtsamateks juhuslikeks sündmusteks on täringu viskamisel silmade tulek, kaardi võtmine kaardipakist, mündi viskamine, võit loteriil, inimese eluiga jne.

Juhuslikud sündmused:

<u>Kindlaks sündmuseks</u> **Ω nimetatakse sündmust, mis katse tulemusena alati toimub, näiteks visatud kivi kukub alati maa peale tagasi (ei kehti kosmoses); iga elusolend hukub temperatuuril 1000 ºC.**

<u>Võimatu sündmus</u> **Φ on sündmus, mis vaadeldava katse tulemusena kunagi ei toimu, näiteks 7 silma saamine täringuviskel või varahommikune temperatuur akna taga –275 ºC.**

Juhuslikud sündmused:

Kui sündmuse A toimumisest järeldub sündmuse B toimumine, siis tähistame

$$A \subset B$$

s.t. A kuulub B-sse. Sel korral nimetatakse sündmust A ka sündmuse B alamsündmuseks või osasündmuseks ja sündmust B sündmuse A ülemsündmuseks.

Näiteks, kui sündmus A tähendab kõiki TT kursusele registreerinuid ja B – kõiki TT kursusele registreerinud tütarlapsi, siis ilmselt $B \subset A$.

Juhuslikud sündmused:

Sündmusi A ja B nimetatakse võrdseteks või samasteks või identseteks (ekvivalentseteks), kui kehtivad samaaegselt seosed

$$A \subset B$$

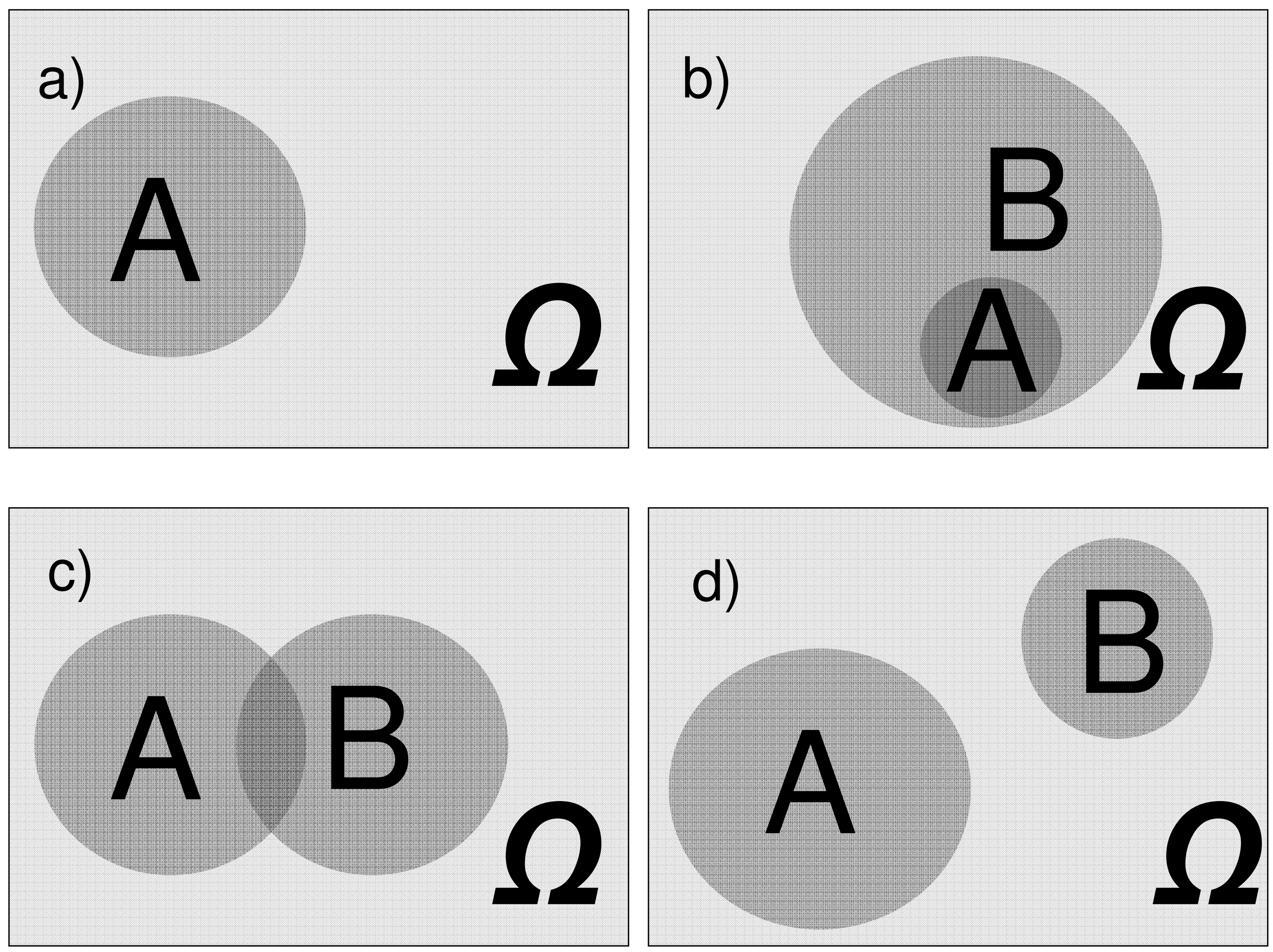
$$B \subset A$$

sel korral kirjutatakse

$$A = B$$

Venni diagrammid:

Graafiliselt kujutame edaspidi sündmusi punktihulkadena Venni diagrammi näol.



Tehted sündmustega:

Sündmuste vahelised elementaartehted on järgmised:

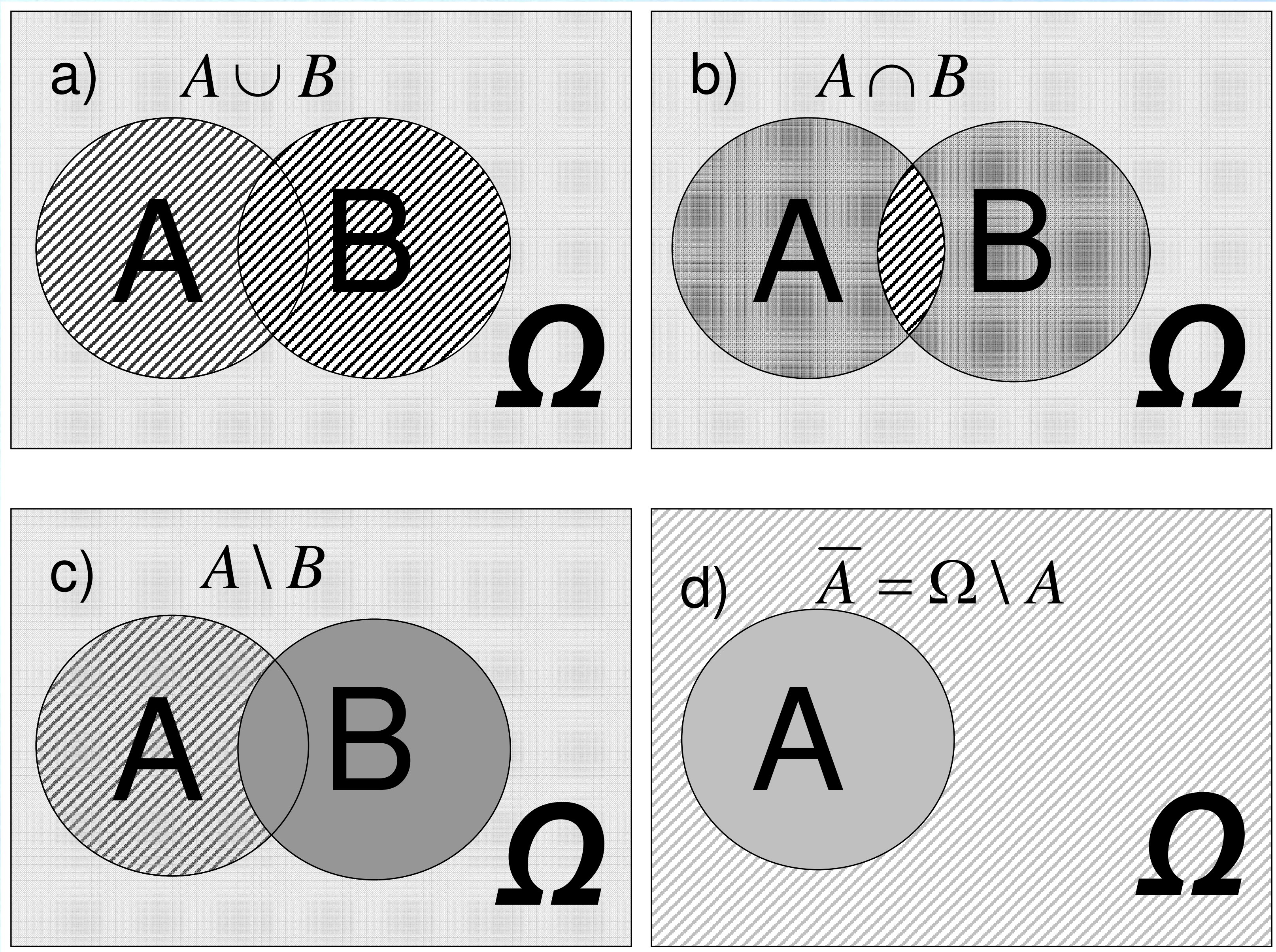
a) summa (või): $A \cup B$ – toimub kas A, B või nii A kui ka B.

b) korrutis (ja): $A \cap B$ – toimuvad nii A kui ka B.

c) vahe: $A \setminus B$ – toibub A, aga B ei toimu.

d) vastandsündmus: $\overline{A} = \Omega \setminus A$ – ei toimu sündmust A.

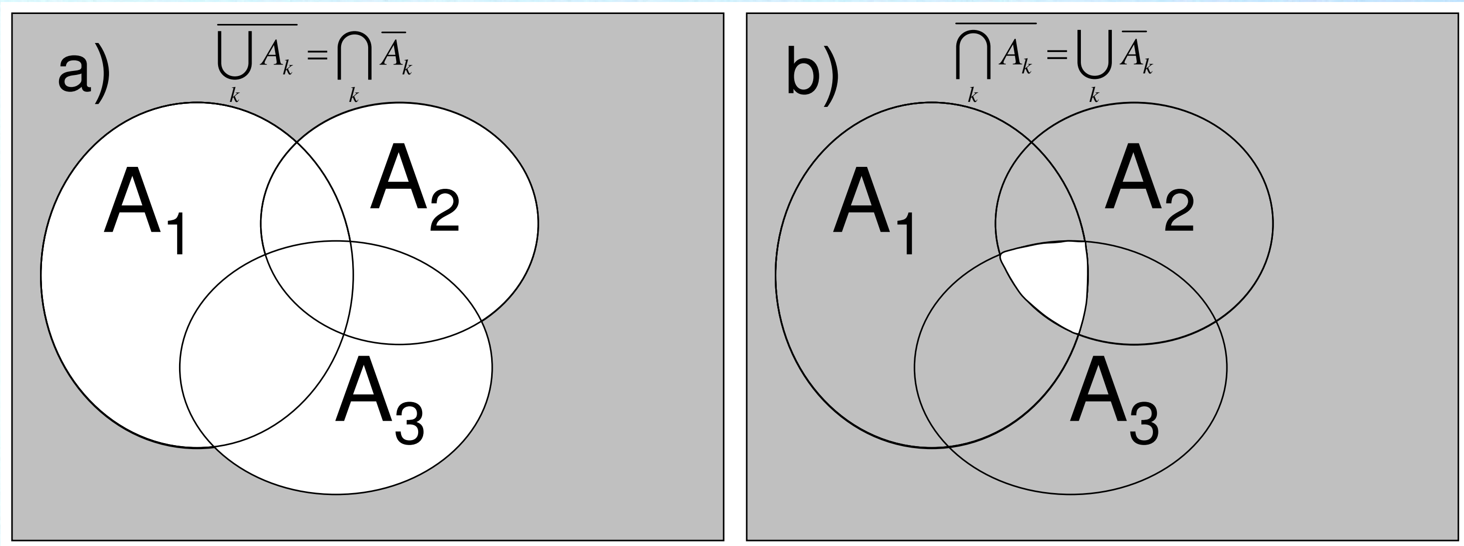
Tehted sündmustega:



Tehted sündmustega:

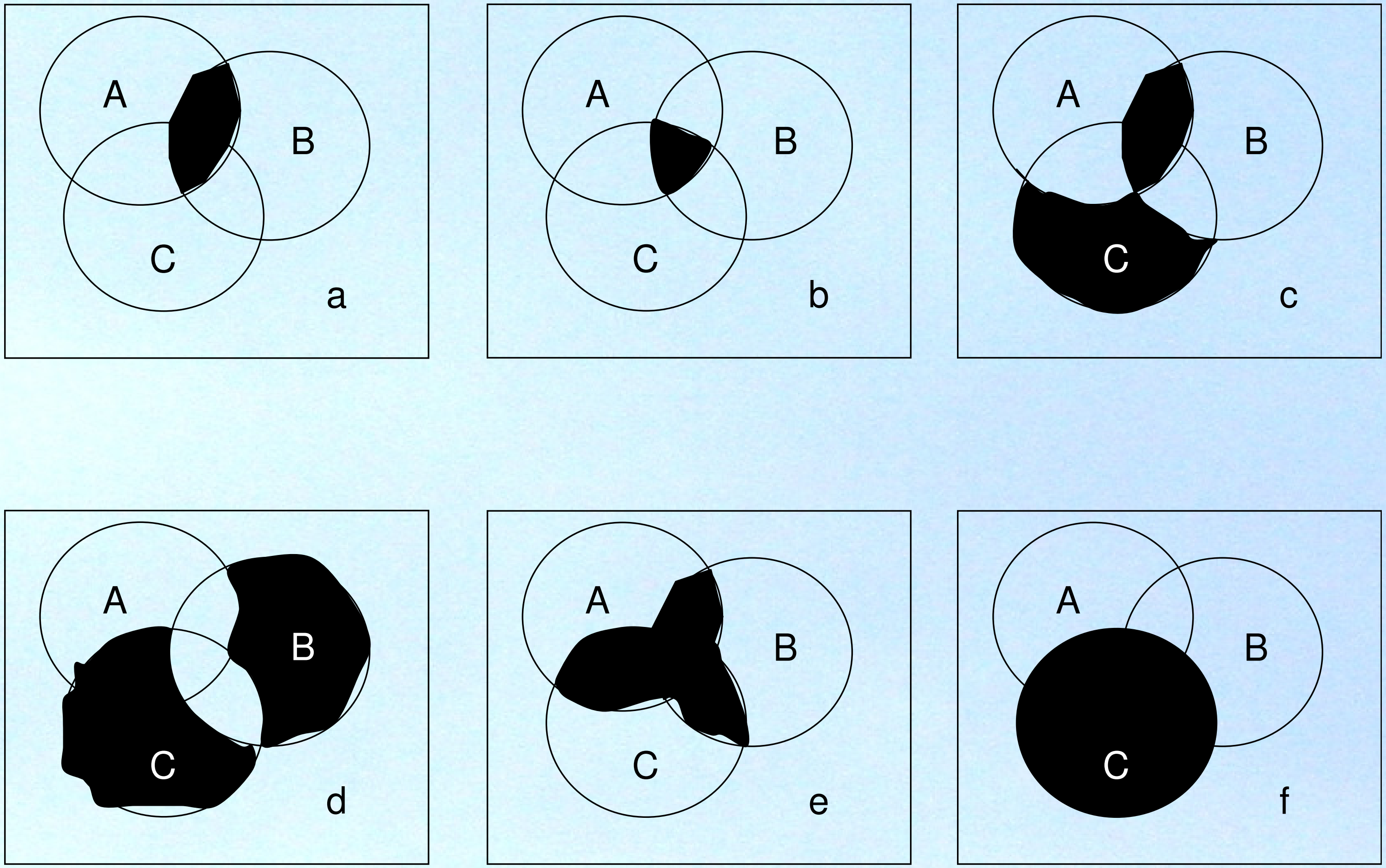
duaalsusseosed ehk Morgani seadused:

$$\overline{\bigcup_k A_k} = \bigcap_k \overline{A_k}, \qquad \overline{\bigcap_k A_k} = \bigcup_k \overline{A_k}.$$



Ülesanded:

Esitada Venni diagrammil mustaks värvitud ala HT (hulgateooria) tehete abil.



