

Jumalat ma usun. Kõik ülejäänud peavad esitama andmed.

W. Edwards Deming,
USA statistikaproffessor

SATISTIKA

TES0020

Õppekava:
Rakenduslik majandusteadus TAAB

AKO SAUGA

Tänane loeng

- Sissejuhatus (ülevaade aineprogrammist)
 - Aine eesmärgid, sisu
 - Õpimeetodid, õppetöö korraldus
 - Õppematerjalid
 - Hindamine
- Sissejuhatus statistikasse
 - Mõõtmine ja mõõteskaalad
 - Andmekogumismeetodid
 - Statistiliste vaatluste liigitus
 - Vead mõõtmisel
 - Andmete esitus

Ako Sauga lühi CV

- Haridus
 - 2006 PhD ökoloogias, Tallinna Ülikool
 - 2001 MSc füüsikas, Tallinna Ülikool
 - 1983 diplom teoreetilise füüsika erialal, Tartu Ülikool
- Amet
 - Statistika ja ökonomeetria dotsent
Majandusanalüüsi ja rahanduse instituut
 - 1996 - 2007 Audentese Ülikool, lektor, dotsent, IT osakonna juh.
 - 1983 - 1996 TTÜ Füüsika instituut, laborant, assistent
- Loengukursused
 - TTÜ-s statistika, ökonomeetria, finantsökonomeetria.
 - Audentese Ülikoolis (1996-2007) majandusmatemaatika, statistika, kvantitatiivsed meetodid majanduses, statistika ja tõenäosusteooria (IT erialale), modelleerimine ja optimeerimine (logistika erialale), statistika sotsiaalteadustele, informaatika.
 - Tallinna Ülikoolis füüsikaline maailmapilt, stohhastilised protsessid.

Ako Sauga kontaktandmed

- E-post ako.sauga@taltech.ee
- Ruum SOC-480
- Koduleht <https://www.sauga.pri.ee/>
- Individuaalsed konsultatsioonid
 - teisipäeviti 14:00 – 15:00;
 - registreerimine kodulehel www.sauga.pri.ee

**ÕPPEAINE EESMÄRGID, ÕPPEMEETODID,
ÕPPETÖÖ KORRALDUS**

Suurandmed

Iga päev koguneb maailmas andmeid
16 438 356 000 gigabaiti.

Igal aastal ca 40 ZB (zetabaiti) so $40 \cdot 10^{12}$ GB.

Kes oskavad andmetega midagi peale hakata, neil on

EELIS

Miks on vaja statistikat?

Selleks, et

- teha kokkuvõtteid tohutuist andmekogumest;
- teha paremaid otsuseid;
- vastata olulistele sotsiaalsetele küsimustele;
- hinnata kampaaniate, programmide, uuenduste tõhusust.

Õppeaine eesmärgid, õpitulemused

Statistika õpetamisega taotletakse, et õppija

- omab ülevaadet majandusteoorias ja ärinduses kasutatavatest statistika mõistetest ja meetoditest;
- oskab töödelda, esitada ja interpreteerida statistilisi andmeid;
- suudab analüütiliselt mõelda.

Peale ainekursuse läbimist üliõpilane:

- kasutab korrektselt kirjeldava ning järeldava statistika termineid ja meetodeid;
- sõnastab põhjendatult uurimisküsimused ja neist lähtuvalt valib sobivad andmeanalüüsi meetodid;
- analüüsib tarkvara abil statistilisi andmeid, kasutades kirjeldava ja järeldava statistika meetodeid;
- tõlgendab ja hindab kriitiliselt andmestikke ja statistilise andmeanalüüsi tulemusi;
- sünteesib ja sõnastab järeldused, tuginedes andmeanalüüsi tulemustele.

Õppeaine sisu

- Statistika põhimõisted. Vaatlusteooria alused.
- Keskmised, variatsioonnäitavud ja nende kasutamine.
- Indeksid ja indeksanalüüs.
- Tõenäosusteooria elemente.
- Tõenäosusjaotused (binoomjaotus, Poissoni jaotus, normaaljaotus, t -jaotus).
- Valikuuringud ja järelduste tegemine üldkogumi kohta.
- Hüpoteeside statistiline kontrollimine (t -test, χ^2 - test, F -test, dispersioonanalüüs).
- Korrelatsioon- ja regressioonanalüüs.
- Aegread: elementaaranalüüs, tasandamine, kompleksanalüüs ja prognoosimine.

Õppemeetodid

- Iga nädal üks teema. Detailne kalenderplaan on laiendatud ainekavas (vt Moodle).
- Loengud
 - põhinevad õpikul;
 - erinevad näited, analüüsid, tõlgendamine;
 - demod;
 - mõned tõestused, arvutused.
- Praktikumid, õppejõud **Jelena Matina**
 - praktikumides kaasa oma sülearvuti
 - ülesanded paberil ja Excelis;
NB! ei õpetata Excelit, vaid statistikat Excelis
- Iseseisev töö
 - töö õpikuga;
 - näidete uurimine;
 - interaktiivsed demod;
 - iseseisev ülesannete lahendamine nii paberil kui tabelarvutuses;
 - **testid ja kontrollülesanded** õpikeskkonnas Moodle.

Põhiõpik

A. Sauga "[Statistika õpik majanduseriala üliõpilastele](#)", TTÜ Kirjastus 2017, 766 lk.

Pdf versioon (2020, teine, parandatud väljaanne) ja õpiku juurde kuuluvad andmefailid TTÜ Raamatukogu [digikogus](#)

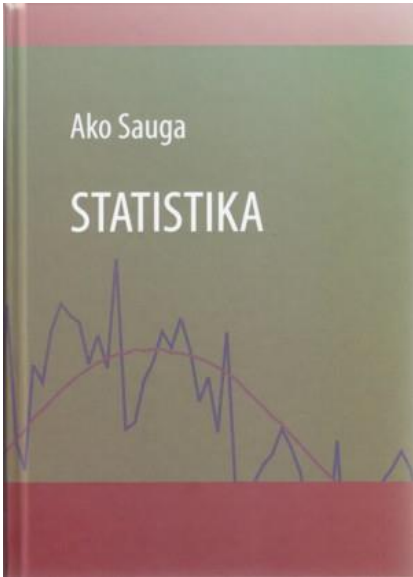
2017. a parim eestikeelne kõrgkooliõpik.

