

Trivariate Density Statlet



Revised: 10/9/2017



Summary	1
Data Input.....	3
Statlet	4

Summary

The **Trivariate Density Statlet** displays the estimated density function for 3 columns of numeric data. It does so using either a 3-dimensional contour plot or a 3-dimensional mesh plot. The joint distribution of the 3 variables may either be assumed to be multivariate normal or be estimated using a nonparametric approach.

Sample StatFolio: *trivariate density.sgp*

Sample Data

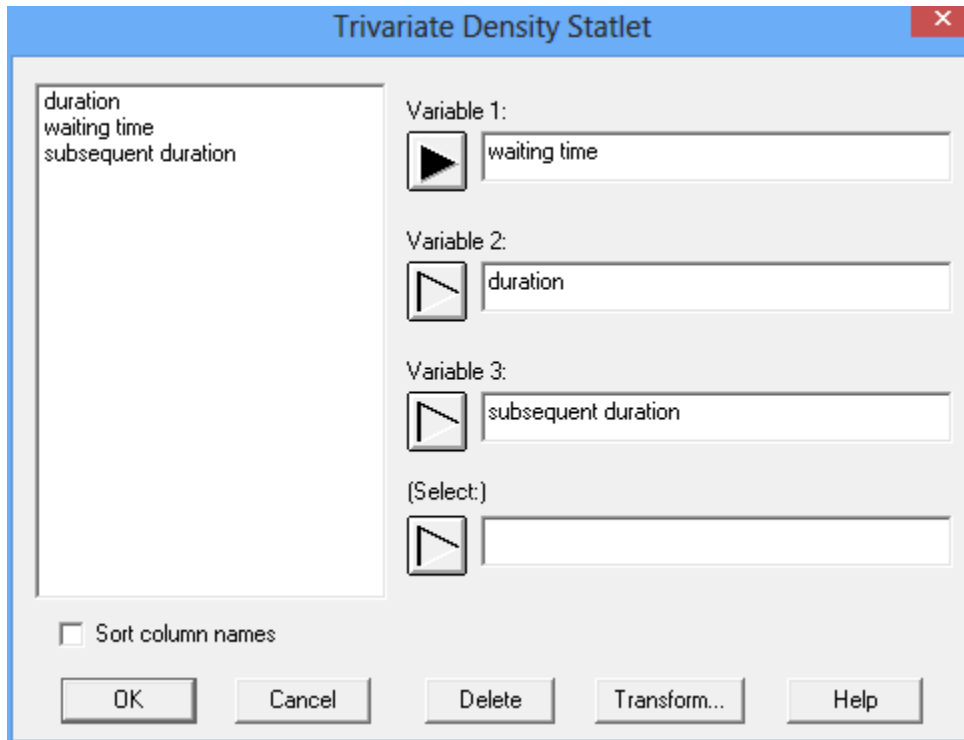
The file *old faithful.sgd* contains data describing 272 consecutive eruptions of the Old Faithful geyser in Yellowstone National Park. The first several rows of that file are shown below:

<i>duration</i>	<i>waiting time</i>	<i>subsequent duration</i>
minutes	minutes between eruptions	minutes
3.600	79	1.800
1.800	54	3.333
3.333	74	2.283
2.283	62	4.533
4.533	85	2.883
2.883	55	4.700
4.700	88	3.600
3.600	85	1.950
1.950	51	4.350
...

The first column contains the duration of the eruption. The second column contains the elapsed time until the next eruption occurred. The third column contains the duration of the subsequent eruption.

