IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Measures of Central Tendency – Mean, Median, Mode

Mrs. Pranjal Prasad Tarte,

Assistant professor
Computer Science Department
P.V.G'S College of Science and Commerce, Pune, India

Abstract : **In any research, enormous data is collected and, to describe it meaningfully, one needs to summarize the same. The bulkiness of the data can be reduced by organizing it into a frequency table or histogram. Frequency distribution organizes the heap of data into a few meaningful categories. Collected data can also be summarized as a single index/value, which represents the entire data. These measures may also help in the comparison of data**

INTRODUCTION

Central tendency

By means of classification and frequency curve we get an idea about the shape of frequency distribution. In most of the frequency distributions we observe that all class-frequencies are not the same. Initially frequency is small in magnitude, later on it increases, and it reaches to maximum in the field in the middle part of the data and then falls down. In the other the frequency curve is bell- shaped. Here we note properties of observations are not uniformly spread. However, most of the observations get clustered in the central part of the data. This property of observations is described as Central tendency.

Measures of Central Tendency

Central Tendencies in Statistics are the numerical values that are used to represent mid-value or central value in a large collection of numerical data or the Statistical measure identifies a single value as a representative of the collected data. These obtained numerical values are called Central or Average Values in Statistics, Such a value is of great significance because it depicts the nature or characteristics of the entire data, which is otherwise very difficult to observe.
Some of the most commonly used measures of central tendency are:

1) Mean or Arithmetic mean
2) Median
3) Mode

1) Mean or Arithmetic mean ($\bar{X}$)

It is the most commonly used measure of central tendency and widely applicable average.
The arithmetic mean, also known as the average, is a statistical measure that represents the central tendency of a set of numbers. It is calculated by summing up all the values in a dataset and dividing the sum by the total number of values.
To calculate the arithmetic mean, follow these steps:

1) Add up all the numbers in the dataset.
2) Count the total number of values in the dataset.
3) Divide the sum by the total number of values.
4) Mathematically, the arithmetic mean can be represented as:

Arithmetic Mean = (Sum of all values) / (Total number of values)
For example, let's calculate the arithmetic mean of the following dataset: 5, 7, 9, 12, and 15.
Step 1: Add up all the numbers: 5 + 7 + 9 + 12 + 15 = 48
Step 2: Count the total number of values: There are 5 values in the dataset.
Step 3: Divide the sum by the total number of values: 48 / 5 = 9.6
Therefore, the arithmetic mean of the dataset is 9.6.

Example of grouped frequency distribution
Q. Marks obtained by the students are given below find out the Arithmetic mean for grouped frequency distribution

Marks obtained (xi)	Frequency (Fi)	FiXi
10	5	50
20	10	200
30	12	360
40	21	840

$\bar{X} = \sum FiXi / \sum Fi$ i=1, 2,…., n
= 1450/48
= 30.20
Arithmetic mean is 30.20

1.1) Merits and Demerits of Arithmetic Mean

Merits of Arithmetic Mean:

1) Although Mean is the most general way to calculate the central tendency of a dataset however it cannot always give the correct idea always, especially when there is a large gap between the datasets.
2) It is applicable only for quantitative data
3) It is unduly affected by the extreme observations.
4) It cannot be computed for the frequency distribution with open end classes.
5) It cannot be determined graphically.
6) Sometimes Arithmetic mean cannot be an actual observation in a data.

Demerits of Arithmetic Mean:

1) It is too much affected by the extreme values
2) Mostly it does not correspond to any value of the set of observations.
3) It cannot be calculated for frequency distribution with open end classes.
4) It does not convey any information about the spread or trend of data.
5) It is not a suitable measure of central value in case of highly skewed distribution.

