



## Data manipulation and bivariate plots

### Contents

1. INTRODUCTION.....	1
2. ASSIGNING LABELS TO CATEGORIES OF ORDINAL VARIABLES.....	2
3. COMPUTE VALUES FOR A NEW VARIABLE.....	3
4. CREATING A SUBSET OF THE DATA.....	4
5. BIVARIATE PLOTS.....	6
6. CALCULATE AND SAVE COVARIANCES AND MEANS FOR SUBGROUPS.....	9

### 1. Introduction

In this example, use is made of the Fitness/Cholesterol data described in detail in the previous section. The data set contains 8 variables, these being Group, Age, Height, Weight, %Fat, Strength, Trigl and Cholest.

The categorical variable Group has 4 values, these being 1, 2, 3, and 4, denoting weight lifters ( $n_1 = 17$ ), students ( $n_2 = 20$ ), marathon athletes ( $n_3 = 20$ ) and coronary patients ( $n_4 = 9$ ) respectively.

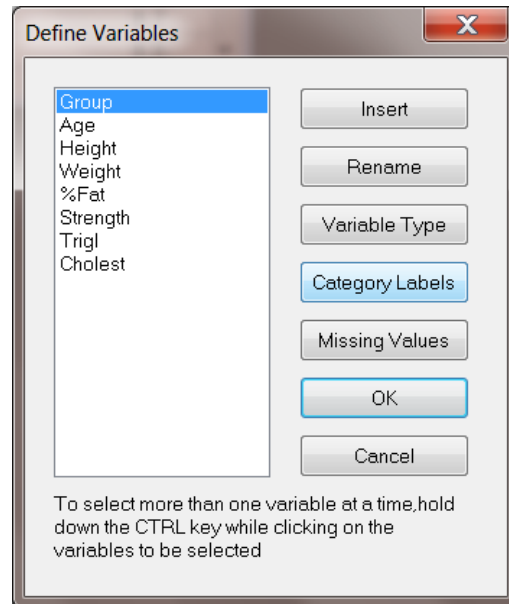
The user interface will be used to

- Assign labels to the categories of Group
- Insert a new variable called Totchol into the data set
- Assign the value Trigl + Cholest to the new variable Totchol
- Create a data subset consisting of the 20 marathon athletes
- Display bivariate plots for the marathon athlete subgroup
- Calculate covariances and means for groups 1 to 3

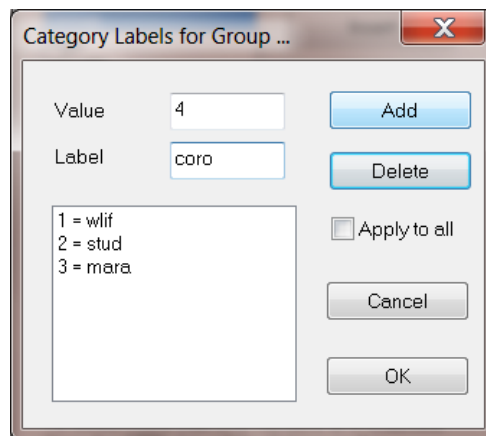
## 2. Assigning Labels to Categories of Ordinal Variables

To display the fitness/cholesterol data, the **File, Open** option is used. Select **Files of type \*.Isf** from the drop-down list box. The file **fitness.Isf** can be found in the **Prelis examples** folder.

From the **Data** menu, select **Define Variables** and highlight the variable **Group** as illustrated in the next image.



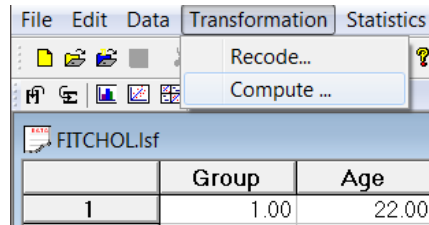
Once this is done, click on the **Category Labels** button.



