Essentials of Data Science with R Software-1

Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Lecture 66

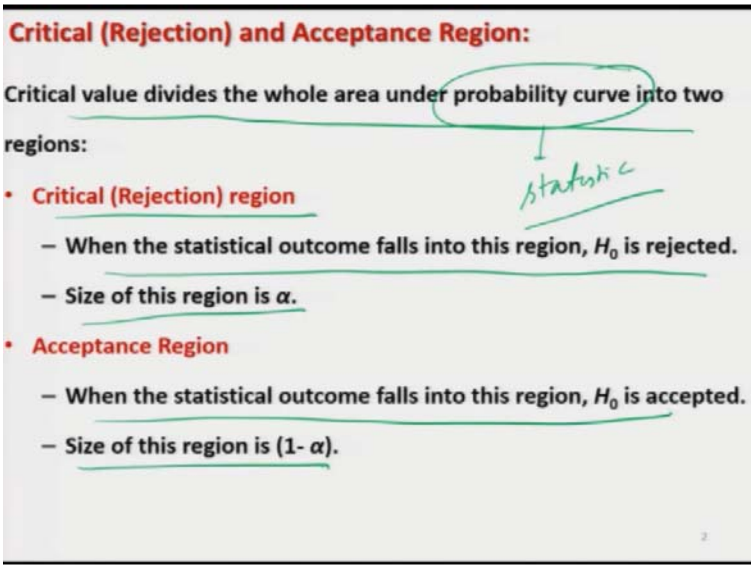
Test Procedures for One Sample Test for Mean with Known Variance

Hello friends. Welcome to the course essentials of data science with R software-1 in which we are trying to understand the basic fundamentals of probability theory and statistical inference and you can recall that in the last long lecture we had initiated a discussion on the topics of test of hypothesis and we had discussed different types of concept definitions and I tried my best to interconnect them to give you a theoretical background.

Now, in this lecture we are going to implement them and now in this lecture I will try to take up an example where I would try to show you how you can conduct a test of hypothesis for the mean of a normal distribution. And here in this case I am going to assume that the variance σ^2 is known.

So, if you remember we had done the similar thing earlier also. So, in this lecture I will assume the variance is known and in the next lecture I will try to show you ah that how to implement it when the variance is unknown. So, we begin our lecture and I expect that you have taken a good revision of the earlier lecture otherwise there may be some problems, you may not understand couple of things because I am assuming that you have revised the lecture.

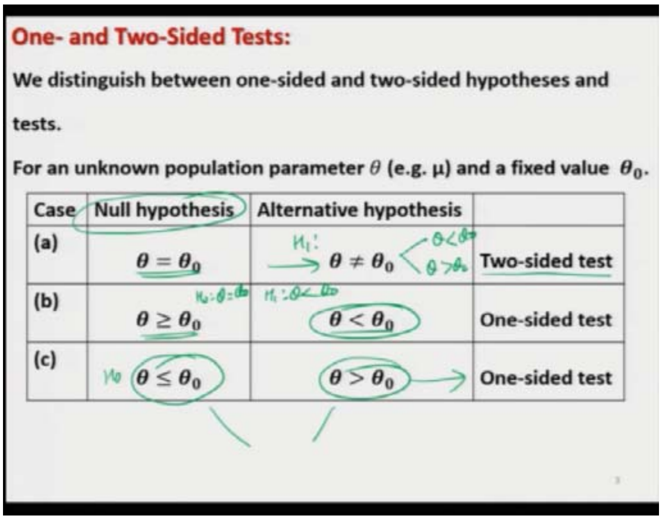
(Refer Slide Time: 01:38)



So, let us begin our lecture. So, now just for your quick revision our basic objective is this that we want to define a critical region that means we want to divide the space of $X_1, X_2,\ldots, X_n$ into two parts such that in one region I can say the hypothesis is going to be accepted and in another region I can say that the hypothesis is going to be rejected. So, we have critical values which divide the whole area under the probability curve into two regions. Probability curve of what?

You will see there will be some statistics so that will be the probability curve of those statistic that may be like t distribution, Chi-square distribution, F distribution, normal distribution and so on. So, one is the critical region or this is also called the region of rejection, so in this region when the statistical outcome falls then we say that H_0 is rejected and the size of this reason is α and another is acceptance region when the statistical outcome falls in this region, we say that H_0 is accepted and the size of this region is $1 - \alpha$.

(Refer Slide Time: 02:43)



So, now we can have different types of this hypothesis that we can divide them into two types of definition which are one sided and two sided test of hypothesis and then we will try to develop the test. So, first we try to understand that under what type of condition we can say test is one sided or two-sided. It depends whether our hypothesis are one-sided or two-sided and based on that we try to define the critical regions.

One thing I can inform you that at this moment, I am going to show you that what are the basic fundamentals of the test of hypothesis, that how are you really going to implement it using the statistical fundamentals but surely when you are working in the data science then these decisions are going to be taken based on some software outcome.

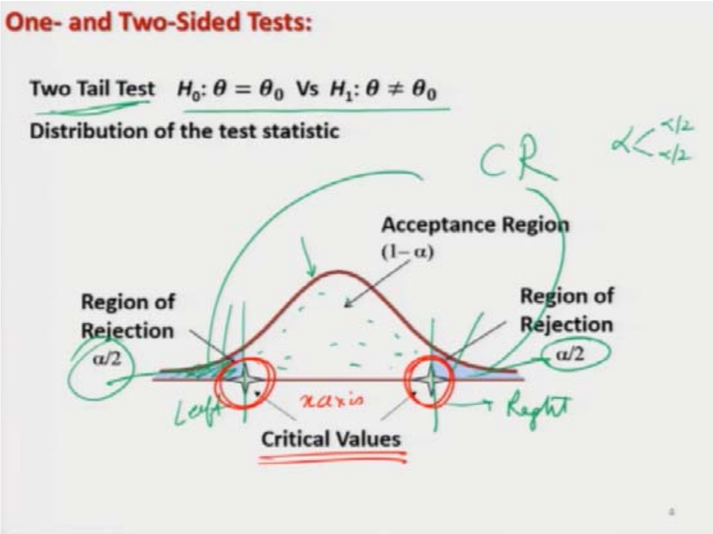
So, when these rules are implemented over the soft year then the decision rule become little bit different, the way we try to decide whether the null hypothesis is going to be accepted or rejected that may be little bit different but it is very important for you to understand what is really happening there and in order to understand that thing, you need to understand this basic fundamental that how the critical regions are divided, how do they look in one sided, two sided test and believe me when you will really conduct the test of hypothesis on a real set of data that will be a job of fraction of second that is my promise to you.

But it is very important for you to understand what exactly you are doing or the software is doing. So, now let me try to give you an idea about the one sided and two sided test. So, suppose your null hypothesis is like that H_0: $\theta = \theta_0$, where θ is some parameter and θ_0 is some known value. For example θ can be the mean of a normal population and θ_0 can be some pre-assigned value of the mean. So, and suppose the alternative here is $H_1: \theta \neq \theta_0$. So, now you can see here we have two option whether $\theta < \theta_0$ or $\theta > \theta_0$.

So, that is why this is called as two-sided test and you will see that the critical region in this case will be divided in two sides of the distribution also. And similarly, if you try to take here another here a test like as here $H_0 : \theta \geq \theta_0$ versus $\theta < \theta_0$ or that's can this can also be H_0 $\theta = \theta_0$ versus say H1 θ $\theta \leq \theta_0$, right.

These are actually one sided test right because you can see here we are trying to take the decision only in the one side and the corresponding critical regions are also going to be on the one side of the region and similarly if you try to take here another option that $H_0 : \theta \leq \theta_0$ and alternative here is $\theta > \theta_0$ then this is also a one sided test and you can see here in both the cases both the null hypothesis and alternative hypothesis they are disjoined, they are mutually exclusive.

(Refer Slide Time: 05:57)

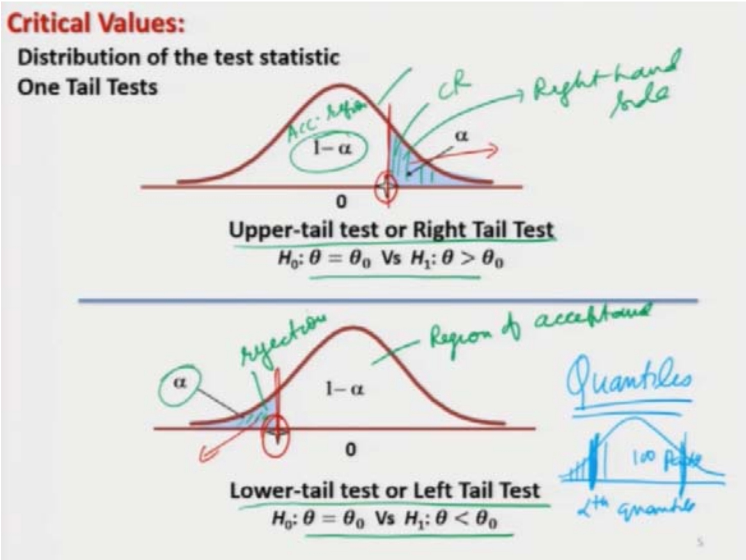


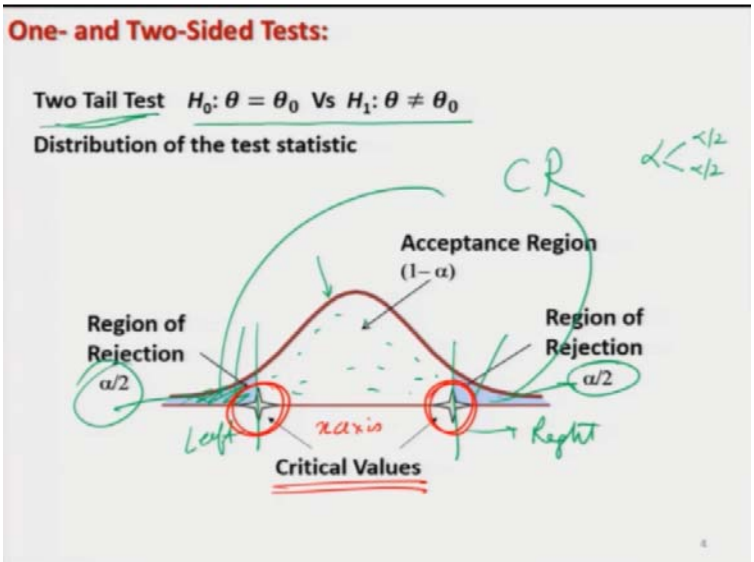
So, now what you will see that when we are trying to conduct a two-sided test then the critical region will lie on the both side of the distribution of the statistics. Distribution of the statistics means we are going to find here are statistics which will follow distribution like a normal Chi-square t, F etc. So, this red line this is going to indicate the distribution of that statistic. So, for example, in case you if you are trying to take here two-sided test this is also called as Two Tail Test because the critical regions which are defined here this region and this region and they are in the blue shade. So, they are on the both sides of the distribution, on the left hand side as well as on the right hand side.

