DOI: https://doi.org/10.3126/cognition.v5i1.55408

CONCEPTUALIZING CENTRAL TENDENCY AND DISPERSION IN APPLIED STATISTICS

Indra Mali Malakar[1]

Abstract

In any research to summarize the huge collected data for describing its characteristics, frequency distribution or histogram are used to reduce the heap of data. Thus, the collected data can be summarized as a single index or value representing the entire data that facilitate data comparison. Statisticians use a variety of numerical measures or indices in order to summarize data in a concise yet informative manner. The other way of representing the bunch of data is the tabular, diagrammatic and graphical approaches that visually illustrate the unorganized data. However, these techniques can not be enough to explain the quantitative data. Thus, statistical analysis is required to determine the various numerical measures that describe the inherent characteristic of frequency distribution. There are many methods of describing the characteristics of the collected data. The averages are the measures that condense a huge unwieldy set of numerical data into single numerical values which are representative of the entire distribution. Measures of dispersion show the information about the amount of variability or deviation or scatteredness or variation of the data from the average or central value. Thus both measures are used to interpret or analyze the data in academic and quantitative research.

Keywords: Mean, median, mode, statistical analysis & standard deviation.

Introduction

The significant objective of statistical analysis in any research is to explain the characteristics of a frequency distribution by determining various numerical measures. For analyzing and interpreting the main traits of the frequency distribution, it is necessary to determine the numerical measures like averages, measures of dispersion, skewness, kurtosis, etc. An enormous data is gathered in any kind of research to describe it meaningfully and one needs to summarize it for interpretation and analysis. The large size of data can be decreased by organizing it into a frequency table or histogram (Manikandha, 2011). In reality, the frequency distribution manages the mass of data into a few meaningful categories. Similarly, these collected data can be summarized as a single value or index for representing the entire data for interpretation. After the data have been classified and tabulated, the next process is to analyze it but these tabular, diagrammatic, and graphical approaches are the visual illustration of the unorganized data. However, these techniques are not sufficient enough to describe the quantitative data in detail. Thus there is a necessity to determine the various numerical measures that describe the inherent characteristics of frequency distribution which is the important objective of statistical analysis.

The averages are the measures that condense a huge unwieldy set of numerical data into single numerical values which are representative of the entire distribution. Averages have the typical fact that all other items of the distribution congregate around the averages and these values are in the central part of the frequency distribution that gives the idea about the concentration of the values. So they have also termed "measures of central tendency" (Nepal & Malakar, 2022). Therefore, central tendency is defined as "the statistical measure that identifies a single value as representative

1_Asst. Professor and PhD Scholar of Population Studies,
Patan Multiple Campus, TU, Patan Dhoka, Lalitpur, Nepal, Email: indraamali1982@gmail.com

of an entire distribution (Gravetter & Wallnau, 2000). To help in decision-making based on the data available and to explain the frequency distribution concisely, the measures of central tendency are necessary and they have the indispensable role to facilitate the comparison of various distributions of variables as well as to establish the degree of relationship among the variables. Therefore, averages are the first constraint for statistical analysis of the frequency distribution of the data (Mishra, 2010). The main goals of central tendency are to get the single value that represents the entire data, to facilitate comparison, to present the salient features of a mass of complex data, and to help decision-making.

Measures of dispersion provide information about the amount of variability or deviation or scatteredness or variation of the data from the average or central value. Thus the study of scatteredness or variation of the data from the averages of a distribution is known as the study of the measure of dispersion. Spiegel (2002) defined variation or dispersion of data as "the degree to which numerical data tend to spread about average values". L.R Connor (1932) defined "dispersion as a measure of the extent to which the individual items vary". The main goals of measures of dispersion are to find out the reliability of an average, to control the variation of the data from central distribution, to provide a comparison of variability of two or more than two distributions, to facilitate the use of other statistical measures for further analysis of data as regression and correlation analysis for further analysis of distribution and to help in devising a system of quality control. A limitation of a measure of central tendency is that the loss of information in condensing the whole data into a single point is substantial however this loss is partially recovered by supplementing it with a measure of dispersion (Anilkumar & Samiyya, 2013). If there is no difference between various values and average then there is no dispersion (Joshi & Singh, 2013).Therefore, dispersion shows the width of the distribution (Swinscow & Campbell, 2003). Thus, both measures are equally significant to explain the characteristics of a frequency distribution in any research.

