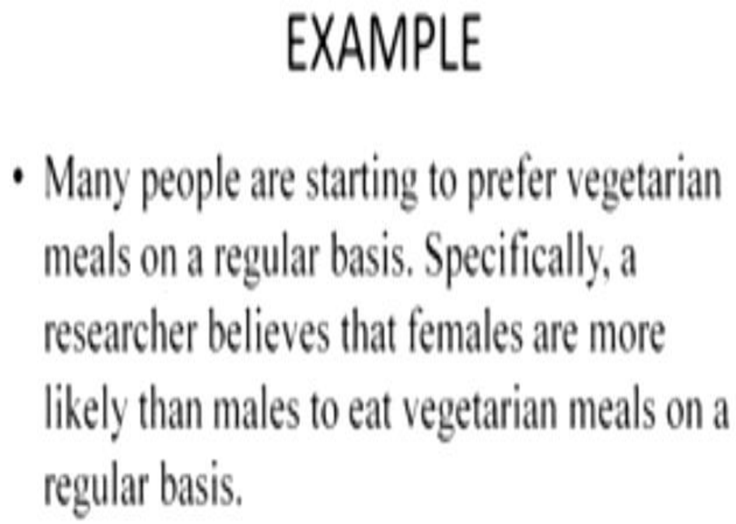
Marketing Research and Analysis-II (Application Oriented)
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 10
Hypothesis Development – II
(with a real life case)

Welcome everyone to the session of Marketing Research and Analysis II. So as we were discussing in the last session about the tail of a hypothesis and the direction of a test. So we discussed about how to develop a two-tailed or two-directional hypothesis. Now today, we will discuss about a one-tailed test and a hypothesis which leans towards the left or to the right okay.

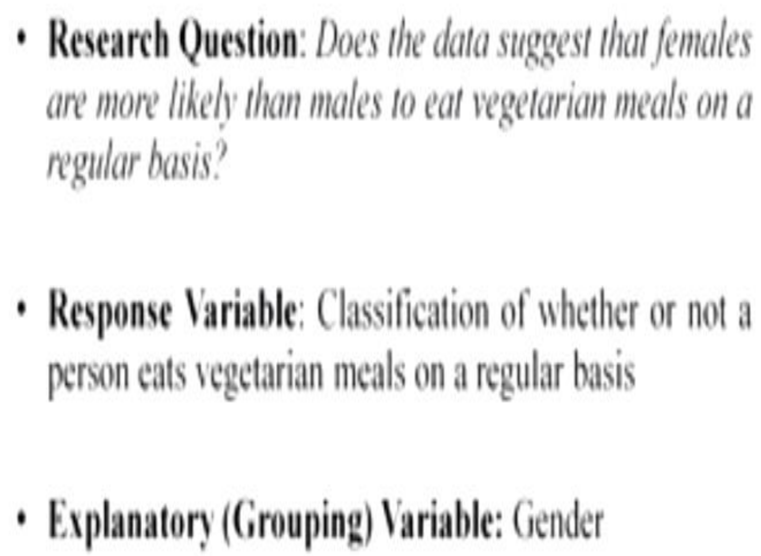
(Refer Slide Time: 01:03)



So let us look at this example first. So it says many people are starting to prefer vegetarian meals on a regular basis. Specifically, a researcher believes that females are more likely than males to eat vegetarian meals on a regular basis. So the question here is that means this researcher is feeling that now-a-days people are more interested to go for vegetarian food and especially females. So what is the research question in this condition and then we will see the hypothesis right.

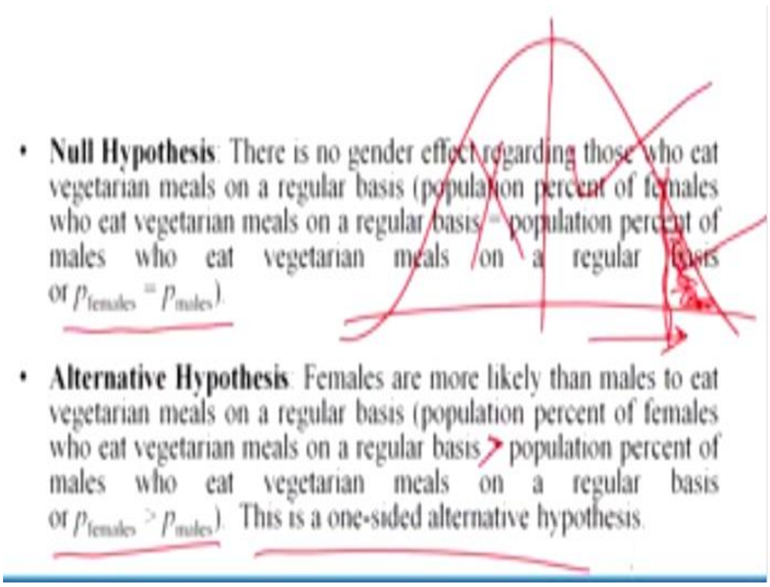
So let us see the research question. You may develop one in this meanwhile, you can think of it and then we are going towards the exactly what actually it happens.

(Refer Slide Time: 01:44)



So research question. So as it says does the data suggest that females are more likely than males to eat vegetarian meals on a regular basis. The question it is asking is can we conclude from the data that females are preferring more vegetarian meals than males on a regular basis. So how would you do it. The classification is done on the persons eating the vegetarian meals on a regular basis. What is the explanatory variable, the gender, whether it is female or male.

