## 1. Introduction

Welcome to the series of E-learning modules on stem and leaf plot. In this module, we are going cover the definition, utility, construction and analysis of stem and leaf plot as a graphical tool of representation of data.

By the end of this session, you will be able to:

- Explain stem and leaf plot
- Explain the uses of stem and leaf plot
- Explain the construction of stem and leaf plot from a given data
- Explain the analysis and interpretation of the data

A collection of information for a specific purpose is called a data set, which forms the base for any statistical analysis.

The data thus collected for analysis purpose can be shown in the form of graphs, charts and tables, which helps in making the data analysis a creative and easy process.

There is a variety of age-appropriate techniques of data analysis. The seven types of data display that are most popularly used in data analysis are:

Line plots, Histograms, Stem and leaf plots, Two-way tables, Box plots, Scatter plots, Time-series plots.

It is important to keep the following points in mind while doing a data analysis:

- For any data set, there is more than one way to plot the data.
- It is valuable to plot the data in several ways as it gives a lot of insight on various aspects of the data.
- Creative insights are an asset to good graphing and data analysis so that everyone can interpret the data and only then, the graph is useful.

In this module, let us discuss the use and construction of a stem and leaf plot in the analysis of the data.

#### What is a stem and leaf plot?

A Stem and Leaf Plot is a type of graph that is similar to a histogram but shows more information. It helps in summarizing the data to find its distribution and extract detail about the individual values. Stem and Leaf Plots are great organizers for large amount of information.

The Stem-and-Leaf Plot summarizes the shape of a set of data (the distribution) and provides extra detail regarding individual values. The data is arranged by place value.

The digits in the largest place are referred to as the stem and the digits in the smallest place are referred to as the leaf (leaves). Let us take the number 145 as an example. Here the 14 digits are in the largest place, so it is called as the stem and 5 is in the smallest place, which is called as leaf.

When should we use a stem and leaf plot?

A stem and leaf plot is a very common tool to represent and analyze large amount of numbers. Series of scores on sports teams, series of temperatures or rainfall over a period of time, series of classroom test scores are examples of when Stem and Leaf Plots could be used.

Definition:

A stem-and-leaf plot uses the digits of data values to organize a data set. Each data value is broken into a stem (digit or digits on the left) and a leaf (digit or digits on the right). A stemand-leaf plot shows how data are distributed.

What are the parts of a stem and leaf plot?

Title: The title briefly describes the information that is contained in the stem and leaf plot graph.

Stem: The digits in the largest place are referred to as the stem. Each data value is broken into a stem they are digit or digits on the left. The "stem" is the left-hand column which contains the tens digits.

Leaf: The digits in the smallest place are referred to as the leaf (leaves). Each data value is broken into a leaf they are digits of the value on the right. The leaves are displayed to the right of the stem.

Key: The key explains what the stems and leaves represent. It shows how to interpret the digits. Every stem and leaf plot must have a key. Sometimes the numbers may repeat itself.

Legend: The legend provides additional information for that documents where the data came from and how the measurements were gathered.

#### Construction of Stem and Leaf Plot

How do we construct a Stem and leaf plot from a given data?

In constructing the stem-and-leaf plot, we pool counts within a class of numbers, and we retain the actual numbers to make the graph, rather than using dots or Xs, so that we can continue to use the graph for determining other summaries of the data, such as the median.

The following 6 steps are generally practiced while constructing the graph.

Step 1: Ordering the data: In this step the collected data is arranged according to a numerical order generally ascending order.

Step 2: Choosing the stem & leaves: In this step the data so collected is separated into a stem and a leaf

Step 3: Write the stems to the left of the vertical line: In this step, the stems are written in ascending order with a vertical line to its right.

Step 4: Write the leaves for each stem to the right of the vertical line: In this step, the leaf is written on the plot next to the stem, until all the values in the data set are completed.

Step 5: Rearrange the leaves: In this step, we create a new plot; the leaves are rearranged so that they are ordered from smallest to largest.

Step 6: Add title, legend and key: In this step, we try to give a title and legend to the graph so that it helps in identifying the five Ws of the data (Who, Where, What, When and Why) that are important for understanding and interpreting the data.

