



Subject: Environmental Sciences



Production of Coursework
- Content for Post-Graduate Courses



Paper No: 14 Statistical Applications in Environmental Sciences
Module: 8 Measures of Dispersion I


Development Team

Principal Investigator & Co- Principal Investigator	Prof. R.K. Kohli Prof. V.K. Garg & Prof. Ashok Dhawan 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	Measures of Dispersion I
Module Id	EVS/SAES-XIV/8
Pre-requisites	Module 1-7
Objectives	To give an introduction to different measures of dispersion
Keywords	Range, quartile deviation, mean deviation, standard deviation



A Gateway to All Post Graduate Courses

Module 8: Measures of Dispersion -I

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

1. Learning Objectives

In this module, a complete introduction to measures of dispersion and different measures of dispersion are discussed with examples. Properties of different measures are also discussed with their merits and demerits. Through this module, one can easily understand about which method to use under what type of conditions. The topic of measures of dispersion is covered in two modules. This module will cover the absolute measures of dispersions. Other topic of relative measures will be covered in the module “Measures of Dispersion- II with Skewness and Kurtosis”. Questions with answers are included to give an in-depth knowledge of the topic.

2. Introduction

In this module, an important measure that is used to see how each observation varies from the mean value. When one is interested to know about the variation of observations from the mean value. The measures that are used to detect this variations in the observation are called Measures of Dispersions.

In the module “Central Tendency Measures –I” and “Central Tendency Measures –II” we discussed about different measures of central tendency. We discussed about the mathematical averages and positional averages. The basic purpose of these measures is to find out a single value that represent the whole dataset. Also concentration of observations about the central part of the data were observed. But these measure will not take into account of the fact that whether two or more different dataset have same mean values but it does not mean that the observations are same.

For example, we have two dataset given below that have same mean value but the observations are not same.

Dataset I: 8, 9, 10, 11, 13, 15

Dataset II: 3, 4, 6, 14, 16, 23

Hence the total sum of both datasets are same i.e. 66 and the mean is also same i.e. 11. Now the question is how we can say that mean is the representation of the data? The answer is central tendency measure gives you just an idea about the concentration of the observation around a central value but it does not say anything about the variation of these observations from the central part. In dataset I, the values are very close to each other and also to the mean value so we can say that 11 is the correct representation of the dataset but in dataset II, observations are very scattered and also very far away from the mean value. Hence, one can see that 11 as a mean value is not represent the whole data set in a correct manner.

Hence, one should not rely just on the central tendency measure to take any opinion about the observations but also one should also think about the dispersion or variation among the observation. In this module, different type of measure will be discussed that consider variation of observation from different values like difference between highest and smallest values, absolute difference of observations from a particular quantity etc. but mainly the variation of observations from its mean is considered the most in the literature.

Definitions

Some authors have defined the measures of dispersion as:

“Dispersion is the measure of variation of the items” by A.L. Bowley.

“Dispersion or spread is the degree of the scatter or the variation of the variable about a central value” by B.C. Brooks and W. F. L. Dicks.

In literature, the dispersion is also considered as synonym for heterogeneous in the data. As heterogeneous is basically used to understand the extent of variations among observations. Dispersion can only be zero if all the observations have same values. The dispersion is more when the difference between the observations is very large. So one can say that if the variation is small like in data set I then it is considered as insignificant but if the variation is large as shown from the data set II then it is considered as significant.

Also dispersion is termed as second ordered means as central tendency measures is the first ordered means where one can see the tendency of the values around the middle of the data. As measures of dispersion are basically used to discuss about the variation, scatterness of the observations from the central tendency measure. The measures of central tendency and measures of dispersion both together discuss about the characteristics of the data set but they do not able to demonstrate that, to what extent the observations deviate from the central value i.e. whether equal number of observations are dispersed from the central tendency or whether the data is symmetrical about the mean or not. It also give us an idea of how many observations have their value below the mean value or above the mean value. If one is interested to know about the concentration of the observations around a central tendency measure then it is essential to study two more measures. These are Skewness and Kurtosis. These two measure are considered as a supportive measures for better understanding the characteristics of the data. Skewness and Kurtosis is discussed in the module “Measure of Dispersion –II with Skewness and Kurtosis”.

Importance of measures of dispersions

- (a) The main motive behind using these measures to check the authenticity of the central tendency measures. It is used to see whether the value of central tendency measure are reliable or not.
- (b) The second important thing about the measures of central tendency is that these are also used to compare the two datasets through relative values.
- (c) This measure is also useful in identifying the reasons behind the variations in order to control them but these are not helpful in give the exact reason behind variations in the observations. For

example, in electronic industry quality of an item cannot be judged only by through whether an item is produced under defined limits but also through the variations between the characteristics of observations.

- (d) Measures of dispersions are also used to help for further analysis of the data like correlation, regression, testing of hypothesis and ANOVA etc

As there are many measures of dispersion, the ideal measure prevails the following characteristics:

- (a) The values of these measures should be rigid.
- (b) It should be calculated on all the observations.
- (c) The method should be easily calculated and understood by non-mathematical background person.
- (d) The measure should be used for further algebraic treatment.
- (e) The measure should not be affected by extreme values and fluctuation of the observations.

The measures of dispersion are further categorized into two types. These are shown in the following flowchart

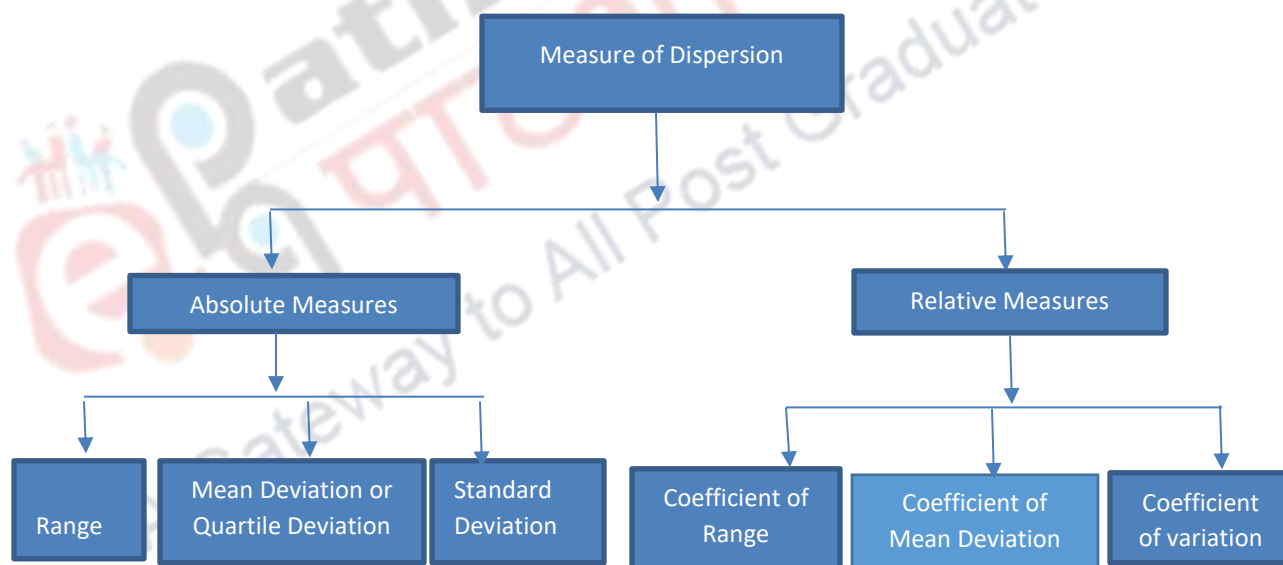


Figure 1

From the Figure 1, the measure of dispersion is categorized into two measures: Absolute Measures and Relative Measures. As dispersion measure is used to detect the deviation of the observations from the central tendency. If the measures of dispersion express the dispersion of the observations in the original units then the measures are called absolute measures of dispersion. One can only compare the variations between two series if both series are in same units otherwise comparison is not meaningful. To

overcome this problem, those measures that give the dispersion values in terms of ratio and percentage are called relative measures.

A relative measure of dispersion is the ratio of absolute measure of dispersion with its appropriate average. These measure are independent of units and these are termed as coefficient of dispersion. One important thing, while the calculation of relative measures, is that the units of absolute measure and the appropriate average must be same.

In this module, absolute measures will be discussed with examples. Relative measures will be discussed in the module “Measures of Dispersion II with Skewness and Kurtosis”.

Range

Range is considered as the simplest absolute measure of dispersion. It is evaluated by just taking the difference between the maximum value and minimum value in the data set.

Range = Maximum value- Minimum value

Ques 1: Calculate the range of the following dataset:

23, 45, 56, 52, 64, 35, 42, 51, 76, 65.

Ans: Maximum value is 76 and minimum value is 23. Hence

Range = $76 - 23 = 53$.

Merits of Range

These are the merits of the range measure

- (i) Range is the simplest among all the absolute measures of dispersion.
- (ii) Range is defined in a rigid manner and easily calculated. Due to this property it is widely used in the industry as a quality control tool.
- (iii) Range is computed within second hence one can get a complete picture of variability in the dataset in a short time.

Demerits of Range

There are few demerits of range also. These are:

- (i) As range is evaluated just from the difference of maximum and minimum value of the dataset. Hence it does not depend on all the observations. So it is possible that range will be same for two dataset that have same maximum and minimum value but different intermediate values that make no sense.
- (ii) As we already discuss that range depend on just two values and both of these values are extreme values. Hence range is affected a lot by extreme values.
- (iii) Also range will not take into account the shape of the distribution. It may be possible that range will be same for the dataset whether it is symmetric or skew symmetric.

- (iv) Range is affected by the change in the observations.
- (v) Range cannot be evaluated from the grouped frequency distribution with open end class.

Quartile Deviation

This measure of dispersion is related with range and it removes the drawbacks of range upto some extent. It is defined as the difference between third quartile and first quartile divided by 2. In other words,

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2}$$

where Q_1 is the first quartile;

Q_3 is the third quartile.

Also $Q_3 - Q_1$ is know as the interquartile range.

Ques 2. Find out the quartile deviation from the following data set related to the marks obtained by 15 students in statistics.

60, 67, 56, 78, 92, 55, 72, 54, 49, 59, 37, 84, 83, 69, 62

Ans Arrange the observations in the ascending order

37, 49, 54, 55, 56, 59, 60, 62, 67, 69, 72, 78, 83, 84, 92

Now compute the first quartile $\frac{N+1}{4}$ th term i.e. 55

Similarly third quartile as 12th term i.e. 78.

Quartile deviation is $(78 - 53)/2 = 25/2 = 12.5$.

Merits of Quartile Deviation

There are some merits of quartile deviation. These are

- (i) Quartile deviation is easy to understand like range.
- (ii) As quartile deviation is based on first and third quarter hence it removes the limitations of the range.
- (iii) As quartile deviation consider only first quarter and third quarter value. These values are derived by neglecting first and last 25% of the observations respectively. Hence it is not affected by the extreme values.
- (iv) Quartile deviation is more useful to consider in case of skewed symmetrical data.
- (v) Quartile deviation can be evaluated for open end class interval as it is not depend on the extreme values.

Demerits of Quartile deviation

There are some demerits of quartile deviation. These are

- (i) Quartile deviation is evaluated on just half of the observations as it neglect first and last 25% of the observations. Hence it is not considered as a fully appropriate measure for detecting variation among observations.
- (ii) Quartile measure is based on quartiles which are positional averages. It represents the deviation of observations among quartiles that is just a distance on a scale.
- (iii) Quartile deviation is very much influenced by the change in the observations. A change in single observations lead to change in the value of quartile deviation.
- (iv) Quartile deviation just give an idea about the deviation like range. So it is not the best measure for measuring dispersion.

Mean deviation

Mean deviation is computed as the absolute value of difference of observations from a central tendency i.e. mean, median and mode. Mostly, mean deviation is calculated by taking deviation of observations from mean and median.

Let x_1, x_2, \dots, x_n be the n observations to compute mean deviation. First assume a number A and take absolute deviation of each observations from the value A i.e. $|x_1 - A|, |x_2 - A|, \dots, |x_n - A|$. As some of the observations are more than A and some of them are less so modulus is taken to accumulate the total deviation of observation from a point A .

Now the sum of deviated observations from A is $\sum |x_i - A|$ and the arithmetic mean of these observation will give mean deviation about A .

$$\text{Mean deviation about } A = \frac{1}{n} \sum |x_i - A|$$

If A is chosen as mean value then the deviation is called mean deviation about mean and if A is chosen as median then the deviation is called mean deviation about median.

The above formula is used for simple series data. For ungrouped frequency distribution, the formula will be

$$\text{Mean deviation about } A = \frac{1}{N} \sum f_i |x_i - A|$$

For grouped frequency distribution, the formula will be

$$\text{Mean deviation about } A = \frac{1}{N} \sum f_i |m_i - A|$$

where N is the total frequency;

f_i is the frequency corresponding to the i th observation;

m_i be the midpoint of the i th class interval.

Ques 3. Calculate the mean deviation about mean for the following observations.

30, 40, 50, 55, 60, 65, 70, 80, 90, 100

Ans First find out the A.M. of the observations. It is 64. Now take absolute deviation of each observation from 64. The values are 34, 24, 14, 9, 4, 1, 6, 16, 26. Take the sum of these observation and divide it by 10 i.e. number of observations.

The answer is $134/10=13.4$.

Ques 4. Calculate the mean deviation about mean for the following dataset.

Observations	frequency
38	7
43	9
46	10
49	6
51	4
54	8
67	3
78	5

Table 1

Ans. First compute A.M. of the data using the formula $\frac{\sum f_i x_i}{N}$, where $N = \sum f_i$. The values are shown in the Table 2. After that subtract this mean value from each observation and take modulus. Now multiply this absolute value with the corresponding frequency. Now compute the A.M of these absolute observation. Hence 8.22 is the mean deviation about mean.

Observations	frequency	$f_i x_i$	deviation from mean $x_i - \bar{x}$	absolute deviation about mean $ x_i - \bar{x} $	$f_i * x_i - \bar{x} $
38	7	266	-12.653846	12.65384615	88.57692308
43	9	387	-7.6538462	7.653846154	68.88461538
46	10	460	-4.6538462	4.653846154	46.53846154
49	6	294	-1.6538462	1.653846154	9.923076923
51	4	204	0.34615385	0.346153846	1.384615385
54	8	432	3.34615385	3.346153846	26.76923077
67	3	201	16.3461538	16.34615385	49.03846154
78	5	390	27.3461538	27.34615385	136.7307692
	N = 52	$\sum f_i x_i = 2634$			427.8461538

	mean	$\frac{\sum f_i x_i}{N} = 50.65384615$		$\frac{\sum f_i x_i - \bar{x} }{N} = 8.227810651$
--	------	--	--	--

Table 2

Ques 5 Calculate the mean deviation about mean for the following dataset.

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 3

Ans: First compute A.M. of the data using the formula $\frac{\sum f_i m_i}{N}$, where $N = \sum f_i$ and m_i is the mid point of the i th class interval. The values are shown in the Table 4. After that subtract this mean value from midpoints and take modulus. Now multiply this absolute value with the corresponding frequency. Now compute the A.M. of these absolute observation. Hence the mean deviation about mean for grouped frequency distribution dataset is 6.556.

