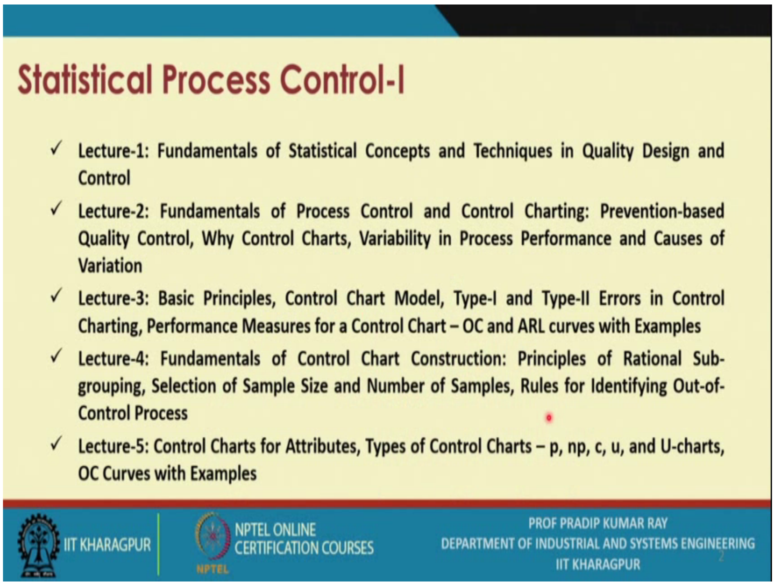
Quality Design and Control

Prof. Pradip Kumar Ray
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 16
Statistical Process Control-I

So, in the 4th week we will be discussing one of the most important topics under quality design and control that is the statistical process control. Now, when we talk about statistical process control this topic covers a number of subtopics in fact and that is why you know this statistical process control these aspect I will be discussing in 2 weeks time. In the first week that means, this is the current 4th week I will be discussing a certain basics are related to the statistical process control these I am calling it statistical process control part one and in the subsequent week that means, in the next week I will be discussing the part 2 of the statistical process control.

(Refer Slide Time: 00:43)



Now, under part one of statistical process control we have 5 lecture sessions and these lecture sessions are fundamentals of statistical concepts and techniques in quality design and control. So, you must have certain the knowledge in the fundamentals of statistical concepts of techniques.

In the next lecture I will discuss fundamentals of process control and control charting. In next lecture that means, lecture 3 I will talk about the basic principles control chart model type 1 and type 2 errors in control charting performance measures for a control chart, operating characteristic curve or OC curve and ARL curves average run length curves with examples. So, we will be citing examples.

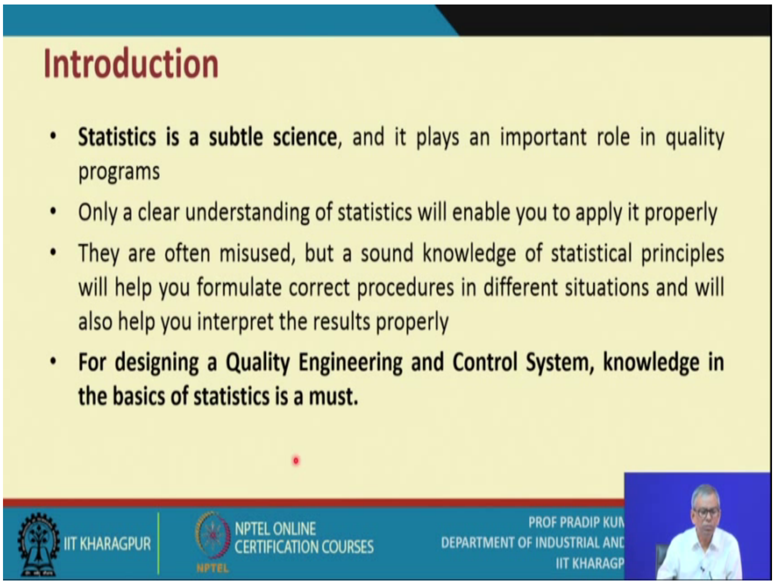
In lecture 4, I will discuss the fundamentals of control chart construction and related to construction the principles of rational subgrouping selection of the sample size and the number of samples rules for identifying out of control process. And in the last lecture that is the lecture 5, I will be discussing control charts for the attributes this is a type of control chart. Types of under this category we have 5 types of control charts each one I will discuss like p, chart, np chart, c chart, u chart and capital U-charts and OC curves with examples I will also discuss. So, this is the broad coverage.

Now, today at this point in time I will be discussing the fundamentals of statistical concepts and techniques in quality design and control.

(Refer Slide Time: 03:26)

You know we have been telling you all the time that the tools and techniques which you use for quality design and control they are primarily dependent on the statistical methods and the statistical approaches. So, before we start learning all those tools and techniques, you must have the knowledge in the fundamentals of a certain statistical concepts and techniques. So, this is the background and this is the requirement.

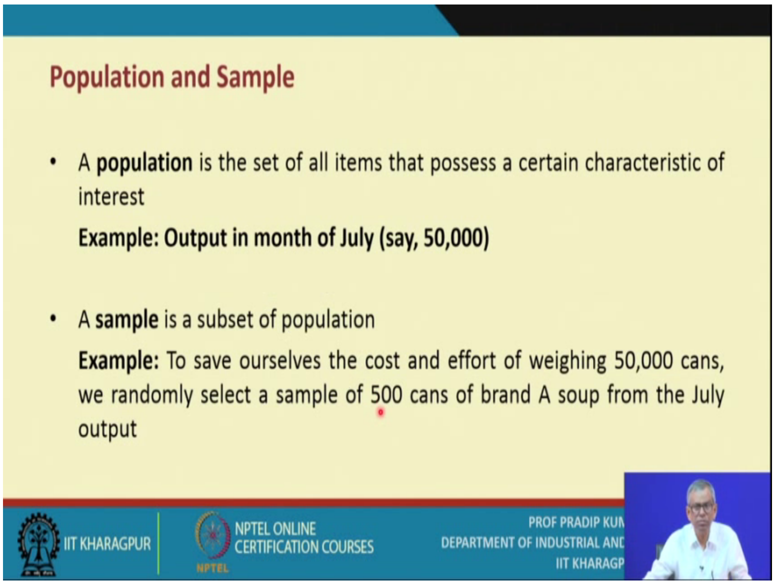
(Refer Slide Time: 04:12)



Now, let me just tell you certain basics related to the statistical analysis like the statistics. What is statistic? Statistics is a subtle science and it plays an important role in quality programs. So, that is point to be noted that is a subtle science. Only a clear understanding of statistics we will enable you to apply it properly. So, that is one number 2; that means, if we want to apply you know the quality related tools and techniques. So, understanding of statistics is a must.

