GEO-A-CC-3-07-P- STATISTICAL METHODS IN GEOGRAPHY LAB

2.BASED ON FREQUENCY TABLE ,MEASURE OF CENTRAL TENDENCY , DISPERSION WOULD BE COMPUTED AND INTERPRET HISTOGRAM AND FREQUENCY POLYGON

List of West Bengal districts ranked by literacy rate

LITERACY RATE OF WESTBENGAL (%) 2011 Excluding 0-6 yrs population

1.	**Purba Medinipur**	87.66
2	**Kolkata**	87.14
3	**North 24 Parganas**	84.95
4	**Howrah**	83.85
5	**Hooghly**	82.55
6	**Darjeeling**	79.92
7	**Paschim Medinipur**	79.04
8	**South 24 Parganas**	78.57
9	**Bardhaman**	77.15
10	**Nadia**	75.58
11	**Cooch Behar**	75.49
12	**Dakshin Dinajpur**	73.86
13	**Jalpaiguri**	73.79
14	**Bankura**	70.95
15	**Birbhum**	70.90
16	**Murshidabad**	67.53
18	**Maldah**	62.71
19	**Uttar Dinajpur**	60.13

HONOURS COURSE: CORE SUBJECTS

2.14 GEO-A-CC-3-07-P – Statistical Methods in Geography Lab ✧ 30 Marks / 2 Credits

A laboratory notebook, comprising class assignments of the following is to be prepared and submitted. The exercises are to be drawn in pencil with photocopied representation of source materials where necessary. All texts are to be handwritten.

1. Construction of data matrix with each row representing an areal unit (districts / blocks / *mouzas* / towns) and corresponding columns of relevant attributes [15]
2. Based on the above, a frequency table, measures of central tendency, and dispersion would be computed and interpreted using histogram and frequency curve [15]
3. From the data matrix, a sample set (20%) would be drawn using random, systematic, and stratified methods of sampling and the samples would be located on a map with an explanation of the methods used [15]
4. Based on the sample set and using two relevant attributes, a scatter diagram and linear regression line would be plotted and residual from regression would be mapped with a short interpretation [15]
5. Viva-voce based on laboratory notebook (5 Marks)

References

Acevedo, M.F. 2012. Data Analysis and Statistics for Geography, Environmental Science and Engineering, CRC Press.

Harris, R., Jarvis, C. 2011. Statistics for Geography and Environmental Science, Prentice Hall.

McGrew Jr., J.C., Lembo Jr., A.J.,Monroe, C.B. 2014. An Introduction to Statistical Problem Solving in Geography, 3rd ed, Waveland Press.

Pal, S. K., 1998. Statistics for Geoscientists: Techniques and Applications, Concept Pub Co.

Rogerson, P.A. 2015. Statistical Methods for Geography: A Student's Guide, 4th ed, Sage.

Sarkar, A. 2015. Practical Geography: A Systematic Approach, 3rd ed, Orient Blackswan.

TABLE-1.5 (a)

Work Participation Rate (main + marginal) in West Bengal, 2001

State/District	Work Participation Rate Total
(1)	(2)
West Bengal	36.8
Burdwan	35.5
Birbhum	37.4
Bankura	44.7
Midnapore	39.0
Hooghly	36.9
Purulia	44.5
North 24 Parganas	33.4
South 24 Parganas	32.5
Kolkata	37.6
Howrah	33.7
Nadia	35.1
Murshidabad	34.2
Uttar Dinajpur	38.3
Dakshin Dinajpur	40.8
Malda	40.7
Jalpaiguri	38.3
Darjeeling	35.4
Cooch Behar	39.0

Frequency Distribution Table

Tabulation

After the data is arranged in a frequency table it will become much easier to handle it for further statistical analysis and it ca also be easily referred to in anywhere in the text. The raw data for this purpose can be transformed into grouped as well as ungrouped “Frequency Distribution Tables”.

Frequency Distribution Table

When we collect the data from the field it is found not in any order. We can’t refer it anywhere in The text as it does not carry any meaning in that form. Through tabulation we can make it appear more meaningful and also handy to manage. The raw data is converted into small groups and number of observations falling in each group are recorded. Observations falling in a group are considered as similar. By classifying the data into groups in tabulation we remove the minor differences in the data and retain the major differences. In a frequency distribution table we have two columns. First column gives the range of the group known as class and the second gives the frequency of each class i.e. number of observations falling in each class.

There are two types of frequency distributions: (a) Ungrouped and (b) Grouped.

Frequency Distribution Table

Grouped Frequency Distribution Table

Most of the time we have to handle the data which is continuous by nature, like: rainfall, agricultural
production, income etc. Such data occurs in frictions also. The range of the continuous data is also
large. In such cases, instead of the fixed number of the variable the classes are formed into some
ranges, known as classes and the number of observations, known as frequency, falling in each class
is tabulated. A hypothetical frequency distribution table of the grouped data of the daily rainfall of
90 days of a season of an area may look like the one given below

Distribution of Daily Rainfall of 90 days of an area(In mm)

Daily rainfall (in mm)	**Number of days (f)**
20-30	**5**
30-40	**6**
40-50	**11**
50-60	**18**
60-70	**19**
70-80	**15**
80-90	**13**
90-100	**2**
100-110	**1**
Total	**90**

Size of family (X)	Number of families(
1	2
2	14
3	22
4	24
5	18
6	14
7	6
Total	100

Ungrouped Frequency DistributionTable

In an ungrouped frequency distribution the classes consist of the fixed number and is used for the
data which is discontinuous by nature and can't occur in fractions; like size of the family, number of
schools, number of floods in a year to a river etc. The range of the discontinuous data, generally, is
not very large. An ungrouped frequency distribution table may look like the one given below:

Frequency Distribution Table

In the above frequency table, the values of the variable are tabulated for smaller group of the values
of the variables which are known as classes. Every class has two values known as **class limits:**
Lower class limit as well as **Upper class limit**. The difference between the upper limit and the lower limit of any class is known as **class interval**. In the present case the first class has the lower limit as 20.0mm
and the upper limit as 30.0 mm and the class interval of 10.0 mm..
In the second class the lower limit is 30.0 mm and the upper limit is 40.0 mm, and so on. All the class intervals of the above frequency
distribution table are equal. We notice that upper limit of every class become the lower limit of the next class. So, it should not be counted at two places. The convention is that any value less than the
upper limit should be included in the class itself. However, the values equal to the upper limit a class should go to the next class where it is the lower limit. So in every class the lower limit is included in
the class but not the upper limit.

Distribution of Daily Rainfall of 90 days of an area(In mm)

Daily rainfall (in mm)	Number of days (f)
20-30	5
30-40	6
40-50	11
50-60	18
60-70	19
70-80	15
80-90	13
90-100	2
100-110	1
Total	90

Frequency Distribution Table

In a grouped frequency distribution table number of classes and the class intervals are very important and are related to each other. If our class intervals are large, the number of classes will be less. On the contrary if the class intervals are small, number of classes will increase. A good frequency distribution table maintains balance between the two. Very large number of classes will lose the advantage of summarising the data. A very small number of classes like; 2 , 3 or 4 will result in significant loss of information.

There are suggestions regarding the number of classes, one such suggestion traditionally referred in the books is that the number of classes of a frequency distribution table, k , should be determined by the formula:

$K = 1 + 1.33 \log N$ which is hardly in practice.

Even when it is found to have class interval not in rounded form, the class intervals of multiple of five or ten are preferred due to practical reasons.

Unequal Class interval

The difference between upper limit and the lower limit of a class is known as the class interval which may be equal or may not be equal for all the classes. Class intervals are commonly of equalsize. In some cases, however, the equal class intervals are not required also. For example, the tabulation of urban settlements whose size in India varies from below 5000 population to 12442000 (highest population of Mumbai 2011) population, uses unequal class intervals due to the range of variations in data. For a range of 12437000, if we use equal class intervals of 5000 each we require 12437000/5000= 2488 (after rounding) classes. This as cumbersome as the data itself, no simplification in handling and interpretation.On the other hand if we take 10 classes of class interval of 10,00,000.0 population, we heavily loose the details as the very first class from below 5000 to 10,00,000.0 (below million cities) will have 7882 towns out of total 7935 towns in India in 2011 (99.3 %). This is as bad as having no information.

In such cases where the range of data is too large, for example population of towns, income of individuals in a society, land holdings among farmers etc. we are forced to go for unequal class intervals in such a manner that class intervals are smaller to begin with the smaller values and become larger and larger as we procced to the higher values. Indicating that smaller differences can’t be ignored at lower end but same differences are not equally important as we move to higher values where only higher differences matter. Thus Census of India classifies the towns in the form of unequal class intervals as given below:

Size class distribution of towns India 2011

Size class of towns	Population Class interval	Number of towns (2011)
Class VI TOWN	Below 5000	499
Class V	5000-10000	2188
Class IV	10000-20000	2238
Class III	20000-50000	1912
Class II	50000-100000	600
Class I	100000 and above	496
	TOTAL	7933

Either for equal or for unequal class interval, the choice of the class intervals is crucial. For equal class intervals one has to decide about the number of class intervals only. Range of data divided by number of classes will determine the class intervals. Often, the researchers marginally alter it also to suit their convenience. For example if the class interval as per calculations are found to be 19.73 one can change it to 20.0 for the ease of computations and interpretations. There are no hard and fast rules regarding number of classes. The guiding principle is that they should not be too many or too less. Commonly their number lies between 9,10 to 12, 15. For unequal class intervals, number of classes are generally less as each class represents a category of the data and there should not be larger number of categories to avoid confusion. For example, in the case of census classification of towns of India, class intervals correspond to well recognized six classes of towns. What is more important in such cases is the understanding of the researcher To convert the data into meaningful categories

Calculation Table for histogram, frequency Polygon and Curve ,Ogive

Class Limit % of literate	Class Boundary % of literate	Tally	Frequency No of literate (f)	Class Width (i)	Midvalue (x)	fx	Lower class boundary	Cumulative Frequency	
								Less than	More than

Calculation Table for histogram, frequency Polygon and Curve ,Ogive

Class Limit % of literate	Class Boundary % of literate	Tally	Frequency No of literate (f)	Class Width (i)	Midvalue (x)	fx	Lower class boundary	Cumulative Frequency	
								Less than	More than
56-62	55.5-62.5		2	7	59	118	55.5	0	19
63-69	62.5-69.5		2		66	132	62.5	2	17
70-76	69.5-765		7		73	511	69.5	4	15
77-83	76.5-83.5		5		80	408	76.5	11	8
84-90	83.5-90.5		3		87	261	83.5	16	3
							90.5	19	0
			Total(N)= 19			**Total=fx1422**			

HISTOGRAM

A frequency distribution table arranges the data into some ordered form which helps us in understanding the distributional properties of the data in a much better way than the raw data. For example , after transferring the data into a frequency distribution form, we can easily see as to how many observations are found in the middle of the values and how many on the either side of it. We can also see the inequalities in the distribution and other important socially important characteristics of the data. These characteristics become more visible if we plot the distribution of the data on a “Histogram”.