Class Interval	mid value (m_i)	Frequency f_i	$f_i m_i$	deviation from mean $m_i - \bar{x}$	absolute deviation about mean $ m_i - \bar{x} $	$f_i * m_i - \bar{x} $
0.0-5.0	2.5	4	10	-12.12121212	12.12121212	48.48484848
5.0-10.0	7.5	6	45	-7.12121212	7.12121212	42.72727273
10.0-15.0	12.5	8	100	-2.12121212	2.12121212	16.96969697
15.0-20.0	17.5	5	87.5	2.878787879	2.87878787	14.39393939
20.0-25.0	22.5	7	157.5	7.878787879	7.87878787	55.15151515
25.0-30.0	27.5	3	82.5	12.87878788	12.87878788	38.63636364
		$N = 33$	$\sum f_i m_i = 482.5$			216.3636364
			$\bar{x} = \frac{\sum f_i m_i}{N} = 14.6212$			$\frac{\sum f_i m_i - \bar{x} }{N} = 6.556$

Table 4

Merits and Demerits of Mean deviation

Merits

There are some merits of mean deviation measure. These are

- Mean deviation is defined in a rigid manner.

- (ii) Mean deviation is an easy method for a non-mathematical background person.
- (iii) Mean deviation is based on all the observations so it is better than range and quartile deviation measures.
- (iv) Mean deviation is less influenced by the extreme values than range, standard deviation.
- (v) Mean deviation is appropriate for comparison purpose because it is derived from the deviation of observations from central values i.e. mean, median or mode.
- (vi) Mean deviation is basically an average of absolute deviation from central value. Hence it remove biasness to some extent that occur due to observations while calculating and provide an accurate value for dispersion.

Demerits

There are few demerits of mean deviation measure. These are

- (i) The major drawback of mean deviation measure is that while computing it we ignore the signs of deviation and take absolute of it. So one cannot get any information about the concentration of the observations above the mean or below the mean. So this measure is not useful for further mathematical treatment.
- (ii) Mean deviation cannot be computed for open end classes distributions.
- (iii) Mean deviation is not an accurate measure of dispersion for highly skewed data.

Although mean deviation has some limitations. Due to its simplicity and easy to understand features it has great importance in accountancy, economics, forecasting and business sectors.

Standard Deviation

Karl Pearson in 1823 first introduced this measure and after that it is the most widely used measure of dispersion till date. Standard deviation is the measure that prevails all the features that other measures lack of. So basically it is considered as an ideal measure of deviation. It is also know as square root of variance, root mean square deviation etc and denoted by the Greek letter σ (sigma). Standard deviation value is high or low when there is more or less variation among observations respectively.

Let x_1, x_2, \dots, x_n be the n observations and \bar{x} be the average value, then the standard deviation (S.D.) is evaluated as

$$S. D. = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

For frequency data, the formula is

$$S. D. = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{N}}$$

For grouped frequency data, the formula is

$$S. D. = \sqrt{\frac{\sum f_i (m_i - \bar{x})^2}{N}}$$

where N is the total frequency;

f_i is the frequency corresponding to the i th observation;

m_i be the midpoint of the i th class interval.

Ques 6: Calculate standard deviation of the following dataset.

30, 40, 50, 55, 60, 65, 70, 80, 90, 100

Ans. First calculate the mean of the observations i.e. 64. Now subtract each value from 64 and take square. Add the square for all observation i.e. $(30 - 64)^2 + (40 - 64)^2 + \dots + (100 - 64)^2$ which is 4290. Divide this sum by 10 and take a square root.

Standard deviation value is $\sqrt{\frac{4290}{10}} = 20.712$.

As the observations are very far away from the mean value. Hence the standard deviation is coming out to be very high.

Ques 7: Calculate the standard deviation for the following dataset.

Observations	frequency
38	7
43	9
46	10
49	6
51	4
54	8
67	3
78	5

Table 5

Ans. Calculate the mean of the observations as shown in Table 6. Subtract the mean from the observations as shown in 4th column. Now take square of these observations as shown in the 5th column in Table 6. Column 6 shows the product of frequencies with values from column 5. Now take sum of the observations in column 5 and divide it by N i.e. total frequency. Take square root of the value 125.22 and the require standard deviation is 11.190.

