How to Easily Calculate Mean, Median, and Mode from a Stem-and-Leaf Plot

Authored by
stats writer

December 5, 2025

Introduction to Stem-and-Leaf Plots and Descriptive Statistics

A Stem-and-Leaf Plot is a fundamental tool in exploratory data analysis. Unlike simple bar graphs or histograms, this type of plot offers a dual benefit: it displays the distribution of a quantitative dataset while simultaneously retaining the integrity of the original numerical values. This unique graphical representation achieves its effectiveness by systematically splitting each observation into a **stem**, usually representing the leading digits, and a **leaf**, which is typically the final digit. This preservation of raw data is highly advantageous when calculating essential measures of central tendency, such as the mean, median, and mode, as it allows for direct extraction of the necessary numbers.

The structure of the Stem-and-Leaf Plot naturally organizes the data in ascending order, which is a critical prerequisite for determining the median, and makes identifying repeated values, crucial for the mode, nearly instantaneous. However, calculating the mean requires reconstructing the original numerical values from the stems and leaves before performing the summation and division. Understanding how to properly decode the plot is the critical first step toward deriving these key descriptive statistics, which help summarize the characteristics and location of the data distribution.

The illustrative image below demonstrates the basic concept where the stem forms the primary numerical component and the leaf provides the precise trailing digit. For example, if the stem is '2' and the leaf is '3', the resulting observation is 23. If the stem is '10' and the leaf is '5', the observation is 105. It is essential to always review the plot's legend (or key) to confirm the specific scaling used, as sometimes the leaves may represent tenths or hundreds, rather than just the ones place.

Points Scored	
Stem	**Leaf**
1	1, 2, 2, 3
2	2, 3, 6, 6
3	0, 1, 2
4	2, 5, 6
5	1, 2
6	1

Translating the Plot into a Comprehensive Dataset

Before any statistical calculations can be performed, the numerical observations must be extracted directly from the Stem-and-Leaf Plot. This process involves pairing each leaf with its corresponding stem and converting that pair back into a single number. For a plot where the stem represents the tens digit and the leaf represents the units digit, every row must be systematically read to generate the complete list of individual data points. This extraction step is non-negotiable for achieving accuracy in the subsequent calculations, particularly the mean, which utilizes every value in the set.

The organized structure of the plot greatly simplifies this translation. Because the stems are listed vertically in ascending order and the leaves are listed horizontally in ascending order, the final extracted list of raw observations will already be sorted numerically. This pre-sorting saves considerable time and minimizes errors, especially when dealing with large volumes of data. Once the complete, sorted list of observations is finalized, we are prepared to apply the formal statistical formulas for the mean, median, and mode.

This tutorial will use a specific example dataset--plant heights--to walk through the precise steps required to find all three measures of central tendency. The detailed breakdown will demonstrate how to transition from the graphical plot to the necessary raw numbers, and subsequently how to apply the statistical principles accurately.

Case Study: Analyzing Plant Height Dataset

We will now proceed with a practical example involving a dataset represented by the following Stem-and-Leaf Plot. This plot visually summarizes the height (in centimeters, for instance) of 19 different plants, providing a clear visual representation of the data distribution.

Plant Height (inches)	
Stem	**Leaf**
1	1, 2, 3
2	6, 6
3	0, 1, 2
4	2, 5, 6
5	1, 2
6	1
7	8
8	2, 2
9	3, 4

Upon careful examination of the plot, we can systematically retrieve the 19 individual height measurements. For instance, the stem '1' paired with leaves '1, 2, 3' translates to the values 11, 12, and 13. The stem '8' paired with leaves '2, 2' translates to 82 and 82. Performing this reconstruction across all stems provides us with the complete, ordered dataset used for all calculations:

11, 12, 13, 26, 26, 30, 31, 31, 42, 45, 46, 51, 52, 61, 78, 82, 82, 93, 94.

With the raw data extracted, we can now proceed to calculate the three primary measures of central tendency. Since the total count of observations (sample size, N) is 19, this odd number is important when determining the position of the median.

Methodology: Step-by-Step Determination of the Mean

The mean, or arithmetic average, is the measure of central location most sensitive to the values of every single observation in the dataset. To calculate the mean from a Stem-and-Leaf Plot, we must first confirm that we have reconstructed all the individual values, and then apply the standard formula: the sum of all values divided by the total number of values (N).

Calculating the Mean: This process requires two distinct mathematical operations. First, summation of all 19 values; second, division by the sample size, which is N = 19.

Summation of Observations: We add every value identified from the plot:
$$ text{Sum} = 11 + 12 + 13 + 26 + 26 + 30 + 31 + 31 + 42 + 45 + 46 + 51 + 52 + 61 + 78 + 82 + 82 + 93 + 94 $$

The total sum is 906.

