

Preliminary analysis of ordinal data

Contents

| | | |
|----------|--------------------------------------|----------|
| 1 | INTRODUCTION | 1 |
| 1.1 | THE POLITICAL ACTION SURVEY | 1 |
| 1.2 | DATA SCREENING | 2 |
| 1.3 | MISSING VALUES | 4 |
| 1.3.1 | <i>Listwise Deletion</i> | 5 |
| 1.3.2 | <i>Pairwise Deletion</i> | 6 |
| 1.3.3 | <i>Imputation</i> | 7 |
| 1.4 | USING INTERACTIVE LISREL | 10 |
| 1.4.1 | <i>Get Yourself a PSF File</i> | 10 |
| 1.4.2 | <i>Imputation and Data Screening</i> | 10 |

1 Introduction

This example is the first of a set of examples extracted from a note by K.G. Jöreskog first posted on the SSI website in 2005 with the title “*Structural Equation Modeling with Ordinal Variables using LISREL*”.

Observations on an ordinal variable represent responses to a set of ordered categories, such as a five-category Likert scale. It is only assumed that a person who selected one category has more of a characteristic than if he/she had chosen a lower category, but we do not know how much more. Ordinal variables are not continuous variables and should not be treated as if they are. It is common practice to treat scores 1, 2, 3, ... assigned to categories as if they have metric properties but this is wrong. Ordinal variables do not have origins or units of measurements. Means, variances, and covariances of ordinal variables have no meaning. The only information we have are counts of cases in each cell of a multiway contingency table. To use ordinal variables in structural equation models requires other techniques than those that are traditionally employed with continuous variables.

1.1 The Political Action Survey

To illustrate the data analysis in this paper I use the Political Action Survey which was a cross-national survey designed and carried out to obtain information on conventional and unconventional forms of political participation in industrial societies (Barnes & Kaase, 1979). The data was made available by the Zentralarchiv

fur Empirische Sozialforschung, University of Cologne. The data was originally collected by independent institutions in different countries. Neither the original collectors nor the Zentralarchiv bear any responsibility for the analysis reported here. These are the questions that were used in the USA. In Britain, the same questions were used with Congress in Washington replaced by Parliament. In the other countries the corresponding questions were used in other languages.

The first Political Action Survey was conducted between 1973 and 1975 in eight countries: Britain, West Germany, The Netherlands, Austria, the USA, Italy, Switzerland, and Finland. New cross-sections including a panel were obtained during 1980-81 in three of the original countries: West Germany, The Netherlands, and the USA. All data was collected through personal interviews on representative samples of the population 16 years and older.¹

The Political Action Survey contains several hundred variables. For the present purpose of illustration, the six variables representing the operational definition of *political efficacy* will be used. The conceptual definition of political efficacy is *the feeling that individual political action does have, or can have, an impact upon the political process* (Campbell, *et al.*, 1954). The operational definition of political efficacy is based on the responses to the following six items:

| | |
|----------|---|
| NOSAY | People like me have no say in what the government does |
| VOTING | Voting is the only way that people like me can have any say about how the government runs things |
| COMPLEX | Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on |
| NOCARE | I don't think that public officials care much about what people like me think |
| TOUCH | Generally speaking, those we elect to Congress in Washington lose touch with the people pretty quickly |
| INTEREST | Parties are only interested in people's votes but not in their opinions |

Permitted responses to these statements were:

- AS agree strongly
- A agree
- D disagree
- DS disagree strongly
- DK don't know
- NA no answer

These responses were coded 1, 2, 3, 4, 8, 9, respectively.

1.2 Data Screening

Most raw data from surveys are downloaded from large files at data archives and stored on media like diskettes or tapes for analysis. The data file may contain many variables on many cases. Before doing more elaborate analysis of the data, it is important to do a careful data screening to check for coding errors and other mistakes in the data. Such a data screening will also reveal outliers and other anomalies and detect if there are specific patterns of missing values in the data. The data screening gives a general idea of the character and quality of the data.

PRELIS automatically does such a data screening by determining for each variable the distinct data values present in the data and the number of each. If a variable has more than 15 distinct values, PRELIS will group them in intervals and determine the number in each interval.

The data on the efficacy variables for the USA are available in the file **EFFICACY.RAW**. The data values have been recorded with one space before each number, *i.e.*, with two columns per variable.

There are two ways of running LISREL:

- From syntax files (in this case PRELIS syntax files)
- Using the point and click Windows interface.

In the following I illustrate the examples using syntax files. In Section 1.4 I describe briefly how one can do the same thing using Interactive LISREL.

A simple PRELIS input file for screening the data is as follows (**ORD11.PRL**):

```
EFFICACY: PRELIS Run 1
!Data Screening of Political Action Data for the USA:
!Cross-Section Data - Variables: 136 - 141 (Political Efficacy Variables)

Data Ninputvariables = 6
Labels
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
Rawdata=EFFICACY.RAW
Output
```

Note that the number of cases (records of data) need not be specified; PRELIS determines the sample size, all distinct data values for each variable and the absolute and relative frequency of occurrence of each value. PRELIS also gives a bar chart showing the percentage of each data value. The output file shows that PRELIS has correctly determined

- that there are 1719 cases in the data,
- that there are six distinct values on each variable, coded 1, 2, 3, 4, 8, 9,
- the distribution of these values.

The results are presented in compact form in Table 1.

Table 1: Univariate Marginal Distributions

| | Frequency | | | | | | Percentage | | | | | |
|----------|-----------|-----|-----|-----|----|----|------------|------|------|-----|-----|-----|
| | AS | A | D | DS | DK | NA | AS | A | D | DS | DK | A |
| NOSAY | 175 | 518 | 857 | 130 | 29 | 10 | 10.2 | 30.1 | 49.9 | 7.6 | 1.7 | 0.6 |
| VOTING | 283 | 710 | 609 | 80 | 26 | 11 | 16.5 | 41.3 | 35.4 | 4.7 | 1.5 | 0.6 |
| COMPLEX | 343 | 969 | 323 | 63 | 9 | 12 | 20.0 | 56.4 | 18.8 | 3.7 | 0.5 | 0.7 |
| NOCARE | 250 | 701 | 674 | 57 | 20 | 17 | 14.5 | 40.8 | 29.3 | 3.3 | 1.2 | 1.0 |
| TOUCH | 273 | 881 | 462 | 26 | 60 | 17 | 15.9 | 51.3 | 26.9 | 1.5 | 3.5 | 1.0 |
| INTEREST | 264 | 762 | 581 | 31 | 62 | 19 | 15.4 | 44.3 | 33.8 | 1.8 | 3.6 | 1.1 |

These results agree exactly with those given in the Political Action Codebook, pp. 174-180. Thus, we have a solid base to continue our analysis. If something had been wrong in the data, it would have been detected by this kind of data screening.

In the Political Action Survey the responses *Agree Strongly*, *Agree*, *Disagree*, *Disagree Strongly*, *Don't Know*, *No Answer* to the political efficacy items were coded 1, 2, 3, 4, 8, 9, respectively. One can assign category labels to category codes by including the following line in the input file (see file **ORD11.PRL**):

```
CLabels NOSAY - INTEREST 1=AS 2=A 3=D 4=DS 8=DK 9=NA
```

1.3 Missing Values

Obviously, the responses *Don't Know* and *No Answer* cannot be used as categories for the ordinal scale that goes from *Agree Strongly* to *Disagree Strongly*. The usual way to deal with such responses is to declare them as missing values and include some treatment of missing values in the analysis. A lengthy discussion of this issue is beyond the scope of this paper. LISREL provides *multiple imputation* for continuous normally distributed variables with data missing at random (MAR). One can also estimate a LISREL model directly from raw data with full information maximum likelihood (FIML) under multivariate normality and MAR. Although very powerful under the assumptions made, these procedures should not be used routinely to solve missing data problems. They cannot (should not) be used with categorical variables.

One should try to take the mechanism that generates the missing data into account. Why are data missing on a particular variable? Does the probability of a missing value on this variable depend on the values of the variable itself? If so, MAR does not hold. Does the probability of a missing value on this variable depend on other variables? If so, MAR may hold. But if one can find other variables that can be used to predict the missing values these can be taken into account.

One general procedure to do this is the matching procedure described in Section 1.3.3. This can be used with ordinal or continuous variables with any distribution and does not require MAR. For continuous variables and data missing completely at random (MAR), this procedure, also called *similar response pattern imputation*, was evaluated by Brown (1994) and was found to work well as compared with several other procedures, including listwise and pairwise deletion.

In the following, I consider these three procedures in turn. I do not recommend that listwise and pairwise deletion be used to solve the missing data problem unless you have MCAR, and even with MCAR, I do not recommend

computing covariance or correlation matrices with pairwise deletion. *I include the listwise and pairwise procedures here as descriptive devices as they provide further insight into the missing data problem.*

1.3.1 Listwise Deletion

Missing values do not seem to be a serious problem in this example. As seen in Table 1, the percentage of *Don't Know* answers varies from 0.5% for COMPLEX to 3.6% for INTEREST. More people answer *Don't Know* for TOUCH and INTEREST than for the other items. The percentage of *No Answer* responses varies from 0.6% for NOSAY and VOTING to 1.1% for INTEREST. Listwise deletion means that all cases with *Don't Know* and *No Answer* responses will be excluded in the analysis. This is illustrated in the following input file (**ORD12.PRL**):

```
EFFICACY: PRELIS Run 2
!Data Screening of Political Action Data for the USA:
!Cross-Section Data - Variables: 136 - 141 (Political Efficacy Variables)
!Listwise Deletion of Missing Values
Data Ninputvariables = 6 Missing = 8,9
Labels
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST Rawdata=EFFICACY.RAW
CLabels NOSAY - INTEREST 1=AS 2=A 3=D 4=DS
Output
```

This gives the following results.

Number of Missing Values per Variable

| NOSAY | VOTING | COMPLEX | NOCARE | TOUCH | INTEREST |
|-------|--------|---------|--------|-------|----------|
| 39 | 37 | 21 | 37 | 77 | 81 |

Distribution of Missing Values

Total Sample Size(N) = 1719

| Number of Missing Values | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------------------|------|-----|----|----|---|---|---|
| Number of Cases | 1554 | 106 | 26 | 18 | 4 | 2 | 9 |

The distribution of missing values over variables are given first. It is seen that there are only 21 missing values on COMPLEX whereas there are 77 and 81 on TOUCH and INTEREST, respectively. As we already know that most of the missing values on TOUCH and INTEREST are *Don't Know* rather than *No Answer* responses, it seems that these items are considered by the respondents to be more difficult to answer.

Further down in the output is the distribution of missing values over cases. It is seen that there are only 1554 out of 1719 cases without any missing values. The other 165 cases have one or more missing values. With listwise deletion this is the loss of sample size that will occur. Most, or 106, of the 165 cases with missing values have only one missing value. But note that there are 9 cases with 6 missing values, *i.e.*, these cases have either not responded or have responded *Don't Know* to all of the six items. These 9 cases are of course useless for any purpose considered here.

The rest of the output (not shown here) gives the distribution of the 1554 cases of the listwise sample over the four ordinal categories for each variable. It is seen that most people answer either *agree* or *disagree*. Fewer people answer with the *stronger* alternatives.

1.3.2 Pairwise Deletion

A more comprehensive data screening can be done by pairwise deletion. To do so, add the specification Treatment = Pairwise to the Data line and put MP (for *missing patterns*) on the Output line (see file **ORD12.PRL**).

This gives the following results:

Effective Sample Sizes
Univariate (in Diagonal) and Pairwise Bivariate (off Diagonal)

| | NOSAY | VOTING | COMPLEX | NOCARE | TOUCH | INTEREST |
|----------|-------|--------|---------|--------|-------|----------|
| NOSAY | 1680 | | | | | |
| VOTING | 1658 | 1682 | | | | |
| COMPLEX | 1670 | 1674 | 1698 | | | |
| NOCARE | 1655 | 1656 | 1675 | 1682 | | |
| TOUCH | 1620 | 1627 | 1635 | 1622 | 1642 | |
| INTEREST | 1619 | 1621 | 1632 | 1622 | 1598 | 1638 |

This table gives the univariate and bivariate sample sizes. Thus, there are 1680 cases with complete data on NOSAY but only 1638 cases with complete data on INTEREST. There are 1658 cases with complete data on *both* NOSAY and VOTING but only 1598 cases with complete data on *both* TOUCH and INTEREST.

The same kind of information, but in terms of percentage of missing data instead, is given in the following table.

Percentage of Missing Values
Univariate (in Diagonal) and Pairwise Bivariate (off Diagonal)

| | NOSAY | VOTING | COMPLEX | NOCARE | TOUCH | INTEREST |
|----------|-------|--------|---------|--------|-------|----------|
| NOSAY | 2.27 | | | | | |
| VOTING | 3.55 | 2.15 | | | | |
| COMPLEX | 2.85 | 2.62 | 1.22 | | | |
| NOCARE | 3.72 | 3.66 | 2.56 | 2.15 | | |
| TOUCH | 5.76 | 5.35 | 4.89 | 5.64 | 4.48 | |
| INTEREST | 5.82 | 5.70 | 5.06 | 5.64 | 7.04 | 4.71 |

The next lines give all possible patterns of missing data and their sample frequencies. Each column under Pattern corresponds to a variable. A 0 means a complete data and a 1 means a missing data.

Missing Data Map

| Frequency | PerCent | Pattern |
|-----------|---------|-------------|
| 1554 | 90.4 | 0 0 0 0 0 0 |
| 16 | 0.9 | 1 0 0 0 0 0 |
| 12 | 0.7 | 0 1 0 0 0 0 |
| 1 | 0.1 | 1 1 0 0 0 0 |

| | | | | | | | |
|----|-----|---|---|---|---|---|---|
| 4 | 0.2 | 0 | 0 | 1 | 0 | 0 | 0 |
| 11 | 0.6 | 0 | 0 | 0 | 1 | 0 | 0 |
| 31 | 1.8 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0.1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 2 | 0.1 | 1 | 1 | 0 | 0 | 1 | 0 |
| 1 | 0.1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 4 | 0.2 | 0 | 0 | 0 | 1 | 1 | 0 |
| 1 | 0.1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 32 | 1.9 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0.1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0.1 | 1 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0.1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 5 | 0.3 | 0 | 0 | 0 | 1 | 0 | 1 |
| 2 | 0.1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 1 | 0.1 | 0 | 0 | 1 | 1 | 0 | 1 |
| 1 | 0.1 | 1 | 0 | 1 | 1 | 0 | 1 |
| 14 | 0.8 | 0 | 0 | 0 | 0 | 1 | 1 |
| 4 | 0.2 | 1 | 0 | 0 | 0 | 1 | 1 |
| 4 | 0.2 | 0 | 1 | 0 | 0 | 1 | 1 |
| 2 | 0.1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0.1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0.1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 2 | 0.1 | 0 | 1 | 1 | 1 | 1 | 1 |
| 9 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 |

Thus, there are 1554 cases with no missing data, there are 16 cases with missing values on variable 1 only, and 1 case with missing values on both variable 1 and 2, etc. Note again that there are 9 cases with missing values on all 6 variables.

