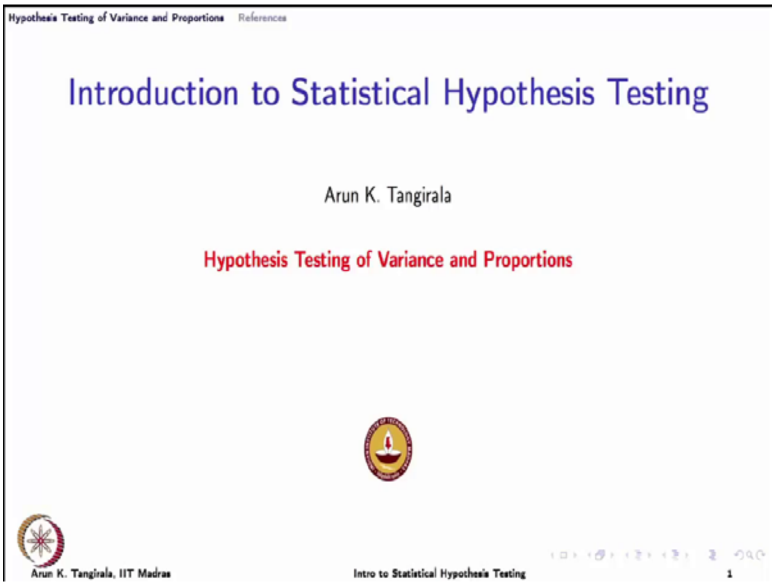
Introduction to Statistical Hypothesis Testing
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

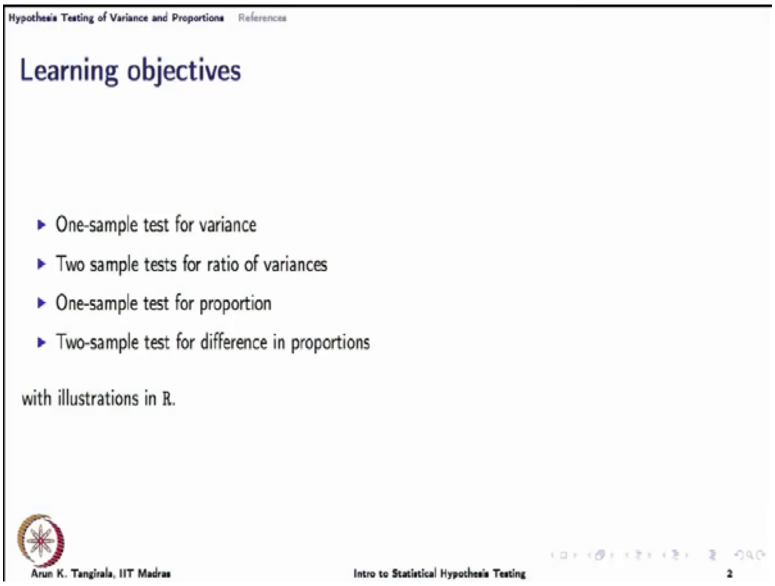
Lecture – 12
Hypothesis Testing of Variance and Proportions

(Refer Slide Time: 00:09)



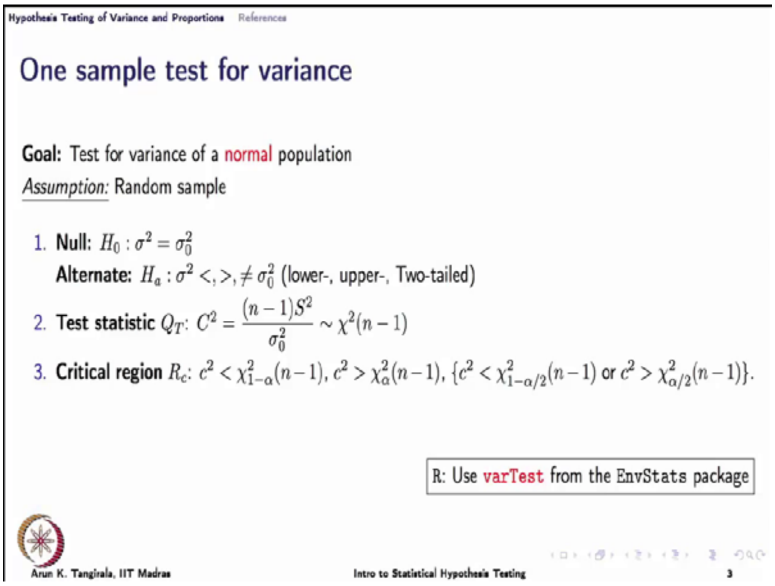
Welcome back to the lectures on Introduction to Statistical Hypothesis Testing. Until now we have that is in this unit, we have looked at the basics of hypothesis testing and we have also learned how to test for means under different conditions both for a single population and two populations. In this lecture, what we are going to focus on is hypothesis testing on of variance and proportions; by variance, we mean variability. Again here, as before we would be looking at the single population and the two-population case that is the one sample test and the two-sample test.

(Refer Slide Time: 00:58)



To be a specific, we will begin with one sample test for variance and then move on to two-sample test for ratio of variability; followed by one sample test for proportion and the two-sample test for difference in proportion. So, you see that when we want to compare variability, we are looking at ratios; whereas, when we are looking at comparing proportions, we are looking at differences. So, you may want to think has to why we do not look at either difference in variability as a parameter to be tested or the ratio of proportions as a parameter to be tested. After all we want to compare and comparison can be done either by ratio or difference, but there is some specific reason as to why we do this and think about it and see if you can answer these questions. And as usual as in the previous lecture, I will show you how to conduct this test in R.

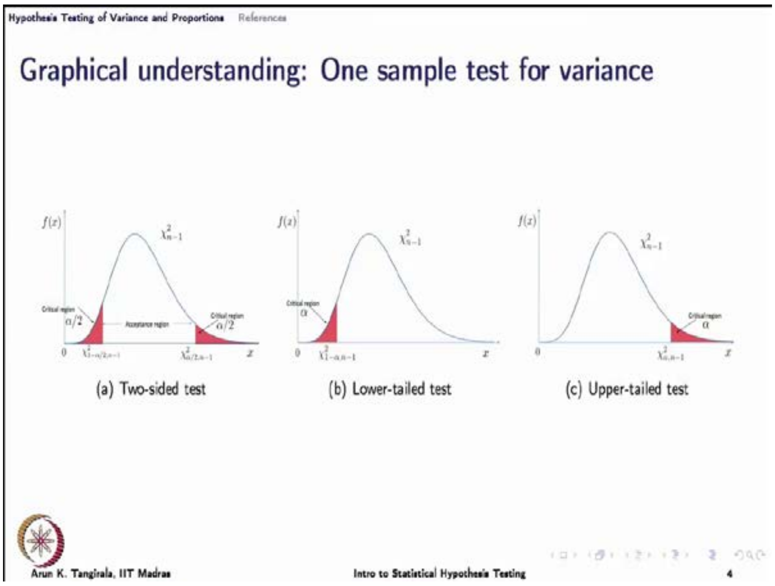
(Refer Slide Time: 02:03)



Now, once again before we proceed through this lecture, it may be useful for you to have the lecture on the sampling distributions handy with you, so that you can quickly look up the distribution of the test statistics that we are going to use in each of these cases. Now we begin with the one sample test for variance the goal is to test for the variability of a normal population, under the assumption of random sample. So, we have data coming from a Gaussian distribution and as usual we assume random sample. And we have the null as per the test; the null is sigma square is equal to a postulated value. And the alternative hypothesis is one of the three possibilities. The test statistics as we know earlier is this normalized sample variance sample variance is S square based on the n minus 1 in the denominator and the normalized statistics is n minus 1 S square by sigma naught square.

This normalized statistics as we know follows a chi square distribution with n minus 1 degrees of freedom. But the main thing to remember is the big assumption that we are making that the data come from a normal population; otherwise this sampling distribution for the normalized statistic or the test statistic that we have C square is not necessarily chi square that is very important point to remember. And then, the critical region is as usual the three possibilities depending on the three corresponding alternative hypothesis that we have. And once again, we have the alpha determining the critical value which is alpha is nothing but your type I error. And you can use in order to conduct this one sample test for variance, there is a no built in routine in R, in the sense in the base package of R let may be more specific. However, they are exists a package called env stats environmental statistics essentially, env stats package that you can install which carries this var test routine, and which helps you conduct this one-sample test for variance. Otherwise, of course, you can always write a small script that will compute the test statistics from the data and compare it with the critical value. It is no big deal actually to write a small script.

(Refer Slide Time: 04:44)

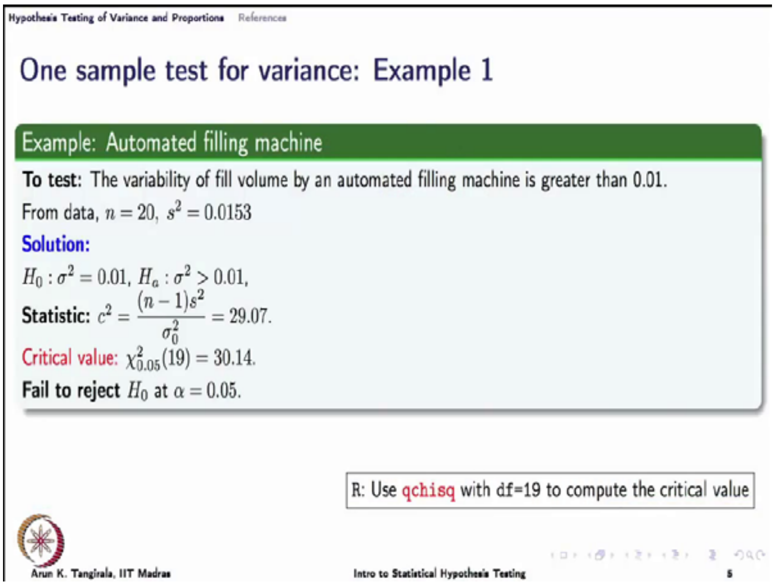


Before we proceed with an example, it may be useful to go through this graphically illustration for this one sample test for variance. As we have just said and we have been discussing throughout this course, there are three possibilities for the alternative hypothesis. The difference that is the type the sigma square is not equal to the postulated value, which leads to the two-sided test and then sigma square being less than the postulated value leading to the lower-tail test and then you have the upper-tail test for the sigma square greater than the postulated value as the alternative hypothesis. As before that as we did in the case of testing of means, we have shaded the critical region in red, but the difference that you see between the case of testing of means and testing of variance is that now the PDF of the test statistic is not symmetric.

As a result, particularly for the two-sided test case, one has to calculate the two critical values at the left and the right critical values separately. For the case of sample mean, we did not have done this because the sample mean has a Gaussian distribution and that is the main purpose of this illustration, but the rest of the application and the concepts remain the same. And there is a small thing that we should notice we have denoted chi square with the degrees of freedom indicated in appearing the subscript whereas in our lectures we have indicated the degrees of freedom in the parenthesis, but that is the minor notational difference that hopefully you should be able to live with.

So, in the two-sided test, what is important is of course, the alpha that is the type of error is distributed symmetrically to the left and right side of the postulated value that does not change, so that is a point to remember. What we are saying is it is equally likely that the observed statistics may fall to the left or to the right of the postulated value, but which is also the case we remains, but what is important is because of the nature of the chi square distribution, the critical value or the quantiles that you are going to calculate have to be calculated separately. Of course, you can use the qchisq routine in R with the specification of alpha appropriate quantile. In this case, if you are going to compute the left critical value you are going to specify alpha by 2, because that is what the qchisq expects for the probability and with the appropriate degrees of freedom, which is n minus 1 that is all to it, but otherwise the rest of the story remains the same.

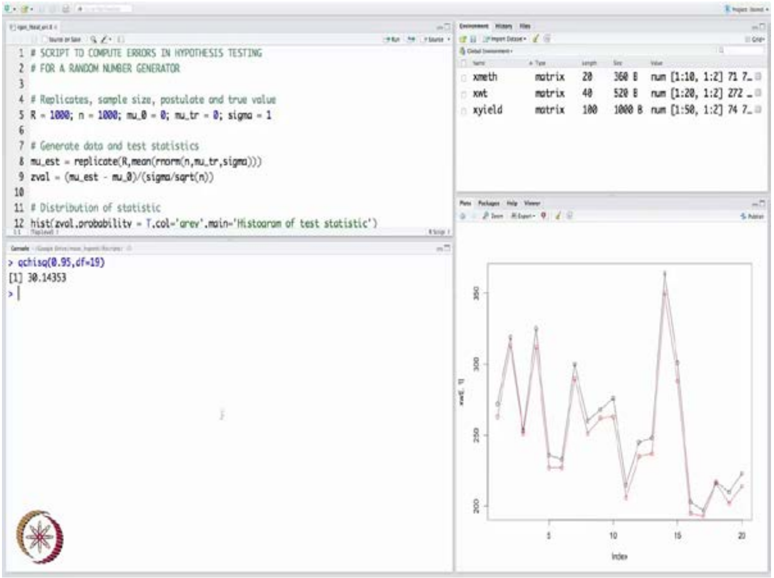
(Refer Slide Time: 07:38)



Let us now go through two quick examples; one is the example from our motivating lecture on the automated filling machine, where we were interested in testing the variability of the fill volume. There is a filling machine that fillings bottles and it is a fact that there is going to be variability in the volume that of the liquid being filled in the bottle, but we want this variability to be really low. And the filling machine is acceptable if the variability is less than 0.01. So, if as an end user I may say that I would like the variability to be not less than 0.01 or I would like to test for variability being greater than 0.01. If the null hypothesis is rejected then that would mean that I would reject also this automated filling machine meeting my requirements.

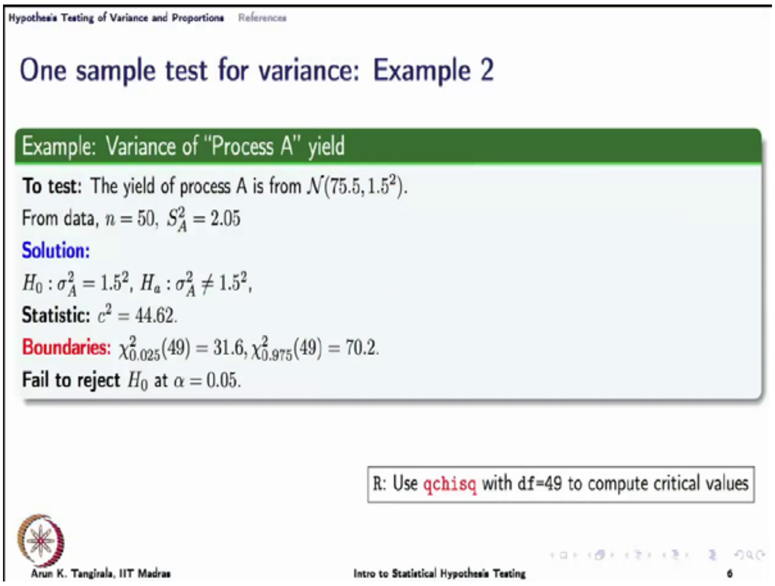
So, what we do is we collect twenty bottles at random that are filled by this automated machine this is an example from the book by Montgomery and Runger, where the directly the variability that is the sample variance is reported which is reported as 0.0153, and now the null hypothesis is as expected sigma square 0.01 and the alternate hypothesis is that sigma square is greater than 0.01. So, we compute the test statistic which is n minus 1 times square by sigma naught square, we know all of these values and status test statistic works out to be 29.07. The critical value now remember this is a one-sided test, in fact, this is a upper-tail test and therefore, we can use qchisq to compute the critical value or you can look up a statistical table.

(Refer Slide Time: 09:41)



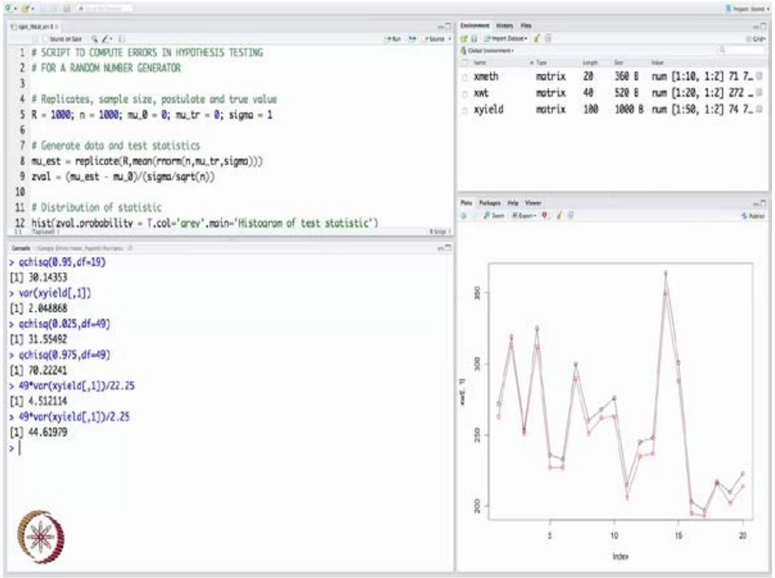
Let see how to do this in R that is how to compute the critical value. So, what we are looking at is qchisq, and remember, this is an upper-tail test, therefore the probability that I am looking at to be fed to qchisq is 1 minus alpha, because it is an upper-tail test, and alpha is 0.05. Therefore, I should ask for 0.95 that is the quantile correspond to this probability with degrees of freedom being 19, because I have twenty observations in the sample and that is a critical value for you which is what I have reported here 39.14. Now that it is, so we have the test statistic following to the left of the critical value that that is the extreme value that we are willing to tolerate if the null hypothesis is true for the test statistics is 39.14. Whereas the observed statistics is not as extreme as we are willing to tolerate, therefore we fail to reject the null hypothesis, very simple. In other words, we do say that there is not enough evidence to believe that the variability is greater than 0.1 you can say so or essentially you do say that fine this automated filling machine does meet my requirements.

(Refer Slide Time: 11:26)



Let us move on now to the second example, which is got to do with the yield data that we discussed earlier in the hypothesis testing of means. Now, we are looking at the yield of this process A, and asking if the variability of the yield. Remember, there is a process, it could be reactor or it could be even an agricultural methodology where you are looking at crop yield, but in this case, this is a chemical reactor. And you keeping making same product across batches, what you are interested in knowing is if the variability across batches is a postulated value right. When we where testing for means we assumed that the variability is known that was necessary in that example because we wanted to work to such an example. But here now we are saying that I have data and I will estimate the variability and ask if whatever I have assumed to be correct holds or not. Therefore, now the null hypothesis is sigma square A is 1.5 squares that is 2.25 is the postulated variability. We have as before in the previous example that we worked out in mean k testing of means k we have 50 observations in the sample and the estimated variability is 2.05. In fact, we have this yield data with us.

(Refer Slide Time: 12:48)



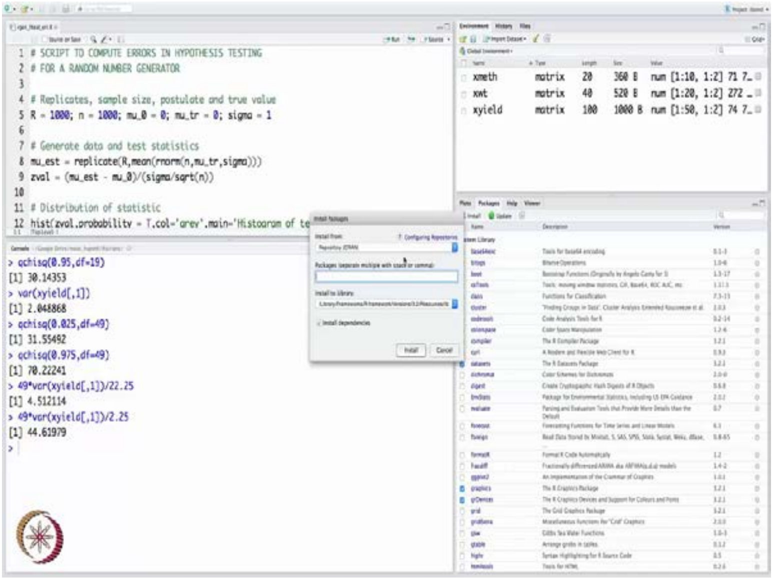
Let us actually go to R and ask so if you look at the environment here, we have this or the workspace in R studio. We have already the variables loaded. So, let us ask actually what is the yield? Variance of the process A which is contained in the first column of x yield, and you get this. What I have reported is a rounded off to the second decimal; this is the estimated variance all right, but you should check if this variance var routine in r does it use n minus 1 or n, you have to be careful. Because you have small samples here, I leave it to you whether it is going to use n minus 1 or n; let say we assume it uses n minus 1 for the sake of discussion. Now what we are asking that is that the postulated value is 2.25, whether I can treat this observed statistic as good as the postulated value given that I have 20 observations in my sample, alright.

Now, what we can do is of course, you can compute the critical value this is a two-tailed test you can say or two-sided test, and now the type one error is distributed equally to the left of that and right of the postulated value. And I need now remember we said for a two-tail of variance, I need to compute the left and right critical values individually. So, let us compute the left and right critical values individually and then we will also see how to do all of these using the var test routine from the env stats package. So, in order to compute the left critical value, once again we use the q chi square. And now we supply 0.025 as a probability, because it is alpha by 2 on the left hand side, which is the probability of making a type one error and degrees of freedom here is 49. So, this is the left critical value. And then to determine the right critical value is apply here 0.975, this now constitutes the acceptable region thirty one point something and seventy point something is our acceptable region, which is what I have reported here in the slide for you 31.6, and 70 point I just rounded it off to the first decimal.

Now the question is whether the statistic falls within the acceptable region. We only computed the variance; we did not compute the test statistic. We can go back and compute the test statistic, and see if it agrees with what I have shown on the slide. So, what we shall do is we know that test statistics is n minus 1 times S square by the sigma square postulated value, therefore the test statistic is 49 times variance of the process a divided by 2.25 all right. So, this is a value that one obtains, and that is exactly the value that we have seen, and now it is straightforward to determine the outcome of the hypothesis test. The observed statistics falls within acceptable region, we do not say well within a acceptable region because does not matters as long as it belongs to the acceptable region, we do not reject the null hypothesis, and consequently we fail to reject the null hypothesis that alpha equals 0.05.

One can also use the p value method, and the p value is calculated using the same procedure as we have done for the means. I leave it to calculate the p value, but I will show you how to use the var test routine from the env stats package to conduct this hypothesis test. Because, there is no built in routine in R to compute or the conduct the hypothesis test for one sample test for variance, and that is a reason we turn to this an env stats packages.

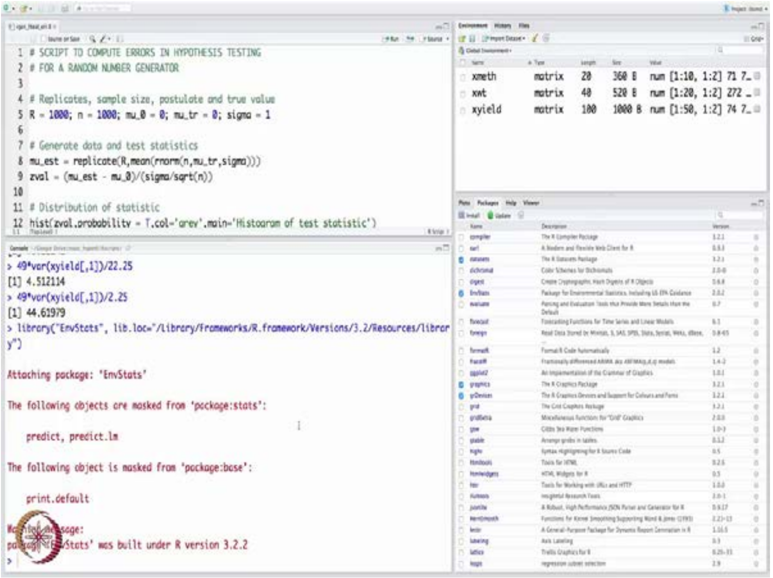
(Refer Slide Time: 17:18)



The env stats package is a user contributor package is does not come with the base installation of R, so one has to manually install it. And installation of user contributed packages in R especially through R studio is a breeze, all you have to do is go to packages and it will show you the list of the packages that have be installed already in your system. And if you want to install new packages you can actually go click on install and all you have to do is point to the repository, there are a number of links that also R studio will bring up in preferences. I have be pointed it out to a repository; there is all these repositories are all mirrored, there is a central repository and the central repository is mirrored worldwide. Therefore, you should not have an issue, you can point one that is a nearest to you or that is more reliable nearest does not necessarily mean reliable all the time.

And once you said the repository perfectly then you can actually type the package that you want. At the moment, I do not have an internet connection; otherwise I would have shown a live demo. However, use type for example, env stats, and it would show you like this right and that is it. So, it will show you the package and then you make sure that you click on install dependencies and you installed the package, all the dependencies would be installed. What do you mean by dependencies; dependencies for this package to run whatever package you are installing it may need a few other packages or may not be if we do not know. So, when you say install dependencies, it automatically looks up all the other packages that have to be installed for this package of interest to run and that is it you are done. In this case, I have done already this exercise and the env stats package has been install, and all I have to do now is go back to the packages and then make sure that it is loaded.

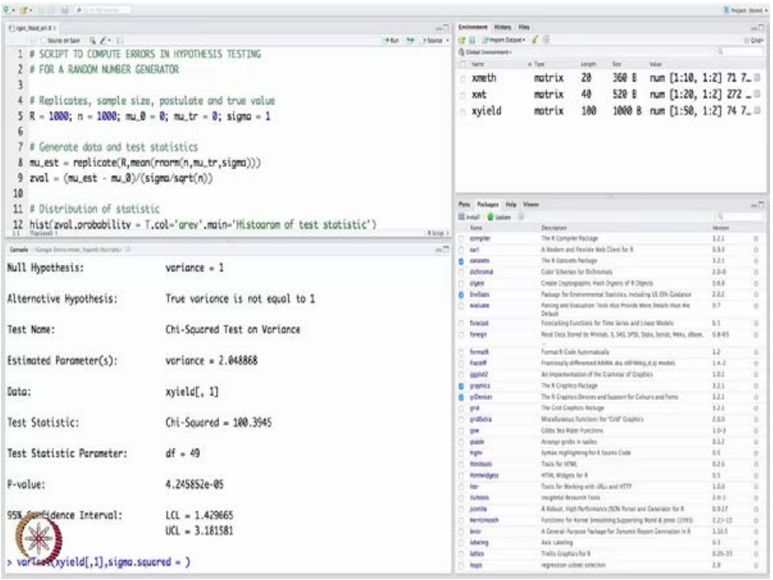
(Refer Slide Time: 19:33)



So, when I just select the package, it automatically loads a package and then also shows; what are the routines that have the same name in other packages. So, what this means is when you install a new package there are going to be routines that share the same name as other routines in other packages, and therefore it would amount to essentially over loading. And it is says basically that these following objects are masked form package stats; for example, if you take predict or predict dot lm; these routines are also present in a stats package. What it says is these are also present in this env stats package, therefore, when you use predict or predict dot lm, then it would use the once from env stats package. If you want to go back to the stats one, you can either unload this package by going back to the package list and deselecting the package or there is another alternative to do it which will not discuss now that something that you should remember. Whenever you load a new package not installed, when you load a package, installation does not do much for you in terms of conflicting of routines, but when you load a particular library then a there is a possibility that you can have a conflict. And therefore, you should read this message carefully to know for yourself as to what packages are going to be in a kind of a conflict.

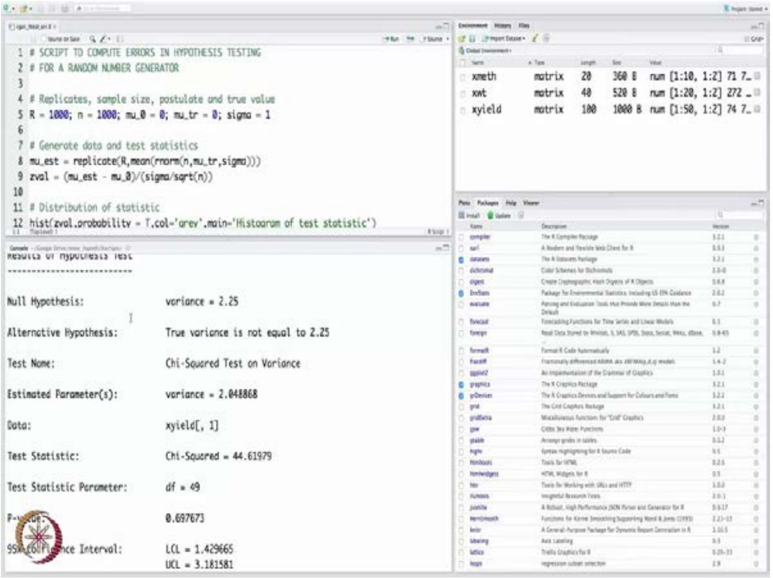
Anyway there is also another package it says routine it says print dot default, which is in the base package of r that also has the same name for a routine in this env stats package and that makes a difference to the way things are printed on the screen perhaps or printed on the plot. And it is only printing on the screen, it is not plot dot default, sorry, therefore, it may make a difference in terms of how things are displayed on the screen things meaning the results of some running some routines, alright.

(Refer Slide Time: 21:48)



So, let us actually now get back to the var test routine in order to conduct the hypothesis test for variance. And if you have loaded nice thing about R studio is it will show you the matching routines. It showed me var test as a matching routine and the syntax is also shown here. All I have to do is supply the data and I am going to do that, and I need to specify the alternative. In this case, the alternative is two-sided which is the default, and significance level has to be specified and that is about it; we do not have to do much. Simply run this var test, and of course, what we need to specify sorry we need to specify the default value, so which is supplied through this sigma dot squared that is essentially we need to specify the postulated value. And it is asking for sigma dot squared that means, it is asking for sigma square naught. And in our example the sigma squared is 2.25, there you go.

