FIFTEEN

Latent Class Models for Contingency Tables with Missing Data

Christopher Winship, Robert D. Mare, and John Robert Warren

1. INTRODUCTION

Missing data is a common problem in many types of data analysis. In this paper, we show how to deal with missing data in loglinear analyses of frequency tables.[1] Our approach is based on two ideas: (1) Latent class models can be adapted to contingency tables with missing data by defining variables that are latent (missing) for some cases and are manifest (observed) for others; and (2) latent class models can be viewed as loglinear models for tables in which some cells are unobserved or partially observed. Using our approach, we can retain the loglinear model framework and notation and deal with missing data through a modest extension of the standard model. Flexible software for latent class models, such as DNEWTON (Haberman, 1989) and LEM (Vermunt, 1996) is required, but the conceptual extension of elementary loglinear models is straightforward.[2]

By explicitly incorporating missing data into the analysis of a contingency table, one can address two concerns. First, a researcher may be worried about the possible loss of statistical power or precision of estimation that results when observations with missing data are excluded from an analysis. If many cases have missing data on at least one variable, exclusion of these cases from the analysis may substantially reduce the sample and create unacceptably large standard errors. One may want to incorporate cases with missing data into the analysis so that the information associated with these cases can be used to obtain more precise estimates. When the loss of statistical power is the only problem, incorporating missing data into a loglinear analysis is usually straightforward.

Second, one may be concerned that the exclusion of missing data may result in inconsistent parameter estimates in loglinear models. This

is likely to occur if there is a systematic mechanism producing the missing data. In this case one should develop a model of the missing data process jointly with the substantive model of interest. Later, we discuss more precisely the types of missing data processes that lead to inconsistent estimates. Although correcting problems of this type can be difficult, it may be essential if one is to make appropriate substantive conclusions.

The next section of the paper presents an example that illustrates that how one deals with missing data affects both the precision of parameter estimates and the substantive conclusions that one draws. The subsequent section briefly discusses one conventional approach to dealing with missing data, namely, adding a category to a variable for missing data. We show that this procedure typically leads to inconsistent estimates. Next we show how a contingency table can be extended to incorporate missing data so that loglinear models for partially observed data can be applied. We then consider alternative assumptions about missing data and the models that these assumptions imply. We examine different models for our first example, and then present a more complex empirical example. We then discuss various problems in estimation and identification.

2. EXAMPLE 1: PRENATAL CARE AND INFANT MORTALITY

Panel (a) of Table 1 presents data on the relationship between prenatal care and infant mortality in two clinics. These data were first analyzed in Bishop, Fienberg, and Holland (1975). Little and Rubin (1987)

Table 1. Contingency Table with Partially Classified Observations

		Survival (S)	
Clinic (C)	**Prenatal Care (P)**	**Died**	**Lived**
(a) Completely Classified Cases			
Clinic A	less	3	176
	more	4	293
Clinic B	less	17	197
	more	2	23
(b) Partially Classified Cases (Clinic Missing)			
	less	10	150
	more	5	90

Source: (a) Bishop et al. (1975), Table 2.4-2; (b) artificial data from Little and Rubin (1987, Table 9.8, p. 187).

supplement these data with the hypothetical data in Panel (b) of the table, which contain 255 infants whose clinic ID is missing. A researcher who wishes to analyze the combined data in the two panels faces two problems. First, data are missing for 255 out of 970 cases. If all cases with missing data were omitted, this would substantially reduce statistical power. This is a particularly serious issue because the response variable, infant mortality, measures a rare event.

Second, assumptions about the true values of the missing data may markedly affect the estimated effect of prenatal care on infant mortality. Table 2 illustrates a range of possible outcomes under various extreme assumptions. All other possible assignments of the missing data are less

Table 2. Mortality Rates for Data in Table 1 Under Alternative Assumptions about Missing Data

Assumption	Clinic	Level of Prenatal Care: Less	More	Difference
1. Observed data collapsed over clinic		5.4	2.6	2.8
2. Complete data	A	1.7	1.3	0.4
	B	7.9	8.0	0.1
Missing are all:				
3. Clinic = A	A	3.8	2.3	0.5
4. Clinic = B	B	7.2	5.8	1.4
5. If care = less, clinic = A	A	3.8	1.3	2.5
If care = more, clinic = B	B	7.9	5.8	2.1
6. If care = more, clinic = A	A	1.7	2.3	−0.6
If care = less, clinic = B	B	7.2	8.0	−0.8
7. If survival = died, clinic = A	A	6.9	2.3	4.3
If survival = lived, clinic = B	B	6.4	1.8	4.6
8. If survival = lived, clinic = A	A	1.3	1.0	0.3
If survival = died, clinic = B	B	12.6	23.3	−10.7
9. If care = less & survival = died, or care = more & survival = lived, clinic = A	A	6.9	1.0	5.9
If care = more & survival = died, or care = less & survival = lived, clinic = B	B	11.8	23.3	−11.5
10. If care = less & survival = died, or care = more & survival = lived, clinic = B	A	0.9	3.0	−2.1
If care = more & survival = died, or care = less & survival = lived, clinic = A	B	12.0	1.7	10.3

extreme in that they yield estimates that fall within the range of those reported in Table 2.

Estimates of the effect of prenatal care vary across a wide range. If clinic status is ignored (assumption 1), the estimated difference in mortality rates by level of prenatal care is large (5.4% − 2.6% = 2.8%). However, if only complete data are used and estimates are computed within clinics (assumption 2), the estimated differences are very small (0.4% for Clinic A and 0.1% for Clinic B). Assuming that all missing data are from Clinic A (assumption 3) or Clinic B (assumption 4) produces somewhat higher estimates of the effect of the difference (0.5% and 1.4%, respectively). Assuming more complex patterns of missing data increases the range of possible effects. Under assumptions 5–10, the pattern of missing data is the opposite for the two clinics. For example, assumption 5 is that those who in fact were in Clinic A but are missing information on their clinic received less prenatal care, and that those who in fact were in Clinic B but are missing this information had more prenatal care. Across these assumptions for Clinic A, the difference in the mortality rates between the two levels of prenatal care ranges from −2.1% to +5.9%, and for Clinic B it ranges from −11.5% to +10.3%. The estimates of the percentage difference effects are greatest for the most complex assumed patterns for the missing data, that is, assumptions 9 and 10, where the missing data pattern is a function of both level of care and survival status and the effect of level of prenatal care is the opposite in the two clinics. Given the wide range of estimates in Table 2, it is impossible to infer for either clinic whether more prenatal care is beneficial. What this exercise does show, however, is that the most extreme estimates of the effect of prenatal care occur when we assume that the mechanism generating missing data differs between the two clinics.

3. CONVENTIONAL APPROACHES TO MISSING DATA

The most common approach to incorporating missing data into a loglinear analysis is to add a "missing" category to the variables with missing data. This is analogous to creating a dummy variable in a regression analysis to indicate that respondents are missing on a variable. In both cases, inconsistent estimates generally result (Little and Rubin, 1987). We consider the loglinear case here. The inclusion of a missing data category may affect both estimates of associations among the variables that contain no missing data and also estimates of associations involving variables that have missing data. The top panel of Table 3 presents a hypothetical table

Table 3. Hypothetical Data for Showing the Effects of Using a Missing Data Category

No Missing Data ($N = 3{,}600$)						
Marginal Associations						
	$X=0$	$X=1$				
$Z=0$	1,000	800				
$Z=1$	800	1,000				
Odds Ratio	1.56					
Conditional Associations						
	$Y=0$		$Y=1$			
	$X=0$	$X=1$	$X=0$	$X=1$		
$Z=0$	800	400	200	400		
$Z=1$	400	200	400	800		
Odds Ratio	1.0		1.0			
With Missing Data Conditional Associations						
	$Y=0$		$Y=1$		$Y=$ Missing	
	$X=0$	$X=1$	$X=0$	$X=1$	$X=0$	$X=1$
$Z=0$	640	320	160	320	200	160
$Z=1$	320	160	320	640	160	200
Odds Ratio	1.0		1.0		1.56	

with three variables, X, Y, and Z, in which no data are missing. In this table X and Z are marginally associated with an odds ratio of 1.56, but, within levels of variable Y, X, and Z are independent.

Now suppose that 20% of the data on Y are missing purely at random. The lower panel of Table 3 arrays these data and distinguishes between missing and nonmissing data by adding a third category to Y. If we exclude the missing data from the sample, we get the same result as with the original data – X and Z are conditionally independent. However, the standard errors of any estimates will be larger because of the smaller sample. Consider the association between X and Z for those cases in which Y is missing. For these cases, X and Z are conditionally dependent with an odds ratio of 1.56, the same odds ratio as in the marginal table between X and Y in the original data. This is because the missing category contains cases in which both $Y = 0$ and $Y = 1$. In fact, because the data are missing completely at random, it contains the same proportion of cases with $Y = 0$ and $Y = 1$ as in the full sample. Thus, the conditional association of X and Z for those with missing data is equal to the marginal association between X and Z in the original table, not their conditional relationship. If one uses both the missing and nonmissing cases and estimates a common value for the conditional association between X and Z, the estimate will lie between the marginal and the true conditional association between X and Z. Because it is impossible to control properly for Y for the cases with missing values on Y, including a separate category for cases that are missing on Y will not give a consistent estimate for the association of X and Z conditional on Y.[3]

4. LATENT CLASS MODELS

The most common use of latent class models is to account for the associations among observed categorical variables with one or more latent categorical variables. One examines alternative models for a cross-classification of the observed variables and typically assumes that the observed variables are conditionally independent within categories of the latent variable(s) (Goodman, 1974; Dayton and Macready, 1980; McCutcheon, 1987; Clogg, 1988; Hagenaars, 1988, 1993). By definition, the latent dimensions of tables in latent class models have missing data for every observation. In contrast, in latent class models for missing data, data are observed for some cases and unobserved for others. Thus, in the approach discussed here, the latent class models have latent *cells* rather than latent dimensions.

In our discussion and analysis of contingency tables with missing data, we use several configurations of the data. We distinguish among three configurations: (1) observed table, (2) complete data table, and (3) expanded table. In the observed table, each variable that has missing data contains an additional category for missing observations. The complete data table is the cross-classification of observations that have no missing data. The expanded table, which is only partially observed, consists of the substantive variables cross-classified by a set of dichotomous variables that indicate whether observations are missing on each of the original variables. We term these variables *missing indicators*. The expanded table represents not only how the substantive variables in the data are related to each other, but also how the missing indicators relate to each other and to the substantive variables.

We illustrate the relationships among these tables using the infant mortality data shown in Table 1. The data contain four variables: clinic (C), prenatal care (P), survival status (S), and a missing on clinic indicator (M) coded not missing = 0 and missing = 1. The 2^4 table of C by P

by S by M is the expanded table. It contains two 2^3 subtables. One subtable table consists of the cross-classification of C by P by S on the data with no missing values on C. This is the complete data table and is equivalent to subtable (a) in Table 1. The other subtable is a 2^3 subtable of the incomplete data. This is subtable (b) in Table 1 cross-classified by the unobserved clinical status of these individuals. Thus, this 2^3 table for the incomplete data is latent. The union of the complete data table with the two-way margin for the incomplete data is the observed table, which is simply Table 1.

5. LOGLINEAR–LATENT CLASS MODELS FOR MISSING DATA

Missing data can be incorporated into an analysis by applying loglinear models to the expanded table. Because neither table is directly observed, we require a latent class model. Assuming that observations are obtained under a multinomial sampling scheme, we see that a loglinear model for the expanded table in the infant mortality example is

$$\begin{aligned}\log \pi_{ijkl} = \lambda &+ \lambda_i^M + \lambda_j^C + \lambda_k^P + \lambda_l^S + \lambda_{ij}^{MC} + \lambda_{ik}^{MP} + \lambda_{il}^{MS} \\ &+ \lambda_{jk}^{CP} + \lambda_{jl}^{CS} + \lambda_{kl}^{PS} + \lambda_{ijk}^{MCP} + \lambda_{ijl}^{MCS} + \lambda_{ikl}^{MPS} \\ &+ \lambda_{jkl}^{CPS} + \lambda_{ijkl}^{MCPS},\end{aligned}$$

where π_{ijkl} denotes the probability that an individual falls into the cell for the ith level of M $(i = 0,1)$, the jth level of C $(j = 0,1)$, the kth level P $(k = 0,1)$, and the lth level of the S $(l = 0,1)$. Here, λ is determined by the constraint that $\sum_{ijkl} \pi_{ijkl} = 1$, and the remaining λ's are parameters, some of which may be zero. (For example, λ_{jk}^{CP} denotes the clinic by prenatal care association for a contrast involving the jth clinic and the kth prenatal care categories.) We assume that these parameters satisfy the typical constraint that the sum of all parameters of a particular type is zero.

Let f_{ijkl} represent the observed frequency in the $ijkl$th cell. When this cell is not directly observed, we denote the hypothetical frequency by f^*_{ijkl}. If $i = 0$ when clinic is known and $i = 1$ when clinic is missing, then we observe the collapsed tables $f_{1+kl} = f^*_{10kl} + f^*_{11kl}$, but not the f^*_{10kl} or f^*_{11kl} separately. The log-likelihood function for the model is

$$\log L = \sum_{jkl} f_{0jkl} \log(\pi_{0jkl}) + \sum_{kl} f_{1+kl} \log(\pi_{10kl} + \pi_{11kl}),$$

where the first summation is for cells that are not missing and the second summation is for cells that contain missing data. The terms contained in the first summation follow the form for likelihood terms in a loglinear model (e.g., Agresti, 1990, p. 166), whereas the terms in the second summation follow the form for likelihood terms in a latent class model (Andersen, 1980, p. 260). Our missing data model, therefore, combines elements of loglinear and latent class models.

6. TYPES OF MISSING DATA MODELS

The discussion of missing data patterns in Table 2 showed that estimates of the effect of prenatal care depend critically on assumptions about how the missing data are distributed. Having shown how to include missing data into a loglinear analysis through the use of the expanded table, we can examine specific models.

A. MCAR Models

The most restrictive model assumes that data are missing completely at random (MCAR) (Little and Rubin, 1987). This means that the missing indicators are assumed to be independent of all other variables. Consider the hypothetical data in Table 3 where there are three variables, X, Y, and Z, and there is missing information on Y for some cases. Let M be the missing data indicator for Y. Then the MCAR model for these data is equivalent to (M) (XYZ), where we have used the parentheses to indicate an arbitrary set of associations among X, Y, and Z. The models $(M)(XY)(YZ)$, $(M)(XZ)(Y)$, and $(M)(XYZ)$ are all examples of models in which the data are assumed to be MCAR. In each of these cases, whether data are missing, as indicated by M, is assumed to be independent of the other variables in the model. When data are MCAR, they can be omitted from the analysis without affecting parameter estimates. However, because omitting missing data reduces the sample size, it reduces the precision of estimated parameters involving variables that have no missing data.

B. MAR Models

A less restrictive and usually more realistic assumption is that whether a variable is missing is a function of the values of other observed variables; that is, data are missing at random (MAR), conditional on the observed data (Little and Rubin, 1987). With MAR data, whether a variable is missing for a particular case is random conditional on the observed values on the other variables. In terms of our hypothetical data in Table 3, MAR models are of the form $(MXZ)\ (XYZ)$, where, as before, the associations among the variables inside the parentheses are arbitrary. These models are MAR because M is conditionally independent of the variable it pertains to, Y. Typically, the inclusion of MAR data has very small effects on the estimated relationships between the variable that has missing data and the other variables. However, dropping cases with missing data may affect the estimated associations among the other variables in the model. Thus, incorporating cases with missing data into the analysis may affect both the precision and consistency of the estimates for the relationships among variables that do not have missing data (Winship and Mare, 1989).

C. NINR Models

When the probability that a variable has missing data is associated with the variable itself conditional on the other variables in the model, this requires a model for nonignorable nonresponse, or NINR. NINR models are required when it is likely that some survey respondents refuse to answer a question. For example, persons with unusually high or low incomes may be less willing to divulge their incomes than persons with incomes in the middle of the income distribution. Elsewhere (Winship and Mare, 1989) we examined data on whether individuals had ever been arrested. Here, one would expect that persons who had been arrested would be less likely to answer the question than those who had not been arrested.[4] In our hypothetical data, any model that assumes dependence between M and Y is a NINR model. For example $(MY)\ (XY)\ (YZ)$, $(MY)\ (MX)\ (XY)\ (YZ)$, and $(MYZ)\ (XY)\ (YZ)$ are all NINR models.

When data are in fact subject to nonignorable nonresponse, omitting observations with missing values results in estimates that are not only inefficient but inconsistent as well. Often the bias can be considerable. For example, in Table 2, assumptions 3 through 10 are consistent with a variety of NINR models. Depending on what assumptions one makes about the true values of the missing data, one gets quite different estimates for the effect of Prenatal Care on Survival. Unfortunately, NINR models can often be difficult to estimate.

7. FITTING MISSING DATA MODELS TO INFANT MORTALITY DATA

We can illustrate the fitting of alternative missing data models with the data in Table 1. Table 4 presents the estimated likelihood ratio G^2 and

Table 4. Goodness of Fit of Selected Models for Infant Mortality Data

Model	G^2	df	BIC
Complete Data			
1. *SC PC*	0.08	2	−13.06
MCAR Models			
2. *M S PC*	24.81	7	−23.33
3. *M SP PC*	20.02	5	−14.37
4. *M SC PC*	7.98	5	−26.41
5. *M SC PC SP*	7.84	4	−19.67
MAR Models			
6. *MP S PC*	20.13	5	−14.26
7. *MP PC SP*	15.33	4	−12.18
8. *MP SC PC*	3.30	4	−24.21
9. *MP SC PC SP*	3.16	3	−17.47
10. *MS PC SP*	17.83	4	−9.68
11. *MS SC PC*	5.79	4	−21.72
12. *MS SC PC SP*	5.65	3	−14.98
13. *MS MP PC*	17.94	4	−9.57
14. *MS MP SP PC*	13.55	3	−7.08
15. *MS MP SC PC*	1.57	3	−19.06
16. *MS MP SC PC SP*	1.38	2	−12.37
NINR Models			
17. *MC S PC*	20.13	5	−14.26
18. *MC PC SP*	15.33	4	−12.18
19. *MC SC PC*	2.26	4	−25.25
20. *MC SC PC SP*	2.00	3	−18.63
21. *MC MP S PC*	20.13	4	−7.38
22. *MC MP PC SP*	15.33	3	−5.30
23. *MC MP SC PC*	0.39	3	−20.24
24. *MC MP SC PC SP*	0.39	2	−13.36
25. *MC MS PC*	17.94	4	−9.60
26. *MC MS PC SP*	13.55	3	−7.08
27. *MC MS SC PC*	1.64	3	−18.99
28. *MC MS SC PC SP*	1.40	2	−12.35

Bayesian information criterion (BIC) (Raftery, 1995) statistics for the complete data and selected missing data models. In this analysis, we regard infant mortality as the response variable and clinic and prenatal care as the explanatory variables. Thus, we consider only models that include the association between prenatal care and clinic (PC). The G^2 values here should be viewed with caution. Because several of the frequencies in Table 1 are very small, the G^2 statistics may not follow a X^2 distribution (Agresti, 1990).

