Mine Automation and Data Analytics

Prof. Radhakanta Koner

Department of Mining Engineering

IIT (ISM) Dhanbad

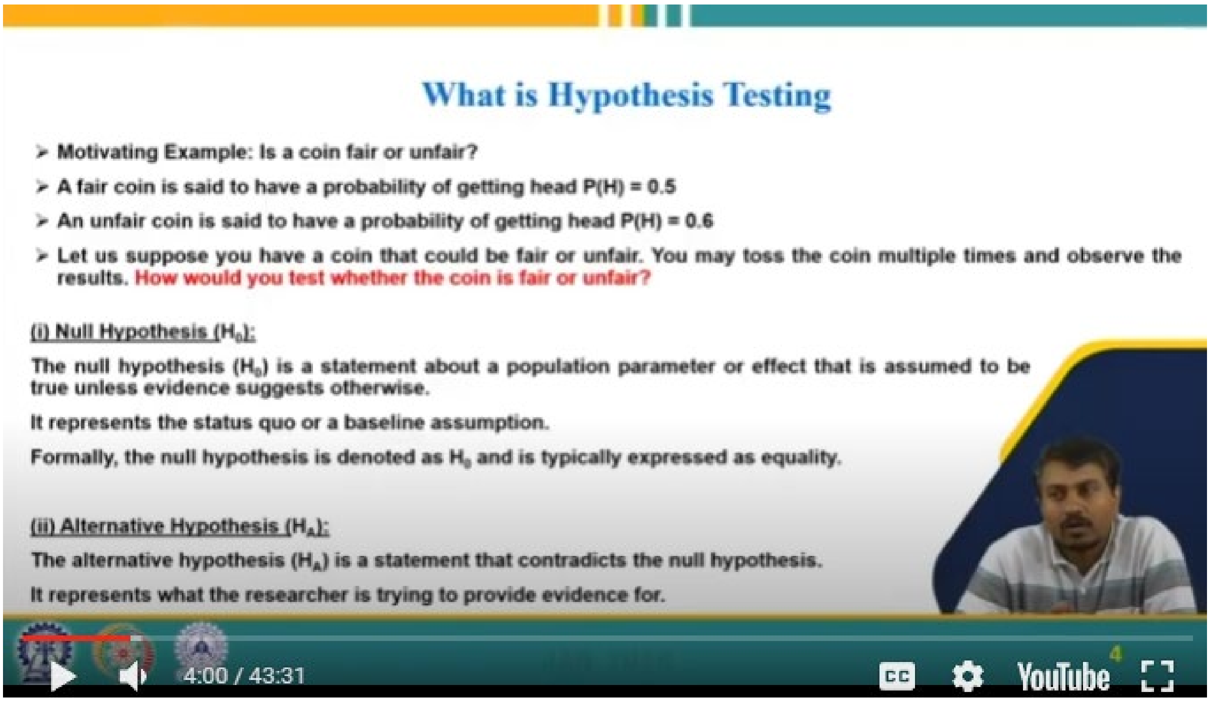
Week - 9

Lecture 41: Hypothesis Testing - I

Music Welcome to my course, Mine Automation and Data Analytics. Today, we will discuss hypothesis testing. For the last few lessons, you have seen that we debated the statistical model and probability. We are all trying to analyze the data and reach a fruitful conclusion. So, in Mine Automation, we expect a large amount of data about the different kinds of operations, other products, sensors, data, and so on. So here is the data we are getting, the expected data, some product specifications, or the machine when we are procuring. The company claims that this machine will serve you in this way.

When you are testing the efficacy of this machine, you are also measuring its performance. So, based on this measurement and the company's claim, does it fit with the expected prediction of the company that this machine will perform this way or a particular supply when you are ordering? The company is telling this that this small amount of error might be there, and you are checking and have observed whether the error is within that range. So, hypothesis testing is being used for this particular exercise. So today, in this lesson, we will discuss the basic framework of hypothesis testing, part 1.

So, in this lesson, we will discuss the concept of hypothesis testing, what hypothesis testing is, and the idea of size and the power of a test. Then, we will discuss the Neyman Pearson hypothesis testing paradigm, and then we will discuss different types of hypothesis testing composites. Lastly, we will complete this lecture with motivational examples for hypothesis testing. So hypothesis testing is all about testing the quality of some data and the product's quality, whether it is in line with the expectation, meeting the expectation, or deviating from it. So, to match with that, hypothesis testing is required. Here, we start with a straightforward example, the coin example that we have used several times, and most statistical books widely refer to this particular example.



