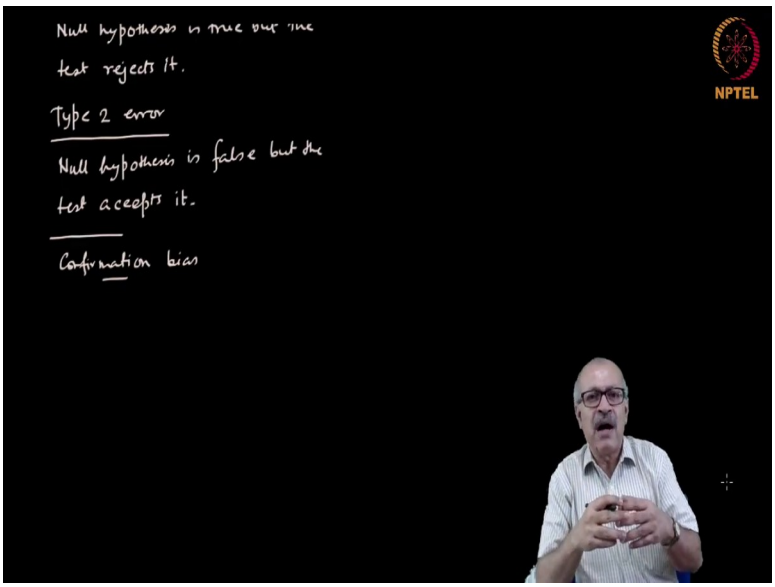
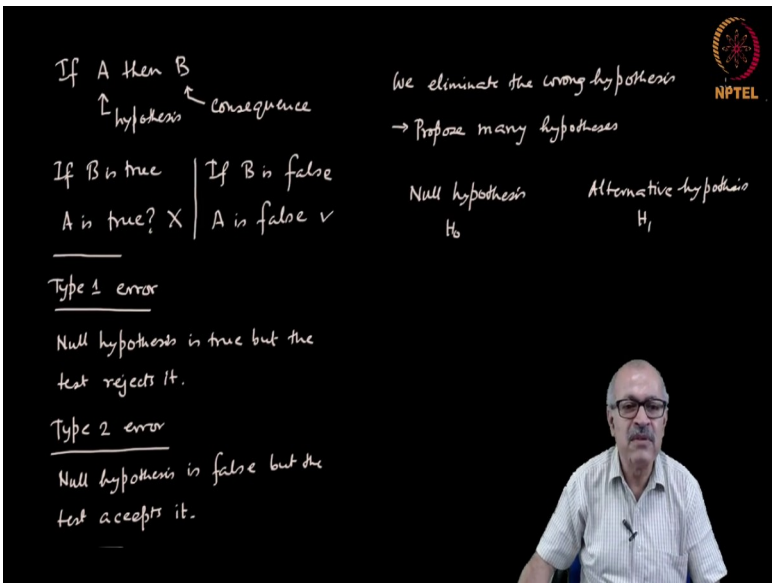
Research Methodology
Prof. Soumitro Banerjee
Department of Physical Sciences
Indian Institute of Science Education and Research, Kolkata

Lecture - 42
Issues in Hypothesis Testing Part 02

(Refer Slide Time: 00:17)