For the complete data, model 1, which includes all one-way effects and the two-way associations between Survival Status and Clinic (*SC*), and between Prenatal Care and Clinic (*PC*), fits the data extremely well ($G^2 = 0.08$, df = 1), implying that Prenatal Care and Survival Status are conditionally independent within clinics. The data, therefore, suggest that the level of prenatal care does *not* affect the probability that an infant will survive (Bishop et al., 1975; Little and Rubin, 1987). If we include the missing data, then a much larger class of models is available. Among MCAR models, the (*M*) (*SC*) (*PC*) model (model 4), which is analogous to the best-fitting complete data model, fits the data well ($G^2 = 7.98$, df = 5). This is an adequate fit, assuming that G^2 does in fact follow a X^2 distribution for these data.

Although the fit of the MCAR model is reasonable, it is nonetheless of interest to examine MAR and NINR models for these data. Considering these models as a whole, one can draw a number of general conclusions. First, models that do not contain both the (*SC*) and (*PC*) associations fit the data poorly. Second, for all three types of missing data models, the (*SP*) association appears to be statistically insignificant. Thus, for these data, our analysis is consistent with the analysis of the complete data alone. The data support the assumption that prenatal care and survival status are conditionally independent.

Across the three types of missing data models, however, there are some interesting differences. MAR models fit better than their corresponding MCAR models by the G^2 criterion. Whereas the (*M*) (*SC*) (*PC*) model has a G^2 of 7.98 on 5 degrees of freedom, the two comparable MAR models where *M* is assumed to be a function of Level of Prenatal Care (8) or Survival Status (11) have G^2 of 3.30 and 5.79, respectively (df = 4). Differencing these amounts from the MCAR G^2, we get G^2 statistics of 4.68 and 2.19 (df = 1). Assuming that G^2 follows a X^2 distribution, we see that the first of these differences is significant at the 0.05 level, indicating that a significant improvement in fit is achieved by assuming that *M* is associated with level of prenatal care (*P*). Adding the (*MS*) term to the (*MP*) (*SC*) (*PC*) model lowers the G^2 to 1.57 (Model 15), but this decrement is not statistically significant. The NINR models also fit the data better by the G^2 criterion. A model in which the likelihood of having missing data is simply a function of the clinic, that is, (*MC*) (*SC*) (*PC*), gives a G^2 of 2.26 on 4 degrees of freedom (Model 19). This model has a lower G^2 with the same degrees of freedom as the MAR model (*MP*) (*SC*) (*PC*) (Model 8), although a nested comparison between these two models is obviously impossible. The (*MC*) (*SC*) (*PC*) model implies that the likelihood of missing data on clinic is only associated with which clinic a mother was served by. Adding an (*MS*) term to the model lowers the G^2 to 1.64 (df = 3) (Model 27). Alternatively, adding an (*MP*) term to the (*MC*) (*SC*) (*PC*) model lowers the G^2 to 0.39 (Model 24), an extremely good fit. Neither of the changes, however, is significant at the 0.05 level, assuming a X^2 distribution for G^2. With considerations of fit, parsimony, and plausibility, we conclude that Model 19 is the most satisfactory model for the data.

It is certainly possible to conceive of more complex NINR models. Table 2 illustrated how various assumptions about the pattern of missing data lead to substantial differences in estimates for the effect of Level of Prenatal Care on Survival. Cases 5 through 10 in Table 2 are all equivalent to saturated NINR models, that is, models that fit the data perfectly and that have zero degrees of freedom. Many of these models are also unidentified and, by definition, cannot be tested against the data. All of these models assume that the missing mechanism has a three-way or higher interaction with the clinic and another variable. For example, cases 5 and 6 assume a three-way interaction among *M*, *C*, and *P*. Cases 9 and 10 assume a four-way interaction among *M*, *C*, *P*, and *S*. These models all assume that in the two clinics different types of data are likely to produce missing data on clinic. Without some assumptions we cannot rule out the estimates associated with these allocations. If, however, we assume that the missing data mechanism is the same in the two clinics except for the rate at which missing data occurs, then this rules out these possibilities. The (*MC*) (*SC*) (*PC*) model represents this assumption.

8. EXAMPLE 2: INTERGENERATIONAL EDUCATIONAL MOBILITY

A common use for loglinear models is the analysis of intergenerational social mobility. These analyses typically focus on cross-classifications of parents' and offsprings' socioeconomic characteristics derived from retrospective reports by the offspring. A problem for these analyses is missing data on the parents' characteristics. In this example, we examine the association between father's schooling and offspring's educational attainment, using data from the 1994 General Social Survey (GSS) and the Survey of American Families (SAF). The 1994 GSS, a cross-section survey of the

Table 5. Offspring's Educational Attainment by Father's Educational Attainment

	Offspring's Schooling		
Father's Schooling	**<12**	**12**	**>12**
<12	46	135	113
12	12	75	145
>12	4	30	206
Missing	23	29	18

Source: 1994 General Social Survey, respondents age 18 and older with a sibling interviewed in the Survey of American Families.

