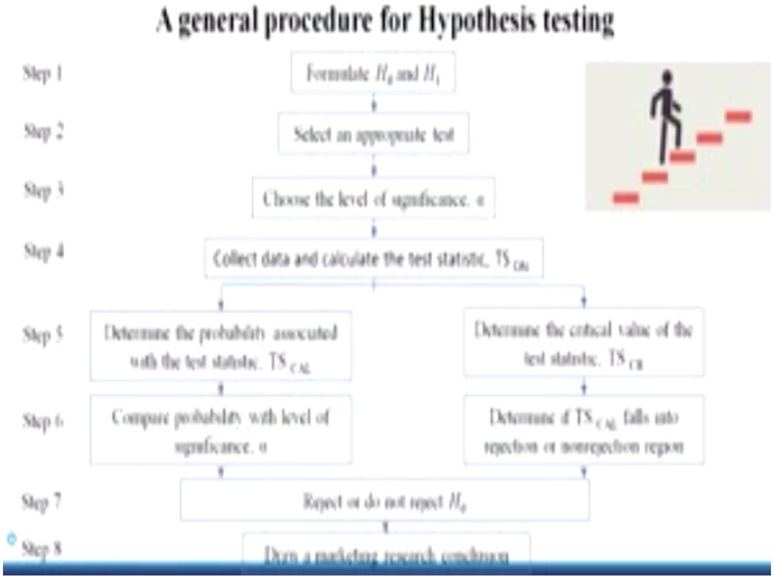
Marketing Research and Analysis-II (Application Oriented)
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 9
Hypothesis Development – I
(with a real life case)

Welcome everyone to the lecture series of Marketing Research and Analysis II. In the last class, we had discussed about research question and hypothesis development. So we just learned that research question basically comes from the exploratory study, so where the researcher puts in a question when he does not have any knowledge and hypothesis somewhere, where you have some knowledge about the subject of study and then you try to develop some propositions.

You assume something so that you can later on try to prove it, test it and prove it, so but before that, the very important question that comes to mind is this how to develop this hypothesis. So let us look at the slides, this was from the last class if you can see.

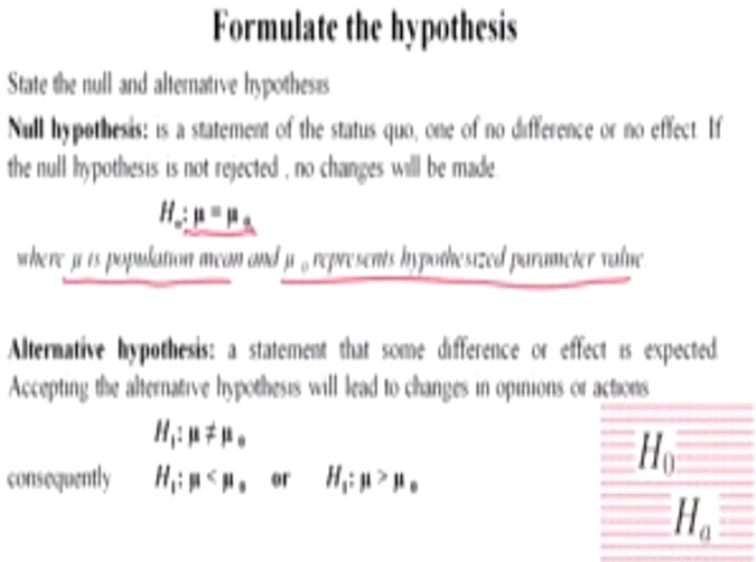
(Refer Slide Time: 01:22)



This is how the hypothesis looks like. So you formulate the hypothesis at the beginning, so the H0 stands for null hypothesis H1 stands for the alternate hypothesis. Then once you test and develop the hypothesis, which I said how this is very important because if you go wrong in it then maybe you are testing the wrong hypothesis, so maybe I will show you again little bit. Then after that, we select the right test. So what are the tests?

So there are various tests to test your hypothesis, so which test is the most appropriate test in your case that needs to be learned. Then comes the level of significance, now what is this level of significance, I will explain and it is very important during a hypothesis test. Finally we calculate the statistics and then compare it with some table value and then say whether the hypothesis is to be rejected or not rejected and finally draw a conclusion out of it. This is what we had started.

(Refer Slide Time: 02:27)



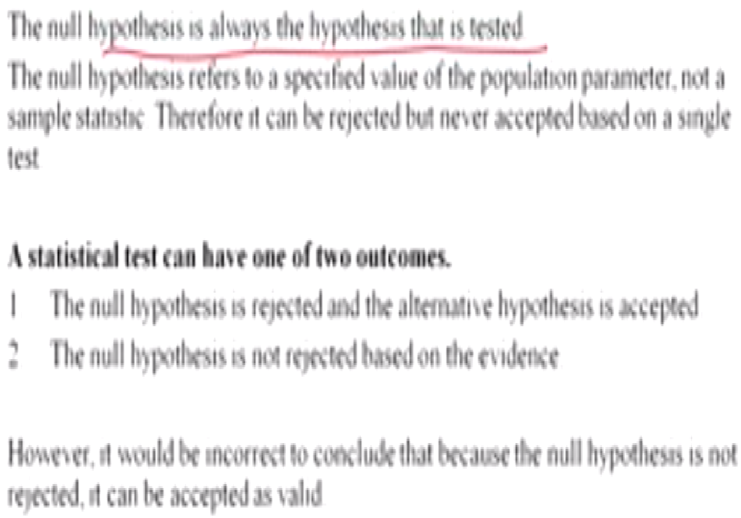
So null hypothesis is as I had said is a statement of the status quo, means something that is going on and you want to maintain that. So if I remember you should go back to Newton's law of inertia where it says everybody continues to be in a state of rest and continuous moving until and unless something external force is applied over it. Similarly the null hypothesis is the same thing. It says it does not change until and unless something happens.

So the question is when you make the null hypothesis, so if in this case you see, so we are saying there are 2 things, one is the μ which is the population mean and μ_0 is a hypothesized parameter value. Now this hypothesized parameter value for example when I will explain you about one sample test or something then you are comparing the sample mean with the hypothesized population mean and you try to see whether actually there is a significant difference exists or it does not exist.

