

Kursus: Andmeanalüüs/Kvantitatiivsed ...

Seminar V: Üldistav statistika – statistilised olulisustestid.

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?



Kuidas teada saada, mida võib valimi põhjal väita üldkogumi kohta?

Selleks, et teada saada, milliseid üldistusi (ja kas üldse) saab valimi põhjal üldkogumi kohta teha, saab peale kogutud andmete esmast analüüsi kasutada **üldistava statistika** meetodeid, mis võib jagada kahte suurde rühma :

- Vahemikhinnangud e usaldusintervallid
- Statistilised olulisustestid

PS! Samu meetodeid saab kasutada ka eksperimentaalses uuringutes, et kontrollida, kas saadud erinevused jms on tekkinud gruppide erineva mõjutamise tulemusena või võivad olla juhuslikud.

Meetodi valik lähtuvalt analüüsi eesmärgist ja andmete tüübist

Milline analüüsimeetod valida?	Parameetrilised meetodid (eeldus: arvtunnused)	Mitteparameetrilised meetodid (järjestus- või nimitunnused aga ka arvtunnused)
1 grupp (keskmine tase/ osakaal)	K.st: \bar{x} , s, jne Ü.st: vahemikhinnangud (μ , σ)	K.st: sagedustabel, % Ü.st: vahemikhinnangud
2 gruppi ERINEVUSED	K.st: \bar{x}_1 \bar{x}_2 Ü.st: t-test	K.st: risttabel Ü.st: χ^2 -test
3 või enam gruppi ERINEVUSED	K.st: \bar{x}_1 \bar{x}_2 \bar{x}_3 ... Ü.st: ANOVA	K.st: \bar{x}_1 \bar{x}_2 \bar{x}_3 ... Ü.st: Kruskal-Wallis test
2 või enam tunnust SEOSSED		

Statistilise olulisustesti põhisammud:

E I: Analüüsisin olemasolevaid andmeid kirjeldava statistika meetodite abil ning leidsin midagi „huvitavat“ (nt. erinevuse või seose jne)

E II: Tekkis küsimus: „Kas võib üldistada?“

$\bar{U} \rightarrow V \rightarrow \bar{U}$ (v eksperimentaalne disain)

I. **Õige olulisustesti valik** (lähtuvalt probleemist ja andmetüübist)

II. **Valitud olulisustesti eelduste kontroll:**

ei 
jah 

III. **Hüpoteesid:** sisukas hüpotees H_1 :

nullhüpotees H_0 :

Olulisuse nivoo α

(„Kui väike peab olema H_0 kehtimise tõenäosus, et me võiks ilma suurema riskita ta mittekehtivaks tunnistada?“)

IV. **Arvutused**

eesmärgiks hinnata H_0 kehtimise tõenäosust p

(„Kui suur on tõenäosus, et olukorras, kus H_0 kehtib, tekkis valmis olnud erinevus v seos juhuse tõttu?“)

($p = Sig = olulisuse\ tõenäosus$)

V. **Otsus tulemuse kohta:**

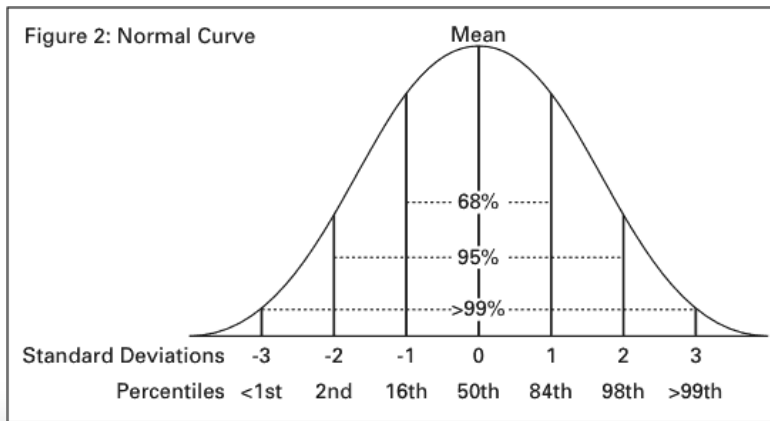
$p > \alpha$ H_0 jääb kehtima - statistiliselt mitte oluline (ei üldista)

$p \leq \alpha$ H_1 tõestatud - statistiliselt oluline (võib üldistada)