Õpikus on ligi 200 näidet, millest enamikuga on võimalik tutvuda ka õpiku juurde kuuluvates tabelarvutuse failides, kus on arvutused koos selgitavate kommentaaridega.

Lahendamiseks on üle 400 ülesande, millest ligikaudu pool tuleb lahendada tabelarvutuses, kasutades ülesannete failides olevaid andmeid.

Valdav enamik näidetest ja ülesannetes kasutatud andmetest pärineb reaalsest elust.

TTÜ Raamatukogus 69 eksemplari.



Töö õpikuga

Iseseisev lugemine: mõisted, definitsioonid, valemid, näited. Loengutele vastavate peatükkide numbrid on kirjas laiendatud ainekavas ja Moodles.



õpiku näite juures tähendab, et vastava näite andmed ja arvutused on olemas Excelis:
[paralleelselt tuleb lugeda õpikut ja uurida samu näiteid Excelis.](#)



koht õpikus, kus näpunäited vajalike statistika-alaste tabelarvutuse funktsioonide kasutamise kohta.

Õpiku lõpus on tabelarvutuse funktsioonide loetelu.

Iga peatüki lõpus on ülesanded. Ülesandeid on kaht tüüpi:

- paberil lahendamiseks;
- tabelarvutuses lahendamiseks (andmed failides).
- Õpiku lõpus on kõigi ülesannete vastused.
- Praktikumides lahendatakse läbi ca 100 ülesannet.
- Iseseisvaks lahendamiseks jääb ca 300 ülesannet.

Ülesanded paberil lahendamiseks

- Vajalikud, et paremini mõista statistiliste suuruste olemust, osata neid tõlgendada.
- Algul arvutus valemi järgi kalkulaatori abil.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \quad x_i = 3; 5; 8$$

- Siis alles Excelis, kasutades vastavat Exceli funktsiooni (põhineb samal valemil)

STDEV.P

x_i 100 erinevat väärtust

Küsimuste-vastuste foorum

Moodles

- Küsimusi võib esitada nii loengumaterjali kui ka ülesannete kohta.
- Küsimist ei tohi peljata: see näitab, et üliõpilane tegeleb õppematerjaliga.
- Vastata võivad kõik. Vastajad saava lisapunkte, mida arvestatakse lõpphinde panekul.

Iseseisvaks tööks vajalik tarkvara

- Mõni tabelarvutusprogramm

- MS Excel

- Kõik TTÜ üliõpilased saavad tasuta kasutada Microsoft Office paketti, vt

- <https://taltech.atlassian.net/wiki/spaces/ITI/pages/38994454/Microsoft+Office+kodukasutus+Microsoft+Office+private+usage>

- vabavara LibreOffice Calc

- <https://www.libreoffice.ee/>

- vabavara OpenOffice

- <http://www.openoffice.org/et>

Tabelarvutuse funktsioonid on neis ühesugused.

- Interaktiivsete demodega töötamiseks vajalik vabavara Wolfram Player

- <https://www.wolfram.com/player/>



Õppematerjalid Moodle

- <https://moodle.taltech.ee/course/view.php?id=262>
- Kursuse nimetus
TES0020 – Statistika (Ako Sauga; TAAB)
- Võtmesõna **mediaan**

Moodles on

- õpiku failid näidete ja ülesannetega;
- testid, kontrollülesanded;
- interaktiivsed demod;
- foorumid (teated õppejõult, küsimused-vastused, varia);
- kontrolltööd.

Testid ja kontrollülesanded

- Iga teema kohta test, kokku 15 testi.
 - Kokku annavad testid 15% lõpphindest.
 - Iga test on avatud **üks nädal**
 - avatakse esmaspäeval kell 8:00;
 - suletakse pühapäeval kell 23:59.
- 5 kontrollülesannet (teemad 4, 6, 10, 13 ja 15) Moodles.
 - Kokku annavad 10% lõpphindest.
 - Moodles on ülesande tekst ja saab alla laadida andmefaili. Excelis tuleb sooritada vastavad arvutused ning vastused kirjutada testi.
 - Vastuste kontrollimine käib automaatselt.
 - Ülesande esitamiseks on aega 2 nädalat.

Lähemalt vt laiendatud ainekavast.

Vajalik PIDEV TÖÖ semestri vältel!

Kontrolltööd

1. kontrolltöö (distantilt) on avatud Moodles K, 26.veebr.- P, 2. märts (5. õppenädal).

- Üks ülesanne, mille tekst kuvatakse Moodle testis.
- Leida etteantud Eesti Statistikaameti andmebaasi tabelist vajalikud andmed ning nende põhjal leida Excelis testis nimetatud näitaja (mõni keskmine, variatsioonnäitav vms).
- Arvuline vastus kirjutada Moodles vastavasse lünka. Vastust kontrollitakse automaatselt.
- Andmete otsimiseks ja arvutamiseks aega **1 tund**.
- Annab 5% lõpphindest.

2. kontrolltöö praktikumis 12. õppenädalal.

- 3 ülesannet, mille tekstid kuvatakse Moodle testis.
- Arvutused tuleb teha paberi peal, kasutada kalkulaatorit.
- Vastused kirjutada Moodles vastavasse lünka. Vastuseid kontrollitakse automaatselt.
- Ülesannete lahendamiseks aega 30 minutit.
- Annab 15% lõpphindest.

