**Business Statistics**
**Prof. M. K. Barua**
**Department of Management Studies**
**Indian Institute of Technology- Roorkee**

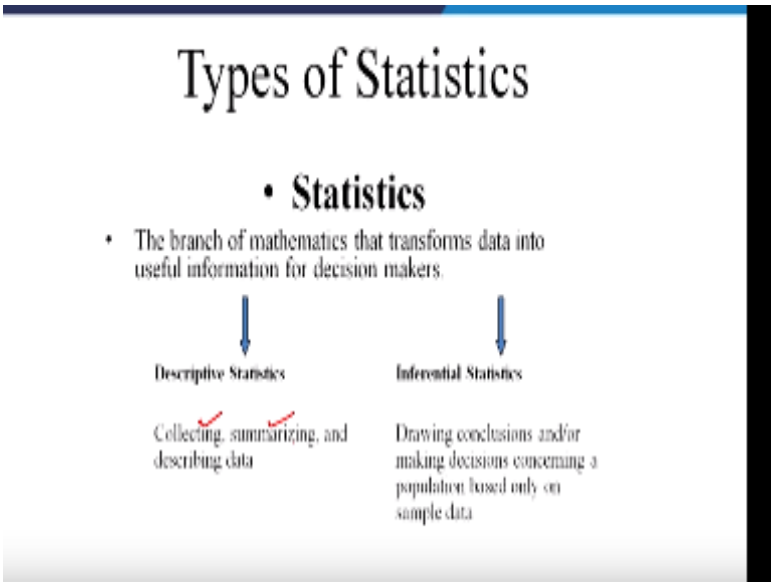
**Lecture – 02**
**Types of Statistics, Types of Data and Sources of Data, Population vs Sample**

Good afternoon friends in previous session of this particular course we discussed the importance of business statistics which would study business statistics why it should be studied. This course is useful for whom and as I said in previous session as well that the course is useful for B tech students, BBA students MBA students, part time professionals and those who are dealing with data.

We have also seen in previous session about types of statistics what care one should take if a person is dealing with secondary data. We have also seen definitions of data raw data, data set, elements and so on. In this class we are going to talk about descriptive statistics and inferential statistics. So first we will look at what are different types of descriptive statistics available to us. So descriptive statistics as I said in previous session as well.

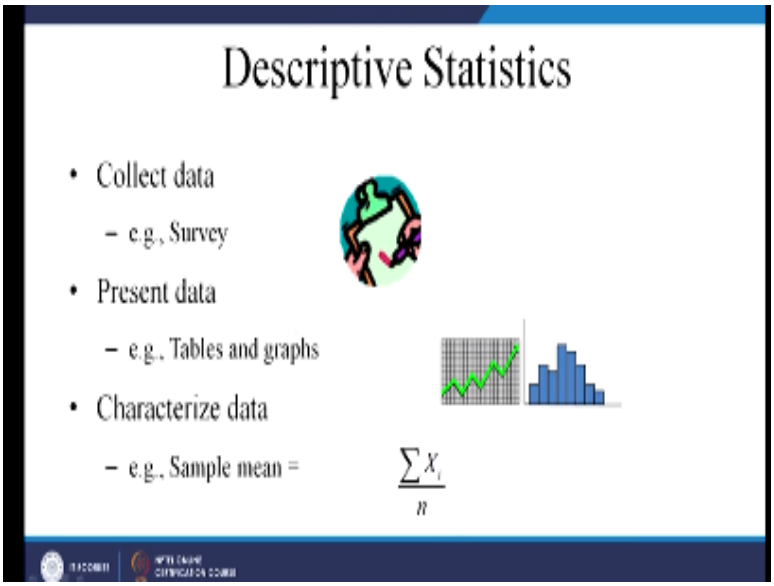
That it describes something offer product or of an individual or of an organization. So we collected data in descriptive statistics.

**(Refer Slide Time: 01:54)**



We summarize them and we analyse data after that we make a particular decision. Right so we will discuss more about descriptive statistics.

