Chapter 164

Lag Plots

Introduction

A *lag plot* is used to help evaluate whether the values in a dataset or time series are random. If the data are random, the lag plot will exhibit no identifiable pattern. If the data are not random, the lag plot will demonstrate a clearly identifiable pattern. The type of pattern can aid the user in identifying the non-random structure in the data. Lag plots can also help to identify outliers.

This procedure will generate all lag plots up to a user-specified maximum lag k in a single run.

Definition of Lag

For data values Y_1, Y_2, …, Y_N, the k-period (or k^{th}) lag of the value Y_i is defined as the data point that occurred k time points before time i. That is

$$\text{Lag}_k(Y_i) = Y_{i-k}$$

For example, $\text{Lag}_1(Y_2) = Y_1$ and $\text{Lag}_3(Y_{10}) = Y_7$.

Lag Plot Patterns

The pattern in the lag plot can help you identify possibly hidden patterns in the data. The following sections give some examples of random and non-random data structures and their resulting lag plots.

Random Data

Random data gives rise to lag plots with no pattern. The points in the lag plot appear scattered from left to right and top to bottom.
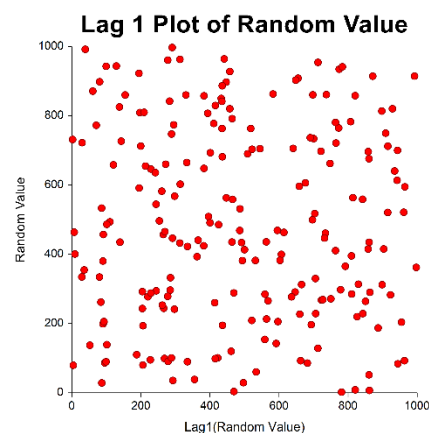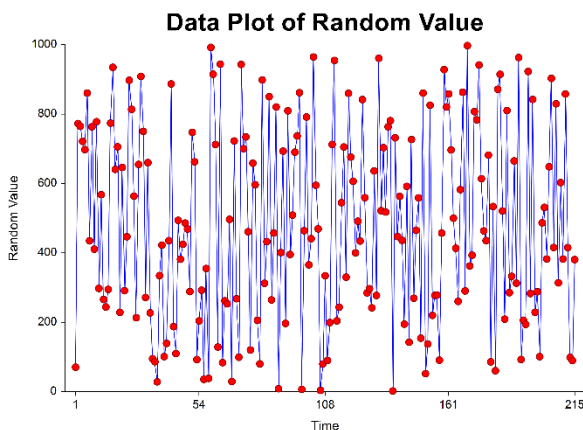
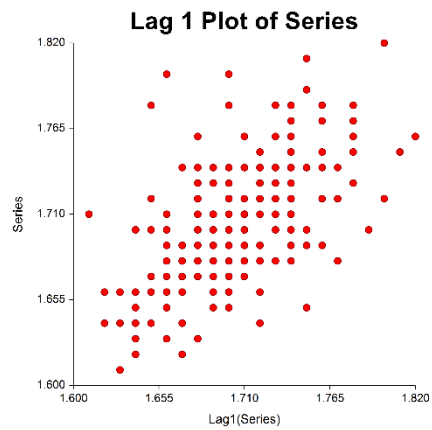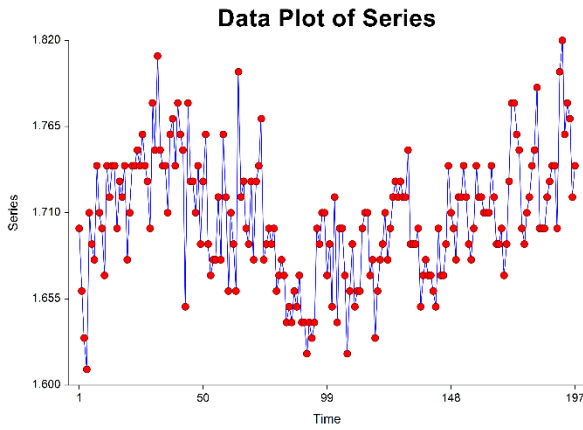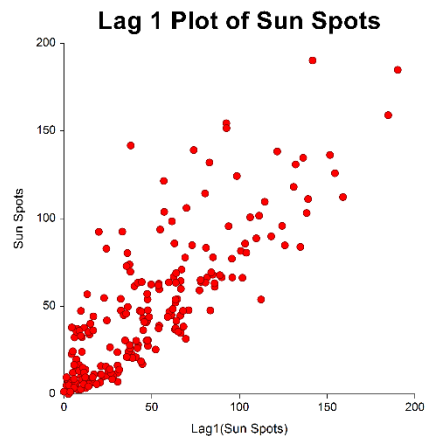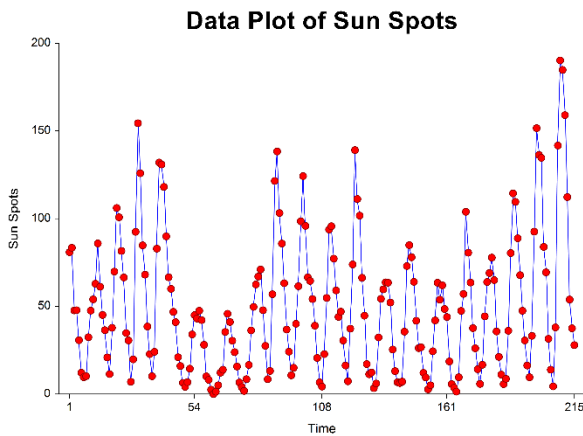## Data with Autocorrelation

