

Kvantitatiiviset menetelmät

- Pienryhmiin ilmoittautuminen alkaa ke 2.2. klo 9.00

- **Ryhmä 1: Jussi Kinnunen:**
Ensimmäinen kokoontuminen to 24.2. klo 14-16, paikka päärautaud IV SPSS-harjoitukset: ti 29.3. klo 11-13 ja to 7.4. klo 15-19
- **Ryhmä 2: Mikko Mattila**
Ensimmäinen kokoontuminen to 24.2. 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 (täynnä)
- **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

Monimuuttujamenetelmät

- Yhden muuttujan menetelmät (univariate statistics):
 - keskiluvut ja hajontaluvut
- Monimuuttujamenetelmät:
 - kahden (bivariate) tai useamman (multivariate) muuttujan välisten suhteiden tarkastelu
 - Useimmiten menetelmissä oletetaan kausaaliasetelma (esim. regressioanalyysi)
 - Aina ei tarvitse tehdä tätä oletusta (esim. korrelaatiokertoimet, ryhmittelyanalyysi)
 - Soveltuva menetelmä riippuu muuttujien mittaustasosta

© Mikko Mattila 2005

2

Soveltuvan menetelmän valinta

		SELITETTÄVÄ MUUTTUJA	
SELITTÄVÄ MUUTTUJA	Laatuero- tai järjestysasteikko	Laatuero- tai järjestysasteikko	Välimatka- tai suhdelukuasteikko
	Välimatka- tai suhdelukuasteikko	- Ristiintaulukointi - Log-lineaariset mallit	- Varianssianalyysi
		- Logistinen regressio - Multinomiaalinen regressio	- Regressioanalyysi

© Mikko Mattila 2005

3

Korrelaatiokertoimet

- Korrelaatio kuvastaa kahden muuttujan välisen lineaarisen riippuvuuden astetta
- Voimakas korrelaatio tarkoittaa sitä, että tiedettäessä toisen muuttujan arvo, voidaan toisen muuttujan arvo päätellä melko täsmällisesti
- Soveltuvan korrelaatiokertoimen valinta riippuu muuttujien mittaustasosta
- Vaihtoehtoina esim.: Pearsonin tulomomenttikorrelaatiokerroin, kontingenssikerroin, Yulen Q, phi, Cramerin V, lambda, epävarmuuskerroin, Spearmanin rho ...

© Mikko Mattila 2005

4

Tulomomenttikorrelaatiokerroin

- Pearsonin tulomomenttikorrelaatiokerroin on yleisimmin käytetty kerroin
- Soveltuu kahden vähintään välimatka-asteikollisen muuttujan lineaarisen riippuvuuden tarkasteluun
- Merkitään r_{xy}
- Arvo vaihtelee välillä $-1 \dots +1$

© Mikko Mattila 2005

5

Tulomomenttikorrelaatiokerroin

- Huomioitavia seikkoja:
 - korrelaatiokerroin on symmetrinen eli $r_{xy} = r_{yx}$
 - korrelaatio ei kerro mitään kausaalisuudesta
 - jos muuttujien välillä epälineaarinen suhde, korrelaatiokerroin ei välttämättä mittaa sitä
 - yksittäiset poikkeavat havaintoarvot voivat vaikuttaa kertoimeen "vääristäen" tulosta
 - korrelaatiokertoimen tilastollisen merkitsevyyden arviointi tärkeää

© Mikko Mattila 2005

6

Esimerkki

Correlations

		D1 LEFT-RIGHT PLACEMENT	D11 AGE EXACT	D29 INCOME HH - FIN
D1 LEFT-RIGHT PLACEMENT	Pearson Correlation	1.000	-.039	.197**
	Sig. (2-tailed)	.	.247	.000
	N	862	862	810
D11 AGE EXACT	Pearson Correlation	-.039	1.000	-.215**
	Sig. (2-tailed)	.247	.	.000
	N	862	1022	925
D29 INCOME HH - FIN	Pearson Correlation	.197**	-.215**	1.000
	Sig. (2-tailed)	.000	.000	.
	N	810	925	925

