

Multiple imputation

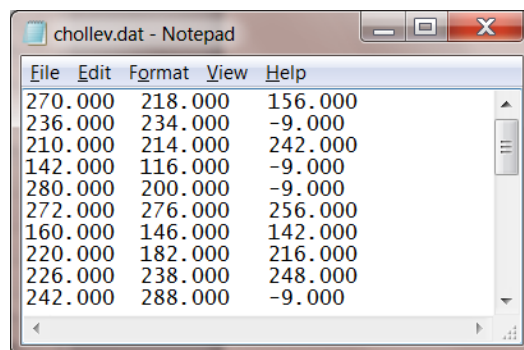
Contents

1. Introduction 1
2. Use of the MCMC Option to obtain more than one imputed data set 5

1. Introduction

Ryan and Joiner (1994, Table 9.1) report serum-cholesterol levels for $n = 28$ patients treated for heart attacks. Cholesterol levels were measured 2 days and 4 days after the attack. For 19 of the 28 patients, an additional measurement was taken 14 days after the attack.

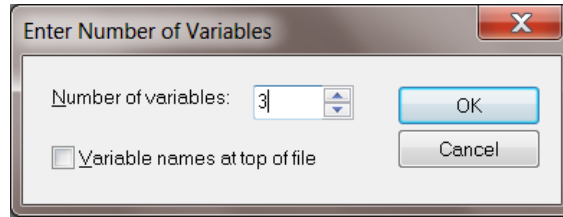
Ten rows of the data set, **chollev.dat** (**Prelis examples** folder), are shown below



File	Edit	Format	View	Help
270.000	218.000	156.000		
236.000	234.000	-9.000		
210.000	214.000	242.000		
142.000	116.000	-9.000		
280.000	200.000	-9.000		
272.000	276.000	256.000		
160.000	146.000	142.000		
220.000	182.000	216.000		
226.000	238.000	248.000		
242.000	288.000	-9.000		

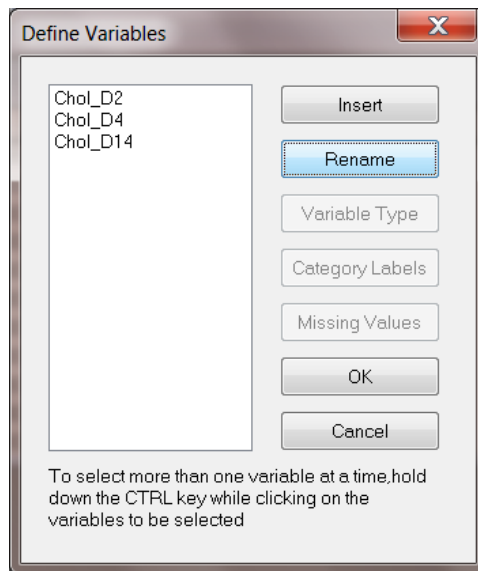
Select the **Import Data...** option from the **File** menu. Select the **tutorial** folder in the **Open** dialog box, select **chollev.dat** and as file type **Free Format Data (*.dat,*.raw)** then click **Open**. A **Save As** dialog is produced by clicking the **Open** button. Save the file as **chollev.LSF**.

Enter the number of variables in the **Number** string field of the **Enter Number of Variables** dialog box.

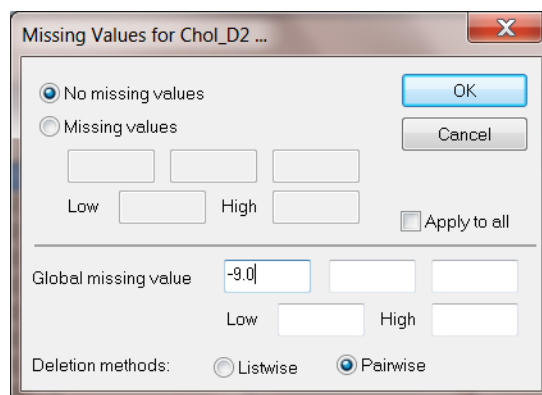


Click **OK** to create a LISREL system data file. From the main menu bar, select the **Define Variables** option from the **Data** menu to obtain the **Define Variables** dialog box.