**(Refer Slide Time: 02:11)**

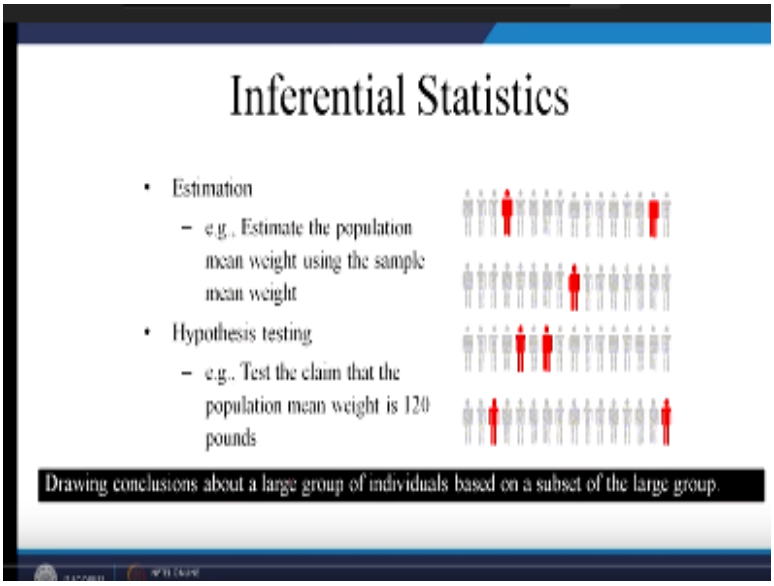



So as I said it just more of data collection so data collection can be through several ways. Or let us say survey can be through several methods. You can have let us say a form to form survey you can have one to one survey you can have survey at the mall. You can have a survey at the office you can have a survey at the home and there are several other methods of data collection. So once you have collected data then your responsibilities to present data in appropriate form.

So that you can infer something out of that data so either you come with tables, graphs, charts, pie charts scatter plots and so on and then try to go for some kind of analysis on data right? For example, you can simply do mean right you can calculate mean mode, median range and there are several other statistical analysis you can perform on data. As far as inferential statistics is concerned as I said in previous classes as well.

That on the basis of some simple we try to know the characteristics of population. And there are two types of inferential statistics.

**(Refer Slide Time: 03:46)**



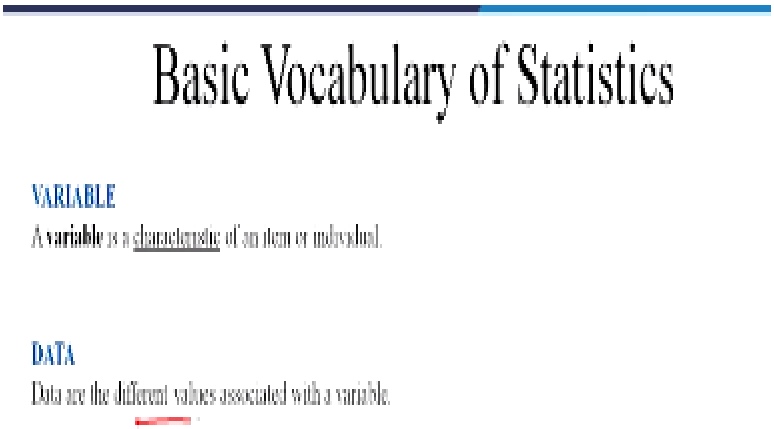

Estimation for example estimation of population mean through simple mean okay I will give you an example. Let us say there are 30 students in your class and you want to know the weight of average weight of the class. So what you should do you should take a sample from that class find out the mean and you should say that this mean is equivalent to the mean of the population right?

So that is estimation the other one is hypothesis testing. So there are two types of inferential statistics in estimation hypothesis testing. So here what we say we test a claim that the population mean weight is 120 pounds. So initially you what we are doing hypothesis testing we prepare a hypothesis we frame a hypothesis about population parameter. We say that because we do not know whether what is the mean weight of the population.

So we assume initially we say that it is 120 pound. Now it may be right or it may be wrong. It will be decided once you take sample from that population find out the weight. Okay so these are two broad inferential statistical tools estimation and hypothesis testing. So we are drawing conclusion about large group of individuals just based on subset of the large group that is subset is nothing but sample of the population.

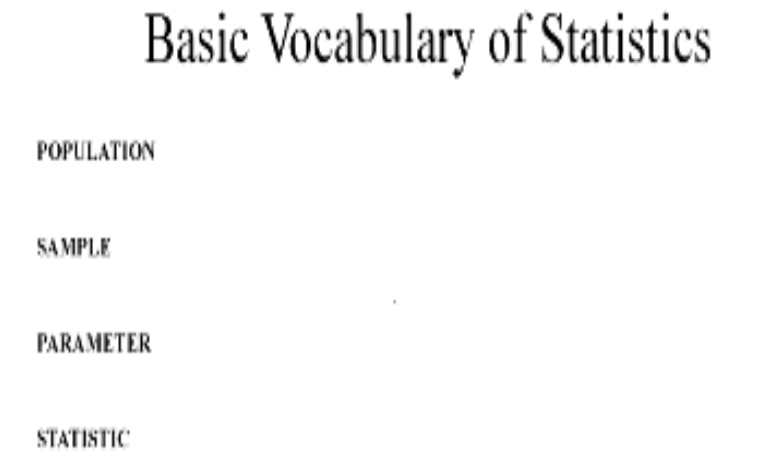
We also have seen a couple of definitions. So what is variable? It is a characteristics of an item or an individual. For example, height of students, weight of students and so on. So all these are nothing think but variables right?