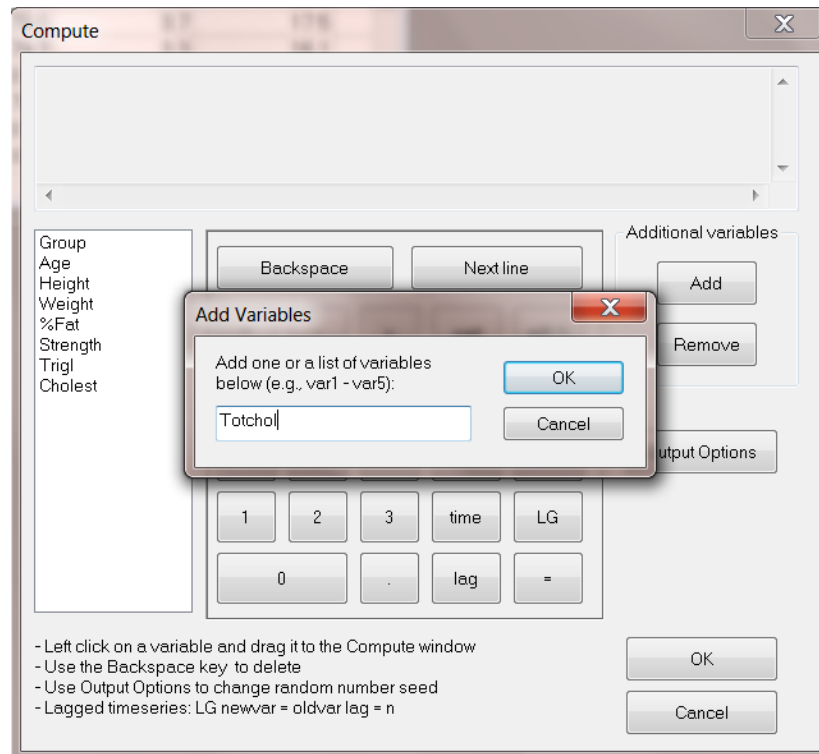
In the value string field, type the value 1 and in the label string field type the value wlif. Click the **Add** button to obtain the code 1 = wlif in the syntax window. Proceed in a similar fashion until 4 = coro has been entered in the code string field. When done, click **OK** to return to the **Define Variables** dialog box.

### 3. Compute Values for a New Variable

To insert and create values for a new variable Totchol, select the **Compute...** option from the **Transformation** menu to activate the dialog box labeled **Compute**.



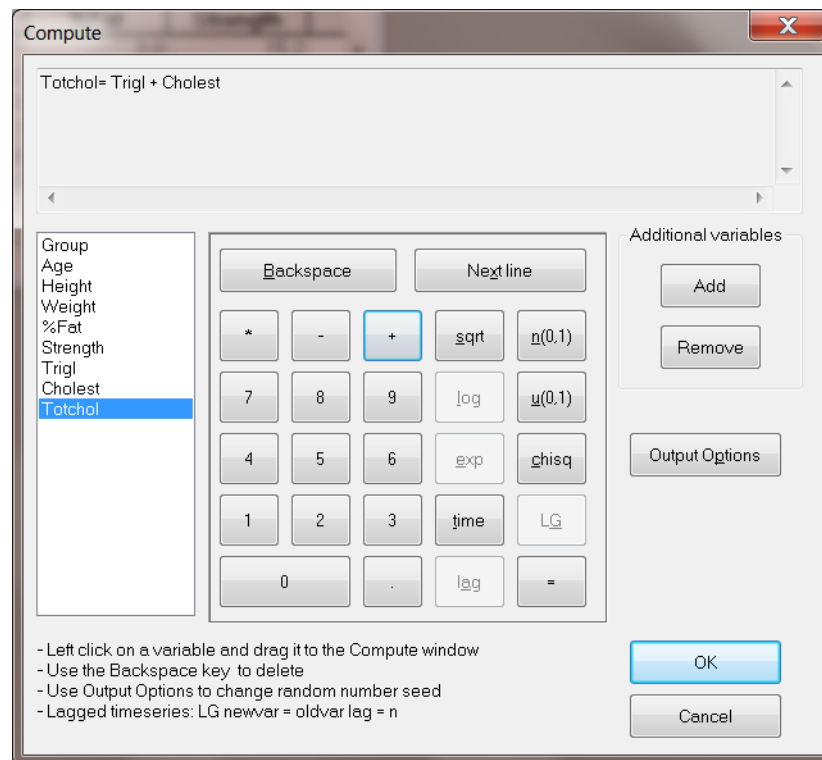
Next click the **Add** button and type Totchol in the text box and, when done, click the **OK** button.



