Bivariate Density Statlet

statgraphics 18® centurion

Revised: 10/10/2017

Summary

The **Bivariate Density Statlet** creates a frequency histogram for 2 columns of numeric data. It is used to visualize the joint distribution of 2 random variables. The controls on the toolbar make it easy to change the definition of the classes into which the data are grouped. The density function of a bivariate normal distribution with the same means, standard deviations, and covariance as the data may be created instead of a histogram. In addition, a nonparametric density estimator may be drawn.

Sample StatFolio: *bivariatestatlet.sgp*

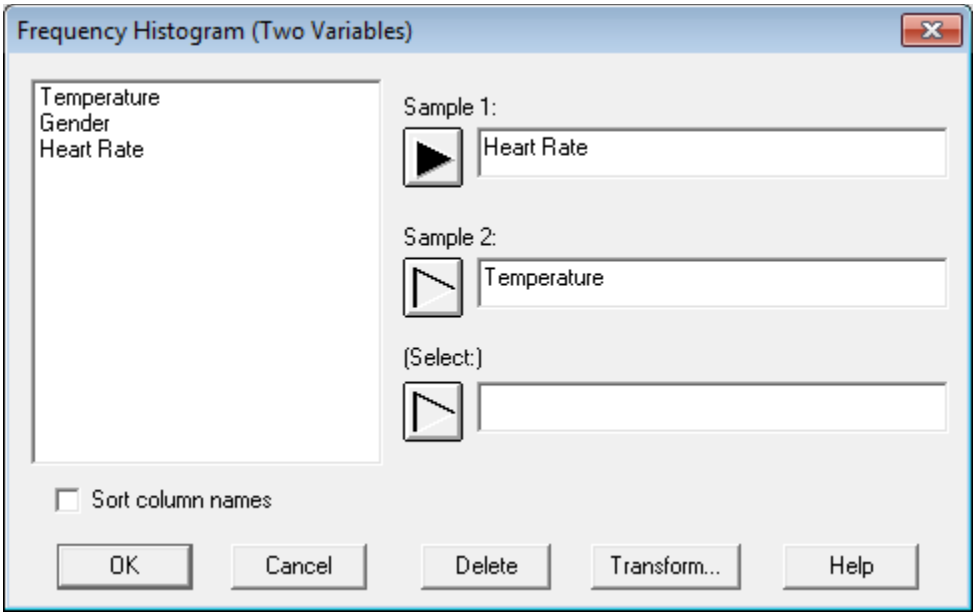
Sample Data

The file *bodytemp.sgd* contains measurements of the body temperature and heart rate of 130 individuals. The first several rows of that file are shown below:

Temperature	*Gender*	*Heart rate*
98.4	Male	84
98.4	Male	82
98.2	Female	65
97.8	Female	71
98	Male	78
97.9	Male	72
99	Female	79
98.5	Male	68
98.8	Female	64
98	Male	67
…	…	…

Data Input

The data to be analyzed are specified on the following data input dialog box:

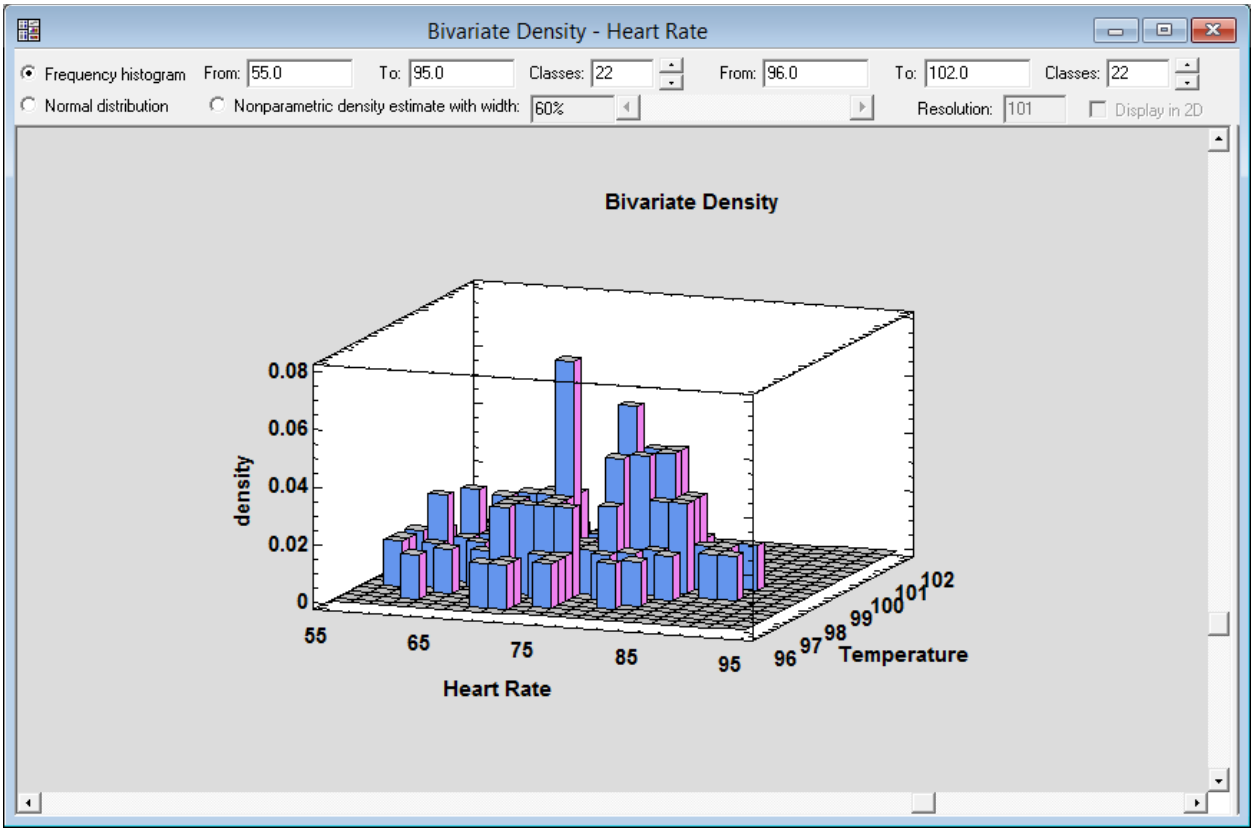


Each row of the datasheet is assumed to contain a single observation. Enter the names of the columns that contain:

- **Sample 1:** numeric column containing the *n* observations for the first variable.
- **Sample 2:** numeric column containing the *n* observations for the second variable.
- **Select:** subset selection.

Statlet

The output of this procedure is displayed in a dynamic Statlet window:



The following options are available:

Frequency histogram

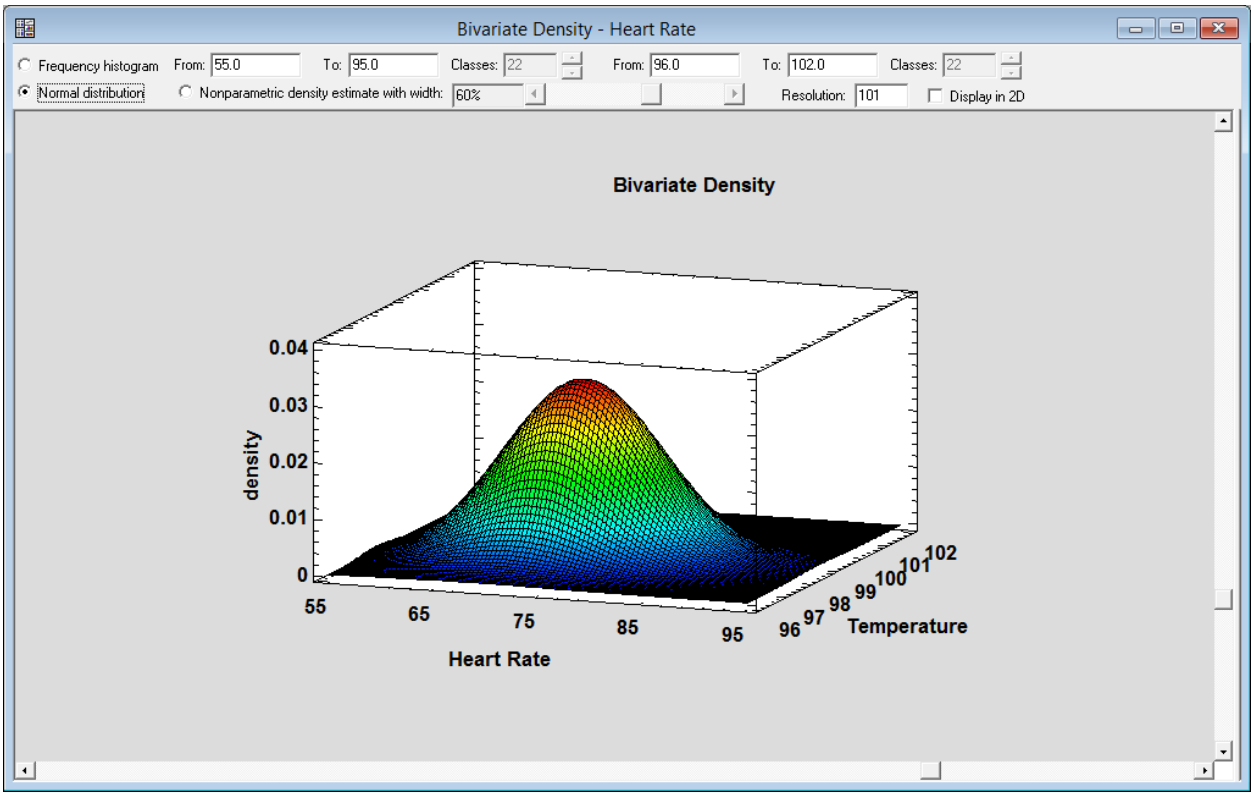
This option divides the data for each variable into *k* adjacent non-overlapping intervals and plots bars of height proportional to the number of observations that fall within all pairs of intervals. To be considered within an interval, an observation must be greater than the lower limit of the interval and less than or equal to the upper limit.