Point to be noted: To get a good picture of the data, the final plot must be made quite neatly. To encourage this, it is often convenient to use graph paper, and to write each digit in a separate box on the graph paper.

### 3. Types of Stem and Leaf Plot

Now let us discuss the types of stem and leaf plot.

There are two types of stem and leaf plot they are Split stems and Double stem plots.

Split Stems...

Are stem and leaf plots that helps in understanding the spread of the data, which might reveal additional patterns. One way to do this is to split the stems into a smaller number of equally sized units.

For example let us take the counts of sit ups done by the students of a class and group it in units of 5 ones (30-34, 35-39, 40-44, etc)

(30-34 can be represented as 3 in the stem and other digits up to 34 in the same stem in leaf) (Now, when we have to represent data points between 35-39 we can see that the stem is 3 which is as same as the stem above. Hence, we use a black dot which represents the same value of stem as above)

(Similarly this is we have represented the data sets between 40-44)

(Between 45-49) (Between 50-54) (Between 55-59) (Between 60-64) And the key is 34 sit-ups 35 sit-ups

Double stem-and-leaf plots:

A double stem and leaf plot is called as a back-to-back stem and leaf plot. This is sometimes used to compare two sets of data or rounded and truncated values of the same data. In a back-to-back plot, the same stem is used for the leaves of both plots.

For example if we have a data set for two classes class 1 and class 2 we will represent the data in a single stem and leaf plot graph as follows:

The data to the left of the stem represents data points of class 2 and data on the right of the stem represents data points of class 1. The common stems are 3, 4, 5, and 6.

|3|4 represents 34 sit ups of class1

0 |5| and 05 represents 50 sit ups of class 2

Note that back-to-back stem-and-leaf plots require writing the left-hand numbers backwards.

# 4. Representation of Stem and Leaf Plot

How to make stem and leaf plot using one-digit number? Sometimes the data will contain numbers from 0 to 9. Let us see how to represent them in a stem and leaf plot. Consider the following data. 11 4 6 13 52 8 5 3

Now we know how to represent the two digit number but when it comes to one digit number, First of all arrange them in ascending order

3 4 5 6 8 11 13 52

The one digit number can be split as 0 AND 3 keeping 0 in the tens place and 3 in the ones place. When done so we get the following stem and leaf plot.

So the stem- plot is formed as

KEY:

0|3 means single digit 3 and 1|1 means double digit 11.

How to make stem and leaf plot using three-digit number? Now let us check what happens when a three-digit number is taken as data. Consider the following data: 4 58 396 24 56 8 472 125 200 50

First of all arrange them in ascending order

4 8 24 50 56 58 125 200 396 472

Now, when we take the three digit number, arrange both hundreds and tens place in the stem and only the units place leaf.

So, 125 can be separated as 12 in the stem place and 5

In the leaf place, because 12 is at the larger place and 5 is at the smaller place. Similarly, we have separated the digits 200, 396 and 472.

So, the stem and leaf plot is formed as

And the key is

0|4 means single digit 4, 2|4 means double digit 24 and 12|5 means three digits 125.

Let us discuss the stem and leaf plot with decimal.

Now we can draw the same graph using decimal numbers.

Here also the stem and leaf plot has no difference. It only looks at the last digit for the leaves and the digits before it for the stem. However, just put a legend to show what type of numbers we are using.

For example consider the data:

6.8 4.9 5.2 3.6 6.2 6.3

As usual arranging in ascending order 3.2 3.6 4.9 5.2 6.2 6.3 6.8 Here, as before the ones digits will be for the stem values, and the tenths will be for the leaves.

Hence the plot will be

The key 3|2= 3.2

How do we interpret a Stem and leaf plot?

The stem and leaf plot can be used in getting various information like symmetry, skewness, outliers, clusters and gaps, of a data.

It is also used in calculating the various measures of central tendency like mean, median or mode of a data

- The mean or average is the sum of the numbers divided by the total number of data points.
- The median is the middle number when the numbers are written in order. If there are even data points than the average of the middle two numbers are taken.
- The mode is the number that occurs most frequently. A set of data can have more than one mode or no mode. A mode can be identified by looking at which value is having the most scores.

Let us discuss the advantages of a stem and leaf plot.