**. Correlation is significant at the 0.01 level (2-tailed).

© Mikko Mattila 2005

7

Osittaiskorrelaatiokerroin

- Osittaiskorrelaatiokerroin kertoo kahden muuttujan riippuvuuden kun yhden tai useamman muun muuttujan vaikutus on poistettu (vakioitu)
- Merkitään $r_{xy.z}$
- Esim. jäätelön kulutus ja hukkumiskuolemat korreloivat vahvasti, mutta niiden osittaiskorrelaatiokerroin on nolla, kun lämpötila on vakioitu
- Osittaiskorrelaatiokerrointa käytetään suhteellisen harvoin, koska sama informaatio saadaan regressioanalyysin avulla kätevämmän

© Mikko Mattila 2005

8

Spearmanin järjestyskorrelaatio

- Periaatteessa tarkoitettu järjestysasteikollisille muuttujille, mutta usein järkevää käyttää myös korkeammalla mittaustasolla
- Molempien muuttujien havainnot järjestetään suuruusjärjestykseen ja annetaan havainnoille uudet arvot niin, että pienin saa arvon yksi, toiseksi pienin kaksi jne.
- Lasketaan tavallinen Pearsonin korrelaatiokerroin järjestyslukuista
- Spearmanin korrelaatiokerroin ei ole yhtä herkkä poikkeamille lineaarisuudesta kuin Pearsonin kerroin

© Mikko Mattila 2005

9

Ristiintaulukointi

- Ristiintaulukointia käytetään tutkittaessa kahden tai useamman muuttujan jakaumaa ja keskinäistä riippuvuutta
- Muuttujat mitattu joko luokittelu- tai järjestysasteikolla
- Usein käytetään myös välimatka- tai suhteasteikon muuttujia, mutta ne on ensin luokiteltava
- Tutkitaan, onko tarkastelun kohteena olevan selitettävän muuttujan jakauma erilainen selittävän muuttujan eri luokissa

© Mikko Mattila 2005

10

Esimerkki: keskusteluaktiivisuus

- Esimerkki: eroavatko miehet ja naiset toisistaan siinä, kuinka usein he keskustelevat politiikasta ystäviensä kanssa?
- Absoluuttisen lukumäärät eivät osoita suoraan sukupuolien välisiä eroja → on laskettava prosenttijakaumat

	Mies	Nainen
Usein	43	29
Silloin tällöin	323	298
En koskaan	108	174

© Mikko Mattila 2005

11

Esimerkki: keskusteluaktiivisuus

Taulukko 2. Aktiivisuus keskustella poliittisista asioita ystävien kanssa sukupuolen mukaan (%).

	Mies	Nainen
Usein	9	6
Silloin tällöin	68	60
En koskaan	23	35
Yhteensä (n) $\chi^2=18,4$; vapausasteita=2; $p<0,01$	100 (474)	100 (501)

© Mikko Mattila 2005

12

Prosenttijakaumien oikea suunta

- Ristiintaulukkoa tehdessä on tärkeää laskea prosenttijakaumat oikeaan suuntaan
- Prosentit lasketaan selittävän muuttujan luokissa
- Viime kädessä tutkimusongelma ratkaisee suunnan

© Mikko Mattila 2005

13

Ristiintaulukon esittäminen

- Tavanomainen ratkaisu: selittävä muuttuja sarakkeille, selitettävä riveille
- Taulukkoon tulee liittää yhteenlasketut tiedot prosenteista, koska ne helpottavat tulkintaa
- Tapausten lukumäärät (n) ilmoitettava
- Taulukossa ilmoitettava tilastollisen testin tulos

© Mikko Mattila 2005

14

Ristiintaulukon merkitsevyyden testaus

- Voidaanko otoksen tulosten perusteella päätellä tarpeeksi luotettavasti, että havaitut erot pätevät myös perusjoukossa?
- Testausmenetelmänä χ^2 -riippumattomuustesti
- Perustuu havaittujen ja odotettujen frekvenssien erotukselle