Lähemalt vt laiendatud ainekavast.

Eksam

- Eksam koosneb kahest osast:
 - Teooria test Moodles
 - valikvastustega küsimused;
 - abimaterjali kasutamine **ei ole ette nähtud**;
 - 15% lõpphindest.
 - Ülesannete lahendamine
 - andmeanalüüs Excelis;
 - tulemused, järeldused sisestada Moodle testi;
 - Exceli fail koos arvutustega üles laadida Moodles;
 - abimaterjali kasutamine **on lubatud**;
 - 40% lõpphindest.

Lõpphinne

Lõpphinde koostis:

Testid	15%
Kontrollülesanded	10%
1. kontrolltöö	5%
2. kontrolltöö	15%
Eksam	55% (15% test + 40% ülesanded Excelis)
KOKKU	100%

91% - 100%	suurepärase (5)
81% - 90%	väga hea (4)
71% - 80%	hea (3)
61% - 70%	rahuldav (2)
51% - 60%	kasin (1)
0% - 50%	puudulik (0)

KÜSIMUSI?

Statistika mõiste

- Arvandmed
- Praktiline tegevus
- Teadus

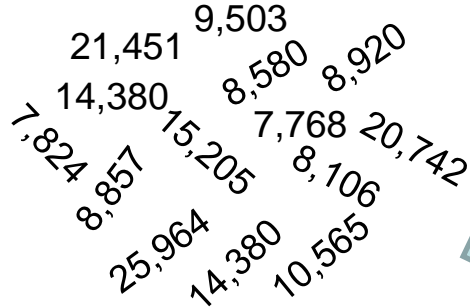
Statistika on teadus massnähtuste kvantitatiivse uurimise meetoditest.

0,43 0,12 3 34,9
5,2·10⁻⁴ 0,174 553 401 3,12 645
670 67 3 231 3400 2,8·10⁵
1000 1,3 98 54 25 500 000
57000 15,2 66 512 5780

Jagunemine

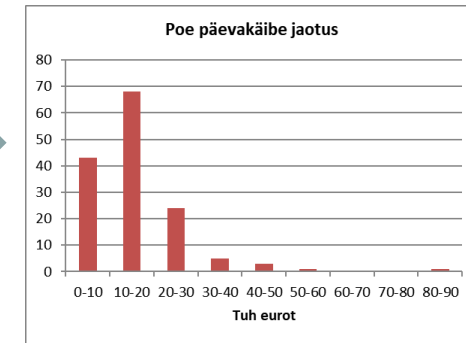
KIRJELDAV STATISTIKA

Statistilise informatsiooni kompaktne ja ülevaatlik esitamine: sobivad näitarvud, tabelid, diagrammid.