Data Input

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

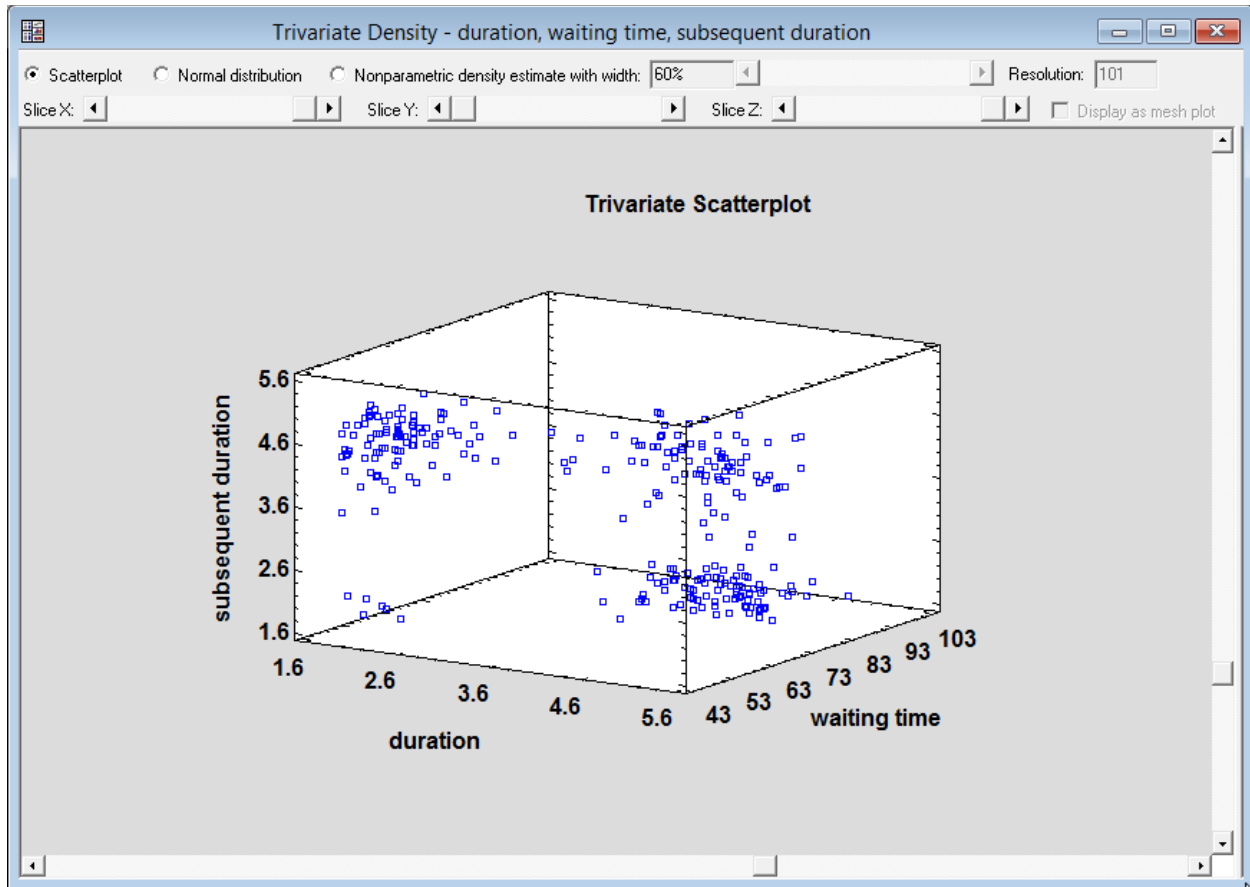


Each row of the datasheet is assumed to contain a single observation. Enter the names of the columns that contain:

- **Variable 1:** numeric column containing the n observations for the first variable.
- **Variable 2:** numeric column containing the n observations for the second variable.
- **Variable 3:** numeric column containing the n observations for the third variable.
- **Select:** subset selection.

Statlet

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



The following options are available:

Scatterplot

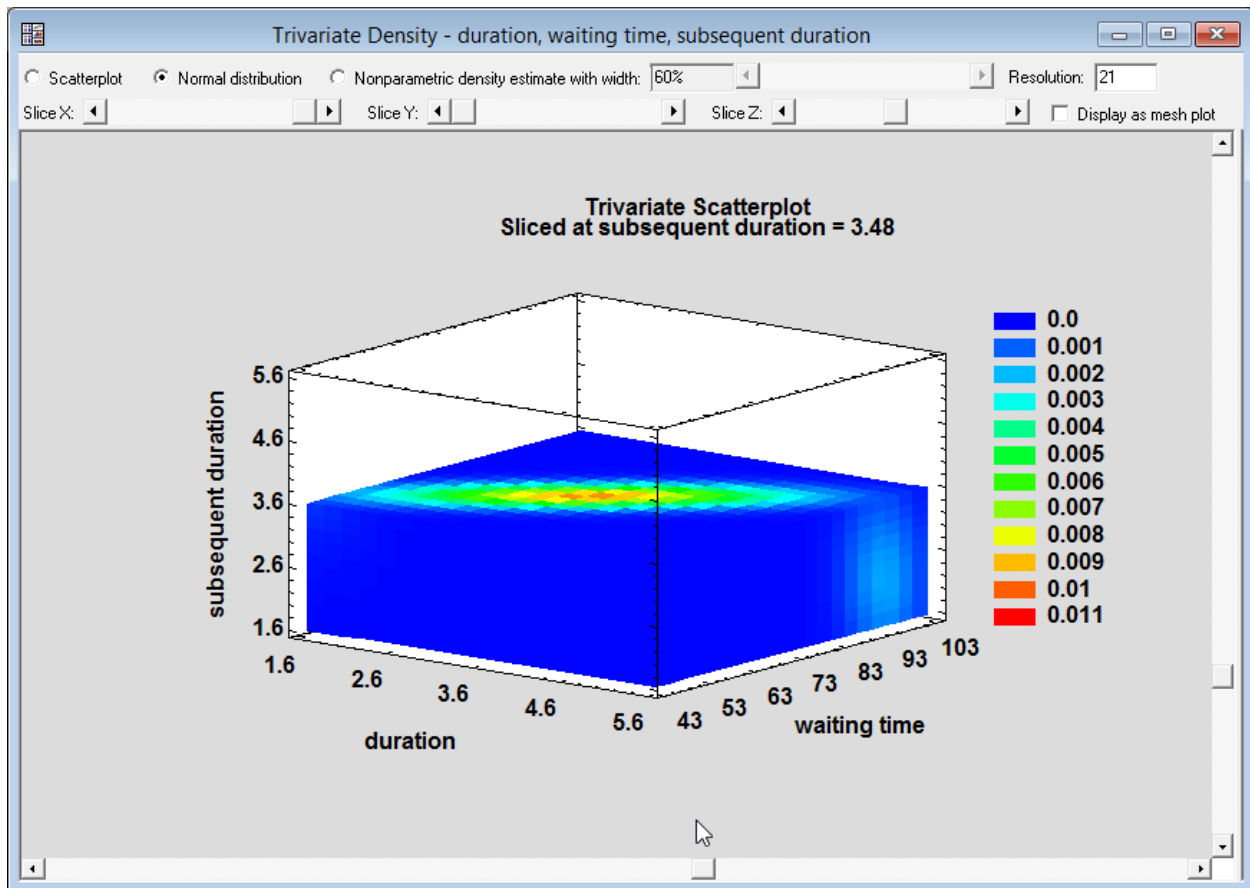
Each observation is displayed as a point in the space defined by the 3 variables. The sample data show a number of regions in which the points appear to cluster.