**(Refer Slide Time: 05:48)**



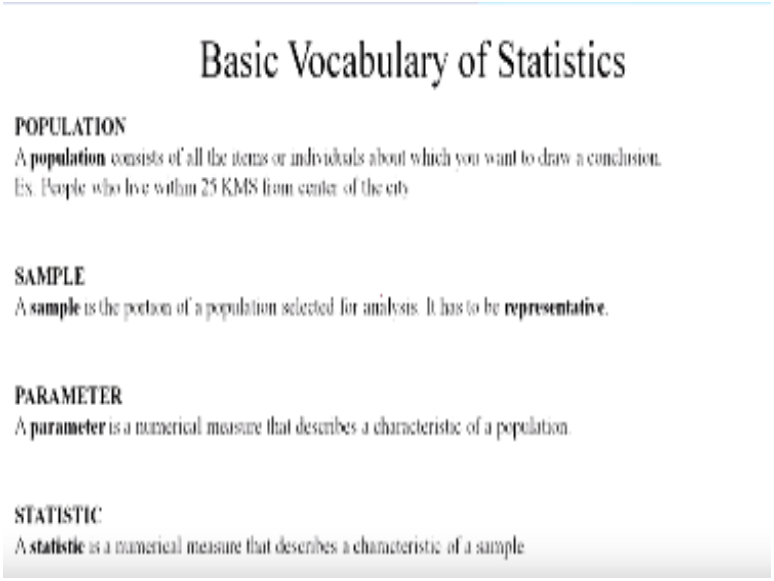

Data once again are different values are associated with a variable or with variables.

**(Refer Slide Time: 05:52)**




So let us look at some basic vocabulary of statistics. So you got population there is something called sample there is something called parameter and there is something called statistics. So what do you think? What is population? Just think for a while what is population?

**(Refer Slide Time: 06:19)**




So population is basically all the items or individuals about which you want to draw conclusion. For example, let us say you want to know the income of those people who live in 25 kilometres radius from the centre of the city. And the age of those people should be between 18 to 25 right? So all those people who are in this age group and live in this 25 kilometres of the radius that would be the population.

Let us say that population for simplicity is 100 so all this one and that individuals are population. Okay what does sample is the proportion of population selected for analysis. So well select some of the people who are here from 100 people so that we can know income of the entire population. So how to select sample it should be a representative sample. For example, in this 100 in this set of 100 people let us say there are 75 males and 25 females.

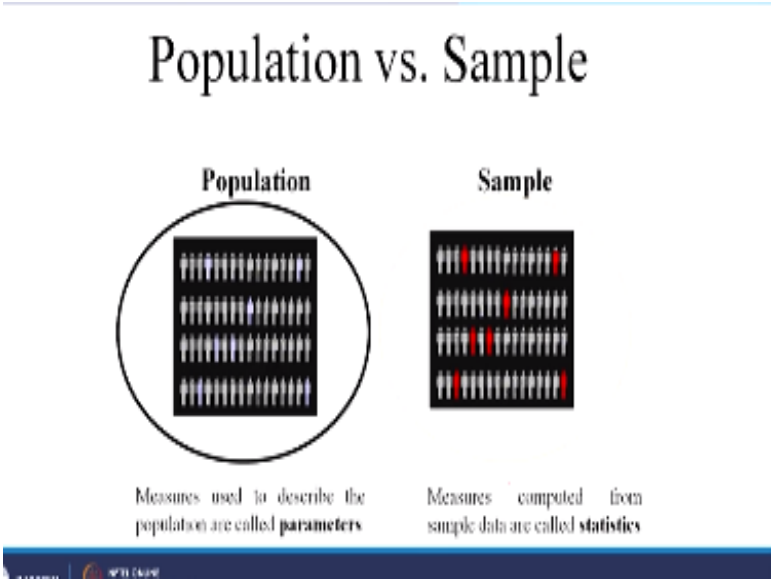
And let us say if you are taking a sample of for example let us say 20. So if you are taking 20 sample over here then how many should be male? Because 75% are male so 75 .75*20 right? So whatever is the number you will get right. Similarly, 0.25*20 so these many females you should select and these many males you should select. So that is why you will have a representative sample.

A representative sample means the sample should have the same characteristics which the population has got right? Parameter is a numerical measure that describes characteristics of the

population which the characteristics of the population not the sample. So it could be let us say population mean or it could be standard deviation of the population. Statistics is a measure that describes characteristics of simple.

So sample mean x bar sample standard deviation. This how we represent sample standard deviation. So population is about so parameter is about population and statistics is about sample right?

**(Refer Slide Time: 09:26)**




This is just one more slide on population sample. So this is your entire population so population measures use to describe the population are called parameters. So those measures which we use to describe population are called parameters and those measures which we used to describe sample are called statistics right? So this is your sample right.

