



TALLINNA ÜLIKOOL

Andmeanalüüs: statistiline andmestik ja kirjeldav statistika

IFI7041

Loeng: Kairi Osula
Seminar: Taivo Tuuling

Kursuse korraldus

- **Loeng** – 6 nädalat
 - 31.01
 - 7.02
 - 14.02
 - 21.02
 - 28.02
 - 7.03
- Loengu slaidid ja muud materjalid on avalikus veebis:
www.tlu.ee/~kairio/7041



TALLINNA ÜLIKOOL

Kirjeldav statistika

Kursuse korraldus

- **(OK) Kirjalik avatud küsimustega test 40%**
 - **(OK) Rühmatöö 10%**
 - **(TT) Kodused (is.) tööd 50%**
- „A“ - suurepärase 91-100%
„B“ - väga hea 81-90%
„C“ - hea 71-80%
„D“ - rahuldav 61-70%
„E“ - kasin 51-60%
„F“ - puudulik 0 – 50%
- Positiivse hinde saamiseks peavad olema tehtud pos.tulemusele mõlemad osad (test & is.töö)
 - Hinded sisestab praktikumi juhendaja Taivo Tuuling



TALLINNA ÜLIKOOL

Kirjeldav statistika

Kursuse korraldus

- **10%** - rühma ettekanne ühes loengus
- NB! Kõik rühmad ei mahu viimasesse loengusse!!!
- Ühes rühmas 4-5 üliõpilast
- **Teema** – huvitav küsimus, mis loengus tekkis.
- Ettekande aeg 4-5 minutit.
- Slaidid pdf ja saata mulle e-mailile (kairio@tlu.ee)
- **Seda osa saab järgi vastata, aga siis olete te üksi rühmas ja mina annan huvitava teema!**
- **40%** - avatud küsimustega kirjalik test
- Põhieksamitel on **materjalide kasutamine lubatud**
 - 25.04 & 23.05
 - Kõik mis on paberandjal, arvutis on avatud 1 fail (slaidid), telefone ei kasuta
- **Järeleksamitel** me materjale ei kasuta
- Avatud küsimused, iga kord erinevad, mitu varianti
- **Kirjutamiseks aega 1 tund**
- Tulemused kahe päeva jooksul, tööd saab pärast näha (eelneval kokkuleppel)



TALLINNA ÜLIKOOL

Kirjeldav statistika

Lisalugemist

- "Uuri ja kirjuta"
 - Autorid: S. Hirsjärvi, P. Remes, P. Sajavaara
 - Kirjastus: Medicina
 - Ilumisaasta: 2005
- "Uurija käsiraamat"
 - Autor: A.Kidron
 - Kirjastus: Mondo
 - Ilumisaasta: 2008



TALLINNA ÜLIKOOL

Kirjeldav statistika

Lisalugemist

- "Andmete analüüs ja tõlgendamine sotsiaalteadustes"
 - Autor: Liina Mai Toeding
 - Kirjastus: Tartu Ülikooli kirjastus
 - Ilumisaasta 2007
- "Andmeanalüüs sotsiaalteadustes"
 - Autor: Liina Mai Toeding
 - Kirjastus: Tartu Ülikooli kirjastus
 - Ilumisaasta: 1998

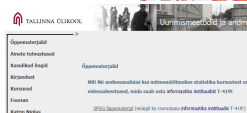


TALLINNA ÜLIKOOL

Kirjeldav statistika

Lisalugemist

- "Statistilise andmetöötluse pakett SPSS 14.0" Põhikursus
 - Autor: Katrin Niglas
 - Kirjastus: Tallinna Ülikooli kirjastus
 - Ilumisaasta: 2008



TALLINNA ÜLIKOOL

Kirjeldav statistika

Uuringutüpe saab klassifitseerida järgmiselt:

Eesmärgi
(purpose) alusel

- Avastav
- Kirjeldav
- Analüütiline
- Ennustav

Uuringu
lähenemise
(approach) alusel

- Teoreetiline
- Empiiriline
- Rakenduslik

Protsessi alusel

- Kvalitatiivne
- Kvantitatiivne
- Kombineeritud



TALLINNA ÜLIKOOL

Kirjeldav statistika

Protsessi alusel

Protsessi alusel

- Kvalitatiivne
- Kvantitatiivne
- Kombineeritud

KVANTITATIIVNE UURING

- Kui palju? (probleemi kirjeldamine)
- Miks? (probleemi põhjuste tuvastamine)

- Tulemused üldistatakse üldkogumile, mõnikord küsitav (või mittevajalik)
- Üldistamine tugineb arvulisele argumentile

- Seoste uurimine (ei näta põhjuslikku seost)
- Hüpoteeside testimine
- Eelneb mahukas planeerimisfaas

KVALITATIIVNE UURING

- Kuidas? (probleemi kirjeldamine)
- Milleks? (probleemi põhjuste tuvastamine)

- Tulemuste analüüs tömahukas
- Uuritavate hulk väike (üldistamine pigem ennustamise, seaduspärasuste/trendide väljatoomine)
- uuritakse tõlgendusi, hoiakuid ning arvamusi (koos põhjendustega)
- Saab töötada välja efektiivseid lahendusi konkreetsetele isikutele, mille laiemat mõju saab hiljem uurida kvantitatiivsete meetoditega



TALLINNA ÜLIKOOL

Kirjeldav statistika

Uuringutüpe saab klassifitseerida järgmiselt:

Eesmärgi
(purpose) alusel

- Avastav
- Kirjeldav
- Analüütiline
- Ennustav

Uuringu
lähenemise
(approach) alusel

- Teoreetiline
- Empiiriline
- Rakenduslik

Protsessi alusel

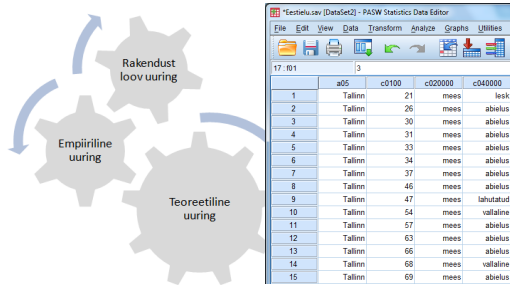
- Kvalitatiivne
- Kvantitatiivne
- Kombineeritud



TALLINNA ÜLIKOOL

Kirjeldav statistika

Uuringu lähenemise põhjal



TALLINNA ÜLIKOOL

Kirjeldav statistika

Teoreetiline uurimus

Eesmärk Olemasoleva teabe/uuringute analüüs
Erinevate teooriate kriitiline analüüs (võrreldakse ühe eeliseid teistega)

Autor peab näitama, milline on tema panus uute teadmiste otsingul, kasutamisel, süstematiseerimisel ja hindamisel.

