K.T.S.P.Mandal's

SAHEBRAOJI BUTTEPATIL MAHAVIDYALAYA

Rajgurunagar, Tal-Khed, Dist-Pune,410505

Department of Commerce

F.Y.B.COM

SEM-I

Subject

Business Mathematics & Statistics-I

Chapter 4

Measures of Central Tendency

And

Measures of Dispersion

Unit 4

Measures of Central Tendency and Measures of Dispersion

Introduction-

A measure of central tendency is a number used to represent the center or middle of a set of data values. The mean, median, and mode are three commonly used measures of central tendency. Measures of Dispersion A measure of dispersion is a statistic that tells you how dispersed, or spread out, data values are. One simple measure of dispersion is the range, which is the difference between the greatest and least data values. Standard Deviation Another measure of dispersion is standard deviation, which describes the typical difference (or deviation) between a data value and the mean.

Frequency Distribution

A frequency distribution is an overview of all distinct values in some variable and the number of times they occur. That is, a frequency distribution tells
how **frequencies** are **distributed** over values.
Frequency distributions are mostly used for summarizing categorical variables. That's because metric variables tend to have many distinct values. These result in huge tables and charts that don't give insight into your data. In this case, histograms are the way to go as they visualize frequencies for *intervals* of values rather than each distinct value.

Data-

Any bit of information that is expressed in a value or numerical number is data.

Attributes-

An **attribute** refers to the quality of a characteristic. The theory of attributes deals with qualitative types of characteristics that are calculated by using quantitative measurements. Attributes have value that are described by words rather than numbers.

Variables-

A variable is any characteristics, number or quantity that can be measures or counted.
A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a **data item**. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye color and vehicle type are examples of variables. It is called a variable because the value may vary between data units in a population, and may change in value over time.

Classification of Data

a) Types of classification of data-
1) One way classification- Classified by single characteristics
2) Two way classification- Classified by two characteristics
3) Multi Way classification- Classified by more than two characteristics

Methods of Data classification

1) Inclusive Method

When the data are classified in such a way that both lower and upper limits of a class interval are included in the interval itself, then it is said to be the inclusive method. In this method the upper limit of a previous class less by 1 from the lower limit of the next class interval.

Class Intervals	Frequency
0-4	5
5-9	12
10-14	16
15-19	6

2) Exclusive method

When the data are classified in such a way that the upper limit of a class interval is the lower limit of the succeeding class interval(i.e. no data point falls into more than one class interval) then it is said to be the exclusive method.

Class Intervals	Frequency
0-5	5
5-10	12
10-15	16
15-20	6

Important terms relating to frequency distribution

1) **Class limits**- Every class has lower limit and upper limit i.e. upper limit and lower limit. Example- class 10-20 so Lower limit is 10 and Upper limit is 20
2) **Class Intervals**- The difference between the upper value and lower value is the class interval which can be normally same for all classes Example Class 10-20 so class interval(Difference) 10
3) **Mid value/ Class mark/ Mid point**- It is the mid value of the class

$$\frac{\text{Lower limit + Upper limit}}{2}$$

Example Class 10-20 so

Mid value = 10+20/2 = 30/2 =15 is mid value

4) **Class frequency**- The number of observation corresponding to each class gives the class frequency
5) **Open End classes**- The class with one unspecified limit is called as open-end class i.e. Above10 or below 20 .
6) **Class boundaries**- The class boundaries are the numbers up to which the actual magnitude of observation in the class can be extended. The class boundaries are also called as actual limit or extended limits

Class	Boundaries
10-19	9.5-19.5
20-29	19.5-29.5

7) **Class width** –It is an actual length of the class interval class width is always calculated by taking difference between upper boundaries and lower boundary

Class	Boundaries	Class width
10-19	9.5-19.5	10

8) **Cumulative frequency Distribution**-

Technically, a cumulative frequency distribution is the sum of the class and all classes below it in a frequency distribution. All that means is you're adding up a value and all of the values that came before it.

1) **Less than cumulative frequency**-

It is obtained by adding successively the frequencies of all the previous classes including the class against which it is written. The cumulate is started from the lowest to the highest size.

