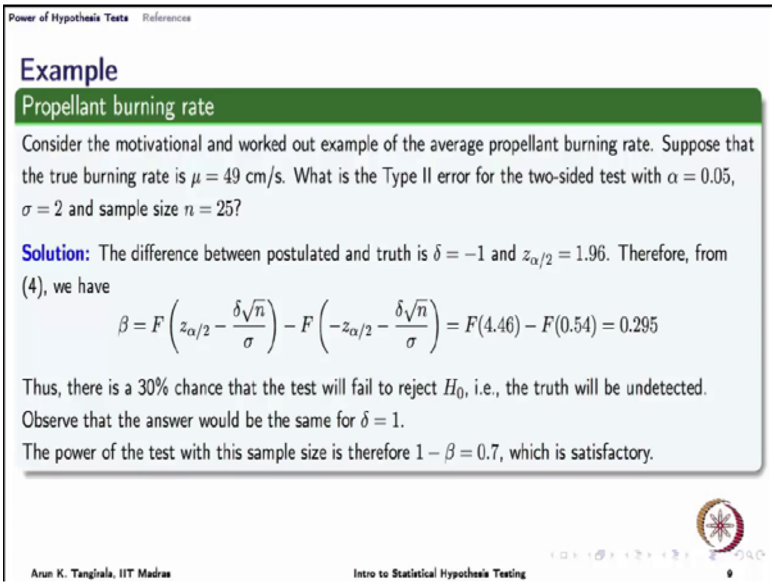
Introduction to Statistical Hypothesis Testing
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 19
Factors affecting hypothesis test

(Refer Slide Time: 00:01)



We will look at an example of the propellant burning rate, our favorite example from the book by Montgomery and Runger. We had conducted a hypothesis test with the postulated value being 50 centimeters per second. And, we did a hypothesis test with sample size being 25, and the standard deviation was specified as 2 centimeters per second. Now, let us ask the question, with the same sample size, with the same data that we had, would we have, what is the power, what is the probability of rejecting the h naught when the true average burning rate is 49 centimeters per second. We never asked this question earlier. There, we were only testing if the postulated value for the average burning rate with the data as evidence should hold, or should be rejected.

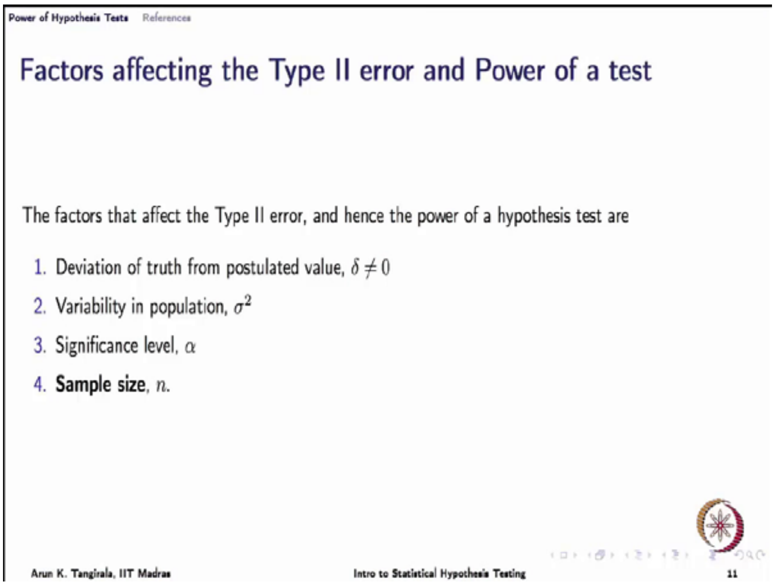
But now, we are saying, we are asking this question that, if the truth is 49 centimeters per second, can my data detect it? Rather, my test statistics can detect it? So, delta, in this case, is minus 1, right. Delta is the truth minus postulated value, 49 minus 50; and alpha as usual is 0.05. And, let us assume that, we are looking at the probability of Type II error. The question asked, Type II error, but actually, we are asking for probability of Type II error; for the 2-sided test, that also has to be specified. We know that, for the 2-sided test, we look at z alpha by 2, and alpha is 0.05, implies z alpha by 2 is 1.96. Therefore, from the expression for beta, we can calculate beta to be 0.295. Now, how did we get this? Well, we just plugged in the value of z alpha by 2; delta is minus 1; n is 25 and sigma is 2 and all of this is given. We can either look up the standard charts; maybe, you should make it a habit if you are writing the exam, to look up the standard charts, because in the exam you will not have access to a computer to find, to use p-norm, or any other software package.

So, all you look up in the standard chart is the probability that, the random variable will take on values up to 4.46, and likewise here, 0.54. Therefore, you get 0.295. What does this mean? There is a, that, there is a 30 percent chance that, we will fail to reject h naught, because power is 1 minus 0.295; assume it is 0.3 for discussion purposes. Power of the test denotes my ability, or the probability of rejecting h naught when it is false. So, obviously, beta denotes by definition, the probability of failing to reject h naught when it is false. What kind of false, thing we have here? Well, the truth is minus 1 in deviation from the postulated value. What does this mean? If I were to repeat this calculation for a different truth, let us say, 48 centimeters per second, or 40 centimeters per second, or even 30 centimeters per second, and so on. I would get different values of beta, right. Suppose, the truth was that, I was testing, sorry, I was calculating the beta instead of mu 1 being 49, suppose, mu 1 was 40 centimeters per second; that means, we want the test to detect a very large difference between the postulated and the truth.

Intuitively, what you think? Larger the deviation, will the probability, will higher be the probability of detecting the falsehood of h naught? What do you think? Obviously, it should, because, larger the deviation, more pronounced will be the effect, and that you should see in beta. You should also be able to see quantitatively; that means, beta should take on lower and lower values; that means, the probability of making an error, what kind of error, rejecting h naught when the truth is really far away from postulated value should be lower and lower, as the truth is getting farther and farther away from the postulated value. So, that is what we mean by sensitivity of the test to deviations from the truth. But, it is no big deal for any hypothesis test to detect large deviations. What we want the hypothesis test to be able to detect even small deviations with small sample sizes. Now, that is a conflicting requirement; pretty much like what we had in confidence interval. We said having narrow confidence interval and high degree of confidence are both conflicting; likewise here, to be able to detect small deltas, with small sample sizes, is going to be pretty tough, because, it is conflicting. If I want to be able to detect small delta for a fixed everything else, meaning alpha and sigma, I would need larger values of n, and we will see that expression shortly.

So, the question is therefore, how can we improve the power of a hypothesis test? In this case, power is 0.7, which is OK. Now, can I increase the sample size to improve the power? For that, we need to understand the relation between the sample size, delta and the power.

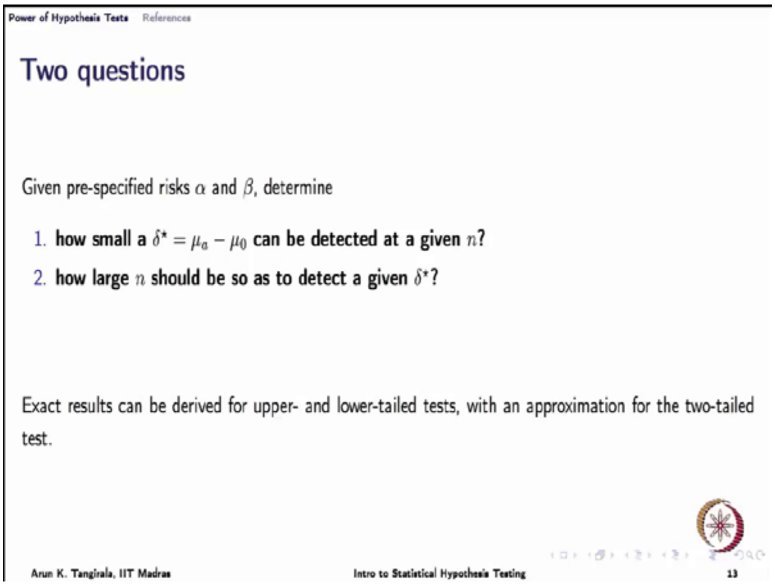
(Refer Slide Time: 05:55)



Let us do that now. We have talked about this earlier as well, what factors affect a hypothesis test. Now, we are asking what factors affect the Type II error. One is a delta, as we have just discussed, larger the delta, easier for the test to detect; but the challenge is in detecting smaller deltas. And then, the variability in population, we do not have a control over that. Significance level, alpha, we cannot really go too low. And, the only factor perhaps, if anything, is the sample size. Therefore, we shall ask these 2 questions.

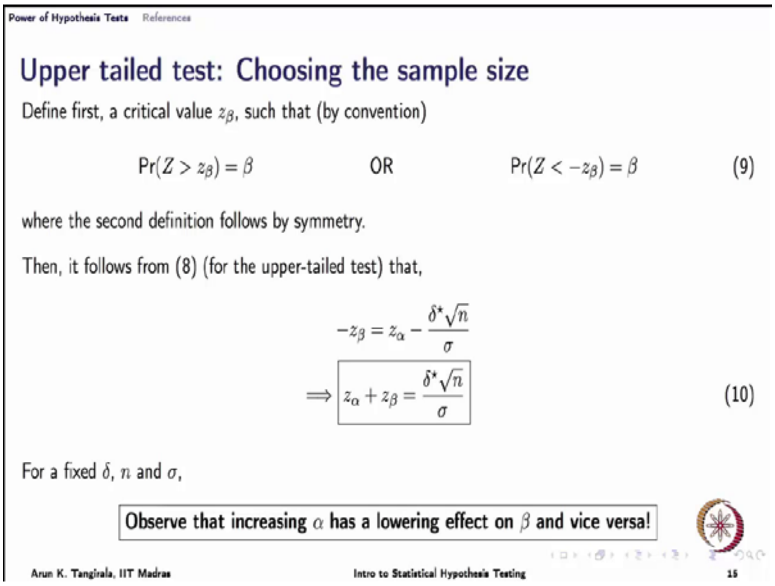
How small a delta can be detected for a fixed sample size? I have decided, let us say that, I will collect 50 observations, or in this case, the propellant burning rate, for the propellant burning rate testing I have chosen to collect 20, work with 25 observations; not samples, 25 observations.