Näited Filosoofilised küsimused
Puhta matemaatika teoreemid
Ajaloolised uuringud
Arvutikasutuseetika (loogiline arutelu ja seisukohad).
Tarkvara võrdlev analüüs

Referaat (?)
Seminaritöö kui referaat
(teoreetiliste uuringute KÕIGE välismise kiht)



TALLINNA ÜLIKOOL

Kirjeldav statistika

Rakenduslik uurimus

Näited Personaalne arendusprojekt (või selle osa)
Õppematerjalide loomine

Probleemi analüüs (vajadused, eesmärgid, olemasolev teave)

Disaini protsess (tööjaotus, ajakava, meetodid)

Disaini protsessi resultaat (rakenduse visandid, vaheversioonid, lõplik rakendus)

Hindamine e. evalvatsioon (rakenduse testimine, hindamine lähtuvalt standarditest, kasutajate tagasiside)

ADDIE mudel



TALLINNA ÜLIKOOL

Kirjeldav statistika

Empiiriline uuring

- Uurimisprobleem
 - küsimus, hüpotees, eesmärk
- Valikumeetod
 - juhuslik valim, üks juhtum, mitu juhtumit
- Andmekogumis meetod(id)
 - struktureeritud ankeet, struktureerimata intervjuu, ..
- Andmeanalüüsi meetodid
 - statistilised meetodid, kodeerimine
- Tulemused/järeldused
 - kirjeldused, empiirilised üldistused, seaduspärasused,...



K. Niglas



TALLINNA ÜLIKOOL

Kirjeldav statistika

Uuringutüüpide omavaheline seotus

- Igas uuringu põhiskeemis on teatud alamtappid, mis vajalik rakendada kõrvalolevatele uuringutüüpidele omaseid mõtlemis- või tegutsemisviise.
 - näiteks on hea arendusuuringu lahutamatuks osaks
 - valdkonnaga seotud teooriate läbitöötamine
 - empiirilise andmestiku kogumine
 - ja analüüs vajaduste selgestegemise ja/või rakenduse testimise etapis



TALLINNA ÜLIKOOL

Kirjeldav statistika

Empiiriline uuring

- Uurimisprobleem
 - küsimus, hüpotees, eesmärk
- Valikumeetod
 - juhuslik valim, üks juhtum, mitu juhtumit
- Andmekogumis meetod(id)
 - struktureeritud ankeet, struktureerimata intervjuu, ..
- Andmeanalüüsi meetodid
 - statistilised meetodid, kodeerimine
- Tulemused/järeldused
 - kirjeldused, empiirilised üldistused, seaduspärasused,...



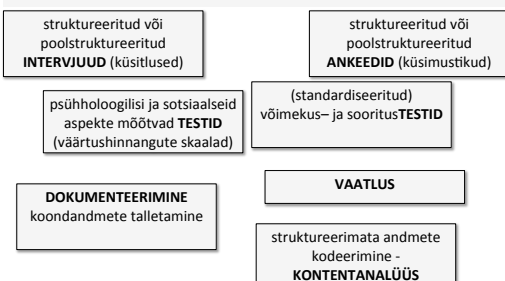
K. Niglas



TALLINNA ÜLIKOOL

Kirjeldav statistika

Andmekogumismeetodid



TALLINNA ÜLIKOOL

Kirjeldav statistika

Pilootuuring

- Testib idee toimet **reaalsetes tingimustes**, kuid **vähendatud mahus**.
 - Küsimustiku uuringueelne kontrollimine
 - Vähendab probleemide ja vigade tekkimist reaalse andmekogumise käigus
 - Vähe vastajaid

N: reklaamlakatid, loterii, soodushind

esmalt ühes kaupluses, reklaami väljapanek ühes geograafilises piirkonnas

Nende põhjal tehakse omakorda järeldused täiendamisevajaduse ning edasise rakendamise kohta kogu sihtturul.



TALLINNA ÜLIKOOL

Kirjeldav statistika

Küsimustik

- **Hea küsimustik on:**
 - selge sõnastusega,
 - kergesti ja üheselt mõistetavate küsimustega,
 - kompaktna ja kiiresti vastatav,
 - vormistusest korrektne;
 - koostatud nii, et oleks minimeeritud vastajate ja andmete töötajate poolt potentsiaalselt tehtavate vigade hulk.
- **Hoiduda tuleks sellistest küsimustest nagu:**
 - suunavad küsimused;
 - teaduslikult täpse, kuid pika ja keeruka sõnastusega küsimused;
 - mitmeti mõistetavad küsimused;
 - ärritavad küsimused (sh küsimusega mittesobivad vastusevariandid).

Küsimustik

- **Küsimuste/mõõdikute kavandamisel mõtle ja otsi infot järgneva kohta:**
 - kas saab uuritavat nähtust mõõta otse või läbi indikaatori(te)?
 - kas antud nähtust on eelnevates uuringutes mõõdetud ning kas vastav kirjandus on usaldusväärne?
 - kas saab juba olemasolevaid mõõdikuid kohandada või tuleb välja töötada uued mõõdikud/küsimused?
 - (pööra tähelepanu: kultuurilised ja kontekstuaalsed erinevused; võrreldavus eelnevate uuringute tulemustega, jne)
 - kas peaks kasutama piloteerimist ning järelkontrolli?

