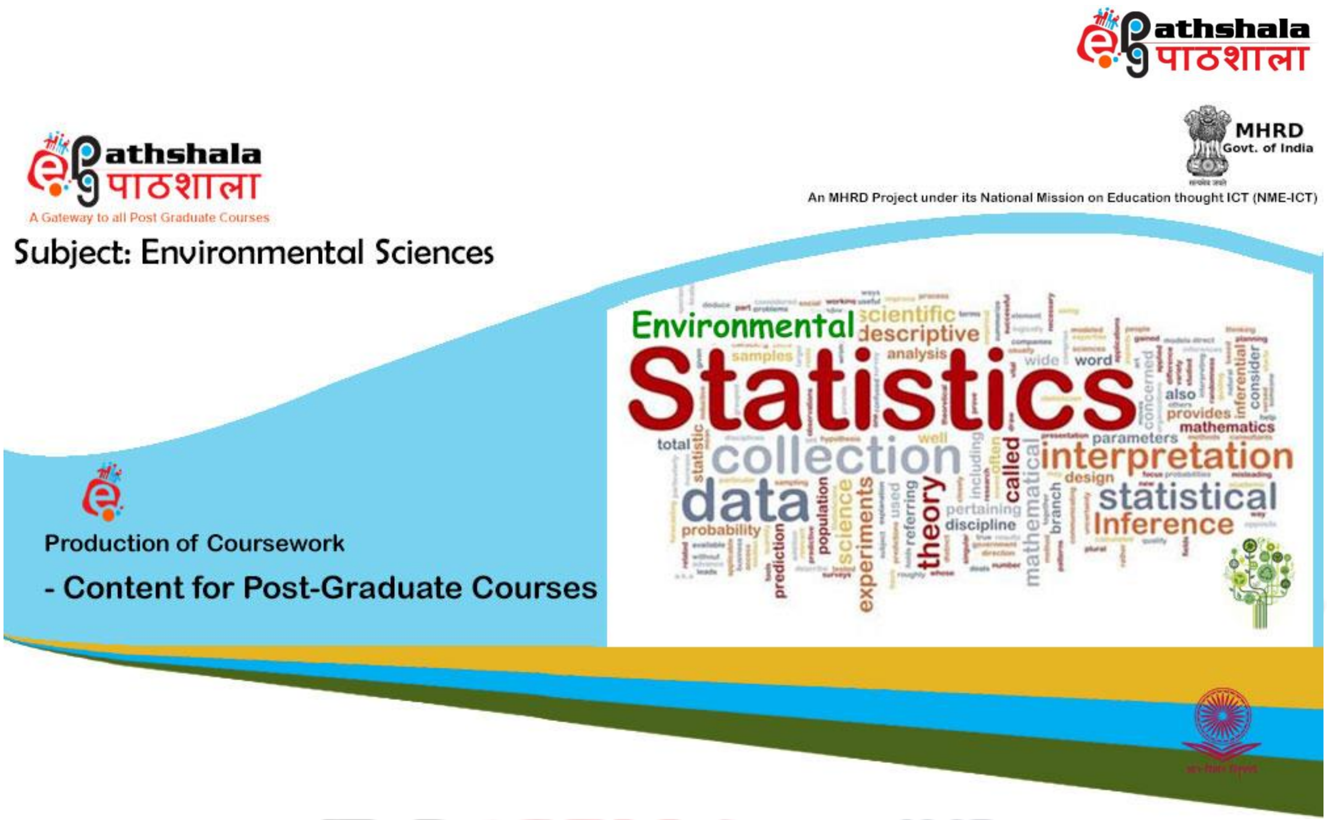

Paper No: 14 Statistical Applications in Environmental Sciences

Module: 7 Central Tendency Measures-II

e-Pathshala पाठशाला — A Gateway to all Post Graduate Courses — MHRD Govt. of India

Development Team

Principal Investigator & Co- Principal Investigator	**Prof. R.K. Kohli** **Prof. V.K. Garg &Prof.AshokDhawan** **Central University of Punjab, Bathinda**
Paper Coordinator	**Dr. Harmanpreet Singh Kapoor,** **Central University of Punjab, Bathinda**
Content Writer	**Dr. Harmanpreet Singh Kapoor** **Central University of Punjab, Bathinda**
Content Reviewer	**Prof. Kanchan Jain,** **Panjab University, Chandigarh**
Anchor Institute	**Central University of Punjab**

