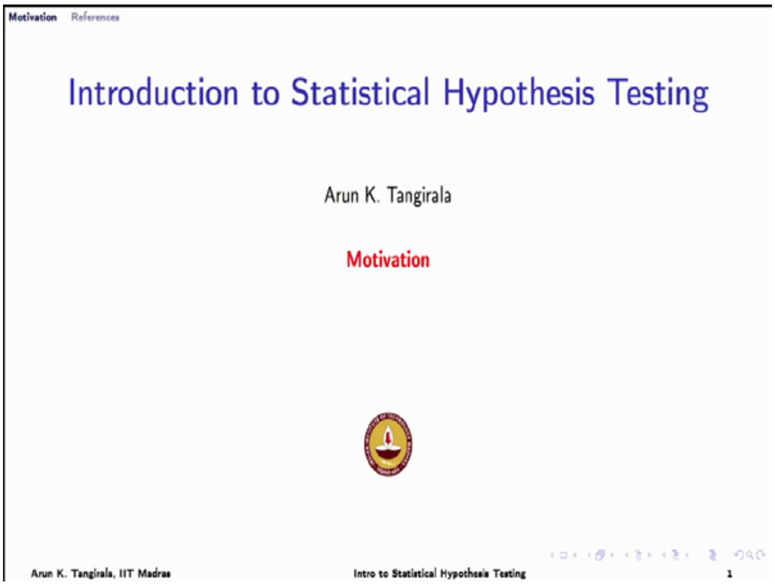
Introduction to Statistical Hypothesis Testing
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

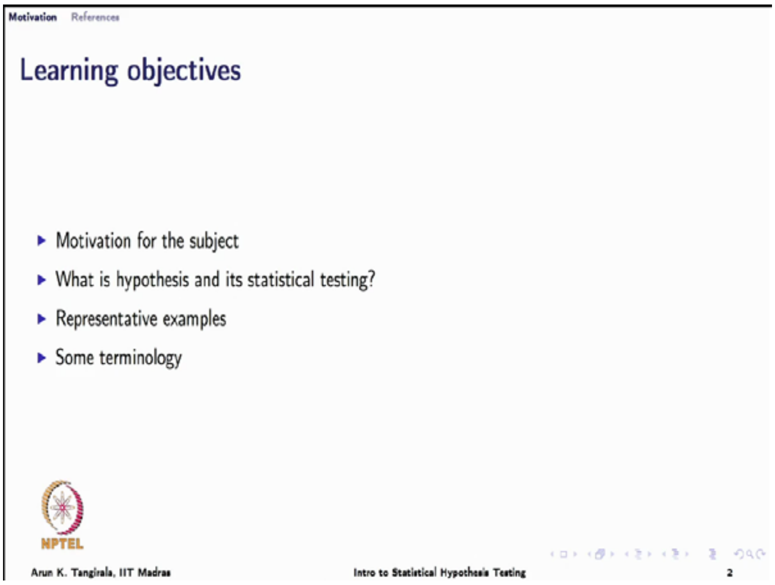
Lecture - 01
Motivation

(Refer Slide Time: 00:09)



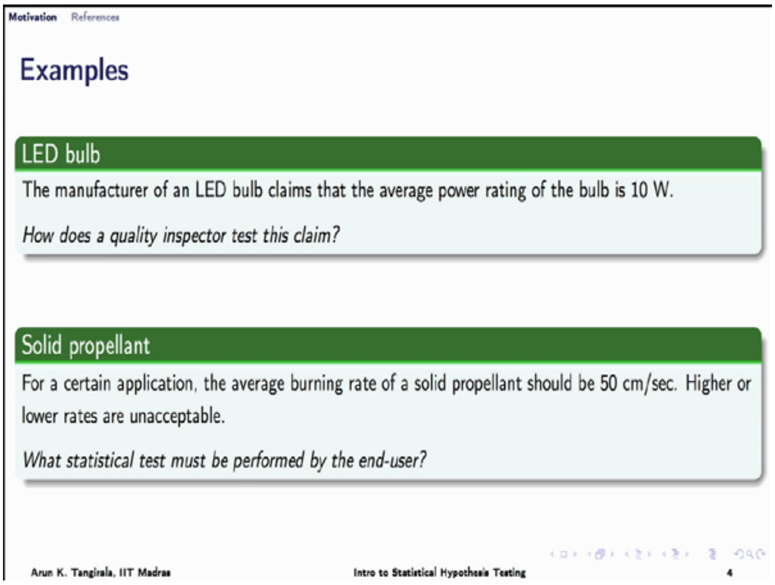
Hello friends. Welcome to the opening lecture on the course on Introduction to Statistical Hypothesis Testing. This is a 10-hour course and what will do in this lecture is primarily look at some motivating examples.

(Refer Slide Time: 00:28)



Get introduced to hypothesis testing, that is the problem statement itself and some philosophy of how the hypothesis is tested statistically. We will try to understand this with a use of few representative examples and also I will briefly talk about the course plan. Before we march forward, I just wanted to say that this hypothesis testing, the subject of hypothesis testing is critical to every step of or every exercise of data analysis as you will see through representative examples. It is not just correlation analysis or linear regression and so on. Any kind of data analysis at some stage involves hypothesis testing and it is with that reason and motivation that an exclusive 10-hour course is being offered. So, for those of you who want to get deep into a data analysis, it is important to go through this course and clearly understand the concepts underneath the subject of hypothesis testing.

(Refer Slide Time: 01:49)



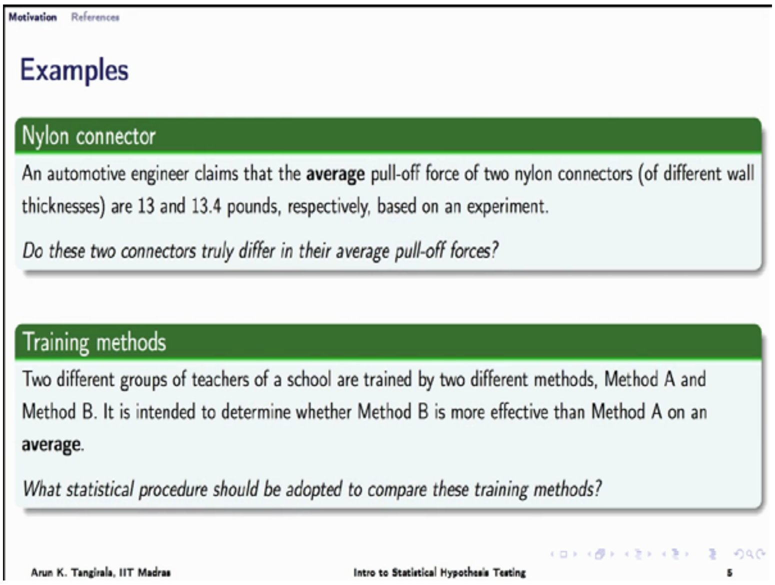
So, let us begin with a few examples and try to understand what is hypothesis testing? Let us take this example of the case, where the manufacturer of an LED bulb, today we all use LED bulbs to light up our homes, they are suppose to be energy or power efficient. So, there is this manufacture of a LED bulb, who claims that the average power rating of the bulb that is manufacturing is 10 watts, that is the power that it consumes. Now, as a quality inspector it is either you or me, how do we test this claim that the manufacturer is making? Of course, as I said early on the subject is concerned with statistical data analysis therefore; we are not going to test this claim using first principles that is fundamental knowledge and so on. The idea is to collect some data from the manufacturing process and then subject the data through statistical analysis with the goal of testing this claim.

This is probably one situation that we commonly run into. Although we talk of LED bulb, in this example, there are many manufacturing processes that make different kinds of products and that is always a concern of quality there, both for the manufacturer and the consumer. So, how do you actually go about testing a quality of those products in all such situations? Now, in a different example, let us say that in a certain application I am looking at the average burning rate of a solid propellant. Now, this solid propellants of course, are used in many different applications. The one particular application that we can think of is an escape system for the air crew that means the crew driving an aircraft. What we would be interested in here is the average burning rate, because this determines how quickly or slowly the ejection system or escape system for the aircrew would work, too slow would not be good obviously, and too fast also would not be good because both can be end up causing injury to the pilot and of course, to the life itself. So, we are interested in testing for this particular solid propellant, if the average burning rate is 50 centimeter per second.

What this means, as we just said is higher or lower rate are unacceptable. Now, how do we statistically test this kind of requirement that is necessary on the solid propellants? So, there is some supplier of a solid propellant and I would like to subject that solid propellant, those samples that I get from the supplier to a statistical test and check if on an average the burning rate is 50 centimeter per second. Now, the key word is here on an average, it clearly implies that when I look at each specimen that I get for testing, I may not get an a burning rate of exactly 50 centimeters per second. Some could give me 51 or some could give me 49 point 5 and so on and both the supplier and the end user, that is you know very well that it is impossible to maintain or to achieve the same burning rate for every specimen that I receive from the manufacturer or even as a manufacturer I would not be able to guarantee that. That is because every process has some inherent variations that are beyond our control.

We try to control the factors that are within our control, but otherwise there are always going to be factors that are beyond our control and these uncertainties are going to be present in an every process. As a result, it is not possible to achieve this kind of target, exact target for every specimen, be it the LED bulb or the solid propellant or any other product for that matter. However, what we hope is on the average the errors are ironed out and we achieve this target of 50 centimeters per second. So, question now is, how do we statistically test this whether the manufacturer or the specimen supplied by the manufacturer meets this requirement? Now, in both these cases here we are taking of averages, but there are other situations where that is averages of a single products, but there are other situations where may be we want to compare the averages of 2 different products or products of 2 different specifications and so on. So, as an example let us look at this nylon connector example.

(Refer Slide Time: 07:11)



This Nylon connectors are used in automobile industry extensively and an automotive engineer claims that he designed this nylon connector, claims that the average pull-off force of 2 different nylon connectors, that is of 2 different wall thicknesses have pull-off force of 13 and 13.4 pounds, respectively. Based on some experiments are the engineers has connected. Now, keeping the numbers 13 and 13.4 apart, the point that the engineer is trying to make is that these or test rather is that the different wall thicknesses that the engineer has chosen has resulted in 2 different pull-off forces.

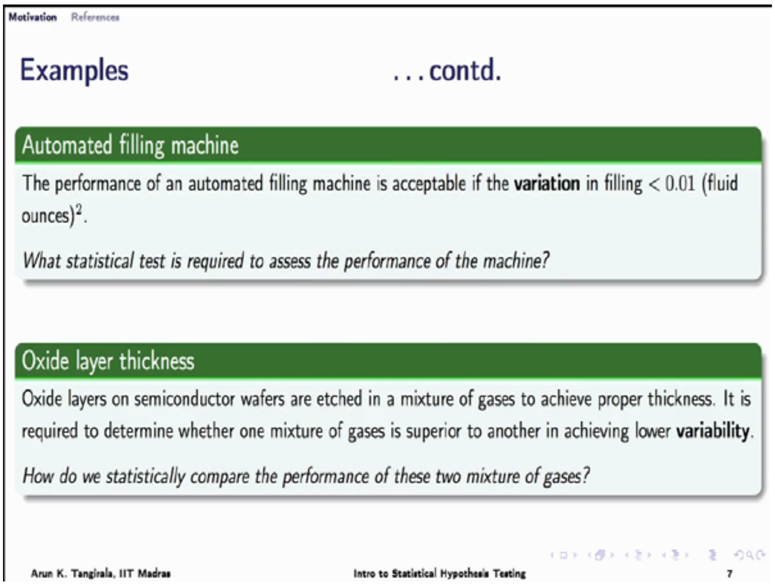
Again, how do we determine whether this claim is correct? For example, do this average pull off forces really differ as the engineer believes so, or is just by chance in that experiment that the engineer has obtained a pull-off force of 13 and 13.4, it is very likely. The major challenge and all of this, whether it is comparing averages or looking at average of an individual product or a random variable and so on, is that we do not get an opportunity to look at all the possibilities. Here for example, we are asking if the difference between pull-off forces is just by chance or is it that truly there is a difference between pull-off forces. We would able to answer this provided ideally, if we had the access to all possible situations corresponding to this wall thicknesses, that is I make all possible nylon connectors of this 2 different thicknesses and then examine their pull-off forces, but that is not possible in practice, I can only look at a few specimens which the collection of, which we call as a sample in statistical data analysis and based on this samples I am supposed to answer this question. Whether truly the difference between the pull-off forces is by chance or that there is systematic difference.