(Refer Slide Time: 06:42)



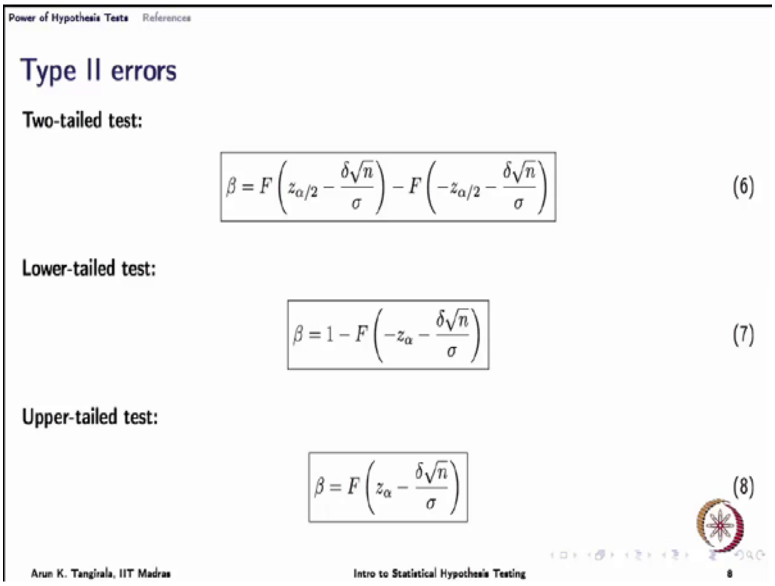
How small a delta can I detect with a sample size of 25, for example? And, likewise, how large a sample size should be, to detect a given delta? And, we know; we have just discussed that, intuitively, there is an inverse relationship. If the smaller the delta is, that means, the more sensitivity I want for the hypothesis test, the larger should be the sample size, and vice versa. As the sample size decreases, the deviation that I can detect best becomes larger and larger, alright. So, what we shall do is, we shall derive certain results, and it turns out that, we can derive exact results for the upper and lower-tailed tests, but unfortunately, only approximate results for the two-tailed test.

(Refer Slide Time: 07:41)



So, we will define a critical value z beta, in the same way as we did for z alpha. There, we said probability, z alpha is such that, probability z greater than z alpha is alpha. Likewise, we shall define z beta such that, the probability of z greater than z beta is beta, or by symmetry that, probability of z less than minus z beta is beta. So, both are equivalent. Now, what we do is, we, for a given beta, we can find z beta, because that you can read off from the standard Gaussian distribution chart, or using R.

(Refer Slide Time: 08:31)



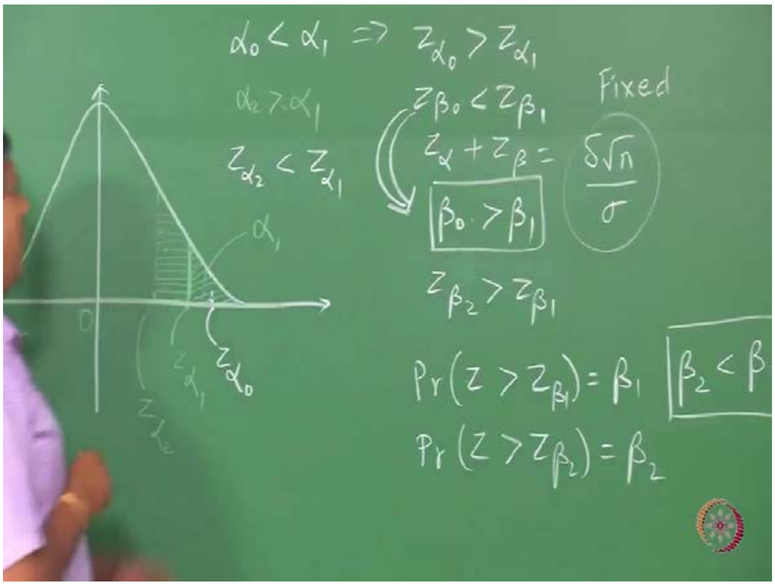
Now, because we are looking at the upper-tailed test, we pick the expression that we had in equation 8, which says that, beta is this CDF; Cumulative Distribution Function, the probability of finding z, some random, standardized, Gaussian random variable less then z alpha minus delta root n by sigma. Now, because we have defined beta to be probability of z less than minus z beta, naturally, the argument that we have in F becomes minus z beta. So, all you have to do is connect equation 9 and equation 8; that is all. Then, you see that, the z alpha minus delta root n by sigma is nothing, but minus z beta. And, as a result, we have z alpha plus z beta equaling delta star root n by sigma. They, we have introduced delta star, because this is now some kind of an optimal delta, or it is kind of a delta that exactly satisfies some equation.

And therefore, we have used the star. This is nothing, but actually, the smallest deviation that we can detect. This is the answer to the first question. So, z alpha plus z beta is delta star root n by sigma. This is the fundamental result that we have, in answering both these questions that we had raised earlier; how small delta should, can be, how small a delta that can be detected for a given n, and how large n should be for a given delta or delta star.

So, equation 10 gives you the fundamental relation that is used in hypothesis testing and answering the 2 questions that we raised earlier, but, in addition, it also gives us a very important insight into a point that we made early on in this lecture, and even in the course that we have made earlier, which is that, if I increase the Type I error, I can lower the Type II error, or, if I try to decrease the Type I error, I end up increasing the Type II error. Let us understand how equation 10 helps us in getting that insight.

So, let us now, for discussion purposes, we freeze delta, n and sigma, that is, we freeze the deviation of truth, or the postulated value from the truth, and of course, sample size we have fixed and sigma is fixed; for all these 3 terms, fixed to their respective values, increasing alpha has a reverse effect on beta. How do we understand that from equation 10? Let us look at that this way. When I increase alpha, that means, I am increasing the Type I error. What, let us say, alpha is 0.05 presently, and I increase it to 0.1. What happens to z alpha? By definition, it actually moves closer to the origin. Let us understand this graphically.

(Refer Slide Time: 11:31)



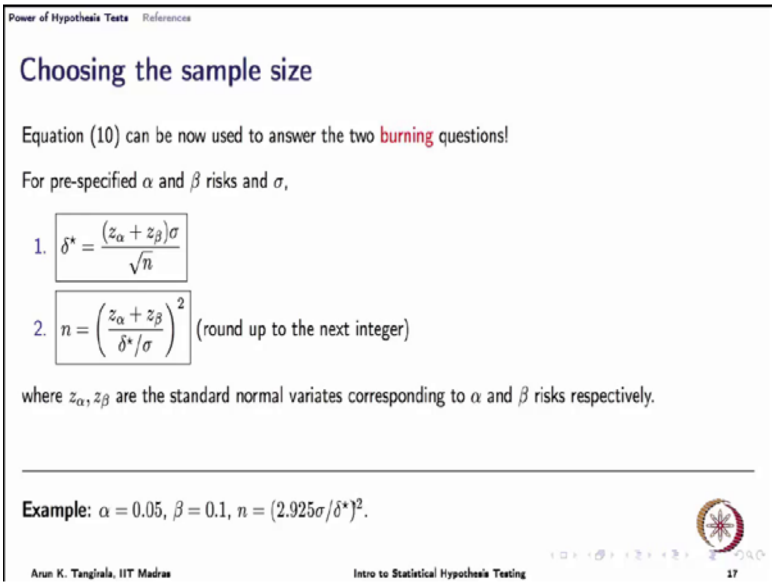
So, here, we have a standard Gaussian distribution, alright; and then, we have this, here is, let us say, your z alpha, initially. We know that, the area to the right of z alpha is alpha. Now, let us say, this is alpha 11. And now, I choose alpha 2, such that, alpha 2 is less than, sorry, greater than alpha 1, which mean then, let us call this as z alpha 1. Now, let us say, alpha 2 is such that, here, you have z alpha 2. And now, this area is also included in addition to this shaded area. So, what has happened is, increasing the Type I error has moved z alpha closer to the origin, because z is a, let us say, standard Gaussian random variable. And, we have here; the expression z alpha plus z beta is delta root n by sigma. So, this right-hand side here, is fixed for a specified value of delta and sigma and for a given sample size. What this means is, if z alpha increases, z beta has to come down. So, what happens by this relation here is, when I move from alpha 1 to alpha 2, z alpha 1, z moves from z, the critical values moves from z alpha 1 to z alpha 2; and since, z alpha 2 is less then z alpha 1, z beta 2 is greater than z beta 1.

Now, we know from definition that, the probability, the way we define beta, is the probability of some standard Gaussian random variable being greater than z beta. Now, you can check when you are looking at these 2 probabilities. So, if I plug in z beta 1 here I get beta 1 and likewise, z beta 2, I get here beta 2. So, the question is, whether beta 2 is less than, or greater than beta 1. We know that, z beta 2 is greater than z beta 1, and therefore, beta 2 has to be less than beta 1. So, when I increase alpha 1, that is, the Type I error, from alpha 1 to alpha 2, I have produced a reduction in the Type II error.

And, in a similar way, you can argue that, decreasing alpha 1, alpha from alpha 1 to some other value, let us say, alpha naught, such that alpha naught is less than alpha 1, would shift this critical value to the right; that means, take it farther away from the origin. Say, this is z alpha naught; it has moved away from the origin; moving away from the origin would mean that, z alpha naught is greater than z alpha 1, right. Since, z alpha naught is greater than z alpha 1, z beta naught has to be less than z beta 1. And, which would imply that, beta naught is greater than beta 1.

So, when you try to reduce the Type I error, the Type II error increases; this is exactly the kind of what (Refer Time: 15:29) between alpha and beta. It is not that alpha plus beta equals 1. Again, I have said this earlier, but there is a relation that, such that, in terms of z alpha and z beta, that is a thing. So, if you are looking at relation between alpha and beta, you do not have that, but, you have relation between the corresponding variates; z alpha plus z beta is fixed. So, if you try to reduce one, the other shoots up, and vice versa, for a fixed delta n and sigma; that is what you should remember.

(Refer Slide Time: 16:03)

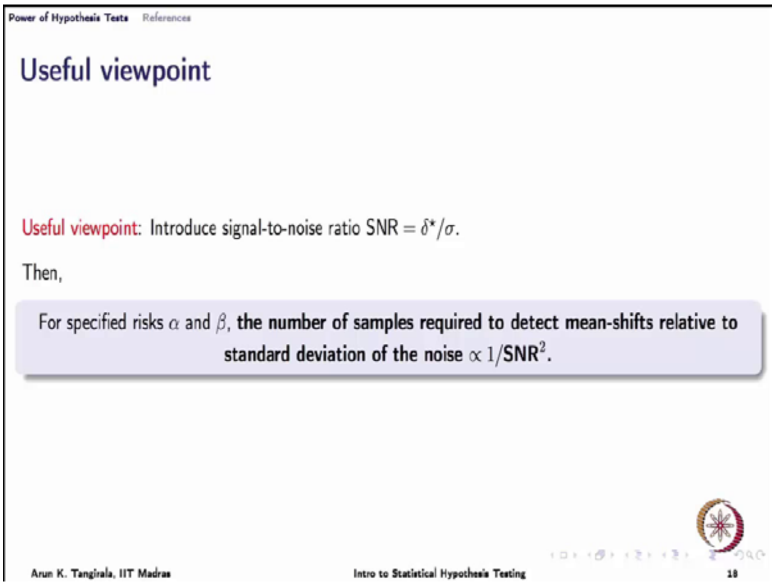


So, let us get back now to the 2 burning questions and answer them. The first question was, how small a deviation delta I can detect for a given sample size and of course, sigma, alpha and beta? And, the answer is pretty straightforward. All you do is, you rewrite this expression 10 in terms of known quantities. We know, since alpha and beta are specified, z alpha and z beta, you can read off from the charts; sigma is known, n is known and therefore, you have this expression.

And, the second one, which has got to do with, how large a sample size should be, for a given delta and sigma and alpha and beta? Once again, pretty clear from equation 10, as to how arrive at the answer. Now, there is no guarantee that you will get an integer to be conservative, you round up to this answer to the next integer not to the lower one; that means, you take the ceiling, not the floor. And of course, just to reemphasize, z alpha and z beta are the standard normal variates corresponding to your risks alpha and beta, very good.

So, for example; assume alpha is 0.05 and beta is 0.1; that means, we want a power of 0.9, the probability that, a probability of rejecting h naught when it is false, we want it to be quite high, 0.9. Then, you can plug in this value here. And, of course, we have not specified delta and sigma. There is a reason we have not done that. Under these conditions, this is the sample size formula. Now, you can use this for a known sigma and delta. Now, normally, you may not know sigma, you may not know delta. As an experimentalist, you, what you want, probably be able to answer is, if I know the relative values of delta and sigma, or the relative values of delta square and sigma square, if I know that, then maybe I can actually define, also arrive at the answer for the sample size. Now, this is something that is also asked in parameter estimation.

(Refer Slide Time: 18:19)



So, if you were to define the signal to noise ratio as delta by sigma, this signal to noise ratio is different from the usual signal to noise ratio that one encounters in signal processing, or data analysis in parameter estimation and so on. This signal to noise ratio has got to do with, what is the signal that I want to detect which is a deviation delta in presence of uncertainty, which is characterized by sigma. So, you may also define signal to noise ratio as delta square by sigma square; absolutely fine. But, if you have to define signal to noise ratio as delta by sigma, then, it is pretty clear that, the number of samples, should not be number of samples, should be number of observations, strictly speaking. So, the sample size that is required to detect mean shifts, this expression is only valid for known sigma or specified delta by sigma.

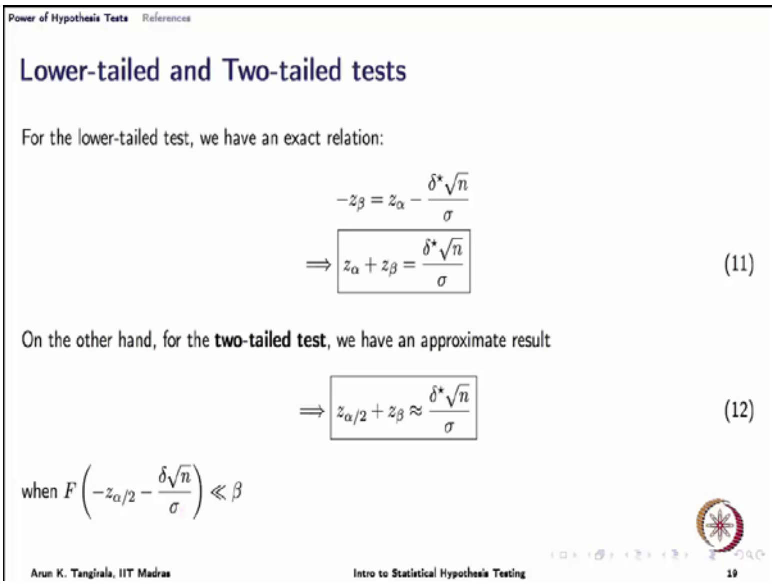
The sample size is inversely propositional to SNR square. Of course, if you are to define

SNR as delta square by sigma square, then, you would probably say, inversely propositional to the SNR. What does it mean? As the signal to noise ratio increases, that means, let us say, I have some level of randomness or variability, which basically is, probably due to noise, the delta should be much higher than the noise level for me to be able to detect it.

Let me give you a simple example. Suppose, you know, we all actually drive vehicles. Let us say, I take a 2-wheeler, or a 4-wheeler; generally, we hear some engine noise; we are not talking about really very posh vehicles and high class vehicles, but a vehicle that a common man uses. We do hear certain noise. Let us say, I am driving an automobile. I do hear some engine noise. Now, from time to time, an alert driver would listen to all the sounds in the automobile, to be able to detect any faults. Now, suppose, the fault, that is, the deviation of the current situation from normalcy, which we would call as, let us say, some abnormality, has a deviation of delta; that means, suppose, some component of the automobile has gone wrong and a new level of noise, a new kind of noise sets in.

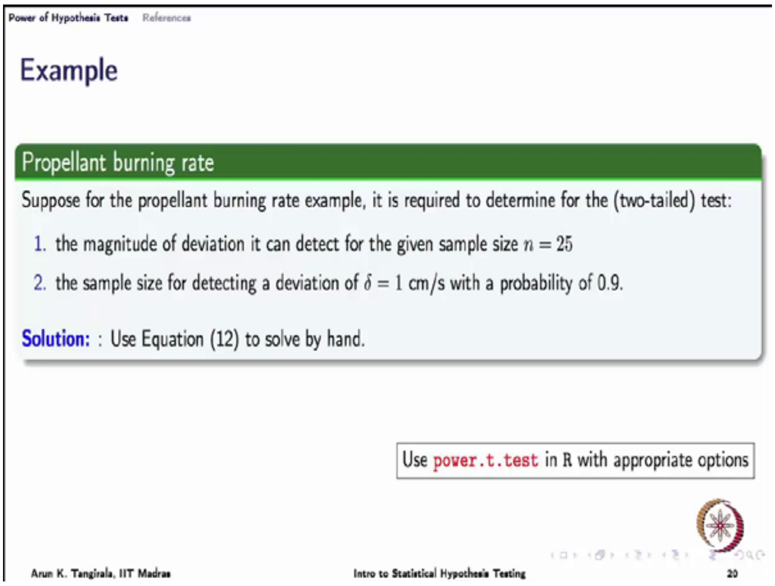
Suppose, this faulty component is making, is causing a noise, that is well within the band of the regular noise that I see, I will not be able detect it; but, if it makes larger noise, that means, delta is much greater than sigma, then, I can easily detect the noise. It is exactly the same scenario here, in hypothesis testing. The deviation of the truth from the postulated value should be large enough compared to the uncertainty, or the regular variability that you see, for you to be able to detect that something has gone away from what I postulate, or what I considered to be normal, to be correct; that is the story to remember. And, this is a very fantastic view point that one can take into account. This view point also is actually nicely discussed in the book by Ogunnaike.

(Refer Slide Time: 22:02)



Now, we have derived the expression for the upper-tailed test case. In a similar way, you can also derive the expression, same expression that you get for the lower-tailed test. But, as I had remarked earlier, for the two-tailed test, one has to make an approximation. Why? Because, if you look at the expression that we have for beta for the two-tailed test, unfortunately, there are 2 terms; and these two, presence of 2 terms makes it difficult to arrive at a closed form expression for the sample size, delta and sigma, like we have done for the one-sided test. Therefore, under some approximations, that is, what kind of approximation? Assume that, this second term is much, much less than beta. In that case, you will have only a single term and therefore, you, you should be able to write an approximate term, where, an expression, where you have z alpha by 2 plus z beta is approximately equal to delta root n by sigma. In such situations, only this expression should be used. So, you should be able to really test that out, before you use the expression. Notice that, unlike in the single-tailed test, we have z alpha by 2. So, that is another change that you should watch out for.

(Refer Slide Time: 23:20)

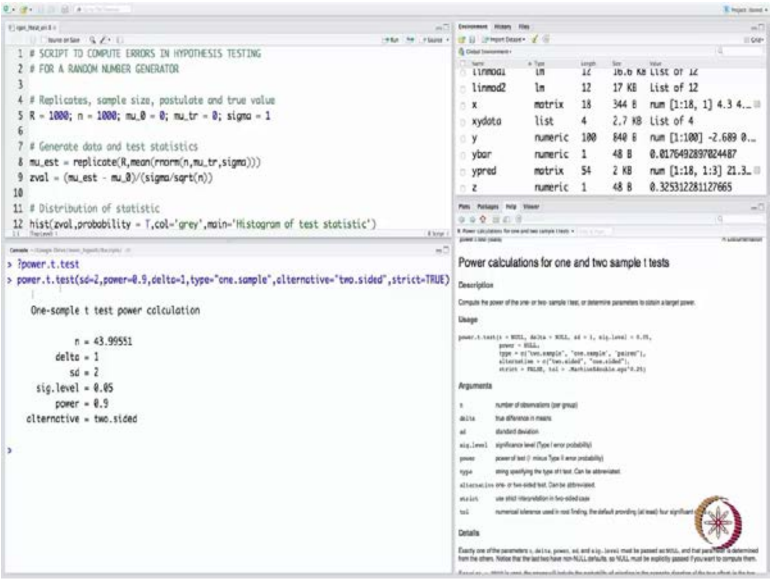


So, let us look at the example we have here. Suppose, again, the same propellant burning rate example. Now, I want to determine for the two-tailed test, 2 things; one, the magnitude of deviation that the test can detect for a given sample size, let us say 25 and the sample size that is required for detecting a deviation of 1 centimeters per second, with a probability of 0.9; that means, beta is 0.1. Earlier, we had the power of test; this probability of 0.9 is nothing but the power.

