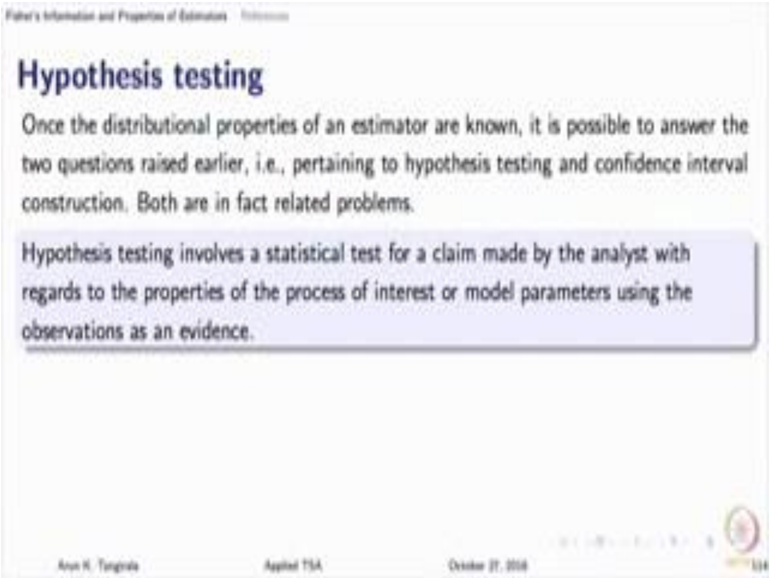
Applied Time - Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 93
Lecture 40B - Goodness of Estimators 2 -8

Let us now talk about hypothesis testing.

(Refer Slide Time: 00:14)



Which is concerned and this is one of the most critical and core tasks of any data analysis, it does not matter whether you are doing time series modeling or machine learning or a pattern recognition whatever may be the data analysis exercise that you are looking at. At some point in your data analysis you would be testing some kind of a hypothesis, because the objective of any data analysis exercise although in the beginning is exploratory later on it becomes confirmatory. You must have heard of these two terms in data analysis one is called exploratory data analysis other is confirmatory data analysis.

Exploratory is I do not know anything I am just exploring the data; I want to define a problem, I want to do something with this data. Confirmatory is you begin with the postulate and then there is a certain objective and the direction to your data analysis. So, at some point you have to ascertain whether the postulate that you have made is confirm.

And again there because of the uncertainty you can never answer this question with 100 percent certain.

So, what you are doing in hypothesis testing is you are using data as an evidence, you first begin with the postulate and you search for evidence in the data; what kind of evidence are you searching for if a whether this evidence is refuting the postulate that you are making. So, it is a different kind of search, you are not searching for evidence that supports your claim. Yes, in that process you may think you are doing that, but in hypothesis testing typically you make a postulate which is called the null hypothesis which is a default one which will hold when even when the data is not given; that is you make the claim for example that the mean of some process is something.

It is either up to you or someone else to refute that to show evidence that this claim does not hold. So, what does one do collects data and then like a detective searchers through the data searches for evidence to see if the null hypothesis can be refuted. If there is insufficient evidence the benefit of doubt goes back to null hypothesis. That is exactly like to do in a coat of law where the default hypothesis is that accused is innocent. Does not matter what has happened, to begin with the accused is always innocent. And the reason is as we say in a court of law at least in most judicial systems across the globe the premise and the principle is that the innocent should not be punished even as the culprit may goes scot free.

So, we want to make sure that the benefit of doubt is always given to the accused. Then it is of course the lawyer who is actually trying to prove that no the accuse is not innocent it is a burden of that lawyer to prove that to provide sufficient evidence that will pull the judge away from that null hypothesis. So, it takes quite a bit of power and that is why the power of hypothesis test is actually rest on your ability to provide evidence and the method of testing, everything put together to pull the coat or any data analysis exercise from the default null hypothesis, because the default is null hypothesis. So, it is always a tendency to go back on it and it is a fall back option.

So, that is of course when the truth is at the null hypothesis is false. So, your ability to reject the null hypothesis when it is false determines the power of the hypothesis test. What I am going to show you very quickly is a procedure for hypothesis testing. And you have to ask yourself why on earth we are talking about this in time series modeling.

