Chapter 23

Fitting a Latent Trait Model for Missing Observations to Racial Prejudice Data

Martin Knott and Panagiota Tzamourani

London School of Economics and Political Science

1. Introduction

This paper will show the use of a latent variable model for binary data which allows information on the latent variable to be extracted from the pattern of the missing data. We use a modified version of the logit-probit model for binary data (Bartholomew, 1987) proposed by Knott, Albanese and Galbraith (1990). The model will be applied to data on racial prejudice from the 1991 British Social Attitudes Survey (Brook, Prior and Taylor, 1992).

2. Model for Binary Data with Missing Values

Suppose that $X_1,\ldots,X_k$ are k binary variables (items) taking values 0 and 1. In addition we shall allow $x_i = 9$ to denote a missing value for the i'th item. This might be „no answer recorded“ or „no-opinion expressed“.

Let x_{vi} be the value of the i'th item X_i for the v'th individual, $v = 1,\ldots,N$. The row vector $\mathbf{x}_v = (x_{v1},\ldots,x_{vk})$ is called the *response pattern* for the v'th individual.

In what follows we assume a two-dimensional common factor $\theta = (\theta_a,\theta_e)$, where θ_a measures attitude and θ_e measures the tendency to express an opinion. θ_a and θ_e have independent N(0,1) distributions. The assumption of independence is necessary to keep clear the distinction between attitude and expression and allows a simple understanding of the model. The posterior distributions θ_a and θ_e though may be correlated. Bock and Aitkin (1981) and Bartholomew (1988) have shown that in one-factor models the form of the distribution of θ does not affect much the results of the analysis and the N(0,1) is a convenient choice.

We shall further assume that individuals behave independently and that the choices made by an individual to respond with approval, disapproval or neither of these are conditionally independent between items given the individual's value of θ. Conditional independence implies that the latent vector θ is sufficient to explain all the association between the X_i's for an individual. Then the joint conditional probability $g(x|\theta)$ is

$$g\left(x|\theta\right) = \prod_{i=1}^{k} g_i\left(x_i\middle|\theta\right), \tag{1}$$

where $g_i(x_i|\theta)$ is the conditional probability of response x_i for the i'th item.

We break down the modelling of the response function into two layers. For each item,

$$p\left(X_i = 1 \middle| \theta, X_i \neq 9\right) = \pi_{ai}\left(\theta_a\right) \tag{2}$$

and

$$p\left(X_i \neq 9 \middle| \theta\right) = \pi_{ei}\left(\theta_a, \theta_e\right).$$

It follows that

$$p\left(X_i = 1 \middle| \theta\right) = \pi_{ai}\left(\theta_a\right)\pi_{ei}\left(\theta\right) \tag{3}$$

$$p\left(X_i = 0 \middle| \theta\right) = \left(1 - \pi_{ai}\left(\theta_a\right)\right)\pi_{ei}\left(\theta\right)$$

and

$$p\left(X_i = 9 \middle| \theta\right) = 1 - \pi_{ei}\left(\theta\right) .$$

The probability of a missing value for item i, $X_i = 9$, is allowed to depend on both factors θ_a and θ_e. By allowing the underlying attitude θ_a to affect the probability of expression of attitude one may hope to recover information about attitudes from the pattern of non-expression of opinion.

We define such a model, which allows information to be gleaned from the pattern of missing values, as

$$\operatorname{logit}\left(\pi_{ai}\left(\theta\right)\right) = \sigma_i^a + \beta_{ai}^a \theta_a \tag{4}$$

$$\operatorname{logit}\left(\pi_{ei}\left(\theta\right)\right) = \sigma_i^e + \beta_{ai}^e \theta_a + \beta_{ei}^e \theta_e$$

$$\text{for } i = 1,2,\ldots k .$$

In order to fit the model we create k „pseudo-items“ which take the value of 1 if X_i is 1 or 0 and the value of 0 if X_i is 9. The model is then fitted to 2k items with the constraint that $\beta_{ei}^e = 0$ for the „original“ k items. The model is a special case of the Bock and Aitkin (1981) multidimensional model, for two dimensions, where some of the parameters are constrained.

