Exploratory analysis of political survey data

Contents

1. Introduction 1
2. Importing Data into PRELIS 1
3. Data Screening 2
4. Assigning Labels to Categories 3
5. Define Missing Values 5
6. Imputation of Missing Values 6
7. Homogeneity Tests 7
8. Graphical Display of Univariate Distributions 9
9. Import Data and Compute Polychoric Coefficients 9

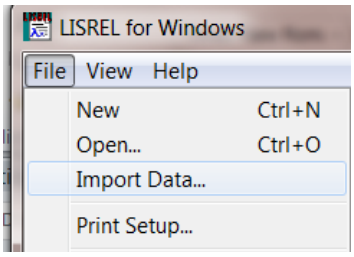
1. Introduction

The SPSS file **dataex7.sav** contains six variables on a subset of cases from the first cross-section of a political survey. Use is made of a number of dialog boxes to

- import **dataex7.sav** and save it as a LISREL (*.**lsf**) spreadsheet file
- do data screening
- assign labels (names) for categories of the ordinal variables
- define missing values
- impute missing values and calculate polychoric coefficients and the asymptotic covariance matrix
- perform a homogeneity test for pairs of ordinal variables
- graphical display of the univariate distributions of the variables.

2. Importing Data into PRELIS

From the **File** menu, select **Import Data…**



The user is prompted to select the drive and path where the data file to be imported is stored (in this case the **PRELIS examples** folder).

We would like to import an SPSS for Windows (*.**sav**) file named **dataex7.sav**. From the **Files of Type** drop-down menu, select SPSS for Windows (*.**sav**). Once **dataex7.sav** is selected, click **Open** to proceed.

Hint: Click (left mouse button) on the **List Files of Type** drop-down list box and type in the letter "S" to go directly to all files types starting with the letter "S".

Select a file name and folder in which the LISREL system data file should be saved. Note that the extension has to be *.**lsf**. When done, click **Save** to display the system file as a spreadsheet.

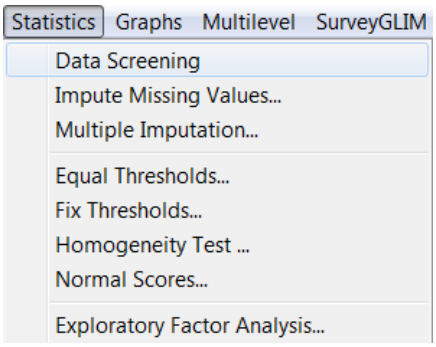
DATAEX7.LSF

	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
1	2.00	2.00	1.00	1.00	1.00	1.00
2	2.00	3.00	3.00	3.00	2.00	3.00
3	3.00	2.00	2.00	3.00	3.00	3.00
4	3.00	3.00	2.00	3.00	2.00	3.00
5	2.00	2.00	1.00	2.00	2.00	2.00
6	2.00	2.00	1.00	1.00	2.00	1.00

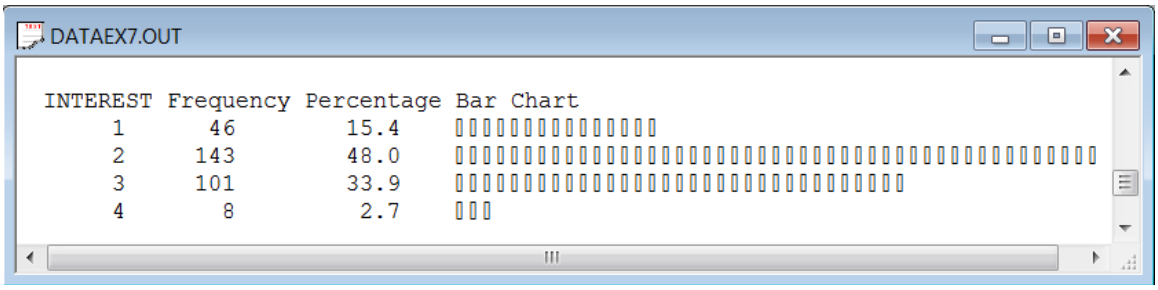
The file **dataex7.lsf** contains data for six variables, these being NOSAY, VOTING, COMPLEX, NOCARE, TOUCH and INTEREST.

3. Data Screening

The simplest form of data screening is performed by selecting **Data Screening** from the **Statistics** menu:



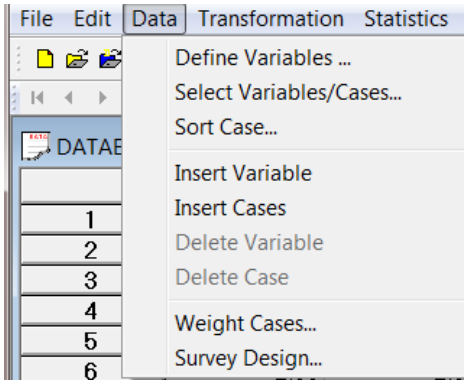
For each variable, **dataex7.out** lists the number and relative frequency of each data value, and it gives a bar chart showing the distribution of the data values. For the variable INTEREST, it is as follows:



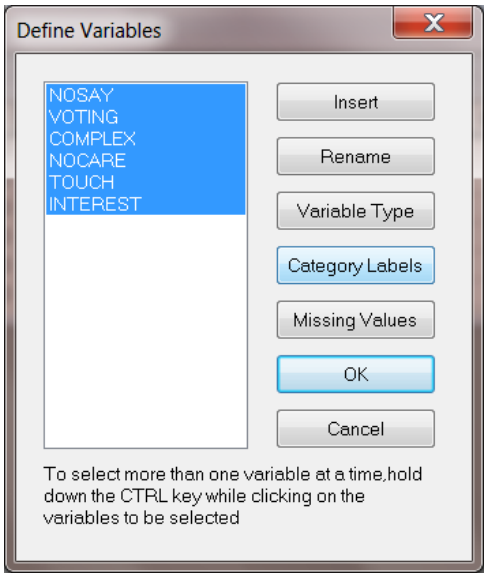
4. Assigning Labels to Categories

Each of the 6 political efficacy items has four distinct numerical values. To assign category labels AS, A, D and DS to the numerical values 1, 2, 3, and 4 (AS = Agree strongly, A = Agree, D = Disagree, DS = Disagree Strongly) we proceed as follows.

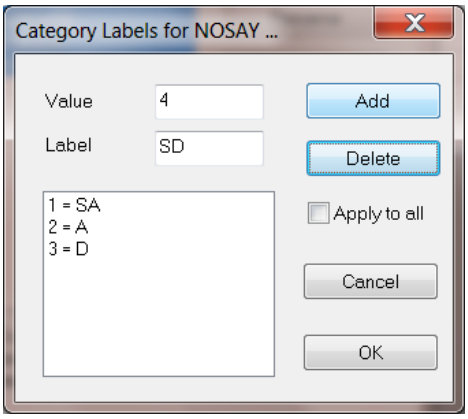
Select the **Define Variables** option from the **Data** menu to activate the **Define Variables** dialog box.



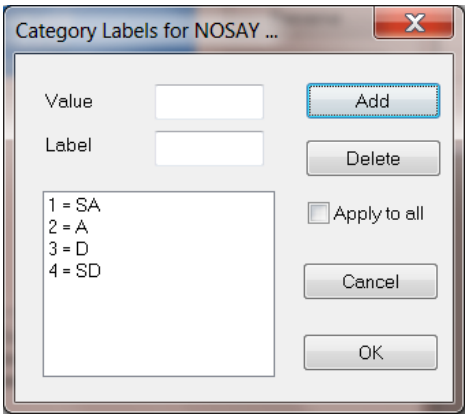
Select all six variables by holding down the **CTRL** key on the computer keyboard while clicking on each variable (left mouse button). Alternatively, one can drag the cursor over the six variables with the left mouse button held down. The selected variables will be highlighted as shown below:



Click **Category Labels** to obtain the **Labels** dialog box:



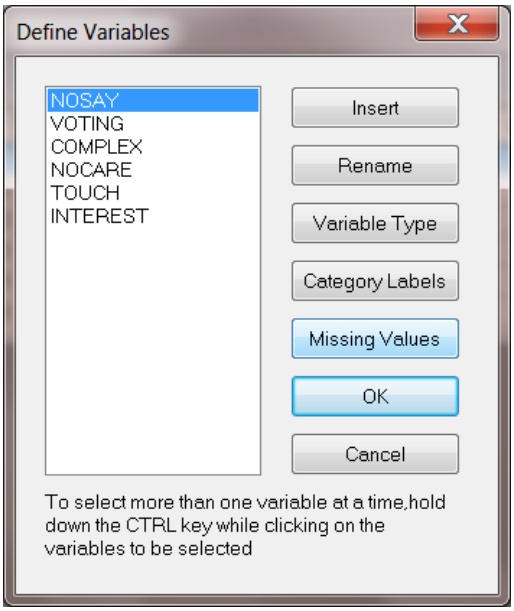
Use this dialog box to enter a numerical value and its corresponding label, then click **Add** to display the result in the string field at the bottom. Finally, when the last value and label (4 and SD) have been entered, click **Add** to obtain the following dialog box.



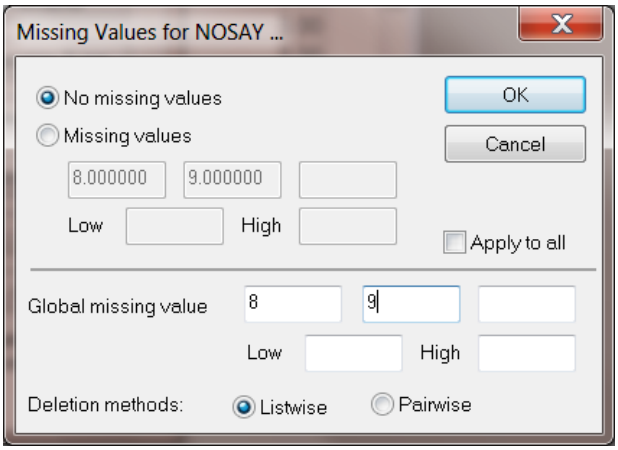
Any of these choices can be deleted by highlighting it (left mouse button) and then clicking on the **Delete** button. When done, click **OK** to return to the **Define Variables** dialog box.