U.S. population, included a module on the socioeconomic characteristics of persons related to GSS respondents, including parents, children, spouses, and siblings. The SAF was a telephone survey administered to one randomly selected sibling of the GSS respondents (Mare and Hauser, 1994). Table 5 cross-classifies father's and respondent's schooling as reported by respondents in the GSS. These data were restricted to persons age 18 and older who had a sibling who was interviewed in the SAF. Of the 836 persons included in the table, 70, or 8.4%, failed to report their father's educational attainment. This may be construed as a relatively modest amount of missing data, and some researchers would simply omit observations in which father's schooling is missing. This decision, however, may have a big effect on one's estimates and inferences. One way of seeing the consequences of omitting missing data is to examine the estimated distribution of father's schooling and association between father's and offspring's schooling under several alternative hypothetical patterns of missing data. Table 6 shows the four local odds ratios for the 3 × 3 table of father's schooling by offspring's schooling under four scenarios: (1) data missing completely at random; (2) missing observations on father's schooling are all drawn from respondents whose fathers have less than 12 years of schooling; (3) missing observations are all drawn from respondents whose fathers have more than 12 years of schooling; and (4) missing observations are all drawn from the same level of schooling as the respondent. Under these alternative assumptions, the proportions of persons whose fathers have less than 12 years of schooling ranges from approximately 33% to approximately 40%. The estimated local odds ratios under alternative assumptions about missing data vary substantially. For example, the local cross-product ratio between whether a father is a high school graduate versus a dropout and the corresponding contrast

Table 6. Local Odds Ratios and Distributions of Father's Schooling Under Alternative Assumptions about Missing Data

	Local Odds Ratios		Distribution of Father's Schooling		
	<12/12	**12/>12**	**<12**	**12**	**>12**
1.		Observed Data			
<12/12	2.13	2.31			
			0.384	0.303	0.313
12/>12	1.20	3.55			
2.		Missing Data All Taken from Diagonal			
<12/12	4.43	1.67			
			0.354	0.293	0.352
12/>12	0.87	5.36			
3.		Missing Data All Taken from <12			
<12/12	2.62	2.42			
			0.409	0.257	0.332
12/>12	1.20	3.55			
4.		Missing Data All Taken from >12			
<12/12	2.13	2.31			
			0.328	0.257	0.413
12/>12	0.35	1.96			

Source: Table 5.

for offspring varies between 2.13 for data missing at random and 4.43 for data missing exclusively from persons who have the same schooling level as their fathers.

We can examine the association between father's and offspring's schooling while taking account of missing data by using variants of the missing data models discussed earlier. Although data may be missing completely at random, it is more likely that whether data are missing on father's educational attainment is associated with a respondent's own educational attainment and possibly father's educational attainment itself. The former problem may arise because better educated respondents are more likely to be cooperative and conscientious in answering survey questions. The latter problem may arise if respondents perceive that it is desirable to have a better educated father or if individuals with more poorly educated fathers are less likely to know their father's schooling. Some offspring of fathers with low levels of educational attainment may misreport their father's schooling, but others may simply not report it.

By itself, Table 5 provides limited information with which to investigate the impact of missing data on our estimates of the distribution of father's education or of the association between father's and offspring's education. That better educated respondents may be more likely to report their father's schooling can be investigated with this table. If this is the only systematic source of missing data, and if father's and offspring's educational attainments are associated, this idea can be represented as a saturated MAR model for this table. However, the idea that whether data are missing on father's schooling is associated with father's schooling itself requires a model of nonignorable nonresponse (NINR), which is not identified from these data. To investigate nonignorable nonresponse requires a variable that is associated with father's schooling but not with whether father's schooling is missing. Such variables are difficult to find because most characteristics of an individual that are associated with father's schooling are also associated with the individual's propensity to report father's schooling. A solution to this problem is to use the responses to the same item in an independent interview conducted with a person related to the original respondent. The SAF asked a sibling of each GSS respondent to report on father's schooling. Table 7 cross-classifies GSS respondent's report of father's schooling, GSS respondent's report of his or her own schooling, and the SAF respondent's – that is, GSS respondent's sibling's – report of father's schooling. This table includes categories for missing data on both reports of father's educational attainment and can be used to examine a variety of models for missing data.

These models can be regarded as applying to an *expanded table* that includes separate dimensions for the substantive variables of interest and for whether these variables are missing. The $3 \times 3 \times 3 \times 2 \times 2$ expanded table has the following five dimensions: GSS respondent's educational attainment (O), GSS respondent's report of father's educational attainment (F_G), SAF respondent's report of father's educational attainment (F_S), whether F_G is missing (M_G), and whether F_S is missing (M_S). For example, from Table 7 we can identify GSS respondents who have missing data on father's schooling classified by their own schooling and their siblings' reports of father's schooling. We do not know the educational attainment of these individuals' fathers (although a large fraction of these individuals have the same father as their sibling and, for them, their father's education can be inferred). Thus, for a given level of own schooling and sibling's report of father's schooling, GSS respondents who have missing data on their own report of father's schooling are distributed in an unknown way across categories of their father's schooling. Our models for missing data are based on selected interactions among the five dimensions of this expanded table.

Table 7. GSS Respondent's Educational Attainment by Father's Educational Attainment Reported by GSS Respondent by Father's Educational Attainment Reported by SAF Respondent

	Father's Schooling (SAF)											
	<12 Offspring's Schooling			12 Offspring's Schooling			>12 Offspring's Schooling			Missing Offspring's Schooling		
Father's Schooling (GSS)	<12	12	>12	<12	12	>12	<12	12	>12	<12	12	>12
<12	33	104	88	1	11	13	0	2	2	12	18	10
12	2	11	16	7	49	109	1	13	16	2	2	4
>12	0	0	3	1	9	23	3	21	177	0	0	3
Missing	12	17	5	4	6	6	0	3	2	7	3	5

Many logically possible models may be fit to the frequencies in Table 7. We limit the range of possible models through the following substantive considerations. First, because of the well-known correlation between the socioeconomic positions of parents and offspring, father's and offspring's educational attainment are associated. Indeed, this association provides the substantive interest in this table. Inasmuch as the GSS and SAF respondents' reports of father's educational attainment (F_G and F_S) apply to the same individual in most families, they are two reports of the same trait and thus both of these measures are associated with offspring's schooling. Thus, all models should include the F_GO and the F_SO associations.[5] Second, inasmuch as most siblings share the same father, their reports of father's schooling are likely to be strongly associated. Thus, all of our models include the F_GF_S association. Third, siblings' propensities to fail to report father's schooling may be associated, either because of a shared reluctance to provide this information or a shared ignorance of their father's schooling. For most models, therefore, we include the M_GM_S association. Fourth, it is an empirical question whether data are missing on father's schooling is associated with respondent's own educational attainment and with father's schooling itself. Thus, we examine alternative models with and without the OM_G, OM_S, F_GM_G, and F_SM_S associations. Finally, we assume that whether a person reports father's schooling is *conditionally* independent of his or her sibling's reported level of father's schooling, given the association between each sibling's reported level of father's schooling. Thus, we assume the absence of the F_GM_S and the F_SM_G associations. These are the key restrictions for identifying NINR models for these data.