The parameter β_{ai}^a is called a discrimination parameter because as a coefficient of θ_a its size determines the effect which a given change in θ_a has on the response function $\pi_{ai}(\theta_a)$. For two individuals a given distance apart on the θ_a scale, the bigger the absolute value of β_{ai}^a, the greater the difference in their probabilities of giving a positive response, and thus the easier it is to discriminate between them on the evidence of their responses to item i.

The parameter σ_i^a gives the probability, π_i, of a positive response to item i by a median individual, that is,

$$\pi_i = p\left(X_i = 1 \middle| \theta_a = 0\right) = \frac{\exp\left(\sigma_i^a\right)}{1 + \exp\left(\sigma_i^a\right)} . \tag{5}$$

If, for example, σ_i^a is large and positive (negative), the probability of a positive (negative) response for a median individual will be close to 1.

The coefficient β_{ai}^{e} gives information on how the attitude affects the probability of expressing an opinion. If β_{ai}^{e} is positive, then an individual who approves item i is more likely to express an opinion. The model thus allows the user to obtain not only information on attitude available from the 0's and 1's in the response patterns, but also any information about attitudes available from the pattern of missing values.

The coefficient β_{ei}^{e} is the discrimination parameter of the item for the latent factor θ_e. For two individuals a given distance apart on the θ_e scale, the bigger the absolute value of β_{ei}^{e}, the greater the difference in their probabilities of giving a response.

The interpretation of these parameters will be clarified with the application given in section 6.

3. The computer program TWOMISS

The TWOMISS program fits the above model using a modified E-M algorithm (Bock and Aitkin, 1981) proposed by Bartholomew (1987, chapter 6).

TWOMISS can also fit a simpler model where the $\pi_{ai}(\theta)$ depends only on θ_a and $\pi_{ei}(\theta)$ depends only on θ_e. There is no link between attitude and non-response but a dependence of non-response between items is modelled.

The assumed N(0,1) distributions of θ_a and θ_e are approximated by Gauss-Hermite quadrature points and their corresponding normalized weights (Straud and Sechrest, 1966). The program allows the number of points to be chosen; usually we choose them to be 16. Input to the program is rather complicated but completely described in the manual (Albanese and Knott, 1992).

4. Goodness-of Fit

If the sample size N is large compared with the number of possible response patterns a chi-squared (χ^2) or log-likelihood ratio (G^2) goodness-of-fit test can be carried out on the observed frequencies of the response patterns.

Even with large N there are often many small expected frequencies so that pooling becomes necessary. If s_0 is the number of frequencies after pooling the degrees of freedom are (s_0 - the number of unconstrained parameters - 1). Often there may be no degrees of freedom left to judge the goodness-of-fit.

One may also look at the percentage of G^2 explained by the model, i.e.,

$$\% \, G^2 = \left(L_1 - L_0\right) / L_0$$

where L_1 is the loglikelihood at the model and L_0 is the loglikelihood at the saturated model. This may be particularly useful when measuring the improvement to the fit due to an additional latent variable.

Formal tests are hard to justify, but if k is not too large, the goodness-of-fit of the model may be judged by comparing the observed and expected frequencies of the response patterns.

It is often informative to compare the observed and fitted values of the one-, two- and three-way marginal frequencies.

5. Measurement of the Latent Variable

One reason to fit a latent variable model is to obtain a score of the latent variable for each respondent. Different methods of scoring each response pattern, and hence each individual, have been proposed. One way of scoring is to use the estimated *posterior mean* of each latent variable given the response pattern **x**. For the model in section 2 the posterior means are more suitable as measures of the latent variables than the component score (Bartholomew 1984), since these take account of non-response.

6. Data on Racial Prejudice

The items analysed here are the following questions from the British Social Attitudes Survey 1991 (Brook, Prior, and Taylor, 1992).

1. Thinking of black people - that is, people whose families were originally from West Indies or Africa - who now live in Britain. Do you think there is a lot of prejudice against them in Britain nowadays, a little, or hardly any?
2. Do you think most white people in Britain would mind or not mind if a suitably qualified person of Asian origin were appointed as their boss? If „would mind“, a lot or a little?
3. And you personally? Would you mind or not? If „would mind“, a lot or a little?
4. Do you think that most white people in Britain would mind or not mind if one of their close relatives were to marry a person of Asian origin? If „would mind“, a lot or a little?

