Paper 201-31

Performing Latent Class Analysis Using the CATMOD Procedure

David M. Thompson, Dept. of Biostatistics and Epidemiology,
Univ. of Oklahoma Health Sciences Center, Oklahoma City, OK

ABSTRACT

Latent class analysis (LCA), which is currently unavailable in SAS, has attracted the interest of clinical professionals and others who must place clients in diagnostic or other categories when a gold standard for doing so is poorly defined. This paper demonstrates a SAS approach to performing LCA by combining PROC CATMOD's loglinear modeling facility with conventional DATA steps.

LCA, a categorical analog to factor analysis, posits membership among unobserved classes to explain the pattern of association observed in a multi-dimensional contingency table. The SAS macro estimates two types of parameters: (1) the prevalence of each latent class and (2) the probabilities, conditional on class membership, that an individual demonstrates a specific response to an observed variable.

The program arrives at estimates using a classic expectation-maximization algorithm. Maximization steps specify a loglinear model in PROC CATMOD while DATA steps recalculate expected values for the latent class parameters. The program also calculates goodness of fit statistics, tracks iteration histories, and reports parameter standard errors.

In its current version, the program demonstrates how to employ information on four binary observed variables to estimate the structure of a latent variable with two hypothesized classes. PROC CATMOD's flexibility in loglinear modeling potentially permits estimation of larger models, and of models whose indicators demonstrate some local dependence even after positing a latent structure. The primary challenge to these expansions lies in the writing of more flexible DATA steps to calculate expected parameter values.

Keywords: Latent class analysis, PROC CATMOD, SAS macro, EM algorithm, diagnosis

INTRODUCTION

Latent class analysis is unavailable in SAS, one of the statistical packages most widely used by biomedical researchers. Investigators who wish to perform latent class analysis must write their own programs or obtain dedicated software that is sometimes expensive. This paper illustrates an approach to LCA that uses conventional PROC and DATA steps. In its current version, the program demonstrates how to employ information on four binary manifest variables to estimate the latent structure of a single unobserved variable with two hypothesized classes.

Latent class analysis is a categorical analog to factor analysis. Factor analysis attributes the covariance structure of a sample with multiple variables to unobserved factors. Similarly, latent class analysis posits unobserved classes to explain association in a multi-dimensional contingency table. The approach estimates two types of population parameters: (1) the prevalence of each latent class, the number of which the analyst must specify *a priori*; (2) the probabilities, conditional on latent class membership, that an individual will demonstrate a specific response to an observed variable.

The literature on latent class analysis customarily refers to the observed variables as "manifest indicators." When the manifest indicators in question are binary, these conditional probabilities are equivalent to sensitivities and specificities, but are estimated in the absence of an observable "gold standard." Latent class analysis has received attention for its value in validating diagnostic decisions in the absence of a gold standard for decision-making (Rindskopf & Rindskopf, 1986; Faraone & Tsuang, 1994; Formann & Kohlmann, 1996; Hui & Zhou, 1998; Albert, McShane, & Shih, 2001). A related area of research applies latent class analysis to assess inter-rater agreement when a definitive standard does not exist to support the use of more conventional statistics (Espeland & Handelman, 1989; Formann, 1994).

Rindskopf and Rindskopf (1986) illustrate how latent class analysis assigns class membership based on observed pattern of association (Figure 1) among four binary measures that clinicians collected in a sample of 94 patients with suspected cardiac disease. The four measures are an elevated Q-wave in the electrocardiogram (EKG) tracing, a history of chest pain, lipid profiles whose patterns are the opposite of those normally observed, and an abnormal elevation in a serum enzyme (CPK) whose presence suggests cardiac muscle damage. These data present a complex pattern of association and independence (Figure 1).

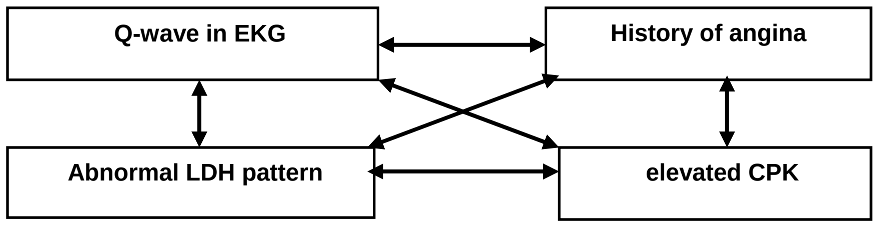

Figure 1. Observed variables demonstrate a complex pattern of association (after Rindskopf & Rindskopf, 1986).