5. Define Missing Values

Suppose that the data set contains values 8 and 9, which are actually codes to indicate missing responses. From the **Define Variables** dialog box, we select the **Missing Values** option to activate the **Missing Values** dialog box shown below:



Click the **Missing values** radio button. Enter the values 8 and 9 in the first two string fields as shown below. Check the **Apply to all** check box and the **Listwise** deletion radio button. When this is done, click **OK** to return to the **Define Variables** dialog box. Click **OK** to exit this dialog box and return to the PSF window.



Note:

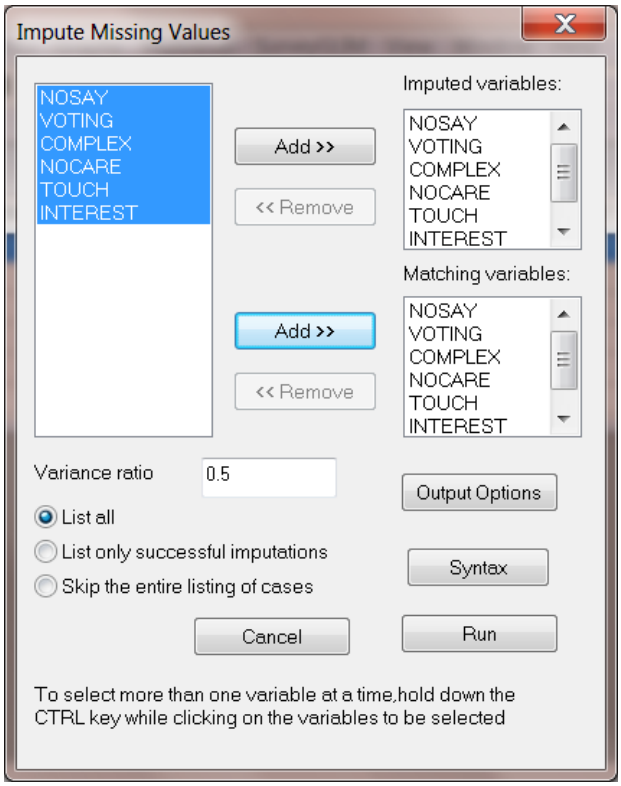
It is important to select **Save** from the **File** menu for the changes to take effect.

6. Imputation of Missing Values

Missing values can be handled by choosing pairwise or listwise deletion. PRELIS offers another possible way to handle missing values, namely by imputation, i.e. by substitution of real values for missing values. The value to be substituted for the missing value for a case is obtained from another case that has a similar response pattern over a set of matching variables. This method is recommended for ordinal variables. Multiple imputations are based on maximum likelihood procedures and is recommended for continuous variables.

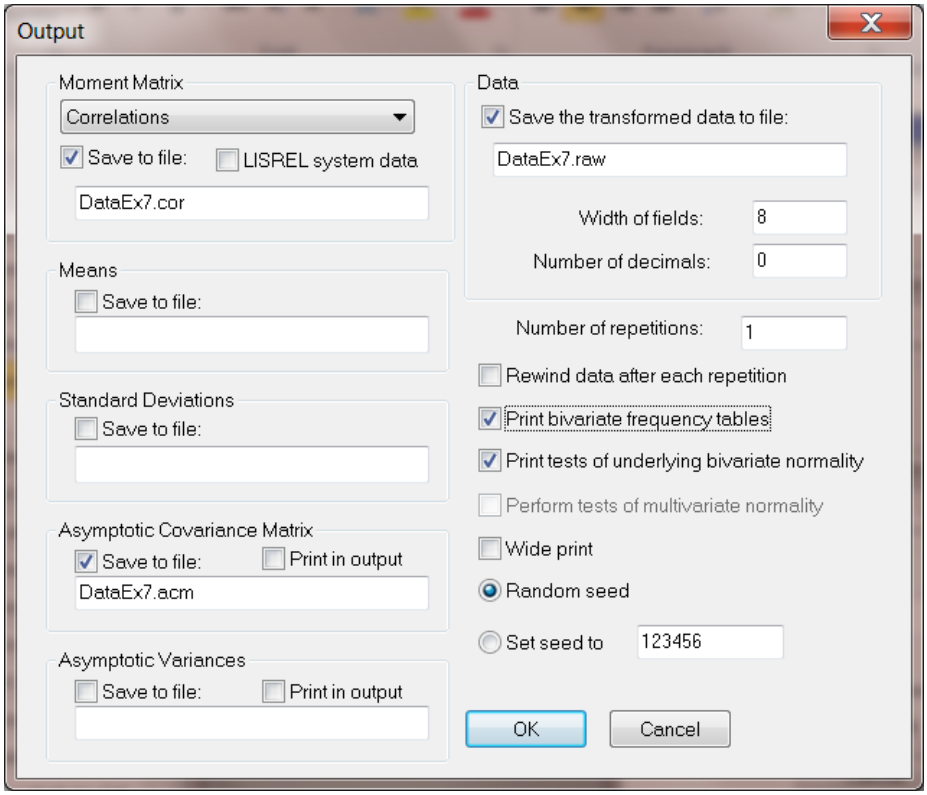
Select the **Impute Missing Values…** option from the **Statistics** menu. By clicking on **Impute Missing Values…**, the dialog box shown below appears.

