UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE A: TRABALHO DE PESQUISA

POLYMISS: A COMPUTER PROGRAM FOR FITTING A ONE - OR TWO-FACTOR LOGIT-PROBIT LATENT VARIABLE MODEL TO POLYNOMOUS DATA WHEN OBSERVATIONS MAY BE MISSING

MARTIN KNOTT
M. TERESA ALBANESE

SÉRIE A, Nº 34
PORTO ALEGRE, OUTUBRO DE 1993

POLYMISS: A COMPUTER PROGRAM

FOR FITTING A ONE- OR TWO-FACTOR LOGIT-PROBIT LATENT VARIABLE MODEL

TO POLYTOMOUS DATA WHEN OBSERVATIONS MAY BE MISSING

Martin Knott[1] and Maria Teresa Albanese[2]

1 – Department of Statistics
London School of Economics and Political Science
Houghton Street
London WC2A 2AE.

2 – Departamento de Estatistica,
Universidade Federal do Rio Grande do Sul
Av. Bento Gonçalves, 9500
91500 Porto Alegre, Brasil.

USER DOCUMENTATION

SUMMARY

1-		Introduction	3
2-		Basic model for polytomous data	3
3-		Basic model for polytomous data with missing observations	5
4-		A modified E-M algorithm for one-factor model	6
5-		Goodness-of-fit	7
6-		Measurement of the latent variable	8
7-		Introduction to **POLYMISS**	9
8-		Description of input channels	10
	8.1-	Input format for channel 5	10
	8.2-	Input format for channel 3	12
9-		Description of output channels	13
	9.1-	Output of channel 8	13
	9.2-	Output of channel 7	13
	9.3-	General comments	13
10-		Example	15
	10.1-	One factor	15
	10.2-	Two factors	25
11-		References	35

Annexes:

1	The COUNT program	36
2	The ZSCORE program	39

1- Introduction

Latent variable modelling of polytomous data would be appropriate, for example, in attitude scaling where n individuals answer each of p items. We write the number of different responses for item i as $c_i, i = 1, ..., p$.

The **POLYMISS** program fits one and two logit-probit latent variable models to polytomous data which may have missing observations (Sections 2,3 and 4); provides information on goodness-of-fit (Section 5); and gives scores or fitted values of the latent variable(s) (Section 6).

The fitted latent variables can be interpreted as summarising the associations between the observed polytomous responses to the p items; i.e. as summarising what the items (or subsets of the items) measure in common. If the model does not fit the data then the calculated scores will not adequately summarise the structure of the associations.

Whether or not the model fits the data calculated scores may provide a useful partial summary of the data.

2- Basic Model for Polytomous Data

Let c_i denote the number of categories of item i (we shall also use the word *variable*) which are labelled $1, 2, ..., c_i$ $(i = 1, 2, ..., p)$ and indexed by s.

Suppose that $X_1, ..., X_p$ are p polytomous variables taking c_i values, so that

$$X_{i(s)} = \begin{cases} 1, & \text{if the response falls in category } s; \\ 0, & \text{otherwise.} \end{cases}$$

The $c_i \times 1$ vector with these elements is denoted by x_i and, obviously, $\sum_s x_{i(s)} = 1$ the full response pattern for an individual is denoted by $x' = (x_1', x_2', ..., x_p')$ of dimension $\sum_i c_i$.

We shall suppose that the responses are controlled by a latent variable Z. For each item there is a set of response functions $\pi_{i(s)}(z)$ defined by

$$P[X_{i(s)} = 1|z] = \pi_{i(s)}(z)$$

for $(s = 1, ..., c_i;\ i = 1, ...p)$ so that $\sum_s \pi_{i(s)}(z) = 1$

We shall assume conditional independence, that is the joint conditional probability $g(x|z)$ is

$$g(x|z) = \prod_{i=1}^{p} g_i(x_i|z) \tag{1}$$

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

Conditional independence is the assumption that the latent vector Z is sufficient to explain all the association between the responses given to different items by an individual.

As the X's are polytomous, the conditional probability function of x_i given z is the multinomial given by

$$g(x_i|z) = \prod_{s=1}^{c_i} (\pi_{i(s)}(z))^{x_{i(s)}} \tag{2}$$

for $i = 1, ..., p$, where $\pi_{i(s)}(z) = P[X_{i(s)} = 1|z]$, the probability than the answer of an individual to item i falls into category s.

Consequently, from (1) and (2) the joint probability function of x can be written as

$$f(x) = \int_{R_z} h(z) \prod_{i=1}^{p} \prod_{s=1}^{c_i} \pi_{i(s)}(z)^{x_{i(s)}} dz \tag{3}$$

where R_z is the range space of Z and $h(.)$ is the prior density of Z.

One may obtain many different models by specification of the latent vector prior density $h(.)$ and the shape of the response function $\pi_{i(s)}(z)$. POLYMISS is concerned with the one- and two-latent logit-probit response function as defined below.

One factor model

In the one factor model a single latent variable (factor) Z_a, explains all the association between the responses to different items.

We use the **logit-probit** response function that is, the response function for category s of item i is given by

$$\text{logit}(\pi_{i(s)}z_a) = a_{0i(s)} + a_{1i(s)}z_a,$$

or

$$\pi_{i(s)}z_a = \frac{\exp(a_{0i(s)} + a_{1i(s)}z_a)}{\sum_{r=1}^{c_i}\exp(a_{0i(r)} + a_{1i(r)}z_a)}. \tag{4}$$

where $\sum_s \pi_{i(s)}(z_a) = 1$ with $i = 1, 2, ..., p$ and $s = 1, 2, ..., c_i$. Z_a is distributed as N(0,1) and the parameters $a_{0i(s)}$ and $a_{1i(s)}$ are referred to as **difficulty** and **discriminating** parameters, of item i and category s respectively. These parameters determine the position and shape of the response function.

As expressed in (4), $\pi_{i(s)}(z)$ is over-parametrized, so without loss of generality we may fix the location of a set $a_{1i(s)}$, as we please.

In **POLYMISS** we have chosen to take $a_{1i(1)} = 0$. The order in which the categories are labelled is arbitrary, but the one labelled 0 will be called the reference category for that variable.

To facilitate the interpretation of $a_{0i(s)}$ we put $z_j = 0$ for all j,and thus obtain the response probability for the "median" individual. Let this be denoted by

$$\pi_{i(s)}(0) = \frac{\exp(a_{0i(s)})}{\sum_{r=1}^{c_i}\exp(a_{0i(r)})}. \tag{5}$$

Since the origin of the a's is arbitrary we may again set $a_{0i(1)} = 0$ and then

$$\pi_{i(s)}(0) = \frac{\exp(a_{0i(s)})}{1 + \sum_{r=2}^{c_i}\exp(a_{0i(r)})} \tag{6}$$

for $i = 1, 2, ..., p$ and $s = 1, 2, ..., c_i$.

The interpretation of the $\pi_{i(s)}$'s is the probability of a positive response of a median individual for category s of item i.

The discriminating power of an item i is indicated by the spread of the $a_{1i(s)}$ as functions of s. A large spread produces larger differences between the corresponding response probabilities and so a better chance of discriminating between individuals a given distance apart on the z-scale on the evidence of x_i.

The a_1's are weights in the component score (see Section 6). Here we are looking at the relative influence which each observed variable (item) has in determining the value of the component. An item will be an important determinant if all the $a_{1i(s)}$ for a given i are large. It is the average level of the a_i's rather than their dispersion which counts.

Two factor model

If there are two factors (latent variables) then the response function of category s of item i in equation 4 becomes

$$P(X_{i(s)} = 1|z_{a1}, z_{a2}) = \pi_{i(s)}(z_{a1}, z_{a2})$$

where

$$\text{logit}(\pi_{i(s)}(z) = a_{0i(s)} + a_{1i(s)}z_{a1} + a_{2i(s)}z_{a2}$$

or

$$\pi_{i(s)}(z) = \frac{\exp(a_{0i(s)} + a_{1i(s)}z_{a1} + a_{2i(s)}z_{a2})}{\sum_{r=1}^{c_i} \exp(a_{0i(r)} + a_{1i(r)}z_{a1} + a_{2i(r)}z_{a2})} \tag{7}$$

for $i = 1, 2, ..., p$, $s = 1, 2, ..., c_i$ where z_{aj} is the jth latent variable (factor), $j = 1, 2$. We assume that Z_{a1} and Z_{a2} are independent normal with mean zero and variance one.

By analogy with one factor model, POLYMISS sets $a_{0i(1)} = 0$ and $a_{1i(1)} = 0$, for $i = 1, 2, .., p$. It follows that the response function is given by

$$\pi_{i(s)}(z) = \frac{\exp(a_{0i(s)} + a_{1i(s)}z_{a1} + a_{2i(s)}z_{a2})}{1 + \sum_{r=2}^{c_i} \exp(a_{0i(r)} + a_{1i(r)}z_{a1} + a_{2i(r)}z_{a2})}$$

for $i = 1, 2, ..., p$, $s = 2, 3, ..., c_i$.

3- Basic Model for Polytomous Data with Missing Observation

The following model is an extension for polytomous data of the simplest model for binary data proposed by Knott at al.(1990). Observations may be missing for a variety of reasons, for example, no opinion was expressed for that item, either because the response was 'don't know', or because the response was not recorded. A full analysis would take these reasons into account, possibly by treating the observable variables as polytomous - having several categories to allow for different types of missing values. We shall let $X_i = 9$ denote a missing value (non-response or 'no-opinion expressed') for the i'th item (observable variable).

For a fairly general model we could allow a two-dimensional common factor $Z = (Z_a, Z_e)$ where Z_a, Z_e are given independent $N(0, 1)$ distributions. For example, the factor Z_a might summarise attitude and Z_e might summarise the tendency to express an opinion. As was the case for the model for polytomous data for complete response (no observations missing), we shall assume that individuals behave independently, and we assume that the choices made by an individual to respond with approval, disapproval or not to respond at all are conditionally independent between items given the individual's value of Z. We break down the modelling of the response function into two layers.

For each item,

$$P(X_{i(s)} = 1|Z, X_{i(s)} \neq 9) = \pi_{ai(s)}(z_a)$$

and

$$P(X_{i(s)} \neq 9|Z) = \pi_{ei(s)}(z_a, z_e).$$

It follows that

$$P(X_{i(s)} = 1|Z) = \pi_{ai(s)}(z_a)\pi_{ei(s)}(z).$$

$$P(X_{i(s)} = 0|Z) = (1 - \pi_{ai(s)}(z_a))\pi_{ei(s)}(z).$$

and

$$P(X_{i(s)} = 9|Z) = 1 - \pi_{ei(s)}(z).$$

Thus for this family of models if a response $1, 2, ..c_i$ has been observed, the response function is the same at that for the models for complete responses (section 2), but the probability of a missing value (9) is allowed to depend in the most general case on both factors Z_a and Z_e. For example one can within this family of models allow the underlying attitude Z_a to affect the probability of a response, and so one may hope to recover information about attitudes from the pattern of non-response (non-expression of opinion).

It is probably not worth using the missing responses in any of these models if the only response pattern with missing information has every item coded 9. We consider the simplest sub-models of the family.

This is a simple model, in which for each item the probability of a missing response 9 is assumed constant over all individuals, independently of other items. For one-factor model

$$\text{logit}(\pi_{ai(s)}(z)) = a_{0i(s)} + a_{1i(s)}z_a$$

$$\text{logit}(\pi_{ei(s)}(z)) = e_{0i(s)} \tag{8}$$

for $i = 1, 2, ..., p$ and $s = 2, 3, ..., c_i$.

