How to Calculate Median from Frequency Table (With Examples)

Authored by
stats writer

December 4, 2025

Determining the median value from a frequency table is a fundamental skill in descriptive statistics. The median represents the **central tendency** of a data set, serving as the value that divides the ordered distribution into two equal halves. Unlike the mean, the median is robust against outliers, making it a preferred measure for skewed distributions. When working with large sets of raw data, organizing them into a frequency table significantly streamlines the calculation process, though it requires an understanding of how to use **cumulative frequency** to locate the precise middle point.

The core principle involves first identifying the total number of observations (N), which is the sum of all frequencies. This total dictates the position of the median. Once the positional value is calculated (N/2 or (N+1)/2), you must refer to the cumulative frequency column to find the interval or specific data value where this positional value falls. The data value associated with that position is defined as the median. This methodology ensures accurate identification of the 50th percentile mark within the organized distribution. We will explore the rigorous steps and practical examples necessary to master this crucial statistical technique, covering both discrete data and the complexities of grouped data.

Understanding the Role of Cumulative Frequency

When dealing with raw data, determining the median involves simple ordering. However, a frequency table organizes data based on how often each value occurs, making the use of **cumulative frequency** indispensable for locating the median position efficiently. Cumulative frequency is defined as the running total of frequencies; it tells us how many observations fall below the upper boundary of a given class or specific data value. This running total provides a quick map of the entire data distribution, allowing statisticians to pinpoint the exact location of the 50th percentile without listing every single value individually.

The first crucial step in calculating the median from any frequency distribution, whether discrete or continuous, is always the establishment of the cumulative frequency column. This column is built by adding the current frequency to the sum of all preceding frequencies. The final value in this column must always equal N, the total number of observations (the sum of all frequencies). This N value is critical because it dictates the location of the median observation, which is mathematically defined by its position in the ordered data array.

While the simplest approach for small frequency tables (where the data is discrete) is conceptually listing the values, as demonstrated in the initial examples, the robust, scalable method relies entirely on cumulative totals. This statistical tool is especially powerful when dealing with very large data sets or data that has been grouped into class intervals, where listing individual observations is impossible or highly impractical. Mastering the calculation of cumulative frequency is synonymous with mastering the calculation of the median from tabulated data.

Step-by-Step Calculation for Discrete Data

To precisely locate the median in a discrete frequency table, a structured, two-phase approach is required. This process is necessary to transition from aggregated frequency counts back to the central position within the data. These steps supersede the need to manually write out every single data point, making the calculation process much faster and less error-prone, especially for tables with large frequencies.

The core method relies on finding the positional value of the median observation and then matching that position against the cumulative frequency column. The following ordered steps outline the professional procedure for calculating the median of ungrouped, discrete data:

Calculate the Total Number of Observations (N): Sum all the frequencies (f) in the table. This gives N, the total sample size.
Determine the Positional Value: Calculate the position of the median item. For statistical purposes, the median position is generally determined by the formula: (N + 1) / 2. This formula identifies exactly which observation in the ordered list represents the median.
Calculate the Cumulative Frequencies (CF): Add a new column to the frequency table for the running total of frequencies.
Locate the Median Class/Value: Find the first cumulative frequency (CF) value that is greater than or equal to the positional value calculated in Step 2. The corresponding data value (X) in that row is the median.

It is important to remember the distinctions between odd and even sample sizes (N). While the (N+1)/2 rule handles both, understanding the conceptual difference is key. If N is odd, the median position will be an integer, locating a single observation. If N is even, the position will end in .5 (e.g., 10.5), indicating that the median is the average of the two observations surrounding that half position (i.e., the 10th and 11th values). In discrete frequency tables, this usually means identifying the data value corresponding to the CF that first exceeds N/2.

Example 1: Median Calculation for Discrete Data (Odd N)

Consider the scenario of analyzing the performance of seventeen soccer teams. The data is presented below in a frequency table, summarizing the number of wins achieved by teams in a specific league. Since the total number of teams (N) is 17, an odd number, we expect the median to correspond directly to a single observation within the data set. This example provides a clear illustration of how the positional formula works when N is odd, resulting in an exact integer position.

Wins	Frequency
0	2
1	3
2	4
3	1
4	2
5	3
6	2

To apply the formal method, we first determine the required position. The total frequency N is 17. Using the median position formula: Position = (N + 1) / 2. This yields (17 + 1) / 2 = 9. Therefore, the median value is the score corresponding to the **9th observation** in the ordered list. To locate this position quickly without listing all 17 values, we would typically add a **cumulative frequency** column to the table.

Using the cumulative frequency technique, we look for the row where the cumulative frequency first equals or exceeds 9. The original data shows the following distribution: 0 wins (frequency 2), 1 win (frequency 3), and 2 wins (frequency 4). The cumulative frequencies are: 2 (for 0 wins), 2+3=5 (for 1 win), and 5+4=9 (for 2 wins). Since the cumulative frequency reaches 9 exactly at the value X=2, the 9th observation must be 2.

If we were to manually arrange all of the individual values from smallest to largest, as shown below, we confirm the position:

Values: 0 (1st), 0 (2nd), 1 (3rd), 1 (4th), 1 (5th), 2 (6th), 2 (7th), 2 (8th), **2** (9th), 3 (10th), 4, 4, 5, 5, 5, 6, 6

Identifying the value directly in the middle of the ordered list confirms that the 9th value is indeed **2**. Thus, the median number of wins for a team in this league is **2**.

Example 2: Median Calculation for Discrete Data (Even N)

In this second example, we examine the household size data collected from 20 different households in a specific region. The resulting frequency table is provided below. Since the total number of observations, N, is 20 (an even number), the median will not correspond to a single observation but rather the average of the two central values. This necessity for averaging the middle pair is a defining characteristic of calculating the median for distributions with an even total frequency.

Household Size	Frequency
1	2
2	1
3	5
4	5
5	3
6	2
7	1
8	1

We must first calculate the positional values using the total N=20. The two central positions are N/2 and (N/2) + 1. This gives us 20/2 = 10, and 10 + 1 = 11. Therefore, we need to find the value of the **10th observation** and the **11th observation**. We locate these values using the **cumulative frequency** column. Based on the table provided, the cumulative frequencies quickly show us where these positions fall.

The cumulative frequency builds as follows: CF reaches 2 at X=1; CF reaches 3 at X=2; CF reaches 8 at X=3 (since 5 households have size 3); and CF reaches 13 at X=4 (since 5 households have size 4). Since the 10th and 11th positions both occur within the range covered by the value X=4, both the 10th and 11th observations are equal to 4.

Arranging all the individual values from smallest to largest provides a visual confirmation of this positioning:

Values: 1, 1, 2, 3, 3, 3, 3, 3, 4 (9th), **4** (10th), **4** (11th), 4, 4, 5, 5, 5, 6, 6, 7, 8

Since there are two values located directly in the middle, 4 and 4, we calculate the average of these two values to find the median: (4 + 4) / 2 = **4**. The median household size is **4**. This concludes the demonstration for discrete data, highlighting how the method adapts seamlessly whether the total frequency N is odd or even.

Calculating Median for Grouped Frequency Data

A significant challenge in statistics arises when dealing with grouped frequency tables, where data is clustered into class intervals (e.g., 10-19, 20-29). In this scenario, we do not know the exact values of the individual observations; we only know the frequency within each range. Consequently, we cannot determine the median directly as we did with discrete data. Instead, we must estimate the median value using **interpolation** within the identified **median class**. The assumption underlying this calculation is that the observations within the median class are evenly distributed across that interval.

The formula for calculating the median (M) from a grouped frequency table is substantially more complex than the discrete method, demanding precision in identifying the components:

M = L + * w

Where each term plays a specific role: **L** is the lower boundary of the median class; **N/2** is the median position (half the total frequency); **CF** is the cumulative frequency of the class immediately preceding the median class; **f** is the frequency of the median class itself; and **w** is the width of the median class interval. Finding the median class--the class whose cumulative frequency first exceeds N/2--is the critical first step in applying this formula accurately.

The process of interpolation essentially uses the ratio of the distance between the required median position (N/2) and the preceding cumulative total (CF) relative to the frequency of the median class (f). Multiplying this ratio by the class width (w) determines how far into the median class boundary (L) we must move to find the estimated median. This methodology provides the best statistical estimate of the center point when dealing with aggregated data, ensuring that the calculated median adheres to the definition of splitting the distribution into two equal halves.

Applying the Grouped Data Formula

To successfully apply the formula for grouped data, a systematic approach is necessary. First, calculate N (total frequency) and N/2 (the median position). Second, construct the **cumulative frequency** column. Third, identify the median class by locating the first CF greater than or equal to N/2. Fourth, extract the values for L, CF (preceding), f, and w based on the identified median class.

It is paramount to use the true class boundaries for L, especially when class intervals appear separated (e.g., 10-19 and 20-29). The lower boundary (L) for the median class must be adjusted to bridge the gap between classes (e.g., 19.5). Failure to use the true lower class boundary will lead to an inaccurate estimate of the median.

Once all variables are correctly identified and substituted into the formula, the final calculation yields the interpolated median value. This result is always a point estimate falling within the range of the median class. This detailed methodology ensures that, even when faced with aggregated data in a frequency table, we can still provide a robust measure of the central tendency for the underlying data set.

Key Differences Between Discrete and Grouped Median Calculation

While both methods aim to find the 50th percentile, the approaches for discrete (ungrouped) and grouped frequency tables differ fundamentally due to the nature of the data itself. Discrete data allows us to identify the exact observation number corresponding to the median position. For example, in a discrete table, if the 10th position is the median, and the 10th observation is '5', the median is exactly 5. The cumulative frequency column is used merely for location identification.

In contrast, grouped data necessitates a process of estimation. Once the median class is identified, the median is calculated through **interpolation**, assuming linear distribution within that interval. We are not finding an existing data point; we are estimating where the 50% mark would fall if we knew all the raw values. This distinction is crucial for understanding the precision of the resulting statistic.

Furthermore, the formulas differ significantly. For discrete data, we primarily use the positional formula (N+1)/2 and rely on the cumulative frequency column to look up the corresponding X-value. For grouped data, we use the complex interpolation formula involving the lower boundary (L) and the class width (w), transforming the positional calculation into a continuous estimate. Understanding these differences allows the statistician to select the appropriate method, ensuring the validity of the central tendency measure derived from the specific type of frequency table presented.

Conclusion: Importance of the Median in Data Analysis

The ability to calculate the median from a frequency table, whether discrete or grouped, is vital for accurate data analysis. The median provides a robust measure of **central tendency**, offering a perspective often more insightful than the mean, particularly in fields like economics, demographics, or testing, where distributions are frequently skewed by extreme values. By utilizing the structured methodology involving **cumulative frequency**, analysts can quickly process large volumes of data organized in tables, translating frequency counts back into meaningful positional statistics.

Through the systematic application of positional formulas and, where necessary, the interpolation technique for grouped data, we ensure that the estimated center point accurately reflects the distribution. This rigorous approach is a cornerstone of descriptive statistics, allowing researchers to draw reliable conclusions about the characteristics of their data set. Mastering these calculation methods ensures clarity and precision when summarizing large amounts of information.

Ultimately, the median serves as a powerful diagnostic tool. When comparing the median, mean, and mode, their relative positions can reveal the skewness of the distribution. A well-calculated median, derived efficiently from a frequency table, is thus an essential component in generating a complete and unbiased summary of any numerical data collection.