Welcome students and viewers unit 1 basic concepts and statistics. The module that I'll be taking up today is introduction to statistics module #1. The brief outline of this particular module is an introduction followed by definitions of functional classification of Statistics. The scope of statistics and its limitations. The learning outcomes students would be able to understand the field of statistics cite commonly used definitions, differentiate statistical methods based on function, and describe the scope and limitations of statistics. Have you ever wondered? How do experts forecast the weather? How do insurance companies decide the risk associated with us? How do experts claim stock prices might rise or fall? Have you ever wondered how pharmaceutical companies decide if a drug is effective? How do you so often find relatable ads popping up while surfing the Internet? Or how do teams decide what strategies to use against which team? Well, I'm hoping at least by the end of this course you will have some idea as to how these questions are answered. Let's begin. Statistics as a discipline. Is that which concerns with the collection, organization, analysis, interpretation and presentation of data? Statistics can be basically classified into a plural form and singular form. When you talk about plural form, you're talking about numerical data that people attribute two. Somebody says the unemployment rate is 2.5%. Somebody says inflation rate is 5.6%, but these are facts and figures. But these facts and figures have to eventually arise from some kind of data set. And that brings us to the second part. That's the singular form of statistics, which basically concerns statistical methods. And these methods are a variety of types. It can be relating to collection of data. It can be relating to an analysis, interpretation or graphical presentation of data for that matter. So let's begin. In the beginning of definitions, we have the definition of the plural sense. Statistics may be defined as the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated, or estimated according to a reasonable standard of accuracy collected in a systematic manner for a predetermined purpose and placed in relation to each other. This is the definition of statistics in the plural sense provided by Professor Horace Secrist. In the singular sense. We have a couple of definitions here, one from Croxton and Cowden, and the other from Wallace and Roberts. The first one goes statistics may be defined as the science of collection, presentation, analysis, and interpretation of numerical data. This one is by Croxton and Cowden, and the second one is by Wallace and Roberts, which goes as follows. Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainty, and we all know that life is very uncertain similarly. When we involve in any kind of activity, be it business or be it consumer choice, there is always an uncertainty associated with it. Hence the importance emerges of statistics. Let's classify statistics now based on a couple of different measures. Statistics can be classified into descriptive statistics or inferential statistics. Descriptive statistics basically describes the sample data. Assume you had to describe in particular an individual. You would basically talk about that person's height, person’s weight, etc., right? Maybe the color of his eyes even. Similarly, when you want to describe a data set, you have something called Descriptive Statistics. We'll see what that is in a couple of minutes. On the other side, you have inferential statistics. Basically, you infer and make inferences about a population when you have sample information with you. I know it's a bit confusing right now. Let's hold that thought for a bit and we'll get there eventually. So what is population and sample? Can you see the large circle on your screen with all those red and orange dots there, right? Suppose that is your whole population and it consists of a number of individuals. And samples are basically subsets or subunits that you pick out from this population based on some signs of criteria. Now in this paper we don't actually deal with the way we choose the samples, but

we deal with what kind of analytical tools we have. Once we have chosen the samples. Let's talk about population. The population is the entire set of information, or subjects or units that you are planning to study, and the statistics that you emerge out of these populations are called as parameters and the alphabets that we use to describe these parameters are Greek alphabets, for example mu. On the other side, suppose we do not have information about the population, but we use a sample. Now, the statistics that we find from this sample are called sample statistics and we denote them with English alphabets as denoted here X Bar. So then let's take an example. Consider we would like to determine the age of individuals working for a small firm. In this case, the population will consist of all individuals working in that form. Assume this population is 100. We randomly decide to select10 individuals from this population to actually find the average age. Thus the sample will consist of 10randomly selected individuals working in that particular form. Once again, come back to the definition. What is the classification? We spoke about statistics basically divided into descriptive statistics and inferential statistics. Descriptive statistics basically tells you and describes the particular sample data for you, and we use these sample statistics to make inferences about