- A histogram is a collection of a set of rectangles with bases equal to the class interval of each classof the corresponding frequency distribution and the height of the rectangle will be equal to the corresponding frequencies of each class. Histogram is a non-cumulative frequency graph, it is drawn on a natural scale in which the representative frequencies of the different class of values are represented through vertical rectangles drawn closed to each other. Measure of central tendency, mode can be easily determined with the help of this graph.

Uses of histogram:

1. Represents the data in graphic form.
2. Provides the knowledge of how the scores in the group are distributed. Whether the scores are piled up at the lower or higher end of the distribution or are evenly and regularly distributed throughout the scale.
3. Frequency Polygon. The frequency polygon is a frequency graph which is drawn by joining the coordinating points of the mid-values of the class intervals and their corresponding frequencies.

HISTOGRAM

How to draw a Histogram:

Step—1:

Represent the class intervals of the variables along the X axis and their frequencies along the Y-axis on natural scale.

Step—2:

Start X axis with the lower limit of the lowest class interval. When the lower limit happens to be a distant score from the origin give a break in the X-axis n to indicate that the vertical axis has been moved in for convenience.

Step—3:

Now draw rectangular bars in parallel to Y axis above each of the class intervals with class units as base: The areas of rectangles must be proportional to the frequencies of the corresponding classes.

In this graph we shall take class intervals in the X axis and frequencies in the Y axis. Before plotting the graph we have to convert the class into their exact limits.

c.i.	f
19.5—24.5	2
24.5—29.5	2
29.5—34.5	5
34.5—39.5	10
39.5—44.5	6
44.5—49.5	2
49.5—54.5	3

Histogram plotted from the data.

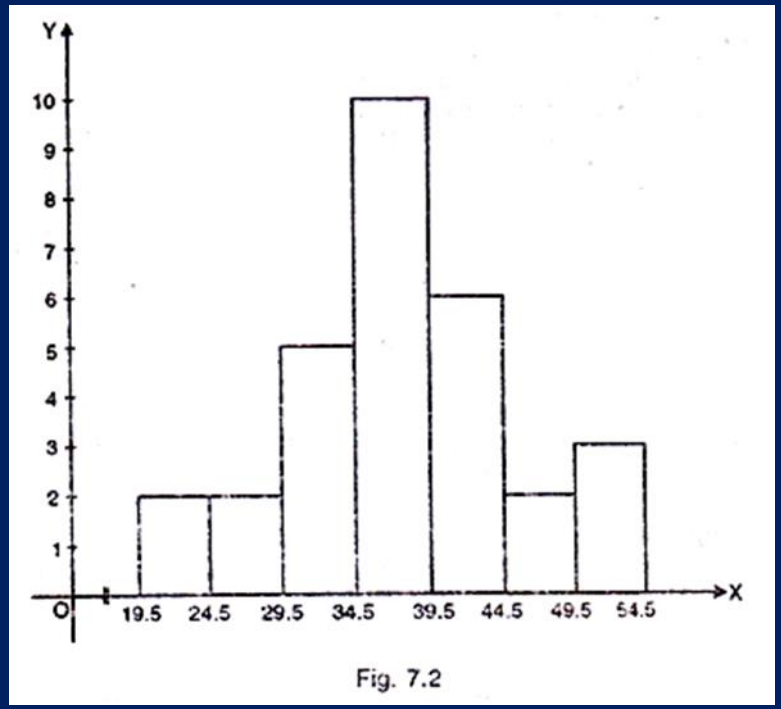


Fig. 7.2

HISTOGRAM

Advantages of histogram:

1. It is easy to draw and simple to understand.

2. It helps us to understand the distribution easily and quickly.

3. It is more precise than the polygene.

Limitations of histogram:

1. It is not possible to plot more than one distribution on same axes as histogram.

2. Comparison of more than one frequency distribution on the same axes is not possible.

3. It is not possible to make it smooth.

Frequency Polygon / Curve

A frequency polygon is almost identical to a histogram, which is used to compare sets of data or to display a cumulative frequency distribution. It uses a line graph to represent quantitative data. Statistics deals with the collection of data and information for a particular purpose. The tabulation of each run for each ball in cricket gives the statistics of the game. Tables, graphs, pie-charts, bar graphs, histograms, polygons etc. are used to represent statistical data pictorially. Frequency polygons are a visually substantial method of representing quantitative data and its frequencies. Let us discuss how to represent a frequency polygon.

Frequency Polygon" by joining the middle points of the upper sides of each bar. To show the pattern of change as a gradual process the polygon is converted into a smooth curve also, which is known as"Frequency Distribution Curve" or only frequency curve.

<u>Steps to Draw Frequency Polygon</u>

To draw frequency polygons, first we need to draw histogram and then follow the below steps:

Step 1- Choose the class interval and mark the values on the horizontal axes

Step 2- Mark the mid value of each interval on the horizontal axes.

Step 3- Mark the frequency of the class on the vertical axes.