For two-factors model

$$\text{logit}(\pi_{ai(s)}(z)) = a_{0i(s)} + a_{1i(s)}z_{a1} + a_{2i(s)}z_{a2}$$

$$\text{logit}(\pi_{ei(s)}(z)) = e_{0i(s)} \tag{9}$$

for $i = 1, 2, ..., p$ and $s = 2, 3, ..., c_i$.

The model says that every individual has the same probability of expressing an opinion on a given item, and that non-expression gives no information about attitudes. There is no allowance for clustering of non-response over several items for the same respondent.

4- A Modified E-M Algorithm for One-Factor Model

The **POLYMISS** program fits one or two factor logit-probit latent models described in sections 2 and 3, using a modified E-M algorithm proposed by Bartholomew (1987, Chapter 7). The main results for the one factor latent variable model follow.

Even though the latent variable Z_a is distributed as $N(0,1)$, it is proposed as an approximation that Z_a assumes values z_{a1}, z_{a2}, ..., z_{ak} with probabilities $h(z_{a1})$, $h(z_{a2})$, ..., $h(z_{ak})$ chosen so that the joint probability function

$$f(x_u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_u|z_a)h(z_a)dz_a$$

for the response pattern $x_u, u = 1, 2, ..., n$ can be approximated with high accuracy by Gauss-Hermite quadrature, *i.e*,

$$f(x_u) = \sum_{t=1}^{k} g(x_u|z_{at})h(z_{at})$$

$u = 1, 2, ..., n$, where z_{at} is a tabled quadrature point (node) and $h(z_{at})$ is the corresponding weight (see Straud and Sechrest,1966).

The quadrature weights, $h(z_{at})$, are approximately the normalized, i.e., $\sum_{t=1}^{k} h(z_{at}) = 1$, values of the probability density of a $N(0,1)$ random variable at the points z_{at}, which are chosen to give the best approximation to the marginal probability function $f(x_u)$. This approximation becomes more accurate as the number of quadrature points increases. From the maximization of

$$L = \sum_{u=1}^{n} \log f(x_u)$$

we obtain, for $v = 0, 1$ and $i = 1, 2, ..., p$, $s = 2, .., c_i$

$$\frac{\partial L_z}{\partial a_{vi(s)}} = \sum_{t=1}^{k} z_{at}^{v}\{\sum_{u=1}^{n} x_{ui(s)}h(z_{at}|x_u) - \pi_{i(s)}(z_{at})\sum_{u=1}^{n} h(z_{at}|x_u)\}$$

$$\frac{\partial L_z}{\partial a_{vi(s)}} = \sum_{t=1}^{k} z_{at}^{v}\{R_{it(s)} - N_t\pi_{i(s)}(z_{at})\} \tag{10}$$

where

$$R_{it(s)} = \sum_{u=1}^{n} x_{ui(s)} h(z_{at}|x_u), \tag{11}$$

$$N_t = \sum_{u=1}^{n} h(z_{at}|x_u), \tag{12}$$

and $h(z_{at}|x_u)$ is the conditional probability of Z_{at} given the response pattern x_u.

The estimation of the parameters is performed by choosing any starting values for $\{a_{0i(s)}\}$ and $\{a_{1i(s)}\}$ followed by repeated applications of (10), (11) and (12) over the set of items, using an E-M algorithm defined as

E-step: Calculate the values of $R_{it(s)}$ and N_t using equations (11) and (12).

M-step: Obtain improved estimates of the $\{a_{0i(s)}\}$ and $\{a_{1i(s)}\}$ solving equation (10), using the values of $R_{it(s)}$ and N_t from the E-step.

The E-M cycles are continued until convergence is obtained. In this case the number of values that the latent variable assumes is fixed and the set of values constitutes the distribution of Z_a.

5- Goodness-of-fit

If the sample size n is large compared with $\prod_{i=1}^{p} c_i$ (number of possible response patterns) a chi-squared or log-likelihood goodness-of-fit test can be carried out on the observed frequencies of the response patterns. Often, there are many small expected frequencies so that pooling becomes necessary. Since the number of degrees of freedom in the unpooled case for a single factor variable is $\prod_{i=1}^{p} c_i - 2\sum_{i=1}^{p}(c_i - 1)$ and for two factors is $\prod_{i=1}^{p} c_i - 3\sum_{i=1}^{p}(c_i - 1)$, then situations may occur where there will be no degrees of freedom to judge the goodness-of-fit.

When a formal test cannot be carried out and p 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. An additional check may be done by comparing the observed and fitted values of the one-, two- and three-way marginal frequencies or an analysis of residuals.

6- Measurement of the Latent Variable

We might wish to score each response pattern, and hence each individual, using scores which measure the latent variables.

Basic model for polytomous data

We could for example, score the response 11232 by using the **total score**, giving the score 9. In this case we would be considering that all items are equally important, and scoring categories of response with their integer labels in a fairly arbitrary way.

Another way of scoring the response patterns (Bartholomew's (1987) **component score**) is to use the sum of the parameter estimates $\hat{a}_{1i(s)}$ of the model. The response pattern 11232 would be scored $\hat{a}_{11(1)} + \hat{a}_{12(1)} + \hat{a}_{13(2)} + \hat{a}_{14(3)} + \hat{a}_{15(2)} = \hat{a}_{13(2)} + \hat{a}_{14(3)} + \hat{a}_{15(2)}$ since $\hat{a}_{1i(1)}$ was set equal to zero for all $i = 1, 2, .., p$. This method of scoring is more informative and give a better scoring than just using the total score. We also may use the estimated **conditional mean** of the latent variable given the response pattern x, that is, $\hat{E}(Z|x)$ which can be coupled with the use of the conditional standard deviation given the response pattern as a measure of the accuracy of the latent variable. Knott and Albanese (1992) or Albanese (1990) investigated the relation between component scores and conditional means for binary data and found that the latter maintains the advantages of the component score, but it is more stable. These results should hold for more than 2 categories. The component score is strongly dependent on the values of $a_{1i(s)}$ while the conditional mean depends on $\pi_{i(s)}$ (equation 4), which does not vary much for different choices of $\hat{a}_{1i(s)} > 3$.

If we want to use the latent score in further analysis then the conditional mean is more informative and reliable than the component score, specially when one or more $\hat{a}_{1i(s)}$ are large (bigger than 3.0).

Basic models for polytomous data including missing observations

If the data include missing observations, the conditional mean is more suitable as a measurement of the latent variable than the total score or the component score, since these latter do not cope with missing observations.

7- Introduction to POLYMISS

The computer program **POLYMISS** fits one and two factor logit-probit latent variable models to polytomous data, which may have missing observations. The parameters are estimated by a marginal maximum likelihood procedure using a modified E-M (expectation-maximisation) algorithm (Section 4).

POLYMISS also calculates the proportion of missing observations for each observed variable and the asymptotic standard deviations of the parameters. Observable variables may be responses to items in a questionnaire.

Bartholomew's component score (1980) and/or the estimated conditional mean and standard deviation of each latent variable given the response pattern can be calculated for each response pattern (and hence for each individual). For the two factor model the first and second conditional means and their standard deviations are given.

The goodness-of-fit of the model is measured by the value of a likelihood ratio (LR) statistic, which should be interpreted with caution when the number of observed variables is large compared with the sample size. When the LR statistic is inappropriate the user may judge the goodness-of-fit from a comparison between the observed and expected first, second and third order margins.

Input can be either individual response patterns or their frequency distributions as presented in the next section.

POLYMISS is a program written in standard FORTRAN 77 for micro-computers, or large size computers. The execution time of this program can increase rapidly with the number of observed variables and the number of quadrature points.

POLYMISS program suite contains a main routine POLY and the following subroutines, which must be compiled and linked together: EM1, EM2, PCOUNT, QUAD, PHILIK, ENER, VARIANCE, POSMEAN, MARGIN and INV. The routines described in the annexes, COUNT and ZSCORE run separately.

The initial control parameters which define the dimension of the matrices are

$NRP = 1500$ corresponding to the maximum sample size if the data are given as individual response patterns, or the maximum number of different response patterns if the frequency distribution of the distinct response patterns is given. The frequency distribution of the distinct response patterns can be obtained using the COUNT program;

$NV = 20$, the maximum number of variables;

The maximum number of categories for each item was set to be equal to 5.

Let NC be the total number of categories then $M1 = 2*(NC - NV)$ and $M2 = 3*(NC - NV)$, the maximum number of parameters to be estimated (for the one and two factor models respectively).

The annex of POLYMISS gives two programs: COUNT and ZSCORE.

COUNT is designed to provide the frequency distribution of the responses given by N individuals to P items, which take values $1, 2, .., c_i$ (maximum equal to 5).

ZSCORE is designed to provide the scoring of the latent variable for each individual in the order given in the input file. As the output from the POLYMISS program displays the response patterns in increasing order according to the conditional mean, ZSCORE needs to be used when the latent scores for response patterns in the input file order are the input data in further analysis.

8- Description of the Input Channels

8.1- Input format for channel 5 (filename: POLY.INP)

The file for this channel contains all the control parameters and the data set as described below. Control parameters are read in FREE format, but the data set is read in FIXED format.

<u>Line 1:</u>

Title: Title of data set (at most 70 characters).

<u>Line 2:</u>

N, P, NQ

N: number of individuals in sample
P: number of observed variables (items)
NQ: number of quadrature points (8, 16, 24, 32, 48)

<u>Line 3:</u>

C(1) C(2) ... C(P)

C(i): number of categories (2, 3, 4, 5)

<u>Line 4:</u>

NFAC, INPUT, FREQ, DISPLAY, MTER, LOUT8, ERRC

NFAC : 1 or 2 (number of factors)
INPUT : 0 or 1
0 The initial parameter estimates are set in the program, that is,

One-factor

$$
\begin{array}{llcll}
0 & 0.5 & = & \hat{a}_{01(2)} & \hat{a}_{11(2)} \\
0 & 0.5 & = & \hat{a}_{01(3)} & \hat{a}_{11(3)} \\
. & . & = & \ldots & \ldots \\
0 & 0.5 & = & \hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} \\
0 & 0.5 & = & \hat{a}_{02(2)} & \hat{a}_{12(2)} \\
0 & 0.5 & = & \hat{a}_{02(3)} & \hat{a}_{12(3)} \\
. & . & = & \ldots & \ldots \\
0 & 0.5 & = & \hat{a}_{02(c(2))} & \hat{a}_{12(c(2))} \\
. & . & = & \ldots & \ldots \\
0 & 0.5 & = & \hat{a}_{0p(2)} & \hat{a}_{1p(2)} \\
0 & 0.5 & = & \hat{a}_{0p(3)} & \hat{a}_{1p(3)} \\
. & . & = & \ldots & \ldots \\
. & . & = & \ldots & \ldots \\
0 & 0.5 & = & \hat{a}_{0p(c(p))} & \hat{a}_{1p(c(p))}
\end{array}
$$

Two-factors

$$
\begin{array}{lllclll}
0 & 1.0 & 1.5 & = & \hat{a}_{01(2)} & \hat{a}_{11(2)} & \hat{a}_{21(2)} \\
0 & 1.0 & 1.5 & = & \hat{a}_{01(3)} & \hat{a}_{11(3)} & \hat{a}_{21(3)} \\
. & . & . & = & \ldots & \ldots & \ldots \\
0 & 1.0 & 1.5 & = & \hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} & \hat{a}_{21(c(1))} \\
0 & 1.0 & 1.5 & = & \hat{a}_{02(2)} & \hat{a}_{11(2)} & \hat{a}_{21(2)} \\
0 & 1.0 & 1.5 & = & \hat{a}_{02(3)} & \hat{a}_{11(3)} & \hat{a}_{21(3)} \\
. & . & . & = & \ldots & \ldots & \ldots \\
0 & 1.0 & 1.5 & = & \hat{a}_{02(c(2))} & \hat{a}_{11(c(2))} & \hat{a}_{21(c(2))}
\end{array}
$$

.	.	.	=	...	...	...
.	.	.	=	...	...	...
0	1.0	1.5	=	$\hat{a}_{0p(2)}$	$\hat{a}_{1p(2)}$	$\hat{a}_{2p(2)}$
0	1.0	1.5	=	$\hat{a}_{0p(3)}$	$\hat{a}_{1p(3)}$	$\hat{a}_{2p(3)}$
.	.	.	=	...	...	...
0	1.0	1.5	=	$\hat{a}_{0p(c(p))}$	$\hat{a}_{1p(c(p))}$	$\hat{a}_{2p(c(p))}$

INPUT 1 The initial parameter estimates are given in channel 3.