Similarly, when I look at, let say 2 different training methods that I have doctorate. Let us say as a head of a school for training the teachers of my school, I would like to know if one method is effective than the other. So, let say we have this 2 methods, method A and B and the intention is to determine, whether method b is more effective than method a. On an average, again the reason we are using the term average is because let say a teacher goes through training method a and another teacher goes through training method b and these teachers, respectively in turn go and you implement them in their classrooms. If you were to select one student at random from each of these classes then it may turn out that even though truly method b is effective more effective than method a, the student from the teacher who went through method b can perform poorer than the student who has learnt from the teacher who went through method a. So, that does not mean that method b is less effective than method a because as we discussed earlier, in the case of LED bulbs or solid propellant, there are going to be inherent variations in each observation, but we are looking at is collectively which we call as the samples, space or populations. Collectively if there is a big difference between these two populations, in this case of course, the populations here correspond to method b and method a, that what you mean by population is all possibilities.

So, once again what statistical procedure should be adapted to compare this training method? So, you can see slowly the kind of examples that we are looking at are not restricted to one discipline that is the point number one and point number two these are the examples or representatives of situations that we encounter even in daily life. It is not possible for me to completely list or exhaustible list all possible situations, but you can now quickly relate to what you would encounter even in your daily life. When you go to a shop to buy two different products of course, you may not have an opportunity to perform statistical test there, but in your professional life or in your research you would encounter these kinds of situations more than often and we will go through few more representative examples and you will hopefully be more convinced that these kinds of situations are frequently and statistical data analysis comes to your rescue.

Let us move on and look at another kind of example, where I am not interested in averages, but I am interested in what is known as variability. We have been saying that there are going to be variations in processes, again if you take the LED bulb, the power rating of a sink of one bulb, the actual power that is consumed by one bulb may be slightly different from the power consumed by another bulb from the same manufacturing process under the same manufacturing conditions and that is what we mean by inherent variations. In this example, we have an automated filling machine and I would like to see, if this performance of the filling machine is acceptable and let say we call it acceptable.

(Refer Slide Time: 13:12)

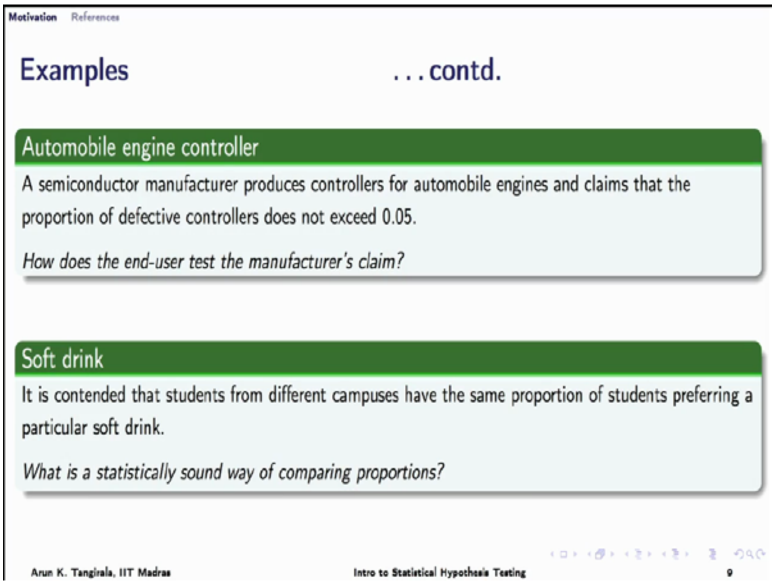


If the variation in filling this, filling could be bottles, it could be soft drink bottles, milk bottles or whatever the variation should not be more than 0.01. We call it acceptable, if the variation in the filling between two bottles is less than 0.01 appropriate units.

Why do not we want higher variation because if the variation is quite high, what this means is some bottles are going to be under filled and some bottles are going to be over filled, both of which are not good for the manufacturer and for the end user. Again what statistical test can be performed to assess the performance of this machine? So, now, we are taking about variability, earlier we talked about average. In another example, just in continuation of this variations that we want to test here like we did for the averages we look now at comparing variations across 2 different situations. Now, the example is that of, with the setting in a semiconductor manufacturing industry or semiconductor wafers. We know that semiconductors have oxide layers on them and these oxide layers are actually edged in an environment where there is a mixture of gases so as to achieve proper thicknesses. So, that is a medium in which the oxide layers are edged on to the semiconductor wafers. Now, what we would like to know is I have 2 different mixtures of gases as an engineer or as a manufacture. I would like to know which mixture of gases is providing me better thicknesses. Now, here in this case I would like to focus on the variability in the sense, whether I choose this mixture of gases or another mixture of gases from specimen to specimen there is going to be variation in the thicknesses and I would like to obviously minimize, so that I can claim that almost every product has a similar amount of thicknesses.

Here, the question is of course, not on comparing thicknesses, but comparing variations in the thicknesses. Of course, you can set up such a problem as well, where you compare average thicknesses as well. So, there are many problems that we could discuss, specifically in this example we want to ask the question. How do I test or how do I determine which mixture of gases is giving me lower variability or if the manufacturer claims something that one mixture of gases is giving him lower variability than the other? How as an engineer or as a tester you would like to, you would go about testing this claim? So, in all of this, we use the term statistical. You should understand that is a keyword there, which means we are relying on some statistics and we will technically define what is statistics later on. But inherently it means, we are relying on data, that is point number 1 and point number 2 is that there are going to be uncertainties and we have to somehow deal with this uncertainties in a systematic manner. Now, let us look at different situations, where we are not testing means, we are not testing or comparing variations, but we are testing what is known as proportions. This is also a very common situation that we can encounter.

(Refer Slide Time: 17:10)



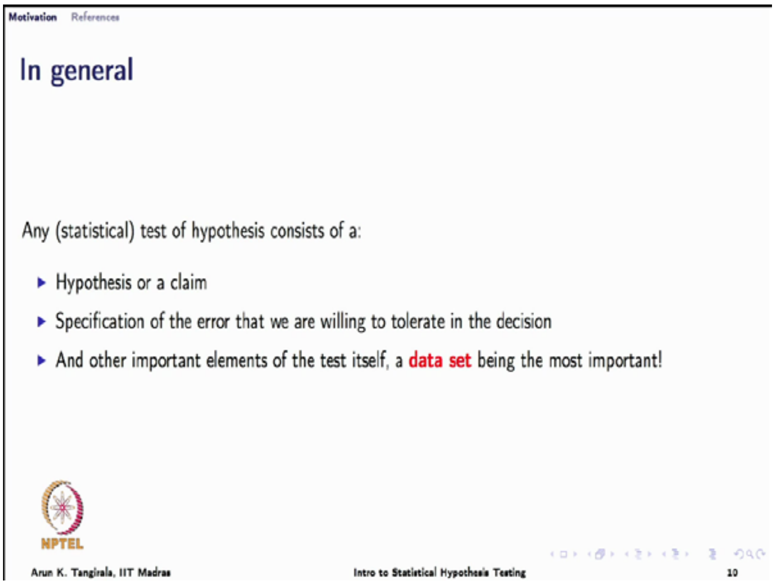
Again, let us look at a semiconductor manufacturer, who produces controllers for automobile engines and claims that the proportion of defective controllers does not exceed 0.05. Obviously, not every now on then, but occasionally any manufacturer would end up producing defective item or an item that deviates from the targets specifications. Now, as an end user how would you test this manufactures claim? Again, you would collect data and then go through a systematic procedure for testing this hypothesis that is been put forward by the manufacturer. So, you can see slowly that in all of these examples, there is a postulate or there is a claim and this postulate or claim is what we technically term as hypothesis.

So, in this case the hypothesis is that, at least put forth by the manufacturer that the proportion of the defective controllers does not exceed 0.05. Now, slowly a question emerges in our minds as to, does it matter whether I am looking at proportions? Whether I am looking at variations? Whether I am looking at comparison of variability or comparison of averages and so on? Does it really matter in terms of the procedures that I form? It turns out that the procedures thankfully remain the same and that is a beauty. In this course, that is why once we have through the foundations what we will do is, we will go through the generic procedure. In this lecture also, I will give you a quick review of what procedure is involved a generic procedure and then see how this procedure is applied to each situation. So, the moment you understand the philosophy of the generic procedure you are ready to become an expert in hypothesis testing. It is a procedure that most are really worried about or confused about and we will spend some time on that of course, in this lecture as well as more so in the coming lectures.

Now as with the previous cases we may end up in a situation where we are comparing proportions. So, here is an example which is got nothing to do necessarily with engineering here, it is got to do with human behavior or human psychology and human satisfaction so on. So, let us look at this example, we have students from 2 different campuses, it is easy to relate this example and it is contended that the students from 2 different campuses have the same proportion of students preferring a particular soft drinks. So, that some soft drinks we do not want really worry about the names of the soft drinks. There is some soft drink that the students are interested in naturally and the contention here is that the same proportions of students prefer this soft drink across 2 different campuses.

Now, once again, how would you go about testing this? In your mind, you would have already now started building the procedure. You would randomly select individuals from 2 different campuses, just ask the question as to what soft drink he or she, this particular student would prefer and note down the data and then come back and analyze the data. The key there is randomly selecting the individual, obviously we do not want to have a biased opinion; we do not want to form a biased opinion. Therefore, it is very important that you select the student at random and that is the key, one of the key things in collecting data for hypothesis testing. We will keep emphasizing the need for random sample that is a technical word that is used to indicate that there is no systematic bias whatsoever in selecting the student in this example or selecting this specimens in the previous example where we are looking at proportion of defective controllers or in all the previous examples as well. So, slowly now we understand, that we are relying on statistics, we are relying on data and this data has to be collected randomly and then some kind of procedure, statistical procedure has to be adopted to answer all these questions.

(Refer Slide Time: 21:52).

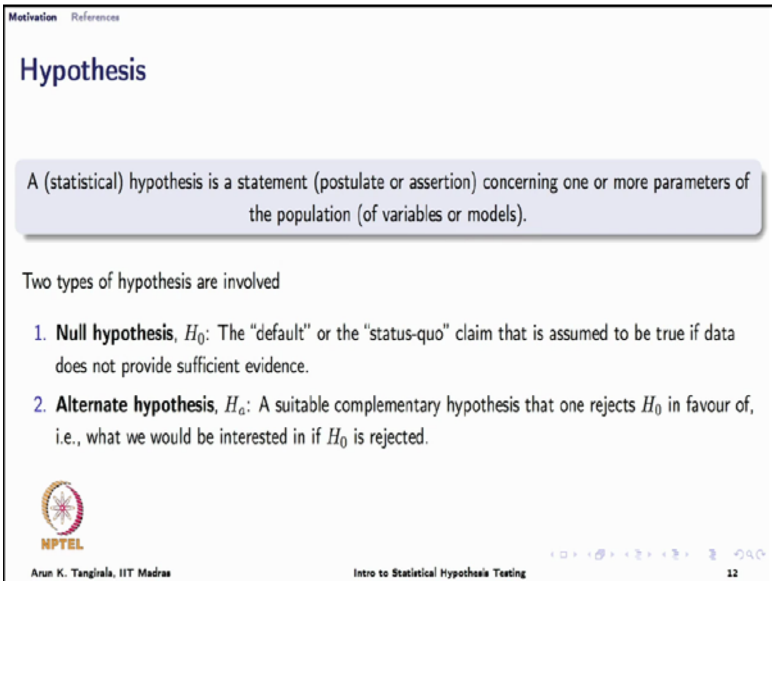


Now, in all of this, if you take the common things, it is clear that any statistical test of hypothesis consists of few basic elements and the core element is a hypothesis itself. Obviously, if there is no hypothesis, is nothing to test and there is no need to sit through this courses as well. So, at the core we have hypothesis or a claim. We will quickly define what is hypothesis and what we have not talked about in the previous example is something that is very important to hypothesis testing, which is the specification of the error that we are willing to tolerate or accept in the final decision. Inherently, what this means is, in every hypothesis test there is going to be some inaccuracy, that is let us say I come to a conclusion for the previous example that we have seen defective controllers. Let us say, I collect randomly a few specimens from the process as it is manufacturing this controllers and I carry out my data analysis, hypothesis testing and I come to a conclusion that the manufacturers claim cannot be rejected.

What is the manufacturers claim in the previous example, that the proportion of defective controllers does not exceed 0.05? Now, it does not mean that our decision to reject the manufacturers claim is 100 percent accurate or is accurate. There is a chance that I must have made an error in the test or finally, in not rejecting the manufactures claim. The truth may be that indeed that the proportion of defective controllers is greater than 0.05 and it is that fact, that we have to admit in all hypothesis test, that there is going to be an error in the final decision, and why is this error occurring; simply because I am not looking at the entire population.

So, in this example entire population would mean all this specimens that the manufacturer has ever produced, which is impossible to access or even analyze. We rely on what are known as samples, that is a subset of this population and as we will see later on there are a few factors that affect the final decision and some of this factor depend on the data, some of these factors depend on the statistics that we choose and the distributions so. But, predominantly it depends on the inherent uncertainties in the data itself which are beyond our control and which we cannot have a complete picture of. Therefore, in every hypothesis test it becomes important to stay upfront, how much error we are willing to tolerate. We will call that as a significance level, but we will come to that. So, then of course, there are this other important elements of the test, which is the data set. Of course, the key is how you sample, that is how you obtain this data and as I have said earlier it is important to randomly sample the specimens or the students and so on. So, that you do not introduce any bias at the stage of data and positions. So, data is of course, a core, but the most important thing is the hypothesis therefore, it is important to understand what hypothesis is technically and how do we frame these hypothesis in a statistical appropriate manner.

(Refer Slide Time: 25:41).



So, let us begin quickly with what is hypothesis? This lecture we will only focus on hypothesis definition. In subsequent lectures, we look at technicalities as well the other elements of hypothesis testing. So, what is a hypothesis? Generally, followed definition is, hypothesis is a statement which is either postulate or assertion concerning one or more parameters of the population; where again here population should not be thought of as a dictionary meaning of population of individuals and so on. Here, in statistics, population refers to all possibilities and this entire population may be characterized by certain parameters such as mean, variance and so on and this terms, mean and variance come into picture because again we treat the different possibilities in the population as kind of a random that is they are unpredictable and there those technical terms will also become clear later on.

Now, what we mean to say here is that, in every hypothesis testing or there is a statement that we may concern one or more parameters. It is good to begin with a single parameter problem and may be in this course, perhaps we are not going to look at simultaneously testing of many parameters, but we are going to look at different parameters of the population. So, if you relate this to the previous examples then you will see that the parameters that we are talking about are mean, variance, proportion and so on. Now, there are 2 different kinds of hypothesis that are used in a hypothesis testing. One is called the Null hypothesis and the other is called the Alternate hypothesis. Now, that may be the beginning of confusion for many as to why there should be 2 different kinds of hypothesis when we have only one definition for a statistical hypothesis. The reason is as follows, the philosophy in hypothesis testing is before I collect the data I make a certain claim, whether it is a manufacturer or the end user and so on; whatever, there is a claim that is made by someone before the data is collected even if the claim is not made, there is a default understanding. So, common example that is given is that of a court of law. In most countries, the default assumption is every citizen in the country is innocent which may seem a bit absurd to begin with, but the benefit of doubt always goes to the citizen. So, the default claim is that when a lawyer is arguing a case in the court, the lawyer, the judge and everyone in the room remembers that before the evidence is presented and the cases argued upon the default hypothesis or the claim is that, the accused is innocent.

So, the benefit of doubt always goes to the innocent that is how most courts operate.

They say that essentially if you look at the Indian court of law and as I said many other countries that the innocent should not be punished, the guilty may goes caught free, but we are more worried about the innocent. Unnecessary, we should not punish the innocent so, therefore, the default is that this person who is accused, standing in the box is innocent and it is the prosecutors burden to provide enough data or as we call as evidence to show that this null hypothesis does not stand that means, this null hypothesis that accused innocence has to be rejected and if you recall the judge always gives out a sentence saying that sufficient evidence.

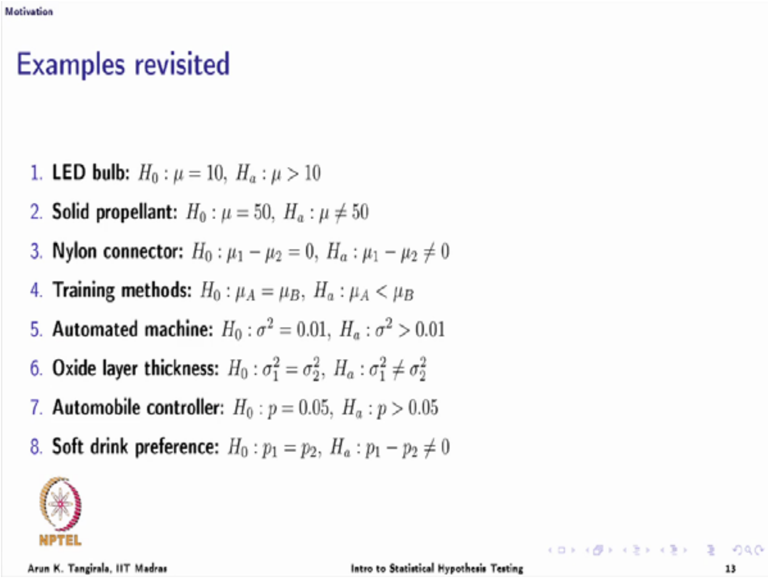
For example, sufficient evidence has not been provided to say that this person is guilty never does the judge give out the claim saying; yes, I am convinced that the accused is innocent. It is very hard to verify or set the innocence of an individual, it is very hard right, but what is relatively easier is to disprove the innocence and the prosecutor is trying hard to provide evidence based on his or her faith that yes, the accused is innocent. To provide evidence that yes, this person who is being called innocent is not innocent and a judge is also looking at both sides, both the lawyer, who is defending the accused and the prosecutor evidence. Both evidences are being examined upon and finally, the judge comes to a conclusion. So, what the judges actually performing in a court room is a hypothesis test. So, therefore, there is this null hypothesis, that is kind of a status-quo, if sufficient evidence is not provided then null hypothesis stays. The alternative hypothesis is a complimentary thing that typically constitutes, what we want to fall back on if the null hypothesis is rejected, that is suppose we say that going back to the previous example of defective controllers, the manufacturer claims that the proportion of defective controllers is less than 0.05. Now, as an end user of these controllers, let us say I am the engineer in an automotive industry, who is receiving who is ordering these controllers from the manufacturer, I would be interested in knowing whether this proportion of the defective controllers is greater than 0.05.

For example, I want to refute the manufacturers claim. So, the intention or the purpose will determine what the alternative hypothesis is. The null hypothesis is typically of an equality kind of hypothesis and will come to that. So, in that situations a null hypothesis would be the proportion of defective controllers is 0.05, which is the critical thing and the alternative hypothesis would be the proportion of defective controllers is greater than

0.05. So, as end users I randomly sample the specimens and then I subject it to a statistical test and see if there is enough evidence to reject the null hypothesis. The moment it is rejected it goes in favor of the alternative hypothesis and then I can say that the manufactures claim is rejected.

So, the alternative hypothesis is the key and that is where a lot of confusion arises, null hypothesis is relatively easier to frame.

(Refer Slide Time: 33:14).



So, if you look at all the examples here for our discussion and for your understanding, I have given the null hypothesis and the alternative hypothesis. We will not go through all of them, but we will just randomly select a few of this and talk about the null and the alternative hypothesis. So, let us look at the solid propellant example, the thing that we want to test there, remember what we said is we do not want lower or higher burning rates and therefore, then if I do not find enough evidence that is in the absence of sufficient evidence, the null hypothesis is that the average burning rate is 50 centimeter per second.

When I find enough evidence in the data to reject this null hypothesis, what would I like to be rejected in favor of and I would like it to be rejected in favor of, that is not I would like it to be, but if the null hypothesis is rejected, it has to be rejected in the context of the alternative hypothesis, which is that the burning rate is not 50. So, the moment I reject the mu equals 50, if it becomes in favor of the alternative hypothesis and I come to the conclusion that this solid propellant is not suited for the application that I am looking at of course, it is possible that the alternative hypothesis can be less than kind of or greater than kind of situations and we have quite few situations like that. So, for example, let us look at the training methods example, where we were looking at whether method B is more effective than method A for training the teachers. Again, the null hypothesis that both the status-quo is that both methods are identical.

Now, the alternative hypothesis here is that what do I want to test, I want to test that method B is more effective than method A and the mu here refers to the average performance. Let us say or the average score that the teacher score or the students or the teachers score after going through the training sections. If method B is more effective on an average I believe that this mu will be an indicator of the effectiveness of the method and I would like to test if mu b greater than mu a. So, you can see typically what I want to test goes and sits in the alternative hypothesis and this status-quo in all these 8 examples are all of equality type and there is a reason for this to be equality type. I will not go into the reason right now, but that is a reason that keeps bugging our minds. I have seen in many students asking this question as to why the null hypothesis should be of equality type and of course, and I was learning also I had a similar question.

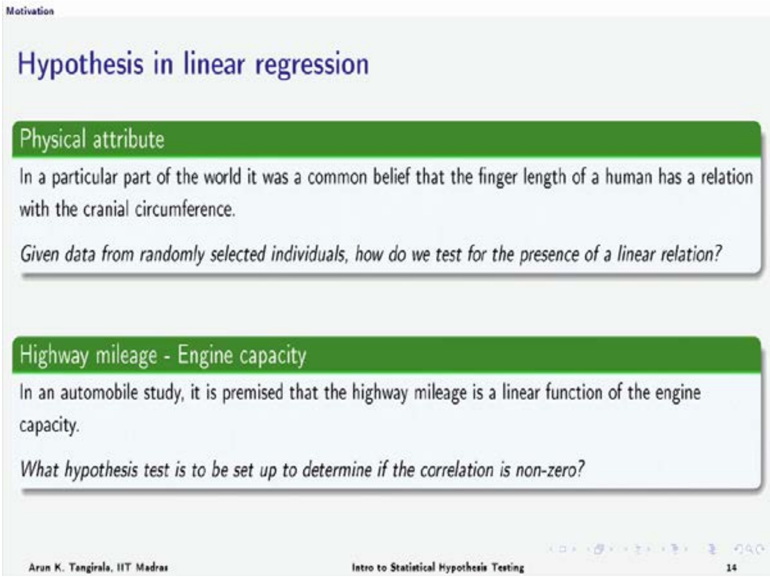
Now, let us look at another example here of that of automated filling machine, we said that we would like to know if the filling machine performance of it is acceptable. Now, here there are 2 possibilities, for me as an end user I am really worried whether I should buy this automated machine or not. Now, it depends on what I want, if I am more inclined towards rejecting, that is buying this or not buying this automated filling machine. Then the alternative hypothesis becomes sigma square greater than 0.01, but if I am as a manufacturer I want to test whether my claim, that it is actually less than 0.01 is correct, then it becomes difficult because that is the situation here that you have to look at. The critical value is 0.01 and as an end user and manufacturer we find if the variance does not exceed 0.01. So, the null hypothesis is actually taking care of that what I should be more worried about even as a manufacturer or as an end user, whether I am making filling machines or buying filling machines that do not give me acceptable performance.

So, if the null hypothesis holds then both the manufacturer and the end user are happy; that means, in the sense they are content with it. What I should be worried about either ends manufacturer, end user is, whether the filling machine is giving me unacceptable performance and that is where the alternative hypothesis is that sigma square is greater than 0.01. You may be tempted to write the alternative hypothesis as sigma square less than 0.01, but it does not serve any purpose because then both the null hypothesis and the alternative hypothesis would mean the same thing by the machine or keep going with the manufacturing process, no worries, it should be complimentary, the alternative hypothesis should always be complimentary to the null hypothesis, such that if one is rejected then the other should be the situation. So, here remember that is only caution that one as to exercise in framing the alternative hypothesis. So, let us look at one final example where we are looking at soft drink preference. Again here, this is of say proportional comparison of proportions. Remember, we wanted to know rather the contention is that the same proportions of students across 2 different campuses have the same preference for the particular soft drink. So, the null hypothesis again the status-quo is that, it is correct that on both campuses we have the same proportion of students not the number of students, but proportion of students preferring this soft drink.

Now, suppose that is not the case then the alternative hypothesis should be not there. There is a difference we are not interested at this moment in knowing whether one campus has larger proportion of students preferring this soft drink over the other. If that is the case, then the alternative hypothesis would be difference. So, for example, if I want to know that students on campus two or the campus two has a larger proportion of students preferring the soft drink than in campus one. Then the alternative hypothesis would be that p 1 is less than p 2, but here I am not interested in that the only contention that I want to test is that there is a difference between the proportions where you see again. Here the alternative hypothesis is framed on what I want to test and the null hypothesis it is easy to frame. It is always of an equality sign, what you want to test typically goes and sits in the alternative hypothesis and you can relate this even to the court of law where the person is brought in has been accused and what is being tested for is whether the person has committed that particular crime and so on. So, the alternative hypothesis is that the person is not innocent and the null hypothesis is that the person is innocent always equality type of sign.

So, there are two more situations that we will consider in this course and which are very common in data analysis.

(Refer Slide Time: 41:06)

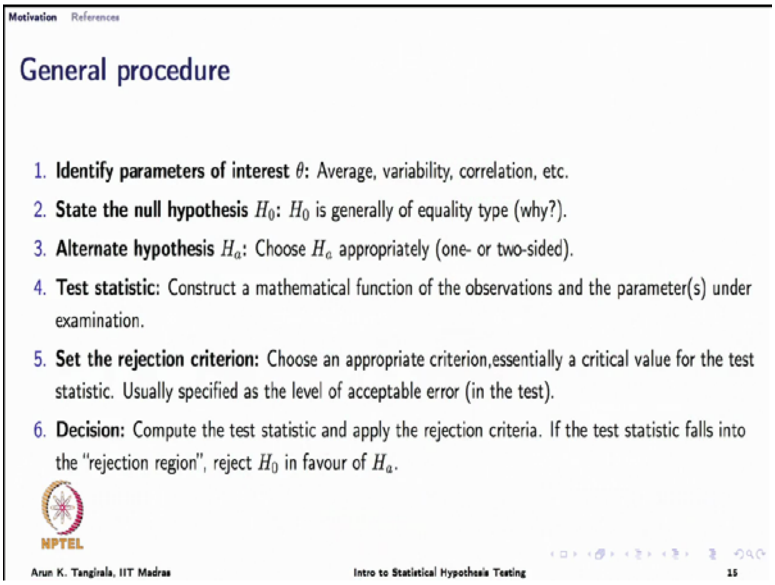


Where you encounter hypothesis test and this is in the context of linear regression does not mean that a non-linear regression you do not encounter hypothesis test, but that is beyond the scope of this course. Linear regression is something that we carry out routinely in data analysis, where we are trying to explain one variable as a linear function of another variable for prediction purposes. So, let us look at this simple example where in 18th and 19th century, there were in a particular part of the world it was a common belief that the finger length of human being as a relation to the cranial circumference, that is a circumference of our skull or head. Now, it sounds a bit weird, but that is the belief and maybe it is not, we do not know. One way is to argue this biologically go in to biology and find out, whether this is true there is a relation and so on, but there is a lot of hard work of course, it is scientifically correct way of looking at it.

The other sort or the approaches is to use statistics, collect data randomly, select individuals record, their finger lines and the cranial circumferences and see if there is a relation in particular. We would like to, if this is a linear relation there could be a non-linear relation too. So, given data from randomly selected individuals how do we test for the presence of a linear relation? So, very quickly linear regression is something that all of us are familiar with you knows trying to fit a line essentially between 2 variables and a line is characterized by slope and intercept and in this case we are asking whether the slope is 0; that would be the null hypothesis. Now, you think of what the alternative hypothesis is and in another example again related to linear regression. Suppose, in an automobile study, I am interested in knowing whether the highway gasoline mileage is linearly related to the engine capacity, is it correlated with it? Again here we can ask, what hypothesis test has to be setup to determine if the correlation is non zero and correlation is actually technical term that we are using later on.

We learn what correlation is, it means linear dependence. So, we would like to know if there is a linear dependence of the highway mileage on the engine capacity and that is very useful thing to know. For an automobile manufacturer again, you will have to randomly collect data and then perform a statistical test. Linear regression and correlation are very closely type to each other as we will know in this case again think of what would be the null hypothesis and what would be the alternative hypothesis. I am not going to give you the answer right now, but think of it and then of course, if you have questions you can always ask on the forum. So, let me quickly go through in few minutes, we will wind up the lecture quickly, go through the procedure and we will go through this procedure in much detail in the coming lectures.

(Refer Slide Time: 44:26)



So, the first step in statistical hypothesis test is to identify the parameter, whether you are going to test for average variance proportion and so on and then of course, as we went through just now, state the null hypothesis and the alternative hypothesis that is critical. If you state the alternative hypothesis wrongly your test may go in vain, remember that and then we chose what is known as a test statistic.

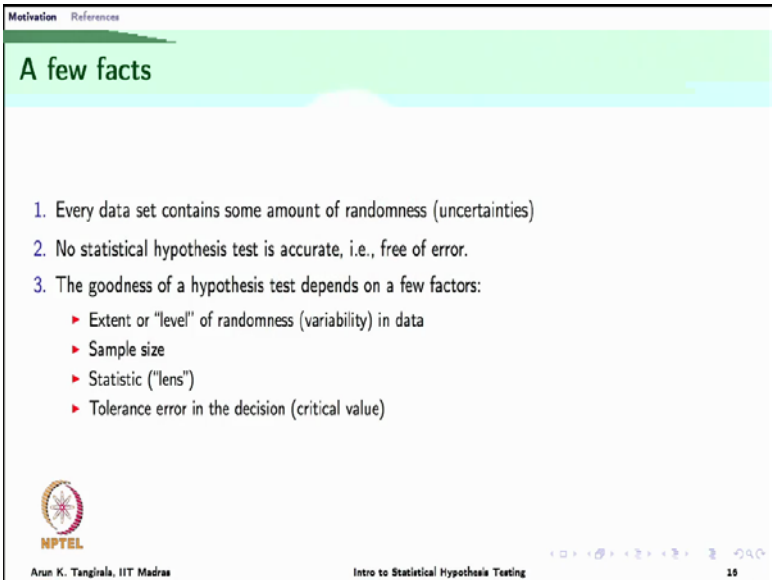
I do not want to go into the technicalities of this term, but an analogy that is usually given is that you have data with you and you have a null and alternative hypothesis. So, now, you are like a detective trying to find evidence in the data to see whether you should reject or not reject the null hypothesis that is technical term. We never say accept the null hypothesis like I explained earlier. So, here test statistic is like a lens that the detective use as to search for evidence. It basically goes into the data extracts the information that you want relevant to the null hypothesis and comes up with some number at the moment think of it as a number that eventually come out with that is some mathematics and statistics involved, which we will know gradually, but eventually the outcome of this exercise you will get a number and now you set a rejection criteria that is now you ask whether this number that I have obtained from the test statistic, is it acceptable or not?

As a simple example going back to the solid propellant case, the null hypothesis is that the average burning rate is 50 centimeter per second and the alternative hypothesis is not. So, what do I do? I would randomly collects specimens and put them together, what is known, what we call as a random sample and run an experiment where I determine the average burning rate for each specimen, note them down, not the average, but the burning rate for each specimen, note them down and then take the average of the burning rates for each specimen.

Let us say I have collected 20 specimens. So, I have 20 different burning rates with for each of that is for corresponding to these 20 specimens and take the average of these 20 readings. If that average turns out to be, let us say 48.5 that is a number that I get and that is kind of a statistic that I am generating from the data. Now is this 48.5 to be considered close enough to 50, 50 is the testing the ideal number that I should have got, but I will not get it because I have not looked at the entire population, I only taken a sample. So, is the difference between 48.5, which is average that I have obtained for the sample and 50, which is the target that I have is that difference by chance or is that a systematic difference between that and there you set a rejection criterion. Remember, we said earlier that we will make an error whatever we do even if I collect 100 samples, we are going to end up with some error may be lesser error than that with 20 samples. So, one has to specify the level of acceptable error at this stage and then call the shot that is whether the difference between 48.5 and 50 is to be considered systematic or by chance and then you make a decision. So, that is the general procedure for a statistical hypothesis testing.