© Mikko Mattila 2005

15

Ristiintaulukon merkitsevyyden testaus

- Nollahypoteesi: ei eroja selittävän muuttujan eri luokissa
- Jos p-arvo on riittävän pieni, voidaan päätellä, että erot ovat tilastollisesti merkitsevät
- Huom.: tilastollinen merkitsevyys eri asia kuin sisällöllisesti tärkeä ero

© Mikko Mattila 2005

16

Ristiintaulukon elaboraatio

- Löydettyä kausaalisuhdetta täsmennetään ja varmennetaan tuomalla analyysiin uusia muuttujia

	Alle 35 v.		35-59 v.		60 v. täyttäneet	
	M	N	M	N	M	N
Usein	4	2	10	8	17	9
Silloin tällöin	64	58	71	62	72	59
En koskaan	33	40	19	31	11	32
Yhteensä (n)	100 (183)	100 (184)	100 (194)	100 (199)	100 (96)	100 (118)
	$\chi^2=2,8$; vapausast.=2; p=0,24		$\chi^2=7,8$; vapausast.=2; p=0,02		$\chi^2=14,4$; vapausast.=2; p<0,01	

© Mikko Mattila 2005

17

χ^2 -riippumattomuustesti

	Mies	Nainen	Yhteensä
Usein	43 35,0 9 %	29 37,0 6 %	72
Silloin tällöin	323 301,9 68 %	298 319,1 60 %	621
En koskaan	108 137,1 23 %	174 144,9 35 %	282
Yhteensä n	100 % 474	100 % 501	975

- Nollahypoteesi: ei eroja selittävän muuttujan eri luokissa
- Tarvitaan tieto odotetuista frekvensseistä

© Mikko Mattila 2005

18

χ²-riippumattomuustesti

	Mies	Nainen	Yhteensä
Usein	43 35,0 9 %	29 37,0 6 %	72
Silloin tällöin	323 301,9 68 %	298 319,1 60 %	621
En koskaan	108 137,1 23 %	174 144,9 35 %	282
Yhteensä n	100 % 474	100 % 501	975

• Odotetut frekvenssit saadaan kaavasta:

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{N}$$

• E_{ij} = odotettu frekvenssi

• $O_{i.}$ = i:n rivin kokonais-

summa
• $O_{.j}$ = j:n sarakkeen kokon-

naissumma

N = havaintojen määrä

Esim. $(621 \cdot 474) / 975 = 301,9$

© Mikko Mattila 2005

19

χ²-riippumattomuustesti

	Mies	Nainen	Yhteensä
Usein	43 35,0 9 %	29 37,0 6 %	72
Silloin tällöin	323 301,9 68 %	298 319,1 60 %	621
En koskaan	108 137,1 23 %	174 144,9 35 %	282
Yhteensä n	100 % 474	100 % 501	975

• χ²-testisuure saadaan kaavasta:

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

• Eli:

$$(43-35,0)^2/35,0 +$$

$$(29-37,0)^2/37,0 +$$

...

$$(174-144,9)^2/144,9$$

$$= 18,4$$

© Mikko Mattila 2005

20

χ²-riippumattomuustesti

- χ²-testisuure sai esimerkissä arvon 18,4
- Sitä vastaavan p-arvon tietämiseksi tarvitaan vielä tieto taulukon vapausasteista (degrees of freedom, d.f.)
- Se saadaan kaavasta:
(rivien määrä - 1) * (sarakkeiden määrä - 1)
- Vapausasteet on siis (3-1) * (2-1) = 2
- Kun tiedetään χ²-luku ja vapausasteet, p-arvo voidaan etsiä tilastollisesta χ²-jakaumataulukosta
- Onneksi tätä ei tarvitse tehdä käsin vaan SPSS-ohjelma laskee kaiken automaattisesti

