THE USE OF LATENT CLASS ANALYSIS IN MEDICAL DIAGNOSIS

David Rindskopf
Educational Psychology, CUNY Graduate Center, New York, NY 10016

KEY WORDS: Missing Data, Categorical Data, Logistic Regression, Measurement Error

The evaluation of medical signs, symptoms, and tests for the purposes of diagnosis is usually framed within the context of estimating the sensitivity and specificity of the indicator. Sensitivity is the probability that a person with the disease will be positive on the indicator; specificity is the probability that a person without the disease will be negative on the indicator. The estimation of sensitivity and specificity depends on knowing who does and does not have the disease; that is, there must be a "gold standard" for diagnosis. Rindskopf and Rindskopf (1986) applied latent class analysis to this problem, and showed that sensitivity and specificity could be estimated, under some conditions, even without a gold standard. The current paper briefly reviews those findings, and then discusses extensions of latent class methods for diagnosis. These methods have attracted much attention in the medical statistics community in the past 15 years; a search of Pub Med (September, 2002) resulted in over 200 references for the key word term "latent class"; a selection of these is included in the Reference section.

The next section of the paper gives a brief overview of latent class analysis, followed by a simple example from the literature. Next, I discuss a model with a predictor of latent class, using as an example data on children's wheeze measured at four ages, with asthma predicted by whether or not the mother smokes. Then I discuss a new conceptual model that adds floor and ceiling effects to logistic regression. This model is similar to models with error of measurement, but with a different interpretation. Finally, I present some implications of these models and their utility in medical diagnosis.

Latent Class Analysis: A Brief Overview

Latent class analysis hypothesizes the existence of one of more unobserved (underlying, latent) categorical variables to explain the relationships among a set of observed categorical variables. In the medical diagnosis context, the observed variables are signs, symptoms, or test results, usually dichotomized into a binary classification (positive and negative). The latent variable is true status on the disease; while this is often dichotomous (disease present or absent), it may not be (e.g., heart attack, congestive heart failure, or no heart problem). Sometimes there is more than one latent variable; each might correspond, e.g., to the presence or absence of a particular disease. In this paper, the examples are all cases in which there is only one latent variable.

The observed data are usually presented in terms of a crosstabulation of the observed variables. The statistical model has two kinds of parameters. Unconditional probabilities are the probabilities of being in each latent class (i.e., each level of the latent variable, if there is only one such variable; or each combination of levels, if there is more than one). Conditional probabilities are the probabilities of having a particular result on a test, given membership in a specific latent class. There is a set of conditional probabilities for each observed variable; these are assumed to be independent, so that the latent variable explains the relationships among the observed variables. The sensitivities and specificities of the observed measures are conditional probabilities, and thus are parameters in the model.

Models can be tested using any of a number of fit statistics appropriate for categorical data, including various statistics with chi-square distributions, and adjusted fit statistics such as Akaike's or Schwartz's Information Criteria. Care must be taken to be certain that a model is identified; that is, that all parameters are estimable. This can be accomplished through algebraic proof, or (more commonly) using numerical techniques. Sometimes more than one model is found that fits the data. Nested models can be compared to see if the extra parameters are needed, although there is some disagreement about when this is appropriate.

Once models are found that fit the data, the parameters can be interpreted. Another goal is to use Bayes' Theorem to examine how well individuals can be assigned to latent classes. In medical situations, this is the process of diagnosis. Some response patterns may be easier to classify than others. One can also determine whether it is possible to simplify the classification rules, e.g. by counting number of positive results or symptoms.

A Simple Example: Myocardial Infarction

The following example is summarized from Rindskopf and Rindskopf (1986). It illustrates the main points about the use of latent class analysis in medical diagnosis.

Data come from a study of patients admitted to an emergency room suffering from chest pain (Galen & Gambino, 1975). Each of four indicators was scored as either indicating a myocardial infarction (MI; commonly known as heart attack) or not indicating MI. The indicators included history, EKG (inverted Q-wave), and two blood tests (CPK and LDH). The data set consists of counts of the number of patients in each of the 16 possible patterns of indicators.

The data were consistent with a simple 2-class model (LR=4.29, df=6, p=.64), where the classes represented those with and without MI. The data were inconsistent with several other possible models, including the model of complete independence of indicators, and a quasi-independence model.

Table 1 contains the parameter estimates for the model. The unconditional probabilities of being classes 1 and 2 are about .46 and .54 respectively. To determine what these classes mean, one must examine the conditional probabilities. For each indicator, in class 1 there is a relatively high prob

Table 1. Parameter Estimates for MI Data, 2-class unrestricted model

		mi	no mi
		0.4578	0.5422
cpk	1	1.0000	0.1956
cpk	0	0.0000	0.8044
ldh	1	0.8279	0.0269
ldh	0	0.1721	0.9731
his	1	0.7914	0.1951
his	0	0.2086	0.8049
qwa	1	0.7669	0.0000
qwa	0	0.2331	1.0000

ability of the indicator being positive; for class 2, there is a relatively low probability of the indicator being positive. Therefore, class 1 represents those with MI, and class 2 represents those without MI. The indicators vary in their sensitivities and specificities; for example CPK has a high sensitivity, but low specificity, while Q wave on an EKG has a high specificity but lower sensitivity. History is modest on both measures, while LDH is similar to EKG results. Sometimes a graphical presentation of results is easier to examine; Figure 1 shows a plot of sensitivities by specificities, similar to ROC plots in signal detection theory. The ideal indicator would be at the point (1,1) in the upper right-hand corner. The closer an indicator is to that ideal point, the better the indicator is.

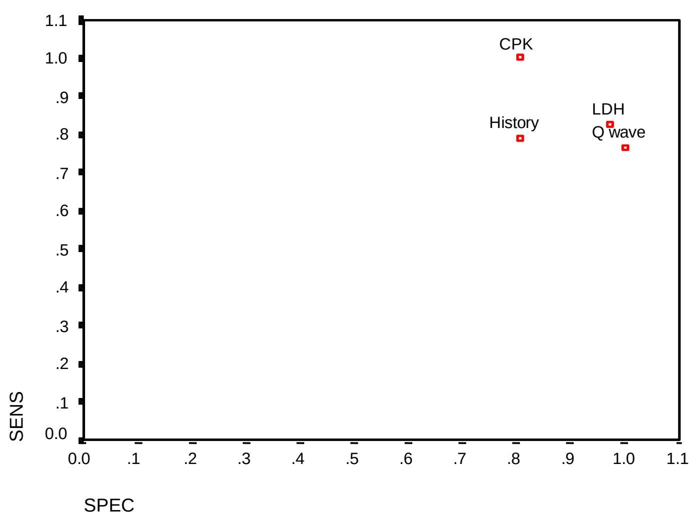

Figure 1. Plot of sensitivity and specificity of indicators in MI data, similar to ROC plot.

The actual statistical process of diagnosis is the assignment to latent classes. The usual procedure is to use all observed variables, as shown in Table 2. The largest probability of error is for patients who were positive on CPK and history, but negative on LDH and Q-wave.

In medical diagnosis, unlike the typical situation in latent class analysis, one might be interested in class assignment on the basis of only a subset of the observed variables. For example, one might want to use only history and EKG results, which are available more quickly than the LDH and CPK, which are based on blood tests. Technically, of course, this is no problem; one merely applies Bayes' Theorem. Results are shown in Table 3. If Q wave is positive, history is irrelevant; the patient had an MI. If both are negative, the patient probably did not have an MI. If Q wave is negative but history is positive, there is quite a bit of uncertainty.

Table 2. Assignment to latent class for MI data, with error probabilities for each pattern of indicators

C	L	H	Q	F	1	2	Diag	Err
1	1	1	1	24	1.00	0.00	1	0.00
1	1	1	0	5	0.99	0.01	1	0.01
1	1	0	1	4	1.00	0.00	1	0.00
1	1	0	0	3	0.89	0.11	1	0.11
1	0	1	1	3	1.00	0.00	1	0.00
1	0	1	0	5	0.42	0.58	2	0.42
1	0	0	1	2	1.00	0.00	1	0.00
1	0	0	0	7	0.04	0.96	2	0.04
0	1	0	0	1	0.00	1.00	2	0.00
0	0	1	0	7	0.00	1.00	2	0.00
0	0	0	0	33	0.00	1.00	2	0.00

Note: C= CPK, L=LDH, H=History, Q=Q wave. Response patterns with zero observed frequency are omitted.