**(Refer Slide Time: 09:51)**

| | Population | Sample |
|---|---|---|
| Definition | Complete enumeration of items is considered | Part of the population chosen for study |
| Characteristics | Parameters | Statistics |
| Symbols | Population size = N<br>Population mean = $\mu$<br>Population S.d = $\sigma$ | Sample size = n<br>Sample mean = $\bar{x}$<br>Sample S.d = s |

So definition if you take then this complete enumeration of items is population complete enumeration and part of it is sample. So this parameter this s statistics for example these are different symbols we use for population and sample. So this is capital and population size small and sample size population mean, sample mean standard deviation and sample standard deviation.

**(Refer Slide Time: 10:22)**

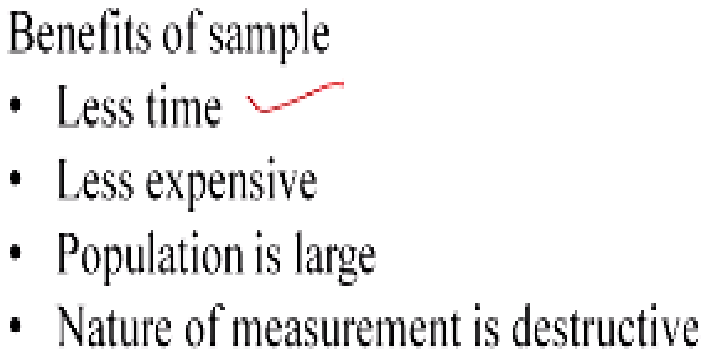
Which is better : Samplings or complete enumeration?

So which is better at the end of the day should you go for sampling or just complete enumeration or population there are several factors which will decide what decisions you should take. Let us say if your population is small let us say there are only five students in your class and if you want

to measure mean weight of the class and you can take weight of each individual and then just divide it by a 5 right?

So there is no need of sampling in that case because your population size is its too small right? But generally we take samples when your population is large. And in real life you ever large populations. So if you take samples from those populations or if you take a sample from a population then there are several advantages.

**(Refer Slide Time: 11:25)**




It saves time you can easily take few samples from population and perform whatever analysis you want to perform. Let us say if you want to measure height measure it. If you want to measure weight measure it right? It is of course less expensive because you are approaching only few ones from that population. It is good to go for sampling when your population is large as I have already mentioned.

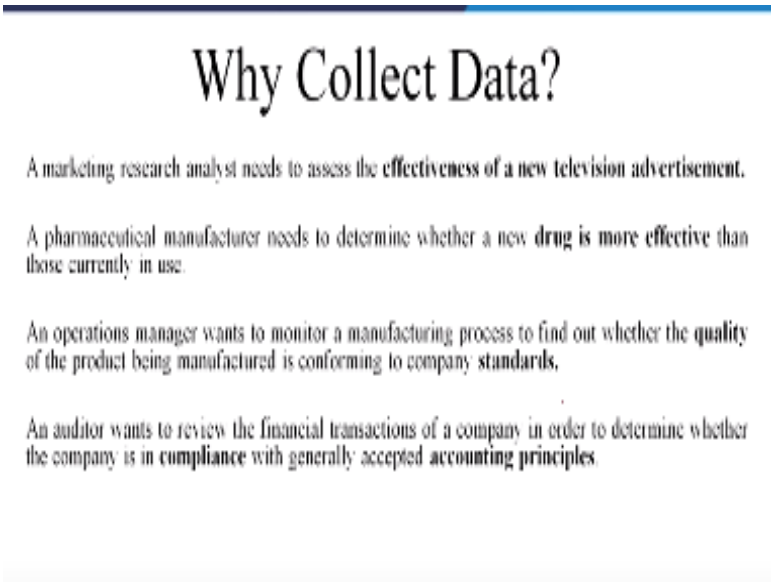
Now there is there is something a situation where the product is destructive in nature then you should go for sampling not the complete enumeration. So let us say I take you an example I will give you an example. Let us say you are manufacturing perfume and you are packing perfume in different bottles. So before supplying perfume to customers or let us say to your distributor are two retailer what you should do you should test the product.

You should test it for its quality so how to test quality of a perfume. You should you know take each and every bottle and spray it right? But if you do it then you would be testing or you would be wasting an entire perfume in testing itself. So there would not be any perfume to sell to the dealer or to distributor or to customer. So in that case if you produce let us say a 100 bottles every day you take a sample of 1 or 2 bottles and tested for quality.

So in situation like these you should go for sampling right? Another example let us say if you manufacture bricks and let us say you want to supply it to a builder for construction purposes. So if you want to check the strength of the bricks so what you would do you would not crush every brick is it not. Otherwise what you would supply to building you know contractor when that owner of the building.