Description of Module	
Subject Name	**Environmental Sciences**
Paper Name	Statistical Applications in Environmental Sciences
Module Name/Title	**Central Tendency Measures II**
Module Id	EVS/SAES-XIV/7
Pre-requisites	Module 1-6
Objectives	Basic introduction to different positional averages like Median and Mode
Keywords	Median, mode, positional averages, decile, percentile

Module 7: Central Tendency Measures- II

- **Learning Objectives**
- **Introduction.**
- **Positional Averages**
- **Relationship between Mean, Median and Mode**
- **Summary**
- **Suggested Readings**

1. Learning Objectives

In this module, a complete explanation about different types of positional averages of data will be discussed. This module helps us to learn different methods of central tendency measures and it properties. Through this module, one can learn about which method to use under what type of conditions. Questions with answers are included to give an in-depth knowledge of the topic.

2. Introduction

As we already discussed in the module "Central Tendency –I" that data contain those elements that take different values for the same objective. In this scenario, the measure of central tendency is used to give you an idea about the data. Central tendency is a measure that provides a single value that represents a group of values. A.M., G.M. and H.M. are discussed as a mathematical average methods. In this module, we will discuss another branch of central tendency measure that is positional averages. Positional averages are measures that provide values that depends on the requirement of the observers.

For example, one is interested in the mid value of the data for that median is used, other is interested in the value that repeat more in the data set for that mode is calculated.

We will discuss all these measure and other measures that lie in this categories. Some of them are quartiles, deciles and percentile.

In the next section, we will start with the definition of Median and we will see with examples how to calculate it for different data set. After that we will discuss other positional average measure with their definitions and formula. Merits and demerits and relationship are also discussed between them for better understanding.

3. Positional Averages

Median

Median is the most commonly used word in our practical life. Generally, one think of median as the value that split the item into two equal parts. In statistics, we also want to find out that value that lies in the middle of the data. To find out the median value, first arrange the observations in ascending or descending order then take the middle value of the data that value represents the median value. For example, if one has 11 observations in the data set then to find out the median. One has to arrange the observations in ascending and descending order then take 6^{th} value in the arranged series as median value. Now, if one has 12 observations in the data sets then the method is different from the previous method. As one can observe that the median lies between 6th and 7th value in the arranged data so arithmetic average of these 6th and 7th observations are taken to find out the median value. The formula and methods to find out the median are given in steps. Basically the calculation of median involve two things

(a) To locate the position of middle value
(b) To calculate its value

Based on the data type, median is evaluated as follows:

(a) **Simple data**:

(i) First arrange the given observations $x_1, x_2, \ldots, x_n$ in ascending or descending order.

(ii) If n is an odd value i.e. not divisible by 2 then the middle value is evaluated as $\left(\frac{n+1}{2}\right)$th value in the data.

(iii) If n is an even quantity .i.e. divisible by 2 then the median value is the arithmetic of $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th value.

Ques 1: Find out the median of the following dataset:

N	1	2	3	4	5	6	7	8	9
X	25	36	21	56	43	53	63	49	71

Table 1

Ans. Ordered series is 21,25,36,43,**49**,53,56,63,71.
n is odd here so the median value is 49. One can easily see that 49 lie in the middle of the data set.

Ques 2: Find out the median of the following dataset:

N	1	2	3	4	5	6	7	8	9	10
X	25	36	21	56	43	53	63	49	71	67

Table 2

Ans Here we add a new item in the data. n is an even value for finding out the median value again arrange the observations in the ascending order.
21, 25, 36, 43, **49, 53,** 56, 63, 67, 71
Here the median value lies between the 5th and 6th value in the ordered data set. After taking Arithmetic Mean (A.M.) of these two values the median value is calculated as 51.

(b) **Ungrouped frequency data:**

If the data observations are available in the form of ungrouped frequency. The steps to find out the median are:

(i) Arrange the observations in increasing or decreasing orders

(ii) Compute cumulative frequency of the observations

(iii) Then use the formula $\frac{N+1}{2}$ to calculate the middle value, where N is the total frequency.