So, both these are they are the critical region or the region of rejection and the area in between which is here in the white shade, this is the acceptance region. So, what we try to do that in order to get a good test I can say you try to take the probability of Type One Error as α and you try to divide it into two parts $\alpha/2$ and you try to take the $\alpha/2$ region on the left hand side here you can see and the $\alpha/2$ region on the right hand side of the distribution here like this one and these values which you are trying to obtain here which are trying to partition the total area under the curve into these three different regions so these are the value on the x-axis right they are called as critical values.

And if you remember earlier in the lecture, I had used this word critical values, so these are the critical values actually. So, critical values are simply the values on the x-axis corresponding to which they are trying to divide the region into acceptance region and rejection region.

(Refer Slide Time: 07:55)





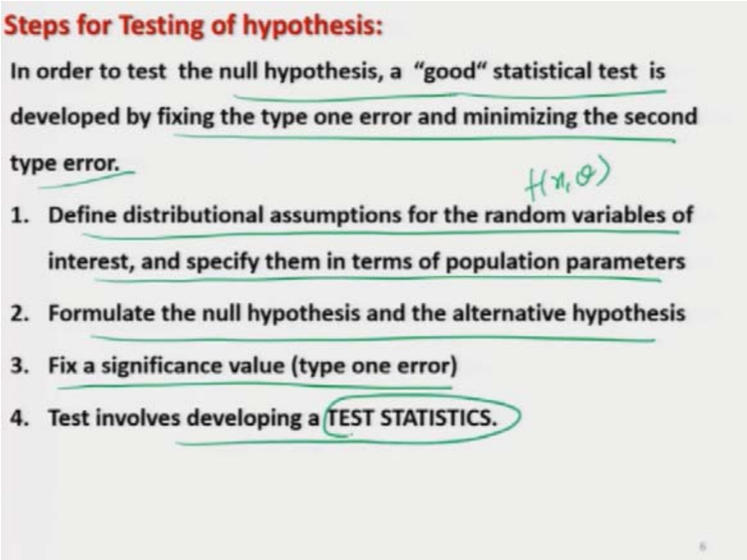
Now, similarly if you try to take here the one sided test, so in this case if you are trying to take the level of significance or the probability of Type One Error as α, then in case if you are trying to take a hypothesis like this one H_0 $\theta = \theta_0$ versus H_1: $\theta > \theta_0$ then the critical region is going to lie on the right hand side of the curve. So, this is the curve of the distribution of the test statistics. So, you can see here this blue shaded area is the α and the remaining area here is $1 - \alpha$. So, this is the critical region and this is the acceptance region.

So, this is called as upper tail test or right tail test also because the critical region is lying on the right-hand side of the distribution. And similarly in case if you are trying to consider the hypothesis H_0 $\theta = \theta_0$ versus H_1: $\theta < \theta_0$ then the critical region is going to lie on the left hand side of the curve. And the size of this area is going to be alpha, so this is also the level of significance that we need to fix before we try to conduct the experiment. So, since the critical region is lying on the left hand side of this distribution, so this is also called as lower tail test or left tail test.

So, what we try to do here that we simply try to compute the value of the test statistics and we try to see that in this curve, this one or this one depending on the nature of hypothesis where it is lying in the area of region of acceptance or rejection. So, this is for example here the region of acceptance and this here is the region of rejection. Now, I would like to have your attention here, in case if you try to see what are these critical values the critical values in both these cases they are the value on the x-axis which are trying to partition the total area into two parts such that the area in the upper tail test is α on the right hand side of the distribution and in the case of left hand or left tail test, the area is on the left hand side of the distribution. Now, if you try to see what are those critical values can you recall the definition of Quantiles?

You are simply trying to divide the total area into here 100 parts and then you are trying to find out here the α Quantile on the either on the left hand side or you can similarly find out here the $1 - \alpha$ Quantile or in this case you can find out the value of the $\alpha/2$ with Quantile. So, now either you try to use the table or you try to use the software that is up to you right. So, but this α probability and these critical values these are very easily obtainable from the software.

(Refer Slide Time: 11:00)

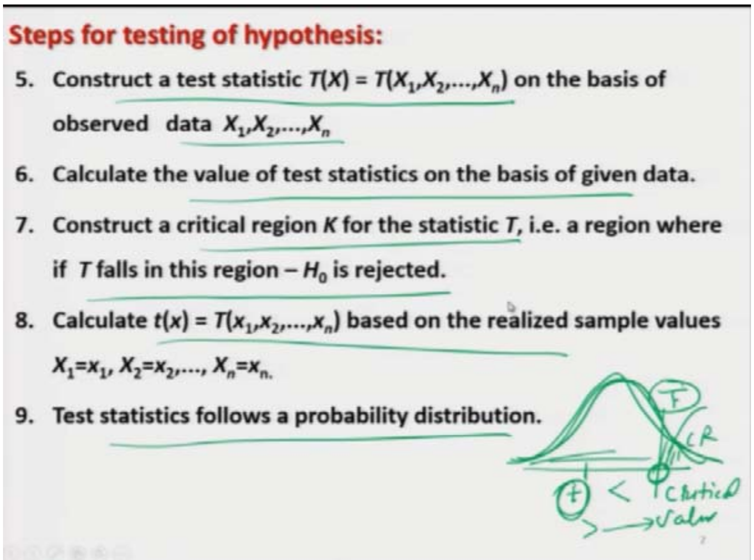


So, now what are the steps for test of hypothesis? So, in case if you want to test a null hypothesis and that is going to be done by a good statistical test. Good statistical test means where the whether you have the best critical reason or the test is most powerful test or uniformly most powerful test that means the test in which the power of the test is maximum. So, they are obtained by developing by fixing the Type One Error and minimizing the Type Two Error because they are obtained by the Neyman Pearson lemma although we are not giving here the proof of these tests but that is the way they have been obtained either using the Neyman Peasron lemma or the likelihood ratio test, if you want to go into more details you can pick up any statistics books and you will find such derivations over there. So, now the first step is define distributional assumption for the random variable of interest and specify them in the form of population parameters.

So, you need to define here f(x θ), that whether the population is normal or binomial or Poisson and what are the corresponding parameters in which you are interested. Then you try to formulate the null hypothesis and alternative hypothesis. How to formulate them? You have now learned that the null hypothesis is formulated in such a way such that the Type One Error is more serious than the Type Two Error and then we try to fix a significance level or equivalently the probability of Type One Error and then after that one can use the Neyman Pearson lemma or the likelihood ratio test to develop a test statistic.

So, this test statistics is a function which is dependent only on the sample observations and we can obtain the value of such test statistics and we can see whether this test statistics is going to lie in the region of acceptance or rejection. So, obviously when you are trying to conduct different types of test of hypothesis, depending on the test of hypothesis this test statistics is going to be different. For example the test statistics for conducting a test of hypothesis for the mean when σ^2 is known and when σ^2 is unknown they are different. The test statistics for testing about the variance of the normal population that is going to be different than the test statistic that is used for the mean of the normal population.

(Refer Slide Time: 13:26)

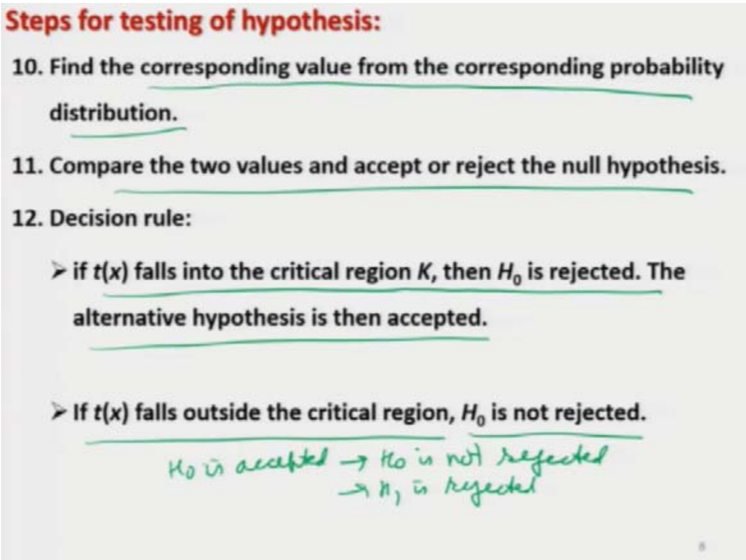


So, it means we have to somehow construct the test statistics and then we have to find out its value on the basis of observed data and we have to calculate the value of the test statistics. Now, we have to construct a critical region for this statistics t. So, that is the region where if t falls in this region then H_0 is rejected. So, now we are simply going to do one thing that we are going to find out the value of the test statistics on the basis of given set of data and we are trying to just see whether the value is lying in the region of acceptance or rejection.

For example if this is the distribution of t and suppose this is your here critical region and you find out here the value of small t suppose small t comes out to be here, so this is here the critical value. So, obviously if this value of t is lying in this region that mean this value is smaller than the critical value then you are going to accept the hypothesis and if this value lies in the critical reason that means t is greater than the critical value, then you are going to reject the hypothesis.

As simple as that and then for that we also need that what is the distribution of t so that we can create this type of curve. So, this test statistics follows a certain distribution and usually you will see that it is going to be one of the sampling distribution like as Chi-square, t or F or that can be normal also depending on the requirement.

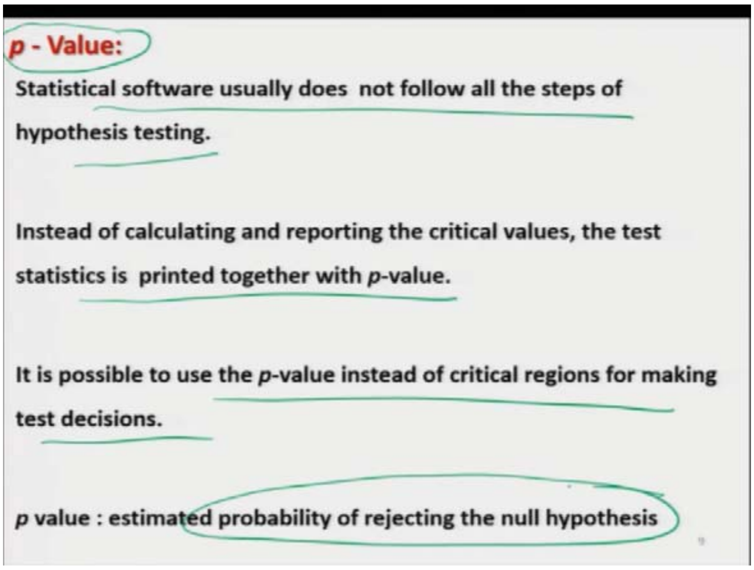
(Refer Slide Time: 14:53)



So, then what are we going to do we are simply going to find out the corresponding value of the statistics on the basis of data as well as on the basis of given distribution and then we try to compare them and after comparison if the calculated value of the test statistics is going to lie in the region of acceptance, we accept H_0 and if the calculated value of the test statistic, that is going to lie in the region of rejection then we reject the H_0. So, the decision rule becomes very simple that if T(X) that is the value of the statistics on the basis of given sample of data falls in the critical region then H_0 is rejected and automatically obviously the alternative hypothesis is then accepted.

The other possibility is that if t(x) falls outside the critical region that means it is lying in the region of acceptance, then H_0 is not rejected, H_0 is accepted. So, when I say that H_0 is accepted, this automatically means that H_0 is not rejected as well as H_1 is rejected because only one event can occur at a time. So, this is a simple language and similarly if I say that H_1 is accepted that means H_0 is rejected.

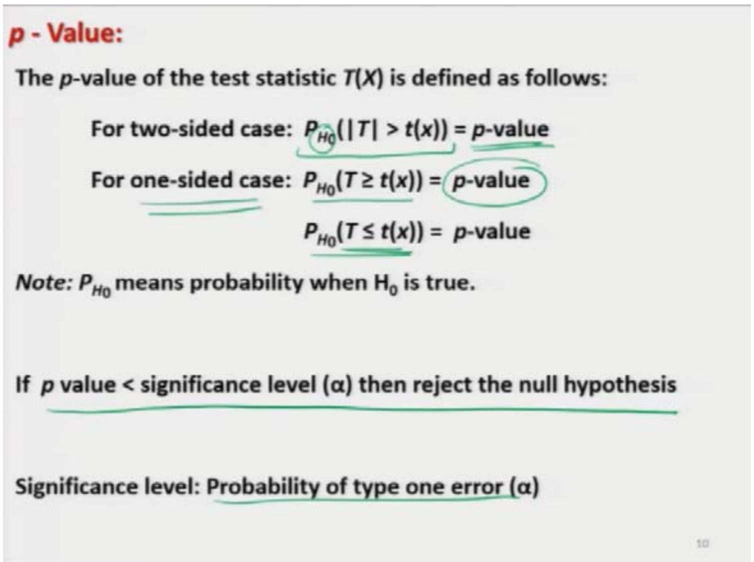
(Refer Slide Time: 16:11)



So, now there is one thing more which is very important nowadays particularly when you are trying to work with the software in data science. This is p-value, small p-value so this statistical software usually does not follow the steps of the hypothesis testing what I have just explained you but the basic fundamentals instead they try to calculate and report the critical values and they give you the results different types of details and they will find out the so called p-value. This is the name actually this is the most popular name and they try to use this p-value to help us in taking a decision whether the hypothesis is going to be accepted or not.

So, in this case we do not need to look into the critical reasons, we do not need to look into the probabilities we do not need to compute the Quantiles from the software but we simply have to look into the p- values. What is this p-value? This is the estimated probability of rejecting the null hypothesis.

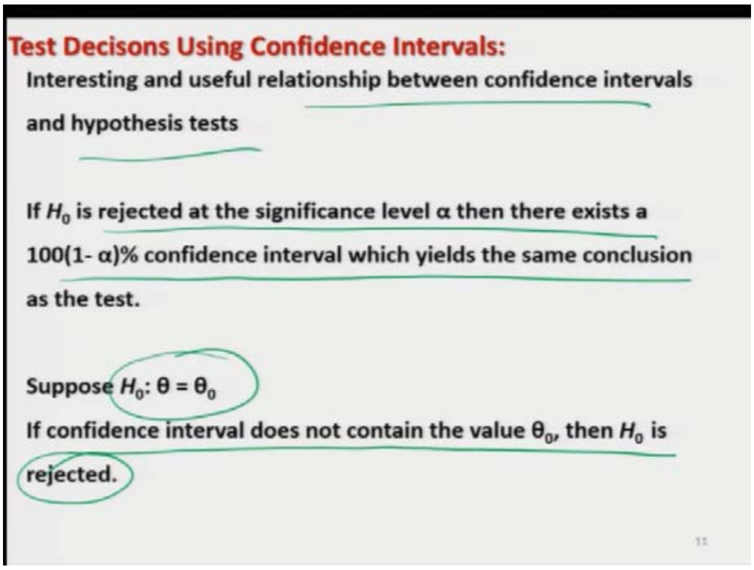
(Refer Slide Time: 17:13)



So, obviously if you are trying to have a one sided or two headed case then this p-values can be defined accordingly, for example, for the two sided case the p-value is defined by this probability that absolute value of t is greater than the calculated value of the statistics under H_0 when I am writing here p and underscore H_0, that means H_0 is true that is under H_0, this probability will give you the value of p-value and in case if you are trying to deal with one sided test of hypothesis, then the probability that the t that is the statistics is greater than or equal to the calculated value t(x) on the which is obtained on the basis of given set of data under H_0 that is when H_0 is true, this probability is called as p-value and similarly once you are trying to get a left hand tail test, then this probability is defined by t less than equal to t(x).

So, now what is the decision rule that is very simple, for a given set of data the software will compute the required statistics and it will also compute the value of p-value and if p-value is less than the significance level that is α then we reject the null hypothesis. This is the very simple rule that you have to keep in mind and what is the significance level, this is the probability of Type One Error that is indicated by α. So, always remember one thing, reject the null hypothesis if p-value is less than α that is all.

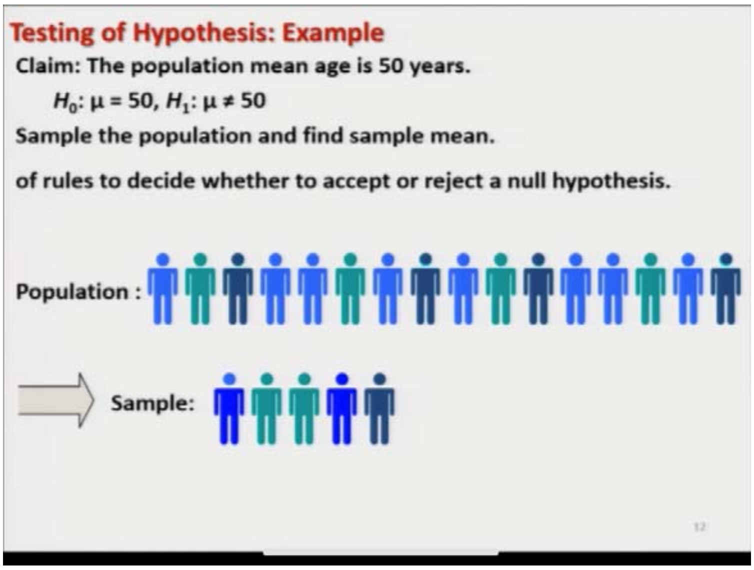
(Refer Slide Time: 18:41)



So, now there is another thing, you already have done the confidence interval. So, the decisions about the confidence intervals can also help in test of hypothesis and the decision to accept or reject a hypothesis can also be taken on the basis of confidence interval. So, there is a very useful relationship between the confidence interval and test of hypothesis. So, this is like this that if H_0 is rejected at α level significance then there exist a $100(1 - \alpha)\%$ confidence interval which is the same conclusion as the conclusion of the test.

So, for example, suppose we are interested in testing a hypothesis H_0 $\theta = \theta$, θ_0 can be μ it can be variance σ^2 or it can be any function of the parameters. So, the rule is very simple, if confidence interval does not contain the value θ not which is specified, then H_0 is rejected and that is why you will see that in many of the software, instead of giving the p-value they also give you the value of the confidence interval and then you have to simply see whether the value θ_0 is lying within the confidence interval or not and based on that you can take a call whether H_0 is accepted or rejected.

(Refer Slide Time: 19:57)

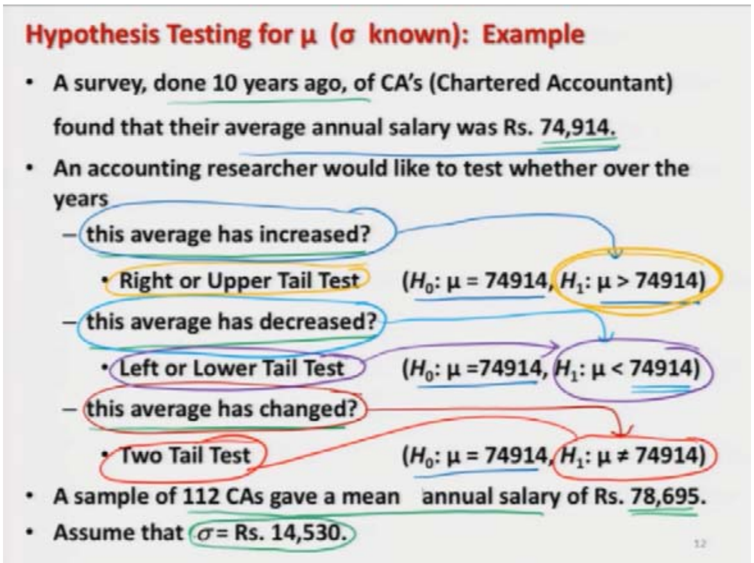


Now, let us try to take a very simple example and with this simple example I want to show you a couple of things. That number one, how are we really going to work and there is a big confusion among the many students. They think that if they try to change the hypothesis, the alternative hypothesis because for example if I say H_0 $\theta = \theta_0$ is the null hypothesis then the alternative hypothesis can be constructed like H_1: $\theta < \theta_0$ or $\theta > \theta_0$ or $\theta \neq \theta_0$.

So, many times I have seen in my experience that people get confused and they try to think that if they try to change the alternative hypothesis that instead of taking θ greater than θ not if they try to consider θ less than θ not or θ is not equal to θ not, the statistical conclusions are going to be changed. This is wrong, this is myth and this is what exactly I want to show you here with this simple example. I will take the same data set and in this example I will give you in detail that what are the steps involved in the test of hypothesis.

How are you going to take a conclusion and this process will be true for all the test of hypothesis that we are going to do in this course and actually you will also learn which are not included in the course also. So, let us try to understand this thing and so I am going to repeat the same example three times. So, you have to just understand how the things are happening and you will see that it is not difficult at all, provided you understand the basic fundamentals.

(Refer Slide Time: 21:50)



So, let us come to our slides and try to consider this simple example so now i am going to consider here an example and through this example, I will try to develop the test of hypothesis for the mean of a normal population when σ is known. So, the example is like this, suppose a survey was conducted 10 years back to know the salaries of the chartered accountants and it was found that the annual salary was Rs.74,914 well that's a hypothetical value the salaries of chartered accountants are much higher.

