

MTTTP1, luento 11.9.2018

KERTAUSTA

- Populaatio
tutkimusobjektien muodostama joukko, johon tilastollinen tutkimus kohdistuu, koko N
- Populaation yksikkö
tilastoyksikkö, havaintoyksikkö
- Otos
populaation osajoukko, koko n
- Tilastoyksikön ominaisuudet
tilastollisia muuttujia

- Empiirinen havaintoaineisto (data)
saadaan mittaamalla tilastoyksiköiden ominaisuuksia
- Tilastolliset analyysimenetelmät
välineitä havaintoaineiston tutkimiseksi ja
johtopäätelmien tekemiseksi
- Tilastollinen analyysi
kuvailevaa analyysia
tilastollista päättelyä

- Tilastollisen tutkimuksen työvaiheet

- 1 Suunnittelu

- tutkimuskohteen & aiheen valinta
tilastoyksikkö
muuttujat
- tutkimuksen suorittamisen suunnittelu
kyselylomake
otantamenetelmä
koejärjestely jne.

- 2 Aineiston hankkiminen ja tallennus analysointia varten

- suunnittelun havaintoaineiston hankinta
- tallennus ja muokkaus analysointia varten

3 Aineiston kuvailu

- kuvailevan tilastotieteen keinoin aineiston sisältämän tiedon esittely ja tutkiminen

4 Tilastolliset mallit ja testaukset

- populaatiosta tehtyjen väittämien testaukset aineiston (otoksen) perusteella
- todennäköisyysteoriaan perustuvien tilastollisten mallien sovittaminen havaintoaineistoon

5 Raportointi

- johtopäätelmien teko ja niiden esittäminen ja tulkinta

Ks. Harjoitustyön ohjeet

<http://www.sis.uta.fi/tilasto/mhttp1/syksy2018/htyop118.pdf>

3 HAVAINTOAINEISTO JA HAVAINTOMATRIISI

Aineiston hankinta

- otantatutkimus, päättely populaatiosta satunnaisesti populaatiosta tehdyn otoksen (satunnaisotoksen) perusteella
- kokeellinen tutkimus, päättely populaatiosta saatujen tulosten perusteella

Esim. 3.1. Päätelytilanteita

[http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/luen
torunko.pdf#page=7](http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/luen
torunko.pdf#page=7)

a) Puolueen kannatuksen arviointi, esim.

<https://yle.fi/uutiset/3-10387592>

Muodostetaan luottamusväli todelliselle kannatukselle.

b) Halutaan arvioida suomalaisten naisten keskipituutta. Lasketaan otoksesta keskipituus ja arvioidaan virhettä, joka liittyy päätelyyn. Tässä voidaan muodostaa keskipituudelle luottamusväli.

Otantamenetelmät (tapoja satunnaisotoksen tekemiseen)

- yksinkertainen satunnaisotanta YSO
- systemaattinen otanta SO
- ositettu otanta OO
- ryväotanta RY

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=7>

Sopiva aineisto voi olla olemassa, se voidaan saada myös yhdistelemällä eri lähteistä.

Analysoitavassa aineistossa

- n tilastoyksikköä, $a_1, a_2, a_3, \dots, a_n$
- p muuttujaa, $x_1, x_2, x_3, \dots, x_p$

Havaintomatriisi on $n \times p$ –taulukko, jossa muuttujien arvot jokaiselta tilastoyksiköltä muodossa:

	x_1	$x_2 \dots$	$x_j \dots$	x_p
a_1	x_{11}	$x_{12} \dots$	$x_{1j} \dots$	x_{1p}
a_2	x_{21}	$x_{22} \dots$	$x_{2j} \dots$	x_{2p}
..				
a_j	x_{j1}	$x_{j2} \dots$	$x_{jj} \dots$	x_{jp}
.				
a_n	x_{n1}	$x_{n2} \dots$	$x_{nj} \dots$	x_{np}

Havaintomatriisissa n riviä ja p saraketta, sarake muodostaa kyseisen muuttujan jakauman.