(iv) After this, locate which cumulative frequency value is equal to $\frac{(N+1)}{2}$ or just higher to it and corresponding value of observation is the median value.

Ques 3: Find out the median of the following observations:

Observations	Frequency
45	5
76	6
56	7
89	8
93	9

Table 3

Ans First order the observations in ascending order as

Observations	Frequency	C.F.
45	5	5
56	7	12
76	6	18
89	8	26
93	9	35
	N = 35	

Table 4

Next, compute the cumulative frequency (C.F.) as shown in the third column of Table 4. As N is 35 so using the formula $\frac{N+1}{2}$ i.e. 36/2 is 18. In the third column, 76 observation has cumulative frequency equal to 18. So 76 is the median value of the dataset.

Let us suppose the N is more than 35 i.e. 37 then using the formula $\frac{N+1}{2}$ the value is 19 then the median value will be 89. Hence one can understand which value to choose for the median depending on the total frequency N.

(c) Grouped frequency data:

If the observations are given in the grouped frequency table then one has to follow these steps to find out the median value. These are:

(i) First compute the value of $\frac{N}{2}$.

(ii) Now use this formula

$$\text{Median} = l_1 + \frac{\left(\frac{N}{2} - F\right)}{f} \times c$$

where,

l_1 is the lower class interval;

N is the total frequency;

f is the frequency value;

c is the width of the class interval;

F is the cumulative frequency preceding the class containing the value of $\frac{N}{2}$.

One should refer to the module "Diagrammatic and Graphical Representation of the Data-I" to recall the terms used above.

Ques 4: Find out the median value of the following data set:

Observations	Frequency
0-10	7
10-20	5
20-30	4
30-40	6
40-50	8
50-60	10

Table 5

Ans First find out the cumulative frequency of the observations as shown in the third column in Table 6. Now $\frac{N}{2}$ is 20 which lies in the fourth row.

Observations	Frequency	C.F.
0-10	7	7
20-30	5	12
20-30	4	16
30-40	6	22

40-50	8	30
50-60	10	40
	N = 40	

Table 6

Here l_1 is 30, f is 10, F is 16 and c is 10.

Now use the formula given above

$$\text{Median} = 30 + \frac{20 - 16}{6}\ x\ 10 = 36.667$$

Hence, one can evaluate the median for the group frequency data.

Now, will discuss the merits and demerits of the median.

Merits and Demerits of the median

Every measure has some merits and some demerits. One should keep this in mind before applying the method on the data.

Merits: The following are some of the important merits of median.

(i) Median is also a measure which is rigidly defined.
(ii) One can understand this method easily. Even for a non-mathematical background person can understand this with little efforts.
(iii) As median is considered as positional average of central tendency so extreme observations have no impact on it. Also one can find the median value for open end class and unequal classes.
(iv) Median can be calculated if extreme values are missing in the data set.
(v) The major advantage of this measure is that it can be located graphically and sometime even by inspection.
(vi) It is the considered as the best measure of central tendency in case of qualitative data. Although the qualitative data have no quantity but still can be arranged in increasing and decreasing order for example qualitative observations in termed of ranks can be ordered.

Demerits

(i) As comparative to other measure, median can be calculated if the observations are ranked either in increasing or decreasing order which is not required in other measures of central tendency.
(ii) Median does not take into account the magnitude of the observations. It only divide the whole dataset into two equal parts.
(iii) As Median is a positional average so it is not treated for further algebraically purpose.

(iv) Median like most of the measures of central tendency may or may not be representative value of the series.

(v) It is not suitable where one has to assign weight to the observations then median will not be used.

Quantiles

As we discussed about median which is a positional average that divide the data into two equal parts. There are some other measures that are also used to divide or partition the dataset into fixed number of parts. All these measures of such types are called **Quantiles.** Some of them are Quartiles, Percentile, and Deciles.

Quartiles

Quartiles are those positional averages which divide the total number of observations into four equal parts. As we know that three points are required to divide the data into four equal parts. In quartiles, these three parts are called (a) First Quartile (b) Second Quartile (c) Third Quartile.

