

## CHAPTER 8. ADDITIONAL TUTORIALS AND ASSOCIATED DATA SETS

Several additional tutorials are under development and when completed will be accessible from our website. These tutorials include:

### Tutorial #5: Using Latent GOLD 4.0 with the Known Class Option

**DEMADATA = 'DEPRESS2.SAV'**

In this tutorial, we illustrate the use of the 'known class' feature in Latent GOLD 4.0 to take into account additional information on a subset of cases which allows us to classify them into a particular class with probability one. In this case, the information comes from a physician's diagnosis of the patient as 'Depressed' or merely 'Troubled', corresponding to 2 of the 3 latent classes.

### Tutorial #6: Estimating a Random Intercept Regression Model

**DEMADATA = 'CRACKERS.SAV'**

(SOURCE: KELLOGG COMPANY STUDY)

In this tutorial, we illustrate the use of continuous factors (CFactors) to control for the 'level effect' in ratings data. A latent class regression model is estimated where the dependent variable is ratings of 15 crackers on taste, and 12 predictors correspond to different attributes of the crackers. Different classes are identified that show different taste preferences, controlling for their overall rating level. These data are based on a paper by Popper et. al. The use of CFactors requires the Advanced version of Latent GOLD 4.0.

## DATA SETS AND EXAMPLE LGF FILES

Below are descriptions of 80 data sets that have been analyzed using Latent GOLD 4.0. Some of these are the subjects of tutorials, whereas others are used in our courses and workshops. These data sets are included in the demo version of Latent GOLD 4.0. For some data sets we also prepared lgf files illustrating the most important Latent GOLD 4.0 Basic and Advanced features. The data and lgf files are also available separately on our website at [http://www.statisticalinnovations.com/products/latentgold\\_datasets.html](http://www.statisticalinnovations.com/products/latentgold_datasets.html).

### 8.1. Dichotomous, Nominal, or Ordinal Indicators: Cluster and DFactor Models, as well as Models with Continuous Factors (IRT Models)

#### Dichotomous indicators

##### 1. hannover.sav:

- 5 dichotomous indicators
- survey data on pain related to rheumatic arthritis
- cluster or DFactor model
- used in Kohlman and Formann (1997), Magidson and Vermunt (2001), and in the Latent GOLD 2.0 user's manual (Vermunt and Magidson, 2000a)

##### 2. political.sav

- 5 dichotomous indicators on political involvement and tolerance
- 3 (nominal) covariates
- 3-cluster model, 2-DFactor model, or 2-cluster model with a local dependency
- data from Political Action Survey
- used in Hageaars (1993) and Vermunt and Magidson (2000a, 2000b)

##### 3. landis77.sav

- dichotomous rating (presence/absence of carcinoma in the uterine cervix) of 118 slides by 7 pathologists
- see also landisreg.sav for other data structure
- 3-cluster, 2-DFactor model, CFactor model (2PLM), CFactor model with equal effects (Rasch), or combination of 2-cluster and Rasch model
- sparse table: use bootstrap p value
- used as illustration in Agresti (2002), Magidson and Vermunt (2003a, 2004), and Vermunt and Magidson (2005a). Original data in Landis and Koch (1977).

## CHAPTER 8. ADDITIONAL TUTORIALS AND ASSOCIATED DATA SETS

### 4. heinen2.sav

- 5 dichotomous indicators of gender roles (male sample)
- same data set in other format in heinen2reg.sav
- 3-cluster, 3-level 1-DFactor model, and various types of IRT models
- used as illustration in Heinen (1996)

### 5. heinen\_mf.sav

- same data as heinen2.sav but now for males and females
- gender can be used as a covariate, possibly affecting indicators (item bias)
- data can also be used for unrestricted multiple group analysis (with female2 as known-class indicator)
- see also SMABS 2004 workshop transparencies

### 6. vdheijden.sav

- 3 dichotomous indicators of youth delinquency
- ethnic group and age group are covariates
- used by Van der Heijden et al. (1992) to illustrate logit-restricted latent budget analysis, which is a LC cluster model with covariates

### 7. depression.sav

- 5 depression indicators and covariate sex
- 3-cluster, 3-level 1-DFactor, or 2-cluster model with a CFactor model
- used in Magidson and Vermunt (2001) and Schaeffer (1988)

### 8. knowclass.sav

- simulated data set based on the 3-cluster solution obtained with the depression.sav data set
- information on known class membership generated using 3 mechanism: MCAR, MAR (depending on the sum of all item responses), NMAR (depending class membership itself)
- in the NMAR case known-class yes/no should be used as covariate

### 9. lcamis.dat

- 5 dichotomous indicators
- example of LC model with missing data on indicators
- simulated data set

### 10. lifestyle.sav

- data on a large set of lifestyle activities (dichotomous indicators) and a few covariates (source: The Polk Co.)
- demo data set in Latent GOLD 2.0 and used in Magidson and Vermunt (2003b)

## 11. store.sav

- 5 dichotomous items related to consumer behavior
- standard LC cluster model
- used in Dillon and Kumar (1994)

## 12. coleman.sav

- classical data set of Coleman
- 2 indicators, membership of and attitudes toward leading crowd, measured at two occasions
- 2-DFactor model (unrestricted or restricted)
- analysed by Goodman (1974) and Agresti (2002, table 12.8)

## 13. gss94.sav

- data from the 1994 General Social Survey
- 3 attitudes toward abortion indicators, and covariate gender
- 2-cluster model, LC Rasch (two-cluster with equal effects), parametric Rasch (CFactor with equal effects)
- data taken from Agresti (2002, table 10.13)

## 14. financial.sav

- data on ownership of 4 financial products
- taken from Paas (2002)

## 15. hadgu.sav

- 6 measures (tests) for diagnosing chlamydia trachomatis (most common sexually transmitted disease), where one test (culture) is a gold standard and can therefore be used a Known-Class indicator
- 2-class model with local dependencies modeled with a CFactor (with equal effects across tests)
- used by Hadgu and Qu (1998) with the purpose is to determine the sensitivity and specificity of the various tests

## Polytomous indicators

## 16. judges.dat

- trichotomous ratings of three judges, that can be treated as ordinal
- 3-cluster, 3-level 1-DFactor model, and CFactor/IRT (partial credit) model, possibly with equal effects across indicators
- used in Dillon and Kumar (1994), and in the Latent GOLD version 2.0 user's manual (Vermunt and Magidson, 2000a)

## CHAPTER 8. ADDITIONAL TUTORIALS AND ASSOCIATED DATA SETS

### 17. gss82white.sav

- 2 dichotomous and 2 trichotomous indicators that can be treated as nominal or ordinal
- data from General Social Survey '82, white sample
- the purpose of the analysis is to construct a typology of survey respondents
- 3-cluster or 2-factor model
- used in McCutcheon (1987), Magidson and Vermunt (2001, 2004), and Vermunt and Magidson (2005a)

### 18. gss82.sav

- same indicators as in gss82white.sav, but for full sample (whites and non-whites)
- several covariates that can be treated as active or inactive
- used in Magidson and Vermunt (2004)

### 19. elliot.sav

- marijuana use of children (13 years of age in 1976) in 5 consecutive years (trichotomous ordinal response variable)
- see also elliotreg.sav for other data structure
- standard cluster model with time-specific indicators and sex as covariate
- use bootstrap p value because of sparseness
- references to data set: Elliot et al. (1989), Vermunt and Hagenaaars (2004), and Vermunt, Rodriguez and Ato (2001)

### 20. heinen3.sav

- 5 trichotomous indicators of gender roles that can be treated as nominal or ordinal
- cluster, order-restricted cluster, DFactor, and various types of IRT models.
- bootstrap p value
- used as illustration in Heinen (1996)

### 21. environment.dat

- 6 trichotomous items measuring attitudes towards environmental issues
- there are two underlying dimensions: willingness (item 1-3) and awareness (items 4-6)
- data used by Croon (2002)

### 22. internet99.sav

- data on internet use (source: Mediamark Research Inc. 1999)
- relationship between internet usage and several demographic covariates
- used in Magidson and Vermunt (2003b)