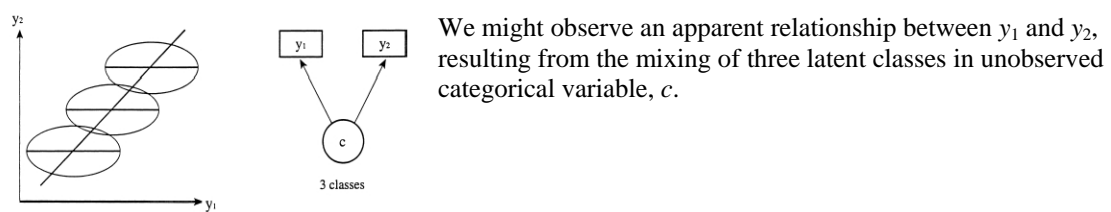
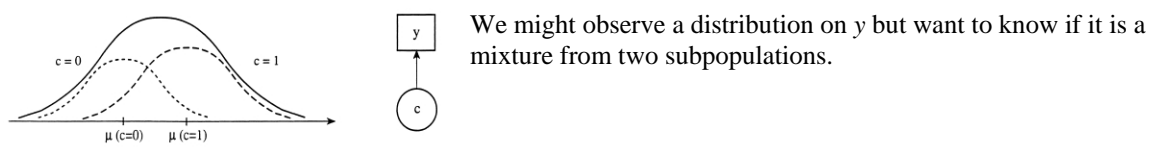


## QuaSSI Backgrounder: Latent Class and Finite Mixture Models

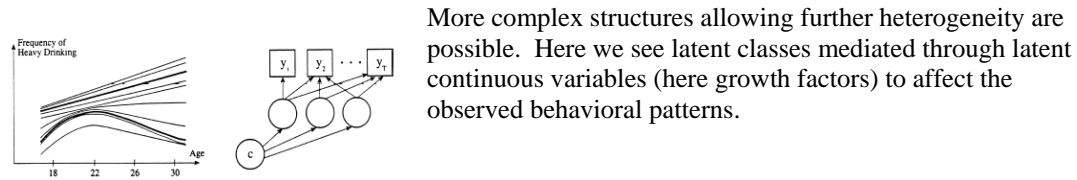
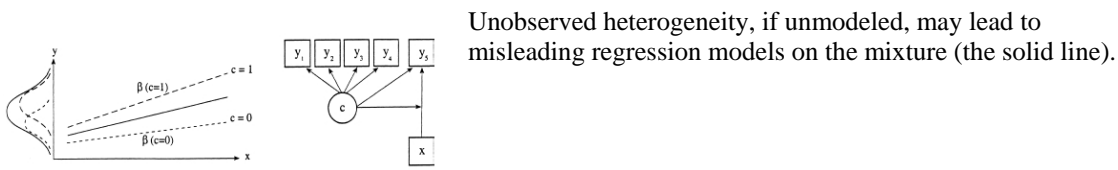
Prepared by Tatiana V. Vashchilko, QuaSSI Predoctoral Fellow

### What is Latent Class Analysis?

- **Other terms:** Latent class and similar models are also known as *finite mixture* or *unobserved heterogeneity models*. The term *latent class analysis* is reserved by most authors for mixture models in which observations are categorical, often summarized in a contingency table.
- **Assumption:** The sample under study is drawn from the population that consists of several, say  $K$ , unobserved groups. That is, every observation in the sample can be thought of as being drawn from one of those  $K$  subpopulations (or latent classes), and  $K$  could be known or unknown.
- **Basic Goals:** Identify or confirm latent classes from observed variables. This can have further application for evaluating classifiers (like diagnostic tests) or for data reduction or data-mining.



- **Extended Goals:** Model the unobserved heterogeneity to improve inference.



Figures are from Bengt O. Muthén. 2001. "Latent Variable Mixture Modeling" in *New Developments and Techniques in Structural Equation Modeling* ed. By Marcoulides, George A. and Randall E. Schumacker. Mahwah, N.J. Lawrence Erlbaum Associates, Inc. (Available at PSU through [www.netlibrary.com](http://www.netlibrary.com)).