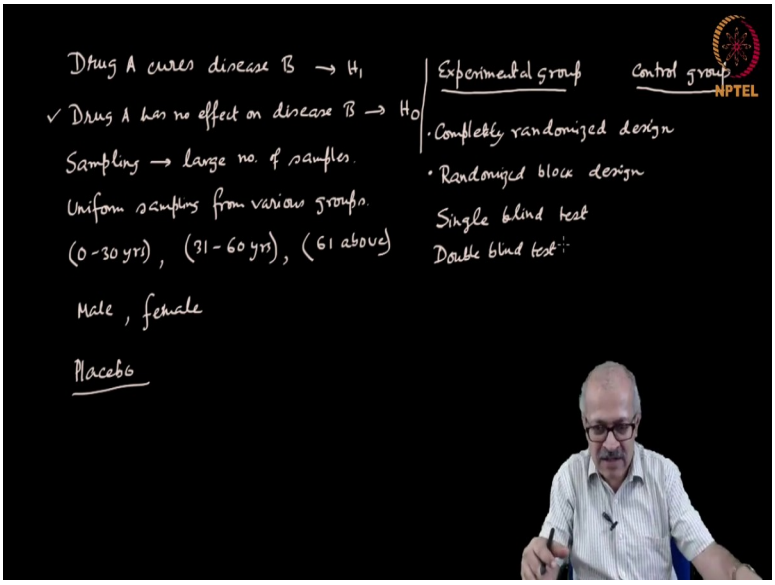


(Refer Slide Time: 00:50)



Now, for people uninitiated with the procedure of scientific testing. If we ask him or her to test a hypothesis, the kind of problem that happens, let me illustrate that with an example. We will take that example through, when we talk about hypothesis testing.

(Refer Slide Time: 01:12)



Suppose somebody has proposed a hypothesis that drug A cures disease B. You might put different things in place of drug A, different things in place of disease B. Drug A could be chloroquine, drug disease B could be malaria or any such. When you think in terms of that, you might place something that you know of. So, this is your hypothesis, the alternative hypothesis.

Now if you ask an uninitiated person, somebody who has not learnt how to do scientific experiments, if you ask how do you test, he or she will simply say that: administer this drug A on some patients of disease B. And if they get cured, it is a cure of the disease. No, it does not work. It does not work because, the people who have been afflicted with disease B; if you administer the drug A, some of them will not get cured even if the drug A is a proper drug, because of various other factors. Or if drug A is not a cure in reality, still some people will get cured because of their own body resistance or some other factors. And because of that, you cannot say by applying that on some people, depending on whether they get cured or not.

This is something that we will hear very often even from doctors: I have applied it on this person, this person, and this person, and they got cured. Therefore, I think this is the

cure. No, there has to be a scientific test for it. And this cannot be taken as cure if a few people get cured just because it is not a statistically sufficient sample. So, there has to be a statistical test done in order to test the hypothesis.

I said that whenever there is a hypothesis there is also a null hypothesis. In this case what will be the null hypothesis? 'The drug A has no affect on disease B' – that is the null hypothesis. Now, in testing a hypothesis, a scientist always has to start by believing the null hypothesis. It is a very common mistake by starting by believing the alternative hypothesis. Very common mistake. That is what normally results in confirmation bias.

So, you have to start by believing the null hypothesis and then do the test. And finally, check whether you have sufficient ground, sufficient reason, to reject the null hypothesis, and then you can embrace the alternative hypothesis. So, we start always by accepting the null hypothesis. If we have to accept something, accept the null. Then the possibility of error is less. So, we accept the null hypothesis.

And then 'accept' means, you do not announce that this is true; the acceptance is in the process of planning the experiment. Because, if you accept the other one and plan the experiment, there is a possibility of confirmation bias. Now we have to do the sampling. In sampling the first thing, obviously, is to collect a sufficient number of samples. So, a large number of samples.

'Sample' in case of this particular example that I have taken would mean, the samples of people afflicted with disease B. Now, there can be male as well as female, there can be children, young people, adults, somewhat aged, geriatric population afflicted with disease B. And it is possible that the body will react differently to the disease B and whatever causes the disease B.

And therefore, in order to do a proper sampling; you have to sample uniformly from all these possible groups. A proper sampling would mean a uniform sampling from various groups. For humans, we depend on the kind of disease, depend on our preliminary analysis of the disease. There has to be sufficient ground for doing that and you have to state that in the paper.

Nevertheless, suppose that you divide up to 30 years as one group, 31 to 60 as another group and 61 above as another group. Then if you draw 100 samples in total, they have

to be equally drawn from these three groups. And not only that, there would be male and female. So, within this group the number that you draw, that also has to be divided into male and female, this group male and female, etc., provided the disease is not sex dependent or age dependent.

There are some geriatric diseases. There is no point in drawing samples from 0 to 30 years. Or there are some female diseases; there is no point in drawing samples from the male group. So, I am not considering that. But, if you can identify that there are different types in the population; then all I am meaning is that, you have to draw equally from all these types.