(Refer Slide Time: 23:04)



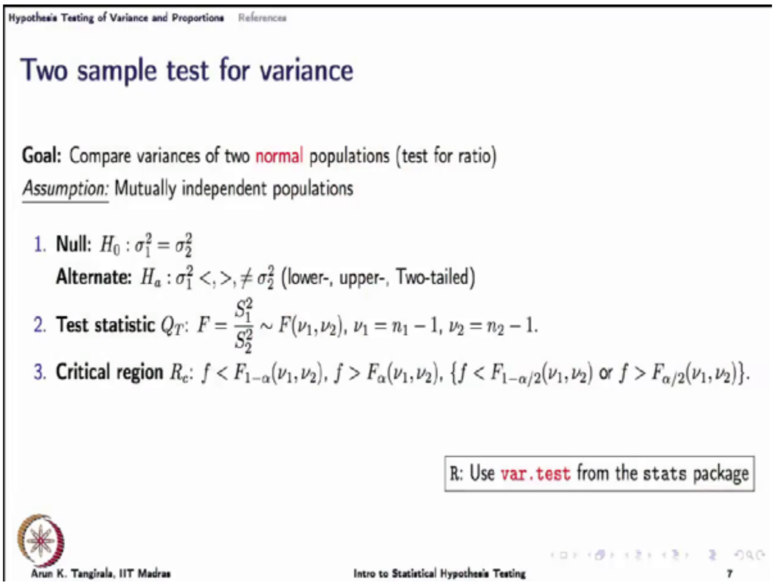
So, now, it reports in a very nice way. In fact, printing of the results that you see in my opinion is much more neatly formatted compare to what you see as a kind of a cluttered output that you get from t dot test or var dot test as I will show you that is the once that are in the base package. Here the authors have taken special efforts to ensure that the results are formatted in a nice way before they have displayed all right. So, the null hypothesis, it is a giving you the alternative hypothesis true variance is not equal 2 and the name of the test is a chi square test estimated parameters variance is 2.05 that is this is not variance actually this is your test, sorry this is the variability, but it is also going to also report the test statistic for you, sorry about that. So, the estimated parameter is what we have estimated earlier using the var routine, and the data of course, that you are supplied it is reporting. The test statistic is what exactly we have computed earlier degrees of freedom is 49. And the p-value is computed for us, and the p-value is quite high when we know that the p-value quite high the interpretation is what I have observed is not more extreme than that I can tolerate.

If the p value was low, what we mean by low is lower than alpha 0.05, then we have hit upon test statistic realization or a sample that is showing an extreme value more extreme than what I am willing to tolerate. In this case the p-value is high, therefore, we do not reject the null hypothesis. And finally, you also see the confidence intervals here LCL stands for lower confidence you can say bound upper confidence bound here; lower confidence level or upper confidence level as you want to think. And the interpretation of this confidence interval is essentially the same as that we had in the testing of means case. This is the two-sided test, this is a 95 percent confidence interval because alpha is 0.05; what this means is the true variability is likely well is going to be in this region spanned by LCL and UCL values with a 95 percent confidence; we do not say 95 percent probability because the truth is a deterministic quantity. So, we say that from the random sample I have constructed what is known as an interval estimate for the truth and I am 95 percent confident; that means, there are 5 percent there is 5 percent chance that the truth can reside outside this interval; that means, I would have missed the true - the capturing the truth.

And as I had explained in the testing of means case, the way to use this confidence region for hypothesis testing is in the two-sided test for instance is to see if the postulated value false within this interval. If it does then that postulated value is also a likely truth and therefore, there is we do not reject the null hypothesis or we say we do not have enough evidence to reject the null hypothesis. At the truth fallen outside this or this interval then we would have rejected the null hypothesis, and that would have shown up in your p-value that would have also shown up in your critical values. Everywhere the results would be consistent. In essence, we have been learning three different ways of conducting the hypothesis test; one using the critical value method, other using the p-value method, and the third is the using the confidence interval method.

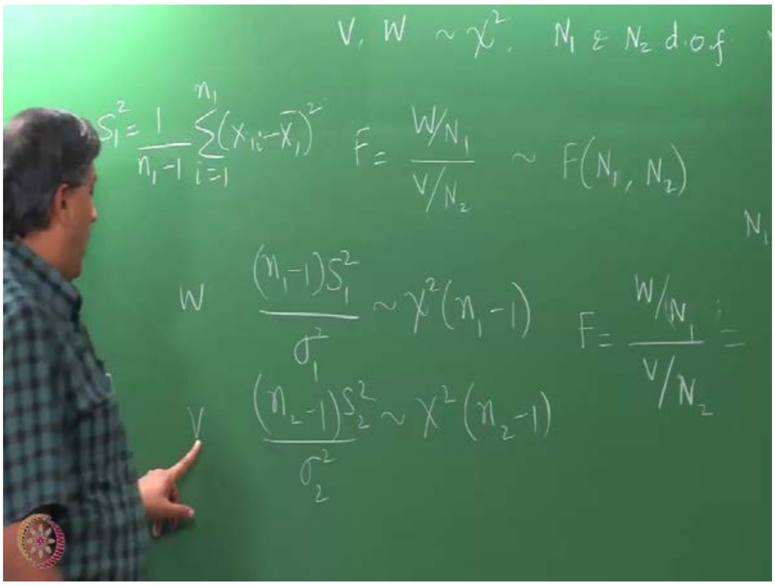
And as I mentioned earlier, we will know the technicalities of how to construct a confidence interval what is the meaning of a confidence interval in the following lecture, very good. So, now, we are sure that at least all the three give me the same result. It is a good way of corroborating our results; sometimes we may make an error in reading the numbers. Therefore, it is good to go through all these three and check if all the 3 different ways of testing the hypothesis, give me the same answer. In this case, I fail to reject the null hypothesis.

(Refer Slide Time: 27:37)



So, let us now proceed to the next case, which is the case of two-sample test for variance. Again the story is the same, similar as we had before; we are going to compare variances of two normal populations; that means, populations following a Gaussian distribution, but we are going to test for ratios. And the null is as usual sigma one square equals sigma two square and the alternatives is one of the three possibilities. Of course, there is a big assumption here and I just forgot to mention that which is that the population that we are comparing at mutually independent like the once that these the same assumption like the one that we made for comparing means. In fact, in the mean comparison, we had also this case of pair test where the populations were not independent, but before that we had discussed for the difference in means as a parameter to compare the means of two populations. There also we are assuming that the populations are independent, alright. So, with this assumption, with these null hypotheses now and the alternative hypothesis determine by the application, the test statistic is now the ratio of the sample variances; and this ratio of sample variances is known to follow F distribution when the null hypothesis is true. Let me give you a brief background on that may be a minute or so, as to why this follows an F distribution.

(Refer Slide Time: 29:06)

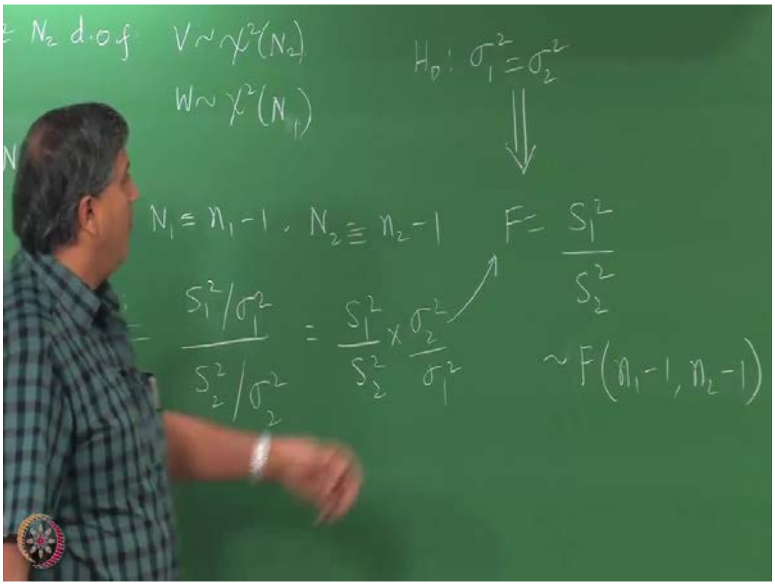


So, quickly going to the theory of a ratio of variables that follow a chi square distribution; from the theory of probability distribution, we can say that if I have 2 variables w chi square distribution. Let say I have 2 variables V and W have a chi square distribution each with let us say sum N 1 and N 2 degrees of freedom respectively. In other words, to be more specific say V has a chi square distribution with N 1 degrees of freedom, and W also has a chi square distribution with N 2 degrees of freedom. When I have a set of two random variables such as these then the ratio of W over N 2 and V over N 1 this ratio let say define this by F follows an F distribution with N 2 comma N 1 degrees of freedom. What we mean by again degrees of freedom here is that the numerator contributes to the randomness enough the denominator also contributes to the randomness enough. So, you have two different dimensions itself from where the variability of through of F can be affected; and in each dimension, you have this many degrees of freedom N 2 and N 1.

Now, how is this related to what we use in the ratio of variances? Well, we know that when I have samples drawn from Gaussian population and I compute the sample variance. Let us say I have a random sample drawn from a Gaussian population, let say from population 1 where this is assume that this random sample contains N 1 observations then this statistic N 1 minus 1 times S 1 square over sigma 1 square has a chi square distribution with N 1 minus 1 degrees of freedom. May be to keep things consistent we could also change here N 1 N 2 to N 1, it is just a notational thing, so that you can view more comfortable with what we are going to do next all right. So, we know this already, but N 1 minus 1 times S 1 square or sigma 1 square has a chi square distribution with N 1 minus 1 degrees of freedom where S 1 square is our sample variance with the N 1 minus 1 in the denominator or a divisor it is an unbiased estimator of the variable variance of X 1, which is sigma 1 square, alright. Likewise, we have for the second population N 2 minus 1 S 2 square by sigma 2 square following a chi square distribution with N 2 minus 1 degrees of freedom. Now of course, what you have to remember is this results is to assuming that one that the population the respective population follow Gaussian distribution; and two, that you have obtained a random sample; only under this 2 assumptions this is right.

Now, this fits in very nicely here what I want to do is, I want to compare the variances; one way of comparison is ratio, other way of comparison is different as you done in the means, but this results we know already, therefore, we would like to exploit this result. Here it says if W has a chi square distribution with N 1 degrees of freedom then W over N 1 likewise over v over N 2 follows an F distribution. Now, all I have to do is think of this as W and think of this as V, alright.

(Refer Slide Time: 33:42)

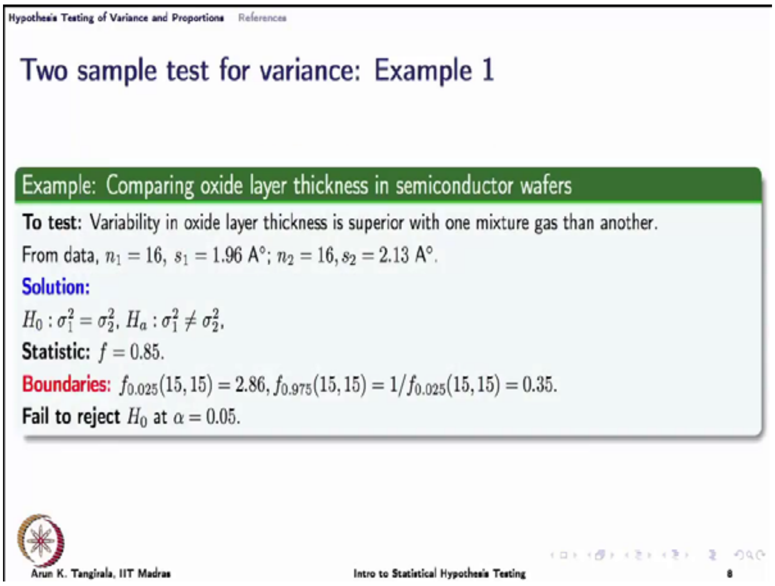


So, then what do I have here, if I were to construct a statistic with W over N 1 over V over N 2 and call that as F then we know W is this n 1 minus 1 times S 1 square or sigma 1 square and therefore, sorry n 1 minus 1 here, because n 1, but we know that N 1 is n 1 minus 1 and N 2 is n 2 minus 1 because W has n 1 minus 1 degrees of freedom and V has n 2 minus 1 degrees of freedom sorry for the confusion, but these are lower cases n's and these are upper cases N's that is all we have to keep in mind that is it. So, W over N 1 would be W over n 1 minus 1 and that would be S 1 square over sigma 1 square likewise you would have S 2 square over sigma 2 square. Therefore, the F is actually S 1 square over S 2 square times sigma 2 square over sigma 1 square then why do we have only S 1 square over S 2 square as the test statistic.

Well, do you have an answer to that? Well, the answer to it is pretty straightforward; we are always testing conducting a hypothesis test assuming the null hypothesis to be true. So, when the null hypothesis is true, what is the null hypothesis, the null hypothesis is that h naught is that sigma 1 square equals sigma 2 square. And we conduct hypothesis test anchoring ourselves to this truth, and when this is the case then what we have is under this assumption, if this is true then this F becomes S 1 square over S 2 square. This has now F distribution with small n 1 minus 1 comma small n 2 minus 1 degrees of freedom that is a story behind using this F statistic for comparing variances. Hope now it is clear as to why we use the statistic, but you should always remember that this is the original result. And because the null hypothesis of equality type, we have F equals S 1 square over S 2 square coming up as a test statistic. Imagine if the null hypothesis was not of the equality type was of the inequality type, then it would have been difficult for us to use this result because then the F value the test statistic becomes pretty complicate. So, once again we do realize that setting the null hypothesis always to be of equality type simplifies the way we conduct or hypothesis test, alright.

So, getting back to now two-sample test, test a generic procedure, construct this test statistic as S 1 square or S 2 square under the null hypothesis is being true, the ratio that is this F S 1 square or S 2 square follows an F distribution with n 1 minus 1 and n 2 minus 1 degrees of freedom and then you determine the critical region accordingly. Once again you can use Q F in R to determine the critical values of course, you have to specify there two parameters which are the degrees of freedom mu 1 and mu 2.

(Refer Slide Time: 37:40)

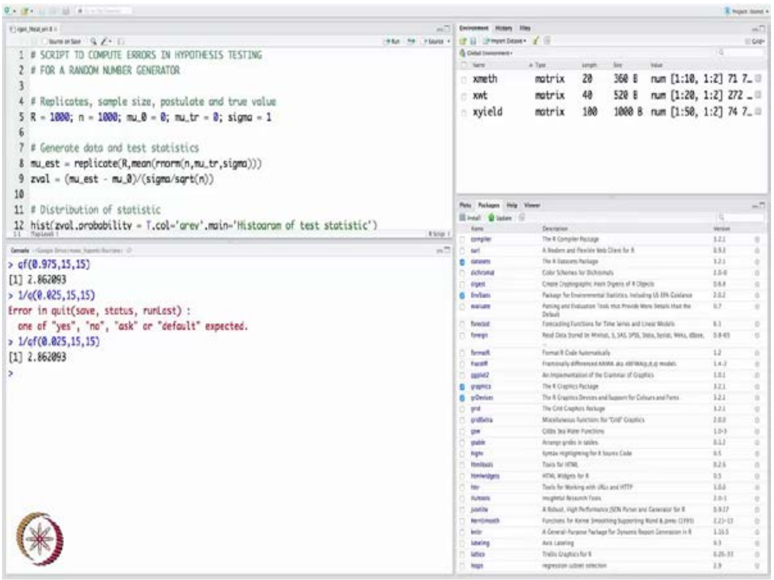


So, let us go through 3 examples to understand how these two-sample test for variability are conducted. The first one again is from our motivation lecture, where we want to compare the variability in the oxide layer thickness of a semiconductor wafer when I use two different mixtures of gases you may recall that we are discussed this example. So, what we have done is or the manufacturer has done is that has used 2 different gases mixture of gases for the environment when the oxide was a layers were being formed. And then about 16 of specimens from each of the scenarios were drawn randomly, and the variability was estimated. And, I have just taken this value from the book by Montgomery and Runger; the S 1 and S 2 are the estimated standard deviations they are not estimated variances. Both have the same sample size 16 specimens each and now our null hypothesis is clearly sigma 1 square equal sigma 2 square and alternative hypothesis is it is not. And the statistic works out to be 0.85, if you put through the number that is your S 1 square over S 2 square.

And the boundaries are computed using your Q f or you can look up a table. Now remember alpha is 0.05 this is a two-tailed test therefore, I am going to look up the quantiles to the left and the right of the postulated value in the F distribution. The F distribution or the density function is an asymmetric one pretty much looks like the chi square one you can think of f has been the counter part for chi square as the T is been the counterpart for the Gaussian. Remember when we go from Gaussian to T, we say more or less t looks like a Gaussian, it is symmetric, but the shape is affected by the so called degrees of freedom there. Likewise here, when we move from chi square to F distribution, we have an additional dimension, therefore, we have two degrees of freedom and the F looks pretty much similar to chi square except that it is shape is influenced by both degrees of freedom compared to your chi square. Otherwise you know once you have understood that analogy it is fairly simple to work with.

So, now you can either look up a table or you use Q F to determine these critical values. The nice thing about an F distribution is if when you are working a let say for with a two-sided test specifically the right critical value shares a relation with left critical value in that. For example, here the f that you wanted to compute that is a critical value that you wanted to compute to the right is computed by specifying the alpha which is 0.975 not alpha, but 1 minus alpha by 2. You would specify that in your Q f and also feed the degrees of freedom this is equal to 1 over what you have computed of the left critical value. So, look there is a complimentary here, you want to compute Q f.

(Refer Slide Time: 41:11)

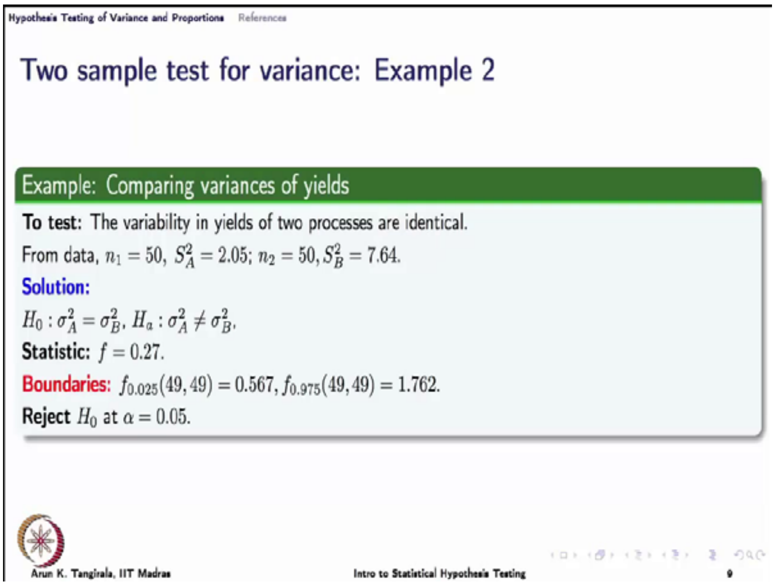


Let me show you an R, you want to compute here q f the probability being 0.975, and the degrees of freedom being 15 comma 5 this is the value that you get is a right critical value which is equal to let us actually say one over q 0.025 comma 15 comma 15 sorry q f. So, the values are the same right. What this means is with the f distribution things are easier then we what we thought so before because it is a complicated one by the way the f distribution does have an analytical expression that looks pretty intermediating, we do not need that at least when we are using the software or when we are looking at tables. So, what we have seen here is I just need to know one of the bounds and the other bound is already known which is nice all right.

