

Kvantitatiiviset menetelmät

- Pienryhmiin ilmoittautuminen alkaa ke 2.2. klo 9.00
 - **Ryhmä 1: Jussi Kinnunen:**
Ensimmäinen kokoontuminen to 24.2. klo 14-16, paikka ??
SPSS-harjoitukset: ti 29.3. klo 11-13 ja to 7.4. klo 15-19
 - **Ryhmä 2: Mikko Mattila**
Ensimmäinen kokoontuminen to 24.1 klo 10-12, U40 sali 13
SPSS-harjoitukset ma 14.3. klo 16-19 ja to 17.3. 15-18
 - **Ryhmä 3: Minna Tiili**
Ensimmäinen kokoontuminen to 24.2. klo 10-12, päärak. sl 16
SPSS-harjoitukset ke 30.3. klo 16-19 ja to 31.3. klo 9-12
 - **Ryhmä 4: Hanna Wass**
Ensimmäinen kokoontuminen pe 25.2. klo 10-12, U37 sh 1
SPSS-harjoitukset 17.3. klo 9-12 ja pe 18.3. 9-12

© Mikko Mattila 2005

1

Perusjoukko ja otos

- Havaintoyksikkö on empiirisen mittauksen kohde
- Perusjoukko on kaikkien havaintoyksiköiden muodostama kokonaisuus
- Otos on perusjoukkoa pienempi havaintoyksiköjoukko, johon mittaus kohdistetaan

© Mikko Mattila 2005

2

Otos ja näyte

- Otos on havaintoyksikköjen joukko, johon kaikilla havaintoyksiköillä on tiedossa oleva nollaa suurempi todennäköisyys tulla valituksi
- Edustava otos kuvaa mahdollisimman hyvin perusjoukon ominaisuuksia
- Näyte on havaintoyksikköjen joukko, jonka valitsemiseen ei käytetä satunnaisuutta
- Näytteen edustavuudesta ei ole mitään takeita

© Mikko Mattila 2005

3

Tilastollinen päättely

- Kuvaileva tilastoanalyysi tiivistää informaatiota muuttujien ominaisuuksista ja niiden välisistä suhteista
 - Ei ole tarkoitus tehdä yleistyksiä perusjoukosta
- Tilastollisen päättelyn avulla voidaan tehdä johtopäätöksiä perusjoukosta
 - Kuinka hyvin otoksen avulla mitatut tulokset voidaan yleistää koko perusjoukkoa koskeviksi tuloksiksi?

© Mikko Mattila 2005

4

Luottamusväli ja -taso

- Luottamusväli kertoo mille välille perusjoukon tunnusluvun arvo sijoittuu tietyllä todennäköisyydellä
- Esim. 95% todennäköisyydellä suomalaisten Nato-kannatus sijoittuu 40-50% välille

© Mikko Mattila 2005

5

Luottamusväli ja -taso

- Luottamustaso kertoo, millä todennäköisyydellä perusjoukkoa kuvaava tunnusluku on tietyllä luottamusvälillä
 - Edellisessä esimerkissä 95% on luottamustaso
- Luottamusväli tietämiseksi täytyy tietää luottamustaso ja päinvastoin
- Mitä korkeampi luottamustaso, sitä pidempi luottamusväli ja päinvastoin

© Mikko Mattila 2005

6

Otantajakauma

- Luottamusvälin laskemiseksi tarvitaan tietoa otantajakaumasta
- Otantajakauma on tunnusluvun jakauma, joka saadaan ottamalla kaikki määrätyn kokoisen otokset perusjoukosta
 - Esim. otetaan kaikki mahdolliset 1000 hengen otokset suomalaisista ja lasketaan Nato-kannattajien osuus jokaisesta otoksesta
- Tunnusluku voi olla osuus, keskiarvo, keskihajonta, korrelaatiokerroin, regressiokerroin jne.

© Mikko Mattila 2005

7

Keskivirhe

- Otantajakauman ominaisuudet voidaan määrittellä laskennallisesti tilastotieteen menetelmin
- Luottamusvälin laskemiseksi tarvitaan tieto tunnusluvun keskivirheestä
- Tunnusluvun keskivirhe kuvaa sen otantajakauman hajontaa

© Mikko Mattila 2005

8

Keskivirheen laskeminen

- Jos kyse on osuudesta (esim. kuinka monta prosenttia suomalaisista kannattaa Nato-jäsenyyttä), se saadaan kaavasta:
- S=keskivirhe, p=kannattajien prosenttiosuus, q=vastustajien prosenttiosuus, n=otoskoko.