Data with autocorrelation gives rise to lag plots with linear patterns that follow the diagonal. As the level of autocorrelation increases, the points cluster more tightly along the diagonal.
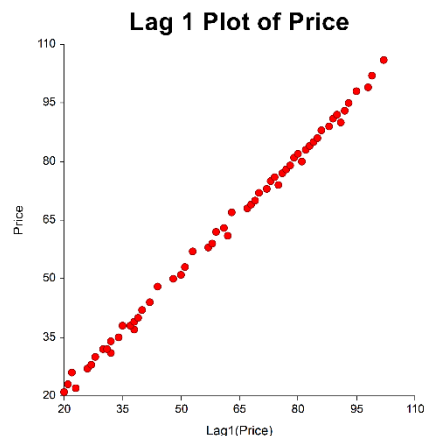
Data with Weak Autocorrelation
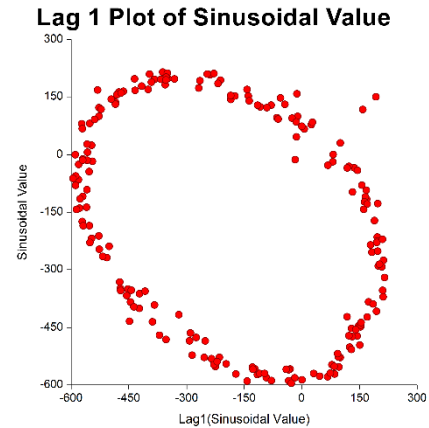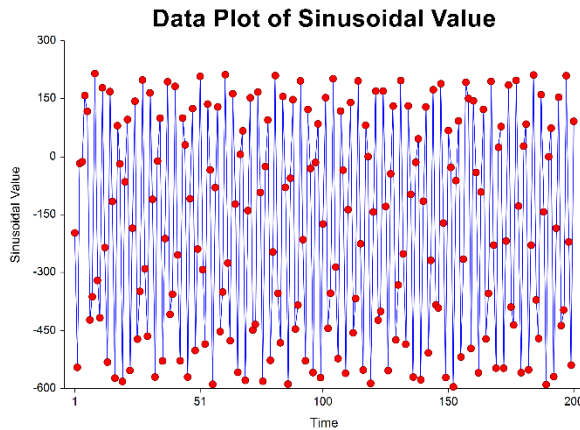

Data with Moderate Autocorrelation

Data with High Autocorrelation

Sinusoidal Data

Single-cycle sinusoidal data gives rise to lag plots with circular or elliptical patterns. Values lying off the ellipse should be considered as potential outliers.

Data Structure

Each lag plot is constructed from a single numeric variable. If multiple variables are entered, a separate lag series is created for each variable. A data label variable can also be entered.

Procedure Options

These options are used to define the lag plots.

Variables Tab

This panel specifies which variables are in the lag plot and how the plot is drawn.

Variables

Variable(s)

Specify one or more columns of data for the lag plot. If more than one column is specified, a separate set of lag plots will be generated for each column.

