Contents

Data-based graphs: Univariate graphs

Pie Chart

A pie chart display of the percentage distribution of a variable may be obtained if the variable does not have more than 15 distinct values. A pie chart is a graphic representation of percentages or frequencies by means of a circle that is subdivided into slices in such a way that the areas of these slices are proportional to the percentages or frequencies. Pie charts may be customized by using the graph editing dialog boxes as shown in this section.
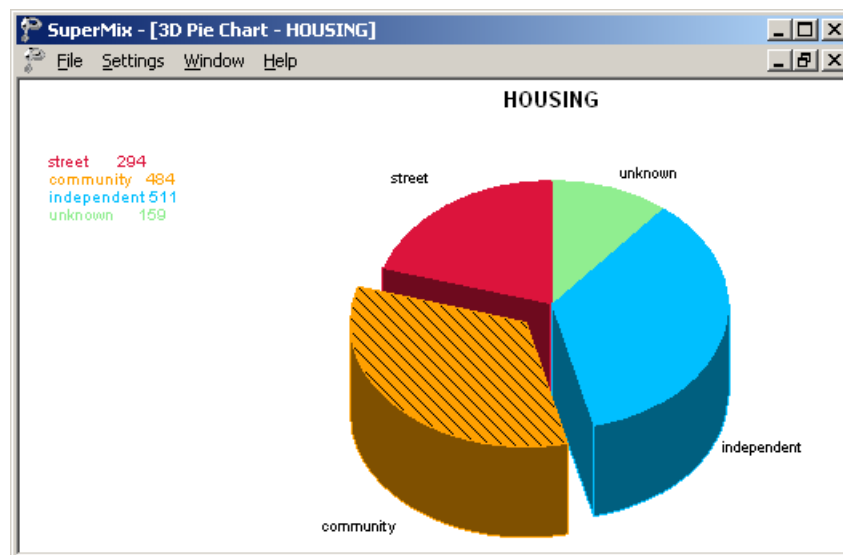

The data from the McKinney Homeless Research Project study (see Section 3.7) were used to produce a pie chart representing the variable HOUSING as given in

Examples\Primer\Graphics\sdhouse.ss3. Recall that subjects' housing status was recorded at 4 time points. Over all time points, 511 subjects reported living in independent housing, 294 subjects were living on the street, 484 lived in community housing, and 159 cases were classified as unknown.

Creating a pie chart

To create the pie chart shown above, start by opening the spreadsheet **sdhouse.ss3**. Next, right-click on the column header of the variable HOUSING and select the **Column Properties** option from the pop-up menu.
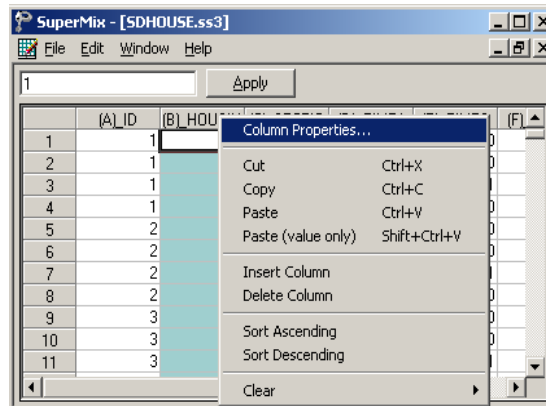

The **Column Properties** dialog box is displayed. This dialog box is used to define the type of variable (categorical or continuous) and to provide labels for the categories of nominal and ordinal variables. Indicate HOUSING as a categorical variable by clicking the appropriate radio button, and enter the labels for each category in the **Label** column as shown below. Click **OK** to return to the spreadsheet window and save the changes to the spreadsheet using the **File**, **Save** option.
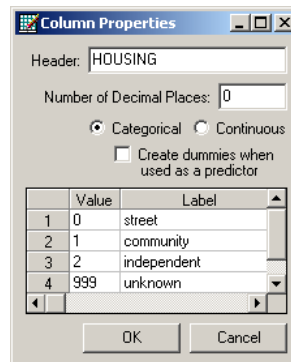