Presented with these data, the latent class analysis algorithm arrives at a classification scheme (Figure 2) that explains or “explains away" (Goodman, 2002, p.4) the associations that are otherwise observed among the four binary indicators. Once the analysis has constructed appropriate latent classes, the observed indicators are independent, conditional on a subject’s class membership.

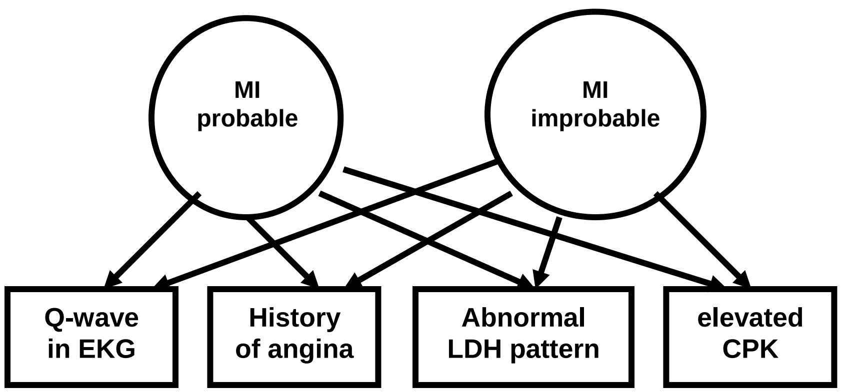

Figure 2. Conditional on latent class membership, observed variables are independent.

The identity or significance of the latent classes is informed by theoretical or substantive knowledge. For example, the Rindskopfs (1986) presume that they detected two latent classes that were related to the prevalence of myocardial infarction.

Investigators who wish to use SAS to perform latent class analysis must currently resort to its matrix language, PROC IML, or use lesser-known SAS procedures. IML modules that perform latent class analysis include one by the author (Thompson, 2003) and latent class regression macros developed at the Johns Hopkins School of Public Health (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997). Other researchers have applied latent class models to assess diagnostic accuracy by using SAS PROC NLIN, which performs nonlinear regression using weighted least squares estimation (Engels, Sinclair, Biggar, Whitby, & Goedert, et al., 2000; Blick & Hagen, 2002).

LOGLINEAR PARAMETERIZATION OF THE LATENT CLASS MODEL

Two equivalent sets of expressions describe the parameters of the latent class model. One parameterization is probabilistic. The other, which lends itself to use with PROC CATMOD, expresses the latent class model as a log-linear model wherein the manifest indicators (A, B, C, and D) are locally independent, given membership in the latent class X (Haberman, 1979; Espeland & Handelman, 1989; Heinen, 1996). The approach should, in principle, apply to PROC GENMOD, which can also estimate loglinear models.

In place of the probabilistic expression of locally independent joint probabilities

$$\pi^{ABCDX} = \pi^{A|X}\,\pi^{B|X}\,\pi^{C|X}\,\pi^{D|X}\,\pi^{X},$$

the loglinear parameterization defines the log probabilities (Heinen, 1996, equation 2.15, p. 51):

$$\ln \pi^{ABCDX} = \lambda + \lambda_i^{A} + \lambda_j^{B} + \lambda_k^{C} + \lambda_l^{D} + \lambda_t^{X} + \lambda_{it}^{AX} + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX}$$

$$\pi^{ABCDX} = \exp(\lambda + \lambda_i^{A} + \lambda_j^{B} + \lambda_k^{C} + \lambda_l^{D} + \lambda_t^{X} + \lambda_{it}^{AX} + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX})$$

Following Heinen (1996, equation 2.17, pp. 52-53), conditional probabilities are calculable in terms of loglinear parameters. For example,

$$\begin{aligned} P(A{=}i \mid X{=}t) &= P(AX) / P(X) \\ &= \Sigma_b\, \Sigma_c\, \Sigma_d\, \pi^{ABCDX} \;/\; \Sigma_a\, \Sigma_b\, \Sigma_c\, \Sigma_d\, \pi^{ABCDX} \\ &= \exp(\lambda_i^{A} + \lambda_{it}^{AX}) \;/\; \Sigma_a \exp(\lambda_i^{A} + \lambda_{it}^{AX}) \end{aligned}$$