Step 4- Corresponding to the frequency of each class interval, mark a point at the height in the middle of the class interval

Step 5- Connect these points using the line segment.

Step 6- The obtained representation is a frequency polygon.

Frequency Polygon /Curve

Marks in Mathematics	40–45	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–95	95–99
No. of students	1	3	2	4	5	6	10	8	5	6	2	1

c.i.	f.
34.5—39.5	0
39.5—44.5	1
44.5—49.5	3
49.5—54.5	2
54.5—59.5	4
59.5—64.5	5
64.5—69.5	6
69.5—74.5	10
74.5—79.5	8
79.5—84.5	5
84.5—89.5	6
89.5—94.5	2
94.5—99.5	1
99.5—104.5	0

Advantages of frequency polygon:

1. It is easy to draw and simple to understand.
2. It is possible to plot two distributions at a time on same axes.
3. Comparison of two distributions can be made through frequency polygon.
4. It is possible to make it smooth.

Limitations of frequency polygon:

1. It is less precise.
2. It is not accurate in terms of area the frequency upon each interval.

Uses of frequency polygon:

1. When two or more distributions are to be compared the frequency polygon is used.
2. It represents the data in graphic form.
3. It provides knowledge of how the scores in one or more group are distributed. Whether the scores are piled up at the lower or higher end of the distribution or are evenly and regularly distributed throughout the scale.

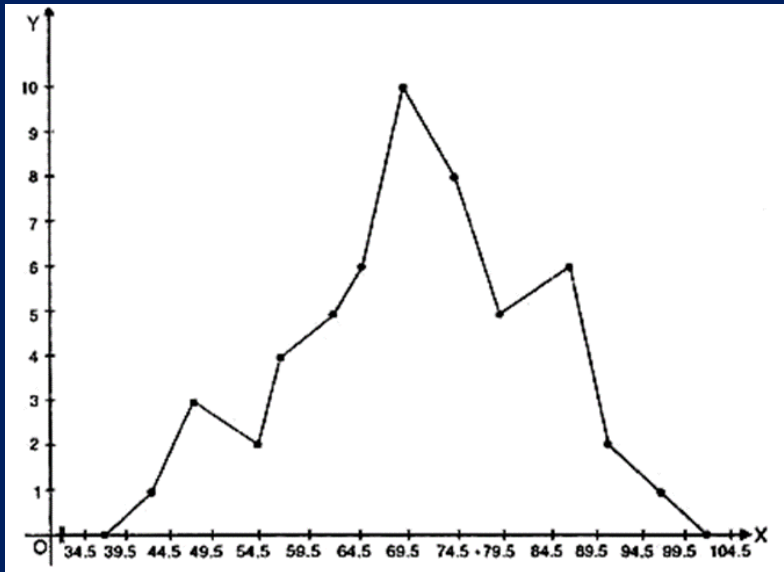


Fig. 7.3 Frequency polygon plotted from the data.

Frequency Polygon /Curve

c.i. (with exact limit)	f	Smoothed frequency
34.5—39.5	0	0 + 0 + 1 ÷ 3 = .33
39.5—44.5	1	0 + 1 + 3 ÷ 3 = 1.33
44.5—49.5	3	1 + 3 + 2 ÷ 3 = 2.00
49.5—54.5	2	3 + 2 + 4 ÷ 3 = 3.00
54.5—59.5	4	2 + 4 + 5 ÷ 3 = 3.67
59.5—64.5	5	4 + 5 + 6 ÷ 3 = 5.00
64.5—69.5	6	5 + 6 +10 ÷ 3 = 7.00
69.5—74.5	10	6 +10 + 8 ÷ 3 = 8.00
74.5—79.5	8	10 + 8 + 5 ÷ 3 = 7.67
79.5—84.5	5	8 + 5 + 6 ÷ 3 = 6.33
84.5—89.5	6	5 + 6 + 2 ÷ 3 = 4.33
89.5—94.5	2	6 + 2 + 1 ÷ 3 = 3.00
94.5—99.5	1	2 + 1 + 0 ÷ 3 = 1.00
99.5—104.5	0	1 + 0 + 0 ÷ 3 = .33