Suppose that we would like to impute the missing values on each variable using all the other variables as matching variables. Cases with missing values will be eliminated after imputation.



Hold the **CTRL** key down while selecting each of the 6 variables (left mouse button). Click the top **Add>>** button to add the selected list of **Imputed variables** and then click the bottom **Add>>** button to add the selected list of **Matching variables**.

To obtain moment matrices, click **Output Options** in the dialog box shown above. The **Output** dialog box appears.



From the **Type of Moment matrix to be analyzed** drop-down list box, select **Correlations**. Select **Save to File** and type in the name of the file.

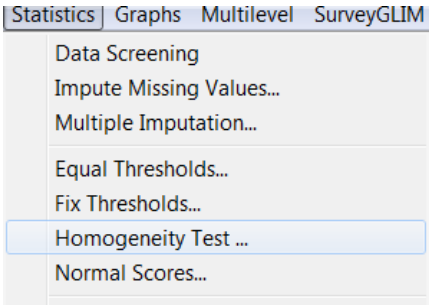
To save the newly created data set check the **Save the transformed raw data to file** check box and enter a file name. Note that if the file extension is ***.lsf**, the file will be saved as a LISREL system data file. Also check the **Save to file** check box under **Asymptotic Covariance matrix** and enter **dataex7.acm** in the string field. Click **OK** to return to the **Impute Missing Values** dialog box. If satisfied with the selections made, click **Run** to invoke PRELIS. On execution, an output file named **dataex7.out** is produced. Alternatively, click **Syntax** to create a PRELIS syntax file.

7. Homogeneity Tests

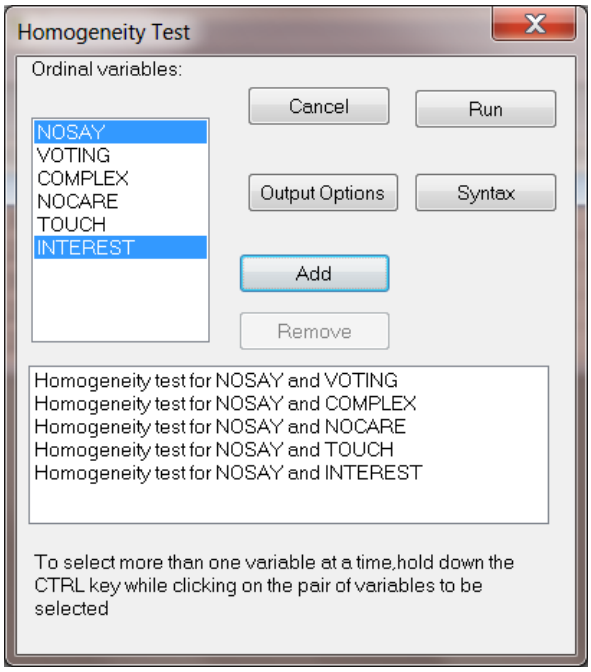
The homogeneity test is a test of the hypothesis that the marginal distributions of two categorical variables with the same number of categories (k) are the same. The χ^2 statistic for testing this hypothesis has k - 1 degrees of freedom.

The homogeneity test is particularly useful in longitudinal studies to test the hypothesis that the distribution of a variable has not changed from one occasion to the next.

To test for homogeneity, select **Homogeneity Test** from the **Statistics** menu to open the **Homogeneity Test** dialog box.



Hold the **CTRL** keyboard button down while selecting two variables from the list of **Ordinal variables** as shown in the dialog box below.



Once two variables are highlighted, the **Add** button is activated, and the selection can be entered by clicking on this button. The dialog box above shows that the variables NOSAY and INTEREST are highlighted. By clicking **Add**, this test will be added to the 4 tests shown in the bottom window of the dialog box.

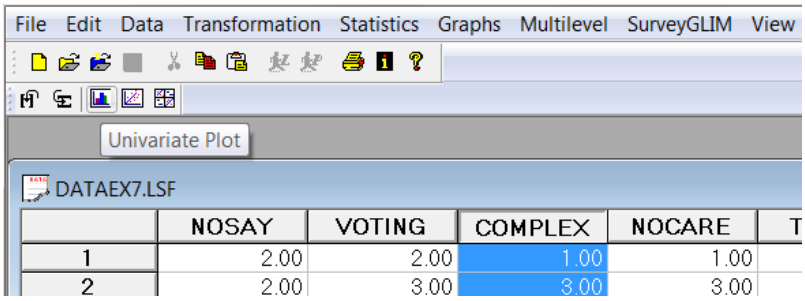
When all the tests are entered, click **Output Options** for additional output or, if not required, the **Run PRELIS** icon button to start the PRELIS analysis. A summary of the test results is given below:

Homogeneity Tests

Variable	vs.	Variable	Chi-Squ.	D.F.	P-Value
NOSAY	vs.	VOTING	53.220	3	0.000
NOSAY	vs.	COMPLEX	134.792	3	0.000
NOSAY	vs.	NOCARE	29.270	3	0.000
NOSAY	vs.	TOUCH	94.407	3	0.000
NOSAY	vs.	INTEREST	51.844	3	0.000

8. Graphical Display of Univariate Distributions

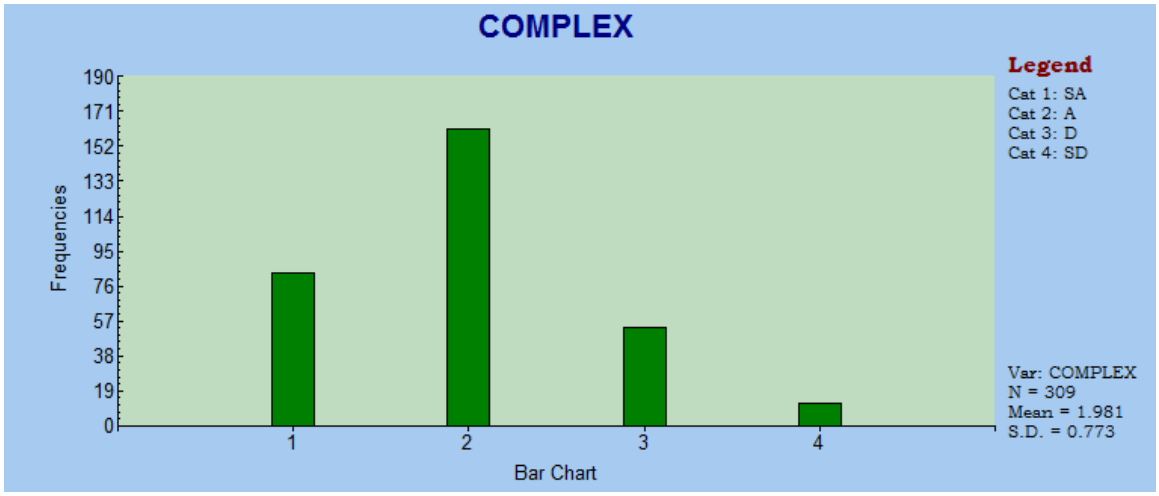
A bar chart of the univariate distribution of each variable may be obtained by clicking on a variable name to highlight the column of data values as shown below.



	NOSAY	VOTING	COMPLEX	NOCARE	T
1	2.00	2.00	1.00	1.00	
2	2.00	3.00	3.00	3.00	

Once this is done, click the **Univariate Plot** icon button (third icon from the right on the toolbar) to create the bar chart. Alternatively, one may select the **Univariate** option from the **Graphs** menu.

The bar chart given below shows that the majority of respondents selected category 2 (Agree), whilst only a small number selected the "Disagree Strongly" category.



9. Import Data and Compute Polychoric Coefficients

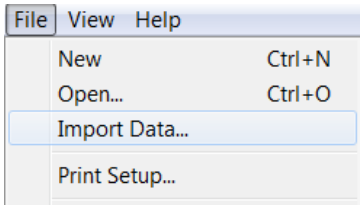
LISREL provides the user with numerous features when importing data as an **.lsf** file. The user can, for example, delete variables, select cases and compute various sample statistics interactively by making use of a set of menus.

As an illustration, an SPSS data set, **data100.sav**, which contains two continuous and four ordinal variables, will be imported. The interactive graphical user interface environment will be then used to

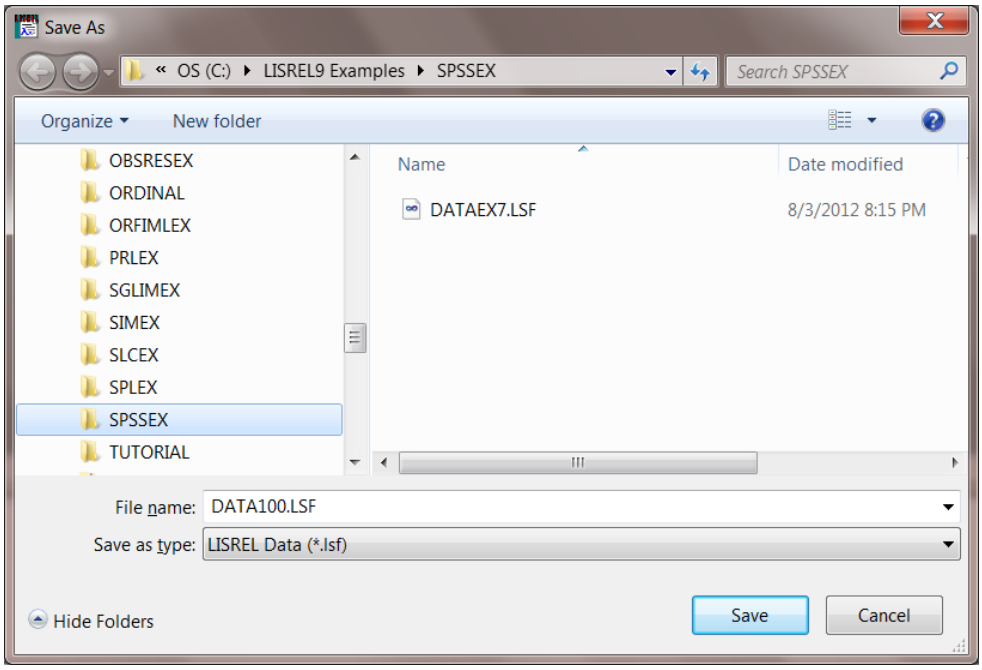
- delete the continuous variables
- select odd numbered cases (that is cases 1, 3, ...)
- calculate a matrix of polychoric correlation coefficients

- save this and the new data set to external files.