Objectives

The main objective of this study is to analyze the concept of the measures of central tendency which include mean, median, and mode, and dispersion which includes standard deviation. The specific objectives are to provide the basic notion of the measure of central tendency and dispersion and to assess the application of these measures in applied statistics to the readers.

Methodology

This paper methodologically uses content analysis as a secondary data technique. The purpose of this method is, it helps the researcher to draw a comparative view from the secondary data produced by various writers to apply measures of central tendency and dispersion according to the different situations. The secondary sources have been grabbed from prior studies, relevant journals, and so on. Furthermore, the techniques like descriptive cum comparative research design and content analyses have been used in the study. The collective and descriptive information has been presented in different tables and some hypothetical examples have been given to show the application of using these measures of central tendency and dispersion in a different context.

Analysis and Interpretation of data

The most commonly used measures of central tendency are mean, median, and mode. The choices of the appropriate measure of central tendency rely on the level at which a variable is measured. The table is presented below for the measurement of central tendency:

Table 1: Measures of central tendency and level of measurement

Measure	Level of measurement	Example	Method of analysis
Mean	Interval/ratio	Temperature ^{0}C/^{0}F (no absolute zero-interval). Height, age (ratio).	Correlation/OLS regression
Median	Interval/ratio Ordinal	Size of household (ratio). Job grade, age groups, education (ordinal).	Correlation (ordinal). OLS regression (ordinal).
Mode	Interval/ratio Ordinal Nominal	Job grade, age groups, education (ordinal). Religion, ethnicity, occupation (nominal).	Chi-square test Proportion test

Argyrous, 1997 & Gurung, 2013.

To see how each of these measures of central tendency is calculated, the use of tables with hypothetical results has been shown below:

Table 2: Gender of student

Gender	Frequency
Male	10
Female	14
Third Gender	2
Total	26

Table 2 reveals the nominal level of measurement and the distribution of cases between males, females and third gender. At this level, the only measure of central tendency that is appropriate is the mode (M_o). Therefore, the mode is only the measure of central tendency that can be calculated for nominal data. A simple way to determine the modal value or category is to inspect the raw frequency table that is enough for finding mode. For example, the category for gender of students that has the highest frequency is male, with 14 responses (Table 2) i.e. the modal value (M_o) = Male.

The next hypothetical example is as follows: 30 survey participants were asked to indicate their level of agreement with the statement "Regular physical exercise is important for mental health". The result is shown below:

Table 3: Frequency distribution of "Agreement Level"

Agreement level	Frequency
Strongly disagree	4
Disagree	4
Neither agree nor disagree	10
Agree	15
Strongly agree	7
Total	40

Table 3 reveals the ordinal level of measurement and the distribution of cases of strongly disagree,

disagree, neither agree nor disagree, agree and strongly agree. At this level, the measure of central tendency can be mode or median. As the highest frequency is Agree, with 15 responses, the mode is Agree. In this case, the median can also be used for calculating the ordinal data. Since, there are 40 values; the middle value is at the 20th position. At the 20th position, Agree lies in it, so, the median is Agree (Bhandari, 2022).

The hypothetical example for ratio data is; a survey of 26 students was asked about their weight. The result is shown below:

Table 4: Frequency distribution of "Weights"

Weights in Kg	Frequency
0-10	4
10-20	6
20-30	8
30-40	5
40-50	3
Total	26

Table 4 reveals the ratio level of measurement and the distribution of weights according to the difference of 10 between the weight classes starting from (10-20) to ending at (40-50). In this hypothetical example, mode, median and mean can be used for calculating ratio data. The mode is 24, median is 23.75 and mean is 23.85 in this hypothetical example (Bhandari, 2020).

The importance of measures of central tendency is as follows:

- The measures of central tendency allow researchers to determine the typical numerical point in a set of data.
- The measures of central tendency determine the center value lying in the distribution of data of any sample which are being distributed on a range from the lowest value to the highest value.
- It is common to hear people describe measures of central tendency as "the average" score or point in a particular group because it describes what is typical, normal, usual, or representative. Although from a statistical perspective "the average" refers to the arithmetic mean, the concept of "average" is an easy way to think about what measures of central tendency say about data (Pederson, 2017).