On the other hand, the alternative hypothesis is something where we are trying to see that a statement that some difference is expected, that means as I said in the last class also, the alternate hypothesis is the one which the researcher is more interested generally in almost all the cases. There are few cases where the researcher is also interested to accept the null hypothesis, but there are very limited cases, generally when we talk about in a marketing or market condition, we will always think of accepting the alternate and try to disprove the null hypothesis, that is the most important thing that a researcher thinks of.

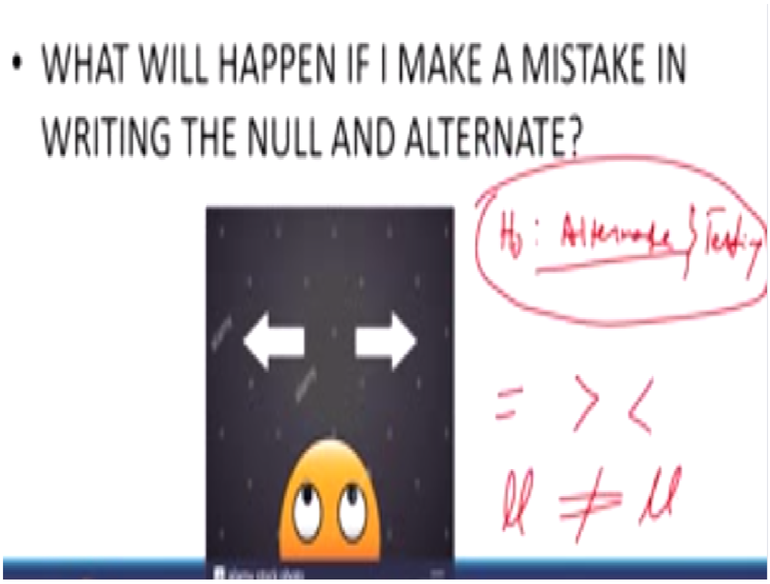
Similarly as I said so in this alternate hypothesis what is happening? If you can see μ is not equal to μ_0. So the population mean is not equal to the hypothesized value, that means μ can be less than μ_0 or μ can be greater than μ_0, so this is the case that we are talking about.

(Refer Slide Time: 04:35)



As I said the null hypothesis is the one that is always tested.

(Refer Slide Time: 04:36)



So if you will see here I put in a humorous picture the person is thinking where to move. So what happens if I make a mistake in writing the null and alternate. So if you make a mistake that means suppose you write the H0 as the one which you are writing for the alternate. Suppose you make this mistake, then what is happening is you are actually testing this one, you are testing this hypothesis. So ultimately, the inference that you will draw will be according to this one, so that means you have already made a big mistake.

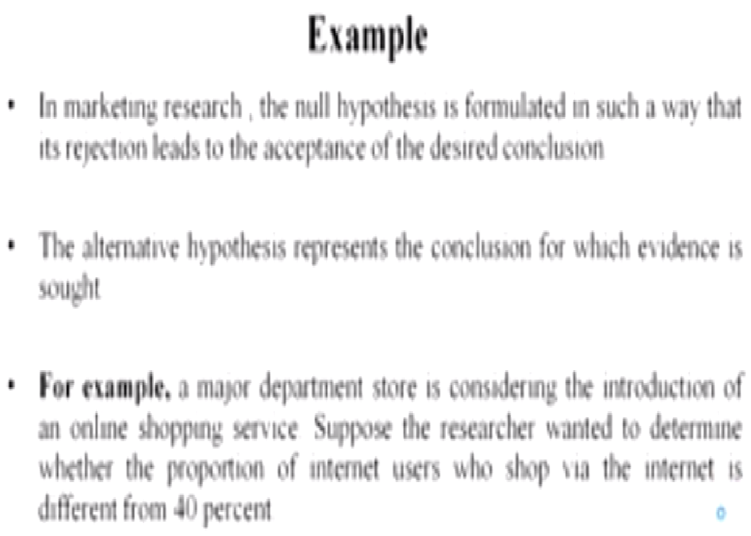
So what you should have been rejecting may be should you have accepted or vice versa it has happened. So you need to be very careful. So to do that also I have had said one thing since most of the time, since the null hypothesis an equal to case, no change, that is equal to, so a = b condition. There can be another statistical condition which can be greater than or less than. So the easier way to write a hypothesis or develop a hypothesis try to see what exactly do you want to check and the one you want to check accordingly or you want to disprove because the intention is to disprove the null hypothesis.

So what is the researcher actually wanting, he wants to know, suppose let us say a company is making an advertisement and he wants to test the effect of advertisement on sales. So the null hypothesis says advertisement has got no effect correct because at this moment, we cannot say anything, so what is the researcher wanting. The researcher actually wants to see the null hypothesis should be wrong that means advertisement should make some effort.

So that should make some effort, could be sometimes this is a bad advertisement, it will have a negative effect; it is a positive advertisement or a well-accepted advertisement, it will have a positive effect. So in that case you are going for the not equal to case, that means the change in sales is different, that means the initial state that you had prior is different to the let us say the later one, let say okay. So this condition is what we are basically checking.