From the **File** menu, select **Import Data**:

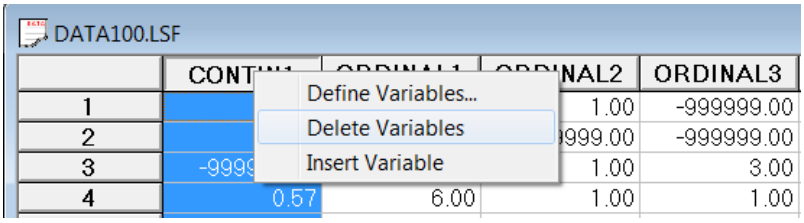


The user is prompted to select the drive and path where the data file to be imported is stored. From the **List Files of Type** drop-down list box select SPSS for Windows (*.**sav**) and then select the SPSS for Windows file named **data100.sav.** Select a file name and folder in which the PRELIS data should be saved. Note that the extension has to be *.**lsf**. When done, click **Save**.

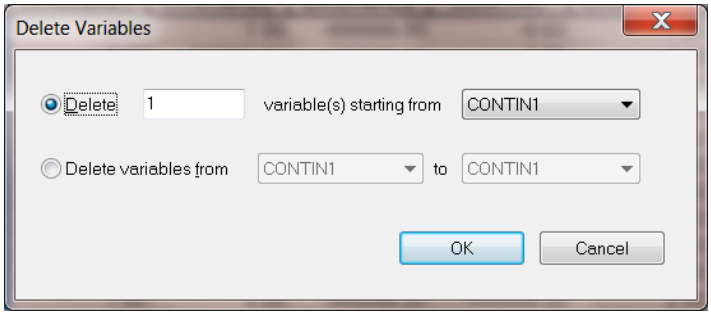


The LISREL system data file (spreadsheet) will be displayed in the LSF window. Note that the file **data100.lsf** contains data values for six variables, these being CONTIN1, ORDINAL1, ORDINAL2, ORDINAL3, CONTIN2 and ORDINAL4.

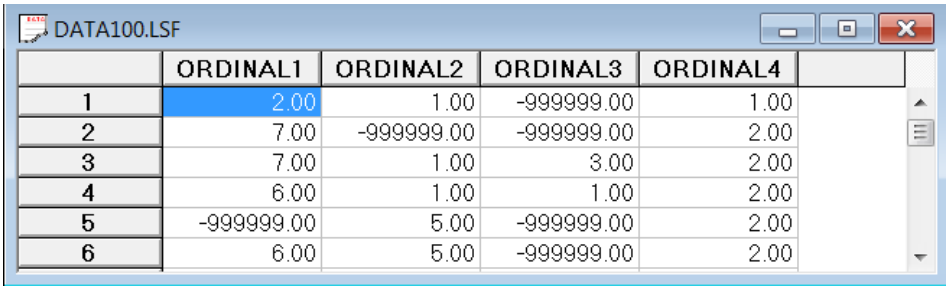
Suppose that we are only interested in the ordinal variables, then CONTIN1 and CONTIN2 may be deleted from the data set. To delete the variable CONTIN1, click on the rectangle containing the name CONTIN1. The column containing the data values will change in color as shown below. This variable may now be deleted by right-clicking on the column header CONTIN1 to select the **Delete Variables** option from the popup menu or, alternatively, by selecting the **Delete Variable** option from the **Data** menu.



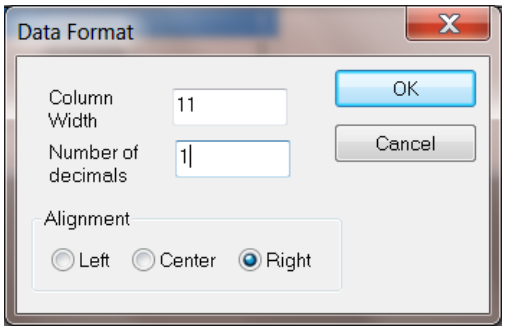
Click the **OK** button of the **Delete Variables** text box.



The variable CONTIN2 is subsequently deleted by following the procedure described above. Note that one should use the **File, Save** option to ensure that these changes take effect.

