“It is better to be roughly right than precisely wrong”- John Maynard Keynes

CHAPTER I: Introduction

Why study statistics? is a question countless students ask. Is it simply a HilCoE requirement you have to complete? One simple answer is that statistics have become a common part of every day life and, therefore, deserve attention. A quick glance at the newspaper yields statistics that deal with crime rates, birth rates, death rates, average income, average rainfall, and so on. By a common definition, therefore, statistics consist of facts and figures.

Statistics is used in almost all fields of human endeavor. It is in one form or the other has become a common language by which agronomists, chemists, educators, *programmers,* geneticists, political scientists, military strategists, medical researchers, sport managers etc. read some portion of the world in which we live. That is why we say:***” mathematics or the body of mathematical applications called statistics- is the language in which we read the universe.”***

The word statistics is derived from the Italian word ***stato*** which means “state” and ***statista*** refers to a person involved with the affairs of state. Therefore, statistics originally meant the collection of facts that is useful to the ***statista***. Statistics in this sense was used in Italy (in the 16^{th} century) and spread to France, Holland and Germany.

The term “Statistics” has been derived from the Latin word “STATUS” which means political states. Initially statistics was concerned with economical, political and demographical particulars of a country. Slowly development of theory of probability added importance to statistics. The development of modern statistics started after industrial revolution. Now statistics became a universally applicable science. In any field where ever information can be measured statistics acts as a required tool for further action.

The subject statistics is primarily concerned with making decisions about various properties of some populations of interest such as stock market trends, unemployment rates, interest rates, inflation rates over the years and so on. For example consider the following statistical statements:

- The crime in Nazareth has gone by 12% of what it was last year.

- The average salary of instructors at HiLCoE is 2500 birr per month.
- The rate of inflation in South Africa is expected to remain below 3% per year for the next ten years.
- Of all the students who join higher institution in Ethiopia less than 60% actually graduate.
- Majority of Americans consider that Japanese cars are superior in quality than American cars.

All the above statements represent statistical conclusion in some form. These conclusions help us in formatting specific policies and attitudes with respect to diverse areas of interest. Statistics then is the branch of science that deals with numbers or figures describing the state of affairs of various situations with which we are generally and specifically concerned.

In sum, the subject of Statistics deals primarily with numerical data gathered from surveys or collected in experiments. Its objective is to summarize such data so that the summary gives us a good indication about some characteristics of a population or phenomenon that we wish to study.

1.1 Definition of statistics

Statistics is a branch of mathematics that deals with the scientific methods of collecting, organizing and analyzing data so as to make valid conclusions and reasonable decisions on the basis of such analysis.

The raw materials for any statistical analysis are ***data*** which are the numerical information from which we make analysis and interpretation.

1.2 Steps of statistical investigation

To ensure that our conclusions are meaningful, it is necessary to subject our data to scientific analyses so that rational decisions can be made. Hence, the field of statistics is concerned with the following basic steps:

- The situation being investigated should be carefully and fully defined
- Proper collection of data
- Organization and classification of data

- Presentation of data
- Analysis of data
- Interpretation of data.

1.3 Importance and Limitations of Statistics

1.3.1. Importance: Statistics is no longer confined to the domain of mathematics. It has spread to most of the branches of knowledge including social and natural sciences. One of the reasons for its phenomenal growth is the variety of different functions to it. Some of the most important functions of statistics are:

- ***Statistics represents facts in the form of numerical data***. Statistical expressions are numerical expressions, which are in precise and definite form, which make it easier to understand. For example, statement like "crime is very high in Awassa" does not convey as precise meaning as the actual figures related to crime for the current year and their comparison to similar figures of the previous year.
- ***Statistics condenses and summarizes a mass of data in to few presentable, understandable and precise figures*.** The raw data, as usually available, is huge and haphazard messy. It is generally not possible to draw any conclusion from the raw data as collected. Hence it is necessary and desirable to express this data in few numerical values. For example the average salary of employees at Addis Ababa University is derived from a mass of data from surveys.
- ***Statistics facilitates classification and comparison of data***. Arrangement of data with respect to different characteristics facilitates comparison and interpretation. Moreover the absolute figures themselves do not convey any significant meaning. It is their comparison that helps us draw conclusions. Statistical devices such

as averages, percentages, ratios etc. are the tools that can be employed for the purpose of comparisons.

- ***Statistics helps the central management and government in formulating policies***. Different governmental policies regarding import and export trade, taxation, planning, resource allocation and so on are formulated on the basis of data regarding these elements. Many other policies are based upon statistical forecasts made by statisticians, such as policies regarding housing, employment, industrial expansion, food grain production and so on.
- ***Statistics helps in predicting future trends.*** Statistical methods are highly useful tools in analyzing the past data and predicting some future trends. For example the sales of a particular product for the next year can be computed by knowing the sales for the same product over the previous years, the current market trends and the possible changes in the variables that affect the demand of the product.

1.3.2. Limitations: The field of statistics, though widely used in all areas of human knowledge and widely applied in a variety of disciplines such as business, economics and research, has its own limitations. Some of these limitations are:

- ***Statistics does not deal with individual values.*** It only deals with aggregate (collective) values. For example the marks obtained by one student in a class do not carry any meaning in itself unless they can be compared with a set standard or with those of other students in the same class or with his marks obtained in earlier periods.
- ***Statistics does not deal with qualitative characteristics.*** It only deals with data which can be quantified. Certain qualitative characteristics such as honesty, integrity, poverty, beauty or goodness can not be measured statistically.
- ***Statistical conclusions are not universally true.*** Since statistics is not the exact science, as is the case with natural sciences, the statistical conclusions are true under certain

assumptions. Also, the field deals extensively with the laws of probability which at best are educated guesses.

"To guess is cheap. To guess wrongly is very expensive"

- ***Statistical interpretation requires a high degree of skill and understanding of the subject.*** In order to get meaningful results, it is necessary that the data be properly and professionally collected and critically interpreted. It requires extensive training to read statistics in its proper context. It may lead to fallacious conclusions in the hands of inexperienced people.
- ***Statistics can be misused.*** Inaccurate or incomplete figures can be manipulated to get desirable references. For example "9 out of 10 dentists' surveyed recommended ***Colgate*** toothpaste", gives the impression that 90% of all dentists recommend ***Colgate.*** This may not be true since the sample size may not be significant or even representative. Accordingly, such statistics can be highly misleading.

1.4 CHARACTERISTICS OF STATISTICS (Home take assignment)

1.5 Terminologies

Scientific research typically begins with general question about a specific group (or groups) of individuals. For example a researcher may want to examine the attitudes toward abortion for men versus women. In this case the researcher is interested to compare the group of men with the group of women. In statistical terminology, the entire group that a researcher wishes to study is called a population. But in collecting data concerning the characteristics of a group of individuals or objects, it is often impossible to observe the entire group.

Example1: - The number of fish in lake Shala.

- The number of residents in China who watch the 19th World Olympics Champion in Beijing.

In such cases, instead of examining the entire group, called ***population***, we examine a small group called ***sample***.

Definition 1:- A ***population*** in statistical terms is the set of all elements that belong to the entire group.

Definition 2:- A summary measure that describes any given characteristic of the population is known as a ***parameter.***

Parameters describe the **P**opulation. Notice that both begin with the letter **P**.

Definition 3:- A ***sample*** is a portion of the total population that is considered for study and analysis.

Definition 4:- The summary measure that describes a characteristic of the given sample is known as ***statistic***.

A **S**tatistic describes the **S**ample. Notice that both begin with the letter **S**.

Example2: - The average amount spent on textbooks by Unity University College students is unknown. To estimate this average 500 students were surveyed, and the average amount is determined to be 160 birr per term. In this scenario, the 160 birr is a *statistic*.

1.6 Types of Statistics

Definition 1:- The phase of statistics, which seeks only to describe and analyze a sample without drawing any conclusion about a population, is called ***descriptive statistics***.

Descriptive statistics is what most people think when they hear the word *statistics.* It includes the collection, presentation and description of sample data. As the name suggests, the descriptive statistics merely describe the data and consist of methods and techniques used in collection, organization, presentation and analysis of data in order to describe the various features and characteristics of such data. These methods can be either graphical or computational. Thus

data can be presented in the form of chart or table in order to show certain trends, proportions, maximum and minimum values and so on. In addition to the organization of data, the field of descriptive statistics is concerned with the analysis of data so that the data can be easily understood. Averages, proportions, and other measures that describe the spread of data around the average are also some of the measures used to describe the data. Note that even if data set has hundreds of scores, the average of those scores provides a single descriptive value for the entire set.

Example 2:- The following data refer to the number of malaria patients who have been treated in Arba Minch Hospital from 1994 to 1998(Ethiopian calendar).

No.	Year (Ethiopian Calendar)	Number of patients
1	1994	5026
2	1995	5873
3	1996	16512
4	1997	6774
5	1998	7580

Table 1:- **The number of malaria patients who have been treated in Arba Minch Hospital from 1994 to 1998 (Ethiopian calendar).**

i) If we calculate the average malaria patients from 1994 to 1998 as

$$\frac{5026+ 5873 +16512 + 6774 + 7580}{5} = 8353$$

then our work belongs to the domain of descriptive statistics.

If we say that the number of malaria patients in the year 1996 was the largest of all the remaining years in the collected data, then again our work belongs to the domain of descriptive statistics.

As we have said above data can be presented in the form of chart or table as the following figure shows.

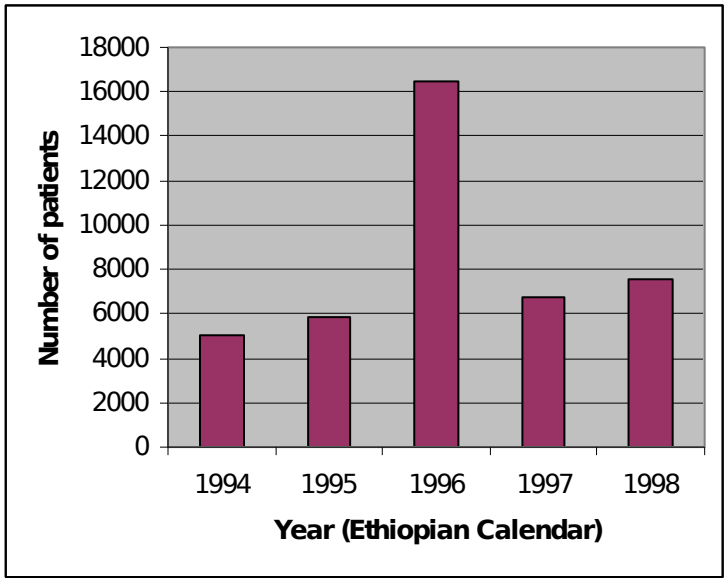


***Fig 1:-* Presentation of data in the form of tables and charts.**

Definition 2:- The phase of statistics, which deals with drawing important conclusions or predictions about a population from analysis of a sample, is called ***inferential statistics***.

The term ***inferential statistics*** refers to the technique of interpreting the values resulting from the descriptive techniques and making decisions and drawing conclusions about the population. That is statistical inference deals with methods of inferring or drawing conclusions about the characteristics of the population based upon the results of the sample taken from the population. The following examples are some of the situations that the field of inferential statistics deals with.

a) About 40% of students in the evening division of the higher institutions of Ethiopia are married. These statistics refer to the entire population of higher institution students in Ethiopia in the evening division. It would be reasonable to assume that these percentages were calculated on the basis of samples taken from the population of all higher institution

students in Ethiopia in the evening division. The students in these samples were asked in order to know as to how many of these students were married. The answers formed the basis for drawing conclusions about the entire population of the required students.

b) There is a definite association between alcoholism and liver cancer. This statement is the result of endless research on many samples taken and studied in order to find out if there was any correlation between alcoholism and liver cancer and based upon the results thus obtained from sample studies, a valid statement about the association of alcoholism with liver cancer in the whole population can be made.

c) In our example 2 given above, if we predict the number of malaria patients to be treated in Arba Minch Hospital in the year 2005(Ethiopian calendar) to be 11500, then our work belongs to the domain of inferential statistics.

Note that the question here is: How do these *sample statistics* relate to *population parameters*? Can we say that the conclusion drawn from the analysis of the sample are exactly the same as the conclusions that would be drawn from the entire population from which the representative sample was taken?

CHAPTER II

COLLECTION OF DATA

“Get the facts first, and then you can distort them as much as you please”- Mark Twain

2.1 Definitions

One of the first problems a statistician faces is that of obtaining data. Data doesn’t just happen; data must be collected. It is important to obtain a “good data” such that the inferences ultimately made will be based on the statistics obtained from the data. These inferences can only be as good as the data.

Definition 1:- *Collection of data* is gathering all the necessary information for the purpose of inquiry.

Example 1:- In science the data resulting from experiments must be collected and analyzed.

In education descriptive data are frequently used to describe test results.

In government many kinds of statistical data are collected all the time.

2.2 Scope of statistical investigation_

Definition 2:- A *population* is the set of all individuals of interest in a particular study.

Definition 3:- A *census* is a survey in which observation are made on the entire population. That is if every element in the population can be listed, or enumerated, and observed, then a census is compiled. A census is 100% survey.

Example 2:- If you wish to estimate the current average cost of text books per term, per student, then the population of interest is the currently enrolled student body. Thus the survey which can be observed is a census.

Although research questions concern an entire population, it usually is impossible for a researcher to examine every single individual in the population of interest. Therefore, researchers typically select a smaller, more manageable group from the population and limit their studies to the individuals in the selected group. In statistical terms, a set of individuals selected from a population is called a sample. A sample is intended to be representative of its population, and a sample should always be identified in terms of the population from which it is selected.

Definition 4:- A *sample* is a set of individuals selected from a population, usually intended to represent the population in a research study.

Example 3: A researcher may want to examine the attitudes toward abortion for women versus men. In this case a researcher may

not able to include all women and men in this particular area of interest.

One problem with using samples, however, is that a sample provides only limited information about the population. Although samples are generally *representative* of their populations, a sample is not expected to give a perfectly accurate picture of the whole population. There usually is some discrepancy between a sample statistic and the corresponding population parameter. That is any inference or conclusion based on limited information may not be accurate. This discrepancy is called *sampling error*, and it creates another problem to be addressed by inferential statistics.

Definition 5:- *Sampling error* is the discrepancy, or amount of error, that exists between a sample statistic and the corresponding population parameter.

2.3 Advantages and Disadvantages of census and sample survey

A census is a 100% survey. Its advantage is that since it considers every member of the entire group it is somehow “accurate” than the sample survey. However, census are seldom used because they are often very difficult and time consuming to compile, and therefore, very expensive. Moreover although research question concern an entire population, it usually is impossible for a researcher to examine every single individual in the population of interest.

On the other hand a statistic obtained from a sample generally will not be identical to the corresponding population parameter. Because the sample is only a small part of the total population, you should not expect the sample value to be equal to the population parameter. A statistic has always some “marginal error,” which is defined as sampling error.

2.4 Sampling Techniques

2.4.1. Introduction

Another concern for researchers using the survey method is who will participate in the survey. For the results to be meaningful, the individuals who make the survey should represent the population under investigation. As we have discussed above, the population consists of all of the people about whom a study is meant to generalize, whereas the sample represents

the subset of people from the population who actually participate in the study. In almost all cases, it is not feasible to survey the entire population Instead; we select a subgroup or sample from the population and give the survey to them. If we are to draw any reliable and valid conclusions concerning the population, it is imperative that the sample be “like” the population – *a representative sample.* When the sample is representative of the population, we can be fairly confident that the results we find based on the sample also hold for the population. In other words, we can generalize from the sample to the population.

2.4.2 Types of Sampling

As we have discussed earlier inferential statistics are a necessary part of conducting research. Suppose a researcher would like to test the effectiveness of a new therapy program for depression. It would be too costly and time–consuming to test the program on all depressed individuals. Therefore, the treatment is tested on a sample of depressed patients. The investigator would like to generalize the findings to the entire population. If the treatment program is effective for the sample of depressed people, it would be great to be able to state with confidence that it will also work for others. It is important to note that inferential statistical methods will allow meaningful generalizations only if the individuals in the sample are representative of the population. To do so we are using two types of sampling methods: Random sampling methods and non random sampling methods.

a) Random sampling methods

One way to ensure that the sample is representative is to use *random sampling*, or *random selection* method. Although there are several formally defined techniques for random sampling, one basic requirement is that all individuals in the population have the same chance to be selected.

Definition: - *Random sampling,* or *random selection*, is a process for obtaining a sample from a population that requires that every individual in the population have the same chance of being selected for the sample. A sample obtained by random selection is called a random sample.

Here we are going to see four types of random sampling methods, namely: simple random sampling, stratified random sampling, systematic random sampling, and cluster random sampling.

i. Simple Random Sampling

Simple random sampling is a technique in which sample is drawn that each and every unit in the population has an equal chance of being included in the sample. In this type of sampling every item is given an opportunity of being selected. Which item is selected is just a matter of chance for no personal bias is at work here.

A random sample may be selected by (i) Lottery method, (ii) Table of Random Numbers.

***Example*:-** Among 500 electric bulbs produced by a company, an expert wants to determine the percentage of defective items by taking a random sample of 30 bulbs.

One way to pick a random sample is to number the bulbs 1 through 500. Write the numbers 1,2,3,.....,500 on cards or pieces of papers, put them in a box, shake the box to mix up the papers and then draw 30 numbers.

Another way of obtaining such a sample is by using table of random numbers. The numbers on the slips of papers would indicate which element in the population is included in the sample. Although this procedure is straight forward, we could run into the problem if the number of elements in the population is large or if the slips of paper are not well mixed. To avoid such problems, statisticians frequently use a table of random numbers in choosing a simple random sample.

Random numbers are developed in such a way that each of the numbers from 0-9 has the same chance of occurring. Consider the following random numbers.

57728	16308	27337
63962	45765	75060

51041	22252	38447
84048	53578	67379
17736	34458	67227

Several points should be noted concerning this procedure.

1. If any previously chosen number comes up again, it should be ignored. For example if 048 were to come up again before 30 bulbs were chosen, it should be ignored.
2. If a number comes up which is not admissible because it doesn't correspond to any item in the population, it also should ignored. For example 728, 962, 736 and so on must be ignored.
3. It makes no difference whether you read the table downward, upward, to the left, to the right or diagonally. As long as you proceed from number to number in a systematic fashion, the table will work perfectly.

Merits of random sampling method: Sampling through random numbers is a probability sampling and as such it is free from any bias arising out of personal judgment or discretion of the investigator. The results therefore are more representative of the population.

ii. Systematic Random Sampling

Another popular method of sampling is called *systematic sampling*. In this method, it is assumed that the elements of the population arranged in some natural sequential order. Then we select a (random) starting point and select every k^{th} element for ur sample. For instance, people lining up to. Buy Theater tickets are "in order." To generate a systematic sample of these people (and ask questions regarding topics such as age, smoking habits, income level, etc.), we would include every 5^{th} person in line. The "starting" person could be selected at random from the first five.

The advantage of systematic sample is that it is easy to get. However, there are dangers in using systematic sampling. When the population is repetitive or cyclic in nature, systematic sampling should not be used. For example, consider a fabric mill that produces dress material. Suppose the loom that produces the material makes a mistake every 15^{th} meter, but we check only

every 14^{th} meter with an automated electronic scanner. In this case, a random starting point may or may not result in detection of fabric flaws before a large amount of fabric is produced.

In general we can use the following three steps in using systematic sampling method.