## Classics and General Treatments

- Lazarsfeld PF, Henry NW. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Goodman, L. A. (1974). "Exploratory latent structure analysis using both identifiable and unidentifiable models." *Biometrika*, 61, 215-231.
- McCutcheon AC. (1987). *Latent class analysis*. Beverly Hills: Sage Publications.
- Aitken, M. and D.B. Rubin. (1985). "Estimation and hypothesis testing in finite mixture models", *Journal of the Royal Statistical Society*, B, 47, 67-75.
- Clogg, C. C. (1981). Latent class models for measuring. In R. Langeheine & J. Rost (Eds.) *Latent trait and latent class models*. New York: Plenum.
- Langeheine, R. & Rost, J. (Eds.) (1988). *Latent trait and latent class models*. New York: Plenum.
- Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Sage Publications.
- Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (Ch. 6; pp. 311-359). New York: Plenum.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, California: Sage.
- J. Rost & R. Langeheine, Eds. (1997) *Applications of latent trait and latent class models in the social sciences*. New York: Waxmann.

## Mixture Models and LCA in General Methods Texts

- A Agresti. 2002. *Categorical Data Analysis* (2<sup>nd</sup> ed). Chapter 13.
- A Gelman, JB Carlin, HS Stern, DB Rubin. 2004. *Bayesian Data Analysis* (2<sup>nd</sup> ed). Chapter 18.
- AC Cameron and PK Trivedi. 2005. *Microeconometrics: Methods and Applications* Chapter 18.

## Introductions and Related Resources Available Online

- The Methodology Center at the College of Health and Human Development at Penn State University. <http://methodology.psu.edu/lca/>
- The Center for Integrated Latent Variable Research at the University of Maryland. <http://www.cilvr.umd.edu/>
- Lecture notes on LCA from G. David Garson's (NC State) class on *Topics in Multivariate Analysis*, <http://www2.chass.ncsu.edu/garson/pa765/latclass.htm>
- *See also the section on Software below.*

## Recent extensions of LCA and mixture modeling approaches (\*@ Penn State):

- GR Hancock, KM Samuelsen, eds. 2007. *Advances in latent variable mixture models*. Greenwich, CT: Information Age Publishing, Inc.
- YW Teh, MI Jordan, MJ Beal, DM Blei. 2006. "Hierarchical Dirichlet processes." *Journal of the American Statistical Association*.
- KL Nylund, T Asparouhov, B Muthén. 2006. Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling*.
- J Guo, M Wall, Y Amemiya. 2006. Latent class regression on latent factors. *Biostatistics*.
- FG Garre, JK Vermunt. 2006. "Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation." *Behaviormetrika*.
- CC Yang. 2006. "Evaluating latent class analysis models in qualitative phenotype identification." *Computational Statistics and Data Analysis*.
- \*Chung, H, Y Park, ST Lanza. 2005. "Latent transition analysis with covariates: Pubertal timing and substance use behaviours in adolescent females." *Statistics in Medicine*.
- \*ST Lanza, LM Collins, JL Schafer, BP Flaherty. 2005. "Using Data Augmentation to Obtain Standard Errors and Conduct Hypothesis Tests in Latent Class and Latent Transition Analysis." *Psychological Methods*.
- \*LM Collins. 2005. "Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model." *Annual Reviews of Psychology*.
- NL Zhang. 2004. "Hierarchical Latent Class Models for Cluster Analysis." *Journal of Machine Learning Research*.
- \*ST Lanza, BP Flaherty, LM Collins 2003. "Latent class and latent transition analysis." In JA Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* Wiley.
- JK Vermunt. 2003. "Multilevel Latent Class Models." *Sociological Methodology*.
- AK Formann. 2003. "Latent Class Model Diagnosis from a Frequentist Point of View." *Biometrics*.
- A McCutcheon, J Hagenaars, eds. 1999. *Advances in Latent Class Modeling*. Cambridge.