(Refer Slide Time: 02:14)

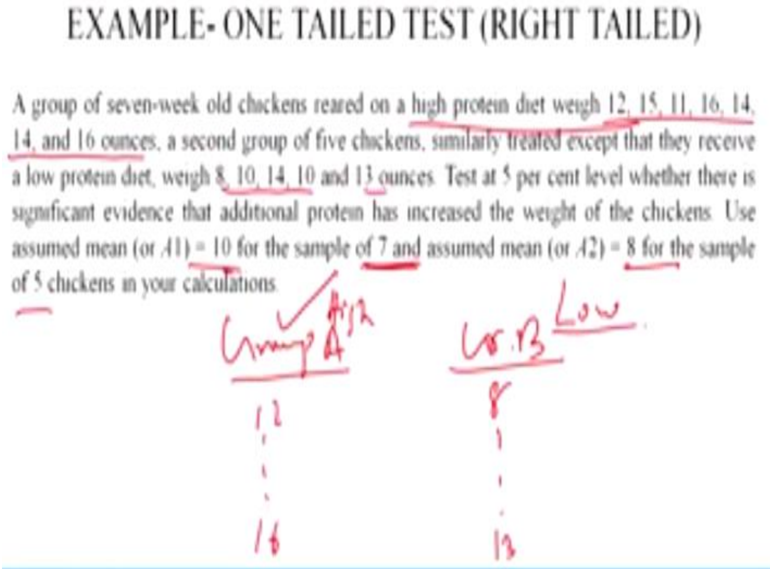


So the null hypothesis in this case is there is no gender effect, obviously we are saying that there is no gender effect. So that means we are saying those who eat vegetarian meals on a regular basis. Now look at this, now here we are saying the proportion of females is equal to the proportion of the males because we do not know at the first instant, so we are saying, the null hypothesis says there is no difference, but the alternate on the other hand says females are more likely than males to eat vegetarian meals.

Now in this case that means what, population percentage of females who eat vegetarian meals is greater than the population percentage of males who eat the vegetarian meals. So in this condition, we are interested for only one side. So if I draw this, how will it look. So suppose I am saying that the percentage of females and the percentage of males, I am saying the percentage of females is more than the males, so my shift of the study is towards this side, is not it?

I am not interested in the left hand side where may be the females number is less than the males, the proportion of females is less than the males. So in such a condition, we are not interested in this, we are interested only in this side. So I am saying this is my critical value the boundary and this is my rejection zone. So the complete rejection zone falls towards the right. So this is a one-sided test, alternative hypothesis.

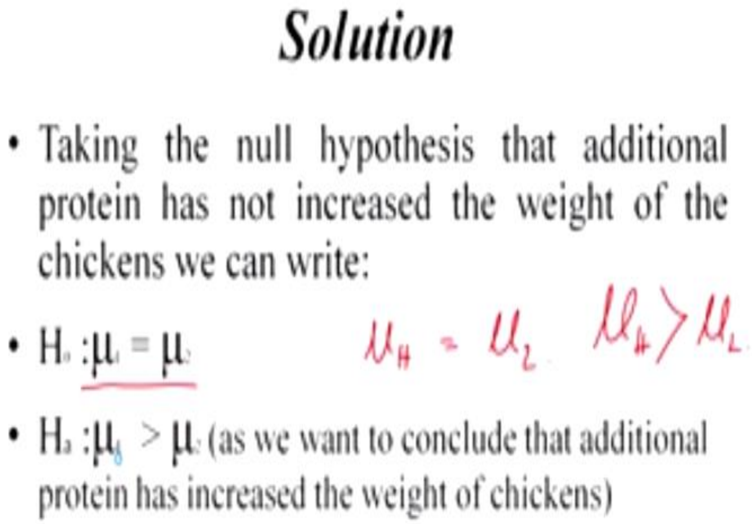
(Refer Slide Time: 03:52)



Let us do one more. A group of 7-week old chickens reared on a high protein diet weighed this following; 12, 15, 11, 16, 14, 14, and 16 ounces, so this is group A, group B. So you had 12, 15, and goes on till 16 ounces. The other side was that this second group of 5 chickens similarly treated except that they received a low protein diet weighed 8 to 13. So he is saying the researcher is asking you the to test at a 5% level of significance whether there is significant evidence that additional protein has increased the weight of the chicken.

So the difference is one was given a high protein diet, the first case, high protein diet, this is high protein, this is low protein okay. So is there any difference between the weight of the chickens when they are treated with a different protein content. So testing at level of significance, he says test at 5% level of significance, whether there is significant evidence that additional protein has increased the weight of the chickens. Use the assumed mean = 10, that means assume the mean to be 10 for a sample of 7 and A2 this side is equal to 8 for a sample of 5. So there were 1, 2, 3, 4, 5, 6, 7 in the first case, sample of 7, and 5 in this case 1, 2, 3, 4, and 5.

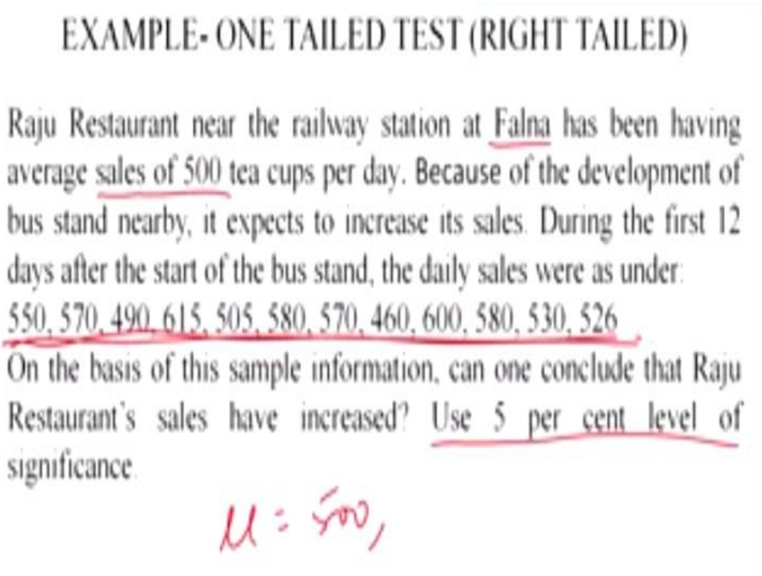
(Refer Slide Time: 05:38)



So first of all let us see. So what is the solution? How do you develop the hypothesis? Taking the null that additional protein has not increased the weight, obviously you cannot claim anything at the moment. So you say mu1 ($\mu 1$) = mu2 ($\mu 2$) that means the weight of chicken with a high protein is equal to the weight of the chicken with a low protein diet. On the other hand, the researcher is actually interested if he has given extra protein that means he needs to see because he has put in some cost into it.