The dialog box below shows that the variable Totchol has been added to the list. We now wish to enter the values  $\text{Totchol} = \text{Trigl} + \text{Cholest}$  for this variable.

The **Compute** dialog box displays the new variable as a function of Trigl and Cholest. To enter this equation, click on Totchol (left mouse button). While holding the mouse button down, drag this item to the position shown in the **Compute** dialog box and release the mouse button. Click on the “=” sign, then drag Trigl to the position next to the “=” sign. When this is done, click on the “+” sign and finally drag Cholest to the position shown. Alternatively, one could also double-click on

a variable name to enter the variable into the **Compute** window. Click **OK** to update the \*.Isf file. The result of this computation is shown below.

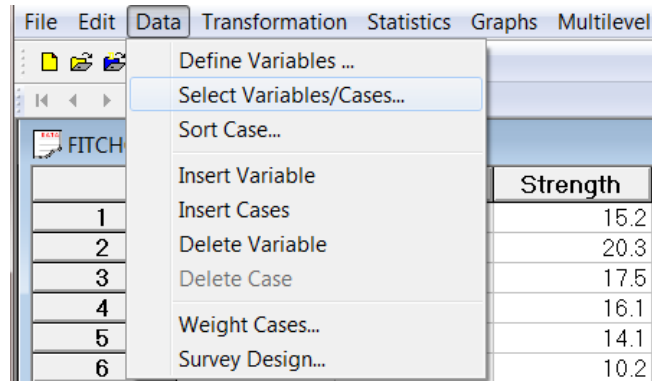


	Height	Weight	%Fat	Strength	Trigl	Cholest	Totchol
1	179.20	107.10	3.00	15.20	0.58	4.44	5.02
2	183.00	112.20	4.60	20.30	1.51	4.88	6.39
3	175.70	78.00	3.70	17.50	1.20	4.33	5.53
4	182.50	79.70	3.30	16.10	0.75	3.66	4.41
5	178.00	81.80	2.70	14.10	0.75	4.57	5.32

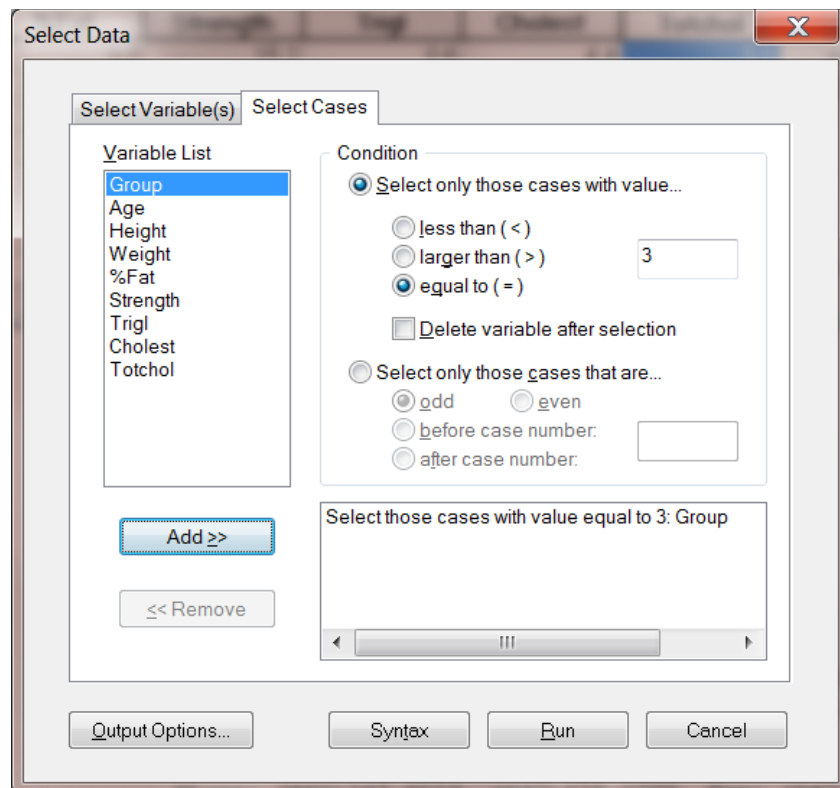
It is also possible to save the data under a new name, e.g. **fitness2.Isf** by clicking the **Output Options** button on the **Compute** dialog box and then checking the **Save the transformed data to file** check box.