Table 3. Using a subset of indicators to make a diagnosis

Q-wave	History	P(MI|Q,H)	Prop.
0	0	.04853	.47
0	1	.44394	.18
1	0	1.00000	.06
1	1	1.00000	.29

Latent class analysis with a categorical predictor of class: Wheeze in children

In some analyses one or more predictors of the latent class is available. Consider data on wheeze in children from the Six Cities study (see, e.g., Cunningham, et al, 1994), for which data from one city are widely used as an example data set. Children were assessed for presence or absence of wheeze each year from age 7 through 10 years. If no other information were available, a latent class analysis similar to that for the MI data would be suitable. In this case, one such additional variable was whether or not the mother smoked. Theoretically, passive smoke might affect asthma in children.

The results of fitting a latent class model with mother's smoking as a predictor of class are contained in Table 4. First, note that at each age at which wheeze was measured, the probability of wheeze was somewhat high in class 1, and very low in class 2; this makes class 1 the class with probable asthma. This class contains about 16 percent of the children in the study; the class without asthma contains about 84 percent of the children. While specificity of wheeze was good (relatively low) at each age, sensitivity was not as high as might be desired, being generally above .50 but less than .75.

Table 4. Parameter estimates for wheeze data

class		1	2
p(class)		0.1607	0.8393
smo	yes	0.4487	0.3290
age7	yes	0.5924	0.0796
age8	yes	0.7222	0.0636
age9	yes	0.7019	0.0542
age10	yes	0.5119	0.0418

Note: LR = 13.6793, df=20, p=0.8464. Data are available online at http://www.statsci.org/data/general/wheeze.html

The probability of the mother smoking was about 45 percent among children with asthma, and 33 percent among children without asthma. To test whether this effect was significant, the model was run again, constraining the conditional probabilities of smoking to be equal in both classes. This model fit the data very well.

For the model with equal rates of smoking in the two classes, Bayes' Theorem was used to assign children to class based on the four observed measures of wheeze. The results are shown in Table 5. In general, as would be expected, the most difficult response patterns to assign were those where wheeze occurred on two occasions and did not occur on the other two. Even in these cases, the probability was relatively high (about .6 to .8) that the child actually had asthma. In fact, because of the approximately equal sensitivities and specificities across ages, it is possible to deduce a simple rule for assignment to classes: If wheeze occurred not at all or only once, the child probably does not have asthma; if wheeze occurred two or more times, the child probably does have asthma. Such a simple rule will not always suffice, even for a two-class model.

Table 5. Assignment to latent classes, wheeze data, for model with no effect of mom's smoking; sorted by modal class

A7	A8	A9	A10	P(1)	P(2)	Class	P(err)
1	1	1	1	0.9961	0.0039	1	0.0039
1	1	1	2	0.9131	0.0869	1	0.0869
1	1	2	1	0.8597	0.1403	1	0.1403
1	2	1	1	0.8693	0.1307	1	0.1307
2	1	1	1	0.9373	0.0627	1	0.0627
1	1	2	2	0.2030	0.7970	2	0.2030
1	2	1	2	0.2165	0.7835	2	0.2165
1	2	2	1	0.1388	0.8612	2	0.1388
1	2	2	2	0.0067	0.9933	2	0.0067
2	1	1	2	0.3832	0.6168	2	0.3832
2	1	2	1	0.2660	0.7340	2	0.2660
2	1	2	2	0.0148	0.9852	2	0.0148
2	2	1	1	0.2822	0.7178	2	0.2822
2	2	1	2	0.0161	0.9839	2	0.0161
2	2	2	1	0.0094	0.9906	2	0.0094
2	2	2	2	0.0004	0.9996	2	0.0004

Note: If positive 0 or 1 time, class = 1; If positive 2 or more times, class = 2

Other contributors to the literature who have proposed similar models include Clogg (1981), Dayton and Macready (1988, 2002), Hagenaars (1990), Vermunt (1997), and Vermunt and Magidson (2002).

Formally, this model is equivalent to a two-group latent class model, with equality of all conditional probabilities across the two groups (mothers who do and do not smoke). There are several advantages of the current formulation over the two-group form of the model. First, it is easier to extend this approach to the case with multiple predictors. Second, this method gives a direct estimate of the relationship between the predictor and the latent variable; this is even more important with multiple predictors. Third, it is easy to extend this model to the case where there are continuous or quasi-continuous predictors (e.g. several levels of a quantitative variable). The following example illustrates a variation on this model.

Logistic regression with floor and ceiling effects

Logistic regression models assume that for a low enough value of the predictor(s), the probability of a response is zero, and that for a high enough value of the predictor(s), the probability of a response is one. In some applications, one or both of these assumptions is likely to be false. For example, in predicting graduation from college using SAT scores, there will undoubtedly be a proportion of students who graduate in spite of low SAT scores, and another group who will not graduate in spite of high SAT scores.