This kind of information is very effective in detecting specific patterns of missingness in the data. In this example there are no particular patterns of missingness. The only striking feature is that there are more missing values on TOUCH and INTEREST. We know from the first run that these are mainly *Don't know* responses.

1.3.3 Imputation

Another way to deal with the problem of missing values is by imputation, *i.e.*, by substituting a real scale value 1, 2, 3, or 4 for the missing values 8 and 9. PRELIS has a procedure for imputing missing values on a variable by matching on other variables. This procedure is based on the idea that if person *a* has a missing value on variable *i* and has the same response pattern as person *b* on a set of matching variables, it is likely that he/she should have the same value on variable *i* as person *b*. Therefore, *b*'s value on variable *i* is substituted for *a*'s missing value on variable *i*. If there are several persons with the same response patterns on the matching variables and with the same values on variable *i*, there is an even stronger case for substituting this value for *a*'s value on variable *i*.

As most of the missing values are on the variables TOUCH and INTEREST, one idea to increase the listwise sample size is to impute missing values on TOUCH and INTEREST by matching on the other variables. This is merely an illustration. To do imputation the following PRELIS input file can be used (see file **ORD13.PRL**):

```
EFFICACY: PRELIS Run 3
Cross-Section Data - Variables: 136 - 141 (Political Efficacy Variables)
Imputation of Missing Values on TOUCH and INTEREST
```

```

Listwise Deletion of Missing Values after Imputation
Saving Imputed Data in EFFICACY.IMP
Data Ninputvariables = 6 Missing = 8,9
Labels
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
Rawdata=EFFICACY.RAW
Impute (TOUCH INTEREST) (NOSAY - NOCARE) XN
CLabels NOSAY - INTEREST 1=AS 2=A 3=D 4=DS
Output RA=EFFICACY.IMP Width = 2 Ndecimals = 0

```

Missing values on TOUCH and INTEREST are only imputed if matching cases are found. Therefore, there may still be missing values on TOUCH and INTEREST after imputation. Note that PRELIS will only impute "legal" values, *i.e.*, values 1, 2, 3, and 4, not values such as 2.5 or 3.4. All the missing values on NOSAY - NOCARE are still there after imputation. In this example, all missing values remaining after imputation are eliminated by listwise deletion and the listwise sample is saved in the file **EFFICACY.IMP**. The specification Width = 2 Ndecimals = 0 means that the data will be written in two-column fields without any decimals. Hence, the file **EFFICACY.IMP** can be read in free format. This file has no missing values in it.

The output lists all imputed cases as follows:

Imputations for TOUCH

| | | | | | |
|-----------|--------------------|---|---------------------------|-----|-----|
| Case 28 | imputed with value | 2 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 29 | imputed with value | 2 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 238 | imputed with value | 2 | (Variance Ratio = 0.454), | NM= | 10 |
| Case 336 | imputed with value | 1 | (Variance Ratio = 0.277), | NM= | 26 |
| Case 418 | imputed with value | 2 | (Variance Ratio = 0.380), | NM= | 76 |
| Case 530 | imputed with value | 2 | (Variance Ratio = 0.360), | NM= | 143 |
| Case 578 | imputed with value | 3 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 600 | imputed with value | 2 | (Variance Ratio = 0.405), | NM= | 31 |
| Case 604 | imputed with value | 2 | (Variance Ratio = 0.493), | NM= | 23 |
| Case 684 | imputed with value | 2 | (Variance Ratio = 0.396), | NM= | 32 |
| Case 951 | imputed with value | 1 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 963 | imputed with value | 2 | (Variance Ratio = 0.358), | NM= | 144 |
| Case 985 | imputed with value | 2 | (Variance Ratio = 0.463), | NM= | 28 |
| Case 1189 | imputed with value | 2 | (Variance Ratio = 0.356), | NM= | 145 |
| Case 1268 | imputed with value | 2 | (Variance Ratio = 0.410), | NM= | 5 |
| Case 1311 | imputed with value | 2 | (Variance Ratio = 0.410), | NM= | 5 |
| Case 1541 | imputed with value | 1 | (Variance Ratio = 0.269), | NM= | 27 |
| Case 1578 | imputed with value | 2 | (Variance Ratio = 0.399), | NM= | 9 |
| Case 1584 | imputed with value | 2 | (Variance Ratio = 0.389), | NM= | 33 |
| Case 1704 | imputed with value | 2 | (Variance Ratio = 0.476), | NM= | 24 |

