

Kursus: Andmeanalüüs/Kvantitatiivsed ...
Tallinn, 28. september. 2007

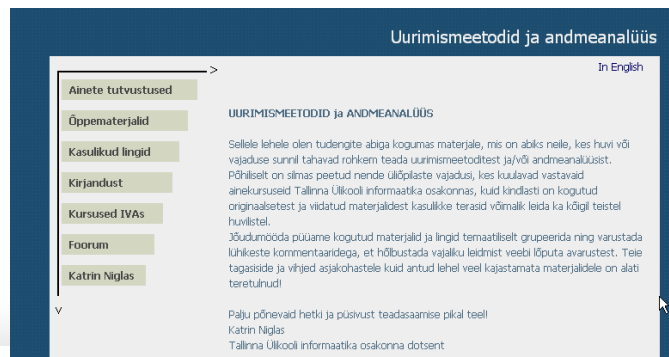
Seminar I: Andmed ja esmane andmetöötlus.

Õppejõud: Katrin Niglas
PhD, dotsent
informaatika osakond



Kui peaks huvi tekkima ...

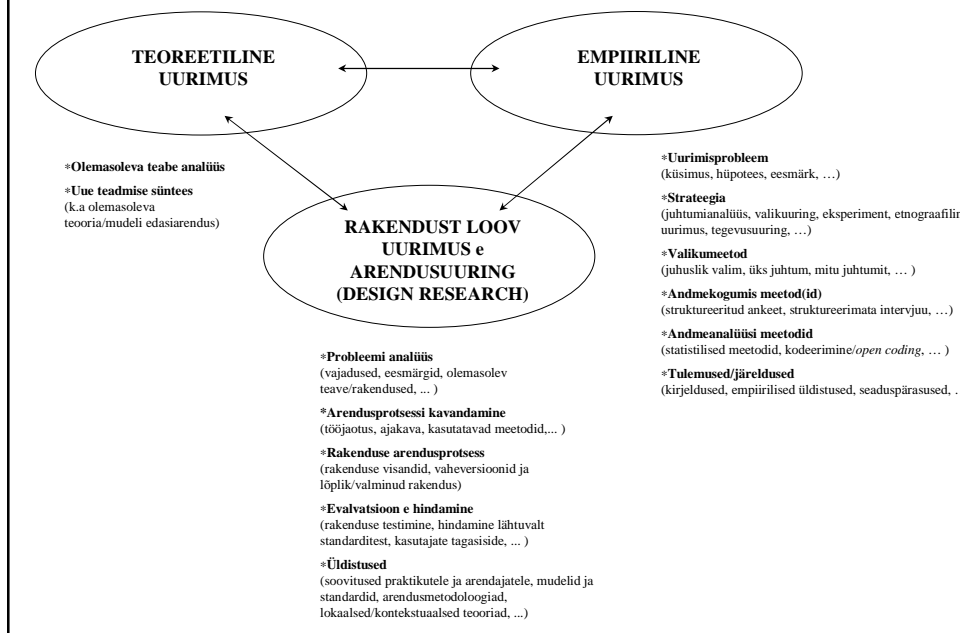
- Kursusest osavõtjatele: IVA's (Kvantitatiivsed.../A... S2007)
- Lisamaterjal kõigile: www.tlu.ee/~katrin/



The screenshot shows a website page with a dark blue header containing the title "Uurimismeetodid ja andmeanalüüs" and a language selector "In English". A sidebar on the left lists navigation options: "Ainete tutvustused", "Õppematerjalid", "Kasulikud lingid", "Kirjandust", "Kursused IVAs", "Foorum", and "Katrin Niglas". The main content area is titled "UURIMISMEETODID JA ANDMEANALÜÜS" and contains a paragraph of text in Estonian. At the bottom of the main content, there is a small bio for Katrin Niglas: "Paju põnevaid hetki ja püüvust teadasaamise pikal teel Katrin Niglas Tallinna Ülikooli Informaatika osakonna dotsent".



Uuringutüüpide üldistatud klassifikatsioon



Empiirilise uurimuse metodoloogilised aspektid, millest moodustub uuringu disain e plaan:

Uurimisprobleem (küsimus, hüpotees, eesmärk, ...)

Strateegia (juhtumianalüüs, (tõenäosuslik) valikuuring, eksperiment, tegevusuuring, ...)

Valikumeetod (juhuslik valim, üks juhtum, mitu juhtumit, ...)