- *From:* Specifies the lower limit of the leftmost interval.
- *To:* Specifies the upper limit of the rightmost interval.
- *Classes*: Specifies the number of intervals or classes *k*.

Note that the vertical axis is scaled so that the volumes of the bars sum to 1, which corresponds to bivariate density functions.

Normal distribution

This adds a bivariate normal density function with means, standard deviations and covariance equal to that of the data.



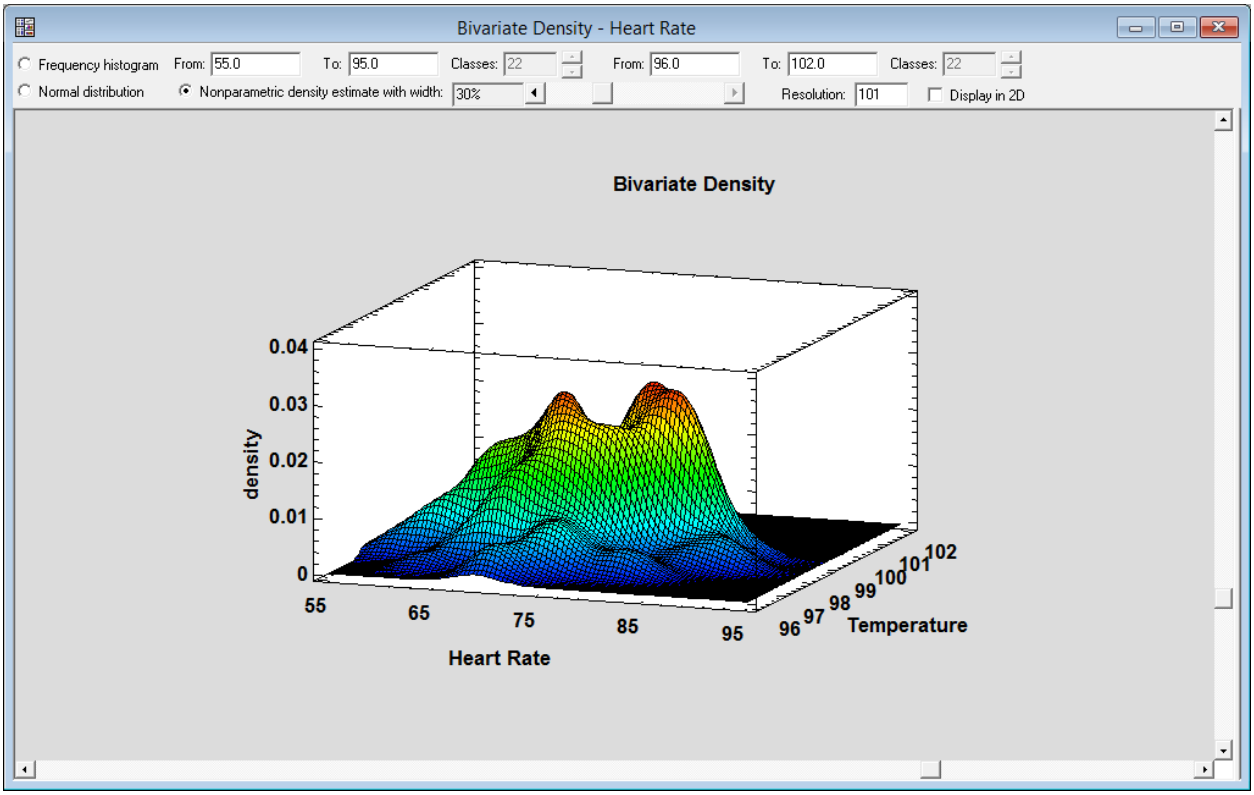
The bivariate normal density function is defined by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho}{\sigma_1\sigma_2}(x_1-\mu_1)(x_2-\mu_2) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} \quad (1)$$

