



Exploratory analysis of Fitness data

Contents

1. INTRODUCTION.....	1
2. DESCRIPTION OF THE FITNESS/CHOLESTEROL DATA.....	1
3. EXPLORATORY DATA ANALYSIS.....	3
4. GRAPHICAL DISPLAYS.....	6
5. REGRESSION ANALYSIS.....	10

1. Introduction

Once a *.lsf file is opened, the LSF toolbar appears and a number of additional dialog boxes become accessible to the user. PRELIS syntax can be generated by making use of these dialog boxes. As an illustration, we use a data set called **fitchol.dat** (Du Toit, Steyn and Stumpf, 1986.)

2. Description of the Fitness/Cholesterol Data

Two of the many factors that are known to have some influence or relevance on the condition of the human heart are physical fitness and blood cholesterol level. In a related research project, four different homogeneous groups of adult males were considered. A number of plasma lipid parameters were measured on each of the 66 individuals and fitness parameters were also measured on three of the four groups.

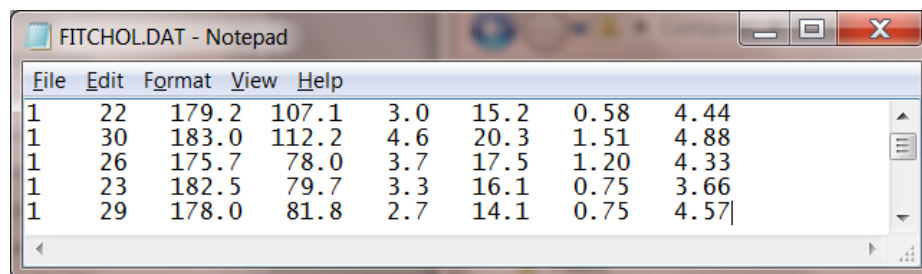
The groups are:

Group	Description
1	Weightlifter (n1 = 17)
2	Student (control; n2 = 20)
3	Marathon athlete (n3 = 20)
4	Coronary patient (n4 = 9)

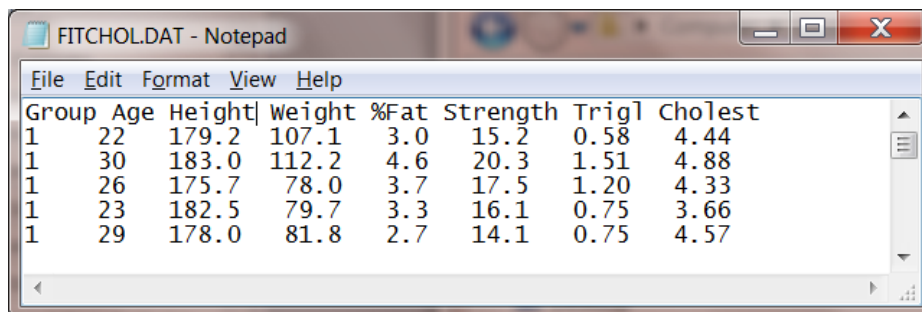
The characteristics that we will consider here are:

Variable	Label
X1 = Age (years)	Age
X2 = Height (cm)	Len
X3 = Weight (kg)	Weight
X4 = Percentage fat	%Fat
X5 = Strength-breast (lb)	Strength
X6 = Triglycerides	Trigl
X7 = Cholesterol (total)	Cholest

The folder name containing the dataset is **Prelis examples** and the file name is **fitchol.dat**. Locate this file and open with Notepad. The first 5 lines of the data set are given below:



Use Notepad (or any other text editor) and insert the variable names on top of the data:

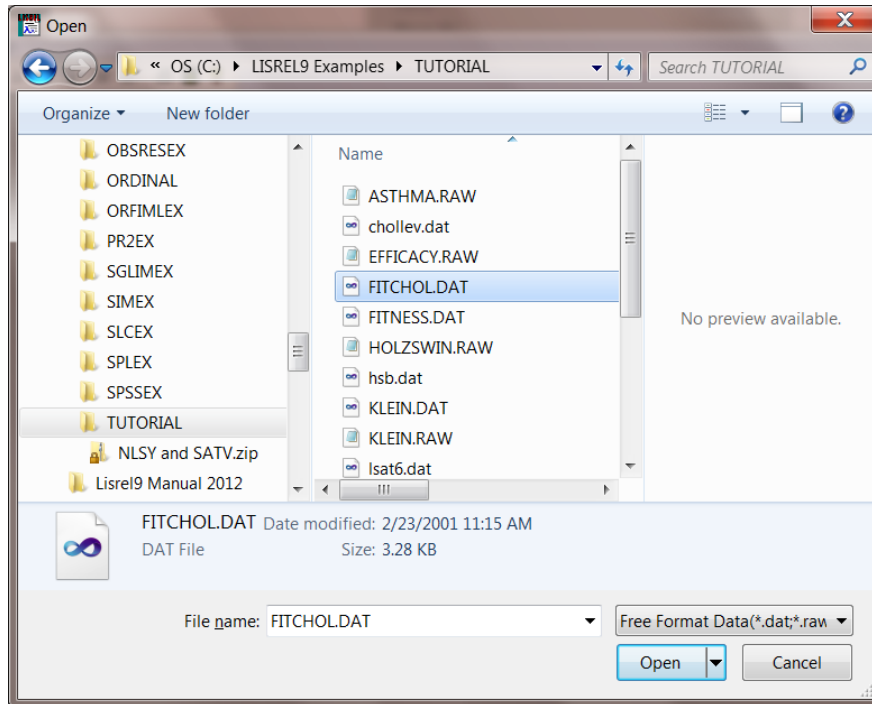


3. Exploratory Data Analysis

From the **File** menu, select **The Import Data...** option

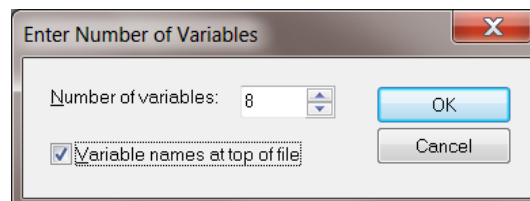
The **Open** dialog box that appears will prompt the user to select the drive and path where the raw data is stored. Locate this file and open with Notepad:

Select **Free Format Data (*.dat, *.raw)** and select the file **fitlchol.dat**. Click **Open** to proceed.



Select a filename and folder in which the LISREL system data should be saved. Note that the extension has to be ***.lsf**.

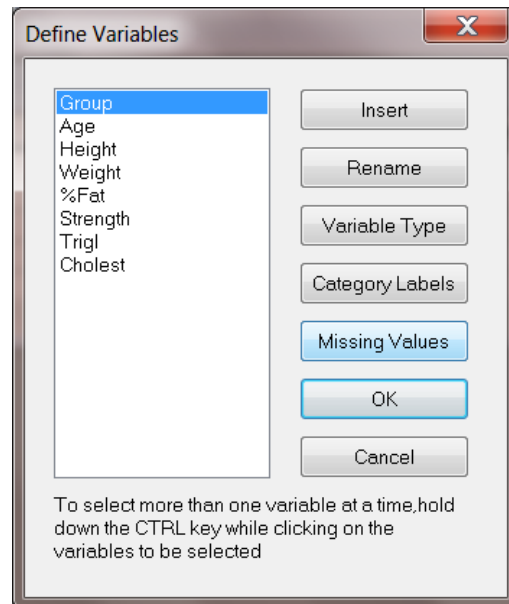
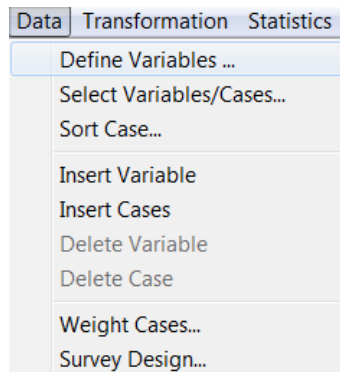
Enter the number of variables in the dataset (see below) and select **Variable names at the top of file**.



