

Kvantitatiiviset menetelmät

Tentti pidetään tiistaina 22.2. klo 16-18 V20 Is 1

Pienryhmäkokoontumisissa tarvitaan EK03-aineiston haastattelulomake. Sen voi tulostaa verkosta. Linkki löytyy kurssin kotisivulta: <http://www.valt.helsinki.fi/staff/mmattila/kvanti/>

© Mikko Mattila 2005

1

Soveltuvan menetelmän valinta

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

© Mikko Mattila 2005

2

Logistinen regressioanalyysi

- Logistinen regressioanalyysi on normaalin regressioanalyysin erityistyyppi
- Käytetään tilanteissa, jossa selitettävä muuttuja voi saada vain kaksi arvoa
- Esimerkiksi:
 - Äänestikö vastaaja viime vaaleissa?
 - Onko vastaaja naimisissa?
- Selitettävä muuttuja koodattava dummy-muuttujaksi

© Mikko Mattila 2005

3

Logistinen regressioanalyysi

- Logistisen regression avulla tutkitaan jonkun tapahtuman tai asian tilan todennäköisyyteen vaikuttavia tekijöitä
- Vaikuttavatko selittävät muuttujat tilastollisesti merkittävästi tapahtuman todennäköisyyteen?
- Kuinka suuri tämä vaikutus on?

© Mikko Mattila 2005

4

Logistinen regressioanalyysi

- Yksinkertaistettuna logistinen regressiomalli on tavallinen regressiomalli, jossa selitettävänä muuttujana on tapahtuman riskin luonnollinen logaritmi
- Riski kuvaa tapahtuman esiintymistodennäköisyyttä
 - Eurovaaleissa 2004 noin 40% äänesti
 - Tällöin yksilön "riski" äänestää on $0,4/(1-0,4)=0,67$
 - Eduskuntavaaleissa 2003 noin 70% äänesti
 - Riski on $0,7/(1-0,7)=2,33$
 - Riskisuhde voi saada arvoja nollan ja äärettömän välillä

© Mikko Mattila 2005

5

Logistinen regressioanalyysi

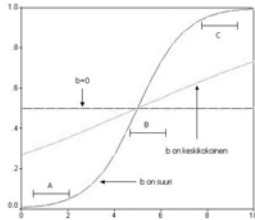
- Koska regressioanalyysi sopii parhaiten tilanteeseen, jossa selitettävän muuttujan arvoja ei ole etukäteen rajoitettu, otetaan riskisuhteesta logaritmi
 - tämä luku voi saada arvoja miinus äärettömän ja äärettömän välillä
- Logistisen regressioanalyysin kaava:

$$\ln \left[\frac{P(Y=1)}{1-P(Y=1)} \right] = a + bx$$

© Mikko Mattila 2005

6

Logistinen käyrä



Logistisessa regressioanalyysissä ei oleteta muuttujien välisen suhteen olevan lineaarinen, vaan seuraavan ns. logistista eli s-käyrää.

© Mikko Mattila 2005

7

Esimerkki

- Vastaaajia pyydettiin valitsemaan kahdesta vaihtoehdosta se, joka vastaa paremmin hänen mielipiteitään:
 - 1. "Muissa maissa valmistettuja tuotteita voidaan tuoda tänne ja myydä täällä, jos ihmiset haluavat ostaa niitä" (40%)
 - 2. "Ulkomaisten tuotteiden myynnille Suomessa pitäisi olla enemmän esteitä, jotta voitaisiin suojella tämän maan ihmisten työpaikkoja" (60%)
- Selittävinä muuttujina sukupuoli, ikä, ylpeys suomalaisuudesta, ideologinen sijoittuminen ja tulot

© Mikko Mattila 2005

8

Esimerkki

Muuttuja	Regressiokerroin	Merkitsevyys
Vakio	-0.00	$p=0,99$
Sukupuoli (nainen=1, mies=0)	0,48**	$p=0,001$
Ikä	0,02**	$p<0,001$
Ylpeys suomalaisuudesta (1=suuri ... 4=heikko)	-0,10	$p=0,33$
Sijoittuminen vasemmisto-oikeisto -ulottuvuudella (1-10)	-0,07	$p=0,11$
Tuloluokka (1-10)	-0,08*	$p=0,01$

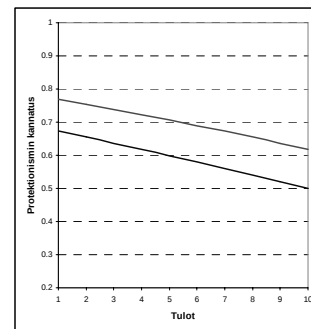
Logistisen regressioanalyysin ennustekykyä voidaan arvioida katsomalla, kuinka hyvin selittävien muuttujien avulla voidaan ennustaa heidän vastauksensa.

Tässä tapauksessa regressiomalli ennustaa 80% oikein.