Can you give me two examples why I should be interested in hypothesis testing in time series analysis? Whatever I teach in this course has a barring and it is an integral part of time series analysis, but you should be convinced more than me. Where do you think these kinds of problems arise in time series analysis?

Student: Testing in.

Sorry.

Student: Think which model is the better one.

But what would be your hypothesis there?

Student: (Refer Time: 04:59) same and like the.

Can you translate that to a statement on parameters? Let me give you a hint; see this is where I require you to think learning concepts is one thing applying them is another thing I am you would have realize that in quizzes also. I would like you to think where hypothesis test occurs and I will give you a hint what is a first step in time series analysis in modeling. What is that? What is the technical statement I mean? I have been gone out of a time series analysis course you should use the technical terms correctly.

Student: stationarity

Ok, one is stationarity very good. So, how would you check for stationarity may be one of the non stationarity you would check is for integrating effects and we say that you know series as integrating effects if the pole is on the unit circle. So, I would fit an AR 1 model and ascertain the null hypothesis is that let us say it is the pole is at 1 and then the alternative hypothesis no it is not at 1. The other standard test is if you are given it is stationary what is this other standard step that you would take?

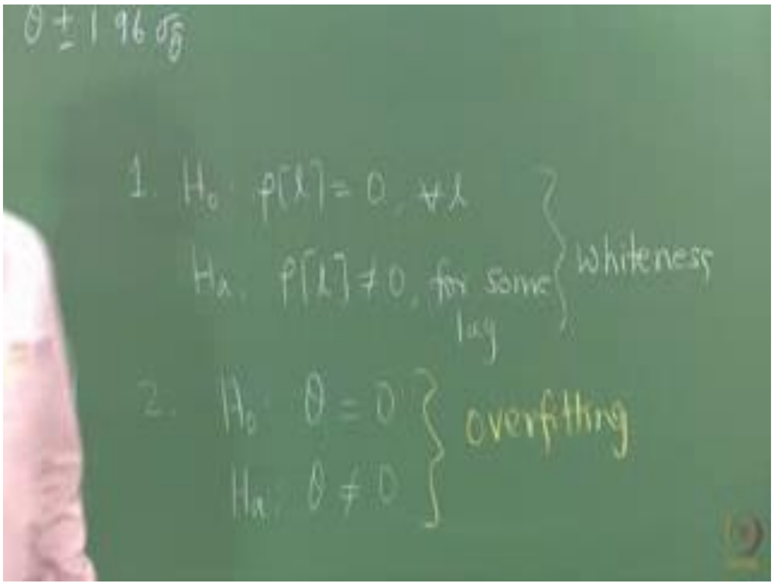
Student: (Refer Time: 06:23).

Yes, so as like many of you been saying I want you to use a correct term we perform what is known as a whiteness stress. What does whiteness stress translate to? This is a statement verbal statement, but what is a statistical statement?

Student: (Refer Time: 06:41).

So, the null hypothesis is that the ACF is 0 at all lags.

(Refer Slide Time: 06:46)



So, this one common hypothesis that we run into is that the ACF or you can say sigma rho is 0 for all lags and the alternate is it does not for some lag, of course it is a tough one to test but will do it. The other kind of hypothesis test that you would run into is concerning modeling, you would this whiteness stress not only appears at the beginning of the time series modeling but also post modeling, where you are testing the whiteness of residuals.

The second hypothesis test that is commonly encountered in time series analysis is that the parameter that I have estimated in the model is 0, the true parameter. Where is that coming from? What is the verbal statement corresponding to that? If I am testing a null hypothesis of this type what is the verbal statement? Now I am giving you the statement it could be that either one parameter or a bunch of parameters in the model are 0, so it is a single parameter. And the alternative is it is not, where does this come from?

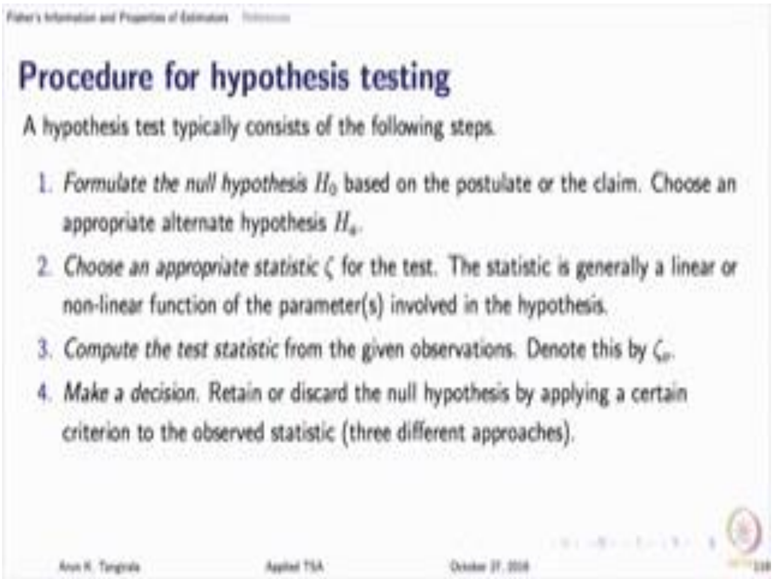
Student: (Refer Time: 08:10).

Ok not bad, but is there a better way of putting it? The verbal statement corresponding to this is that I have over parameterized the model; I have included a parameter that was not necessary. So, these kinds of hypothesis statements come about in when I am testing for over fitting. What does a over fitting mean that I have unnecessarily complicated the model more than what it should have been. For example, the underlying process may be AR 2 I must have fit and AR 3. This could also be applied to PSE, for example PSE of beyond certain lag is 0.

So, there are many different hypothesis test that you would encounter, but predominately these are the once that you would run into apart from the standard hypothesis test on mean being 0 or the variance being of some value and so on. But, by and large you would encounter hypothesis test of this type. So, this is for whiteness. So, we appropriate I use a white chalk for this and this is for over fitting or over parameterization. Elsewhere in spectral domain you can have another hypothesis test which is that, that is the equivalent of whiteness test in the spectral domain would be that this spectral density is equal at all frequencies; that is another hypothesis test.

So, there are number of hypothesis test and you should go back and think as to what are the different kinds of hypothesis, just go through that exercise you do not have to do it manually that is you do not have to simulate something into it, but think go through that thought process and believe me you learn a lot more then you would just by listening to the lectures. So, the first step in hypothesis testing is to formulate the null hypothesis, second is to choose an appropriate statistics as I said.

(Refer Slide Time: 10:15)



Like a detective you go and investigate find out if there is evidence and the statistics is like your lens. And then the first step is very very important you have to formulate the right null hypothesis and the alternative one. And the third step is to compute the test statistic from the given observation. And forth one is to make a decision, whether what you have observed conforms to the null hypothesis or against it; that is what we mean by make a decision.

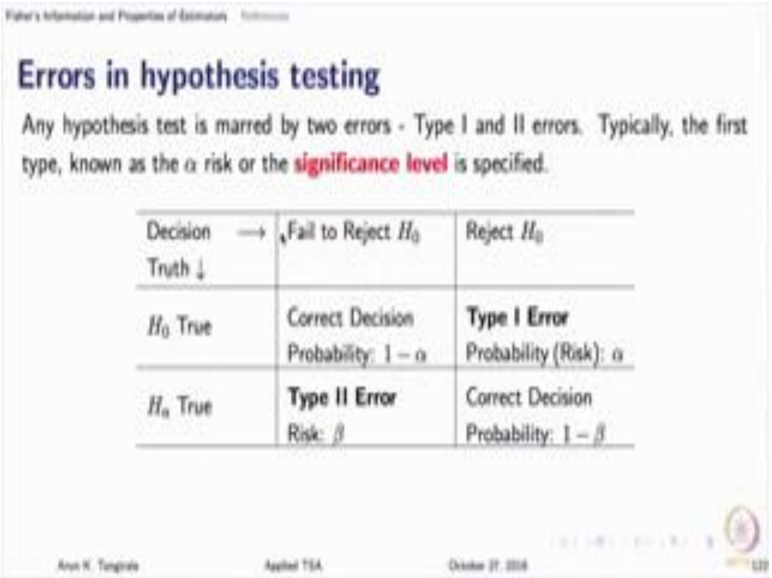
And at stage in fact when you are computing the test statistic there is a very important point their which is that you hold on to the null hypothesis, until you make a decision you assume that the null hypothesis is true then you compute a statistic. For example here, you would compute the statistic, so here it would be sigma of the process. Suppose I am constructing I am conducting a hypothesis test on average which will go through very quickly. There would be a postulate for theta naught and I would compute the statistic assuming that postulate to be true. So, you will always conduct the hypothesis testing given that h naught is true, it is a conditional testing that you doing. You are not doing it unconditional testing, always the hypothesis testing is conditional. You first fix the null hypothesis, this is like in the court of law assume that accused is innocent would he or she would have done any of this, they say no that is not likely.