Division by Sample Size: We divide the sum (906) by the total number of plants (19):

Mean = (11+12+13+26+26+30+31+31+42+45+46+51+52+61+78+82+82+93+94) / 19 = **47.68**.

The calculated mean height for the plants in this dataset is approximately **47.68** cm. This figure represents the gravitational center of the distribution, meaning that if all the data points were perfectly balanced along a number line, 47.68 would be the pivot point. The mean is highly susceptible to outliers, but in this relatively symmetric distribution, it provides a robust measure of the average height.

Methodology: Identifying the Median (Positional Average)

The median is the value that splits the ordered dataset into two halves. Unlike the mean, the median is resistant to extreme values or outliers, making it a very reliable measure of typical central location, especially in skewed distributions. The primary advantage of using the Stem-and-Leaf Plot is that the data is already sorted, fulfilling the necessary first step automatically.

Determining the Median: Since we have an odd number of observations (N=19), the median will be a single value that exists exactly in the dataset. We can locate the position of the median using the formula (N + 1) / 2.

Calculate the Median Position: (19 + 1) / 2 = 10. The median is the 10th observation in the sorted list.

Locate the Value: Starting from the smallest observation (11), we count up to the 10th position.

The sorted list clearly shows the central value. The 10th position is highlighted below, illustrating the exact middle point of the distribution, with nine values preceding it and nine values following it:

11, 12, 13, 26, 26, 30, 31, 31, 42, **45**, 46, 51, 52, 61, 78, 82, 82, 93, 94

The median height turns out to be exactly **45** cm. This means that half of the plants measured are 45 cm or shorter, and half are 45 cm or taller. The closeness of the median (45) and the mean (47.68) suggests that the distribution of plant heights is relatively balanced and not heavily skewed.

Methodology: Locating the Mode (Frequency Average)

The mode is the most straightforward measure to find, representing the value or values that appear most frequently in the dataset. Because the Stem-and-Leaf Plot organizes the data, repeated values are easily identified by observing repeated leaves corresponding to the same stem, or in consecutive entries in the raw data list.

Finding the Mode: We examine the raw, ordered list of 19 plant heights to identify any values that occur more than once.

Systematic Review: Scan the entire dataset for duplications.

Identify Frequencies:

The value 26 appears twice.
The value 31 appears twice.
The value 82 appears twice.

In this specific dataset, we find that the values 26, 31, and 82 each occur twice, which is a higher frequency than any other single value. We highlight these recurring values in the sorted list to confirm the multiple modes:

11, 12, 13, **26**, **26**, 30, **31**, **31**, 42, 45, 46, 51, 52, 61, 78, **82**, **82**, 93, 94

Since there are three values (26, 31, and 82) that share the highest frequency (two occurrences each), this dataset is classified as **multimodal**, specifically trimodal. The modes of the dataset are **26, 31, and 82**. The existence of multiple modes suggests that the data distribution may have three distinct peaks or clusters, reflecting potential subgroups within the plant sample (e.g., small, medium, and large plants).

Interpreting and Contextualizing the Measures of Central Tendency

Having calculated the mean (47.68), the median (45), and the modes (26, 31, 82), we now possess a comprehensive summary of the plant height dataset derived entirely from the Stem-and-Leaf Plot. The purpose of these measures of central tendency is not merely to provide numerical answers but to offer insight into the typical value and shape of the distribution.

The proximity of the mean (47.68) and the median (45) confirms the visual impression provided by the plot: the data is fairly symmetrical, meaning there is no severe pull from extreme values toward one side of the distribution. If the mean were significantly larger than the median, it would indicate a right (positive) skew, typically caused by a few extremely large observations. Conversely, if the mean were much smaller, it would signal a left (negative) skew.

The trimodal nature of the data (26, 31, 82) is the most interesting finding. While the mean and median indicate a central tendency near 46 cm, the modes reveal that the most common individual heights cluster around 26 cm, 31 cm, and 82 cm. This strongly suggests that a simple average (mean) may not fully capture the complexity of this dataset, implying that researchers should investigate potential underlying factors that caused these distinct height groups to emerge.

Conclusion: Utilizing Stem-and-Leaf Analysis for Robust Statistics

The exercise demonstrates that the Stem-and-Leaf Plot is an extremely powerful tool for quickly calculating robust descriptive statistics. By preserving the raw numerical detail while simultaneously ordering the observations, it streamlines the process of finding the mean, median, and mode. Whether dealing with educational examples or real-world research data, mastering the translation of the plot into its component numbers is the key to accurate statistical analysis.

In summary, the mean provides the arithmetic balance, the median provides the exact center of the position, and the mode identifies the most frequent occurrences. Together, these statistics offer a complete and nuanced understanding of where the data is centered and how it is distributed.