Üldine skeem

1. Palun vastata 5-palli skaalal, kus 1 – ei nõustu üldse ja 5 – nõustun täielikult	1	2	3	4	5	Ei oska vastata
Hommikult tuleb kiiresti hea tundega						
Mina lootsin, et meeldib õppida						
Mul on kaeveldamisel hea tunne						
Kiisli õppimise ja sportimise vahel on hea lõõgastamine						
Õpetajad on õpetamisele aetud						

2. Mingeid vabaid, millega alati nõust

Õpetajad näevad oma vabad	
Õpetajad oskavad teha hea teha enda	
Õpetajad ei ole alati hea	
Õpetajad on õpetajad	
Õpetajad hoolivad õpilastest	

3. Töö näide, kuidas õpetajad on sinu arengut toetanud

4. Mida soovid veel teada

Üldised andmed vastaja kohta

Olet: _____

Õpet: _____

Kui oled nõus, et need juhtumid saad seostada sinu arenguga ja ettepanekuid sinu osaka, kirjuta neid

Millest sõltub andmeanalüüsimetodi valik?

Küsimuse tüübib

- Uurimisküsimus: laiem
- Analüüsküsimus: nt. kas kaks gruppi on erinevad/seotud?

Andmete tüübib (väärtuste järjestatavus, skaalavahemike võrdsus)

- **Nimitunnused**
 - Nimitunnuse väärtuseid ei saa järjestada, järjestustunnusel saab
- **Järjestustunnused**
 - Arvunnuse skaalavahemikud on võrdsed, järjestustunnusel mitte
- **Arvtunnused**
 - Arvtunnuse korral saame arvutada keskvaartust, st.hälvet; binaarse tunnuse korral mitte
- **Binaarsed tunnused**

Sihtrühmast

- Uurija teadmised/oskused
- Kellele esitab, kuidas?

Tunnuse tüübid

Arvestame
järjestatavust, skaalavahemike võrdsust

Nimitunnused

Nimitunnuse väärtuseid ei saa järjestada

Järjestustunnused

Järjestustunnuste väärtuseid saab
järjestada
Skaalavahemikud ei ole võrdsed

Intervalltunnused

Skaalavahemikud on võrdsed
Vähe erinevaid väärtuseid
Palju võimalikke väärtuseid

Binaarsed tunnused

Kaks võimalikku väärtust
Järjestatavus ja skaalavahemike võrdsus ei ole probleem

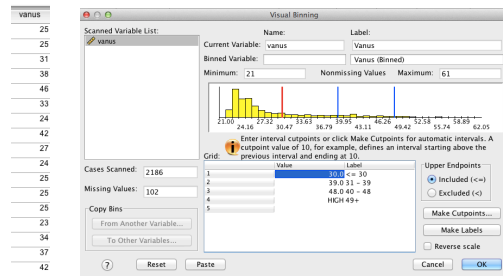
Eeltöö – andmestiku korrastamine

- Puuduvad väärtused - kui palju neid on, mida teha tühjade lahtritega?
- Andmesisestusvead
- Andmete grupeerimine (vajadusel)
- Skaalade pööramine (vajadusel)

Sugu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	mees	503	23.0	23.9	23.9
	naine	1604	73.4	76.1	100.0
	Total	2107	96.4	100.0	
Missing	System	79	3.6		
Total		2186	100.0		

Vanus



Kool

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Eesti Infotehnoloogia Koolid	22	1.0	1.0	1.0
	Eesti Kunstiakadeemia	39	1.8	1.8	2.8
	Eesti Maailikool	239	10.9	10.9	13.7
	Eesti Muusika- ja	26	1.2	1.2	14.9
	Teatriakadeemia				
	Lääne-Viru Rakenduskõrgkool	67	3.1	3.1	18.0
	Mainori Kõrgkool	96	4.4	4.4	22.4
	Sisekaitseakadeemia	56	2.6	2.6	24.9
	Tallinna Pedagoogiline Seminar	89	4.1	4.1	29.0
	Tallinna Tehnikakõrgkool	24	1.1	1.1	30.1
	Tallinna Tehnikakõrgkool	452	20.7	20.7	50.8
	Tallinna Tervishoiu Kõrgkool	60	2.7	2.7	53.5
	Tallinna Ülikool	449	20.5	20.5	74.1
	Tartu Kõrgem Kunstikool	12	.5	.5	74.6
	Tartu Ülikool	555	25.4	25.4	100.0
Total		2186	100.0	100.0	

Õppevaldkond

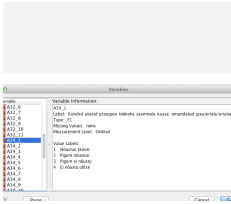
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Haridus	266	12.2	12.2	12.2
	Humanitaaria ja kunstid	230	10.5	10.5	22.7
	Sotsiaalteadused, ärinäus ja õigus	385	17.6	17.6	40.3
	Loodus- ja täppisteadused	319	14.6	14.6	54.9
	Tehnika, tootmine ja ehitus	271	12.4	12.4	67.3
	Põllumajandus	58	2.7	2.7	69.9
	Tervis ja haalo	225	10.3	10.3	80.2
	Tööstus	132	6.0	6.0	86.2
Total		2186	100.0	100.0	

Tegevusala

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Õppin	1164	53.2	53.2	53.2
	Õppin	213	9.7	9.7	62.9
	Õppin ja õppin	570	26.1	26.1	89.0
	ei tööta ega õpi, otsin aktiivselt tööd	62	2.8	2.8	91.8
	ei tööta ega õpi, ei otsi ka aktiivselt tööd	16	.7	.7	92.5
	olen lapsega kodus	111	5.1	5.1	97.6
	teenin aega kaitseväes	1	.0	.0	97.6
	muu	26	1.2	1.2	98.8
	õppin ja otsin aktiivselt tööd	13	.6	.6	99.4
	töötan ja planeerin edasist õpinguid	10	.5	.5	100.0
Total		2186	100.0	100.0	