When done, click **OK** to open the LSF window in which the newly created ***.lsf** file is displayed.

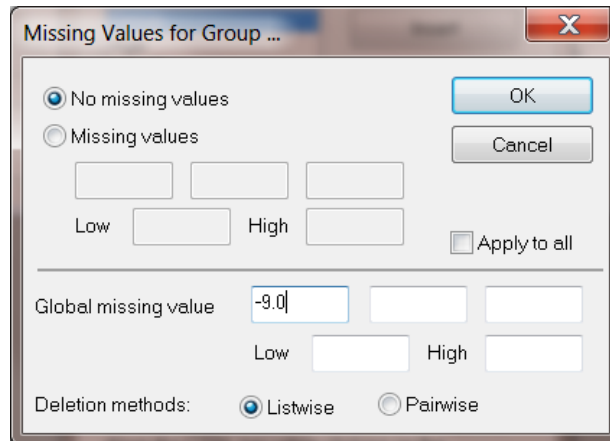
	Group	Age	Height	Weight	%Fat	Strength	Trigl
1	1.00	22.00	179.20	107.10	3.00	15.20	0.58
2	1.00	30.00	183.00	112.20	4.60	20.30	1.51
3	1.00	26.00	175.70	78.00	3.70	17.50	1.20
4	1.00	23.00	182.50	79.70	3.30	16.10	0.75
5	1.00	29.00	178.00	81.80	2.70	14.10	0.75

Since the fitness data set (**fitchol.dat**) contains missing values that are denoted by -9.0, we would like to remove these values prior to doing any exploratory analyses. To do so, select the **Define Variables** option from the **Data** menu to obtain the **Define Variables** dialog box.

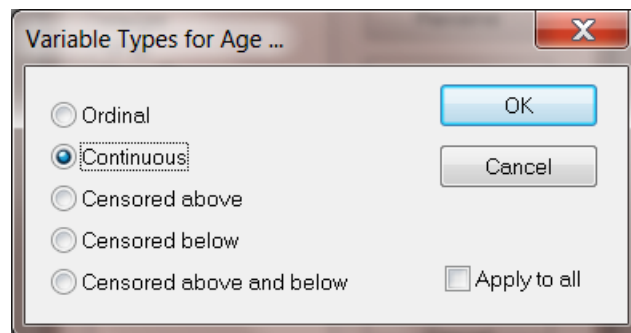


Select the variable in the first row. This action will activate the **Variable Type**, **Category Labels**, **Missing Values** and **Labels** buttons. Click **Missing Values** to obtain the **Missing Values** for dialog box.

On this dialog box, check the **Global Missing values** radio button and then type in the value of -9.0. Click **OK** to return to the **Define Variables** dialog box.



Suppose that we would like to specify all variables (except Group) as continuous. On the **Define Variables** dialog box, select the variables Age to Cholest and then click **Variable Type** to activate the **Variable Type for** dialog box. Check the **Continuous** radio button and then click **OK** to return to the **Define Variables** dialog box.