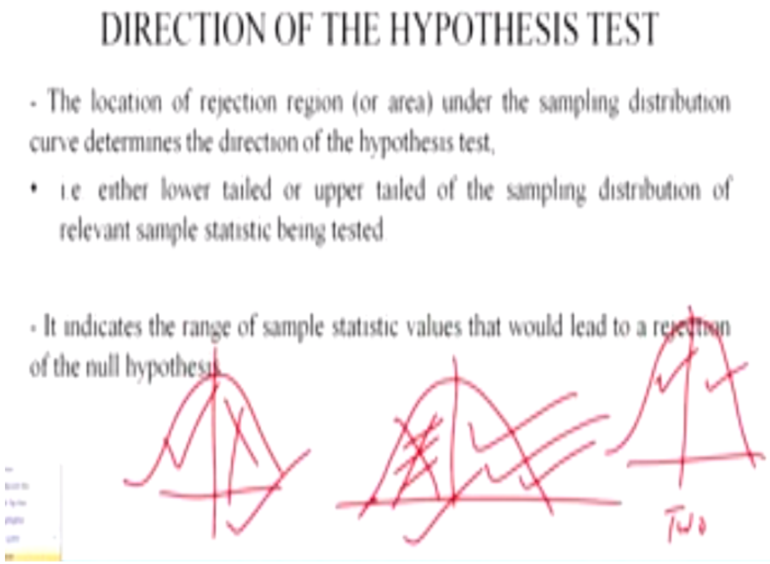
So whenever you are making a hypothesis, kindly be very clear what exactly are you thinking or do you want from this test, the one that you want please write it in the alternate hypothesis simply and the equal to case along with maybe it is an equal to or let us say equal to and greater than or equal to and less than condition, then that is one you take it to the other one, the null hypothesis.

(Refer Slide Time: 07:26)



So forming the null and alternate is actually some you need to be very clear about it. So let us today talk about this thing. So initially when I said you develop your null and alternate, the null sometimes it is also taken as like this, some places it is written as Ha and Hb right H1, H2, the null and alternate. So this way it is written. So the question is after you had the null, then what you do is you need to understand the direction of the test.

(Refer Slide Time: 07:34)

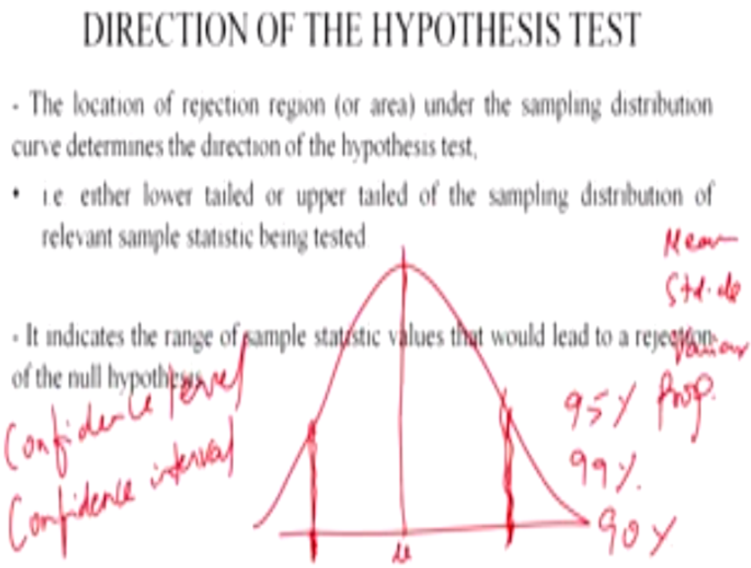


Now what is this direction and why it is important. Direction of the test means what is the side the test is moving, that means suppose you want to see whether the sales is increasing or not only, I want to see whether the sales is, I am not concerned whether the sales is falling or not let us say. Then in such condition my interest let us say in the normal distribution, I am not interested in the negative side, I am not interested so I am crossing this, where I am interested I am moving, whether it is increasing or not, so I am looking at only this side.

In some condition it could be different, it could be like I am interested, whether the price will fall or not, the sales of cars will fall or not, increasing suppose I am not interested, the company is not interested in this part, so is only interested in this part, could be. In some conditions you do not know anything, so you are clueless, so that condition you may go either this side this is also true, this is also true. So this is another kind of where it is a two tailed and even it is a non-directional, in the last class we had discussed.

So these 2 are directional, this is a non-directional. So as you can see what is this directional, let us go.

(Refer Slide Time: 09:35)



The location of the rejection region, so the rejection region. Now let me clear it for your benefit, again I will show you. So the location, this is let us say the normal distribution. So as you know the mean, median and the mode all lie on the same axis okay in the normal distribution. So the question here is what is it saying the location of the rejection region, now what is the rejection region. Now there is a term that you need to be very clear, 2 terms are very important, one called the confidence level and second called the confidence interval.

So what are these 2 things. So to understand properly, confidence level is nothing but the probability that a confidence interval will include the population parameter. Now what does it mean and what is this confidence interval. It is a range. So let us say the range is up to this µch. So I am saying that my class interval that means the range into which the true population will, parameter will fall. What is the true population parameters, population parameters are like for example the mean, the standard deviation, the variance, you can have the proportion.

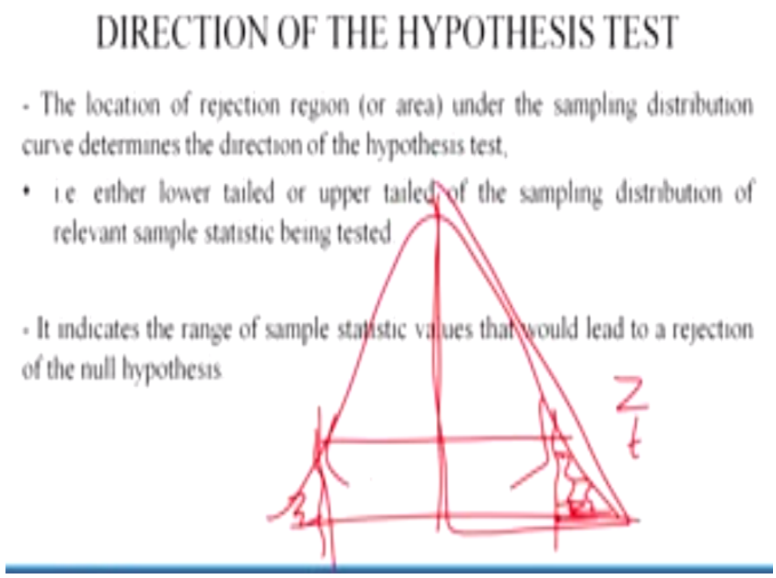
So These are all the parameters that you need to check or you are trying to check. So it says the population parameter falls basically, we are talking about the mean in this case, falls within a standard region. So if it falling within this region, then I am saying well this is my desired limit okay. Now the question what is the probability at which this confidence interval is lying, that changes okay. So I may have a 95% confidence level, I may have a 99% confidence level, I can have 90% confidence level, I can have various confidence levels okay.