Table 8 presents goodness-of-fit statistics for selected models fit to the observed data in Table 7. Model 1 is an MCAR model in that it

Table 8. Goodness of Fit of Selected Models for Educational Mobility Table

Model	G^2	df	BIC
1. F_GF_S, F_GO, F_SO, M_G, M_S (MCAR)	1,470.6	27	1,288.9
2. F_GF_S, F_GO, F_SO, M_GM_S (MAR)	117.9	26	−57.0
3. F_GF_S, F_GO, F_SO, OM_G, M_GM_S (MAR)	73.6	24	−87.9
4. F_GF_S, F_GO, F_SO, OM_S, M_GM_S (MAR)	87.1	24	−74.4
5. F_GF_S, F_GO, F_SO, OM_G, OM_S, M_GM_S (MAR)	51.5	22	−96.5
6. F_GF_S, F_GO, F_SO, F_GM_G, F_SM_S, M_GM_S (NINR)	48.0	22	−100.0
7. F_GF_S, F_GO, F_SO, OM_G, OM_S, F_GM_G, M_GM_S (NINR)	46.2	20	−88.4
8. F_GF_S, F_GO, F_SO, OM_G, OM_S, F_SM_S, M_GM_S (NINR)	28.5	20	−106.1
9. F_GF_S, F_GO, F_SO, OM_G, OM_S, F_GM_G, F_SM_S, M_GM_S (NINR)	20.3	18	−100.8

assumes conditional independence of whether data are missing on the two measures of father's schooling from any of the other dimensions of the table. This model includes parameters for the marginal distributions of whether data are missing on F_G and F_S (M_G and M_S, respectively), but no association between whether data are missing on these two variables. As indicated by both the likelihood ratio G^2 and the BIC statistics (Raftery, 1995), this model fits very poorly.

Models 2–5 are MAR models. Model 2 includes a parameter for the association between M_G and M_S and fits the data much better than Model 1. This suggests that GSS respondents and their siblings both fail to report their father's schooling at a much higher rate than one would expect if their rates of nonresponse were statistically independent. Common family circumstances may determine whether offspring know their father's educational attainment. Models 3, 4, and 5 incorporate parameters for the association between GSS respondent's schooling and whether GSS and SAF respondents' have missing data on father's schooling. Inclusion of both of these associations significantly improves the fit of the model. The OM_G association implies that better educated respondents differ from more poorly educated respondents in their level of cooperation with the survey or their knowledge of their father's schooling. The OM_S association may arise because the table does not include a dimension for SAF respondent's own educational attainment. Given a strong correlation between siblings' educational attainments, we observe an OM_S association when the schooling of SAF respondents is not taken into account.

Models 6–9 are NINR models that include associations between F_G and F_S on the one hand and M_G and M_S on the other. Model 6 includes the F_GM_G and F_SM_S associations, but it excludes the associations between M_G and M_S and GSS respondent's educational attainment (O). This model fits much better than Model 2, the corresponding MAR model ($G_2{}^2 - G_6{}^2 = 69.9$, 4 df, $p < .001$), and it provides provisional evidence of nonignorable nonresponse. Model 6, however, does not fit the data well by the G^2 criterion. A more stringent test for NINR is to estimate the F_GM_G and F_SM_S associations in the presence of the OM_G and OM_S associations. Models 7 and 8 include the F_GM_G and F_SM_S associations, respectively, as well as OM_G and OM_S. Model 7 fits marginally better than Model 5, the corresponding MAR model ($G_5^2 - G_7^2 = 5.3$, 2 df; $p = .071$), whereas Model 8 fits much better than Model 5 ($G_5^2 - G_8^2 = 23.0$, 2 df, $p < .001$). This provides strong evidence for NINR for F_S and somewhat weaker evidence for F_G. Model 9 includes the OM_G, OM_S, F_GM_G, and F_SM_S associations and fits significantly better than Model 5 as well as the three

Table 9. Estimated Association Parameters for NINR Model (Model 9)

Model Terms	λ	SE(λ)	exp(λ)
$M_G M_S$	0.670	0.358	1.954
$F_{G12} F_{S12}$	3.868	0.294	47.847
$F_{G12} F_{S>12}$	3.967	0.570	52.826
$F_{G>12} F_{S12}$	4.476	0.646	87.882
$F_{G>12} F_{S>12}$	8.056	0.778	3,152.654
$F_{G12} O_{12}$	0.075	0.501	1.078
$F_{G12} O_{>12}$	0.704	0.488	2.022
$F_{G>12} O_{12}$	−0.285	0.876	0.752
$F_{G>12} O_{>12}$	1.562	0.836	4.768
$F_{S12} O_{12}$	0.645	0.514	1.906
$F_{S12} O_{>12}$	0.917	0.504	2.502
$F_{S>12} O_{12}$	1.325	0.858	3.762
$F_{S>12} O_{>12}$	1.606	0.831	4.983
$F_{G12} M_G$	−0.197	0.462	0.821
$F_{G>12} M_G$	−2.292	1.469	0.101
$F_{S12} M_S$	−2.156	1.233	0.116
$F_{S>12} M_S$	−2.672	0.951	0.069
$O_{12} M_G$	−0.924	0.326	0.397
$O_{>12} M_G$	−1.620	0.281	0.198
$O_{12} M_S$	−0.883	0.351	0.414
$O_{>12} M_S$	−0.865	0.378	0.421

other NINR models. By the likelihood ratio criterion, Model 9 is the best fitting model ($G_9^2 = 20.3$, 18 df, $p = .316$), although the BIC indicates that Model 8 is slightly more satisfactory.