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Väga oluline	339	15.5	20.8	20.8
	Kõige oluline	807	36.9	49.6	70.5
	Väheoluline	387	17.7	23.8	94.3
	Ei ole üldse oluline	92	4.2	5.7	100.0
	Total	1626	74.4	100.0	
Missing	System	560	25.6		
Total		2186	100.0		

Andmeanalüüsi vahendid

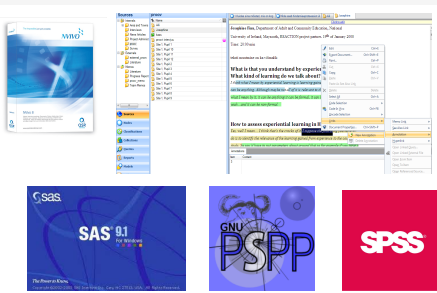


Kulvõrd aitasid praeguse töökoha saamisele kaasa: omandatud (peajerialalised)

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Nõustun täiesti	1016	46.5	82.3	82.3
Pigem nõustun	852	16.1	21.6	83.8
Pigem ei nõustu	146	6.7	8.9	82.8
Ei nõustu üldse	118	5.4	7.2	100.0
Total	1632	74.7	100.0	
Missing	254	25.3		
Total	1886	100.0		

TALLINNA ÜLICOOL Kirjeldav statistika

Andmeanalüüsi vahendid



TALLINNA ÜLICOOL Kirjeldav statistika

Andmeanalüüsi küsimus

KIRJELDAV (esmane analüüs)

- tekst
- tabel
- diagramm

VÕRDLEV (erinevused gruppide vahel)

KORRELATSIION (seosed tunnuste vahel)

TALLINNA ÜLICOOL Kirjeldav statistika

Esmane analüüs

- Eesmärk
 - uuritava nähtuse **süsteemiline kirjeldamine** protsentjaotuste ja keskvaartuste kaudu, tabelite ja graafikute vormis.
- Vaadeldakse vaid **üksiktunnuseid**
 - Kõrvale jäetakse tunnuste omavahelised seosed.

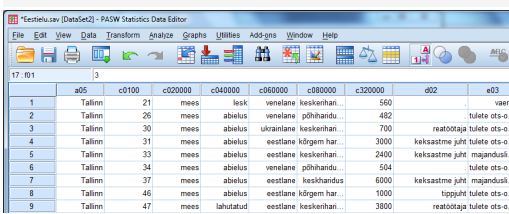
KIRJELDAV (esmane analüüs)

VÕRDLEV (erinevused gruppide vahel)

KORRELATSIION (seosed tunnuste vahel)

TALLINNA ÜLICOOL Kirjeldav statistika

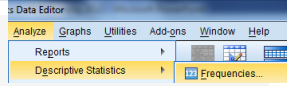
Esmane ülevaade andmetest



„Kui suur osa õpilasi kasutab Facebooki iga päev?“
 „Kas ja kui palju leidub neid õpilasi, kes Facebooki üldse ei kasuta?“
 „Mis on kõige tüüpilisem kasutussagedus ehk millise vastusevariantidest on valinud kõige suurem osa õpilastest?“

TALLINNA ÜLICOOL Kirjeldav statistika

Sagedustabel



- Frequency - vastajate arv
- Percent - osakaal
- Valid percent - osakaal mittepuuduvatest väärtustest
- Cumulative percent - kumulatiivne protsent

		vanuse_grupid			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	kuni 20	43	4.9	5.0	5.0
	21-30	147	16.7	17.0	21.9
	31-40	174	19.7	20.1	42.0
	41-50	162	18.4	18.7	60.7
	51-60	160	18.1	18.5	79.1
	üle 61	181	20.5	20.9	100.0
	Total	867	98.3	100.0	
Missing	System	15	1.7		
Total		882	100.0		

TALLINNA ÜLICOOL Kirjeldav statistika

Sagedustabel

Kas olete poliitikast ... ?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	väga huvitatud	36	4,1	4,1	4,1
	küllaltki huvitatud	308	34,9	35,5	39,6
	mitte eriti huvitatud	350	39,7	40,3	80,0
	üldse mitte huvitatud	174	19,7	20,0	100,0
	Total	868	98,4	100,0	
Missing	keeldus	2	,2		
	ei oska öelda	3	,3		
	System	9	1,0		
	Total	14	1,6		
Total		882	100,0		



TALLINNA ÜLIKOOL

Kirjeldav statistika

Tabeli (ridade) järjestamine

- Internetikasutuse osakaalud 2008.aastal erinevates riikides 6.-17.aastaste laste seas.

Tabel 2A.

Riik	Osakaal
Eesti	93%
Itaalia	45%
Kreeka	50%
Leedu	86%
Läti	83%
Rootsi	91%
Soome	94%
Ungari	88%

Tabel 2B.

Riik	Osakaal
Soome	94%
Eesti	93%
Rootsi	91%
Ungari	88%
Leedu	86%
Läti	83%
Kreeka	50%
Itaalia	45%

- Juhul kui tabelis toodud kategooriad ei ole sisuliselt tähenduslikus järjekorras, siis järjestatakse tabeli read sageduste/osakaalude järgi (Tabel 2B).



TALLINNA ÜLIKOOL

Kirjeldav statistika

Järelduse koostamine

- Tabeli 10. põhjal näeme, et 15 vastajat (38,5%) hindas ettevalmistust ebapiisavaks; 9 vastajat (23,1%) enam-vähem piisavaks ning 15 vastajat (38,5%) täiesti piisavaks.

Tabel 10. Hinnang ettevalmistusajale

	Arv	Protsent
ei olnud piisav	15	38,5%
enam-vähem	9	23,1%
täiesti piisav	15	38,5%
KOKKU	39	100,0%



TALLINNA ÜLIKOOL

Kirjeldav statistika

Järeldustes...

- Ei tohi liialdada ebamääraste väljenditega

- Enamasti
- Sageli
- Suuremas osas
- Harva
- Mõnikord
- Kohati.

Need tekitavad küsimusi, kui sageli, kui harva, mis tingimustel jne.

- Paremad on täpsemad väljendid

- alla poole (46%)
- ligi kolmandikul juhtudest
- peaaegu kolmveerand näidetest jne.



