

Andmeanalüüs: statistiline andmestik ja kirjeldav statistika

IFI7041
Kairi Osula (kairio@tlu.ee)
www.tlu.ee/~kairio/7041

Täna

14.15-15.45 16.00-17.30

Sissejuhatus kursusesse
Eksamist ja praktilistest töödest
Uuringute tüübid
Põhimõisted
Empiirilise uuringu etapid
SPSS
SPSS eeltöö enne statistilist analüüsi

Kursuse materjalid

www.tlu.ee/~kairio/7041

Kairi Osula
FB: Kairi Osula
kairio@tlu.ee

Kasutage suhtlemisel kindlasti kursuse koodi (IFI7041).

Kursuse ülesehitus

Kontakttunde 5x4 (14.15-17.45) ruumis T-302

31.08 - sissejuhatus, SPSS
21.09 - andmete haldamine, tabelid
19.10 - kirjeldavad arvnäitajad, diagrammid
16.11 - diagrammid
7.12 - korrelatsioonanalüüs

1.kodune töö - tähtaeg: 5.11
2.kodune töö - tähtaeg: kirjaliku testi kirjutamise päev

Kursuse hinne

• HINNE

- „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%

Praktilised tööd (50%) Kirjalik test (50%)
2. töö (30p.) - testi kuupäev 2. võimalus -

Positiivse hinne saamiseks peavad olema tehtud **pos.tulemusele (25,5%)** mõlemad osad (kirjalik test & praktilised tööd)

Praktiline töö 1

Rühmatöö (1-3 in. ühes rühmas)
Maksimumtulemus 20 punkti.
Tähtaeg: 5.11.2017 (Kui teile antud tähtaeg üldse ei sobi, andke teada!)
Tähtajast hiljem saadetud töö tuleb teha üksinda ja selle maks.tulemus saab olla 15p.

Uurimisküsimuse püstitamine
Andmete kogumine
Esmane kirjeldav analüüs

Täpne ülesandekirjeldus tuleb 21.09

Praktiline töö 2

Rühmatöö (1-3 in. ühes rühmas)
Maksimumtulemus 30 punkti.
Tähtaeg: testi kirjutamise kuupäev
Tähtajast hiljem saadetud töö tuleb teha üksinda ja selle maks.tulemus saab olla 25p.

Õpitud meetodite kasutamine andmete analüüsimisel

Täpne ülesandekirjeldus tuleb 7.12

Kirjalik test

Võtke kaasa **1 A4** paber - kirjutage sinna, mida vajalikuks peate
Järeleksamil ei ole materjalide (A4 paberi) kasutamine lubatud

Test toimub tavalises klassis, kirjutusvahend ja dokument kaasa
Registreeruge ÕIS-is, kindlasti

Testi pikkus 1A4

Avatud küsimused

Kindlasti küsin - tunnusetüüpide määramist, analüüsitulemustele
järeluste kirjutamist, seoseanalüüsis järeluste tegemist jne.

Kindlasti ei küsi - SPSS-i käske ja seda, mida me rääkinud ei ole!)

Lisalugemist

- "Uuri ja kirjuta"
 - Autorid: S. Hirsjärvi, P. Remes, P. Sajavaara
 - Kirjastus: Medicina
 - Ilmumisaasta: 2005



- "Uurija käsiraamat"
 - Autor: A.Kidron
 - Kirjastus: Mondo
 - Ilmumisaasta: 2008



Lisalugemist

- "Andmete analüüs ja tõlgendamine sotsiaalteadustes"
 - Autor: Liina Mai Tooding
 - Kirjastus: Tartu Ülikooli kirjastus
 - Ilmumisaasta 2015
- "Andmeanalüüs sotsiaalteadustes"
 - Autor: Liina Mai Tooding
 - Kirjastus: Tartu Ülikooli kirjastus
 - Ilmumisaasta: 1998



Lisalugemist

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



Uuringute tüübid

Uuringutüüpide üldistatud klassifikatsioon

Empiirilised uuringud

Teoreetilised uuringud

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

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

Näited Filosoofilised küsimused ja puhta matemaatika teoreemid
Ajaloolised uuringud
Arvutikasutuseetika (loogiline arutelu ja seisukohad), tarkvara võrdlev analüüs
Referaat (?)