- A stem and leaf plot is the most convenient, instant and simple way of representation for small set of data.
- A stem-and-leaf plot provides a convenient way to display the distribution of these data.
- It retains the actual numbers to make the graph, rather than using dots or Xs.
- The graph can be used for determining other summaries of the data.

Let us discuss the limitations of a stem and leaf plot.

- It is difficult to use the stem and leaf plot for describing large data sets.
- There is little flexibility in the choice of stem.
- Does not convey a rapid reading of class frequency
- It cannot be used to represent categorical data.
- The presentation does not look good when there are stems with no leaf.

### 5. Illustration

Illustration:
Make a stem and leaf plot of the Statistical test scores given below.
56, 65, 98, 82, 64, 71, 78, 77, 86, 95, 91, 59,
69, 70, 80, 92, 76, 82, 85, 91, 92, 99, 73
Then complete each question.
a. What type of graph does a stem and leaf plot represent when turned vertically?
b. What was the lowest score on the statistical test?
c. What was the highest score on the Statistical test?
d. In which interval did most students score?

Solution:
Step 1: Ordering the data:

56, 59, 64, 65, 69, 70, 71, 73, 76, 77, 78, 80, 82, 82, 85, 86, 91, 91, 92, 92, 95, 98, 99

Step 2: Choosing the stem and leaves:

In this example, we have 56 as a value in which 5 is the stem and 6 is the leaf. Since the data range from 56 to 99, the stems range from 5 to 9. To plot the data, make a vertical list of the stem.

Step 3: Write the stems to the left of the vertical line:

Step 4: Write the leaves for each stem to the right of the vertical line: Here, the stem and leaf of each data is written.

Step 5: Rearrange the leaves:

In this step, we rearrange the leaves of the data in ascending order.

Step 6: Add title, legend and key:

The title for this graph is statistical score.

Legend is the score of 23 students for the test conducted during the month of May. Key is 8|0 means 80

Now, let us answer few questions related to the same illustration.

- a. What type of graph does a stem and leaf plot represent when turned vertically? A stem and leaf plot when turned vertically represents a histogram.
- b. What was the lowest score on the statistical test? The lowest score is 56
- c. What was the highest score on the Statistical test? The highest score is 99
- d. In which interval did the students score most? In the interval between 90 -100

We often wish to compare the results from two (or more) situations.

If we have exactly two situations to compare where each could be represented, a back-toback stem and leaf plot provides a convenient way to display the data.

Here, we plot two different situations on opposite sides of the stem describing the dependent variable.

For a better understanding, let us take the scores of the same 23 students in another subject say History to compare their performance.

History data: 56, 59, 65, 69, 69, 70, 73, 73, 76, 78, 80, 80, 82, 82, 85, 86, 88, 91, 91, 92, 92, 95, 98

Statistics data: 56, 59, 64, 65, 69, 70, 71, 73, 76, 77, 78, 80, 82, 82, 85, 86, 91, 91, 92, 92, 95, 98, 99

Construction of a back-to-back stem plot is as shown. In this, we can see that we could easily compare statistics marks with the history marks at the same time. As we can see in this data, the more number of students scored in History are in the range of 80 and in statistics, it is in the range of 90.

Key: 6| 5| is read as 56 backwards and | 6|4 is read as 64.

While comparing the scores of two subjects from the above data we can see that the lowest score in both the subjects is 56, the highest score in statistics is 99, and history is 98. We can also see that most students' score in statistics are in the 90-100 interval and in history the maximum students score is in the interval 80-90.

We need to make a note that data with more than two digits can be rounded to two digits before plotting or can be truncated to two digits. To truncate means to cut off. For a stem and leaf plot, you would truncate everything after the second digit.

Let us discuss the example by taking the number 385. This when rounded we will get 39 and when truncated we will get 38.

Let us take an example of the population of the various states and see how to represent the data in a back-to-back stem leaf plot the values by truncating and rounding them. The following table shows the population of six different states.

The rounded and truncated values are as shown in the table.

In this table, the stem of 149, 300, 403, 515, and 657 is common. However, when 299 is rounded off as 30, it is considered in the next step and is retained as 29 only in truncated value.

Here's a summary of our learning in this session:

- What a stem and leaf plot is and its usage
- How to segregate and prepare the data for constructing the graph
- A detail understanding of representing the various kinds of digits in the graph
- Understanding its application