TALLINNA ÜLIKOOL

Kirjeldav statistika

Statistiliste andmete esitamine

- Üldine reegel

Vali esitluseks **diagramm**, kui soovid eelkõige anda kiiret ülevaadet üldtendentsi(de)st ja suundumus(te)st

TEKSTI SEES TOODUD ARVUDENA

Esitlusviis peaks toetama parimat viisi tulemuste sisust kiiret ja õiget arusaamist ning olema kompaktne.

Vali arvulise info edastusviisiks **tekst**, kui korraga on vaja esitada vaid üks-kaks arvulist näitajat

ARVJONONISE e DIAGRAMMINA

TABELINA

Vali esitluseks **tabel**, kui on vajalik anda edasi täpset arvulist infot või kui võrreldavate arvnäitajate suurusjärgud on väga erinevad

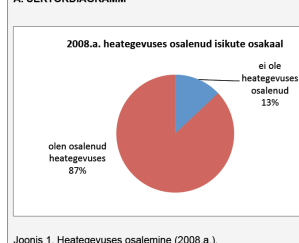


TALLINNA ÜLIKOOL

Kirjeldav statistika

TNS Emori läbiviidud heatugevusliku käitumise uuringu tulemused.

A. SEKTORDIAGRAMM



B. TEKST

2008.a. TNS Emori läbiviidud uuringus heatugevuskultuuri käitumisest selgus, et 87% vastajatest on viimase aasta jooksul heatugevusega ühel või teisel moel kokku puutunud.

Milline nendest on parem viis andmete esitamiseks suuliselt ettekandes ja uurimistöö kirjalikus raportis.



TALLINNA ÜLIKOOL

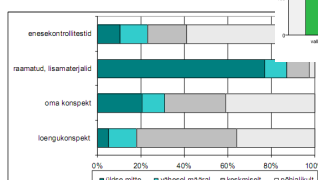
Kirjeldav statistika

Andmete esitamine - tekstina

- Vastajaid vähem kui 100**
 - 23.03.09 toimunud kirjeldava statistika osa eksamil osales 39 üliõpilast.
 - Eksamil osalenud 39-st üliõpilasest ligi kolmandik (15) oli osalenud kõikides loengutes; kõikides praktikumides osalemise vastav arv oli 10.
- Soovi korral võib ülevaatlikkuse tõstmiseks sagedusele sulgudes lisada osakaalu**
 - Uuringus osales 17 inimest, kellest 4 (23%) olid teinud rahalisi annetusi eelmise aasta jooksul.*
- Vastajaid rohkem kui 100**
 - Seisuga 30 aprill 2007 on 1.6% vanemahüvitise saajatest mehed.

Tulpdiagramm

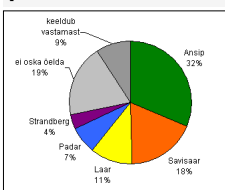
Tulba kõrgus näitab vastajate arvu või protsenti



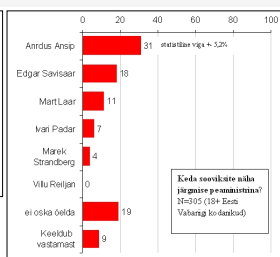
- * Võrdleb erinevaid kategooriaid
- * Pika teksti korral teljed ära vahetada
- * Tulbad võiks paigutada suuruse järjekorda (kui ei ole sisulist järjestust)

Joonis 11. Milliste materjalidega töötasid eksamiks valmistumisel

Keda sooviksite näha järgmise peaministrina?

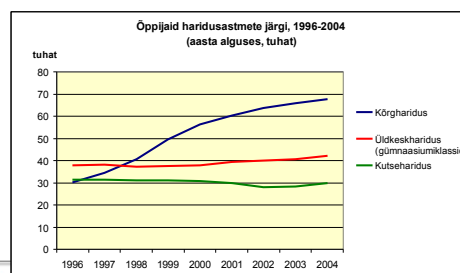


Turu-uuringute AS küsis 16-17.02 telefoni teel 305 valimisõiguslikku Eesti elanikku. Küsitluse valim on representatiivne ning üldistatav valimisõiguslike Eesti elanike suhtes.

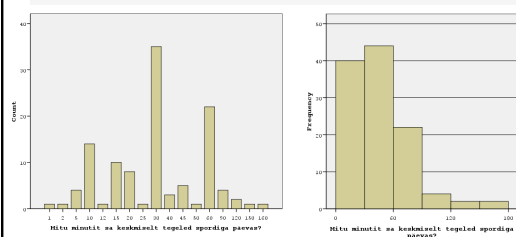


Joondigramm

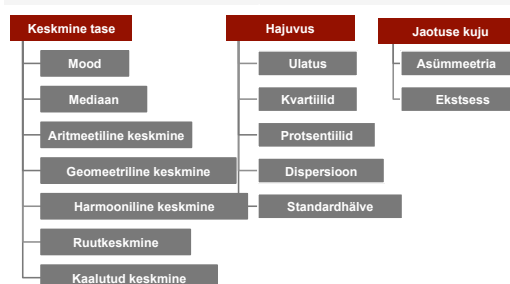
Ajas muutuvate andmete kirjeldamine



Tulpdiagramm ≠ Histogramm



Kirjeldavad arvnäitajad



Aritmeetiline keskmine e keskvärtus

- ...võimaldab suurt hulka numbrilisi andmeid koondada ja välja tuua üldtendents.
- Puuduseks **tundlikkus äärmuslike väärtuste** suhtes, kasutatakse eelkõige väikese hajuvuse korral keskvärtuse suhtes.
- Nt keskmine vanus 44 ei ütle midagi selle kohta, kui palju on alla 20-aastaseid.
- Mood ja mediaan** – muutuvad siis, kui esineb olulisi muutusi andmetes
- Aritmeetiline keskmine** – muutub siis, kui muutub kasvõi üks rea liige
- Keskvärtus on võrreldes teiste näitajatega **kõige stabiilsem**
- Kõigile teada tuntud arv näitaja (kõik teavad ja oskavad arvutada)

Ulatus e haar (*Range*)

- ... maksimaalse ja minimaalse väärtuse vahe e. vahemiku laius, milles andmed paiknevad

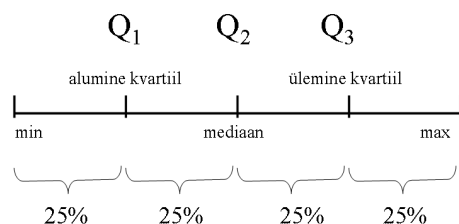
- Milliste maakondade tulemused hajuvad kõige rohkem?