1.2) Real life situations of Arithmetic Mean

1) Exam Grades: Suppose a teacher wants to determine the average score of a class on a math exam. They would add up the scores of all the students and divide the sum by the total number of students to find the arithmetic mean.
2) Temperature: Meteorologists use the arithmetic mean to calculate average temperatures. They sum up the recorded temperatures for a specific period, such as a month, and divide the sum by the number of days in that period to obtain the average temperature.
3) Financial Analysis: When assessing investment returns, the arithmetic mean is often used. For example, an investor might calculate the average annual return of a portfolio over several years to evaluate its performance.
4) Sports Statistics: In sports, the arithmetic mean is used to determine averages for various statistics. For instance, the average points scored per game by a basketball player or the average batting average of a baseball player can be calculated using the arithmetic mean.
5) Population Studies: In social sciences, researchers often use the arithmetic mean to describe characteristics of a population. For instance, the average income of a particular region or the average age of a group of people can be determined using the arithmetic mean.
6) Quality Control: In manufacturing, the arithmetic mean is used to monitor and control product quality. Measurements of product dimensions or weights are taken, and the average is calculated to ensure it falls within acceptable limits.

1.3) Certainly! Some additional points about the Arithmetic Mean:

1) Representation of Central Tendency: The arithmetic mean is one of the most common measures of central tendency. It aims to capture the typical or average value of a dataset. By summing up all the values and dividing by the total count, it provides a single representative value.
2) Sensitivity to Outliers: One important characteristic of the arithmetic mean is that it is sensitive to extreme values or outliers in the dataset. A single outlier can significantly affect the value of the mean. Therefore, when dealing with datasets that may contain outliers, it's important to consider alternative measures, such as the median or trimmed mean.
3) Suitable for Numerical Data: The arithmetic mean is primarily used for datasets consisting of numerical values. It is not applicable to categorical or ordinal data, as the calculation relies on the numerical values of the dataset.

4) Limitations with Skewed Distributions: The arithmetic mean may not accurately represent the typical value in datasets with skewed distributions. Skewness refers to the asymmetry of the data. In such cases, the median, which represents the middle value when the data is sorted, can be a more appropriate measure of central tendency.
5) Sample Mean vs. Population Mean: The arithmetic mean can be calculated for both a sample and a population. When calculating the mean for a sample, it represents the average value of the observed sample data. The population mean, on the other hand, represents the average value of an entire population. Depending on the context and purpose, the sample mean or the population mean may be used.
6) Continuous and Discrete Data: The arithmetic mean can be calculated for both continuous and discrete data. Continuous data refers to measurements that can take on any value within a range (e.g., height, weight), while discrete data refers to specific values (e.g., number of siblings, number of goals scored). The arithmetic mean can handle both types of data.
7) Additivity Property: The arithmetic mean has an additivity property, which means that if you have two separate datasets and calculate the mean for each, then the mean of the combined dataset is equal to the weighted average of the individual means. This property is useful in various statistical calculations.

Overall, the arithmetic mean is a widely used and intuitive measure of central tendency. It provides a useful summary statistic for understanding and analyzing datasets, but it is important to consider its limitations and potential alternatives depending on the characteristics of the data.

2) Median

The median is another measure of central tendency, similar to the arithmetic mean. While the mean represents the average value of a dataset, the median represents the middle value when the dataset is arranged in ascending or descending order.
To calculate the median, follow these steps:

1) Arrange the dataset in ascending or descending order.
2) If the dataset has an odd number of values, the median is the middle value.
3) If the dataset has an even number of values, the median is the average of the two middle values.
4) If the number of observations is odd, then (n + 1)/2th observation (in the ordered set) is the median. When the total number of observations are even, it is given by the mean of n/2th and (n/2 + 1)th observation

i.e. If the number of observations are odd, then (n + 1)/2th observation (in the ordered set) is the median. When the total number of observations are even, it is given by the mean of n/2th and (n/2 + 1)th observation.
For example, let's calculate the median of the following dataset: 3, 5, 7, 12, 15.
Step 1: Arrange the dataset in ascending order: 3, 5, 7, 12, 15.
Step 2: Since the dataset has an odd number of values (5), the median is the middle value, which is 7.
Therefore, the median of the dataset is 7.
The median is particularly useful when dealing with skewed distributions or when the dataset contains outliers. Unlike the mean, the median is not affected by extreme values. It provides a more robust estimate of the central tendency and is less influenced by the presence of outliers.