A details description of the measures of central tendency and dispersion are given below:

Arithmetic Mean

Mean is the most commonly used measure of central tendency. There are different types of mean viz. arithmetic mean, weighted mean, geometric mean (GM), and harmonic mean (HM). The arithmetic mean or simply mean is the most popular and widely used measure of central tendency which is considered the ideal measure of central tendency since it satisfies almost all requisites of ideal measures of central tendency given by Professor Yule (Yule,1925).

Importance of Arithmetic Mean

- It is used to calculate the average value of quantitative data when the distribution does not have very large and very small items and used to obtain an average value of distribution having closed-end class intervals and having extreme items.
- The arithmetic mean is more suitable for average than others, while we want to deal with the quantitative measures such as average bonus, average income, average sales, average

profit, average production, average height, average expenditure, and revenue, etc.

- It uses every value in the data and therefore it is a good representative of the data.
- It is the measure of the central tendency that excellently resists the fluctuation between different samples (Glaser, 2000).

Demerits of mean

The major default of the mean is that it is sensitive to extreme values or outliers especially when the sample size is small (Dawson & Trapp, 2004). Therefore, it is not an appropriate measure of the central tendency for skewed distribution.

Weighted Arithmetic Mean

In simple arithmetic mean, all items in the distribution are equally important. But in practice, this may not be true. The relative importance of some items in distribution is more important than others (Petrie & Sabin, 2009). But in such cases, proper weight is to be given to various values. Then weighted arithmetic mean, usually denoted by $\overline{X}_w$ and is given by $\overline{X}_w = \frac{W_1X_1 + W_2X_2 + W_3X_3 + W_nX_n}{W_1 + W_2 + W_3 \ldots + W_n}$

$\therefore \overline{X} = \frac{\sum WX}{\sum W}$ Where, W = given weight. The hypothetical example is given below:

Table 5: Frequency distribution of an inquiry into the budgets of middle-class families in a family

	Food	Rent	Clothing	Fuel	Others
Expenses on	10%	15%	20%	15%	25%
Index number	90	100	85	65	137

The weighted average is calculated as:

Group	Weights (W)	Index No. (X)	WX
Food	10	90	900
Rent	15	100	1500
Clothing	20	85	1700
Fuel	15	65	975
Others	25	137	3425
	∑W = 85		∑WX = 8500

Therefore, the Weighted mean $(\overline{X})_W = \frac{\sum WX}{\sum W} = \frac{85000}{85} = 100.$

Geometric mean

It is the arithmetic mean of the values taken on a log scale. It is expressed as the nth root of the product of observation. It is given as

Geometric mean (GM)= $\sqrt[n]{(x_1)\,(x_2)\ldots(x_n)}$ and log (GM) = $\frac{\sum(log x)}{n}$. GM is the appropriate measure when values change exponentially and in case of skewed distribution that can be made symmetrical by a log transformation.

Importance and demerits of geometric mean

It is more commonly used in microbiological and serological research. The major disadvantage of it is that, it cannot be used in any of the values that are zero or negative.

Harmonic mean

It is the reciprocal of the arithmetic mean of the observations. Harmonic mean = $\frac{1}{\frac{\Sigma\frac{1}{x}}{n}} = \frac{n}{\Sigma\frac{1}{x}}$. The harmonic mean is suitable when the reciprocals of values are more useful.

Importance of harmonic mean

It is used when we want to determine the average sample size of several groups, each of which has a different sample size (Manikandan, 2011). In the comparison of the arithmetic mean concerning the geometric mean (GM) and harmonic mean (HM), the arithmetic mean is always greater than GM which is also always greater than HM (Norman & Streiner, 2000).

Median

The median is the measure of central tendency which is not calculated, but identified or determined. It is a measure that informs about the value of the middle observation when data have been arranged in ordered series from the lowest to the highest value. Half of the units of a sample (or population) lie above the median and half below the median (Forum for public health in South Eastern Europe, 2013). The median is a positional average that divides the whole distribution of data into lower 50% and upper 50% which is also called the middle value of the given distribution of the values. It is quite different from the mean as the median describes the position of the variable in the distribution.

Importance of median

- It is suitable to measure central tendency (or average) for qualitative characteristics such as knowledge, intelligence, beauty, honesty, talent, good, bad, defective, etc.
- It is also a more appropriate/ suitable (computable) average (or measure of central tendency) for the open-ended classified data (Nepal, 2017).
- The median can be used for ordinal categorical data and interval data. When analyzing the interval data, the median is preferred to the mean when the data are not normally (symmetrically) distributed, as it is less sensitive to the influence of outliers (Cluskey & Lalkhen, 2007).