	Maximum	Minimum	Range
Harjumaa	100	10	90
Hiumaa	95	15	80
Ida-Viru	100	10	90
Jõgevamaa	100	10	90
Järvamaa	100	10	90
Lääne-Vi	100	10	90
Läänemaa	100	15	85
Põlvamaa	100	10	90
Pärnumaa	100	10	90
Raplamaa	95	15	80
Saaremaa	95	10	85
Tartumaa	100	0	100
Valgamaa	100	15	85
Viljandi	100	10	90
Võrumaa	100	0	100

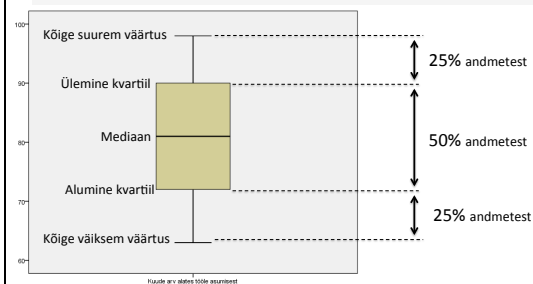
- + lihtsamini leitav
- sõltub äärmistest väärtustest, mis võivad olla ekstreemsed!!

Kvartiilid

- Kvartiilid jagavad variatsioonireala nelja võrdsesse ossa



Karpdiagramm



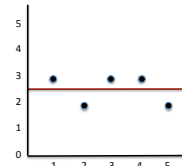
Standardhälve

- Kui palju üksikud tulemused erinevad keskmisest?
- Kui andmed on ühesugused => st.hälve=0
- Mida rohkem nad erinevad => suurem on st.hälve

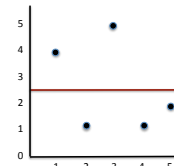
	Maximum	Minimum	Range		Std Deviation
Harjumaa	100	10	90	Harjumaa	21,86
Hiumaa	95	15	80	Hiumaa	17,81
Ida-Viru	100	10	90	Ida-Viru	22,05
Jõgevamaa	100	10	90	Jõgevamaa	21,10
Järvamaa	100	10	90	Järvamaa	18,49
Lääne-Vi	100	10	90	Lääne-Vi	18,81
Läänemaa	100	15	85	Läänemaa	19,58
Põlvamaa	100	10	90	Põlvamaa	19,74
Pärnumaa	100	10	90	Pärnumaa	20,05
Raplamaa	95	15	80	Raplamaa	17,61
Saaremaa	95	10	85	Saaremaa	18,38
Tartumaa	100	0	100	Tartumaa	22,27
Valgamaa	100	15	85	Valgamaa	20,42
Viljandi	100	10	90	Viljandi	20,40
Võrumaa	100	0	100	Võrumaa	21,19

Standardhälve

1.õppejõud
M=2,6
SD=0,55



2.õppejõud
M=2,6
SD=1,82

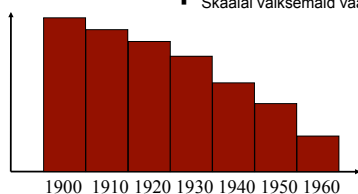


asümmeetria



• Positiivne asümmeetria (skewness)

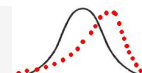
- Jaotus on väljavenitatud paremalt poolt
- Jaotuse "saba" on paremal pool
- Skaalal väiksemaid väärtuseid rohkem



TALLINNA ÜLIKOOL

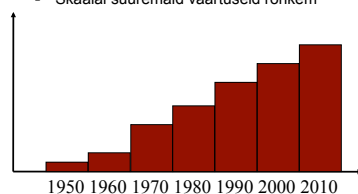
Kirjeldav statistika

asümmeetria



• Negatiivne asümmeetria (skewness)

- Jaotus on väljavenitatud vasakult poolt
- Jaotuse "saba" on vasakul pool
- Skaalal suuremaid väärtuseid rohkem



TALLINNA ÜLIKOOL

Kirjeldav statistika

Tulemuste esitamine

- Uurimuses osalejate kommunikatsiooniga rahulolu määra hindamiseks arvutati välja üldise kommunikatsiooniga rahulolu keskmine näitaja $M=4,82$ ($SD=0,94$), mida suurem on saadud tulemus, seda kõrgem on rahulolu tase (1 – väga rahulematu, 7 – väga rahul). Keskmise põhjal võib väita, et uuringus osalenud inimesed on oma organisatsioonide kommunikatsiooniga keskmisest rahuleavamad.

- Tabel 1. Materiaalse töökeskkonnaga rahulolu erinevused tulenevalt soost

tunnus	välite nr	valimigrupp	aritmeetiline		
			N	keskmine	standardhälve
sugu	5.	naine	28	4,68	0,6
		mees	6	5	0
	7.	naine	28	4,64	0,5
		mees	6	5	0
	10.	naine	28	4,81	0,4
		mees	6	5	0

TALLINNA ÜLIKOOL

Kirjeldav statistika

Andmeanalüüsi küsimus

- tekst
- tabel
- diagramm

KIRJELDAV
(esmane analüüs)

- tekst
- tabel
- diagramm

VÖRDLEV
(erinevused gruppide vahel)

- keskväärtuste kaudu
- proportsioonide kaudu

KORRELATSIOON
(seosed tunnuste vahel)

TALLINNA ÜLIKOOL

Kirjeldav statistika

Erinevuste uurimine

• Keskvväärtuste arvutamise kaudu

sissetulek_eur	sugu	koolis	haridus	perekonnaseis	leibkond	vanus	füüsiline_vorm
5424,8	Mees	14	Kutseharidus + keskhharidus	Üksik (pole olnud abielus)	3	24	Hea
2069,9	Mees	16	Ülikooliharidus, kraadiharidus	Vabaabielus (elan koos p...)	2	28	Rahuldav
1443,7	Naine	12	Keskharidus	Üksik (pole olnud abielus)	2	31	Väga hea
1383,6	Mees	16	Rakenduslik kõrgharidus	Vabaabielus (elan koos p...)	2	33	Hea
1383,6	Naine	17	Ülikooliharidus, kraadiharidus	Üksik (pole olnud abielus)	1	31	Hea
1164,3	Mees	15	Kutseharidus + keskhharidus	Lahutatud / Elan lahul	1	35	Hea
1164,3	Mees	12	Kutseharidus + keskhharidus	Vabaabielus (elan koos p...)	2	36	Väga hea
1164,3	Mees	15	Ülikooliharidus, kraadiharidus	Ametlikult abielus	4	58	Hea

- Milline on vastajate keskmine sissetulek? Kas mehed või naised teenivad keskmiselt rohkem?
- Kuidas sõltub sissetuleku suurus omandatud haridustasemest?
- ...

TALLINNA ÜLIKOOL

Kirjeldav statistika

Erinevuste uurimine

• Keskvväärtuste kaudu

sissetulek_eur	sugu	koolis	haridus
5424,8	Mees	14	Kutseharidus + keskhharidus
2069,9	Mees	16	Ülikooliharidus, kraadiharidus
1443,7	Naine	12	Keskharidus
1383,6	Mees	16	Rakenduslik kõrgharidus
1383,6	Naine	17	Ülikooliharidus, kraadiharidus
1164,3	Mees	15	Kutseharidus + keskhharidus
1164,3	Mees	12	Kutseharidus + keskhharidus
1164,3	Mees	15	Ülikooliharidus, kraadiharidus

Report

Keskmine sissetulek pereliikme kohta viimasel kuul

Milline on Teie praegune...	N	Mean	Median	Minimum	Maximum	Std. Deviation	Skewness
Algharidus	25	5123,60	3367,00	1200	14076	3835,789	,979
Põhiharidus	128	4994,63	4349,50	200	14134	3391,010	,944
Kutseharidus (ilma keskhariduseta)	35	5681,46	4892,00	200	14076	3487,758	,530
Keskharidus	243	6122,84	5773,00	400	22320	3731,763	,853
Kutseharidus + keskhharidus	232	6592,94	5650,50	440	83867	6319,241	6,124
Rakenduslik kõrgharidus	131	6980,74	6500,00	250	21390	3842,535	,016
Ülikooliharidus, kraadiharidus	163	8528,12	8370,00	1807	32600	4411,045	1,359
Total	957	6570,93	5865,00	200	83867	4685,427	5,277

TALLINNA ÜLIKOOL

Kirjeldav statistika

Erinevuste uurimine

- Kui keskvaartust arvutada ei saa

- Tunnused: nimitunnus, binaarne tunnus, järjestustunnus

sissetulek_eur	sugu	koolis	haridus	perekonnaseis	leibkond	vanus	füüsiline_vorm
5424.8	Mees	14	Kutseharidus + keskharidus	Üksik (pole olnud abielus)	3	24	Hea
2069.9	Mees	16	Ülikooliharidus, kraadiharidus	Vabaabielus (elan koos p...	2	28	Rahuldav
1443.7	Naine	12	Keskharidus	Üksik (pole olnud abielus)	2	31	Väga hea
1383.6	Mees	16	Rakenduslik kõrgharidus	Vabaabielus (elan koos p...	2	33	Hea
1383.6	Naine	17	Ülikooliharidus, kraadiharidus	Üksik (pole olnud abielus)	1	31	Hea
1164.3	Mees	15	Kutseharidus + keskharidus	Lahutatud / Elan lahus	1	35	Hea
1164.3	Mees	12	Kutseharidus + keskharidus	Vabaabielus (elan koos p...	2	36	Väga hea
1164.3	Mees	15	Ülikooliharidus, kraadiharidus	Ametlikus abielus	4	58	Hea

- Millise haridustasemega vastajate seas on enam lahutatud isikuid?



TALLINNA ÜLIKOOL

Kirjeldav statistika

Erinevuste uurimine

- Kui keskvaartust arvutada ei saa

- Millise haridustasemega vastajate seas on enam lahutatud isikuid?

- võrdlev sagedustabel e RISTTABEL

haridus	perekonnaseis
Kutseharidus + keskharidus	Üksik (pole olnud abielus)
Ülikooliharidus, kraadiharidus	Vabaabielus (elan koos p...
Keskharidus	Üksik (pole olnud abielus)
Rakenduslik kõrgharidus	Vabaabielus (elan koos p...
Ülikooliharidus, kraadiharidus	Üksik (pole olnud abielus)
Kutseharidus + keskharidus	Lahutatud / Elan lahus
Kutseharidus + keskharidus	Vabaabielus (elan koos p...
Ülikooliharidus, kraadiharidus	Ametlikus abielus

Milline on Teie praegune perekonnaseis? Perekonnaseis Cross-tabulation

Count		Perekonnaseis					Total
		Üksik (pole olnud abielus)	Ametlikus abielus	Vabaabielus, elan koos perekonnaga	Lahutatud / Elan lahus	Laht	
Milline on Teie praegune haridustase?	Algharidus	15	5	3	2	0	25
	Peiteharidus	87	15	32	1	2	137
	Kutseharidus (ilma kraadiharidusega)	7	16	10	4	0	36
	Kraadiharidus	58	101	68	28	2	253
	Kutseharidus + kraadiharidus	33	115	64	32	9	243
	Rakenduslik kõrgharidus	26	52	37	15	6	136
	Ülikooliharidus, kraadiharidus	32	81	27	19	11	170
Total		256	284	231	89	30	1060