Whether the coin we are tossing is head or tail, the expected outcome of this toss is a 50-50 chance. So, based on observation, does this 50-50 chance prevail or something else? So, to establish the fact that the coin is fair, the coin is fair, which means the probability of getting a head or the likelihood of getting a tail is 0.5, or exactly 0.5, half.

So when the coin is unfair, it is not that the probability of coming head will be 0.5; it will not be like that. For an unfair coin, the likelihood of coming head might be 0.3, 0.2, 0.4, maybe 0.6, maybe 0.7, like that. So, that is an unfair coin. The fair coin always has the chance that the probability of a head will come to precisely 0.5. So, to test this particular statement, is the coin we have fair or unfair? So, what do we have to do if this is my question, intention, and investigation? We have to toss the coin multiple times, and we have to note the result; we have to observe it and note the result. And then, based on the observed data, the observed outcome that we got after several times we toss the coin, we have stated already, we have to analyze. Now, how would we test whether the coin is fair or unfair? So, there are mathematical techniques that test hypotheses. So, in hypothesis testing, we come across two standard terms. One is the null hypothesis, and the other is the alternative hypothesis.

So, the null hypothesis H0 or H0 is a statement about the population parameter or the effect that is assumed to be valid unless evidence suggests otherwise. That is, it represents the status quo or a baseline assumption. So formally, the null hypothesis is denoted by H0 and is typically expressed as equality. For example, H0 is equal to means H0 is mu is equal to 0.5 like that. The alternative hypothesis comes in a particular situation when we want to reach some new conclusion or make some new decision based on the observed outcome. That observed outcome suggests something new, and that new contradicts the null hypothesis, the null hypothesis statement. So, it represents what the researcher is trying to provide evidence for. So, this null hypothesis is crucial for the scientific and engineering communities to reach new conclusions and discover new truths about the phenomena that were unknown to us. We have some preconceived or preliminary ideas, but the test and process suggest something else, which represents something new.

So, the alternative hypothesis always gives us some new insight into the process, the phenomena, and the object. So, let us represent the hypothesis testing using the simple data we are discussing. That is the outcome of the head and tail, and we are tossing the head coin. The number of times we have observed now we want to test whether the coin is fair or unfair. So, let us decide whether the coin is fair.

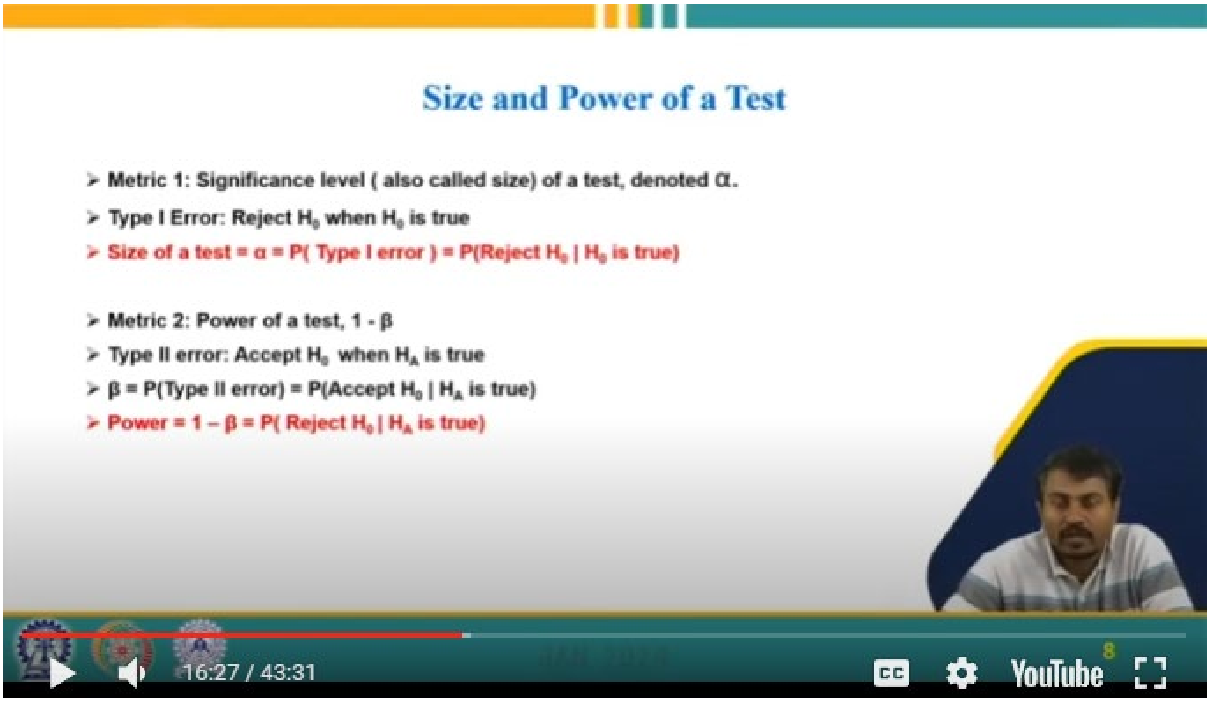
So the null hypothesis is that p h0 e head equals 0.5, and the probability of coming head is 0.5. For a fair coin example, the alternative hypothesis is that the likelihood of a head is equal to 0.6. This is one of the most important statistical analysis methods we are using, and it has many applications, a broad area of applications in engineering problem simulation. So, the null hypothesis is the testing we are conducting, representing the assumption to be tested. We assume that the coin is fair. The alternative hypothesis represents the researcher's claim or the possibility of an effect or difference, the new truth of this analysis. So, hypothesis testing aims to gather evidence from sample data to decide whether to reject the null hypothesis in favor of the alternative hypothesis.

So the thing is, if the data suggest strongly that it matches the alternative hypothesis, we then have to reject the null hypothesis, or else there might be a picture that we are not getting a substantial amount of data or observed data that we can convincingly reject the null hypothesis. Instead, we have to state and accept that condition k; under these circumstances, the null hypothesis is accepted and cannot be rejected like that. We are so accepting or rejecting the null hypothesis. So, for example, is a coin fair or unfair? So suppose we toss a coin three times; the possible outcome is head, then again head, then again hate possibly or head, then head, then tell, or so and so forth. So, there are eight possibilities of sequence.

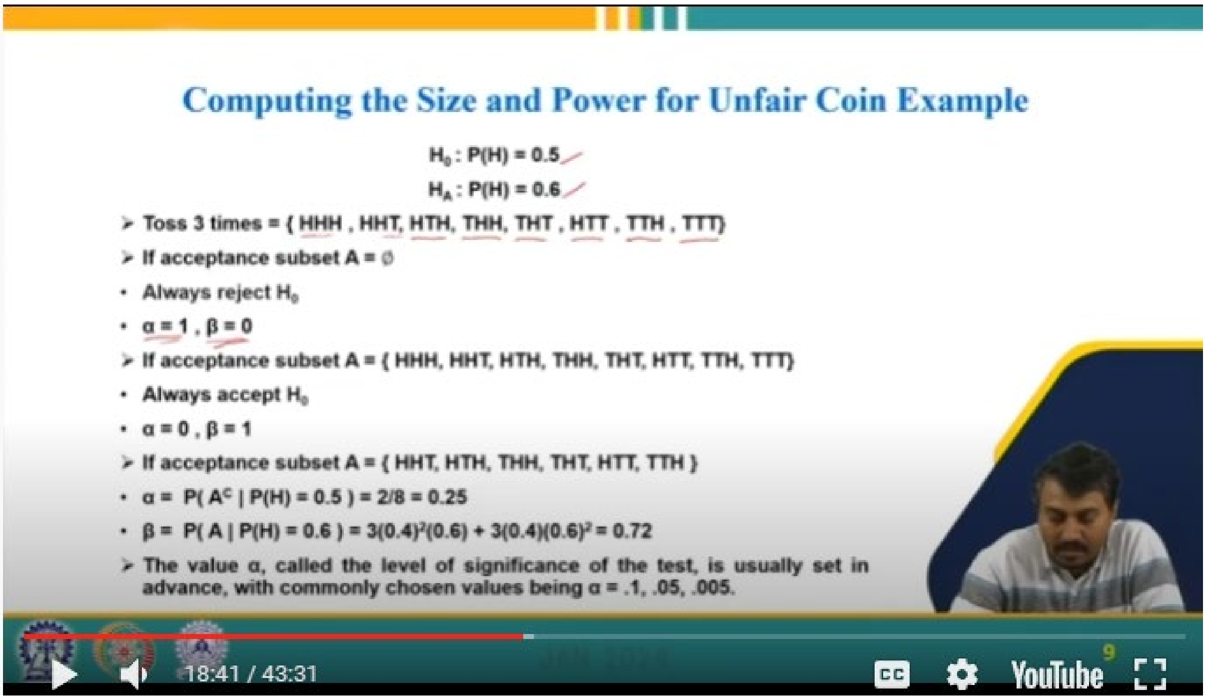
So, for some outcome now, let us state our problem. For some outcomes, we will accept the null hypothesis that is H0 or H0 or others; for others, we will reject the H0. So, let A be the set of all outcomes for which we accept the null hypothesis, which is H0. So, every acceptance of set A corresponds to a test; it must match the test result. So, another concept that we want to delete is a test's size and power.