So you should take a couple of bricks crush them for testing the strength. And then on the basis of that you say that bricks are of good quality and then you supply her right? So situation again like this you should use sampling right? Why to collect data there are several applications. In fact, in previous class I discussed about this.

**(Refer Slide Time: 14:22)**




So let us take a there is a marketing manager he wants to know the effectiveness of a new tv advertisement which he has just prepared. So whether it is better than previous advertisement how many people are watching new advertisement after watching how many people are coming

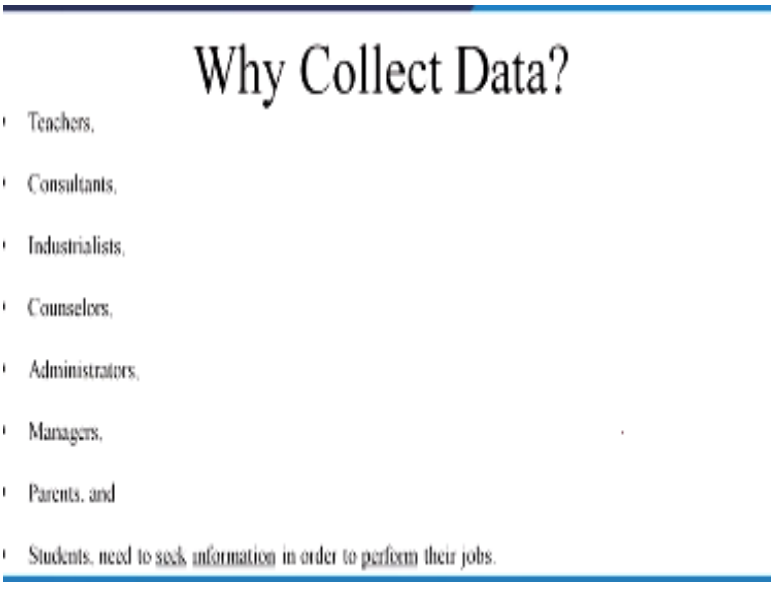
to market for buying the product. For which advertisement has been designed. Let us say a pharmaceutical manufacturer wants to know.

The effectiveness of new drug which led pharmaceutical company has recently prepared. So how this new drug is more effective than previous drug. So using data manufacturer of pharmaceutical drugs can make a decision. An operator at let us in a manufacturing process right are at an assembly line and wants to know what is the quality of the product which are being manufactured.

Are they being manufactured as per standards of the company or not? Another example is let us say an auditor wants to review the financial transaction of a company in order to determine whether the companies in compliance with generally accepted accounting principles or not. So if not then why not? How company can follow the standards. So and there are 100s of such applications not under 100s there are 1000s of such applications.

So all these people there are few names I have written all of them are using data for their jobs.

**(Refer Slide Time: 16:24)**

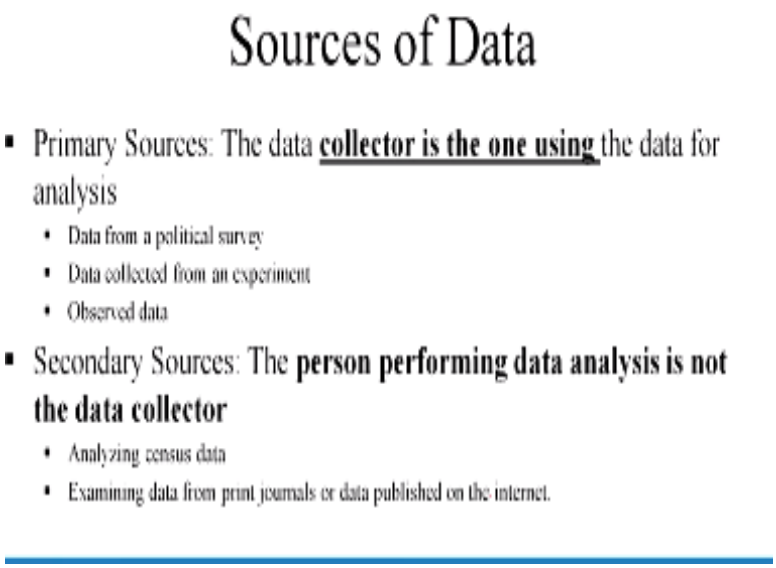



Be a teacher or a consultant or industrialist how teacher is using data. So let us say if you are teaching to a class then you can find out what is the mean marks of the class in midterm examination. If mean marks are let us say out of 25 it is let us say 10 then you need to take a

corrective action right. So like this consultants, industrialist, counsellors, administrators, managers, parents.

Parents can see how their children are performing in different subjects. And they can take corrective measures for their children and they can go for a coaching class and so on. Right? So students and almost all of us use data to perform our jobs okay.