Percentile

Percentile are those positional averages that divide the data into 100 equal parts. Similarly as quartiles they are termed as first, second,…, ninty nine percentile.

Deciles

Deciles are those positional averages that divide the data into 10 equal parts and they are also termed as first, second,…,ninth deciles.

Now in the next section, we will discuss about the methods to calculate different positional averages in the given data set of different types.

Methods to calculate positional averages

As median is also a positional average. There is a lot of similarities between the methods to calculate median and quantiles. One can observe this similarity from the formula given below:

(a) **Simple Series**: Let $x_1, x_2, \dots, x_n$ be the simple series and n denotes the number of observations then similarly for evaluating median, first order the observations in increasing and decreasing series. After that the different type of quantiles can be found as follows:

Quartile

First Quartile is $\frac{n+1}{4}$, Second Quartile is $\frac{2(n+1)}{4}$ and Third Quartile is $\frac{3(n+1)}{4}$.

As second quartile formula is same as median so we can say that second quartile is also known as median.

Similarly for **Percentile** and **Decile**, the methods are

k^{th} Decile is $\frac{k(n+1)}{10}$, where k=1,2,…,9 and

m^{th} Percentile is $\frac{m(n+1)}{100}$, where m=1,2,..,99.

Hence one can evaluate quantiles for simple data using the above methods and interpret them accordingly.

(b) Ungrouped Frequency Distribution: As the observations are available with frequency and the total frequency is termed as N.

Quartile

i^{th} Quartile is $i(N+1)/4$, where i=1,2,3.

Similarly for **Percentile** and **Decile**, the methods are

k^{th} Decile is $\frac{k(N+1)}{10}$, where k=1,2,…,9

and

m^{th} Percentile is $\frac{m(N+1)}{100}$, where m=1,2,..,99.

Hence one can evaluate the quantiles for ungrouped frequency data.

(c) Grouped Frequency Distribution: For evaluating quantiles for grouped frequency distribution, one must follow these steps.

(i) Divide total frequency i.e. N/q where q can be 4, 10 and 100 for quartile, decile and percentile.

(ii) Similarly as we did for median, locate the class interval whose cumulative frequency is equal to this ratio or just greater than this value.

(iii) Select that class interval and use the formula given as

$$i\text{th Quartile} = l_1 + \frac{\frac{iN}{4} - F}{f} \text{ x c} \qquad (1)$$

where $i = 1,2,3$.

$$k\text{th decile} = l_1 + \frac{\frac{kN}{10} - F}{f} \text{ x c} \qquad (2)$$

where $k = 1,2,\dots,9$

$$m\text{th percentile} = l_1 + \frac{\frac{mN}{100} - F}{f} \text{ x c} \qquad (3)$$

where $m = 1,2,\dots,99$

l_1 is the lower class interval;

N is the total frequency;

f is the frequency value;

c is the width of the class interval;

F is the cumulative frequency preceding the class containing the value of $\frac{N}{q}, q = 4{,}10$ and 100.

Ques 5: Calculate the first quartile, ninth decile and 50th percentile from the following data set of score obtained by 10 candidates out of 200 in state government examination.

78, 97, 112, 134, 152, 145,165,107,143, 132

Ans Arrange the data in ascending order

S.No.	Observations	S.No.	Observations
1	78	6	134
2	97	7	143
3	107	8	145
4	112	9	152
5	132	10	165

Table 7

Now apply the formula of to find out the first quartile i.e. $\frac{10+1}{4} = 2.75$ th term. It is evaluated as second term +0.75 (value of 3rd term- value of 2nd term) which is equal to 97+0.75 (107-97) =104.5.

Ninth decile value will be computed in the same manner and the value is 9.9th term. Its value is 9th term+0.9 (vaule of 10th observation-value of 9th observations) which is equal to 152+0.9 (165-152)=163.7.

50th percentile is evaluated as 5.5 which is equal to median of the data. Hence it is the average of 5th and 6th term. So 50th percentile value will be (132+134)/2 i.e 133.

Similarly, one can evaluate the value for other quantiles.

Ques 6: Calculate the first quartile, ninth decile and 50th percentile from the following data set of score obtained by 100 students (out of 200) in state government examination.