Table 9 presents estimates of two-way association parameters and partial odds ratios from Model 9, which reveal the systematic nature of missing data on father's educational attainment. First, the estimated partial odds ratio for the two missing data indicators ($M_G M_S$) indicates a positive association between siblings' propensities to fail to report father's educational attainment. The SAF respondent is almost twice as likely to have missing data on father's schooling if his or her sibling in the GSS fails to report father's schooling than if the sibling reports father's schooling. Second, more highly educated GSS respondents are much more likely to report their father's schooling than their more poorly educated counterparts. Relative to GSS respondents who were high school dropouts, for

example, those who had more than a high school degree had odds of not reporting father's schooling that are only one fifth as great. Finally, missing data on father's educational attainment is associated with the level of father's education. For example, for SAF respondents whose fathers had more than a high school degree, the odds of missing data are only approximately 7% of the odds for SAF respondents whose fathers were high school dropouts. Both the parameter estimates and the goodness-of-fit tests provide more evidence of nonignorable nonresponse for SAF respondents' reports of fathers' schooling than for GSS respondents' reports. This may occur because GSS respondents' own educational attainments are controlled in our models, whereas SAF resondents' attainments are not.

9. ESTIMATION AND IDENTIFICATION

A. Estimation

The estimates reported in this paper were calculated by using DNEWTON, a flexible program for the estimation of latent class models, including models with latent cells such as the ones discussed in this paper (Haberman, 1989). Further discussion of DNEWTON is provided in the Technical Appendix. A good alternative to DNEWTON is LEM (Vermunt, 1996), which has similar capabilities. The development of several user-friendly latent class programs has substantially reduced the burden of estimating latent class models. Although modern software makes the estimation of loglinear models for missing data feasible, several estimation problems nonetheless are common. These include (1) failure of the program to converge; (2) estimates on the boundary of the parameter space; (3) cells with small expected frequencies (< 5), and (4) poor model fit. These important issues, which are problems for latent class models more generally, are discussed in the introductory chapter of this volume.

B. Identification

Consider the hypothetical data shown in Table 3 consisting of variables X, Y, and Z but collapsed over Z. If we add to this a variable M indicating whether information on Y is missing for a case, we then have a 2^3 table of X by Y by M. This is shown in Table 10 both as a 2×3 subtable of the observed table and as a 2^3 expanded table that is partially latent. In a complete eight-cell table, a fully saturated model has seven parameters

Table 10. Hypothetical Data from Table 3 Collapsed Over Z

(a) Observed Table

	$Y=0$	$Y=1$	$Y=$ Missing
$X=0$	960	480	360
$X=1$	480	960	360

(b) Expanded Table

	$M=0$		$M=1$		
	$Y=0$	$Y=1$	$Y=0$	$Y=1$	Total
$X=0$	960	480	?	?	360
$X=1$	480	960	?	?	360

(plus a grand mean). However, Table 10 has only six cells – four for the X–Y complete data subtable and two for the categories of X among respondents with missing data on Y. Therefore, a model with at most five parameters is identified. If one fits a marginal parameter for each of the dimensions X, Y, and M, then, among hierarchical models, the most complex five-parameter model can include at most two two-way interactions. The potential models are as follows: (1) (MX) (XY) (MAR Model); (2) (MY) (XY) (NINR Model); and (3) (MX) (MY), which is not identified.

To see that the (MX) (MY) model is not identified, note that in panel (b) of Table 10 one can observe the association between X and Y for the complete data. The (MX) (MY) model assumes that the partial X–Y association (conditional on M) is zero, which can be tested by using the complete data. That we can test that the X–Y interaction is zero means that a parameter for this association is identified, irrespective of whether other parameters of the (MX) (MY) model are not identified. Such a test would use 1 degree of freedom, leaving only 4 degrees of freedom for estimating the five parameters of the (MX) (MY) model. Thus the (MX) (MY) model cannot be identified.

MAR models are usually identified, even if data are missing on several variables (Little and Rubin, 1987, pp. 171–94). A sufficient condition for identification of MAR models is that some observations are complete on all variables and that these observations represent all possible combinations of the variables in the model (Fuchs, 1982).

General rules for the identification of NINR models have not yet been developed. Some guidance is available, however, from results on the identifiability of NINR models for two-way tables and from rules for identification of latent class models. Little and Rubin (1987, pp. 238–9) summarize the identifiability of models for two-way ($J \times K$) tables in which one dimension of the table has missing observations. Denote the dimensions of the table by X and Y. Let all observations be present for X, but some observations on Y be missing. Let a third variable, M, denote whether data are missing on Y. Among the several loglinear models that can be fit to the X, Y, M table, NINR models are those that include the M–Y interaction. The only NINR model that is potentially identifiable is (MY) (XY), that is, a model in which the fully observed variable is associated with the partially observed variable, but is independent of whether data are missing on the partially observed variable. This model is identified if $J \geq K$, that is, if the number of categories of the fully observed variable is at least as large as the number of categories of the partially observed variable. In a 2×2 table, $J = K$ and the model is just identified.

These results suggest that NINR models are identified if (1) for every partially observed variable, there exists a fully observed variable that is conditionally independent of the missing data indicator; and (2) the number of categories of the partially observed variable does not exceed the number of categories of the fully observed variable. The fully observed variable plays a role in identifying the model that is analogous to an instrumental variable in a structural equation model, and the condition is equivalent to assuming that parameter for the two-way interaction between the fully observed variable and the missing indicator is zero. This restriction can sometimes also be met by assuming that higher-order interaction terms are zero. Baker and Laird (1988), for example, estimate a model in which two fully observed variables both affect the missing indicator, but the model is identified because the three-way interaction among the variables is assumed to be zero.

An intuitive way of understanding the identifiability of some NINR models is to note that they are often similar to standard latent class models that are analogous to factor models (e.g., Goodman, 1974). Consider the NINR (MC) (SC) (PC) model for the Little–Rubin data. If C were missing on *every* observation, the model would be a standard latent class model in which C is a latent "factor" and M, S, and P are its observed "indicators." The model assumes that M, S, and P are conditionally independent given C. If C were missing on every case, however, the model would not be identified without either another indicator for C or additional restrictions on the parameters. In the pure latent variable case, we observe only the two-way relationship between P and S, which is insufficient for estimating the associations between both S and C and also P and C.

But when data are missing for only some of the cases, we observe the relationship between S, P, and C in the complete data. Thus we can estimate both the S–C and P–C associations simply from the complete data. Only the relationship between M and C has to be estimated indirectly. When C is partially observed, therefore, this NINR model is identified.