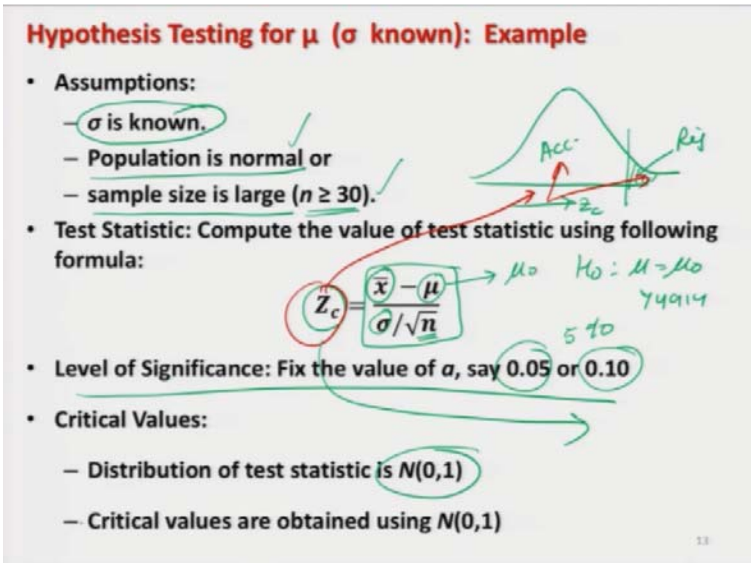
Okay, so now an accounting researcher want to test and wants to know what happened after that survey whether this average value have increased now or they have decreased. So, the researcher wants to know whether this average value has increased over a period of years or this average value has decreased or this average value has changed. So, now if you try to see depending on the objective you are going to choose the right form of the hypothesis. So, obviously the null hypothesis is going to be that the average annual salary was Rs.74,914 rupees.

So, H_0 μwill be $H_0 : \mu = 74{,}914$ and this will remain in the other two cases also. Now, try to look at the question, so if you try to see here in this question whether this average has increased or not, this can be indicated by the alternative hypothesis that now what is happening that the we are assuming that the salary has increased so $H_1 : \mu > 74914$. And similarly in case if the objective is to know whether this average has decreased then this can be represented by the alternative hypothesis here H_1 $\mu$is smaller than 74914 and in case if somebody wants to test here whether this average has changed, change means whether the salary has increased or decreased anything can happen. So, this question can be translated by the alternative hypothesis H_1 $\mu$is not equal to 74914.

Now, in this case what you have to just keep in mind that when you are trying to take the alternative hypothesis of this type H_1 $\mu$is greater than 74914 or $\mu$is greater than say μnot then this is going to be a right tail test or upper tail test that you already have understood what are these values. In case if you are trying to take the alternative hypothesis like h1 $\mu$is less than μnot or H1 is $\mu$is less than 74,914 then your the test is left tail test or lower tail test and the critical region will lie on the left hand side of the distribution.

And in case if you are trying to take here this type of alternative that H_1 $\mu$is not equal to 74,914, then this will become a two tailed test and the critical region will lie on the left hand side as well as right hand side of the distribution. So, this is what I wanted to show you that how you can construct the hypothesis and how the tests are going to be affected by that one. So, now what is happening, a sample of 112 charted accountants has been taken, their annual salaries have been obtained and their average value has been obtained as Rs.78,695. Now, because in this case I want to illustrate the test procedure when σ^2 is known, so we are simply assuming here that suppose the value of σ is known to be Rs.14,530.

(Refer Slide Time: 25:42)



So, now you see in case if I want to develop a test statistics for conducting this type of test, I need to make some assumption and you will see that here the information or the knowledge that you learned or the topics that you learn during the distribution of test statistics or the sampling distribution Chi-square, t, F they are going to help you. If you do not know you will have no idea how this tests are coming. So, we are assuming here that σ is known, population is normal and the sample size is large that is and greater than or equal to 30.

Now, I am 100 percent confident that you can understand that what is the meaning of population is normal and what is the meaning of n is greater than equal to 30. Recall your discussion when we discuss the normal distribution and t distribution for say more than 30 degrees of freedom and this concept I will use once again when I will try to develop the test statistic when σ^2 is not known. So, now after this we have to develop the test statistics. So, we try to use the Neyman Pearson lemma or the likelihood ratio test and based on that the outcome comes out to be that the test statistic is of this form, $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

And now you know what is this value. You have used this value many, many times in the past and you know that distribution of this quantity here is N(0, 1). So, now what is the rule you have been given the data. So, from the data you have to simply obtain the value of sample mean. You have to give here the specified value of μthat is μnot, if my hypothesis is $H_0 : \mu = \mu_0$.

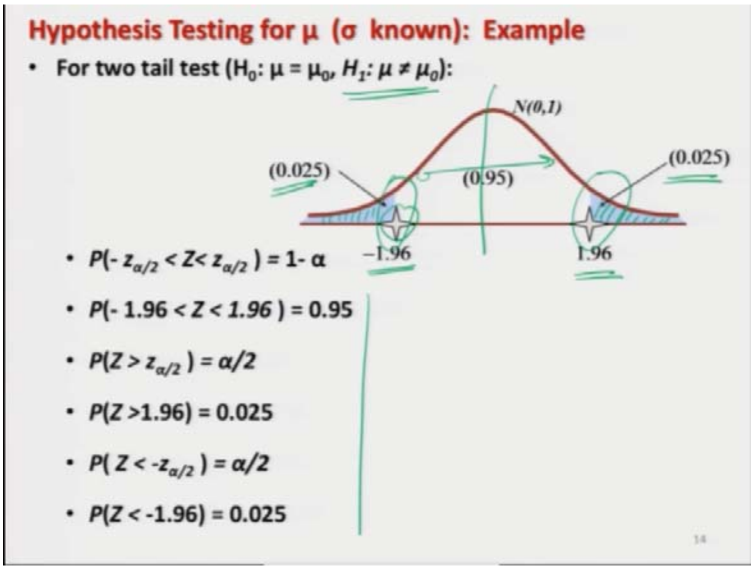
For example in this case, this will be 74,914, Rs.74914 and then a standard deviation is given to you, sample size n is given to you. So, you simply try to substitute those values and try to obtain the value of Z_c and now what you have to see, that suppose I take here the right tail test like this one. So, this is here the acceptance region and this is here the rejection region and this is here on the x-axis, these are the value of Z_c.

So, you simply have to see this value which you are trying to obtain here where this lies in acceptance region or in rejection region and that's all. So, this I will try to show you but anyway for our test procedure now what we have to do, we have to fix our level of significance. That means the fix the value of α. So, usually you will see in practice people are trying to fix α is equal to 0.05 or 0.10 and remember one thing this α is the same what you learned in the confidence interval as $1 - \alpha$ to be the confidence coefficient.

The reason for fixing this 0.05 or 0.10 was simply because these tables were computed manually, so they had prepared the table for some collected values of α but now with the help of a software, you can choose any value of the α and this is also called as 0.05 is called as 5 percent level of significance and similarly 0.10 that is also expressed in terms of percentage. So, this is another condition that we try to express the level of significance in the form of percentage also. Now, you have to simply obtain the critical values.

Now, the question is from where are you going to obtain the critical values? And for that what you have to do, you have to find the distribution of Z_c. Now, how to find the distribution of Z_c? What is the distribution of Z_c? That you already have done that $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ follows a standard normal distribution that is normal with mean 0 and variance 1, so I can find out the critical values from N(0, 1). That is all.

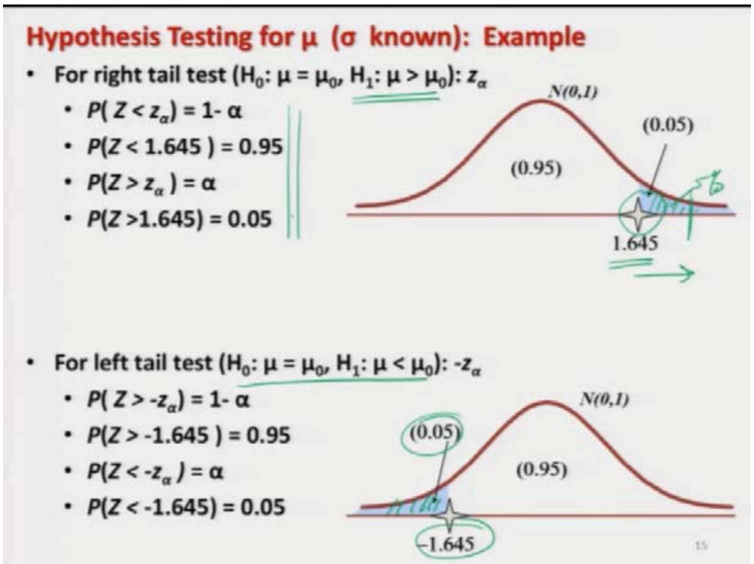
(Refer Slide Time: 29:34)

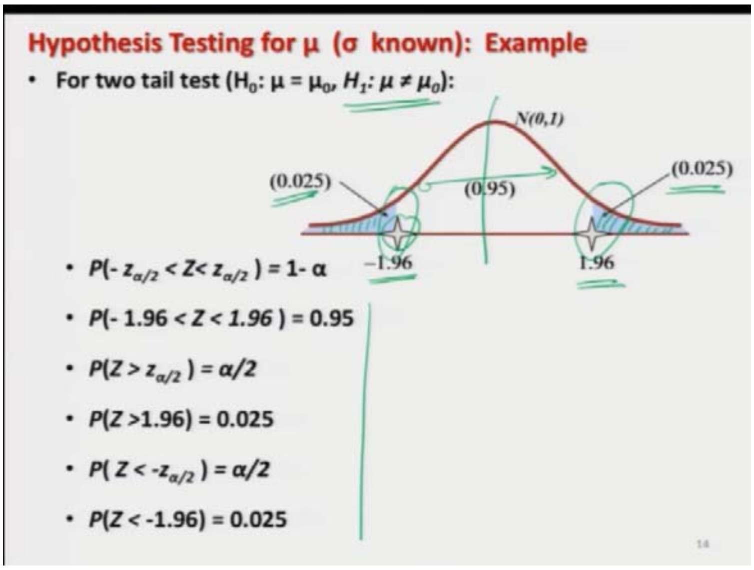


And for example, in this case, if you try to take here α is equal to 5 percent then the case of two-sided test, the critical values are going to be on the left-hand side or on the right-hand side and this α/2 will become simply 0.025, on left hand side and right hand side both and if you try to look from the tables, the value on the x-axis at which this area is 0.025 is − 1.96 and this area is 0.025 it is 1.96. Actually, that can be obtained because this distribution is symmetric around mean.

So, now what you can see here that this area in between this comes out to be 0.95. Okay so and now here are some values which I have given you that if you try to look from the books, you will get these values from there but anyway means you can compute them directly on the software. So, these are the values that is that lies between − 1.96 and + 1.96 and so on.

(Refer Slide Time: 30:38)

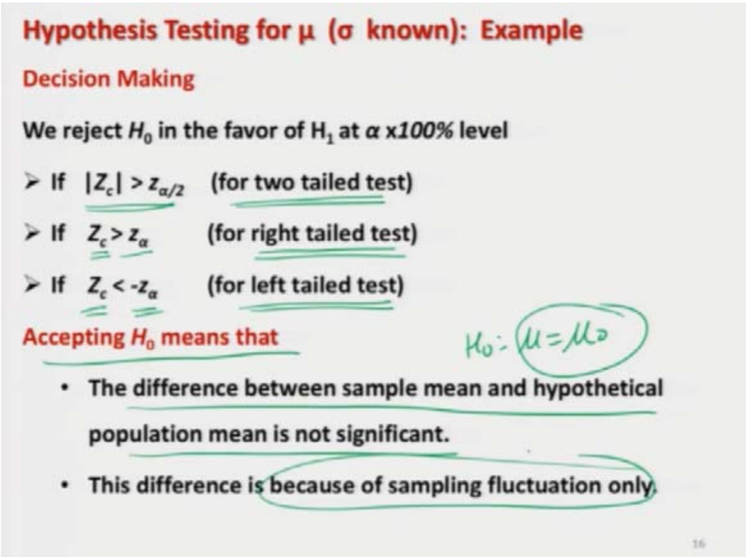


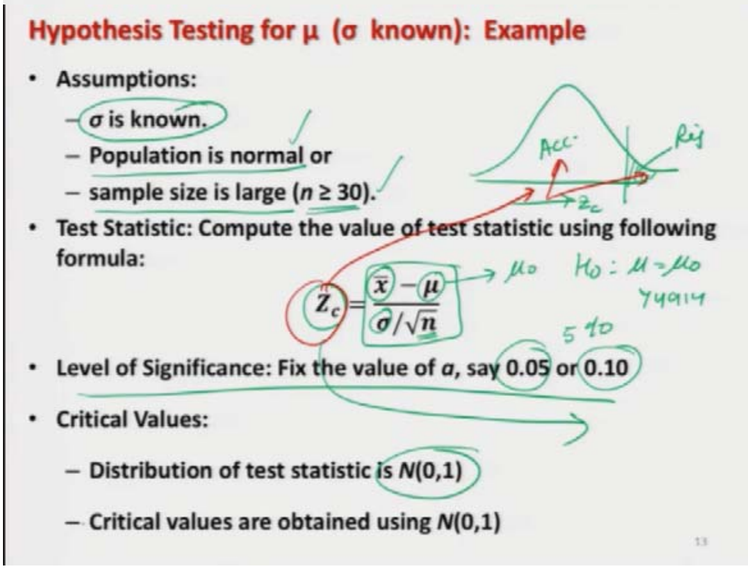


And similarly, if you are trying to deal with the alternative hypothesis like H_0: $\mu = \mu_0$ versus H_1: $\mu > \mu_0$, then the critical reason is going to be lie on the right-hand side and the critical value that you can obtain from the table is now 1.645 because this is the 5 percent area of the total area. So, you can see here when this area is 5 percent, this value is 1.645 and when this area becomes half that is 2.5 percent this value will shift to this new value which is obtained here in this case 1.96.

So, similarly in case if you want to consider the case H_0 $\mu = \mu_0$ versus H_1: $\mu < \mu_0$, then in this case the critical region is going to lie on the left hand side of the distribution of normal 0,1 and this area is going to be simply α. This is 0.05 and corresponding to which the value on the x-axis is −1.645. So, this value which I have given you here probability that Z is less than 1.6545 or greater than 1.645 etc. they are available in the tables but you can obtain directly from the software also.

(Refer Slide Time: 31:49)



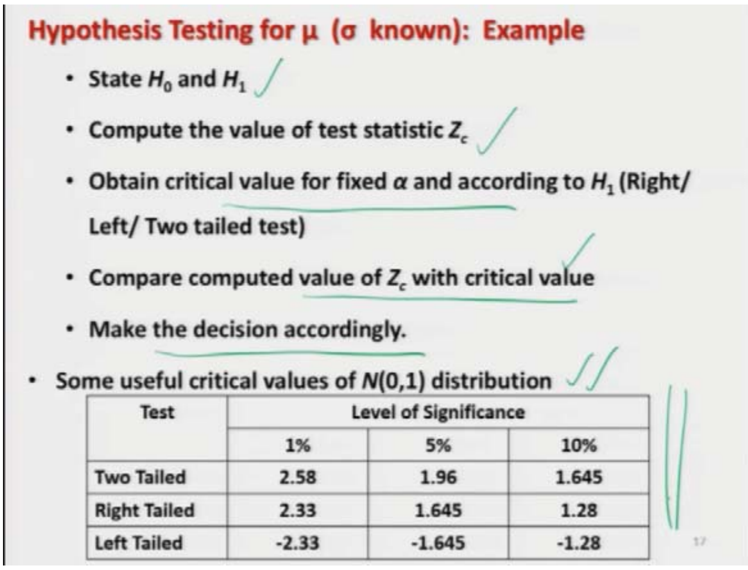


Now, how are you going to make a decision? It is very simple, reject H_0 in the favor of H_1 at α level of significant or say 100 α percent level of significance. Now, depending on the nature of alternative hypothesis, you can take the call. For two scale tests you are simply trying to say absolute value of Zc that you have obtained on the basis of given sample of data, you can see here this is your here Zc right that you have obtained on the basis of given sample of data.

So, that is going to be a numerical value just try to see that absolute value of Zc is greater than Z α/2 or not. If yes then the hypothesis is rejected. In the case of right tail test, what you have to see that the calculated value that is Zc is greater than Z α and in the case of left tail test what you have to see that Zc is less than ⁻Z α. So, simply have to actually see whether the value of Zc, this is the calculated value of Z statistics is lying in the region of acceptance or rejection. And when we try to say accept H_0, what is the meaning of this?

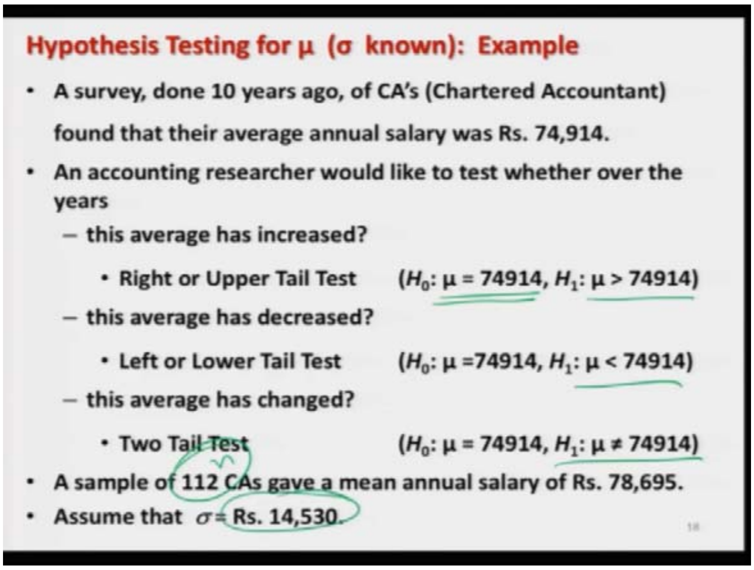
This means that the difference between the sample mean and the hypothetical population mean is not significant because you are trying to say that here H_0 μequal to μ0 is accepted that means the difference between the μand μ0 is not very high and possibly you can accept it at say α level of significant. That means you are allowing some error to be there but that error you know how much error can be there. And the difference between the two value can be attributed that it is due to the sampling fluctuations random errors etc.

(Refer Slide Time: 33:28)



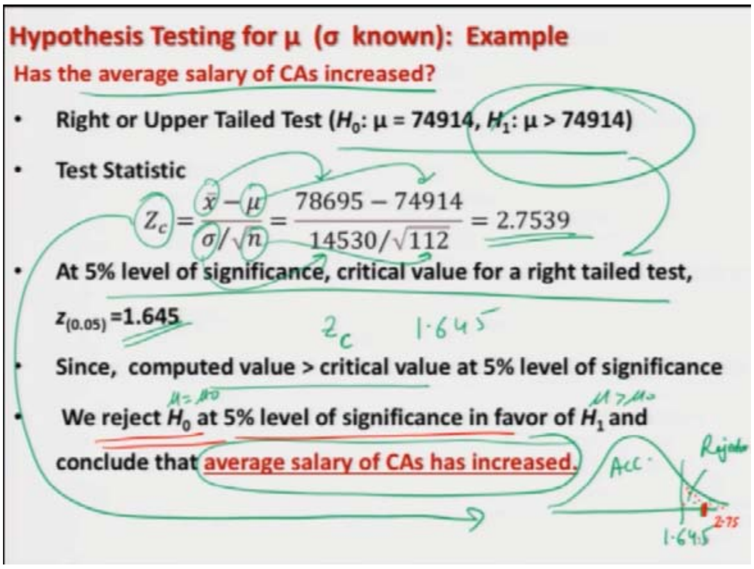
So, now what are the steps that we are going to follow? So, the first step is going to be state your null hypothesis and alternative hypothesis, compute the value of the test statistics. In this case it is Zc, obtain that critical values for a fixed level of significance α and depending on the nature of the test whether it is one tail, two tails, left tail, right tail, two sided etc. you can decide the direction of the critical region then try to compute the value of Zc on the basis of given set of data and try to compare it with the critical value and then you try to make the decisions accordingly. Some useful value of the normal zero one distribution I have given here but they are means available in all the books and you can also compute them directly on the in the software also.

(Refer Slide Time: 34:14)



So, now I come back to my this example. This is the same example which I just considered. So, now here I have μequal to 74914 and then I have taken here three possible alternative hypothesis and small n is going to be here 112 and σ is equal to here 14530.

(Refer Slide Time: 34:34)



So, now I try to implement all these rules and try to see what happens and how are you going to interpret it. You see once you do the algebra with your own hands manually then you will understand what is really happening. In case if you are getting the value directly from the software, it will be fast but possibly at least if you ask me I cannot understand what is really happening and for mathematics I always prefer to do the algebra with my own hand, on my own paper. So, now first I try to consider the first example.

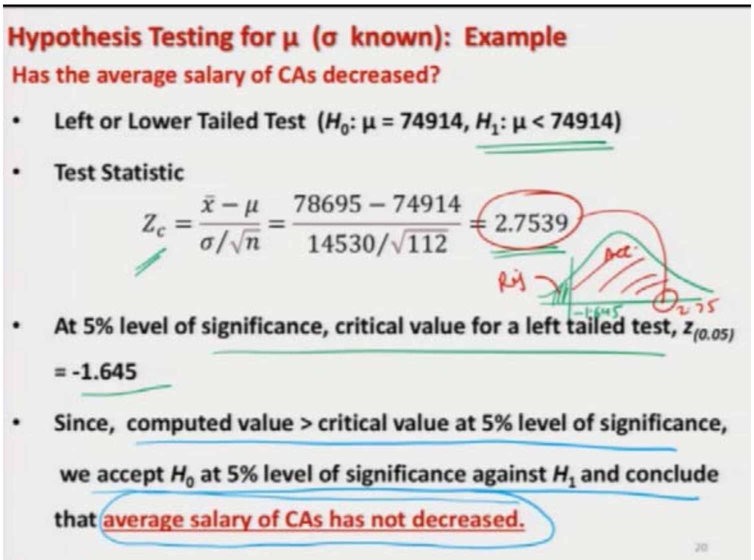
Has the average salary of the chartered accountants increased? So, in this case we have this set of hypothesis H_0: μ =74914 versus H_1: μ > 74914. So, the first rule is now compute your here Zc, now the value of x bar is given to be here 78695, the value of μis given here to be 74914, the value of σ is given here 14530 and small and the sample size here is 112 and if you try to compute it this will come out to be 2.7539.

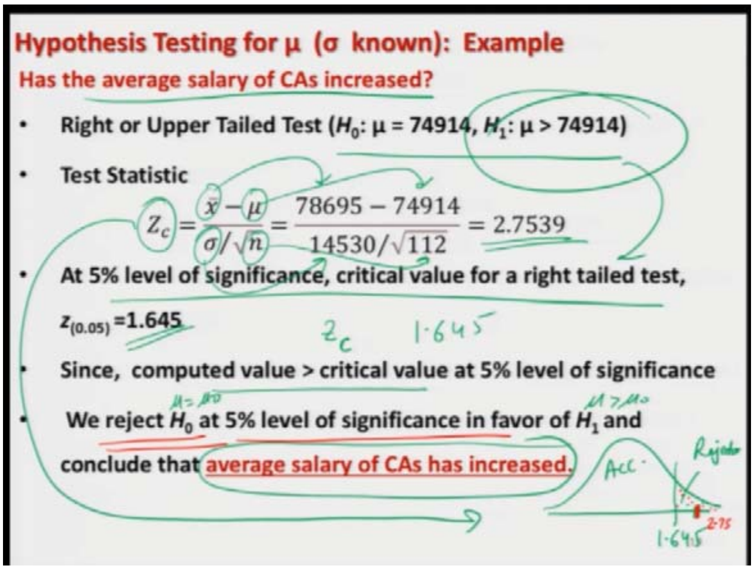
And now in case you try to see the value of Z_c from the table that is at a 5 percent level of significance, the critical value for a right tail test because you can see here this is a right tail test this will come out to be 1.645. So, now you simply have to compute this value of Zc and you have to compare it with 1.645 and you have to see where it lies. So, this computed value is greater than the critical value at 5 percent of level of significance.

Why? You can see here this is your here curve and somewhere the critical value here is 1.645, so this is the area which is the area of rejection and this is the area of your acceptance and now you can see here the value of Z_c what you have computed here this is lying somewhere here. I can use here a different pen. It is lying somewhere here this is 2.75 and so on. So, this value is lying in the region of rejection. So, I can see here that we reject H_0 at 5 percent level of significance in favor of H_1. That means H1 is going to be accepted.

Now, the next question is what is the meaning of this statement that we are rejecting H_0? What is your here H_0? μ is equal to μ_0 and H_1 here is say $\mu$is greater than μ_0. That means it is indicating that the average salary of the charted accountants has increased. So, remember this statement that it is giving you a conclusion that average salary of the chartered accountants has increased.

(Refer Slide Time: 37:13)

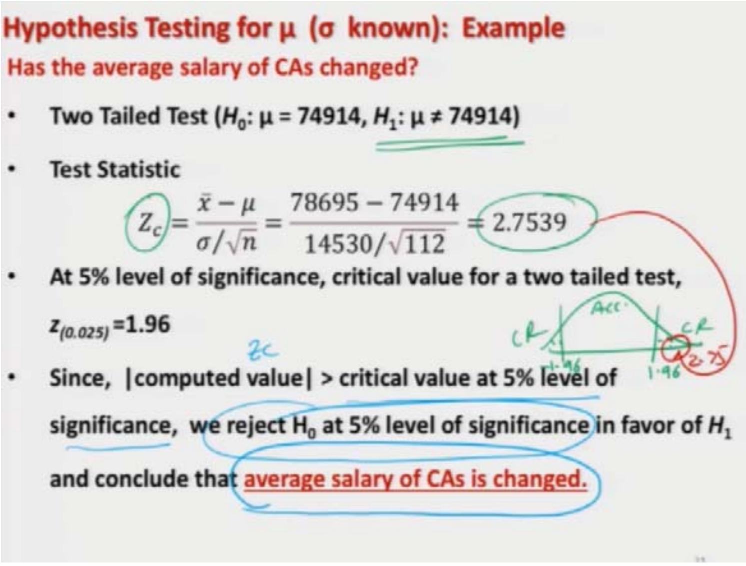


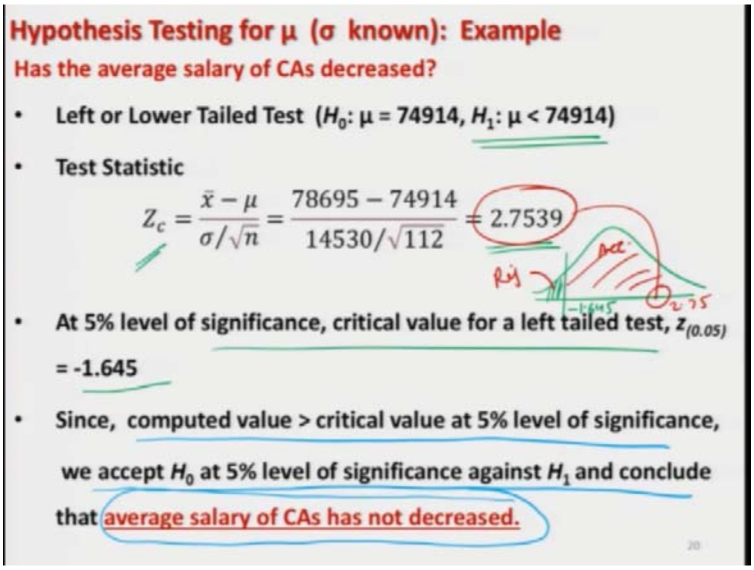


Now, we try to do the same thing. We have this hypothesis where H_1 is μless than 74914 and we try to compute the value of here Z_c which is again going to be the same value but now what we have to do? We have to take the critical region on the left hand side. So, at 5 percent level of significant the critical value is coming out to be $^-$1.645 and now if you try to see what is the value here for this here Z_c, this is 2.7539.

So, this is your here acceptance region and this is your here rejection region. So, this 2.75 is lying somewhere here. So, you can see that here in this case the value of Z_c is lying in the region of this acceptance. So, I can say here that in this case the computed value is greater than the critical value at 5 percent level of significance, so we accept H_0 at 5 percent level of significance against H_1 and now what we conclude? We conclude that the average salary of chartered accountants has not decreased. Now, you can compare the difference between the two statement that average salary of the chartered accountants has increased and the average salary of the chartered accountants has not decreased.

(Refer Slide Time: 38:36)

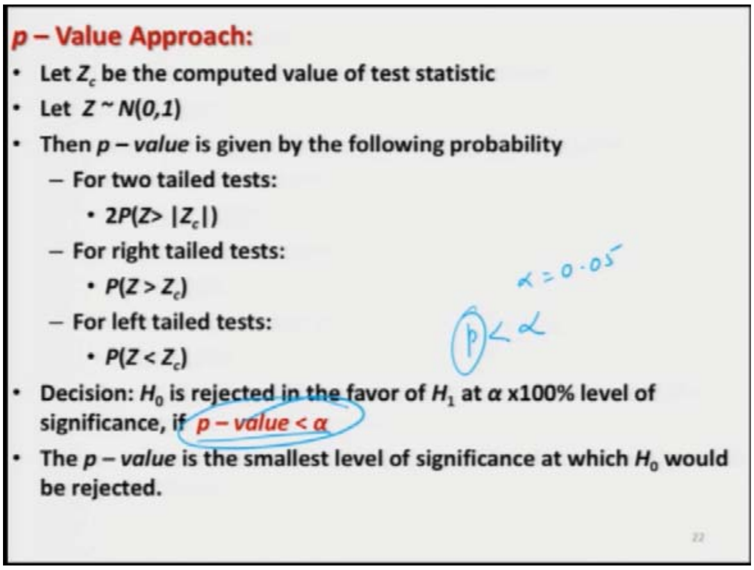




So, you can say that the conclusion is the same. The soul of the conclusion is the same. Now, we try to take here the third option. The third option is where we are trying to take the alternative hypothesis to be two-sided, μ ≠ 74914 and then we try to compute the value of here Zc that will coming out to be the same but in this case what is going to happen, the critical regions are going to lie on the both hand side. So, in this case, this is coming out to be −1.96 and here it is 1.96 which are here the critical region. This is here the critical region, this is here the critical region and this is here the acceptance region and now you can see where this 2.75 is lying. This is lying somewhere here, 2.75.

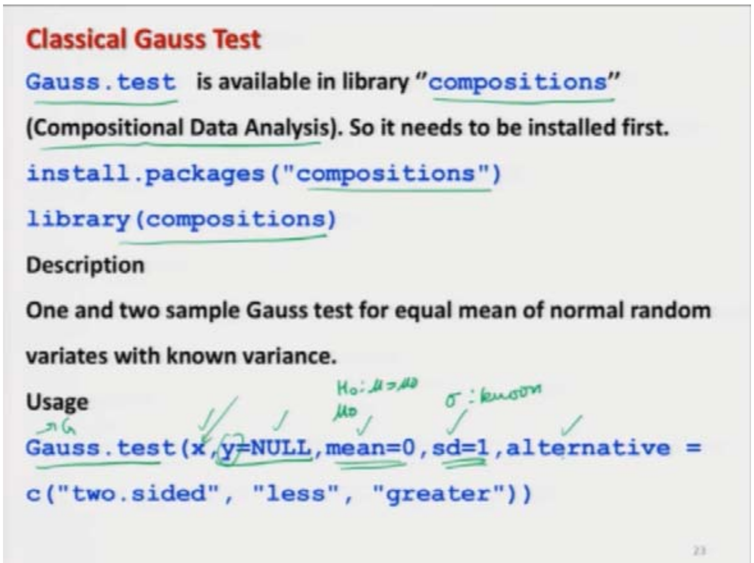
So, this line in the region of rejection, so now we can conclude here that since the absolute value of this computed value Z_c this is greater than the critical value at 5 percent level of significance, so we reject H_0 at 5 percent level of significance in favor of H_1. So, now if you try to translate this statement, this is indicating that we can conclude that average salary of chartered accountants is changed right and if you try to compare this statement with this statement that average salary of the chartered accountant has not decreased or the average salary of the chartered accountants has increased, you can see here that you are getting the same conclusions at least in the soul and the way you are trying to interpret it that is only changing. So, either you try to take the alternative hypothesis to be one-sided or two-sided the final outcome is the same.

(Refer Slide Time: 40:13)



So, do not get confused that you are trying to change or you or do not try to conclude if you think that the conclusions are going to be changed. The same thing I can do in a software also, I simply have to compute here the p-value and based on that I simply have to compare if the p-value is coming out to be less than α or not. So, α here for example I have taken to be 0.05, so whatever is the value of p that can be computed on the basis of software and we can compare it here. So, remember one thing, reject H_0 if this p-value is smaller than α that is the simple rule that you are going to follow but here I cannot show you because I am doing all the calculations manually but definitely when I try to go to the software I can show you this outcome.