Marks	Frequency
113	13
166	12
124	18
146	16
137	14
143	9

156	18

Table 8

Ans First order the observations in ascending order and compute cumulative frequency

Marks	Frequency	C.F.
113	13	13
124	18	31
137	14	45
143	9	54
146	16	70
156	18	88
166	12	100
	N = 100	

Table 9

Cumulative frequency is shown in third column in Table 9.

For first quartile the formula is $\frac{N+1}{4}$i.e. $\frac{100+1}{4} = 25.25$ th term. It is evaluated from the cumulative frequency that contain the term or just than the term. So the firs quartile value is 124.

Ninth decile value will be computed in the same manner and it is 90.9th term. It can be seen from the Table 9 that 166 value's frequency hold the frequency of ninth decile. So the ninth decile is 166. Similarly 50th percentile is evaluated as 50.5 which is equal to median of the data. Hence by using the formula the value of the median is 143.

Hence one can compute the positional average of different types for ungrouped frequency distribution.

Ques 7: Calculate the first quartile, ninth decile and 50th percentile from the following data set of score obtained by 500 students (out of 200) in state government examination.

Class Marks	Frequency
80-100	30
100-120	170
120-140	150
140-160	110
160-180	40

Table 10

Ans Calculate the cumulative frequency first.

Class Marks	Frequency	C.F.
80-100	30	30
100-120	170	200
120-140	150	350
140-160	110	460
160-180	40	500
	N = 500	

Table 11

For first quartile, first compute $\frac{N}{4}$ i.e. 125th item and see where this value lies in the third column of Table 11. Here the item lies in the class 100-120 then by using the formula given in equation (1)

$$\text{First quartile} = 100 + \frac{125-30}{170} \text{ x } 20 = 111.176.$$

Similarly, for ninth decile locate the value of $\frac{9N}{10}$ i.e. 450th term in the third column of Table 11 and use equation (2) to solve this

$$\text{Ninth decile} = 140 + \frac{450-350}{110} \text{ x } 20 = 141.819.$$

Also, for 50th percentile first evaluate the ratio i.e. $\frac{N}{2}$ i.e. 250th term and locate the class which contain this term from Table 11 and use the equation (3) as

$$\text{50th percentile} = 120 + \frac{250-200}{150} \text{ x } 20 = 120.67.$$

Hence the values of first quartile is 111.176, ninth decile is 141.819 and 50th percentile is 120.67.

Mode

The origin of the word Mode is from a French letter la mode and its meaning is the most famous thing or thing that prevails as a trend in the society. In statistics, it is considered as a term that has highest frequency value in a series. In other words, mode is a value that occur with maximum frequency in a dataset.

"The mode of a distribution is the value at the point around which the items tend to the most heavily concerntrated. It may be regarded as the most typical of a series of values" by Cowden and Croxton.

Thus, it is clear from the definition that the mode is a value that has the greatest engrossment of values. One thing should keep in mind while evaluating in mind that the value that occurs a lot must be mode because it may be possible that the engrossment may be over some other value and also there may be one or two points where engrossment of values occur. Hence there is no single mode in a dataset. When a data has single mode it is called unimodal when two mode then it is called bi-modal and more than two then multimodal.

Mode has a great importance in the practical life. As every firm is interested to know the purchasing pattern of the customer which brand and item customers use to buy more and at what price and quantity. For example, mobile company want to know the price at which customer use to buy handsets. The size of shoe, color and design that has maximum demand in the market. Hence, the importance of mode can be understood that it is measure that can be applied for quantitative data and qualitative data.

Method to calculate mode

(a) **Simple series:** It is very easy to determine the value of mode from a simple series. It can be determined by counting the value that has the maximum repetition in the data set. Hence the value that has maximum repetition is the modal value.

Ques 8: Calculate the mode from the following dataset.

23,34,34,56,55,43,34,35,34,43,34

Ans Count the frequency of each observation here 34 occurs the maximum time i.e. 5 and 43 occurs 2 times rest all the observations appear only one time in the series. Hence mode is 34.

(b) **Ungrouped frequency distribution:** Calculation of mode for ungrouped distribution is similar as simple series. The only difference is that here the observations are already given in the frequencies and there is no requirement to count them. So it is very easy to find mode in this case.