Ülesanded:

Uurida Venni diagrammi järgi, et kui

$$A \cap B \subset \overline{C}$$

ja

$$A \cup C \subset B$$

järelikult

$$A \cap C = ?$$

Ülesanded:

Münti visatakse 3 korda. Kirjuta välja variandid, mille puhul:

A. kulli visatakse täpselt 2 korda
B. kulli visatakse vähemalt 2 korda
C. kull ei tulnud enne, kui tuli kiri*
D. esimesena visati kull

leia järgmised sündmused:

$$\overline{A}$$

$$A \cup (C \cap D)$$

$$A \cap \overline{D}$$

* **s.t. kiri tuli ka.**

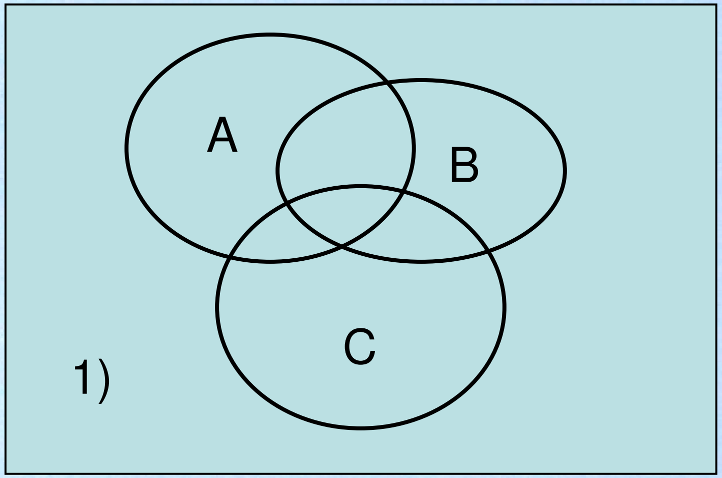
Ülesanded:

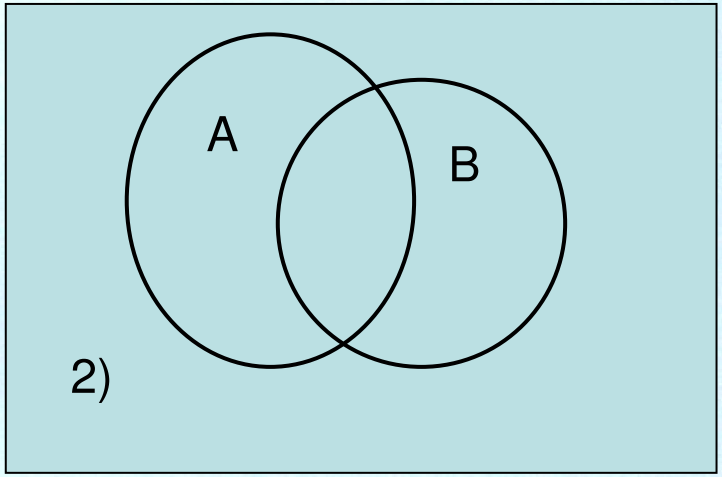
Joonista Venni diagrammid järgmistele tehetele:

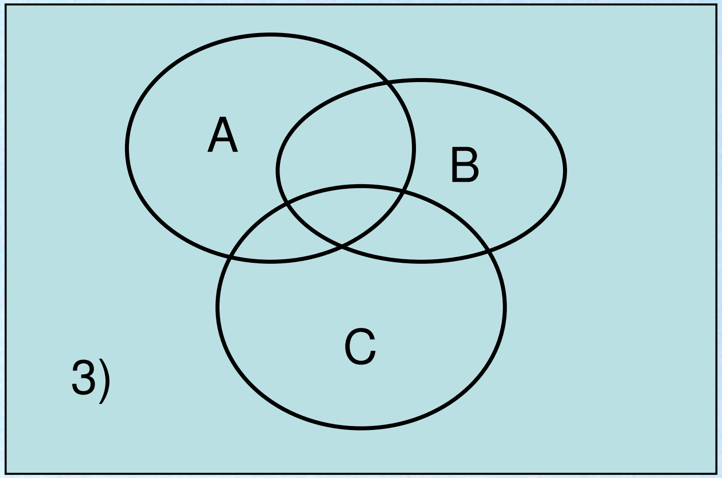
1) $(A \setminus B) \cup C$

2) $(A \cap B) \cup (A \setminus B)$

3) $(A \setminus C \setminus B) \cap (A \cap C)$







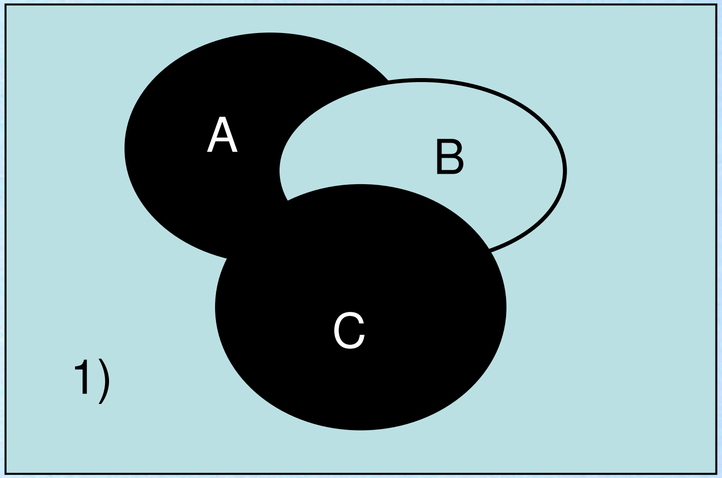
Ülesanded:

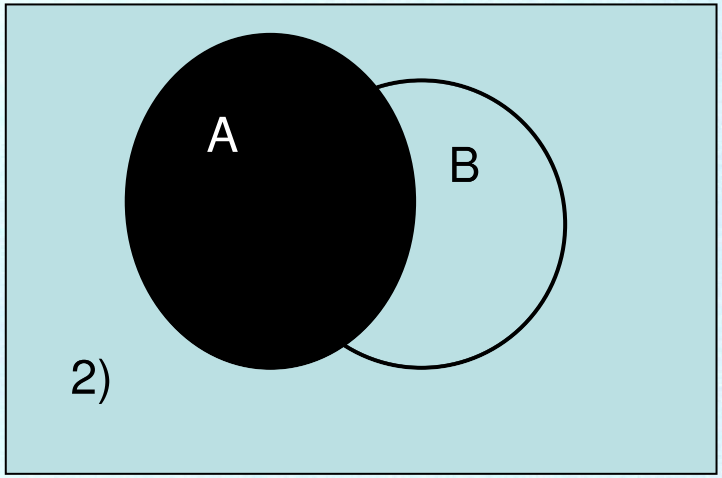
Joonista Venni diagrammid järgmistele tehetele:

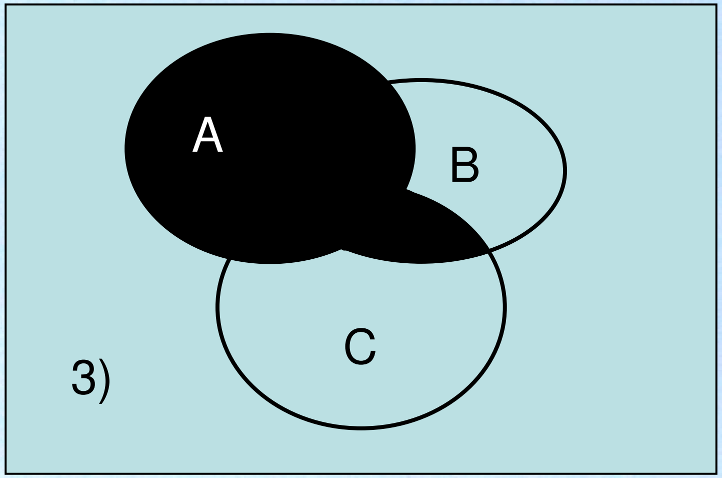
1) $(A \setminus B) \cup C$

2) $(A \cap B) \cup (A \setminus B)$

3) $(A \cup B) \cap (A \cup C)$







Kodune töö

Enam-vähem igal nädalal on Moodles kodune test. Testid annavad punkte eksamihindesse ning eksamile pääsemise eelduseks on testidest summaarselt vähemalt 50% punktide saamine.

Testide küsimused "enesekontrolliks" on harjutamiseks, nende eest punkte ei saa.

Selle nädala test sulgub 19.02.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

Juhuslikes

Sündmuse tõenäosus:

Sündmuste tõenäosuse klassikaline definitsioon – Sündmuse *A* tõenäosuseks *P*(*A*), nimetatakse sündmuse toimumiseks soodsate juhuste arvu *m*(*A*) suhet kõigi võimalike juhuste arvusse *n*, kus juhused moodustavad elementaarsündmuste süsteemi.

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

Näide 1.5. Kausis on 4 kollast ja 5 punast ploomi. Kausist võetakse juhuslikult 1 ploom, kui suur on tõenäosus, et see ploom on kollane? Kausis on kokku 4 + 5 = 9 ploomi, seega ploomivalikus on 9 erinevat võimalust, kollaste ploomide ehk soodsate juhuste arv on 4, seega kausist juhuslikult kollase ploomi võtmise tõenäosus on *P*(*A*) = 4 / 9 = 0,44.

Sündmuse tõenäosus:

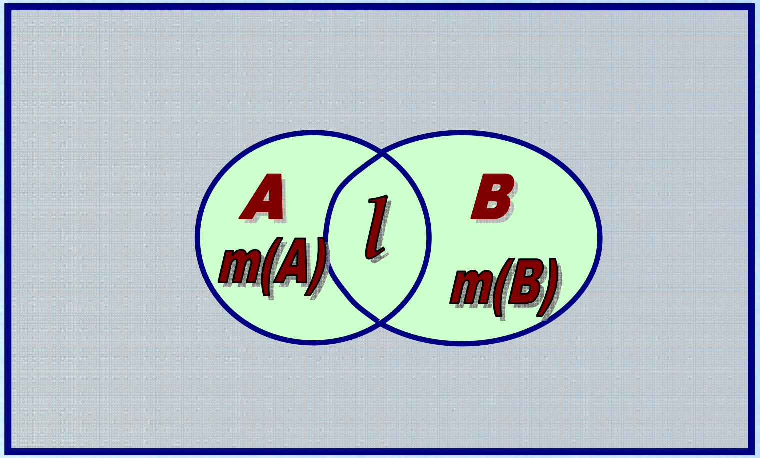
Sageli pole siiski tõenäosuse klassikalist definitsiooni võimalik rakendada ja tuleb leida sündmuse suhteline sagedus

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

Piirjuhul läheb sündmuse sagedus sündmuse tõenäosuseks:

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

Sündmuse tõenäosus:



1.4.1. Tõenäosuste liitmisreegel

Kahe sündmuse A ja B korral kehtib võrdus

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,.$$

Tõestus:

Lähtudes joonisest 1.4,

$$P(A) = \frac{m(A)}{n},\; P(B) = \frac{m(B)}{n},\; P(A \cap B) = \frac{l}{n}$$

ja

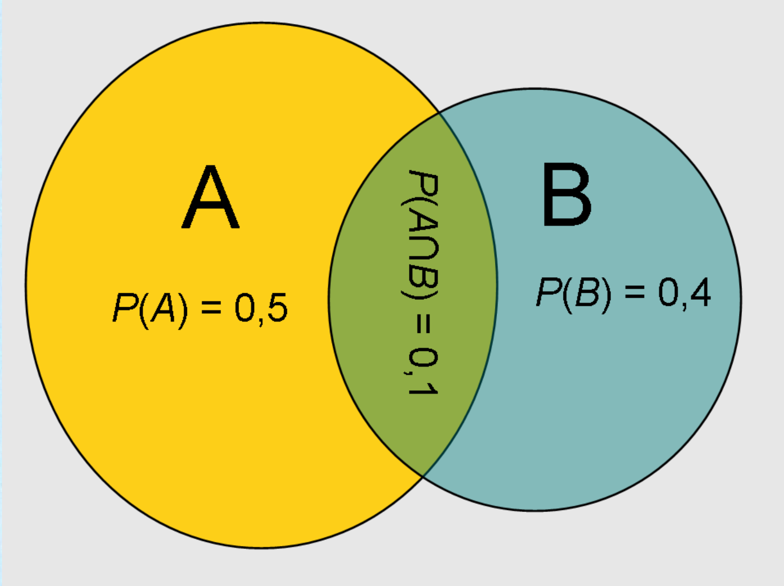
$$m(A \cup B) = m(A) + m(B) - l\,,$$

(liige „$-l$“ seetõttu, et A ja B ühisosa loeti 2 korda), siis

$$P(A \cup B) = \frac{m(A) + m(B) - l}{n} = \frac{m(A)}{n} + \frac{m(B)}{n} - \frac{l}{n} = P(A) + P(B) - P(A \cap B)\,.$$

Sündmuse tõenäosus:

Tõenäosuste korrutamisreegel



Kahe juhusliku sündmuse A ja B korral kehtivad võrdused

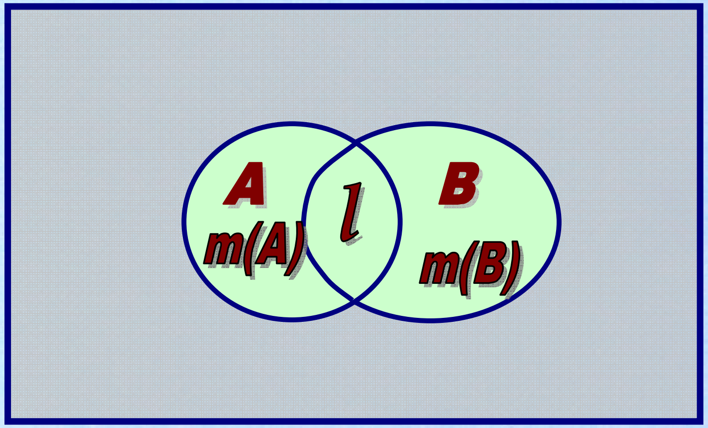
$$P(A \cap B) = P(A)P(B|A),$$

$$P(A \cap B) = P(B)P(A|B),$$

kus $P(A|B)$ on sündmuse A tinglik tõenäosus sündmuse B suhtes. Ühe sündmuse A tõenäosust, mis on leitud tingimusel, et teine sündmus B toimus, nimetatakse sündmuse A tinglikuks tõenäosuseks sündmuse B suhtes ja tähistatakse $P(A|B)$.

Tinglik tõenäosus:

Lähtume joonisest 1.4:



$$P(A) = \frac{m(A)}{n}, \quad P(B) = \frac{m(B)}{n}, \quad P(A \cap B) = \frac{l}{n}.$$

Tingliku tõenäosuse mõiste järgi

$$P(B \mid A) = \frac{l}{m(A)}, \quad P(A \mid B) = \frac{l}{m(B)}$$

järelikult

$$P(A \cap B) = \frac{l}{n} = \frac{m(A)}{n} \cdot \frac{l}{m(A)} = P(A)P(B \mid A)$$

Analoogiliselt

$$P(A \cap B) = \frac{m(B)}{n} \cdot \frac{l}{m(B)} = P(B)P(A \mid B)$$

Sõltumatute sündmuste hulkade korral ei sõltu sündmuste hulga A tõenäosus sellest, kas on toimunud sündmuste hulk C, ning vastupidi. Seega on $P(A \mid C) = P(A)$.
See on ka tingimuseks sündmuste sõltumatuse kohta.

Tinglik tõenäosus:

Näide 1.6. Olgu Ω kõik TT kursusele registreerunud, *A* on noormehed ja *B* on füüsikud.

17.01.13 seisuga oli registreerunuid kokku 95, noormehi neist 71. Füüsikuid oli 28. Füüsikutest noormehi oli 23.

LEIA:

1. $P(A)$; 2. $P(B)$;

3. $P(A \cap B)$; 4. $P(A \cup B)$;

5. $P(B \mid A)$; 6. $P(A \mid B)$.

7. Kontrolli, kas hulgad *A* ja *B* on omavahel sõltumatud.

8. Kontrolli, et kehtivad võrdused

$$P(A \bigcap B) = P(A)P(B \mid A) = P(B)P(A \mid B).$$

Tõenäosus sündmuste ruumis Ω:

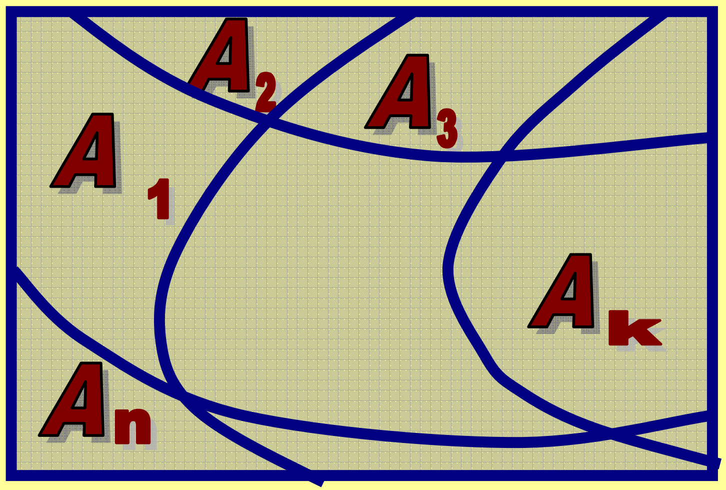
Defineerime esmalt sündmuste täissüsteemi $\{A_1,\ A_2,\ldots,\ A_n\}$.

(1) sündmused välistavad vastastikku üksteist

$$A_i \cap A_k = \Phi \quad (i \neq k).$$

(2) Sündmused moodustavad täissüsteemi

$$\bigcup_{k=1}^{n} A_k = \Omega$$



Kui üksteist välistavad sündmused $\{A_1,\ A_2,\ldots,\ A_n\}$ moodustavad täissüsteemi ruumis **Ω**, siis kehtib

$$\sum_{i=1}^{n} P(A_i) = 1$$

Täissüsteem tähendab, et igal katsel leiab aset mingi sündmus (üks ja parajasti üks sündmus). Välistatud on olukord, kus katses üldse midagi ei juhtu.

Ülesanded:

Ülesanne 1.4. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 2/3$ ja $P(B) = 1/6$ ning $P(A \cap B) = 1/9$. Leida tõenäosus $P(A \cup B)$?

Ülesanne 1.5. Olgu E ja F kaks sündmuste hulka, mille kohta on teada, et neist vähemalt ühe toimumise tõenäosus on ¾. Mis on tõenäosus, et ei E ega ka F ei toimu?

Ülesanne 1.6. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 0,3$ ja $P(B) = 0,4$ ning $P(A \cap B) = 0,2$. Leida tõenäosus $P(\overline{A} \cap B)$?

Ülesanded:

Ülesanne 1.7. Vaatame sündmuseid A, B ja C, mis võivad olla mingi eksperimendi tulemused. Kontrolli, kas tõenäosus, et ainult sündmus A toimub (B ja C ei toimu) on arvutatav valemiga $P(A) = P(A \cup B \cup C) - P(B) - P(C) + P(B \cap C)$.

Ülesanne 1.8. Olgu A ja B kaks sündmuste hulka, mille tõenäosused on $P(A) = 0{,}4$ ja $P(B) = 0{,}5$ ning $P(A \cap B) = 0{,}1$. Leida tõenäosus, et juhuslik sündmus kuuluks hulka A või B, kuid mitte mõlemasse korraga.

Ülesanne 1.9. Ülesanne 1.8. algandmete põhjal kontrolli, kas sündmused A ja B on omavahel sõltumatud.

Bayesi valem:

Tingtõenäosuse definitsioon oli

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Siin $P(A|B)$ on Tõenäosus A saamiseks tingimusel, et sündmus B on toimunud. Selles kontekstis nimetatakse sündmust B ka hüpoteesiks.