With the variable HOUSING defined as nominal and labels assigned to its categories, proceed to make the pie chart. First, select the **File**, **Data-based Graphs**, **Univariate** option from the pop-up menu as shown below.
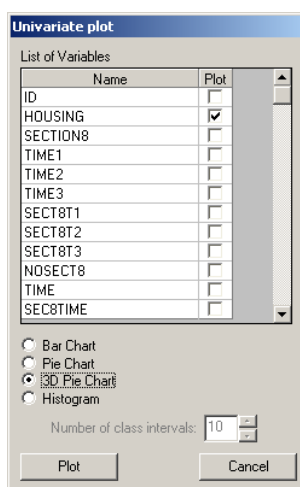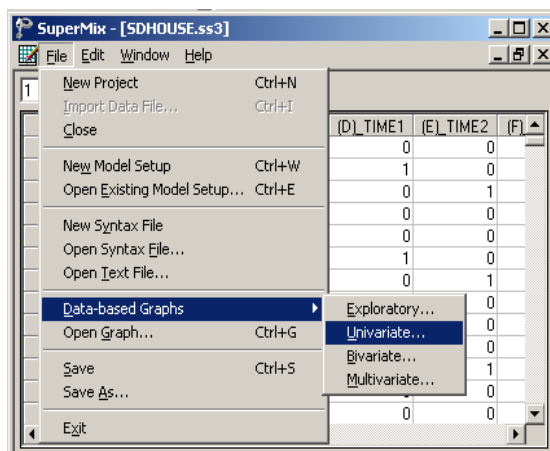
The **Univariate plot** dialog box is displayed. Select the variable HOUSING as the variable to be plotted, and request a **3D Pie Chart** by clicking the radio button next to this option. Click the **Plot** button to display the pie chart.
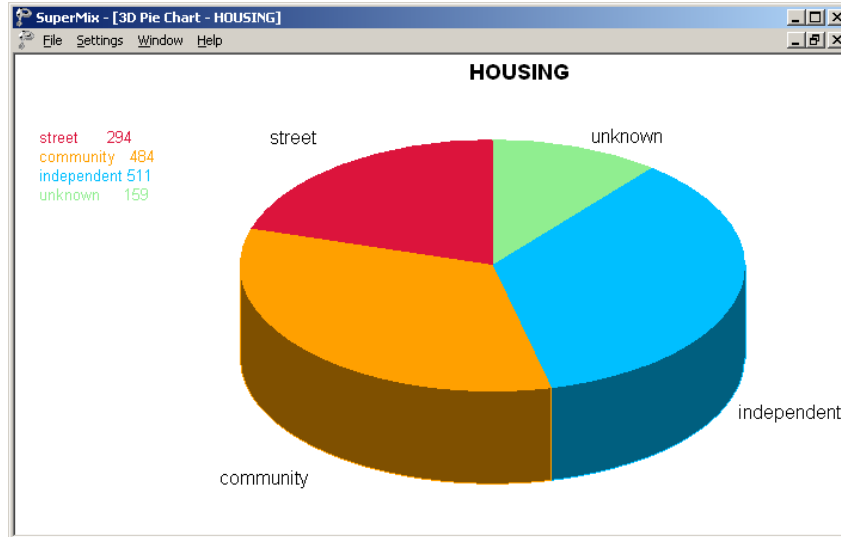
The attributes of the pie chart can be changed by clicking on areas of the graph. For example, clicking on the legend box in the top right corner of the graph will activate the **Legend Parameters** dialog box. This dialog box can be used to change the attributes of the legend box and its contents.
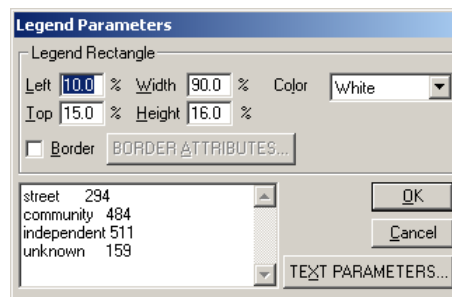