Normal distribution

This option plots a trivariate normal density function with means, standard deviations and covariances equal to that of the data. It may be plotted in two formats:

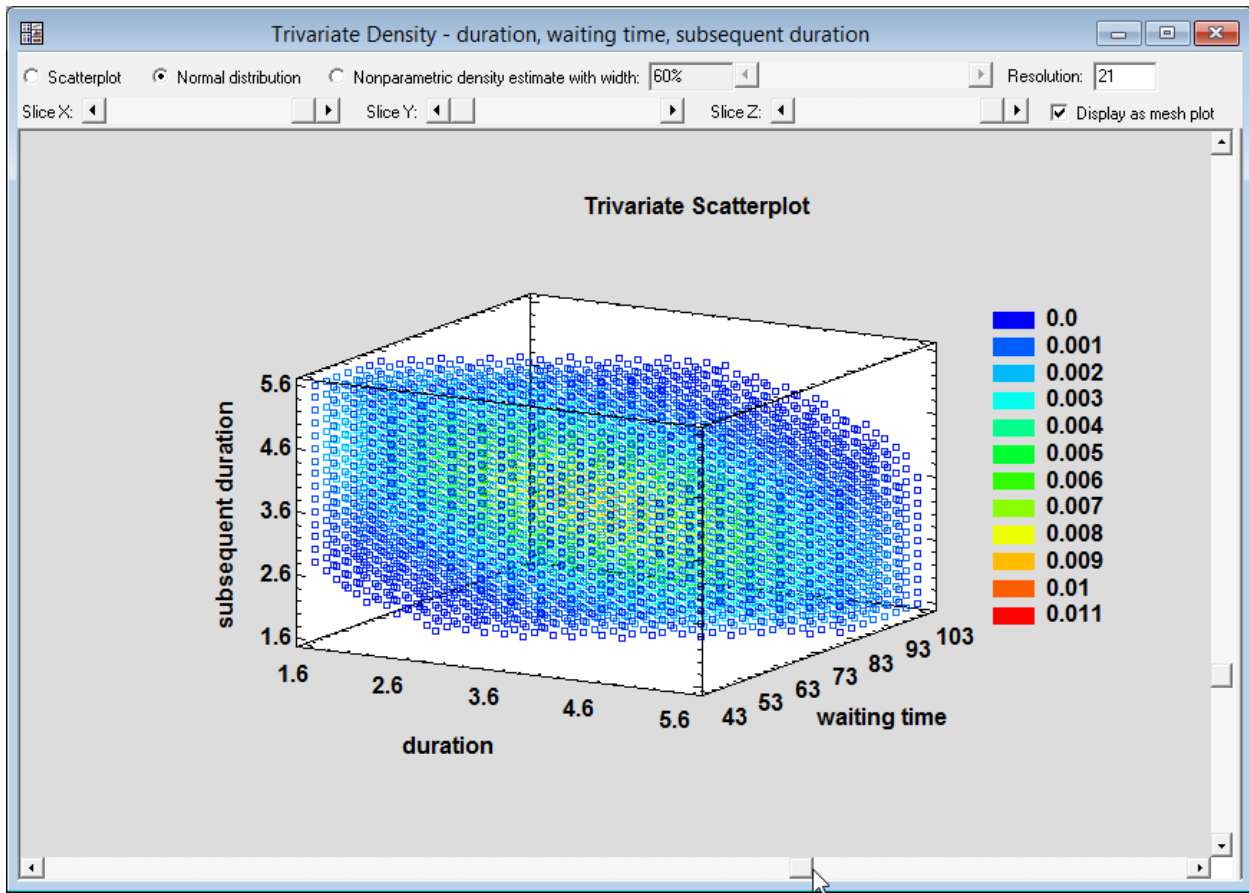
1. As a 3-dimensional contour plot.
2. As a 3-dimensional mesh plot.

When plotted as a contour plot, contours are shown on the front, side and top faces of a cube as shown below:



The 3 sliders on the toolbar control the values of X, Y and Z at which the contour levels are drawn. In the above plot, the centroid of the data is located in the region shown as red.