Then you would come to the conclusion that whatever as happened whether this person as sum weapon at home or carrying some weapon in the hand is only just by chance. If the defendant lawyer is able to prove that it was only by chance that the accused was caught up in all of this otherwise no no no no he was very innocent [FL]. Then the judge would have to go by that, but if the prosecutor is able to prove that no all of this is cannot by chance even as accused is innocent. There was a systematic plan there, and therefore the person deserves to be called as a criminal or culprit and so on.

So, that is exactly what is happening in hypothesis testing you hold on to the null hypothesis compute the statistic and then you ask how likely is the observed value given that the null hypothesis is true. As simple examples suppose r norm claims at the mean is 0 you generate 100 observations you compute the sample mean. Suppose you get minus 2.1 as a sample mean, and let us say then you compute the standardized statistic and it turns out to be fairly high. You say that; yes, I agree that when this true mean is 0 it is possible that I can get a large value for this statistic, but that probability is very low. Therefore, I would actually refused the null hypothesis rather than actually blaming the realization saying that it is a low probability realization; that is the basic idea and hypothesis testing.

So, remember that no hypothesis test can result in a perfect decision. And therefore, there are two types of errors in hypothesis testing depending on the combination of the truth and the decision. And this table essentially summarizes that you must have seen this table in the videos as well.

(Refer Slide Time: 13:53)

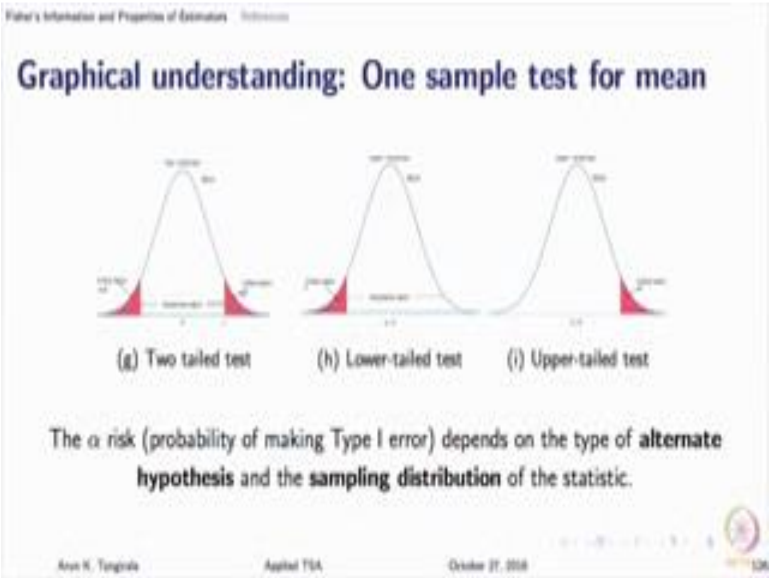


Basically, if the truth is that h naught is true and you failed to reject h naught then you have made a correct decision cool. But, if you have rejected h naught when it is true then that is an error. And there is a chance that you will do it that is what I meant earlier. No, hypothesis testing is fool proof. This probability of making this kind of an error is denoted by alpha is called type one error, h naught is true and you have rejected. On the other hand h naught could be false, the mean may not be 0 or may not be what you are postulated, maybe theta is not 0 and you have failed to reject h naught; that is also an error you should have rejected h naught. That is another error in the probability of making that type two error is denoted by beta.

Generally, this alphas and betas are fighting with each other, but do not think that alpha plus beta is 1; that is wrong. And the of course, if h naught is false and you have rejected h naught that it is a correct decision and the probability of that is naturally 1 minus beta. So, these are the two types of errors that you will encounter in any testing like any kind of decision making process. And normally in hypothesis testing what one needs to specify is one of these two errors, you have to say that conduct a hypothesis test with this level of tolerance I would like that kind of qualifies a goodness of hypothesis test.

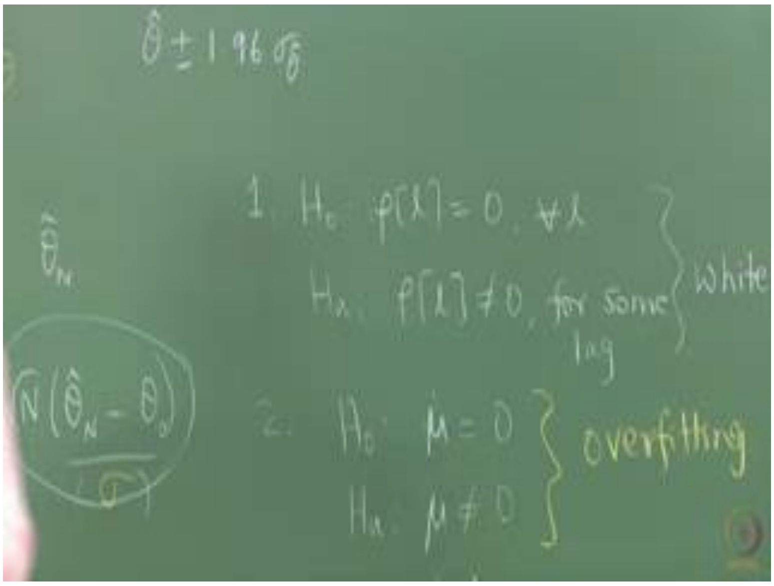
So, you have to specify what goodness you want from the test. And typically one specifies alpha, the type one error because I will not go into detail you should see through the lectures calculating type two error is a lot more complicated for the user and it depends on what the actual truth is. Beta which is type two error has to be calculated by stating the true truth not a postulated truth and it can change with that. So, that is a painful thing. It is lot easier to specify alpha because you are assuming h naught to be true, therefore it is fixed and you are only going to calculate what is a probability of rejecting h naught it is a typical values of 0.05 or 0.01, and this is same alpha that appeared in confidence region.

(Refer Slide Time: 16:27)



So, just to illustrate quickly if you are looking at hypothesis testing of a mean there are three possibilities for the alternative hypothesis: one possibility is like this write in place of theta you may have mu.

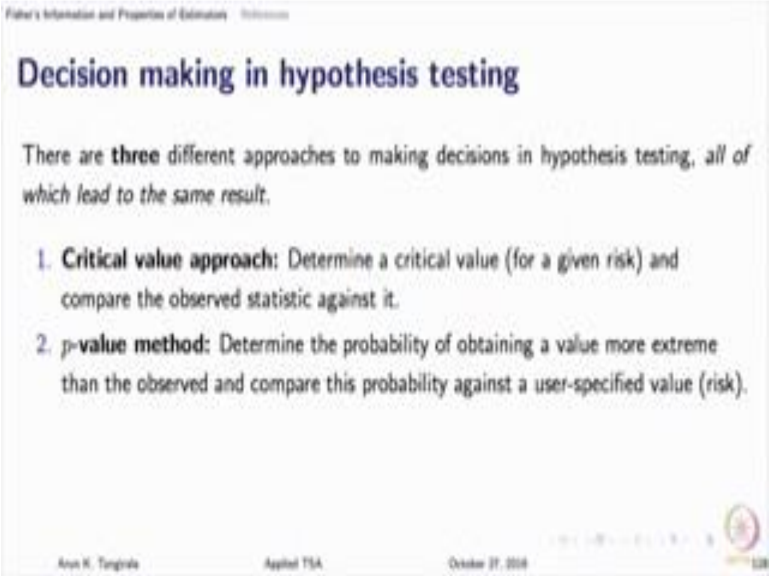
(Refer Slide Time: 16:35)



The other alternate is that mu is greater than 0 and the third alternate is it is less than 0. So, depending on the alternate that is why I said you have to be careful in farming the null hypothesis. The region of so called rejection you can change, so called critical regions can change. So, you have accordingly a two tail test, lower tail test, and an upper tail test. Of course, I have written it the other way round here, but anyway. So, if you have an alternate hypothesis of type mu naught equal to 0 you run into a two tailed test.