2.1) Merits and demerits of Median

Merits of Median:

1) Median is not influenced by extreme values because it is a positional average.
2) Median can be calculated in case of distribution with the open-end intervals
3) Medians can be located even if the data are incomplete.
4) Median can be located even for the qualitative factor such as ability, honesty etc.
5) Median can be located visually in case of discrete series.

Demerits of Median:

1) A slight change in series may bring drastic change in median value.

2.2) Real life situations of median

1) Income Distribution: The median is frequently used to analyse income distribution within a population. It provides a measure of the middle-income value, indicating the income level that separates the higher-earning and lower-earning individuals or households. The median income is often used as an indicator of the overall economic well-being and inequality within a society.
2) Housing Prices: When analysing housing prices in a particular area, the median price is commonly used to determine the typical or central price point. It helps understand the price range that most buyers or sellers fall within, providing insights into the affordability of housing options.
3) Test Scores: In educational settings, the median is often used to interpret test scores. By calculating the median score, educators and researchers can identify the middle performance level in a group of students. This information helps gauge the overall academic performance, evaluate the effectiveness of teaching methods, or compare different schools or districts.
4) Health Data: The median is frequently employed in healthcare and medical research to analyse various health-related measurements. For instance, it can be used to determine the median age of patients diagnosed with a specific disease, the median duration of a certain medical condition, or the median response time for a particular treatment.

5) Travel Time: When estimating travel time or commute duration, the median is often used to represent the typical or average time it takes to reach a destination. By considering the median travel time, transportation planners and commuters can gain a better understanding of the expected duration and plan accordingly.
6) Population Age: The median age is commonly used to describe the age distribution within a population. It represents the age at which half of the population is younger and the other half is older. The median age is useful for studying demographic trends, making policy decisions related to healthcare, retirement, and social services, and understanding the overall age structure of a society.

2.3) Here are a few key points about the median:

1) Robust to Outliers: The median is less sensitive to outliers compared to the arithmetic mean. Even if the dataset contains extreme values, the median remains unaffected.
2) Suitable for Skewed Distributions: Unlike the mean, the median is a suitable measure of central tendency for datasets with skewed distributions. It provides a better representation of the typical value when the data is not symmetrically distributed.
3) Applicable to Ordinal Data: In addition to numerical data, the median can also be calculated for ordinal data. Ordinal data represents categories with a specific order but does not necessarily have fixed numerical values. The median provides a meaningful measure of central tendency for such data.
4) Easy to Understand: The median is a relatively easy concept to understand and interpret. It represents the middle value in the dataset, which can be useful for understanding the distribution and characteristics of the data.
5) It's important to note that the choice between using the mean or the median depends on the nature of the data and the research question at hand. Both measures have their own strengths and limitations, and the appropriate measure should be selected based on the specific context and goals of the analysis.
6) Waiting Times: In various service industries, such as restaurants, hospitals, or customer support centres, the median waiting time is often used to evaluate the efficiency of service delivery. By estimating the median wait time, businesses can assess customer satisfaction, identify potential bottlenecks, and make improvements in their operations.