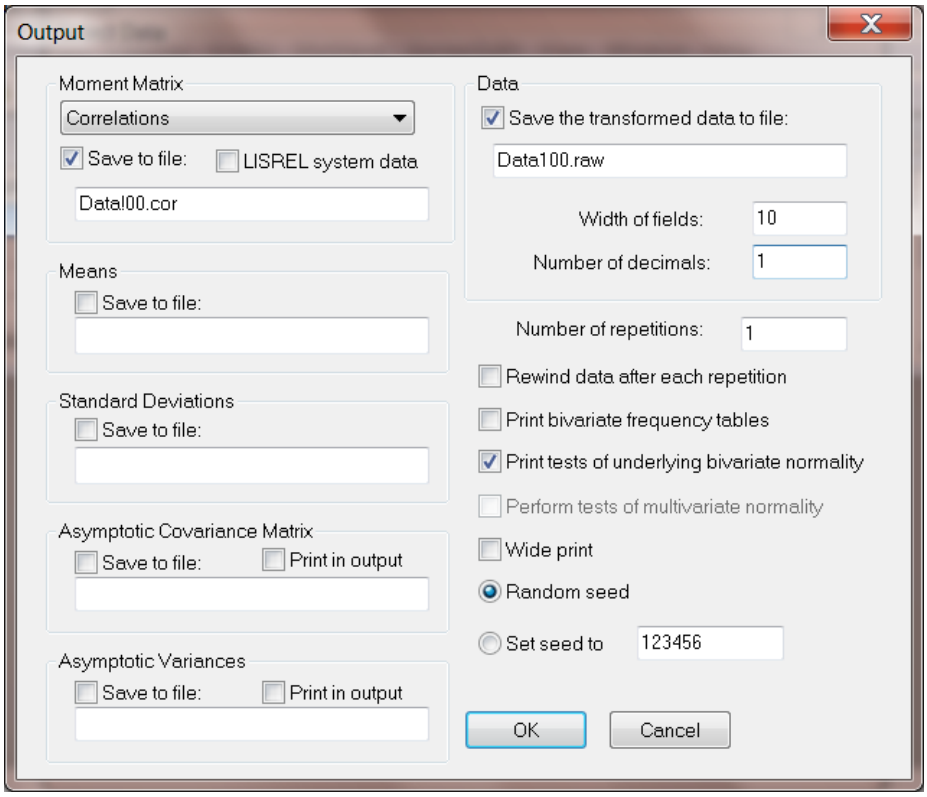


The resulting data set consists of 4 variables and 100 cases. Note that a ####### symbol in a data field indicates a missing value, or a value that contains too many digits to be displayed correctly. One can view such values by changing the **Width** and **Number of decimals** defaults by using the **Edit** menu to open the **Data Format** dialog box.



The next step in the data analysis procedure is to select the odd-numbered cases. Select the **Select Variables/Cases** option from the **Data** menu to obtain the dialog box shown below.

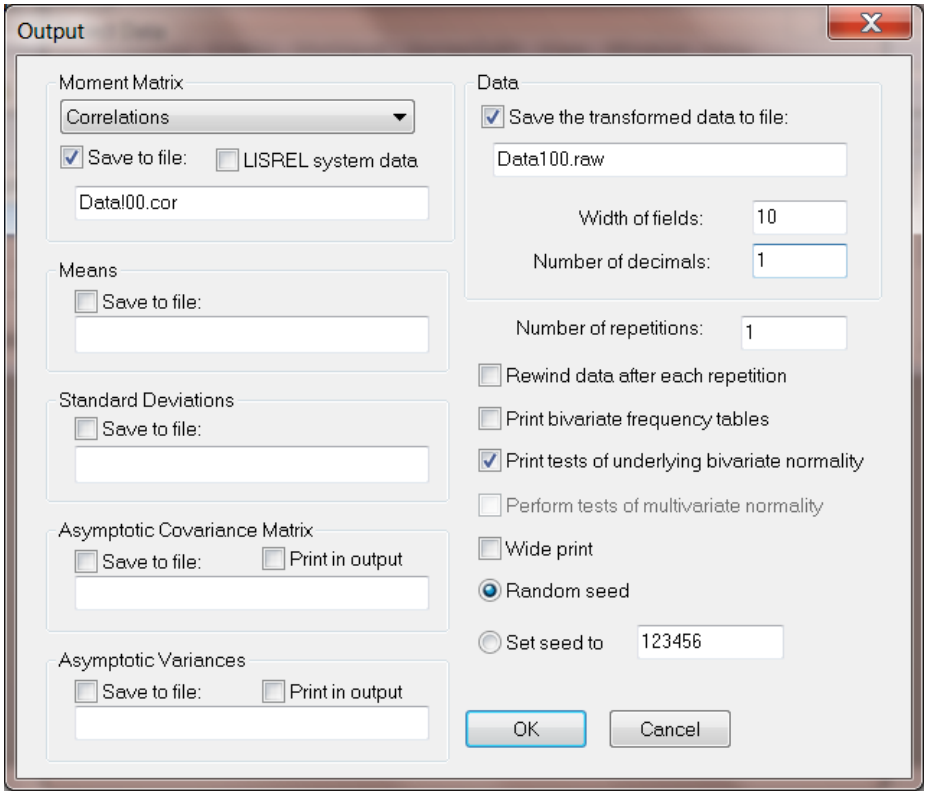
Click the **Select only those cases** radio button and then click the **odd** radio button. This instruction is transferred to the text box by clicking the **Add** button. In doing so, the words Select those case numbers that are odd appears in the text box. Any command in this box can be removed by highlighting the specific command before clicking **Remove**. Once done, click **Output Options** to obtain the **Output** dialog box.