© Mikko Mattila 2005

21

Esimerkki

- Tutkitaan suomalaisten poliittisen sijoittumisen ja EU-kannatuksen suhdetta ristiintaulukoinnilla
- Poliittisen sijoittumisen alkuperäinen skaala (1-10) on uudelleenkoodattu seuraavasti:
 - 1-4 = vasemmisto
 - 5-6 = keskusta
 - 7-10 = oikeisto
- EU-kysymyksessä kysytty on Suomen jäsenyyttä EU:ssa mielestäsi 1) Hyvä asia, 2) Huono asia, 3) Ei hyvä eikä huono asia

© Mikko Mattila 2005

22

Esimerkki

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	25.991 ^a	4	.000
Likelihood Ratio	25.790	4	.000
Linear-by-Linear Association	10.345	1	.001
N of Valid Cases	827		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 51.00.

© Mikko Mattila 2005

23

Esimerkki

Q17 EU MEMBERSHIP - GOOD/BAD * D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT Crosstabulation

		LEFT-RIGHT PLACEMENT - RECODED 3 CAT			Total
		(1 - 4) Left	(5 - 6) Centre	(7 - 10) Right	
Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count 78	116	134	328
		% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	36.6%	32.4%	52.3%
	A bad thing	Count 54	97	47	198
		% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	25.4%	27.1%	18.4%
	Neither good nor bad	Count 81	145	75	301
		% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	38.0%	40.5%	29.3%
Total		Count 213	358	256	827
		% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	100.0%	100.0%	100.0%

© Mikko Mattila 2005

24

Esimerkki

Chi-Square Tests

D10 SEX		Value	df	Asymp. Sig. (2-sided)
Male	Pearson Chi-Square	19.823 ^a	4	.001
	Likelihood Ratio	19.441	4	.001
	Linear-by-Linear Association	5.311	1	.021
	N of Valid Cases	379		
Female	Pearson Chi-Square	14.068 ^b	4	.007
	Likelihood Ratio	13.802	4	.008
	Linear-by-Linear Association	5.201	1	.023
	N of Valid Cases	448		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 26.89.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 24.12.

© Mikko Mattila 2005

25

•Seuraavaksi tutkitaan onko poliittisen sijoittumisen ja EU-kannatuksen suhde samanlainen miehillä ja naisilla eli tehdään ristiintaulukon elaborointi

Q17 EU MEMBERSHIP - GOOD/BAD * D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT * D10 SEX Crosstabulation

D10 SEX	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT			Total
				(1 - 4) Left	(5 - 6) Centre	(7 - 10) Right	
Male	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	47	45	78	174
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	42.0%	36.3%	59.1%	45.9%
		A bad thing	Count	34	31	26	91
	Q17 EU MEMBERSHIP - GOOD/BAD		% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	30.4%	23.0%	19.7%	24.0%
		Neither good nor bad	Count	31	65	28	114
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	27.7%	40.7%	21.2%	30.1%
Female	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	112	135	132	379
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	100.0%	100.0%	100.0%	100.0%
	Q17 EU MEMBERSHIP - GOOD/BAD	A bad thing	Count	31	67	56	154
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	30.7%	30.0%	45.2%	34.4%
	Q17 EU MEMBERSHIP - GOOD/BAD	Neither good nor bad	Count	20	66	21	107
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	19.8%	29.6%	16.9%	23.9%
Total	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	50	90	47	187
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	49.5%	40.4%	37.9%	41.7%
	Q17 EU MEMBERSHIP - GOOD/BAD	A bad thing	Count	101	223	124	448
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	100.0%	100.0%	100.0%	100.0%
	Q17 EU MEMBERSHIP - GOOD/BAD	Neither good nor bad	Count	112	135	132	379
			% within D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT	100.0%	100.0%	100.0%	100.0%

© Mikko Mattila 2005

26

Esimerkki

- Edellisen taulukon mukaan vaikuttaisi siltä, että naiset suhtautuvat EU:hun miehiä epäileväisemmin poliittisesta sijoittumisesta riippumatta → tehdään tästä oma taulukko

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.310 ^a	2	.000
Likelihood Ratio	15.348	2	.000
Linear-by-Linear Association	15.209	1	.000
N of Valid Cases	964		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 106.86.