Mode

Mode is the value (observation) in the series which repeats (occur) a maximum number of times or it is the value (observation) that has the highest frequency. Mode is the most frequently occurring value, whose repetition is maximum i.e. mode is the value, whose frequency is maximum.

Importance and defects of mode

- There are many situations in which A.M. and Median (Md) fail to reveal the characteristics of data such as most common stock, most common wage, most common height, the most common size of shoe, size of T-shirts and other ready-made garments we have in mind mode and not the A.M. or Median.
- It is not based on all observations and is not suitable for further mathematical treatment.

Standard Deviation

Standard deviation is defined as the positive square root of the average of the squares of the deviations of the items from their arithmetic mean. In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values (Bland & Altman, 1996).It is also termed the root mean squared deviation from the mean of the data. It was first suggested by Karl Pearson in 1893. It is usually denoted by σ (sigma) of the Greek Alphabet. Since, it satisfies most of the requisites of a good measure of dispersion, it is regarded as the best measure of dispersion (or ideal measure of dispersion). It is a more powerful and widely used measure of dispersion than others. The reason for using standard deviation (SD) as the best measure of dispersion is that if the observations are from a normal distribution, than 68 percent of observations lie between mean ± 1 SD, 95% of the observations lie between mean ± 2 SD and 99.7% of the observations lie between mean ± 3 SD (Manikandan, 2011). It should be well-defined, easy to calculate and understand, and should be based on all the items as well as it should not be affected by the personal bias of the researcher or extreme values in the series (Bhardwaj & Sharma, 2013).

Computation process of mean, median, mode and standard deviation

The mean is computed by adding all the values in the data set divided by the number of observations in it. Mean is given by the formula for ungrouped and grouped data as follows: For ungrouped data, $\bar{X} = \frac{\Sigma X}{n}$ and $\bar{X} = \frac{\Sigma fX}{N}$ and for grouped data, $\bar{X} = \frac{\Sigma fm}{n}$ where Σ (upper Greek letter sigma) refers to summation, x refers to the individual value, f refers to the recurrence of values, n is the number of observation, m is the midpoint of class interval and N is the sum of frequencies denoted by Σf (Sundar & Richard, 2006). It should be noted that the mean computed from the frequency distribution is not accurately the same as that computed from the raw data (Sundaram et.al, 2010). Median is calculated by using different formulas for ungrouped and grouped data as follows: (M_d)= value of $(\frac{n+1}{2})^{th}$ item for ungrouped data. For grouped data median is given as: (M_d) = value of $(\frac{n}{2})^{th}$ item and $n = \Sigma f$ and $M_d = L + \frac{h}{f}(\frac{n}{2} - c.f)$ for grouped data. Where n is total frequency, n/2 = position of median class, L = lower limit of the median class, f = frequency of median class, h = class size of median class or width of the median class, c.f = less than c.f preceding the c.f of the median class.

The mode is calculated for ungrouped data by inspection. If the ungrouped data has the highest frequency, then the value corresponding to the highest frequency is the mode. In continuous series or grouped data, the class interval in which the mode lies, corresponding to the highest frequency can be obtained by inspection. Then, the mode can be calculated by using the formula: Mode (M_0) = $L + \frac{\Delta_1}{\Delta_1+\Delta_2} x\, h$, for $\Delta_1 + \Delta_2 \neq 0$, where, L = Lowest value of the modal class, Δ_1 = Difference between highest frequency and its preceding frequency $(f_1 - f_0)$. Δ_2 = Difference between the highest frequency and its succeeding frequency $(f_1 - f_2)$. h = size of modal class interval/height of modal class interval. In case of the frequency distribution is irregular or the value of frequency distribution is haphazard, for finding the mode, then, the grouping method is applied and if the mode is ill-defined then an empirical method is applied to find the mode which is given by Mode = 3Median – 2 Mean.

Standard deviation is calculated for grouped and ungrouped data as follows: For ungrouped data,

S.D = $\sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$ for individual series, S.D = $\sqrt{\frac{\Sigma fX^2}{n} - (\frac{\Sigma fX}{n})^2}$ for discrete series and S.D = $\sqrt{\frac{\Sigma fm^2}{n} - (\frac{\Sigma fm}{n})^2}$ for continuous series.

