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

DEMODATA = '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

DEMODATA = '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 an 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.

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 Hagenaars (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).

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 illustrated 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 diagnozing 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)

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 Hagenaars (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)

23. USelection2000.sav

- National Election Studies election survey data set 2000.T
- relationship between vote and ratings of Bush and Gore
- Source: Burns et. al. (2001), ICPSR Study Number: 3131

8.2. Single Response Variable: Mixture of Univariate Distributions (using Cluster or Regression Module)

24. galaxy.dat

- velocities of 82 galaxies diverging away from our own galaxy
- mixture of univariate normals
- set bayes constants off and increase number of start sets to reproduce results reported by McLachlan and Peel (2000)

25. enzyme.dat

- enzymatic activity in the blood among a group of 245 individuals
- mixture of univariate normals
- used as illustration in McLachlan and Peel (2000)

26. acidity.dat

- acidity index measured in a sample of 155 lakes in north-central Wisconsin
- mixture of univariate normals
- used as illustration in McLachlan and Peel (2000)

27. candy.dat

- single count variable: number of packages of hard candy purchased in a week
- example of simple mixture model
- can be specified with regression or cluster with number of packages as count
- used in Dillon and Kumar (1994), Magidson and Vermunt (2004), and in Latent GOLD 2.0 user's manual Vermunt and Magidson (2000a).

28. candy_trunc.txt

- single truncated count variable: number of packages of hard candy purchased in a week among consumers
- example of simple mixture model for truncated counts
- can be specified with regression or cluster

- used in Dillon and Kumar (1994)

29. nov2002.sav

- results of statistics exam November 2002
- a 2 class binomial (exposure equal to 20) or normal mixture separates perfectly the students who pass and the ones that do not pass the exam.

30. sids.dat

- data of 100 counties in north Carolina concerning children suffering from sudden infant death syndrome: number of deaths and population at risk
- mixture of Poisson rates
- example used by Böhning (2000) to illustrate disease mapping.

8.3 Continuous, Count and Mixed-Scale Indicators: Cluster Models and Models with Continuous Latent Variables (Factor Analysis and Generalized IRT)

31. iris.dat

- 4 continuous indicators: measures taken on 150 irises
- LC cluster model or mixture model clustering
- true specie is known and can be compared with cluster solution (use true as inactive covariate)
- illustrates different specifications of within cluster variance-covariance matrix
- classical data set from Fisher

32. kmeans.sav

- simulated data set to illustrate LC clustering with continuous variables and compare it with K-means clustering
- different specification of the error variances
- used in Magidson and Vermunt (2002a, 2002b)

33. diabetes.sav

- 3 continuous indicators
- example of LC clustering
- clinical classification can be compared with LC cluster classification
- used in Fraley and Raftery (1998), Vermunt and Magidson (2002), and Magidson and Vermunt (2004)

34. cancer.dat

- clustering based on pre-trial "covariates" collected before for a prostate cancer clinical
- eight are treated as continuous and four as categorical indicators
- example of LC clustering with mixed mode data
- used as illustration in Hunt and Jorgensen (1999), McLachlan and Peel (2000), and Vermunt and Magidson (2002)

35. srcddata.txt

- continuous outcome variable read# (child's reading recognition) measured at 4 occasions (many missing data) on 405 children
- covariates: child's gender (male=1), mother's age in years at Time 1, child's age in years at Time 1, child's cognitive stimulation at home, and child's emotional support at home
- longitudinal data for specifying growth model: cluster model with one or two CFactors
- data used in Vermunt and Magidson (2005c) and made available at http:/ww.duke.edu/curran/
- file also contains an ordinal outcome variable anti# (child's antisocial behavior) measured at four time points, which was however not used in Vermunt and Magidson (2005c)

36. abortion_cluster.sav

- same data as abortion.sav example (see below), but here in standard rectangular data format instead of repeated measures format
- indicators are binomial counts: number of agrees out of 7 abortion situations measured at 4 occasions
- source: McGrath and Waterton (1986)

8.4. Latent Class and Random-effects Regression Modeling

Mixture Regression Models for Single Response

37. follman.dat

- effect of poison on survival
- dichotomous dependent variable survival can be treated as nominal, ordinal, or binomial count, since all are equivalent for dichotomous variables
- using logdose as a numeric class-independent predictor yields a non-parametric random effects logistic regression model
- used in by Follmann and Lambert (1989), Formann (1992), and Agresti (2002) to illustrate non-parametric random-effects logistic regression