This is a fundamental concept for hypothesis testing, and we must understand it in more detail. So metric 1 is the significance level, also called the test size, and the alpha denotes it. Here is the type 1 error that we are subdividing into two parts: the type 1 error and the type 2 error. So, the type 1 error is that we reject the null hypothesis when the null hypothesis is true. So the size of the test that is alpha is the probability of type 1 error, probability of type 1 error; this kind of error that the null hypothesis is true, but we are rejecting it.

What is the probability? That is similar to the probability of rejecting a null hypothesis when the null hypothesis is true. Another is metric 2, which is the power of the test, which is one minus beta. So, the type 2 error is that we accept the null hypothesis when the alternative hypothesis is true. We accept the null hypothesis when the alternative hypothesis is true. So the beta is the probability of type 2 error is equal to the likelihood of accepting the H0 null hypothesis when the alternative hypothesis is true, or the power is one minus beta is equal to the probability of rejecting the null hypothesis when the alternative hypothesis is true.



So these are the two representations of the size and power of the hypothesis testing: alpha and one minus beta. So, let us compute the size and power of the unfair coin example. So here, the H0, the probability of a coming head, is 0.6. So, the likelihood of H0 is 0.5, and the probability of the alternative hypothesis, that is, the probability of coming head, is 0.6. So we toss the coin three times. So this is the total sample space head, head, head, head, head, head and tail, then head, tail, head, tail, head, head, tail, head, tail, head, tail, tail, tail, tail, tail, tail, tail, total eight number of possibilities in the sample space. So if acceptance of set A is null, the acceptance of set is null.



So, we always have to reject the null hypothesis. We have to because it will not match. So, in that case, alpha is equal to 1, and beta is equal to 0. The acceptance of set A is the same as this total outcome. Then, we always have to accept the null hypothesis.

So, in that condition, the alpha is 0, and the beta is 1. Now, another example of the same problem: if the acceptance of set A is out of this 8, we have taken only 6, out of this total 8, we have taken only 6 of this that is from this to this. So, we have removed the head, head, and tail on this left side. So if the acceptance of set is this A, then alpha is the probability of A complement provided the likelihood of a coming head is 0.5, is two by 8 or 0.25, and beta is the probability of A for the head chances of the future head is 0.6. So that is 0.72 in total. So, the value alpha is called the level of significance of the test, and usually, we set in advance that this level of relevance is your probability and your chance. We have commonly chosen this value alpha is equal to 0.1, 0.05, and 0.005. So, these are the most famous values we use to test the hypothesis. So, what is the Neumann-Pearson paradigm of hypothesis testing? So here, H naught is a null hypothesis on the distribution of x, and H a is the alternative hypothesis. Now, the test is defined by an acceptance set A. So if the sample falls in A, accept the null hypothesis; otherwise, reject the null hypothesis. So, the two errors are type 1 errors, rejecting the null hypothesis when the null hypothesis is true.

Type 2 error accepts the null hypothesis when the alternative hypothesis is true, and the accurate metric, which is the size of the test alpha, is equal to the probability of rejecting the null hypothesis when the null hypothesis is true. The power one minus beta equals the probability of rejecting the null hypothesis when the alternative hypothesis is true. Type of hypothesis testing. So, this is the simple hypothesis testing. We have the composite hypothesis testing. So, let us discuss this first with simple hypothesis testing.