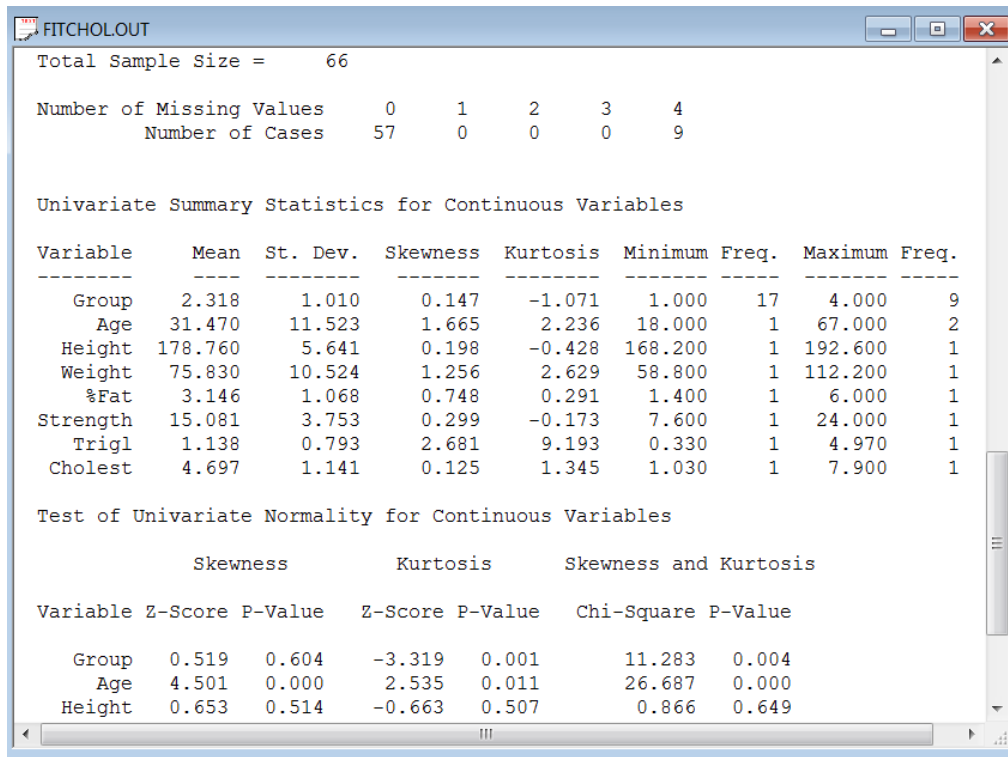


The dialog box may also be used to change the variable labels by clicking on a variable and then **Rename**.

Before fitting an appropriate model to data, it may be wise to run a data screening procedure. The PRELIS data screening procedure provides information on the distribution of missing values, univariate summary statistics and test of univariate normality for continuous variables. The procedure also provides information on the distribution of variables over a number of class intervals.

To do data screening, select **Statistics** from the main menu bar and then the **Data Screening** option from the **Statistics** menu.

A selection of the output file obtained is shown below:

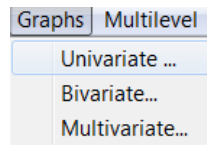


4. Graphical Displays

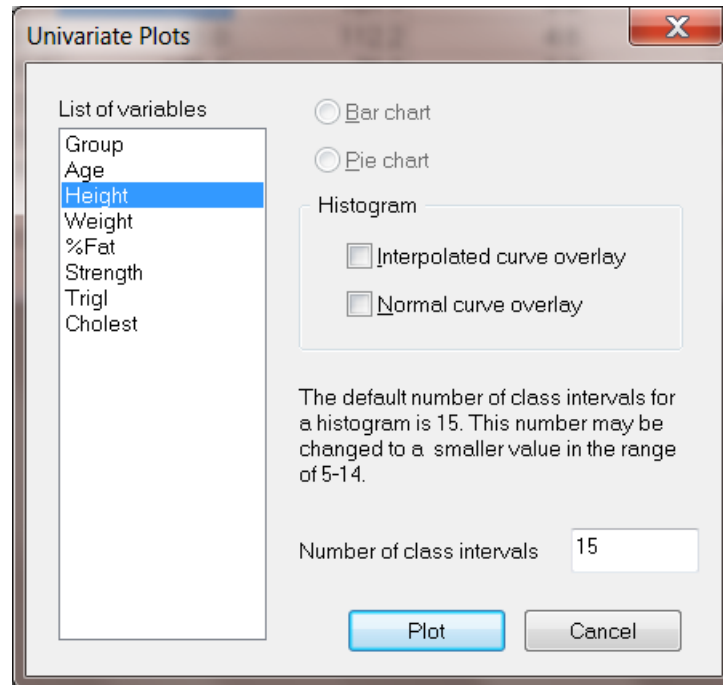
To produce a histogram of the data in the **Height** column, click on the rectangle containing the name **Length**. The column containing the data values will change in color as shown below.