2.4) Certainly! Some additional points about the median:

1) Handling Skewed Distributions: The median is particularly useful when dealing with datasets that have skewed distributions. Skewness refers to the asymmetry of the data. In such cases, the mean may be influenced by extreme values, while the median provides a more robust estimate of the central tendency. For example, in a dataset with a long tail on one side, the median will be less affected by the extreme values in the tail.
2) Unequal Distribution of Values: The median is especially valuable when the dataset contains unequal distribution of values. It is not influenced by the magnitude of the values, but rather their relative position. Therefore, the median can give a representative measure even when the dataset has significant variations in values.
3) Ordinal Data: The median can be applied to ordinal data, which represents categories with a specific order but does not necessarily have fixed numerical values. For example, if you have a survey with responses like "strongly disagree," "disagree," "neutral," "agree," and "strongly agree," you can calculate the median to determine the central tendency of the responses.
4) Data with Outliers: The median is resistant to outliers, which are extreme values that deviate significantly from the other values in the dataset. Since the median is based on the middle value(s), it is not affected by outliers. This makes it a useful measure when you want to estimate the central tendency without the influence of extreme values.
5) Calculating the Median: If the dataset has an odd number of values, the median is simply the middle value. If the dataset has an even number of values, the median is the average of the two middle values. For example, in the dataset 2, 4, 6, 8, the median is the average of 4 and 6, which is 5.
6) Visualization: The median can be used to divide a dataset into two equal halves. In a box plot, the median is represented by a line dividing the box into two parts. This visualization provides insights into the distribution of the data and helps identify any Skewness.
7) While the median is a useful measure of central tendency, it does have limitations. It may not provide a complete picture of the dataset, as it only considers the middle value(s) and not the range or variability of the data. Therefore, it is often used in conjunction with other descriptive statistics and measures to gain a comprehensive understanding of the dataset.

3.1) Mode

The mode is another measure of central tendency that represents the most frequently occurring value(s) in a dataset or the observation with maximum frequency. Unlike the mean and median, which focus on the average or middle values, the mode identifies the value(s) that appear most frequently.

Example:

1. The mode of {4, 2, 4, 3, 2, 2} is 2 because it occurs three times, which is more than any other number.
2. Find out the mode of following observations.

X	10	12	14	16	18
F	2	12	23	16	8

Mode is 14.

3.1) Merits and Demerits of Mode

Merits of Mode:

1) Mode is the term that occurs most in the series hence it is not an isolated value like Median nor it is value like mean that may not be there in the series.
2) It is not affected by extreme values hence is a good representative of the series.
3) It can be found graphically also.
4) For open end intervals it is not necessary to know the length of open intervals.
5) It can also be used in the case of Quantitative phenomena.
6) With only just a single glance on data we can find its value. It is the simplest.
7) It is the most used average in daily life, such as average marks of a class, average number of students in a section, average size of shoes, etc.

Demerits of Mode:

1) Mode cannot be determined if the series is bimodal or multimodal.
2) Mode is based only on concentrated values; other values are not taken into account in-spite of their big difference with the mode. In continuous series only the lengths of class intervals are considered.
3) Mode is most affected by fluctuation of sampling.
4) Mode is not so rigidly defined. Solving the problem by different methods we won't get the same results as in case of mean.
5) It is not capable of further algebraic treatment. It is impossible to find the combined mode of some series as is in case of mean.

3.2) Real life situations of mode:

1) Clothing Sizes: In the fashion industry, the mode is used to identify the most commonly worn clothing sizes. By determining the mode size for a particular garment, manufacturers can ensure they produce enough inventory to meet customer demand and minimize stock outs.
2) Product Preferences: Businesses often use the mode to analyse customer preferences and identify the most popular products or features. For example, an e-commerce company may determine the mode colour, size, or style of a product to inform their inventory management and marketing strategies.
3) Voting and Elections: The mode is relevant in analysing election results. It helps identify the winning candidate or option by determining the mode number of votes. If multiple candidates or options have the same highest frequency of votes, it indicates a tie or the need for further resolution methods.
4) Traffic Patterns: The mode is useful in studying traffic patterns and identifying the most common or peak times for traffic congestion. By analyzing traffic data and identifying the mode time periods, urban planners and transportation authorities can make informed decisions about infrastructure improvements and traffic management strategies.
5) Internet Usage: Internet service providers and website administrators often analyze internet usage patterns to optimize their services. The mode can be used to identify peak usage periods, such as the mode time of day or day of the week when internet traffic is highest. This information helps in capacity planning and resource allocation.
6) Medical Diagnosis: In medical research and healthcare settings, the mode is employed to identify the most prevalent disease or medical condition within a population. It helps healthcare professionals and researchers understand the most commonly occurring ailments, allocate resources accordingly, and develop appropriate preventive measures.
7) Survey Responses: When analysing survey data, the mode can be used to identify the most frequent response to a particular question. This information helps researchers and businesses understand the prevailing opinions, preferences, or attitudes of the survey respondents.
8) Stock Market Analysis: The mode is sometimes used in stock market analysis to identify the most frequently occurring stock price within a given period. It can provide insights into the price levels at which a particular stock is heavily traded, indicating potential support or resistance levels.