(Refer Slide Time: 40:55)

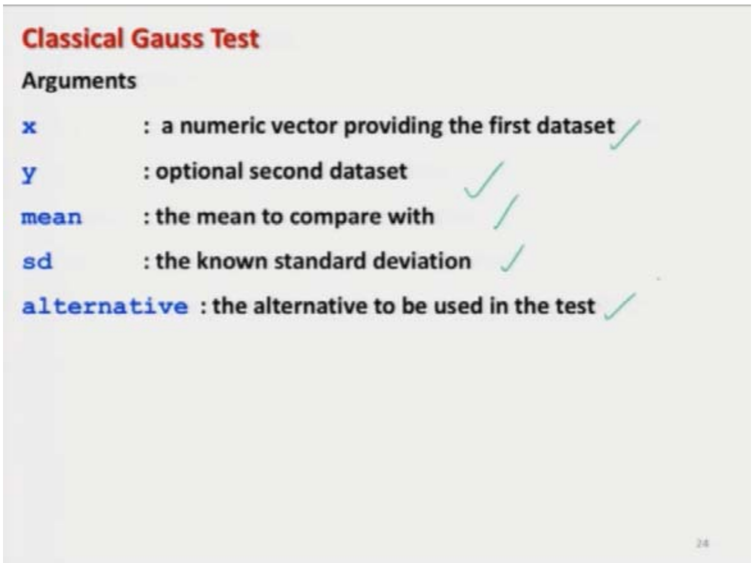


So, now let me try to show you how you can conduct such a test in the R software. So, there is a test whose name is Gauss test, here this G here is capital and this is available in the library say compositions. So, this is a library or a package that has been given by the authors of the book compositional data analysis and so we need to install it first.

So, we so try to install the package compositions and then you try to upload it using the library command and after that the command for one sample, two sample this test what we have just done this is the Gauss test is like this gauss dot test where G is going to be in the capital letters and then inside parenthesis you can see here there are commands here x, y, mean, standard deviation and alternative.

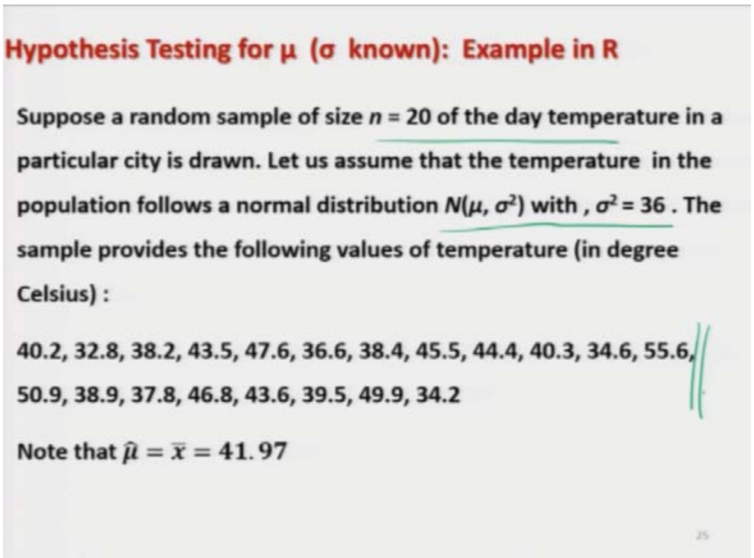
Well because we are dealing here only with the one sample test, so that is why you can see here in this case y is going to be null when we will consider the two sample test then the data about the second sample will be entered in the place of here y. So, in the present case we have the data only for the x and this here mean this is the value of μ0 from H_0 $\mu = \mu_0$ and sd is the value of σ which is here known and alternative here is like depending on the H_1 whether this is two sided or less than type or greater than type. So, you can see here it is very simple.

(Refer Slide Time: 42:28)



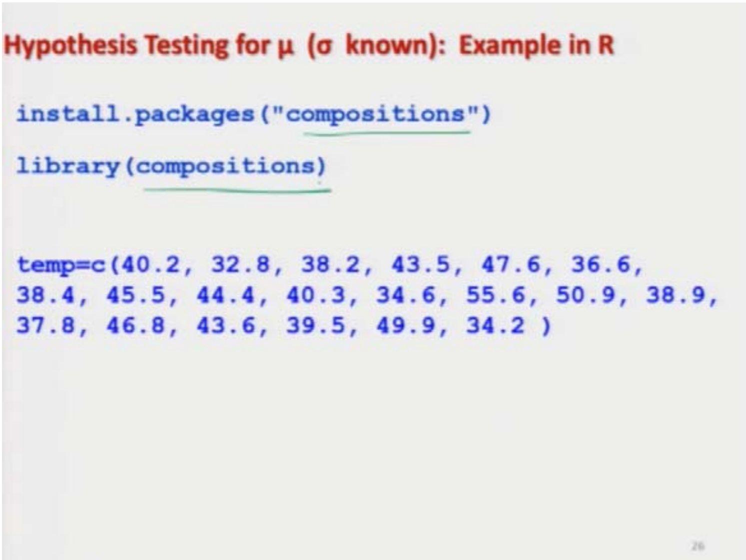
So, now let me try to take an example and you can see here this is the same thing which I just explained you.

(Refer Slide Time: 42:36)



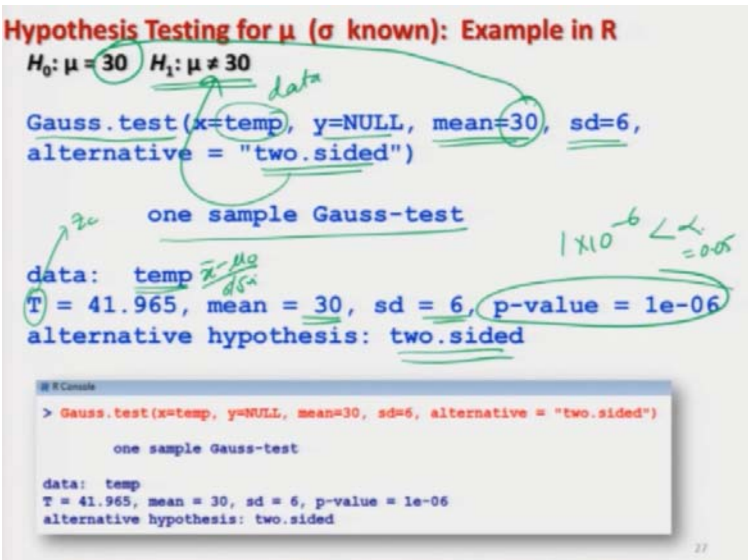
So, now let me take the same example that I took in the case of confidence interval. So, you can compare the values which are coming in the case of this confidence interval and through this test of hypothesis. So, in which so the example was that we have collected a sample of 20 days temperature in a particular city and the population of the temperature is assumed to be normal $\mu\sigma^2$ where σ^2 is known to be 36 and the 20 values on the temperature is recorded in degree Celsius. In this case if you remember we had computed the point estimate μhat as x bar and that will come out to be 41.97 degree Celsius.

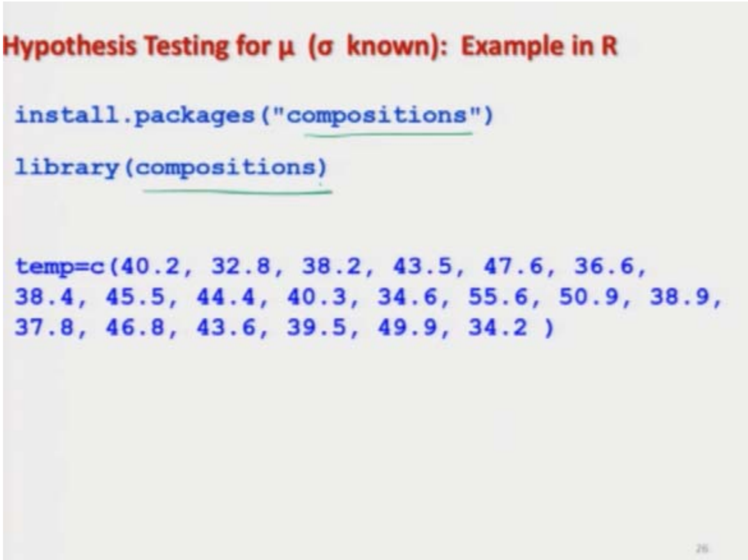
(Refer Slide Time: 43:17)



So, I try to install this package first compositions and that and then I load it and then the data I have stored in the variable temperature.

(Refer Slide Time: 43:28)





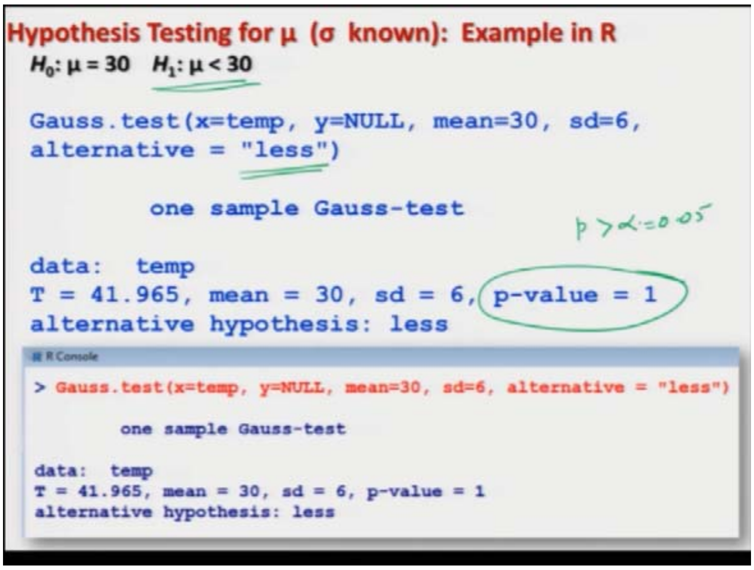
And now if you try to see what is really happening how are you going to give the data here? So, now I am writing considering here the first test that I want to suppose test H_0 μequal to 30 that means the average temperature if you try to look here this data somewhere here is 40, 32, 38 etc. and we want to test the hypothesis that H_0 $\mu$is equal to 30 that means the average temperature is whether close to 30 degrees or not.

So, the alternative I am taking that $\mu$is not equal to 30. So, now I have to give the command here like this Gauss dot test x equal to the temp. Temp is the data, it is not temporary right. It is temp because you have given the data here as a temp. And then y here is null because this is a one sample test, now the mean here is 30.

This 30 is coming from where? Here and then, what is here this sd? Sd is coming from here because in the question it is given that σ^2 is equal to 36. So, and then alternative is a two sided which is coming from here but you have to remember one thing, this the way it is expressed you have to follow the same rule. It is within the double quotes two dot sided and if you try to see its output it will look like this this will give you the heading one sample Gauss test, the data here is temp, the value of that Z statistics this is that is indicated here as a t.

