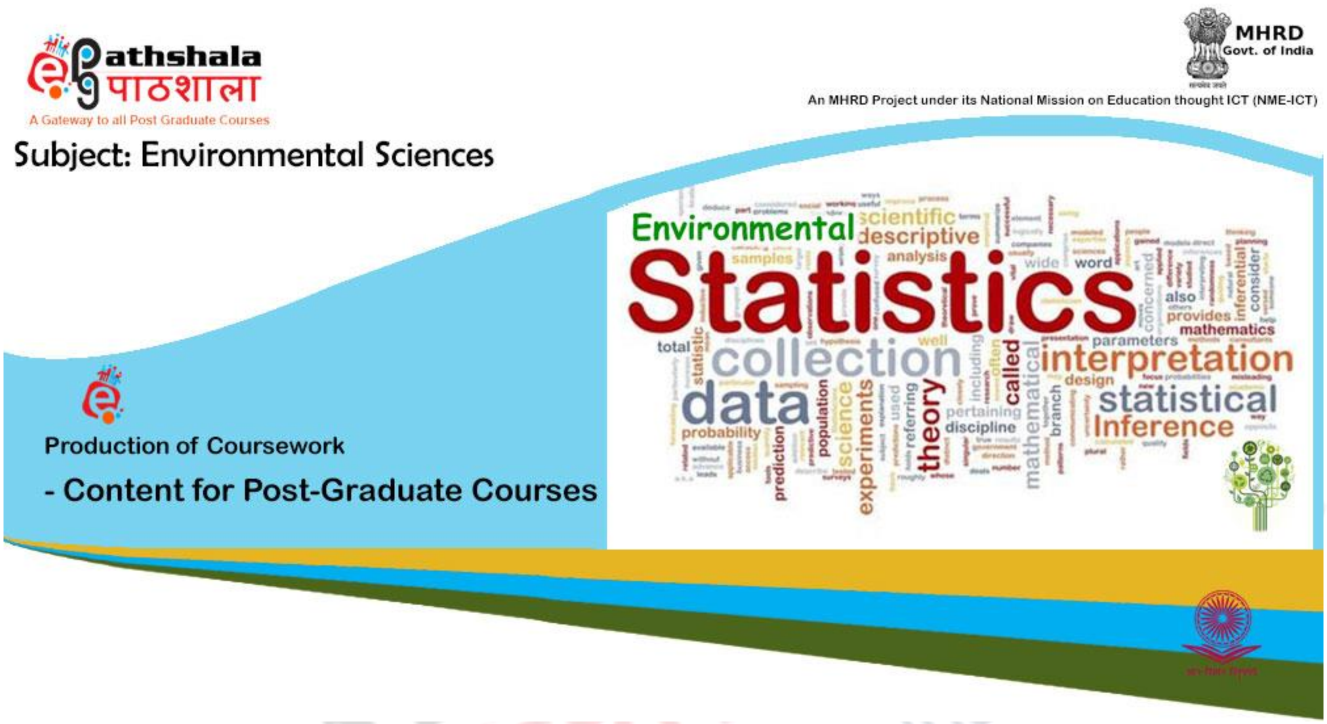

A Gateway to all Post Graduate Courses
MHRD
Govt. of India
An MHRD Project under its National Mission on Education thought ICT (NME-ICT)
Subject: Environmental Sciences
Production of Coursework
- Content for Post-Graduate Courses
Environmental
Statistics
collection
data
interpretation
statistical
Inference

Paper No: 14 Statistical Applications in Environmental Sciences

Module: 6 Central Tendency Measures-I

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 I**
Module Id	EVS/SAES-XIV/6
Pre-requisites	Module 1-5
Objectives	Basic introduction to mathematical averages like A.M., G.M. and H.M.
Keywords	Arithmetic Mean, Geometric Mean and Harmonic Mean

Module 6: Central Tendency Measures- I

- **Learning Objectives**
- **Introduction.**
- **Mathematical Averages**
- **Summary**
- **Suggested Readings**

1. Learning Objectives

In this module, a complete explanation about different types of measures of central tendency of any data will be discussed. This module will help to understand different methods of central tendency measures and it properties. Through this module, one can learn about which method is to be used under what type of conditions. The topic of central tendency measures is covered in two modules. This module will cover the mathematical averages measure of the data. Other topic of positional average will be covered in the module "Central Tendency Measure- II". Questions with answers are included to give an in-depth knowledge of the topic.