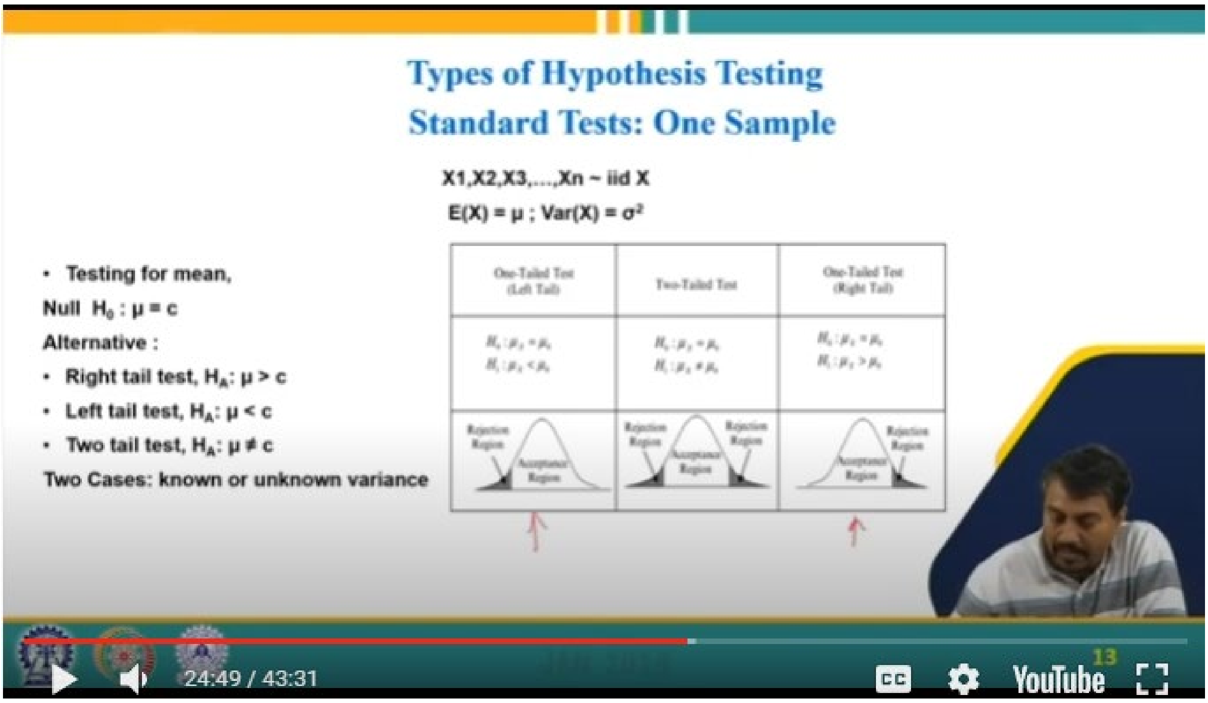
A hypothesis that completely specifies the distribution of the sample is called a simple hypothesis. For example, tossing the coin's probability of a coming head is 0.5, and the likelihood of a future head is 0.8 we defined. So, it precisely pinpoints the phenomena and their correctness.

For the normal distribution mu comma, in 3 samples, we are now telling mu is equal to 1, mu is equal to minus one, etcetera. So, this is a simple null versus a simple alternative. So, this is one example of simple hypothesis testing. Now, let us see composite hypothesis testing. A hypothesis that does not specify the distribution of the sample is called a composite hypothesis.