© Mikko Mattila 2005

9

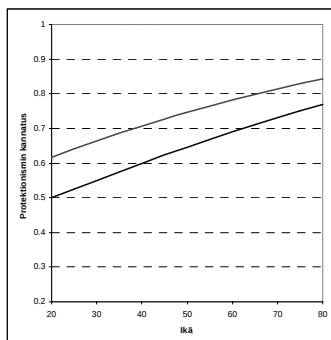
Esimerkki



© Mikko Mattila 2005

10

Esimerkki



© Mikko Mattila 2005

11

Multinomiaalinen logistinen regressio

- Normaalin logistisen regressioanalyysin laajennus
- Selitettävä muuttuja voi saada kolme tai useampia vaihtoja
- Esimerkiksi vastaaja äänesti joko hallituspuolueen ehdokasta, oppositiopuolueen ehdokasta tai ei äänestänyt lainkaan

© Mikko Mattila 2005

12

Esimerkki2

- Eurobarometri 2001 (55), tutkitaan neljää maata
- "EU:n laajentumisen jälkeen jäsenvaltioiden määrä kasvaa 15:stä yli 25:teen. Pitäisikö mielestänne silloin EU:n päätökset tehdä
 - 1. kaikkien maiden yksimielisyydellä, vai
 - 2. enemmistöpäätöksillä"
- Selittävät muuttujat: ikä, ideologinen tausta (1-10), tietämys EU:sta (1-10), sukupuoli, tuloluokka (1-10), politiikasta keskustelun aktiivisuus
 - kolme luokkaa 1=usein, 2=silloin tällöin, 3=ei koskaan
 - muodostettu kaksi dummy-muuttujaa KESKUSEI ja KESKSILL

© Mikko Mattila 2005

13

Esimerkki2

Q31 EU DECISION MAKING AFTER ENLARGEMENT * INRA COUNTRY CODE Crosstabulation

			INRA COUNTRY CODE				Total
			GER	NET	FIN	SWE	
Q31 EU DECISION MAKING AFTER ENLARGEMENT	Unanimously by all member states	Count	311	252	375	359	1297
		% within INRA COUNTRY CODE	36.4%	26.4%	40.0%	39.9%	35.6%
	By a majority of the member states	Count	544	702	563	541	2350
		% within INRA COUNTRY CODE	63.6%	73.6%	60.0%	60.1%	64.4%
Total	Count		855	954	938	900	3647
	% within INRA COUNTRY CODE		100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	50.414 ^a	3	.000
Likelihood Ratio	51.831	3	.000
Linear-by-Linear Association	6.011	1	.014
N of Valid Cases	3647		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 304.07.

© Mikko Mattila 2005

14

Esimerkki2

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	ikä	-.007	.004	2.657	1	.103	.993
	vas/oik	.085	.040	4.451	1	.035	1.089
	tietämys	.071	.044	2.569	1	.109	1.073
	NAINEN	-.349	.152	5.274	1	.022	.706
	tulot	.048	.021	5.170	1	.023	1.050
	KESKUSEI	.713	.305	5.449	1	.020	2.039
	KESKSILL	.108	.179	.366	1	.545	1.114
	Constant	-.168	.398	.178	1	.673	.845

a. Variable(s) entered on step 1: D11, D1R, Q15, NAINEN, D29_FIN, KESKUSEI, KESKSILL.

© Mikko Mattila 2005

15

Esimerkki2

Classification Table^a

		Predicted		
		Q31 EU DECISION MAKING AFTER ENLARGEMENT		Percentage Correct
Observed		Unanimously by all member states	By a majority of the member states	
Step 1	Q31 EU DECISION MAKING AFTER ENLARGEMENT	57	247	18.8
	By a majority of the member states	39	458	92.2
Overall Percentage				64.3

a. The cut value is .500

Selittävien muuttujien avulla pystytään ennustamaan 64% vastaajien mielipiteistä.

© Mikko Mattila 2005

16

Esimerkki2

	Suomi	Ruotsi	Alankomaat	Saksa
Vakio	-0,17	-0,41	1,15**	0,01
Ikä	-0,01	-0,01	-0,00	0,01
Vas/oik -skaala	0,09*	0,10*	-0,03	-0,02
Tietämys EU:sta	0,07	0,04	-0,07	0,07
Sukupuoli (1=nainen)	-0,35*	-0,41**	-0,24	-0,05
Keskustelee usein	0,71*	0,59*	1,16**	0,67*
Keskustelee silloin täl.	0,11	0,42*	0,43*	0,43*
Tuloluokka	0,05*	0,03	0,03	-0,03
Oikein luokiteltujen osuus	64,3%	62,3%	75,0%	64,6%

© Mikko Mattila 2005

17

	SDP	Keskusta	Kokoomus	Vasemmistoliitto	Vihreät
Vakio	-1,80 (1,07)	-6,11** (1,10)	-7,08** (1,21)	-2,43 (1,72)	-2,58 (1,40)
EU-jäsenyys hyvä	0,47 (0,30)	-0,44 (0,28)	0,90** (0,32)	-0,69 (0,53)	0,34 (0,40)
EU-jäsenyys huono	-1,10* (0,50)	0,23 (0,34)	-0,88 (0,53)	-0,21 (0,55)	-0,90 (0,68)
Vas.-oik. sijoittuminen	-0,24** (0,06)	0,29** (0,06)	0,59** (0,08)	-0,76** (0,12)	-0,20* (0,09)
Kiinnostus politiikasta	-0,14 (0,20)	0,19 (0,16)	0,43* (0,18)	0,81* (0,32)	0,27 (0,26)
Ikä	0,02** (0,01)	0,01 (0,01)	0,01 (0,01)	0,03 (0,02)	-0,05** (0,01)
Sukupuoli (nainen)	0,15 (0,27)	0,08 (0,24)	-0,39 (0,26)	0,62 (0,45)	1,40** (0,42)
Oikein luokiteltujen osuus (%)	87,3	85,0	86,3	95,0	93,4
Nagelkerken R ²	0,14	0,12	0,33	0,43	0,22

© Mikko Mattila 2005

18