So at what confidence level are you talking about? So when you are talking about the different confidence levels, automatically the mean will start shifting. So what is it saying, the location of the rejection region. So if this is my acceptance region, what is my rejection region. My rejection is this spot. If it is a 2 tail test, it is this part. Suppose it is a 1 tail test and I am not bothered about this side, for example forget it then, I am only bothered about this side, this rejection region. So I am only concerned about this much.

So let us be very clear about it. So what it saying is it determines the direction of the hypothesis test, that is either lower tail, lower tail is this one, this part or upper tail, this part, the right end and the left end, left end is the lower upper end is right or the sampling distribution of the relevant sample statistic being tested. It indicates the range of sample static values that would lead to a rejection of the null hypothesis.

Now the parameters that you are using, the values of these parameters will decide whether we would accept or reject the null hypothesis. As I have already said, you are not testing the alternate hypothesis, you are always testing the null hypothesis. So my null hypothesis is saying if anything is lying within this zone, let us say this zone okay, if it is lying within this zone, then I am not rejecting my null hypothesis, but something is lying here in this rejection places let us say let us say.

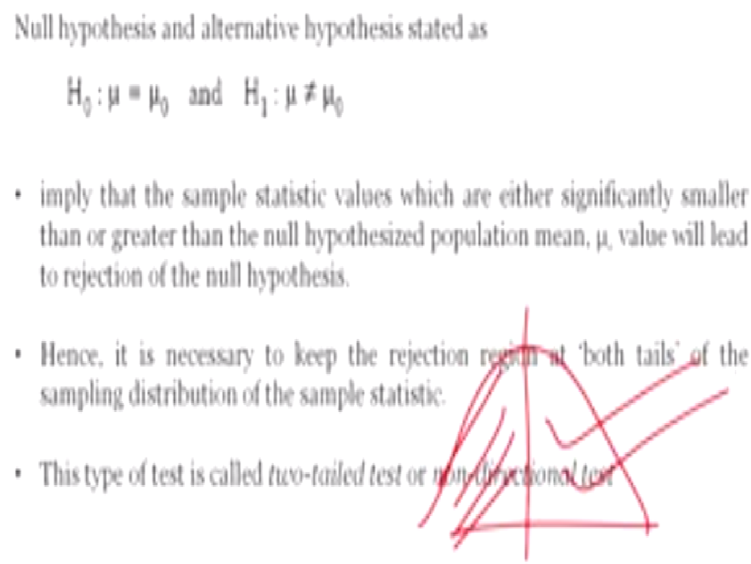
(Refer Slide Time: 13:32)



So let me again make a new diagram for your clarity. So what I am doing now, so I am saying my rejection zone is this place. Suppose a 2-tail test, it is these 2 sides. If it is only 1 tail, then only I am bothered about may be one side, whichever right or left. So I am saying here. So well that will decide whether, suppose my μ falls beyond this, my statistics, my calculated value which we calculate z value or the t value right side.

We say if it the static calculated value comes beyond this region, then we will say it is falling away from the accepted zone which is up to this much and so therefore it is to be rejected. So the null hypothesis is rejected and the alternate is then accepted in that condition.

(Refer Slide Time: 14:26)

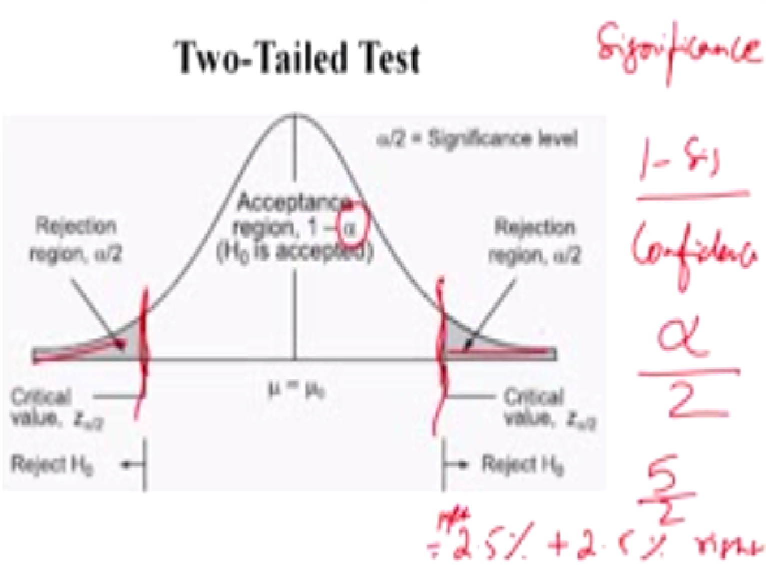


So what it says for example for the null, $\mu = \mu_0$. We have already looked at this earlier also and in the alternate, we are saying μ is not equal to μ_0. So it implies that the sample statistic values which are either significantly smaller than or greater than, the null hypothesized population mean μ_0 will lead to the rejection of the null hypothesis, that is what I said. So if we go back and see if it is falling beyond this limit, this acceptance limit, then you are saying it is to be rejected.

Hence it is necessary to keep the rejection region at both tails of the sampling distribution, but it is not always true because if it is only when you are clueless about the direction, you do not know which direction the thing should move, for example suppose the mean of the population is 10 and you are trying to see whether the mean of the class is coming less than 10 or more than 10, you have no idea, in that condition, it can be even 8 or it can be even may be 13 or 12.

So in such conditions when I am not sure of it, I will go for a 2-tail test, but suppose I am sure if the students of this section are more than the mean of the class, then in that case I will never test for the left tail where it is going, this part I will not check, the left tail, I will only check for the right, actually it is true that the class mean, the average mark or whatever it is, is more than the μ which is the class population's mean. So this type of test is called a two-tailed.