$$S = \sqrt{\frac{pq}{n}}$$

© Mikko Mattila 2005

9

Keskivirheen laskeminen

- Jos kyse on keskiarvosta (esim. mikä on suomalaisten keskipalkka) saadaan keskivirhe kaavasta:
- Kaavassa S=keskivirhe, s=muuttujan keskihajonta otoksessa, n=otoskoko

$$S = \frac{s}{\sqrt{n}}$$

© Mikko Mattila 2005

10

Luottamusvälin laskeminen

- 95% luottamusväli lasketaan kaavasta: otoskeskiarvo $\pm 1,96 \cdot$ keskivirhe
 - Esim. otoksessa 45% kannattaa Natoa ja keskivirhe on 1,57
 - Silloin Nato-kannatus on välillä $45\% \pm 1,96 \cdot 1,57$ eli 41,9%-48,1%
- Jos halutaan 99% luottamusväli, käytetään kertoimena 2,58

© Mikko Mattila 2005

11

ISSP 2000 / esimerkki

- q32: Oletteko jäsenenä jossakin ryhmässä, jonka tärkeimpänä tavoitteena on luonnon- ja ympäristönsuojelu?
- q33: Oletteko viimeksi kuluneen vuoden aikana
 - a) Allekirjoittaneet jonkin ympäristönsuojelua koskeneen adressin tai vastaavan?
 - b) Lahjoittanut rahaa jollekin ympäristönsuojelujärjestölle tai -ryhmälle?
 - c) Ottanut osaa mielenosoitukseen tai marssiin jonkun ympäristöasian vuoksi?

© Mikko Mattila 2005

12

ISSP 2000 / esimerkki

	Kyllä	Ei	N
Jäsenyys	5,4%	94.6 %	1500
Adressi	21,6%	78.4 %	1495
Lahjoitus	23,7%	76.3 %	1495
Mielenosoitus	1,1%	98.9 %	1489

© Mikko Mattila 2005

13

ISSP 2000 / esimerkki

Jäsenyys	5,4%	4.3 - 6.5	3.9 - 6.9
Adressi	21,6%	19.5 - 23.7	18.9 - 24.3
Lahjoitus	23,7%	21.5 - 25.9	20.9 - 26.5
Mielenosoitus	1,1%	0.6 - 1.6	0.4 - 1.8

© Mikko Mattila 2005

14

Eurobarometriesimerkki

"In political matters people talk of "the left" and "the right". How would you place your views on this scale?

LEFT 1 2 3 4 5 6 7 8 9 10 RIGHT

MAA	KESKIARVO	N	KESKIHAJON TA
Suomi	5,58	862	1,91
Espanja	4,60	795	2,03
Tanska	5,44	954	1,94

© Mikko Mattila 2005

15

Eurobarometriesimerkki

- Lasketaan Suomalaisten vastaajien 95% luottamusväli:

$$S = \frac{s}{\sqrt{n}} = \frac{1,91}{\sqrt{862}} = 0,065$$

$$5,58 - 1,96 * 0,065 = 5,45$$

$$5,58 + 1,96 * 0,065 = 5,71$$

© Mikko Mattila 2005

16

Eurobarometriesimerkki

MAA	95%:n luottamusväli
Suomi	5,45 – 5,71
Espanja	4,46 – 4,74
Tanska	5,32 – 5,56

© Mikko Mattila 2005

17

Otoksen ja perusjoukon koko

- Otokseen kasvaessa luottamusväli pienenee eli pystytään tekemään tarkempia arvioita perusjoukosta
- Luottamusvälin puolittamiseksi otoskoko täytyy nelinkertaistaa
- Perusjoukon koko ei vaikuta tilastollisten yleistysten tarkkuuteen
- Samaa otoskoko voidaan käyttää väestömäärältään eri kokoisissa maissa

© Mikko Mattila 2005

18

Otoskoon vaikutus

Esimerkissä on laskettu eri otoskokojen vaikutus keskivirheeseen ja luottamusväliin. Oletuksena on, että 50% vastaajista on vastannut "kyllä" kysymykseen.

Otoskoko	Keskivirhe	95%:n luottamusväli
50	7.1	36 – 64
100	5.0	40 – 60
200	3.5	43 – 57
500	2.2	46 – 54
1000	1.6	47 – 53
2000	1.1	48 – 52
10000	0.5	49 – 51

© Mikko Mattila 2005

19

Hypoteesien tilastollinen testaus

- Hypoteesien testauksen perusperiaatteiden ymmärtäminen tärkeää, koska kaikki monimuuttujamenetelmät sisältävät "automaattisesti" tilastollisia testejä
 - ristiintaulukko → χ^2 -testi
 - regressioanalyysi → t-testi, F-testi
 - varianssianalyysi → F-testi
 - jne.

© Mikko Mattila 2005

20

Hypoteesien tilastollinen testaus

- Huom. tutkimushypoteesin ja tilastollisen hypoteesin ero
- Tutkimushypoteesi: paremmin koulutetut henkilöt osallistuvat poliittisesti enemmän kuin vähemmän koulutusta saaneet
- Em. hypoteesia voidaan operationalisoida, mitata ja tutkia monella eri tavalla
- Tilastollinen hypoteesi liittyy aina jonkin muuttujan jakaumaa tai muuttujien välistä yhteyttä kuvaavaan tunnuslukuun

© Mikko Mattila 2005

21

Hypoteesien tilastollinen testaus

- Hypoteesien testauksen viisi vaihetta:
 - 1) hypoteesien valinta
 - 2) tilastollisen testin valinta
 - 3) merkitsevyystason valinta
 - 4) testin suorittaminen
 - 5) päätöksen tekeminen
- Kyseessä hypoteesin testauksen "oppikirjamalli", todellisuudessa nämä vaiheet ovat tutkijan päässä, eikä niitä muuta kuin poikkeustapauksissa kirjata esim. tutkimusraporttiin

© Mikko Mattila 2005

22

Hypoteesien valinta

- Nollahypoteesi H_0
 - Teoriasta johdetun oletuksen vastainen hypoteesi
 - Esim. miesten ja naisten keskipalkka on sama
- Vastahypoteesi H_1
 - Miesten ja naisten keskipalkan ovat erisuuret (kaksisuuntainen testaus)
 - Naisten keskipalkka on miesten keskipalkkaa pienempi (yksisuuntainen testaus)

© Mikko Mattila 2005

23

Tilastollisen testin valinta

- Riippuu hypoteesin sisällöstä
- Erilaisia testejä on kymmeniä
- Moneen määrälliseen menetelmään sisältyy automaattisesti testit
 - esim. Pearsonin korrelaatiokertoimen yhteydessä testataan, eroaako ko. kerroin tilastollisesti merkitsevästi nollasta

© Mikko Mattila 2005

24

Merkitsevyystason valinta

- Kutsutaan myös riskitasoksi
- Määrittää todennäköisyyden sille, että nollahypoteesi hylätään, vaikka se on tosi
- Yleensä käytetään 5%, 1% ja 0,1% tasoja
- Em. tasot ainoastaan sopimuksenvaraisia

© Mikko Mattila 2005

25

Testin suorittaminen

- Käytännössä aina tilasto-ohjelmiston avulla
- Tuloksena p-arvo, joka määrittää virheen todennäköisyyden, jos nollahypoteesi hylätään

© Mikko Mattila 2005

26

Päätöksen tekeminen

Jos p-arvo alle valitun merkitsevyystason nollahypoteesi hylätään

- Tällöin puhutaan tilastollisesti "merkitsevää" tuloksesta (ei "merkittävästä")
- Merkitään taulukoihin usein tähdillä:
* ($p < 0,05$), ** ($p < 0,01$), *** ($p < 0,001$)
- Sisällöllinen tulkinta tärkeä! → p-fetismin vaara

© Mikko Mattila 2005

27

Päätöksen tekeminen

- Päätöstä tehtäessä on mahdollisuus kahteen virheeseen:
- Jos nollahypoteesi hylätään, vaikka se on tosi → hylkäämisvirhe (engl. type I error)
- Jos nollahypoteesi hyväksytään, vaikka se on epätosi → hyväksymisvirhe (type II error)
- Hyväksymis- ja hylkäämisvirheiden todennäköisyydet toisistaan riippuvaisia → jos hylkäämisvirheelle asetetaan vaativa raja (esim. 0,1%), kasvaa hyväksymisvirheen todennäköisyys ja päinvastoin

© Mikko Mattila 2005

28

Hypoteesien testauksen kritiikki

- Tekninen: ovatko otokset todella satunnaisotoksia?
- Tieteenfilosofinen: nollahypoteesi useimmiten triviaali
- Käytännöllinen: testit tehdään väärin tai niiden "ritualistinen" käyttö
- Pragmaattinen vastakritiikki: testauksessa ongelmia, mutta siitä huolimatta parempi käyttää niitä, kuin luopua kokonaan tilastollisen merkitsevyysarvioinnista.

© Mikko Mattila 2005

29

Aineistona koko perusjoukko?

- Onko tilastollisessa testauksessa mieltä, jos aineistona on koko perusjoukko?
- Asia on jonkin verran kiistanalainen, mutta valtavirran mielipide on, että testaaminen on hyödyllistä myös tässä tapauksessa
- Esim. tutkitaan Helsingin yliopiston professoreiden tieteellistä tuottavuutta ja havaitaan ero alemmista sosiaaliluokista ja korkeimmista sosiaaliluokista tulleiden välillä
- Vaikka kyseessä ei ole otos, voi erot kuitenkin johtua satunnaisista tekijöistä, koska jos kaikki profat jaettaisiin satunnaisesti kahteen ryhmään, myös näiden ryhmien välillä olisi todennäköisesti pieniä eroja → tilastollinen testi kertoo, onko ero niin suuri, että on epätodennäköistä, että yhtä suuri ero saataisiin jakamalla perusjoukko satunnaisesti ryhmiin

© Mikko Mattila 2005

30