For all questions we selected only the cases answered by those classified as „white“ - 1408 respondents. The last three questions were given to a randomly selected half of the sample of „whites“. The other half was given the same questions but referring to „black“ or „West Indian“ people instead of „Asians“. For the purpose of this analysis we added the two halves. For question 1, „a lot“ was coded as 1, „a little“ and „hardly any“ were coded as 0, and „don′t know“ and „not answered“ were coded as 9. For the other three questions „mind a lot“ was coded as 1, „mind a little“ and „not mind“ were coded as 0, and „don't know“ and „not answered“ were coded as 9. So 1 records the existence of racial prejudice, 0 absence or a small presence, and 9 denotes the „missings“.

By applying the model described in section 2 to these questions we are aiming to fit the observed marginal frequencies by the means of two factors, one which we hope will measure an attitude towards „non-white“ people or racial prejudice and another which will measure a tendency to express an opinion.

Part of the output from TWOMISS appears in Tables 1 and 2.

Item	σ_i^a	*s.d.*	β_{ai}^a	*s.d.*	β_{ai}^{a*}	π_i
1	0.06	0.06	0.37	0.08	0.35	0.51
2	-4.50	1.59	4.87	1.82	0.98	0.01
3	-5.13	0.65	2.62	0.47	0.93	0.01
4	-0.65	0.09	1.39	0.17	0.81	0.34

Table 1: Parameter estimates relative to attitude

Item	σ_i^a	*s.d.*	β_{ai}^e	*s.d.*	β_{ei}^e	*s.d.*	β_{ai}^{e*}	β_{ei}^{e*}
1	4.66	0.38	0.02	0.29	1.47	0.28	0.01	0.83
2	5.96	1.11	0.28	0.45	3.05	0.77	0.09	0.95
3	6.34	0.85	-0.37	0.43	2.52	0.51	-0.14	0.92
4	7.67	3.00	1.25	0.71	4.21	1.97	0.28	0.93

Table 2: Parameter estimates relative to non-response

The σ_a^i coefficients for $i = 1, \ldots, 4$ determine the probability of a 1 as the response from a median individual. The β_{ai}^a further indicate which items discriminate better between individuals with a similar position on the underlying attitude scale. We assume that the higher the position of an individual on the attitude variable the more likely he is to respond positively, i.e., 1, to each item. Thus the β_{ai}^a 's must all be positive. The β_{ai}^a coefficients can also be thought of as factor loadings, especially when their standardised form is considered:

$$\beta_{ai}^{a*} = \beta_{ai}^{a} \big/ A \tag{6}$$

where

$$A = \left(1 + \beta_{ai}^{a^2}\right)^{1/2} .$$

We also use a similar standardisation for β_{ai}^e and β_{ei}^e.

In Table 1 the discrimination parameter estimates range from 0.37 to 4.87. Standard errors (s.d.) are given in the table, but may not be well estimated. Notice that β_{a1}^a is close to zero indicating that this item does not „load“ as high on the underlying factor as the other items do. The first item does not seem to measure racial prejudice. This can be explained by the fact that the question asked for item 1 is too abstract, it does not quantify the terms „a lot“ or „a little“ prejudice, whereas the other items are quite specific, they refer to particular circumstances. Since σ_1^a is almost zero the median individual has a close to 50% probability to say that there is prejudice in Britain against black people and about the same probability of saying the opposite. It seems that people may not understand the question, because it is so vague, and they may be answering almost at random.

Item 2 seems the most discriminating item since it has the highest β_a^a coefficient. For items 2,3,4 σ_i^a are negative corresponding to the low probability for a positive response for the median individual, i.e., the median individual has a small probability of saying that people would have something against a „non-white“ person.

The coefficients β^{e}_{ai} given in Table 2 show the dependence of the probability of expressing an opinion (scored 0 or 1) on the attitude latent variable. The positive sign of the coefficients for items 1 and 2 suggest that people placed on the „positive“ side on the underlying attitude variable would be likely to respond. But the coefficients are very small and compared to their standard errors so they are not very informative. The larger coefficient of item 4 suggests firmly that people who would answer positively to this item would be more likely to respond. The negative sign of the coefficient of the third item though suggests that a person who would be placed on the „positive“ side of the underlying attitude and thus responding 1 to item 3 would probably not express an opinion. As item 3 is the only item that refers to one's own feelings we could infer that people would not admit that they are prejudiced themselves but would more easily say that society or other people are. But since the coefficient is also close to zero this argument should not be taken too far.