Observations (x_i)	Frequency (f_i)	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i * (x_i - \bar{x})^2$
38	7	266	-12.65384	160.1198225	1120.838757
43	9	387	-7.653846	58.58136095	527.2322485
46	10	460	-4.653846	21.65828402	216.5828402
49	6	294	-1.653846	2.735207101	16.4112426
51	4	204	0.346153	0.119822485	0.479289941
54	8	432	3.346153	11.19674556	89.5739645
67	3	201	16.34615	267.1967456	801.5902367
78	5	390	27.34615	747.8121302	3739.060651
	N = 52	$\sum f_i x_i = 2634$			$\sum f_i * (x_i - \bar{x})^2 = 6511.769231$
	Mean	$\frac{\sum f_i x_i}{N} = 50.653$			$\frac{\sum f_i (x_i - \bar{x})^2}{N} = 125.2263314$
					$\sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}} = 11.19045716$

Table 6

Ques 8. Calculate the standard deviation for the following dataset.

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 7

Ans. First calculate the mid point of the class intervals then calculate the mean of the observations as shown in Table 8. Subtract the mean from the midpoint as shown in 5th column. Now take square of these observations as shown in the 6th column in Table 8. Column 7 shows the product of frequencies with values from column 6. Now take sum of the observations in column 7 and divide it by N i.e. total frequency. Take square root of the value 57.62 and the require standard deviation is 7.59.

Class Interval	mid value (m_i)	Frequency (f_i)	$f_i m_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
0.0-5.0	2.5	4	10	-12.1212	146.9234894	587.6939578
5.0-10.0	7.5	6	45	-7.1212	50.71148944	304.2689366
10.0-15.0	12.5	8	100	-2.1212	4.49948944	35.99591552
15.0-20.0	17.5	5	87.5	2.8788	8.28748944	41.4374472
20.0-25.0	22.5	7	157.5	7.8788	62.07548944	434.5284261
25.0-30.0	27.5	3	82.5	12.8788	165.8634894	497.5904683
		N = 33	$\sum f_i m_i = 482.5$			$\sum f_i * (m_i - \bar{x})^2 = 1901.515152$
			$\bar{x} = \frac{\sum f_i m_i}{N} = 14.6212$			$\sum \frac{f_i(m_i - \bar{x})^2}{N} = 57.62167126$
						$\sqrt{\sum \frac{f_i(m_i - \bar{x})^2}{N}} = 7.59089397$

Table 8

Merits and Demerits of Standard deviation

Merits

As standard deviation is considered as the best absolute measure of dispersion. These are the merits of standard deviation measure

- (i) Standard deviation is rigidly defined measure.
- (ii) Standard deviation is based on all the observations.
- (iii) Standard deviation is evaluated by squaring of deviation of observations from the mean. This makes the standard deviation for further mathematical treatment.
- (iv) Standard deviation has less influenced by the change in the sample observations.
- (v) It is also possible to compute the combined standard deviation of two or more dataset from individual standard deviation.
- (vi) Standard deviation is used for further statistical analysis like skewness and correlation.

Demerits

Although standard deviation is considered as the best absolute measure of dispersion but it is not free from demerits. These are

- (i) Standard deviation is very difficult to understand for a non-mathematical background person.
- (ii) Standard deviation gives more weight to extreme values and less weight to values close to mean. As while squaring the deviation this bias will be increased.

Although, standard deviation has some demerits or limitations but this is the widely used method for the absolute measure of dispersion.

3. Summary

In this module, one category of measures of dispersion i.e. absolute measures of dispersion are discussed in detail. These are range, quartile deviation, mean deviation and standard deviation. We first give an introduction about these measures. Questions and answers are also included for better understanding of the topic. Merits and demerits of each measure are discussed for better understanding and comparison purpose. Other branch of dispersion measure that is relative measure of dispersion will be discussed in the module “Measures of Dispersion- II with Skewness and Kurtosis”.

4. 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.