Click on the **Text Parameters** button to change the font and size of text of legend as shown below.
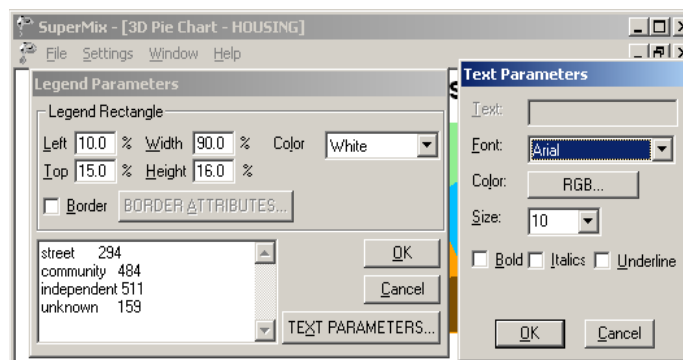
The appearance of the pie chart can be changed by clicking anywhere on the pie chart to activate the **Pie Chart Parameters** dialog box as shown below.
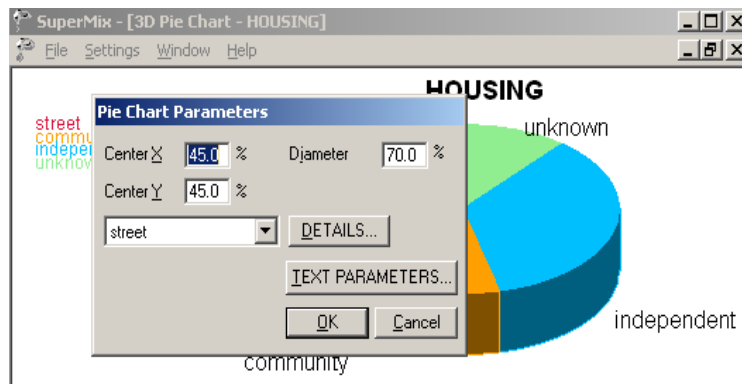
To move a slice of the pie chart, select the category of interest and click the **Details** button. Change the value in the **Explosion** field, for example to 0.2, change the hatch style and color if wanted and click **OK** on both the **Pie Slice Parameters** and **Pie Chart Parameters** dialog boxes to obtain the pie chart shown below.
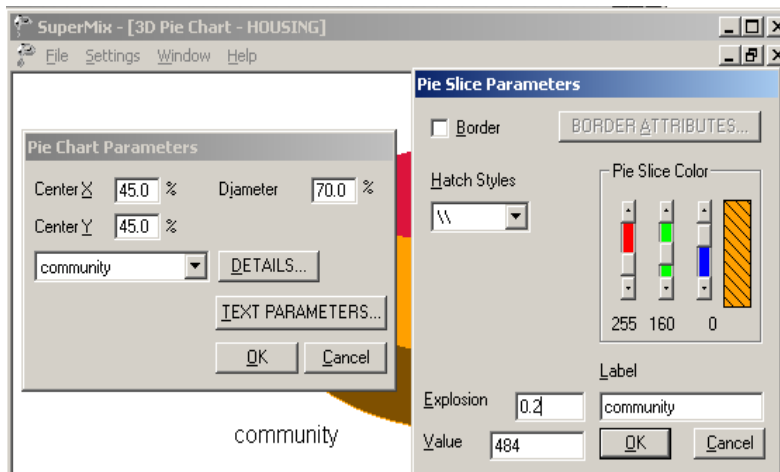
The color of each pie slice can be changed by moving each of the three color sliders shown above.
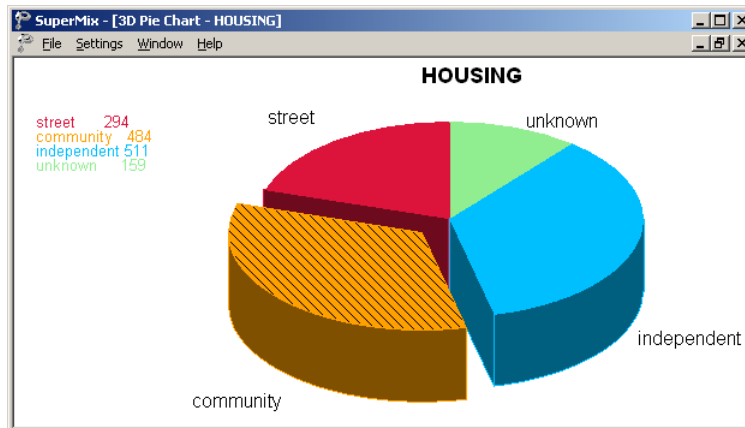
Bar chart