38. fabric.dat

- number of faults in a bolt of fabric of a certain length
- random-effects Poisson regression with log length as predictor
- data used by Aitkin (1996) and McLachlan and Peel (2000)

39. beta.dat

- meta analysis of 22 clinical trials of beta-blockers for reducing mortality after myocardial infarction
- dependent is a binomial count
- observations within a clinic are dependent
- LC regression model with random intercept (3 classes) and fixed treatment effect
- data used by Aitkin (1999) and McLachlan and Peel (2000)

40. dmft.sav

- dental health trial on prevention of tooth decay among 797 Brazilian children
- dependent variable: # of decayed, missing or filled teeth (DMFT)
- explanatory variables are: Treatment (1 = no treatment; 2 = oral health education; 3 = school diet enriched with rice bran; 4 = mouth rinse with 0.2% NaF solution; 5= oral hygiene; 6 = all four treatments), Ethnic group (1= brown; 2 = white; 3 = black) and Gender (1 = male; 2 = female)
- Poisson or binomial count regression with overdispersion, using a LC regression, a zero-inflated regression, or random-intercept regression model
- data analyzed by Skrondal and Rabe-Hesketh (2004, section 11.2)
- see also SMABS 2004 workshop transparencies

41. cace.sav

- to illustrate "complier average causal effect" model using Known-Class option
- compliance is known for the treatment group but unknown (latent) for the control group
- LC regression in which treatment has an effect in the compliance class, and in which compliance (yes/no) is predicted using covariates
- data analyzed by Skrondal and Rabe-Hesketh (2004) and can be obtained from the ICPSR website (under JOB #2739)

42. long1.sav

- continuous dependent: firstjobcens0 or firstjobtrunc0 is the prestige of the first academic job (minus 1 to get the censoring/truncation at 0 instead of 1)
- various predictors
- censored normal regression, censored-inflated normal regression, or truncated normal regression
- data used by Long (1997, chapter 8)

43. long2.sav

- count dependent: number of articles in last 3 years of PhD
- various predictors (two copies of each in file)

- Poisson regression, zero-inflated Poisson regression, random-intercept Poisson regression, and zero-inflated random-intercept Poisson regression
- data used by Long (1997, chapter 9)

44. runshoes.dat

- count dependent: number of running shoes for a sample of runners
- predictors: runs per week, miles run per week, distance runner
- truncated Poisson count regression model
- used in textbook "Analyzing Categorical Data" by Jeffrey S. Simonoff

Two-level and Multiple Response Data Sets

45. bang.txt

- contraceptive use (dichotomous outcome)
- data from 1989 Bangladesh Fertility Survey (Huq and Cleland 1990)
- women nested within districts
- predictors: number of children, age in years (centered), and urban (0=rural)
- data obtained from multilevel modeling website

46. snijdersbosker.sav

- performance of pupils on a language test (continuous outcome)
- data taken from the Snijders and Bosker (1999) book on multilevel analysis
- children nested within school
- pupil-level predictors: IQ, SES (both overall centered)
- school-level predictors: school_IQ, school_SES, groupsize (centered), combination classes (yes/no)
- see also SMABS 2004 workshop transparencies

47. conjoint.sav

- rating-based conjoint example
- simulated data
- full factorial design (2*2*2) with 8 replications
- LC regression with ordinal dependent, 3 predictors (product attributes) and 2 covariates (individual characteristics)
- used in Magidson and Vermunt (2003b) and in Latent GOLD 2.0 user's manual (Vermunt and Magidson 2000a)

48. crackers.sav

- data from a consumer taste study sponsored by the Kellogg company, where consumers rated their liking of 15 crackers on a nine-point liking scale.
- an independent trained sensory panel evaluated the same crackers in terms of their sensory attributes (e.g. saltiness, crispness, thickness, etc.), yielding ratings on 12 flavor, texture, and appearance dimensions
- LC regression analysis with a random intercept