Andmekogumis meetod(id) (struktureeritud ankeet, struktureerimata intervjuu, ...)

Andmeanalüüsi meetodid (statistilised meetodid, kodeerimine/*open coding*, ...)

Tulemused/järeldused (kirjeldused, empiirilised üldistused, seaduspärasused, ...)

Andmete kogumine - *measurement*

Tavapärased andmekogumismeetodid:

- dokumenteerimine / koondandmete talletamine

- struktureeritud või poolstruktureeritud intervjuud (küsitlused)
- struktureeritud või poolstruktureeritud ankeedid (kirjalikud küsimustikud)
- (standardiseeritud) võimekus- ja sooritustestid
- psühholoogilisi ja sotsiaalseid aspekte mõõtvad testid (nt väärtushinnangute skaalad)

- struktureerimata andmete kodeerimine - kontentanalüüs



Andmete kogumine II

Küsimuste/mõõdikute kavandamine -
operatsionaliseerimine:

- uuritav nähtus/fenomen vs indikaatorid/mõõdikud
- kas antud nähtust/fenomeni on eelnevates uuringutes mõõdetud ning kas vastav kirjandus on usaldusväärne ja väliidne?
- kas saab juba olemasolevaid mõõdikuid kohandada või tuleb välja töötada uued mõõdikud/küsimused?
 - kultuurilised ja kontekstuaalsed erinevused!
 - võrreldavus (võrdlevad ja longituud uuringud)
- kas peaks kasutama konsultante?
- kas peaks kasutama piloteerimist ning järelkontrolli?



Andmete kogmine III

Hea küsimustik on

- selge sõnastusega,
- kergesti ja üheselt mõistetav,
- kiiresti vastatav,
- vormistuselt korrektne;
- koostatud nii, et oleks minimeeritud vastajate ja andmete töötajate poolt potentsiaalselt tehtavate vigade hulk.

Andmete kogmine IV

Hoiduda tuleks sellistest küsimustest nagu:

- suunavad küsimused;
- teaduslikult täpse, kuid pika ja keeruka sõnastusega küsimused;
- halvasti või mitmeti mõistetavad küsimused;
- ärritavad küsimused (sh küsimusega mittesobivad vastusevariandid);
- kategoorilist laadi ja eitust sisaldavad küsimused (!?);
- kirjalikes (eriti posti teel korraldatud) küsitlustes pole soovitatav esitada avatud küsimusi (!?)

Andmetabel (korrektne, kuid mitteprofessionaalne)

Sugu	Sünniaeg	Pikkus	Kaal	Keskmine hinne	Hobi	Tähtkuju
N	3.04.1981	160	48	4,31	sport	Jäär
N	14.11.1979	162	53	3,26	Muusika	Skorpion
M	18.02.1980	169	60	3,67	Ujumine	Veevalaja
M	24.01.1980	162	53	4,38	Magamine	Veevalaja
N	23.08.1980	165	55	5	Muusika	Neitsi
N	23.04.1980	169	54	4,31	sport	Sõnn
N	19.08.1980	168	56	4,44		Lövi
N	21.05.1980	169	57	4,75	Kassid	Sõnn
M	4.08.1980	179	76	3,38	Söömine	Lövi

Andmetabel (professionaalne)

ID	A05	C0100	C0200	C0400	C3200	D02	E03	E05
1	6	21	1	4	560	9	4	4
2	6	26	1	2	482	9	3	4
3	6	30	1	2	700	1	3	2
4	6	31	1	2	3000	3	3	3
5	6	33	1	2	2400	3	2	2
6	6	34	1	2	504	9	3	3
7	6	37	1	2	6000	3	2	2
8	6	46	1	2	1000	4	3	3
9	6	47	1	6	3800	1	3	3

Andmetabel – kõige tähtsamad reeglid

Statistiline andmestik: objekt – tunnus – väärtus – skaala
 ↓ ↓ ↓
Tabel: rida – veerg – lahter

- Õige andmetabel on “askeetlik”
- Igal tunnusel unikaalne nimi (veeru pealkiri esimeses reas); ei kasutata mitut veergu ühendavat pealkirja!
- Igas lahtris ainult üks ühik infot!
- Ühes veerus ainult üht tüüpi andmed (arvude vahele muud ei sisestata; puuduvad väärtused kodeeritakse või jäetakse tühi lahter)
- Kui ühe küsimuse puhul on lubatud valida mitu vastusevarianti, siis iga variant annab andmetabelis eraldi tunnuse (veeru)



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)



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?