An entire course is about going through this steps a particularly that of computing the test statistic and setting up the rejection criterion and analyzing the resulting error. How do we choose the sample size to minimize the error are and then of course, how do we what is the impact of that on a final decision is what we are going to predominantly learn in this 10 hour course. Before I conclude, let me state a few facts maybe some of which I have already stated.

(Refer Slide Time: 48:39)

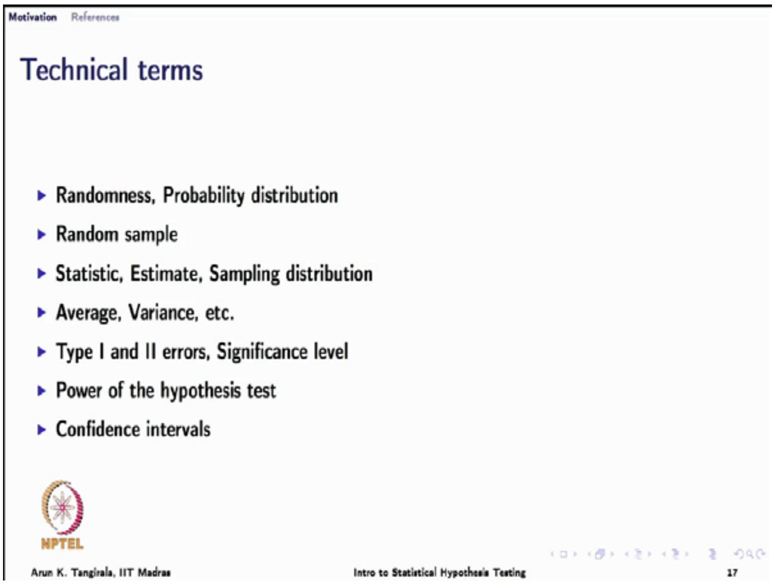


The first fact in hypothesis test to remember is every data set contains some amount of randomness is very important. That is the reason why we are challenged by this hypothesis, that is the main challenge if everything is deterministic there is no challenge in hypothesis test and secondly,

As I mentioned earlier no statistical hypothesis test is accurate there is always going to be some error and the goodness of a hypothesis test that is how low an error we can achieve depends on a few factors. One is the extent or the level of randomness in the data, how much uncertainty do we have? Right now it is all qualitative, but very soon will quantify what we mean by level of randomness and the sample size? That means, how many specimens do I collect? How many individuals I may surveying and so on and the statistic itself that is the lens that I am using to search for the evidence if I choose better lens of a higher quality that means, more rigorous test statistic I may get a different result.

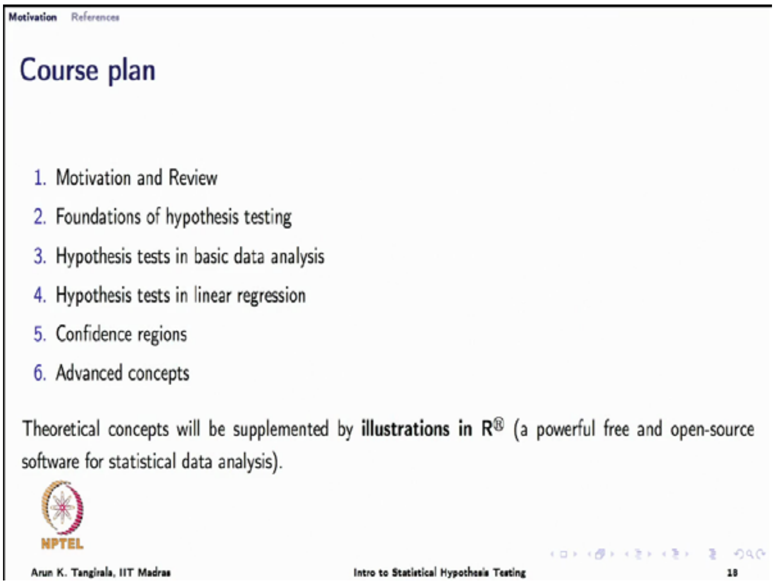
So, that is very important and there of course, there are some technicalities there which will learn later on and then the tolerance error is the acceptable error that I have for the final decision that I have make that also as an impact on the goodness of the hypothesis test. Later on we will use the term called power of a hypothesis test which will characterize the goodness of a test hypothesis test. So, in this course we are going encounter a number of technical terms I am not going to read out all of them this is just for more for your reading.

(Refer Slide Time: 50:22)



But some of the things that we have already encounter. For example, random sample, randomness statistic, average significance level and so on. So, understanding this technical term is very important and we will define each of these technical terms as we go along and of course, if you have any questions you can actually ask them on the forum. I want to conclude this lecture with a few seconds discussion on the course plan.

(Refer Slide Time: 50:49)



We will initially go through a review. We have already gone through hopefully quite a few motivating examples and now you have to really motivate to go through this course and we will go through review of probability and statistics concepts, all relevant to hypothesis testing and then go through foundations that is understand the generic procedure a lot more in the detail and use that procedure to learn how to apply this in basic data analysis. What we mean by basic data analysis? Is what the kind of examples that we have seen, testing of the mean, comparison of means, testing variances, compare them and so on and then also look at hypothesis test in linear regression finally, there is something that we have not talked about known as a confidence regions. I will talk about it later constructing confidence regions for a parameter that I am estimating is equivalent to performing a hypothesis test. So, that is an alternative way of computing the hypothesis test at this point I do not want to talk about it and then look at a few advance concepts. So, that is the course plan for us and here are a few references. Of these references in particular, what we will closely follow are the books by Montgomery and Runger.

(Refer Slide Time: 52:13)

It is a fantastic book to have titled: Applied Statistics and Probability for Engineers and the other book by Ogunnaike; this also an excellent book on Random Phenomena: Fundamentals of Probability and Statistics for Engineers.

Now, although these 2 books are for Engineers, if you take a look into the contents of the book, lot of examples are non-engineering type example. So, please do not get carried away by the title of this book. They are pretty much universal, all the concepts and principles that are contained in these books and other books are pretty much universal as you can see that a lot of emphasis on using statistics for engineers, but it does not mean statistics is not used else were. Statistics is a subject that is used in almost every field where ever you are collecting data and you are analyzing that experimental data. So, again do not get carried away by the titles, read those books. They are excellent in terms of content and also the diversity of the examples that are presented.

So, hope to see you again in the next lecture and of course, throughout the course and hope that you are going to have a lot of fun and joy learning the subject.

See you.