So, good, now we have the left and the right critical bounds. Here of course, we have 0.35 and 2.86 so that is left critical value and the right critical value. Remember both chi square and f distributed variables are always non-negative; that means, the range of values at the random variable whether it has f distribution or a chi square distribution always runs from 0 to infinity. There is known the negative part of the real axis is excluded unlike in the Gaussian or the T distribution cases, very good. Now, we ask if the observed statistic false within the acceptable region and it does, and therefore, we fail to reject the null hypothesis. The observed statistic is 0.85, and what we have here is the bounds being 0.35 and 2.86.

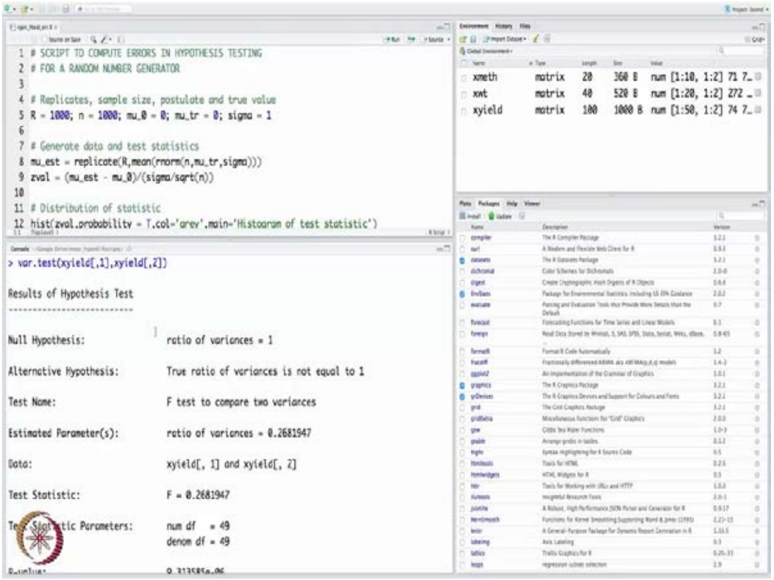
Now there may be come confusion here as to what is a left and what is a right by the notation that I have used here. Do not confuse this f subscript notation that I have used with the values that we given q f; we usually in q f we gave the complimentary probability, and therefore, you have to understand this 2.86 to be the right critical value, obviously, and 0.35 to be the left critical value. So, do not worry about that fortunately for us, one is the reverse of the reciprocal of the other good.

(Refer Slide Time: 43:57)



So, now, let us move to the second example, where now we are going to compare variability of yields again it is a same follow through example you have used in means also in the one-sample test for variability we have used this a example. Now, we have 2 processes as we had in the case of comparison of means there we compared means now we are going to compare variability. Remember, the nice thing about comparing variability is I do not need to know the knowledge of the true means that is quite important to know. The variability now in the yields of two processes are identical is what we want to test; in other words that they are not identical you can say so. Therefore, the alternative hypothesis is that the variability's are not identical or not equal. We can go to now where dot test routine which comes to the base package in stats and run through this example, then come back and see if you get the same numbers.

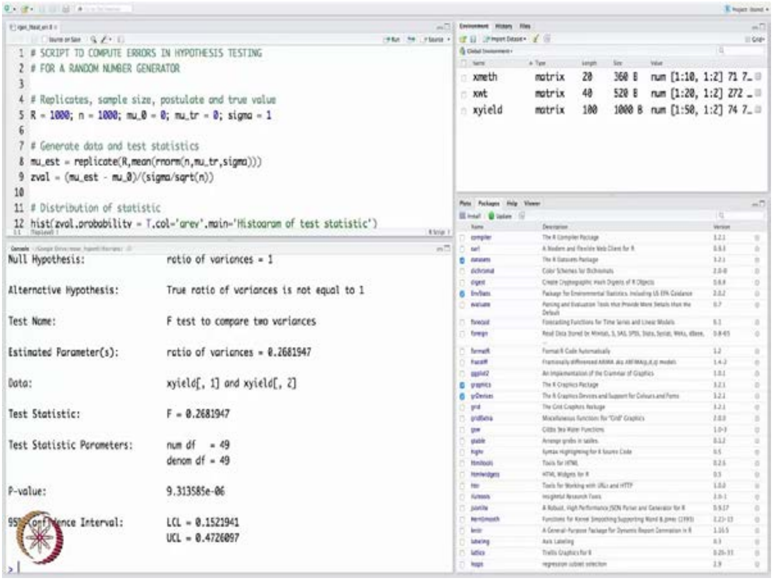
(Refer Slide Time: 45:00)



So, let us go back to R, clear the screen and then now we use the var dot test and supply the data for the 2 reactors that is the yields from 2 reactors. And now, we have to supply the postulated value that is the ratio we are saying that the ratio should be all to one that is what we are saying and that is our null hypothesis and the alternative is that it does not. So, it is a two-sided test and that is about it we do not have to specify anything else. You can look up here other options that are available, so ratio is that of the postulated value here we are saying it is equal. In many application, we want this a probably test that the variability in one process is twice the variability in the other for example, or some gamma times or k times the variability in the other one, in which case we have to specify what is that ratio. Here the ratio that we are testing for is one therefore, I do not have to specify, it is a default value alternative is two-sided conf dot level is our alpha, which is 0.05. So, I do not have to specify anything here.

Let us look at the output of the var dot test. Now, let me point out here because we have loaded the n stats package the output has been formatted in the nice way for us had we not loaded that then the output would have been cluttered then the numbers would not change, but it is just the ease with which we can read the output is much better. Now, the null hypothesis is that the ratio of variances is one alternative hypothesis is it is not and the test name is that we are using an F test and estimated parameter is a just ratio of variances.

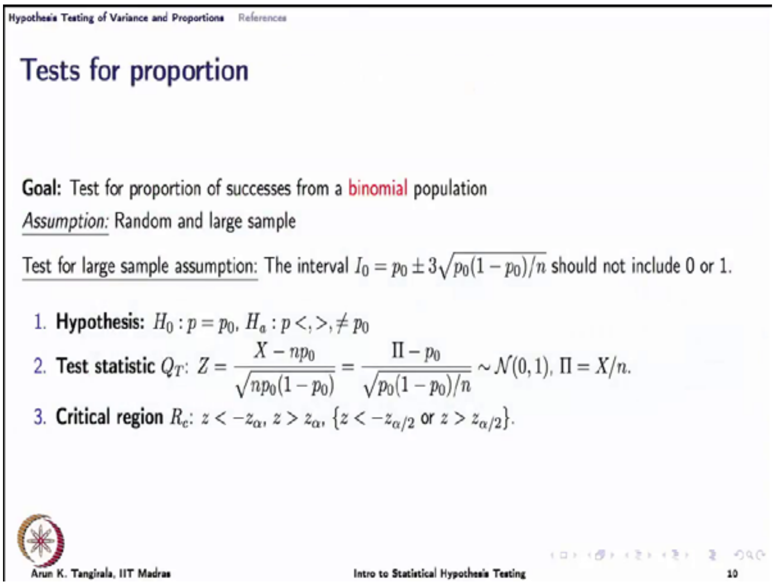
(Refer Slide Time: 46:51)



You can check cross check what you can do is, you can compute manually variance of the first column of x yield by the variance of the second column of x yield, it should give you the same value. And the test statistic works out to be this, value here, let see if that is what is reported in the slide for us yeah it is roughly rounded off to the second decimal 0.27, very good. And then, we have here numerator and denominator degrees of freedom 49 and the p-value is reported. Now you observe that the p-value is quite low. We do not really worry about how low it is we only ask if this p-value is lower than alpha or not. It is lower than alpha therefore, what is the decision, the decision is to reject the null hypothesis that means, we do say no, there is no enough evidence to believe that these two reactors or two processes give me the same variability; one has a lower variability than the other. Of course, then one has to ask which one a as the lower variability. So, I leave that you to conduct that hypothesis test; suppose you postulate that process a, a yields with lower variability then that process b then go ahead and formulate your null hypothesis of course, that is always a, but your alternative hypothesis and check if that alternative hypothesis is rejected or not.

Now, we can also use a confidence interval way of testing this hypothesis. The lower and the upper confidence limits are given for us. And it does not contain the postulated value which is one, which means that there is a 95 percent there is a way chance that the truth is not 1, you can say so. Well that is a that is a statistically it is not a correct way of stating that let me put it in a different way; there is not enough evidence at least with 95 percent confidence I can say that there is not enough evidence to believe that the ratio variances is one that is a more statistically appropriate way of stating the outcome of this results. All 3 therefore, now tell us, in fact, we have not computed the lower and upper bounds you can do that or I have given that in the slide for you. The lower and upper bounds are given and for the critical values and the observed statistic falls outside this acceptable region, therefore, we reject the null hypothesis. Again all the three point to the same result which means we are doing things correctly reject the null hypothesis at alpha equals 0.05. I leave it as a homework exercise for you to see which process yields me lower variability. The data is with you we are gone to post it any ways. So, you can play around with the data.

(Refer Slide Time: 50:06)

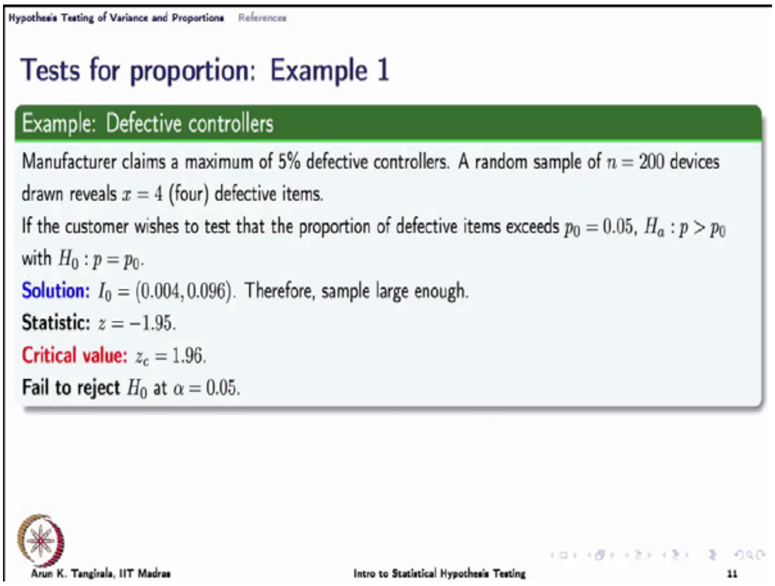


So, the final case that we have is testing for proportion again one-sample and two-sample test. This is the one sample test for proportion of success from a binomial population and assumption is that we have a random sample as usual and that the sample size is large. The sample size large assumption can be tested based on the sample size, which is n and the postulated probability. If this interval p naught plus or minus 3 times root p naught times one minus p naught by n includes 0 or 1 then you can say that the large sample assumption is violate. Where does this comes from, well, it comes from a confidence interval kind of construction we will not go into the details, but this the practically speaking what you should do is in order to use this test statistic that we have given here, first you should always check whether this interval contains 0 or 1. The basic idea if you recall from the sampling distribution is we are using an approximate result here the approximate result is that the z statistic that we have return here which is x minus n p naught by square root of n p naught times 1 minus p naught follows an approximate Gaussian distribution.