2. Introduction

In real life, we collect data from population that has same characteristics for a particular objective. The data that are collected contain elements may have different information. Now how we can say anything about the nature of the data. In this scenario, we use a measure that provides an idea about the data. This measure is called 'central tendency measure'. Central tendency is a measure that provides a single value that represents a group of values. However, it should satisfy some certain conditions.

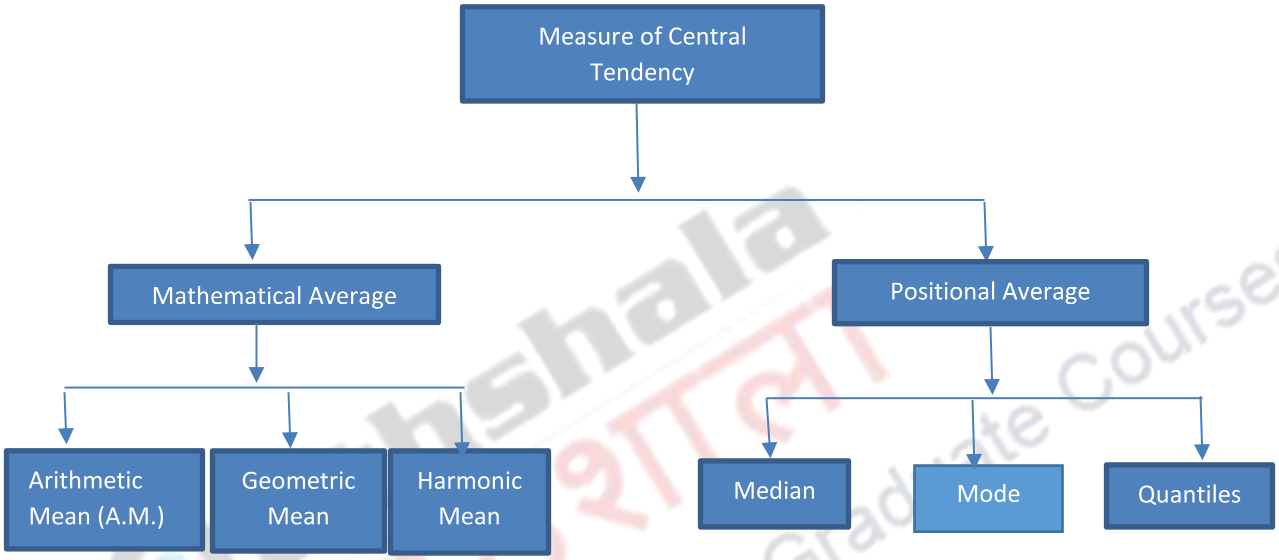
There are some properties that are expected from a central tendency measure:

(i) It should be defined in a rigid manner such that its meaning should be unambiguous.
(ii) It should be calculated on the entire observations in the data.
(iii) It should be calculated in a little time frame and in an easy manner.
(iv) It should possess mathematical properties so that we can further use it.
(v) It should not be influenced by the extreme values of the data.
(vi)It must be sensitive to the small changes in the data values.

Although while applying the measure on the data, one should keep in mind that the data should be collected from homogeneous group. If the data is not from the homogenous group, then the value that we get may lead to incorrect decision. For example, while calculating the average height of the students in a class one should first categorize the gender at first place after that calculate its mean value on category basis. If students are not categorized according to the gender then the average height of students lead to unrepresentative form. Hence one can solve this type of problem by categorizing heterogeneous data into homogenous one.

In Statistics, three measures are considered generally that is mean, median and mode. Mean measure is further divided into three categories (i) Arithmetic Mean (ii) Geometric Mean and (iii) Harmonic Mean.

The relationship between different measures are shown below:



From the above measures, the arithmetic mean, median and mode are widely used. On the other hand, geometric and harmonic mean are used in particular situations.
In the following section, one can see the complete description of these measures in detail.

3. Mathematical Averages

3.1 Arithmetic Mean (A.M.)

Arithmetic Mean is among the most widely used measure of central tendency for a single representation of observations. The formula for the evaluation of the arithmetic mean (A.M) is given as

$$\text{Arithmetic Mean (A. M.)}=\frac{\text{sum of all obervations}}{\text{number of observations}}$$

Let $x_1, x_2, \ldots, x_n$ be n observation is a data set. The Arithmetic mean of the data is denoted as $\bar{x}$.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

For example, let 23, 24, 25, 35, 32 be 5 class size of different sections of 9th class in a private school. We want to know about the average class size in 9th class. To calculate this we first calculate the sum of all class sizes and divide it by no of sections. The arithmetic mean is 27.8≈ 28. It means that there is approximately 28 students on average basis in each class.

This method cannot be applied if the no. of the elements in the data set are very large. For the convenience purpose frequency tabulation method is used. One can recall the tabulation methods from the previous modules.

Let $x_1, x_2, \ldots, x_n$ have frequencies $f_1, f_2, \ldots, f_n$ respectively that means that in the complete data set x_1 appears f_1 times and x_2 appears f_2 times and so on.

Hence the formula for A.M. in this case will be

$$\bar{x} = \frac{\{f_1x_1 + f_2x_2 + \cdots + f_nx_n\}}{f_1 + f_2 + \cdots + f_n}$$

This method of calculating Arithmetic Mean in term of frequencies is known as Weighted Arithmetic Mean. As we are multiplying the x_i's with their corresponding frequencies (f_i) .

An example consists of no. of child in 50 families in a locality is presented below will give you a better understanding of the weighted arithmetic mean.

No. of child	0	1	2	3 or above
Families	10	15	20	5

Table 1

Now to calculate the average number of child in a family. One can use weighted arithmetic mean and the value is $1.4 \approx 1$. Hence on average there is a single child in each family.

Also in real life, we have large amount of data so as we know from the module "Diagrammatic and Graphical Representation of the Data" that in such type of cases, one can construct frequency table. For example, we are interested in the marks of the students spending hours on daily basis to learn through online program. A artificial data was available of 1000 students and information is given in the following table as:

Time Spent (in hours)	No. of students
0-4	690
4-8	285
8-12	20
12-16	5

Table 2

Now if one is interested to find the average no of hours spend on average basis using A.M. then the formula is

$$\bar{x} = \frac{\{f_1x_1 + f_2x_2 + \cdots + f_nx_n\}}{f_1 + f_2 + \cdots + f_n}$$

but here $x_1, x_2, \dots, x_n$ are the midpoint of class interval. So the arithmetic mean of the data can be found as

$$= \frac{2*690+6*285+10*20+14*5}{690+285+20+5} = \frac{3360}{1000} = 3.360$$

Hence from the above example, we can conclude that 3.36 hours or 3 hours and 6 minutes on average are spent by the students on daily basis.

Now we will discuss about the merits and demerits of Arithmetic Mean (A.M.).

As arithmetic mean is widely used in literature due to the following merits

Merits:

(a) A. M. can be easily understood by a person and does not require any particular need to explain it.

(b) The calculation of the A.M. is very simple and does not require any calculations.

(c) It is derived without ordering the data like other measure of central tendency.

(d) It is considered as comparative very stable and good method of central tendency in comparison to other measures.

(e) If one has large amount of data then it is considered as a good method of central tendency as the unusual observations in one direction can be offset by the unusual observations in other direction.

Demerits:

There are also certain shortcoming of any measure and A. M. is not an exception. There are some important demerits of A.M given below:

(a) A.M. cannot be obtained through guesses like other measure mode etc. whether it is not considered as a major demerit but still it is. Also median can be located graphically but A.M. cannot.

(b) A.M. cannot be found for the data when one has missing observation in the data. Also for a small set of data it is highly sensitive the extreme values. In such cases, A.M. will lead to a wrong conclusions.

(c) A.M. cannot be evaluated for the qualitative characteristics that cannot be measured, like sex, level of illness etc. In such cases it is suitable to use other measure of central tendency like median.