Typically, in out modeling and the null hypothesis the whiteness stress we deal with two tail test only, but you should be also be comfortable with the lower tail and upper tail test. And the alpha depends on the nature of alternate hypothesis, those red areas that I have indicated those are probabilities. What you are seeing as curve is the probability distribution of the statistic the sampling distribution. This total area in the red region for the two tail test is alpha, and if the distribution is Gaussian then the areas are distributed symmetrically.

(Refer Slide Time: 17:53)

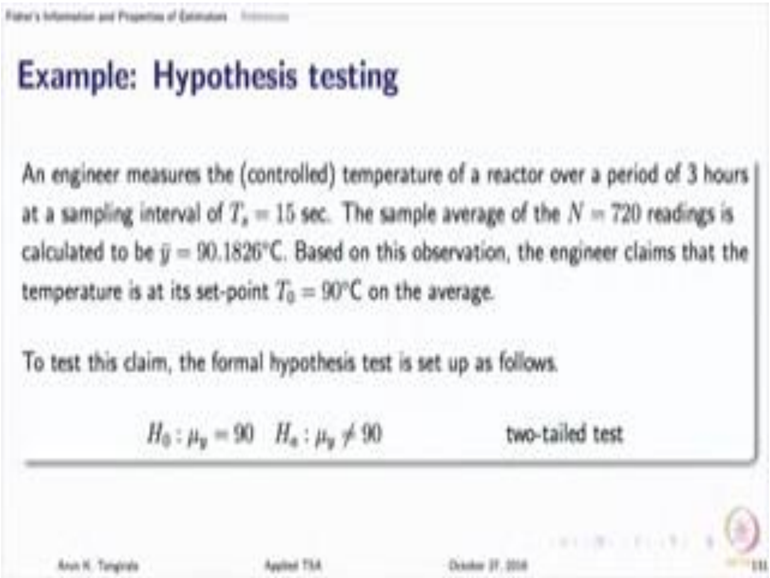


So, the idea in hypothesis test there are three different approaches in hypothesis testing: one is called a critical value approach where you compare the observed statistic with a critical value. That means, what you think is tolerable, if beyond that you say no no no its not possible that h naught is true and I would have obtain this statistic value. And critical value is conducted is, sorry is calculated or determined by first specifying alpha. For example, in a two tail test here the total area in the red is alpha and you would read of from the either from the chart or using r you would figure out this critical value. So, the white region here you can call it as acceptable region; acceptable means if the statistic falls in that region you do not reject the null hypothesis. That is a critical value approach.

The other is a p value approach; you look at the probability of obtaining a value more extreme than what you have observed. Suppose, I go through an example and to understand, but the basic ideas is if the probability of obtaining a value higher when more extreme let me put say higher would been always on the right hand side now more extreme then what you have observed is very low than already you are in extreme region. If the probability of obtaining an extreme value more extreme then what you have observed is high then that means you are in a safe region. So, these p values in general if it is very low, what is very low? In fact, when p value is less than alpha then the null hypothesis is rejected, because alpha is what you are willing to tolerate of obtaining an extreme value more extreme then what you have observed and the p value is the observed probability, observed significance level if that is low as I said in (Refer Time: 19:43) books; if the p value is low the null hypothesis must go that simple rhyming statements.

And third approach is a confidence interval approach. In this, in fact this is a simplest in my opinion construct a confidence region check if the postulated value is in the confidence interval if it is in then you reject the null hypothesis; it is a simplest. You may ask why you have said earlier that the confidence region can miss out on the truth; yes, that is the alpha risk. Yes, it could have miss out the truth in which case the null hypothesis gets rejected even though it is true, that is exactly the alpha risk. So, let me just quickly go through an example.