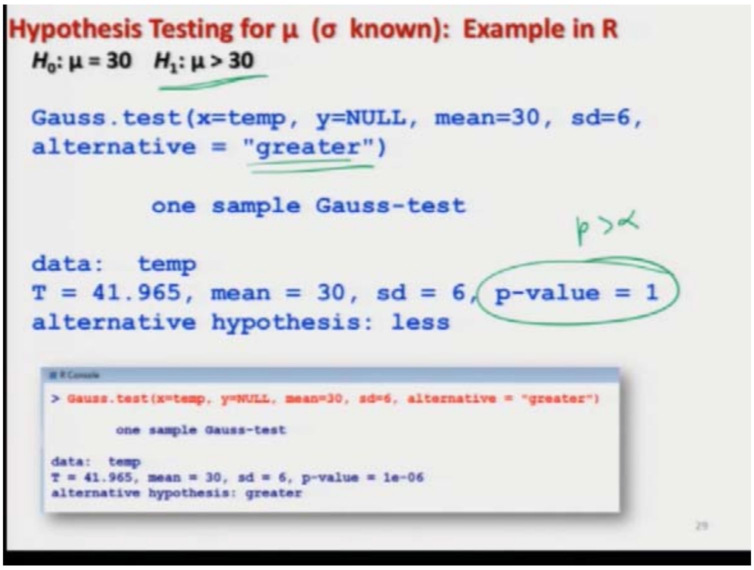
Why this is? Because it is trying to calculate something which is similar to the t statistic that we will see in the next lecture. So, this is the value of here x bar $-$ mu_0 upon σ by root n actually, then the mean here is 30 as the array is 6 and you can see here now this p-value. p-value is coming out to be here 1 into 10 power of $^{-}$6, which is much, much smaller than the value of α if you try to take α to be 0.05 and alternative here is two sided and so you can conclude what is really going to happen. So, in this case the H_0 is rejected and this is here the screenshot.

(Refer Slide Time: 45:36)



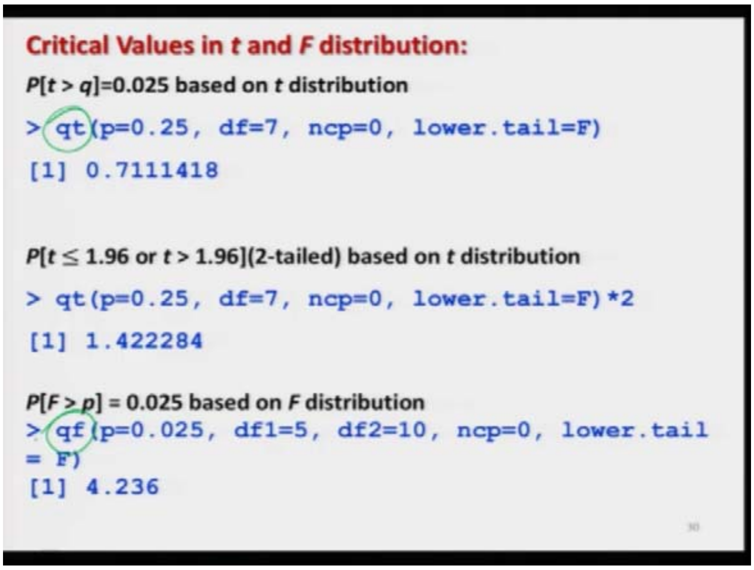
I will try to show you on the R console also and similarly if you try to take the alternative to be here H1 μless than 30, then you have to use the same command but only you have to change the alternative is equal to less inside the double quotes and then you will get here the this type of outcome here. Everything will remain the same except the p-value. Now, you can see here that the p-value is here 1. Do you think that p value is less than alpha? No p-value is greater than alpha, α is equal to 0.05 if you try to take. So, in this case, try to take a conclusion so H_0 is not rejected right. So, here in this case H_0 is accepted and this is the screenshot of the same operation.

(Refer Slide Time: 46:13)



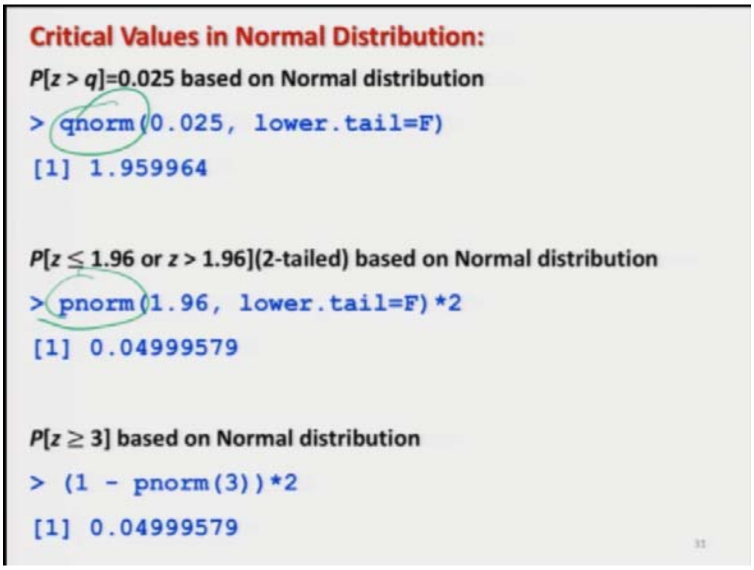
And similarly, if you try to take the alternative here to be H_1: μgreater than 30 then you will have to make only here one change that alternative is equal to greater within the double quotes. And in this case you can see here that p-value is again coming out to be 1 and here in this case p-value is greater than α. So, once again H_0 is not getting rejected. So, this is how you can take the call and this is here the screen shot

(Refer Slide Time: 46:44)



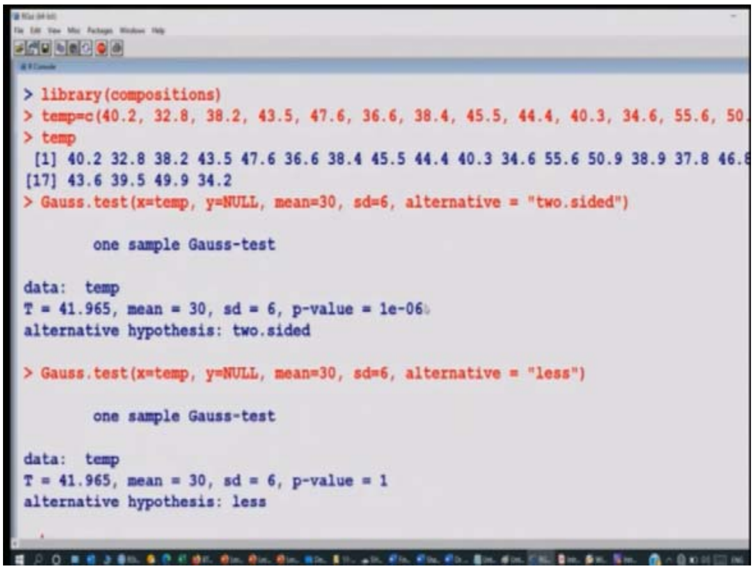
And similarly in these slides, I have just revised the things what I did earlier that if you want to compute the critical values that means you need to find out the Quantiles. So, you know that in the case of t distribution the Quantiles can be computed by the command qt. In the case of F distribution, they can be commuted by the command qf.

(Refer Slide Time: 47:02)



And similarly in the case of normal distribution also, you can compute them by qnorm. So, and the different types of probabilities you can compute by p norm also, so I will not go into that detail because I already had done it but I thought I should at least do it, at least I just inform you so that if you want you can use it.

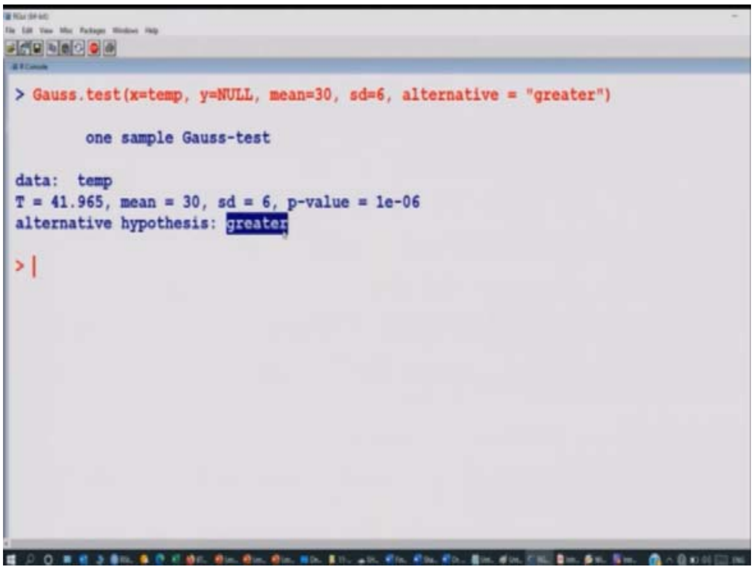
(Refer Slide Time: 47:32)



So, now let me try to come to the R console and try to show you these things on the R console. They are very simple thing but I just want to show you. So, first let me try to load this library well this library is already in my computer, so you do not have to hurry for this thing and now I am entering here the data on the temperature.

So, you can see here this is the data on the temperature and now you try to use this here Gauss test for H_0 µis not equal to µnot, that is two side that you can see here. And you can see here this is here the p-value like this one and if you try to use it for that alternative less than µnot, then you have to simply change the alternative to be less inside the double quotes and you can get here these are the values that you have obtained.

(Refer Slide Time: 48:14)



And if you try to just do it for the H_1: μgreater than mu_0, then you have to give the alternative as a greater and you can see here that this is giving you here the p-value and it is also specified here that means looking at the outcome also, you can decide that what is your alternative hypothesis.

So, now we come to an end to this lecture but you can see at the end, that solving or taking conclusion decisions using the software for the test of hypothesis is very simple. And now you might be wondering that why I have taken such a long lecture for this thing. Do you think that whatever the software outcome you are going to get can you really interpret it if you do not know these fundamentals? And if you try to learn it that okay just try to remember that whether p-value is less than α or not but surely that will be the role of a compounder not of a doctor and we are not here to be the compounder.

So, you can see here that conducting test of hypothesis is very simple. Now, onwards, I have to just inform you that what are the required test statistics for different types of job and after that you have to just follow the steps either in the software or in the manually using the t values or the calculating the statistic, that is very simple.

So, I would now request you try to take some example but try to do them manually first and try to see are you getting the same conclusion which you are going to get in the software and then try to use the same data into the software also and then try to see whether your conclusion which you are making on the basis of the understanding that is matching with this outcome of the software or not.

And try to see into the data and try to make different types of comparison, for example in the same set of data temperature, you can take um this H_0 to be μ is equal to 100 and try to see whether the average is close to 100, well that data says the average is close to something around 35. So, try to practice try to learn more and I will see you in the next lecture with the case when σ^2 is unknown. Till then, good bye.