In this section we consider models in which the asymptotes of the logistic regression equation are not constrained to be zero and one as they are in traditional models. The idea is presented graphically in Figure 2.

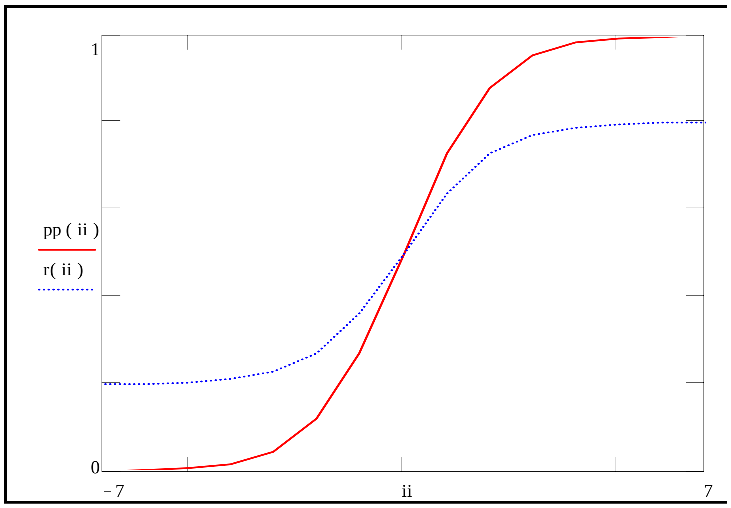

Figure 2. Logistic regression with floor and ceiling effect

This idea has some precedent in related areas. In item response theory (IRT) models, guessing parameters allow items to have a probability greater than zero of being responded to correctly, primarily to adapt to guessing on true-false or multiple choice tests. Finney proposed a probit model with a floor effect in toxicity studies; he called it a model for toxicity with natural mortality. He never seems to have had a ceiling effect for situations in which a certain proportion of animals were not affected by a poison or drug, and his floor effect never found its way into standard computer programs for logistic regression or probit models.

Technically, the model is identical to a model for errors in variables proposed by Ekholm and Palmgren (1982). The interpretation is different, even though the fit of the model is identical. In this case, Ekhlolm and Palmgren would interpret the floor and ceiling effect as solely due to errors of measurement.

The model with floor and ceiling effects (or, equivalently, the logistic regression model with errors of measurement) is also equivalent to a special kind of latent class model. This model has only one (binary) observed indicator, and one or more predictors of the latent variable. With a single continuous predictor having five or more levels, the model is identified, even though traditional latent class models would require four or more indicators. The interpretation of the latent class model would be similar to that of the Ekholm and Palmgren model, rather than the model with floor and ceiling effects.

Table 6 presents the results of an analysis of data on wheeze in coalminers, originally from Ashford & Sowden (1970). These data were analyzed by Ekholm and Palmgren using their errors-in-variables model. Here we should get the same results, but the interpretation is different. The floor effect could also be due to some miners contracting asthma (or wheeze alone) from other causes, and the ceiling effect could be miners whose lungs are particularly resistant to developing wheeze, and would not develop it no matter how long they were miners. Only replicate measurements at the same time period would allow an analysis that would separate these possibilities from errors of measurement.

Table 6 Coal miner data, wheeze, model with floor and ceiling effects

	Latent Class	
	1	2
Marginal prob	0.7052	0.2948
Wheeze status		
yes	0.0000	0.6792
no	1.0000	0.3208
Age group		
1	0.9224	0.0776
2	0.8904	0.1096
3	0.8474	0.1526
4	0.7915	0.2085
5	0.7219	0.2781
6	0.6395	0.3605
7	0.5480	0.4520
8	0.4532	0.5468
9	0.3616	0.6384

In keeping with the different interpretation of this model, the entries in Table 6 are not all the same as in the usual latent class output. The marginal probabilities are the unconditional probabilities of being in each latent class. Wheeze status contains the conditional probability of having wheeze or not, given class membership. The numbers for age group, however, are the reverse of the usual interpretation; they are the conditional probability of latent class membership given age, and must be read in rows instead of columns. For example, miners in age group 1 have a probability of about .92 of being in latent class 1, which represents no wheeze.