The hypothetical examples have been given below for the calculation of mean, median, mode and standard deviation as follows:

Table 6: Sample of five measurements

Sample	4	6	10	12	14

Table 7: Frequency distribution of marks

Marks obtained	10	20	30	40	50	60	70	80	90
No. of persons:	3	6	9	20	25	18	7	3	1

Table 8: Frequency distribution of marks

Marks	1- 5	5– 10	10-15	15-20
Frequency	2	5	10	3

In Table 6, the sample of five measurements is 4, 6, 10, 12, and 14 have been given, then,
For Mean: sample mean $(\bar{x}) = \frac{\sum x}{n} = \frac{4+6+10+12+13}{5} = \frac{45}{5} = 9.$
For median: Position of median (M_d) = value of $(\frac{n+1}{2})^{th}$ item $= \frac{5+1}{2} = \frac{6}{2} = 3^{rd}$ term. The 3rd term is 10, so the median is 10. As the given samples do not repeat for maximum number of times, the mode is not confined in this case.
For standard deviation: S.D = $\sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{(4-9)^2+(6-9)^2+(10-9)^2+(12-9)^2+(14-9)^2}{5}} = 3.71.$ (Table 6).
The computation of mean, median, mode and standard deviation of Table 7 from a frequency distribution is given below:

Marks obtained(X)	Frequency (f)	fX	c.f	X^2	fX^2
10	3	30	3	100	300
20	6	120	9	400	2400
30	9	270	18	900	8100
40	20	800	38	1600	32000
50	25	1250	63	2500	62500
60	18	1080	81	3600	64800
70	7	490	88	4900	34300
80	3	240	91	6400	19200
90	1	90	92	8100	8100
	N = 92				$\sum fX^2$= 231700

For mean: $\sum fx = 4370$, N = 92. Now, $\bar{x} = \frac{\sum fx}{N} = \frac{4370}{92} = 47.5$ i.e. average marks is 47.5.
For median: Position of median (M_d)= value of $(\frac{n+1}{2})^{th}$ item = $\frac{92+1}{2}$ = 46.5th term. In the c.f table, the c.f greater than 46.5th term is 63. So, the corresponding value of 63 is 50. Thus, median is 50.

For mode: The highest frequency is 25 and its corresponding value is 50. So, mode is 50.

For standard deviation: S.D = $\sqrt{\frac{\sum fX^2}{n} - (\frac{\sum fX}{n})^2} = \sqrt{\frac{231700}{92} - (\frac{4370}{92})^2} = 16.19.$
The computation of mean, median, mode and standard deviation for grouped data of Table 8 is given below:

Class interval	Frequency (f)	Mid value (m)	fm	c.f	m^2	f m^2
1-5	2	3	6	2	9	18
5-10	5	7.5	37.5	7	56.25	281.25
10-15	10	12.5	125	17	156.25	1562.5
15-20	3	17.5	52.5	20	306.25	918.75
	N = 20		$\sum$fm= 221			$\sum fm^2$=2780.5

For mean: $\sum fm$= 141, N = 40, now $,\overline{X} = \frac{\sum fm}{N} = \frac{221}{20} =$ 11.05.

For median: Position of median (Md) = value of $(\frac{n}{2})^{th}$ item= $\frac{20}{2}$ = 10. In the c.f table, the c.f greater than 10 is 17. So, L = 10, f = 10, c.f =7, h = 5. Then, $M_d = L + \frac{h}{f}(\frac{n}{2} - c.f) = 10 + \frac{5}{10}(\frac{20}{2} - 7)$ = 11.5.

Discussion

One of the important properties of the mean is that it minimizes error in the prediction of any one value in data set i.e. it is the value that produces the lowest amount of error from all other values in the data set. An important property of the mean is that it includes every value in data set as part of the calculation. In addition, the mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero. Similarly, the median is an important metric to calculate because it gives us an idea of where the "center" of a dataset is located. It also gives an idea of the "typical" value in a given dataset. It turns out that the median is a more useful metric in the following circumstances. When the distribution is skewed and contains outliers, the median is used. Mode tells us what most of the pieces of data are doing within a set of information. It is representative score of group and can be used with non-numerical data.