Latent class prevalences are derived similarly:

$$\begin{aligned} \pi_x &= P(X{=}t) = \Sigma_a\, \Sigma_b\, \Sigma_c\, \Sigma_d\, \pi^{ABCDX} \;/\; \Sigma_a\, \Sigma_b\, \Sigma_c\, \Sigma_d\, \Sigma_x\, \pi^{ABCDX} \\ &= \exp \lambda_t^{X} / \Sigma_x \exp \lambda_t^{X} \end{aligned}$$

MAXIMUM LIKELIHOOD APPROACH TO LATENT CLASS ANALYSIS

Most statisticians credit Lazarsfeld and Henry (1968) with the origins of latent class analysis and Goodman (1974) with the computational breakthroughs that made it practical. Goodman's maximum likelihood approach (1974, pp. 216-218; see also McCutcheon, 1987, pp. 21-27; McCutcheon, 2002, p. 64) remains the standard way to estimating parameters in the latent class model. The joint probability π^{ABCD} of any observed response profile (A=i, B=j, C=k, D=l) can be summed across the hypothesized latent classes (X=t).

$$\pi^{ABCD} = \Sigma_t\ \pi^{ABCDX}$$

The probability of obtaining the observed frequency count n_{ijkl} for the response profile {i,j,k,l} is (after Dayton & Macready, 2000, p. 214):

$$(\pi^{ABCDX})^{n}{}_{ijklt}$$

Accordingly, the likelihood L of obtaining the observed counts for all possible response profiles is

$$L = \Pi_i\, \Pi_j\, \Pi_k\, \Pi_l\, \Pi_t\, (\pi^{ABCDX})^{n}{}_{ijklt}$$

and the corresponding log likelihood is

$$\log L = \Sigma_i\, \Sigma_j\, \Sigma_k\, \Sigma_l\, \Sigma_t\, n_{ijklt} \ln(\pi^{ABCDX}).$$

THE EXPECTATION-MAXIMIZATION (E-M) ALGORITHM

Because only the responses are observed, but latent class membership (X=t) is not, likelihood functions can be complex. A way around this difficulty involves estimating the missing information on latent class membership, then maximizing the likelihood for the provisional but complete 'data.' The approach involves alternating steps that first calculate the likelihood function's *expected* value, and then find the parameter values that *maximize* the function. The expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) first calculates the likelihood function's expected value, given the observed frequency counts for the complete (though provisional) data structure and given provisional values for the model's parameters estimates. Next, it maximizes the likelihood function to update estimates for the latent class model's parameters.

E-M ALGORITHM ACHIEVED THROUGH ALTERNATING PROC AND DATA STEPS

This paper describes a SAS approach to LCA that operationalizes the E-M algorithm by taking advantage of PROC CATMOD's loglinear modeling capabilities. In a series of "handoffs," loglinear LC model parameters are maximized in a PROC step, then their expected values are recalculated in a subsequent DATA step.

In its current version, the program demonstrates how to employ information on four binary manifest variables (a,b,c, and d) to estimate the structure of a single latent variable (x) with two hypothesized classes.

To begin the sequence, raw data (in the form of a contingency table that contains no information on latent class membership) is "augmented" by assigning each response profile to a latent class. Then, in a **maximization step**, PROC CATMOD produces updated parameter estimates by maximizing a likelihood function based on a loglinear model that features conditional independence:

```
ods output
    anova=mlr MaxLikelihood=iters estimates=mu covb=covb;
proc catmod data=two order=data;
  weight count;
  model a*b*c*d*x = _response_ / wls covb addcell=.1;
  loglin a b c d x a*x b*x c*x d*x;
run;
quit;
```

While the program described here uses one of CATMOD's three estimation techniques, namely weighted least squares (WLS), it can just as easily invoke options that employ the iterative proportional fitting (ML=IPF) or Newton-Raphson (ML=NR) algorithms. Similarly, the approach is potentially applicable if PROC GENMOD's options are set to estimate a loglinear model using Fisher scoring, in place of or in combination with Newton-Raphson estimation.