So he is trying to see whether actually protein has helped or not. So he is trying to now say high protein has got a weight larger than the low protein ones. So as we want to conclude that additional protein has increased the weight of chicken, therefore we are saying mu (μ)1 is greater than mu (μ)2 okay. So this is a one-tailed, but right-tailed test.

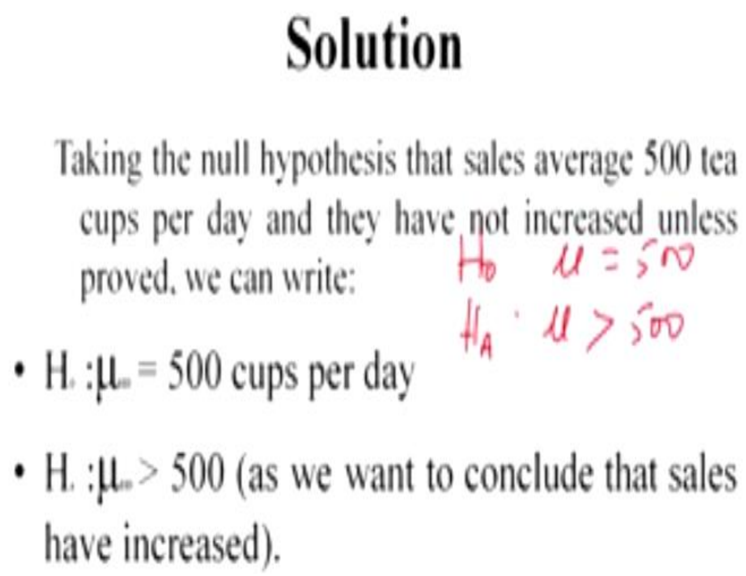
(Refer Slide Time: 06:33)



Similarly one more. Raju Restaurant, a restaurant's name, near the railway station at a place Falna has been having average sales of 500 tea cups per day. Because of the development of the bus stand nearby, the government did something and the improvement has happened, it expects to increase its sales because of whatever the government's developments has happened in that place. During the first 12 days after the start of the bus stand, the daily sales were as under given post the development.

On the basis of the sample, can one conclude that Raju Restaurant sales have increased, yes or no, again he is saying use a 5% level of significance. So what is it saying, earlier the mu (μ) was 500. Now after the test has been done, now it is given certain sample values it has given and he was asking you whether the sample after the development whatever has happened you have taken the sample studies, whether the difference in the pre and post has there been any difference or not. So it is more like a same sample taken twice okay.

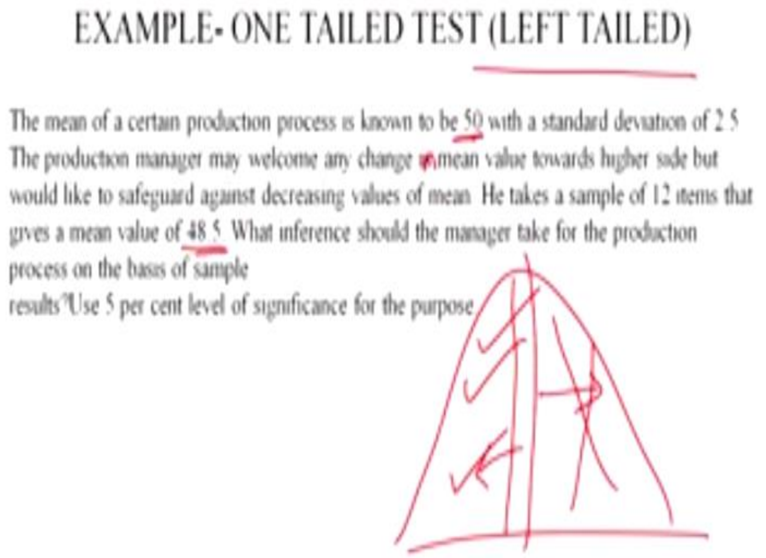
(Refer Slide Time: 08:03)



Solution. So mu (μ) = 500 cups per day, correct. So we are thinking the mu (μ) has as you see taking the null hypothesis average sales is 500 tea cups per day and they have not changed, they have not increased unless proven. So here I have not proven anything, but I have to test the null hypothesis. What is my alternate, now mu (μ) is greater than, because there has been a development. So because there has been a development, the sales has increased.

So mu (μ)H is greater than 500 as we want to conclude the sales have increased. So you need to be very careful, you need to be very clear and careful with this what do you want. I want to know whether it has increased more than 500, just write the alternate first then. So what I am saying my H alternative is that mu (μ) is greater than 500, this is what I want to check, but on the contrary then what is my null, the mu (μ) = 500, there is no change, so this is my null right.

(Refer Slide Time: 09:01)



One more. The mean of a certain production process is known to be 50 with a standard deviation of 2.5. The production manager may welcome any change in the mean value towards higher side, but would like to safeguard against decreasing values of mean. Suppose it increases towards the higher side, it is good. For example, let us say you are a shopkeeper selling bulbs. So if the life of the bulbs is increasing, no issues, it is satisfactory, it is rather a good thing but he is worried whether it should not fall.

If it falls, then consumers will have complaints okay. Same thing here. He takes a sample of 12 items that gives a mean value of 48.5. The sample mean is 48.5. What inference should the manager take for the production process on the basis of the sample results. Again take a 5% level of significance. What you do is you try to write the null hypothesis and the alternate. As you see, this is a left tailed, why?