Discrete or categorical data can also be displayed graphically using a bar chart. A bar chart is a graphic representation of the frequency distribution of discrete or categorical data in which the values or categories are given on the horizontal axis and the frequencies are given on the vertical axis.

In Section 3.5 we considered longitudinal data from the NIMH Schizophrenia Collaborative Study on treatment-related changes in overall severity. An ordinal version of the Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS) was used as the outcome variable. The variable Imps79O from the data set **schizx1.ss3** had four categories: 1 = normal or borderline mentally ill, 2 = mildly or moderately ill, 3 = markedly ill, and 4 = severely or among the most extremely ill. This variable can be represented graphically by either a pie or bar chart. In the graph below, we opted to display this variable in the form of a bar chart that offers the added advantage of retaining the ordinality of the variable in question.

Most of the total of 1603 measurements obtained over the course of the study indicated severe mental illness. The second largest group of measurements indicated mild mental illness. While this graph offers an instructive snapshot of the distribution of the ordinal variable Imps79O, it does not allow us to take the longitudinal nature of this study or the treatment group individuals were assigned to into account. To do so, we need to consider bivariate or multivariate graphics instead of univariate graphs (see Sections 4.4 and 4.5 for more information).
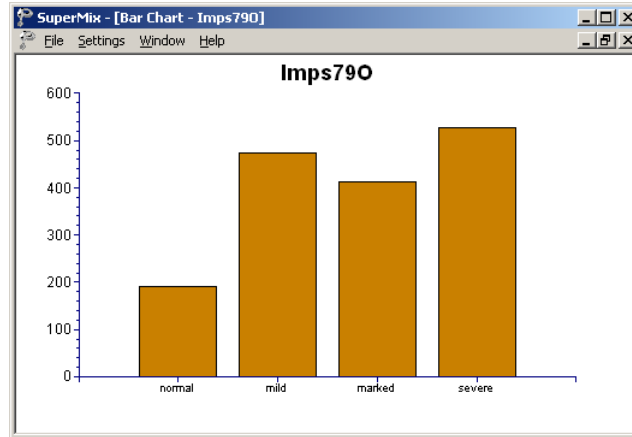
Creating a bar chart

Start by opening **Examples\Primer\Graphics\schizx1.ss3**. Right-click on the column header of the variable Imps79O and select the **Column Properties** option from the pop-up menu to open the **Column Properties** dialog box. This dialog box is used to define the type of variable (categorical or continuous) and to provide labels for the categories of nominal and ordinal variables. Indicate Imps79O as an ordinal variable by clicking the appropriate radio button and enter the labels for each category in the **Label** column as shown below. Click **OK** to return to the spreadsheet window and save the changes to the spreadsheet using the **File**, **Save** option.