## 4. Creating a Subset of the Data

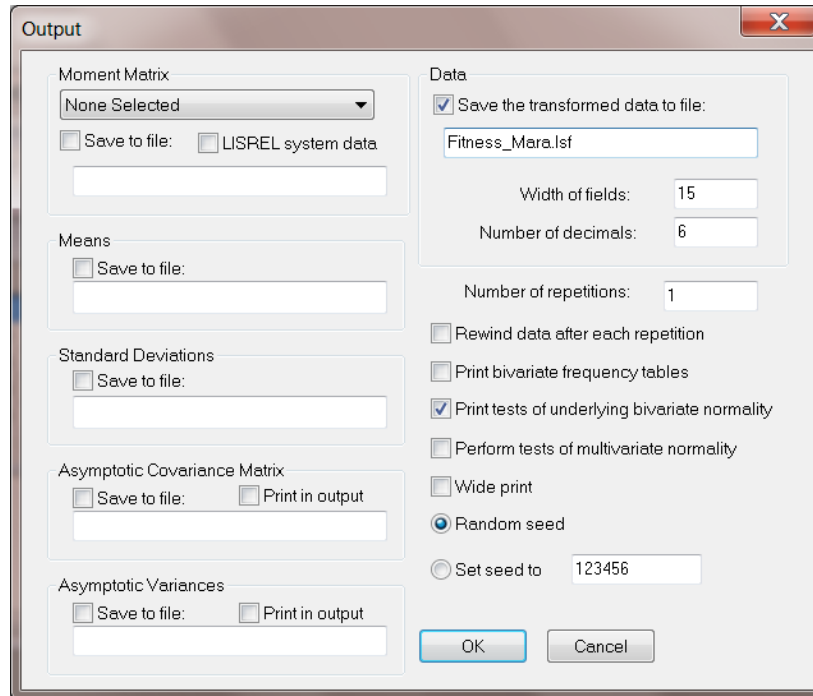
Suppose that we want to focus on the subgroup of marathon athletes (Group = 3). To create a data set consisting of this subgroup, select the **Data, Select Variables/Cases** option as shown below.



The **Select Variables/Cases** dialog box appears.



Highlight the variable **Group** and select the **Select only those cases with value** and **equal to** options. Type in the value of **3** and then click the **Add** button to obtain the words **select those cases with value equal to 3: Group** shown in the window below the **Add** and **Remove** buttons. This syntax can be removed if it is highlighted and the **Remove** button is clicked. Choose the **Output Options** button to save the file under a new name, for example **Fitnessmara.Isf**.



**Note:**

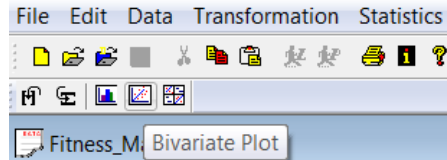
If the file is not saved under a new name, the existing \*.lsf will be overwritten.

When done, click **OK** to start PRELIS. The data subset consisting of Group=3 values only will be created. This spreadsheet, **Fitnessmara.lsf**, is shown below.