i. Comparisons of the measures of central tendency

Of the three measures of central tendency, the mean is the most stable measure. If repeated samples were drawn from a given population, the mean would vary or fluctuate less than the modes or medians. The arithmetic mean is the most appropriate measure in situations in which the concern is for totals of combined performance of a group. When the primary concern is to learn what a typical value is, then, the median would be preferred (Polit & Hungler, 1999). Of the discussed measures of central tendency, the mean is by far the most widely used as it takes every score into account, is the most efficient measure of central tendency for normal distributions and is mathematically tractable making it possible for statisticians to develop statistical procedures for drawing inferences about means. On the other hand, the mean is not appropriate for highly skewed distributions and is less efficient than other measures of central tendency when extreme scores are possible. The geometric mean is a viable alternative if all the scores are positive and the distribution has a positive skew. The median is useful because its meaning is clear and it is more efficient than the mean in highly skewed distribution. However, it ignores many scores and is generally less efficient than the mean. The mode can be informative but should almost never be used as the only measure of central tendency since it is highly susceptible to sampling fluctuations (Weiss, 1989., Lane., 2009 & Statistics, 2009). The level of measurement plays a vital role to determine the appropriate index of central tendency that can be used to explain a variable. In general, the mode is appropriate for nominal measures. The median is appropriate for ordinal measures. The mean is appropriate for interval and ratio measures (Munro, Visintainer & Page, 1986). However, the higher the level of measurement, the greater the flexibility in choosing a descriptive statistic. Variables measured on an interval or ratio scale can use any of the three measures of central tendency, although it is usually preferable to use the mean (Polit & Hungler, 1999). In a skewed distribution, the values of the mode, median and mean differ. In a symmetric distribution, the mean, median and mode coincide. But the mean is typically higher than the median and the mode, in positively skewed distribution ($Mo<Md<\bar{X}$) (Figure I). The mean is lower than the median and the mode, in negatively skewed distributions ($Mo>Md>\bar{X}$) (Figure II). The figures are shown below (Ranchov, 2009).

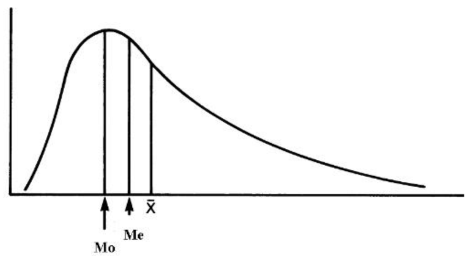

Figure I

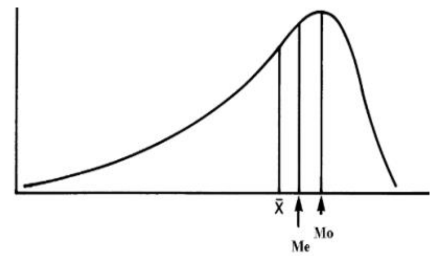

Figure II

ii. Application of central tendency and dispersion

Mean, median, mode and standard deviation are used according to the situation of the study for interpretation and analysis of data. It should be used according to the purpose of the study.

The hypothetical example is given for the usages of the central tendency in different situations. Suppose, five people of ages 6, 6, 8, 10, and 85 years are watching a movie. If we want to analyze the information regarding the type of movie they are watching, then the central tendency can be fruitful and appropriate to determine the characteristic of the information. Now let's find which measures of central tendency can be better to explain the information. If we use mean, then their mean age is (8+6+6+10+85)/5=115/5=23 years, which is the average or mean of the information. This average value of 23 years indicates that, they might be watching an adult movie. As there are four people who are under age eleven, and the old age person of 85 years, we cannot agree with this statement. The mean is not a good representation of the central value in this case. This value is skewed by 85 years.

The other approach is to take the median to determine whether it represents the group. The median is 8 years in this case. This median value exactly gives a far better representation of the age of the people for the analyzing the information, as most of the people are around this age group. The most suitable and possibility to interpret the information regarding the type of movie watching is the cartoon network. Therefore, the median in this example represents a good representation of the central value. The next hypothetical example for the application of mode is given below. Suppose, the shopkeeper sells 25 pairs of shoes a day. The sizes of the shoes are 26,22,23,25,26,21,26,37, 38,39,26,26,37,39,40,26,21,20,26,24,27,26,26,26,39.In this hypothetical example, the maximum number of times the value repeat is 26, so, the mode is 26. This implies that 26 number sizes of the shoe have more demand than other sizes of shoes. Therefore, in this regard, mean and the median are not suitable to represent the entire distribution. Therefore, when we discuss about the most common, then the use of mode is appropriate.