When is that approximation true? Well, when n is large and also that the probabilities are not at the extreme values; extreme values meaning either not too low or not too high whether it is for a success or a failure it does not matter. If the probability of success is too low then the probability of failure is going to be too high because only 2 possibilities. So, it does not matter whether where these probability of success or failure, you should not be work having an extreme scenario there. When that this a case and the sample size is large, approximately z follows a Gaussian distribution, under this approximation only we are going to test this one sample proportion case. For a more detail test of proportions and so on your advice to look up the literature, this is very short course and hypothesis testing, therefore, we do not have the time to discuss further scenarios and discuss non parametric test and so on, alright.

So, let us proceed now the critical region is as usual you use the normal distribution determine the critical values and or the acceptable region based on the specification of type I error and the alternate hypothesis.

(Refer Slide Time: 52:53)

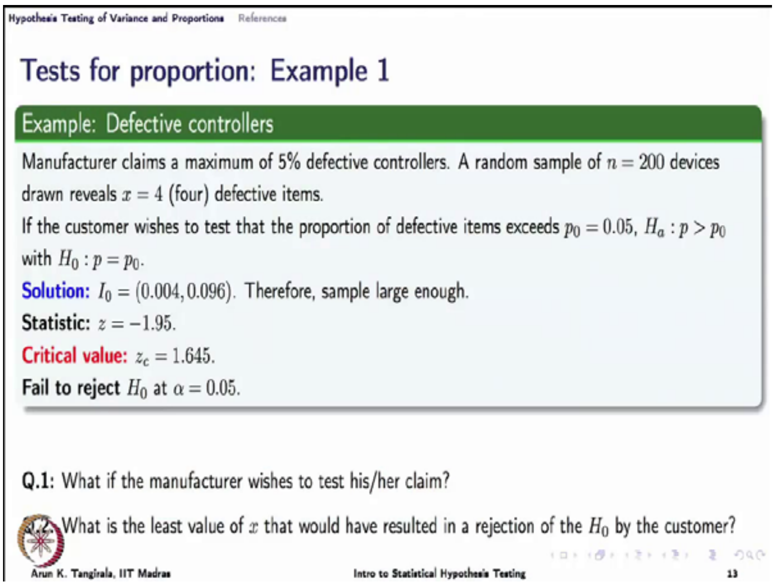


Let us look at an example now - first example. Again this is from our motivation lecture; manufacturer claims a maximum of 5 percent of defective controllers, and random sample of 200 devices are drawn, from where 4 items have found to be defective. Now if the customer wishes to test that the proportion of defective items exceeds 0.05 the alternate hypothesis is that p is greater than 0.05 that is the postulated value or the value that is put forth by the manufacturer. Before we use the previous method remember we have to construct that interval and the construct the interval n in this case is 200, p naught is 0.05, we do not know if it is too small let us not think or make any judgments if 0 or 1 appears in this interval. It is kind of an indication that the probability of proportion or the proportion that we want to test is way too low for this method to be used. Now, we have computed the confidence interval and it turns out that it does not include 0 or 1. Therefore, we can say that the sample is large enough and also that the postulated proportion is not at the extreme value.

On the other hand if I had postulated a value of 0.005 that is the manufacturer had said that there are only 0.5 percent of defective controllers. And I want to use this test you will find out that with this sample size, it is not possible to test that because the confidence interval will include 0 that you can check, you can just go back here and calculate the confidence interval with n equals 200 and p naught equals 0.005. In this case with for this sample size this proportion is not extreme. So, you see what we call as extreme proportion is a relative think, it is relative to the sample size that you have that is very important to remember, good. So, now it allows us to use this method; of course, we would use an example as academician we are always notorious for that we will use examples that allows us to look at simple cases, but this a 10 hours course you are going to work with simple cases so as to convey the idea, alright.

So, getting back to the problem here the observed statistic works out to be minus 1.95. So, here how do you calculate the observed statistic x is 4, n is 200, p naught is 0.05 or you can say you can use the other one the pi that you have is a estimated proportion which is 4 over 200, alright. And then you have a minus p naught there, p naught is a 0.05. So, either way, you can compute your z statistic and it turns out to be minus 1.95 and this is a one-sided test. What kind of a one-sided test, this is an upper tail test and therefore, we compute only one critical value which is to the right of the postulated value. And we have here, we know from you are previous experience at the critical value at alpha 0.05 is it should not actually 1.96 that is a mistake there. So, let us get back now the observed statistic is minus 1.95, as we calculate using this expression here x is 4 and n is 200, p naught is 0.05 or you can use the other expression here which as the pi unit and pi is 4 over 200 and p naught is 0.05.

(Refer Slide Time: 56:51)

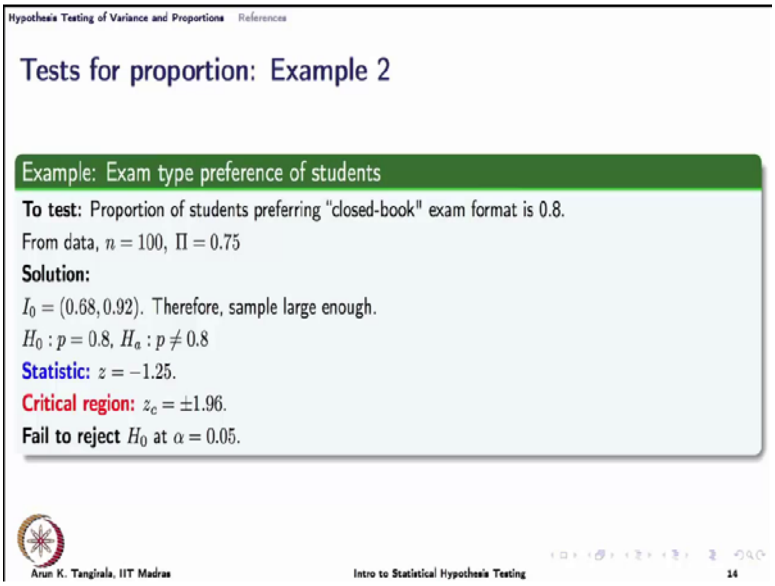


Either way you will get the same value being minus 1.95, and the critical value now is 1.645. Remember this is a one-sided, in fact an upper tail test. And therefore, the critical value is going to be on the right-side and we compute this using q nom of 0.95 because alpha is 0.05 and that works out roughly to be 1.645. And therefore, now the statistic is to the left of the extreme value, and therefore, we fail to the reject the null hypothesis in other words the manufacturer is safe. A manufacturer can claim that yes to the maximum possible proportion of defective controllers in his manufacturing is 5 percent, OK.

Now, what you may want to do is ask if what would be the alternative hypothesis, if it is the manufacturer who is conducting this hypothesis test. As a customer have done it what would be the alternative hypothesis would this null hypothesis be still hold or still would it be rejected. And the other question that may be you want to look at is, what is the least value of x that is what happened is we have draw 200 devices randomly from the process and found that 4 items were defective. What could have been the least value; that means, you can say minimum value of defective items that could have resulted in a rejection of the null hypothesis; obviously, that has to be greater than four, correct; so, as to push the observed statistic to the right of the critical value which is 1.645. How many defective items in other words would we have tolerated in this sample of 200 and still not reject null hypothesis, something to think about.

Of course, there are other question that you could ask what could have been the smallest sample size that could have rejected let the rejection of null hypothesis for example, So, there are so many question that one could ask not only in this hypothesis test, but also other hypothesis test. We will consolidate all or most of these questions in the last lecture, where we will ask what are the things that any hypothesis test is sensitive to. At the moment, we are only learning how to conduct a hypothesis test, alright.

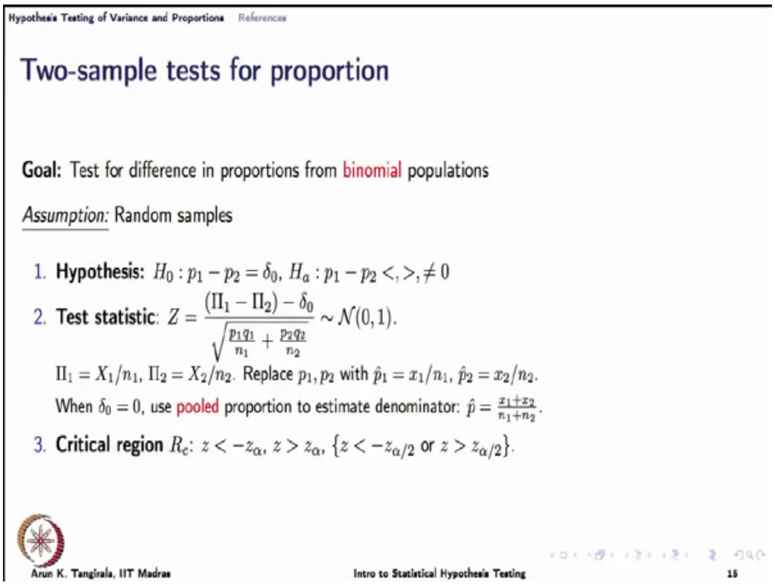
(Refer Slide Time: 59:34)



So, let us now move to the second example, where we are looking at an exam type preference of students and the what we want to test is that the proportion of students preferring a close book exam format is 0.8. So, here we are not asking that more than 80 percent are preferring or less than 80 percent are preferring, we just want to say test the hypothesis test that 80 percent of the students prefer close booking exam on a campus. So, now again I have taken this example from the book by Ogunnaike and from the data you can actually compute the proportion it turns out to be 0.75, the sample size is 100. Once again we do not know if the sample size is large or small and so on for the postulated proportion. So, we construct that interval and once again this interval does not include as 0 or 1; therefore, we say that the sample size is large enough for the postulated probability or proportion.