They are often misused, but a sound knowledge of statistical principles will help you formulate correct procedures in different situations and will also help you interpret the results properly is it Ok. And the point to be noted that is very important for designing quality engineering and control systems that is your goal that is your purpose knowledge in knowledge in the basics of statistics is a must, is it ok.

(Refer Slide Time: 05:30)

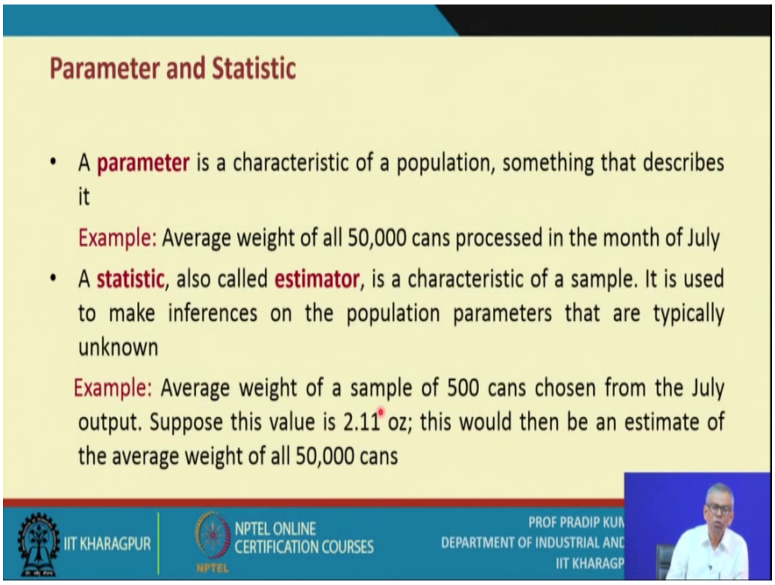


So, with this basic you know the introductory remarks let me the define certain terms and terminologies we frequently come across while we study statistics is it ok, and many such a transfer terminologies we will come across while you analyze data statistically for quality control or quality improvement, for quality engineering purposes.

What is the population? Population is the set of all items that causes a certain characteristic of interest example output in month of July. So, I do not need to elaborate say you produce 50,000 you know the units in a particular month. So, that is your population.

What is the sample? Obviously, the sample sometimes this is also referred to as a subgroup. So, please note it down the sample sometimes in the context of control charting we refer to as a subgroup. So, a sample or a subgroup is a subset of population. What is an example? Example is to save ourselves the cost and effort of weighing 50,000 cans suppose the 50,000 units are the cans we randomly select a sample of 500 cans or brand a soup from the July output, is it clear. That means instead of measuring or instead of measuring the entire you know the population what you try to do; that means, you just select a sample of 500 cans; that means, this is to be, this is to be selected this sample is to be selected and normally; that means, this is just you know 1 percent this is basically 1 percent is not it now. So, that is called the sample.

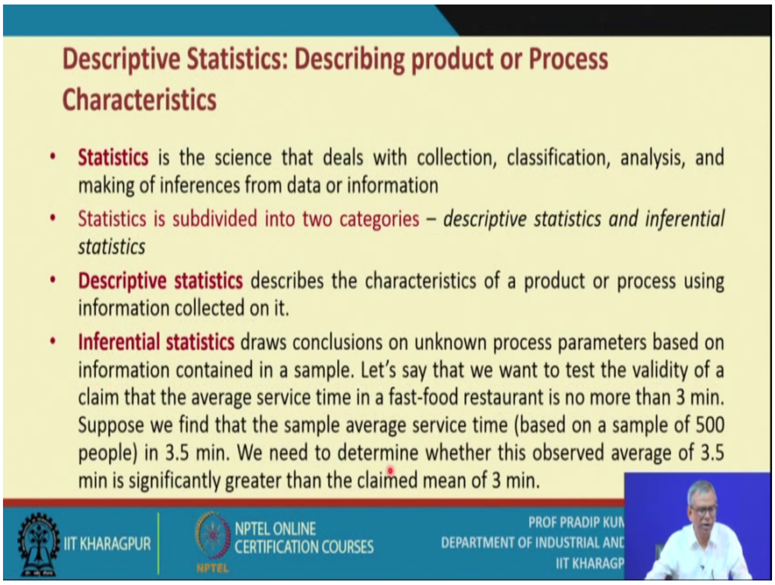
(Refer Slide Time: 07:31)



What is the parameter and what is a statistic? A parameter is a characteristic of a population; that means, whenever you refer to the population we say that how do you identify it characteristics. So, when you are able to identify its characteristic of the population you know it is referred to as a parameter. Example average weight of all 50,000 cans processed in the month of July, so 50,000 in there is a is your the population size. So, what you try to do each and every can you get the data of its weight and then you you know the calculate its average. So, that is the parameter of the population.

What is the statistic? Whenever we say it is a statistic is a statistic or sometimes this is referred as an estimator is a characteristic of a sample. So, it is used to make inferences on the population parameters that are typically unknown; that means, here in order to get the average with respect to the population you have to you know consider all these 50,000 data points. Whereas if you refer to the sample what was the sample size sample size is just 500. So, the 500 data points you collect and you calculate its average value that is 2.11 ounce this would then be an estimate of the average weight of all 50,000 cans. That means, instead of measuring all these I know all the 50,000 data points. So, it may be very very time consuming and it may not be necessary in many cases. So, we refer to the sample and get the value of the sample statistic.

(Refer Slide Time: 09:28)

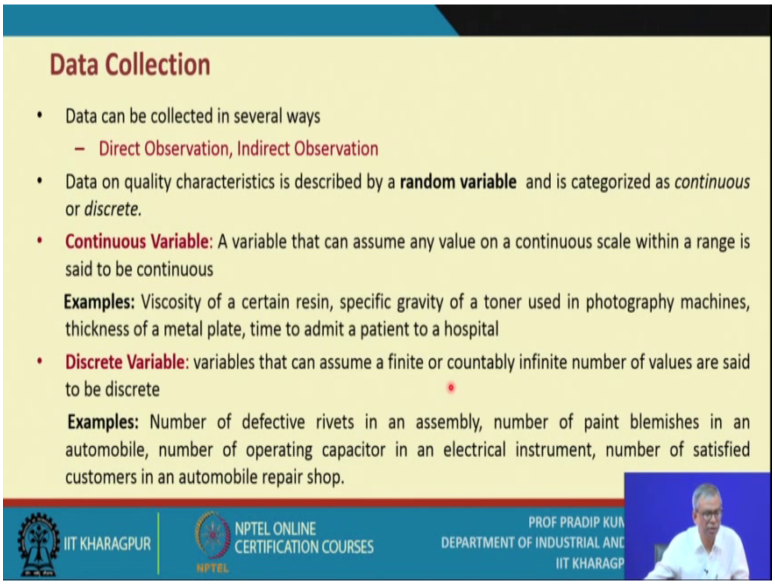


Now, these statistics is a science that deals with collection this is the exact definition of statistics. So, statistics is the science that deals with collection, classification, analysis and making of inferences from the data or information is it. So, this is the definition.