FREQ: 0 or 1
0 Input data are individual response patterns.
1 Input data is the frequency distribution of the response patterns.

DISPLAY: 0 or 1
0 Frequency distribution of response patterns is not displayed prior to fitting the model.
1 Frequency distribution of response patterns is displayed prior to fitting the model.

MTER: Maximum number of iterations. Set to 50 for a first try. If ERRC below is made smaller, MTER may need to increase.

LOUT8: 0 or 1
0 Do not create a special file for the final parameter estimates.
1 Create a special file for the parameter estimates (LI8.OUT). This file may be used as an input file for channel 3 when running the program again for a large number of iterations.

ERRC: Convergence tolerance for the E-M algorithm , for instance 0.00001 with MTER set to 50.

Reading response patterns

If FREQ = 0 then

Line 5, 6,...,N+5 Format(20I1)

IRESP(1,1)	IRESP(1,2)	...	IRESP(1,P)
IRESP(2,1)	IRESP(2,2)	...	IRESP(2,P)
...	...	...	...
IRESP(N,1)	IRESP(N,2)	...	IRESP(N,P)

where IRESP(L,I) = 1,2,...,C(I) or 9 (missing) is the response of individual L, L=1,2,...,N to item I, I=1,2,...,P. Thus, for example, if the number of categories is 3 then IRESP(L,I) is equal to 1, 2, 3 or 9.

If FREQ = 1 then

Line 5: NR Number of different response patterns.

Line 6, 7,...,NR+5: Format(I4,1X,20I1)

RL(1)	IRESP(1,1)	IRESP(1,2)	...	IRESP(1,P)
RL(2)	IRESP(2,1)	IRESP(2,2)	...	IRESP(2,P)
...	...	...	...	...
RL(NR)	IRESP(NR,1)	IRESP(NR,2)...		IRESP(NR,P)

where RL(L), L = 1,2,...,NR is the observed frequency of the response pattern L and IRESP(L,I) = 1,2,...,C(I) or 9 (missing) is the response to item I, I=1,2,...,P of the response pattern L. Thus, for example, if the number of categories is 3 then IRESP(L,I) is equal to 1, 2, 3 or 9.

8.2- Input format for channel 3 (filename: POLY3.INP)

If INPUT = 1, initial parameter estimates are inputted from channel 3. They are read in FREE format and we can use the output estimates in channel 8 as the input in channel 3.

It follows the input format for channel 3:

Line 1: Title or first line from channel 8.

Line 2,3,...,P+1: Free format

Each line corresponds to the parameter estimates of item I, I= 1,2...,P, obtained from fitting the logit-probit model for one or two factors.

One-Factor

$$
\begin{array}{ll}
\hat{a}_{01(2)} & \hat{a}_{11(2)} \\
\hat{a}_{01(3)} & \hat{a}_{11(3)} \\
. & . \\
\hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} \\
\hat{a}_{02(2)} & \hat{a}_{12(2)} \\
\hat{a}_{02(3)} & \hat{a}_{12(3)} \\
. & . \\
\hat{a}_{02(c(2))} & \hat{a}_{12(c(2))} \\
. & . \\
\hat{a}_{0p(2)} & \hat{a}_{1p(2)} \\
\hat{a}_{0p(3)} & \hat{a}_{1p(3)} \\
. & . \\
. & . \\
\hat{a}_{0p(c(p))} & \hat{a}_{1p(c(p))}
\end{array}
$$

Two-Factor

$$
\begin{array}{lll}
\hat{a}_{01(2)} & \hat{a}_{11(2)} & \hat{a}_{21(2)} \\
\hat{a}_{01(3)} & \hat{a}_{11(3)} & \hat{a}_{21(3)} \\
. & . & . \\
\hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} & \hat{a}_{21(c(1))} \\
\hat{a}_{02(2)} & \hat{a}_{12(2)} & \hat{a}_{22(2)} \\
\hat{a}_{02(3)} & \hat{a}_{12(3)} & \hat{a}_{22(3)} \\
. & . & . \\
\hat{a}_{02(c(2))} & \hat{a}_{12(c(2))} & \hat{a}_{22(c(2))} \\
. & . & . \\
\hat{a}_{0p(2)} & \hat{a}_{1p(2)} & \hat{a}_{2p(2)} \\
\hat{a}_{0p(3)} & \hat{a}_{1p(3)} & \hat{a}_{2p(3)} \\
. & . & . \\
. & . & . \\
\hat{a}_{0p(c(p))} & \hat{a}_{1p(c(p))} & \hat{a}_{2p(c(p))}
\end{array}
$$

9- Description of the Output

9.1- Output from Channel 8 (filename: LI8.OUT)

Channel 8 gives the parameter estimates, which may be used directly as input for channel 3 (Input=1) if the program is run again. Line 1 shows the title, number of variables and quadrature points.

9.2- Output from Channel 7 (filename: LI7.OUT)

This output contains all results of the fitting.

(1) All control parameters (sample size, number of variables, etc.) which steer program activity are printed.

(2) The E-M algorithm is an iterative scheme for computing maximum likelihood estimates. Convergence is tested at each iteration by examining the relative change in the value of the likelihood function.

(3) The maximum likelihood search routine performs at most MTER iterations. If the E-M has not converged by then (precision still bigger than ERRC) a message is printed out. We can use the estimates given in LI8.OUT as input for channel 3 and run the program again, thus proceeding until convergence is obtained.

(4) The asymptotic standard deviations of the parameter estimates are obtained by inverting the observed second derivative matrix at the maximum likelihood (ML) solution.

(5) If the standard deviations of the parameter estimates are too large they will come out in the output as '*****'. This is likely to happen when the parameter estimates are very large.

(6) The goodness-of-fit is measured by the value of a likelihood ratio statistic, G-Square, defined as

$$G^2 = 2\sum_{l=1}^{NR} RL(L) \ln \frac{RL(L)}{N * PL(L)}$$

where $RL(L)$ and $N * PL(L)$ are the observed and expected frequencies of the response pattern L, which has an asymptotic χ^2 distribution on $(NR_0 - 2 * (NC - P))$ degrees of freedom for one factor model and $(NR_0 - 3 * (NC - P))$ degrees of freedom for two factors model where NR_0 denotes the number of terms in the summation (bearing in mind that response patterns whose expected parameters estimated (NC: total number of categories and P: number of variables). For large $NC - P$ most of the observed frequencies $RL(L)$ will take the value 0 or 1 and the expected frequency will be very small. In this case the likelihood ratio statistic will be inappropriate. We can use the comparison between the observed and expected first, second and third margin order to check how well the model fits the data. If some responses are missing, the G-square statistic calculated by the program is not appropriate, so the comparison of margins should be used to assess goodness-of-fit.

(7) The RG statistics is defined as

$$RG = \frac{G^2_{(k+1)} - G^2_{(k)}}{G^2_{(k)}}$$

where $G^2_{(k)}$ is the G-square statistics at step k. If RG is smaller than the convergence tolerance (ERRC), the process is said to converge. For two-factor models one may need to take ERRC as small as 0.0000001 to be sure that the iterations have come close to finding the ML estimates. A final decision on whether the ML estimates have been achieved is best made by the user after checking that the derivatives are near to zero.

(8) If the model fits the data, we might be interested in scoring each response pattern, and consider this score a measurement of the latent variable.

We may score the response pattern x_u with the **conditional mean**

$$E(Z|x_u) = \int_z zh(z|x_u)dz$$

the **component score**

$$CS(x_u) = \sum_{i=1}^{p} a_{1i(s)} x_{ui(s)}$$

or the **total score**

$$T(x_u) = \sum_{i=1}^{p} \sum_{s=1}^{c_i} x_{ui(s)}$$

(9) When the response pattern includes any missing response, the component score and the total score are set equal to zero.

(10) The response patterns are ranked in increasing order of the conditional mean of the first latent variable. Also provided are the standard deviations of the first and second latent variables given the response pattern. Large values of the standard deviation may lead one to exclude that response pattern from further analysis of the scores of response patterns, (for example, the conditional standard deviation may be large where more than half of responses are missings).

(11) All item parameter estimates ($\hat{a}_{vi(s)}, v = 0, 1, 2$) are constrained to lie between -10.0 and 10.0, without loss of information. In general this is necessary to ensure that when one or more $\hat{a}_{vi(s)}, v = 0, 1, 2$ get large we still can see how well the model fits the data before the PHILIK routine 'blows up' or the VARIANCE routine sends out a message that the matrix is singular. The derivative for a truncated $\hat{a}_{vi(s)}$ may be large when iteration finish.

(12) Very often when working with a 5 points scale, one or both of the extremes categories have very low observed frequencies for at least one item. In this case, sometimes the iterative procedure diverges, that is, the loglikelihood starts decreasing and after few iterations, increasing again. When this starts happening, the iterative procedure stops and POLYMISS prints out the message: 'Iterative procedure diverged'. Either try a new starting point, or reduce the number of quadrature points.

9.3 - General Comments

(1) For two-factor models it is best to start off with a small number of quadrature points, say 8, and to use the output from that run for the initial parameter values of calculations for larger numbers of quadrature points.

(2) It may be better to compile the FORTRAN with an option /G_FLOATING to avoid overflow errors. This is particularly important for the two-factor models.

(3) Constraints on RAM in small personal computers may require that two-factor models are fitted with only a small number of quadrature points.

(4) The program COUNT in the annexe may be used to preprocess the response patterns to reduce their number. The program ZSCORE of the annexe may be used for scoring responses on the latent variable in the order of response input, which is useful if the scores are to be used for other analysis.

10 - Example

Environment Attitudes Data

The data are the responses in 1990 given by 311 individuals to six of the seventeen items concerning attitude to environment by members of a panel surveyed in each of the years 1989 to 1990 as part of an investigation of British Social Attitudes. For each item, respondents were asked: "How concerned are you about each of these environmental issues?"

(1)- Very concerned

(2)- Slightly concerned

(3)- Not very concerned

(4)- Not at all concerned

(7)- Don't know enough to make up my mind

The six selected items are:

(1)- Lead from petrol

(2)- River and sea pollution

(3)- Transport and storage of radioactive waste

(4)- Air pollution

(5)- Transport and disposal of poisonous chemicals

(6)- Risks from nuclear power station

As the seventeen items, except the original item thirteen, had observed frequency smaller than 10% in category 4; and four of the seventeen had observed frequency smaller than 10% in category 3, these two categories were amalgamated. Thus the new category 3 means not very concerned or not at all concerned.

Some members of the panel failed to respond to the items about environment either completely or in part, leading us to fit the logit-probit model including missing observation described in Section 3. Category 7 was also coded as missing(9) response.

10.1- One Factor

Input format for channel 5:

For the data we are considering in the first line we have ENVIROMENT 6 ITEMS

On the second line we have may set N = 311   P = 6   NQ = 48.

On the third line we may set C(1)=3 C(2)=3 C(3)=3 C(4)=3 C(5)=3 C(6)=3

On the fourth line we may have NFAC = 1, INPUT = 0, FREQ = 0, DISPLAY = 0, MTER = 50, LOUT = 1, ERRC = 0.0000001

From the fifth line we display individual response patterns.