	Group	Age	Height	Weight
1	1.00	22.00	179.20	107.10
2	1.00	30.00	183.00	112.20

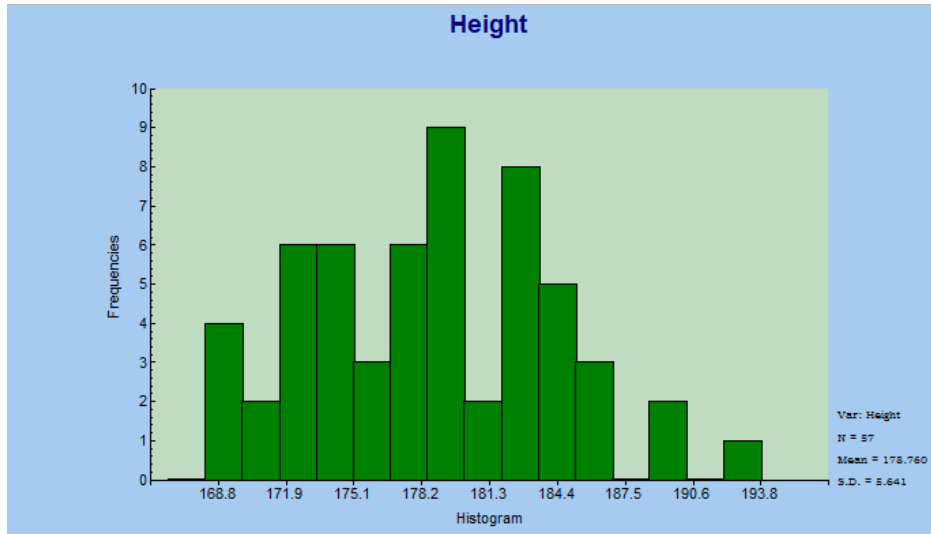
On the main LISREL menu bar, select the **Univariate plot** icon button. Note that, if you move the mouse to this icon, the words Univariate Plot will be displayed in the proximity of the icon button. As an alternative to producing a histogram by clicking on the **Univariate plot** icon button, one may select the **Graphs, Univariate** option on the main menu bar as shown below. Before doing so, make sure that a variable (column) in the spreadsheet is not selected and that the variable is appropriately defined as ordinal or continuous.



To obtain the **Univariate Plots** dialog shown below, make sure that if a column is selected to click on the rectangle containing the corresponding variable label to deselect it.



A histogram of the variable Height is displayed below. The colors of the background and the bars can be changed by double-clicking on the appropriate area. This action will produce a dialog box in which the necessary change may be made. Likewise, one can change the color and appearance of the axes and labels by double clicking on a specific label, for example on **Histogram**. The dialog box allows one to select a font and font size and to change the color of the text. The various dialog box controlling aspects of the graphs are discussed in the *Graphical Users Interface Guide*.