Now, this, the subject of statistics if subdivided into two categories, first one is the descriptive statistics and the second category is inferential statistics. Now, what is descriptive statistics? Now, the descriptive statistics describes the characteristics of product or process using information collected on it. So, that is you know a straightforward you know the definition. What is inferential statistics? Draws conclusions on unknown process parameters, parameters already, parameters means with respect to the population based on information contained in a sample. So, that is essentially inferential statistics. Let us say that we want to tells the validity of a claim that the average service time in a fast food restaurant is no more than 3 minutes this is just I am trying to cite an example.

So, the average you know the service time is 3 minutes suppose we find that the sample average service time based on a sample of 500 people is 3.5 minutes, 3.5 minutes, we need to determine whether these observed average of 3.5 minute is significantly greater than the claimed mean of 3 minutes is it. So, this is the problem I hope that the problem is meant very very clear. Now, whether 3.5 in this case is significantly greater than a value of 3 minutes, so this is the thing.

(Refer Slide Time: 11:38)



So, this way we say that how to answer; obviously, have to apply the inferential concept of inferential statistics like say you know hypothesis testing and all.

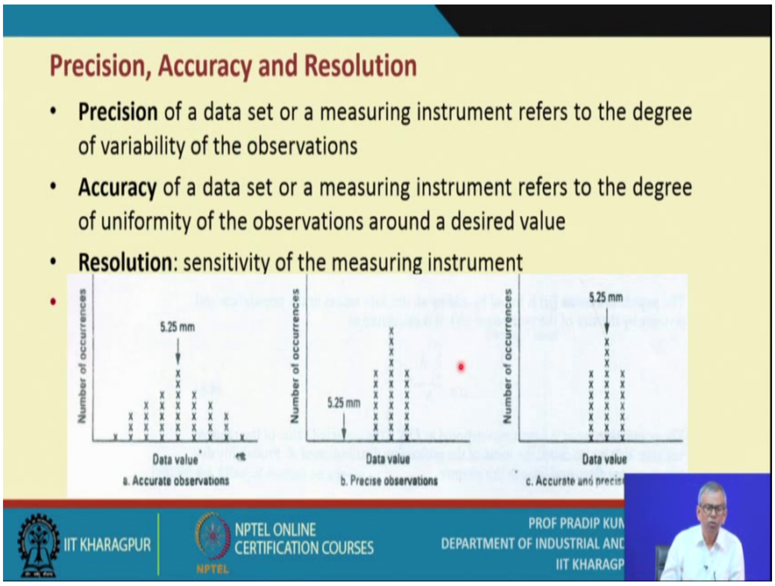
Now, you know any statistical analysis depends on data collection and similarly we say any quality related exercise depends on data collection. So, you will come across different kinds of data and first question is how do you collect this data. There could be several ways for collecting data we say through direct observation you can collect the data or indirect observation. Sometimes know you say that you have to collect data they are referred to as a primary source of data; that means, actually you go to the field and you have observed the systems and you have collected data is it. So, your analysis based on essentially analysis from the data you collect from the primary sources and there could be indirect observations also, for collection of data you can refer to the secondary sources of other data or information.

Now, data on quality characteristics are described by random variable and is categorized as continuous or discrete. So, I think you are already aware of what is it random variable. So, the random variable can be of say either continuous type or the discrete type. So, what is a continuous variable? A variable that can assume any value on a continuous scale any value; that means, integer or fraction within a range is said to be continuous, very clear in fact.

So, what are the examples? Like viscosity of a certain regime specific gravity of a toner used in photography machines, thickness of a metal plate, time to admit a patient to a hospital. So, these are the few examples etcetera etcetera. There are many examples of continuous you know the variables in any exercise on quality it is high, highly likely that you will come across several such continuous variables.

What is a discrete variable? A discrete variable can assume a finite or countably infinite number of values and that is why they refer to as the discrete random variable. Examples number of defective rivets in an assembly it can be only you know either 0 or any other integer. Number of paint blemishes in and out of oil it just cannot be a fraction, number of operating capacitor in an electrical instrument, number of satisfied customers in an automobile repair shop. So, these are the examples.

(Refer Slide Time: 14:52)



Now, whenever we refer to you know the quality related data many a time this data you get by using measuring instruments and, sometimes we will say that the quality is as good as the measuring instrument. That means, the make sure that the data which you get by using a measuring instrument that this data not be influenced by the measurement error is it ok. So, what do you need to do? That means, related to a particular say the measuring instrument, 3 aspects you must know. First is the precision of a data set the second one is the accuracy of a data set that these data set is obtained by using a measuring instrument or a gauge.

And the last one is the resolution of the measuring instrument. So, let me just explain briefly. What is precision? Precision of a data set or a measuring instrument refers to the degree of variability of the observations as I have already told you that any exercise is on quality, is essentially an exercise on variability. So, the degree of variability is referred to as the precision of the data set.

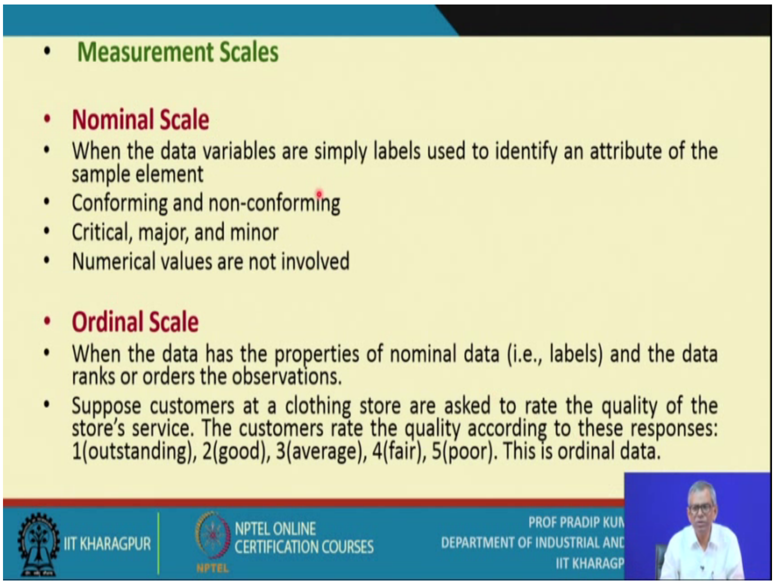
Next one is the accuracy or sometimes you know the accuracy means. Now, the accuracy of a data set or a measuring instrument refers to the degree of uniformity of the observations around a desired value. A desired value may be the standard or the target and accuracy means what is the difference between your observation and the standard you set for the given quality characteristics. So, that is referred to as accuracy.

What is resolution? Resolution refers to sensitivity of the measuring instrument. So, now, these examples are given like say here the data values are plotted and this is the number of occurrences which is like the frequency I know say the distribution. So, what do you find that 5.25 mm is actually your value and, this is the value you get almost at the central location and here we are saw the accurate observations this is accurate; that means, the repeatability is very high repeatability that is the point you must remember and the repeatability of the data points very very high. So, then it is referred to as the precision if the repeatability is very very high.

So, here you know these are the precise observations, these are the precise observations; that means, the most of the data points you know they are occurring within a very narrow band is a very narrow band whereas, here you know this is a wide band is it; that means, the range of the data point is very wide. So, that is why you know the accuracy may be here; that means, we are referring to this particular standard whereas, here there is a standards are quite off from say the data values.

So, but the data values are occurring within a you know the sort or say the close the range that is why here precision is very high, but it may not be accurate here you know the precision is poor, but it is not that precise, but you know the accurate observation is at the center. And here actually the data are you know is very precise as well as accurate. So, these are the 3 examples you have accurate observations, precise observations and what you need both accurate and precise observations.

(Refer Slide Time: 19:32)

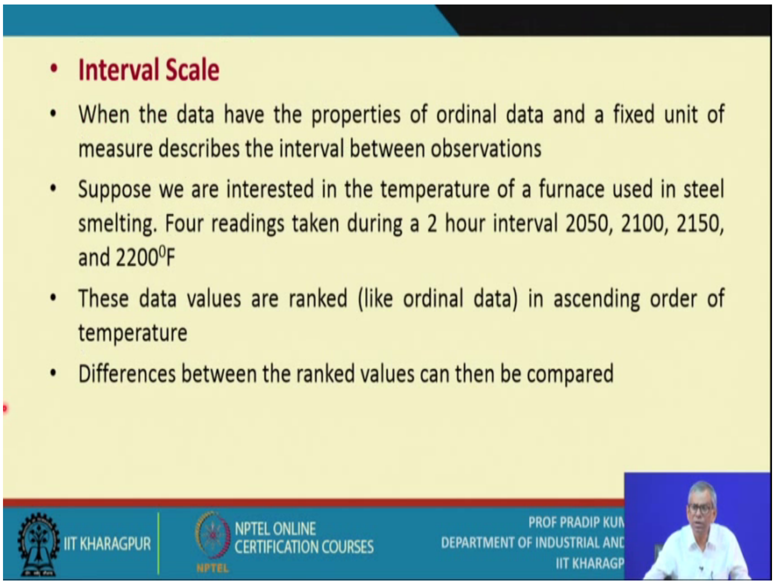


Now, when you collect the data, now you must use certain measurement scales and this you already I presume that you already aware of. Now, there are 4 scales we have; that means, 4 kinds of the data you come across and in the field of quality what you find that you come across one or more such you know the data ok.

So, the first one first scale you use that is the nominal scale. When the data variables are simply levels used to identify an attribute of a sample element, like we say it is conforming or nonconforming, these are nominal values. So, nominal scale you use. As a critical major or minor is it, like say it could be critical failure, it could be a major failure, it could be a minor failure. So, these 3 categories we have. Numerical values are not involved is it all right. So, such data you generate and the corresponding scale is referred to as the nominal scale.

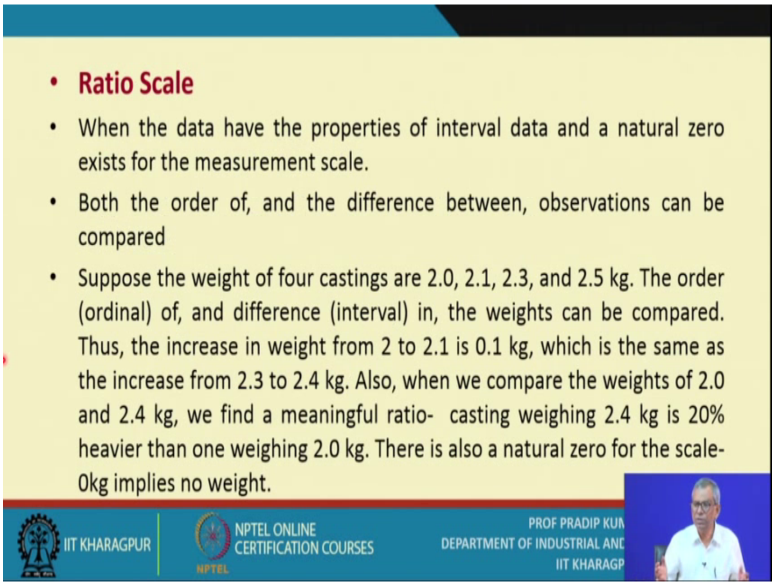
Next one is referred to as ordinal scale. When the data has the properties of nominal data like the levels and the data ranks or orders the observations is it ok. Suppose the customers at a clothing store are asked to rate the quality of the store service is it this is just one example. The customers rate the quality according to these responses like it says if it says 1 that means outstanding, 2 means good, 3 means the average, 4 means the fair, 5 means poor. So, this is an ordinal data is it; that means, if someone say 5; that means, it is always it is a ranking; that means, is a quite you know away from a rank of one; that means, that is the best or the outstanding. So, this is ordinal scale.

(Refer Slide Time: 21:39)



Then the third one is the interval scale when the data have the properties of ordinal data and a fixed unit of measure describes the interval between observations. So, that is very clear. In fact, suppose we are interested in the temperature of furnace used in a steel smelting. So, this is the typical example 4 readings taken during a 2 hour interval of 2050, 2100, 2150 and 2200 degree Fahrenheit. These data values are ranked like ordinal data in ascending order of temperature like we have already done. The differences between the rank values can then be compared is it. So, this is the interval scale.