3.3) Some key points about the mode:

1) Multiple Modes: A dataset can have one mode, known as unimodal, or multiple modes, known as bimodal, trimodal, or multimodal. If there are several values that occur with the same highest frequency, all of them are considered modes.
2) Categorical and Numerical Data: The mode can be calculated for both categorical and numerical data. In categorical data, the mode represents the most common category or class. For numerical data, it represents the value(s) with the highest frequency.
3) Useful with Nominal Data: The mode is particularly useful when dealing with nominal data, which consists of categories that have no inherent order. For example, if you have a dataset of eye colours (blue, brown, green), the mode would indicate the most frequently occurring eye colour.
4) Handling Skewed Distributions: The mode can be helpful in identifying the central tendency of a dataset with skewed distributions. In skewed datasets, the mode can provide information about the peak or cluster of values that occur most frequently, even if the distribution is not symmetric.
5) Missing Mode: It is possible for a dataset to have no mode if all values occur with the same frequency. In other words, if no value appears more frequently than others, the dataset is considered to have no mode.

6) Visual Representation: The mode can be visually represented in a histogram or bar chart by identifying the bar or category with the highest frequency. This helps visualize the most common values in the dataset.
7) It's worth noting that the mode may not always be the most appropriate measure of central tendency, especially when working with continuous or interval data. In such cases, the mean or median may be more suitable. Additionally, the mode may not provide a comprehensive summary of the dataset, as it focuses solely on the most frequent value(s) and does not consider the overall distribution or variability of the data.

Overall, the mode is a valuable measure for identifying the most frequently occurring value(s) in a dataset. It is particularly useful with nominal or categorical data and can provide insights into the central tendency of skewed distributions.

3.4) Certainly! some additional points about the mode:

1) Applicability to Different Data Types: The mode can be calculated for various types of data, including categorical, nominal, ordinal, and discrete numerical data. It is not limited to a specific type of data and can be used to identify the most common category or value within a dataset.
2) Data with No Mode or Multiple Modes: There are cases where a dataset may have no mode, meaning that no value appears more frequently than others. For example, in a dataset with the values [2, 4, 6, 8], there is no mode because all values occur with the same frequency. On the other hand, a dataset can also have multiple modes, indicating that two or more values occur with the same highest frequency. For example, in a dataset with the values [2, 4, 4, 6, 6, 8], both 4 and 6 are modes.
3) Usefulness in Descriptive Statistics: The mode is a fundamental measure used in descriptive statistics. Alongside the mean and median, it provides a comprehensive summary of a dataset's central tendency. By considering the mode, you gain insights into the most prevalent values, which can be valuable for understanding the characteristics of the data.
4) Handling Skewed Distributions: The mode is particularly useful when dealing with skewed distributions. Skewness refers to the asymmetry of the data. In such cases, the mode can help identify the peak or cluster of values that occur most frequently, providing information about the dominant features of the distribution.
5) Frequency Distribution: The mode is closely related to the concept of frequency distribution. A frequency distribution table or graph displays the values in a dataset along with their corresponding frequencies (how often they occur). The mode, in this context, represents the value(s) with the highest frequency in the distribution.
6) Limitations of the Mode: While the mode is a valuable measure, it does have some limitations. For instance, in datasets with continuous or interval data, the mode may not accurately represent the central tendency since it focuses on specific values rather than considering the entire range of values. In such cases, the mean or median might be more appropriate measures to use.
7) Remember that the choice of the mode, mean, or median as a measure of central tendency depends on the type of data, the distribution of values, and the specific objectives of the analysis. It's often beneficial to consider multiple measures together to obtain a more complete understanding of the dataset.