Ques 9: Calculate the mode from the following dataset

Observation	Frequency
34	10
43	15
56	16
76	6
64	11

Table 12

Ans From the second column in Table 12, one can see that frequency of 56 is the maximum among all the observations. Hence 56 is the mode.

(c) **Grouped frequency distribution:** In grouped frequency data, it is difficult to trace the mode value just by looking at the frequency of the observations. One must follows these steps to calculate mode. These are:

(i) First locate the interval that has highest frequency.

(ii) Apply the formula

$$Mode = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \text{ x } c \quad (4)$$

where l_1 is the lower class interval that contain the largest frequency;

f_1 is the highest frequency in the dataset;

f_0 is the preceding frequency of largest frequency i.e. f_1;

f_2 is the succeeding frequency of the largest frequency i.e. f_1;

c is the width of the class interval.

One important thing, while applying this method is that all the class intervals must be uniformly distributed otherwise this method will not lead to correct result. Also this method is not useful in case of multimodal data.

Ques 10: Calculate the mode of the following dataset.

Observation	Frequency
0-10	23
10-20	34
20-30	45
30-40	21

Table 13

Ans From Table 13, one can see that the highest frequency is 45 and the corresponding class interval is 20-30. Now apply the formula given in equation (4)

$$\text{Mode} = 20 + \frac{45 - 34}{90 - 34 - 21} \text{ x } 10 = 23.142.$$

Hence the 23.142 is the mode of the dataset.

Next we will discuss about the merits and demerits of mode.

Merits and demerits

We will discuss about the merits of mode first and then demerits.

Merits

(i) Mode is the representative value of the observation and the major benefit of this method is that mode can obtained just by looking at the dataset in simple and ungrouped dataset.

(ii) Second major benefit of using mode lies in the simplicity that is it is very simple to calculate then other measures of central tendency.
(iii) Mode has no effect of extreme observations.
(iv) Mode can be evaluated for open ended distributions.
(v) Mode can be calculated in case of missing observations.
(vi) Mode can be used for qualitative dataset.

Demerits

(i) The major drawback of this measure is that it is not rigidly defined due to the application different formulae to compute mode and all of them lead to different values.
(ii) Mode value can only be used for further decision making if it is evaluated from a large dataset. For example, mobile companies and shoe companies must require large data set from a particular region to make their policies. One cannot rely on just few observations and change their policies.
(iii) Another important thing about this measure that mode always does not exist. Unlike other measure of central tendency it is not always possible to obtain mode from the dataset.
(iv) Mode value cannot be used for further mathematical treatment.
(v) Mode cannot be determined in grouped frequency distribution like some other measures of central tendency.

4. Relationship between Mean, Median and Mode

Now, we will discuss a little about the relationship between mean, median and mode. There is a relationship between mean, median and mode depending upon the symmetry of the data. If data are symmetric then mean=median=mode otherwise if data are not symmetric i.e. asymmetric then either

$$\text{mean} > \text{median} > \text{mode}$$

or

$$\text{mode} > \text{median} > \text{mean}.$$

Hence, in literature another important exist between three measures. It is called an empirical relation between three measure.

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median}).$$

5. Summary

In this module, measures of positional averages are discussed that is a branch of central tendency of measures. These are median, mode and quantiles. Although median is a part of quantiles but it is studied differently from quantiles due to its importance in literature. In this module, we discussed about different positional averages and their methods and which method to use under which situations. We also discussed how to evaluate them for different types of data like simple, frequency data and group frequency data. Merits and demerits of all three mean and relationship between them are also discussed for better understanding.

6. Suggested Readings

Agresti, A. and B. Finlay, Statistical Methods for the Social Science, 3rd Edition, Prentice Hall, 1997.

Daniel, W. W. and C. L. Cross, C. L., Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition, John Wiley & Sons, 2013.

Hogg, R. V., J. Mckean and A. Craig, Introduction to Mathematical Statistics, Macmillan Pub. Co. Inc., 1978.

Meyer, P. L., Introductory Probability and Statistical Applications, Oxford & IBH Pub, 1975.

Triola, M. F., Elementary Statistics, 13th Edition, Pearson, 2017.

Weiss, N. A., Introductory Statistics, 10th Edition, Pearson, 2017.