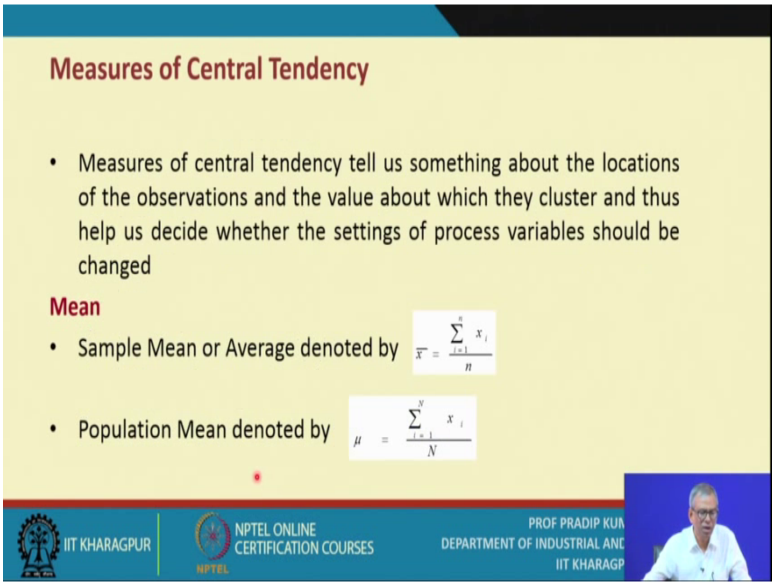
(Refer Slide Time: 22:27)



And the last one is the ratio scale when the data have the properties of interval data is it and a natural 0 exists for the measurement scale is it ok. Then we refer to the ratio scale both the order of and the difference between observations can be compared.

So, here in as example. Suppose the weight of 4 castings are say 2.0, 2.1, 2.3 and 2.5 kg fine. The order of and the difference that is the interval in the weights can be compared. There is a increase in weight from 2 to 2.1 is 0.1 kg which is the same as the increased from 2.3 to 2.4 kg. Also when we compare the weights of 2 and 2.4 kg we find a meaningful ratio casting weighing 2.4 kg is 20 percent heavier than one weighing to kg is it. So, this is just I am describing you already may be knowing all these concepts and all. So, there is also a natural 0 for the scale, 0 kg implies no weight is it. So, this is automatically it is; that means, the concept is there when I say that you know the 2 point kg is the weight it means I have in my mind that there may not there could be a value called 0 that means, no weight at all.

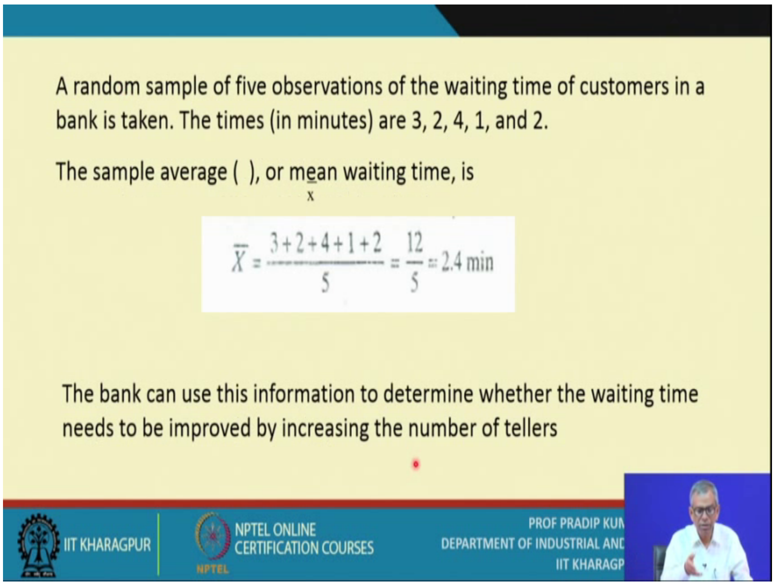
(Refer Slide Time: 24:00)



Now, when you collect the data first thing what you have to do; that means, by analyzing the data you have to have some meaningful information, is it ok. So, the meaningful information means at the first instance you must know that what is the central tendency of the data, central tendency and the second one is you know, but the dispersion, how the dispersed the data points are. So, these two are very very important I have already referred to you know the accuracy and the precision, is it ok.

So, the central tendency the data, so for that you need to you may have several measures. So, the measures of central tendency may tell us something about the locations of the observations and the value about who is the cluster and thus help us decide whether the settings of the process variables should be changed. So, like say mean we have the sample mean or average these are the notations like X bar we say and sigma X i divided by n, n is the number of data points and when we refer to the population. So, the mean is denoted by you know the notation is mu is it sigma X i divided by capital N, capital N is the population size.

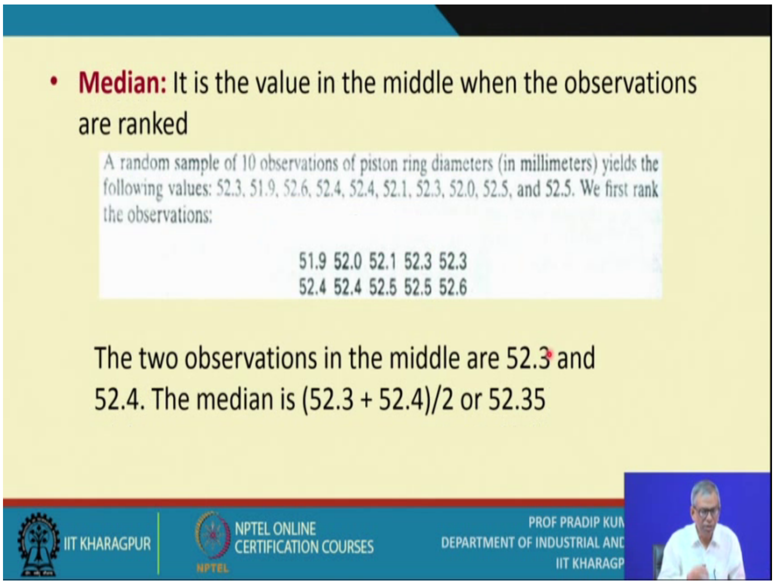
(Refer Slide Time: 25:31)



Then like this is one example right we take a random sample of 5 observations of the waiting time of customers in a bank. So, this is the data. So, what are these timings? These timings are 3 2 4 1 and 2. What is the sample average? That is X bar that is this 3 plus 2 plus 4 plus 1 plus 2 divided by 5 that is 2.4 mm. So, it is specifically the arithmetic mean. Now, say we know a simple arithmetic mean.