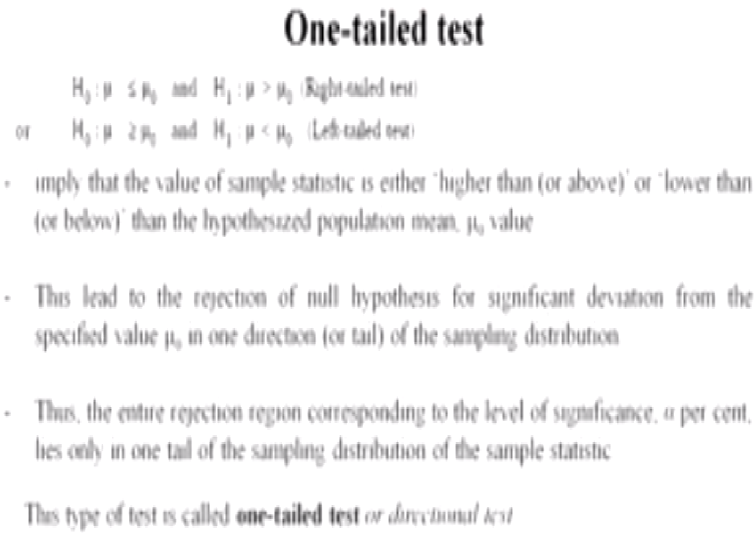
(Refer Slide Time: 16:14)



Now let us look at this, it is I think very clear. So up to this is my acceptance region, H0 is the acceptance. Now what is this alpha, this alpha is nothing but it is called as the significance level. Significance level alpha is very important because the researcher has to decide at what level of significance does he want to test his hypothesis. Does he want to test his hypothesis at 1%, 0.01%, 10%, 5%, it can vary anything. So that means what, 1 minus significance or alpha is my confidence, is it clear. Now look at this rejection zone.

So since this is a 2-tailed test, so what has happened, my alpha has been split into 2 parts because there can be rejection in both sides of the distribution, so that is why what we have done suppose it would have been 5%, let us say, my significance alpha is only 5%, there we would take 2 sides of the distribution, we will divide it by 2 so that becomes 2.5% to the left and + 2.5% to the right, this is right side, this is left side, so this is the left end. So either way it can fall. So if it goes beyond these values, this side and this side, then you are rejecting.

(Refer Slide Time: 18:03)

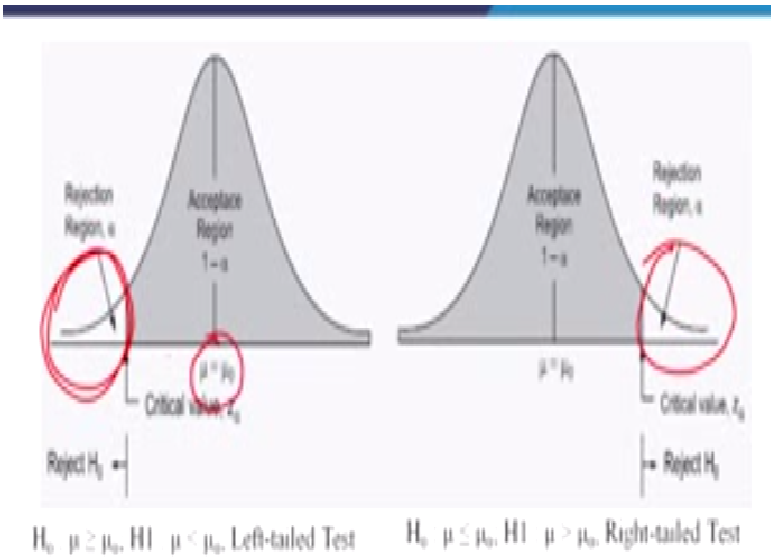


Now coming to the one tail. So in this case, what we are saying is as I said earlier also, there can be a right-tailed test and left-tailed test. So if μ is less than equal to μ_0 and let us say for a null hypothesis, μ is less than equal to μ_0, so the is equal to case is always attached with null if you can observe it right, and alternate is μ is greater than μ_0, that means the population mean is greater than the hypothesized mean. Similarly in the other case for a left-tailed test, μ is greater than equal to μ_0 and the alternate is μ is less than.

So replace the value of the sample statistic is either higher or lower than the hypothesized population mean which is the μ_0 okay. This lead to the rejection of null hypothesis for significant deviation from the specified value of the sampling distribution. I do not get into things, this is very simple okay, very very simple, you just need to understand whether the value that I have counted, whether my calculated value is falling within this limit of acceptance or it is falling beyond my acceptance, to this side or that side that you have to check.

Thus the entire rejection region corresponding to the level of significance alpha lies only in one tail in case of a one-tailed test. So as we can go back and check, suppose this is two tailed may be I have a one tail okay now this is a one tailed. Let me finish this, so this type of test is called a one-tailed.

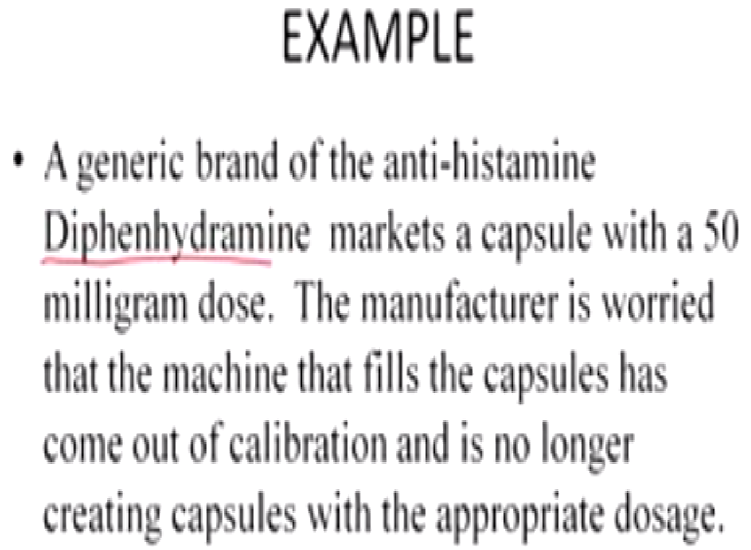
(Refer Slide Time: 19:22)



Now let us look at this case. Now if you see this here the rejection zone is lying here correct, and in this case, the rejection zone is lying here. So what has happened. The alpha which is the significance level, this significance level now did not be distributed divided by 2 because it is moving to only one side, so the entire alpha 5% may be suppose we take a 95% confidence level so 1 – 0.95 minus is equal to 5% which is my significance level, so that is now coming to this end.