The PROC step's provisional estimates of loglinear latent class model parameters are passed to an **expectation step**. This data step (detailed below) transforms the loglinear parameters into probabilities, then uses them to update expected values for joint probabilities π^{ABCDX} and posterior probabilities for latent class membership $\pi^{X|ABCD}$.

```
data four;
/*read in estimated model parameters (lambda1-lambda9) that were output from PROC
step, then restructured in intervening data step.*/
  set mu;
/*vector of CATMOD's loglinear parameter estimates*/
    array mu [9] lambda1-lambda9;
/*vector of variable names*/
    array vars [4] a b c d;
/*vector of conditional and LC probabilities*/
    array p [10] pa_x1 pb_x1 pc_x1 pd_x1
                 pa_x2 pb_x2 pc_x2 pd_x2
                 px1 px2;
/*vector of joint probabilities*/
    array pjoint [2] piabcdx1 piabcdx2;
    do a=0 to 1;
      do b=0 to 1;
        do c=0 to 1;
          do d=0 to 1;
            do x=1 to 2;
              do var=1 to 4;
                value=vars[var];

/*conditional probabilities for latent classes 1 and 2*/
       p[4*(x-1)+var]
       = exp(mu[var]*(-1)**value + mu[var+5]*(-1)**(value+x))

        /(exp(mu[var]*(-1)**value + mu[var+5]*(-1)**(value+x))
    + exp(-mu[var]*(-1)**value - mu[var+5]*(-1)**(value+x)));
```

```
/*latent class probabilities*/
    p[8+x]= exp(mu[5]*(-1)**(x-1))
        / (exp(mu[5]*(-1)**(x-1)) + exp(-mu[5]*(-1)**(x-1)));

/*joint probabilities for each class*/
    pjoint[x]=p[4*x-3]*p[4*x-2]*p[4*x-1]*p[4*x]*p[8+x];

/*unconditional predicted response probabilities (across both classes)*/
    piabcd=piabcdx1+piabcdx2;

/*Posterior probabilities (pix1_abcd) are the probabilities that an individual
resides in latent class X=t, given observed responses A,B,C, and D*/
    pix1_abcd=piabcdx1/piabcd;
    pix2_abcd=piabcdx2/piabcd;

    if x=2 and var=4 then output;
                end;
              end;
          end;
        end;
      end;
   end;
drop lambda1-lambda9 value var x;
run;
```

Once this **expectation step** has calculated updated expected values for the model parameters, the new values are used to revise counts in a complete contingency (conditional) contingency table. These expected counts (which are products of observed counts and updated expected posterior probabilities) are passed, at the top of the iteration loop, to the next **maximization step**. The process repeats continues until it converges on stable estimates of latent class parameters. At this point, the predicted contingency table has a structure that demonstrates local independence among all manifest indicators, conditional on membership in a latent class.

CALCULATION OF MODEL PARAMETERS FROM CATMOD'S LOGLINEAR ESTIMATES

The DATA step that performs the expectation step in the E-M algorithm must convert the PROC step's loglinear parameter estimates into probabilities, then use these to update expected values for joint probabilities π^{ABCDX} and posterior probabilities for latent class membership $\pi^{X|ABCD}$. This section illustrates the mathematical details that underlie the DATA step.

When manifest indicators are binary (e.g., A=0,1) and the model posits two latent classes (X=1,2), conditional probabilities are calculated as, for instance:

$$P(A=0 \mid X=1) = \exp(\lambda_i^{A}+\lambda_{it}^{AX}) \; / \; \Sigma_a \exp(\lambda_i^{A}+\lambda_{it}^{AX})$$

$$=\exp(\lambda_0^{A}+\lambda_{01}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{01}^{AX}) + \exp(\lambda_1^{A}+\lambda_{11}^{AX})]$$

Latent class prevalences are calculated:

$$P(X=1) = \exp \lambda_1^{X} / \Sigma_x \exp \lambda_t^{X}$$

$$= \exp \lambda_1^{X} / [\exp \lambda_1^{X} + \exp \lambda_2^{X}]$$

Unless the number of loglinear parameters is restricted, models cannot estimate unique solutions. "Effect coding," where related parameters sum to zero, appropriately restricts the parameters. Specifically,

$$\lambda_0^{A} + \lambda_1^{A} = 0, \text{ therefore } \lambda_0^{A} = -\lambda_1^{A}$$

$$\lambda_1^{X} + \lambda_2^{X} = 0, \text{ therefore } \lambda_1^{X} = -\lambda_2^{X}$$

$$\lambda_{01}^{AX} + \lambda_{11}^{AX} = 0, \text{ therefore } \lambda_{01}^{AX} = - \lambda_{11}^{AX}$$

