Find the Mode of Grouped Data (With Examples)

Authored by

Mohammed looti

October 31, 2025

In the realm of data analysis, working with massive datasets is a common challenge. To manage this complexity, analysts often organize raw observations into **grouped data**. This vital organizational process condenses voluminous information into manageable categories, simplifying interpretation. However, calculating measures of **central tendency**, such as the **mode**, requires a specialized mathematical approach when dealing with grouped data. This comprehensive guide details the precise methodology for accurately estimating the mode of grouped data, supplemented by practical, step-by-step calculations.

The **mode** is a core measure within **descriptive statistics**, identifying the value or category that appears most frequently in a collection of data points. For ungrouped, raw data, determining the mode is trivial: a simple count of occurrences suffices. The challenge arises when data is aggregated into **class intervals**, as the granularity of individual observations is lost. Since we only know the total count (frequency) within each interval, we must turn to estimation techniques.

Consider a typical scenario where data has already been aggregated. Without the original, individual observations, we cannot calculate the exact mode. Instead, we must employ a sophisticated formula that leverages the relationship between class frequencies to pinpoint a reliable estimate of where the distribution peaks.

Range	Frequency
1-10	2
11-20	7
21-30	10
31-40	3
41-50	1

While the true, precise mode remains hidden within the class boundaries, the estimation formula is highly effective. It provides a robust approximation that accurately reflects the highest concentration of data points in the distribution.

The Formula for Estimating the Mode of Grouped Data

Because the exact value of the mode cannot be pinpointed directly from grouped data, statisticians rely on an estimation formula. This formula is ingeniously designed to consider the **frequency** of the class containing the mode (the **modal class**) relative to the frequencies of the classes immediately surrounding it. Essentially, it applies a weighted average, biasing the estimated mode towards the side of the modal class that has a higher concentration of data, thereby maximizing precision.

The standard formula used globally for calculating the mode of grouped data is represented as follows:

Mode of Grouped Data = L + W

Accurate application of this formula requires a deep understanding of each constituent variable. Below is a detailed breakdown of the components:

L: Represents the **lower limit** of the designated **modal class**. This is the starting value of the interval identified as having the highest frequency.
W: Denotes the **width of the modal class**. This value is derived by calculating the difference between the upper limit and the lower limit of the modal interval. For example, in a class interval of 21-30, the width is 9 (30 - 21).
Fm: Is the **frequency** of the **modal class** itself. This is the count of observations falling specifically within that interval.
F1: Refers to the **frequency** of the class interval that immediately precedes (comes before) the **modal class**.
F2: Signifies the **frequency** of the class interval that immediately succeeds (comes after) the <a href="https://en.wikipedia.org/wiki/Modal_class.

Identifying the Modal Class: The Essential First Step

Before any mathematical substitution can take place, the calculation of the grouped data mode hinges entirely on correctly identifying the **modal class**. The modal class is defined as the specific **class interval** within a **frequency distribution** that exhibits the highest count, or frequency. Since this interval contains the largest number of data points, it logically follows that the true mode is located somewhere within its boundaries.

To identify this crucial starting point, one must simply scan the frequency column of the distribution table and locate the largest number. The class interval corresponding to this maximum frequency is the modal class.

Referring back to the initial example provided earlier, we can quickly locate the modal class:

Range	Frequency
1-10	2
11-20	7
21-30	10
31-40	3
41-50	1

In this particular dataset, the frequency column clearly shows that the class interval 21-30 has the maximum frequency. Consequently, **21-30** is unequivocally identified as the **modal class**. This identification is absolutely foundational; any error at this stage will invalidate the subsequent calculation.

Example 1: Estimating the Mode for Student Exam Scores

We will now apply the grouped data mode formula to a real-world scenario involving student performance. Suppose a teacher has compiled the exam scores of 40 students and organized them into the following **frequency distribution**. Our objective is to estimate the most frequently occurring exam score.

Exam Score	Frequency
51-60	4
61-70	8
71-80	15
81-90	8
91-100	5

The first step, as always, is to identify the **modal class**. A quick examination of the frequency column reveals that the class interval 71-80 contains the highest frequency, which is 15. Therefore, **71-80** serves as our modal class.

Having established the modal class, we meticulously extract all necessary parameters for the formula:

L: The **lower limit** of the modal class (71-80) is **71**.
W: The **width of the modal class** is determined by 80 - 71, which equals **9**.
Fm: The maximum **frequency** of the modal class (71-80) is **15**.
F1: The **frequency** of the preceding class (61-70) is **8**.
F2: The **frequency** of the succeeding class (81-90) is **8**.

We now substitute these specific values into the mode estimation formula and proceed with the calculation:

Mode = L + W
Mode = 71 + 9
Mode = 71 + 9
Mode = 71 + 9
Mode = 71 + 9
Mode = 71 + 4.5
Mode = **75.5**

The final calculation yields an estimated modal exam score of **75.5**. This figure effectively approximates the peak of the distribution and indicates the score around which the majority of students clustered, providing clear insight into the class's performance trend.

Example 2: Calculating the Mode for Basketball Player Performance

To further solidify the methodology, let us examine a second scenario. Consider a dataset showing the number of points scored per game by 60 professional basketball players, organized into a **frequency distribution**. Our goal is to determine the estimated mode for points scored.

Points Scored	Frequency
1-10	8
11-20	25
21-30	14
31-40	9
41-50	4

Adhering to our established protocol, we begin by locating the **modal class**. By reviewing the frequency column, it is evident that the interval 11-20 has the highest count of 25 players. Consequently, **11-20** is designated as the modal class for this performance data.

We proceed to gather the necessary data points from the distribution table based on the identified modal class:

L: The **lower limit** of the modal class (11-20) is **11**.
W: The **width of the modal class** is calculated as 20 - 11, resulting in **9**.
Fm: The **frequency** of the modal class (11-20) is **25**.
F1: The **frequency** of the preceding class (1-10) is **8**.
F2: The **frequency** of the succeeding class (21-30) is **14**.

Finally, we input these values into the grouped data mode formula to execute the precise calculation:

Mode = L + W
Mode = 11 + 9
Mode = 11 + 9
Mode = 11 + 9
Mode = 11 + 9
Mode = 11 + 5.464285713
Mode = **16.46** (rounded to two decimal places)

The estimated mode for the points scored per game is approximately **16.46**. This value signifies the score range where the highest density of player performance lies, offering a critical summary statistic for player analysis.

Conclusion: Interpreting the Grouped Data Mode

Estimating the **mode** for **grouped data** is an indispensable technique within **descriptive statistics**. It is particularly useful when analysts are confronted with large, complex datasets or when access to the original individual observations is restricted. Although the result is an estimate rather than a definite exact value, the method provides exceptionally meaningful and robust insights into the most frequently occurring range within a distribution, especially when data is summarized in **frequency distributions**.

The procedure, while reliant on a specific formula, is highly systematic and straightforward once the foundational step--identifying the **modal class** and its surrounding frequencies--is completed accurately. By carefully following the logical steps and applying the formula with precision, analysts can confidently derive the estimated mode, thereby significantly enhancing their comprehension of underlying data trends and enabling more informed statistical interpretations.

This estimation methodology enjoys widespread application across diverse disciplines, including quality control, epidemiological studies, market analysis, and educational evaluation. It serves as an efficient and powerful tool for summarizing the central tendency of both categorical and continuous data that has been organized into defined class intervals.

Additional Resources

To further enhance your understanding of grouped data analysis and other statistical operations, we recommend exploring the following related tutorials: