CHAPTER 1

Basic Concepts In Statistics

1.1 Statistical Concepts

Our life is full of events and phenomena that enhance us to study either natural or artificial phenomena could be studied using different fields one of them is statistics. For example, the applications of statistics are many and varied as follows:

-People encounter them in everyday life

-Reading newspapers or magazines,

-Listening to the radio, or watching television.

Since statistics is used in almost every field of endeavor, the educated individual should be knowledgeable about the vocabulary, concepts, and procedures of statistics.

Definition 1.1.1

Statistics is a branch of science dealing with collecting, organizing, summarizing, analysing and making decisions from data.

Statistics is divided into two main areas, which are **descriptive** and **inferential** statistics.

A Descriptive Statistics

Suppose that a test in statistics course is given to a class at KSU and the test scores for all students are collected, then the test scores for the students are called data set (the definition of this term will be discussed deeper in section 1.2). Usually the data set is very large in the original form and it is not easy to use it to draw a conclusions or to make decisions while it is very easy to draw conclusions from summary tables and diagrams than from such original data. So reducing the data set to form more control by constructing tables, drawing graphs and provide some numerical characteristics for which is a simple definition to introduce descriptive statistics.

Definition 1.1.2

Descriptive statistics deals with methods for collecting, organizing, and describing data by using tables, graphs, and summary measures.

B Inferential Statistics

The set of all elements (observations) of interest in a study is called a population, and the selected numbers of elements from the

population is called a sample. In statistical problems we may interest to make a decision and prediction about a population by using results that obtained from selected samples, for instance we may interest to find the number of absent students at PY on a certain day of a week, to do so, we may select 200 classes from PY and register the number of students that absent on that day, then you can use this information to make a decision. The area of statistics that interest on such decision is referred to inferential statistics.

Definition 1.1.3

Inferential statistics deals with methods that use sample results, to help in estimation or make decisions about the population.

During this section, we will clarify the meaning of population, sample, and data. Therefore, the understanding of such terms and the difference between them is very important in learning statistics. For example, if we interest to know the average weights of women visited diet section in a hospital during specified period of time, then all women who visited that section represents the study population.

Definition 1.1.4

A population is the set of all elements (observations), items, or objects that bring them a common recipe and at least one that will be studied their properties for a particular goal. The components of the population are called individuals or elements.

Remark

Note that a population can be a collection of any things, like Ipad set, Books, animals or inanimate, therefore it does not necessary deal with people.

Any collection of things, including a joint gathering recipe at least one to be examined for a particular purpose, called a statistically population (or population as a matter of shortcut). The components of the population are called individuals or elements.

Example 1

a. In a study of the average number of students in secondary schools in Riyadh city, where there are different stages of the students, such as first, second and third secondary, as well as there are male and female, but they all gathered, including prescription study in high school. Therefore, we find that high school students in Riyadh make up a population.

b. In a study of the evolving condition of the patients in a hospital, where there are many people of different types of diseases, but they all bind them recipe disease, so patients that in the hospital make up a population.

c. In a study to determine the technical condition of the aircraft of the Gulf Cooperation Council (GCC), where travel aircraft, military training aircraft, and Helicopters ...., but they are all characterized by their ability to fly, so the aircraft in the Gulf Cooperation Council (GCC) are population

Note that a population can be a collection of any things, like set of trees, people, animals or inanimate (books, cars, metal...). Therefore it does not necessary deal with a people.

Definition 1.1.5

A sample is a subset of the population selected for study.

Referring to the example of interest to know the average weight of women that visited diet section, in this case the registered weights of some women represent a sample.
In practical life there are many ways to get a sample from the population under study, for example; face-to-face interview, online electronic questionnaires, paper questionnaires and using telephones.

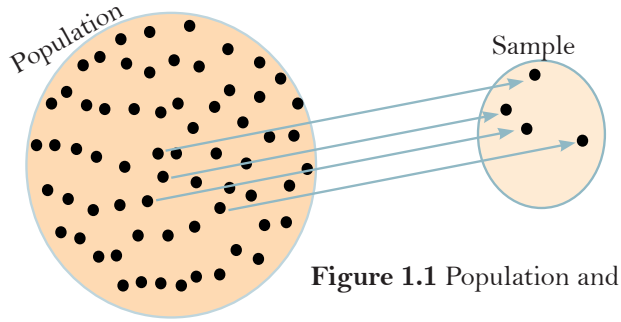

Figure 1.1 Population and Sample

Let us discuss an example on determining the population and the sample for a study

Example 2

If we take a class of the students in Stat 140 course at PY, then: all the students registered in the course represent the population, and any class of them is represent a sample.

1.2 Variables and Types of Data

Basic terms that will be used frequently in this section, and they are very important tools in statistical problems, such terms are, an **element**, a **variable** and their types, a **measurement**, and a **data set,** Therefore to understand such terms, it is necessary to illustrate the following definitions.

Definition 1.2.1

An element (or member of a sample or population) is a specific subject or object about which the information is collected.

Example 1 below discuss the definition of an element numerically

Example 1

The following table gives the number of snake bites reported in a hospital in 3 cities (A, B, C).

City	Number of Snake Bites
A	10
B	17
C	11

Each one of the cities is a member, that is; city A is a member, city B is a member, and also city C is a member.

Definition 1.2.2

A variable is a characteristic under study that takes different values for different elements.

For example, if we collect information about income of households, then income is a variable .These households are expected to have different incomes; also, some of them may have the same income. Note that a variable is often denoted by a capital letter like X, Y, Z, ... and their values denoted by small letters for example $x, y, z, ...$

Definition 1.2.3

The value of a variable for an element is called an observation or measurement.

The following is an example to explain the difference in the meaning between variable and the measurement.

Remark

Any study is based on a problem or phenomenon such as heavy traffics, accidents, rating scales and grades or others. The researcher should define the variables of interest before collecting data.

Example 2

Referring to example 1, we see that the variable (let X for example) is the number of snake bites and each one of the number of bites 10, 17, 11 represents an observation or measurement. Where we have $X(A)=10$, $X(B)=17$ and $X(C)=11$.

We know that the variable is a characteristic under study that takes different values for different elements. In statistics, we have two types of variables according to their elements; first type is called **quantitative** variable and the second one is called **qualitative** variable.

When a subject can be measured numerically such as (the price of a shirt), then the subject in this case is quantitative variable. The following definition provides us with this concept.

Definition 1.2.4

Quantitative variable gives us numbers representing counts or measurements.

When a subject cannot be measured numerically such as (eye color), then the subject in this case is qualitative variable. The following definition provides us with this concept.

Definition 1.2.5

Qualitative variable (or categorical data) gives us names or labels that are not numbers representing the observations.

Remark

Quantitative variables give us quantitative data and inquires about the phrase "how much", while the qualitative variables give us the qualitative data and inquires about the phrase "what or what is".

The following examples illustrates the two type of variables

Example 3

The following table shows some examples of the two types of variables

Quantitative variable gives us quantitative data	Qualitative variable gives us qualitative data
The age of people in years 19, 2, 45, 23, 88, ...	The gender of Organisms Male, Female .....
Number of children in family 5, 2, 4, 1, 14, ...	Results tossed a coin twice *HH, HT, TH, TT* (*H*=Head, *T*=Tail)
The heights of buildings in meters 15, 5.6, 12.7, 105, 27, ...	Eye color of people Black, Brown, Blue, Green, ...
The weights of cars in tons (ton=1000 Kg) 2.35, 1.65, 2.05, 2.10, 1.30, ...	Religious affiliation Muslim, Christian, Jew, ...
The speed of a car going on a main road in Km 110, 105, 85, 120, 90, ...	The pressure in a boiler High, Moderate, Low

Moreover, the variables measured in quantitative data divided into two main types, **discrete** and **continuous**. A variable that assumes countable values is refer to discrete variable, otherwise the variable is a continuous one. Accordingly, we provide the following definitions.

Definition 1.2.6

Discrete variables assume values that can be counted.

In following we illustrate some examples on a discrete variable

Example 4

-The number of children in a family, , where we have 1,2,3, ... or k children.

-The number of students in a classroom, where we have 21, 25,32,18 and so on.

-Number of accidents in a city, where we have 1,2,3,... or k accidents.

The other type of quantitative variable is the continuous variable which is assumed uncountable values, and offer us the following definition.

Definition 1.2.7

Continuous variables assume all values between any two specific values, i.e. they take all values in an interval. They often include fractions and decimals.

In the following we illustrate some examples on a continuous variable

Example 5

- Temperature: For example the temperature in Riyadh city in last summer was between 15 and 56, i.e. the temperature $t \in [15, 56]$.

- Age: For example the age of a horse is between 0 (Stillborn) and 62 years (Said the oldest horse was 62 years, but the middle age of a horse is 30 years), i.e. the age of a horse $x \in [0, 62]$

- Height: For example the height of a student in a Country is between 110 cm (person elf) and 226 cm (person giant), i.e. the height of a student $x \in [110, 226]$

Variables classified according to how they are categorized or measured. For example, the data could be organized into specific categories, such as major field (mathematics, computers, etc.), nationality or religious affiliation. On the other hand, can the data values could be ranked, such as grade $(A,\ B,\ C,\ D,\ F)$ or rating scale (poor, good, excellent), or they can be classified according to the values obtained from measurement, such as temperature, heights or IQ scores. Therefore we need to distinguish between them through the measurement scale used. There are four levels of measurement **scales**; **nominal**, **ordinal**, **interval**, and the **ratio** level of measurement, the difference between these four levels is explained in the following definitions.

Definition 1.2.8

The nominal level of measurement classifies data into mutually exclusive (disjoint) categories in which no order or ranking can be imposed on the data.

The following examples include nominal level of measurements in different cases.

Example 6

- Gender: Male, Female.
- Eye color: Black, Brown, Blue, Green, ...
- Religious affiliation: Muslim, Christian, Jew, ...
- Nationality: Saudi, Syrian, Jordanian, Egyptian, Pakistani, ...
- Scientific major field: statistics, mathematics, computers, Geography, ...

When the classification takes ranks into consideration, the ordinal level of measurement is preferred to be used. The following definition provided us this concept.

Definition 1.2.9

The ordinal level of measurement classifies data into categories that can be ordered, however precise differences between the ranks do not exist.

The following examples include some ordinal level of measurements.

Example 7

Grade $(A,\ B,\ C,\ D,\ F)$: Grading technique is the most common example on ordinal level. For example we find that the system of appreciation in Saudi universities are (in descending order) $A^+,\ A,\ B^+,\ B,\ C^+,\ C,\ D^+,\ D,\ F$.

- Rating scale (bad, good, excellent and so on ...): To test the quality of the canned product, we find that the state of the tested object either excellent or good or bad.

- Ranking of football players: A football player can be ranked in first grade, second grade, third grade, ...

- Ranks of university faculty members: Academic ranks usually classified as professor, associate professor, assistant professor , and instructor.

The third level of measurement is called interval level. The following definition provided us this concept.

Definition 1.2.10

The interval level of measurement orders data with precise differences between units of measure. (in this case there is no meaningful zero). On the other hand, the resulting measurement values belong to an interval of the real numbers.

Example 8

- **IELTS:** An International English Language Testing System. ILETS is an international system to test the English language in order to study and work. The degree x to which the grant will be between zero and 9, i.e. $x \in [0,9]$

-**TOEFL:** Test of English as a Foreign Language. TOEFL is a standardized test of English language proficiency for non-native English language speakers wishing to enroll in some universities in the world.

- **SAT** score and IQ test: The SAT is a standardized test widely used for college admissions in some universities in the world. It is a good predictor of a student's performance in the first year of college. Degree to which the grant will be between zero and 2400, i.e. $x \in [0,2400]$

- **Temperature:** When the degrees of temperatures are measured in Celsius or Fahrenheit, then the values that we obtain from absolute zero (-273.15 but without this degree) extends to millions as is the case in the sun and stars.

Note that if we compare between two temperature degrees, like 30°C and 60°C we can't say that 60°C is as high as twice the degree 30°C; but we can say there is a 30°C difference between them. In the sense that we can not be compared to some of the quantities to others in this case. On the other hand, if the temperature of something equals to zero, that does not mean it does not have a temperature.

That mean in the Celsius temperature the zero means there is a temperature and it is very cold that is the zero does not mean nothingness. The data at this level do not have a natural zero starting point. The measurements that rely (or that adopt) zero as starting point called ratio level and offered us the following definition