49. USselection2000reg.sav

- National Election Studies election survey data set 2000.T
- same data as USselection2000.sav, but now in regression format
- Source: Burns et. al. (2001), ICPSR Study Number: 3131

Various Cluster and IRT Data Set in the Form of Multiple Records per Case

50. landisreg.sav

- dichotomous rating (presence/absence of carcinoma in the uterine cervix) of 118 slides by 7 pathologists
- dependent variable "rating" can be treated as nominal, ordinal, or binomial count since all are equivalent for dichotomous variables.
- LC regression model with rater as nominal predictor. Specifying the rater effect as class independent yields a LC Rasch model. Class dependent yields a standard LC model.
- variable "sumscore" can be used as inactive covariate to see how the latent classification is related to the sum of the ratings.
- the file contains dummies for the raters to change the coding scheme.
- a copy of the predictor rater (rater_) is included to specify a two-dimensional model (LC factor model).
- sparse table: use bootstrap p value
- used as illustration in Agresti (2002) and Magidson and Vermunt (2003a, 2004) and Vermunt and Magidson (2005a). Original data in Landis and Koch (1977).

51. heinen2reg.sav (male sample)

- 5 dichotomous indicators of gender roles
- same data as heinen2.sav, but other data structure
- 3-class regression model: item effect class-independent yields a LC Rasch model; item effect class-dependent yields a standard LC model
- used as illustration in Heinen (1996)

52. heinenreg_mf.sav

- same as heinenreg2.sav but now for males and females (also the same as heinen2_mf.sav but in other format)
- gender can be used as covariate, predictors, or in gender-item interaction (item bias)
- standard LC, restricted LC, LC Rasch, and IRT models

- see also SMABS 2004 workshop transparencies

53. colemanreg.sav

- same data as coleman.sav but in a different format
- item characteristics are included as predictors to test several assumption
- predictors: item, member, attitude, time1, time2, member1, member2, attitide1, and attitude2
- best model is a 2-factor like structure with a member and a attitude factor
- analysed by Goodman (1974) and Agresti (2002, table 12.8)

54. gss94reg,sav

- same data as gss94.sav but in a different format

55. financialreg.sav

- same data as financial.sav: ownership of 4 financial products
- taken from Paas (2002)

Longitudinal Data for LC Growth Modeling

56. abortion.sav

- data from the British Social Survey (McGrath and Waterton, 1986)
- the dependent "number of times that one agrees with abortion out of 7 situations" should be treated as binomial count
- year is a class-dependent (random, level-1) predictor and religion a class-independent (fixed, level-2) predictor
- the data file contains dummies for the time and religion categories to use dummy instead of default effects coding
- the data file also contains an incremental coding of the time categories and time squared to play with the time effect
- used by Vermunt and Van Dijk (2001) to illustrate the connection between LC regression and random-coefficients, mixed, hierarchical, or multilevel models, as well as in Magidson and Vermunt (2004).

57. elliotreg.sav

- marijuana use of children (13 years of age in 1976) in 5 consecutive years (trichotomous ordinal response variable)
- LC growth model with time as nominal/ascending/class-dependent predictor and sex as covariate (see Vermunt and Hagenaars, 2004)

- possible to include random intercept (CFactor)
- references to data set: Elliot et al. (1989) and Vermunt et al. (2001)
- see also SMABS 2004 workshop transparencies

58. rats.dat

- growth of rats in first weeks
- LC growth model for continuous outcome variable
- reference to data set: Gelfand et al. (1990)

66A. growth1.sav

- LC Growth Model based on 59 epileptics who were randomly assigned to either an anti-seizure medication or placebo.

Event History and Transition Data

59. jobchange.dat

- LC regression model for event history data (piece-wise exponential survival model)
- data from 1975 Social Stratification and Mobility Survey Japan (see, Yamaguchi, 1991)
- the event of interest is first inter-firm job change
- event should be treated as Poisson count with an exposure variable
- time, categorized in 3 intervals, is a class-independent nominal predictor
- single covariate firm size (either nominal or linear with extra dummy for government)
- used as illustration in Vermunt (2002a)

60. empltran.dat

- discrete-time event history or survival model with multiple outcomes
- two predictors/covariates: cohort and sex
- exposure time should be used as replication weight
- used in Blossfeld and Rohwer (1995)