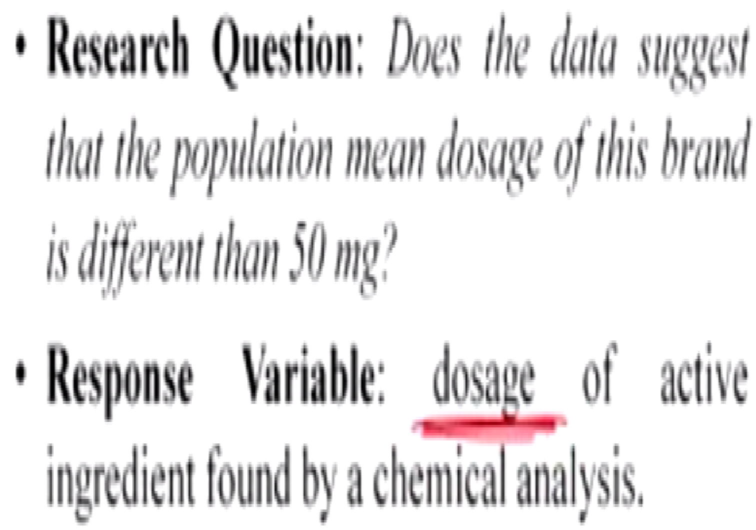
So now for example, you see now here μ is greater than equal to μ_0, μ is greater than equal to μ_0 and H1 is my alternate hypothesis μ is less than μ_0. So when μ is less than the μ_0 that means this population mean is I am saying less than the hypothesized mean, then it is a left-tailed test, and when it is more it is to the right-tailed test. I think this is clear, one-tailed, two-tailed. So let me now go to some examples. How to develop hypothesis. See you have understood by now the direction of the test, which tail you are trying to check and the significance level which is said is nothing but 1 minus the confidence level. So that is very clear. It is the researcher who decides, nobody else tells him or her what is the significance level he should be taking. So now this is the example. Now kindly read the example with me.

(Refer Slide Time: 21:03)



A generic brand of a drug now markets a capsule with a 50 milligram dose. What is they saying? It markets a capsule with 50 mg dose. The manufacturer is worried that the machine that fills the capsule has come out of calibration, it may be worn and it has been old machine or something, and is no longer creating capsules with the appropriate dosage. So in such a condition, what is the research question first let see. Suppose this is the condition, so you can again go back to the case and see so this is the condition, what is the research question. You can think on your own what is the research question by the time I will show you. So kindly try to develop your own research question and the hypothesis.

(Refer Slide Time: 21:58)



So research question does the data suggest that the population mean dosage of this brand is different than 50 mg. The question is, he is asking a question. Does the data whatever data you have collected of different samples, does it suggest that the population mean is different than 50 mg. So what is the response variable you are trying to check. The dosage of active ingredient found by a chemical. So what is the dosage, this is my response, I am interested in checking the dosage.

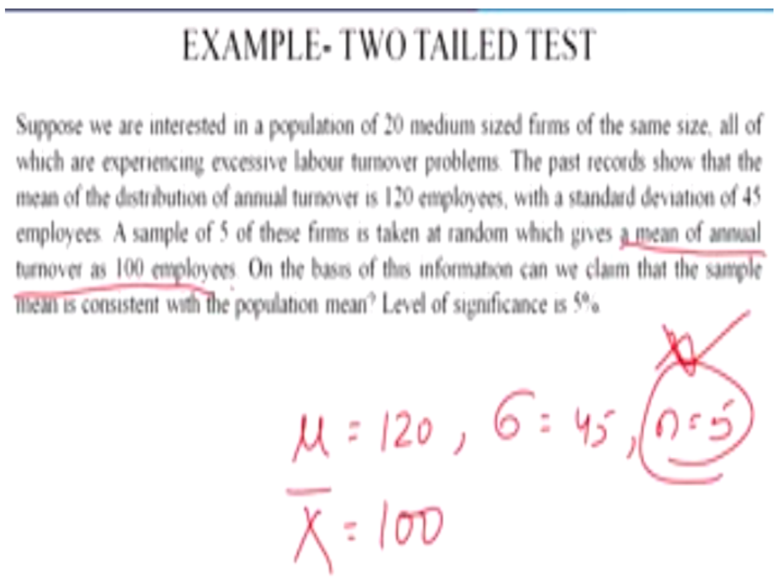
(Refer Slide Time: 22:32)

$\mu = 50 mg.$

- **Null Hypothesis**: On the average, the dosage sold under this brand is 50 mg (population mean dosage = 50 mg).
- **Alternative Hypothesis**: On the average, the dosage sold under this brand is not 50 mg (population mean dosage ≠ 50 mg). $\mu \neq 50 mg.$
- This is a two-sided alternative hypothesis.

Now the null and alternative. So on the average, the dosage sold under this brand is 50 mg. So population mean is equal to 50 mg. So μ = 50, this is my null hypothesis. I am saying that my dosage under this brand is 50 milligram. What is my alternate, is not 50 milligram, obviously on the average the dosage sold under this brand is not 50 milligram. So my μ is not equal to 50 mg. So this is a 1 or 2-tailed hypothesis. As it is not equal to, so it can be less than, it can be greater than, so therefore it is a 2-tailed hypothesis.

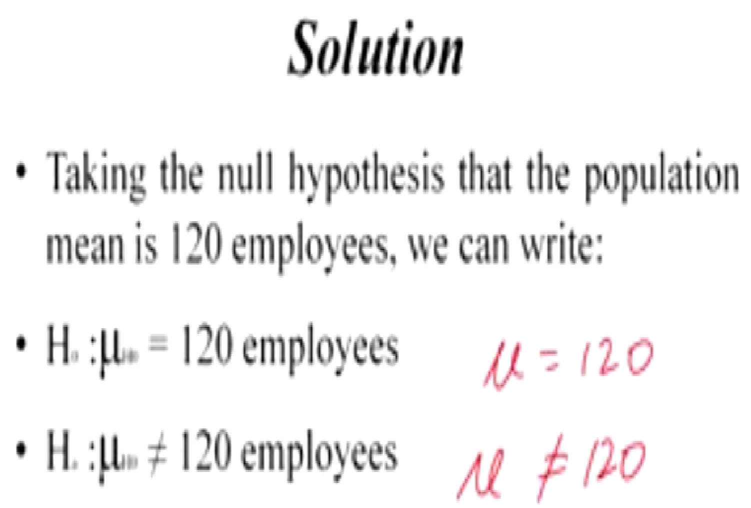
(Refer Slide Time: 23:22)



Now I would suggest kindly do this exercise with me. Suppose there are 2-3 examples which we will cover. Suppose we are interested in a population of 20 medium sized firms of the same size all of which are experiencing labour turnover problems. The past record shows that the mean of the distribution is 120 employees, how much is the mean, $\mu = 120$ with a standard deviation of 45. The sample is marked as n, $n = 5$. A sample of this 5 firms gives, actually $n = 5$ in this case means 5 firms we have taken and gives a mean of annual turnover as 100 employees, now what is this mean of, now out of this 5 firms.

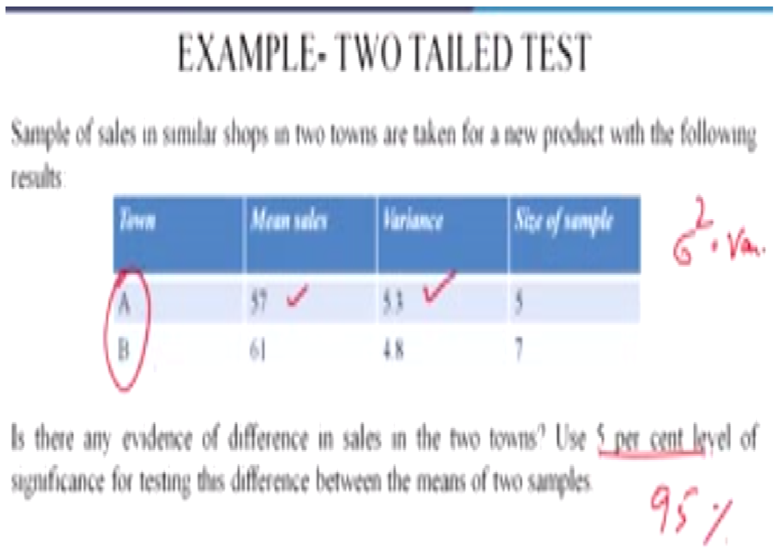
So this is the sample mean, so sample mean is 100. The population mean was 120, the sample mean was 100, taking this $n = 5$. On the basis of this information, can we claim that the sample mean is consistent with the population mean or is it significantly different or is it not different. The level of significance if 5%, now you need to suppose you are a researcher and you are doing some research in HR or something, then you need to test this hypothesis. So first of all kindly formulate the hypothesis.

(Refer Slide Time: 24:51)



The solution taking the null hypothesis that the mean is 120, we can write μ = 120 employees, so μH0 is what did you say here, which gives a mean of 100, so now it wants to check whether this μ is correct or not. So what it is trying to say, the null hypothesis is checking. So if it is not equal to 120, then we can say there is a difference between the population and the sample mean. So that is what he is trying to check. So either μ is equal to 120 or μ is not equal to 120. So if it is 100 or 90 or 130 is a different case okay.

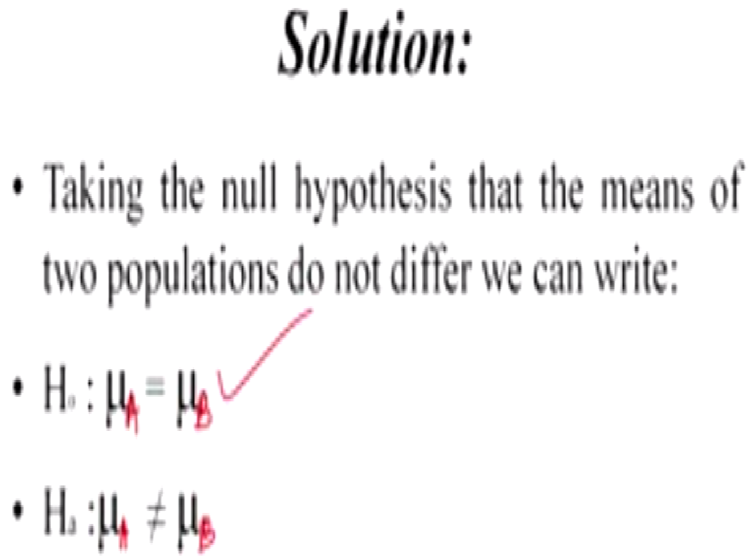
(Refer Slide Time: 25:33)



Similarly another example. Sample of sales in similar shops in 2 towns let say A and B are taken for a new product. Town A has a mean sales of this much, variance is this much. What is variance, we all you know sigma square, standard deviation square is equal to variance. So size of the sample is 5. B mean sale of 61, 4.8 and 7. Now is there any evidence of difference in sales in the 2 towns, is the sales of these 2 towns different, use a 5% level of significance. So if I am saying 5% level of significance, what is my confidence level then 95% okay.

For testing the difference between the samples of the 2 places of 2 samples for significance for testing the difference between the means of 2 samples sample A and sample B. Now kindly proceed and please develop the hypothesis.

(Refer Slide Time: 26:32)



So taking the null hypothesis that the means of the 2 populations do not differ, obviously when you do not know whether it does a difference or not, how can you claim anything. So you say there is no difference. So the first condition, so null hypothesis is that $\mu 1$ that the mean of town A, I could have said μ_A and μ_B also, is same. Similarly, otherwise I would say μ_A is not equal to μ_B okay. So either it could be less or it could be more, so when I do not know, again it is a 2-tailed test.