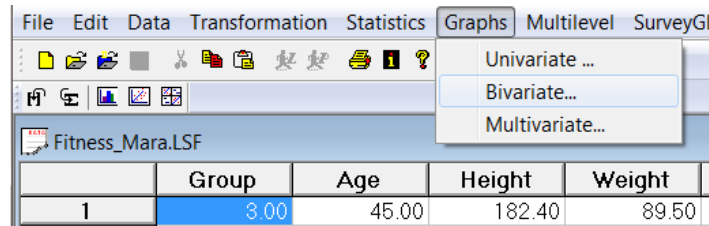
	Group	Age	Height	Weight	%Fat	Strength	Trigl
1	3.00	45.00	182.40	89.50	5.30	21.70	0.86
2	3.00	23.00	176.40	58.80	1.90	10.50	0.84
3	3.00	27.00	169.10	70.70	5.80	24.00	1.15
4	3.00	37.00	173.70	76.40	3.60	16.60	1.12
5	3.00	18.00	190.70	70.70	2.70	13.10	0.72

## 5. Bivariate Plots

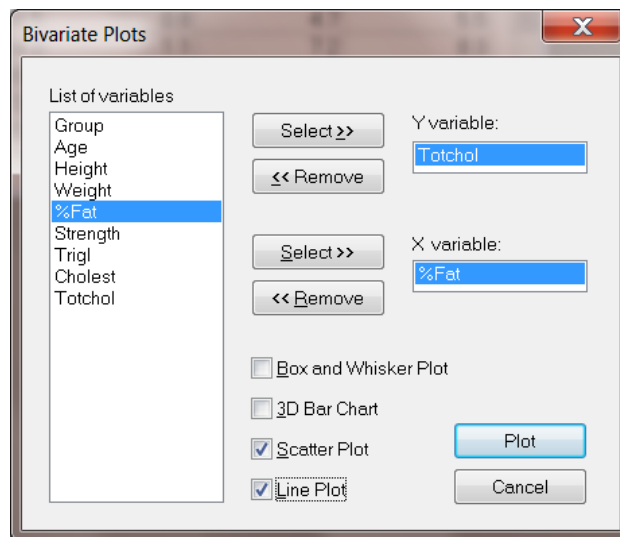
A bivariate plot can be obtained by clicking on the **Bivariate** icon button shown next to the **Univariate** icon button on the second tool bar box below.



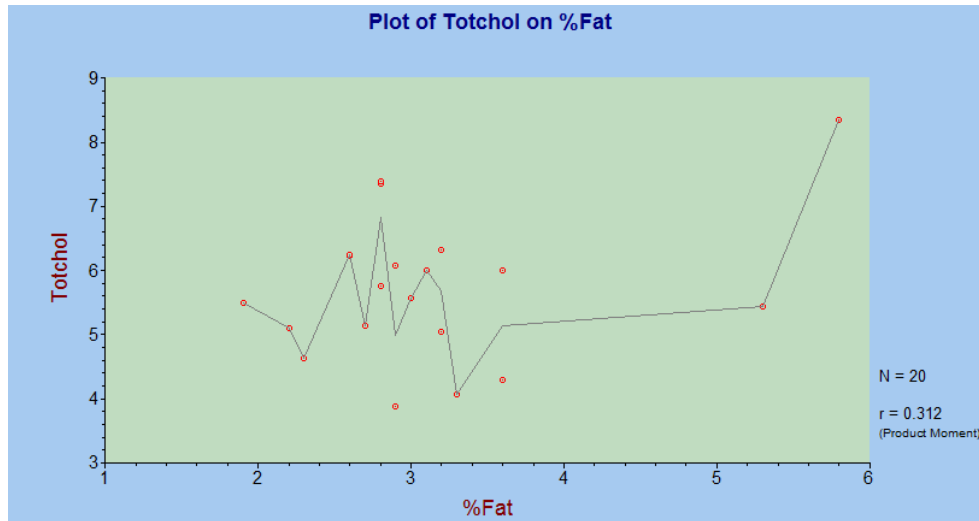
One can also obtain a bivariate plot of two variables by selecting the **Graphs, Bivariate** option to obtain the **Bivariate Plots** dialog box.