Klass	Sagedus
0-10	43
10-20	68
20-30	24
30-40	5
40-50	3
50-60	1

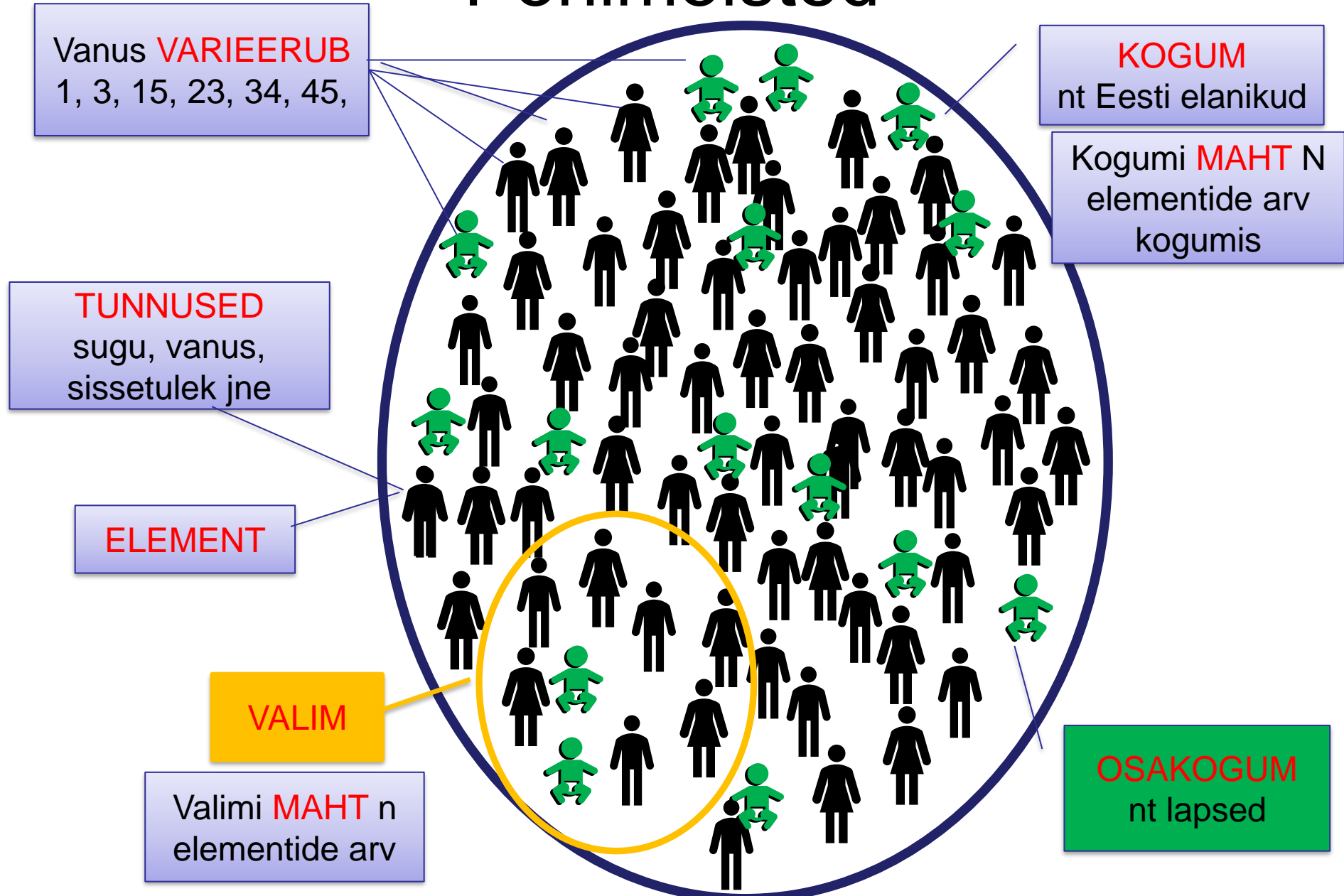
aritmeetiline keskmine 15,8 tuh €



JÄRELDAV STATISTIKA

Vaatlustulemusi kasutatakse hinnangute ja prognooside tegemiseks (veel) vaatlemata objektide ja situatsioonide jaoks. Meetodid põhinevad tõenäosusteoorial.

Põhimõisted



Mõõtmine

- Mõõtmine on objektide võrdlemine.
- Korraga saab võrrelda ainult **kaht** objekti omavahel.



- Kui objekte on palju, valitakse välja üks (etalon) ning teisi võrreldakse sellega.



Erinevad mõõtmistasemed

Juku ja Kalle on erineva pikkusega

erinevus

Juku on pikem kui Kalle

järjestus

Juku pikkus on 1,03 Kalle pikkust

arvuline suhe

INFORMATIIVSUS

Hindeskaalad

- Nimi- ehk nominaalskaala
 - objektide **eristamiseks**;
 - sugu, rahvus, huvid, kaubakood, ettevõtte tegevusala.
- Järjestusskaala
 - võimaldab objekte **järjestada** mingi tunnuse alusel;
 - nt ettevõtted: väikesed, keskmised, suured;
 - küsitlus: "poolt", pigem poolt kui vastu", "pigem vastu kui poolt", "vastu,,";
 - intervallid skaalajaotuste vahel **pole võrdsed**.
- Intervallskaala
 - skaalajaotuste intervallid on võrdsed.

Nimiskaala numbriline kodeerimine

Olgu meil 5 ettevõtet.

Ettevõte	Tegevusala
A	Info ja side
B	Ehitus
C	Ehitus
D	Info ja side
E	Töötlev tööstus

Andmebaasis nimiskaala väärtused kodeeritakse numbritega.

Info ja side 1
Ehitus 2
Töötlev tööstus 3

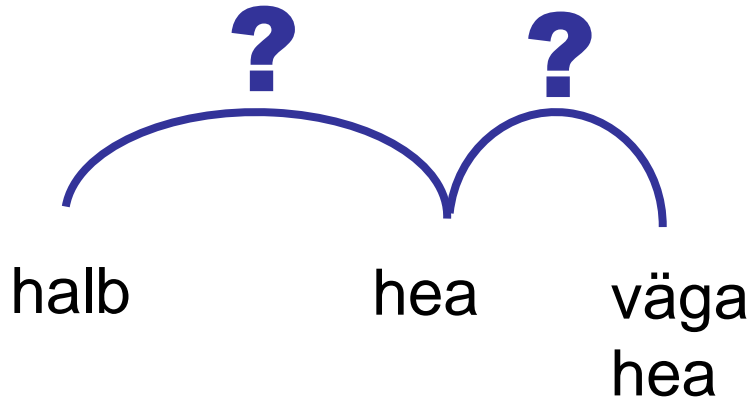
Ettevõte	Tegevusala
A	1
B	2
C	2
D	1
E	3

Kas saab leida tegevusala aritmeetilist keskmist?

Kas saab leida tegevusala aritmeetilist keskmist?

numbriline kood \neq arv

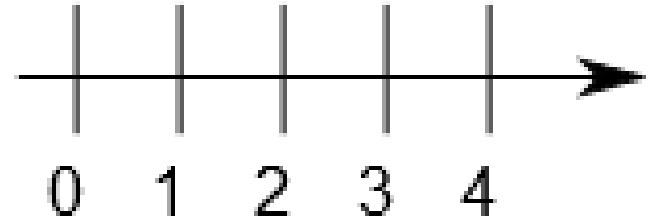
Järjestus- ja intervallskaala



Järjestusskaala

halb < hea < väga hea

hea - halb \neq väga hea - hea



Intervallskaala

$$2 - 1 = 3 - 2$$

Järjestusskaala numbriline kodeerimine

puudulik kasin rahuldav hea väga hea suurepärane

Kas saab leida aritmeetilist keskmist?

Kodeerimine

0

1

2

3

4

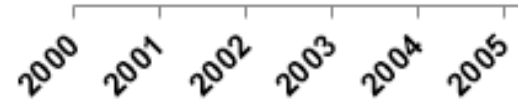
5

Kas saab leida aritmeetilist keskmist?

Intervalliskaalade jagunemine

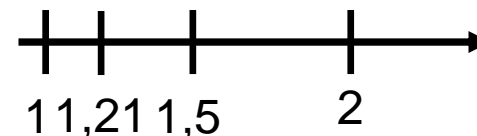
Kuidas nullpunkt määratud?

- Vahemikskaala
 - nullpunkti asukoht kokkuleppeline;
 - ajaskaala, Celsiuse skaala temperatuuri mõõtmiseks,
 - võib leida vahesid, ei tohi leida suhteid.
- Suhteskaala
 - nullpunkt fikseeritud absoluutselt;
 - objekti pikkus, kaal, töötajate arv, käive, mingi tegevuse jaoks kulunud aeg.



Kuidas arväärtused esinevad?

- Diskreetne skaala
 - näiteks loendamisel saadud 0, 1, 2, 3, 4
- Pidev skaala
 - iga kahe arvu vahel võib leida veel arvusid.