Imputations for INTEREST

| | | | | | |
|----------|--------------------|---|---------------------------|-----|-----|
| Case 12 | imputed with value | 1 | (Variance Ratio = 0.292), | NM= | 31 |
| Case 28 | imputed with value | 1 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 48 | imputed with value | 2 | (Variance Ratio = 0.337), | NM= | 138 |
| Case 67 | imputed with value | 3 | (Variance Ratio = 0.309), | NM= | 6 |
| Case 332 | imputed with value | 2 | (Variance Ratio = 0.335), | NM= | 139 |
| Case 418 | imputed with value | 2 | (Variance Ratio = 0.429), | NM= | 73 |
| Case 469 | imputed with value | 2 | (Variance Ratio = 0.333), | NM= | 140 |
| Case 510 | imputed with value | 2 | (Variance Ratio = 0.287), | NM= | 17 |
| Case 530 | imputed with value | 2 | (Variance Ratio = 0.331), | NM= | 141 |
| Case 578 | imputed with value | 3 | (Variance Ratio = 0.000), | NM= | 1 |
| Case 598 | imputed with value | 2 | (Variance Ratio = 0.329), | NM= | 142 |

```

Case 653 imputed with value 2 (Variance Ratio = 0.424), NM= 74
Case 680 imputed with value 3 (Variance Ratio = 0.375), NM= 84
Case 796 imputed with value 3 (Variance Ratio = 0.465), NM= 4
Case 951 imputed with value 1 (Variance Ratio = 0.000), NM= 1
Case 1156 imputed with value 2 (Variance Ratio = 0.419), NM= 75
Case 1227 imputed with value 2 (Variance Ratio = 0.327), NM= 143
Case 1232 imputed with value 2 (Variance Ratio = 0.325), NM= 144
Case 1249 imputed with value 3 (Variance Ratio = 0.435), NM= 27
Case 1368 imputed with value 2 (Variance Ratio = 0.414), NM= 76
Case 1477 imputed with value 3 (Variance Ratio = 0.266), NM= 7
Case 1516 imputed with value 2 (Variance Ratio = 0.323), NM= 145
Case 1556 imputed with value 3 (Variance Ratio = 0.497), NM= 6
Case 1704 imputed with value 2 (Variance Ratio = 0.240), NM= 24

```

Here NM is the number of matching cases and the Variance Ratio is the ratio between the variance of variable *i* for the matching cases and the total variance of variable *i* for all cases without missing values.

After imputation, the distribution of missing values over variables is:

Number of Missing Values per Variable After Imputation

| NOSAY | VOTING | COMPLEX | NOCARE | TOUCH | INTEREST |
|-------|--------|---------|--------|-------|----------|
| 39 | 37 | 21 | 37 | 57 | 57 |

Compared with the previous output we have gained 20 cases on TOUCH and 24 cases on INTEREST. These are the 44 cases just listed.

The distribution of missing values over cases is now:

Distribution of Missing Values

Total Sample Size (N) = 1719

| Number of Missing Values | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------------------|------|----|----|----|---|---|---|
| Number of Cases | 1589 | 80 | 17 | 18 | 4 | 2 | 9 |

The listwise sample size after imputation is 1589. This was 1554 in the previous output. Altogether we have gained 35 cases with complete data. Of these 35 cases, 26 had 1 missing value and 9 had 2 missing values. All these missing values were imputed.