Kasutame korrutamisvalemit:

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

Siit saab

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Bayesi valem:

Korrutamisvalem on:

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Esitame täistõenäosuse *B* jaoks:

$$P(B) = P(A \cap B) \qquad + \; P(\overline{A} \cap B) =$$

$$= P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})$$

Eelneva põhjal saame <u>Bayesi valemi</u>:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}$$

Ülesanded:

Ülesanne 1.10. Leida tõenäosus, et klassis, kus on *n* tudengit, ei esine korduvaid sünnipäevi?

Enne arvutamist teha kõhutunde pealt pakkumine, et kui suur peaks olema *n*, et see tõenäosus oleks suurem kui 0,5, s.t. et tõenäosus, et klassis on korduv sünnipäev oleks suurem, kui et ei ole korduvat sünnipäeva.

Oletame, et tudengeid on kaks. Esimese tudengi sünnipäeval pole millegagi korduda. Tõenäosus $P(A_1A_2)$, et teise tudengi sünnipäev ei kattuks esimese tudengi sünnipäevaga on:

$$P(A_1A_2) = 1 - \frac{1}{365}$$

Ülesanded:

$$P(A_1A_2) = 1 - \frac{1}{365}$$

$$P(\mathrm{A}_3 \mid A_1A_2) = 1 - \frac{2}{365}$$

$$P(A_1A_2A_3) = P(A_3 \cap A_1A_2) = P(A_3 \mid A_1A_2) \cdot P(A_1A_2) =$$

$$= (1 - \frac{2}{365}) \cdot (1 - \frac{1}{365}) \approx 0{,}9917$$

$$P(A_n \mid A_{n-1}) = 1 - \frac{n-1}{365}$$

$$P(A_n) = P(A_n \cap A_{n-1}) = P(A_n \mid A_{n-1}) \cdot P(A_{n-1}) =$$

$$= (1 - \frac{n-1}{365}) \cdot P(A_{n-1})$$

Ülesanded:

$$P(A_n) = P(A_n \cap A_{n-1}) = P(A_n \mid A_{n-1}) \cdot P(A_{n-1}) =$$

$$= (1 - \frac{n-1}{365}) \cdot P(A_{n-1})$$

Rakendades saadud valemit järjest alates n = 4, saame kergesti üldvalemi:

$$P(A_n) = P(A_n \cap A_{n-1}) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4) \cdot \ldots \cdot P(A_n) =$$

$$= (1 - \frac{1}{365}) \cdot (1 - \frac{2}{365}) \cdot (1 - \frac{3}{365}) \cdot (1 - \frac{4}{365}) \cdot \ldots \cdot (1 - \frac{n-1}{365})$$

Vastav joonis demos „korduvad_synnip2evad.mcd“

Ülesanded:

Ülesanne 1.11. Täringut visatakse 2 korda. *A* on „silmade summa on 4“, *B* on „vähemalt ühe täringu silmade arv on 3“. Kas *A* ja *B* on sõltumatud sündmused?

Ülesanne 1.12. Kaardipakist (52 kaarti) võetakse kaks kaarti. Olgu S1, et esimene kaart on poti mastist. Olgu S2, et teine kaart on potimastist. Arvuta:

a) $P(S1);\ P(S2 \mid S1);\ P(S2 \mid \overline{S1})$

b) Mis on tõenäosus, et sõltumata esimese kaardi mastist on teine kaart potimast;

c) Kas S1 ja S2 on sõltumatud sündmused?

Ülesanded:

Ülesanne 1.13. Sinul diagnoositakse üks vähelevinud haigus. Sa tead, et selle haiguse saamise tõenäosus on ainult 1%. Tähistame, et D tähendab, et „sul on see haigus“ ja T tähendab, et „test ütleb, et sul on see haigus“. On teada, et see test pole täiuslik, ehk $P(T \mid D) = 0{,}98$ ning $P(\overline{T} \mid \overline{D}) = 0{,}95$. Testi tulemus oli positiivne, kuid mis on tõenäosus, et sul tõesti on see haigus?

Paneme HT valemitena kirja olemasoleva info:

$P(D) = 0{,}01$, järelikult $P(\overline{D}) = 0{,}99$

$P(T \mid D) = 0{,}98$

$P(\overline{T} \mid \overline{D}) = 0{,}95$, järelikult $P(T \mid \overline{D}) = 0{,}05$

Küsitakse $P(D \mid T) = ?$

Ülesanded:

Paneme HT valemitena kirja olemasoleva info:

$P(D)=0{,}01$, järelikult $P(\overline{D})=0{,}99$

$P(T \mid D)=0{,}98$

$P(\overline{T} \mid \overline{D})=0{,}95$, järelikult $P(T \mid \overline{D})=0{,}05$

Küsitakse $P(D \mid T)=?$

Leiame kõigepealt tõenäosuse, et test annab positiivse tulemuse:

$$P(T)=P(T\cap D)+P(T\cap \overline{D})=P(T \mid D)\cdot P(D)+P(T \mid \overline{D})\cdot P(\overline{D})=$$

$$=0{,}98\cdot 0{,}01+0{,}05\cdot 0{,}99=0{,}0098+0{,}0495=0{,}0593$$

Kasutame nüüd tingliku tõenäosuse arvutamise reeglit:

$$P(D \mid T)=\frac{P(D\cap T)}{P(T)}=\frac{0{,}0098}{0{,}0593}=0{,}165$$

Ülesanded:

Ülesanne 1.14. Arendame nüüd eelmist ülesannet edasi. Testi järgi oled sa haige. Et kontrollida, et tõesti oled sa haige, teed sa veel ühe testi. See osutus ka positiivseks. Mis on nüüd tõenäosus, et sa tõesti oledki haige?

Lahenda ise. Mis on erinevus võrreldes alltoodud eelmise ülesande tingimustega?

Paneme HT valemitena kirja olemasoleva info:

$P(D) = 0{,}01$, järelikult $P(\overline{D}) = 0{,}99$

$P(T \mid D) = 0{,}98$

$P(\overline{T} \mid \overline{D}) = 0{,}95$, järelikult $P(T \mid \overline{D}) = 0{,}05$

Küsitakse $P(D \mid T) = ?$

Ülesanded:

Ülesanne 1.15. Seadmel on ülekuumenemise eest hoiatav lamp, mis peaks süttima, kui seade kuumeneb üle. Tähistame *W*-ga sündmust, et „hoiatustuli süttib“ ning *F*-ga, et „seade on ülekuumenenud“. Seadme ülekuumenemise tõenäosus on 0,1; tõenäosus, et kui seade on tõesti üle kuumenenud, siis süttib hoiatuslamp, on 0,99. Tõenäosus, et lamp süttib, kui seade pole ülekuumenenud, on 0,02. a) Leia tõenäosus, et hoiatustuli on süttinud; b) Leia tingimuslik tõenäosus, et kui lamp süttib, et siis on seade üle kuumenenud.

Lahendus: Lahenda ise (osa vastused on antud)

Lambi süttimise tõenäosus on 11,7%

Lambi süttides on seade ülekuumenenud 84,6% juhtumitest

Kodune töö

Enam-vähem igal nädalal on Moodles kodune test. Testid annavad punkte eksamihindesse ning eksamile pääsemise eelduseks on testidest summaarselt vähemalt 50% punktide saamine.

Testide küsimused "enesekontrolliks" on harjutamiseks, nende eest punkte ei saa.

Selle nädala test sulgub 26.02.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

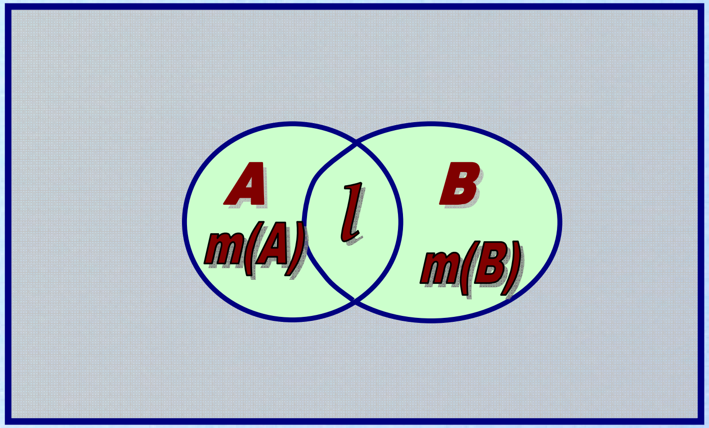
Tagasiside

Tinglik tõenäosus:

Sõltumatute sündmuste hulkade korral ei sõltu sündmuste hulga A tõenäosus sellest, kas on toimunud sündmuste hulk C, ning vastupidi. Seega on $P(A \mid C) = P(A)$.

See on ka tingimuseks sündmuste sõltumatuse kohta.

Hulgad on omavahel sõltumatud parajasti siis, kui ühte hulka kuulumise teadasaamine ei mõjuta tõenäosust teise hulka kuuluda.



Kombinatoorika:

Permutatsioonid on n elemendilise hulga elementidest moodustatud n-elemendilised järjestatud osahulgad.

Permutatsioonide arv leitakse valemiga $P_n = n!$

Kombinatsioonid *n*-elemendist *k*-kaupa on n-elemendilise hulga k-elemendilised osahulgad.

Arvutusvalem on: $C_n^k = \dfrac{n!}{k!(n-k)!}$.

Variatsioonid n elemendist k kaupa on n-elemendilise hulga k-elemendilised järjestatud osahulgad.

Arvutusvalem on: $V_n^m = \dfrac{n!}{(n-m)!}$

Kombinatoorika:

<u>**Liitmise reegel**</u> **- kui mingi elemendi A saab valida *n* erineval viisil ja elemendi B saab valida *m* erineval viisil, siis elemendi "kas A või B" saab valida**

***m + n* erineval viisil.**

<u>**Korrutamise reegel**</u> **- kui mingi elemendi A saab valida *n* erineval viisl ja elemendi B saab valida *m* erineval viisil, siis elementide paari "A ja B" saab valida**

***m • n* erineval viisil.**

<u>**NB! A ja B valikud peavad olema sõltumatud!!!**</u>

Näide 2.5. Martil on kuus "Jaguari", üheksa "Volvot" ja kolm päevinäinud "Moskvitši". Töölesõiduks on tal valida

$6 + 9 + 3 = 18$ auto vahel;

Näide 2.6. Jürkal on kooliminekuks võimalik valida nelja ülikonna, kolme jope ja viie paari kingade vahel. Jürkal on valikuvõimalusi kokku

$4 \cdot 3 \cdot 5 = 60.$

Näide 2.7. Maril on kooliminekuks võimalik valida nelja kleidi, kolme mütsi ja viie paari kingade vahel. Samas iga müts ei sobi iga kleidiga ning iga kleit iga kingaga. Seetõttu on valikuvõimalusi oluliselt vähem kui 60.

Kombinatoorika:

Ülesanne 2.1. Klassis on 6 poissi. Kui mitmel viisil saab neid kehalise kasvatuse õpetaja tunni algul rivvi panna? Kas ta jõuab kõik järjestused 45 minutiga läbi proovida, kui iga ümberjärjestuse peale kulub 5 sekundit?

Ülesanne 2.2. Klassi kokkutulekule saabus 32 vana klassikaaslast ning iga tulija kallistas varem kohalejõudnut. Mitu kallistust tehti kokku?

Ülesanne 2.3. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on 2 õpiku võtmiseks?

Ülesanne 2.4. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on 2 matemaatikaõpiku võtmiseks?

Ülesanne 2.5. Riiulis on 5 matemaatika ja 7 füüsika õpikut. Mitu võimalust on ühe matemaatikaõpiku ja ühe füüsikaõpiku võtmiseks?

Kombinatoorika:

<u>Ülesanne 2.6.</u> Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on mõlemad viiekroonised?

<u>Ülesanne 2.7.</u> Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on need erineva nominaalväärtusega?

<u>Ülesanne 2.8.</u> Martil on taskus viis viiekroonist ja neli kümnekroonist rahatähte. Kui suur on tõenäosus, et kahe kupüüri juhuslikul võtmisel on need sama nominaalväärtusega?

<u>Ülesanne 2.9.</u> Lumivalgekesel on 7 sõpra. Mitu erinevat seltskonda saab Lumivalgeke endale külla kutsuda?

<u>Ülesanne 2.10.</u> Viking Lotol on 48 numbrit (kuuli) ja võidud on 3, 4, 5 ja 6 õiget numbrit (6 on ‚Jackpot'). Tuletada vastavad võidutõenäosused:

KOMBINATOORIKA VALEMID

Permutatsioonid *P*

123 213 312
132 231 321

$$P_n = n!$$

(**korduvustega** ***PK***)

122 212 221

$$PK_{n(\alpha,\beta,\dots,\lambda)} = \frac{P_n}{P_a \cdot P_b \cdot \dots \cdot P_l} = \frac{n!}{a! \cdot b! \cdot \dots \cdot l!}$$

Variatsioonid *V*

12 21 31 41
13 23 32 42
14 24 34 43

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

(**korduvustega** ***W***)

aaa aab aba abb
baa bab bba bbb

$$W_n^m = n^m$$

Kombinatsioonid *C*

12 13 14
23 24 34

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

(**korduvustega** ***G***)

aaa aab abb bbb

$$\Gamma_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m! \cdot (n-1)!}$$

Ülesanne 2.11. Kui mitu erinevat kolmest saiast koosnevat lõunaoodet saab kohvikus tellida, kui menüü sisaldab 11 erinevat nimetust saiasid?

Ülesanne 2.12. Kui mitmel erineval viisil saab nimes TEELE tähti selliselt ümber paigutada, et kolm tähte E ei satuks kõrvuti?

Ülesanne 2.13. Auto registreerimisnumber koosneb kolmekohalisest arvust (ka arv 031 loetakse kolmekohaliseks) ja kolmetähelisest sõnast. Mitu autot on võimalik registreerida, kui kasutatakse 25-tähelist tähestikku?

<u>Ülesanne 2.14.</u> Ühiselamu toas elab 3 üliõpilast. Neil on kokku 4 erinevat tassi, 5 erinevat alustaldrikut ning 6 erinevat teelusikat. Kui mitmel viisil saavad nad katta laua hommikuseks kohvijoomiseks, kui nõude paigutus laual pole oluline?

<u>Ülesanne 2.15.</u> Kui mitmel viisil saab jagada 36-kaardilise kaardipaki kaheks osaks nii, et mõlemad osad sisaldaks 2 ässa?

<u>Ülesanne 2.16.</u> Viis sõpra otsustavad teha lauajalgpalli turniiri, nii et mängitakse paarides ja kõik võimalikud paarid mängiksid kõigi võimalike paaridega korra läbi. Kui oleks neli mängijat, siis oleks kolm mängu (1, 2 – 3, 4), (1, 3 – 2, 4) ja (1, 4 – 2, 3). Mitmest mängust koosneb turniir 5-kesi mängides? Mitu mängu tuleks mängida sama süsteemi järgi, kui oleks 8 sõpra?

Kodune töö

Testide küsimused "enesekontrolliks" on harjutamiseks, nende eest punkte ei saa.

Selle nädala test sulgub 05.03.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

Tagasiside

Juhusliku suuruse jaotused

Juhuslik suurus on suurus, mis katse tulemusel omandab ühe või teise, varem mitteteadaoleva väärtuse mingist väärtuste hulgast. Viimast hulka nimetatakse juhusliku suuruse võimalike väärtuste hulgaks.

Juhuslikukuks suuruseks on näiteks:

- **täringu viskamisel saadud silmade arv,**
- **telefonikeskjaama saabuvate kutsungite arv,**
- **spordivõistlustel kaugushüppe võitja tulemus,**
- **elektripirni põlemiskestus,**
- **gaasimolekuli kiirus,**
- **välisõhu temperatuur (mingil kindlal ajahetkel) Tartus Raekoja Platsil.**

Juhusliku suuruse jaotused

Juhuslikud suurused jaotuvad kahte klassi:

- **diskreetsed**
- **pidevad**

Diskreetsed suurused saavad omada ainult ettemääratud väärtuseid, näiteks täringuvise silmasid 1 – 6, mündivise kulli ja kirja ning suvalise digitaalmõõtevahendi näit (digitaalne kaal resolutsiooniga 1 gramm võimalikud mõõtetulemused grammides on täisarvud).

Juhuslikku suurust nimetatakse pidevaks, kui tema võimalike väärtuste hulk on arvtelje (lõplik või lõpmatu) vahemik. Füüsikalised suurused ise on üldiselt pidevad, näiteks temperatuur, mass, takistus jne.

- **Diskreetimine teisendab pidevaid andmeid diskreetseteks.**
- **Interpoleerimine teisendab diskreetseid andmeid pidevateks.**

Juhusliku suuruse jaotused

Juhuslikke suurusi iseloomustatakse jaotusseadustega. Öeldakse, et jaotusseadus ehk jaotus ehk tõenäosus-jaotus on antud, kui on määratud:

1) **juhusliku suuruse võimalike väärtuste hulk,**
2) **eeskiri, mis võimaldab leida tõenäosust, et juhusliku suuruse väärtus on mingis piirkonnas.**

Jaotusseaduse enamlevinumateks ja praktilisteks vormideks on jaotustabel, jaotustihedus ja integraalne jaotusfunktsioon.

Kui juhuslikul suurusel on antud jaotusseadus, siis öeldakse, et juhuslik suurus on antud jaotusega või juhuslik suurus allub antud jaotusele.

Juhusliku suuruse jaotused

Vaatleme diskreetset juhuslikku suurust X võimalike väärtustega x_1, x_2, ..., x_n. See, et juhuslik suurus X omandab ühe neist väärtustest, näiteks $X = x_2$, on tüüpiline juhuslik sündmus mingi tõenäosusega. Tähistame tõenäosused

$$P(X = x_k) = p_k; \qquad (k = 1,\ 2,\ \ldots n),$$

mida nimetame diskreetse juhusliku suuruse X **tõenäosusfunktsiooniks**. Ilmselt kehtib

$$\sum_{k=1}^{n} p_k = 1, \qquad (3.1)$$

kuna sündmused $X = x_k$ on ainuvõimalikud, üksteist välistavad ja moodustavad täissüsteemi.

Hulka X: x_1, x_2, ..., x_n võib vaadelda ka elementaarsündmuste ruumina, ainult et klassikalist võrdtõenäosuse eeldust ei tehta.

Juhusliku suuruse jaotused

Diskreetse juhusliku suuruse jaotusseaduse lihtsaimaks esitamisvormiks on tabel, kus on loetletud võimalikud väärtused ja neile vastavad tõenäosused.