(d) It may be possible that a value calculated through A.M. is difficult to interpret. For example, one can recall the value of A.M. evaluated from Table 1 which is 1.4 that has no interpretation as child is considered in unit not in fraction. So the 1.4 child is meaningless.

(e) It is possible that it may lead to false conclusions.

At the last one can say that A.M. is considered as a simplest measure of central tendency but it will mislead you when data are of heterogeneous form i.e. non-homogeneous.

3.2 Geometric Mean

Geometric Mean is another method to find out the mean of n observations. It is evaluated through the following method.

Let us suppose $x_1, x_2, \dots, x_n$ are n observations, then

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

If $x_1, x_2, \dots, x_n$ have frequencies $f_1, f_2, \dots, f_n$ respectively, for evaluating the G.M. the method is

$$G.M. = \left(x_1^{f_1}, x_2^{f_2}, x_3^{f_3} \dots, x_n^{f_n}\right)^{\frac{1}{N}}$$

where, $N = \sum f_i$ is the sum of all observation’s frequency and is called the total frequency.

Hence, one can use the above method to find the G.M. of the observations. It is difficult to find out the G.M. by using this formula or one can use the above formula when the no of observations is very small. In this case, one can use log on both side of the formula. Then the modified form the above formula are given below respectively.

$$log G.M. = \frac{1}{n}\sum \log(x_i) \qquad (1)$$

$$\log G.M. = \frac{1}{N}\sum f_i \log(x_i) \qquad (2)$$

Now, if one is using the formula of equation (1) and (2), then one has to take antilog at the end to find out the G.M. value.

$$G.M. = Antilog\left(\frac{1}{n}\sum \log(x_i)\right) \qquad (3)$$

$$G.M. = Antilog\left(\frac{1}{N}\sum f_i \log(x_i)\right) \qquad (4)$$

Steps for the calculation of the G.M. for simple data using equation (3)

(i) Calculate log (base 10) of the observations

(ii) Take the sum of all log of observations i.e. $\sum \log(x_i)$.

(iii) To find the log G.M., divide $\sum \log(x_i)$ by total no. of observations.

(iv) Take antilog (base 10) of $\frac{1}{n}\sum \log(x_i)$ and this is G.M. of the required data.

Ques 1: Find out the G.M. of the following data given as 34,57,37,43,52,67,54,56.

Ans In Table 3, the log values are shown in second column.

Observations	$log(x_i)$
34	1.5314789
57	1.7558749

37	1.5682017
43	1.6334685
52	1.7160033
67	1.8260748
54	1.7323938
56	1.748188
No. of obs =8	13.511684
log G.M.	1.6889605
G.M.	48.86079

Table 3

From the above data, the required value is calculated as 48.86079.

Now to find the G.M. of the data when the data set has frequencies with them. There are further two cases

(a) Discrete case (Ungrouped Frequency Series)
(b) Continuous case or Class Interval (Grouped Frequency Series)

In the first case, observations are given with their corresponding frequencies. In the second case, as data is given in the class interval. In these two cases the G.M. is calculated in a different manner.

(i) Discrete Case:

G.M. for the discrete case is considered first. One has to follow some steps given below to calculate the G.M.

Steps are:

(i) Calculate log (base 10) of the observations.
(ii) Multiply the log values with the corresponding frequencies values and take their sum i.e $\sum f_i \log(x_i)$.
(iii) Divide the term i.e $\sum f_i \log(x_i)$ by N that is the total frequency.
(iv) Take anitlog (base 10) of $\frac{1}{N}\sum f_i \log(x_i)$. This will give you the value of G.M for frequency data.

Ques 2: Find out the G.M. of the following data given as

Observations	frequency
38	7
43	9
46	10

49	6
51	4
54	8
67	3
78	5

Table 4

Ans

Observations	Frequency	$log\ (x_i)$	freq*$log(x_i)$
38	7	1.5797836	11.05848518
43	9	1.6334685	14.7012161
46	10	1.6627578	16.62757832
49	6	1.6901961	10.14117648
51	4	1.7075702	6.830280704
54	8	1.7323938	13.85915008
67	3	1.8260748	5.478224408
78	5	1.8920946	9.460473013
	N=52		88.15658428
	log (G.M.)		1.695318928
	G.M.		49.58141632

Table 5

In the Table 5, data values are given in the first column and their corresponding frequencies in the second column. Logarithm of observations are calculated in the third column. Fourth column represents the values of product of logarithm values corresponding to their frequencies. 88.15658428 represents $\sum f_i \log(x_i)$ and 1.695318928 is calculated by dividing 88.15658428 by N i.e. 52 and 49.5841632 is the anitlog of 1.695318928. Hence the G.M. is calculated as 49.5841632.

(ii) Continuous case or Class Interval:

If the data is of continuous form then one has to first construct frequency table using class intervals. Then only one can find out the G.M. for the continuous data. There are few steps for calculate G.M. in these cases. These steps are:

(i) First calculate the mid points (m_i) of the class intervals.

(ii) Calculate log (base 10) of the mid values of each class interval i.e. $log(m_i)$.

(iii) Multiply the log values with the corresponding frequencies values and take their sum i.e $\sum f_i \log(m_i)$.

(iv) Divide the term i.e $\sum f_i \log(m_i)$ by N that is the total frequency.

(v) Take anitlog (base 10) of $\frac{1}{N}\sum f_i \log(m_i)$. This will give you the value of G.M for frequency data.

Ques 3: Find the G.M. of the following data.

Class Interval	Frequency
0.0-5.0	4
5.0-10.0	6
10.0-15.0	8
15.0-20.0	5
20.0-25.0	7
25.0-30.0	3

Table 6

Ans

Class Interval	mid value (m_i)	Frequency	$log(m_i)$	$log(m_i)$*freq
0.0-5.0	2.5	4	0.397940009	1.591760035
5.0-10.0	7.5	6	0.875061263	5.25036758
10.0-15.0	12.5	8	1.096910013	8.775280104
15.0-20.0	17.5	5	1.243038049	6.215190243
20.0-25.0	22.5	7	1.352182518	9.465277627
25.0-30.0	27.5	3	1.439332694	4.317998081
		N = 33		35.61587367
			Log G.M.	1.079268899
			G.M.	12.00242219

Table 7

In the Table 7, first column represents the class intervals and their corresponding frequencies are given in the third column. Mid values are calculated in the second column and logarithm of mid values are calculated in the fourth column. Fifth column represents the values of product of logarithm values corresponding to their frequencies. 35.61587367 represents $\sum f_i \log(m_i)$ and 1.079268899 is calculated by dividing 35.61587367 by N i.e. 33 and 12.00242219 is the anitlog of 1.079268899. Hence the G.M. is calculated as 12.00242219.

Hence, one can use the above steps for calculating the G.M. in continuous case.

As we are using the G.M. as a measure of central tendency one must know where to use this measure to get more appropriate information. Hence it is essential to discuss about the merits and demerits of this measure.

Merits of G.M.

(a) It is defined in a rigid manner.
(b) It is based on all the values of the data and cannot be calculated for data with missing values.
(c) G.M. can be calculated in those cases where only the total product of observations and number of observations are known without having knowledge of individual values.
(d) G.M. can be used for further mathematical treatment like rations and percentages. Therefore it can be used for further analysis like index number etc.
(e) The main advantage of G.M. is that it is not affected much by extreme large and small values and it gives more weight to small values in comparative to A.M. which gives more weight to large values.

Demerits

There are few demerits of G.M.. These are

(a) Due to the complexity involved in the calculation of G.M. . It is difficult for a non-mathematical person to understand it.
(b) The calculation of G.M. requires the knowledge of logarithm that is difficult for a non-mathematical background person.
(c) G.M. cannot be calculated for the data which contain a value zero.
(d) G.M. is not useful for the data that have different signs that is some of the observations are positive and other are negative. In such case, G.M. value has no meaning.
(e) G.M. value may not be the actual value of the observations like other measure of central tendency.
(f) G.M. is not advisable to use in those cases where small observations must be given more weights and higher observation must be given less weights.
Another measure of central tendency is Harmonic Mean (H.M.). This measure of central tendency has a great importance in science and mathematical field.