File: POLY.INP

```
ENVIRONMENT 6 ITEMS
    311 6 48
    3 3 3 3 3 3
    1 0 0 0 50 1 0.0000001
    111111
    119991
    211112
    219193
    ......
    ......
    ......
    312122
    211223
    223323
    123323
```

Output from channel 7

```
    *** PROGRAM POLYMISS ***

MAXIMUM LIKELIHOOD ESTIMATION OF A 1 FACTOR LOGIT/PROBIT MODEL
FOR POLYTOMOUS DATA

ENVIRONMENT 6 ITEMS

NUMBER OF OBSERVED VARIABLES = 6

NUMBER OF CASES SAMPLED = 311

NUMBER OF DIFFERENT RESPONSE PATTERNS = 105

NUMBER OF QUADRATURE POINTS USED = 48

MAXIMUM NUMBER OF ITERATIONS PERMITTED = 50

CONVERGENCE TOLERANCE FOR THE RELATIVE LIKELIHOOD VALUE = 0.00000010

A LIKELIHOOD RATIO TEST OF OBSERVED AND EXPECTED
FREQUENCIES OF RESPONSE VECTORS IS TO BE CARRIED OUT

ASYMPTOTIC STANDARD DEVIATIONS ARE OBTAINED FROM THE
INVERSE OF THE OBSERVED SECOND DERIVATIVE MATRIX
```

MARGINS

ITEM	CATEGORY	
ITEM 1		
	1	0.598
	2	0.331
	3	0.058
	MISSING	0.013
ITEM 2		
	1	0.772
	2	0.183
	3	0.026
	MISSING	0.019
ITEM 3		
	1	0.717
	2	0.196
	3	0.061
	MISSING	0.026
ITEM 4		
	1	0.633
	2	0.322
	3	0.029
	MISSING	0.016
ITEM 5		
	1	0.717
	2	0.193
	3	0.061
	MISSING	0.029
ITEM 6		
	1	0.508
	2	0.312
	3	0.161
	MISSING	0.019

INITIAL ESTIMATES OF ITEM PARAMETERS

	CATEGORY	A(0,I,J)	A(1,I,J)
ITEM 1			
	2	0.000	1.000
	3	0.000	1.000
ITEM 2			
	2	0.000	1.000
	3	0.000	1.000
ITEM 3			
	2	0.000	1.000
	3	0.000	1.000
ITEM 4			
	2	0.000	1.000
	3	0.000	1.000
ITEM 5			
	2	0.000	1.000
	3	0.000	1.000
ITEM 6			
	2	0.000	1.000
	3	0.000	1.000

ITER	PROP	LOGLIKELIHOOD
1	0.33470	-1572.86346
1	0.17443	-1488.09535
2	0.11813	-1440.70171
3	0.08336	-1411.20712
4	0.05870	-1392.16919
5	0.03856	-1380.39615
6	0.02463	-1373.16625
7	0.01576	-1368.65376
.	.	.
.	.	.
.	.	.
44	0.00001	-1359.41265
45	0.00000	-1359.41136
46	0.00000	-1359.41025
47	0.00000	-1359.40928
48	0.00000	-1359.40844
49	0.00000	-1359.40771
50	0.00000	-1359.40708

*** MAX. NO. OF ITERATION EXCEEDED WITH RG STATISTICS 0.0000023

MORE ITERATIONS NEEDED TO ACHIEVE CONVERGENCE. CHECK THE DERIVATIVES OF LOG-LIKELIHOOD FOR FURTHER INFORMATION ON CONVERGENCE.

*** ITERATIONS FINISHED ***

NUMBER OF ITERATIONS IS 50

% OF G-SQUARE EXPLAINED 50.0926

LOGLIKELIHOOD VALUE -1359.4071

LIKELIHOOD RATIO STAT. 213.9481

DEGREES OF FREEDOM -11

DERIVATIVES OF LOGLIKELIHOOD

ITEM	CATEGORY	PARAMETER	DERIVATIVE
1	2	1	-0.01038007
1	2	2	-0.01607449
1	3	1	-0.00762531
1	3	2	-0.01175137
2	2	1	-0.01003292
2	2	2	-0.02018897
2	3	1	-0.00671159
2	3	2	-0.01343167
3	2	1	-0.01973129
3	2	2	-0.02904877
3	3	1	-0.01313931
3	3	2	-0.02032499
4	2	1	-0.02638349
4	2	2	-0.04884083
4	3	1	-0.01862773
4	3	2	-0.03445092
5	2	1	-0.01262074
5	2	2	-0.01632689
5	3	1	-0.00801406
5	3	2	-0.01065429
6	2	1	-0.02480291
6	2	2	-0.02726217
6	3	1	-0.01646732
6	3	2	-0.01903431

MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS AND STANDARD DEVIATIONS

ITEM	CAT	A(0,I,J)	S.D	A(1,I,J)	S.D	PHI(I,J)
1	2	-0.697	0.174	1.363	0.260	0.322
1	3	-2.978	0.399	2.042	0.435	0.033
2	2	-2.105	0.328	1.993	0.408	0.108
2	3	-6.406	0.984	3.713	0.768	0.001

3	2	-2.346	0.441	3.180	0.683	0.087
3	3	-5.634	1.062	4.914	1.016	0.003
4	2	-1.239	0.315	2.888	0.625	0.225
4	3	-8.130	1.546	6.118	1.202	0.000
5	2	-2.574	0.532	3.455	0.852	0.070
5	3	-5.105	0.996	4.603	1.096	0.006
6	2	-0.317	0.176	1.474	0.311	0.384
6	3	-1.792	0.351	2.677	0.482	0.088

FIRST ORDER OBSERVED AND EXPECTED MARGINS

ITEM I1	CAT J1	OBS	EXPECT	OBS-EXP	((O-E)**2)/E
1	1	186	185.67	0.33	0.0006
1	2	103	103.22	-0.22	0.0005
1	3	18	18.11	-0.11	0.0006
2	1	240	238.01	1.99	0.0166
2	2	57	58.85	-1.85	0.0582
2	3	8	8.14	-0.14	0.0023
3	1	223	221.18	1.82	0.0150
3	2	61	62.10	-1.10	0.0194
3	3	19	19.72	-0.72	0.0264
4	1	197	196.45	0.55	0.0016
4	2	100	100.34	-0.34	0.0012
4	3	9	9.21	-0.21	0.0049
5	1	223	221.49	1.51	0.0103
5	2	60	60.99	-0.99	0.0162
5	3	19	19.51	-0.51	0.0135
6	1	157	155.56	1.44	0.0133
6	2	98	98.10	-0.10	0.0001
6	3	50	51.34	-1.34	0.0348

SECOND ORDER OBSERVED AND EXPECTED MARGINS

ITEM I1	ITEM I2	J1	J2	OBS	EXPECT	OBS-EXP	((O-E)**2)/E
1	2	1	1	170	160.50	9.50	0.5620
1	2	1	2	14	20.33	-6.33	1.9707
1	2	1	3	1	1.26	-0.26	0.0538
1	2	2	1	61	65.74	-4.74	0.3471
1	2	2	2	38	30.58	7.42	1.8028
1	2	2	3	1	4.91	-3.91	3.1163
1	2	3	1	9	8.71	0.29	0.0097
1	2	3	2	3	7.19	-4.19	2.4402
1	2	3	3	5	1.86	3.14	5.3102
1	3	1	1	152	156.22	-4.22	0.1140
1	3	1	2	26	20.96	5.04	1.2095
1	3	1	3	5	3.71	1.29	0.4458
1	3	2	1	60	55.68	4.32	0.3358
1	3	2	2	30	33.03	-3.03	0.2782
1	3	2	3	9	11.86	-2.86	0.6883
1	3	3	1	8	6.44	1.56	0.3786

1	3	3	2	5	7.30	-2.30	0.7269
1	3	3	3	5	3.90	1.10	0.3121
1	4	1	1	149	143.66	5.34	0.1987
1	4	1	2	34	37.99	-3.99	0.4188
1	4	1	3	1	1.04	-0.04	0.0019
1	4	2	1	39	45.34	-6.34	0.8866
1	4	2	2	59	50.54	8.46	1.4144
1	4	2	3	3	5.67	-2.67	1.2609
1	4	3	1	6	4.92	1.08	0.2355
1	4	3	2	7	10.52	-3.52	1.1763
1	4	3	3	5	2.37	2.63	2.9044
1	5	1	1	154	156.59	-2.59	0.0429
1	5	1	2	23	19.50	3.50	0.6269
1	5	1	3	4	4.21	-0.21	0.0101
1	5	2	1	59	55.66	3.34	0.2007
1	5	2	2	29	33.06	-4.06	0.4986
1	5	2	3	12	11.51	0.49	0.0205
1	5	3	1	8	6.39	1.61	0.4036
1	5	3	2	7	7.65	-0.65	0.0547
1	5	3	3	3	3.54	-0.54	0.0828
1	6	1	1	114	115.25	-1.25	0.0135
1	6	1	2	54	50.35	3.65	0.2650
1	6	1	3	16	16.50	-0.50	0.0150
1	6	2	1	35	34.47	0.53	0.0081
1	6	2	2	36	39.57	-3.57	0.3223
1	6	2	3	29	27.19	1.81	0.1209
1	6	3	1	6	3.84	2.16	1.2139
1	6	3	2	7	6.92	0.08	0.0008
1	6	3	3	5	6.99	-1.99	0.5670
2	3	1	1	197	193.74	3.26	0.0549
2	3	1	2	32	32.48	-0.48	0.0070
2	3	1	3	9	5.67	3.33	1.9492
2	3	2	1	25	22.34	2.66	0.3164
2	3	2	2	24	24.81	-0.81	0.0264
2	3	2	3	6	10.19	-4.19	1.7201
2	3	3	1	1	0.83	0.17	0.0336
2	3	3	2	4	3.61	0.39	0.0413
2	3	3	3	3	3.48	-0.48	0.0663
2	4	1	1	185	175.27	9.73	0.5402
2	4	1	2	51	57.48	-6.48	0.7310
2	4	1	3	2	1.44	0.56	0.2221
2	4	2	1	10	16.82	-6.82	2.7669
2	4	2	2	44	36.15	7.85	1.7045
2	4	2	3	2	4.93	-2.93	1.7424
2	4	3	1	1	0.57	0.43	0.3345
2	4	3	2	2	4.77	-2.77	1.6112
2	4	3	3	5	2.67	2.33	2.0407
2	5	1	1	196	194.26	1.74	0.0157
2	5	1	2	31	30.36	0.64	0.0134
2	5	1	3	8	6.51	1.49	0.3417
2	5	2	1	24	22.16	1.84	0.1521
2	5	2	2	25	25.31	-0.31	0.0037
2	5	2	3	7	9.68	-2.68	0.7410

