Chapter 196

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 Plot of Sepal Length
―――

Count	Stem		Leaf
6	7 (5-9)	|	677779
7	7 (0-4)	|	0122234
22	6 (5-9)	|	5555566777777778889999
32	6 (0-4)	|	00000011111122223333333334444444
31	5 (5-9)	|	5555555666666777777778888888999
30	5 (0-4)	|	000000000011111111122223444444
18	4 (5-9)	|	566667788888999999
4	4 (0-4)	|	3444

―――

Data Structure

The data are contained in a single column.

Height Dataset (Subset)

Height
64
63
67
.
.
.

Stem-and-Leaf Plot Options

Range Multiplier

The data values are split into two groups: regular values which 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 (Ua and La) which are the data values just less than U + R x IQR and just greater than L - R x IQR.

4. The stem-and-leaf plot is constructed on those values between La and Ua. Any other values are treated separately as outliers.

Common values are 1.5 to 3. If you want to include most all 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 70 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 | 0001122224444445555677778899

could be split into two branches

    2 | 000112222444444
    2 | 555677778899

or five branches

    2 | 00011
    2 | 2222
    2 | 444444555
    2 | 67777
    2 | 8899

Max Leaves per Line

Only about 70 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 truncated is displayed.

For example, the line

    2 | 0001122224444445556 + 5

means that the last five leaves were not displayed.

Example 1 – Create a Stem-and-Leaf Plot

This section presents an example of how to create a 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 Stem-and-Leaf Plots procedure options
- Find and open the **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
>
> Data Variable(s) **1**
>
> Report Options (*in the Toolbar*)
>
> Variable Labels .... **Column Labels**

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

Stem-and-Leaf Plot

Stem-and-Leaf Plot of Sepal Length
───

Count	Stem	Leaf
1	7 |	9
5	7 |	67777
1	7 |	4
4	7 |	2223
2	7 |	01
7	6 |	8889999
10	6 |	6677777777
12	6 |	444444455555
13	6 |	2222333333333
12	6 |	000000111111
10	5 |	8888888999
14	5 |	66666677777777
13	5 |	4444445555555
5	5 |	22223
19	5 |	0000000000111111111
11	4 |	88888999999
6	4 |	666677
4	4 |	4445
1	4 |	3

Stem-and-Leaf Parameters
───

Parameter	Value
Example	'5 | 36' represents the values 53 and 56.
Stem Units	10
Leaf Units	1
Truncated to Nearest	1
Range Multiplier	1.5
Low Outliers	None
High Outliers	None

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.

Count

This is the number of leaves on the line.

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.

Leaf

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 at the bottom shows various parameters used in constructing the plot.