The bank can use this information to determine whether the waiting time needs to be improved by increasing the number of tellers. So, on an average the waiting time is 2.4 minutes. So, that is the information.

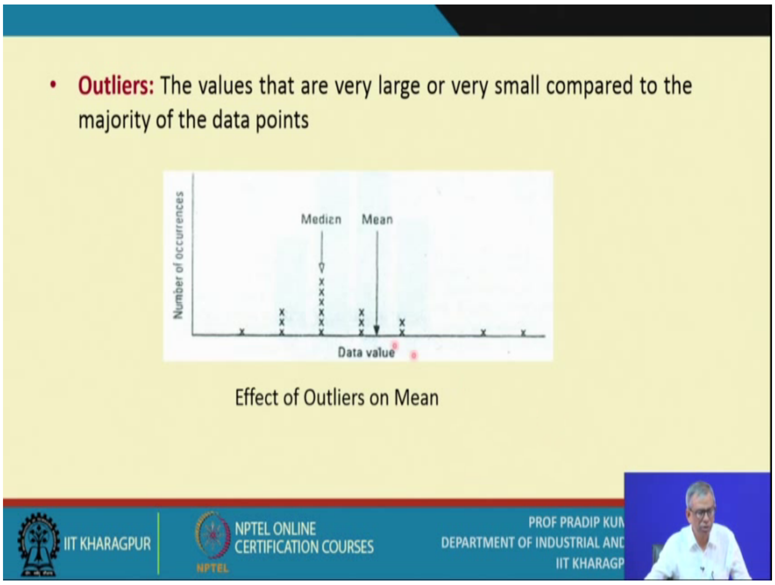
(Refer Slide Time: 25:19)



Now, the next one is the median. It is the value in the middle when the observations are ranked its very simple as it a random sample of pain observations and just giving you one example. Random sample obtain observations of piston ring diameters in millimeters yields the following values. So, what are these values likes a 52.3, 51.9 and so on and so forth. We first rank the observations; that means, the first the lowest values 51.9, next one is 52.0 and then 52.1 and so on like this and the largest one is this is the smallest one is the largest one that is 52.6. So, you just arrange them in ascending orders.

The two observations in the middle are 52.3 this one, this is the fifth one and this is the sixth one 52.4. So, the mean is 52.3 plus 52.4 divided by 2 that is 52.35, is it ok. So, we have you know the 10 values; that means, the even number.

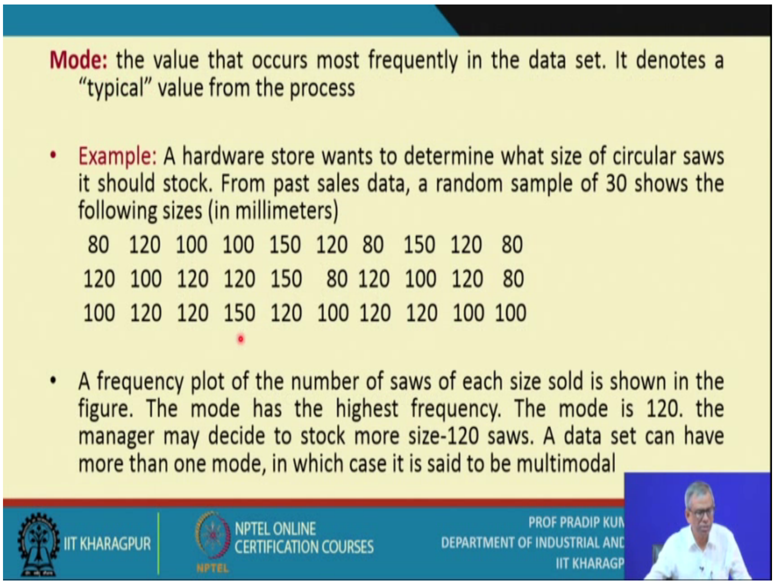
(Refer Slide Time: 27:37)



So, what is an outlier? The values that are very large are very small compared to the majority of the data points is it ok. Like here what you have like say these are the data values. So, we have calculate the mean somewhere here and medium is somewhere here is it. So, you have group the data is it ok.

Somewhere here you find the 2 data points, that is quite different they are very large, similarly this value could be very small compared to the majority of the data points. So, they may be refer to like this point this point as well as this point these 3 points may be referred to as outliers is it means the characteristic is totally different from the other, from that of the other points in this data set.

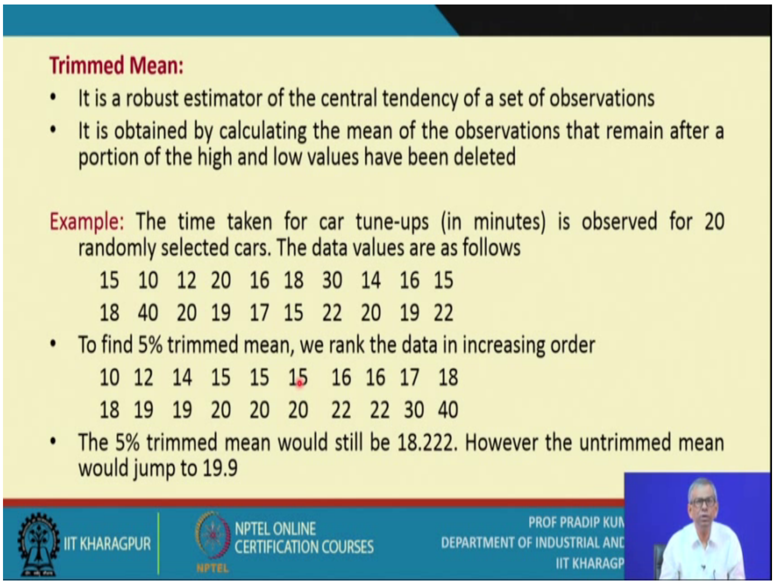
(Refer Slide Time: 28:32)



What is the mode? The mode is the value that occurs most frequently in the data set. So, it is known to you say denotes the typical value from the process. So, this is an example. A hardware store wants to determine what size of circular saws it should stock. From past sales data, a random sample of 30 shows the following sizes. So, these are the data points you have collected. So, measuring unit is millimeters a frequency plot of the number of saws of each size sold he is shown in the figure you already know is it ok, it is what how to draw the frequency the distribution and even, though, so that frequency plot of the number of saws is shown.

The mode has the highest frequency the mode is 120, is it ok. So, like say here you know for each value you know the frequency is counted that how many times it has occurred. You find that the maximum number of times 120 has occurred. The manager may decide to stock more, more size 120 saws a data set can have more than one mode in which case it is. So, be multimodal.

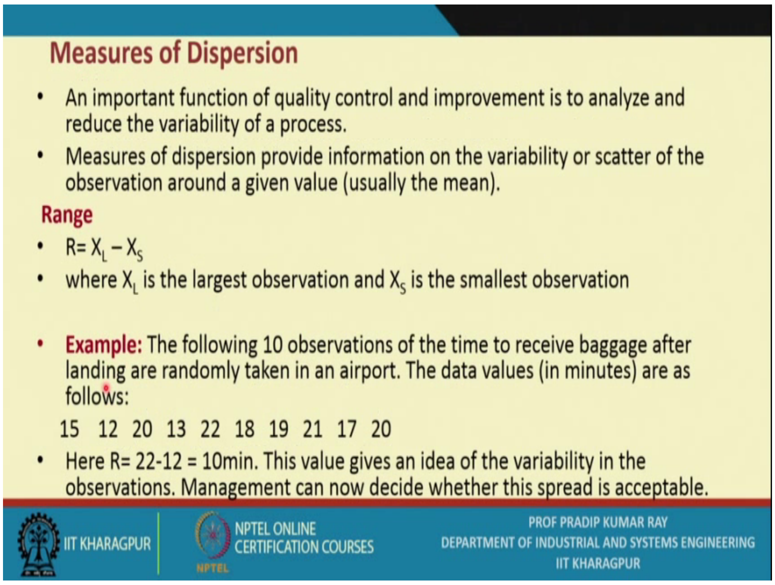
(Refer Slide Time: 30:00)



Now, next one is whenever we refer to such data set many a time you know when you collect data it is like say 100 data points on a particular quality characteristics we have collected it may so happen that the entire data set may contain you know one or more or several a number of say outliers. So, what you need to do; obviously, you know you must not consider the outliers. So, you have to remove the outliers from the data set. So, any remove the outliers from the data set and with the remaining you know the data points you calculate mean it is referred to as the trimmed mean I think it is clear.

So, it is a robust estimator of the central tendency of a set of observations. It is obtained by calculating the mean of the observations that remain after a portion of the high and low values have been deleted. Now, what are these high and low values? They referred to as the outliers, that means the total number of for the data points minus the number of data points considered outliers is it ok. So, with these values I will calculate the mean and this mean is referred to as a trimmed mean. So, this is an example is given, like the time taken for car tune ups these are the data sets and to find 5 percent trimmed mean we rank the data in increasing order is it rank them from 10 to 40. So, the lowest value is twenty and the highest value is 40. So, the 5 percent trimmed mean would still be 18.222; however, the untrimmed mean who would jump to 19.9 reading my point; that means, 5 percent of the trimmed mean; that means, 5 percent of the data points we will just remove is it ok; that means, high value as well as the low values is it ok, choice is yours.

(Refer Slide Time: 32:08)



Now, the next one is. So, these are the already I have mentioned what are the measures of central tendency the location is it means the data are located around certain point. So, that is why they refer to actually the central tendency.

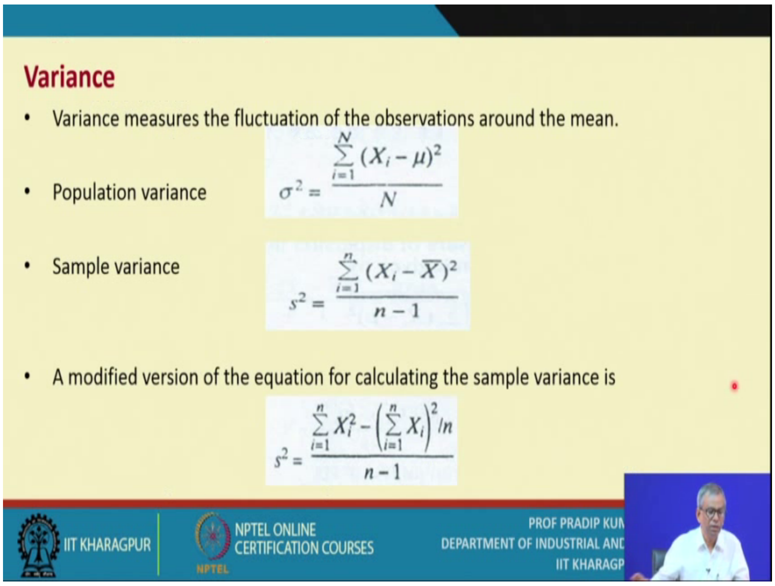
The next one is that the data are dispersed around a certain value definitely. So, the level of dispersion must be done. So, that is why the measures of dispersion; that means, it has a strong relationship with the variability. So, sometimes is referred as a measure of variability. So, there could be different measures for dispersion.

So, an important function of quality control and improvement is to analyze and reduce the variability of the process. This we have been telling you all the time. Measures of dispersion provide information on the variability or the scatter, sometimes this is referred to as the scatter if you remember that we have a scatter plot is it as a part of the 7 tools of quality management is it ok. So, please refer to the scatter plot, the scatter actually when you get the scatter plot you get an idea the initial idea of the level of variability.

So, along a given value usually the mean, along the mean how the data points are dispersed. So, a mean is, the first measure is the range, range is just one of the measures of dispersion ranges is X L minus X S where X L is the largest observation in the data set and X S is the smallest observation. Example, these 10 observations are taken, these are the 10 observations. So, what is the range? So, the maximum value is 22 and minimum is

12 so the range is 10 minutes. So, this value gives an idea of the variability in the observations. Management can now decide whether this plate is acceptable or not.