3.3 Harmonic Mean

Harmonic mean is calculated by first taking the reciprocal of the observations and then taking the reciprocal of arithmetic of these reciprocal observations.

Let us suppose $x_1, x_2, \dots, x_n$ are n observations in the dataset. Then

$$H.M.=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}\dots.+\frac{1}{x_n}}=\frac{n}{\sum\frac{1}{x_i}}.$$

When observations are given with frequencies i.e. $x_1, x_2, \dots, x_n$ with $f_1, f_2, \dots, f_n$ are the corresponding frequency values then the H.M. is evaluated by using the following formula.

$$H.M.=\frac{f_1+f_2+\cdots+f_n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}\dots.+\frac{f_n}{x_n}}=\frac{N}{\sum f_i/\,x_i}$$

where, $N=\sum f_i$ is the sum of all observation's frequency and is called the total frequency.

Harmonic mean is basically used in the case where rates and ratio are involved like per hour, per month, per litre, per min etc. It is used for to find out the average of different types of rates. These rates are basically used to explain the relationship between two units of opposite nature like as you increase the speed of a vehicle then this shortens the travelling time. Hence H.M. is used to find out the average speed in kilometer per hour for a given distance.

In the following example, H.M. is evaluated to find out the average speed of 6 cars given in kilometer per hour.

Ques 4: Calculate the Harmonic Mean (H. M.) for the series 56, 45, 67, 58, 53, 42.

Ans

KMPH	Reciprocal of observations
56	0.017857143
45	0.022222222
67	0.014925373
58	0.017241379
53	0.018867925
42	0.023809524
N=6	0.114923566
H.M.	52.20861322

Table 8

In Table 8, the reciprocal of observation are given in the second column. 0.114923566 represents the total of reciprocal of the observations and 52.2086 is the H.M. of the data.

Similarly, one can keep the following steps in mind to find out the H.M. for observations given with frequencies. These are

For ungrouped frequency data. Steps are

(i) Find out the reciprocal of the observations.
(ii) Multiply the reciprocal observations with the corresponding frequencies values and take their sum i.e $\sum f_i/x_i$.
(iii) Divide N by the term i.e $\sum f_i/x_i$ where N is the total frequency.
(iv) This will give you the value of H.M. for discrete frequency data.

For grouped frequency data. Steps are

(i) First calculate the mid points (m_i) of the class intervals.
(ii) Multiply the reciprocal observations with the corresponding frequencies values and take their sum i.e $\sum f_i/m_i$.
(iii) Divide N by the term i.e $\sum f_i/m_i$ where N is the total frequency.
(iv) This will give you the value of H.M. for grouped frequency data.

Hence, one can evaluate the H.M. for three different situations i.e. simple, discrete frequency and grouped frequency.

Now, we will discuss the merits and demerits of the H.M.

Merits

(a) H.M. is defined in a rigid manner.
(b) H.M. is evaluated on all the observations in the data.
(c) H.M. gives less weight to large values and more weight to small values so this measure is appropriate in those case where such condition is required.
(d) H.M. is used to measure relative changes and is widely used for the averaging of ratio and rates.

Demerits

(a) H.M. is very difficult measure to be understood by a non-mathematical background person.
(b) H.M. is comparatively difficult to calculate than other measure.
(c) H.M. cannot be evaluated if the value of the observation is zero.
(d) H.M. cannot be evaluated when some of the observations are positive and some are negative in the data set.
(e) H.M. value does not exist in the data set.

Relationship between A.M., G.M. and H.M.

A.M. is always greater than or equal to G.M. and G.M. is greater than or equal to H.M. for a given observations. In mathematical term

$$A.M. \geq G.M. \geq H.M.$$

These measures are only equal when all the observations are same in the data.

4. Summary

In this module, three measures of mathematical averages are discussed that is a branch of central tendency of measures. It is generally termed as “mean” in the literature. In this module, we discussed about which measure of mean is more suitable to which situations. We also discussed how to evaluate different types of means like A.M., G.M. and H.M. 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.

5. 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.