Statistical Methods, part 1
Module 3: Observational studies, start

Dan Hedlin

Stockholm University

November 2012

Stockholm University

Selecting model in LCA

- Situation: number of latent variables and levels (classes) of those are given
- Biemer (2011, p 168 and following pages)
- Marijuana, three indicators, latent variable is real use, and grouping variables are age, race and sex (very American)
- Criteria for best model:

1. ‘Overall fit’. Likelihood ratio statistic compared to a chi-square distribution, greater p-value than 0.01. Motivation: large sample, with 0.05 perhaps no model would fit.
2. Smallest BIC. Motivation: BIC suits complex models and large samples

- Instead of 1, dissimilarity index < 0.05 would have been conceivable, in particular if chi-square asymptotics did not work well.

Difference in LR-statistics

- It is common to study the difference in likelihood ratio statistics for nested models. (they must be nested for this to work).
- The difference for simpler model – more complex model, is also asymptotically chi-square distributed under H_0, which says that simpler model is true.
- More complex model preferred if it adds something of value, ie difference must be statistically significant

- Base model: {XGRS }{XA}{XB}{XC}
- This is a local independence model with three grouping variables G, R, S
- Alternative models:

1. {XGRS }{XA GA RA SA}{XB GB RB RS}{XC GC RC SC}, first order interaction added to all A, B and C
2. Local dependence:

   {XGRS }{XA}{XB}{XC}[AB}{BC}

- ... And many other models
- All fitted to three periods of the NHSDA, 1994, 1995 and 1996.
- It was desirable to have the same model for all three years, mostly for practical reasons
- Model 2 fitted best in terms of criteria 1 and 2. Base model did not fit. Model 3 did not have smallest BIC.
- Surprising that model 3 did not make an improvement.

```
* LEM Input Statements  for Past-Year Marijuana Use Example
*
man 6
lat 1
dim 2 3 4 2 2 2 2
lab X R G S A B C
mod XGRS    {XGRS}
    A|XGRS {XA GA RA SA}
    B|XGRS {XB GB RB SB}
    C|XGRS {XC GC RC SC}
rec 192
rco
sta A|X [.9 .1 .1 .9]
sta B|X [.9 .1 .1 .9]
sta C|X [.9 .1 .1 .9]
wma XA xa.wma     * Write marginal XA to compute A error probabilities
wma XB xb.wma     * Write marginal XB to compute B error probabilities
wma XC xc.wma     * Write marginal XC to compute C error probabilities
dat mrj94wtd.txt  * File containing 1994 weighted data
* dat mrj95wtd.txt  File containing 1995 weighted data
* dat mrj96wtd.txt  File containing 1996 weighted data
```

Figure 4.9 LEM input statements for fitting model 1 to the data in Table 4.6.

Note:

- Sample size > 5000 or so. Will give small p-values for any model. If model appears to fit data, do not reject it even if $p<0.05$. Recall that 'all models are wrong but some are useful' (George Box)
- Very large p-value ($p>0.70$ or so) may suggest that the model is over-fitted.

- If Pearson chi-square statistic and LR statistic differ a lot, the LR statistic is generally preferred (Everitt 1977).
- There are rules of thumb for tables with small counts (e.g. expected values in each cell >4 or even only larger than 1 (Everitt 1977, p. 40)
- However, no rules of thumb for large counts

How important are p-values?

- Some researchers are rather uninterested in tests and p-values.
- For example, Goldstein (1995, p 32-33):

  A parameter value = 0 is implausible and uninteresting. Also, with large samples most null hypotheses will be rejected.
- These researchers are more focused on model fit (by appearance), usefulness and confidence intervals. Whether this strategy is more or less subjective is debatable.

Four types of observation

1. Experiment. Treatments are manipulated and the response is observed for each treatment.
2. Controlled observation. Like an experiment but you cannot or do not wish to influence the events. Pre-determined rules for what to note and how. E.g. observations on what pupils in a classroom say.

3. Uncontrolled observation. Like above but not planned ahead of time but still well documented.
4. Rumours, hearsay, undocumented observations, documentation lost, etc

- Note of course that a sociologist or social anthropologist may record rumours rigorously if the rumours are the object of research

Are experiments superior?

- Problem with observational studies: risk for confounding. You never know for sure.
- One can view research in the dimensions

1. accuracy
2. if the results can be generalised
3. If the results are broad or narrow

Surveys and experiments usually lack in 3, they tend to be narrow. A qualitative study in 1. All may have a problem with 2.

Experiments on cancer

- Experiments on breast cancer screening (mammography)
- Health Insurance Plan experiment. Randomised with control group (Freedman 2009, p. 4):

HIP, early 1960s

	Group size	Died in breast cancer, number	Rate (per 1000 women)	All other than breast cancer, rate
Screened	20200	23	1.1	21
Did not participate	10800	16	1.3	38
Total	31000	39	1.3	27
Control	31000	63	2.0	28

Conclusions?