Earlier, we had calculated the power of a test, when the truth was 1 centimeters per second away from the postulated value, but there, we had a 0.7 power, or a 70 percent chance. I want to raise it to 90 percent. So, what do you think, the sample size would be lower, or higher than what we had used previously. Earlier, we had used 25. What do you think would be the answer? Now, as a simple homework, you can actually work out this problem by hand. What I will do is I will show you how to at least answer; we can answer both these questions using R. I will show you only the second one. Again, the first one, you can work out in R, by using this routine called power dot t dot test that is available in the stats package in R. By the way, there is also a p w r package exclusively for calculating power of hypothesis tests, for all kinds of hypothesis tests; mean, test variance, comparison of 2 population variances and so on.

Let me show you how this power dot t dot test works in R. Remember, now, we have been, we have derived the expressions for the known sigma case. So, ideally, I should be using the normal distribution; that means, what do we mean by the normal distribution? The z that we have been used should be replaced by a t, if I want the answers to correspond to what power dot t dot test gives me. In fact, the standard deviation is given, therefore; I should use the normal distribution. I am ok. But, if the standard deviation was not given, then, all you have to do is, go back to the expressions and given in the, and rework the expressions in terms of the t distributed statistic. More or less, the expressions would look the same, except that in place of z alpha you will have a t alpha with n minus 1 degree freedom and so on. So, the power dot t dot test assumes that the standard deviation has been estimated from data. Therefore, the answer that you get from power dot t dot test need not exactly correspond to what you get from your hand worked solution, for both these questions; because, the question here assumes standard deviation is known which is 2 centimeters per second, alright. Nevertheless, it will give you reasonably close answers.

(Refer Slide Time: 26:28)



Let us look at R here and show you how the power dot t dot test works. The first thing always, is to pull up the help on power dot t dot test and it will show you the arguments that you have to supply. What is, what does this routine do for you? Same thing that we have done, spoken until now, specify delta, specify sigma, alpha and beta, it would actually calculate the n for you; or, specify n, alpha, beta and sigma; it will calculate the delta for you.

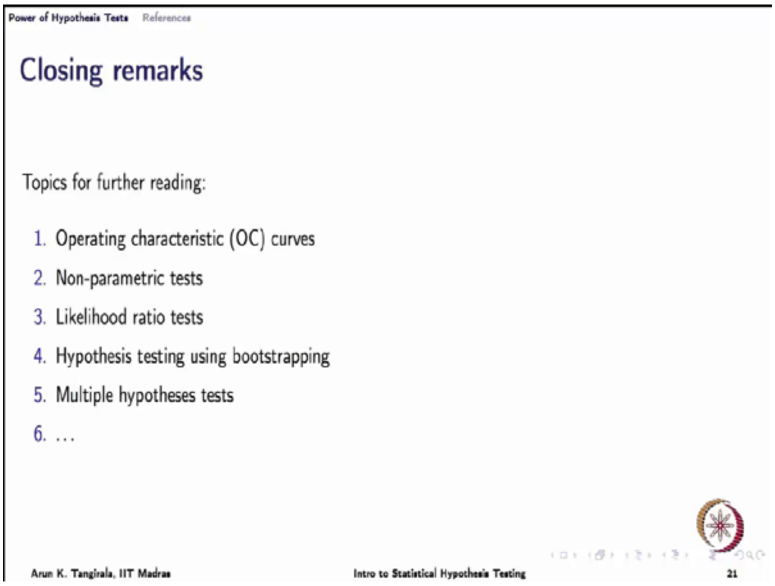
Let us answer the second question that we asked in the example. We said, I would like to detect the deviation of 1 centimeter per second from the postulated value, with a probability of 0.9, which means power is 0.9, and also in a, when we are conducting a two-tailed test, and that, given sigma is 2 centimeters per second; and of course, delta is known; alpha is 0.05. So, let us see how to do that. So, what you do is, here, you specify the standard deviation, power; these are all the arguments to the function; delta is 1, and remember, this is a 1 sample test. So, you should specify that, and the alternative is two-sided. This strict is for the calculations of the two-tailed approximations.

Remember that, we have worked out an approximate answer, approximate relation between z alpha, z beta and delta n by sigma. But, we do not have to do that when we are using a computer. We can actually calculate beta, without using that analytical expression, in a more accurate manner using the computer. But, if you are solving it by hand let us say, in an exam you can use the approximate expression that is given earlier in the slides. So, now, what are we asking for here is the sample size for a specified delta, power, alpha and sigma. So, the answer comes out to be 44.

Remember, we said, we will round it off to the upper integer. Now, as I just remarked earlier, this calculation is based on a t statistic, not a z statistic, that is not on a standard Gaussian statistic. Therefore, one should expect slightly higher answers for n; obviously, why, because, when do we use the t statistic, when sigma is unknown. So, that is another piece of information that I have to extract from data. Therefore, the power should come down, or for a given power, I will require a few more samples than what I would require for a Gaussian distribution. There are both ways of looking at it. In other words, the s d that I have passed on to power dot t dot test, it assumes to be an estimated value. If I want actually tell that it is a fixed, known value, unfortunately, there is no provision there. You will have to go back and write a code which uses the exact expression, but we will not do that. It is OK to actually get larger sample size, rather than getting a smaller sample size.