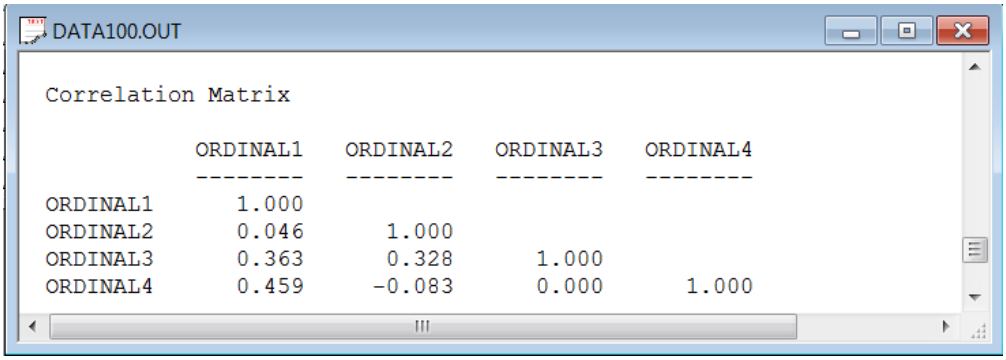
Enter 10 in the **Width of fields** string field and 1 in the **Number of decimals** string field. Using the **Moment matrix** drop-down list box, select **Correlations**. Check the **Save to File** check box and type in the name of the file (**data100.cor**). Save the data set to be created by checking the **Save the transformed data to file** check box and entering a file name. For the present example, the name **data100.raw** is used.

Notes:

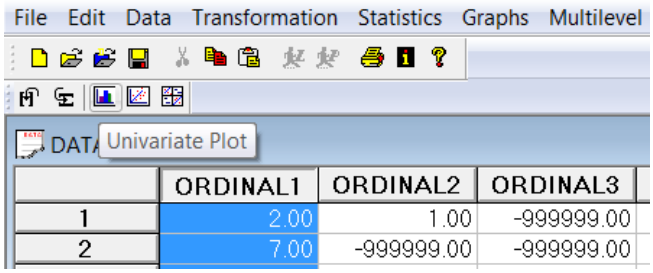
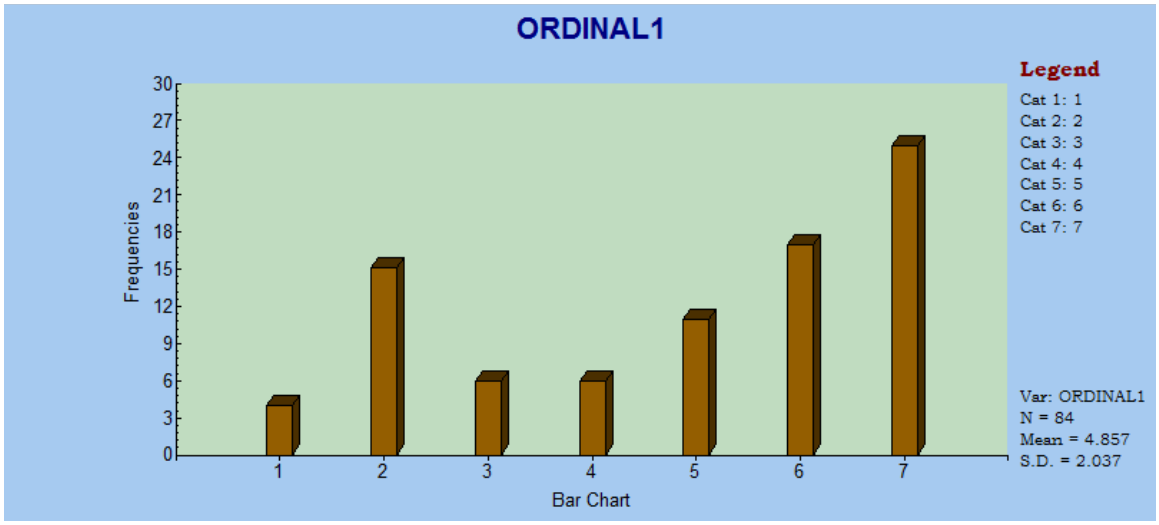
- Tests of multivariate normality are only performed on continuous variables.
- PRELIS computes a polychoric correlation matrix if the variables are ordinal.
- When saving data under a filename with the extension ***.lsf**, a LISREL system data file with that name is created.



Click **OK** when done. This action will return control to the **Select Data** dialog box. If satisfied with the selections made, click **OK** to invoke the PRELIS program. On execution, an output file named **data100.out** is produced. A selection of the output is given.



By closing the output file to return to the LSF window, bar charts can be produced for each of the ordinal variables. Do so by clicking on the rectangle containing the name of the variable and then on the **Univariate Plot** icon button.



A bar chart similar to the one shown next can be obtained by clicking on a vertical bar in the graph to invoke the **Bar Graph Parameters** dialog box. Change the bar type from **Vertical Bars** to **Vertical 3D Bars** to obtain the chart below.