The β^{e}_{ei}'s, as the coefficients of the latent „tendency to respond“ factor, show the importance of each item in discriminating individuals along this factor. For example, a 9 for item 4 will push an individual further down on the $\theta_e|X$ scale than a 9 for item 3. (see Tables 5 and 6).

The loglikelihood ratio statistic $G^2 = 40.25$ and the statistic $\chi^2 = 29.75$ on 3 degrees of freedom does not appear to indicate a good fit. The degrees of freedom are obtained by subtracting from the number of frequencies after pooling (24) the number of unconstrained parameters (20) and 1. The total number of possible observed patterns is 3^4, i.e., 81. However, since the 4-dimensional table is sparse the goodness-of-fit statistics are not reliable.

i	*j*	*obs*	*exp*	*obs-exp*	$(obs - exp)^2 / exp$
1	1	707	706.78	0.22	0.00
2	1	169	165.01	3.99	0.10
2	2	266	264.93	1.07	0.00
3	1	39	46.66	-7.66	1.26
3	2	59	58.15	0.85	0.01
3	3	72	72.22	-0.22	0.00
4	1	297	292.79	4.21	0.06
4	2	195	196.31	-1.31	0.01
4	3	56	57.20	-1.20	0.03
4	4	520	520.84	-0.84	0.00

Table 3: Response (1,1) to items (i,j)

i	*j*	*k*	*obs*	*exp*	*obs-exp*	$(obs - exp)^2 / exp$
1	2	3	32	38.65	-6.65	1.14
1	2	4	126	123.85	2.15	0.04
1	3	4	31	37.62	-6.62	1.17
2	3	4	47	49.06	-2.06	0.09

Table 4: Response (1,1,1) to items (i,j,k)

The $\%G^2 = 94.8$ suggests that the model explains a large part of the association between the items and this is reinforced by the the one-, two- and three-way margins, given in Tables 3 and 4, which are all small.

The posterior means $E(\theta_a|X)$ and $E(\theta_e|X)$, given in Tables 5 and 6 are the scores of the individuals on the latent variables θ_a and θ_e respectively.

We may assume that people project their own views on society and so the views they express for other people or the society reflect their own beliefs. The expected posterior mean $E(\theta_a|X)$ is a measure of their racial prejudice.

$E(\theta_e|X)$ places individuals on the „tendency to respond“ scale according to the pattern of 9's in their responses.

The estimated standard deviations (s.d.) of the conditional distributions of θ_a and θ_e provide a measure of the information in the scoring. These are rather large. We observe that $E(\theta_a|X)$ is generally less informative for the response patterns that contain two or three 9's and becomes more accurate for the response patterns with no 9's.

resp pattern	*obs*	*exp*	*E(θa|X)*	*s.d.*	*E(θe|X)*	*s.d*
0009	18	17.7	-0.95	0.87	-1.40	0.47
0909	10	10.3	-0.74	0.90	-1.93	0.39
1009	15	13.9	-0.67	0.83	-1.44	0.47
9009	2	3.0	-0.66	0.83	-1.71	0.42
0000	379	375.7	-0.61	0.78	0.18	0.87
9909	3	3.7	-0.52	0.90	-2.18	0.43
0900	10	8.5	-0.50	0.83	-1.21	0.52
9000	13	10.5	-0.46	0.75	-0.70	0.67
1909	8	8.8	-0.44	0.88	-1.96	0.38
0099	1	1.7	-0.42	0.77	-1.90	0.37
1000	331	330.8	-0.40	0.74	0.18	0.88
0090	3	2.3	-0.27	0.70	-1.10	0.56
1900	6	7.9	-0.25	0.81	-1.25	0.52
9099	1	0.6	-0.25	0.74	-2.11	0.40
0999	4	3.5	-0.24	0.92	-2.41	0.47
9999	4	3.0	-0.01	0.96	-2.75	0.49
0001	131	140.2	0.07	0.62	0.16	0.89
1999	5	3.6	0.08	0.92	-2.44	0.46
9001	5	4.6	0.16	0.59	-0.75	0.68
1001	158	156.4	0.20	0.58	0.15	0.89
0091	5	1.2	0.28	0.57	-1.17	0.57
0901	7	4.2	0.37	0.75	-1.34	0.54
1091	2	1.5	0.40	0.54	-1.19	0.58
0010	2	2.2	0.46	0.52	0.15	0.89