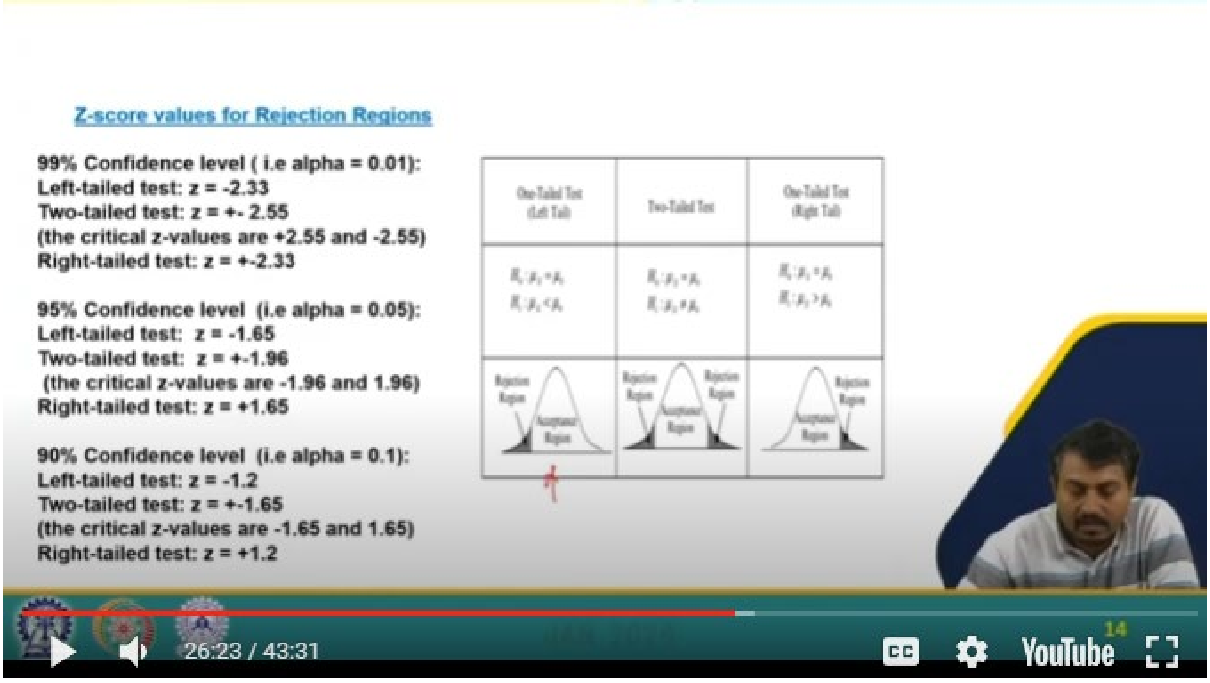
So, the example coin toss. So null is the probability of a coming head is 0.5, and the coin is fair. That is simple. Alternatively, the likelihood of coming head is not equal to 0.5; the coin is unfair; it is a composite.

The standard distribution mu comma three samples. So null mu is equal to 0, and some effect is not present. An alternative, mu, is greater than 1, and the effect is present, which is composite. So this is the picture versus null hypothesis versus the composite null versus the composite alternative. And also simple null versus simple alternative.

So, type of hypothesis testing. So here we have given you the figures for three kinds of tests. This is one tail test that is left tail. This is also a one-tail test on the right tail and a two-tail test on both sides. So, the shaded region is the rejection region. The shaded region is the rejection region.



So for the right tail, one tail test when the null hypothesis mu equals C. So, for the right tail, mu is more significant than C, and for the left, mu is less than C; for the two-tail test, mu is not equal to C. So, in these two cases, we have known or unknown variance. And here, for this example, the expectation of x for this variable of x, x1, x2, x3 is mu, and the variance is sigma square. So, these are the data from the table. This is basically for different levels of alpha for 0.01, 0.05, and 0.1. So, for a 99% confidence level alpha, that is 0.01. For the left tail test here, z is minus 2.33. Here, the z is plus 2.33, and for the two-tail test, it is plus minus 2.55 or 2.57 or, in a greater degree, approximately 2.55. For the 95% confidence, alpha is equal to 0.05.



Similarly, the z for the left tail test is minus 1.65 for the proper tail test plus 1.65; for the two tail test, the critical value is between minus 1.96 and 1.96. For the 90% confidence level that is 0.1, alpha is equal to 0.1, we have a left tail test minus 1.2, which is plus 1.2, and in between minus plus minus 1.65 is basically for the two tail test. So, when we tested one sample, we had a known variance. So, the sigma equals C, and the null hypothesis is H0. So the right tail test H alternative A, sigma is more significant than C. The left-tail test t is a sigma less than C; for the two-tail test, the alternative hypothesis is that the sigma is not equal to C. Let us compare with the mean for the two samples where the expectation for one is mu 1, the other expectation is mu two, and the variance for the first one is sigma 1 square and the second is sigma 2 square.

So, the testing to compare the means null hypothesis H0, mu 1 is equal to mu 2, alternative mu 1 is not equal to mu two. For the variance, sigma 1 equals sigma 2; the alternative hypothesis is that sigma 1 is not equal to sigma 2. The goodness of fit testing. So, this concept is essential for the model and how your model fits the data, and this is called the goodness fit of testing. So, does the sample follow a specific distribution? So, if the distribution is following, then it will fit into the model. So, for the integer sample space Xi of 0, 1, 2, is it the distribution Poisson? Is the distribution normal for the continuous sample Xi from minus infinity to infinity? Based on the data, its pattern, whether it fits with the model, represents the goodness of fit testing.