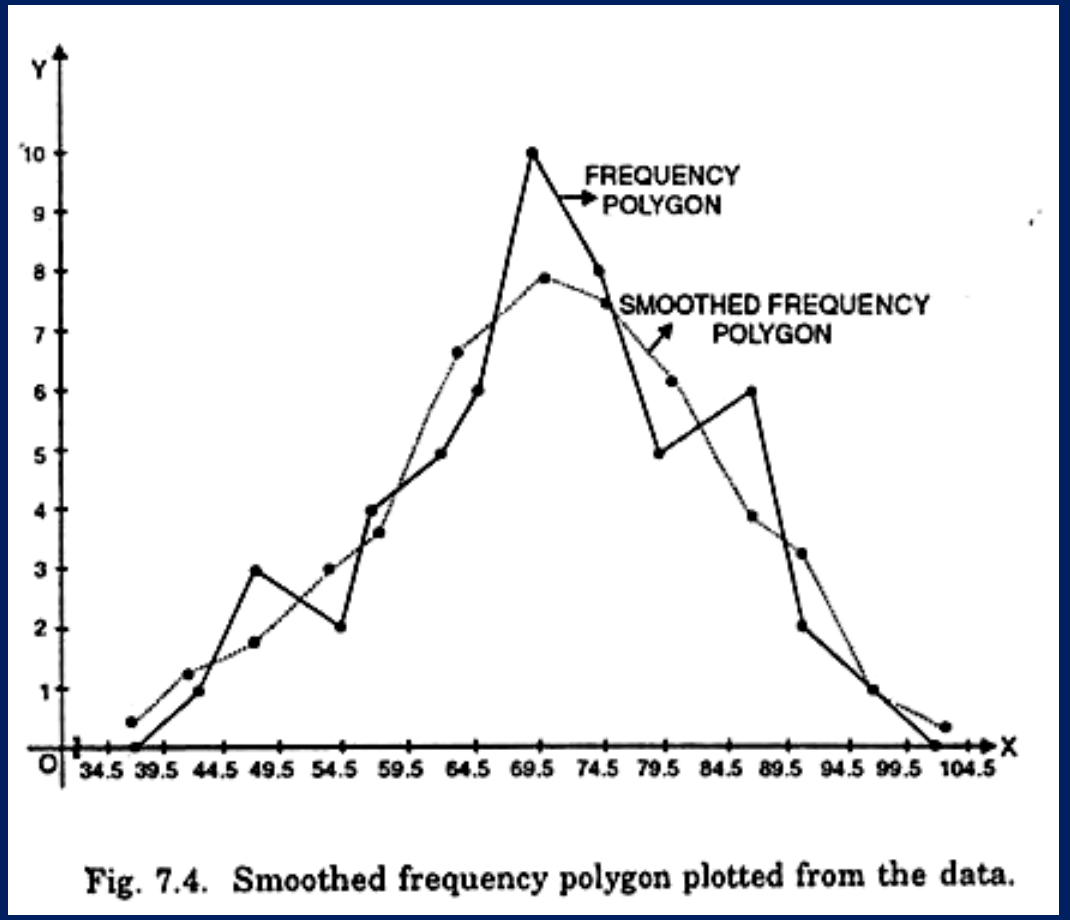


Fig. 7.4. Smoothed frequency polygon plotted from the data.

Measure of Central Tendency

Measures of Central Tendency

Measures of central tendency are numerical descriptive measures which indicate or locate the center of a distribution or data set.

In layman's term, a measure of central tendency is an AVERAGE. It is a single number of value which can be considered typical in a set of data as a whole.

For example, in a class of 40 students, the average height would be the typical height of the members of this class as a whole.

"Add the numbers, divide by how many numbers you've added and there you have it-the average amount of minutes you sleep in class each day."

Measure of Central Tendency

According to Prof Bowley "Measures of central tendency (averages) are statistical constants which enable us to comprehend in a single effort the significance of the whole."

The main objectives of Measure of Central Tendency are

1) To condense data in a single value.
2) To facilitate comparisons between data.

There are different types of averages, each has its own advantages and disadvantages.

Requisites of a Good Measure 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

Uses of Central Tendency:

The central tendency is needed for the following reasons:

Average provides the overall picture of the series. We cannot remember each and every facts relating to a field of enquiry.
2. Average value provides a clear picture about the field under study for guidance and necessary conclusion.
3. It gives a concise description of the performance of the group as a whole and it enables us to compare two or more groups in terms of typical performance.

Measure of Central Tendency

MEASURES OF CENTRAL TENDENCY

MEAN, MEDIAN MODE & PARTITION VALUES (CALCULATION & COMPUTATION)

Mean :

- **Mean :** Mean is the average value of all homogeneous distribution . In other word mean is the sum of all measurements divided by the number of observations in the data set.
- Types of mean: 1. Arithmetic mean
  2. Geometric mean
  3. Harmonic Mean

Calculation for Arithmetic mean:

A. For ungrouped Data:

Formula: $\sum X / n$

where X= variables ,
N = No. of total observations

Example: 2 5 7 8 9 9 12 15 35 54 68 69 78 91 95 95 96 98 98

Mean = $\sum X / n$ = 944/19 = 49.68

Measure of Central Tendency

MEAN
CONTINUED ...

B. For Grouped Data:

I) Long process

Formula: $\sum f X / n$

where X= variables ,

n = No. of total observations

Example:

Class Boundary	Mid value of class (X)	Frequency (f)	Frequency time of mid value (fx)
0 to 10	5	2	10
10 to 20	15	5	75
20 to 30	25	7	175
30 to 40	35	9	315
40 to 50	45	11	495
50 to 60	55	8	440
60 to 70	65	7	455
70 to 80	75	5	375
80 to 90	85	2	170
90 to 100	95	4	380
Total / ∑		60	2890

$$\text{Mean} = \sum f X / n = 2890/60 = 48.16$$

Measure of Central Tendency

Sum of all of the numbers of a group, when divided by the number of items in that list is known as the Arithmetic Mean or Mean of the group. For example, the mean of the numbers 5, 7, 9 is 4 since 5 + 7 + 9 = 21 and 21 divided by 3 [there are three numbers] is 7

Advantages and disadvantages of arithmetic mean

Advantages

It is rigidly defined.
It is easy to calculate and simple to follow.
It is based on all the observations.
It is determined for almost every kind of data.
It is finite and not indefinite.
It is readily put to algebraic treatment.
It is least affected by fluctuations of sampling.

Disadvantages

The arithmetic mean is highly affected by extreme values.
It cannot average the ratios and percentages properly.
It is not an appropriate average for highly skewed distributions.
It cannot be computed accurately if any item is missing.
The mean sometimes does not coincide with any of the observed value

Measure of Central Tendency

MEAN
CONTINUED ...