(Refer Slide Time: 34:15)

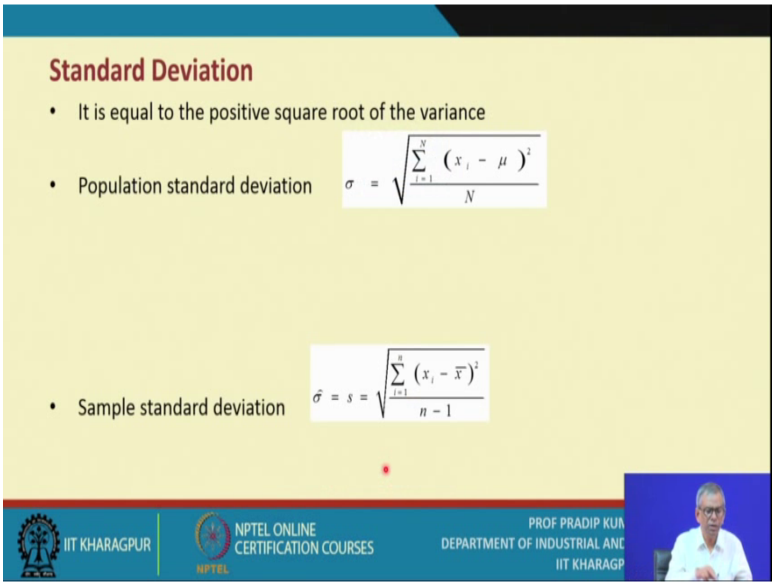


Now, the next important measure is the variance. So, how do you measure variance? That means, you try to measure the fluctuations of the observations or the variations of the observations around the mean. So, that is we refer to as the variance. So, there could be population variance when the population size is N. So, the notation is sigma square X i is the individual sample points, mu is the population mean and the difference is to be squared the squared of the differences and then must we you know they sum them up all the individual points and the square of the deviation actually and then you get the average by dividing it by capital N.

When you get the sample? The sample variance the standard notation is s square and this is the expression here we say that n minus 1 data points we will collected, suppose the sample size is say 50 so obviously, this will be n minus 1; that means, 50 minus 1 40 nine. So, there is the reason for this one and when you say n minus we compute the sample variance with n minus 1 we say it is an unbiased estimate later on, I will explain it in detail that why the sample variance with n minus 1 is it ok, with n minus 1 is preferred.

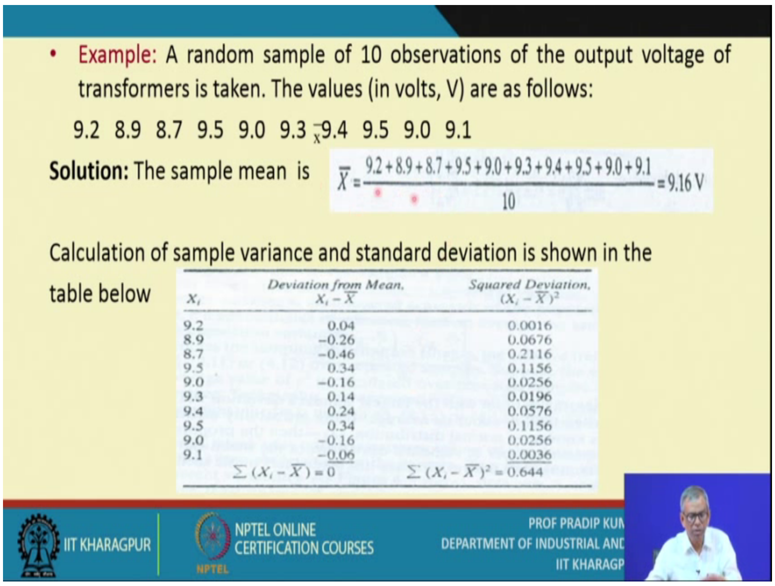
A modified version this you can modify it these expression; however, the equation for calculating the sample variance is this one. So, like this one this is just if you can modify this one ultimately you get these expression, sigma X i square i equal to 1 to n and this is sigma X i, i equals to 1 to 1 square divided by n divided by n minus 1. So, this is just the derivation.

(Refer Slide Time: 34:19)



Then what is the standard deviation? Standard deviation is this body strips square root of the variance so obviously, the population standard deviation is this positive square root notation is sigma. And what is the sample standard deviation? That is sigma hat that is this one, is it ok.

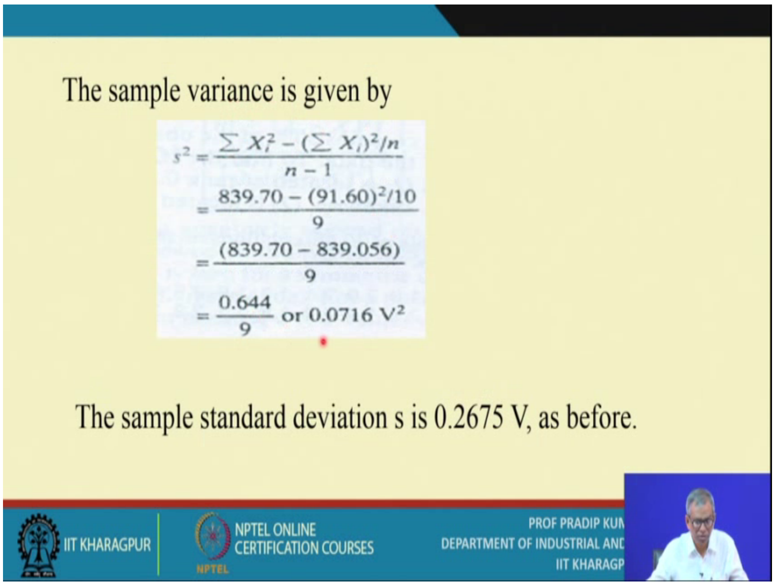
(Refer Slide Time: 36:37)



So, this is an example given. So, you have these data points, you have these data points you calculate the sample mean 10 data points we have there is 9.16 V.

Now, the sample variance how do you calculate? This is the procedure given. So, all the detailed steps are given, is it ok. So, as far the formulations, as far the expressions you get for the samples standard deviation and sample variance so you compute like this. That means individual X i’s values are given. So, X i’s will be a values from the X bar. So, that is to be subtracted then that this deviation is this square terms you add them you get these values and then you calculate the sample variance, is it ok. So, this is the sample variance the sample standard deviations are square positive square root that is this one. So, this way also you can calculate X i’s bar this is another way. So, in the 2 approaches there is another formula the second formula you can use and use this one.

(Refer Slide Time: 37:40)



So, sample standard deviation is this one as before. So, the today just refer to you know two possible measures in descriptive statistics, one is you know by the measures of central tendency and the second one is the measures of say the dispersion. So, two such measures we have discussed one is the range and we will tell you later on that under in which condition we will prefer range and which condition you must not prefer and then we have calculated the standard deviations and the variance for a set of data points. And the third measures we are going to discuss in the next lecture sessions that is referred to as interquartile range, is it ok. So, we you are already exposed to some you know the fundamentals of the statistical methods and the some more we will discuss in the next lecture sessions.

Thank you.