2	5	3	1	2	0.80	1.20	1.8013
2	5	3	2	3	4.15	-1.15	0.3192
2	5	3	3	3	2.95	0.05	0.0008
2	6	1	1	142	138.89	3.11	0.0697
2	6	1	2	66	49.84	-3.84	0.2111
2	6	1	3	30	24.69	5.31	1.1405
2	6	2	1	12	13.11	-1.11	0.0942
2	6	2	2	27	23.79	3.21	0.4317
2	6	2	3	15	20.81	-5.81	1.6214
2	6	3	1	1	0.56	0.44	0.3491
2	6	3	2	3	2.58	0.42	0.0694
2	6	3	3	4	4.84	-0.84	0.1470
3	4	1	1	175	173.22	1.78	0.0183
3	4	1	2	46	43.95	2.05	0.0959
3	4	1	3	1	0.46	0.54	0.6430
3	4	2	1	17	16.36	0.64	0.0248
3	4	2	2	41	40.86	0.14	0.0005
3	4	2	3	2	3.87	-1.87	0.9063
3	4	3	1	4	1.81	2.19	2.6444
3	4	3	2	9	12.95	-3.95	1.2044
3	4	3	3	6	4.64	1.36	0.3964
3	5	1	1	199	190.51	8.49	0.3787
3	5	1	2	18	20.78	-2.78	0.3730
3	5	1	3	4	3.49	0.51	0.0750
3	5	2	1	20	22.67	-2.67	0.3141
3	5	2	2	35	28.15	6.85	1.6671
3	5	2	3	5	9.48	-4.48	2.1199
3	5	3	1	2	2.62	-0.62	0.1467
3	5	3	2	7	10.49	-3.49	1.1620
3	5	3	3	10	6.04	3.96	2.5984
3	6	1	1	143	136.99	6.01	0.2634
3	6	1	2	62	62.08	-0.08	0.0001
3	6	1	3	17	17.84	-0.84	0.0398
3	6	2	1	9	12.83	-3.83	1.1450
3	6	2	2	34	26.50	7.50	2.1228
3	6	2	3	15	21.57	-6.57	2.0000
3	6	3	1	2	1.73	0.27	0.0417
3	6	3	2	1	7.00	-6.00	5.1471
3	6	3	3	16	10.60	5.40	2.7449
4	5	1	1	175	173.78	1.22	0.0086
4	5	1	2	15	14.65	0.35	0.0086
4	5	1	3	4	2.34	1.66	1.1766
4	5	2	1	46	43.73	2.27	0.1181
4	5	2	2	39	40.70	-1.70	0.0713
4	5	2	3	11	13.01	-2.01	0.3095
4	5	3	1	1	0.43	0.57	0.7620
4	5	3	2	5	4.66	0.34	0.0242
4	5	3	3	3	3.85	-0.85	0.1886
4	6	1	1	126	127.80	-1.80	0.0253
4	6	1	2	51	51.51	-0.51	0.0050
4	6	1	3	17	13.35	3.65	0.9968
4	6	2	1	27	24.89	2.11	0.1796
4	6	2	2	46	42.42	3.58	0.3017

4	6	2	3	24	31.10	-7.10	1.6196
4	6	3	1	1	0.37	0.63	1.0451
4	6	3	2	1	2.60	-1.60	0.9825
4	6	3	3	7	6.06	0.94	0.1452
5	6	1	1	141	137.32	3.68	0.0987
5	6	1	2	64	62.21	1.79	0.0517
5	6	1	3	17	17.69	-0.69	0.0272
5	6	2	1	12	11.63	0.37	0.0115
5	6	2	2	29	25.79	3.21	0.4001
5	6	2	3	17	22.40	-5.40	1.2998
5	6	3	1	1	2.10	-1.10	0.5794
5	6	3	2	4	7.27	-3.27	1.4718
5	6	3	3	14	9.76	4.24	1.8410

OBS	EXPECT	E(Z1/X)	SD1	CSCORE	TOTAL	RESPONSE PATTERN
96	84.711	-0.944	0.671	0.000	6	111111
1	2.579	-0.923	0.681	0.000	0	111191
1	0.001	-0.799	0.739	0.000	0	119991
1	0.000	-0.557	0.814	0.000	0	999991
14	16.743	-0.450	0.538	1.363	7	211111
27	23.316	-0.418	0.529	1.474	7	111112
2	3.256	-0.283	0.491	1.993	7	121111
3	1.341	-0.272	0.487	2.042	8	311111
4	3.867	-0.133	0.449	2.677	8	111113
9	8.264	-0.101	0.440	2.837	8	211112
4	6.569	-0.091	0.438	2.888	7	111211
3	2.130	-0.037	0.424	3.180	7	112111
1	1.346	-0.006	0.416	3.356	8	221111
1	1.689	0.011	0.412	3.455	7	111121
1	0.042	0.097	0.407	0.000	0	921112
1	0.087	0.102	0.419	0.000	0	112119
3	1.904	0.104	0.390	4.040	9	211113
6	3.397	0.136	0.383	4.251	8	211211
1	0.001	0.137	0.398	0.000	0	931111
1	0.043	0.150	0.430	0.000	0	111921
5	5.044	0.152	0.379	4.361	8	111212
2	1.753	0.193	0.371	4.654	8	112112
1	0.216	0.202	0.369	4.718	10	311113
1	0.987	0.215	0.367	4.818	8	211121
1	1.152	0.217	0.366	4.830	9	221112
2	1.478	0.230	0.364	4.929	8	111122
2	0.393	0.230	0.364	4.929	9	311211
4	3.503	0.330	0.347	5.724	9	211212
1	0.556	0.346	0.344	5.857	9	112113
1	0.003	0.355	0.467	0.000	0	219193
1	0.375	0.367	0.341	6.033	10	221113
5	0.707	0.391	0.337	6.244	9	221211
2	1.132	0.396	0.337	6.291	9	211122
3	0.923	0.402	0.336	6.342	8	111221
1	1.080	0.403	0.336	6.354	9	121212
1	0.345	0.435	0.331	6.635	8	112121
1	0.069	0.440	0.383	0.000	0	111293
1	0.003	0.464	0.411	0.000	0	212199

1	0.493	0.497	0.324	7.220	10	212113
1	0.065	0.504	0.323	7.280	10	111133
2	0.955	0.519	0.321	7.430	9	212211
1	0.520	0.520	0.351	0.000	0	921129
1	0.000	0.523	0.475	0.000	0	299913
1	0.010	0.524	0.321	7.481	9	211311
1	0.452	0.526	0.320	7.494	10	211123
1	0.031	0.527	0.370	0.000	0	229211
2	1.479	0.531	0.320	7.541	9	112212
2	0.045	0.536	0.319	7.591	10	113113
1	0.880	0.547	0.318	7.705	9	211221
4	1.031	0.548	0.318	7.717	10	221212
1	0.000	0.560	0.502	0.000	0	199292
1	0.402	0.578	0.316	8.010	10	222112
1	0.053	0.584	0.315	8.070	10	121132
2	0.534	0.588	0.315	8.108	9	112122
2	0.156	0.616	0.313	8.396	11	321212
1	0.689	0.650	0.311	8.744	10	112213
1	1.666	0.665	0.311	8.904	10	212212
1	0.490	0.667	0.311	8.920	11	221213
1	0.266	0.704	0.310	9.314	10	112123
2	0.632	0.726	0.310	9.534	10	122212
1	0.040	0.745	0.310	9.732	10	213121
1	0.616	0.752	0.310	9.809	10	121222
1	0.255	0.780	0.311	10.101	10	122122
1	0.910	0.781	0.311	10.107	11	212213
1	0.110	0.785	0.311	10.150	11	312122
2	0.901	0.808	0.312	10.382	11	211223
1	0.126	0.831	0.313	10.620	11	212132
1	0.000	0.843	0.492	0.000	0	129299
1	0.926	0.858	0.314	10.897	11	222212
1	0.124	0.864	0.315	10.958	11	121232
4	1.265	0.868	0.315	10.996	10	112222
1	0.936	0.885	0.316	11.172	11	221222
1	0.034	0.894	0.316	11.255	11	132212
1	0.337	0.928	0.318	11.515	10	122221
1	0.194	0.921	0.318	11.530	12	211233
1	0.000	0.940	0.567	0.000	0	299299
1	0.885	0.990	0.322	12.199	11	112223
2	2.260	1.007	0.323	12.359	11	212222
1	0.669	1.008	0.324	12.375	12	221223
1	0.000	1.043	0.349	11.559	0	393131
2	0.647	1.062	0.327	12.878	11	222221
1	1.064	1.074	0.328	12.989	11	122222
1	0.468	1.079	0.329	13.038	12	312222
1	0.235	1.112	0.331	13.347	12	112233
1	1.875	1.136	0.333	13.562	12	212223
1	0.101	1.143	0.359	0.000	0	292222
1	0.216	1.178	0.336	13.933	12	113223
7	2.537	1.225	0.340	14.352	12	222222
1	0.117	1.275	0.359	0.000	0	222229
1	0.612	1.306	0.347	15.031	13	322222
1	0.193	1.331	0.350	15.242	13	213232

4	2.763	1.370	0.353	15.555	13	222223
1	0.347	1.417	0.357	15.926	13	123223
3	0.240	1.484	0.363	16.445	14	213233
1	1.154	1.518	0.366	16.703	14	222233
1	0.094	1.525	0.366	16.751	14	332222
1	0.026	1.598	0.429	0.000	0	222933
3	0.738	1.761	0.382	18.438	15	223233
1	0.070	1.868	0.389	19.157	14	123323
1	0.032	1.875	0.389	19.206	15	313323
1	0.028	1.995	0.397	19.982	15	332322
1	0.511	2.081	0.424	20.520	15	223323
1	0.175	2.275	0.424	21.654	16	232333
1	0.173	2.520	0.459	22.919	17	333323
2	0.288	2.786	0.507	24.067	18	333333
311	216.624					

10.2- Two Factors

Input format for channel 5:

For the data we are considering in the first line we have ENVIRONMENT 6 ITEMS

On the second line we have may set N = 311 P = 6 NQ = 8.

On the third line we may set C(1)=3, C(2)=3, C(3)=3, C(4)=3, C(5)=3, C(6)=3

On the fourth line we may have NFAC = 2, INPUT = 1, FREQ = 1, DISPLAY = 0, MTER = 50, LOUT = 1, ERRC = 0.00001.

From the fifth line we display individual response patterns.

File: POLY.INP

```
ENVIRONMENT 6 ITEMS
    311 6 48
    3 3 3 3 3 3
    2 1 1 0 50 1 0.00001
    105
    2 113113
    1 393131
    1 213121
    1 111133
    . ......
    . ......
    . ......
    96 111111
    3 211191
    1 211123
```

Input format for channel 3

Since for the previous number of iterations the convergence was not obtained, we run TWOMISS again renaming the output of channel 8 (LI8.OUT) as POLY3.INP and using it as the input for channel 3.