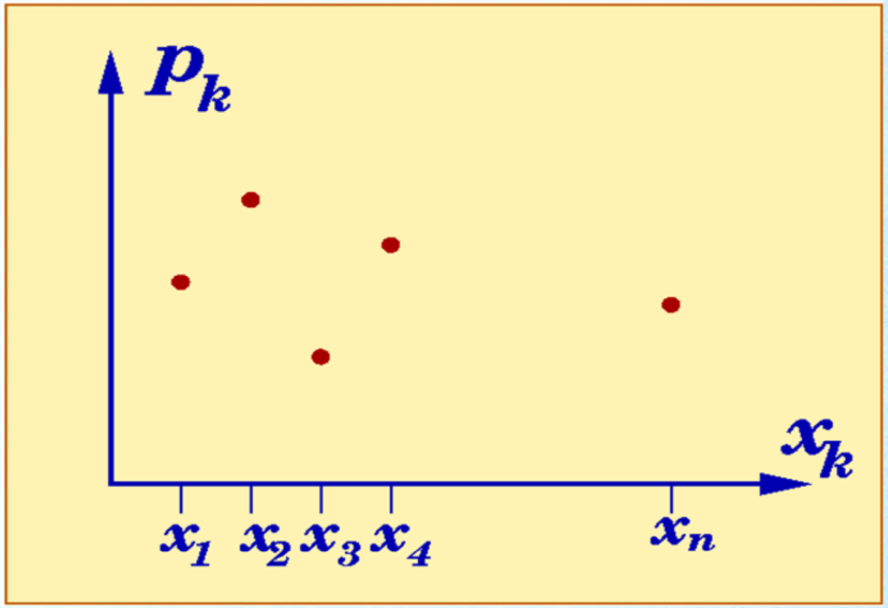
X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

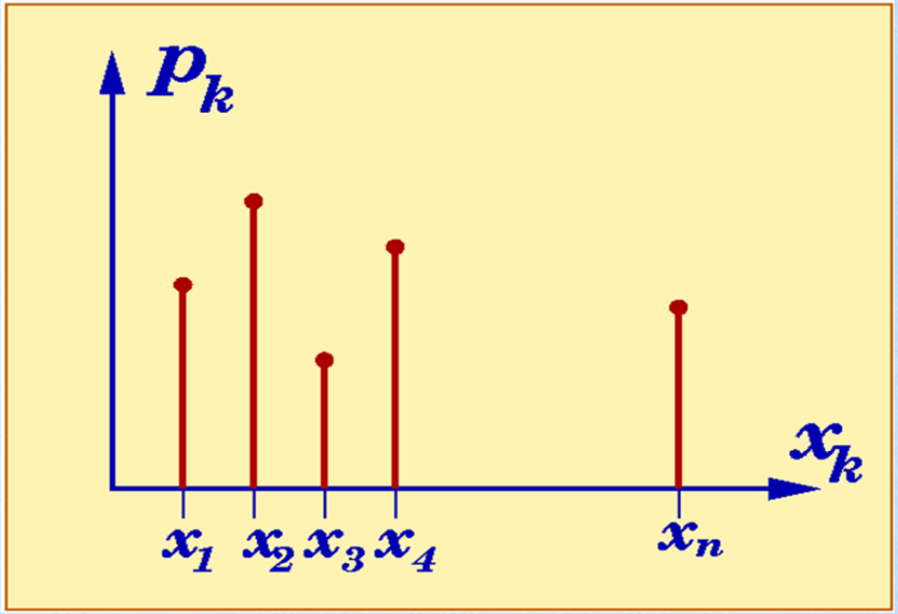
Niisugust tabelit nimetatakse juhusliku suuruse **jaotustabeliks.**

Praktikas sobib kasutada jaotustabelit, kui võimalike väärtuste arv ei ole suur. Mõnedel juhtudel saab jaotusseaduse anda ka valemi kujul

$$p_k = P(X = x_k) = g(x_k).$$

Juhusliku suuruse jaotused





Juhusliku diskreetse suuruse tõenäosusjaotus.

Tõenäosus erineb nullist ainult määratud, diskreetsetel väärtustel. Näiteks täringu viskel väärtus 2,9 ei ole määratud ja tema tõenäosus on seega null.

Juhusliku suuruse jaotused

Näide 3.1. Teelõigul on 4 valgusfoori. Igaüks neist fooridest lubab liiklusvahendi liikumise tõenäosusega p. Leida liiklusvahendi poolt peatuseta (kuni esimese peatuseni) läbitud fooride arvu jaotustabel. Oluline lisaeeldus on, et foorid vilkugu üksteisest sõltumatult!

Olgu juhuslik suurus X liiklusvahendi poolt peatuseta läbitud fooride arv. Ilmselt X võimalikud väärtused on 0 (ei läbi ühtki foori), 1, 2, 3 ja 4. Tähistame tõenäosuse, et liiklusvahend läbib konkreetse foori p-ga, ja tõenäosuse et ei läbi konkreetset foori q-ga: $q = 1 - p$.

Binoomjaotus ehk Bernoulli jaotus

Vaatleme n sõltumatut katset, kui igal katsel toimub sündmus A konstantse tõenäosusega p. Sündmuse A toimumise tõenäosus üksikkatsel on p, mittetoimumise tõenäosus aga on $q = 1 - p$. Kui teeme n katset ja saame mingis konkreetses järjestuses sündmuse A toimumise k korda, ning järelikult, vastandsündmuse $\overline{A}$ toimumise $n - k$ korda, siis on sellise tulemuse saamise tõenäosus $p^k q^{n-k}$. Kokku on erinevaid võimalusi saada erinevas järjestuses k soodsat katset võrdne kombinatsioonide arvuga n liikmest k kaupa, s.o C_n^k-ga. Seega

$$P(X = k) = P_n(k) \equiv C_n^k p^k q^{n-k} .$$

Binoomjaotus ehk Bernoulli jaotus

Ülesanne 3.1. Testis on 10 küsimust, igal küsimusel on 4 valikvastust, millest üks on õige. Kui vastata huupi, siis tulemus allub Bernoulli jaotusele, kus õige vastuse tõenäosus on $P(1) = ¼$. Mis on tõenäosus, et testis saame k õiget vastust? Mis on tõenäosus, et testis saame huupi vastates 5 või 6 õiget vastust?

Ülesanne 3.4. Pood võtab müüki 1000 lampi. Tõenäosus, et juhuslikult valitud lamp on katki, on
$p = 0{,}1$ %. Tähistame X-ga katkiste lampide arvu. Leida tõenäosus, et

a) kõik lambid on terved;

b) üks lamp on katki;

c) katki on rohkem kui 2 lampi.

Geomeetriline jaotus

Olukord nagu eelmisel Bernoulli katsete juhul, et sündmuse A toimumise tõenäosus üksikkatsel on p, mittetoimumise tõenäosus aga on $q = 1 - p$. Nüüd on katseseeria lõpmatu, ja küsitakse, milline on tõenäosus, et esimesed $x - 1$ korda tulevad $\overline{A}$, s.o. sündmust ei toimu, ja soodne sündmus A leiab aset x-ndal katsel.

Geomeetrilise jaotuse tõenäosusjaotus on:

$$P(X = k) = q^{k-1} p$$

Geomeetriline jaotus

Ülesanne 3.5. Langevarju hüpetel on alati on tõenäosus, et langevari ei avane. Oletame, et see tõenäosus on $p = 0{,}1$ %. Arvutada tõenäosus, et a) langevari ei avane 5-ndal hüppel; b) langevari ei avane 1000-ndal hüppel.

Ülesanne 3.6. Oletame, et metsas lõkke süütamisel on tõenäosus, et tule saab süüdatud ühe tikuga p = 25%. Arvutada tõenäosus, et a) lõke sütib esimesest tikust; b) lõke sütib viiendast tikust; c) lõke sütib kahekümnendast tikust.

Ülesanne 3.7. Petsi juures on iga nädal raju pidu, mille käigus käib ta iga kord ka autoga tanklas varusid täiendamas. Oletame, et on 5 % tõenäosus, et jäädakse politseile vahele. Arvutada tõenäosus, et a) ta jääb politseile vahele esimesel peol; b) ta jääb vahele kuu aja pärast 5. peol; c) ta jääb vahele täpselt aasta pärast 52. peol? Teeme eelduse, et pärast politseile vahele jäämist tabab Petsi meelemuutus ning rajud peod lõpevad.

Poissoni jaotus

Vaatleme veel kord katseseeriat, milles on n katset ja sündmuse A tõenäosus igal katsel on p. Tõenäosus, et katseseerias sündmus A toimub k korda, avaldub Bernoulli valemiga. Kui n on küllalt suur ja p väga väike, siis valem praktiliseks arvutamiseks ei sobi, kuna on tegemist väga suurte ja väga väikeste arvude korrutamisega. Leiame sobivama valemi.

Eeldame, et korrutis np on konstantne ja tähistame

$$\lambda = np ,$$

täpsemini

$$\lambda = \lim_{\substack{n \to \infty \\ p \to 0}} np .$$

Poissoni jaotus

Poissoni jaotuseks nimetatakse jaotust

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad (k = 0,\ 1,\ \ldots)$$

Jaotust nimetatakse samuti väikeste arvude seaduseks ehk harva esinevate sündmuste seaduseks, sest see kirjeldab harva esinevaid nähtusi.

Poissoni jaotus annab tõenäosuse saada hästi pikas katseseeria pikkusega n soodne sündmus k-l korral, kui üksiksündmuse tõenäosus on $p = \lambda / n << 1$.

Hinnanguliselt võib binoomjaotust lähendada Poissoni jaotusega, kui $p < 0{,}1$. See aga tähendab, et fikseeritud λ korral peab võtma katseseeria pikkuse tingimusest $n \geq \lambda / p = 10\lambda$. Kui $\lambda = 10$, siis näiteks on usaldusväärne valida n = 100.

Poissoni jaotus

__Ülesanne 3.8.__ **Kõnekeskus saab minutis keskmiselt 20 kõnet. Leia tõenäosus, et veerand minuti jooksul tuleb a) 3 kõnet; b) 5 kõnet; c) 7 kõnet; d) 10 kõnet? Leia tõenäosus, et ühe sekundi jooksul tuleb 2 kõnet?**

__Ülesanne 3.9.__ **Kaubamaja tegi kindlaks, et nõudmine teatud fotoaparaadi järele allub Poissoni jaotusele. Ühe nädala jooksul ostetakse keskmiselt kaks aparaati. Kui kord nädalas uuendatakse kaubamaja laoseisu nii, et poes oleks 4 aparaati, siis kui suur on tõenäosus, et nõudmine ületab pakkumise?**

__Ülesanne 3.10.__ **Tehases on 10000 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?**

__Ülesanne.__ **Poodi tuuakse iga päev 2000 pakki piima. Iga päev leitakse keskmiselt 3 piimapakki, mis lasevad piima läbi. Kui suur on tõenäosus, et ühe päeva jooksul ei leita poest mitte ühtegi lekkivat piimapakki?**

Kodune töö

Selle nädala test sulgub 12.03.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Testi sulgumise järel avaneb “test harjutamiseks”, seda saab harjutada piiramatult kuni semestri lõpuni.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

Tagasiside

Tänane teema:

jaotustihedus ja jaotusfunktsioon, keskväärtus ja standardhälve

jaotusseadused

Elementaarsündmuseks nimetatakse juhusliku katse tulemust.

Diskreetse juhusliku suuruse esinemistõenäosus on avaldatav valemiga

$$p_k = \frac{Soodsate\ elementaarsündmuste\ hulk}{Kõigi\ elementaarsündmuste\ hulk}$$

$$0 \le p_k \le 1$$

$$\sum_k p_k \equiv 1$$

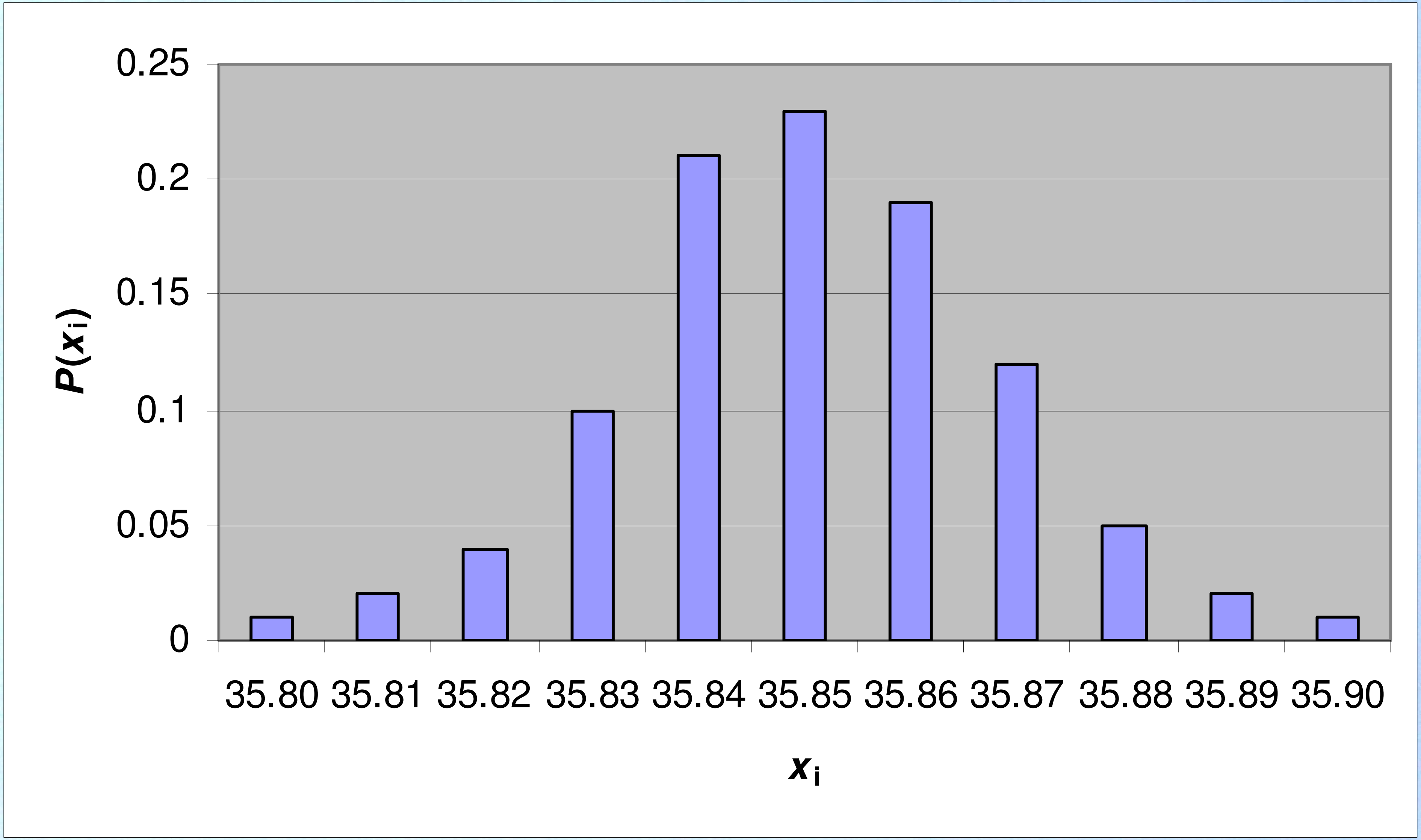
Diskreetsete suuruste <u>jaotusfunktsioon</u> ehk kumulatiivne tõenäosusjaotus on defineeritud valemiga:

$$F(k) = \sum_{i \le k} p_i \qquad 0 \le F(k) \le 1$$

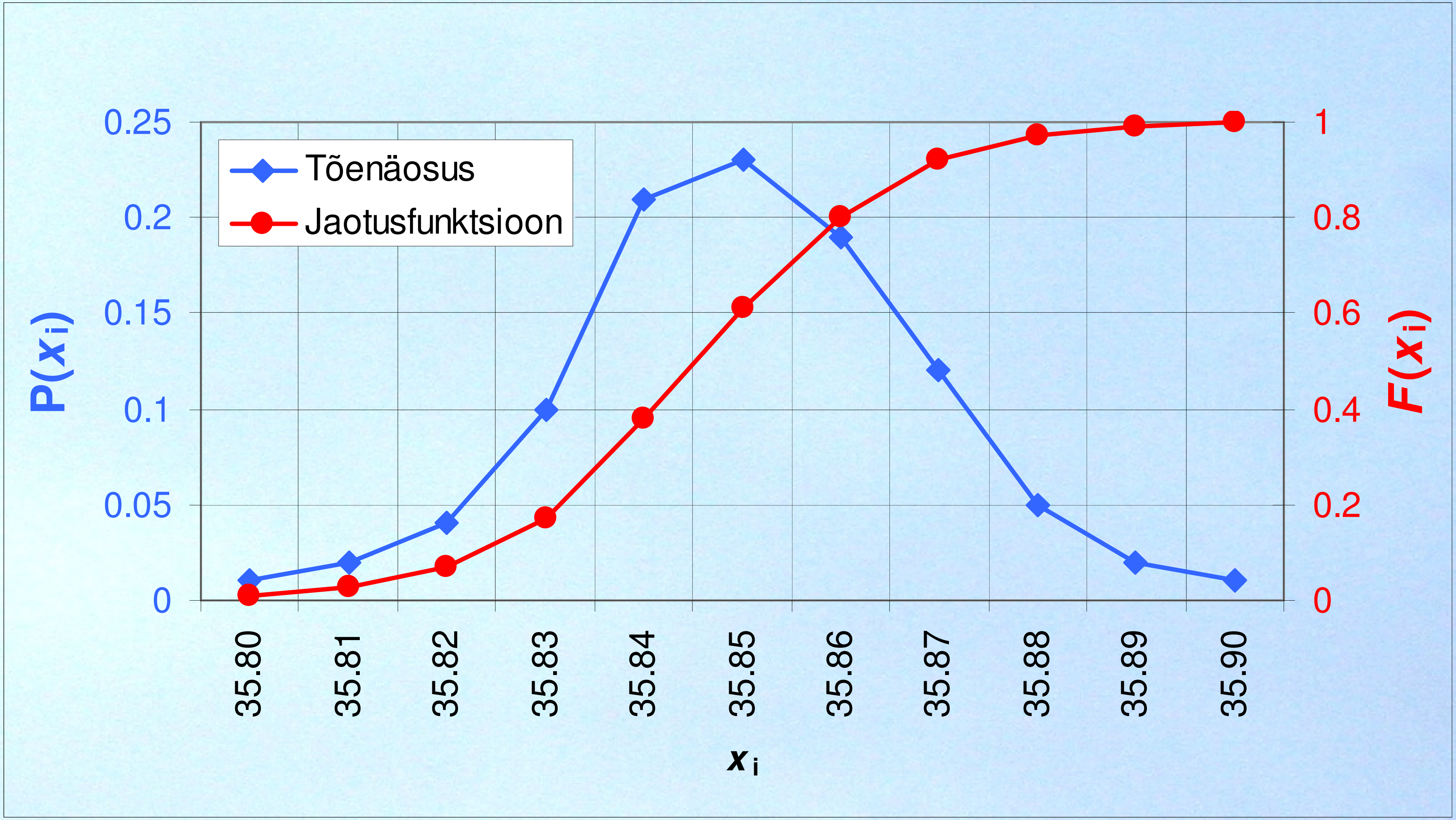
Tabel 2.1. Mõõtmisel saadud diskreetsed arvväärtused x_i, nende esinemiste arv m_i, tõenäosus $P(x_i)$ ja jaotusfunktsioon $F(x_i)$.

x_i	m_i	$P(x_i)$	$F(x_i)$
1	2	3	4
35.80	1	0.01	0.01
35.81	2	0.02	0.03
35.82	4	0.04	0.07
35.83	10	0.1	0.17
35.84	21	0.21	0.38
35.85	23	0.23	0.61
35.86	19	0.19	0.8
35.87	12	0.12	0.92
35.88	5	0.05	0.97
35.89	2	0.02	0.99
35.90	1	0.01	1

jaotusseadused



jaotusseadused



$$F(k) = \sum_{i \le k} p_i = P(X \le x_i)$$

jaotusseadused

Kuna tõenäosusjaotuses ei saa olla negatiivseid väärtuseid, siis jaotusfunktsiooni väärtused ei saa samuti väheneda:

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

Kui aga muuta x väärtust tema võimalikes piirides, muutub $F(x)$ alates 0 kuni 1, s.t. jaotusfunktsioon rahuldab võrratust $0 \leq F(x) \leq 1$.