$\lambda_{02}^{AX} + \lambda_{12}^{AX} = 0$, therefore $\lambda_{02}^{AX} = -\lambda_{12}^{AX}$

The DATA step illustrated earlier employs effect-coding, under which expressions for conditional probabilities reduce to:

$$P(A=0 \mid X=1) = \exp(\lambda_0^{A}+\lambda_{01}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{01}^{AX}) + \exp(\lambda_1^{A} + \lambda_{11}^{AX})\,]$$

$$= \exp(\lambda_0^{A}+\lambda_{01}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{01}^{AX}) + \exp(-\lambda_0^{A}-\lambda_{01}^{AX})\,]$$

$$P(A=1 \mid X=1) = \exp(\lambda_1^{A}+\lambda_{11}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{01}^{AX}) + \exp(\lambda_1^{A} + \lambda_{11}^{AX}]$$

$$= \exp(-\lambda_0^{A}-\lambda_{01}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{01}^{AX}) + \exp(-\lambda_0^{A}-\lambda_{01}^{AX})\,]$$

$$P(A=0 \mid X=2) = \exp(\lambda_0^{A}+\lambda_{02}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{02}^{AX}) + \exp(\lambda_1^{A} + \lambda_{12}^{AX})\,]$$

$$= \exp(\lambda_0^{A}+\lambda_{02}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{02}^{AX}) + \exp(-\lambda_0^{A}-\lambda_{02}^{AX})\,]$$

$$P(A=1 \mid X=2) = \exp(\lambda_1^{A}+\lambda_{12}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{02}^{AX}) + \exp(\lambda_1^{A} + \lambda_{12}^{AX})\,]$$

$$= \exp(-\lambda_0^{A} - \lambda_{02}^{AX}) \; / \; [\exp(\lambda_0^{A}+\lambda_{02}^{AX}) + \exp(-\lambda_0^{A}-\lambda_{02}^{AX})\,]$$

Expressions for latent class prevalences reduce to:

$$P(X=1) \quad = \exp \lambda_1^{X} / [\exp \lambda_1^{X} + \exp \lambda_2^{X}]$$

$$= \exp \lambda_1^{X} / [\exp \lambda_1^{X} + \exp -\lambda_1^{X}]$$

$$P(X=2) \quad = \exp \lambda_2^{X} / [\exp \lambda_1^{X} + \exp \lambda_2^{X}]$$

$$= \exp -\lambda_1^{X} / [\exp \lambda_1^{X} + \exp -\lambda_1^{X}]$$

SIMULATION STUDIES

The series of tables illustrates the current program's performance when applied to 1000 data sets simulated to have a specified latent structure. Because the program is sensitive to randomly assigned starting values, it produced usable estimates for 406 of the simulations. It yielded estimates for latent class prevalences (Table 1) and conditional response probabilities associated with either of two pre-specified latent classes (Tables 2 and 3).

Latent Class Prevalences					
p(X=1)			p(X=2)		
Sim.	LCA	SE	Sim.	LCA	SE
0.5	0.4970	0.0882	0.5	0.5030	0.0882

Table 1. Latent class prevalences estimated for 406 simulated datasets.

Conditional Probabilities (where X=1)											
p(A|X=1)			p(B|X=1)			p(C|X=1)			p(D|X=1)		
Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE
0.9	0.7846	0.1543	0.9	0.7859	0.1670	0.1	0.2032	0.1678	0.1	0.1987	0.1667

Table 2. Conditional response probabilties estimated for latent class X=1 in 406 simulated datasets.

Conditional Probabilities (where X=2)											
p(A|X=2)			p(B|X=2)			p(C|X=2)			p(D|X=2)		
Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE	Sim.	LCA Mean Est.	SE
0.2	0.2531	0.1372	0.2	0.2381	0.1239	0.8	0.7593	0.1306	0.8	0.7771	0.1255

Table 3. Conditional response probabilties estimated for latent class X=2 in 406 simulated datasets.