ENVIRONMENT 6 ITEMS FACTORS=2 QUAD. POINTS = 8

-0.789	1.372	0.767
-2.761	0.690	1.556
-3.702	3.792	1.376
-5.391	1.176	2.718
-2.423	1.478	3.244
-13.418	0.969	10.185
-1.441	2.789	1.939
-6.625	2.295	4.551
-2.796	1.884	3.619
-5.775	1.889	5.270
-0.344	0.940	1.040
-2.045	0.907	2.855

Output from channel 7

The output from channel 7 is called **POLY.LI7**, and is as follows

*** PROGRAM POLYMISS ***

MAXIMUM LIKELIHOOD ESTIMATION OF A 2 FACTOR LOGIT/PROBIT MODEL
FOR POLYTOMOUS DATA

ENVIRONMENT 6 ITEMS

NUMBER OF OBSERVED VARIABLES = 6

NUMBER OF CASES SAMPLED = 311

NUMBER OF DIFFERENT RESPONSE PATTERNS = 105

NUMBER OF QUADRATURE POINTS USED = 8

MAXIMUM NUMBER OF ITERATIONS PERMITTED = 20

CONVERGENCE TOLERANCE FOR THE RELATIVE LIKELIHOOD VALUE = 0.00001000

A LIKELIHOOD RATIO TEST OF OBSERVED AND EXPECTED
FREQUENCIES OF RESPONSE VECTORS IS TO BE CARRIED OUT

ASYMPTOTIC STANDARD DEVIATIONS ARE OBTAINED FROM THE INVERSE OF THE OBSERVED SECOND DERIVATIVE MATRIX

MARGINS

CATEGORY

ITEM 1

1	0.598
2	0.331
3	0.058
MISSING	0.013

ITEM 2

1	0.772
2	0.183
3	0.026
MISSING	0.019

ITEM 3

1	0.717
2	0.196
3	0.061
MISSING	0.026

ITEM 4

1	0.633
2	0.322
3	0.029
MISSING	0.016

ITEM 5

1	0.717
2	0.193
3	0.061
MISSING	0.029

ITEM 6

1	0.508
2	0.312
3	0.161
MISSING	0.019

INITIAL ESTIMATES OF ITEM PARAMETERS

	CATEGORY	A(0,I,J)	A(1,I,J)	A(2,I,J)
ITEM 1				
	2	-0.789	1.372	0.767
	3	-2.761	0.690	1.556
ITEM 2				
	2	-3.702	3.792	1.376
	3	-5.391	1.176	2.718

ITEM 3

2	-2.423	1.478	3.244
3	-13.418	0.969	10.185

ITEM 4

2	-1.441	2.789	1.939
3	-6.625	2.295	4.551

ITEM 5

2	-2.796	1.884	3.619
3	-5.775	1.889	5.270

ITEM 6

2	-0.344	0.940	1.040
3	-2.045	0.907	2.855

ITER	PROP	LOGLIKELIHOOD
1	0.00006	-1334.98172
2	0.00005	-1334.96888
.	.	.
.	.	.
13	0.00001	-1334.89994
14	0.00001	-1334.89725
15	0.00001	-1334.89487

*** ITERATIONS FINISHED ***

NUMBER OF ITERATIONS IS 15

% OF G-SQUARE EXPLAINED 54.5416

LOGLIKELIHOOD VALUE -1334.8949

LIKELIHOOD RATIO STAT. 214.6738

DEGREES OF FREEDOM -19

ITEM	CATEGORY	PARAMETER	DERIVATIVE
1	2	1	-0.01050002
1	2	2	-0.01024560
1	2	3	-0.01430309
1	3	1	-0.01054095
1	3	2	-0.00112718
1	3	3	-0.01835918
2	2	1	-0.01080448
2	2	2	-0.00734269
2	2	3	-0.01832090
2	3	1	-0.00915167
2	3	2	-0.00340594

2	3	3	-0.01699027
3	2	1	-0.00000579
3	2	2	0.00000185
3	2	3	-0.00000257
3	3	1	0.02434294
3	3	2	1.11965727
3	3	3	0.05484791
4	2	1	-0.03187189
4	2	2	-0.02679377
4	2	3	-0.05496643
4	3	1	-0.02845926
4	3	2	-0.01485578
4	3	3	-0.04892992
5	2	1	0.02353714
5	2	2	0.04038334
5	2	3	0.02962763
5	3	1	0.02243193
5	3	2	0.03100196
5	3	3	0.03114831
6	2	1	-0.00208683
6	2	2	0.02285534
6	2	3	-0.01557234
6	3	1	-0.00897947
6	3	2	0.01138884
6	3	3	-0.01519632

MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS AND STANDARD DEVIATIONS

ITEM	CAT	A(0,I,J)	S.D	A(1,I,J)	S.D	A(2,I,J)	S.D	PHI(I,J)
1	2	-0.80	0.20	1.38	0.35	0.74	0.40	0.30
1	3	-2.76	0.39	0.63	0.50	1.54	0.68	0.04
2	2	-3.84	1.78	3.97	2.47	1.31	0.96	0.02
2	3	-5.34	1.07	0.90	1.52	2.65	0.98	0.00
3	2	-2.43	0.66	1.49	0.68	3.24	1.43	0.08
3	3	-13.42	519.49	0.97	1.36	10.19	317.51	0.00
4	2	-1.43	0.42	2.73	0.84	1.84	0.72	0.19
4	3	-6.47	1.53	2.06	1.87	4.38	1.23	0.00
5	2	-2.84	0.79	1.94	0.87	3.66	1.38	0.06
5	3	-5.81	1.48	1.97	1.34	5.30	1.66	0.00
6	2	-0.35	0.19	0.96	0.30	1.02	0.45	0.38
6	3	-2.03	0.44	0.94	0.61	2.81	0.77	0.07

FIRST ORDER OBSERVED AND EXPECTED MARGINS

ITEM I1	CAT J1	OBS	EXPECT	OBS-EXP	((O-E)**2)/E
1	1	186	185.23	0.77	0.0032
1	2	103	102.77	0.23	0.0005
1	3	18	19.00	-1.00	0.0529
2	1	240	238.94	1.06	0.0047
2	2	57	57.13	-0.13	0.0003
2	3	8	8.93	-0.93	0.0965
3	1	223	218.50	4.50	0.0926
3	2	61	61.61	-0.61	0.0061
3	3	19	22.89	-3.89	0.6601
4	1	197	195.99	1.01	0.0052
4	2	100	99.67	0.33	0.0011
4	3	9	10.34	-1.34	0.1735
5	1	223	218.81	4.19	0.0802
5	2	60	61.83	-1.83	0.0540
5	3	19	21.36	-2.36	0.2612
6	1	158	154.72	3.28	0.0697
6	2	97	96.09	0.91	0.0087
6	3	50	54.20	-4.20	0.3251

SECOND ORDER OBSERVED AND EXPECTED MARGINS

ITEM I1	ITEM I2	J1	J2	OBS	EXPECT	OBS-EXP	((O-E)**2)/E
1	2	1	1	170	164.60	5.40	0.1772
1	2	1	2	14	14.15	-0.15	0.0016
1	2	1	3	1	2.90	-1.90	1.2483
1	2	2	1	61	58.98	2.02	0.0694
1	2	2	2	38	38.14	-0.14	0.0005
1	2	2	3	1	3.67	-2.67	1.9433
1	2	3	1	9	12.29	-3.29	0.8820
1	2	3	2	3	4.11	-1.11	0.2976
1	2	3	3	5	2.24	2.76	3.4072
1	3	1	1	152	151.72	0.28	0.0005
1	3	1	2	26	21.24	4.76	1.0669
1	3	1	3	5	7.50	-2.50	0.8358
1	3	2	1	60	56.86	3.14	0.1729
1	3	2	2	30	33.76	-3.76	0.4194
1	3	2	3	9	9.50	-0.50	0.0261
1	3	3	1	8	7.11	0.89	0.1121
1	3	3	2	5	5.82	-0.82	0.1147
1	3	3	3	5	5.59	-0.59	0.0622
1	4	1	1	149	145.26	3.74	0.0963
1	4	1	2	34	34.13	-0.13	0.0005
1	4	1	3	1	2.86	-1.86	1.2131
1	4	2	1	39	40.34	-1.34	0.0444
1	4	2	2	59	56.21	2.79	0.1388
1	4	2	3	3	4.57	-1.57	0.5409
1	4	3	1	6	7.87	-1.87	0.4449
1	4	3	2	7	8.06	-1.06	0.1385
1	4	3	3	5	2.77	2.23	1.7957
1	5	1	1	154	153.25	0.75	0.0037

1	5	1	2	23	20.42	2.58	0.3260
1	5	1	3	4	6.20	-2.20	0.7800
1	5	2	1	59	55.42	3.58	0.2312
1	5	2	2	29	33.84	-4.84	0.6911
1	5	2	3	12	10.54	1.46	0.2025
1	5	3	1	8	7.33	0.67	0.0617
1	5	3	2	7	6.78	0.22	0.0074
1	5	3	3	3	4.35	-1.35	0.4187
1	6	1	1	115	113.45	1.55	0.0212
1	6	1	2	53	47.69	5.31	0.5911
1	6	1	3	16	20.51	-4.51	0.9930
1	6	2	1	35	34.43	0.57	0.0094
1	6	2	2	36	41.61	-5.61	0.7566
1	6	2	3	29	24.74	4.26	0.7320
1	6	3	1	6	4.84	1.16	0.2757
1	6	3	2	7	5.55	1.45	0.3794
1	6	3	3	5	8.24	-3.24	1.2759
2	3	1	1	197	188.98	8.02	0.3399
2	3	1	2	32	32.01	-0.01	0.0000
2	3	1	3	9	11.80	-2.80	0.6644
2	3	2	1	25	24.09	0.91	0.0341
2	3	2	2	24	25.74	-1.74	0.1177
2	3	2	3	6	5.83	0.17	0.0052
2	3	3	1	1	1.21	-0.21	0.0356
2	3	3	2	4	2.67	1.33	0.6600
2	3	3	3	3	4.82	-1.82	0.6867
2	4	1	1	185	181.04	3.96	0.0867
2	4	1	2	51	49.49	1.51	0.0463
2	4	1	3	2	4.58	-2.58	1.4496
2	4	2	1	10	9.06	0.94	0.0978
2	4	2	2	44	44.35	-0.35	0.0028
2	4	2	3	2	2.80	-0.80	0.2280
2	4	3	1	1	2.11	-1.11	0.5836
2	4	3	2	2	3.91	-1.91	0.9324
2	4	3	3	5	2.77	2.23	1.8050
2	5	1	1	196	190.92	5.08	0.1350
2	5	1	2	31	31.36	-0.36	0.0041
2	5	1	3	8	9.74	-1.74	0.3119
2	5	2	1	24	22.35	1.65	0.1215
2	5	2	2	25	25.58	-0.58	0.0131
2	5	2	3	7	7.55	-0.55	0.0394
2	5	3	1	2	1.31	0.69	0.3582
2	5	3	2	3	3.69	-0.69	0.1304
2	5	3	3	3	3.66	-0.66	0.1196
2	6	1	1	142	138.43	3.57	0.0919
2	6	1	2	66	65.03	0.97	0.0144
2	6	1	3	30	30.87	-0.87	0.0243
2	6	2	1	12	12.31	-0.31	0.0078
2	6	2	2	27	27.31	-0.31	0.0035
2	6	2	3	15	16.41	-1.41	0.1205
2	6	3	1	2	0.99	1.01	1.0384
2	6	3	2	2	1.89	0.11	0.0066
2	6	3	3	4	5.88	-1.88	0.6011

3	4	1	1	175	167.93	7.07	0.2980
3	4	1	2	46	46.41	-0.41	0.0036
3	4	1	3	1	0.65	0.35	0.1863
3	4	2	1	17	17.13	-0.13	0.0009
3	4	2	2	41	40.36	0.64	0.0101
3	4	2	3	2	3.13	-1.13	0.4103
3	4	3	1	4	5.90	-1.90	0.6099
3	4	3	2	9	10.33	-1.33	0.1724
3	4	3	3	6	6.29	-0.29	0.0132
3	5	1	1	199	190.12	8.88	0.4148
3	5	1	2	18	19.59	-1.59	0.1288
3	5	1	3	4	2.47	1.53	0.9475
3	5	2	1	20	21.28	-1.28	0.0770
3	5	2	2	35	30.08	4.92	0.8045
3	5	2	3	5	8.47	-3.47	1.4204
3	5	3	1	2	1.78	0.22	0.0264
3	5	3	2	7	10.57	-3.57	1.2041
3	5	3	3	10	9.87	0.13	0.0016
3	6	1	1	143	135.78	7.22	0.3836
3	6	1	2	62	62.98	-0.98	0.0151
3	6	1	3	17	15.53	1.47	0.1397
3	6	2	1	10	13.25	-3.25	0.7971
3	6	2	2	33	26.56	6.44	1.56539
3	6	2	3	15	20.62	-5.62	1.5308
3	6	3	1	2	1.70	0.30	0.0514
3	6	3	2	1	4.08	-3.08	2.3278
3	6	3	3	16	16.66	-0.66	0.0260
4	5	1	1	175	170.34	4.66	0.1274
4	5	1	2	15	15.63	-0.63	0.0252
4	5	1	3	4	4.35	-0.35	0.0280
4	5	2	1	46	44.20	1.80	0.0736
4	5	2	2	39	40.76	-1.76	0.0759
4	5	2	3	11	11.83	-0.83	0.0586
4	5	3	1	1	0.76	0.24	0.0784
4	5	3	2	5	4.45	0.55	0.0690
4	5	3	3	3	4.84	-1.84	0.6980
4	6	1	1	126	125.81	0.19	0.0003
4	6	1	2	51	48.49	2.51	0.1303
4	6	1	3	17	17.91	-0.91	0.0463
4	6	2	1	28	25.71	2.29	0.2047
4	6	2	2	45	44.14	0.86	0.0168
4	6	2	3	24	27.90	-3.90	0.5459
4	6	3	1	1	0.71	0.29	0.1165
4	6	3	2	1	1.92	-0.92	0.4375
4	6	3	3	7	7.51	-0.51	0.0350
5	6	1	1	142	136.48	5.52	0.2233
5	6	1	2	63	62.14	0.86	0.0118
5	6	1	3	17	15.97	1.03	0.0667
5	6	2	1	12	11.68	0.32	0.0089
5	6	2	2	29	25.69	3.31	0.4252
5	6	2	3	17	23.26	-6.26	1.6853
5	6	3	1	1	2.08	-1.08	0.5618
5	6	3	2	4	5.47	-1.47	0.3943