One of the most popular methods is the chi-square goodness fit of testing. Observation. So, in all examples, the question is reasonably posed in a statistical hypothesis testing framework. So, in most cases, the null hypothesis or the alternative hypothesis is composite, and in all cases, the confidence of the testing is very, very important. So, how do we quantify this confidence? So, with the help of an alpha value concept or a notion called a p-value, it quantifies confidence.

The p-value is the probability value, defined as the probability of getting a result that is either the same or more extreme than the actual observations. The p-value is the marginal significance level within the hypothesis testing that represents the probability of occurrence of the given event. So, the p-value is used as an alternative to the rejection point to provide the minor significance at which the null hypothesis would be rejected. And if the p-value is small, then there is more substantial evidence in favor of the alternative hypothesis.

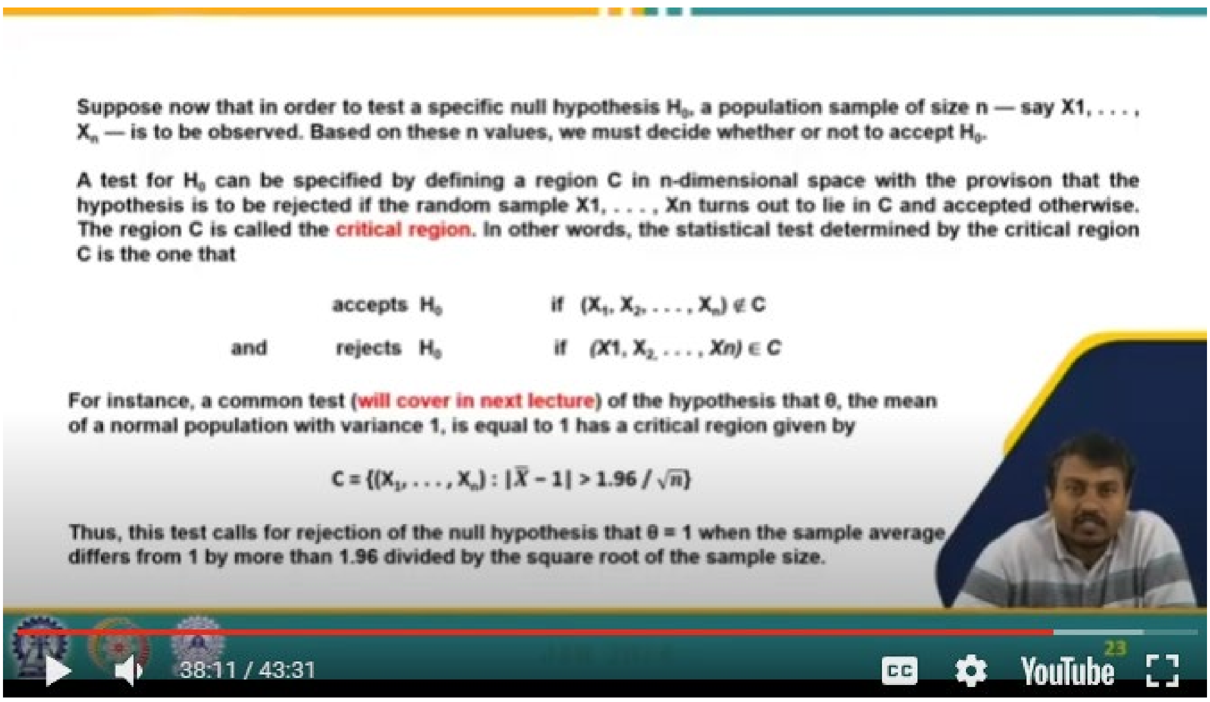
So, let us discuss this one example. This example is familiar to mining engineers and those working in the industry. Let's assume a construction firm will procure many cables that the company has guaranteed to have an average breaking strength of at least 7000 pounds per square inch PSI. So, to verify this claim, the company procuring this material has decided to take a random sample of 10 cables to determine their breaking strength. So, they will then use the result of this experiment to ascertain whether or not they accept the cable manufacturer's hypothesis that the population mean is at least 7000 pounds per square inch. So, a statistical hypothesis is usually a statement about a set of parameters of a population distribution, and it is called a hypothesis because it is yet to be known whether or not it is true.

