

Seminar VI: Seoste analüüs

(kirjeldava ja üldistava statistika meetodid)

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?

Meetodi valik lähtuvalt analüüsi eesmärgist ja andmete tüübist

Milline analüüsimeetod valida?	Parameetriselised meetodid (eeldus: arvtunnused)	Mitteparameetriselised meetodid (järjestus- või nimetunnused 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	K.st: Pearsoni r (korrelatsioonikordaja) Ü.st: $H_0: r_{\text{ÜK}}=0$	K.st: Spearmani ρ Ü.st: $H_0: \rho_{\text{ÜK}}=0$ K.st: risttabel Ü.st: χ^2 -test

Seosekordajad

Korrelatsioonikordaja r väärtuste tõlgendamine:

- $r = +1$ tähendab maksimaalse tugevusega positiivset seost
- $r = 0$ tähendab seose puudumist
- $r = -1$ tähendab maksimaalse tugevusega negatiivset seost

- $|r| < 0.30$ olematu, väga nõrk
- $0.30 < |r| < 0.70$ keskmise tugevusega
- $0.70 < |r|$ tugev

PS! Samade põhimõtete järgi saab tõlgendada ka teisi seosekordajaid!

Seosekordajad

Korrelatsioonikordaja r väärtuste arvutamine:

$$r = \frac{\sum_{i=1}^N \frac{x_i - \bar{x}}{st.h._x} * \frac{y_i - \bar{y}}{st.h._y}}{N}$$



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?“

Ü → V → Ü (v eksperimentaalne disain)

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

II. **Valitud olulisustesti eelduste kontroll:**

ei
jah

III. **Hüüpoteesid:** 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:**

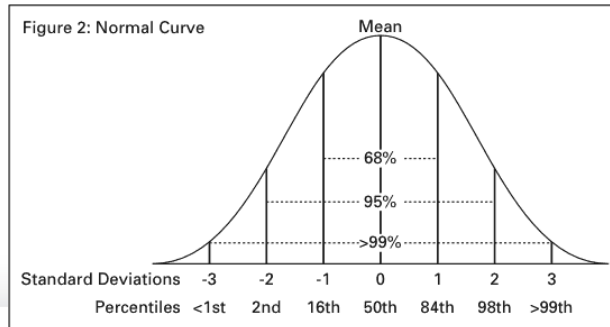
$p > \alpha$ H_0 jääb kehtima - statistiliselt mitte oluline (ei üldista)

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

VI. **Järelduse sõnastamine**

Korrelatsioonikordaja statistilise olulisuse testimine

$$st.viga_r = \frac{1-r^2}{\sqrt{N}} \quad H_0 : r_{\text{ÜK}} = 0 \Rightarrow st.viga_r = \frac{1}{\sqrt{N}}$$



Statistilise olulisuse sisuline tõlgendamine elulisse konteksti

		VALIM	
		s u u r	v ä i k e
E R I S 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: - (!?)