Because as I said, the researcher is not interested if it increases. If it increases, he says it is good, so I am not worried, but what I am worried is it should not be less, if it is less, then it can be dangerous for me. So I am interested in this zone, so this zone I am not interested. So this is the left-tailed test.

(Refer Slide Time: 10:39)

Example

- A study was interested in determining if an exercise program had some effect on reduction of Blood Pressure in subjects with abnormally high blood pressure. μ_1 μ_2
- For this purpose a sample of $n = 500$ patients with abnormally high blood pressure were required to adhere to the exercise regime.
- A second sample $m = 400$ of patients with abnormally high blood pressure were not required to adhere to the exercise regime.
- After a period of one year the reduction in blood pressure was measured for each patient in the study.

Now let us come to the last example for one tailed. Now here in this case, the researcher is interested in determining if an exercise program had some effect on the reduction of blood pressure on people who had an abnormally high blood pressure. So 2 groups were selected, one n = 500 patients with high blood pressure and they were given an exercise program. In the second sample where m = 400, the patients also had abnormally high blood pressure, but they were not asked to go through the exercise program.

After 1 year, the reduction in blood pressure was measured for each patient in the study. So here let us say the reduction in blood pressure in this case is mu(μ) 1 and the reduction in blood pressure in the second case is let us say mu(μ) 2 okay.

(Refer Slide Time: 11:27)

We want to test:

$$H_0 : \mu_1 \leq \mu_2$$

The exercise group ***did not*** have a higher average reduction in blood pressure

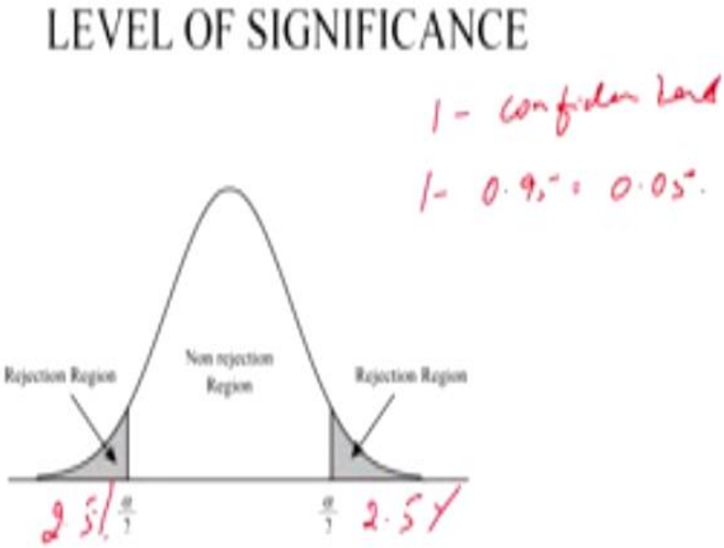
vs $$H_A : \underset{500}{\mu_1} > \underset{400}{\mu_2} \rightarrow \text{no exercise.}$$

The exercise group ***did*** have a higher average reduction in blood pressure

So here we are saying a null hypothesis says that the exercise group did not have higher average reduction in blood pressure, either is equal to or even less than. So what is the alternate that you are interested in, to see whether the exercise had an effect or not effect, and primarily you want to see the exercise had an effect.

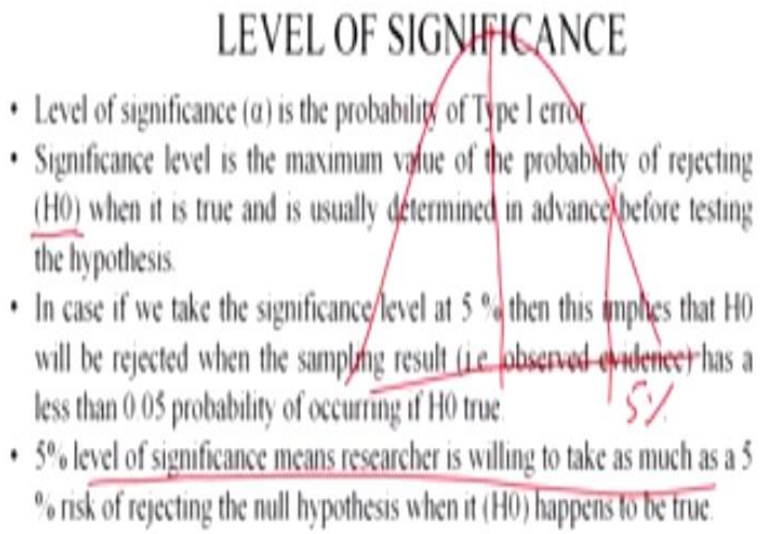
So let us bring it there, so versus the alternate hypothesis in which you are saying the average reduction of blood pressure for the first group which had 500 people is more than the average reduction in blood pressure of the second group which did not exercise and there were 400 hundred people and they did not exercise. So this is what we are interested in. So the exercise group did have a higher average reduction in blood pressure. So now I think you are clear with the tail.

(Refer Slide Time: 12:20)



Now coming to something very important as I had earlier also said, what is this level of significance, I have already explained but still. So level of significance I had said is nothing but 1 minus confidence level. So this is for example you are working at a 1 minus let us say 95% confidence level right. So this is equal to your alpha (α), alpha (α) is the significance level = 5%, and when it is a two-tailed test, as I had earlier said it is distributed on 2 sides, 2.5 here 2.5 here okay, 2.5%, 2.5% okay.

(Refer Slide Time: 13:03)



So level of significance is the probability of a Type I error. Now let us come to this word of Type I error. What is this Type I error, first I will explain. Significance level is the maximum value of the probability of rejecting null hypothesis, H0 is the null hypothesis when it is true and is usually determined in advance before testing the hypothesis. Just take this in case if you take the significance level at 5%, this implies that H0 or the null hypothesis will be rejected when the sampling result has a less than 0.05 probability of occurring is H0 is true.

Now I will explain through a diagram, 5% level of significance means researcher is willing, this is important, this is more clear you see, the researcher is willing to take as much as a 5% of rejecting the null hypothesis when it happens to be true. Okay that means let me explain through a diagram. Now suppose this is my distribution okay, here I am saying this is my 5%, 5% level of significance. So that means what, suppose the null hypothesis is true, take in case of a product.

A manufacturer of let us say laptops. He is manufacturing laptops and he is saying I am interested to work at a 95% confidence level, that means suppose all the products are good still he is ready to reject 5% of the laptops right, although they might be good. The product is good, still he is ready to reject them because of some error in sampling. So this is what is a risk that the producer takes.

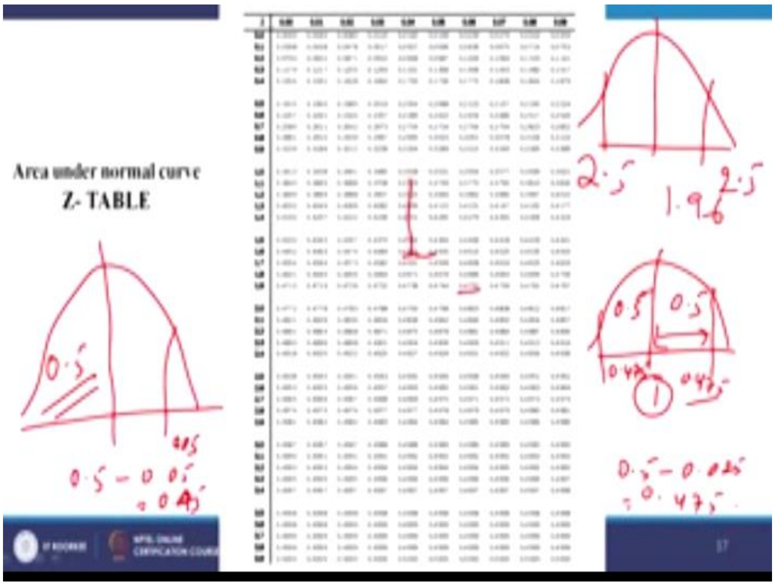
(Refer Slide Time: 14:51)

SUMMARY OF CERTAIN CRITICAL VALUES FOR SAMPLE STATISTIC 'Z'

Rejection Region	Level of significance, α per cent				
	10%	5%	1%	0.5%	0.02%
One-tailed region	+/- 1.28	+/- 1.64	+/- 2.33	+/- 2.58	+/- 2.88
Two-tailed region	+/- 1.645	+/- 1.96	+/- 2.58	+/- 2.81	+/- 3.08

So if you can see summary of certain critical values now for a one-tailed region and a two-tailed region, the values are different. The table values for the level of significance is different. So 10% if you see for a one-tailed region it is 1.28 and for a two-tailed region it is 1.645. I will show you how to read a table also. For 5%, one-tailed is 1.64, two-tailed is 1.96; 1% it is 2.33, 2.58 and it goes on okay.

(Refer Slide Time: 15:21)



So this is how you should see. For example as I said to you I am interested to see for let us say 95% confidence for a two tailed, let us say 95%, so 2.5 here, 2.5 here, how to check it. So what you do is 5% if you divide, suppose the entire area of the distribution is equal to 1, so this side is 0.5, this side is 0.5 right. So now you are cutting down 2.5, so that is equal to if I subtract it, so that means the acceptance area this much is equal to 0.475 this side and here also 0.475.

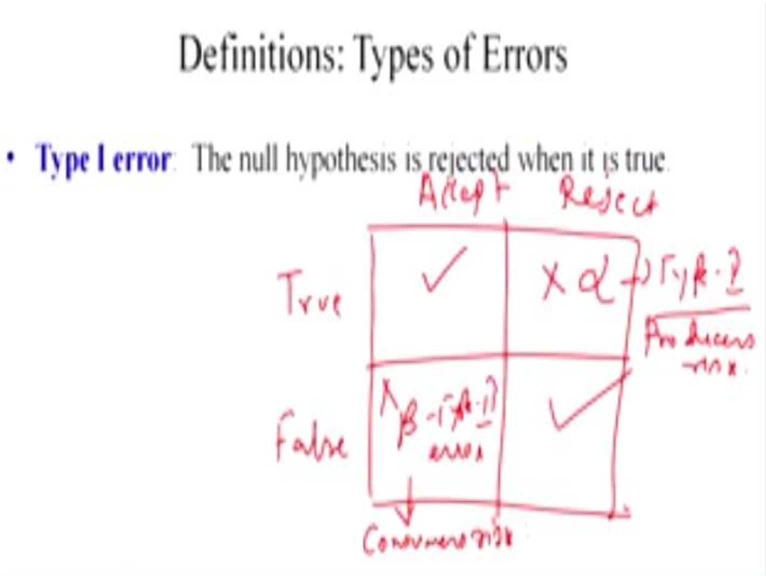
So you divide it, you subtract it, 0.5 – 0.025 so that is equal to this value right, 0.475 okay. So just search 475, so 475 if you search here is here right, so this is 475, so what is the value 1.9 and against the vertical side it is 6.06, so this becomes 1.96. Similarly if somebody is going for a one-tailed test, now what happens in a one-tailed test, let us say, so right or left does not matter. Let us say I am interested for the right-tailed test only. So this is 0.5 as it is and I am cutting the rejection zone at 0.05 okay.

So now, I am interested for this area only. So here, so I am interested to check what is it in a 0.5 – 0.05 so that is equal to 0.45 okay. Now find out 0.45 here. So zero 0.45 is somewhere here, so it is 1.6, you can say 4495, so this is 1.64. So this is how you should check the one-tailed value and the two-tailed value, the area under the curve.

(Refer Slide Time: 17:31)

So now coming to why this is important the significance, what is the merit of significance and how it effects? So there are some errors that a researcher can make. Now what are the errors?