2) **More than cumulative frequency distribution:**

It is obtained by finding the cumulate total of frequencies starting from the highest to the lowest class. The less than cumulative frequency distribution and more than cumulative frequency distribution for the frequency distribution are given below:

Class	**Frequency**	**L.C.F.**	**M.C.F.**
0-10	10	10	100(90+10)
10-20	20	30(10+20)	90(70+20)
20-30	30	60(30+30)	70(40+30)
30-40	40	100(40+60)	40
	N=100		

9) Relative Frequency-

$$\text{Relative Frequency} = \frac{\text{Class Frequency}}{\text{Total Frequency}}$$

Histogram-

A histogram is a graphical display of data using bars of different heights. In a histogram, each bar groups numbers into ranges. Taller bars show that more data falls in that range. A histogram displays the shape and spread of continuous sample data

Frequency Polygon-

A frequency polygon is a graph constructed by using lines to join the midpoints of each interval, or bin. The heights of the points represent the frequencies. A frequency polygon can be created from the histogram or by calculating the midpoints of the bins from the frequency distribution table.

Ogive Curve

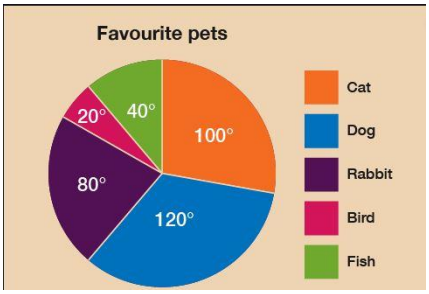
Cumulative Frequency curve also known as ogive curve .It is the graphical representation of a cumulative frequency distribution. Following are the types of ogive curve

a) **Less than ogive curve-** In this method start with the upper limits of the classes and go on adding the frequencies. When these frequencies are plotted , get a rising curve means upward sloping. For less than ogive curve it require less than cumulative frequency. For drawing a less than ogive curve we should take a L.C.F on Y axis and to take a upper limit on X axis.

b) **More than ogive curve-** In this method, start with the lower limit of classes and add frequency from the bottom. When these frequencies are plotted, get a declining curve means downward sloping. For more than ogive curve it require more than cumulative frequency. For drawing a less than ogive curve we should take a M.C.F. on Y axix and to take a lower limit on X axis

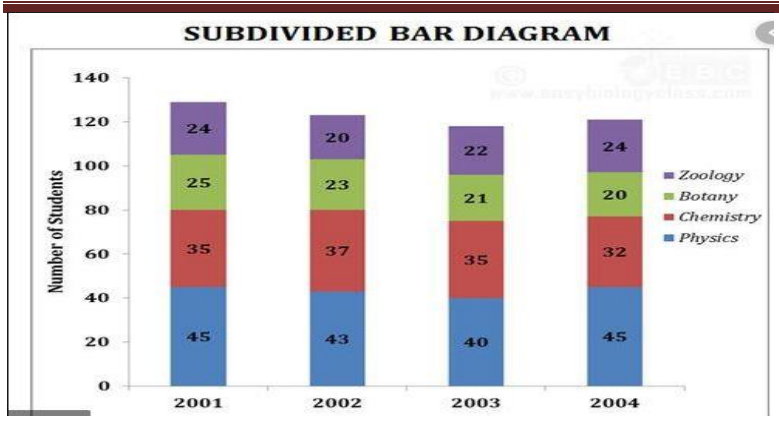
Pie Chart-

A **pie chart** (or a **circle chart**) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.



Sub Divided Bar Diagram-

A sub-divided or component bar chart is used to represent data in which the total magnitude is divided into different or components.



Measures of Central Tendency

A measure of central tendency is a number used to represent the center or middle of a set of data values. The mean, median, and mode are three commonly used measures of central tendency.

Requisites or Objectives of Ideal Measures of Central Tendency

1. It should be rigidly defined.
2. It should be simple to understand & easy to calculate.
3. It should be based upon all values of given data.
4. It should be capable of further mathematical treatment.
5. It should have sampling stability.
6. It should be not be unduly affected by extreme values.

Arithmetic Mean-

The average of a set of numerical values, as calculated by adding them together and dividing by the number of terms in the set. $A.M.=\frac{\text{Sum of the observations}}{\text{Number of Observatins}}$

Merits of A.M.:

1. It is simple to understand and easy to calculate.
2. It is affected by the value of every item in the series.
3. It is rigidly defined.
4. It is capable of further algebraic treatment.
5. It is calculated value and not based on the position in the series.

Demerits of A.M.:

1. It is affected by extreme items i.e., very small and very large items.
2. It can hardly be located by inspection.
3. In some cases A.M. does not represent the actual item. For example, average patients admitted in a hospital is 10.7 per day.
4. A.M. is not suitable in extremely asymmetrical distributions.