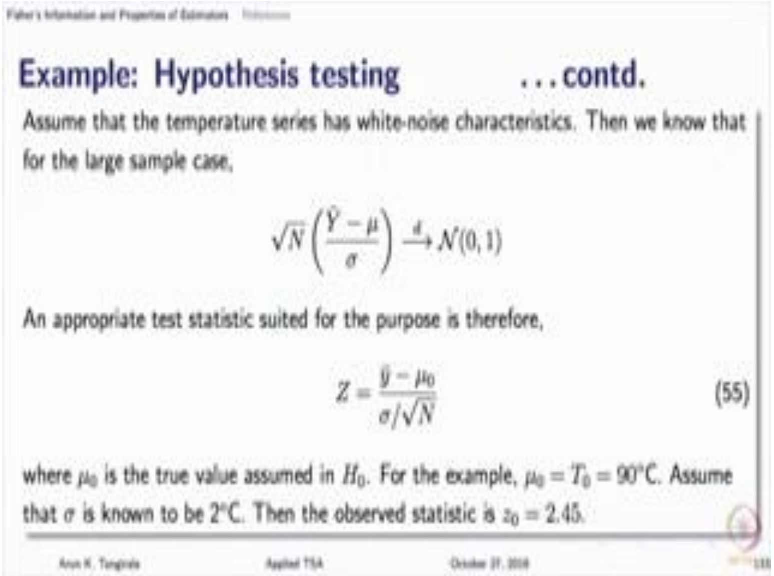
(Refer Slide Time: 20:26)



Here is a temperature we are looking at a reactor and there is an engineer controlling reactor temperature is design a controller and the engineer claims that on the average the temperature is at the set point which is 90 degree Celsius, let us say on the average because at each point it is not possible to maintain.

Now obviously, the boss wants this to be tested. So, seven twenties observation readings are obtained at a sampling interval of 15 seconds that does not matter here sampling interval. And now you want to conduct a hypothesis test of this form, it is a two tail test.

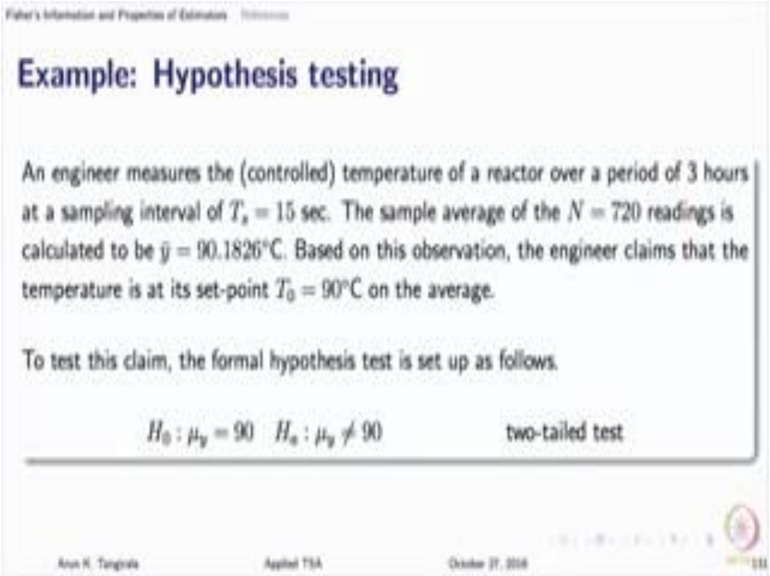
(Refer Slide Time: 21:04)



So, the first step is to compute the statistic to use the lens and you use the sample mean. Please remember that if I change the statistic to sample median then everything changes, the procedure remains the same, but the outcome can change.

So, we know that the sample mean follows a Gaussian distribution; in fact standardized one follows the standard Gaussian. And therefore, we choose to work with this standardized statistic y bar minus mu naught by sigma by root n. We assume that sigma is given as 2. What is n here in this example? 720, and sigma is let us say given to be 2 it is known from some historical data. What is a postulated value? 90.

(Refer Slide Time: 21:47)

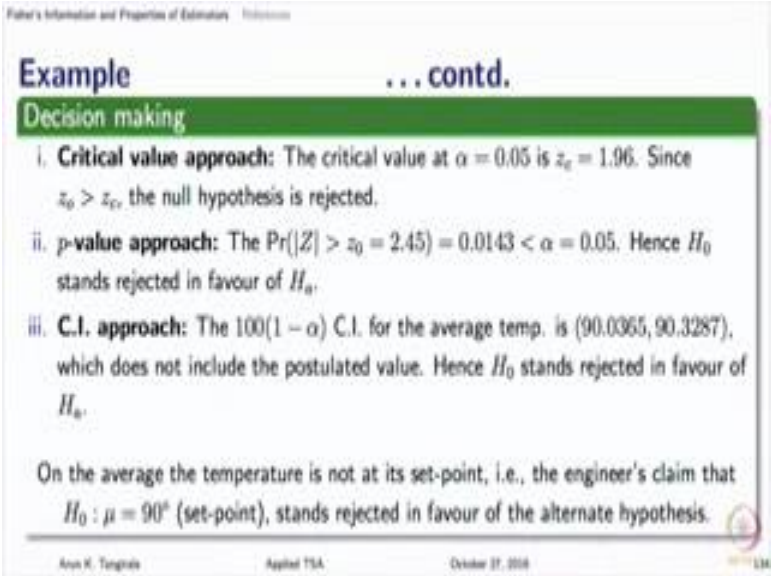


Observed y bar is also given which is 90.1826.