Tõenäosus sündmuse väärtuse sattumiseks vahemikku $[x_1; x_2]$, on võrdne funktsiooni $F(x)$ väärtuste vahega:

$$P\{x_1 \leq x \leq x_2\} = F(x_2) - F(x_1)$$

$$F(k) = \sum_{i \leq k} p_i$$

jaotusseadused

$$P\{x_1 \le x \le x_2\} = F(x_2) - F(x_1)$$

Pideva suuruse korral võib väärtused x_1 ja x_2 võtta võimalikult teineteise lähedusest. Siit tuleneb, et kui võtta $x_1 \to x_2$, siis

$$\lim_{x_1 \to x_2} \left[F(x_2) - F(x_1)\right] = 0.$$

Seega saame väita, et pideva suuruse korral on mingi konkreetse väärtuse tõenäosus võrdne nulliga. Näiteks tõenäosus, et ühe mõõtesilindri maht on täpselt 1 liiter, on null. Seega on mõistlik vaadata pidevate suuruste korral hoopis tõenäosust, et sündmus satuks vahemikku [x_1; x_2].

jaotusseadused

Jaotustihedus *f*(*x*) on tuletis jaotusfunktsioonist *F*(*x*):

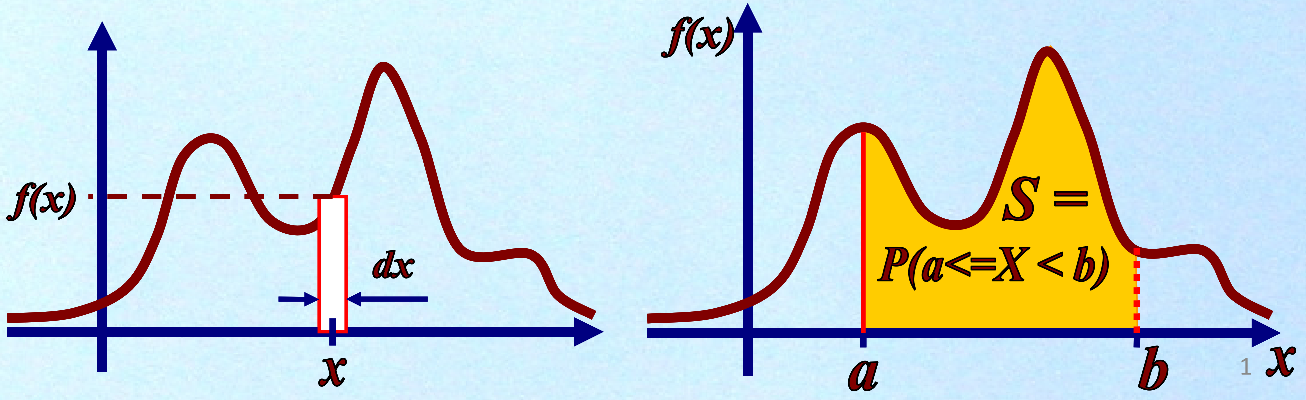
$$f(x) = \frac{dF(x)}{dx}$$

Teistpidi on jaotusfunktsioon määratud integraal jaotustihedusest:

$$F(x_0) = \int_{-\infty}^{x_0} f(x)dx$$

Kuigi matemaatiliselt on jaotustihedus ja jaotusfunktsioon üksteisest tuletatavad, on mõlemad siiski tarvilikud, sageli lihtsustab nende valemite kasutamine oluliselt paljude esmapilgul väga keeruliste ülesannete lahendamist.

jaotusseadused



Geomeetriline interpretatsioon jaotustihedusest ja jaotusfunktsioonist.

$$dp = dF(x) = f(x) \cdot dx$$

$$p(a,b) = P(a \le X \le b) = \int_a^b dP = \int_a^b f(x) \cdot dx =$$

$$= \int_{-\infty}^{b} f(x) \cdot dx - \int_{-\infty}^{a} f(x) \cdot dx = F(b) - F(a)$$

jaotusseadused kokkuvõtlikult

Jaotustihedus *f*(*x*) on: $$f(x) = \frac{dF(x)}{dx}$$

Jaotustiheduse normeerimis- ja mittenegatiivsustingimus:

$$F(\infty) = \int_{-\infty}^{\infty} f(x)dx \equiv 1 \qquad f(x) \geq 0$$

Jaotusfunktsioon *F*(*x*) on: $$F(x_0) = \int_{-\infty}^{x_0} f(x)dx$$

Vahemiku tõenäosus on: $$P\left\{x_1 \leq x \leq x_2\right\} = F(x_2) - F(x_1)$$

Ülesanne 3.11. Juhusliku suuruse X jaotusfunktsioon on

$$F(x) = x^2; \qquad 0 \le x \le 1$$

Arvutage a)vastav jaotustihedus; b) tõenäosus $P(0{,}25 \le X \le 0{,}75)$

Ülesanne 3.12. Defineeri mingi jaotustihedus $f(x)$, mis oleks vahemikus (–1; 2) mittekonstantne ning mujal võrdne nulliga. Leia selle funktsiooni jaotusfunktsioon ning vahemiku (0; 1) tõenäosus.

Ülesanne 3.13. Radioaktiivse aine pooldumist kirjeldab valem $f(t) = \frac{1}{t_0}\exp(-t/t_0)$, kus parameeter t_0 on antud aine poolestusaeg (ajaintervall, mille jooksul jääb ainet $e \approx 2{,}718$ korda vähemaks. Tseesium 137 poolestusaeg on 43 aastat. Leia radioaktiivse aine pooldumist kirjeldav tihedusfunktsioon. Palju kulub aega, et esialgsest Cs ainehulgast jääks järgi ainult 1%?

Ülesanne 3.14. Ruumi temperatuurikontrollisüsteem hoiab ruumi temperatuuri vahemikus 20 °C kuni 24 °C. Temperatuuri tsüklilise muutumise tulemuseks on arkussiinustemperatuurijaotus, keskväärtusega 22 °C. Mitu protsenti ajast on ruumi temperatuur vahemikus 21 °C kuni 23 °C? Mitu protsenti ajast on ruumi temperatuur alla 20,5 °C või üle 23,5 °C?

Juhusliku suuruse arvkarakteristikud: keskväärtus

Diskreetse juhusliku suuruse *x* keskväärtus on:

$$m = \sum_{k=1}^{n} x_k p_k$$

Pideva juhusliku suuruse *x* keskväärtus on:

$$m(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

Pideva juhusliku suuruse *x* n-nda astme x^n keskväärtus on:

$$m(x^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

1. Leia täringuviske keskväärtus.

2. Leia ühtlase jaotuse jaoks *x* ja x^2 keskväärtus.

Juhusliku suuruse arvkarakteristikud: keskväärtus

Defineerime siinkohal veel mõned keskmisi väärtuseid kirjeldavad juhuslike suuruste arvkarakteristikud:

- Mediaan on arv, millest suuremaid ja väiksemaid väärtuseid on variatsioonireas ühepalju, ehk siis jaotusfunktsiooni kaudu väljendades $F(x_{mediaan}) \equiv 0.5$.
- Mood on tunnuse kõige sagedamini esinev väärtus.
- Kaalutud keskmine on arv, mis saadakse, kui aritmeetilise keskmise arvutamisel antakse erinevatele väärtustele erinevad kaalud. Kaaluks võivad olla näiteks mõõtetulemuste standardhälbed.

Juhusliku suuruse arvkarakteristikud: keskväärtus

Näide 3.7. Oletame, et üks firma koosneb juhatajast ja 9-st töötajast. Töötajate kuupalk on 500 EUR, juhata kuupalk on 5500 EUR. Keskmine palk selles firmas on

$$m = \frac{9 \cdot 500 + 1 \cdot 5500}{10} = 1000$$

Samas mediaanpalk selles firmas on keskmisest palgast poole väiksem, 500 EUR.

Üks teine firma koosneb neljast insenerist, kuupalgaga 1200 EUR ning koristajast, kuupalgaga 200 EUR. Keskmine palk selles firmas on

$$m = \frac{4 \cdot 1200 + 1 \cdot 200}{5} = 1000$$

Samas mediaanpalk selles firmas on keskmisest palgast suurem, 1200 EUR.

Juhusliku suuruse arvkarakteristikud: dispersioon

Juhusliku suuruse dispersiooni definitsioonvalem on:

$$D[X] = m\left[(X - m[X])^2\right]$$

Diskreetse juhusliku suuruse *x* dispersioon on:

$$D(x) = \sum_i (x_i - m)^2 p_i$$

Pideva juhusliku suuruse *x* dispersioon on:

$$D(x) = \int_{-\infty}^{\infty} (x - m)^2 \; f(x)\, dx$$

Steineri valem dispersiooni arvutamiseks:

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

Dispersiooni ruutjuurt nimetatakse standardhälbeks:

$$s(x) = \sqrt{D(x)}$$

Keskväärtus ja dispersioon kokkuvõtlikult

$$m(x) = \sum_{k=1}^{n} x_k p_k$$

$$m(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$m(x^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

$$F(x_{mediaan}) \equiv 0.5$$

$$D[X] = m\left[(X - m[X])^2\right]$$

$$D(x) = \sum_i (x_i - m)^2 p_i$$

$$D(x) = \int_{-\infty}^{\infty} (x - m)^2 f(x)\, dx$$

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

$$s(x) = \sqrt{D(x)}$$

Ülesanne 3.15. Jaotusfunktsioon on defineeritud järgmiselt:

$$f(x) = \begin{cases} 5x, & 1 \le x \le 2 \\ 0, & \text{mujal} \end{cases}$$

Tee funktsiooni graafik, hinda silma järgi, kui suur võiks olla keskväärtus. Arvuta antud funktsiooni täpne keskväärtus, kanna see joonisele.

Ülesanne 3.16. Kolmnurkjaotust võib defineerida ka järgnevalt: piirkondades $x < 0$ ning $x > b$ on tema väärtus 0; vahemikus [0; b] on tema väärtus ühtlaselt kasvav ($f(x) = cx$). Arvuta koefitsient c, jaotusfunktsioon, keskväärtus ja mediaan ning standardhälve. Leia tõenäosus, et sündmus satuks piirkonda ($m \pm \sigma$); ($m \pm 2\sigma$). Joonista graafik ning kanna sellele keskväärtus ja mediaan ning viiruta piirkond ($m \pm 2\sigma$).

Ülesanne 3.17. Kolmnurkjaotuse jaotustihedus on defineeritud järgnevalt: piirkondades $x < 5$ ning $x > 8$ on tema väärtus 0; vahemikus [5; 8] on tema väärtus ühtlaselt kasvav ($f(x) = cx$). Leia koefitsient c; jaotusfunktsioon; keskväärtus, median; standardhälve; tõenäosus, et sündmus satuks piirkonda ($m(x)$, 12)?

Kodune töö

Selle nädala test sulgub 19.03.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Testi sulgumise järel avaneb “test harjutamiseks”, seda saab harjutada piiramatult kuni semestri lõpuni.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

Tagasiside

EKSAMI AJAD (viimane korraline loeng peaks olema 29. mai):

20. maiks lõpetajatel hinded väljas!!!

Moodle testi tagasiside:

Tagasiside

Test on <u>3. aprill</u> ja annab 20% koondhindest!

Järgmine nädal kordame seniõpitut!

Väga soovitatav on juba järgmiseks loenguks ette valmistada A4 paber valemitega, mida ka kontrolltööl kasutada.

Tänane teema:

jaotustihedus ja jaotusfunktsioon, keskväärtus ja standardhälve; pidevad jaotused (ühtlane jaotus, eksponentjaotus, normaaljaotus, Studenti jaotus)

jaotusseadused kokkuvõtlikult

Jaotustihedus *f*(*x*) on: $f(x) = \frac{dF(x)}{dx}$

Jaotustiheduse normeerimis- ja mittenegatiivsustingimus:

$$F(\infty) = \int_{-\infty}^{\infty} f(x)dx \equiv 1 \qquad f(x) \geq 0$$

Jaotusfunktsioon *F*(*x*) on: $F(x_0) = \int_{-\infty}^{x_0} f(x)dx$

Vahemiku tõenäosus on: $P\left\{x_1 \leq x \leq x_2\right\} = F(x_2) - F(x_1)$

Keskväärtus ja dispersioon kokkuvõtlikult

$$m(x) = \sum_{k=1}^{n} x_k p_k$$

$$m(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$m(x^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

$$F(x_{mediaan}) \equiv 0.5$$

$$D[X] = m\left[(X - m[X])^2\right]$$

$$D(x) = \sum_i (x_i - m)^2 p_i$$

$$D(x) = \int_{-\infty}^{\infty} (x - m)^2 \, f(x) \, dx$$

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

$$s(x) = \sqrt{D(x)}$$

<u>Ülesanne 3.17.</u> Kolmnurkjaotuse jaotustihedus on defineeritud järgnevalt: piirkondades $x < 5$ ning $x > 8$ on tema väärtus 0; vahemikus [5; 8] on tema väärtus ühtlaselt kasvav ($f(x) = cx$).

Leia

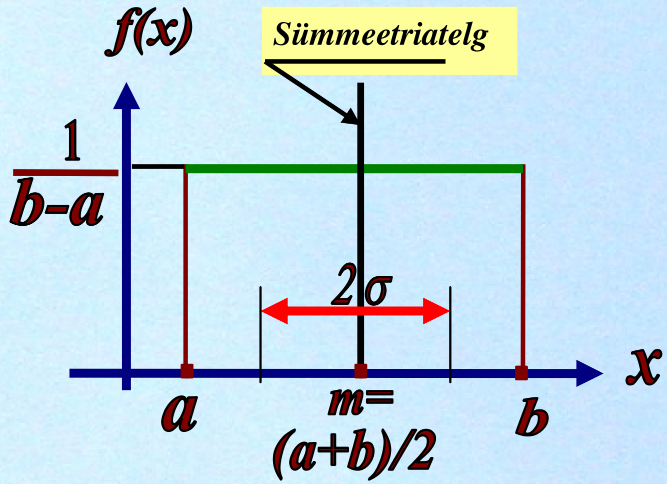
- koefitsient c;
- jaotusfunktsioon;
- keskväärtus,
- mediaan;
- standardhälve;
- tõenäosus, et sündmus satuks piirkonda $(m(x), 12)$?

enamkasutatavad jaotusseadused – ühtlane jaotus

Ühtlane jaotus

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

$$\sigma = \frac{1}{\sqrt{3}}\frac{b-a}{2} \approx 0.577\frac{b-a}{2}$$

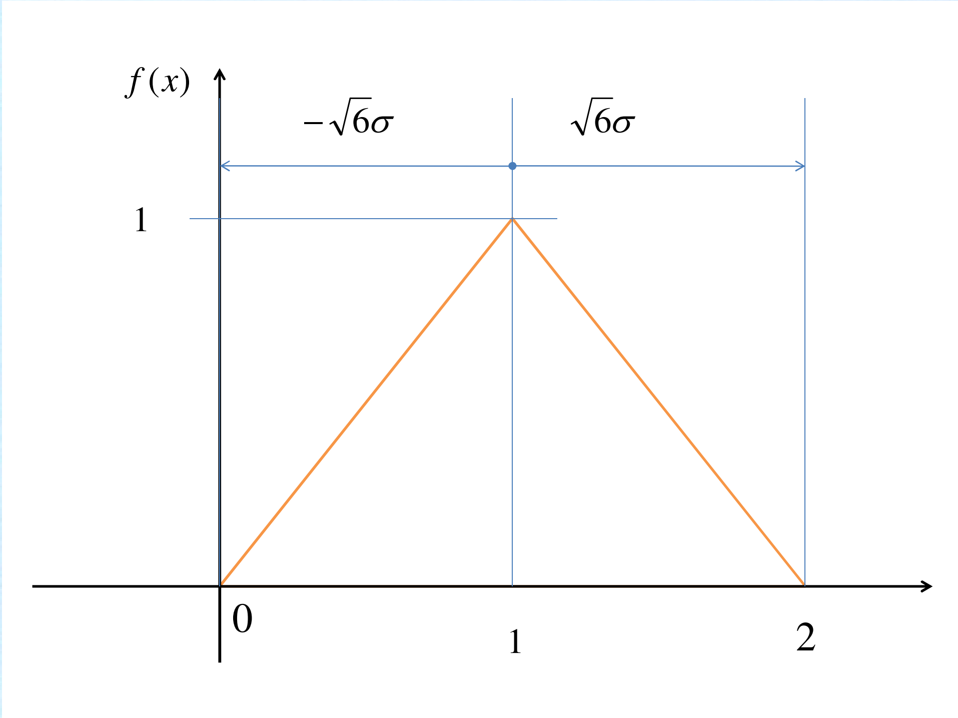


Ühtlase jaotuse näiteks on mõõtevahendi resolutsioonist tingitud määramatus, seda eriti selgelt digitaalnäiduga seadmetel. Kui näiteks digitaaltermomeetri resolutsioon on 0,1 ℃ ning näiduks on 22,6 ℃, siis tegelik temperatuur on vahemikus (22,55 ℃ – 22,65 ℃). Kuna meil pole mingi alust eeldada, et temperatuuril esineks mingisugused eelistatumaid väärtuseid, siis on kõige mõistlikum eeldada, et mõõtevahendi resolutsioonist tingitud määramatus on kirjeldatav ühtlase jaotusena.