The analysis that follows in Section 2 will be based on the data on these 1589 cases stored in the file **EFFICACY.IMP**. The univariate distributions of the six political efficacy items estimated from these data are given in Table 2.

Table 2: Univariate Marginal Distributions

| | Frequency | | | | Percentage | | | |
|----------|-----------|-----|-----|-----|------------|------|------|-----|
| | AS | A | D | DS | AS | A | D | DS |
| NOSAY | 163 | 492 | 811 | 123 | 10.3 | 31.0 | 51.0 | 7.7 |
| VOTING | 275 | 658 | 580 | 76 | 17.3 | 41.4 | 36.5 | 4.8 |
| COMPLEX | 314 | 905 | 309 | 61 | 19.8 | 57.0 | 19.4 | 3.8 |
| NOCARE | 241 | 656 | 636 | 56 | 15.2 | 41.3 | 40.0 | 3.5 |
| TOUCH | 261 | 857 | 446 | 25 | 16.4 | 53.9 | 28.1 | 1.6 |
| INTEREST | 260 | 740 | 560 | 29 | 16.4 | 46.6 | 35.2 | 1.8 |

1.4 Using Interactive LISREL

Here I describe briefly how to do data screening and imputation using Interactive LISREL.

1.4.1 Get Yourself a LSF File

To use Interactive LISREL, the first order of business is to get a LISREL system file (LSF file) for the raw data in file **EFFICACY.RAW**. To do this, go through the following steps:

- Select Import Data in Free Format in the File menu.
- Go folder where the file **EFFICACY.RAW** is and choose **Free Format Data (*.raw)** under **Files of Type**. Open **EFFICACY.RAW**.
- In the **Enter Number of Variables** dialog box, specify 6 variables and click **OK**.
- LISREL then shows the data in a spread sheet. This is the way the LSF file is displayed.
- To complete the specification of the LSF file, go to the **Data** menu.
- Select **Define Variables** and type the names of the variables: NOSAY, VOTING, etc.
- Still in the **Define Variables** dialog box, select **Variables Type**. Select all variables and define them as 'Ordinal.' Click **OK**.
- Still in the **Define Variables** dialog box, select **Category Labels**. Select all variables and define 1 = AS, 2 = A, etc. Click **OK**.
- Still in the **Define Variables** dialog box, select **Missing Values**. Specify 8 and 9 as global missing values. As deletion method choose listwise or pairwise. Click **OK**.
- Still in the **Define Variables** dialog box, click **OK**.

1.4.2 Imputation and Data Screening

Once the LSF has been defined, it can be displayed by selecting **Open** in the **File** menu and choosing LISREL Data (*.lsf) under **Files of Type**. Then double clicking the filename will display the LSF file in spreadsheet form.

To do imputation and data screening follow the following steps:

1. With the LSF file displayed, select the **Statistics** menu and then **Impute Missing Values**.
2. In the **Impute Missing Values** dialog box, select TOUCH and INTEREST and click the first **Add** button. This defines the variables to be imputed. Then select NOSAY - NOCARE and click the third

Add button. This defines the matching variables. Select **List only successful imputations** and then click the **Output Options** button.

3. To save the imputed data as a LSF file, select the **Data** dialog box within the **Output** dialog box and check **Save the transformed data to file**. Then type a file name with suffix LSF, for example **EFFICACY.LSF**. Then click **OK**.
4. You will be returned to the **Impute Missing Values** dialog box. To generate a PRELIS syntax file click the **Syntax** button. Otherwise, click the **Run** button.

A much easier way to generate the file **EFFICACY.LSF** after imputation is to add the specification **RA=EFFICACY.LSF** on the Output line in PRELIS Run 3, see file **ORD13A.PRL**.

The generated file **EFFICACY.LSF** corresponds exactly to the file **EFFICACY.IMP** and the results in the output file should be the same as for the second PRELIS run described in Section 1.3.3. As will be illustrated in the next section, the file **EFFICACY.LSF** can be used for further analysis using interactive LISREL.