So, this is a more conservative estimate than what you would get compared to using a standard Gaussian distributed statistic, right. In fact, if you look up the book by Montgomery, the answer is 42, that is, if you had to actually work with standard Gaussian, the statistic that is, if you assume sigma to be known. It is ok. So, here, it says 44; you would get 42. So, that extra degree samples are required to match the degrees of freedom that you are losing when you are estimating the variance; that is all. And, you can play around with this, and I have also said that, you can use the p w r package to conduct hypothesis tests and, sorry, to compute powers and also determine the samples sizes and deltas for various hypothesis tests.

(Refer Slide Time: 30:44)



Alright, then, now this brings us to the close of this lecture and the close of this course. And, there are of course, there are many topics that we have not really discussed; but remember this is about 11 to 12 hour course max. So, more or less an introductory course on hypothesis testing obviously, it is not possible to talk about everything, but hopefully, this course has given all that is required to get you going and read on several topics.

For example, we have not talked about operating characteristic curves. What are this OC curves? They are essentially a plot of the power versus the delta, or the standardized delta for different sample sizes. How would you use those charts? Well, you would refer to an OC chart for particular type of test, and then for a specified delta, and for a specified power, you would just traverse and pick the sample size; or, for a given sample size and a given power, you can determine what is the deviation that you can detect; essentially, a charted way of all the possible values of n and delta for a given alpha, beta and sigma, right. Maybe, that is not required when you have a computer, and that is why I also have not discussed about this OC curves, but you can read quite a bit on them. And then, we have assumed, in most of the cases, for the variance test, proportion test, and so on, that, particularly variance test and the t test, that, data falls out of a Gaussian distribution; but that need not be the case, in which case, one may have to adopt tests, what are known as non-parametric test. What we mean by non-parametric here is, the distribution for the data, and hence for the statistic, has not been parameterized.

So, for arbitrary distributions, how do you conduct hypothesis tests? Then, you have one of the most reputed tests and powerful test known as the Likelihood Ratio Test which we have not discussed. The Likelihood Ratio Test rests on the concept of likelihood which requires some deliberations. Likelihood was largely popularized by Fisher. It was introduced by Fisher and then, has become very popular. The maximum likelihood estimation is a very powerful method for estimating parameters. So, the core idea in Likelihood Ratio Test is, to use an estimator that maximizes the likelihood. In other words, we use an m l e, and then, set up a test statistic that is a ratio of 2 likelihood functions; that is the basic idea. What are those 2 functions? One, likelihood evaluated under the postulated hypothesis conditions; that means, restricted conditions. The other likelihood being unrestricted, meaning I do not know any of the parameters, I have optimized them; I have found the parameters by maximizing the likelihood.

So, it is the ratio of those two likelihoods that forms the statistic, and one can show that, for the conditions that we have discussed in this course, the Likelihood Ratio essentially takes the same form as the ones that we have discussed under various scenarios; that means, whatever test statistics that we have considered for one sample test for mean, or with the variance unknown, or variance known, they all, they essentially have the same distribution as the distribution of the Likelihood Ratio. So, therefore, we have not missed much, as far as what we have discussed in this course is concerned, and Likelihood Ratio Test. But, for many other scenarios that we have not talked about in this course, Likelihood Ratio Tests are used widely, because they are very powerful; that means, the power of a Likelihood Ratio Test is quite high.

And the other one that we have not talked about, but I have briefly mentioned at sometime, is the method of hypothesis testing using Bootstrapping. We have assumed that all the statistics that we deal with in hypothesis tests, have known sampling distributions. But, that depends on the estimator that we are dealing with, the kind of estimation problem that we are dealing with. In many cases, it is very difficult to obtain the sampling distribution. Therefore, one uses the modern methods of Monte Carlo simulations or Bootstrapping methods. In fact, Bootstrapping allows you to generate artificial realizations from the given data, and so, that you can actually now compute your statistic over all this artificially generated data, from a single data, also, sometimes called a surrogate data, and empirically compute the distribution of the statistic, and then, come back to your hypothesis testing. You can read a lot on Bootstrapping in the literature. It is now becoming increasingly popular in every field. So, there are nice books on Introduction to Bootstrapping. If you are interested, you should go and read them.

And, the other thing that we have not talked about, are multiple hypothesis tests. There may be many parameters that we want to simultaneously test, and in which case, we will have to consider the alpha, Type I error for each of those parameters, or we may consider a collective Type I error. So, there is something called a Bonferroni test, and several other concepts associated with the tests of multiple hypotheses. Again, that is outside the scope of this course and then, there is, of course, many other things that we have not talked about.

All in all, I sincerely hope along with the TAs that you have learnt a lot and you have had a really very productive and fruitful time listening to these lectures, working out the assignments. And, I would also like to take this opportunity to thank the NPTEL staff, both the technical and the administrative staff, for their wonderful cooperation and all the work behind the scenes. So, if you have any questions, maybe feedback, please do feel free to write to us, or post them on the Forum and as usual, we will be happy to answer them. Good luck with the exam, and hopefully, see you on the live hangout that we plan to have after this lecture.

Thank you. Bye.