When there is a need of total picture of a distribution, then there is a need of dispersion. The hypothetical example is given below.

Table 9: Score secured by the students in different tests

Tests	1	2	3	4	5	**Mean**	**Median**	**S.D**
Student A	20	30	40	50	60	40	40	14.14
Student B	0	10	40	60	90	40	40	32.86

In table 15, the mean and median of both students are equal. In this example, if we use mean and median to analyze the consistency of the data or who is better in the study on the basis of test, then it is difficult to say who is better using mean and median as the mean and median are equal. The graph of the score secured by the two students A and B can be shown in the figure III as follows:

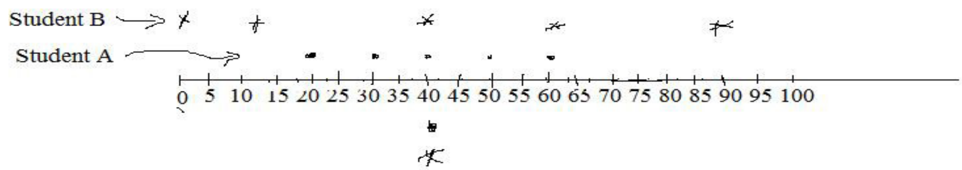

Figure III

The marks of student B have been scattered more far from the mean than that of a student of A in the graph. This also shows that score of student A is close to the mean indicating the consistency of the score. This implies that A's performance is more consistent than that of B or performance of student A is better than student B. If we determine the standard deviation of both students then the standard deviation of the score of student A is 14.14 and that of student B is 32.86. This also evidenced that in comparison to the scores secured by the students, the less value of standard deviation gives better consistency of the data or less spread out of data and more value of standard deviation gives less consistency of data or more spread out of data. So if we have a less value of standard deviation implies our data is more consistent and vice-versa. Thus, a measure of central value alone does not show the whole characteristics of the distribution of data unless the variation or the dispersion of the individual values is considered. Therefore, the standard deviation is used to check the consistency of the data or the spread out or variability of the data. But these methods of data analysis like measures of central tendency and dispersion do not make inferences about a population based on data obtained from a sample (Pandit, 2010).

Central tendency is a solitary central worth that attempt to portray a bunch of information by identifying the central value inside that set of information containing 'n' perceptions or things. Mean is frequently called 'math normal' however there are others likewise the median and mode and when the distribution is more slanted mean isn't all around to estimate the arrangement. Similarly, these measures of central tendency and dispersion can also be used according to the level of measurement and the type of variables as nominal, ordinal, interval, and ratio (Bhattacharya, 2017).

Table 10: The types of variables and use of central tendency and dispersion

Type of variables	The measure of central tendency	Measure of dispersion
Nominal	Mode	1 to 4 valid percent
Ordinal	Median	Minimum, maximum
Interval/ratio (not skewed)	Mean	Standard deviation
Interval/ratio (skewed)	Median	Standard deviation

Source: Bhattacharya, 2017.

Conclusion

This study analyzes the basic concept of measures of central tendency and dispersion. Measures of central tendency are the measures that indicate the approximated center of the distribution whereas the measures that describe the spread of the data are measures of dispersion. These measures include mean, median, mode, standard deviation, etc. These tools can be used in the interpretation of data in quantitative and academic research. From the above discussion, we can conclude that measures of dispersion are capable of algebraic treatment and are free from personal bias. These tools can be suitable for any researchers to analyze the collected data in more precise way for bringing the conclusion. There are many other statistical tools to measure the distribution of data. However, the researcher has discussed only the measures of central tendency and dispersion in this research. From the above study, the readers can achieve certain knowledge of measures of central tendency and dispersion as well as their application in various situations.

REFERENCES

Anilkumar, P. D., & Samiyya, N.V. (2013). Refining measure of central tendency and dispersion. *IOSR journal of mathematics* (IOSR-JM). e-ISSN: 2278-5728. Volume 6, issue 1.

Argyrous, G., & Argyrous, G. (1997). *Statistics for social research*. London: Macmillan.s

Bhandari, P. (2022, November 17). *Ordinal Data | Definition, Examples, Data Collection & Analysis.* Scribbr. Retrieved March 26, 2023, from https://www.scribbr.com/statistics/ordinal-data/.

Bhardwaj, A., & Sharma, K. (2013). Comparative study of various measures of dispersion. *Journal of advances in mathematics*. Vol1, No.1.