TALLINNA ÜLIKOOL

Kirjeldav statistika

Andmeanalüüsi küsimus

- tekst
- tabel
- diagramm

KIRJELDAV
(esmane analüüs)

- tekst
- tabel
- diagramm

VÕRDLEV
(erinevused gruppide vahel)

- keskväärtuste kaudu
- proportsioonide kaudu

KORRELATSIOON
(seosed tunnuste vahel)

- tekst
- tabel
- diagramm

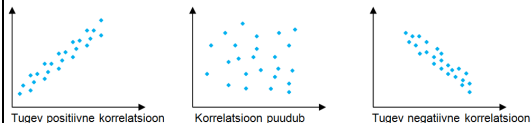


TALLINNA ÜLIKOOL

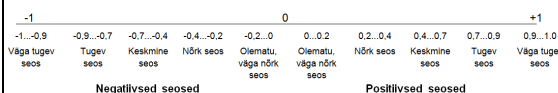
Kirjeldav statistika

Korrelatsioonanalüüs

Seose visuaalne hindamine



Seose analüütiline hindamine

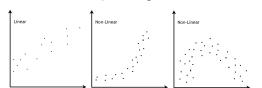


TALLINNA ÜLIKOOL

Kirjeldav statistika

Korrelatsioonikordajad

- Pearson's r
- Standardiseeritud kahe tunnuse vahelise seose kordaja
- Pearsoni kordaja puudused
 - lineaarne seos: tunneb punktipiile, mis on venitatud piki sirget.



- tundlik erandite suhtes: paar üksikut erandit väikeses valimis võivad kahekordistada kordaja väärtust.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$



TALLINNA ÜLIKOOL

Kirjeldav statistika

KORRELATSIOONANALÜÜS

- Kuidas on seotud vanus ja majapidamistöökdeks kuluv aeg?

Correlations			
		vanus	Tunde majapidamistöökdeks (tööpäeviti)
vanus	Pearson Correlation	1	,198*
	Sig. (2-tailed)	,	,000
	N	882	873
Tunde majapidamistöökdeks (tööpäeviti)	Pearson Correlation	,198**	1
	Sig. (2-tailed)	,000	,
	N	873	873

** . Correlation is significant at the 0.01 level (2-tailed).

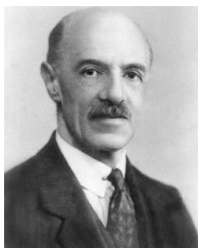


TALLINNA ÜLIKOOL

Kirjeldav statistika

Korrelatsioonikordajad

- **Spearman's ρ**
- MPAR korrelatsioonikordaja
- Astakkorrelatsioonikordaja
 - intervalltunnused ei vasta normaaljaotusele (ka erandlikud väärtused)
 - Järjestustunnus(ed)
 - asendab väärtused järjekorranumbritega ning kasutab Pearsoni kordaja valemit => Spearmani kordaja < Pearsoni kordaja (üldjuhul)



korrelatsioonanalüüs

- Spearmani kordaja

Correlations			
	sissetulek viimasel kuul	Tunde majapidamistöödeks (tööpäeviti)	
Spearman's rho	Correlation Coefficient	1,000	-.357**
	Sig. (2-tailed)	.	.000
	N	829	821
Tunde majapidamistöödeks (tööpäeviti)	Correlation Coefficient	-.357**	1,000
	Sig. (2-tailed)	.000	.
	N	821	873

** . Correlation is significant at the .01 level (2-tailed).

Korrelatsioonikordajad

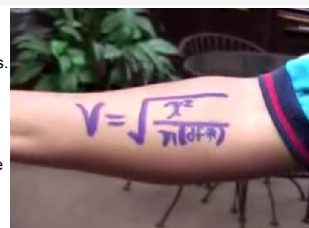
- **Kendall's τ**
- MPAR korrelatsioonikordaja
- Astakkorrelatsioonikordaja
 - kui on vähe andmeid ja palju sarnaseid väärtuseid



Korrelatsioonikordajad

- **Cramer's V**
 - Nimitunnuste seose tugevuse uurimiseks.
 - Kordaja ei näita seose suunda, ainult tugevust.

N: Eriala ja tööl käimise seos $V=.45$



- **Phi**
 - 2x2 binaarsete tunnuste korral
 - Ruutjuur hii-ruudu väärtuse jagatisest valimi suurusega

Milline kordaja valida?

KORDAJA	ANDMED	LISATINGIMUS
PEARSON	I + I	Seose kuju – lineaarne Erandlikud väärtused puuduvad (ei domineeri)
SPEARMAN	I + I, I + J, J + J	Seose kuju ei ole lineaarne ja jaotusel on erandlikud väärtused (I + I)
KENDALL	J + J	Väike valim ja palju sarnaseid väärtuseid
CRAMER	I + N, I + B, N + B, N + N, N + J, B + J	Tõlgendatakse vaid seose tugevust, mitte suunda
PHI	B + B	Tõlgendatakse vaid seose tugevust, mitte suunda

Tulemuste esitamine

- Tajutud stressi taseme ja enesetõhususe vahelise suhte analüüsil ilmnis oluline negatiivne korrelatsioon, mille kohaselt madalama enesetõhususega õpetajad tajusid stressi kõrgemalt ($r=-.37$, $p=.010$).
- Laste väliskeskonna mängu ja vanemate soovide vahel ilmnis tugev negatiivne seos ($r=-0,713$; $p=0,05$), mille põhjal võib väita, et vanemate soov oma laosi näha teatud tegevustes ja olemuses

Tabel 4. Vanemate uskumuste ja soovide ning laste mängulise käitumise vahelised korrelatsioonikoeffitsiendid (Pearsoni korrelatsioon)

Nimetaja	Laste välismäng	Laste sisemäng	Vanemate uskumused	Vanemate soovid
Laste välismäng	1	.791*	-.0192	-.0713*
Laste sisemäng		1	-.0450	-.0391
Vanemate uskumused			1	0,043
Vanemate soovid				1

*. Olulisustasemeus: $p=0.05$ (kahepoolne hüpotees).

statistika