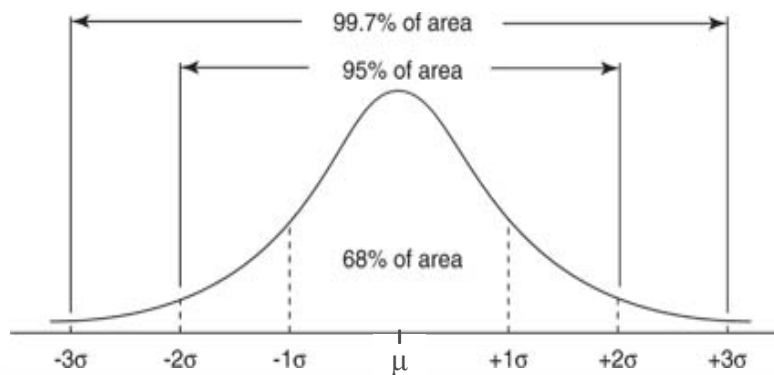
VI. **Järelduse sõnastamine**

NB! Skeem on suuremana
eraldi failis!

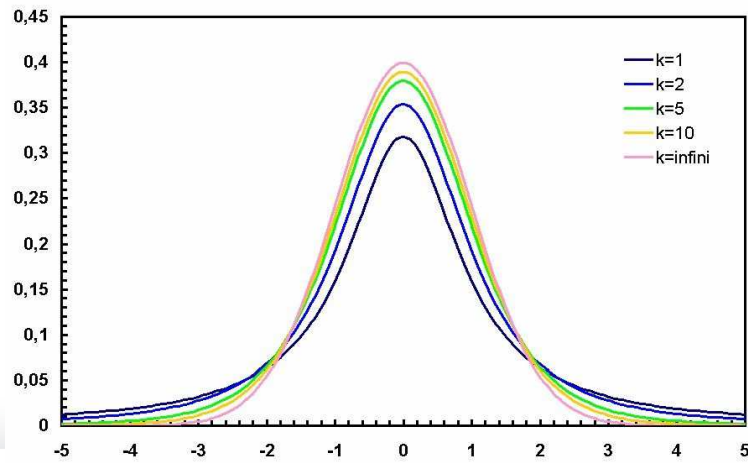
Proportsioonid normaaljaotuskõvera all



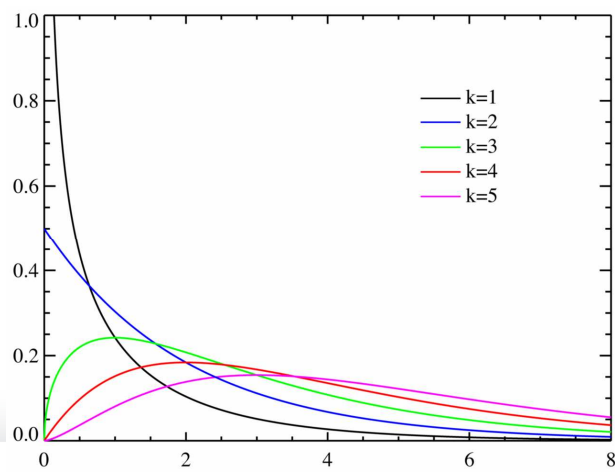
Proportsioonid normaaljaotuskõvera all



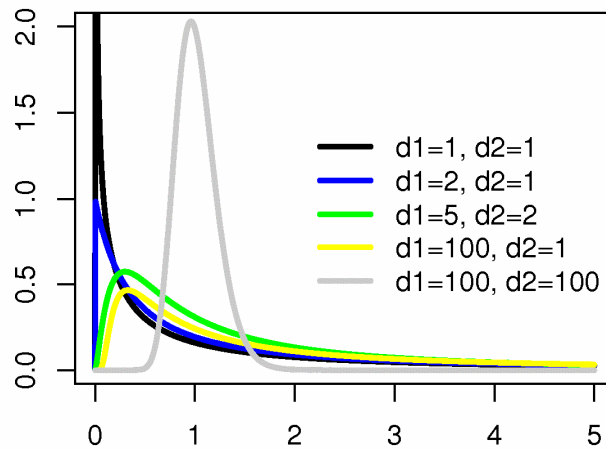
t-jaotuste pere



hii-ruut jaotuste pere



F-jaotuste pere



Statistilise olulisuse sisuline tõlgendamine elulisse konteksti

		VALIM	
		s u u r	v ä i k e
E R I S N E O V U S	s u u r	Stat.olulisus: +	Stat.olulisus: – (?)
		Elul.olulisus: +	Elul.olulisus: + (!?)
	v ä i k e	Stat.olulisus: + (?)	Stat.olulisus: –
		Elul.olulisus: – (!?)	Elul.olulisus: – (!?)