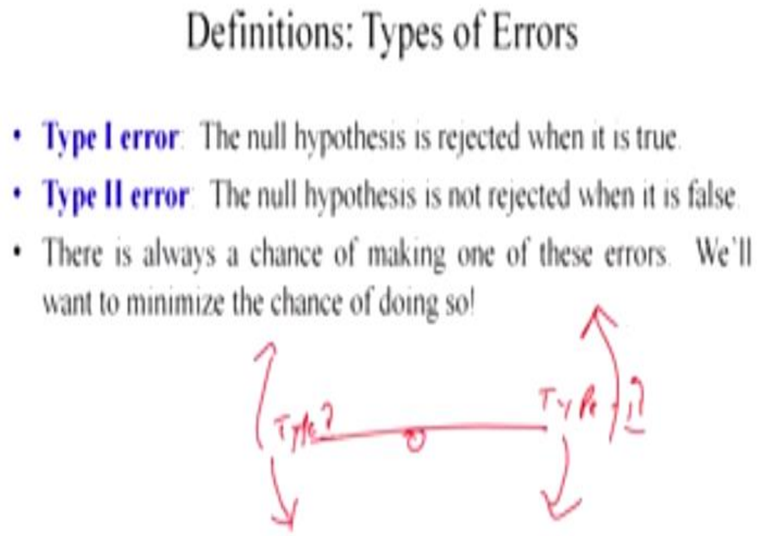
(Refer Slide Time: 17:45)



The errors are suppose let us say you see the null hypothesis is true, the null hypothesis is false, it is accepted, it is rejected; suppose this is the condition. So if it is true and you accept, then it is a good condition, no problem; but if it true and you are rejecting, so this is where you have committed a mistake. So here, the rejection is this alpha (α), we say alpha (α) is the Type I error. False and you have accepted, is a mistake, this is called beta (β) or Type II error, this is also a good condition, it is a false but you have rejected.

So Type I error as I said the null hypothesis is rejected when it is true, the null hypothesis is true but you have rejected it, so it is a Type I error. Type I error is also called as producer's risk, as I explained from the laptop example you can think because he is ready to reject 5% of his good products also. On the other hand, obviously you can now understand, you can really makes sense that it makes sense, that this is called a consumer's because he is accepting a false bad product in the market. So if he is accepting sending a bad product, then it is a consumer's risk obviously.

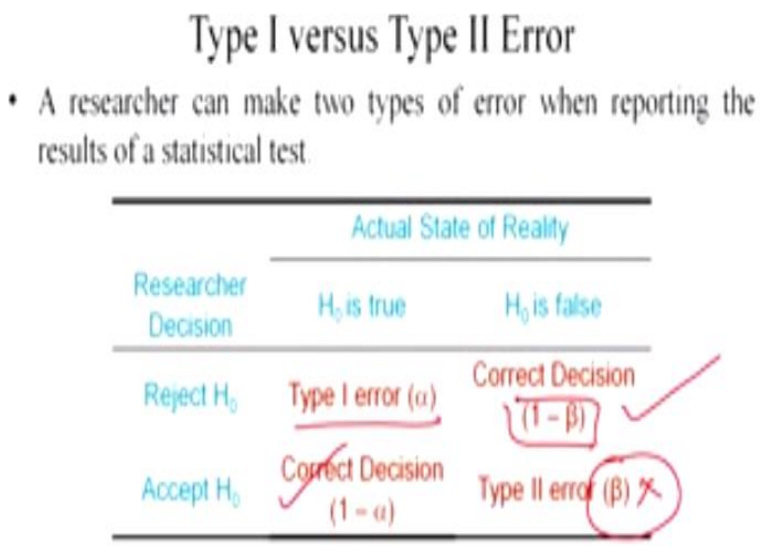
(Refer Slide Time: 19:15)



So now let us see. The null hypothesis is not rejected when it is false or you are accepting a wrong hypothesis which should have been rejected actually. So there is always a chance of making 1 of these errors, you can never come make it 0 right. We will want to minimize the chance of doing so, either you will have some kind of a Type I error or you will have some kind of a Type II error, and if you try to decrease one, the other will increase and vice versa, so it is like a seesaw.

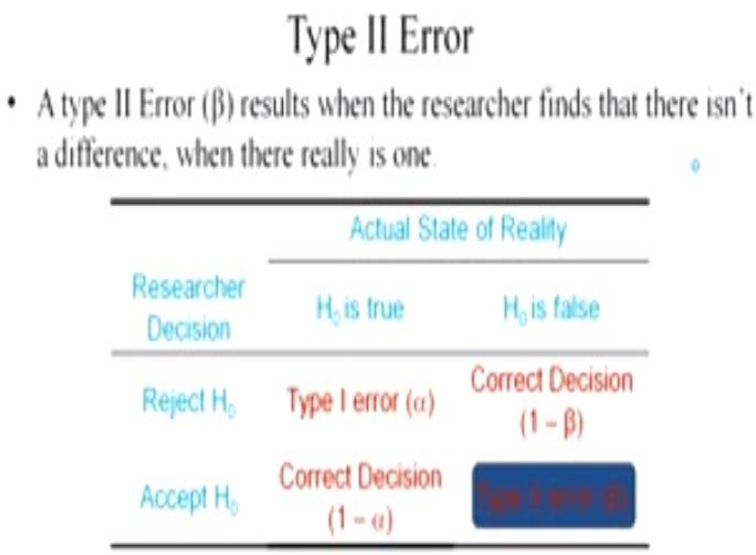
So you can understand this is like a seesaw, this is Type I, this is Type II. So if this grows, then it will come down or if you bring this one down, then this will go up.