Click the **Univariate** option from the pop-up menu displayed when the **File**, **Data-based Graphs** option is selected from the main menu bar.
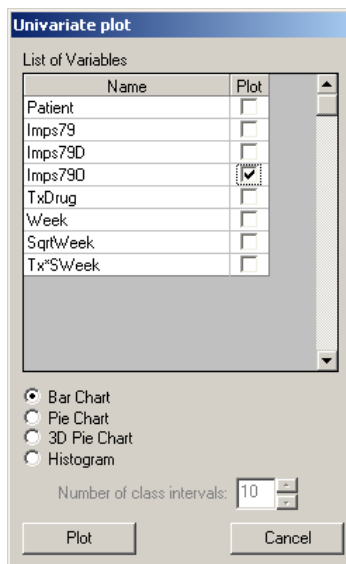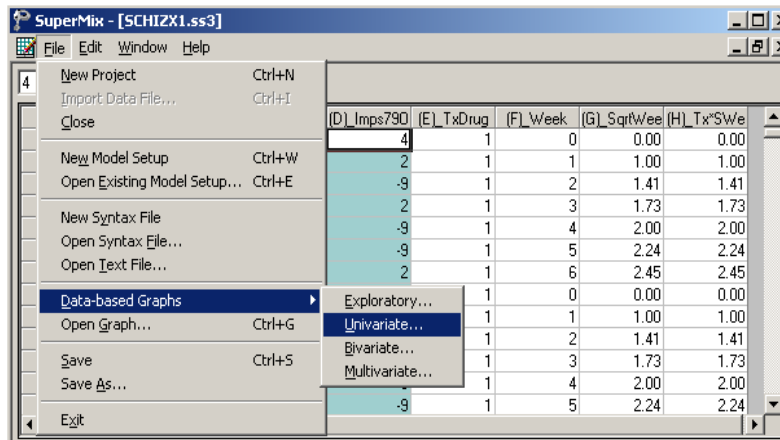
Select the variable Imps79O by checking the box next to this variable in the **Plot** column on the **Univariate plot** dialog box and click **Plot** to display the bar chart.
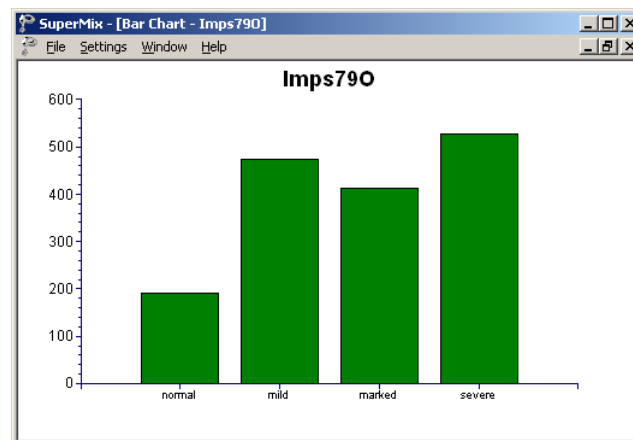
Attributes of the graph are changed by activating the dialog boxes associated with sections of the graph. To change the attributes of the bars, click on a bar in the displayed area to activate the **Bar Graph Parameters** dialog box. This dialog box is usually easy to obtain if the mouse pointer is moved to the top part of the highest bar. The type of bar, color and position of the bars, and the display style can be changed by selecting options from the **Type**, **Bar Color**, **Position** and **Hatch Styles** fields. The **Bar Graph Parameters** dialog box shown below displays the default values of these fields. By clicking the **Data** button, the data plotted – in this case the frequencies associated with each category of the variable Imps79O – will be displayed in a separate window from where it can be copied to the clipboard. Note that if the data contains missing observations, the **Exclude Missing Values** on the **Settings** menu can be used to display a graph with the missing values excluded.
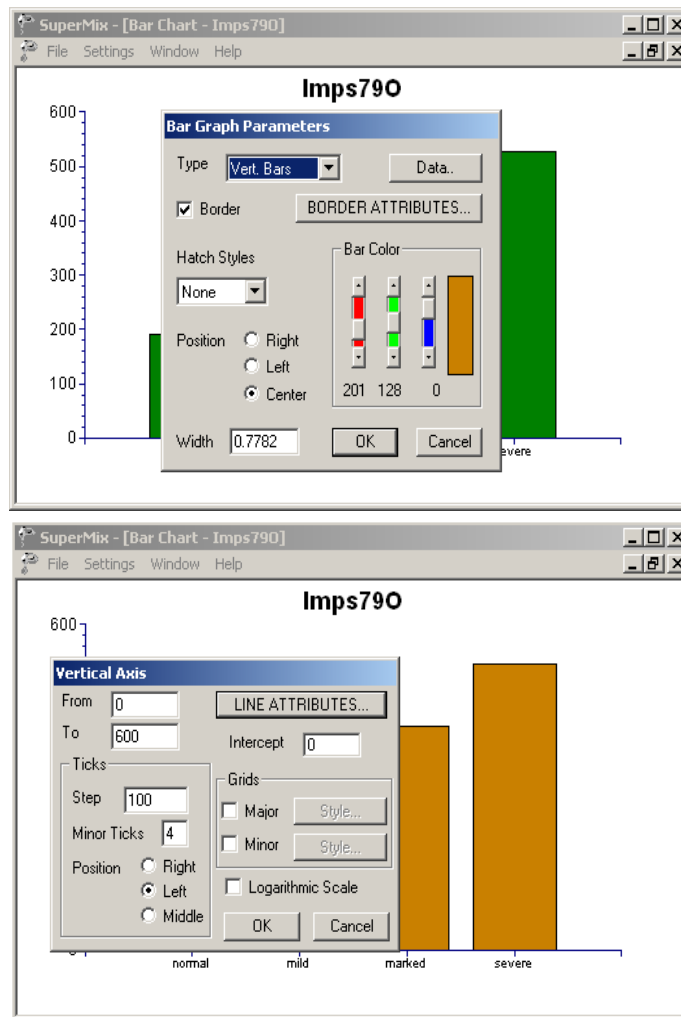
The attributes of the axes can be changed in a similar way: click anywhere on the vertical axis to display the **Vertical Axis** dialog box (shown below), or on the horizontal axis to display the **Horizontal Axis** dialog box. These dialog boxes are used to control the range, tick marks on and scale of each axis. The **Line Attributes** button provides access to the **Line Parameters** dialog box, used to change color, style and width of the axis.