10. CONCLUSION

Conventional methods of dealing with missing data in multivariate models run serious risks. Omitting observations with missing data from an analysis certainly reduces sample size (and thus the precision of estimates) and, at worst, may lead to severe biases in parameter estimates. Simply incorporating missing data by adding categories for missing data to the observed variables generally results in inconsistent estimates. Fortunately, alternatives to the conventional approaches provide powerful methods for investigating the degree to which data are missing systematically and for carrying out appropriate substantive analyses.

The approach illustrated in this chapter[6] is to extend standard analyses of categorical data by recognizing that when some data are missing, the resulting contingency tables have cells that are partially observed. We suggest that one analyze such data by using latent class loglinear models for tables that have a mixture of fully and partially observed cells. This approach enables one to investigate hypotheses about the mechanisms by which missing data come about as well as the substantive relationships of interest. Inasmuch as these models are simply variants of standard loglinear models, one can incorporate missing data by using well-known procedures for specifying multiway interactions in contingency tables. Unlike standard loglinear models for fully observed data, however, these models typically require that the analyst incorporate additional data or make simplifying assumptions to identify the relationships between substantive variables of interest and indicators of whether data are missing.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. New York: J Wiley.
Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland.
Baker, S. G., & Laird, N. M. (1988). "Regression analysis for categorical variables with outcome subject to nonignorable nonresponse," *Journal of the American Statistical Association*, **83**, 62–9.

Bishop, Y. M., Fienberg, S., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.
Clogg, C. C. (1988). "Latent class models for measuring." In R. Langeheine and J. Rost (eds.), *Latent Trait and Latent Class Models*. New York: Plenum, pp. 173–205.
Davis, J. A., & Smith, T. W. (1986). *General Social Surveys, 1972–1986* (MRDF), NORC ed. Chicago: NORC; distributed by Roper Public Opinion Research Center, New Haven, CN.
Dayton, C. M., & Macready, G. B. (1980). "A scaling model with response errors and intrinsically unscalable respondents," *Psychometrika*, **45**, 343–56.
Fuchs, C. (1982). "Maximum likelihood estimation and model selection in contingency tables with missing data," *Journal of the American Statistical Association*, **77**: 270–8.
Goodman, L. A. (1974). "The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: a modified path analysis approach," *American Journal of Sociology*, **79**, 1179–1259.
Haberman, S. J. (1989). "A stabilized Newton–Raphson algorithm for log-linear models for frequency tables derived by indirect observation." In C. C. Clogg (ed.), *Sociological Methodology 1988*, Vol. 18. **City: pub**, pp. 193–211.
Hagenaars, J. (1988). "Latent structure models with direct effects between indicators," *Sociological Methods and Research*, **16**, 379–405.
Hagenaars, J. (1993). *Loglinear Models with Latent Variables*. Newbury Park, CA: Sage.
Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
Mare, R. D., and Hauser, R. M. (1994). "A Survey Module on Families and Social Mobility," unpublished grant proposal to National Science Foundation. Madison, WI: Dept. of Sociology, University of Wisconsin.
McCutcheon, A. L. (1987). *Latent Class Analysis*. Quantitative Applications in the Social Sciences Paper No. 64. Beverly Hills, CA: Sage.
Park, Taesung, & Brown, M. B. (1994). "Models for categorical data with non-ignorable nonresponse," *Journal of the American Statistical Association*, **89**, 44–52.
Raftery, A. E. (1995). "Bayesian model selection in social research." *Sociological Methodology 1995*, Vol. 25, 111–63.
Vermunt, J. K. (1996). *LogLinear Event History Analysis*. Tilburg, The Netherlands: Tilburg University Press.
Winship, C., & Mare, R. D. (1989). "Loglinear Models with Missing Data: A Latent Class Approach." *Sociological Methodology 1989*, Vol. 19, 331–68.
Winship, C., & Mare, R. D. (1992). "Models for sample selection bias," *Annual Review of Sociology*, **18**, 327–50.

NOTES

1. Fuchs (1982), Little and Rubin (1987), Baker and Laird (1988), Winship and Mare (1989), and Park and Brown (1994) provide a technical discussion of

how missing data can be incorporated into loglinear models. In this chapter, we discuss the key ideas needed for the researcher to incorporate missing data in a loglinear analysis.

2. These programs can estimate loglinear models with arbitary patterns of missing data. As a result, we do not need to consider the issues found in earlier literature as to whether the missing data pattern is monotone and/or whether the likelihood can be factored (Little and Rubin, 1987).
3. Winship and Mare (1989) provide a more extensive analysis of this example. In particular, we show that the estimated effect for the association between X and Z is biased when we add a missing data category to Table 3.
4. Econometric models for sample selection bias are examples of NINR models (Winship and Mare, 1992).
5. A related set of models explicitly distinguishes between the "true" father's schooling and the fallible reports of father's schooling provided by each sibling. This distinction can be incorporated into latent class models. We do not consider these models in this chapter.
6. This research was supported by National Science Foundation Grants SBR-94-11875 and SBR-94-11670, and by the Graduate School of the University of Wisconsin-Madison.

Appendix A: Notational Conventions

The purpose of this appendix is to indicate the similarities among what might, at first reading, appear as highly disparate forms of notation used by the various contributors to this volume. The following discussion will be most accessible to those who are familiar with the basic latent class model. Thus, it is recommended that readers familiarize themselves with the notational styles in the two introductory chapters, Chapters 1 and 2, before reading through this appendix.

As with nearly all statistical models, the latent class model can be expressed by using a variety of different notational conventions. The problem of differing notation is exacerbated with the latent class model because the LCM can be parameterized in two seemingly different, though equivalent, parameterizations: probabilistic and loglinear (see McCutcheon, Chapter 2).

The probabilistic parameterization – the most commonly used parameterization – was first introduced to a broader audience by Goodman, who used the iterative proportional fitting estimation method (1974a, 1974b). Goodman adopted a horizontal "overbar" notational style to represent conditional probabilities; thus, in this notational style, the basic latent class model is represented as

$$\pi_{ijkl}^{ABCD} = \sum_{t=1}^{T} \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X},$$

where the symbol π_t^X represents the latent class, or mixing, proportion $P(X = t)$. This is a relatively widespread and common representation usage, although some authors occasionally elect to use some other character than X to represent the latent variable, and some choose to use some other letter than $t = 1, \ldots, T$ to represent the specific classes of the latent