population using the second classification, which is inferential statistics. Let's look at some measures of descriptive statistics. We have the first one measures of central tendency. Quite a few of you would be familiar with this. The second one is measures of dispersion under measures of central tendency. We have the mean median, mode under measures of dispersion. We have range quartile deviations, mean deviations, variance, and standard deviation. Also a third one that some of you might not be familiar with is the measures of asymmetry we cover under this skewness. Sources or moments. Let's take an example. A sample of 10 individuals is chosen at random. You have a data set for their age. Before you now 25, 51, 41, 24 and 47 and so on and so forth. The total of age comes to 333. The number of observations is 10, so let's see what this sample statistics looks like. Before we head on the just a recollection of facts for you, this particular is this particular data you're looking at is a sample data. Because we've chosen 10 individuals from the population of 100. As we've seen in the previous slides. Therefore, the sample statistics covers all those things that we described in the previous slide. The mean the standard error, the median, the mode, the standard deviation, sample the variance. You could pause here and have a look at these statistics that are there on your screen. We also have other sample statistics like kurtosis, skewness, range and also give you the minimum and maximum values in your data set. Moving on. Let's look at what are some of the inferential statistical methods that we have. The first one is the hypothesis testing under hypothesis testing. We have parametric tests as well as nonparametric tests and what you will see when we do hypothesis testing in the modules to follow is basically distributions like this image that you've seen below right just now. Secondly, we'll be dealing with correlation analysis, and we will also be looking at regression analysis. And just a visual image for you of things to come in the modules to follow. So this is an overview of the process that I was just describing some time ago. The first step is the population you use sampling methods to find the sample . Look at your right side the sample box there. From the sample you find sample statistics and from sample statistics you find parameters and from parameters you describe the population. So ultimate goal here is to describe the population and you might understand this that information about the population is not that. Easily available and if even if it was, it would be time consuming and also very expensive and therefore researchers and research students use something called a sampling techniques to find a small subset of that population. And find the sample statistics which are then used to predict the population parameters in a way to describe them. Population now this process of moving from sample to sample statistics is

descriptive statistics. As we've seen a couple of slides ago and the process of moving from statistics to parameters is basically inferential statistics. What is the scope of statistics well? A few to list are here. Finance! You look at your TV screen, when you look at the finance news channel and you see a lot of those Red Arrows and green arrows popping up right? So when someone does says that they expect a particular stock price to rise or fall? How do you really know what's going to happen? Well, he probably bases that on some kind of statistics business analytics businessmen, an firms and large companies use statistics almost every day to look at. They are expenditure patterns. Look at the revenue patterns. Basically objective here to be maximizing profit or minimizing their cost in education. Would look at enrollment ratios. Will look at drop dropout ratios and so on and so forth. In marketing strategies we look at. Different elements in aspects of how people want to market their products and the demand for their particular products. In sports, people decide which player to play in their teams in a particular condition or not. Under econometrics we combine statistics and economic data to analyze and predict outcomes using models. Sciences also use statistics to a large extent, and of course planning and governance has statistics as an internal part to it. Let's look at the limitations. However, after all these benefits that I've listed down to you, and the things that we've spoken about. Statistics does not study qualitative phenomena. That means if you had something like gender, male female or you had color coding, it would not capture that. As a result, we have to use something called coding and therefore we have to be expressing this data qualitative data in numerical form and will be having a look at this in the models to follow as well. It does not study individual units. For example, if you look at a subset and find out one single element there, it won't be studying that it looks at Aggregates. Statistical laws are not exact. Because statistics is based on probability and everything. Any event that occurs is probabilistic in nature and therefore statistical laws are not exact and finally lights liable to be misused. Now this is a question for you to think about cause with any kind of responsibility that we as humans have. Statistics is a powerful tool in terms of determining facts but very well can be misused. Therefore the question of morality and ethics. Here are my references. Thank you very much for joining me in this video. Have a nice day.