Now, when you look at it as a layman 90.1826 seems to be its very close to 90 [FL] I mean what is a big difference only 0.1826 cannot you ignore it is a Friday why are you troubling me or a Thursday Friday is anyway already include in the week end. So, you say 0.1826 (Refer Time: 22:09) that does not matter, but that is not how statistics looks at, it looks at this 90.1826 as been how many observation have gone into that value, the more the observations then the more you are suspicious that the claim is not correct, because few are observations can produce more deviations you have seen that. And also it depends on the variability of the process if the process is not varying too much you should not except too much of a deviation.

So, statistics does a very fantastic job of examining the distance between the observed value and the postulated value and ask if this distance is significant, but not on subjective bases, on statistical bases; that is the basic another prospective of hypothesis testing. So, all you are now doing is calculating this observed statistic and all the three approaches give you the same decision which is that the null hypothesis should be rejected. Why?

(Refer Slide Time: 23:10)



The critical value at alpha equals 0.05 it is a two tail test is 1.96, we know that the 95 percent probability interval is 1.96. And the observed statistic is more than that, which means your tolerance itself is 1.96 and the observed statistic even more, all of this conditioned on the null hypothesis being true. Therefore you say no no no its not possible, if the null hypothesis where to be true at least I should have obtained the observed statistic between plus or minus 1.96 that has not happened, therefore I reject the null hypothesis.

Then the other one is the probability, that is which is the p value approach now. The p value approach looks at the probability of obtaining a value more extreme than that I have observed and it turns out to be 0.015 in this a case which is less than alpha, and therefore we reject the null hypothesis. So, we say that already the observed value is extreme enough, and therefore I will reject it. Up to what p values will I tolerate 0.05 that is what it say- If the probability of obtaining a value more extreme then what I have obtain is greater than 0.05 then I would not have rejected the null hypothesis.

And the third approach is a confidence interval which is I field a new test. You simple construct the 100 times 1 minus alpha this 95 percent that we talked about now we are able to reinterpret that in the context of hypothesis testing. So, this 95 percent confidence interval is now given. And the quick search is for the truth and you clearly see the interval does not contain the truth; therefore the null hypothesis is rejected. Initially, for beginners in hypothesis testing it is a good idea to go through all this three approaches until you pick your favorite approach.

And although each approach gives you the same decision they have different flavors to that. So, the hypothesis testing for you and this is you would conduct hypothesis testing for any other parameter, the procedure is the same. You try and work out a couple of example somewhere from the net or from the video lectures and you would be more confident.

I just conclude the; I am just going to skip the confidence, I will just conclude the part on hypothesis testing by saying here we are able to obtain the sampling distributions and therefore we are able to compute confidence regions or p value or critical value and so on. In practice obtaining f of theta hat may be very difficult as I already had been saying depending on whether hat is coming out of a linear or a non-linear estimator. If you have a complicated estimation problem and you have to conduct hypothesis testing then you have to use a modern approach of goods rapping or surrogate data analysis and so on, where you generate artificial realizations and then you compute the either t he distribution or the critical value or the confidence region or the p value one how those three from your artificial realizations empirically and then you conduct your hypothesis test.

That is what you will see in most of the papers that come out today, because today the computational power is increasing we can effort to work with non-linear estimators. Earlier we said non-linear estimator will keep them aside because it is so difficult to compute those, but today you know we can compute very complicated stuff. So, that is why people have now turned to goods rapping method.

So, with this we come to a close on the properties of the estimators. Next week we will start off with estimators, methods of moments least square, we have already discussed MLE and then we will talk about Bayesian. That will conclude the estimation methods, we will take those estimation methods and apply it to modeling; estimating model parameters, estimating auto covariance’s power spectral densities and so on. And that will kind of conclude the course.