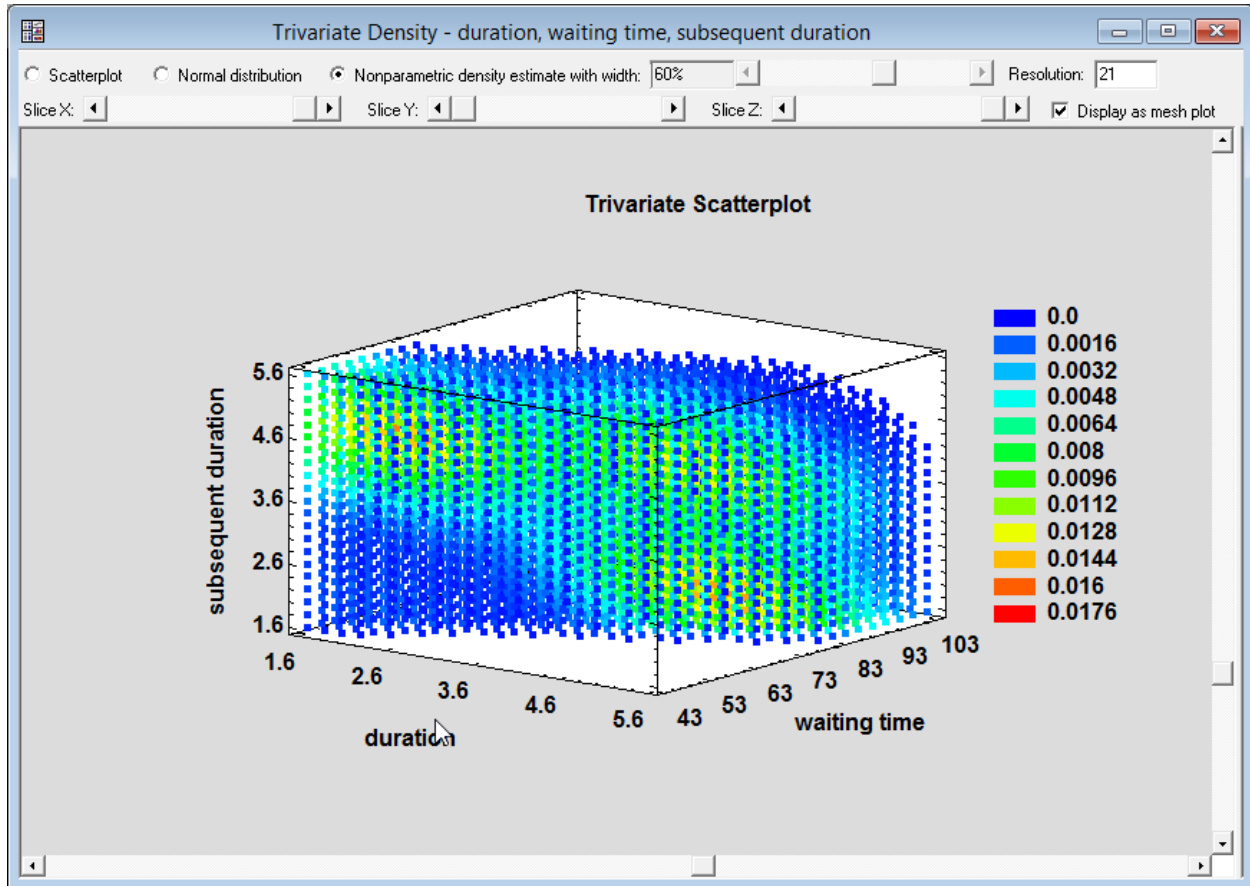
When plotted as a mesh plot, a grid of points is displayed throughout the region:



The *Resolution* setting controls the number of points plotted along each axis.

Nonparametric density estimate

This option plots the density function without assuming any particular distributional form. It is created by counting the number of observations which fall within a window of fixed size moved throughout the range of the data.



The *width* parameter controls how much smoothing is performed, with larger widths giving smoother results.

In the above plot, you can detect 3 areas of red and yellow which are locations in which the points tend to cluster. One area of high density corresponds to eruptions of short duration, which tend to be followed quickly by eruptions of longer duration. After longer duration eruptions, the waiting time tends to be long, with the duration of the subsequent eruption being sometimes short and sometimes long.