Arendusuuringud

Empiiriline uuring

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

K. Niglas

Uurimisprobleem

Küsimus, hüpotees, eesmärk

Valikumeetod - keda uurida?

juhtum(id), valim

Üldkogum/valim

VALIM
Uurimiseks valitud väike objektide grupp

ÜLDKOGUM POPULATSIION
Kõik objektid, kelle kohta meie poolt püstitatud järeldused, oletused, prognoosid kehtivad.

Valimi moodustamine

Tõenäosuslikud valimid <i>(probability sampling)</i>	Mittetõenäosuslikud valimid <i>(non-probability sampling)</i>
VEA HINDAMINE	VEA HINDAMINE
Vea hindamine TÄPNE Matemaatilised valemid Tõenäosusteooria Viga hinnatakse suuremaks väiksemate gruppide korral Vahemikhinnang Olulisustestid Dispersioonanalüüs Regressioonanalüüs	Viga hinnatakse suuremaks kuna tüüpilised või mugavad objektid võivad osutada ebatüüpilisteks või mitteesinduslikeks. Üks-ühele tehtavaid üldistusi üldkogumile teha ei saa. Võimaldab näha töö tulemustes üldisemalt kehtivaid tendentse.

Kihtvalim

Alamosad

Rühmad:

vanuse_soo ...

Rühma seest

valitakse

esindajad

juhulikult,

garanteerides

kõigi

rühmade

võrdse

esindatuse.

	Mudel protsent	Saadud koostululem protsent	Kaalutud koostululem protsent
ELUKOHT			
Linn	68,7	71,6	68,7
Maa	31,3	28,4	31,3
SUGU			
Mees	46,5	45,9	46,5
Naine	53,5	54,1	53,5
VANUS			
15-19	10,1	14,9	10,1
20-29	18,1	19,2	18,1
30-39	17,5	16,7	17,5
40-49	18,6	16,2	18,6
50-59	15,2	15,1	15,2
60-74	20,5	17,9	20,5
RAHVUS			
Eestlased	85,9	85,2	85,9
Mitte-eestlased	14,1	14,7	14,1
REGIOON			
LOODE	45,0	43,5	45,0
KIRDE	18,2	19,1	18,2
KAGU	21,0	20,7	21,0
EDELA	10,5	10,9	10,5
LAANE	5,3	5,7	5,3
HARIDUS			
Põhharidus ja vähem	30,7	29,6	30,7
Keskharidus	54,4	52,0	54,4
Kõrgharidus	14,9	18,4	14,9

Andmete kogumine

Andmete kogumine

- Töös püstitatud hüpoteesi(de) kontrollimiseks on vaja andmeid, fakte.
- Andmete kogumine
 - reaalsel tegeelikust fikseerivaid ning uurivaid meetodeid.
- Kvantitatiivsed meetodid**
- Kvalitatiivsed uurimismeetodid**
- Objektivsemad tulemusi saab, kui neid meetodeid kasutatakse koos.
- Erinevate meetodite kombineerimine võimaldab teha üldistusi ja ühtlasi näha konteksti.

kvantitatiivsed

kvalitatiivsed

- Kui palju?** (probleemi kirjeldamine)
- Miks?** (probleemi põhjused)
- TULEMUSED üldistatakse** üldkogumile, mõnikord on see küsitav (või mittevajalik)
- Üldistamine tugineb arvulisele argumentidele
- Seoste uurimine** (ei näita põhjuslikku seost)

Kas ülikoolide sisseastumiseksamite kaotamine põhjustas madalamad õpitulemused gümnaasiumis või vastupidi?