Kolmnurkjaotus

Vaatame ühte erijuhtu kolmnurkjaotusest, mis on sümmeetriline ning asub lõigul [0; 2], seega on tema jaotustihedus on antud funktsiooniga

$$f(x) = \begin{cases} x; \ 0 < x \leq 1 \\ 2 - x; 1 < x \leq 2 \\ 0; \quad \text{mujal} \end{cases}$$



Et ülalkirjeldatud kolmnurkjaotus on sümmeetriline, siis ilmselt on tema keskväärtus

$$m = \frac{0+2}{2} = 1$$

enamkasutatavad jaotusseadused – kolmnurkjaotus

Kolmnurkjaotus

$$m(x) = \frac{0+2}{2} = 1$$

Kolmnurkjaotuse ruudu keskväärtus on:

$$m(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2 \cdot x\,dx + \int_1^2 x^2 \cdot (2-x)\,dx = \frac{x^4}{4}\Big|_0^1 + \frac{2x^3}{3}\Big|_1^2 - \frac{x^4}{4}\Big|_1^2 =$$

$$= \frac{1}{4} + \frac{16-2}{3} - \frac{16-1}{4} = \frac{3+56-45}{12} = \frac{14}{12}$$

ja kolmnurkjaotuse dispersioon ning standardhälve on:

$$D = m(x^2) - m^2 = \frac{14}{12} - (1)^2 = \frac{2}{12}$$

$$\sigma = \sqrt{D} = \sqrt{\frac{2}{12}} \approx 0{,}41$$

enamkasutatavad jaotusseadused – eksponentjaotus

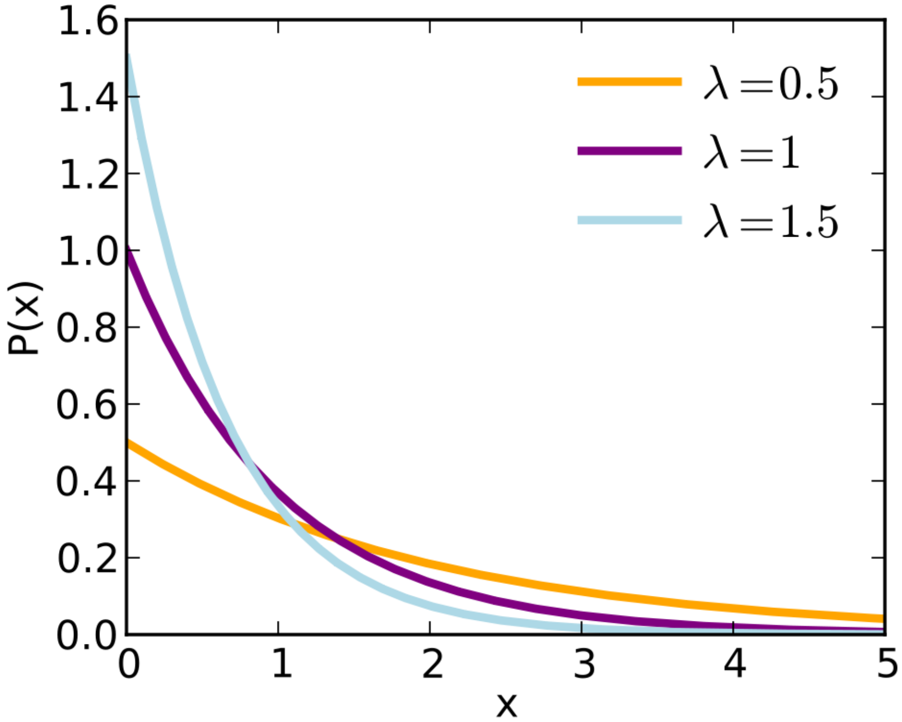
Eksponentjaotus

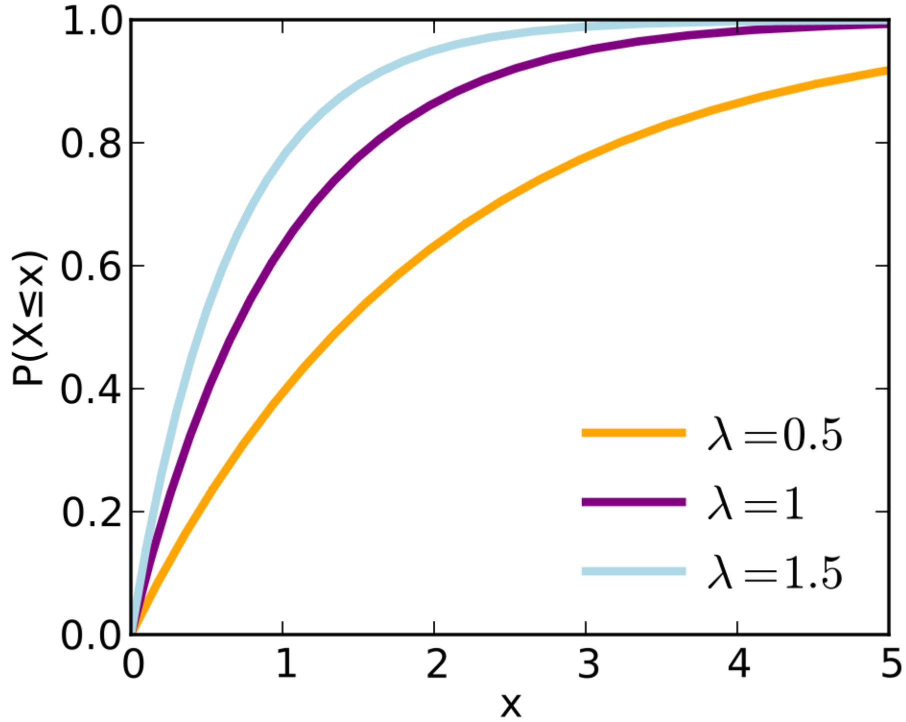
Eksponentjaotuse saate läbi arvutada radioaktiivse aine pooldumist kirjeldavas näites.

$$f(t) = \lambda \exp(-\lambda t)$$

$$F(t) = 1 - \exp(-\lambda t)$$

Siin $\lambda = 1 / t_0$, eksponentjaotuse keskväärtus ja standardhälve on mõlemad võrdsed parameetriga t_0.





enamkasutatavad jaotusseadused – eksponentjaotus

Hõõglambi eluiga allub eksponentjaotusele. Tootja väidab, et hõõglambi keskmine eluiga on 1000 tundi. Kui pikk on keskmise lambi eluiga? Kui suur on tõenäosus, et ühe lambi eluiga on üle 5000 tunni?

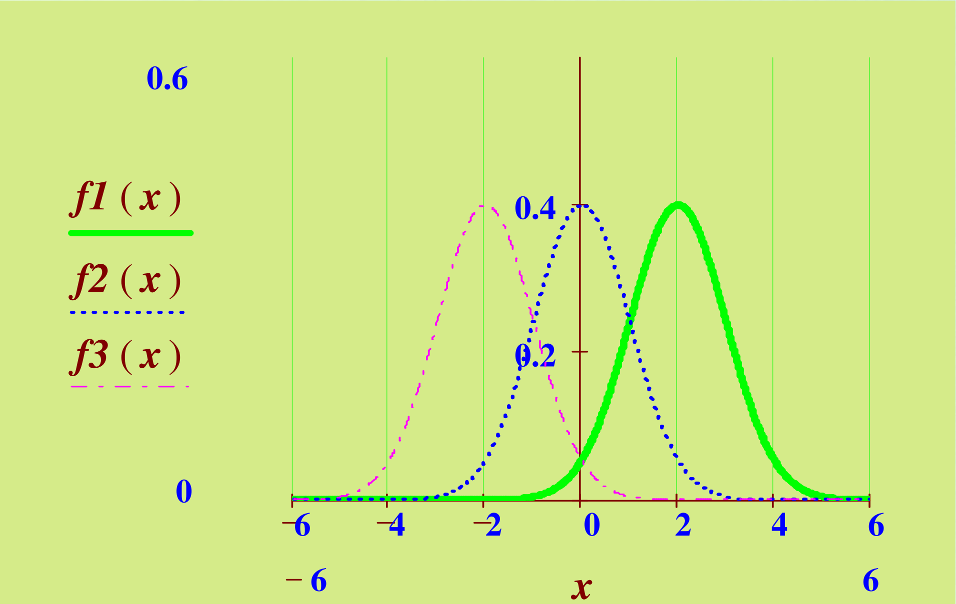
Lahendus – Mathcadiga – (lambi_eluiga.mcd).

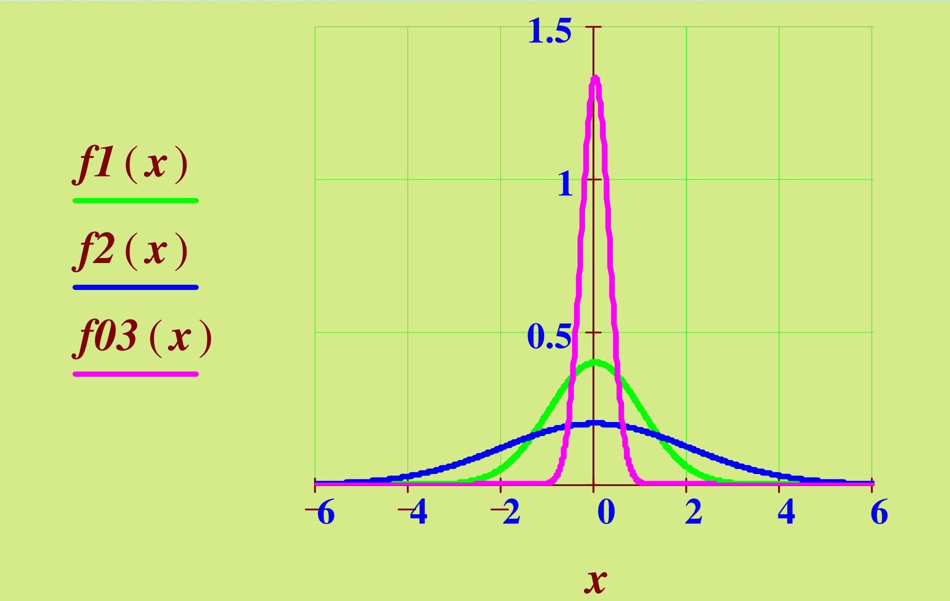
enamkasutatavad jaotusseadused – normaaljaotus

Normaaljaotus

Juhusliku suuruse jaotust nimetatakse normaaljaotuseks ehk Gaussi jaotuseks, kui jaotustihedus on

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right),$$





enamkasutatavad jaotusseadused – normaaljaotus

Normaaljaotus

Nagu teistegi jaotusseaduste puhul, on ka normaaljaotuse puhul mingisse intervalli sattumise tõenäosususe leidmiseks vaja teada tema jaotusfunktsiooni. Praktikas ei ole normaaljaotuse integreerimine just lihtne ülesanne. Selles loengukursuses me ei tõesta, et normaaljaotuse jaotusfunktsioon on:

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-a}{\sigma\sqrt{2}}\right)\right]$$

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,dt$$

Siin erf(*x*) tähistab veafunktsiooni. Praktiliste ülesannete lahendamisel kasutatakse valmisprogrammide abi.

enamkasutatavad jaotusseadused – normaaljaotus

Normaaljaotus

Mathcad'is on normaaljaotuse parameetritega a ja σ jaotustihedus $f(x)$ kohal x arvutatav funktsiooniga dnorm(x, a, σ) ning jaotusfunktsioon $F(x)$ kohal x funktsiooniga pnorm(x, a, σ).

Excel'is on jaotustihedus $f(x)$ arvutatav käsuga normdist(x, a, σ, 0) ning jaotusfunktsioon $F(x)$ käsuga normdist(x, a, σ, 1).

Näide:

Pliiatsitehases toodetavate pliiatsite pikkus allub normaaljaotusele keskväärtusega 180,0 mm ning standardhälbega 0,7 mm. Kui suur on tõenäosus, et toodangu seast juhuslikult valitud pliiatsi pikkus on vähem kui 179 mm?

normdist(179; 180,0; 0,7; 1) = 0,076563726 = 7,66%.

Normaaljaotus

Normaaljaotuse korral satub vahemikku ± 1σ ligikaudu 68%, vahemikku ± 2σ ligikaudu 95% ja vahemikku ± 3σ ligikaudu 99,7% sündmustest.

Normaaljaotust, kus keskväärtus a = 0 ja standardhälve σ = 1 nimetatakse standardiseeritud normaaljaotuseks N(0; 1).

Üleminekuvalem suvaliselt normaaljaotuselt X ~ N(a; σ) standardiseerutud normaaljaotuseks on:

$$\frac{X-a}{\sigma} \sim N(0;1)$$

<u>Ülesanne 3.18.</u> Olgu teada, et 20-aastaste noormeeste pikkused alluvad normaaljaotusele keskväärtusega a = 183 cm ja standardhälbega σ = 7 cm. Kui suur osa noormeestest on

- **lühemad kui a – 2 * σ = 1,69 m;**
- **pikemad kui 2 m;**
- **pikemad kui 2,1 m;**
- **pikkuse vahemikus 180 cm kuni 190 cm?**

<u>Ülesanne 3.19.</u> Ühe gümnaasiumi sisseastumiskatseks on komplekstest. Testi tulemused alluvad normaaljaotusele keskväärtusega *m* = 500 punkti ja standardhälbega σ = 100 punkti. 108-le kohale oli 427 kandidaati. Toomas sai katsetel 573 punkti. Kas sellest peaks piisama gümnaasiumisse pääsemiseks?

Studenti jaotus

Inglise teadlane Student, alias W.S. Gosset tõestas, et:

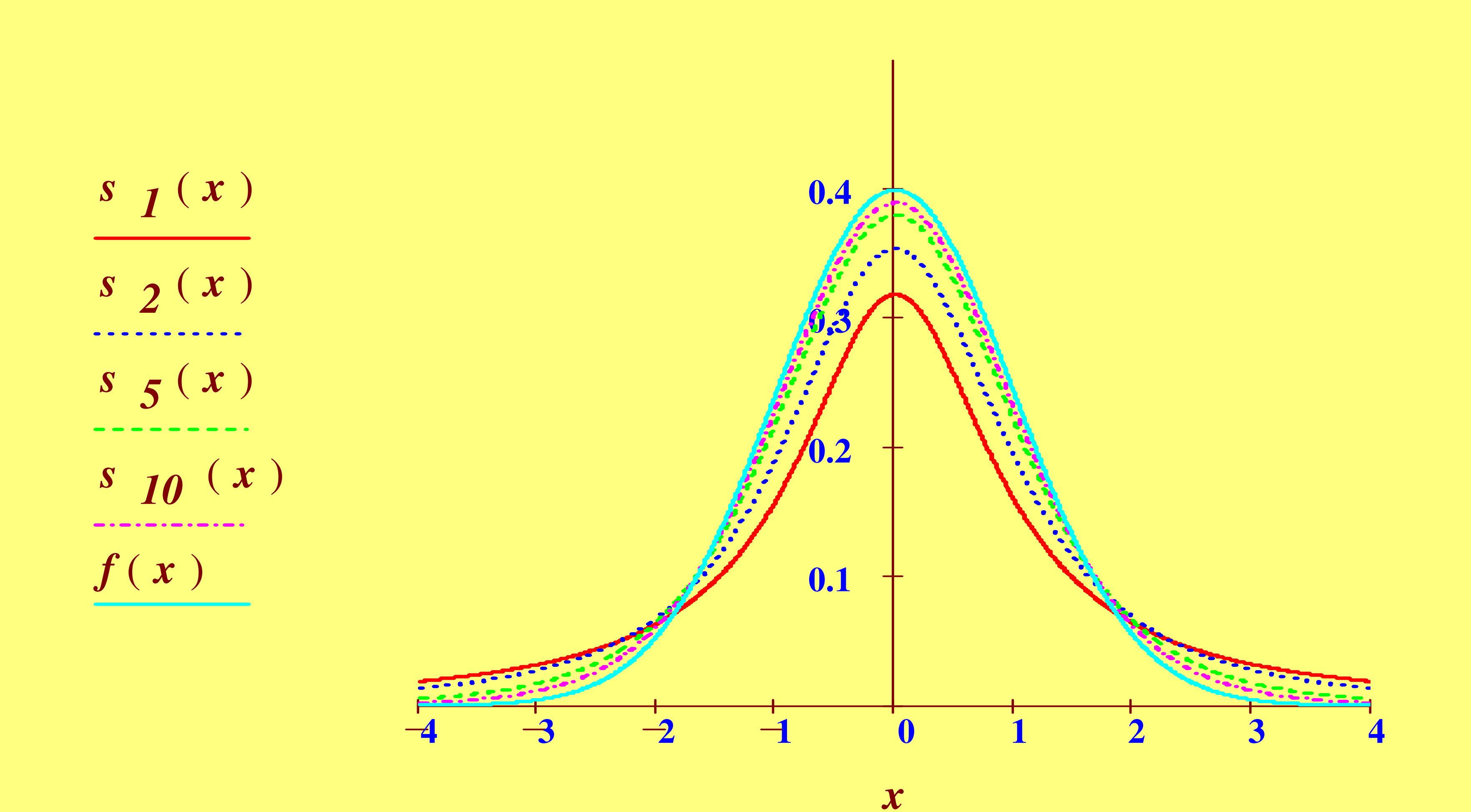
Kui X on normaalne juhuslik suurus, siis juhuslik suurus

$$T = \frac{\bar{x}_N - m[X]}{u(\bar{x}_N)}$$

allub jaotusele tihedusega

$$s_{N-1}(t) = C_{N-1}\left(1 + \frac{t^2}{N-1}\right)^{-N/2} .$$

enamkasutatavad jaotusseadused – Studenti jaotus



enamkasutatavad jaotusseadused – Studenti jaotus

Studenti jaotuse oluline omadus on, et juhusliku suuruse T **jaotusfunktsioon** $s_{N-1}(t)$ **sõltub vaid ühest parameetrist (nn vabadusastmete arvust)** $N-1$ **ja ei sõltu üldse juhusliku suuruse (mõõdetava suuruse)** X **konkreetsest dispersioonist ega keskväärtusest** (küll on aga oluline X normaalsus). Selles mõttes on tegu universaalfunktsiooniga, mis iseloomustab suvaliste ühesuguse normaaljaotusega juhuslike suuruste summa üldisi omadusi.

