Shamim Reza

Mr. Mohammad Manjur Alam (Manju)
Assistant Professor
Department of Computer Science and Engineering
International Islamic University Chittagong.
Mobile: 01715966144
Email: manjuralam44@yahoo.com

CHAPTER ONE

INTRODUCTION OF STATISTICS

Section –A (Mid-term Exam: 30 Marks)

1. **Preliminary idea of Statistics:** Origin, History and Development Statistics, Definition of Statistics, Characteristics, Function, Limitations, Necessity & importance of Statistics, The role of Statistics in Engineering, Population and Sample, Variable and Constants, Different types of variables,
Parameter, Statistic, Scale of measurement, Statistical data, collecting engineering data, Preparation of Questionnaire and Schedule, Presentation and Classification of Data, Construction of Frequency distribution, Graphical presentation of Frequency distribution.

** **Write down the origin of statistics.**

Ans: The word **"Statistics"** have been derived from the Latin word **'Status'** or the Italian word **'Statista'** or the German word **'Statistik'** or the French word **'Statistiques'**. Each of which means **"Political State"** or a Government.

** **History of statistics**

Ans: Gottfried Achenwall used the word "statistic" at a German University in 1749 which means that political science of different countries. In 1771 W. Hooper (Englishman) used the word statistics in his translation of Elements of Universal Erudition written by Baron B.F Bieford, in his book statistics has been defined as the science that teaches us what is the political arrangement of all the modern states of the known world. There is a big gap between the old statistics and the modern statistics, but old statistics also used as a part of the present statistics.

During the 18th century the English writer have used the word statistics in their works, so statistics has developed gradually during last few centuries. A lot of work has been done in the end of the nineteenth century.

At the beginning of the 20th century, William S Gosset was developed the methods for decision making based on small set of data. During the 20th century several statistician are active in developing new methods, theories and application of statistics. Now these days the availability of electronics computers is certainly a major factor in the modern development of statistics.

Shakespeare used a word Statist is his drama Hamlet (1602). In the past, the statistics was used by rulers. The application of statistics was very limited but rulers and kings needed information about lands, agriculture, commerce, population of their states to assess their military potential, their wealth, taxation and other aspects of government.

**** Definition of statistics.**

Ans. Different authors defined statistics in a number of ways. Among those some of the important definitions are given below:

Dr. A.L. Bowely defined "Statistics are numerical statement of facts in any department of enquiry placed in relation to each other".

R.A. Fisher defined, "The science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data".

According to Croxton and Cowden, "Statistics may be defined as the science of collection, presentation, analysis and interpretation of numerical data".

**** Most essential functions of statistics:**

1. It simplifies much of figure.
2. It facilitates classification and comparison of data.
3. It helps in determining the relationships two or more phenomenon.
4. It helps in formulating and testing suitable hypothesis.
5. It helps a central management in formulating suitable future policy.
6. It helps in predicting future trends.

**** Limitations of Statistics:**

1. Statistics deals with aggregate of items and not with individual item or measurement.
2. Statistics deals only with quantitative characteristics.
3. Statistical laws hold good only for the averages.
4. It plays only an auxiliary rule in summarizing a fact.
5. Statistics can be misused.

**** Characteristics of statistics.**

Ans: The characteristics of statistics are given below:

1. Statistics should deal with aggregate of individual rather than with individual alone.
2. Statistics should be expressed as numerical figure.
3. Statistics should have the property of being varied by multiplicity of causes.
4. Statistics are collected or estimated should be reasonable of accuracy.
5. Statistics should be obtained by pre-determined purpose.
6. Statistics are collected is a systematic manner.

**** Importance of statistics.**

Ans: The importances of statistics are given below:

1. Statistics of wealth and manpower are important for development and planning.
2. Statistics are invaluable in business and commerce.
3. Statistics helps the planner to estimate the revenue income and expenditure of the country.
4. Agriculture Statistics may play a key role in agriculture development.
5. In industry, statistics is widely used to provide quality control.
6. Statistics is usually used in education and psychology too.

****Relation between Computer and Statistics.**

Statistics is defined as the science of collecting, organizing, and interpreting numerical facts, which we call data. It is very important for a student of computer sciences. As Computer sciences also deals with organization and interpretation of numerical facts. In fact most of the principles of computer sciences are based on concepts of statistics.

The computer can process large amount of data quickly and accurately. For processing the large amount of data some of the important statistical packages that have been used are – SPSS, SAS, STRATA, S-plus and MINITAB.

****Define population with example.**

The totality of all elements under the study or discussion is called population. The "population" in statistics includes all members of a defined group that we are studying or collecting information.

Example: If we measure the heights and weights of a group of person then it is called population.

Populations are two types: (i) Finite population;

(ii) Infinite population.

Finite population: A Population is called finite population if its elements are countable. Example: Number of students in a university.

Infinite population: A population is called infinite population if its elements are not countable.

Example: Number of fishes in the Bay of Bengal.

****Define Sample with example.**

Ans: A representative part of population is called sample.

Or a part of the population is called **sample.**

Example: If we measure the height and weight of IIUC student, then that of the CSE/EEE/ETE department student are sample.

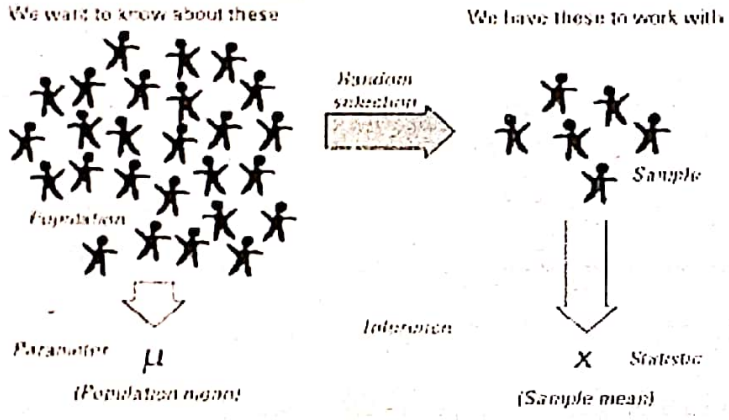
****What is the difference between a population and a sample?**

The main difference between population and sample are as follows:

Population	Sample
# The totality of all elements under the study or discussion is called population.	A part of the population is called sample.

# A **population** includes each element from the set of observations that can be made.	A **sample** consists only of observations drawn from the population.
# Population may be finite or infinite	Sample must be finite.
# Collecting data from every element of a population is not easy.	Collecting data is easy.
# All registered voters in our country	All registered voters in Chittagong district

**** Figure of population and sample**



Define Variable with example.

Any phenomenon which varies from individual to individual is called variable. Variables are represented by symbols (e.g., x, y, or z).

For example Age, weight, height, sex, income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are examples of variables.

Variables can be classified as **qualitative** (categorical) or **quantitative** (numeric).

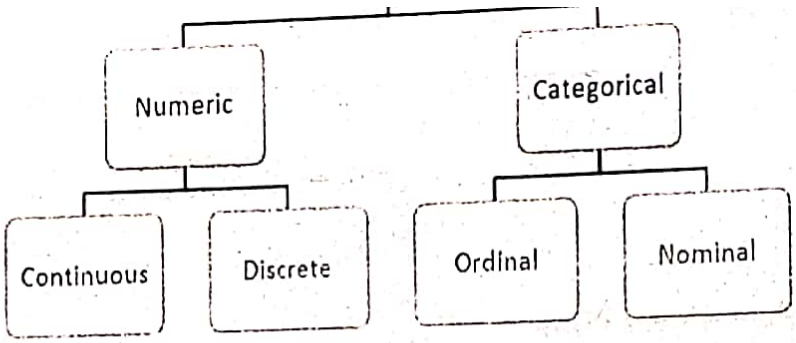
Qualitative variable:
A variable which cannot be expressed as numerically is called **qualitative** (categorical) variable.
Examples: Hair color, gender, field of study, college attended, political affiliation, status of disease infection, sex, business type, eye colour, religion and brand.

Quantitative variable:

A variable which can be expressed as numerically then it is called quantitative variable.

Examples: Height, age, crop yield, GPA, salary, temperature, area, air pollution index (measured in parts per million), etc.

Quantitative variables can be further classified as **discrete** or **continuous**.



**Difference between qualitative variable and quantitative variable.

Qualitative variable	Quantitative variable
Cannot be measured numerically	Can be measured numerically.
It is not countable	It is countable
Qualitative variable can be measure by using nominal and ordinal scale	Quantitative variable can be measure by using interval and ratio scale
The value of qualitative variable is generally discrete.	The value of quantitative variable is discrete and continuous
Algebraic expression meaningless	Algebraic expression meaningful
Skin colour, hair colour, religion, gender, merit, education, character are the example of qualitative variable.	GPA, age, weight, height, income, temperature etc are the example of quantitative variable.

**Difference between Discrete variable and continuous variable.

**** Define constant.**

Ans: A number that is not changing. It is usually denoted by a, b, c or d.

Example: Number of finger in a hand or leg.

Variable	Constant
A variable is always subject to change.	A constant will not change, ever.
it is denoted by X,Y,Z,U,V	It is denoted by a, b ,c or d
Variable are qualitative and quantitative	Constant has no such classification
age, weight, height, salary etc are variable	Total number of days in a week, Number of finger in a hand or leg are constant

**** What do you mean by data?**

Ans: A set of observations obtained from a particular enquiry is called data.

Example: Income of workers or examination marks of a students.

According to sources data are:

1. Primary data;
2. Secondary data;
3.

Primary data: Data collected by the investigator himself/ herself for a specific purpose.

Example: Data collected by different govt., public, private organizations, research bodies, research scholars, NGO's for their official records and research purpose from the field directly are primary data.

Secondary data: When an investigator uses the data which has already been collected by others, such data are called secondary data.

This data can be obtained from journals, report, Internet, Books/ Magazines, Newspapers, Office statistics the government statistics service, the office of national statistics, centre for applied social surveys.

Methods of collecting primary data:

1. Through interview
2. Through questionnaire
3. Through schedule
4. Through local agent.
5. Through observations
6. Through experimentation.

What is the difference between Primary and Secondary Data?

Primary Data

1. Primary data are always original as it is collected by the investigator.

2. Suitability of the primary data will be positive because it has been systematically collected.

3. Primary data are expensive and time consuming.

4. Extra precautions are not required.

5. Primary data are in the shape of raw material.

6. Possibility of personal prejudice.

Secondary Data

1. Secondary data lacks originality. The investigator makes use of the data collected by other agencies.

2. Secondary data may or may not suit the objects of enquiry.

3. Secondary data are relatively cheaper.

4. It is used with great care and caution.

5. Secondary data are usually in the shape of readymade products.

6. Possibility of lesser degree of personal prejudice.

**** Define Questionnaire. Write down the characteristics of a good Questionnaire.**

Ans: In any survey the information are collected according to some predetermined question. A set of question for any survey constitutes the questionnaire.

Or, A set of printed or written questions with a choice of answers, devised for the purposes of a survey or statistical study.

Characteristics of a good Questionnaire:

1.Questions worded simply and clearly, not ambiguous or vague

2.Write an introduction to the questionnaire

3. Question should be logically arranged.

4.Question should be as few as possible.