1. Calculate the ratio N/n, where N is the number of elements in the population and n is the number of elements in the sample. (sample size)
2. Let K be the largest integer less than or equal to N/n. Randomly select one element from the first K elements.
3. Select every K^{th} element thereafter.

Example:- Consider the example on the 500 electric bulbs.

N= 500, n = 30 ⇒ N/n = 500/30 = 16 2/3
The largest integer less than or equal to 16 2/3 is 16 ⇒ K= 16

Then from the first 16 bulbs, choose one randomly. Suppose the 9^{th} bulb is chosen , then we shall have the 9^{th} , the $(9+16)^{th} = 25^{th}$, $(25+16)^{th} = 41$, $(41+16)^{th} = 57^{th}$, and so on.

- As we have discussed above we don't use this method if there is a hidden irregularity.

Example: - If every 16th bulb is defective due to a regularly recurring failure of the machine, then the sample will be misleading.

iii. Stratified Random Sampling

When sampling very large populations, sometimes it is possible (and helpful) to divide the population into subpopulations on the basis of some characteristic. These sub populations are called ***strata.*** These smaller, easier-to-work-with strata are sampled separately. One of the sample designs that starts by stratifying the sampling frame is the *stratified random sampling method*.

Thus a stratified random sample is a sample obtained by stratifying the sampling frame and then selecting a fixed number of items from each of the strata by means of simple random sampling method.

For example, in the population of all undergraduate college students, some strata might be freshmen, sophomores, juniors, and seniors. Other strata might be male and female, day time students and evening students, and so on. In the method of stratified sampling, the population is divided into at least two distinct strata. Then a (simple) random sample of a certain size is drawn from each stratum, and the information obtained is carefully adjusted or weighted in all resulting calculations. Note that in using stratified random sampling method the groups or strata are often sampled to their actual percentages of occurrence in the overall population.

For example suppose in the population of 10,000 students, we want to use a sample of 1000 students. If we decide to use stratified random sampling method and divide the group into male and female students, where the number of female students in the population is 6,500 and that of male students is 3,500, then first we have to decide that the proportion of male and female students in the sample should be the same as that of the population. Thus in this case the number of female and male students to be included in the sample will be 650 and 350 respectively.

Example: - In a factory 30% of the workers are illiterate, 55% can only read and write and 15% are college graduates. In order to obtain some information about their salary, a sample of 400 workers is selected. Here the population is divided into three different groups. Illiterate, those who can only read and write and college graduates. Thus we have three strata. Because of proportionality, we must select:

$\frac{400}{100} \times 30 = 120$ from illiterates

$\frac{400}{100} \times 55 = 220$ from those who can only read and write

$\frac{400}{100} \times 15 = 60$ from college graduates

STRATA

I (30%)	R&W (55%)	C .G 6(15%)	Total
120	220	60	400

Therefore, we use stratified random sampling to select 120 from the illiterates, 220 from those who can only read and write and 60 from college graduates.

There is also a formula to determine the number of elements in each stratum.

$$n_i = \frac{n \times N_i}{N}$$

Where :- n – the total sample size

n_i – stratum sample size

N_i – stratum population size

N – the total population size

Example:- The city council of Addis Ababa has 250 seats. If the seats are shared to the 5 different zones of the city proportional to the population and the number of population is 3,000,000, determine the number of population size taken from each zone.

n= 250, N_1 = 650,000, N_2 = 850,000, N_3 = 560,000

N = 3,000,000 N_4= 500,000, N_5 = 440,000

$n_1 = \frac{250 \times 650{,}000}{3{,}000{,}000} = 54$ $n_2 = \frac{250 \times 850{,}000}{3{,}000{,}000} = 71$

$n_3 = \frac{250 \times 560{,}000}{3{,}000{,}000} = 47$ $n_4 = \frac{250 \times 500{,}000}{3{,}000{,}000} = 42$ $n_5 = \frac{250 \times 440{,}000}{3{,}000{,}000} = 37$

iv. Cluster Random Sampling

Cluster sampling is a method used extensively by government agencies and certain private research organizations. In cluster sampling we begin by dividing the demographic area into sections. Then we randomly select the sections or clusters. Every member of the cluster is included in the sample. For instance, in conducting a

survey, of school children in a large city, we would first randomly select 5 schools and then included all the children from each selected schools.