**(Refer Slide Time: 17:22)**




If you look at sources you can classify sources in many ways internal sources, external sources. Primary, secondary you can have some quantitative, qualitative sources and so on. So if you look at primary sources means if I have to collect data for my problem or fire my project if I am collecting data personally if not personally if I even if I am using let us say field workers for data collection then that would be called primary source of data.

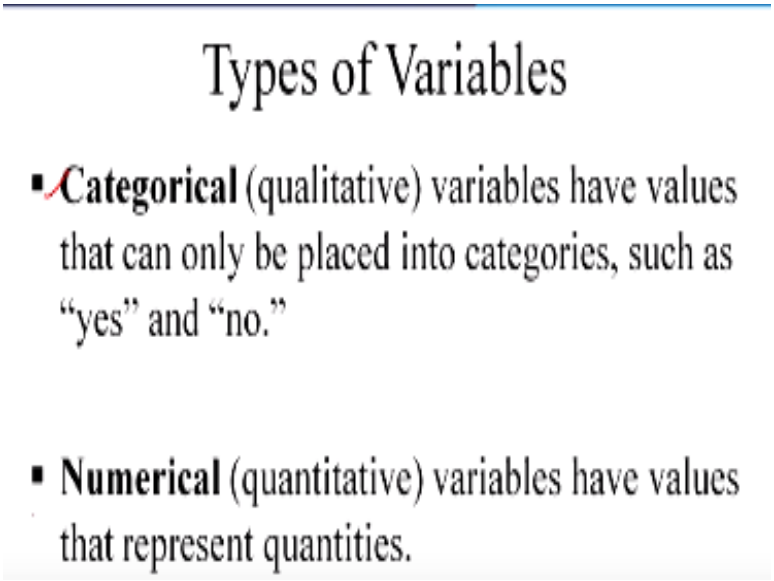
So you can have examples like this data collected by a political party data collected from an experiment. So if you are performing an experiment in laboratory then you would collect data from that experiment. So that would be a primary data you are observing something. Let us say you are observing number of vehicles passing from an intersection. So how many trucks passed from 9:00 AM to 10:00 AM how many two wheeler passed from 9:00 AM to 10:00 AM.

So these are observational data generally you should go for secondary sources initially. If you think that secondary sources are not reliable not updated not accurate then you should go for some primary sources. What about secondary sources? Those sources of data wherein some other persons or other people or other agencies other organisations have collected data for their work their research work their project work their assignment work or some other purpose.

So that is secondary data so as I said if there is a research project or any project or any work you need data then try to search from secondary sources initially and there are several secondary sources you have got let us say newspapers you have got magazines you have got library and you have a different government sources different business associations and so on. So these are the two primary sources of data.

Now let us look at types of variables what is a variable? Variable is characteristics of an individual in which we are interested. For example, height of a student right? Or let us say weight of students in your class. So you can know variables in two forms.

**(Refer Slide Time: 20:35)**



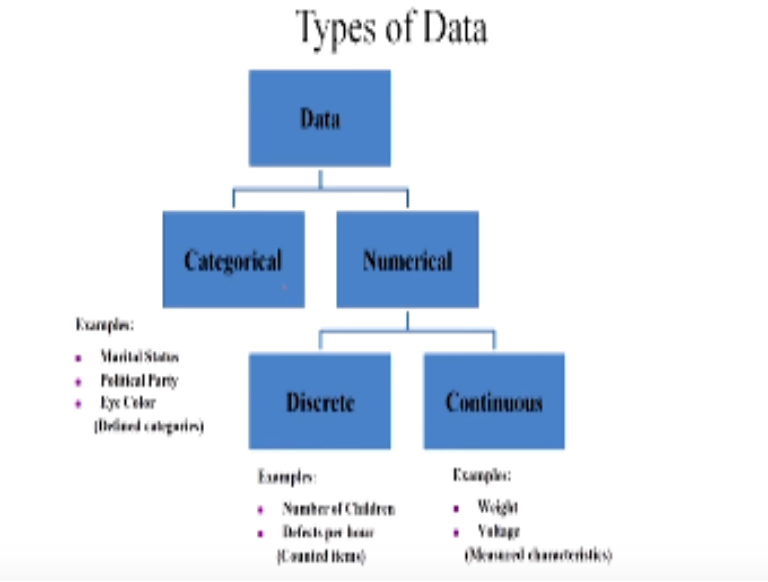

Either it can be a categorical or numerical so either you call it say non-numerical. Non-Numerical and numerical okay or you can call it Quantitative Quanti and these are qualitative. Right so they are one and the same thing. Or you can call it metric data so either you call it numerical quantitative metric one and the same thing. This is non-metric data right so what are

non-metric data variables have values that can only be placed into categories such as such as yes no.

Right let us say male female it is nothing metric in it right let us say you would have seen people having 10 cups of coffee every day. So you can call them heavy coffee drinker. You know what medium coffee drinker right? So heavy medium they are nothing but categorical variables. You can say let us say group 1 group 2 categorical variable. BE first year class BE second year class categorical variable right.

Of course you can quote those categorical variables like let us say yes means 1 and no means 0. You can do coding also once you got categorical variables, numerical variables are variables have values that represent quantitative quantities right? So there would be some numbers for example let us say 13.5 kg right this is nothing but a numerical variable. Okay?

**(Refer Slide Time: 22:52)**



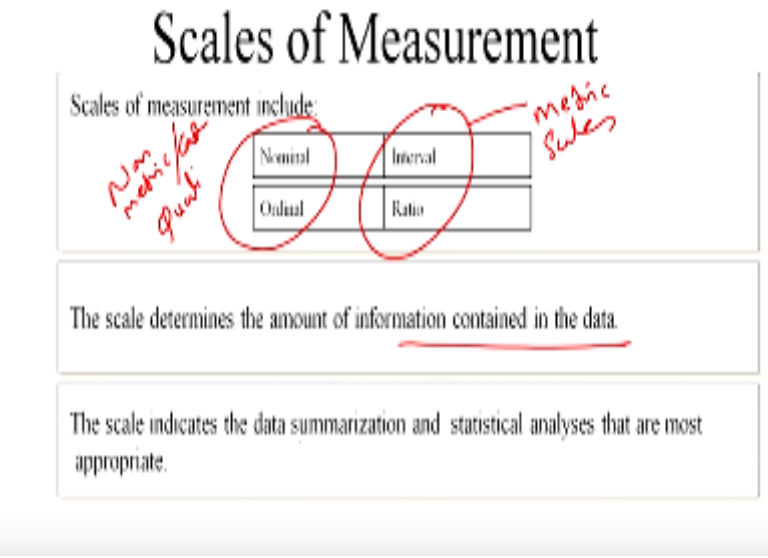

Now types of data so as I said you can have categorical and numerical we have already talked about categorical are some more examples. So marital status of a person married, bachelor divorcee. So there are three categories of this particular variable right political parties party x y z and so on right? If you look at numerical data numerical data can be either can be discrete data or continuous data.

So those data which you can count are called discrete data. For example, number of houses in a city or let us say number of children in a house let us say number of cars in a house number of TV sets in a house defects per hour on an assembly line per hour number of pet animals in a house so all these things can be counted so that is discrete data. You can have continuous data as well.

Weight now weight can be let us say you need to measure the weight you cannot count the weight. So weight can be let us say 13 let us say 130 pounds right or let us say voltage 1.5, resistance 2.7 ohms right so let us say another example let us say Sensex and we did yesterday let us say 5000.13 points. So these are these are the kind of continuous data. Okay let us look at couple of scales because you whenever you go for data collection.

You need to use some scales just like you know but a weighing machine for measuring weight you have got you can measure height of a person in terms of centimetres or inches whatever you want. Similar let us say if you want to measure attitude of a person towards certain product then you need to have some scale right. So let us look at these scales we will look at 4 basic scales.

**(Refer Slide Time: 25:30)**

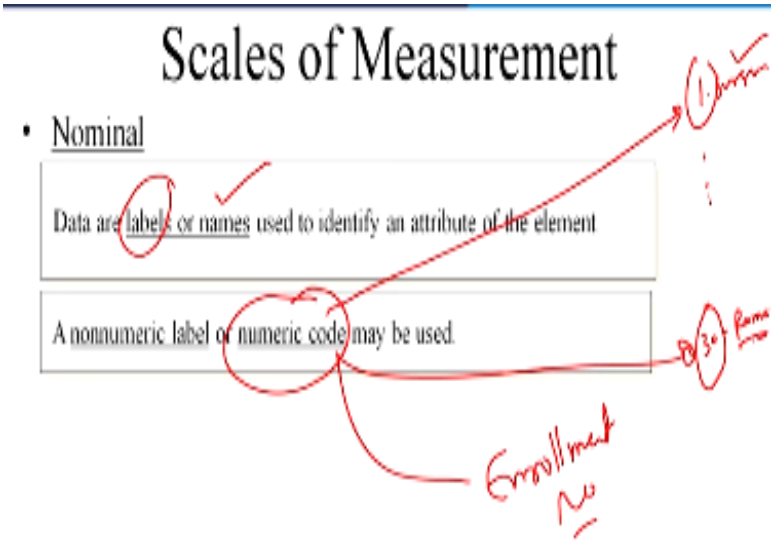



We have got nominal scale ordinary scale, interval scale and the fourth one is a ratio scale. So 2 of these scales are categorical scales you can call them non-metric right or you can call it categorical or qualitative right? One and the same thing these 2 are metric scales right metric

scales. So what is a nominal scale? A nominal scale is a scale which we use to identify a particular person a product or a group or an educational institute or any other thing right?

So the scale determines the amount of information contained in the data of course. So the scale indicates data summarization and statistical analyses that are most appropriate. So first we will have we will use a particular scale well collect data and then we will perform some statistical analysis.

**(Refer Slide Time: 26:49)**

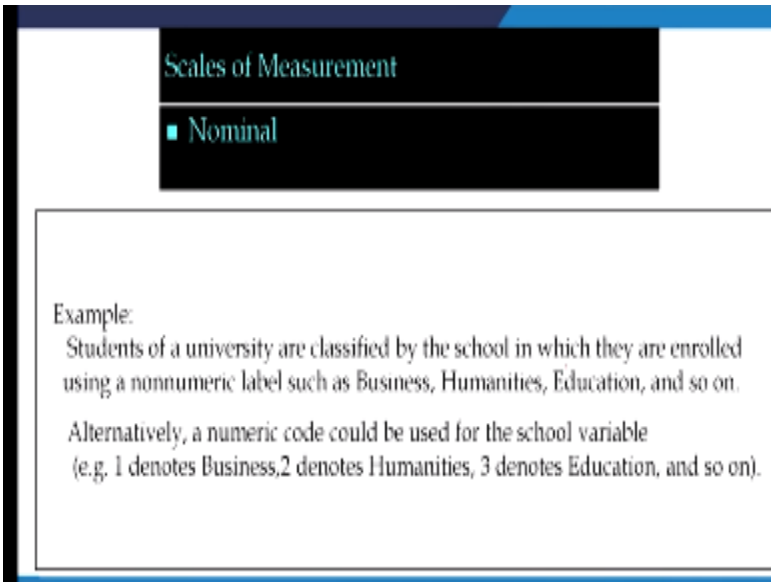



So nominal scale as I said are used for identification of something. So you let us say in your class you have 30 students so different names right? Let us say Ramesh, Suresh, John and so on right? So they are nothing but that that is nothing but a nominal scale we are using some labels do identify an object or a person or a group. Okay so that is nominal scale so once you have got a nominal scale you can assign them some numeric code as well.

For example, there are let us say 30 students in your class serial number 1 Suresh right serial number 3 Ramesh right so these are nothing but numeric code. Or nothing you can have there is something called enrolment number right? Enrolment number so you have got your label and you can have some enrolment number for this particular label right? So this 30 does not mean that this fellow Ramesh is 30 times intelligent then Suresh rate keep in mind.

So we have just numbered these variables just for identification purpose. So this is a nominal scale. Let us look at an example.

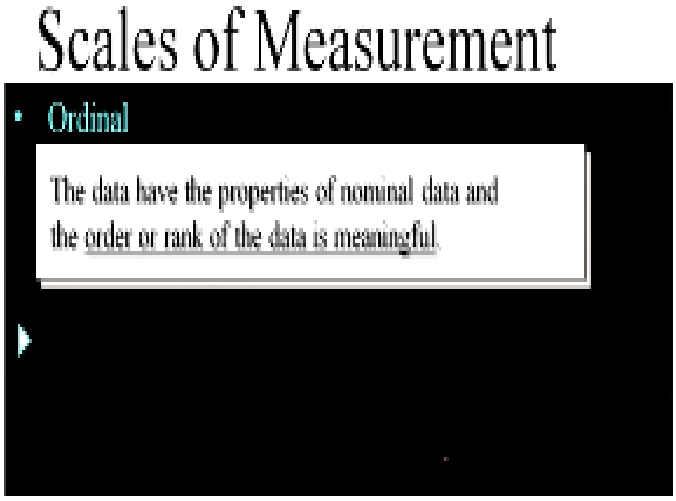
**(Refer Slide Time: 28:26)**




The students of a university are classified by the school in which they are enrolled using a non-numerical label. So let us call them business, departments, students, humanity students and education students and so on right. So you can have mechanical engineering students, electrical engineering, civil engineering chemical materials and so on right? And you can then number as well right.

Let us say for the in this case so we have numbered businesses as 1 humanity as two and so on right? So you can always how a nominal scale and it can be a numbered also right?

**(Refer Slide Time: 29:17)**




The second one is ordinary scale so the data how the properties of nominal scale and the order rank of the data is meaningful in case of ordinary scale. So for the timing let me stop here and let me summarize what we did in the last 30 minutes. We did talk about what is inferential statistics what is descriptive statistics? We have seen different types of scales. We have also seen different types of data.

We have seen discrete data, continuous data. We have also seen different sources of data. As I said initially you should try to go for secondary data but if you do not think if you think that secondary data are not good enough for your work then go for primary data. So with this let me stop here. In next class I will discuss about remaining scales namely ordinary scale, interval scale and ratio scale. Thank you.