Table 5: Posterior means and observed and expected frequencies of the response patterns

We also fitted the data resulting from the alternative dichotomization, i.e., from amalgamating the „mind a little“ with the „mind a lot“ answers. Results were similar in terms of the order of magnitude of the coefficients. The discrimination parameter of item 2 though was very large and the two-way margins not as small as the ones of the model fitted above. For these data (as for others we have modelled) the dichotomisation needs to be carefully chosen.

resp pattern	*obs*	*exp*	$E(\theta_a \mid X$	*s.d.*	$E(\theta_e \mid X)$	*s.d*
1010	2	2.8	0.56	0.51	0.15	0.89
1901	5	5.3	0.57	0.73	-1.37	0.54
9901	1	0.8	0.58	0.72	-1.70	0.43
0011	3	2.8	0.80	0.47	0.14	0.90
1011	5	4.1	0.88	0.45	0.14	0.90
0109	1	0.6	0.97	0.45	-1.69	0.44
0100	21	21.7	0.98	0.43	0.14	0.90
1109	1	1.0	1.04	0.43	-1.70	0.43
1100	36	33.5	1.04	0.42	0.14	0.90
9991	1	0.3	1.10	0.80	-2.09	0.38
1991	1	1.0	1.11	0.79	-1.90	0.36
0101	45	52.5	1.22	0.42	0.13	0.90
9101	1	2.3	1.27	0.42	-0.84	0.71
1101	98	88.9	1.29	0.43	0.13	0.90
0110	5	3.0	1.46	0.47	0.13	0.91
0191	1	1.1	1.49	0.55	-1.29	0.59
1110	6	5.6	1.55	0.49	0.13	0.91
9119	1	0.0	1.56	0.50	-1.95	0.36
9191	1	0.3	1.59	0.59	-1.70	0.45
0911	1	0.4	1.60	0.64	-1.47	0.54
1191	2	2.1	1.61	0.60	-1.30	0.60
0111	20	15.4	1.82	0.55	0.13	0.91
9111	1	0.8	1.90	0.56	-0.86	0.72
1111	26	32.9	1.94	0.57	0.13	0.91

Table 6: Posterior means and observed and expected frequencies of the response patterns

7. Conclusion

This paper illustrates the use of a two-dimensional latent trait model for missing observations. It provides estimates of the item parameters with respect to a latent attitude and simultaneously allows some information on the latent attitude to be recovered from the pattern of missing observations. It also provides scores of the observed response patterns both on the latent attitude and on the „tendency to respond“ latent factor. The results though are only indicative and should be interpreted with caution.

Acknowledgements

This research was supported by the ESRC Analysis of Large and Complex Datasets (ALCD) programme (Grant Reference H519255002). We thank Jane Galbraith and the two referees for their comments on an earlier version of this paper.

References

Albanese, M.T. and M. Knott (1992). TWOMISS: A computer program for fitting a one- or two-factor logit-probit latent variable model to binary data when observations may be missing. Technical report, Statistics Department, London School of Economics and Political Science.

Bartholomew, D.J. (1984). Scaling binary data using a factor model. *Journal of the Royal Statistical Society, Series B 46* (1), 120-123.

Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis*, Volume 40 of *Griffin's Statistical Monographs and Courses*. London: Charles Griffin & Company Ltd.

Bartholomew, D.J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology 41*, 101-107.

Bock, R. and M. Aitkin (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika 46* (4), 443-459.

Brook, L., G. Prior, and B. Taylor (1992). British social attitudes, 1991 survey. Technical report, Social & Community Planning Research.

Knott, M., M.T. Albanese, and J. Galbraith (1990). Scoring attitudes to abortion. *The Statistician 40*, 217-223.

Straud, A. and D. Sechrest (1966). Gaussian Quadrature Formulas. Prentice Hall.