- B. For Grouped Data:
  II. (short Process/ coding Process/ Assume mean Process)

Formula:

$$\text{Mean} = M' + \frac{\sum f X'}{n} \times i$$

Where: M′= Mid value of the score,
f= frequency
X′= different of values from assume mean,
n = No. of total observations,
i = Class interval Example : as follows

Class Boundary	Frequency (f)	X'	Fx'
0 to 10	2	-4	-8
10 to 20	5	-3	-15
20 to 30	7	-2	-14
30 to 40	9	-1	-9
40 to 50	11	0	0
50 to 60	8	1	8
60 to 70	7	2	14
70 to 80	5	3	15
80 to 90	2	4	8
90 to 100	4	5	20
Total / ∑	60		19

: Mean = 45+(19/ 60) x 10
=48.16666

Measure of Central Tendency

Geometric Mean

In Mathematics, the Geometric Mean (GM) is the average value or mean which signifies the central tendency of the set of numbers by finding the product of their values. Basically, we multiply the numbers altogether and take out the nth root of the multiplied numbers, where n is the total number of values. For example: for a given set of two numbers such as 3 and 1, the geometric mean is equal to √(3+1) = √4 = 2.

In other words, the geometric mean is defined as the nth root of the product of n numbers. It is noted that the geometric mean is different from the arithmetic mean. Because, in arithmetic mean, we add the data values and then divide it by the total number of values. But in geometric mean, we multiply the given data values and then take the root with the radical index for the total number of data values. For example, if we have two data, take the square root, or if we have three data, then take the cube root, or else if we have four data values, then take the 4th root, and so on.

Geometric Mean Formula

The formula to calculate the geometric mean is given below:

The Geometric Mean (G.M) of a series containing n observations is the nth root of the product of the values

.Consider, if x1, x2 …. Xn are the observation, then the G.M is defined as:

$$G.M = \sqrt[n]{x_1, x_2, \ldots x_n}$$

or

$$G.M = (x_1, x_2, \ldots x_n)^{\frac{1}{n}}$$

Question 1 : Find the G.M of the values 10, 25, 5, and 30

Solution : Given 10, 25, 5, 30

We know that,

GM = $\sqrt[n]{\prod_{i=1}^{n} x_i}$

= $\sqrt[4]{10 \times 25 \times 5 \times 30}$

= $\sqrt[4]{37500}$

= 13.915

Therefore, the geometric mean = 13.915

Geometric Mean Properties

Some of the important properties of the G.M are:

The G.M for the given data set is always less than the arithmetic mean for the data set
If each object in the data set is substituted by the G.M, then the product of the objects remains unchanged.
The ratio of the corresponding observations of the G.M in two series is equal to the ratio of their geometric means
The products of the corresponding items of the G.M in two series are equal to the product of their geometric mean.

Application of Geometric Mean

The greatest assumption of the G.M is that data can be really interpreted as a scaling factor. Before that, we have to know when to use the G.M. The answer to this is, it should be only applied to positive values and often used for the set of numbers whose values are exponential in nature and whose values are meant to be multiplied together. This means that there will be no zero value and negative value which we cannot really apply. Geometric mean has a lot of advantages and it is used in many fields. Some of the applications are as follows:

It is used in stock indexes. Because many of the value line indexes which is used by financial departments use G.M.
It is used to calculate the annual return on the portfolio.
It is used in finance to find the average growth rates which are also referred to the compounded annual growth rate.
It is also used in studies like cell division and bacterial growth etc.

Measure of Central Tendency

Harmonic Mean

- Harmonic mean is quotient of "number of the given values" and "sum of the reciprocals of the given values".
- For Ungrouped Data

$$H.M \text{ of } X = \bar{X} = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

- For grouped Data

$$H.M \text{ of } X = \bar{X} = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)}$$

Harmonic Mean Example

Calculate the harmonic mean of the numbers: 13.2, 14.2, 14.8, 15.2 and 16.1

Solution:

The harmonic mean is calculated as below:

AS

$$H.M \text{ of } X = \bar{X} = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

$$H.M \text{ of } X = \bar{X} = \frac{5}{0.3417} = 14.63$$

X	$\frac{1}{X}$
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\sum\frac{1}{X}$ =0.3147

Measure of Central Tendency

Example: Calculate the harmonic mean for the given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
F	2	3	11	20	32	25	7

Solution: Now We'll find H.M as:

$$\bar{X} = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)} = \frac{100}{1.4368} = 69.60$$

Marks	x	f	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		4.5	0.4295

Measure of Central Tendency

Median: The point or the value which divides the data in to two equal parts., or when the data is arranged in numerical order .The data must be ranked (sorted in ascending order) first. The median is the number in the middle. Depending on the data size we define median as: It is the middle value when data size N is odd. It is the mean of the middle two values, when data size N is even

Ungrouped Frequency Distribution
Find the cumulative frequencies for the data. The value of the variable corresponding to which a cumulative frequency is greater than (N+1)/2 for the first time. (Where N is the total number of observations.)
Example 1: Find the median for the following frequency distribution

Merits of Median

1. It is rigidly defined.
2. It is easy to understand & easy to calculate.
3. It is not affected by extreme values.
4. Even if extreme values are not known median can be calculated.
5. It can be located just by inspection in many cases.
6. It can be located graphically.
7. It is not much affected by sampling fluctuations.
8. It can be calculated for data based on ordinal scale.