where μ_1 is the mean of the first variable, μ_2 is the mean of the second variable, σ_1 is the standard deviation of the first variable, σ_2 is the standard deviation of the second variable, and ρ is the correlation coefficient.

Nonparametric density estimate

This option adds a nonparametric density estimate to the plot. It is created by counting the number of observations which fall within a window of fixed size moved across the range of the data.



The estimated density function is given by:

$$f(x) = \frac{(\det S)^{-1/2}}{h^2 n} \sum_{i=1}^{n} W\left(\frac{1}{h^2} \left(X_{1,i} - X_1\right)^T S^{-1} \left(X_{2,i} - X_2\right)^T \right) \quad (2)$$

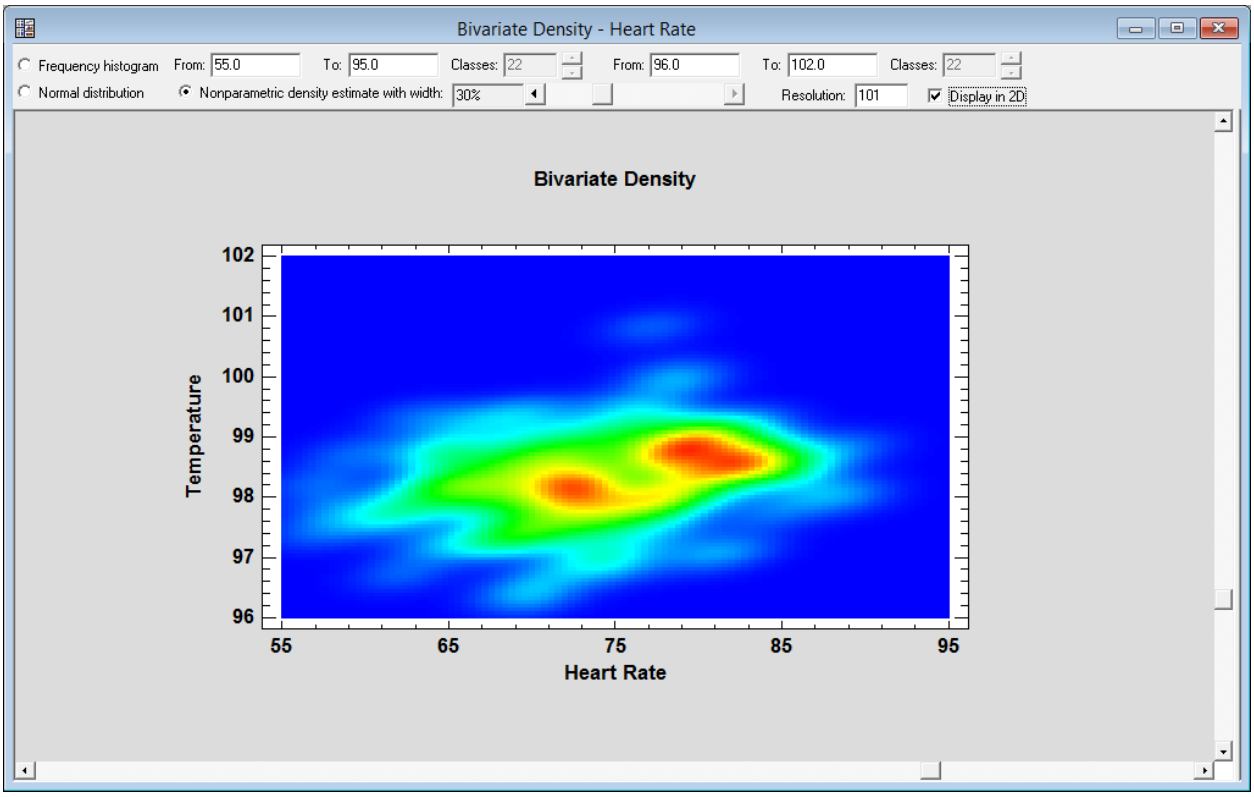
where S is the sample covariance matrix of the 2 variables, h is the width of the window and $W(u)$ is the weighting function defined by