Histogram

The distribution of a continuous variable such as the HDRS ratings in the Reisby data given in **Examples\Primer\Graphics\reisby.ss3** (see Section 3.2 for detailed analyses based on this data) can be graphically depicted by a histogram. Recall that these data are from a study described in Reisby *et. al.*, (1977) that focused on the longitudinal relationship between imipramine (IMI) and desipramine (DMI) plasma levels and clinical response in depressed inpatients over a period of four weeks. The histogram offers a simple way to evaluate the distribution of the variable in question, and is a plot of the frequencies with which values occur (so-called class frequencies), against class intervals.

In the case of the HDRS ratings, we can easily verify that the mean HDRS rating is 17.637, with a standard deviation of 7.19. Ratings range between 0.0 and 39.0. The ratings can be summarized as shown in Table 4.1.

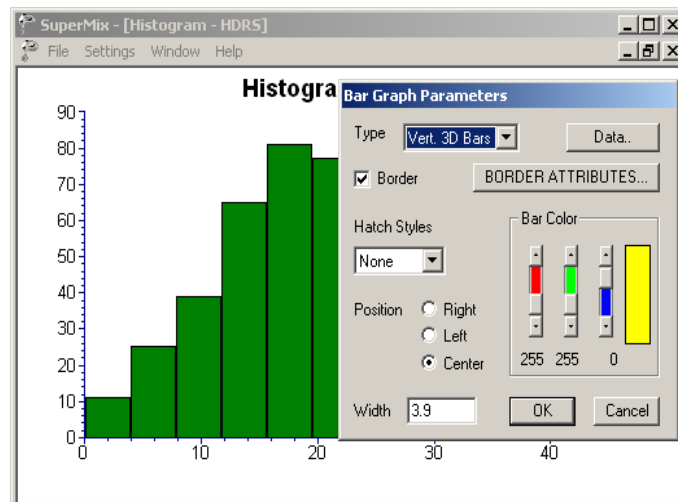**Table 4.1: Frequency distribution of HDRS ratings**