- Kuidas?** (probleemi kirjeldamine)
- Milleks?** (probleemi põhjused)
- TULEMUSTE** analüüs töömahukas
- Uuritavate hulk väike (üldistamine piiratud)
- uuritakse tõlgendusi, hoiakuid ning arvamusi (koos põhjendustega)
- Saab töötada välja efektiivseid lahendusi konkreetsetele isikutele, mille laiemat mõju saab hiljem uurida kvantitatiivsete meetoditega

Kuivõrd piirab öine alkoholimüügikeeld alkoholi tarbimist?

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?

Andmeanalüüs

Põhimõisted

- **Kirjeldav statistika** – andmete esmane kokkuvõtt, ülevaatic analüüs
 - TABELID
 - ARVJONISED
 - KIRJELDAVAD ARVNÄITAJAD
 - SEOSED
- **Üldistav statistika** – milliseid (kas üldse?) saab valimi põhjal üldkogumi (populatsioon) kohta teha?
 - VAHEMIKHINNANG
 - STATISTILISED OLULISUSTESTID

Millest sõltub andmeanalüüsi meetodi valik?

Andmetest

Mis tüüpi on andmed?

Küsimusest

Erinevus, seos

Sihtgrupist

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

Andmeanalüüsi vahendid



SPSS

- 1968.a. Norman H. Nie, Dale H. Bent, C. Hadlai Hull.
- SPSS 16.0 ning vanemad on kasutatavad Windows, Mac, Linux OS-des.
- 2009 IBM
- 2009 PASW (Predictive Analytics SoftWare)
- 2010 IBM SPSS
- 2016 IBM SPSS Statistics 24.0

SPSS



Andmete sisestamine

Andmete sisestamine

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

Andmete sisestamine

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	Sõmine	Lövi

eestielu.sav

	a05	c0100	c020000	c040000	c060000	c080000	c320000	d02	e03
1	Tallinn	21	mees	leak	venelane	keskerh...	560		väer
2	Tallinn	26	mees	abielus	venelane	põhharid...	482		tulete ots-o
3	Tallinn	30	mees	abielus	ukraanlane	keskerh...	700		realtootaja tulete ots-o
4	Tallinn	31	mees	abielus	eeslane	kõrgem har...	3000		keekaastme juht tulete ots-o
5	Tallinn	33	mees	abielus	eeslane	keskerh...	2400		keekaastme juht majandusl...
6	Tallinn	34	mees	abielus	venelane	põhharid...	504		tulete ots-o
7	Tallinn	37	mees	abielus	eeslane	keskharidus	6000		keekaastme juht majandusl...
8	Tallinn	46	mees	abielus	eeslane	kõrgem har...	1000		lippjuht tulete ots-o
9	Tallinn	47	mees	lahutatud	eeslane	keskerh...	3800		realtootaja tulete ots-o
10	Tallinn	54	mees	vallaline	venelane	keskharidus	2200		alamaastme juht tulete ots-o
11	Tallinn	57	mees	abielus	venelane	kõrgem har...	6000		realtootaja majandusl...
12	Tallinn	63	mees	abielus	eeslane	keskerh...	3500		realtootaja tulete ots-o
13	Tallinn	66	mees	abielus	eeslane	keskharidus	2000		realtootaja tulete ots-o
14	Tallinn	68	mees	vallaline	eeslane	kõrgem har...	2480		realtootaja tulete ots-o
15	Tallinn	69	mees	abielus	eeslane	lõpetamata	3500		realtootaja tulete ots-o

Eeltöö – andmestiku korrastamine

- Puuduvad väärtused - kui palju neid on, mida teha tühjade lahtritega?
 - 3%
 - Puuduvate väärtuste defineerimine
 - Puuduvate väärtuste leidmine
 - Puuduvate väärtuste asendamine
-
- Andmesisestusvead
 - Andmete grupeerimine (vajadusel)
 - Skaalade pööramine (vajadusel)