5 6 3 3 14 13.40 0.60 0.0268

OBS	EXPECT	E(Z1/X)	SD1	E(Z2/X)	SD2	TOTAL	RESPONSE PATTERN
2	0.430	-1.667	0.547	1.635	0.050	10	113113
1	0.001	-1.104	0.609	1.644	0.094	0	393131
1	0.053	-0.907	0.566	1.632	0.071	10	213121
1	0.135	-0.784	0.776	1.051	0.551	10	111133
4	4.883	-0.687	0.666	0.349	0.479	8	111113
96	84.683	-0.664	0.817	-0.675	0.804	6	111111
1	2.581	-0.659	0.813	-0.651	0.813	0	111191
1	0.409	-0.632	0.610	0.535	0.376	10	311113
3	2.247	-0.614	0.734	0.003	0.643	8	311111
1	0.003	-0.586	0.679	0.320	0.494	0	931111
1	0.001	-0.580	0.833	-0.554	0.847	0	119991
1	0.964	-0.535	0.646	0.667	0.407	9	112113
3	2.747	-0.525	0.618	0.364	0.456	7	112111
1	0.111	-0.457	0.619	0.436	0.442	0	112119
1	1.730	-0.420	0.593	0.385	0.433	7	111121
1	0.206	-0.415	0.561	1.780	0.383	15	313323
1	0.566	-0.379	0.754	1.106	0.550	10	112123
1	0.549	-0.343	0.505	1.637	0.047	12	113223
2	0.651	-0.329	0.716	2.772	0.373	18	333333
2	1.936	-0.320	0.582	0.423	0.398	8	112112
27	23.200	-0.319	0.730	-0.311	0.708	7	111112
1	0.000	-0.313	0.943	-0.457	0.856	0	999991
1	0.174	-0.281	0.685	2.443	0.546	17	333323
3	1.468	-0.265	0.600	0.303	0.480	9	211113
1	0.003	-0.246	0.653	0.749	0.646	0	219193
1	0.512	-0.234	0.662	0.652	0.381	8	112121
2	1.368	-0.213	0.588	0.438	0.380	8	111122
1	0.040	-0.210	0.658	0.395	0.419	0	111921
1	0.026	-0.065	0.591	0.447	0.362	9	211311
1	0.454	-0.060	0.625	0.594	0.328	10	212113
14	15.503	-0.036	0.697	-0.557	0.727	7	211111
1	0.307	-0.035	0.576	1.399	0.452	12	112233
1	0.740	-0.034	0.593	0.343	0.444	8	211121
1	0.134	0.020	0.655	0.858	0.502	11	312122
2	0.647	0.041	0.632	0.670	0.382	9	112122
1	0.002	0.052	0.599	0.562	0.487	0	212199
1	0.389	0.071	0.600	0.597	0.318	10	211123
3	0.507	0.074	0.557	1.650	0.128	14	213233
1	0.011	0.076	0.556	1.623	0.121	15	332322
1	0.144	0.082	0.554	0.637	0.054	13	213232
1	0.108	0.129	0.590	0.890	0.514	11	212132
9	7.508	0.136	0.628	-0.311	0.650	8	211112
2	0.342	0.144	0.594	0.066	0.574	9	311211
1	0.048	0.167	0.558	0.461	0.420	0	111293
2	0.832	0.191	0.550	0.423	0.369	9	211122
1	0.855	0.252	0.521	0.958	0.534	11	112223
4	5.943	0.261	0.601	-0.368	0.642	7	111211
1	0.000	0.266	0.784	0.427	0.522	0	299913
1	0.061	0.286	0.501	1.637	0.056	16	232333
1	0.555	0.291	0.493	0.590	0.290	10	112213

1	0.018	0.292	0.491	0.580	0.281	10	132211
3	0.716	0.295	0.514	0.409	0.377	8	111221
1	0.036	0.344	0.502	1.404	0.449	14	332222
5	4.271	0.360	0.568	-0.185	0.614	8	111212
1	0.104	0.379	0.439	0.757	0.443	12	211233
2	1.171	0.382	0.457	0.453	0.326	9	112212
1	0.293	0.435	0.424	0.795	0.468	12	312222
1	1.279	0.466	0.455	0.966	0.537	12	212223
1	0.662	0.478	0.329	0.565	0.243	11	212213
4	1.191	0.480	0.340	0.619	0.308	10	112222
2	0.795	0.482	0.434	0.392	0.387	9	212211
2	0.697	0.505	0.310	0.571	0.248	11	211223
1	0.067	0.526	0.436	1.707	0.279	14	123323
1	0.787	0.528	0.398	0.417	0.359	9	211221
6	3.573	0.551	0.555	-0.381	0.629	8	211211
1	0.017	0.553	0.321	0.496	0.259	10	121132
1	1.432	0.560	0.361	0.449	0.319	10	212212
2	1.774	0.602	0.368	0.618	0.319	11	212222
1	0.157	0.633	0.358	0.529	0.254	10	122122
4	3.376	0.643	0.550	-0.166	0.628	9	211212
1	0.226	0.648	0.454	1.640	0.077	13	123223
1	0.000	0.652	0.681	0.181	0.704	0	199292
1	0.000	0.758	0.500	0.270	0.488	0	921129
1	0.127	0.762	0.532	0.071	0.579	10	221113
2	1.430	0.800	0.589	-0.892	0.702	7	121111
1	0.063	0.870	0.529	0.472	0.331	11	121232
1	0.282	0.872	0.571	1.834	0.438	15	223323
1	0.253	0.873	0.533	0.183	0.522	10	222112
1	0.323	0.976	0.569	0.500	0.330	10	122221
1	0.029	1.021	0.602	-0.652	0.674	0	921112
2	0.587	1.034	0.588	0.167	0.523	10	122212
1	0.000	1.044	0.746	0.425	0.819	0	299299
1	1.222	1.087	0.590	-0.906	0.694	8	221111
3	0.886	1.093	0.642	1.734	0.324	15	223233
1	0.676	1.099	0.610	0.179	0.519	10	121222
1	0.000	1.122	0.597	0.175	0.845	0	129299
1	1.145	1.171	0.579	-0.663	0.670	9	221112
1	0.460	1.171	0.597	-0.013	0.559	11	221213
2	0.157	1.186	0.579	-0.132	0.574	11	321212
1	1.126	1.243	0.598	0.552	0.351	11	122222
1	0.636	1.251	0.612	0.473	0.365	12	221223
1	0.026	1.301	0.673	1.408	0.447	0	222933
1	0.432	1.345	0.592	0.748	0.464	13	322222
1	1.784	1.357	0.535	-0.500	0.596	9	121212
1	1.328	1.365	0.664	1.382	0.464	14	222233
1	0.150	1.401	0.715	0.574	0.387	0	292222
1	1.527	1.410	0.662	0.061	0.549	11	222212
2	1.088	1.417	0.603	0.485	0.371	11	222221
1	0.081	1.467	0.525	-0.547	0.634	0	229211
4	3.698	1.509	0.617	0.924	0.538	13	222223
5	2.467	1.518	0.487	-0.674	0.584	9	221211
1	2.022	1.521	0.677	0.090	0.545	11	221222
4	4.207	1.580	0.497	-0.531	0.508	10	221212

1	0.207	1.586	0.609	0.674	0.483	0	222229
7	5.715	1.668	0.592	0.548	0.391	12	222222

311 223.543

Acknowledgment: We are grateful to Department of Statistics, London School of Economics, Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPQ) and Fundação de Amparo a Pesquisa do Rio Grande do Sul (FAPERGS). We are grateful to David Bartholomew and Jane Galbraith for encouragement and help with this documentation. We thank Shing On Leung for giving us copies of his program suite, which allowed us to try his rather different approach to the numerical analysis involved in the estimation of some of the models.

11- References

Albanese, M. T. (1990). *Latent variable models for binary response data.* University of London, London. Ph.D. thesis.

Bartholomew, D.G.(1980). Factor analysis for categorical data. *J. Roy. Statist. Soc., B,* **42**, 293-321.

Bartholomew, D.J.(1987). *Latent variable models and factor analysis.* London: Charles Griffin & Company Ltd.

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

Knott, M. and Albanese, M.T. (1993). Conditional distributions of a latent variable and scoring for binary data. *Brazilian Journal of Probability and Statistics.* To appear.

Leung S. O. (1990). *Estimation and application of latent variable models in categorical data analysis.* External Ph.D. thesis of the University of London.

Shea, B. L. (1984). *FACONE: A computer program for fitting the logit latent model by maximum likelihood.* Department of Statistics, London School of Economics.

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

ANNEX 1

USER INSTRUCTIONS FOR THE COUNT PROGRAM

COUNT is a program written in standard FORTRAN 77 for micro-computers, but it can be easily implemented on large size computers.

COUNT is designed to provide the frequency distribution of the responses given by N individuals to P items, which take values 1, 2, 3, 4, 5 or 9 (strongly agree, agree, neither agree or disagree, disagree, strongly disagree and missing, for example). Missing, coded as 9, means that the response was 'don't know' or it was not recorded. The vector of responses of each individual is called **response pattern**.

The initial parameters, sample size (N) and number of items (P) were set equal to 1500 and 20, respectively. They are defined on the third line of the COUNT program and they are easily changed to any other value.

1- Description of the Input Channel

Input format for channel 5 (filename: COUNT.INP)

The file for this channel contains all the control parameters and the data set as described below. Control parameters are read in FREE format, but the data set is read in FIXED format.

Line 1:

Title: Title of data set (at most 70 characters).

Line 2:

N, P

N : number of individuals in the sample
P : number of observed variables (items)

Reading response patterns

Line 3,4,...,N+2: Format(20I1)

```
IRESP(1,1)  IRESP(1,2)  ...  IRESP(1,P)
IRESP(2,1)  IRESP(2,2)  ...  IRESP(2,P)
...         ...         ...  ...
IRESP(N,1)  IRESP(N,2)  ...  IRESP(N,P)
```

where IRESP(L,I) = 1, 2, 3, 4, 5 or 9 is the response of individual L, L=1,2,...,N to item I, I=1,2,...,P.

