

Kursus: Andmeanalüüs/Kvantitatiivsed ...

Seminar III: Kirjeldav statistika - arvnäitajad.

Õppejõud: Katrin Niglas
PhD, dotsent
informaatika osakond



Millest sõltub analüüsimeetodi valik?

- I. Küsimuse tüübist
e mis tüüpi on küsimus, millele tahame analüüsiga vastust saada – nt Kas kolm gruppi **erinevad**? Kas kaks nähtust on **seotud**?
- II. Andmete tüübist
Kas nimi-, järjestus, arv- või binaarsed tunnused
- III. Sihtrühmast
Kui suurt teadlikkust statistiliste meetodite osas võib eeldada?
Milline esitlusviis on selle rühma puhul kõitev ja sobilik?



Tunnusetüübid /ka skaalatüübid või andmete tüübid/

Võtmeküsimused:

väärtuste järjestatavus ja skaalavahemike võrdsus!

- Nimitunnused (nt rahvus)

! Nimitunnusel ei ole väärtused üheselt järjestatavad, järjestustunnusel on!

- Järjestustunnused (nt haridustase)

! Järjestustunnusel ei ole väärtuste vahemikud võrdsed, arvtunnusel on!

- Arvtunnused /ka intervalltunnused/ (nt laste arv)

- Arvtunnused väheste erinevate väärtustega
- Arvtunnused paljude erinevate väärtustega

! Binaarsel tunnusel on ainult kaks väärtust ja seega järjestamise ja vahede võrdsuse probleemi ei teki!

- Binaarsed tunnused (nt sugu)



Kuidas oma andmeid kokku võtta?

Struktureeritud andmete esmaseks kokkuvõtuks ning ülevaatlikuks analüüsiks saab kasutada **kirjeldava statistika** meetodeid, mis võib jagada kolme suurde rühma:

- Sagedustabelid (sh risttabelid)
- Arvnäitajad
- Arvjoonised e diagrammid



Keskmisega on asi lihtne – seda õpetati koolis!

Sissetulek ühe leibkonnaliikme kohta	
Vastanud leibkondade arv	758
Aritmeetiline keskmine	2300

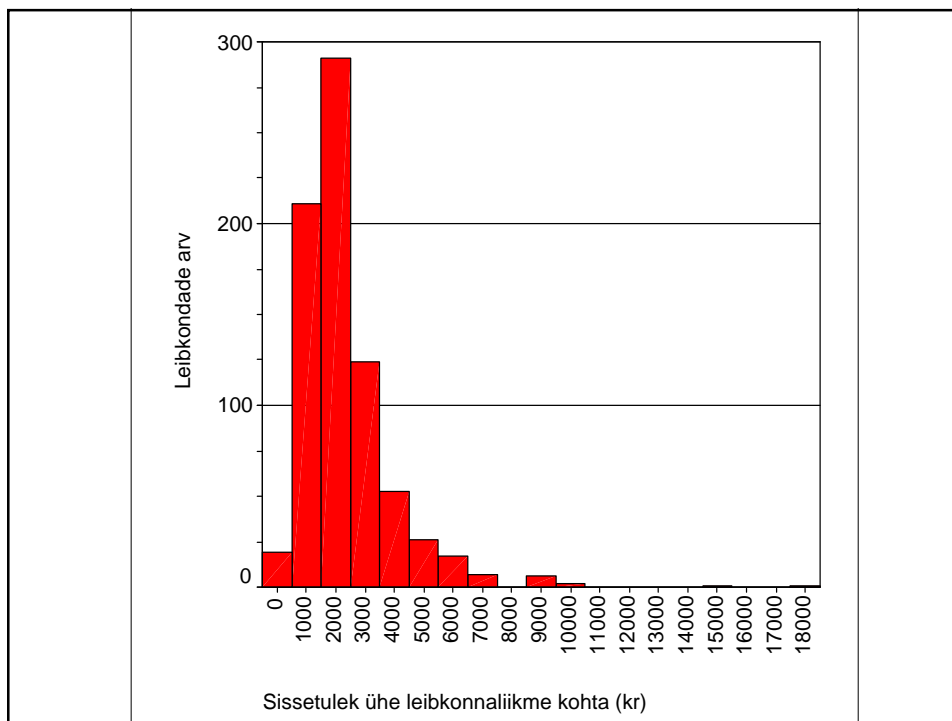
- Perekond Usina sissetulek ühe leibkonnaliikme kohta on 2000 kr
- Kas perekond Usin kuulub madalama või kõrgema sissetulekuga leibkondade **poole** hulka?

