

Latent Class Analysis: Model Selection

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

www.jeroenvermunt.nl

Introduction

- I will discuss model selection, which involves deciding on the number of latent classes
- I will demonstrate how to use the different types of statistics
- I will NOT discuss computation/formulae of these statistics. This is done in the next video
- **WARNING:** You will see that there is no single answer about the number of classes (about which model to select)

Four types of statistics

- Information criteria (BIC, AIC, AIC3)
- Goodness-of-fit tests (L-squared, X-squared)
 - Including bootstrap p values
- Bivariate residuals (BVRs)
- Likelihood-ratio (-2LLdiff) tests
 - Including bootstrap p values

Information criteria

- Weight model fit (log-likelihood value: LL) and model complexity (number of parameters: Npar)
- $-2LL + w * Npar$
- The model with the lower BIC, AIC or AIC3 is the preferred one
- The different variants differ in the value of “w”

Goodness-of-fit tests

- These tests compare the following H0 and H1
 - H0: the model with C classes
 - H1: the “saturated” model
- Chi-squared statistic comparing estimated with observed frequencies
- H0 is accepted when $p > .05$
- In case of sparseness: p-value computed via parametric bootstrap

Bivariate Residuals (BVRs)

- Goodness-of-fit in two-way tables
- Indicate whether local independence assumption holds
- Indicate whether there are residual dependencies between certain pairs of variables
- Rule of thumb: values should be smaller than 3 or 4
- Also: compare them with the BVRs of the 1-class model

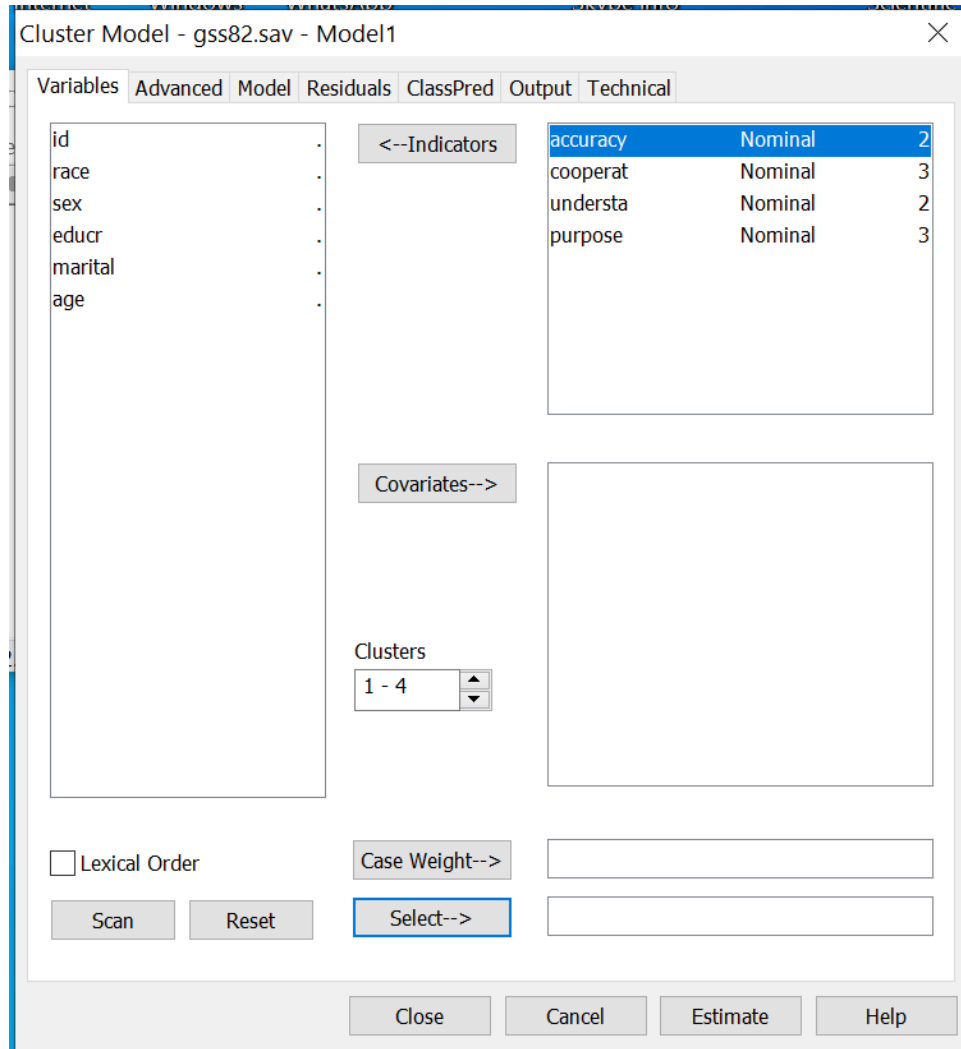
Likelihood-ratio (LR) tests

- These tests compare the following H_0 and H_1
 - H_0 : the model with C classes
 - H_1 : the model with $C+1$ classes
- Statistic is the difference in $-2LL$ between these two models
- But: we cannot use asymptotic/standard p-values
- Vuong-Lo-Mendell-Rubin p-value for robust LR test
- Better: bootstrap p value

Example data set: gss82.sav

- General Social Survey 1982, full sample (see Magidson & Vermunt, 2004)
- Evaluation of surveys by respondent (2 questions): purpose & accuracy
- Evaluation of respondent by interviewer (2 questions): understanding & cooperation
- I will treated these 4 indicators as nominal
- Research question: can we distinguish different types of survey respondents?

Latent GOLD set up with Cluster



Latent GOLD summary table

	LL	BIC(LL)	AIC(LL)	AIC3(LL)	Npar	L ²	df	p-value	Max. BVR	VLMR	p-value
1-Cluster	-4054.8112	8154.0517	8121.6224	8127.6224	6	357.4695	29	2.8e-58	84.2509		
2-Cluster	-3940.6687	7977.6009	7907.3374	7920.3374	13	129.1845	22	3.6e-17	58.4408	228.2850	0.0000
3-Cluster	-3891.4719	7931.0416	7822.9439	7842.9439	20	30.7910	15	0.0095	2.9266	98.3935	0.0000
4-Cluster	-3880.2290	7960.3899	7814.4579	7841.4579	27	8.3050	8	0.40	0.3606	22.4859	0.0206

BIC, BVRs: 3 class model

AIC, AIC3, L2, VLMR: 4 class model

Next video: more details on these statistics

Profile output of 3 class model

	Cluster1	Cluster2	Cluster3	Overall
Cluster Size	0.5677	0.2612	0.1712	
Indicators				
accuracy				
mostly true	0.5959	0.6453	0.0135	0.5091
not true	0.4041	0.3547	0.9865	0.4909
cooperat				
interested	0.9595	0.6413	0.6439	0.8224
cooperative	0.0403	0.2978	0.2507	0.1436
Impatient,Hostile	0.0002	0.0609	0.1054	0.0341
understa				
Good	0.9897	0.3788	0.7383	0.7871
Fair/Poor	0.0103	0.6212	0.2617	0.2129
purpose				
GOOD PURPOSE	0.8863	0.9013	0.1488	0.7640
DEPENDS	0.0563	0.0643	0.2163	0.0858
WASTE OF TIME AND \$	0.0574	0.0345	0.6349	0.1502