



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Otanta-aineistojen analyysi

(78136 , 78405)

Kevät 2010

**TEEMA 3: Frekvenssiaineistojen
asetelmaperusteinen analyysi:
Perusteita**

Risto Lehtonen

risto.lehtonen@helsinki.fi



OHC Survey – Tilastollinen analyysi

- Kysymys:
- **Millä vaihtoehtoisilla tavoilla voidaan reagoida tilastollisen analyysin yhteydessä OHC-tutkimuksen otanta-asetelman ominaisuuksiin?**
 - Ositettu kaksiasteinen ryväotanta
 - Painotus - painomuuttuja
 - OHC-datassa painot = 1
 - Ositus - ositusmuuttuja OSITE
 - Ryvästyminen - ryväsmuuttuja RYVAS



OHC Survey – Tilastollinen analyysi

- Tarkastellaan esimerkinomaisesti kahta vaihtoehtoista lähestymistapaa
- **Asetelmaperusteinen** (*Design-based*) tilastollinen analyysi
 - Binäärinen vaste: Logistinen kiinteiden tekijöiden malli (logit ANCOVA)
- **Malliperusteinen** (*Model-based*) tilastollinen analyysi
 - Binäärinen vaste: Logistinen sekamalli (logit ANCOVA)



Asetelmaperusteinen (*design-based*) analyysi

- Ryväotanta-aineistojen analyysi
 - Aineiston analyysissä luovutaan havaintojen riippumattomuusoletuksesta (*iid*-oletus)
 - Sallitaan, että havainnot voivat korreloida keskenään (rypäiden sisällä)
 - Korreloituneisuuteen reagoidaan asetelmaperusteisilla menetelmillä
 - Asetelmaperusteiset piste-estimaatit ja keskivirheet
 - Asetelmaperusteiset tilastolliset testit
 - Asetelmakerroin (*Design effect, deff*)
 - Sisäkorrelaatio (*Intra-cluster correlation*)



Asetelmaperusteinen analyysi

- Asetelmaperusteinen metodiikka on suositus monissa isoissa tutkimuksissa
 - Terveys 2000 –tutkimus
 - PISA-tutkimukset...

- Kirjallisuutta
 - Chambers R.L. and Skinner C.J. (Eds.) (2004). *Analysis of Survey Data*. Chichester: Wiley.
 - Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester: Wiley. Luvut 5, 7-8



Asetelmaperusteinen analyysi

- **Ohjelmistot**

- SAS: SURVEY-proseduurit
 - SURVEYMEANS, SURVEYFREQ
 - SURVEYREG, SURVEYLOGISTIC

- SPSS
 - Complex Samples -moduli

- Stata
 - SVY-proseduurit

- Muut
 - R-funktioita



Malliperusteinen (*model-based*) analyysi

- Mutkikkaan otanta-asetelman ominaisuuksien mallinnus
 - **Luovutaan iid-oletuksesta**
 - Rypäiden sisäkorreloituneisuuteen reagoidaan mallintamalla
 - Kiinteiden tekijöiden mallit ja GEE-estimointi (*Generalized Estimating Equations*)
 - Sekamallit (*Mixed models*)
 - Monitasomallit (*Multilevel models*)
 - Hierarkkiset mallit (*Hierarchical models*)



Malliperusteinen analyysi

- Kirjallisuutta
 - Demidenko E. (2004). *Mixed Models. Theory and Applications*. New York: Wiley.
 - Diggle P. J., Liang, K.-Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
 - Goldstein H. (2003). *Multilevel Statistical Models*. 3rd edition. London: Arnold; New York: John Wiley & Sons.



Malliperusteinen analyysi

■ Ohjelmistot

■ SAS

- GENMOD
- MIXED
- GLIMMIX

■ SPSS ja Stata

- Vastaavat mallinnusproseduurit

■ Muut

- R-funktioita, LISREL, Mplus...