(Refer Slide Time: 20:08)



So let us take this, Type I versus Type II error. As I had explained, so reject H0, accept H0 is the decision. So Type I error just this is slight shift here, actually this is true, null hypothesis is true, this is null hypothesis is false. In my diagram which I had done, I had brought this portion here and this portion I had taken here, so that is why the difference. So when H0 is true and you have rejected as I had said Type I error, when it is let us say false and you have rejected, so correct decision right. You forget this part at the moment, I will explain this. When it is true and you have accepted, correct decision, but it is false you have accepted, it is a wrong decision or beta (β). So now this beta (β) is a very important thing, what is it saying, you are accepting a false hypothesis.

(Refer Slide Time: 21:08)

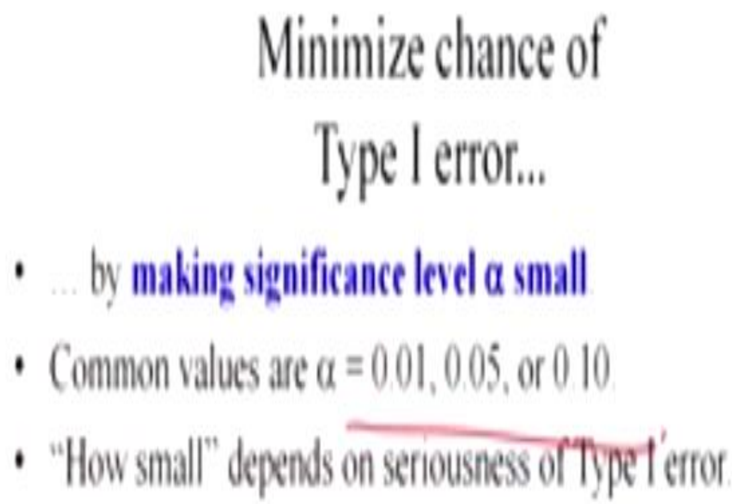


So the probability of a Type I error is determined by the alpha (α) level set by the researcher. Now what is that, let us see. So in this case if you see, the same thing happens here, so Type I right, so we go to the Type II.

(Refer Slide Time: 21:27)

Now let us see, now in this case in the Type I what is happening here, mu(μ) is somewhere here right and your rejection zone is here. So what you have done, you have rejected, this blue is a sign of a good null hypothesis, so you have rejected it, you have taken it to the rejection zone, so that is why this 0.05 alpha (α). It should have been within this limit, but you have actually taken it to the rejection zone, so minimize the chance of Type I error by making significance level small. So if you make a small significance level, your Type I error will be reduced.

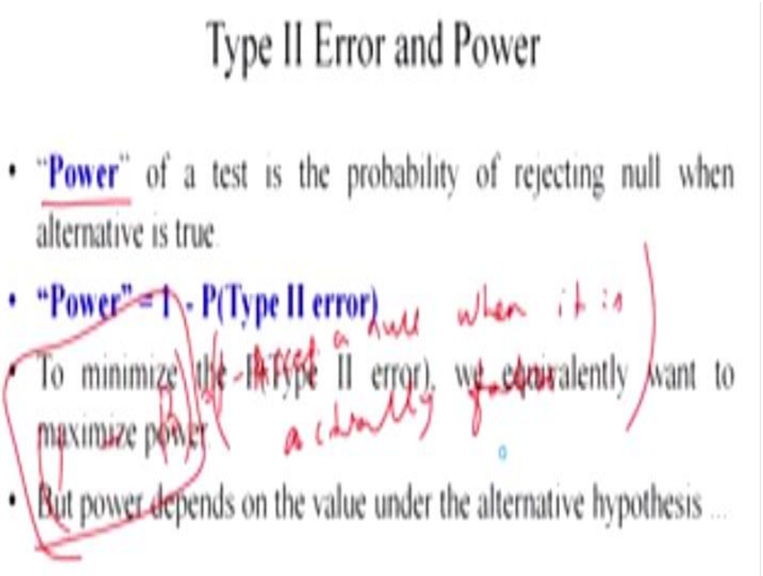
(Refer Slide Time: 22:11)



Common values are this. How small depends on the seriousness of the Type I error, that means what? The Type I error we said is rejecting a right null hypothesis or a good null hypothesis or a null hypothesis which is true you have rejected, that is a Type I error. Now how small this value should be that entirely depends on the researcher to decide, on the seriousness of the Type I error and the researchers objective. Decision is not a statistical one but more of a practical one.

Suppose this value of rejection of a true one you might take it very very low for a case study like a space rocket launch, there you would like your alpha (α) value to be as small as possible, but suppose you are a maker of tables, may be you are not worried and you may be even ready to take a 5% level of significance.

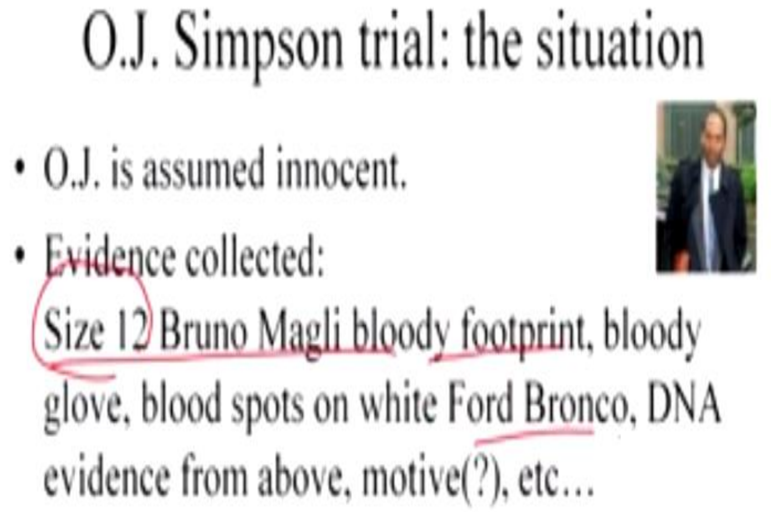
(Refer Slide Time: 23:10)



So Type II error and power. So what is this Type II error and what is this power. Power of a test is the probability of rejecting the null when alternative is true. What is it saying, power of a test is the probability of rejecting the null when the alternative is true, so what was the beta (β), first of all let us remember the beta (β). Beta (β) was we said accepting a null hypothesis when it is actually false, this is what was our beta (β).