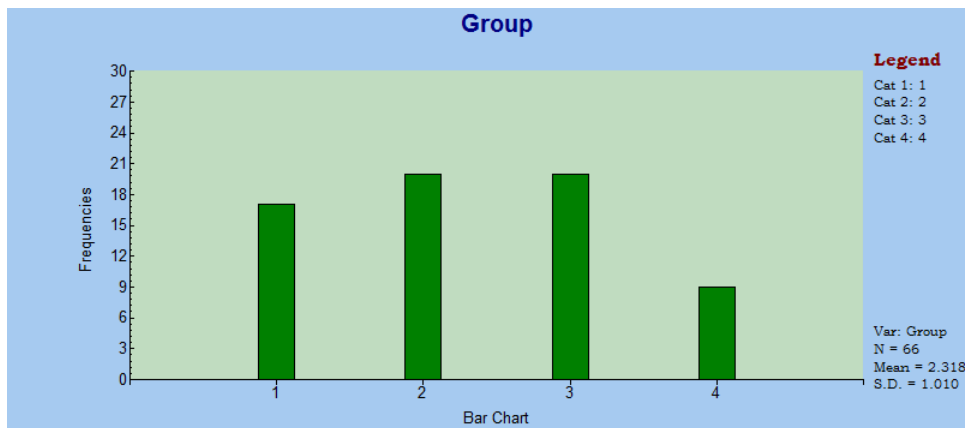


If a variable has less than 16 distinct categories, the **Draw histogram** option produces a bar chart representation of the data. This is the case for the variable Group, which has 4 distinct values, corresponding to

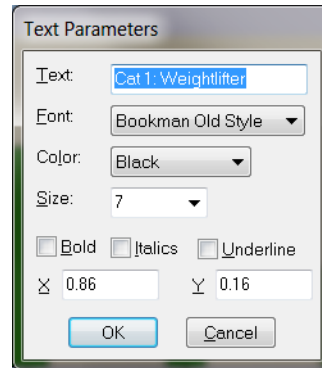
- 1 = Weightlifter
- 2 = Student
- 3 = Marathon athlete
- 4 = Coronary patients

The bar chart below shows that the number of individuals in the first three groups are almost equal ($n = 17, 20$ and 20), but that the sample contains a relatively small number of coronary patients ($n = 9$).

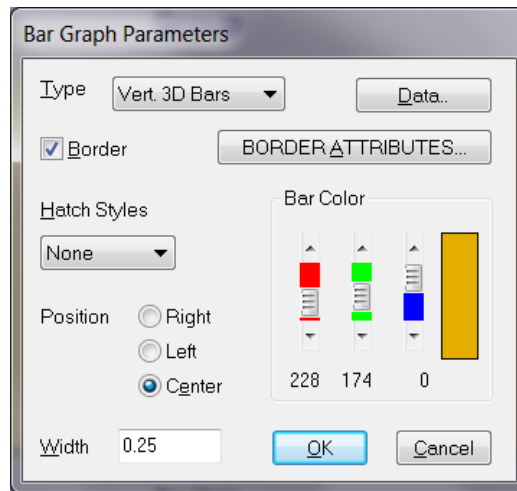
The default legend shows the categories of the variable Group and the values assigned to each category. By double clicking on the Cat1:1 description, the **Text Parameters** dialog box is activated.



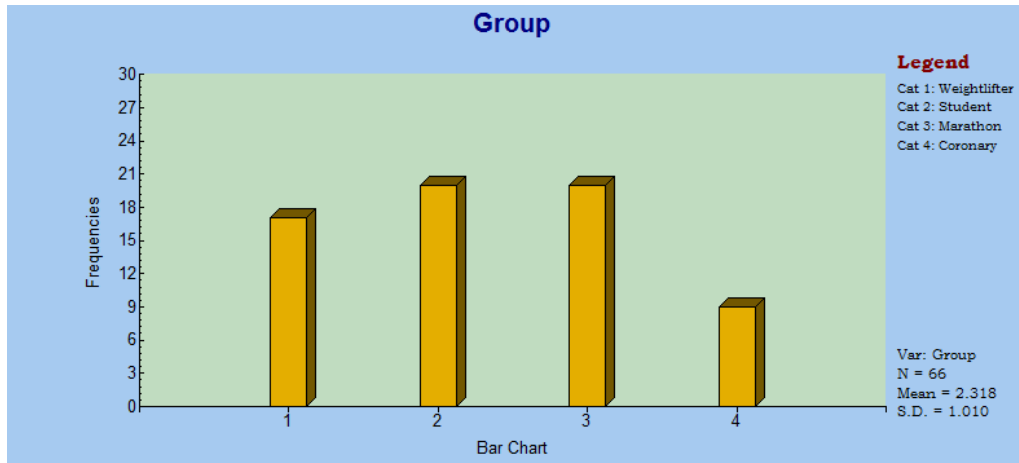
Change Cat1:1 to 1 = Weightlifter and click **OK**. Repeat this procedure to change the descriptions of categories 2, 3, and 4 to Student, Marathon and Coronary, respectively.