Geometric Mean-

The Geometric Mean is a special type of average where we multiply the numbers together and then take a square root (for two numbers), cube root (for three numbers) etc.

Merits

1. It is rigidly defined.
2. It is based on all the observations of the series.
3. It is suitable for measuring the relative changes.
4. It gives more weights to the small values and less weights to the large values.
5. It is used in averaging the ratios, percentages and in determining the rate gradual increase and decrease.
6. It is capable of further algebraic treatment. As such, if the G.M. and the number of items of two or more series are given, we can readily find out the combined G.M. of all the series

Demerits

1. It is not easy to understand by a man of ordinary prudence as it involves logarithmic operations. As such it is not popular like that of arithmetic average.
2. It is difficult to calculate as it involves finding out of the root of the products of certain values either directly, or through logarithmic operations.
3. It cannot be calculated, if the number of negative values is odd.
4. It cannot be calculated, if any value of a series is zero.
5. At times it gives a value which may not be found in the series, and may even be assured or impracticable.

Harmonic Mean-

In mathematics, the harmonic mean (sometimes called the sub contrary mean) is one of several kinds of average, and in particular one of the Pythagorean means. Typically, it is appropriate for situations when the average of rates is desired.

Merits

The harmonic mean has the following merits.

1. It is rigidly defined.
2. It is based on all the observations of a series i.e. it cannot be calculated ignoring any item of a series.
3. It is capable of further algebraic treatment.
4. It gives better result when the ends to be achieved are the same for the different means adopted.
5. It gives the greatest weight to the smallest item of a series.
6. It can be calculated even when a series contains any negative value.
7. It makes a skewed distribution a normal one.
8. It gives a curve straighter than that of the arithmetic and geometric mean.

Demerits

However, the harmonic mean suffers from the following demerits.

1. It is not easy to understand by a man of ordinary prudence.
2. Its calculation is cumbersome as it involves finding out of the reciprocals of the numbers.

3. It does not give better and accurate results when the means adopted are the same for the different ends achieved.
4. Its algebraic treatment is very much limited and not far and wide as that of the arithmetic mean.
5. It is greatly affected by the values of the extreme items.
6. It can not be calculated, if any, of the items is zero.

Median-

Median is the middle number in a sorted list of numbers. To determine the *median* value in a sequence of numbers, the numbers must first be sorted, or arranged, in value order from lowest to highest or highest to lowest.

Characteristics-

1.It is the value of the middle point of the array (not midpoint of range), such that half the item are above and half below it.

2.The value of the media is fixed by its position in the array and doesn't reflect the individual value.

3. Value must be ordered, and may be grouped, for computation.

4. It can be compute when ends are open

5. It is not applicable to qualitative data.

Merits-

1) It is easy to compute and understand.

2) It is well defined an ideal average should be.

3) It can also be computed in case of frequency distribution with open ended classes.

4) It is not affected by extreme values and also interdependent of range or dispersion of the data.

5) It can be determined graphically.

6) It is proper average for qualitative data where items are not measured but are scored.

7)It is only suitable average when the data are qualitative & it is possible to rank various items according to qualitative characteristics.

8) It can be calculated easily by watching the data.

9) In some cases median gives better result than mean.

Demerits :

1) For computing median data needs to be arranged in ascending or descending order.

2) It is not based on all the observations of the data.

3) It cannot be given further algebraic treatment.

4) It is affected by fluctuation of sampling.

5) It is not accurate when the data is not large.

6) In some cases median is determined approximately as the mid-point of two observations whereas for mean this does not happe

Mode-

Mode is highest frequency or most repeated observation. Mode may or may not exist in a series, or if it exists, it may not be unique, or its position may be somewhat uncertain.

Merits of Mode:

1. Mode is the most representative value of distribution, it is useful to calculate model wage.
2. It is not affected by the extreme items in the series.
3. It can be determined graphically.
4. For open-ended classes, Mode can be calculated.
5. It can be located by inspection.

Demerits of Mode:

1. It is not based on all observations.
2. Mode cannot be calculated when frequency distribution is ill-defined
3. It is not capable of further algebraic treatment. Like mean, combined mode cannot be calculated.
4. It is not rigidly defined measure because several formulae to calculate mode is used.