CONCLUSION

This paper presents a use of PROC CATMOD that brings latent class analysis closer to the SAS arena. The approach addresses a fundamental limitation of the EM algorithm, the fact that it does not automatically produce standard errors for the parameter estimates. This prevents one from testing hypotheses or estimating confidence intervals related to the parameter estimates. While relatively complex techniques exist to obtain standard errors for individual parameters (Goodman, 1972; Louis, 1982), PROC CATMOD routinely calculates standard errors when it estimates parameters for loglinear models.

The program calculates standard goodness of fit statistics and iteration histories. The latter are probably of more value in evaluating the adequacy of a latent class model. In fact, the quality of LC parameter estimation is sensitive to initial estimates, and to the true nature of the latent structure. Improvements in the program will cause it solve a latent class problem many times, using randomly selected initial estimates at each iteration. The distribution of parameter estimates so obtained will help the analyst, better than fit statistics, to judge the adequacy of an estimate of a latent structure.

In its current version, the program demonstrates how to employ information on four binary manifest variables to estimate the structure of a single latent variable with two hypothesized classes. However, PROC CATMOD's flexibility in loglinear modeling potentially permits estimation of larger models, and of models whose indicators still demonstrate some local dependence even after positing a latent structure. The primary challenge to these expansions lies in the writing of a more flexible DATA step to calculate expected parameter values.

REFERENCES

Albert, P.S., McShane, L.M., & Shih, J.H. (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors. *Biometrics, 57*, 610-619.

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L. & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association, 92,* 1375-1386.

Blick, J., & Hagen, P.T. (2002). The use of agreement measures and latent class models to assess the reliability of classifying thermally marked otoliths. *Fishery Bulletin, 100,* 1-10.

Dayton, C.M., & Macready, G.B. (2002). Use of categorical and continuous covariates in latent class analysis. In J.A. Hagenaars & A.L. McCutcheon (Eds), *Applied Latent Class Analysis* (pp. 213-233). Cambridge, UK: Cambridge University.

Dempster, A.P, Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Engels, E.A., Sinclair, M.D., Biggar, R.J., Whitby, D., Ebbesen, P., Goedert, J.J., & Gastwirth, J.L. (2000). Latent class analysis of human herpesvirus 8 assay performance and infection prevalence in sub-Saharan Africa and Malta. International Journal of Cancer, 88, 1003-1008.

Espeland, M.A., & Handelman, S.L. (1989). Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics, 45,* 587-599.

Faraone, S., & Tsuang, M. (1994). Measuring diagnostic accuracy in the absence of a “gold standard”. *American Journal of Psychiatry, 151,* 650-657.

Formann, A.K. (1994). Measurement errors in caries diagnosis: Some further latent class models. *Biometrics, 50*, 865-871.

Formann, A.K., & Kohlmann, T. (1996). Latent class analysis in medical research. *Statistical Methods in Medical Research, 5*, 179-211.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215-231.

Goodman, L.A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In: J.A. Hagenaars & A.L. McCutcheon (Eds.). *Applied Latent Class Analysis*. Cambridge: Cambridge University.

Haberman, S.J. (1979). *Analysis of qualitative data, vol. 2: New developments*. New York: Academic Press.

Heinen, T. (1996). Latent class and discrete latent trait models: Similarities and differences. Thousand Oaks, CA: Sage.

Hui, S.L., & Zhou, X.H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research, 7,* 354-370.

Lazarsfeld, P. F., & Henry, N. W. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, B 44,* 226-233.

McCutcheon, A. L. (1987) *Latent class analysis*. Thousand Oaks, CA: Sage.

McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In J.A. Hagenaars & A.L. McCutcheon (Eds), *Applied Latent Class Analysis* (pp. 56-85). Cambridge, UK: Cambridge University.

Rindskopf, R., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine, 5*, 21-27.

Thompson, D. M. (2003). Comparing SAS-based applications of latent class analysis using simulated patient classification data. Unpublished doctoral dissertation, University of Oklahoma, Oklahoma City.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Name	David M. Thompson
Enterprise	Dept. of Biostatistics and Epidemiology, Univ. of Oklahoma Health Sciences Center
Address	CHB 321
City, State, ZIP	Oklahoma City, OK, 73190
Work Phone:	405-271-2229, ext. 48054
Fax:	405-271-2068
E-mail:	dave-thompson@ouhsc.edu
Web:	http://moon.ouhsc.edu/dthompso/