Class interval	Class midpoint	Class frequency	Percentage
[0.0, 3.9)	1.95	11	2.9
[3.9, 7.8)	5.85	25	6.7
[7.8, 11.7)	9.75	39	10.4
[11.7, 15.6)	13.65	65	17.3
[15.6, 19.5)	17.55	81	21.6
[19.5, 23.4)	21.45	77	20.5
[23.4, 27.3)	25.35	49	13.1
[27.3, 31.2)	29.25	19	5.1
[31.2, 35.1)	33.15	8	2.1
[35.1, 39.0)	37.05	1	0.3

The ratings were grouped into 10 classes, each 3.9 in length as indicated in the first column. The second column shows the midpoint of each of the 10 classes, while the third and fourth indicate the frequency and percentage of HDRS ratings that fall within a given class.

The graphical representation of the information in Table 4.1 is obtained by first selecting the **File**, **Data-based Graphs**, **Univariate** option, then selecting the variable HDRS and clicking the **Histogram** radio button, keeping the default **Number of class intervals** 10 as shown below.
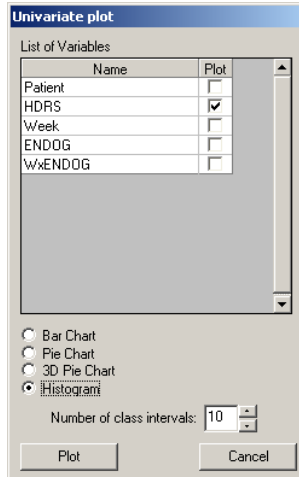
The appearance of the histogram is changed to display vertical 3D bars by clicking on the histogram to activate the **Bar Graph Parameters** dialog box, and by changing the default settings to those shown.

The 3D bar chart of the distribution of HDRS ratings is shown below.
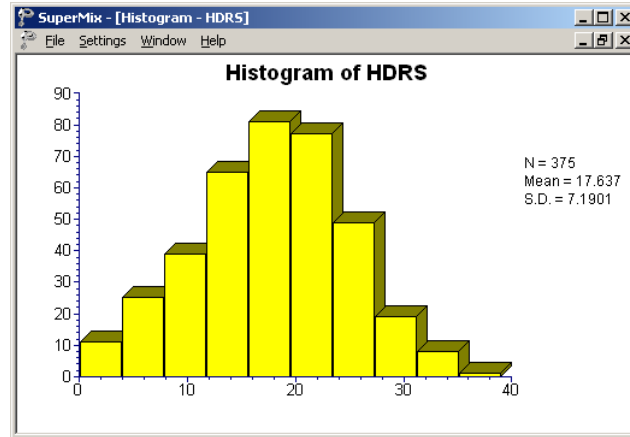
A histogram usually has the following characteristics:

o   The class frequencies are plotted on the y-axis and the class intervals according to a convenient scale on the x-axis.
o   With the class intervals as a basis, rectangles are drawn so that the area of each rectangle is proportional to the corresponding class frequency.
o   It is common practice to leave spaces that are equal to class width on the left-hand side of the smallest class boundary and on the right-hand side of the largest class boundary.

From the histogram above, we conclude that the HDRS ratings are reasonably symmetrically distributed around the mean value of 17.637. In addition, very little data are found at the extreme left or right of the presentation. In general, the bell-shaped curve with short tails for the ratings is very similar to the familiar bell curve of a normal distribution. Treating the HDRS ratings as a normally distributed continuous variable in our previous analyses thus seems reasonable.