Kvalitatiivsed ja kvantitatiivsed tunnused

- Nimiskaalas ja järjestusskaalas mõõdetud tunnused – **kvalitatiivsed** tunnused.
- Intervallskaalas mõõdetud tunnused – **kvantitatiivsed** tunnused.
- Kvantitatiivse tunnuse võib teisendada järjestus- või nimitunnuseks.
 - NB! Kaotame informatsiooni!
- Vastupidine pole võimalik.

Statistiliste näitajate ning analüüsimeetodite **valik** sõltub sellest, millist skaalat on tunnuse mõõtmiseks kasutatud.

Milliste skaaladega on tegemist?

- Sugu: mees
- Perekonnaseis: abielus
- Laste arv: 2
- Haridus: kesk
- Vanus: 25 aastane
- Amet: lukksepp
- Kategooria: III
- Telefoni nr: 666-7777

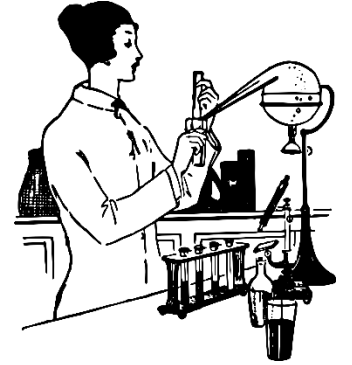
Vastusevariandid:

nimiskaala,
järjestusskaala,
intervallskaala

Andmekogumismeetodid

- Eksperiment

- Manipuleeritakse ühe või mitme tunnusega (**sõltumatud tunnused**) ja mõõdetakse **sõltuva** tunnuse väärtusi.



- Vaatlus

- Uurija ei sekku.
- Kui huvipakkuvate tunnustega manipuleerimine on praktilistel või eetilistel kaalutlustel võimatu.
- **Põhiline meetod** majandus-, äri- ja ühiskonnastatistikas.



Vaatluste liigitus I

- Otstarbe järgi:

- primaarsed,
- sekundaarsed.

- Andmete hankimise viisi järgi:

- otsene vaatlus
- küsitlus

- suuline
- ankeetvaatlus
- korrespondentvaatlus

- dokumentaalvaatlus (andmed kirjalikult või elektroonselt, näit onlain andmebaas)



Vaatluste liigitus II

- Vaatlusobjekti hõlmamise ulatuse järgi
 - monograafiline,
 - võrdlev-monograafiline;
 - äärmuste vaatlemine (parim ja halvim ettevõtte);
 - põhimassi vaatlus;
 - hõlmatakse peamine osa, kõrvaline jäetakse välja;
 - väljavõtteline vaatlus;
 - väljavõtukogum ehk valim;
 - kõikne statistika.
- Vaatluse sageduse järgi
 - pidev (mingi aja jooksul pidevalt);
 - korduv (kindla ajavahemiku järel);
 - ühekordne.

Vead mõõtmistel

Tahtlikud (andmetega manipuleerimine)

- Kreeka võltsis eelarvedefitsiidi suurust, et 2000. a. liituda euroalaga;
- Ungari valitsus esitas ebaõigeid majandusandmeid enne 2006. a. valimisi;
- Swissair esitas 2000. aasta aruandes ebaõigeid andmeid. Aastal 2001 pankrot.

Mittetahtlikud

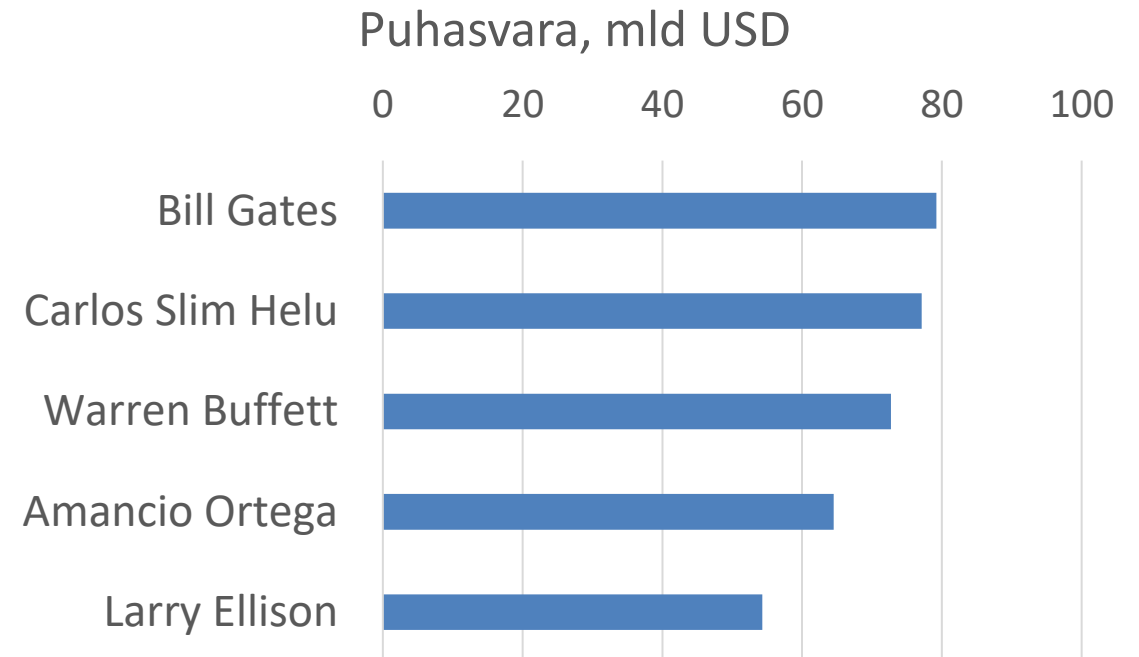
- Süstemaatiline viga
 - ebatäpne mõõtmisvahend (kell, kaal);
 - halvasti sõnastatud küsimus ankeedis.
- Juhuslik viga
 - mõjutab mõõtmistulemust kord ühes, kord teises suunas.
- Ekse
 - jäme viga, enamasti põhjustatud inimlikest eksimustest;
 - näiteks jäeti sisestamata üks arvus esinev 0 (või on üks 0 ülearu).

Andmete esitamine

Andmete esitamiseks ja visualiseerimiseks kasutatakse **tabeleid** ja **diagramme** (arvjooniseid).

Näide: maailma viis rikkaimat inimest 2015. a algul.

Isik	Puhasvara, mld USD
Bill Gates	79,2
Carlos Slim Helu	77,1
Warren Buffett	72,7
Amancio Ortega	64,5
Larry Ellison	54,3

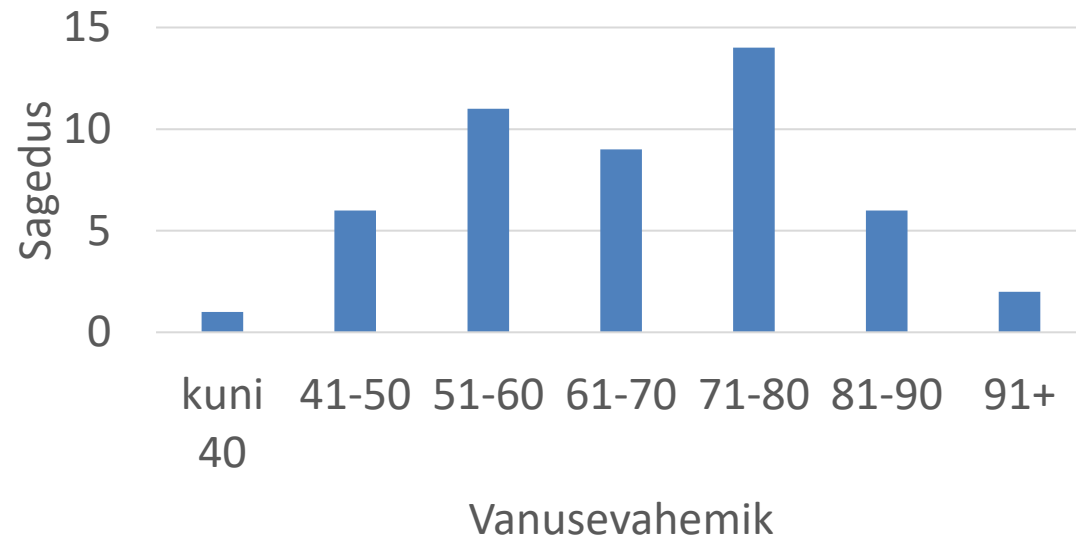


Sagedusklassid

Kui erinevaid väärtusi väga palju, jaotatakse need **sagedusklassidesse**.

Näide: maailma 50 kõige rikkama inimese vanuseline jaotus 2015. a algul.

Vanusevahemik ehk klass	Sagedus
kuni 40	1
41-50	6
51-60	11
61-70	9
71-80	14
81-90	6
91+	2

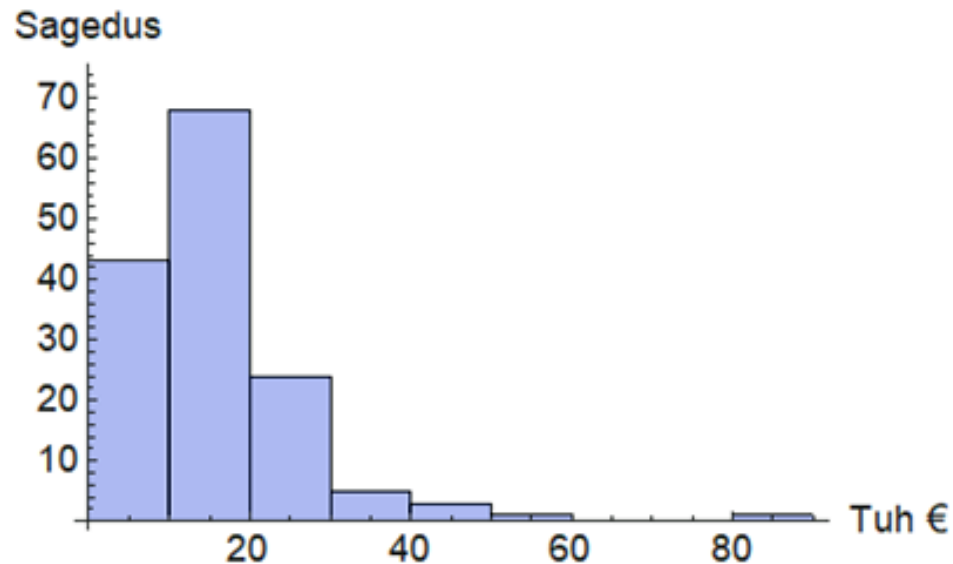


Histogramm iseloomustab intervallskaalas mõõdetud tunnuse jaotust.

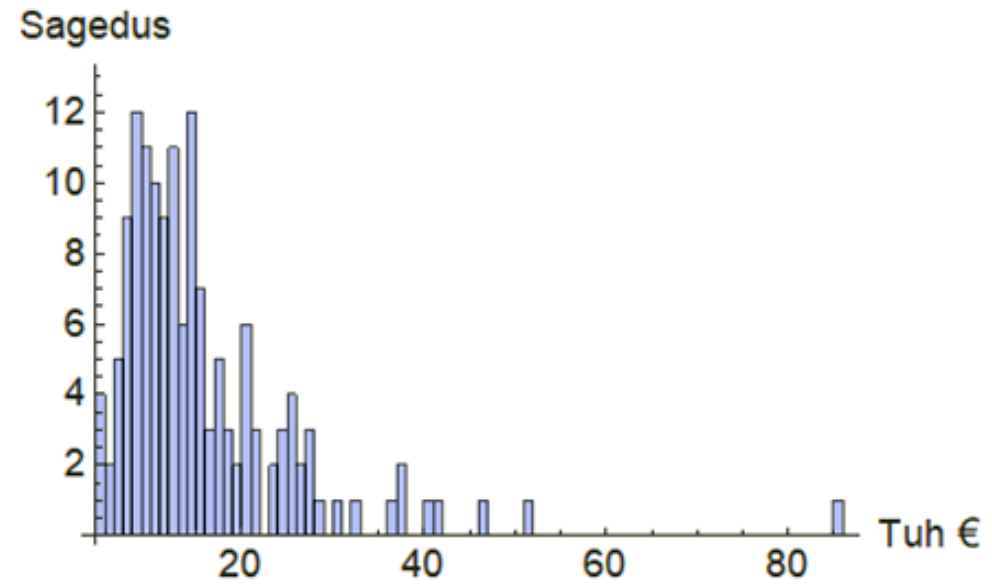
Histogramm ja klassi laius

Histogrammi kuju sõltub klassi laiusest (klasside arvust).

Näide: poe päevakäibe jaotus. 145 päeva, minimaalne 4,1 tuh €, maksimaalne 85,3 tuhat €.



Klassi laius 10 tuh €, klasside arv 9.



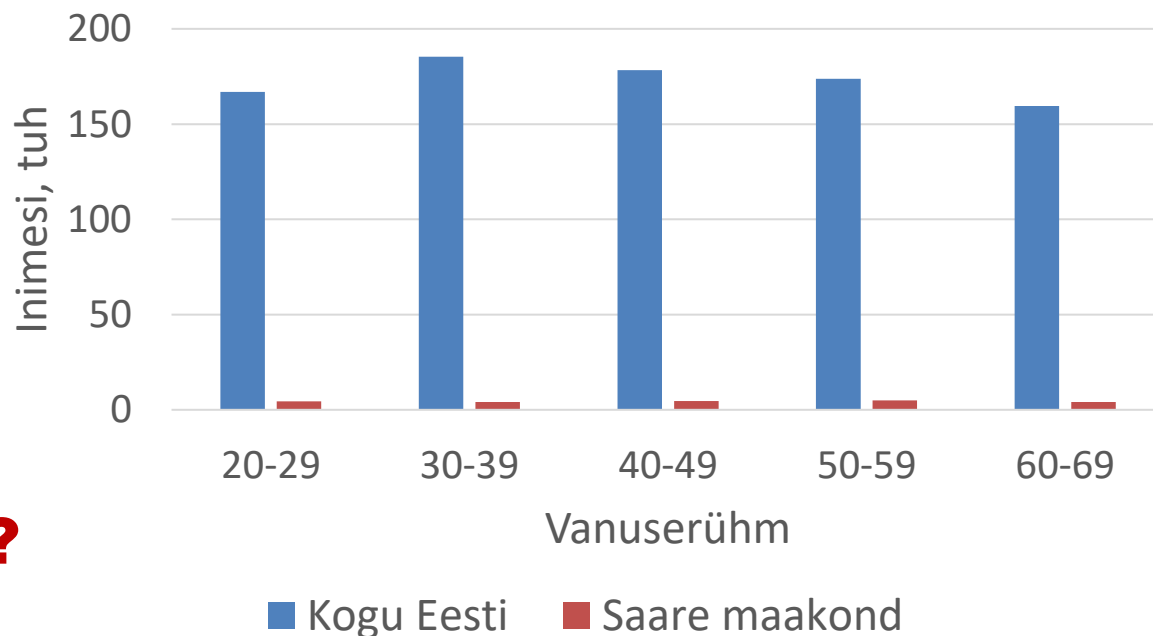
Klassi laius tuhat €, klasside arv 90.

Sageduste võrdlemine, probleem

Vanuserühm	Kogu Eesti	Saare maakond
20-29	166965	4493
30-39	185325	4033
40-49	178287	4624
50-59	173784	4983
60-69	159496	4039

Võrdleme Saare maakonna 20–69 aastaste elanike vanuselist jaotust kogu Eestiga. Aasta 2017.

Mis probleem on selle diagrammiga?

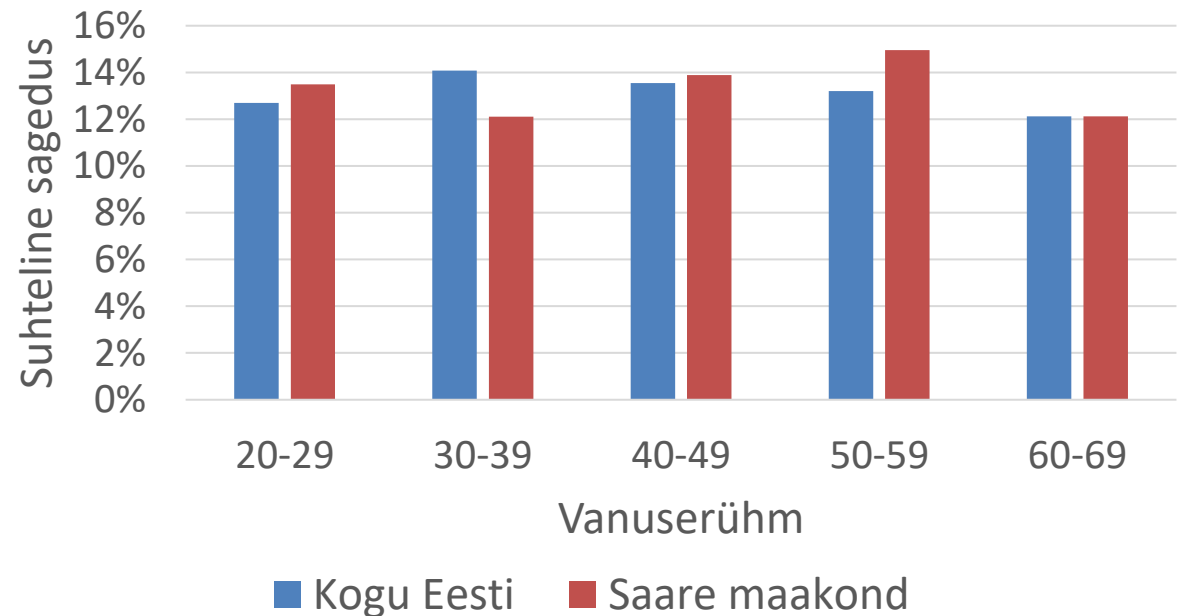


Kuidas probleemi lahendada?