Data Label Variable

The values in this column contain text (or numbers) that are displayed beside each plotted point. The data labels will not be shown unless you activate them by clicking the plot format button and checking "Labels." The size, font, and location options are also set by clicking the plot format button.

Lags

Maximum Lag Plotted

Choose the maximum lag to plot. A separate series is created for each lag that is less than or equal to the value specified here.

Plot Layout

Specify how the lag series will be displayed on the plots. The options are

- **Separate Plot for Each Lag**

  A separate plot is created for each lag series.

- **All Lags Together on a Single Plot**

  A single plot is created with all lag series. Each lag is represented by a different group on the plot.

Format Options

Variable Names

This option selects whether to display only variable's name, label, or both.

Lag Plot Format

Plot Format Button

Click this button to specify the format of the plot.

Example 1 – Creating Lag Plots

This section presents a tutorial on generating 1st and 2nd order lag plots. To run this example, take the following steps using the Sunspot dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Lag Plots window.

1 Open the Sunspot dataset.

- From the **File** menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Sunspot.NCSS**.
- Click **Open**.

2 Open the Lag Plots window.

- Using the Graphics menu or the Procedure Navigator, find and select the **Lag Plots** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the Lag Plots window, select the **Variables tab**.
- Double-click in the **Variable(s)** text box. This will bring up the column selection window.
- Select **Spots** from the list of variables and then click **Ok**. "Spots" will appear in this box.
- For **Maximum Lag Plotted**, enter **2**.
- Leave all other settings at their default values.

4 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Output

This section displays the lag plots. In the output screen you can double-click any of the plots to see it in full size.
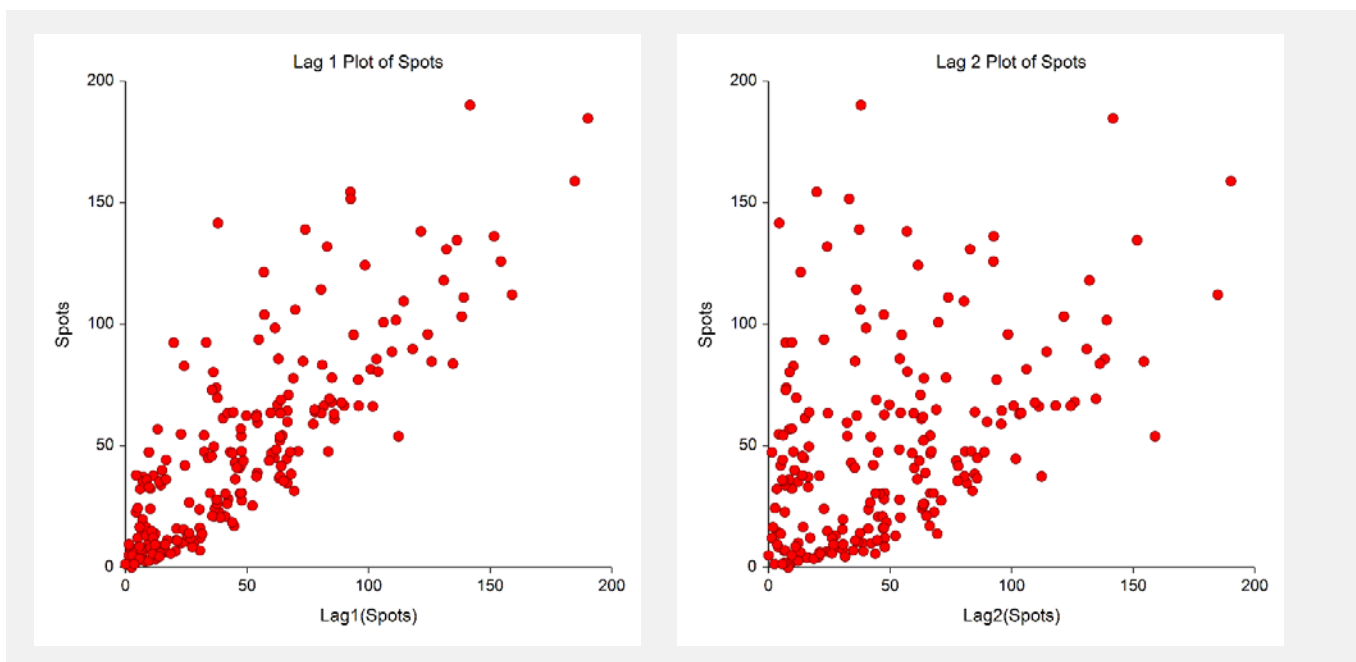
The data variable "Spots" is plotted against the 1st and 2nd order lags in these two plots. The linear pattern in the 1st order lag plot indicates probable autocorrelation.