© Mikko Mattila 2005

27

Esimerkki

Q17 EU MEMBERSHIP - GOOD/BAD * D10 SEX Crosstabulation

		Count	D10 SEX		Total
			Male	Female	
Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	188	172	360
		% within D10 SEX	43.6%	32.3%	37.3%
	A bad thing	Count	105	134	239
		% within D10 SEX	24.4%	25.1%	24.8%
	Neither good nor bad	Count	138	227	365
		% within D10 SEX	32.0%	42.6%	37.9%
Total	Count	431	533	964	
	% within D10 SEX	100.0%	100.0%	100.0%	

© Mikko Mattila 2005

28

Esimerkki

- Seuraavaksi tutkitaan, pysyykö sukupuolien erot EU-kannatuksessa samanlaisina kun vakioidaan vastaajien poliittinen sijoittuminen:

Chi-Square Tests

D1 LEFT-RIGHT PLACEMENT		Value	df	Asymp. Sig. (2-sided)
(1 - 4) Left	Pearson Chi-Square	10.829 ^a	2	.004
	Likelihood Ratio	10.907	2	.004
	Linear-by-Linear Association	7.759	1	.005
	N of Valid Cases	213		
(5 - 6) Centre	Pearson Chi-Square	2.383 ^b	2	.304
	Likelihood Ratio	2.400	2	.301
	Linear-by-Linear Association	.400	1	.527
	N of Valid Cases	358		
(7 - 10) Right	Pearson Chi-Square	8.716 ^c	2	.013
	Likelihood Ratio	8.777	2	.012
	Linear-by-Linear Association	7.823	1	.005
	N of Valid Cases	256		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 25.61.

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 36.58.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 22.77.

29

Q17 EU MEMBERSHIP - GOOD/BAD * D10 SEX * D1 LEFT-RIGHT PLACEMENT - RECODED 3 CAT Crosstabulation

D1 LEFT-RIGHT PLACEMENT	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	D10 SEX		Total
				Male	Female	
(1 - 4) Left	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	47	31	78
			% within D10 SEX	42.0%	30.7%	36.6%
	Q17 EU MEMBERSHIP - GOOD/BAD	A bad thing	Count	34	20	54
			% within D10 SEX	30.4%	19.8%	25.4%
	Q17 EU MEMBERSHIP - GOOD/BAD	Neither good nor bad	Count	31	50	81
			% within D10 SEX	27.7%	49.5%	38.0%
(5 - 6) Centre	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	112	101	213
			% within D10 SEX	100.0%	100.0%	100.0%
	Q17 EU MEMBERSHIP - GOOD/BAD	A bad thing	Count	49	67	116
			% within D10 SEX	36.3%	30.0%	32.4%
	Q17 EU MEMBERSHIP - GOOD/BAD	Neither good nor bad	Count	31	66	97
			% within D10 SEX	23.0%	29.6%	27.1%
(7 - 10) Right	Q17 EU MEMBERSHIP - GOOD/BAD	A good thing	Count	55	90	145
			% within D10 SEX	40.7%	40.4%	40.5%
	Q17 EU MEMBERSHIP - GOOD/BAD	A bad thing	Count	135	223	358
			% within D10 SEX	100.0%	100.0%	100.0%
	Q17 EU MEMBERSHIP - GOOD/BAD	Neither good nor bad	Count	28	47	75
			% within D10 SEX	21.2%	37.9%	29.3%
Total	Q17 EU MEMBERSHIP - GOOD/BAD	Count	132	124	256	
		% within D10 SEX	100.0%	100.0%	100.0%	

© Mikko Mattila 2005

30