8. Home District:	9. Number of brothers and sisters:.....
10. What is your serial number as a son/daughter of your parents? (a) First, (b) second, (c) third,(d) fourth (e) _____	11.Areas of your School at SSC or equivalent level : (a) city, (b) small town (c) village
12. Areas of your School at HSC or equivalent level : (a) city, (b) small town (c) village	13. Did you pass (a) SSC/HSC, (b) O-level/A-level,(c) Madrasa qualifications (d) others
14. SSC and HSC (or equivalent) GPA: SSC:_____ and HSC.....	15. Father's Education: (a) No education (b) Primary, (c) Secondary, (d) Higher Secondary (e) University.
16. Mother's Education: (a) No education (b) Primary, (c) Secondary, (d) Higher Secondary (e) university.	17. What is (or was) your father's occupation: (a) Agriculture (b) Business (c) Service (d) teacher and (e) Others.
18. What is (or was) your mother's occupation: (a) Housewife (b) Business (c) Service (d) teacher and (e) Others.	19. Your father and mother monthly income:.....
20. Economic status of your family:(a) Poor, (b) lower middle class (c) middle class, (d) higher middle class, (e)rich	21.Behind Choosing IIUC: (a) Islamic (b) Tuition (c) Safety and distance (d) Faculty (e)Good private university(f) Scholarship (g) others.
22. How you got information about IIUC: (a) Advertisement, (b) Faculty (c) friends/relatives, (d) internet	23.While attending university, I live: (a) In a hall (b) At home/ Family (c) Mess (d) Relative/friends (e) others
24. Approximately, what was your grade average in your final year of IIUC:.....out of 4.00.	25. Do you discuss your grades with your guardian? (a) Yes (b)No

26.Compared to my friends at university, I am, on an academic basis, performing:(a) Better (b) Same (c)Worse	27. How often did you miss classes: (a) Always (b) Sometimes (C) Never (d) Once every week?
28. Usually I study: (a) Library (b) Class (c) my room (d) If other please Specify:.....	29. Generally, I study.....hours daily.
30.Are you satisfied in IIUC academic system: (a) yes (b) No.	31. Do you seem that tuition fee is higher in IIUC? (a) yes (b) No.
32. What's your dream Career? (a) Subject related job(b) Teacher (c) Banker (d) service (e) others.	

THANK-YOU

**** Types of Questionnaire.**

Ans: There are two types of questionnaire:

1. Unstructured question.
2. Structured question.

Unstructured question: Unstructured question are open-ended, that respondents can answer in their own words.

Example: What is your occupation?

Structured question: A question in which the respondents is given specific limited alternative responses and ask to choose one to his/her own view point.

Example: What is your occupation?

1. Agriculture 2. Business 3. Service 4. Teacher and 5. Others.

Methods of organization of data:

1. Classification
2. Tabulation
3. Frequency distribution

Classification: Classification is the first statistical technique to condense the raw data. It is the process of arranging data in different groups or classes according to their affinities.

Tabulation: Tabulation is a logical and systematic arrangement of statistical data in raws and columns.

Frequency:

The number of times that a given value occurs into each group / class is known as frequency.

For example, if four students have a score of 80 in mathematics, and then the score of 80 is said to have a frequency of 4. The frequency of a data value is often represented by *f*.

Frequency distribution

Arrangement of observations according to frequencies is called frequency distribution.

Write down the construction of a frequency distribution.

Following are the steps for the construction of a frequency distribution:

1. Class limits/Range: The class limits are the lowest and highest values that can be included in the class.

2. Number of classes: $K = 1+3.322 \log_{10}{}^{N}$ where N is the number of observations,

3. Class interval: C.I = Range /K where, Range = H.V - L.V.

4. Class frequency/Tally: The number of observations corresponding to the particular class is known as class frequency of that class.

5. Mid value: (U.L+L.L) /2

6. Cumulative frequency: The number of observation for less then a certain upper limit of the class, if cumulation is done from top of the table.

Assignment-1: The following data refer to the ages of 60 employees of a firm.

33, 41, 21, 25, 36, 38, 35, 36, 35, 37, 42, 30, 35, 37, 36, 38, 30, 54, 40, 48, 15, 28, 51, 42, 25, 41, 30, 27, 42, 36, 28, 26, 37, 54, 44, 31, 36, 40, 36, 22, 30, 31, 19, 48, 16, 42, 32, 21, 22, 40, 43, 42, 39, 38, 37, 33, 49, 47, 46, 48.

(i) Construct a frequency table with suitable class interval.
(ii) Draw histogram, frequency polygon and ogive.

Tips: As like as class lecture.

Assignment-2: Monthly income (in Lac Tk.) of 30 firms in a certain area is given below:

20,19,10,13, 8,15,9,16,18,8,16,17,12,11,10,19,18,17,14,15,12,13,14,12,12.5,18.5,20,15,18,16.

(i) Construct a suitable frequency distribution.

(ii) Draw histogram and ogive.

Assignment-3: Tube light production (hourly) of 30 machines is given below:

9,20,19,10,12,15,9,16,18,8,16,17,12,11,10,19,18,17,14,15,12,13,14,12,12,18,20,15,18,15.

(i) Construct a frequency distribution with suitable class interval.

(ii) Draw ogive curve.

Example 6

The number of calls from motorists per day for roadside service was recorded for the month of December 2003. The results were as follows:

28	122	217	130	120	86	80	90	120	140
70	40	145	187	113	90	68	174	194	170
100	75	104	97	75	123	100	82	109	120
81									

Set up a frequency table for this set of data values.

Solution:

To construct a frequency table, we proceed as follows:

Smallest data value = 28

Highest data value = 217

Difference = Highest value – Smallest value

= 217 – 28

= 189

Let the width of the class interval be 40.

$\therefore$ Number of class intervals $= \frac{189}{40} = 4.7 \approx 5$ (Round up to the next integer)

There are at least 5 class intervals. This is reasonable for the given data.

Step 1: Construct a table with three columns, and then write the data groups or class intervals in the first column. The size of each group is 40. So, the groups will start at 0, 40, 80, 120, 160 and 200 to include all of the data. Note that in fact we need 6 groups (1 more than we first thought).

Class interval	Tally	Frequency
0 - 39		
40 - 79		
80 - 119		
120 - 159		
160 - 199		
200 - 239		

Step 2: Go through the list of data values. For the first data value in the list, 28, place a tally mark against the group 0-39 in the second column. For the second data value in the list, 122, place a tally mark against

the group 120-159 in the second column. For the third data value in the list, 217, place a tally mark against the group 200-239 in the second column.

Class interval	Tally	Frequency
0 - 39	I	
40 - 79		
80 - 119		
120 - 159	I	
160 - 199		
200 - 239	I	

We continue this process until all of the data values in the set are tallied.

Step 3: Count the number of tally marks for each group and write it in the third column. The finished frequency table is as follows:

Class interval	Tally	Frequency
0 - 39	I	1
40 - 79	𝍸	5
80 - 119	𝍸 𝍸 II	12
120 - 159	𝍸 III	8
160 - 199	IIII	4
200 - 239	I	1
	Sum =	31

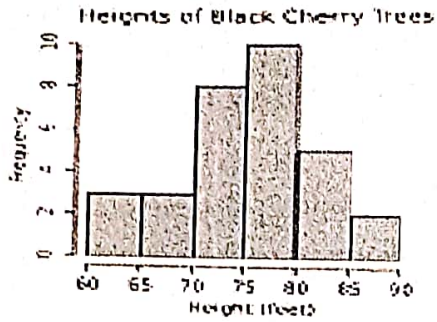
****Define the graphical representation of statistical data. Write different types of graphs and Diagram.**

Ans: In addition to presentation of statistical data through tabular form, one can present the same through some visual aids refer to graphs and diagrams.

Types of Graphs and Diagrams:

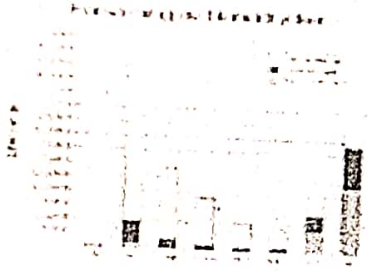
1. Histogram;
2. Bar diagram;
3. Frequency polygon
4. Pie diagram;
5. Scatter diagram;
6. Line diagram;
7. Ogive;
8. Steam and leaf plot.
9. Box plot

Histogram:



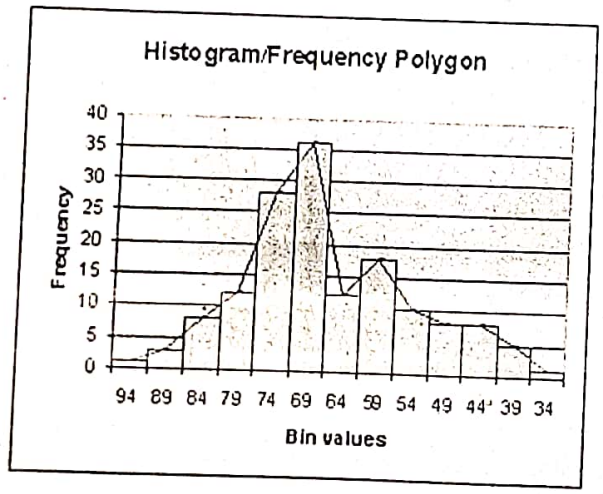
*A histogram consists of tabular frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval.

Bar chart:



A bar chart is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.

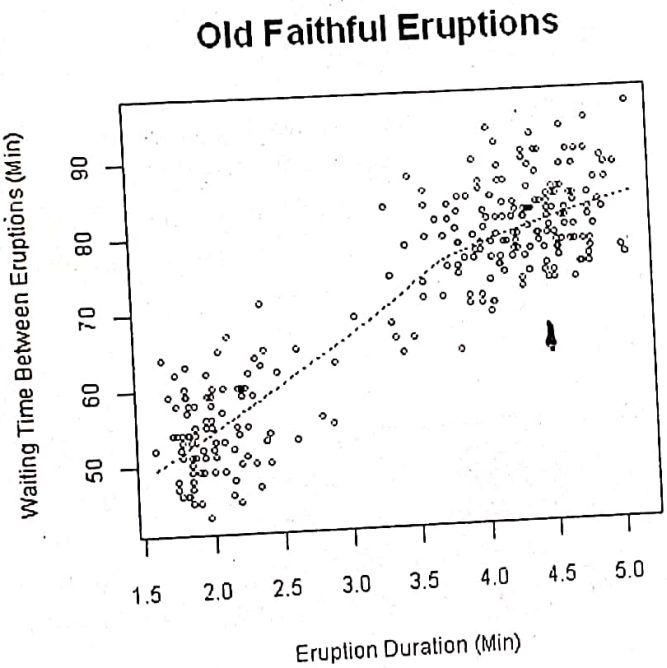
Frequency polygon



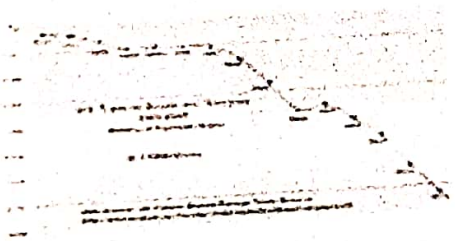
Pie chart:

A pie chart shows percentage values as a slice of a pie.

Scatter diagram:

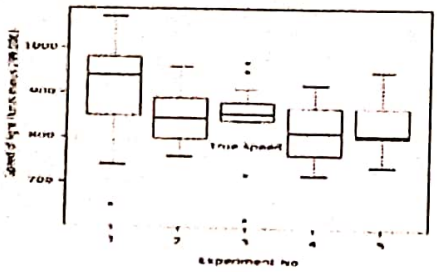


Line chart:



A line chart is a two-dimensional scatter plot of ordered observations where the observations are connected following their order.

Box plot



Cumulative frequency polygon or ogive

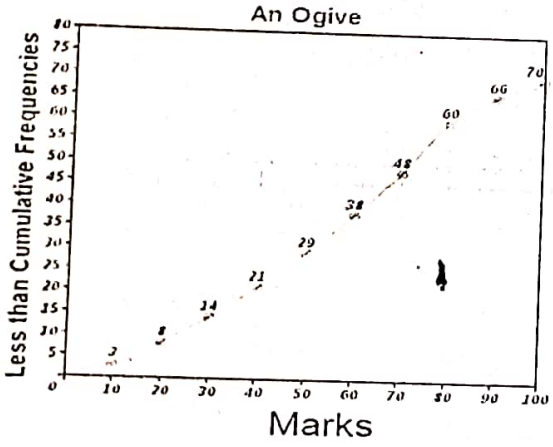


Figure: Ogive by using upper class interval

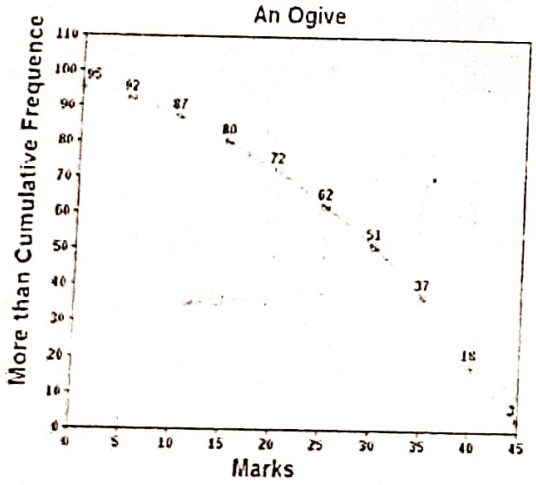


Figure: Ogive by using lower class interval

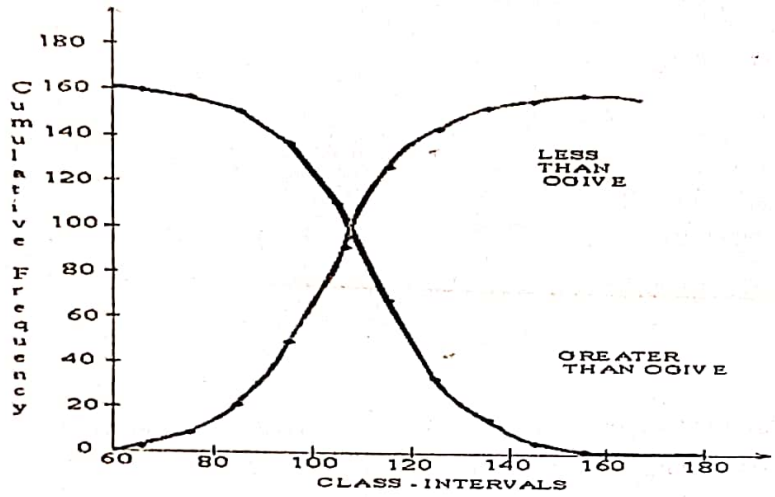


Figure: Ogive by using upper and lower class interval

The difference between graph and diagram:-

Diagram

1) Diagram can be drawn on plain paper and any sort of paper.
2) Diagram is more effective and impressive.
3) Diagram have everlasting effect.
4) Diagram cannot be used as median, mode etc.
5) Diagram can be represented as an approximate idea.