Click on Totchol and then click on the **Select (Y variable)** button. Once this is done, highlight %Fat as shown below, and click on the **Select (X variable)** button. Also select the **Scatter Plot** and **Line Plot** options. To produce a bivariate graph, click **Plot**.



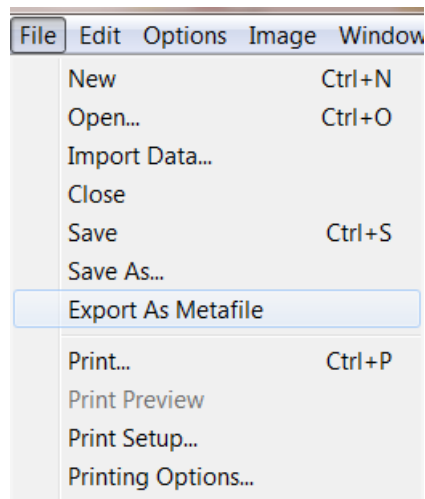
A plot of Totchol on %Fat is obtained, as shown below:



Note that the graph shows the product moment correlation coefficient and the sample size. The line plot is obtained by taking the arithmetic average of all Y-values that correspond to the same X-value. As such, the line plot shows the general trend in the Y-variable with increase in X.

Histograms, bar charts and bivariate plots can be exported as Windows Enhanced Metafiles (\*.emf extension). These files can be imported into any of the Microsoft Office products, and can also be opened with various graphics software packages.

To save a bivariate plot as a \*.emf graphics file, select the **Export as Metafile** option from the **File** menu.



By clicking on **Export as Metafile**, the user is prompted to save the \*.emf file under any name and in any folder.

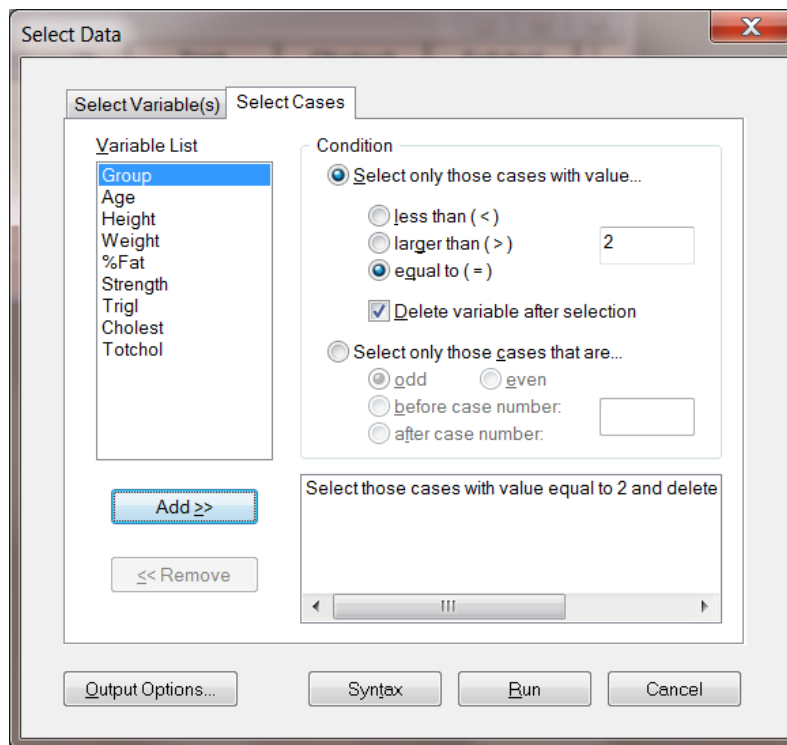


## 6. Calculate and Save Covariances and Means for Subgroups

Suppose that we want to calculate covariances, means and standard deviations for each of the first 3 groups of the fitness/cholesterol data.