The results are consistent with those of Ekholm and Palmgren. The floor is 0; that is, extrapolating to younger ages would give negligible probability of wheeze. The ceiling is about .68, so extrapolating to older miners would give a probability of .68 of wheeze; we would predict that about 32 percent of miners would never get wheeze.

Discussion

Latent class analysis has many advantages over traditional methods of (i) estimating sensitivity and specificity, and (ii) developing diagnostic rules on the basis of medical tests and indicators. The lack of need for a gold standard is the most important reason for preferring latent class models. In addition, one can determine whether certain simple rules for diagnosis (e.g. symptom counts) are reasonable.

Traditional latent class models can be extended in several ways that are useful in medical statistics. First, one can construct and test prediction models for the true (latent) status on the disease. Second, these extensions can allow the estimation of sensitivity and specificity with fewer indicators than a traditional latent class model. Third, with appropriate data, it is possible to separate floor and ceiling effects from measurement error. This requires repeat observations at the same time, and either several times of measurement, or a set of predictors.

References Cited

Ashford, J. R., & Sowden, R. D. (1970). Multivariate probit analysis. *Biometrics, 26*, 535-546.

Clogg, C. C. (1981). New developments in latent class analysis. Pp. 215-246 in D. J. Jackson and E. F. Borgatta (Eds.) *Factor analysis and measurement in sociological research: A multi-dimensional perspective*. London: Sage.

Cunningham J, Dockery D. W., & Speizer F. E. (1994). Maternal smoking during pregnancy as a predictor of lung function in children. *Am J Epidemiol*, 139:1139-52.

Dayton, C. M., and Macready, G. B. (1988). Concomitant-variable latent class models, *Journal of the American Statistical Association*, 83, 173-178.

Dayton C. M. , Macready G. B. (2002). Use of categorical and continuous covariates in latent class analysis. In McCutcheon A, Hagenaars J (eds.), *Advances in latent class modeling,* Cambridge University Press.

Ekholm, A., & Palmgren, J. (1982). A model for a binary response with misclassifications. In R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalised Linear Models,* pp 128-143.. New York: Springer-Verlag.

Finney, D. J. (1952). *Probit analysis.* (2nd ed) Cambridge: Cambridge University Press.

Galen, R. S., & Gambino, S. R. (1975). *Beyond normality*. NY: Wiley.

Hagenaars, J. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Newbury Park, CA: Sage.

Hagenaars J, & McCutcheon A (Eds) (2002). *Applied latent class analysis*. Cambridge, UK: Cambridge University Press

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

Vermunt, J. K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.

Vermunt, J. K. & Magidson, J. (2002). Latent class cluster analysis. In Hagenaars and McCutcheon, eds., *Advances in latent class models*. Cambridge, UK: Cambridge University Press.

Additional 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*(2), 610-9.

Espeland, M. A., & Handelman, S. L. (1989). Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics, 45*, 587-99.

Goetghebeur, E, Liinev, J, Boelaert, M, Van der Stuyft, P. (2000). Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. *Statistical Methods in Medical Research 9*(3), 231-48.

Goldstein, J. M., Santangelo, S. L., Simpson, J. C., Tsuang, M. T. (1990) The role of gender in identifying subtypes of schizophrenia--A latent class analytic approach. *Schizophrenia Bulletin, 16*, 263-275

Jorgensen, P., Jensen, J. (1990). Latent class analysis of deluded patients. *Psychopathology, 23*, 46-51

Sullivan, P. F., Kessler, R. C., & Kendler K. S. (1998). Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *Am J Psychiatry, 155*(10):1398-406

Uebersax, J. S. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association*, 88, 421-427.

Uebersax, J. S. (1994). Latent class analysis of substance abuse patterns. In L. Collins & L. Seitz (Eds.), *Advances in data analysis for prevention intervention research*. NIDA research monograph, No. 142. Rockville, MD: National Institute on Drug Abuse. Available online at http://www.drugabuse.gov/pdf/monographs/142.pdf

Uebersax, J. S., & Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine, 9*, 559-572.

Walter, S. D., and Irwig, L. M. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. *Journal of Clinical Epidemiology, 41*, 923-937

Young, M. A. (1983) Evaluating diagnostic criteria: A latent class paradigm. *Journal of Psychiatric Research, 17*, 285-296.

Young, M. A. and Tanner, M. A. (1983). Recent advances in the analysis of qualitative data with applications to diagnostic classification. In Gibbons, R. D. and Dysken, M. (eds.) *Statistical and Methodological Advances in Psychiatric Research.* New York: Spectrum.