

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

# Otanta-aineistojen analyysi (78405) Kevät 2012 TEEMA 3: Frekvenssiaineistojen asetelmaperusteinen analyysi: Perusteita

Risto Lehtonen

[risto.lehtonen@helsinki.fi](mailto:risto.lehtonen@helsinki.fi)





# Tilastollinen analyysi

- **Millä vaihtoehtoisilla tavoilla voidaan reagoida tilastollisen analyysin yhteydessä havaintojen korreloituneisuuteen?**
- **Esim. OHC-tutkimus**
  - Ositettu kaksiasteinen ryväsoitanta
  - Analyysipainot (OHC-painot = 1)
  - Ositus – ositusmuuttuja OSITE
  - Ryvästyminen – ryväsmuuttuja RYVAS
  - Havaintojen rypäänsisäinen korreloituneisuus (*intra-cluster correlation*)



# Tilastollinen analyysi

- Tarkastellaan kahta vaihtoehtoista lähestymistapaa:
- **Asetelmaperusteinen** (*Design-based*) tilastollinen analyysi
- **Malliperusteinen** (*Model-based*) tilastollinen analyysi
- Molemmissa luovutaan havaintojen riippumattomuusoletuksesta eli **sallitaan havaintojen sisäkorreloituneisuus**



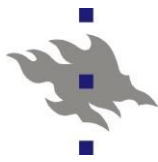
# Asetelmaperusteinen analyysi

- **Asetelmaperusteisessa** (*Design-based*) analyysissa sisäkorrelaatorakenteet otetaan **häiriötekijöinä** (*nuisance effect*), joiden vaikutus ”puhdistetaan pois” analyysin yhteydessä
- Reagoidaan otanta-asetelman ominaisuuksiin:
  - Ositus, ryvästyminen, analyysipainot
- **Kiinteiden tekijöiden** mallit (*fixed-effects models*)
- Yleistetyt lineaariset mallit (*generalized linear models*)
- Asetelmaperusteista metodiikkaa käytetään monissa isoissa tutkimuksissa
  - Terveys 2000 ja Terveys 2010 -tutkimukset
  - PISA-tutkimukset
  - European Social Survey ESS



# Asetelmaperusteinen analyysi – Kirjallisuutta ja ohjelmistoja

- Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Wiley.
- Lumley T. (2010). *Complex Surveys: A Guide to Analysis Using R*. Wiley.
- Chambers R.L. and Skinner C.J. (Eds.) (2004). *Analysis of Survey Data*. Wiley.
- Ohjelmasovelluksia
  - SAS: SURVEY-proseduurit
  - SPSS Complex Samples –moduli
  - Stata: SVY-proseduurit
  - Mplus
  - R-kielisiä funktioita ja ohjelmapaketteja



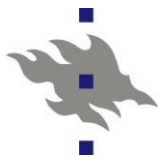
# Malliperusteinen analyysi

- **Malliperusteisessa** (*model-based*) analyysissa rypäiden sisäkorreloituneisuuteen reagoidaan **mallintamalla** korrelaatorakenteita
- Tilastolliset **sekamallit** (*mixed models*)
- Yleistetyt lineaariset sekamallit (*generalized linear mixed models, GLMM*)
  - Tilastolliset sekamallit (*Mixed models*)
  - Monitasomallit (*Multilevel models*)
  - Hierarkkiset mallit (*Hierarchical models*)
  - Kaikki nämä termit viittaavat samaan yleistettyjen lineaaristen sekamallien perheeseen



# · Malliperusteinen analyysi – · Kirjallisuutta ja ohjelmistoja

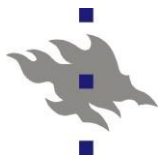
- Demidenko E. (2004). *Mixed Models. Theory and Applications.* Wiley.
- Diggle P. J., Liang, K.-Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press.
- Goldstein H. (2003). *Multilevel Statistical Models.* 3rd edition. Wiley
- Ohjelmasovelluksia
  - SAS: GENMOD, MIXED, GLIMMIX
  - SPSS:n ja Statan vastaavat analyysiohjelmat
  - Mplus ja Lisrel
  - R-ohjelmapaketteja...
- HUOM. Ositukseen reagointi ei luontevaa
- Analyysipainojen käyttö on mahdollista (ainakin SAS)



# Kaksiulotteisten frekvenssitaulujen testit

- Riippumattomuushypoteesin testi
- Homogeenisuushypoteesin testi
  
- Perusteita: [PDF-materiaali](#) Jaetaan luennolla
  
- [Esimerkki](#) Riippumattomuushypoteesi
  - Lehtonen & Pahkinen (2004), Example 7.3
  
- [VLISS](#) Training Key 250 **Test of Independence**
  
- SAS-proseduureja
  - Frekvenssitaulut, testit
  - [SURVEYFREQ](#) (asetelmaperusteinen)
  - FREQ (SRS-oletus, havainnot riippumattomia)





## Esimerkki: PROC SURVEYFREQ

(1) Metodisesti pätevä analyysi  
Asetelmaerusteinen, ryväsotanta

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

(2) Ei-pätevä versio (SRS-oletus)

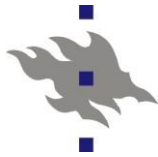
```
proc surveyfreq data=ohc;  
    tables phys*psych3 / chisq;
```



(1) Asetelmaperusteinen analyysi

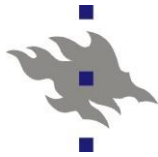
Table of PHYS by PSYCH3

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	



**(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**