Types of Data

Raw Data-

Example -10, 12, 15, 19, 20, 13, 16, 20, 58is a raw data

Ungrouped Data-

Example-

X	5	6	7	8
F	11	25	20	13

Grouped Data-

Example-

X	5-10	10-15	15-20	20-25
F	11	25	20	13

Formulas for solving problems

For Group Data

- **Formula of Mean**- $\frac{\sum X}{n}$
- **Formula of Median** -
  - For Even number series= $\frac{n}{2}$th observation
  - For odd number series= $\frac{n+1}{2}$
- **Formula of Mode** - Most repeated observation

For Ungrouped Data

- **Formula of Mean-** $\frac{\sum fx}{n}$
- **Formula of Median-**

  First calculate LCF then $\frac{N}{2}$th l.c.f.'s observation is a median
- **Formula of Mode-**

Observation that has a highest frequency.

For Grouped Data

- **Formula of Mean-**
  - First calculate mid value as a X
  - The use $\frac{\sum fx}{n}$
- **Formula of Median-**

  -firstly find out median class by $\frac{n}{2}$ th l.c.f.

  -then use $l + \frac{\frac{n}{2} - c.f}{f} X\ h$

  L – lower limit

  c.f.-cumulative frequency of pre median class

  f- frequency of median class

  h- class width

- **Formula of Mode**

Firstly find out modal class i.e. the class that has a highest frequency is called a modal class

Then use- $L + \frac{fm - f1}{2fm - f1 - f2} X\ h$

L means lower limit of modal class

Fm means frequency of modal class

F1 means frequency of pre-modal class

F2 means frequency of post-modal class

h- Means Width

Combined Mean-

The arithmetic averages and the number of items in two or more related groups are known, the combined or the composite mean of the entire group can be obtained by following formula-

$$\frac{n1x1 + n2x2}{n1 + n2}$$

N1 means first group and x1 is a mean of first group

N2 means second group and x2 is a mean of second group

Measures of Dispersion

Introduction-

Dispersion refers to the variability in the size of itemsThe term dispersion is generally used in tow senses; firstly dispersion refers to the variations of the items among themselves. If the value of all the items of series is the same there will be no variation among the various items and the dispersion will be zero.

Definition-

Dispersion is the measure of the variations of the items.

Essentials of a good measure of Dispersion-

1) It should be rigidity defined i.e. for the same data all the methods should give the same answer.
2) It should be based on all the items, so that it would be representative of the data.
3) It should be readily comprehensible and easy to calculate.
4) It should be amenable for further mathematical treatment.
5) It should have sampling stability.
6) It should be unduly affected by extreme items.

Types of Measures of Dispersion-

A) Absolute Measures of Dispersions

The measure which possesses unit of observation is called absolute measures. For example average, height in cm or meters, income in Rs, dollars etc. Thus mean, mode, median are absolute measures. The absolute measures which are commonly used are-

- The range
- Variance
- The standard deviation

B) Relative Dispersions

The measures which do not posses any unit of observation is called relative measures. Relative measures are expressed in ratio or percentage. Thus, dispersion measured as a percentage or ratio of the average is called relative dispersion. It also known as coefficient of dispersion

Relative Measures of Dispersion

- Co-efficient of Variation (C.V.)
- Coefficient of range

Range

This is the simplest method for finding the measure of variability. Range is the difference between the highest and the lowest measurement. Range may be absolute or relative.

Definition

It is defined as the difference between the largest value(L) and the smallest value(S) in the given distribution.

Formula-

1)Range= L-S

L= largest value

S= Smallest value

2)Coefficient of Range= $\frac{L-S}{L+S}$

Merits of Range-

1) It is easy to calculate
2) It is simple to understand
3) It is rigidly defined
4) When items are limited it is useful
5) It gives quick results

Demerits of Range-

1) It is not based on entire-data
2) It gets easily affected by extreme values
3) It is not suitable from a mathematical angle
4) It is suitable for further mathematical treatment and it cannot be computed by open ended closes

Standard Deviation

Standard deviation is denoted by a Greek alphabet sigma(σ). It is defined "the positive square root of the arithmetic mean of the squares of the deviation of all the observation from their arithmetic mean."

Merits of S.D.

1) It is based on all observations
2) It is rigidly defined
3) It is capable of further mathematical treatment
4) It does not ignore algebraic signs of deviations
5) It is not much affected by sampling fluctuations

Demerits of S.D.

1) It is difficult to understand and to calculate
2) It cannot be computed for a distribution with open end class
3) It is unduly affected due to extreme deviations
4) It cannot be calculated for qualitative data.

Coefficient of Variation

The relative measure of SD it is called as coefficient of variation (C.V.).Coefficient of variation is always expressed in percentage.