Esim. CTESTI-aineisto, mikroluokkien verkossa

Muuttujia

IKÄ	Oppilaan ikä 1.11.1974 täysinä vuosina
PITUUS	Oppilaan pituus 1 cm tarkkuudella
PAINO	Oppilaan paino 100 g tarkkuudella.
HENGTIL	Hengitystilavuus 100 cm ³ tarkkuudella
COOPER	Cooperin testin tulos
LEUANVET	Leuanvetojen lukumäärä
KOULU	Oppilaan koulu MYP=0, TKYL = 1
LKASTE	8-luokkaisen koulun I, III, V, VII, Vastaavuus peruskoulun 5, 7, 9 ja lukion 2.
ÄIDINK1, ÄIDINK2, MATEM, KIELI1, KIELI2, KIELI3, LIIKUNTA	Ko. aineitten koulunumerot
SOSRYHMÄ	Isän ammatin koodaus 1-9, Rauhala: 1960-luvun suomalaisten yhteiskunnan sosiaalinen kerrostuneisuus ammatin arvostuksen valossa
KIPPI	Koodaus pääsee kipin= 1, ei pääse =0
PUOLIV	Koodaus pääsee = 1, ei pääse = 0
HAVOPP	Oppilaan numero aineistossa

Tutkimusongelmia?

Esim. PULSSI-aineisto

<http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/luentorunko.pdf#page=102>

Tutkimusongelmia?

Esim. 3.5.

HOTDOG-aineisto

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/lueantorunko.pdf#page=101>

Muuttuja \$/oz ilmoittaa unssihinnan dollareina

1 unssi = 28,35 g = 0,02835 kg, 1 dollari = 0,77€

Kilohinta = $(\$/oz) \times 0,77 / 0,02835$.

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/lueantorunko.pdf#page=11>

Tutkimusongelmia?

Esim. 3.3.

Myytyjä kiinteistöjä

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/lue/ntorunko.pdf#page=10>

Muuttujia P = myyntihinta tuhansina dollareina

S = koko tuhansina neliöjalkoina

Eurohinta = $0,77 \times P \times 1000$

Neliöt = $0,0929 \times S \times 1000$

1 square foot = $0,0929 \text{ m}^2$

Neliöhinta = Eurohinta/Neliöt

Tutkimusongelmia?

Esim. 3.6.

Tampereella 12 kuukauden aikana myytyjä kerrostaloasuntoja, otos 4.6.2012, aineisto Tre_myydyt_asunnot_2012.sav sivulla <https://coursepages.uta.fi/mhttp1/esimerkkiaineistoja/>

Tutkimusongelmia?

4 MITTAAMINEN

- Mittaaminen
menettely (sääntö), jolla tilastoyksikköön liitetään tiettyä ominaisuutta kuvaava luku, mittaluku.
- Mittausvirhettä
mittari epätarkka
häiriötekijät
- Mittarin reliabiliteetin alhainen
toisistaan riippumattomat, samalle tilastoyksikölle tehdyt mittaukset antavat huomattavasti poikkeavia tuloksi

- Mittarin ei validi
ei mittaa sitä ominaisuutta, mitä tarkoitus mitata
(mittari huonosti laadittu)
- Suoraan mitattavissa ja tulkittavissa olevia muuttujia
Esim. Henkilön pituus, paino, kunnan asukasluku,
lasten lukumäärä perheessä
- Eivät suoraan mitattavissa olevia muuttujia, määrittely ei yksikäsitteistä
Esim. Henkilön älykkyys, musikaalinen lahjakkuus,
uskonnollisuus, asenne johonkin; www-sivun
käytettävyys

Esim. Henkilön uskonnollisuutta voidaan mitata kirkossa käyntien määrällä, uskonnollisen kirjallisuuden lukemisella, ...(nk. indikaattorimuuttujien avulla).

Esim. Asenne-/mielipidemittauksissa asennetta/mielipidettä peilaavia väitteitä

Vastaaja valitsee esimerkiksi vaihtoehtoista

- täysin samaa mieltä
- jokseenkin samaa mieltä
- ei samaa eikä eri mieltä
- jokseenkin eri mieltä
- täysin eri mieltä

Muuttujia voidaan luokitella monella tavalla:

- 1) kategorisiin eli kvalitatiivisiin
numeerisiin eli kvantitatiivisiin
- 2) mitta-asteikkojen perusteella
- 3) jatkuva
ei-jatkuva
- 4) selitettävä
selittäjä

1)

Kvalitatiivinen (kategorinen) muuttuja

jakaa tilastoyksiköt tarkasteltavan ominaisuuden suhteen luokkiin

Esim. Henkilön siviilisääty, opiskelijan tutkinto-ohjelma, kaupungin sijaintikunta, vaatteiden kokoluokitus

Kvalitatiiviset muuttujat voidaan koodata numeerisesti, MUTTA numeroarvoilla ei määrällistä tulkintaa; ovat vain luokkien nimiä tai kuvaavat luokkien "suuruusjärjestyksen".