The procedure is illustrated by selecting Group = 2 (2 = students). Select the **Select Variables/Cases** option from the **Data** menu to obtain the **Select Data** dialog box.

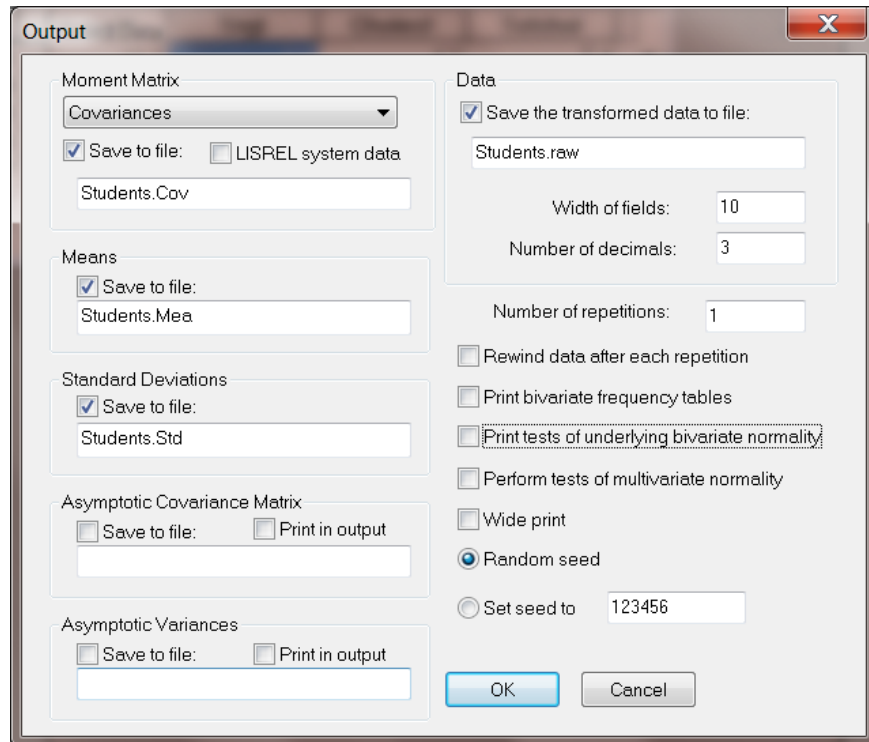
Select the **Select Cases** tab on the **Select Data** dialog box:



Select Group (click on Group with the left mouse button) and then complete the **Select only those cases with value** portion of the dialog box. Since the variable Group will now assume values of 2 only, it should not be included in the calculation of the means and covariances.

If this variable is to be deleted, check the **Delete variables after selection** option. Click the **Add** button to generate the necessary syntax.

Once this is done, the **Output Options** button should be clicked to produce the **Output** dialog box given below.



Complete the dialog box as shown. When done, click **OK** to return to the **Select Cases** dialog box. Click **OK** to start PRELIS.

A portion of the PRELIS output is given below:

Fitchol.PR2.OUT

```
SD 1 = 2
OU MA=CM SM=student.cov RA=student.raw WI=10 ND=3 ME=student.mea SD=student.std
XT XM
```

Total Sample Size = 20

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Freq.	Maximum	Freq.
Age	26.050	4.442	0.592	-0.826	21.000	3	35.000	1
Height	179.760	6.183	0.160	-0.251	168.200	1	192.600	1
Weight	73.070	8.317	0.658	1.146	60.700	1	95.000	1
%Fat	3.260	0.955	-0.243	-0.803	1.400	1	4.700	2
Strength	15.615	3.514	-0.644	-0.167	7.600	1	20.400	2
Trigl	1.051	0.621	2.078	3.730	0.460	1	2.720	1
Cholest	4.508	1.031	0.041	-0.993	2.870	1	6.140	1
Totchol	5.559	1.498	0.808	0.116	3.330	1	8.700	1