Suhtelised sagedused

Vanuserühm			Suhtelised sagedused	
	Kogu Eesti	Saare maakond	Kogu Eesti	Saare maakond
20-29	166965	4493	12,7%	13,5%
30-39	185325	4033	14,1%	12,1%
40-49	178287	4624	13,6%	13,9%
50-59	173784	4983	13,2%	15,0%
60-69	159496	4039	12,1%	12,1%
ELANIKKE KOKKU	1315635	33307		

Selline diagramm sobib.



Vale diagrammi tüüp

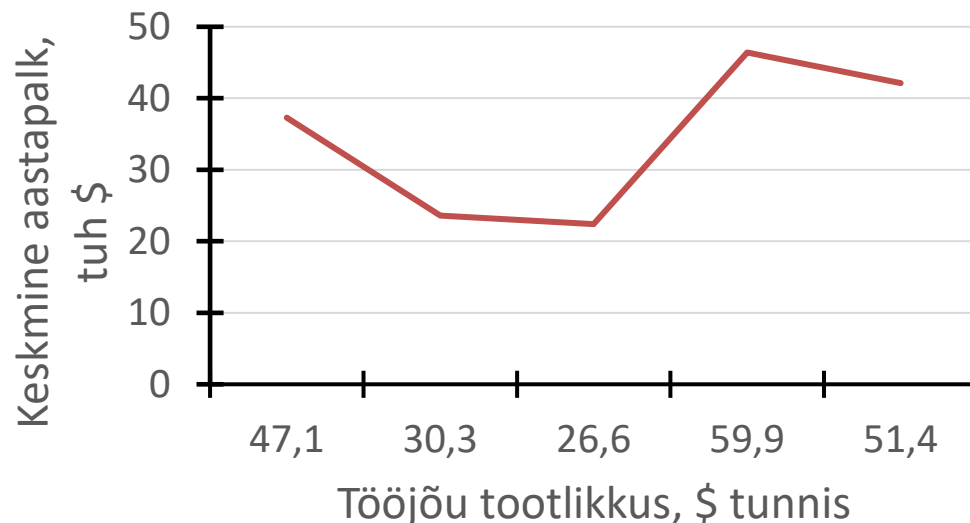
Aasta 2016, OECD andmebaas <https://data.oecd.org>

Riik	Tööjõu tootlikkus, eurot tunnis	Keskmine aastapalk, tuh \$
Hispaania	47,1	37,3
Eesti	30,3	23,6
Läti	26,6	22,4
Saksamaa	59,9	46,4
Soome	51,4	42,1

Soovime uurida, kuidas keskmine palk on seotud tööjõu tootlikkusega.

Mis probleem on selle diagrammiga?

Põhjuseks vale diagrammi tüüp: joon (*line*).



Õige diagramm

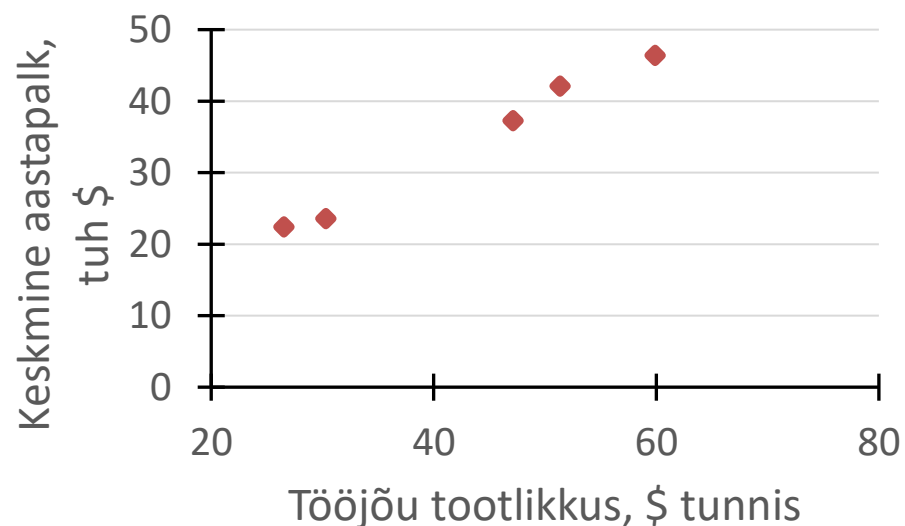
Aasta 2016, OECD andmebaas <https://data.oecd.org>

Riik	Tööjõu tootlikkus, eurot tunnis	Keskmine aastapalk, tuh \$
Hispaania	47,1	37,3
Eesti	30,3	23,6
Läti	26,6	22,4
Saksamaa	59,9	46,4
Soome	51,4	42,1

Soovime uurida, kuidas keskmine palk on seotud tööjõu tootlikkusega.

Näha: kui tootlikkus suurem, on ka palk suurem.

Excelis on X-telg arvteljena ainult hajumisdiagrammil (*scatter*).



Lähem tutvumine arvjooniste ehk diagrammidega on tabelarvutuse failis

ÜL01Arvjoonised.