Kvantitatiivinen (numeerinen) muuttuja

muuttujan arvo mitattaessa reaallinen, mitataan lukumäärää tai mittaus mittayksikköä käyttäen

Esim. Henkilön pituus, opiskelijan ikä, kaupungin asukasluku, vaatteen hinta

2)

Muuttujien mitta-asteikot

Luokittelu- eli laatuero- eli nominaaliasteikko

kvalitatiivinen muuttuja, jonka luokkia ei voida asettaa järjestykseen (esim. paremmuus, suuruus, kovuus)

Esim. Henkilön siviilisääty, opiskelijan koulutusohjelma, kaupungin sijainti

Järjestys- eli ordinaaliasteikko

kvalitatiivinen muuttuja, jonka luokat voidaan asettaa mielekkääseen järjestykseen mitattavan ominaisuuden suhteen

Esim. Asennekysymykset, vaatteiden kokoluokitus

Suhdeasteikko

numeerisen muuttuja, jonka arvo nolla vastaa tarkasteltavan ominaisuuden "häviämistä", absoluuttista nollapistettä

Esim. Henkilön paino (kg) ja pituus (cm), henkilön 100 m juoksuaika (s), asunnon vuokra (€), urheilijan harjoitteluun käyttämä aika päivässä (min)

Intervalliasteikko

numeerisen muuttuja, jonka nollakohta ei suhdeasteikon tapaan määritelty

Esim. Huoneen lämpötila Celsius-asteina.

Absoluuttinen asteikko

suhdeasteikollinen, jossa mittaus kiinnitetyllä mittayksiköllä

Esim. Asunnon huoneiden lukumäärä, perheessä lasten lukumäärä

Esim. 4.2.

Liikuntamäärien mittaus

<http://www.sis.uta.fi/tilasto/mttp1/syksy2018/luentorunko.pdf#page=14>

Kvantitatiivisesti mitattuna:

Harrastan liikuntaa (väh. 30 min/kerta) keskimäärin ____ kertaa viikossa.
Keskimäärin kerralla liikun ____ min.

Kvalitatiivisesti mitattuna:

Harrastan liikuntaa (väh. 30 min/kerta) keskimäärin
____ alle 3 kertaa viikossa
____ 3-4 kertaa viikossa
____ enemmän kuin 4 kertaa viikossa.

Keskimäärin kerralla liikun

____ alle tunnin
____ 1-2 tuntia
____ yli 2 tuntia

Esim. CTESTI-aineiston muuttujien mitta-asteikot

IKÄ	Oppilaan ikä 1.11.1974 täysinä vuosina
PITUUS	Oppilaan pituus 1 cm tarkkuudella
PAINO	Oppilaan paino 100 g tarkkuudella.
HENGTIL	Hengitystilavuus 100 cm ³ tarkkuudella
COOPER	Cooperin testin tulos
LEUANVET	Leuanvetojen lukumäärä
KOULU	Oppilaan koulu MYP=0, TKYL = 1
LKASTE	8-luokkaisen koulun I, III, V, VII, Vastaavuus peruskoulun 5, 7, 9 ja lukion 2.
ÄIDINK1, ÄIDINK2, MATEM, KIELI1, KIELI2, KIELI3, LIIKUNTA	Ko. aineitten koulunumerot
SOSRYHMÄ	Isän ammatin koodaus 1-9, Rauhala: 1960-luvun suomalaisten yhteiskunnan sosiaalinen kerrostuneisuus ammatin arvostuksen valossa
KIPPI	Koodaus pääsee kipin= 1, ei pääse =0
PUOLIV	Koodaus pääsee = 1, ei pääse = 0
HAVOPP	Oppilaan numero aineistossa

Mitta-asteikko vaikuttaa tilastollisen menetelmän valintaan. Numeeristen muuttujien yhteydessä lähes samat menetelmät ja tunnusluvut käyvät kaikille kolmelle mitta-asteikolle.

Suhdeasteikolla muuttujan arvojen suhteilla on mielekäs tulkinta. Intervalliasteikolla voidaan vertailla arvojen eroja, mutta ei suhteita.

Avainkäsitteet:

Havaintomatriisi

Muuttujan jakauma

Otantamenetelmät

Mittaaminen

Kvalitatiivinen muuttuja

Kvantitatiivinen muuttuja

Mitta-asteikot