Example 2 – Combining Multiple Series on a Single Lag Plot

This section presents a tutorial on generating 1st and 2nd order lag series and combining the results onto a single plot. To run this example, take the following steps using the Sinusoidal dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 2** by clicking on Open Example Template from the File menu of the Lag Plots window.

1   Open the Sinusoidal dataset.
- From the **File** menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Sinusoidal.NCSS**.
- Click **Open**.

2   Open the Lag Plots window.
- Using the Graphics menu or the Procedure Navigator, find and select the **Lag Plots** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- On the Lag Plots window, select the **Variables tab**.
- Double-click in the **Variable(s)** text box. This will bring up the column selection window.
- Select **Series** from the list of variables and then click **Ok**. "Series" will appear in this box.
- For **Maximum Lag Plotted**, enter **2**.
- For **Plot Layout**, enter **All Lags Together on a Single Plot**.
- Leave all other settings at their default values.

4   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Output

This section displays the lag plots. In the output screen you can double-click any of the plots to see it in full size.
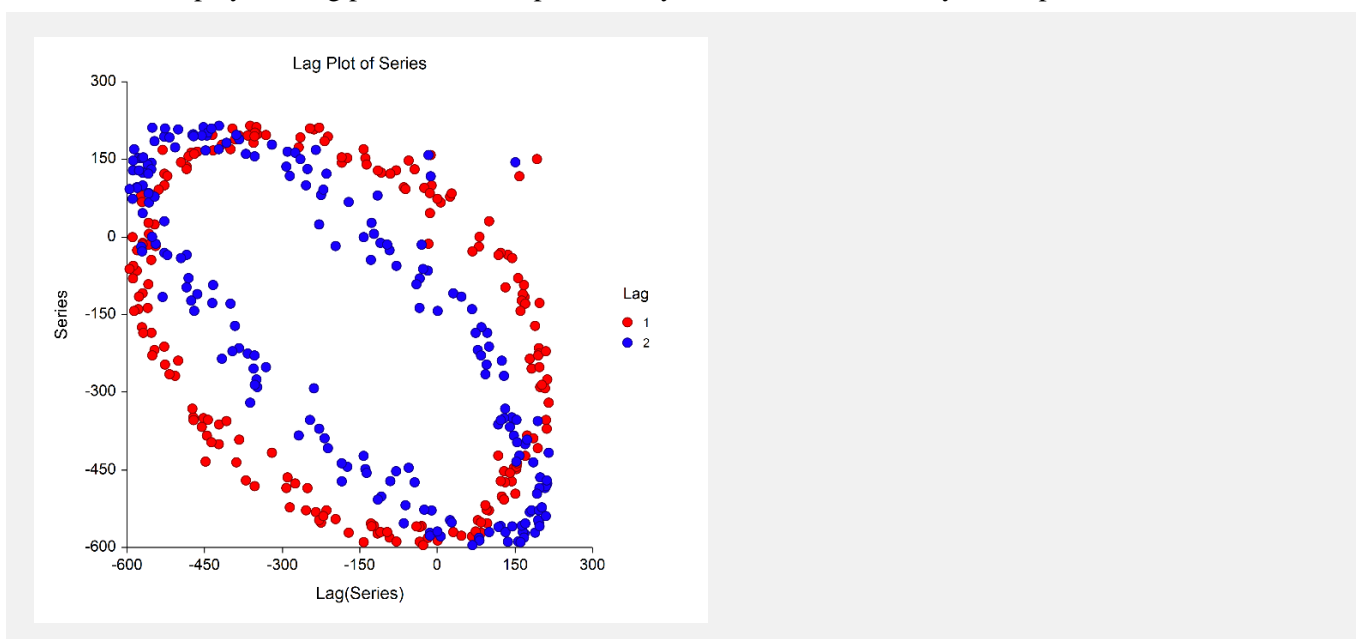
The data variable "Series" is plotted against the 1st and 2nd order lags with the results combined in a single plot. The sinusoidal data exhibits the typical elliptical pattern in the lag plot.