## Software

- **PROC LCA** is a new SAS procedure for LCA and variants, developed by the Methodology Center (College of Health and Human Development) at Penn State. <http://methodology.psu.edu/lca/>
- **poLCA** is a new R package for LCA and variants, developed by political scientists Drew Linzer and Jeff Lewis at UCLA. <http://dlinzer.bol.ucla.edu/poLCA/> (there are other R packages for lca available)
- **gllamm** is a set of Stata libraries (developed by Sophia Rabe-Hesketh, UC-Berkeley) for multilevel and mixed models and includes a variety of LCA and similar models. <http://www.gllamm.org/>
- **Latent GOLD** is a commercial standalone package for LCA and similar approaches. <http://www.statisticalinnovations.com/>
- **MPlus** is a commercial standalone package (developed by Bengt Muthén, UCLA) for models with categorical and continuous latent variables. <http://www.statmodel.com/>
- There are many other specialized programs available. John Uebersax maintains a website on latent class analysis and summarizes many of these here: <http://ourworld.compuserve.com/homepages/jsuebersax/>

## Applications of LCA and Related Mixture Models (\*@ Penn State)

- **Anthropology / Genetics.** GA Satten, WD Flanders, Q Yang. 2001. "Accounting for Unmeasured Population Substructure in Case-Control Studies of Genetic Association Using a Novel Latent-Class Model." *American Journal of Human Genetics*.
- **Criminology / Sociology.** AV D'Unger, KC Land, PL McCall. 2002. "Sex Differences in Age Patters of Delinquent / Criminal Careers: Results from Poisson Latent Class Analyses of the Philadelphia Cohort Study." *Journal of Quantitative Criminology*.
- **Demography / Sociology.** MA Koenig, S Ahmed, MB Hossain, ABMKA Mozumder. 2003. "Women's Status and Domestic Violence in Rural Bangladesh: Individual and Community-Level Effects." *Demography*.
- **Economics.** A Clark, F Etilé, F Pstel-Vinat, C Senik. 2005. "Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries." *Economic Journal*.
- **Economics.** PC Boxall, WL Adamowicz. 2002. "Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach." *Environmental and Resource Economics*.
- \* **Education / Labor Studies.** D Passmore, A Mohamed, C Lin. 2002. "Underemployment and Participation in Vocational Education." *International Journal of Vocational Education and Training*.
- **Linguistics / Communications.** M Weeber, JG Mork, AR Aronson. 2002. "Developing a Test Collection for Biomedical Word Sense Disambiguation." Lister Hill National Center for Biomedical Communications.
- \* **Marketing.** WS DeSarbo, AM Degerati, M Wedel, MK Saxton. 2001. "The Spatial Representation of Market Information." *Marketing Science*
- \* **Political Science / Linguistics.** K Quinn, BL Monroe, M Colaresi, M Crespin, D Radev. 2006. "An Automated Method of Topic-Coding Legislative Speech." Harvard University.
- **Political Science.** S Treier, S Jackman. 2003. "Democracy as a Latent Variable." Stanford.
- **Political Science.** T Blank, P Schmidt. 2003. "National Identity in a United Germany: Nationalism or Patriotism? An Empirical Test with Representative Data." *Political Psychology*.
- **Psychology.** M Eid, R Langeheine, ED Diener. 2003. "Comparing Typological Structures across Cultures by Multigroup Latent Class Analysis: A Primer." *Journal of Cross-Cultural Psychology*.
- **Psychology.** O Laudy, et al. 2005. "Applications of Confirmatory Latent Class Analysis in Developmental Psychology." *European Journal of Developmental Psychology*.
- \* **Public Health / Psychology.** ST Lanza, LM Collins. 2006. "A mixture model of discontinuous development in heavy drinking from ages 18 to 30: The role of college enrollment." *Journal of Studies on Alcohol*.
- **Public Health / Epidemiology.** TJ Croudace, M-R Jarvelin, MEJ Wadsworth, PB Jones 2003. "Developmental Typology of Trajectories to Nighttime Bladder Control: Epidemiologic Application of Longitudinal Latent Class Analysis." *American Journal of Epidemiology*.
- **Sociology.** G Moors, C Wennekers 2003. "Comparing Moral Values in Western European Countries between 1981 and 1999: A Multiple Group Latent-Class Factor Approach." *International Journal of Comparative Sociology*.
- **Sociology.** JK Vermunt, M Kalmijn. 2006. "Random-Effects Models for Personal Networks: An Application to Marital Status Homogeneity." *Methodology*.

*QuaSSI seeks to promote the development of an interdisciplinary community of Penn State faculty, students, and other researchers interested in the improvement and application of quantitative techniques and tools in social science research and training. For more information, please contact the QuaSSI Director, Burt Monroe ([burtmonroe@psu.edu](mailto:burtmonroe@psu.edu)), or visit [qssi.psu.edu](http://qssi.psu.edu).*