(b) Non Random sampling methods

In section 2.4.2 (a) we have discussed different kinds of random sampling techniques. But not all sampling plans, however, involve random selection. In nonrandom (or non probability) sampling, not every unit in the population has a chance of being included in the sample, and the process involves at least some degree of personal subjectivity instead of following predetermined, probabilistic rules of selection. In our discussion we will examine four types of non probability samples:

i) Convenience Sampling: This sampling technique simply uses results or data that are conveniently or readily obtained. In some cases this may be all that is available, and in many cases, it is better than no information. Members of convenience sample are chosen primarily because they are both readily available and willing to participate. For example, a fellow student may ask members of your statistics class to fill out a questionnaire for a study she is doing as an assignment in her *project management* course. So – called – person – on – the – street interviews by television reporters are often identified as "random", although they are usually conducted at shopping areas near the station or at other locations selected primarily because they happen to be convenient for the interviewer and camera crew.

ii)Quota Sampling: This is similar to the stratified probability sample described previously, except that members of the various strata are not chosen through the use of a probability sampling technique. That is in a quota sample the population as a whole is split into a number of groups or strata (as in the case of stratified sampling method) and whoever is charged with drawing the sample is instructed to include a certain number of members of each group. The reason why quota sampling is a form of judgment sampling is that the choice of members from each

group or stratum to be included in the sample is not determined by random selection, but by investigators.

iii)Judgment Sampling: This sample is selected on the basis that the researcher believes the members to be representative of the population. As the result, the representativeness of such a sample is as good as the judgment of the person who has selected the sample.

iv)Purposive Sample: In this sample, members are chosen specifically because they're not typical of the population. For example, the maker of a new pocket calculator may submit prototypes to members of an advanced doctoral seminar in mathematics at Washington Institute of Technology. If these individuals can not figure out how to operate it, the calculator should obviously be simplified. Conversely, if sample of individuals known to have IQs below 80 are able to operate it, it should present no problem to the general population.

Chapter IV

Measures of Central Tendency

4.1. Definition and Purposes of Averages

The goal in measuring central tendency is to describe a group of individuals (more accurately, their scores) with a single measurement. Ideally, the value we use to describe the group will be the single value that is most representative of all the individuals.

Definition: Central tendency is a statistical measure that identifies a single score as a representative for an entire distribution. The goal of central tendency is to find the single score that is most typical or most representative of the entire group.

Usually, we want to choose a value in the middle of the distribution because central scores are often the most representative. In everyday language, the

goal of central tendency is to find the “average” or “typical” individual. This average value can then be used to provide a simple description of the entire population or sample. For example, archeological discoveries indicate that the average height for men in the ancient Roman city of Pompeii was 5 feet 7 inches. Obviously, not all the men were exactly 5 feet 7 inches, but this average value provides a general description of the population. Measures of central tendency also are useful for making comparisons between groups of individuals or between sets of figures. For example, suppose that weather data indicate that during the month of December, Seattle averages only 2 hours of sunshine per day, whereas Miami averages over 6 hours. The point of these examples is to demonstrate the great advantage of being able to describe a large set of data with a single, representative number. Central tendency characterizes what is typical for a large population and in doing so makes large amounts of data more digestible. Statisticians sometimes use the expression “number crunching” to illustrate this aspect of data description. That is, we take a distribution consisting of many scores and “crunch” them down to a single value that describes them all.

Unfortunately, there is no single, standard procedure for determining central tendency. The problem is that no single measure will always produce a typical, representative value in every situation. Therefore, there are three different ways to measure central tendency: the mean, the median, and the mode. They are computed differently and have different characteristics. To decide which of the three measures is best for any particular distribution, you should keep in mind that the general purpose of central tendency is to find the single most representative score. Each of the three measures we shall present has been developed to work best in a specific situation. We will examine this issue in more detail after we define the three measures.