Keskmisega on asi lihtne – seda õpetati koolis!

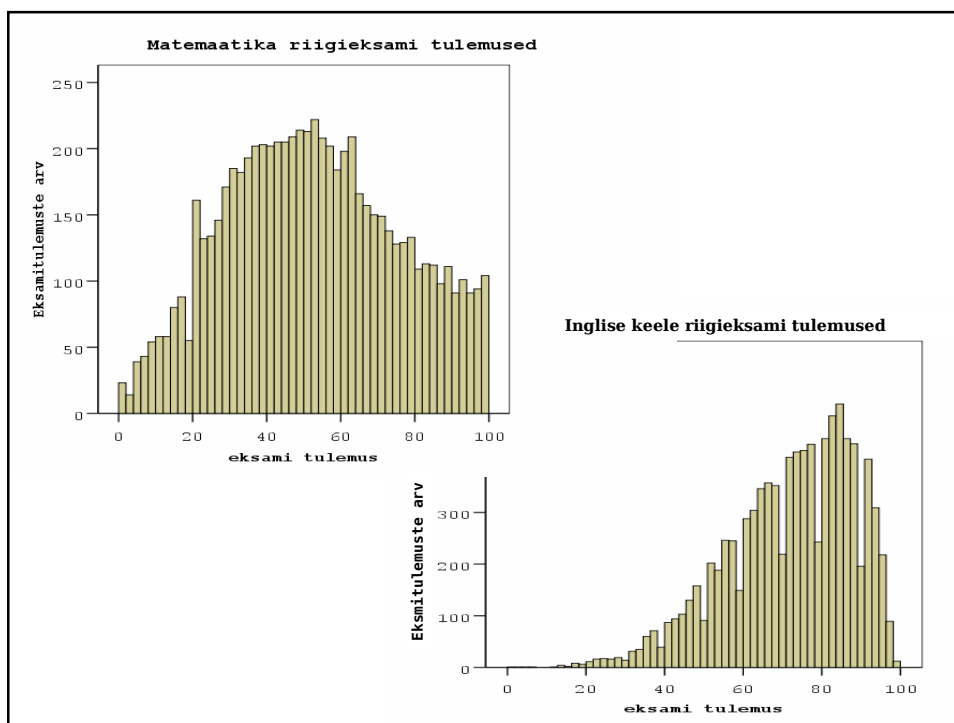
Kuigi järgnevas tabelis toodud teiste arvnäitajate (mediaan ja protsentiilid) põhjal on näha, et perekond Usin kuulub kõrgema sissetulekuga leibkondade poole hulka, oleks korrektne vastus eelmisele küsimusele olnud, et antud andmete põhjal ei ole võimalik esitatud küsimusele vastata!

Sissetulek ühe leibkonnaliikme kohta (EEK)		
N Valid	Mittepuuduvad väärtusi	758
N Missing	Puuduvad väärtusi	252
Mean	Aritmeetiline keskmine (keskväärtus)	2300,44
Median	Mediaan	1851
Mode	Mood	800 (a)
Std. Deviation	Standardhälve	1636,56
Skewness	Asümmeetria	3,13
Range	Ulatus	18000
Minimum	Miinumum	0
Maximum	Maksimum	18000
Percentiles	Protsentiilid	10
		25
		33,3
		50
		64
		66,7
		75
		90