Bhattacharya, S. (2017). Measures of central tendency and measures of dispersion in the graphical demonstration. *International multidisciplinary research journal*. Volume -4, issue-12, December 2017.

Bland, J.M., & Altman, D.G. (1996). *"Statistics notes: measurement error"*. BMJ. **312** (7047): 1654. *doi:10.1136/bmj.312.7047.1654. PMC 2351401. PMID 8664723*.

Cluskey, A.M., & Lalkhen A.G. (2007). Statistics II: Central tendency and spread of data. *Continuing Education in Anesthesia. Critical care & pain*. Volume 7, No. 4.

Connor, L. R. (1932). *Statistics in theory and practice*. PpXvi + 371. 12s. 6d. net.

Dawson, B. & Trapp, R.P. (2004). *Basic and clinical biostatistics*. 4th ed. New York. Mc-Graw Hill.

Forum for public health in South Eastern Europe. (20130. *A handbook for teachers, researchers & health professionals (2nd edition).* Volume II.

Grancharova, G., & Hristova, P. (2004). *Medical statistics* (in Bulgarian). Pleven: Publishing House of the Medical University of Pleven.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences*. 5th ed. Belmont: Wadsworth- Thomson Learning; 2000.

Gurung, B.Y. (2013). Basic techniques of data analysis in demographic study: *Handbook for research methodology; approaches and techniques*. Central Department of Population Studies, Faculty of Humanities and Social Sciences.

Joshi, R.C., & Singh, M. (2013). *Computer-oriented numerical and statistical methods*. 12th editions.

Kelly, I.W., & Beamer J.E. (1986). Central tendency and dispersion. The essential union. *National council of teachers of mathematics*. Vol. 79. No. 1.

Lane, D.M. (2009). *Hyperstat online statistics textbook*. Trimmed mean. Available from URL: http://davidmlane.com/hyperstat/example_1.html. Accessed: April 20, 2009.

Manikandan, S. (2011). *Frequency distribution*. J PharmacolPharmacother; 2:54-6.

Mishra, S.R. (2010). *Quantitative techniques*. 2nd edition.

Munro, B.H., Visintainer, M.A., & Page, E. B. (1986). *Statistical methods for health care research.* Philadelphia: J.B. Lippincott Company.

Nepal .B. & Malakar , M. I. (2022). *Measurement in development*. 2nd edition. Quest publication, Kathmandu.

Nepal, B. (2017). *Statistics for research methodology.* 1st edition. Quest Publication., Kathmandu.

Norman G.R. &Strainers, D.L. (2000). *Biostatistics the bare essentials*. 2nd edition. Department of Clinical and Epidemiology and Biostatistics., Mc University. Hamilton, Ontario, Canada.

Pandit, P. R. (2010*). Mathematical statistics*. Revised edition.n.

Pederson, J. R. (2017). *Measures of central tendency*. In M. Allen (Ed.) The SAGE encyclopedia of communication research methods. Thousand Oaks, CA: SAGE.

Petrie, A., & Sabin, C. (2009). *Medical statistics at a glance*. 3rd ed. Oxford: Wiley-Blackwell.

Polit, D.F., & Hungler, B.P. (1999). Nursing research. *Principles and methods*. 6th rev. ed. Lippincot Williams & Wilkins.

Ranchov, G. (2008). *Biostatistics and biomathematics: Concepts, methods, applications (in Bulgarian)*. Sofia: Eco Print.

Spiegel, M. R. (2002). *Statistics* (Third edition). Newyork.

Statistics.com. Trimean. Available from URL: http://www.statistics.com/resources/glossary/t/trimmean.php. Accessed: April 20, 2009.

Sundar, R. P. & Richard. J. (2006). *Introduction to biostatistics and research method*. 4th ed. New Delhi India. Prentice Hall of India Pvt Ltd.

Sundaram, K. R., Dwivedi, S.N & Sreebuvas, V. (2010). *Medical statistics principles and methods*. 1st edition.

Swinscow, T.D., & Campbell, M.J. (2003). *Statistics at square one*. 10th ed. (Indian).

Weiss, N.A. (1989). *Elementary statistics*. Instuctor's Manual. Reading, Massachusetts: Addison-Wesley Publishing Company.

Yule, G. U. (1925).*"A Mathematical theory of evolution.*Based on the Conclusions of Dr. J. C. Willis, F.R.S.