61. dropout.dat

- school drop-out of brothers at two school levels
- modelled as discrete-time event history model with unobserved heterogeneity to capture dependence between respondent and brother (family effect)
- brother and time (school level) are predictors; father's education can serve as predictor or as covariate
- used as illustration in Mare (1994) and Vermunt (1997)

62. land.sav

- duration time to first serious delinquency
- 411 males from working-class area of London followed from ages 10 through 31
- dependent "first" can be treated as Poisson count or as binomial count. If treated as Poisson count, the exposure can be set to one or one half for the time point at which the event occurs.
- variable "tot" is a risk index that can be used either as predictor or as covariate
- the duration effect (age effect) can be modelled by a quadratic function
- data used as illustration by Land et al. (2001).

63. poulsen.sav

- transitions in brand preference (brand A or other brand) between 5 occasions
- example of mixture transition or mixed Markov model
- predictors are time0 (whether record corresponds to the initial state), ylag_a (previous time point equals brand A), and ylag_oth (previous time point equals brand A). Either the intercept or time0 should be omitted from the model
- data used as illustration by Poulsen (1982)

64. vinken.sav

- timing of four events related to first experience with relationships
- used in Vermunt (2002a)
- Cox model for correlated events
- see also SMABS 2004 workshop transparencies

Longitudinal Data from Repeated Measures/Clinical Trials

65. koch.sav

- repeated measures clinical trial with outcome normal (1) or not normal (0)
- time is a class-dependent predictor, severity a class-independent predictor and treatment is a covariate; this yields a LC growth model in which treatment has an effect on the type of growth curve that one follows.
- an alternative is to use time, severity, treatment, and the treatment-time interaction as class-independent predictors, yielding a standard non-parametric random-effects model.
- used in Agresti (2002) to illustrate random-effects logistic regression. Original data are in Koch et al. (1977).

66. epilep.sav

- randomized controlled trial comparing a new drug with placebo
- outcome variable y is the number of epileptic seizures during the two weeks before each of 4 clinic visits (Poisson count)
- 4 replications per case (4 visits)
- class-independent numeric predictors: treatment, log baseline, log age, visit number, dummy for fourth visit, and treatment log base interaction
- data from Thall and Vail (1990), also used by Rabe-Hesketh et al. (2002)

67. aspartame.dat

- multiple period (5 weeks) crossover trial to test the side effect of aspartame
- the dependent variable is a binomial count; that is, the number of days with a headache out of a total of 7 days (a week).
- the total number of days exposed in a period may be smaller that 7 and the total number of periods may be less than 5 because of drop out.
- predictors are week and aspartame (1= aspartame; 0=placebo)
- covariate: believe as to whether drug will cause and headache
- data used by McKnight and Van Den Eeden (1993) to illustrate models for correlated binomial counts. Can also be modeled as Poisson counts (see Hedeker, 1998).

68. genomics.sav

- multi-visit follow-up of 7 rheumatoid arthritis patients diagnosed as unstable during first visit and assigned to new drug therapy
- blood sample taken during each visit to obtain genetic expressions
- drug effects assessed using IndexZ to see if levels approach those of normals
- source of IndexZ (Source Precision Medicine, Inc.) - patents pending

69. schizophrenia.sav

- effect of drug on severity of schizophrenia
- dichotomious or ordinal dependent variable "severity" measured at 7 occasions (with many missing values)
- can be used for random-effect logistic regression, LC logistic regression, and LC logistics regression with a random intercept
- data used by Hedeker and Gibbon's (1996)

82. allocation.sav

- example of allocation model
- survey question that asks respondents to allocate 100 points over 8 alternatives to indicate the relative importance of each alternative

8.5 Multilevel Latent Class Models and Complex Surveys

Two-level Cluster and DFactor Models

70. mierlo_socmeth.txt

- 5 dichotomous items measuring task variety
- missing values on items (some of which were caused by a mistake made in the recoding of the items)
- employees nested with teams
- simplest variant of multilevel LC model with either GClasses or GCFactors affecting the clusters
- data set taken from dissertation from Van Mierlo (2003), and used by Vermunt (2003)