(Refer Slide Time: 27:07)

EXAMPLE- TWO TAILED TEST

The mean produce of wheat of a sample of 100 fields is 200 lbs. per acre with a standard deviation of 10 lbs. Another sample of 150 fields gives the mean of 220 lbs. with a standard deviation of 12 lbs. Can the two samples be considered to have been taken from the same population whose standard deviation is 11 lbs? Use 5 per cent level of significance

$n_1 = 100$, $\mu_1 = 200$, $\sigma_1 = 10$

$n_2 = 150$, $\mu_2 = 220$, $\sigma_2 = 12$

Third and I think is the last. The mean produce of what of a sample of 100 fields is 200 pounds per acre with a standard deviation of 10. So what is it saying $n_1 = 100$, $\mu_1 = 200$, standard deviation is equal to 10, standard deviation of the first one. Similarly $n_2 = 150$, the $\mu_2 = 220$, and the sigma = 12.Can the samples be considered to have been taken from the same. See why do we do this test, the test of hypothesis is done in order the main idea is to check whether the sample is actually coming from the population that we are intending to check or not.

So is the sample a part of the population or not, to check this only we do the hypothesis testing all the time right. So in this case again, I am taking 5% level of significance. So what is the solution?

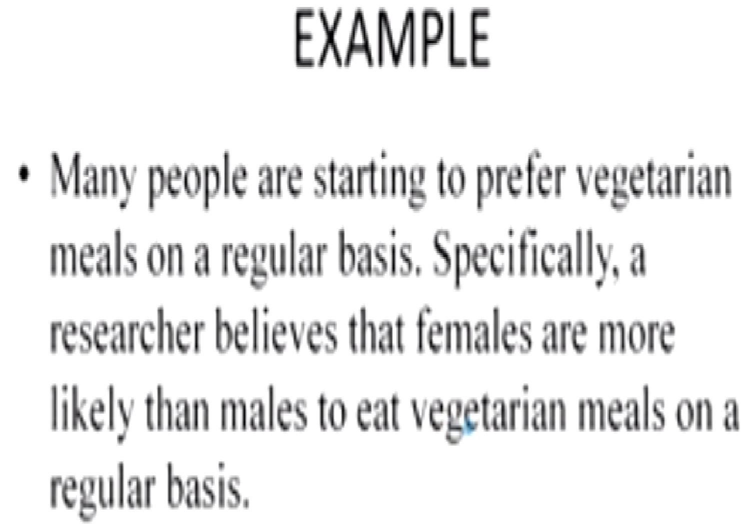
(Refer Slide Time: 28:13)

Solution

- Taking the null hypothesis that the means of two populations do not differ, we can write
- $H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 \neq \mu_2$ → Alt.

The null hypothesis that the means of 2 populations do not differ, so we write again $\mu_1 = \mu_2$ and μ_1 is not equal to μ_2 is my alternate.

(Refer Slide Time: 28:32)



This is a new example that I am starting. So this example you see this is slightly different, now what it is different and how it is different. Well it says many people are starting to prefer vegetarian meals on a regular basis. Specifically researcher believes that females are more likely than males to eat vegetarian meals on a regular basis. What is he trying to say now a days people are fed up with non-vegetarian meals and they are more preferring vegetarian meals and this is more true for females.

So he is trying to make a study on this because this will affect the kind of advertisements or kind of sales a company makes for vegetarian versus non vegetarian foods or even restaurants can make that.

(Refer Slide Time: 29:11)

- **Research Question**: *Does the data suggest that females are more likely than males to eat vegetarian meals on a regular basis?*

- **Response Variable**: Classification of whether or not a person eats vegetarian meals on a regular basis

- **Explanatory (Grouping) Variable:** Gender

So does the data suggest, this is research question right, that females are more likely than males to eat vegetarian, that mean do more number of females consume vegetarian meals in comparison to the males. It is what is the research question. What is the response variable, the classification of whether or not a person eats vegetarian meals on a regular basis, whether you are eating a vegetarian or not and what is your gender. So gender is my explanatory variable.

(Refer Slide Time: 29:42)

- **Null Hypothesis**: There is no gender effect regarding those who eat vegetarian meals on a regular basis (population percent of females who eat vegetarian meals on a regular basis = population percent of males who eat vegetarian meals on a regular basis or $p_{females} = p_{males}$)

- **Alternative Hypothesis**: Females are more likely than males to eat vegetarian meals on a regular basis (population percent of females who eat vegetarian meals on a regular basis > population percent of males who eat vegetarian meals on a regular basis or $p_{females} > p_{males}$). This is a one-sided alternative hypothesis.

Now what is the null hypothesis. There is no effect, first I am saying no, gender has no effect, there is nothing like this that females will be more inclined towards vegetarian meal than the males, there is nothing like it. So what he is saying there is no gender effect regarding those who eat the vegetarian meals on a regular basis. So the proportion of females is equal to the proportion of males, but what is my alternate. The alternate is see this is interesting, females are more likely than males to eat.

I am not interested to know that there are less females than males to eat, I am not interested, to vegetarian meals on a regular basis. So that is for this reason I am only interested that the proportion of females having vegetarian meals in comparison to the proportion of males is greater than the males, but that means, if I do it in a terms of the let us say distribution, now I will say I am not interested for this side, I am not interested for this side, I am only interested whether it is greater than or not. So this is what is the condition of a one-sided alternative or one-tailed test.

Well let me close this session here. What we discussed today is we have started developing hypothesis and also again I am reiterating developing a null hypotheses and alternate is very important because if you do it wrong, then your entire study will be completely error, so there will be lot of error in the study and it will be a wrong study. So kindly understand, have patience in learning these things and understanding how to develop the null hypothesis, what things you should keep in mind.

If you do that, it is very simple, and if you make an error, then it can be costly to a researcher. So I would suggest you take caution and do the development of the hypothesis. Thank you. We will meet in the next session. Thank you so much.