REFERENCES:

1) Swinscow TD, Campbell MJ. 10th ed(Indian) New Delhi: Viva Books Private Limited; 2003. Statistics at square one.
2) Gravetter FJ, Wallnau LB. 5th ed. Belmont: Wadsworth – Thomson Learning; 2000. Statistics for the behavioral sciences.
3) Sundaram KR, Dwivedi SN, Sreenivas V. 1st ed. New Delhi: B.I Publications Pvt Ltd; 2010. Medical statistics principles and methods.
4) Petrie A, Sabin C. 3rd ed. Oxford: Wiley-Blackwell; 2009. Medical statistics at a glance.
5) Norman GR, Streiner DL. 2nd ed. Hamilton: B.C. Decker Inc; 2000. Biostatistics the bare essentials.
6) SundarRao PS, Richard J. 4th ed. New Delhi: Prentice Hall of India Pvt Ltd; 2006. Introduction to biostatistics and research methods.
7) Glaser AN. 1st Indian Ed. New Delhi: Lippincott Williams and Wilkins; 2000. High Yield Biostatistics.
8) Dawson B, Trapp RG. 4th ed. New York: Mc-Graw Hill; 2004. Basic and Clinical Biostatistics.
9) Mogull, Robert G. "Popular Measures of Central Tendency." Mathematics Teacher 83, no. 9 (December 1990): 744–46.
10) Silver, G. L. "Operational measures of central tendency." Applied Mathematics and Computation 186, no. 2 (March 2007): 1379–84.
11) Manikandan, S. "Measures of central tendency: The mean." Journal of Pharmacology and Pharmacotherapeutics 2, no. 2 (2011): 140.
12) Manikandan, S. "Measures of central tendency: Median and mode." Journal of Pharmacology and Pharmacotherapeutics 2, no. 3 (2011): 214.
13) Gonzales, Vera A., and Kenneth J. Ottenbacher. "Measures of Central Tendency in Rehabilitation Research." American Journal of Physical Medicine & Rehabilitation 80, no. 2 (February 2001): 141–46.
14) Chambers, Ll G. "74.10 A Spurious Relation between Measures of Central Tendency." Mathematical Gazette 74, no. 467 (March 1990): 54.
15) Fess, Elaine Ewing. "Assessment of Normative Statements Through Measures of Central Tendency." Journal of Hand Therapy 1, no. 5 (October 1988): 198–99.

16) Farebrother, Richard William. "Neopythagorean Approaches to Measures of Central Tendency and Dispersion." CHANCE 33, no. 1 (January 2, 2020): 26–29.
17) Wilcox, Rand R., and H. J. Keselman. "Modern Robust Data Analysis Methods: Measures of Central Tendency." Psychological Methods 8, no. 3 (2003): 254–74.
18) Fabián, Zdeněk. "New Measures of Central Tendency and Variability of Continuous Distributions." Communications in Statistics - Theory and Methods 37, no. 2 (January 7, 2008): 159–74.
19) Sinova, Beatriz, Sara De La Rosa De Sáa, María Asunción Lubiano, and María Ángeles Gil. "An Overview on the Statistical Central Tendency for Fuzzy Data Sets." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 23, Suppl. 1 (December 2015): 105–32.
20) KLLOGJERI, PELLUMB, and ADRIAN KLLOGJERI. "Use of GeoGebra in teaching about central tendency and spread variability." Creative Mathematics and Informatics 21, no. 1 (2012): 57–64.