71. mierlo_mbr.sav

- same data set as mierlo_socmeth.dat, but without the mistake in the recoding (results are therefore slightly different)
- in addition, 4 individual-level covariates: year of birth (4 levels), number of years in the current job (3 levels), number of working hours per week (3 levels), and gender. The 57 cases with missing values on items and/or covariates can be retained in the analysis (using the include missing all option).
- random-intercept model for the clusters using a GCFactor
- used in Vermunt (2005)

72. cito.dat

- data on mathematical skills on pupils: 18 mathematics test items (correct/incorrect) administered to 2157 pupils
- pupils are nested within 97 schools
- three individual-level covariates (SES, IQ and Gender) and one school-level covariate (CITO)
- multilevel variant of a DFactor model
- used by Fox and Glas (2001) and by Vermunt (2003)

73. meulders.sav

- three-mode three-way data from a psychological "experiment" 101 1st year psychology students to indicate whether when angry at someone they would display 8 behaviors (fly off the handle, quarrel, leave, avoid, pour up ones heart, tell one's story, make up, clear up the matter) in 6 situations (like the other, dislike the other, unfamiliar with the other, other has higher status, other has lower status, and other has equal status other.
- situations are nested within persons
- situations are non exchangeable, therefore use situation as covariate affecting class membership
- GClasses (of persons) affecting intercept of and situation effect on clusters (of persons in situations)
- used in Meulders et al. (2002, Journal of Classification) paper on LC models for three-mode data

Three-level Regression Models

74. immunization.sav

- complete immunization of children in Guatemala (binary response variable)
- individuals (children) nested within families, and families nested within communities
- three-level binary logistic regression using parametric or nonparametric random effects
- 4 individual, 5 family and 2 community level predictors (some are dummies)
- used by Rodriguez and Goldman (2001).

75. tvsfp.sav

- ordinal outcome variable: the tobacco and health knowledge scale (THKS) score defined as the number of correct answers to seven items on tobacco and health knowledge (collapsed into our ordinal categories).
- schools were randomized into one of four conditions combining the factors TV (a television intervention, 1=present, 0=absent) and CC (a social-resistance classroom curriculum, 1=present, 0=absent)
- classes are nested within schools and pupils are nested within classes
- data are from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) and used by Hedeker and Gibbons (1996)

76. socatt.txt

- same data as abortion.sav file, but now with district number and some extra covariates
- repeated measures nested within cases and cases nested within districts
- three-level binomial count regression using either parametric or nonparametric random effects
- source: McGrath and Waterton (1986) and multilevel modeling webpage
- used in Vermunt (2002c) and Vermunt (2004)

77. zugugl.sav

- three well-being items (zufrieden, gut, glücklich) measured at three occasions
- responses are both in 3-point scale and 5-point scale format
- 3-level regression or 2-level IRT model
- data used by Steyer and Partchev (2001) to illustrate their state-trait model for ordinal variables, which is a 2-level IRT model

78. tob3vote.sav

- response variable: voting pro-tobacco by members of the Congress from 1997-2000
- predictors/covariates: party, amount of money member received from tobacco industry (money), and the number of harvest acres in the member's state in 1999 (acres)
- votings/bills nested within members and members nested within states
- 3-level random-effects regression model and 2-level LC model
- used by Luke (2004) in his Sage textbook "Multilevel Modeling"

Complex Survey Options

79. patterson.sav

- standard LC model for 4 dichotomous response variables (vegetable consumption at 4 occasions)
- stratum, PSU, and weight variable
- totvgt1-totvgt4 variables were used by Patterson, Dayton, and Graubard's (2002) in their article on LC analysis of complex sampling survey data
- the variables v1-v6 were used by Vermunt (2002b), who made use of the fact that the 4 occasions were actually 6 different time points with missing values on at least two time points
- data set does also contain some covariates, as well as information on fruit consumption (was not used by the above authors)

80. pattersonreg.sav

- regression format data set based on patterson.sav
- contains the variables totvgt1-totvgt4 used Patterson, Dayton, and Graubard's (2002)
- LC growth model

81. pattersonreg2.sav

- regression format data set based on patterson.sav
- constains variables v1-v6 (vegetables) used in Vermunt (2002b), as well as f1-f6 (fruit)
- can be used to specify a LC growth model for vegetables or for fruit, or a multilevel LC model in which vegetables and fruit consumption are used as indicators of a time-specific latent variable