Use the dialog box to rename VAR1, VAR2 and VAR3 to Chol_D2, Chol_D4 and Chol_D14 respectively. Click **Variable Type** to define these variables as continuous.



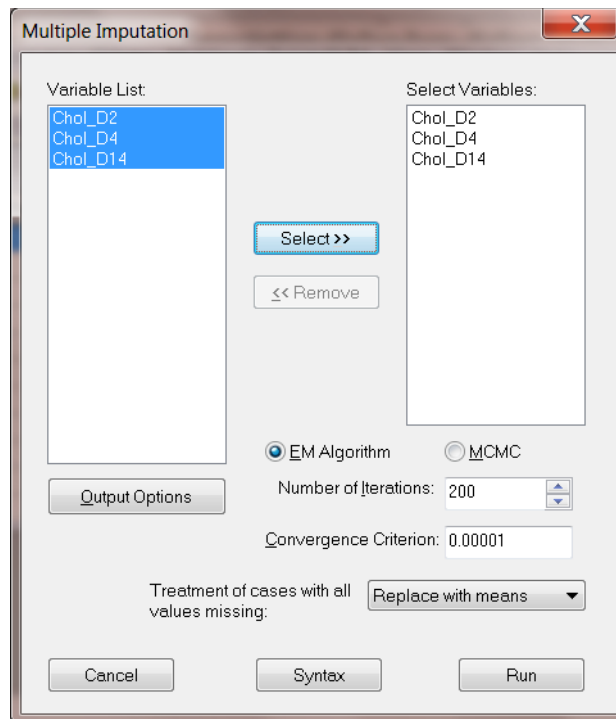
When done, click **Missing Values** to assign the value of -9.000 as the global missing value and select pairwise deletion. Make sure that all the changes are saved by using the **File, Save** option.



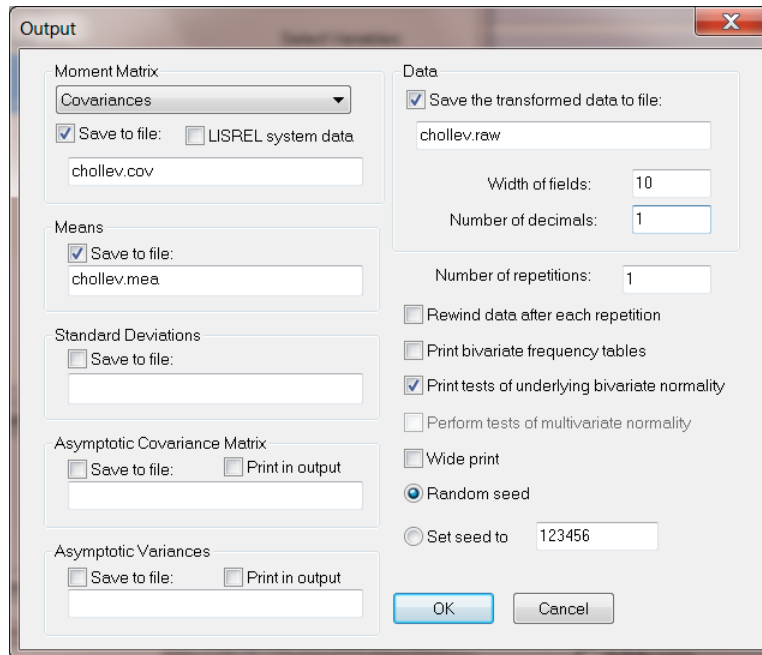
The first five rows of the **.lsf** file are shown below.

	Chol_D2	Chol_D4	Chol_D14
1	270.00	218.00	156.00
2	236.00	234.00	-9.00
3	210.00	214.00	242.00
4	142.00	116.00	-9.00
5	280.00	200.00	-9.00

Select the **Multiple Imputation** option on the **Statistics** menu to obtain the **Multiple Imputation** dialog box. Select all the variables from the **Variables List**.



To impute data using the EM Algorithm, click the **EM Algorithm** radio button. To impute data using the MCMC algorithm, click the **MCMC Algorithm** radio button. Select the number of iterations and the convergence criterion.