The appearance of the vertical bars can be changed by double clicking on one of the bars until the **Bar Graph Parameters** dialog box is activated. Change **Type** to **Vert. 3-D Bars** and use the **bar color** sliders to change to the desired color. Click **OK** when done.

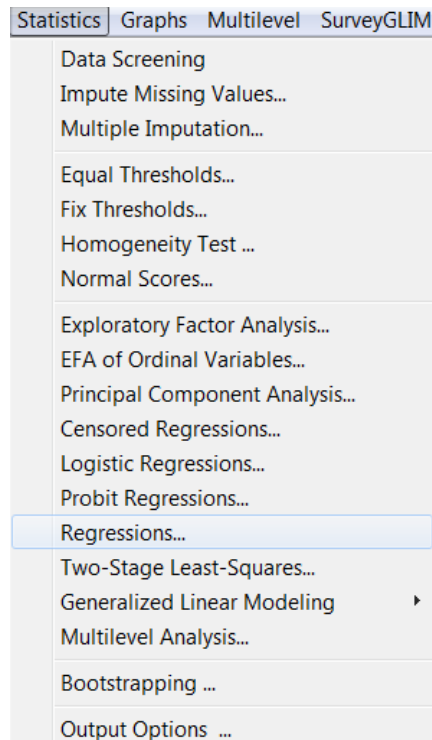


The bar chart shown below reflects these changes in attributes. The bar chart for the variable Group may now be printed. This is done by selecting the **Print** option from the **File** menu. Alternatively, the bar chart can be exported as a Windows metafile (*.wmf) by selecting the **Export as Metafile** option from the **File** menu.



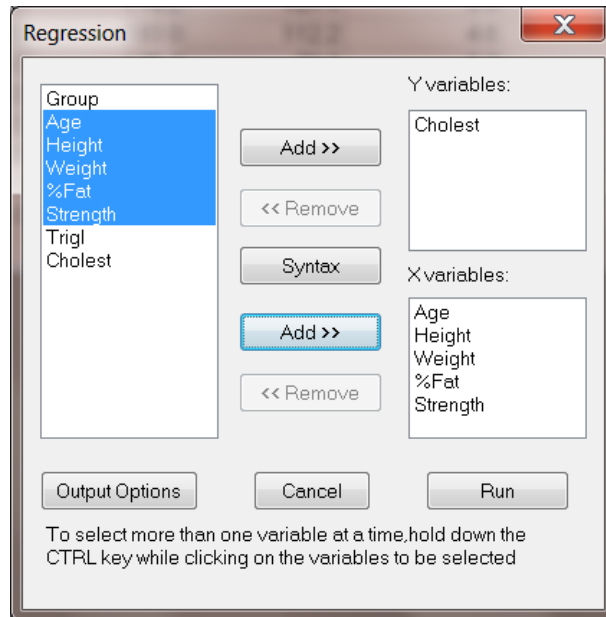
5. Regression Analysis

We may be interested in the prediction of Cholesterol by the variables Age, Height, Percentage Fat and Strength. To obtain the regression equation, select the **Regressions** option from the **Statistics** menu.



The **Regression** dialog box is shown below. This dialog box allows the user to select variables as either Y or X-variables. Click on a variable, for example Cholest, to activate the **Add** button. Click the top **Add** button to add this variable to the list of Y-variables. Select the variables Age, Height,

Weight, %Fat, and Strength and click the bottom **Add** button to add the variable to the list of X-variables. When the selection procedure is completed, click on **Output Options** to activate the **Output** dialog box.



A selection of the output is given below:

