

Kvantitatiiviset menetelmät

HUOM! Tentti pidetään tiistaina 22.2. klo 16-18 Vuorikadulla V20 Is

© Mikko Mattila 2005

1

Muuttujien muunnokset

- Usein empiirisen analyysin yhteydessä tulee tarve muuttaa aineiston muuttujia
- Esim. syntymävuoden muuttaminen iäksi
- Kaksi vaihtoehtoa:
 - Uusien muuttujien luominen
 - Muuttujan uudelleenkoodaus

© Mikko Mattila 2005

2

Uusien muuttujien luominen

- Uusi muuttuja esitetään kaavan avulla
- Esimerkiksi uuden asukastiheysmuuttujan luominen
 - $TIHEYS = ASUKASLUKU / PINTA-ALA$
- Summamuuttujan luominen:
 - $SUMMA = KYS1 + KYS2 + KYS3 + KYS4 + KYS5$

© Mikko Mattila 2005

3

Muuttujien uudelleenkoodaus

- Alkuperäisen muuttujan koodausta muutetaan
 - Vanha koodaus: Mies=1, Nainen=2
 - Uusi koodaus: Mies=0, Nainen=1
- Uudelleenkoodaus voidaan tehdä joko alkuperäiseen muuttujaan tai uuteen muuttujaan (suositeltavampaa)

© Mikko Mattila 2005

4

Muuttujien uudelleenkoodaus

- Uudelleenkoodausta tarvitaan muuttujan "suunnan" vaihtoon (pienet arvot vaihdetaan suuriksi ja päinvastoin)
- Esim. vastaajan terveydentila: "erittäin hyvä" (1), "melko hyvä" (2), "kohtalainen" (3), "melko huono" (4) ja "erittäin huono" (5)
- Uudelleenkoodausta tarvitaan myös vähintään välimatka-asteikollisen muuttujan luokittelussa
 - ikä <35: uusi arvo 1
 - ikä 35- 59: uusi arvo 2
 - Ikä >60: uusi arvo 3

© Mikko Mattila 2005

5

Summamuuttujat

- Summamuuttuja muodostetaan yhdistämällä useita samaa asiaa eri tavalla mittaavia kysymyksiä yhteen muuttujaan laskemalla ne yhteen.
- $SUMMA = KYS1 + KYS2 + KYS3 + KYS4 + KYS5$
- $SUMMA = (KYS1 + KYS2 + KYS3 + KYS4 + KYS5) / 5$

© Mikko Mattila 2005

6

Miksi summamuuttujia?

- Käsitteet moniulotteisia → mittaaminen yhdellä kysymyksellä kattavasti vaikeaa (esim. "konservatiivisuus") → mittauksen validiteetti paranee.
- Mittarin reliabiliteetti paranee → satunnaisvirheen vaikutus vähenee.

© Mikko Mattila 2005

7

Miksi summamuuttujia?

- Mittarin tarkkuus paranee → tarkemmat luokittelut mahdollisia.
- Empiirinen analyysi helpottuu → voi keskittyä yhteen muuttujaan tarkemmin usean muuttujan sijasta

© Mikko Mattila 2005

8

Huomioitavia asioita

- Summamuuttujaa rakennettaessa kiinnitettävä huomiota seuraaviin seikkoihin:
 - Kysymysten on oltava "samansuuntaisia" eli suuri arvo kuvaa jokaisen kysymyksen kohdalla mitattavan käsitteen jompaa kumpaa ääripäätä (jos näin ei ole → "uudelleenkoodaa" poikkeavat kysymykset).

© Mikko Mattila 2005

9

Huomioitavia asioita

- Kysymysten on oltava samalla numeerisella asteikolla (jos näin ei ole → "uudelleenkoodaa" poikkeavat kysymykset).
- Summamuuttujan reliabiliteetti mitattava → yleensä käytetään ns. Cronbachin α -lukua (peukalosääntö: jos α on suurempi kuin 0,7 summamuuttuja on hyväksyttävissä)

© Mikko Mattila 2005

10

Puuttuvat havainnot

- Lähes kaikissa aineistoissa on muuttujia ja havaintoyksiköjä, joista ei ole pystytty mittaamaan kaikki arvoja → puuttuvien havaintojen ongelma
- Seuraukset:
 - Aineiston pieneneminen
 - Tulosten mahdollinen vääristyminen

© Mikko Mattila 2005

11

Puuttuvat havainnot

- Syitä:
 - Ei tarkoitukseen vastata (esim. lasten ikä, jos vastaajalla ei ole lapsia)
 - Kieltäytyy vastaamasta
 - Vastaus epäselvä
 - Vastaaja ei tiedä tai hänellä ei ole mielipidettä
 - Vahinko tai viitseliäisyyden puute
 - Koodaus- tai tallennevirhe
 - Syystä ei ole tietoa

© Mikko Mattila 2005

12

Mitä tehdä?

- Puuttuvien havaintojen poistaminen analyysistä
- Helppo ratkaisu, esim. SPSS-ohjelmassa määritellään tietyt muuttujien arvot puuttuviksi tiedoiksi
- Pienentää aineiston kokoa
- Voi vääristää tuloksia, jos puuttuvat havainnot eivät jakaudu sattumanvaraisesti

© Mikko Mattila 2005

13

Mitä tehdä?

- Muuttujien hylkääminen analyysistä
- Kannattaa harkita, jos muuttujassa suuri määrä puuttuvia havaintoja
- Ongelma, jos muuttuja tärkeä tutkimusongelman kannalta

© Mikko Mattila 2005

14

Mitä tehdä?

- Havaintojen parittainen poistaminen
- Voidaan käyttää jos analyysi perustuu kovarianssi- tai korrelaatiomatriisille
 - Esim. regressio- tai faktorianalyysi
- Korrelaatiokertoimet lasketaan vain kaikista niistä havaintoyksiköistä, joista tiedot molemmista muuttujista
- Parempi vaihtoehto kuin kaikkien puuttuvien havaintojen poistaminen

© Mikko Mattila 2005

15

Mitä tehdä?

- Keskiarvojen käyttö
 - Korvataan puuttuvat havainnot muuttujan keskiarvolla
- Ei pienennä aineiston kokoa
- Pienentää keinotekoisesti muuttujien välisten yhteyksien vahvuutta (esim. korrelaatio)

© Mikko Mattila 2005

16

Mitä tehdä?

- Ryhmäkeskiarvojen käyttö
 - Korvataan puuttuvat havainnot ryhmien keskiarvoilla
 - Esimerkiksi naisvastaajien puuttuvat havainnot korvataan naisvastaajien keskiarvolla jne.
- Lisää ryhmien samankaltaisuutta
- Vahvistaa keinotekoisesti muuttujien välisten yhteyksien vahvuutta (esim. korrelaatio)

© Mikko Mattila 2005

17

Puuttuvien havaintojen koodaaminen

- Puuttuvat havainnot erotettava selkeästi "oikeista" muuttujien arvoista
- Usein käytetään koodeja 0, 9, 99 tai 999
- Tärkeää määritellä tilasto-ohjelmalle puuttuvien havaintojen koodit

© Mikko Mattila 2005

18

Keski- ja hajontaluvut

- Yhden muuttujan jakaumaa voidaan kuvailla:
 - graafisilla menetelmillä,
 - keskiluvuilla,
 - ja hajontaluvuilla.
- Keskiluvut kuvaavat muuttuja arvojen keskimääräistä suuruutta
- Hajontaluvut kuvaavat muuttujan arvojen vaihtelun suuruutta

© Mikko Mattila 2005

19

Soveltuvan keskiluvun valinta

		Muuttujan mittaustaso			
		Luokitteluasteikko	Järjestysasteikko	Välimatkaasteikko	Suhdeasteikko
Keskiluku	Moodi	X	X	X	X
	Mediaani	-	X	X	X
	Aritmeettinen keskiarvo	-	-	X	X
	Geometrisen ja harmonisen keskiarvo	-	-	-	X

© Mikko Mattila 2005

20

Esimerkkiaineisto

Työntekijän havaintonumero	Sukupuoli	Koulutus	Lapsien määrä
Työntekijä 1	Mies	Peruskoulu	0
Työntekijä 2	Nainen	Keskiaste	4
Työntekijä 3	Nainen	Keskiaste	1
Työntekijä 4	Mies	Korkeakoulu	1
Työntekijä 5	Nainen	Keskiaste	2
Työntekijä 6	Nainen	Korkeakoulu	1
Työntekijä 7	Nainen	Korkeakoulu	1
Työntekijä 8	Mies	Peruskoulu	0
Työntekijä 9	Mies	Korkeakoulu	0
Työntekijä 10	Nainen	Keskiaste	2

© Mikko Mattila 2005

21

Moodi eli tyyppi-arvo

- aineiston yleisin muuttujan arvo
- voidaan käyttää millä tahansa mittaustasoilla mitatulla muuttujalla
- esimerkkiaineistossa sukupuolen moodi on nainen ja lapsiluvun yksi
- moodeja voi olla useita: koulutusmuuttujan moodeja ovat keskiaste ja korkeakoulu

© Mikko Mattila 2005

22

Mediaani

- suuruusjärjestyksessä aineiston kesimmäisen havainnon arvo
- Kun muuttujia on parillinen määrä, mediaani on kumpikin kesimmäisistä luvuista (järjestysasteikko) tai kahden kesimmäisen luvun keskiarvo (välimatka- tai suhdeasteikko)
- Äärimmäisen poikkeavat arvot eivät yleensä vaikuta mediaaniin
- Lapsiluku: 0 0 0 1 1 1 1 2 2 4
- Koulutus: P P KE KE KE KE KO KO KO KO

© Mikko Mattila 2005

23

Aritmeettinen keskiarvo

- Yleisin keskiluku
- Käytetään välimatka- tai suhdeasteikoilla
- Keskiarvo = havaintojen summa / havaintojen määrä
- Lapsimäärämuuttujan keskiarvo on $(0+4+1+1+2+1+1+0+0+2)/10=1,2$

© Mikko Mattila 2005

24

Geometrinen ja harmoninen keskiarvo

- Soveltuvat suhdeasteikon muuttujille
- Harvoin käytettyjä
- Yleisin sovellusalue on indeksilaskenta ja kasvuilmiöiden analyysi

© Mikko Mattila 2005

25

Hajontaluvut

		Muuttujan mittaustaso			
		Luokittelusteikko	Järjestysasteikko	Välimatkaasteikko	Suhdeasteikko
	Variaatiosuhde	X	X	X	X
	Vaihteluväli	-	X	X	X
Hajontaluku	Vaihteluvälin pituus	-	-	X	X
	Keskihajonta	-	-	X	X
	Variaatiokerroin	-	-	X	X

© Mikko Mattila 2005

26

Variaatiosuhde

- Käytetään lähinnä luokittelusteikollisten muuttujien yhteydessä
- Kertoo, kuinka suuri osuus havainnoista sijoittuu muuttujan moodiluokkaan
- Kaava: $v = 1 - (\text{havaintojen määrä moodiluokassa} / \text{havaintojen määrä})$
- Vaihtelee nollan (ei hajontaa) ja yhden välillä (suuri hajonta)
- Sukupuolimuuttuja: $v = 1 - (6/10) = 0,4$

© Mikko Mattila 2005

27

Vaihteluväli

- Sopii järjestys-, välimatka- ja suhdeasteikon muuttujille
- Ilmoittaa pienimmän ja suurimman arvon välin
- $W = [x_1, x_n]$
- Koulutus: $W = [\text{"peruskoulu"}, \text{"korkeakoulu"}]$
- Lapsiluku: $W = [0, 4]$

© Mikko Mattila 2005

28

Vaihteluvälin pituus

- Sopii välimatka- ja suhdeasteikon muuttujille
- Ilmoittaa suurimman ja pienimmän arvon erotuksen
- $w = x_n - x_1$
- Lapsiluku: $w = 4 - 0 = 4$

© Mikko Mattila 2005

29

Keskihajonta

- Sopii välimatka- ja suhdeasteikon muuttujille
- Kuvaa, kuinka kaukana yksittäiset havaintojen arvot ovat koko aineiston keskiarvosta

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Varianssi on keskihajonnan neliö

© Mikko Mattila 2005

30

Lapsiluvun keskihajonta

Lapsiluku	(x_i -keskiarvo)	(x_i -keskiarvo) ²
0	-1,2	1.44
4	2.8	7.84
1	-0.2	0.04
1	-0.2	0.04
2	0.8	0.64
1	-0.2	0.04
1	-0.2	0.04
0	-1.2	1.44
0	-1.2	1.44
2	0.8	0.64
keskiarvo=1,2		summa=13.6

$$13,6/10=1,36$$

neliöjuuri 1,36:sta
on noin 1,17
= keskihajonta

© Mikko Mattila 2005

31

Variaatiokerroin

- Käytetään haluttaessa verrata kahden ryhmän keskihajontoja
- Suhteuttaa keskihajonnan keskiarvoon
- $v = \text{keskihajonta} / \text{keskiarvo}$
- Esim. palkkahajonta Suomessa ja Virossa

© Mikko Mattila 2005

32