Demerits of Median

1. It is not based upon all values of the given data.
2. For larger data size the arrangement of data in the increasing order is difficult process.
3. It is not capable of further mathematical treatment.
4. It is insensitive to some changes in the data values.

Measure of Central Tendency

MEDIAN

- Median – the middle value that separates the higher half from the lower half of the data set. The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.

A. For Ungrouped Data:

I) In case of unpaired data:

1st step : Arrange the data ascending or descending order

2nd step : apply the formula & find the place

$$Xm = \frac{n+1}{2}$$

$$Xm = \frac{19+1}{2}$$

$=10^{th}$ place

2 5 7 8 9 9 12 15 35 54 68 69 78 91 95 95 96 98 98

Measure of Central Tendency

MEDIAN CONTINUED ...

II) In case of paired data:

1st step : Arrange the data ascending or descending order

2nd step : apply the formula & find the place

$$Xm = \frac{n+1}{2} \qquad Xm = \frac{20+1}{2}$$

=10.5 th place

2 5 7 8 9 9 12 15 35 54 68

69 78 91 95 95 96 98 98 99

So median=(54 +68)/2

=61

Measure of Central Tendency

MEDIAN CONTINUED ...

B. FOR GROUPED DATA:

$$\text{Median} = L + \frac{\frac{n}{2} - fb}{fm} \times i$$

L = Lower boundary of median class
N = Total frequency
fb = Cumulative frequency up to the lower boundary of median class.
fm = frequency of the median class
I = Class interval

Class Boundary	Frequency (f)	Cumulative frequency (cf)
0 to 10	2	2
10 to 20	5	7
20 to 30	7	14
30 to 40	9	23
40 to 50	11	34
50 to 60	8	42
60 to 70	7	49
70 to 80	5	54
80 to 90	2	56
90 to 100	4	60
	60	

$$\text{Median} = 40 + \frac{\frac{60}{2} - 23}{11} \times 10$$

Measure of Central Tendency

MODE

Mode is the maximum frequency of observation in a set of data.

A. FOR UNGROUPED DATA:

2, 5, 7, 8, 9, 9, 12, 15, 35, 54 68, 69,
78, 91, 95, 95, 95, 95, 95, 95, 96, 98, 98,

So mode = 95

B. FOR GROUPED DATA:

$$\text{Mode} = L + \frac{d1}{(d1 + d2)} x\ i$$

L = Lower boundary of modal class
D1 = Difference of frequency between modal & pre modal class.
D2 = Difference of frequency between modal & post modal class.
I = Class interval

Class Boundary	Frequency (f)
0 to 10	2
10 to 20	5
20 to 30	7
30 to 40	9
40 to 50	11
50 to 60	8
60 to 70	7
70 to 80	5
80 to 90	2
90 to 100	4

$$\text{Mode} = 40 + \frac{(11 - 9)}{((11 - 9) + (11 - 8))} x\ 10$$

Measure of Central Tendency

Mode

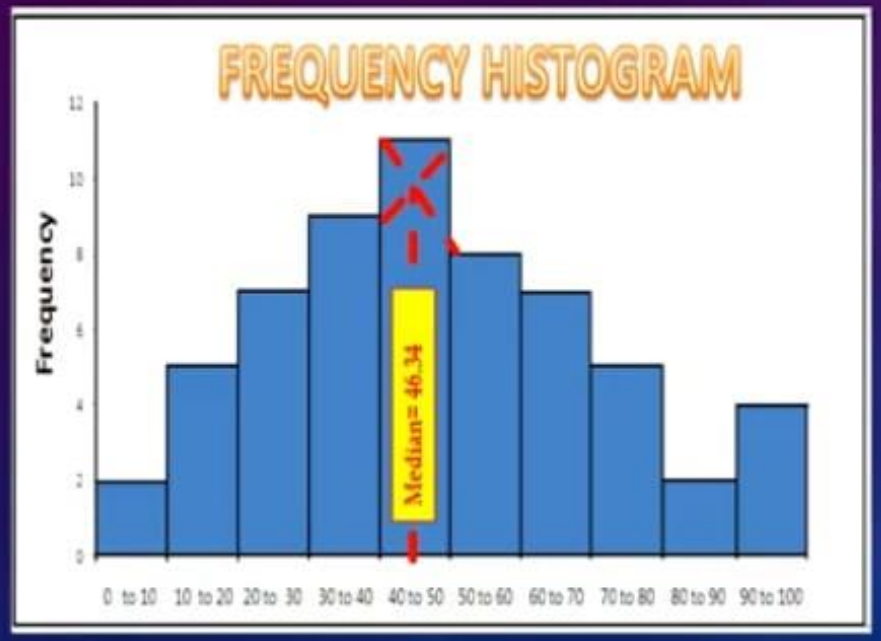
The mode is the number that appears most frequently in a data set