$$W(u) = \frac{1}{2\pi} \exp\left(-u/2\right) \quad (3)$$

A width of 50% is not unreasonable for a small sample but may not give as much detail as a smaller value in larger samples.

Display in 2D

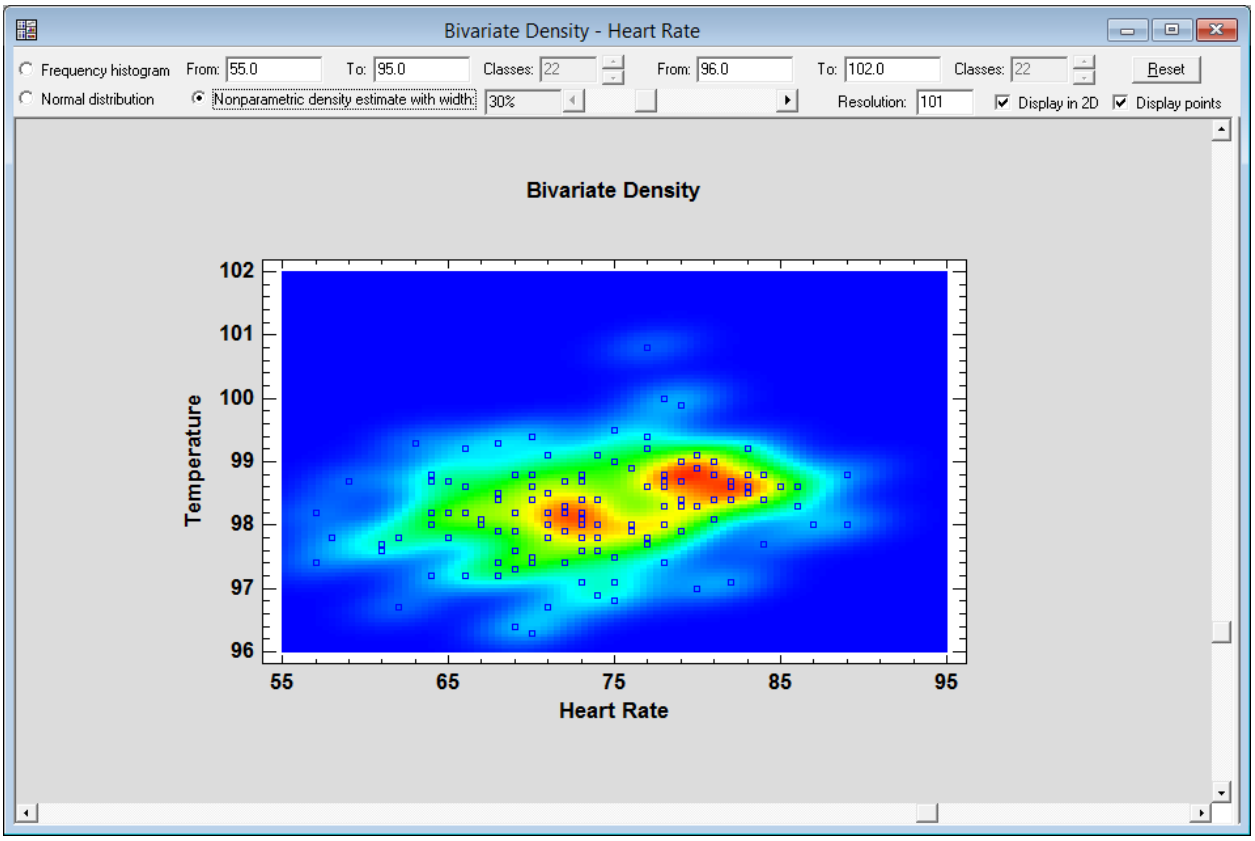
To display the normal distribution or nonparametric density estimate as a two-dimensional contour plot rather than a three-dimensional surface, check the *Display in 2D* box.