Analyysimenetelmiä ja työkaluja

■ Tilastollinen kuvailu

■ Kaksiulotteisten frekvenssitaulujen testit

■ SAS-proseduureja

- Keskiarvot, osuudet
 - SURVEYMEANS (Asetelmaperusteinen)
 - MEANS (SRS-oletus)
- Frekvenssitaulut, testit
 - SURVEYFREQ (asetelmaperusteinen)
 - FREQ (SRS-oletus)

■ Esimerkki

- Lehtonen & Pahkinen (2004), Example 7.3

■ VLISS Training Key 250



Esimerkki: PROC SURVEYFREQ

(1) Metodisesti pätevä analyysi
Asetelmaperusteinen, ryväotanta

```
proc surveyfreq data=ohc;  
  tables phys*psych3 / chisq;  
  strata osite;  
  cluster ryvas;
```

(2) Ei-pätevä versio (SRS-oletus)

```
proc surveyfreq data=ohc;  
  tables phys*psych3 / chisq;
```



(1) Asetelmaperusteinen analyysi
Table of PHYS by PSYCH3

PHYS	PSYCH3	Frequency	Percent	Std Err of Percent
0	1	1785	22.7650	0.6850
	2	1716	21.8850	0.7019
	3	1629	20.7754	0.7435
	Total	5130	65.4253	1.4385
1	1	910	11.6057	0.6078
	2	821	10.4706	0.5323
	3	980	12.4984	0.6330
	Total	2711	34.5747	1.4385
Total	1	2695	34.3706	0.7140
	2	2537	32.3556	0.5863
	3	2609	33.2738	0.6751
	Total	7841	100.000	



(2) SRS-perusteinen analyysi
Table of PHYS by PSYCH3

PHYS	PSYCH3	Frequency	Percent	Std Err of Percent
0	1	1785	22.7650	0.4736
	2	1716	21.8850	0.4670
	3	1629	20.7754	0.4582
Total		5130	65.4253	0.5371
1	1	910	11.6057	0.3617
	2	821	10.4706	0.3458
	3	980	12.4984	0.3735
Total		2711	34.5747	0.5371
Total	1	2695	34.3706	0.5364
	2	2537	32.3556	0.5284
	3	2609	33.2738	0.5322
Total		7841	100.000	

Risto Lehtonen

13



(1) Metodisesti pätevä
asetelmaperusteinen testi

Wald Chi-Square Test

Chi-Square 13.2280

F Value 6.6140

Num DF 2

Den DF 245

Pr > F 0.0016

Adj F Value 6.5870

Num DF 2

Den DF 244

Pr > Adj F 0.0016

Sample Size = 7841

(2) SRS-perusteinen ei-
pätevä testi

Wald Chi-Square Test

Chi-Square 16.3635

F Value 8.1818

Num DF 2

Den DF 7840

Pr > F 0.0003

Adj F Value 8.1807

Num DF 2

Den DF 7839

Pr > Adj F 0.0003

Sample Size = 7841

Risto Lehtonen

14



SAS-sovellukset – Kaksiulotteisten frekvenssitaulujen testit

- **Teema 3:** Frekvenssiaineistojen asetelmaperusteinen analyysi: Perusteita ([PDF-materiaali](#))

- SAS-sovellus – Kaksiulotteisten frekvenssitaulujen testit
 - PROC SURVEYFREQ
 - [Overview](#) [Getting Started](#) [Syntax](#) [Details](#)
 - [Details](#) (SAS Help)



Kirjallisuutta

- Chambers R.L. and Skinner C.J. (Eds.) (2004). *Analysis of Survey Data*. Chichester: Wiley.
- Demidenko E. (2004). *Mixed Models. Theory and Applications*. New York: Wiley.
- Diggle, P. J., Liang, K.-Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- Goldstein, H. (2003). *Multilevel Statistical Models. 3rd Edition*. London: Edward Arnold.
<http://www.cmm.bristol.ac.uk/MLwiN/index.shtml>
- Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester: Wiley.
- OECD (2002a). PISA 2000 Technical Report. Paris: OECD.
<http://www.pisa.oecd.org/>
- Snijders, T. and Bosker, R. (2002). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.