Read the paper - Rindskopf D and Rindskopf W (1986) "The value of latent class analysis in medical diagnosis" Statistics in Medicine, 5, 21-27.

STATISTICS IN MEDICINE, VOL. 5, 21–27 (1986)

THE VALUE OF LATENT CLASS ANALYSIS IN MEDICAL DIAGNOSIS

DAVID RINDSKOPF*

Educational Psychology, CUNY Graduate Center, 33 West 42nd Street, New York, NY 10036, U.S.A.

AND

WALLACE RINDSKOPF

SUMMARY

Assessment of the value of diagnostic indicators such as symptoms and laboratory tests results from calculation of the sensitivity and specificity of the indicators. Knowledge of the rate of occurrence of the disease allows for additional calculations of the error rates in using an indicator. These calculations are accurate only when the data on which they are based are reliable. If the diagnosis, which is used as the criterion for computing the sensitivity and specificity, is not accurate, then the resulting calculations will be in error. We show how a statistical method, latent class analysis, allows for the estimation of the characteristics of indicators even when an accurate diagnosis is unavailable. In addition, the method deals with several indicators at once, and provides a way to combine the information from all the indicators to make a diagnosis.

KEY WORDS Diagnosis Latent class analysis

	Disease +	Disease -
Test +	a	b
Test -	c	d

Sensitivity $= Pr(Test + |Disease+)$

Specificity $= Pr(Test - |Disease-)$

Positive Predictivity $= Pr(Disease + |Test+)$

Negative Predictivity $= P(Disease - |Test-)$

Depending on how data were sampled in order to get a,b,c,d we can estimate these conditional probabilities either directly from the table or using Bayes rule plus additional information about prevalence.

Table I. Data, expected values and assignment to latent classes based on two-class model

Indicator				Frequencies			
Q-wave	History	LDH	CPK	Observed	Expected	Class	Probability of correct classification
yes	yes	yes	yes	24	21·62	2	1·000
no	yes	yes	yes	5	6·63	2	0·992
yes	no	yes	yes	4	5·70	2	1·000
no	no	yes	yes	3	1·95	2	0·889
yes	yes	no	yes	3	4·50	2	1·000
no	yes	no	yes	5	3·26	1	0·580
yes	no	no	yes	2	1·19	2	1·000
no	no	no	yes	7	8·16	1	0·956
yes	yes	yes	no	0	0	—	—
no	yes	yes	no	0	0·22	1	1·000
yes	no	yes	no	0	—	—	—
no	no	yes	no	1	0·89	1	1·000
yes	yes	no	no	0	0	—	—
no	yes	no	no	7	7·78	1	1·000
yes	no	no	no	0	0	—	—
no	no	no	no	33	32·12	1	1·000

The probabilities of correct classification cannot be computed for cells with expected values of zero, and are left blank. Also, the class to which these people should be assigned is undetermined.
The observed frequencies are from Galen and Gambino.[1]

Table II. Parameter estimates for two-class model

	Class 1	Class 2
Unconditional class probabilities	0·542	0·458
Conditional probabilities of indicators, given latent class		
Positive Q-wave	0·000	0·767
Classic history	0·195	0·791
Flipped LDH	0·027	0·828
High CPK	0·196	1·000

1. In Table 1 when the authors refer to Class explain how the 1 and 2 are determined.

   The classes 1 and 2 are determined by calculating the probability that an individual is in one class give their response vector $\mathbf{x}$, i.e. using Bayes Theorem $p(c = k|\mathbf{x})$ is calculated for each $\mathbf{x}$. The predicted class is taken to be the class (i.e. k=1 or 2) corresponding to the largest probability.

2. The authors say on page 23 "The first step in the analysis is to establish that there are actually relationships among the variables. With no relationship among the observed variables, there cannot be more than one latent class." Does this make sense, explain why or why not.

   Here is a very good wrong answer: Actually no. If there are no relationships at all between the observed variables (i.e., r=0), then doesn't that assume that the variables are measuring separate and independent constructs with no possible underlying latent class, particularlay an overall one unifying latent class. If anything, with 4 observed variables with no relationship between them, I would guess these could represent 4 separate latent classes...so it doesn't make sense to me that there can be only one latent class with NO relationship among the observed variables.

Here is the right answer. Yes it makes sense. The basic idea of latent class analysis is that given that we know what class someone is in, the observed variables for people in that class will NOT be correlated. So if the observed variables for the entire sample are uncorrelated, then the entire sample must be in the same 1 class.

3. Based on the results in Table II, which class is the MI class, Class 1 or 2. Explain why you conclude that.

   Table 2 shows the conditional probabilities for each of the indicators. The table reveals that those in Class 1 have low probabilities of being positive for each of the indicators, while those in Class 2 have high probabilities of being positive for each of the indicators. Because the indicators are diagnostics for MI, positive indicators would suggest MI, thus Class 2 would be the patients that have the disease due to the high probability of positive indicators.

4. Additional note: The conditional probabilities in Table II are interpreted as sensitivities and specificities for the 4 different tests.

5. What do you think the authors mean when they talk about "equal gravity of each type of error"? The authors say on the bottom of page 24 "...a primary purpose is to provide guidance in classification of patients as either having or not having MI. To do this, we can use the conditional and unconditional probabilities in a formula related to Bayes' theorem to find, for any pattern of indicators, the probability that a person with that pattern is in a particular latent class. With equal gravity for each of the two possible types of errors (false diagnosis of MI; failure to diagnose MI when present) we could classify people into the most likely category given their particular symptoms. Table 1 shows...the latent class for each observed response pattern (given equal gravity of each type of error) and the probability of correct classification."

6. What is the estimated sensitivity of Positive Q-wave? What is the estimated specificity of High CPK?