Otsene järeldus asjaolust, et juhuslik suurus allub Studenti jaotusele on, et usaldusnivoole p vastav mõõtmistulemuse paiknemise intervall on esitatav kujul

$$P\left[\bar{x}_N - t_{N-1}(p)u(\bar{x}_N) < x < \bar{x}_N + t_{N-1}(p)u(\bar{x}_N)\right] = p.$$

enamkasutatavad jaotusseadused – Studenti jaotus

$$P\left[\bar{x}_N - t_{N-1}(p)u(\bar{x}_N) < x < \bar{x}_N + t_{N-1}(p)u(\bar{x}_N)\right] = p.$$

Siin suurus $t_{N-1}(p)$ on (Studenti) ***t*-kordaja** – parandustegur etteantud usaldusnivoo p ja $N-1$ vabadusastme korral. t-kordaja on samuti ainult ühest parameetrist N sõltuv universaalne usaldusnivoo funktsioon, mis on leitav Studenti jaotusest. Tavaliselt antakse t-kordaja kindlate p väärtuste jaoks tabuleeritult. Mathcadis on t-kordaja leidmiseks funktsioon $qt\left(\frac{1+p}{2}, \nu\right)$; Excelis funktsioon $TINV(1-p, \nu)$.

Antud valemit nimetatakse gi ***Studenti testiks***.

enamkasutatavad jaotusseadused – Studenti jaotus

Vabadusastmete arv $\nu = N - 1$	Osa p protsentides					
	68,27[l]	90	95	95,45[l]	99	99,73[l]
1	1,84	6,31	12,71	13,97	63,66	235,80
2	1,32	2,92	4,30	4,53	9,92	19,21
3	1,20	2,35	3,18	3,31	5,84	9,22
4	1,14	2,13	2,78	2,87	4,60	6,62
5	1,11	2,02	2,57	2,65	4,03	5,51
6	1,09	1,94	2,45	2,52	3,71	4,90
7	1,08	1,89	2,36	2,43	3,50	4,53
8	1,07	1,86	2,31	2,37	3,36	4,28
9	1,06	1,83	2,26	2,32	3,25	4,09
10	1,05	1,81	2,23	2,28	3,17	3,96
11	1,05	1,80	2,20	2,25	3,11	3,85
12	1,04	1,78	2,18	2,23	3,05	3,76
13	1,04	1,77	2,16	2,21	3,01	3,69
14	1,04	1,76	2,14	2,20	2,98	3,64
15	1,03	1,75	2,13	2,18	2,95	3,59
16	1,03	1,75	2,12	2,17	2,92	3,54
17	1,03	1,74	2,11	2,16	2,90	3,51
18	1,03	1,73	2,10	2,15	2,88	3,48
19	1,03	1,73	2,09	2,14	2,86	3,45
20	1,03	1,72	2,09	2,13	2,85	3,42
25	1,02	1,71	2,06	2,11	2,79	3,33
30	1,02	1,70	2,04	2,09	2,75	3,27
35	1,01	1,70	2,03	2,07	2,72	3,23
40	1,01	1,68	2,02	2,06	2,70	3,20
45	1,01	1,68	2,01	2,06	2,69	3,18
50	1,01	1,68	2,01	2,05	2,68	3,16
100	1,005	1,660	1,984	2,025	2,626	3,077
∞	1,000	1,645	1,960	2,000	2,576	3,000

Ülesanne 3.20. Olgu teada, et 20-aastaste noormeeste pikkused alluvad normaaljaotusele, jaotuse parameetrite määramiseks mõõdeti nelja 20-aastase noormehe pikkused. Keskmiseks noormeeste pikkuseks tuli 181 cm standardhälbega 8 cm. Lähtudes Studenti jaotusest, millisesse vahemikku peaks jääma 95% noormeeste pikkused? Kui suur oleks see vahemik siis, kui mõõdetud noormeeste arv oleks olnud 60?

enamkasutatavad jaotusseadused

Tegelikult on erinevaid jaotusseaduseid "mustmiljon" ja neid kõiki igaks juhuks tundma õppida ei ole mõttekas.

Kuidas teada saada, milline jaotusseadus meid huvitav protsessi võiks kirjeldada?

Enamasti leitakse mingit protsessi kirjeldav jaotusseadus kas teoreetilistel kaalutlustel või eksperimendi tulemusena.

<u>Ülesanne.</u> Tudengite eksamitulemust mingil eksamil hinnatakse skaalal 0-st 100-ni. Oletame, et tudeng on eksamist läbi saanud, kui tema tulemus on vähemalt 55 punkti. Oletame, et eksami tulemused on modelleeritavad järgmise jaotustihedusega:

$$f(x) = \begin{cases} \dfrac{x}{4900}; & 10 \le x < 90 \\ \dfrac{90}{4900}; & 90 \le x < 100 \\ 0; & \text{mujal} \end{cases}$$

a) leia tõenäosus, et tudeng kukub eksamilt läbi?

b) leia tõenäosus, et tudeng saab vähemalt 95 punkti?

c) leia eksami keskmine tulemus?

d) leia eksami mediaantulemus?

e) leia eksami tulemuste standardhälve?

Kodune töö

Selle nädala test sulgub 26.03.2013 kell 23:55.

Juba sulgunud teste ei saa hiljem järgi teha!!!

Testi sulgumise järel avaneb “test harjutamiseks”, seda saab harjutada piiramatult kuni semestri lõpuni.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Moodle testi tagasiside:

Tagasiside

EKSAMI AJAD (viimane korraline loeng peaks olema 29. mai):

20. maiks lõpetajatel hinded väljas!!!

Moodle testi tagasiside:

Tagasiside

Test on <u>3. aprill</u> ja annab 20% koondhindest!

Järgmine nädal kordame seniõpitut!

Väga soovitatav on juba järgmiseks loenguks ette valmistada A4 paber valemitega, mida ka kontrolltööl kasutada.

Tänane teema:

jaotustihedus ja jaotusfunktsioon, keskväärtus ja standardhälve; pidevad jaotused (ühtlane jaotus, eksponentjaotus, normaaljaotus, Studenti jaotus)

Juhusliku suuruse arvkarakteristikud

__Ülesanne.__ **Tudengite eksamitulemust mingil eksamil hinnatakse skaalal 0-st 100-ni. Oletame, et tudeng on eksamist läbi saanud, kui tema tulemus on vähemalt 55 punkti. Oletame, et eksami tulemused on modelleeritavad järgmise jaotustihedusega:**

$$f(x) = \begin{cases} \dfrac{x}{4900}; & 10 \le x < 90 \\ \dfrac{90}{4900}; & 90 \le x < 100 \\ 0; & \text{mujal} \end{cases}$$

a) leia tõenäosus, et tudeng kukub eksamilt läbi?

b) leia tõenäosus, et tudeng saab vähemalt 95 punkti?

c) leia eksami keskmine tulemus?

d) leia eksami mediaantulemus?

e) leia eksami tulemuste standardhälve?

Kodune töö

Selle nädala test sulgub 02.04.2013 kell 23:55.

Seekord saab testi korduvalt täita, arvesse läheb parim sooritus!

Juba sulgunud teste ei saa hiljem järgi teha!!!

Tõenäosusteooria ja statistika

LOFY.03.056

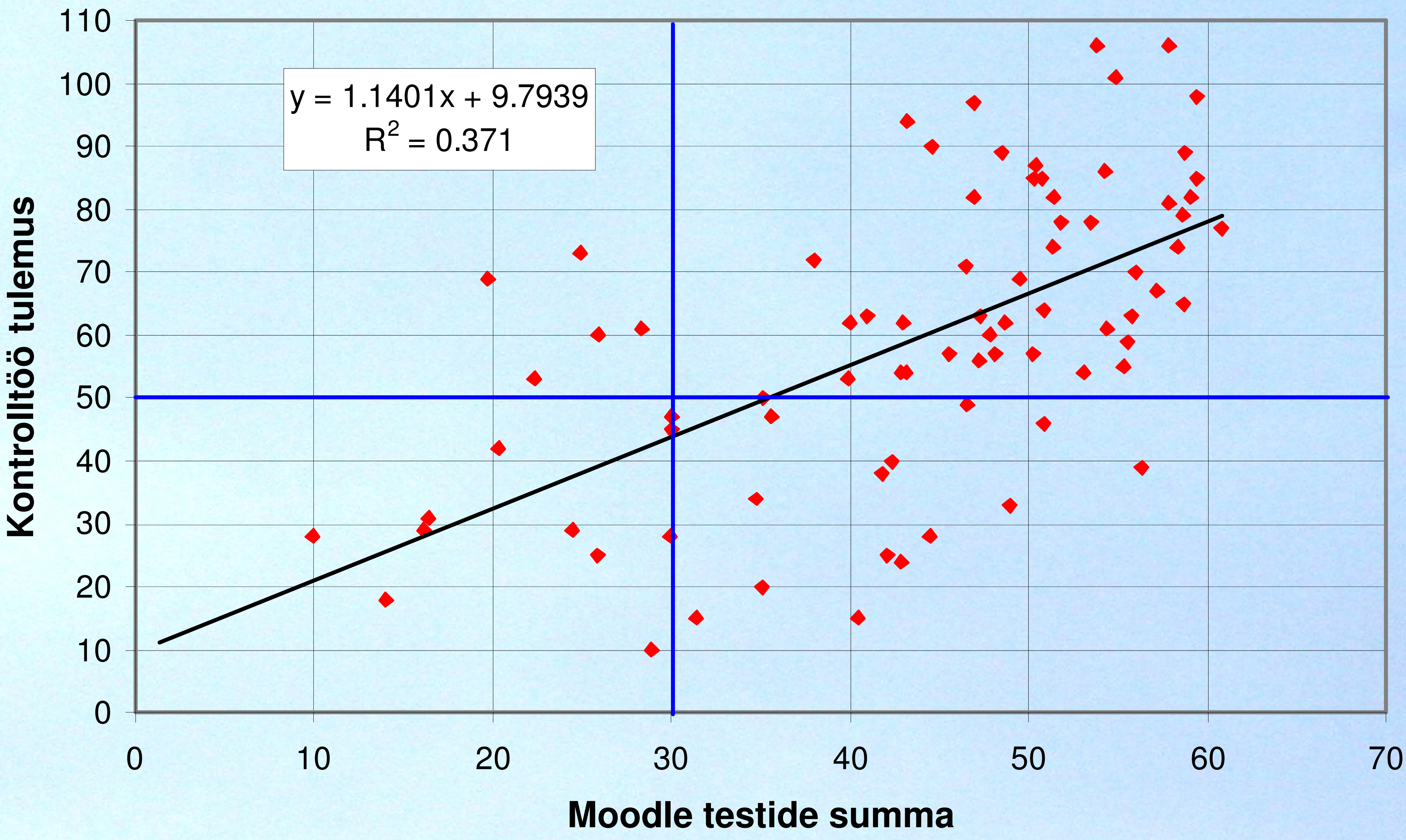
Erko Jakobson, PhD

2013 kevad

Järgnevate loengute esialgne plaan

9.nädal 10.04	Kontrolltöö analüüs, grupitöö tutvustus; tõenäosuse eksperimentaalne hindamine; keskväärtuse ja dispersiooni hindamine mõõtmistulemustest.
10.nädal 17.04	Loengu asemel on kontrolltöö järeltöö
11.nädal 24.04	Kovariatsioon ja korrelatsioon; Z-test ehk normaalsustest; T-test ehk Studenti test.
12.nädal 01.05	Loengut ei toimu (VOLBER)
13.nädal 08.05	Kovariatsioon ja korrelatsioon; Z-test ehk normaalsustest; T-test ehk Studenti test; hüpoteeside kontrollimine.
14.nädal 15.05	Regressioonanalüüs; vähimruutude meetod.
15.nädal 22.05	Regressioonanalüüs; vähimruutude meetod, grupitööde analüüs.
16.nädal 29.05	Grupitööde analüüs; konsultatsioon.

Kontrolltöö tagasiside:



Kontrolltöö tagasiside:

Seoses oodatust oluliselt madalama tulemusega, tuleb korralduses väike muudatus:

Järeltööd võivad teha kõik ning koefitsienti 0,8 ei rakendata. Järeltöö tegijatel läheb arvesse järeltöö hinne, sõltumata töö hindest.

Järeltöö aeg: kolmapäeval, 17. aprillil kell 12-14.

Grupitöö (10% eksamihindest):

Grupitöö on mingi andmebaasi statistiliste seoste ja hüpoteeside otsimine-kontrollimine vabalt valitud arvutusprogrammiga ning saadud tulemuste vormistamine. Grupi suurus – kuni 5 tudengit.

Grupitöö eesmärk:

- **Praktilise kogemuse saamine andmete statistilise analüüsiga. Väga tõenäoliselt sisaldab ka lõputöö statistilist andmetöötlust.**
- **Statistikaprogrammidega tutvumine, sest 3 EAP raames arvutipraktikumi ei jõua teha.**
- **Õpitu rakendamine praktikas.**

Grupitöö (10% eksamihindest):

Grupitöö näide:

Andmebaas – meteomasti 2 m ja 10 m tuule kiirus, resolutsiooniga 2 minutit, 30 päeva pikkune periood suvel.

Hüpoteesid:

- 10 m kõrgusel on tuule kiirus suurem kui 2 m kõrgusel;
- päeval (kell 12 – 16) on tuule kiirus suurem kui öösel (kell 00 – 04);
- päeval on tuule kiiruse varieeruvus suurem kui öösel;
- 2 m kõrgusel on tuule kiiruse varieeruvus suurem kui 10 m kõrgusel;
- korrelatsioon 10 m ja 2 m tuule kiiruse vahel on suurematel tuule kiirustel kõrgem kui väiksemate tuulekiiruste korral;
- jne

Grupitöö:

Analüüsitavad andmed leiavad tudengid soovitatavalt ise, aga vajadusel võib andmebaase saada ka õppejõu käest.

Grupitöö tähtajad:

10. nädalaks (17.04) – grupid moodustatud

11. nädalaks (24.04) – andmebaas valitud ja hüpoteesid (min 2 tk) püstitatud

12. nädala lõpuks (03.05) – esimene versioon valmis

14. nädalaks (15.05) – viimane versioon valmis

1 preemiapunkt, kui grupis on 1 või 2 kolmanda kursuse tudengit ja andmebaas on seotud lõputööga.

Tõenäosuse eksperimentaalne hindamine

tõenäosuse eksperimentaalne hindamine;
keskväärtuse hindamine mõõtmistulemustest;
dispersiooni hindamine mõõtmistulemustest.

Kahjuks veel statistika osa uut konspekti pole ☹

lisamaterjalide osas on see materjal leitav

“II_osa_2012k.pdf”, leheküljed 30 – 42.

Kodune töö

- Tegelege grupitööga, Moodle testi ei ole!
- Järgmine nädal (17.04) loengut ei ole, kuid tuleb Moodle test!

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Järgnevate loengute esialgne plaan

9.nädal 10.04	Kontrolltöö analüüs, grupitöö tutvustus; tõenäosuse eksperimentaalne hindamine; keskväärtuse ja dispersiooni hindamine mõõtmistulemustest.
10.nädal 17.04	Loengu asemel on kontrolltöö järeltöö
11.nädal 24.04	hüpoteeside kontrollimine; T-test ehk Studenti test.
12.nädal 01.05	Loengut ei toimu (VOLBER)
13.nädal 08.05	Kovariatsioon ja korrelatsioon; Z-test ehk normaalsustest; T-test ehk Studenti test.
14.nädal 15.05	Regressioonanalüüs; vähimruutude meetod.
15.nädal 22.05	Regressioonanalüüs; vähimruutude meetod, grupitööde analüüs.
16.nädal 29.05	Grupitööde analüüs; konsultatsioon.

Grupitöö:

Analüüsitavad andmed leiavad tudengid soovitatavalt ise, aga vajadusel võib andmebaase saada ka õppejõu käest.

Grupitöö tegemine on vabatahtlik, saadavad punktid on preemiapunktid! Eksami osakaal muutub 60% → 70%.

Grupitöö viimase versiooni tähtaeg on 14. nädal (15.05)

1 preemiapunkt, kui grupis on 1 või 2 kolmanda kursuse tudengit ja andmebaas on seotud lõputööga.

Grupitöö (10% eksamihindest):

Grupitöö näide:

Andmebaas – meteomasti 2 m ja 10 m tuule kiirus, resolutsiooniga 2 minutit, 30 päeva pikkune periood suvel.

Hüpoteesid:

- **10 m kõrgusel on tuule kiirus suurem kui 2 m kõrgusel;**
- **päeval (kell 12 – 16) on tuule kiirus suurem kui öösel (kell 00 – 04);**
- **päeval on tuule kiiruse varieeruvus suurem kui öösel;**
- **2 m kõrgusel on tuule kiiruse varieeruvus suurem kui 10 m kõrgusel;**
- **korrelatsioon 10 m ja 2 m tuule kiiruse vahel on suurematel tuule kiirustel kõrgem kui väiksemate tuulekiiruste korral;**
- **jne**

keskväärtuse hindamine mõõtmistulemustest

Tehes kokku *N* sõltumatut mõõtmist, saame juhusliku suuruse *X* valimi {x_1, x_2, x_3, ..., x_i, ..., x_{N-1}, x_N}. Selle valimi aritmeetiline keskväärtus

$$\bar{x}_N = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

on parim lähend tõelisele väärtusele piirväärtuse mõttes

$$\lim_{N \to \infty} \bar{x}_N = x_t.$$

Dispersiooni hindamine mõõtmistulemustest

Dispersiooni katselisel määramisel on olukord analoogiline keskväärtuse hindamisega. Meenutame siin dispersiooni definitsiooni

$$D[X] \quad = \quad \sum_{i=1}^{N} (x_i - m)^2 \cdot p_i$$

Kuna summa üle p_i-de peab olema võrdne ühega ning pole põhjust oletada, et näiteks esimesed mõõtmised olid täpsemad kui viimased, siis $p_i = 1/N$, seega

$$D[X] \quad \approx \quad \sigma_N{}^2 \quad = \quad \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - m)^2$$

Alternatiivset valemit dispersiooni arvutamiseks nimetatakse empiiriliseks dispersiooniks:

$$D[X] \quad \approx \quad s_N{}^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - m)^2$$

Dispersiooni hindamine mõõtmistulemustest

$$D[X] \approx \sigma_N{}^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - m)^2 \qquad D[X] \approx s_N{}^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - m)^2$$

Osutub, et nende suuruste keskväärtused rahuldavad seoseid

$$m[\sigma_N^2] = \frac{N-1}{N} d, \qquad m[s_N^2] = d,$$

kus *d* on üksiksuuruse X_k dispersioon. Nagu näeme, lähendab suurus $s_N{}^2$ vähemalt matemaatilise ootuse (keskmise) mõttes tõelist dispersiooni õigesti, kuna suurus $\sigma_N{}^2$ alahindab dispersiooni. Seepärast kasutame edaspidi juhuslikus suuruse dispersiooni hindamisel suuruse *X* empiirilist dispersiooni.

Juhusliku suuruse *X* empiiriline standardhälve on vastavalt

$$s_N = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x}_N)^2}{N-1}} \approx \sqrt{d}\ .$$

Dispersiooni hindamine mõõtmistulemustest

Aritmeetilise keskmise dispersioon on (ilma tõestuseta):

$$D\left[\overline{X}_N\right] = \frac{d}{N}$$

ning järeldusena

$$\sigma(\overline{X}_N) = \sqrt{\frac{d}{N}} \approx \frac{s_N}{\sqrt{N}}$$

Seega: ühesuguselt jaotunud *N* sõltumatu suuruse aritmeetilise keskmise dispersioon on *N* korda väiksem kui üksiksuuruse dispersioon ja standardhälve on ruutjuur *N* korda väiksem kui üksiksuuruse standardhälve.

Dispersiooni hindamine mõõtmistulemustest

Näide: Tudeng ostis karbi (6 tk) mune, suurus M, mis tootja kinnitusel tähendab kaaluvahemikku (53, 63) g. Munade kaalud olid järgmised:

1) 55 g

2) 58 g

3) 54 g

4) 56 g

5) 60 g

6) 53 g

$$m(x) = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$s(x) = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{x}_N)^2}{N-1}} \approx \sqrt{d} \, .$$

$$s(\overline{x}) = \sqrt{\frac{d}{N}} \approx \frac{s(x)}{\sqrt{N}}$$

Leia munade kaalu keskmine väärtus ning standardhälve.

Statistilised hüpoteesid

Statistiliseks hüpoteesiks nimetatakse teatud teineteist välistavate väidete paari üldkogumi(te) või tema parameetrite kohta.

H0 – nullhüpotees, mis tavaliselt väljendab uurijat mittehuvitavat juhtu (üldkogumi vastamine teatud standardile).

Nullhüpoteesi ei ole võimalik tõestada. Selle vastuvõtmine tähendab, et kui uurija tahab mingit erinevust, mõju või seose olemasolu tõestada, siis tuleb tal mõõtmisi jätkata.

H1 – sisukas e. alternatiivne e. konkureeriv hüpotees, mida uurija soovib tõestada (tavaliselt mingi erinevuse, mõju või seose olemasolu).

Statistilised hüpoteesid

Hüpoteeside kontrollimisel püütakse tõestada sisukas hüpotees nullhüpoteesi kummutamise teel.

Selleks arvutatakse valimi andmete põhjal teatud teststatistik, mille teoreetiline jaotus nullhüpoteesi kehtivuse korral on teada.

Juhul kui leitud teststatistiku väärtus on ebatõenäoline, võrreldes tema teoreetilise jaotusega, loetakse nullhüpotees kummutatuks ja sisukas hüpotees tõestatuks.

Kui sisukat hüpoteesi tõestada ei õnnestu, jäädakse nullhüpoteesi juurde, mis võib tähendada nii seda, et

1) olukord vastas nullhüpoteesile; kui ka seda, et

2) valimi maht oli liiga väike sisuka hüpoteesi tõestamiseks.

Statistilised hüpoteesid

Kuna statistiliste hüpoteeside kontrollimisel tehakse valimi põhjal järeldusi üldkogumi kohta, on võimatu vältida vigu. Vigu saab olla kaht liiki.

1. Esimest liiki viga tekib siis, kui võetakse vastu sisukas hüpotees, aga tegelikult on õige nullhüpotees. See on raske viga, mis tähendab, et uurija “tõestas” erinevuse, mõju või seose mida tegelikult ei ole, vaid mis juhuslikult ilmnes mõõdetud valimis.

2. Teist liiki viga tekib siis, kui jäädakse nullhüpoteesi juurde, ehkki tegelikult on õige sisukas hüpotees. See on kergem viga, mis enamasti tähendab, et soovitu tõestamiseks tuleb mõõtmisandmeid juurde koguda.

Statistilised hüpoteesid

Esimest liiki vead on sageli ohtlikud vead, mille tegemist tuleks võimalikult vältida.

Esimest liiki vea tegemise suurimat lubatavat tõenäosust nimetatakse olulisuse nivooks e. riskiprotsendiks.

Olulisuse nivood tähistatakse tähega α.

Olulisuse nivooks valitakse mingi väike arv, sageli 0,1; 0,05; 0,01 (sõltuvalt selles, kui rasketele tagajärgedele võib 1. liiki vea tegemine viia). Füüsikas kasutatakse tavaliselt olulisuse nivood 0,05.

Statistilised hüpoteesid

Ühepoolsed ja kahepoolsed hüpoteesid.

Kahepoolsed hüpoteesid kontrollivad, kas keskväärtused on omavahel erinevad:

Nullhüpotees:	**H0: $z = z_0$**
Sisukas hüpotees:	**H1: $z \neq z_0$**

Ühepoolsed hüpoteesid kontrollivad, kas üks keskväärtus on teisest suurem:

Nullhüpotees:	**H0: $z \leq z_0$**
Sisukas hüpotees:	**H1: $z > z_0$**

Või

Nullhüpotees:	**H0: $z \geq z_0$**
Sisukas hüpotees:	**H1: $z < z_0$**

Statistilised hüpoteesid

Statistiliste hüpoteeside kontrollimine toimub tavaliselt vahemikhinnangute kaudu.

Olgu näiteks nullhüpoteesiks see, et hinnatav parameeter z on võrdne etteantud väärtusega z_0:

Nullhüpotees:	**H0: $z = z_0$**
Sisukas hüpotees:	**H1: $z \neq z_0$**

Leiame usaldusvahemiku (z_a, z_b), nii et

$$P(z_a < z < z_b) = 1 - \alpha.$$

Kui antud väärtus z_0 satub vahemikust (z_a, z_b) välja, siis saab vastu võtta sisuka hüpoteesi, muidu tuleb jääda nullhüpoteesi juurde.

Studenti test

Studenti test ehk t-test on normaaljaotusele alluva juhusliku suuruse keskväärtuste võrdlemiseks.

T-testi puhul kasutatakse t-statistikut ehk Studenti suhet, mille väärtuse saab arvutada valimist järgnevalt:

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)}$$

siin *N* on valimi maht ja *s*(*z*) on standardhälve.

Saab tõestada, et kui üldkogum allub normaaljaotusele, siis t-statistik allub *n* – 1 vabadusastmega Studenti jaotusele. Meenutame, et Studenti jaotus sõltus ainult vabadusastmete arvust ning soovitud usaldusnivoost.

Studenti test

Studenti testi puhul peab sisuka hüpoteesi vastuvõtmiseks olema täidetud tingimus:

ühepoolse hüpoteesi puhul $|t| > t_{\alpha,\nu}$

Kahepoolse hüpoteesi puhul $|t| > t_{\frac{\alpha}{2},\nu}$

Näiteks olulisuse nivoo 0,05 puhul tuleb ühepoolse hüpoteesi kontrollimisel kasutada α = 0,05, kahepoolse hüpoteesi kontrollimisel aga α = 0,025, sest viga võib tulla mõlemas suunas.

Studenti test

Studenti jaotus, Excelis käsk TINV(*P*, v), kus *P* on lubatud esimest tüüpi vea tõenäosus ning v = *n* – 1 on vabadusastmete arv.

v	10%	5%	2.50%	1%	0.50%
1	6.31	12.71	25.45	63.66	127.32
2	2.92	4.30	6.21	9.92	14.09
3	2.35	3.18	4.18	5.84	7.45
4	2.13	2.78	3.50	4.60	5.60
5	2.02	2.57	3.16	4.03	4.77
6	1.94	2.45	2.97	3.71	4.32
7	1.89	2.36	2.84	3.50	4.03
8	1.86	2.31	2.75	3.36	3.83
9	1.83	2.26	2.69	3.25	3.69
10	1.81	2.23	2.63	3.17	3.58
15	1.75	2.13	2.49	2.95	3.29
25	1.71	2.06	2.38	2.79	3.08
50	1.68	2.01	2.31	2.68	2.94
100	1.66	1.98	2.28	2.63	2.87
1000	1.65	1.96	2.24	2.58	2.81
10000	1.65	1.96	2.24	2.58	2.81

Studenti test

Näide: Tudeng ostis karbi (6 tk) mune, suurus M, mis tootja kinnitusel tähendab kaaluvahemikku (53, 63) g. Munade kaalud olid järgmised:

1) 55 g

2) 58 g

3) 54 g

4) 56 g

5) 60 g

6) 53 g

$$m(x) = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{55 + 58 + 54 + 56 + 60 + 53}{6} = 56\ g$$

$$s(x) = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x}_N)^2}{N-1}} = \sqrt{\frac{(-1)^2 + (2)^2 + (-2)^2 + (0)^2 + (4)^2 + (-3)^2}{5}} =$$

$$= \sqrt{\frac{1 + 4 + 4 + 0 + 16 + 9}{5}} = \sqrt{\frac{34}{5}} = \sqrt{6{,}8} = 2{,}608\ g$$

Kontrolli püstitatud hüpoteesid:

1. **Munade mass erines kaaluvahemiku keskmisest (58 g).**
2. **Munad olid kergemad kui kaaluvahemiku keskmine (58 g).**

Studenti test

Näide: Tudeng ostis karbi (6 tk) mune, suurus M, mis tootja kinnitusel tähendab kaaluvahemikku (53, 63) g. Munade kaalud olid järgmised:

$$m(x) = 56 \ g$$

$$s(x) = \sqrt{6{,}8} = 2{,}608 \ g$$

Tõesta hüpoteesid:

1. **Munade mass erines kaaluvahemiku keskmisest (58 g).**
2. **Munad olid kergemad kui kaaluvahemiku keskmine (58 g).**

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)} = \sqrt{6} \cdot \frac{56 - 58}{\sqrt{6{,}8}} = -1{,}879.$$

$$|t| = 1{,}88 < t_{\frac{\alpha}{2},\nu} = t_{0{,}025{,}5} = 3{,}16$$

$$|t| = 1{,}88 < t_{\alpha,\nu} = t_{0{,}05{,}5} = 2{,}57$$

Seega tuleb mõlema hüpoteesi puhul jääda nullhüpoteesi juurde.

Kodune töö

**Moodle testi tuleb nädala lõpuks,
annan sellest eraldi teada.**

Järgmine nädal (1.05) loengut ei ole (volber).

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Järgnevate loengute esialgne plaan

13.nädal 08.05	hüpoteeside kontrollimine;
	χ^2-test; F-test; T-test ehk Studenti test.
14.nädal 15.05	Regressioonanalüüs; vähimruutude meetod.
	χ^2-test; F-test; T-test ehk Studenti test.
15.nädal 22.05	Regressioonanalüüs; vähimruutude meetod,
	Kovariatsioon ja korrelatsioon; grupitööde analüüs.

EKSAMI ajad:

22. mai kell 14 – 16 (3. kursuse tudengitele)

29. mai

5. juuni

Järeleksami aeg:

17. juuni

Grupitöö:

Analüüsitavad andmed leiavad tudengid soovitatavalt ise, aga vajadusel võib andmebaase saada ka õppejõu käest.

Grupitöö tegemine on vabatahtlik, saadavad punktid on preemiapunktid! Eksami osakaal muutub 60% → 70%.

Grupitöö viimase versiooni tähtaeg on 15. nädal (21.05)

1 preemiapunkt, kui grupis on 1 või 2 kolmanda kursuse tudengit ja andmebaas on seotud lõputööga.

Studenti test

Näide: Tudeng ostis karbi (6 tk) mune, suurus M, mis tootja kinnitusel tähendab kaaluvahemikku (53, 63) g. Munade kaalud olid järgmised:

1) 55 g

2) 58 g

3) 54 g

4) 56 g

5) 60 g

6) 53 g

$$m(x) = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{55+58+54+56+60+53}{6} = 56\ g$$

$$s(x) = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x}_N)^2}{N-1}} = \sqrt{\frac{(-1)^2+(2)^2+(-2)^2+(0)^2+(4)^2+(-3)^2}{5}} =$$

$$= \sqrt{\frac{1+4+4+0+16+9}{5}} = \sqrt{\frac{34}{5}} = \sqrt{6{,}8} = 2{,}608\ g$$

Kontrolli püstitatud hüpoteesid:

1. **Munade mass erines kaaluvahemiku keskmisest (58 g).**
2. **Munad olid kergemad kui kaaluvahemiku keskmine (58 g).**

Studenti test

Näide: Tudeng ostis karbi (6 tk) mune, suurus M, mis tootja kinnitusel tähendab kaaluvahemikku (53, 63) g. Munade kaalud olid järgmised:

$$m(x) = 56 \ g$$

$$s(x) = \sqrt{6,8} = 2,608 \ g$$

Tõesta hüpoteesid:

1. **Munade mass erines kaaluvahemiku keskmisest (58 g).**
2. **Munad olid kergemad kui kaaluvahemiku keskmine (58 g).**

$$t = \sqrt{N}\,\frac{\bar{z} - z_0}{s(z)} = \sqrt{6} \cdot \frac{56 - 58}{\sqrt{6,8}} = -1,879.$$

$$|t| = 1,88 < t_{\frac{\alpha}{2},\nu} = t_{0,025,5} = 3,16$$

$$|t| = 1,88 < t_{\alpha,\nu} = t_{0,05,5} = 2,57$$

Seega tuleb mõlema hüpoteesi puhul jääda nullhüpoteesi juurde.

Studenti test

Jätkame konspektist. Analüüsime Tartu temperatuuride andmeid.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Järgnevate loengute esialgne plaan

14.nädal 15.05 **Regressioonanalüüs; vähimruutude meetod.**
χ^2-test; F-test; T-test ehk Studenti test.

15.nädal 22.05 **Regressioonanalüüs; vähimruutude meetod,**
Kovariatsioon ja korrelatsioon; grupitööde analüüs.

EKSAMI ajad:

22. mai kell 14 – 16 ruum 410 (3. kursuse tudengitele)

29. mai kell 12 – 14 ruum 160

5. juuni kell 12 – 14 ruum 160

Järeleksami aeg:

17. juuni kell 10 – 12 ruum 160

Grupitöö:

Analüüsitavad andmed leiavad tudengid soovitatavalt ise, aga vajadusel võib andmebaase saada ka õppejõu käest.

Grupitöö tegemine on vabatahtlik, saadavad punktid on preemiapunktid! Eksami osakaal muutub 60% → 70%.

Grupitöö viimase versiooni tähtaeg on 15. nädal (21.05)

1 preemiapunkt, kui grupis on 1 või 2 kolmanda kursuse tudengit ja andmebaas on seotud lõputööga.

Korrelatsioonanalüüs

Jätkame konspektist: Korrelatsioonanalüüs

Eksami kava:

Eksam on kirjalik, 50% ülesannete lahendamine, 50% teooria. Eksam annab 70% aine koondhindest.

Eksami ajal näidatakse projektoriga töös vajalike jaotuste täiendkvantiilide tabelit.

Näidiseksam.pdf

Studenti test

Jätkame konspektist. Analüüsime Tartu temperatuuride andmeid.

Tõenäosusteooria ja statistika

LOFY.03.056

Erko Jakobson, PhD

2013 kevad

Kokkuvõtvalt seni vaadatud testidest

1. Pearsoni X²-test: $$\mathrm{X}^2 = \sum_{i=1}^{k} \frac{\left(n_i - \tilde{n}_i\right)^2}{\tilde{n}_i}$$

α = CHITEST(χ²; v).

H0 juurde (valim allub võrreldavale teoreetilisele jaotusele) tuleb jääda, kui α > 0,05.

2. F-test dispersioonide võrdlemiseks: $$F = \frac{s_1^{\;2}}{s_2^{\;2}}$$

α = FDIST(*F*; N_1 – 1; N_2 – 1).

H0 juurde (dispersioonid on ühesugused) tuleb jääda, kui α > 0,05.

Kokkuvõtvalt seni vaadatud testidest

3. Studenti t-test:

α = TDIST(t;v;tails).

**H0 juurde (üldkogumite keskväärtused on võrdsed) tuleb jääda,
kui α > 0,05.**

Näidiseksami lahendused