We are trying to go closer to this conclusion but are still determining whether it is true or false. So, the primary problem is to develop a procedure for deciding whether or not the value of a random sample from this population is consistent with the hypothesis or with the manufacturer's claim. So, for instance, consider a normally distributed population with an unknown mean value theta and known variance 1. The statement that theta is less than 1 is a statistical hypothesis we could try to test by observing a random sample from this population. The hypothesis has been accepted if the random sample is deemed consistent with the hypothesis under this consideration. Otherwise, we say it has been rejected.

So here, it is essential to note that in accepting a given hypothesis, we are not claiming that it is true, but rather, we are saying that the resulting data appear to be consistent with it. So, for instance, in the case of average population theta and 1, if the resulting sample size 10 has an average value of 1.25, then although such a result cannot be regarded as being evidence in favor of the hypothesis theta less than 1, it is not inconsistent with this hypothesis which would thus be accepted. On the other hand, if the sample of size 10 has an average value of 3, even though a considerable sample value is possible when theta is less than 1, it is so unlikely that it seems inconsistent with this hypothesis, which would thus be rejected. Consider a population having a distribution of theta where theta is unknown, and suppose we want to test a specific theory about the theta.

We shall denote this hypothesis by H0 and call it a null hypothesis that H0 theta equals 1. So, if f theta is a normal distribution function with a mean of theta and variance equal to 1, there are two possible hypotheses. One, as I said, H0 is theta is equal to 1, or H0 is equal to theta less than equal to 1. So, the first of these hypotheses states that the population is average, with a mean of 1 and a variance of 1. Whereas the second state is H0, here theta is less than equal to 1; it is usually with variance one and means less than or equal to 1.

So, for case 1 a, when it is true, specify the population distribution. In the case of b, H0 theta less than equal to 1, the null hypothesis does not. So, a hypothesis that, when accurate, completely specifies the population distribution is called a simple hypothesis, and one that does not is called a composite hypothesis. So, the first is the example of the simple hypothesis that H0 theta is equal to 1, and the second one, H0 theta less than equal to 1, is the composite hypothesis. So, suppose now that to test a specific null hypothesis H0, a population sample size of n, say x1 up to xn, is to be observed, and based on these n values, we must decide whether or not to accept the null hypothesis H0.



So, a test for H0 can be specified by defining a region C. This C is in n-dimensional space with a provision that the hypothesis will be rejected if the random sample x1 up to xn turns out to lie in the C region and accepts otherwise. So, the region C is called the critical region. In other words, the statistical test determined by the critical region C agrees with the null hypothesis if x1, x2, xn does not lie in the C region and rejects H0 if x1, x2, xn lies in the C region. So, for instance, a standard test of the hypothesis that theta the mean of the average population with variance 1 is equal to 1 as the critical region given by C is x1 up to xn for x bar minus 1 of the mod is more excellent than equal greater than 1.96 divided by root 2 bar n, n is the number of sample that is sample size.

So, this test calls for rejection of the null hypothesis that theta is equal to 1 when the sample average differs from 1 by more than 1.96 divided by the square root of the sample size. So, it is essential to note that when we develop a procedure for testing a given null hypothesis H0 in any test, two types of error can result. The first type 1 error is said to result if the test incorrectly calls for rejecting H0 when it is indeed correct. The second is called type 2 error result if the test calls for accepting a null hypothesis when it is false.

So, the objective of a statistical test of H0 is not to explicitly determine whether or not H0 is true but to determine if its validity is consistent with the resultant data. Hence, with this objective, H0 should only be rejected if the resultant data is doubtful when the H0 is true. So, the classical way of accomplishing this is to specify the alpha value and then require the test to have the property that whenever H0 is true, its rejection probability is never greater than alpha. So, the value alpha, the test's level of significance, is usually set in advance with commonly chosen values that are 0.1, 0.05, and 0.005.

So, in other words, the classical approach of testing H0 is to fix a significance level of alpha and then require that the test have the property that the probability of type 1 error occurring can never be more significant than alpha. So, these are the references. We will cover the next remaining concept in the next lesson. So, these are the things we have covered in this slide.

So, let me conclude in a few sentences. We first defined the null hypothesis test with the hypothesis testing and example. The size and power of the test, then discuss the Neyman Pearson hypothesis testing paradigm, the type of hypothesis testing, standard test 1 sample, standard test 2 sample, and the goodness of fit testing. We have given examples with the significance level alpha with some examples. Thank you.