2- Description of the Output Channel

Output format for channel 7 (filename: COUNT.OUT)

Line 1:

Title

Line 2:

Sample size = N

Number of items = P

Line 3:

Number of different response patterns = NR

Line 4 to N+3:

NR different response patterns of the N individuals according to the following format:

Columns 1 to 4: Format(I4)

RL(L), observed frequency of the response pattern (IRESP(L,I), I=1,P)

Columns 6 to P+5: Format(20I1)

Response pattern (IRESP(L,I),I=1,P)

that is,

RL(1)	IRESP(1,1)	IRESP(1,2)	... IRESP(1,P)
RL(2)	IRESP(2,1)	IRESP(2,2)	... IRESP(2,P)
....	....	....	... .
RL(NR)	IRESP(NR,1)	IRESP(NR,2)	... IRESP(NR,P)

where RL(L) is the observed frequency of the response pattern IRESP(L,I) for L=1,2,...,NR and I=1,2,...,P.

3- Example

Consider a set of simulated data of 11 individuals who have answered 1, 2 3 or 9 (agree, neither agree or disagree, disagree or missing) to each of 3 items.

The **input file** (channel 5) is called **COUNT.INP**, which has the following format:

Simulated data (Line 1, at most 70 characters)
11 3 (Line 2, N and P, free format)
123 (Line 3, IRESP(1,I), I=1,2,3, columns 1 to 3)
112 ...
212
333
912
112 ...
222
313
123
222 ...
222 (Line 13, IRESP(11,I), I=1,2,3, columns 1 to 3)

Then running the program **COUNT** , using as input file COUNT.INP, the frequency distribution of the response patterns are stored in the **output file COUNT.OUT** in the following format:

Simulated data
Sample size = 11 Number of items = 3
Number of different response patterns = 7
2 123
3 222
2 112
1 212
1 333
1 912
1 313

Thus the first four columns from line 4 correspond to the observed frequencies of the 7 different response patterns. After few changes on the first three lines, this file may be used as the input file for the POLYMISS program.

ANNEX 2

USER INSTRUCTIONS FOR THE ZSCORE PROGRAM

ZSCORE is a program written in standard FORTRAN 77 for micro-computers, but it can be easily implemented on large size computers.

ZSCORE is designed to provide the scoring of the latent variable for individual response patterns as given in the input file. As the output from the POLYMISS program displays the response patterns in increasing order according to the conditional mean, ZSCORE needs to be used when the latent scores for response patterns in the input file order are the input data in further analysis.

The initial parameters, sample size (N) and number of items (P) were set equal to 1500 and 20, respectively. They are defined on the third line of the ZSCORE program and they are easily changed to any other value.

1- Description of the Input Channels

1.1 - Input format for channel 5 (filename: ZSCORE.INP)

The file for this channel contains all the control parameters and the data set as described below. Control parameters are read in FREE format, but the data set is read in FIXED format.

<u>Line 1:</u>

TITLE: Title of data set (at most 70 characters).

<u>Line 2:</u>

MODEL, NFAC, N, P, NQ

MODEL : 1, 2 or 3
NFAC : 1 or 2 (number of factors)
N : number of individuals in the sample
P : number of observed variables (items)
NQ : number of quadrature points (8, 16, 24, 32, 48)

Reading response patterns

<u>Line 3,4,...,N+2:</u> Format(20I1)

```
IRESP(1,1)  IRESP(1,2)  ... IRESP(1,P)
IRESP(2,1)  IRESP(2,2)  ... IRESP(2,P)
...         ...         ... ...
IRESP(N,1)  IRESP(N,2)  ... IRESP(N,P)
```

where IRESP(L,I) = 1, 2, 3, 4, 5 or 9 (strongly agree, agree, neither agree or disagree, disagree, strongly disagree or missing for example) is the response of individual L, L=1,2,...,N to item I, I=1,2,...,P.

1.2 - Input format for channel 3 (filename: ZSCORE3.INP)

ZSCORE3.INP contains the parameter estimates after fitting the logit-probit model. They are read in FREE format and we can use the output estimates stored in channel 8 (LI8.OUT) as the input in channel 3.

Line 1:

Title or first line from channel 8.

Models 1 or 2

Line 2,3,...,P+1: Free format

Each line corresponds to the parameter estimates $\hat{a}_{0i(s)}$ and $\hat{a}_{1i(s)}$ (and $\hat{a}_{2i(s)}$ for two factors) of item $i, i = 1, 2..., p$ and $s = 2, ..., c_i$.

One-Factor

$$
\begin{array}{ll}
\hat{a}_{01(2)} & \hat{a}_{11(2)} \\
\hat{a}_{01(3)} & \hat{a}_{11(3)} \\
. & . \\
\hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} \\
\hat{a}_{02(2)} & \hat{a}_{12(2)} \\
\hat{a}_{02(3)} & \hat{a}_{12(3)} \\
. & . \\
\hat{a}_{02(c(2))} & \hat{a}_{12(c(2))} \\
. & . \\
\hat{a}_{0p(2)} & \hat{a}_{1p(2)} \\
\hat{a}_{0p(3)} & \hat{a}_{1p(3)} \\
. & . \\
. & . \\
\hat{a}_{0p(c(p))} & \hat{a}_{1p(c(p))}
\end{array}
$$

Two-Factor

$$
\begin{array}{lll}
\hat{a}_{01(2)} & \hat{a}_{11(2)} & \hat{a}_{21(2)} \\
\hat{a}_{01(3)} & \hat{a}_{11(3)} & \hat{a}_{21(3)} \\
. & . & . \\
\hat{a}_{01(c(1))} & \hat{a}_{11(c(1))} & \hat{a}_{21(c(1))} \\
\hat{a}_{02(2)} & \hat{a}_{12(2)} & \hat{a}_{22(2)} \\
\hat{a}_{02(3)} & \hat{a}_{12(3)} & \hat{a}_{22(3)} \\
. & . & . \\
\hat{a}_{02(c(2))} & \hat{a}_{12(c(2))} & \hat{a}_{22(c(2))} \\
. & . & . \\
\hat{a}_{0p(2)} & \hat{a}_{1p(2)} & \hat{a}_{2p(2)} \\
\hat{a}_{0p(3)} & \hat{a}_{1p(3)} & \hat{a}_{2p(3)} \\
. & . & . \\
. & . & . \\
\hat{a}_{0p(c(p))} & \hat{a}_{1p(c(p))} & \hat{a}_{2p(c(p))}
\end{array}
$$

2- Description of the Output Channel

Output from channel 7 (filename: ZSCORE.OUT)

If the one factor logit-probit model has been fitted then ZSCORE provides for each individual response pattern the conditional mean $E(Z|x)$ and the standard deviation $SD(E(Z|x))$, the component score, and total number of positive responses.

If two factors logit-probit model has been fitted then ZSCORE provides for each individual response pattern the conditional mean for the first factor $E(Z_1|x)$ and $SD(E(Z_1|x))$, the conditional mean for the second factor $E(Z_2|x)$ and $SD(E(Z_2|x))$, and the total number of positive responses.

Publicações do Instituto de Matemática da UFRGS
Cadernos de Matemática e Estatística

<u>Série A: Trabalho de Pesquisa</u>

1. Marcos Sebastiani Artecona - Transformation des Singularités - MAR/89

2. Jaime Bruck Ripoll - On a Theorem of R. Langevin about Curvature and Complex Singularities - MAR/89

3. Eduardo Cisneros, Miguel Ferrero e Maria Inés Gonzales - Prime Ideals of Skew Polynomial Rings and Skew Laurent Polynomial Rings - ABR/89

4. Oclide José Dotto - ε-Dilations - JUN/89

5. Jaime Bruck Ripoll - A Characterization of Helicoids - JUN/89

6. Mark Thompson e V. B. Moscatelli - Asymptotic Distribution of Liusternik-Schnirelman Eigenvalues for Elliptic Nonlinear Operators - JUL/89

7. Mark Thompson - The Formula of Weyl for Regions with a Self-Similar Fractal Boundary - JUL/89

8. Jaime Bruck Ripoll - A Note on Compact Surfaces with Non Zero Constant Mean Curvature - OUT/89

9. Jaime Bruck Ripoll - Compact ε-Convex Hypersurfaces - NOV/89

10. Jandyra Maria G. Fachel - Coeficientes de Correlação Tipo-Contigência - JAN/90

11. Jandyra Maria G. Fachel - The Probality of Ocurrence of Heywood Cases - JAN/90

12. Jandyra Maria G. Fachel - Heywood Cases in Unrestricted Factor Analysis - JAN/90

13. Julio Cesar R. Claeyssen e Tereza Tsukazan de Ruiz - Dynamical Solutions of Linear Matrix Diferential Equations - JAN/90

14. Maria T. Albanese - Behaviour of de Likelihood in Latent Analysis of Binary Data - ABR/91

15. Maria T. Albanese - Measuremente of the Latent Trait Analysis of Binary Data - ABR/91

16. Maria Teresinha Albanese - Adequacy of the Asymptotic Variance-Covariance Matrix Using Bootstrap Jackknife Techniques in Latent Trait Analysis of Binary Data - ABR/91

17. Maria Teresinha Albanese - Latent Variable Models for Binary Response - ABR/91

18. Mark Thompson - Kinematic Dynamo in Random Flows - DEZ/90

19. Jaime Bruck Ripoll e Marcos Sebastiani Artecona- The Generalized Map and Applications - AGO/91

20. Jaime Bruck Ripoll, Suzana Fornari e Katia Frensel - Hypersurfaces with Constant Mean Curvature in the Complex Hyperbolic Space - AGO/91

21. Suzana Fornari e Jaime Bruck Ripoll - Stability of Compact Hypersurfaces with Constant Mean Curvature - JAN/92

22. Marcos Sebastiani Artecona - Une Généralisation de L'Inveriant de Malgrange - FEV/92

23. Cornelis Kraaikamp e Artur Lopes - The Theta Group and the Continued Fraction with Even Partial Quotients - MAR/92

24. Silvia Lopes - Amplitude Estimation in Multiple Frequency Spectrum - MAR/92

25. Silvia Lopes e Benjamin Kedem - Sinusoidal Frequency Modulated Spectrum Analysis - MAR/92

26. Silvia Lopes e Benjamin Kedem - Iteration of Mappings and Fixed Spectrum Analysis - MAR/92

27. Miguel Ferrero, Eduardo Cisneros e Maria Ines Gonzales - Ore Extensions and Jacobson Rings - MAI/92

28. Sara C. Carmona - An Asymptotic Problem for a Reaction-Diffusion Component - JUL/92

29. Luiz Fernando Carvalho da Rocha - Unique Ergodicity of Interval Exchange Maps - JUL/92

30. Sara C. Carmona - Wave Front Propagation for a Cauchy Problem With a Fast Component - OUT/92

31. Marcos Sebastiani Artecona e Iván Pan Pérez - Intersections Transverses dans l'Espace Projectif - OUT/92

32. Miguel Ferrero - Closed Bimodules over Prime Rings: Closed Submodules and Applications to Rings Extensions - DEZ/92

33. Dinara W. X. Fernandez - Método da Máxima Verossimilhança Restrita para Estimação de Componentes de Variância - SET/93

34. Martin Knott e M. Teresa Albanese - Polymiss: A Computer Program for Fitting a One- or Two-Factor Logit-Probit Latent Variable Model to Polynomous Data when Observations may be Missing - OUT/93

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
NÚCLEO DE ATIVIDADES EXTRA CURRICULARES

Os Cadernos de Matemática e Estatística publicam as seguintes séries:

Série A: Trabalho de Pesquisa

Série B: Trabalho de Apoio Didático

Série C: Colóquio de Matemática SBM/UFRGS

Série D: Trabalho de Graduação

Série F: Trabalho de Divulgação

Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:

NAEC - NÚCLEO DE ATIVIDADES EXTRA CURRICULARES
INSTITUTO DE MATEMATICA - UFRGS
AV. BENTO GONÇALVES, 9500 - PRÉDIO 43111
CEP 91509 - 900 AGRONOMIA - POA/RS
FONE: 336 92 22 OU 339 13 55 OU 228 16 33
RAMAL 6197
FAX: 336 15 12