Graph

1) Graph can be drawn only on plain paper.
2) Graph is not more effective and impressive.
3) Graph doesn't have everlasting effect.
4) Graph can be used as median, mode etc.
5) Graph cannot be represented as an approximate idea.

Previous semester question

Spring-2014(CSE):

(a) What is the meaning of the word statistics? Discuss a suitable definition of statistics.

(b) Suppose a medical researcher wants to find the average systolic blood pressure of the employees of a big firm. For this purpose, a sample of 50 employees has been selected randomly from that firm and their systolic blood pressure measurements have been obtained.

(i) What is the population?
(ii) What is the sample?
(iii) What are the difference between (i) and (ii).

(iv) Is the variable qualitative or quantitative?
(v) Is the variable discrete or continuous?

Autumn-12(CSE):

(a) How statistics and computer are related?

(b) Distinguish between (i) Primary and secondary data.
(ii) Population and sample.

(c) Define Questionnaire. Write down the characteristics of a good Questionnaire.

(a) Define qualitative and quantitative variable with example. Identify the nature of the variables (Qualitative or quantitative): Eye Color, Marital status, Blood group, Family member

Autumn-12(ETE):

(a) Define statistics is singular and plural sense.

(b) What do you mean by statistical data? "Statistical data are usually of two types: (i) Primary (ii) Secondary." Explain in brief.

Autumn-12(EEE):

(a) Distinguish between Questionnaire and schedule.

(b) Define qualitative and quantitative variable with example. Identify the nature of the variables (Qualitative or quantitative): Eye Color, GPA, Marital status, Blood group.

(c) Draw Histograms to represent the following frequency distributions:

(b) Define the followings with examples:

(i) Population (ii) Variable (iii) Constant (iv) Qualitative variable

(c) Write down the construction of a frequency distribution. Discuss some important graphs and charts to represent statistical data.

Autumn-10(BBA)

(a) What are the characteristics features of statistics? Define variable with example. Describe the steps for constructing a frequency distribution.

(b) Tube light production (hourly) of 30 machines is given below:

9,20,19,10,12,15,9,16,18,8,16,17,12,11,10,19,18,17,14,15,12,13,14,12,12,18,20,15,18,15.

(i) Construct a frequency distribution with suitable class interval. (ii) Draw ogive curve.

Autumn-09(BBA)

1. (a) Define statistics. (b) Write down the function of statistics. (c) What is data? Define different types of data with examples (d) State the different steps of constructing a frequency distribution.

2. (a) What is classification of data?

(b) The daily wages (in Tk.) of 30 workers in an industry is given below:

70,82,77,95,120,97,100,94,145,147,82,84,140,137,142,139,150,89,136,142,88,76,99,85,106,120, 116, 115,151,130

i) Construct a suitable frequency distribution.

ii) Draw histogram and ogive curve.

Statististics syllabus

Course code : STAT-1201, Course tille : Statistics, Credit Hours:, Contact Hours: 2 W

Objectives: In this course student will learn about' Statistics' in regards to definition of statistics, its necessity, measures of central tendency, dispersion, correlation theory, regression analysis, probability distributions

Section-A (Mid-term: 30 Marks)

1. Preliminaries: Definition of Statistics, Its necessity & importance, Population and Sample, Variable and Constants, Different types of variables, Statistical data, Data Collection and presentation, Construction of Frequency distribution, Graphical presentation of Frequency distribution.

2. Measures of Central Tendency: Arithmetic Mean, Geometric Mean, Harmonic Mean, Median, Mode, Weighted Mean, and Theorems & Problems.

3. Measures of Dispersion: Range, Standard Deviation, Mean Deviation, Quartile Deviation, Variance, Moments, Skewness and Kurtosis, Theorems & Problems.

Reccommended Books: 1.An introduction to the theory of Statistics (R.N. Shill & S.C. Debnath)
2. Business Statistics (M.K. Roy and J.C Paul)