Note: in all of the displays, colors are selected to range from dark blue at 0 to dark red at the maximum value of the estimated normal density function.

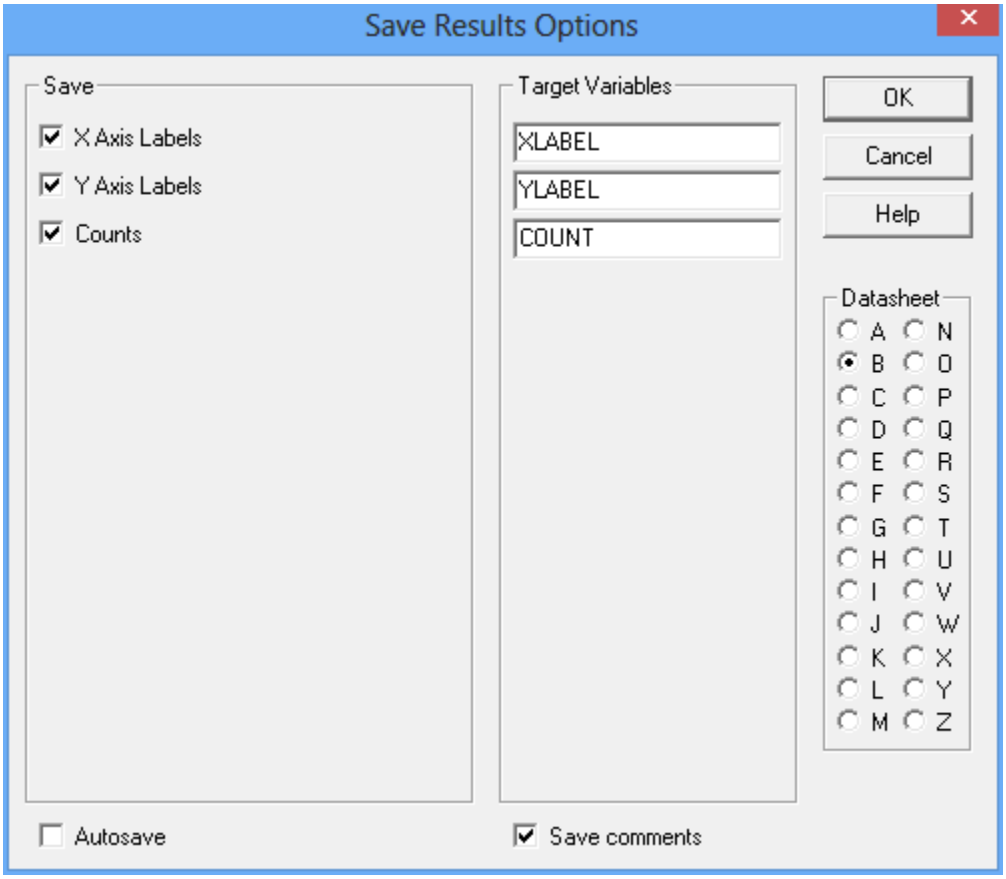
Display Points

Check the *Display points* box to add the data points to the graph:



Save Results

The frequencies displayed by the bivariate histogram can be saved to the DataBook by pressing the *Save Results* button on the analysis toolbar:



- **X Axis Labels** – the intervals corresponding to each division along the x-axis.
- **Y Axis Labels** – the intervals corresponding to each division along the y-axis.
- **Counts** – the frequency or number of observations corresponding to each bar in the histogram.

This output is in the format required to be used as input in the *Crosstabulation* procedure.