Paper No: 14 Statistical Applications in Environmental Sciences

Module: 01 Introduction to Statistics and its Importance

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	**Introduction to Statistics and its Importance**
Module Id	EVS/SAES-XIV/1
Pre-requisites	Nothing
Objectives	To give basic introduction of statistics and also explain its meaning and its importance for other sectors
Keywords	Statistics, data, descriptive, inferential, importance of statistics
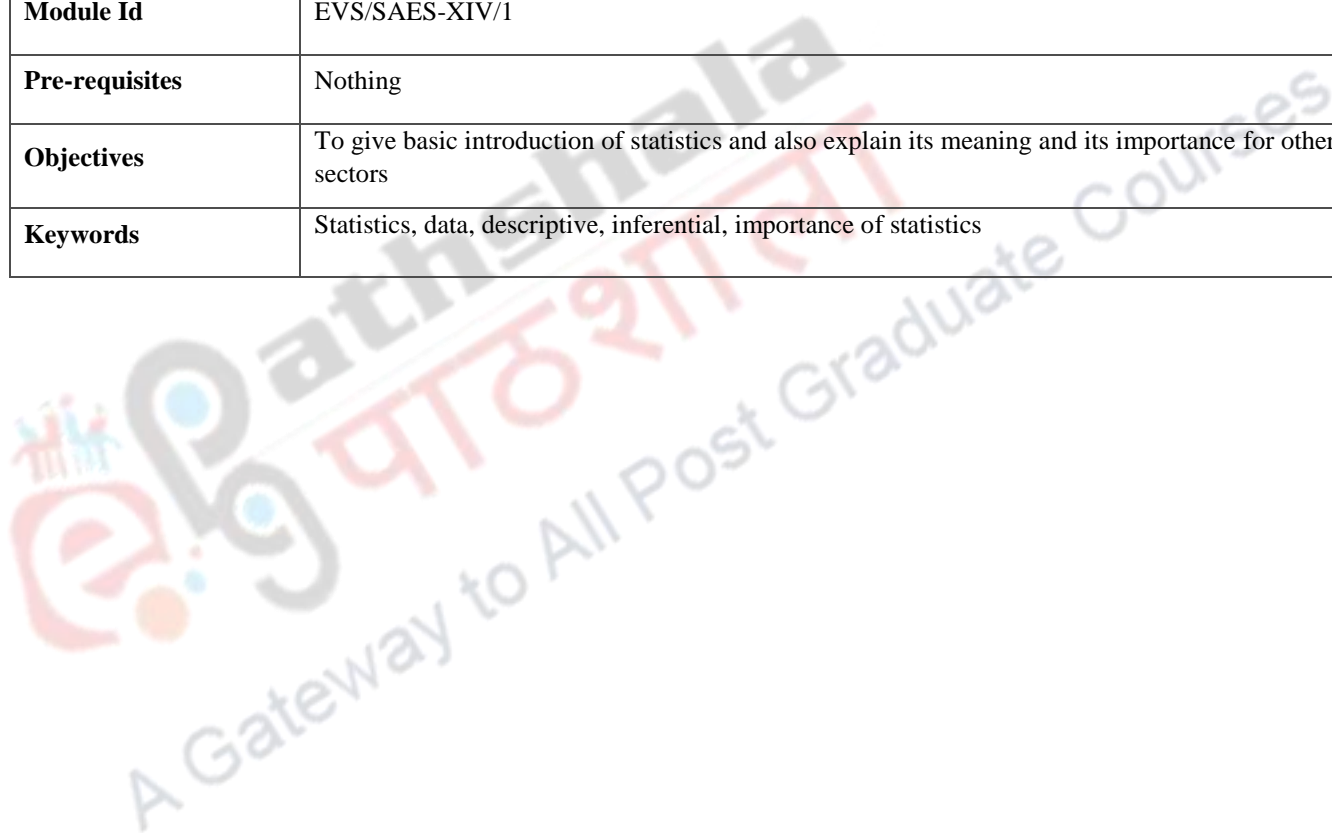
Module 1: Introduction to Statistics and its Importance

- ▪ **Learning Objectives**
- ▪ **Introduction.**
- ▪ **Classification of Statistics**
- ▪ **How Statistics works**
- ▪ **Need of Statistics**
- ▪ **Importance of Statistics**
- ▪ **Limitations of Statistics**
- ▪ **Summary**
- ▪ **Suggested Readings**

1. Learning Objectives

The objective of this module is to give basic introduction of statistics and an attempt has been made to explain its meaning and concept as well as its importance for other sectors. Examples given in this module will help the learner in understanding this topic very easily.

2. Introduction

Statistics is a science that uses different tools and techniques to organize the data in the descriptive form (in the form of table, graphs and pictorial presentation) and extract information from the data that helps in making decisions.

Statistics is considered as a study of collecting, analyzing, interpreting, presenting and organizing the data.

Definition. Statistics consists of a body of methods for collecting and analyzing data (Agresti and Finlay, 1997).

Statistics is basically used to get insight into the data. Statistics helps to get an overall picture of an objective from the data based on graphical presentation or numerical calculation without any constraints on the number of observations. It not only gives graphical representation but also makes inference and predict relations among variables.

For example, opinion polls for election involve designing of questionnaire then collection of data from the field with appropriate sampling methods. Further collected data is analyzed using statistical methods. At the last, the results are present in a understandable manner.

3

Self -Check Exercise

Question: What is Statistics?

Answer: Statistics deals with collection of data related with an objective, its analysis and finally its interpretation that is understandable by all the concerning persons involved with the objective in a direct or indirect manner.

3   Classification of Statistics

Statistical methods have different branches and each branch has its great importance in the literature as well as for practical point of view. The following model explains the relation between these methods.
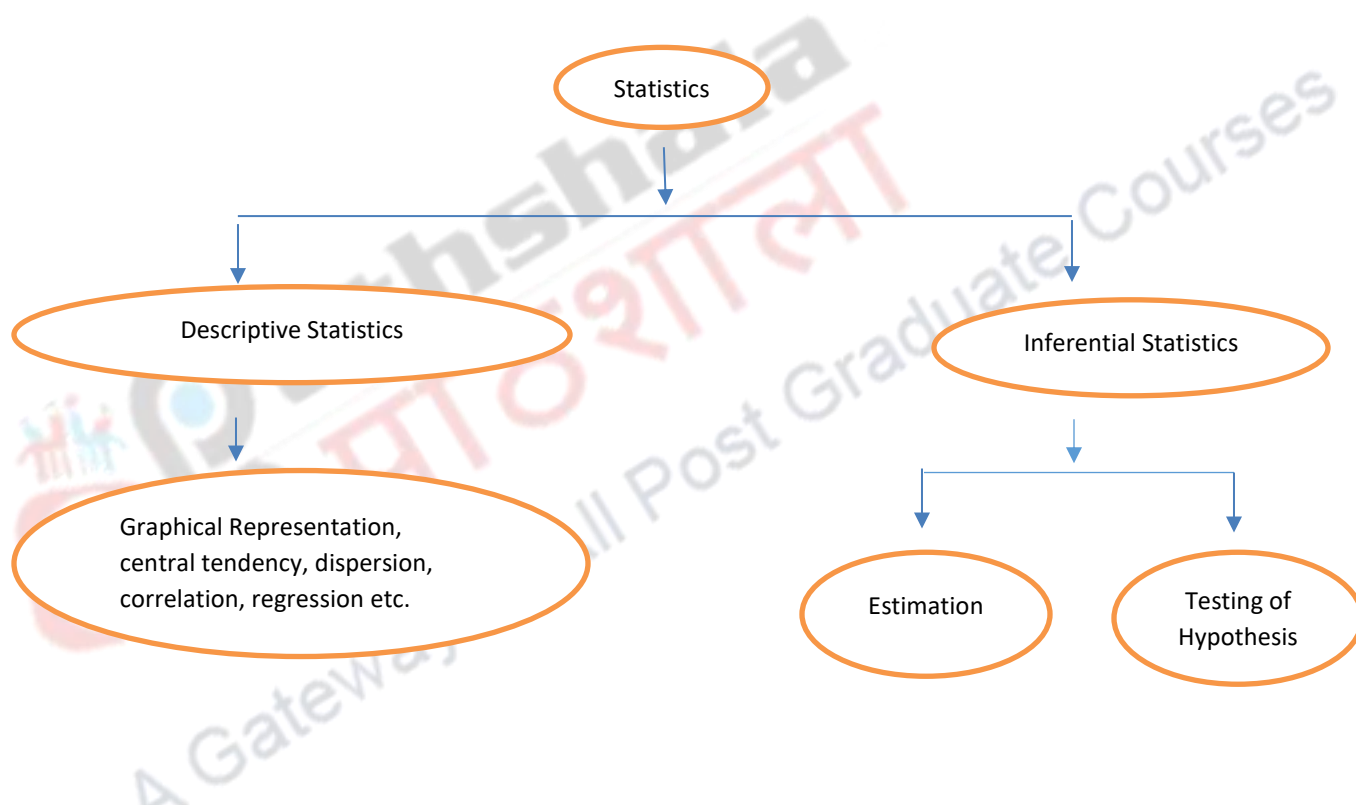
From the above model, one can see that for statistical methods are segmented into two methods. One is descriptive statistics and second is inferential statistics.

3.1 Descriptive Statistics

In Descriptive Statistics, graphical representation methods like histogram, bar chart, pie chart, ogive curves, box plot, line chart etc. are used to present data in graphical form. Descriptive methods also deal with the summary of the data. This includes these following measures:

(a)   Central Tendency Measures

For example mean, median, mode.

(b)   Measure of Dispersion

Variance, standard deviation, range.

There are other measures like skewness, kurtosis, quartiles, quantile, percentile etc. These measures are used to study other characteristics of the data. In descriptive statistics, one can study the scatterplot, correlation between variables, regression analysis.

Basically descriptive statistics deals with quantitatively describing the features of a data available from the sample/population.

3.2 Inferential Statistics

Descriptive statistics is used just to represent the information available in the data. But sometimes it is difficult to collect all the observation in the study for analysis. In this situation, sampling technique is used to collect element of the population in a non-random/random manner. Then inferential statistics is used to extract information from the sample as well to test the significance of an objective based on observations using testing of hypotheses. It is further segmented into two parts: These are

(a) **Estimation:** Estimation techniques is to extract information about the population from the sample.

(b) **Testing of Hypotheses:** This method is used to test the significance of a statement about a problem.

In this section, a brief introduction of statistics methods are given to give a broad picture. It is not expected from a learner to understand each and everything at this stage. In future modules, an attempt has been made to elaborate these methods with examples. For time being, just keep in mind the basic difference between descriptive and inferential statistics.

4   How Statistics works?

In a very simple language, statistics is basically used to find solution for an objective or a problem in different areas for e.g. if a person is interested in calculating the height of the students of college in particular region of a state for study. Now the question is how to approach this problem.

The first thought that come to your mind is how we can say anything about the height among college students as there is variations in heights among the students. Now, what to do?
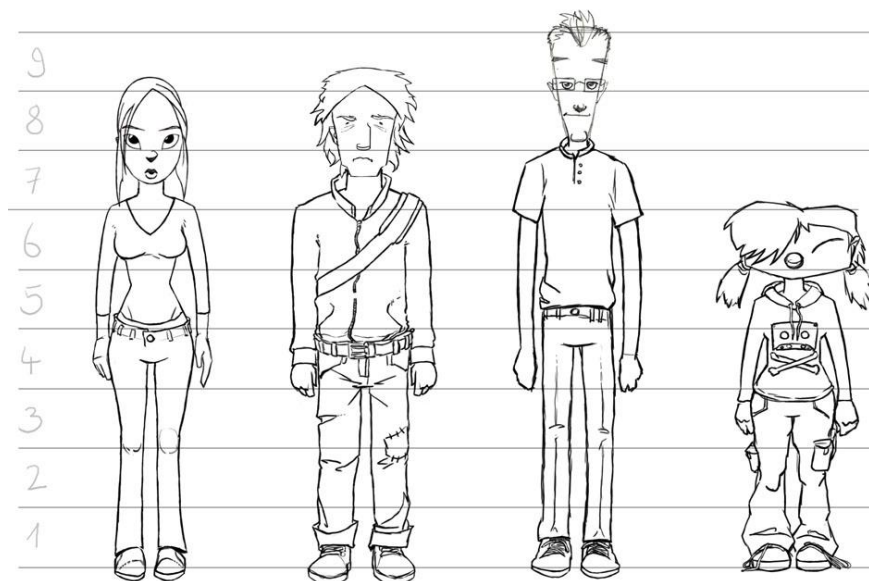
Figure 1

By using statistical methods, tools and techniques, one can find a numerical value that can help a person to find out solution. In this case, one can use very simple method that is measures of central tendency for e.g. mean, median, mode. "Mean" is generally used method in our daily life and is considered as the average of given values of a thing that has same characteristics. So take a mean of the values and see the outcome. The outcome will depend on the observation that we have. If the observations have little variation in term of height then average (mean) value is very close to observations values. If the observations are not homogenous (similar) in terms of height then the average value is different from observations values.

For e.g.

The values of the height of the 8 students in feet are 5' 2'', 5' 4'', 5' 6'', 5' 3'', 5' 6'', 5' 3'', 5' 5'', 5' 6'' and their mean value* is 5'4'' (approx).

If the values of the height of the 8 students are 4' 3'', 4' 9'', 5' 9'', 4'1'', 5' 1'', 4' 8'', 4' 5'',5' 3'' and their mean value is 4' 8'' (approx).

Now from the above example, one can see that there is major difference between the mean values of the two data.

Self-Check Exercise

Question: What is the reason? So what to say?

Answer: When the observations are not same then the mean of the data does not represent the given data. So the output value is different from the values given in the data.

Question: What to do in this situation?

Answer: We have to look for other methods of central tendency measures for relevant answer in this situation. We will study this problem again in further modules. We will discuss it again in the future.

5 Need of Statistics

In the previous section, we discussed a case about height of college's students and how to apply statistical methods with an example. In this section, we will discuss the need of statistics in our life. We are using statistics or basic mathematics intentionally or unintentionally in our daily life. There are many examples like while buying any item from the market people inspect a smallest fraction of an item to check the quality. In statistics, we call this selected item as "sample" and whole lot of an item from which item is selected called as "population".

Nowadays, we are more dependent on information technology that enables us to develop different tools and devices for different purposes. For example, mobile phone manufacturer companies want to increase the credibility of their product. That will increase the branding of their company's name. For these reasons, companies spend a lot on the Research and Development department. To test the reliability of products statistical tools and techniques based on data are used. But with increase in the complexity of the system, new tools and techniques are required which are more efficient and less costly. To fulfill the demand of data analysis consistent research is going on to cater this need. Hence, there is consistent increase in the demand of statistics tools and techniques.

We can see the descriptive presentation of data in day to day life for e.g.one can see the charts and graphs of inflation, opinion polls, economic growth etc on newspapers, television and other communication resources. Figure 2 represents the trend of inflation per year from the year 2001-2006. In this figure, one can observe that in 2002 inflation was less than in 2001.There was an increasing trend of inflation from 2001- 2004 and it declined till 2006.
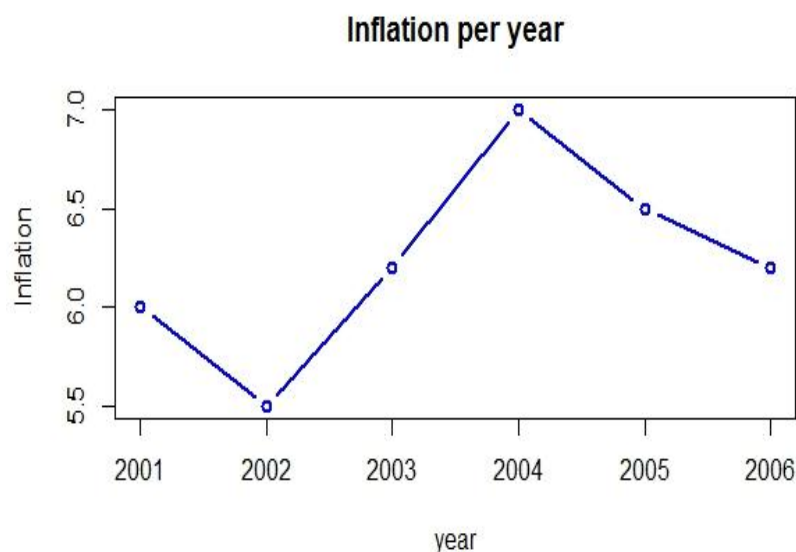
Inflation per year

Figure 2

The purpose of graphical representation helps in (descriptive presentation helps) understanding the objective in a better way. It is easy for a layman person to understand things. In this section, we discussed the need of statistics with an example.

Self-Check Exercise

Question: How statistics affects your daily life activities?

Answer Statistics helps in providing information that is hidden in the data either in descriptive form or inferential form. For example, sampling techniques are used while buying grocery items vegetables and fruits in the market.

Question How statistics estimates the characteristics of population from sample data?

Answer In Statistics, there are different functions (a formula of data values) available to estimate the population characteristics but the one who estimates the population characteristics in more accurate manner is considered. There are some parameters/conditions based on which one can check the optimality of the function.

For example, if one has to study the average expenses of a middle class family on food in a particular small town. One can choose using sampling technique some houses from the list of available houses then collect responses from them. This data is analyzed further to get an estimate of the average expenses on food. One cannot determine

8

the average expenditure on meal of a middle person based on single value. Statistics methods also help to give an interval containing lower and upper values. So, one can predict that expenses on food lie between this interval. In the next section, we will discuss the importance of statistics in different areas.

6. Importance of Statistics

In this section, we will discuss about the importance of statistics in different areas. As Statistics has a great role in different areas like determining the existing status of per capita income, mortality rate, inflation, population density rate etc. This is the reason that statistics holds a great importance in almost every field like Industry, Economics, Mathematics, Astronomy and Nursing and here we will discuss some of them below:

(i) In Industry

In Industry, statistics has its great importance. Industry growth depends on the demand and supply of goods, customer loyalty and market stability. Statistics helps businessman to plan production by taking into account all the possible factors that effects the sale of goods. Quality of products can also be checked by using statistical methods. Hence statistics has great importance in the working of an Industry.

(ii) In Economics

Economics deals with the study of factors that affects the economy of the world. So basically economics depends upon the figures of these factors and with passage of time these figures keep on changing. Economists use statistical tools and techniques to predict these factors with accuracy. In economics, statistical techniques are used for the collection of the data and as well as for analysis purposes. Statistics is also used to compare the factors like demand and supply, import and export, inflation rate, the per capital income etc. Economists use descriptive statistics tools like presenting data in tabular form, histogram, pie charts etc to present their objectives in more attractive as well easily understandable way.

(iii) In Mathematics

Statistics and Mathematics are complementary subjects. In mathematics, we use descriptive statistics as well as inferential statistics like central tendency measures, dispersion measures and estimation, hypothesis testing etc. In statistics, we use integration, derivative and basic algebra for analysis purposes. As mathematical tools and techniques can only give you most reliable results if they are based on strong evidence. Statistical methods are used for collecting these evidences with higher probability of more precisely. For example, to check the viability of a project introduced by a large firm is calculates using both mathematics and statistics techniques. Hence, mathematics and statistics methods are both enrooted with each other. Statistics is also called branch of applied mathematics.

(iv)   In Banking

Statistics has a great importance in banking field. As banks principle duty is to deposit money of depositors and use that money for issuing loans to lenders. There are many factors like rate of interest, economy stability etc. that determine the business of the bank. These factors keep on varying with time. Hence it is very important for banks to predict these factors more precisely as possible. Statistics techniques play an important role in predicting these factors with higher probability. Also statistics helps bank to predict defaulters of loans. For example, if the market will crash then people will lose their confidence in the economy of the country as well as bank. Also lenders will default on their loan amount due to lack of demand of their goods that will decrease their profit. Statistics can lower the bank's risk by predicting factors that effects the market.

(v)   In Administration

Statistical methods are basically used for collection as well as analysis of data and also for the interpretation of the results in different sectors of administration. In government offices, most of the projects for social welfare depend on it successful working. For example, if the government wants to increase the dearness allowance or revise the pay scale of employees due to increase in the living cost. Statistical methods will be used to determine the rate of inflation that affects the cost of living. Also central and state government budgets depends upon the statistical tools that help them in estimating income and expenditure from different sources. So statistics has a great importance in the administration's work.

(vi)   In Accounting and Auditing

Accounting is used by firms to maintain the balance sheets. It is based on data and it does not survive without the authenticity of the data. In accounting, statistics is used for decision making purposes for example how to correctly quantify the assets of the company, liquidity value of the assets, how to calculate depreciation cost of the assets etc. All these decisions are based on the market conditions. For example, all electronic assets value depleted within months due to high new technology development rate.

(vii)   In Natural and Social Sciences

   Statistics has its vital role in natural and social sciences. In natural sciences, for research and development purpose laboratory experiments and field experiments are conducted. Information is gathered in the form of the data. Now to draw conclusion from the data statistical methods are used. For example, a scientists are interested to test the effectiveness of newly developed drug for the heart patients. To test the effectiveness, they collect the information of health status of heart patient after taking dose of new medicine. It is possible

10

that drug is effective for some patients and ineffective for others. So, statistical method especially hypothesis testing (will cover in the coming session) is used to test the significance of newly developed drug for heart patients with some confidence.

In Social Sciences, most of the research work is done with the help of questionnaire and field work. Statistics is used to test the reliability and validity of the questionnaire. It is used to analyze the data obtained from the questionnaire and draw inference from them. For example, most of the social workers collect information through questionnaire. These questionnaire are further processed and stored in the form of data. This data is analyzed further and draw conclusions. In field work, statistics also help social scientists to determine the sample size and to adopt which sampling techniques.

(viii) In Astronomy

Astronomy deals with the densities of heavenly bodies, masses, measurement of distance between stars and planets, their sizes as well as other activities in the space. If errors occur while calculation of these figures that loss is unbearable for all. So, to avoid such type of errors statistical methods are used. For example, from the old days, least squares method is used to find the movements of stars.

(ix) In Demography

Statistics is considered as a backbone of demography. Demography deals with the study of population structure, sex ratio, health status and age group among population. It is difficult to collect information from each individual. Statistical methods are used to determine sample size and sampling technique based on objective. For example, if a person is interested in studying the mortality rate of cancer patients in a state. For collecting data, person must have to determine the resources from where one can collect information. After that appropriate sampling technique is used to collect samples. Based on these values one can estimate the mortality rate of cancer patients in a state.

7. Limitation of Statistics

(i)   Statistical analysis are mostly used for quantitative data. However, there are many tools available that can be used for qualitative analysis.

(ii)   It can only be applied collective and not on single events.

(iii)   Statistics give results on average basis that are true in the long run and not able to give result for a particular case.

(iv)   Statistical results are not percent correct. These are just approximate values and not exact one.

(v)   Statistics approach may not reveal the entire problem as some of the factors are incapable of statistical analysis.

11

(vi)   Statistical methods may lead to wrong decision if not applied in a correct manner.

(vii)   Statistics tools, if applied in a wrong manner, can lead to major loss.

In this section, we discussed about the importance of statistics in different fields. For more information one can refer to the links given below:

http://fen.ege.edu.tr/istatistik/tr/file/ders/14/15-10-14-09-50-11-Ders-Materyal-0.ogretim-C4%B0STAT%C4%B0ST%C4%B0%C4%9EEG%C4%B0R%C4%B0%C5%9E-14.pdf

https://en.wikipedia.org/wiki/List_of_fields_of_application_of_statistics

http://www.emathzone.com/tutorials/basic-statistics/importance-of-statistics-in-different-fields.html

8. Summary: As from above, the statistics has great importance in almost every field whether it is social sciences, natural sciences and industry. The reason behind is as statistics methodology is based on analysis of data (sample or whole population). Due to increase in complexity and volume of data, it is difficult to extract information easily. So, consistent research is going on to develop methods and tools to overcome these problems. These days most of the analysis is done using statistical softwares and programming languages.

9. 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.