And now, the alternative hypothesis as we are mentioned before is that the p is not equal to 0.8. The statistic works out to be minus 1.25 using the same formula that we are used before. The critical region here once again should have been one point yeah now the critical region is 1.96 because now we are looking at a two-sided test, right and the observed statistic falls within the acceptable region, acceptable region is minus 1.96 to plus 1.96. And therefore, we fail to reject the null hypothesis. Here also one can follow the p-value approach, one can use the confidence interval approach, but I leave it as an exercise for you to work out those approaches. Now of the final case of discussion for discussion this lecture is on concerned with the two-sample test for proportions.

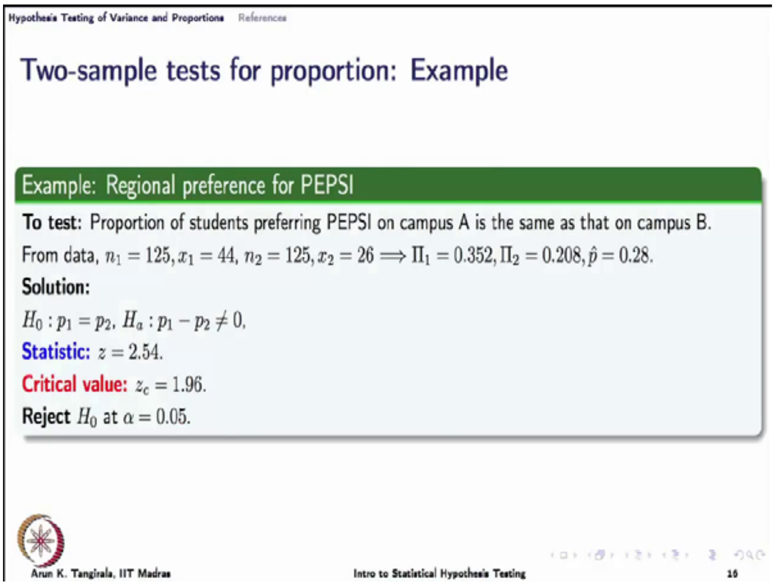
(Refer Slide Time: 61:35)



So, what we want to do is test for differences in proportion from binomial populations, again very specifically we are saying binomial. And assumptions is that we have samples drawn randomly from the respective populations, and as usual by now you should be familiar with the procedure the test statistic is your standardized statistic approximately following a Gaussian distribution. Assuming that once again the sample size is large and as I have pointed out in the lecture on sampling distribution, to compute the denominator of z, you need the values of p 1 and p 2 which we do not know obviously, we have only postulated. So, what we do is we replace this p 1 and p 2 with the respective estimated values ok and that is it. And what we can additionally do is when we believe that both populations have the same proportion that is the postulated value is 0 for the difference in proportion, we can use a pooled proportion idea to estimate the denominator which is for p hat.

So, we need this estimate for 2 things to compute the numerator which is pi 1 minus pi 2 minus delta naught and in the denominator. So, there are 2 scenarios like we had in the comparison of means unequal variances, equal variances. When we had equal variability there we said we could use a pool estimator to estimate the variance; and when they are not equal, we will have to live with the individual estimates the same analogy extends here as well. The critical region as usual that is it.

(Refer Slide Time: 63:27)



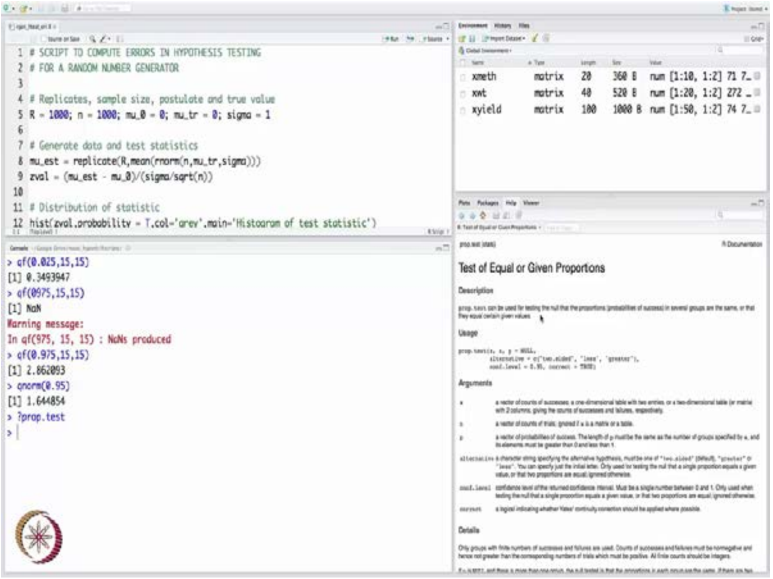
So, let us look at an example and then conclude our lecture. This is the example again from the motivating lecture; in this case, it is a regional preferring for a soft drink called PEPSI. This is then example taken by from the book by Ogunnaike. It is not an advertisement for PEPSI, it is just an example concerned with it. So, what we want to test is that across two campuses, we have the same proportion of students preferring these particular soft drinks. So, 125 individuals from these two or student from these two different campuses were surveyed and asked for their preferences. And from the first campus 44 students preferred this soft drink; whereas from the second campus 26 students prefer these preferences. Clearly, there is a difference you can say, the sample size is the same, you have 44 in one campus and 26 on the other campus. So, there is; obviously, glaring differences never the less it is useful to conduct a hypothesis test on whether this difference is actually significant or not because there is going to be variability that is a reason why we are doing all of these whether it is proportion, variance or mean testing whatever. It is all because of variability uncertainty that I cannot believe just on the phase value of it that this difference being 18 in this case to be considered significant.

So, on the scale of variability that exist among this two campuses we want to ask if this difference of 18 individuals out of 125 preferring this soft drink is to be considered significant or not. So, we compute our pi 1, pi 2 and p hat works out to be 0.28, because we are testing that there is no difference between the proportions we can use the pooled estimator and compute the p and that is how we have computed p that is using x 1 plus x 2 by n 1 plus n 2. X 1 is 44, x 2 is 26, so that works out to be 70, and then you have by 250 right so that is what we have here. So, you get here p hat is 0.28. And the statistics works out to be 2.54, this is a two-sided test. Therefore, the critical of course, should be reported for both sides, but because the observed statistic as turned out to be a positive value I am only reporting the upper bound here. And the observed statistic is more extreme then what I am willing to accept when the null hypothesis is true.

Therefore, I reject the null hypothesis it is most likely the null hypothesis is not correct that is what it means, under the circumstances again as with any hypothesis test. Therefore, we reject the null hypothesis and alpha it could 0.05. Now, once again you can go back and ask well with this data can i ask now which campus has larger proportion of students preferring this soft drink. Well intuitive you can say that it is the first campus that as a larger proportion, but you may want to conduct a hypothesis and check if that hypothesis holds. Of course, the null hypothesis always going to be p 1 equals p 2, the alternate hypothesis going to be either p 1 greater p 2 or p 1 less than p 2 depending on what you suspect to be the campus having a larger proportion of students preferring this soft drink, alright.

So, with this we come to a close, but before I do that let me tell you, you may be wondering why I have not shown these kinds of test in R, why is it not done. Well first of all it is way too simple you do not have to do anything this is a normal normally distributed test statistic I do not have to go to R to use a utility. All I need is a calculated to compute the test statistic, but that is not the main reason, the main reason is the prop dot test, there exist a routine called prop dot test in the base package that can do the same similar kind of test for you.

(Refer Slide Time: 67:52)



So, you can look up the help on test of proportion which allows you to compare across populations, but this is more generic, and also uses what is known as the chi square test this is well known in the literature on test of proportion, where there is a z test that we have seen. But there is also something called a chi square test which is for a generic situation that is maybe you do not have the population following a or the sample size not being large or the test statistic not following a Gaussian distribution. We do not know for example, what distribution they underlying population has and so on. In those cases, you can use this test which is based on the chi square test, test statistic; this was proposed by Pearson. You can still use this for the example that we just discussed except that it works with the squared test statistic instead of the z variant that is the only difference.

So, you are most welcome to use this for any other data for which we want to test the proportion, but remember that it is not using the same test statistic that we are using it is using a squared one. And you can go into the literature to learn; what is the difference between the z-test that we have discussed and the chi square test that is more widely used. The technical term that you will see is the z test corresponds to what is known as the parametric test where as the chi square test corresponds to what is known as the non parametric test and so on. But given the duration of this course, we do not have the luxury of discussing more of this related concept in detail; therefore, hopefully if an advance course is offered will discuss the differences and all the pertaining details. Else please go into the literature, take any standard test book on hypothesis testing even the book by Montgomery or you can this there so many resources on the web, go and check and read as to what is the difference between the chi square test and z-test. For this course, we will restrict over self only to the z-test and we will also evaluate in the examinations only questions based on the z-test on right, alright.

So, now with that we come to a close of this lecture.

(Refer Slide Time: 70:28)

And now, I am sure you have seen a lot of different hypothesis test, and there is high probability that you may be in a confused state that you can work out more examples and get clarity. And of course, listen to this lecture probably more than once. The assignment problems will definitely help you in reinforcing this learning. And what we are going to do next is we are going to actually learn how to construct or high confidence intervals or how to conduct hypothesis test using confidence intervals. And then we will talk of hypothesis test in linear regression which is a short topic. And close this course with a discussion on what factors affect hypothesis test namely the sample size, the variability, the choice of alpha or the alternate hypothesis itself and so on; in particular, we shall focus on the effect of the sample size.

Alright, then see you in the next lecture.