Now if I do it like this, beta (β) is equal to this much, so what is 1- beta (β), so 1-beta (β) is 1- accepting a null hypothesis when it is actually false, that means what, you are accepting a true hypothesis, 1- accepting a null when it is actually false, so this is what is called the power. This 1- beta (β) is called the power of a test, which I had said I will later on explain, if you can see here this one I had said I will explain later on, but power depends on the value under the alternative hypothesis.

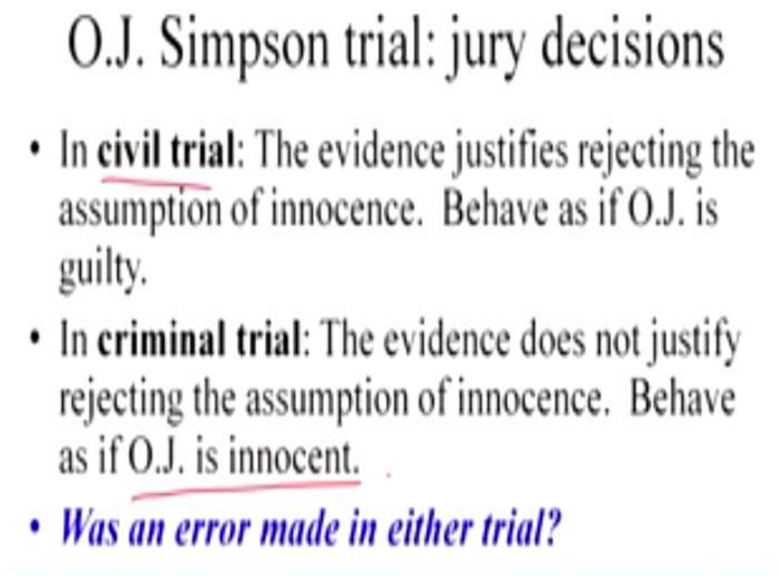
(Refer Slide Time: 24:22)



Now let us take a very classical case actually, this is a very classical case of the Simpson trial, O. J. Simpson who was a player and he faced a trial of murdering his wife, I think murdering his wife or girlfriend whatever it was. So he is assumed to be innocent. As initially when you make the null and alternate, the first thing is you say there is no effect, there is no difference, so similarly you are saying although he might be a criminal but you are saying he is innocent at the first stance, so the null hypothesis says he is innocent generally speaking.

So the evidence was collected that they found some shoes Bruno Magli is a brand of size 12 because he had a large feet which was bloody and there was blood stains, blood stained gloves were there, blood spots were found out on this car, and DNA evidence also was supportive.

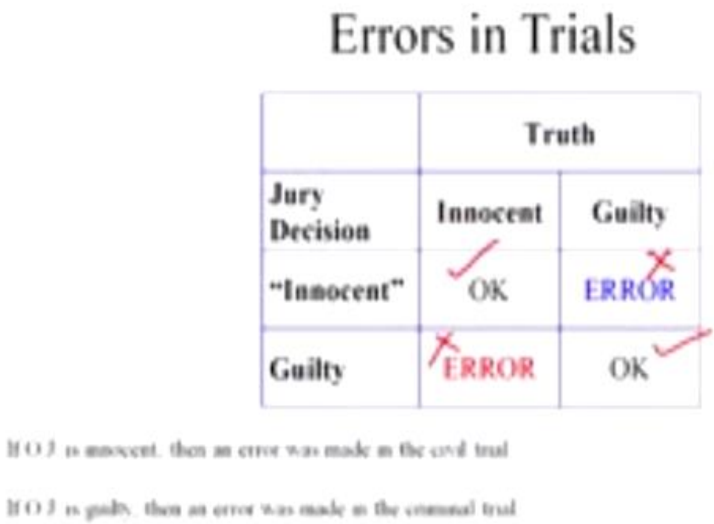
(Refer Slide Time: 25:21)



Now what happened in this case. In civil trial, there are 2 trials basically, the evidence justifies rejecting the assumption of innocence. Behave as if O. J. is guilty. There are 2 ways, there are 2 trials we are taking, civil trial and the criminal trial. In the civil trial, we are saying behave as if O. J. Simpson is guilty, that means you reject the assumption of innocence. In the criminal trial, the evidence does not justify rejecting the assumption of innocence, so you behave as if O.J. is innocent.

So one case we are taking in the criminal trial he is innocent, so before the judge, the criminal lawyer tries to give him a sentence he would first go with the innocent condition right. Was an error made in either trial? Let us see.

(Refer Slide Time: 26:10)



So if the jury decision is innocent and guilty, innocent and guilty, so truth is innocent guilty. So if jury decision he is innocent and he is actually innocent, then it is a good condition, but if he is innocent jury saying but he is actually a guilty, then it is an error right. Again he is saying he is guilty, but actually this is true, that means he is innocent, that means he is not guilty, then it is wrong. So these 4 cases can happen in this case as we explained. If O. J. is innocent, then error was made in the civil trial. If O. J. is guilty, then an error is made in the criminal trial right.

(Refer Slide Time: 26:55)

Power of a test

So what I will do is maybe I will take it over to the next session. So I will explain little more details about the power of a test and we will do that. So today what we have learned now in this session is we have very clearly understood what is this one-tailed test, how to develop a one-tailed test, how you should be understanding the rejection and acceptance zones in a one-tailed test, then what is a significance level and how it is creating problems for example a Type I and a Type II error and we have seen one example of this O. J. Simpson case.

What is the power of a test and why power of a test is important and how it affects some studies, maybe I will explain it in the next session. So I wish I hope you are clear with today's session and I wish you all the luck. Thank you so much.