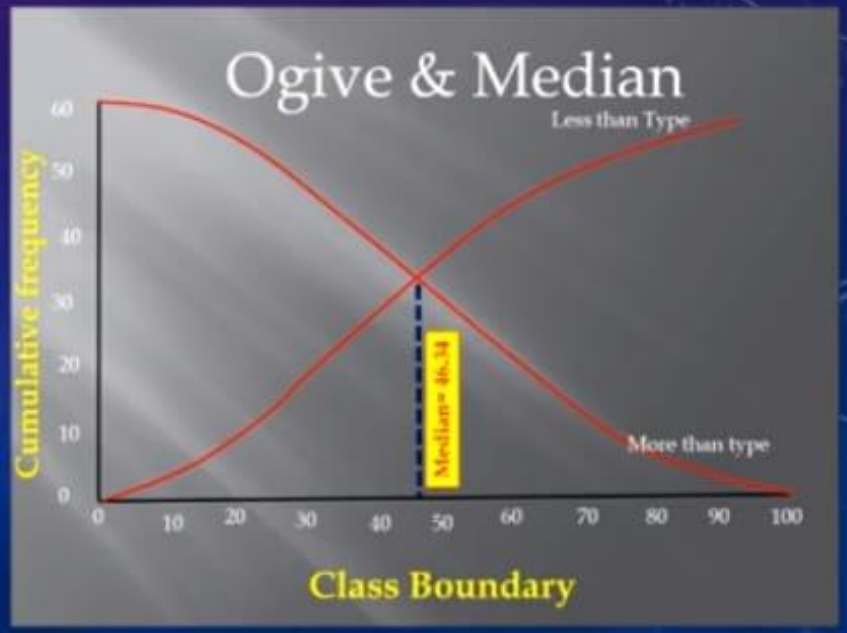
Advantages
Mode is readily comprehensible and easy to calculate. Like median, mode can be located in some cases merely by inspection.
Mode is not at all affected by extreme values.
Mode can be conveniently located even if the frequency distribution has class-intervals of unequal magnitude provided the modal class and the classes preceding and succeeding it are of the same magnitude. Open-end classes also do not pose any problem in the location of mote.
Disadvantages
Mode is ill-defined. It is not always possible to find a clearly defined mode. In some cases we may come across distribution are called bi-modal. If a distribution has more than two modes, it is said to be multimodal.
It is not based upon all the observations.
It is not capable of further mathematical treatment.
As compared with mean, mode is affected to a greater extent, by fluctuations of sampling.

Measure of Central Tendency

MEDIAN
CONTINUED ...

GRAPHICAL REPRESENTATION OF MEDIAN





PARTITION VALUES

If the values of the variate are arranged in ascending or descending order of magnitudes then we have seen above that median is that value of the variate which divides the total frequencies in two equal parts. Similarly the given series can be divided into four, ten and hundred equal pars. The values of the variate dividing into four equal parts are called Quartile, into ten equal parts are called Decile and into hundred equal parts are called Percentile

1 QUARTILES :
Definition. The values of the variate which divide the total frequency into four equal parts, are called quartiles. That value of the variate which divides the total frequency into two equal parts is called median. The lower quartile or first quartile denoted by Q1 divides the frequency between the lowest value and the median into two equal parts and similarly the upper quartile (or third quartile) denoted by Q3 divides the frequency between the median and the greatest value into two equal parts. The formulas for computation of quartiles are given b

Definition,. The values of the variate which divide the total frequency into ten equal parts are called deciles. The formulas for computation are given by

PERCENTILES :
Definition. The values of the variate which divide the total frequency into hundred equal parts, are called percentiles. The formulas for computation are :

PARTITION VALUES

Quartiles(Q), Decile(D) and Percentile(P) are the partition values of any distribution.
Quartiles can be calculated up to three. They are Q_1, Q_2 & Q_3
Deciles can be calculated up to nine. They are D_1, D_2 ..... D_9
Percentile can be calculated up to 99. They are P_1, P_2 ..... P_{99}

Formula:

$$\text{Quartiles} = L + \frac{N/4 - f}{F} \times i$$

Formula:

$$\text{Deciles} = L + \frac{N/10 - f}{F} \times i$$

Formula:

$$\text{Percentiles} = L + \frac{N/100 - f}{F} \times i$$

L = Lower Boundary of Quartile/Decile/Percentile classes
f= Cumulative frequency upto the Quartile/Decile/Percentile classes
F= Absolute frequency of that class
N= No. of total frequency
i= Class interval

PARTITION VALUES CONTINUED ...

- **For Quartiles:**
- $Q_1 = L + \frac{\frac{1XN}{4} - f}{F} X\ i$

So, $Q_1 = 30 + \frac{\frac{60}{4} - 14}{9} X\ 10$

So, $Q_1 = 31.11$

In the same way Calculate

$Q_2 = 46.36364$, $Q_3 = 64.28$

- **For Deciles:**
- $D_1 = L + \frac{\frac{1XN}{10} - f}{F} X\ i$

So, $D_1 = 10 + \frac{\frac{1X60}{10} - 2}{5} X\ 10$

So, $D_1 = 18$

- In the same way Calculate D_2 , D_3, and up to D_9

Class Boundary	Frequency (f)	Cumulative frequency (cf)
0 to 10	2	2
10 to 20	5	7
20 to 30	7	14
30 to 40	9	23
40 to 50	11	34
50 to 60	8	42
60 to 70	7	49
70 to 80	5	54
80 to 90	2	56
90 to 100	4	60
	60	

For Percentile:

$P_{12} = L + \frac{\frac{12XN}{100} - f}{F} X\ i$

So, $P_{12} = 20 + \frac{\frac{12X60}{100} - 7}{7} X\ 10$

So, $P_{12} = 20.86$

In the same way Calculate P_{25} , P_{33}, P_{50} and P_{75}, P_{85} and more.

Remember that value of Median=Q_2=D_5= P_{50}

GRAPHICAL REPRESENTATION OF PARTITION VALUES