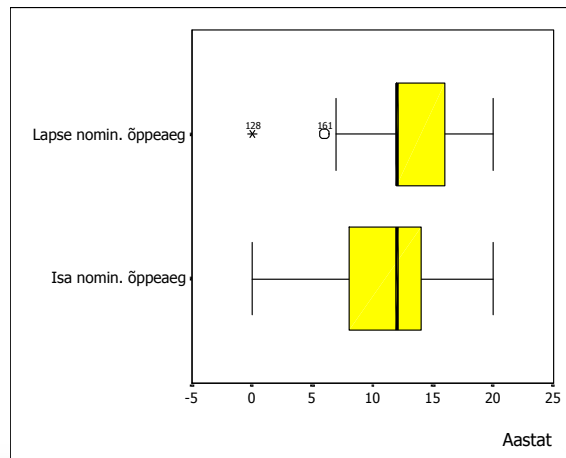
a Multiple modes exist. The smallest value is shown



Riigieksami tulemused (2005, gümnaasium)			
		Matemaatika	Inglise keel
N Valid	Mittepuuduvaid väärtusi	6866	9375
N Missing	Puuduvaid väärtusi	0	0
Mean	Aritmeetiline keskmine (keskväärtus)	52,1	71,9
Median	Mediaan	51	74
Mode	Mood	54	84
Std. Deviation	Standardhälve	23,3	16,0
Skewness	Asümmeetria	0,1	-0,7
Range	Ulatus	100	99
Minimum	Miinumum	0	1
Maximum	Maksimum	100	100
Percentiles	Protsentiid		
	25	34	62
	33,3	40	66
	50	51	74
	66,7	62	81
	75	69	85



Karp-vurrud-diagramm jaotuste võrdlemiseks



Arvnäitajad - kokkuvõte

Arvnäitajate e arvkarakteristikute abil saame iseloomustada tunnuse:

- jaotuse 'keskmist tendentsi' (mood, mediaan, aritmeetiline keskmine)
- jaotuse hajuvust ehk seda, kui võrd erinevad on väärtused/üksiktulemused omavahel (miinimum, maksimum, ulatus, kvartiilid, kvartiilide vahe, standardhälve, dispersioon)
- jaotuse kuju (asümmeetriakordaja, teised arvnäitajad omavahel võrrelduna)
- ...

Mitme tunnuse koos analüüsimisel saab arvnäitajate abil analüüsida:

- gruppide omavahelisi erinevusi
- tunnuste omavahelisi erinevusi
- tunnuste omavahelisi seoseid (NB! Seda õpime hiljem!)
- ...

Kirjeldav statistika – mida oskan juba teha? I

- **Paigutada andmed mõnda tabelisse, mis näitab andmete jaotumist kategooriate või minimaalse ja maksimaalse väärtuse vahel.**
(variatsioonirida, sagedustabel)
- **Illustreerida väärtuste/tulemuste jaotumist diagrammi abil**
(tulpdiaagramm, histogramm, sektordiaagramm, punktdiaagramm, sagedupolügoon, ...)
- **Leida sobiv(ad) arvkarakteristik(ud) iseloomustamaks jaotuse 'keskmist tendentsi'**
(mood, mediaan, aritmeetiline keskmine)

Kirjeldav statistika – mida oskan juba teha? II

- **Leida sobiv(ad) hajuvust iseloomustav(ad) arvkarakteristik(ud) ning illustreerida hajuvust karp-vurrud-diagrammi abil**
(ulatus, kvartiilide vahe, standardhälve)
- **Võrrelda gruppe**
(risttabelid, mitmemõõtmelised sagedustabelid, arvnäitajad mitme grupi kohta erladi, võrdlevad diagrammid: võrdlev tulpdiaagramm, ...)
- **Võrrelda tunnuseid**
(risttabelid, mitmemõõtmelised sagedustabelid, arvnäitajad mitme tunnuse kohta erladi, võrdlevad diagrammid: võrdlev tulpdiaagramm, ...)