To save the imputed data, means, and covariances to external files, click **Output Options** to obtain the **Output** dialog box. Enter the values shown Above. Click **OK** when done to return to the **Multiple Imputation** dialog box.

PRELIS syntax, generated by clicking **Syntax** in the **Multiple Imputation** dialog box, is shown below.

```

L CHOLLEV.PRL
!PRELIS SYNTAX: Can be edited
SY = CHOLLEV.LSF

SE 1 2 3
EM CC = 0.00001 IT = 200 TC = 0
OU MA=CM SM=CHOLLEV.COV RA=CHOLLEV.RAW WI=10 ND=1 ME=CHOLLEV.MEA XM

```

A portion of the PRELIS output is shown below.

```

-----
EM Algorithm for missing Data:
-----

Number of different missing-value patterns=      2
Convergence of EM-algorithm in      4 iterations
-2 Ln(L) =      753.83093

Percentage missing values=  10.71

Estimated Means

Chol_D2      253.9286
Chol_D4      230.6429
Chol_D14     222.2372

Estimated Covariances

                Chol_D2      Chol_D4      Chol_D14
Chol_D2      2276.2910
Chol_D4      1508.4921      2205.9418
Chol_D14      866.3386      1571.5957      2024.5375

```

The percentage of missing cases is 10.71 and the EM procedure converged in 4 iterations.

The first 10 records of the imputed data set, written to a file named **imputed_data.lsf** (the default name), are shown below.

	Chol_D2	Chol_D4	Chol_D14
1	270.00	218.00	156.00
2	236.00	234.00	228.00
3	210.00	214.00	242.00
4	142.00	116.00	146.00
5	280.00	200.00	193.00
6	272.00	276.00	256.00
7	160.00	146.00	142.00
8	220.00	182.00	216.00
9	226.00	238.00	248.00
10	242.00	288.00	272.00

2. Use of the MCMC Option to obtain more than one imputed data set

In contrast to the EM algorithm, the MCMC algorithm is based on random draws from multivariate normal and inverse Wishart distributions. This fact enables the researcher to obtain more than one imputed data set.

This may be accomplished by selecting the MCMC algorithm on the **Multiple Imputation** dialog box. Once this is done, click on **Output Options** to obtain the dialog box shown below.

Output

Moment Matrix
Covariances
 Save to file: LISREL system data

Means
 Save to file:

Standard Deviations
 Save to file:

Asymptotic Covariance Matrix
 Save to file: Print in output

Asymptotic Variances
 Save to file: Print in output

Data
 Save the transformed data to file:
chol_new.lsf
Width of fields: 15
Number of decimals: 6
Number of repetitions: 100
 Rewind data after each repetition
 Print bivariate frequency tables
 Print tests of underlying bivariate normality
 Perform tests of multivariate normality
 Wide print
 Random seed
 Set seed to 8735

OK Cancel

Enter the **Number of repetitions** and enter an integer value into the **Set seed to** field. For this example, the number of repetitions equals 100 and the seed is set to 8735. Note that the data set **chol_new.lsf** will contain $28 \times 100 = 2800$ cases. Click **OK** to return to the **Multiple Imputation** dialog box. If the **Syntax** button is clicked, the PRELIS syntax file shown below is generated.

```
!PRELIS SYNTAX: Can be edited
SY = CHOLLEV.LSF

SE 1 2 3
MC CC = 0.00001 IT = 200 TC = 0
OU MA=CM RA=CHOL_NEW.LSF IX=8735 RP=100 XT XM
```

The contents of the file **chol_new.lsf** are displayed below for cases number 2791 to 2800.

The image shows a screenshot of a spreadsheet application window titled "CHOL_NEW.LSF". The window contains a table with the following data:

	Chol_D2	Chol_D4	Chol_D14
2791	276.00	220.00	188.00
2792	282.00	186.00	182.00
2793	360.00	352.00	294.00
2794	310.00	202.00	214.00
2795	280.00	218.00	202.00
2796	278.00	248.00	198.00
2797	288.00	278.00	235.00
2798	288.00	248.00	256.00
2799	244.00	270.00	280.00
2800	236.00	242.00	204.00