

Chapter 195

Back-to-Back Stem-and-Leaf Plots

Introduction

This procedure generates a stem-and-leaf plot of a batch of data. The stem-and-leaf plot is similar to a histogram and its main purpose is to show the data distribution while retaining the uniqueness of each data value. It is recommended for batches of data containing between 15 and 150 data points.

If the underlying data values have more digits than can be displayed on the chart, the values are truncated, not rounded. For example, 29.87 would be truncated to 29 if two digits were needed or 29.8 if three digits were needed.

Stem-and-Leaf Plots of SepalLength and PetalLength

SepalLength	Stem	PetalLength
977776	7	
4322210	7	
9999888777777776655555	6	6779
7 + 33333333222211111000000	6	0011134
6 + 88887777777666666555555	5	5556666677788899
5 + 4322211111111000000000	5	00001111111223344
99999888887766665	4	5555556667777888899999
4443	4	00000111222334444
	3	55678999
	3	033
	2	
	2	
	1	99

Data Structure

The data are contained in two columns or a single column with another grouping column.

Fruit Yield

Apple	Pear
75	69
63	78
64	59
77	68
55	64
	72

Back-to-Back Stem-and-Leaf Plot Options

Range Multiplier

The data values are split into two groups: regular values which are shown in the stem-and-leaf plot and outliers which are shown at the beginning and end of the plot. This value determines how many of the data values are classified as regular values. This is accomplished as follows.

1. Find the upper and lower hinges (L and U) which are roughly the 25th and 75th percentiles.
2. Calculate $IQR = U - L$.
3. Calculate the upper and lower adjacent values (U_a and L_a) which are the data values just less than $U + R \times IQR$ and just greater than $L - R \times IQR$.
4. The stem-and-leaf plot is constructed on those values between L_a and U_a . Any other values are treated separately as outliers.

Common values are 1.5 to 3. If you want to include most of the data, set this value to 9 or 10.

Branches per Stem

The leaves of the plot are formed from the integers from 0 to 9. When a line becomes too long to display (e.g., if there are more than 35 leaves), the stem may be split into branches. The first branch contains leaves made up of the digits 0 through 4 and the second branch contains leaves made up of the digits 5-9. This option controls the number of branches that are displayed for each stem.

For example, the line (branches = 1)

```
2 | 000112222444444555677778899
```

could be split into two branches

```
2 | 000112222444444
```

```
2 | 555677778899
```

or five branches

```
2 | 00011
```

```
2 | 2222
```

```
2 | 44444555
```

```
2 | 67777
```

```
2 | 8899
```

Max Leaves per Line

Only about 35 leaves can be displayed on a single line. If more than this number occur, the line is truncated to this number of leaves and the number truncated is displayed.

For example, the line

```
2 | 0001122224444445556 + 5
```

means that the last five leaves were not displayed.

Example 1 – Create a Back-to-Back Stem-and-Leaf Plot

This section presents an example of how to create a back-to-back stem-and-leaf plot on the *SepalLength* variable in the *Fisher* dataset.

Setup

To run this example, complete the following steps:

1 Open the Fisher example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Fisher** and click **OK**.

2 Specify the Back-to-Back Stem-and-Leaf Plots procedure options

- Find and open the **Back-to-Back Stem-and-Leaf Plots** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Input Type.....**Two Variables with Response Data in each Variable**

Group 1 Variable.....**SepalLength**

Group 2 Variable.....**PetalLength**

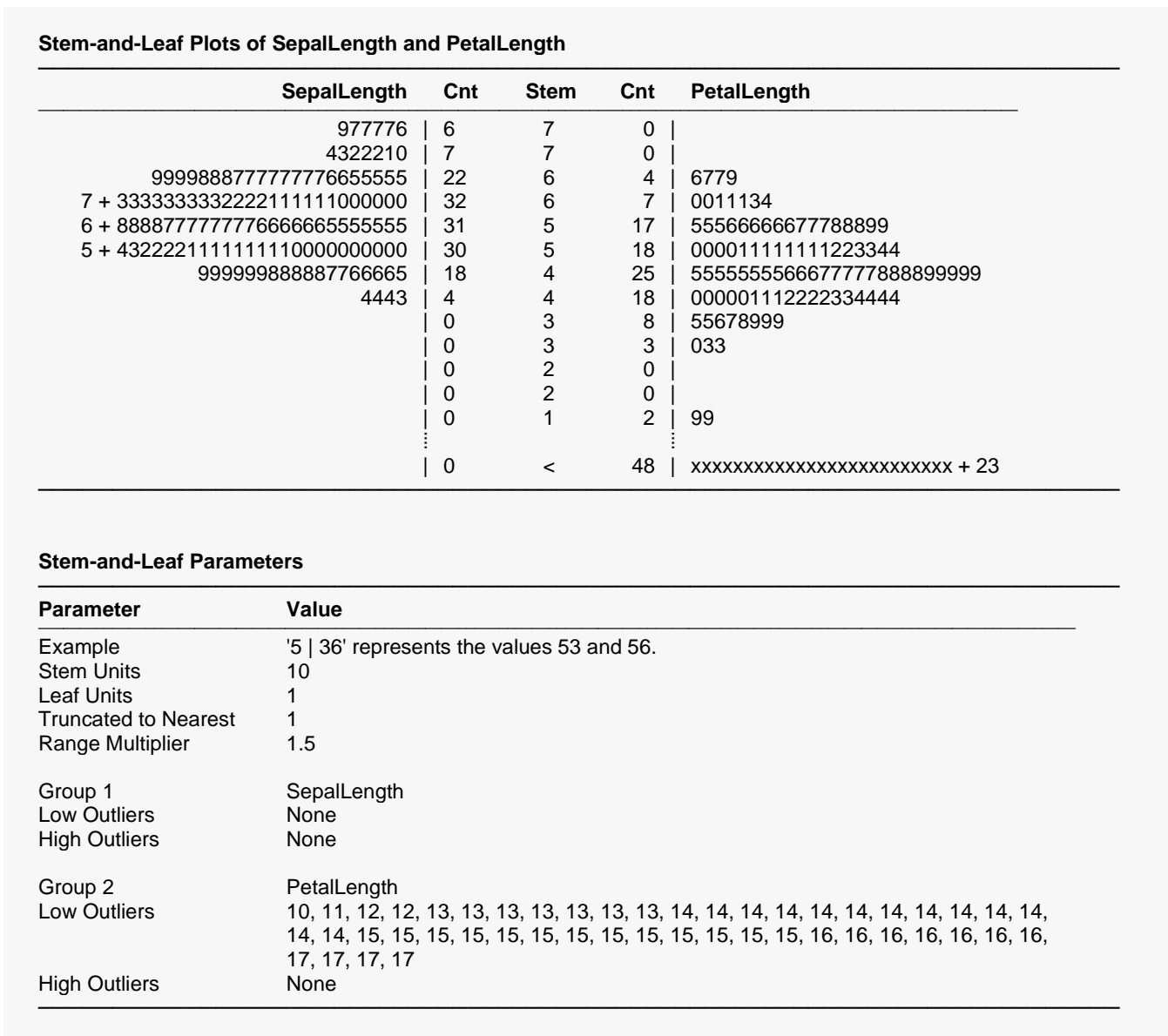
Report Options (*in the Toolbar*)

Variable Labels.....**Column Names**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Back-to-Back Stem-and-Leaf Plot



The stem-and-leaf plot is a type of histogram which retains much of the identity of the original data. It is useful for finding data-entry errors as well as for studying the distribution of a variable.

Counts

These are the number of leaves in each batch (variable or group) of data.

Stem

The stem is the first digit of the actual number. For example, the stem of the number 523 is 5 and the stem of 0.0325 is 3. This is modified appropriately if the batch contains numbers of different orders of magnitude. The largest order of magnitude is used in determining the stem. Depending upon the number of leaves, a stem may be divided into two or more branches. A special set of branch labels is then used to mark the branches.

Back-to-Back Stem-and-Leaf Plots

Leaf (Variable, or Group, Name)

The leaf is the second digit of the actual number. For example, the leaf of the number 523 is 2 and the leaf of 0.0324 is 2. This is modified appropriately if the batch contains numbers of different orders of magnitude. The largest order of magnitude is used in determining the leaf.

Stem-and-Leaf Parameters

This report shows various parameters used in constructing the plot.