So, from this type you have to draw equal number of male and female; from this type you have to draw equal number of male and female; again from here equal number of male and female. So that, when you have collected, the sample represents the variability within the population.

Now, if you apply the drug on the whole sample, all the people that have been collected as sample, then you will not be able to distinguish between the drug A being administered, and drug A not being administered. Therefore, you cannot do that. Therefore, we always divide the sample into two groups: the experimental group, and the control group. One is the experimental group, and the other is the control group.

The experimental group is one on which the experiment is actually conducted. That means, in this particular case, all the people in experimental and control group are patients of disease B. On the experimental group, you apply drug A, and do not apply on the control group, so that after some time, if you examine the result, then you will be able to distinguish the result between the two groups.

It is important: all experiments must contain the control. Because without that, you are unable to really pinpoint the effect of the experiment. This is true in all cases. Now, I am taking a particular example from the drug discovery, but this actually applies to all situations in all fields. So, whatever your field might be, you might think in terms of what the drawing of the experimental group and the control group might be.

But, then, the question that you will face is that, how do you divide the samples into the experimental group and control group? And in that, a very obvious procedure would be a

‘completely randomized design’. You design the experiment in a completely randomized way.

A completely randomized design means, you completely randomly draw and put people in the experimental group and control group. And how do you do that? The standard procedure is that you number each patient or each sample, 1, 2, 3, 4, 5, 6 you would simply number them. And then on a computer you generate that many random numbers. That means, if the number of patients is, say 100, then it is 1 to 100.

You generate random numbers. That way you can completely randomly place the experimental subjects in the experimental group or in the control group. That means, you have labeled it, and generated random numbers and use the random number generated to put the subjects in the experimental group or the control group.

There is another way. This is called the ‘randomized block design’. In this method what you actually do is that, you have drawn from different categories, this category, this category, this category and within that male female. From each category say 0 to 30 years male; within that category you do a completely randomized design. That means, you number them and then generate random numbers and put in the experimental group and control group.

Again go to 0 to 30 years females and do the same thing, 31 to 60 years male, do the same thing. That means, within each category, equal number are placed in the experimental group and in the control group. And that too randomly, without any prior judgment of the experimenter as to who to be put in the experiment group, and who to be put in the control group.

Why is this necessary? Because without that what often happen is that, normally healthy subjects are more probable to be cured and unhealthy subjects are have a larger probability of not being cured. So, if the experimenter has a prior belief that the drug A is a cure of the disease B, then, he would have a tendency of picking and choosing the healthy ones in the experimental group and relatively unhealthy ones in the control group.

This randomized design is either completely randomized design or randomized block design. These eliminate that possibility of personal bias interfering in the selection of the

experimental group and control group. And whenever there are categories within the sample, you should always use the randomized block design, so that at the end of the day, when you have drawn the experimental group and control group, they have equal number from all possible categories.

After that, you administer the drug on the experimental group and do not administer on the control group. And after some time you have to test. How should the test be? Do you go to the people, the patients, and ask them ‘how are you feeling today?’ Can that be a way? No. Because answer to the question, ‘how are you feeling today?’ would also depend on the patient’s mental state on that day.

I mean, even if he may be feeling bad, he may have got a news that elevated his mood and so he will feel that he is doing fine. Whereas, if a person is doing well, but he is psychologically depressed, then he would feel that he is not doing well. So, this cannot be a way of testing. The testing always has to be some kind of a numerical score, something that can be measured. Measured without reference to the patient, that means, without asking the patient. It is not that I ask the patient ‘what is your state of health. put it in the scale of 0 to 10’. No, not like that. You have to measure the blood pressure, you have to measure the temperature, if you draw the blood and count the number of parasites in the blood, and whatever; but something that is measurable. That is important.

Because ultimately you subject the result to a statistical test, and in order to subject any result to a statistical test, there has to be numbers. So, the results from the experimental group and the control group have to be obtained as numbers depending on what the identifier of that disease is. If it is the count of parasites in the blood, then the parasites have to be counted. It is depending on what the character of the disease is.

Now, two crucial issues. Firstly, you have taken the samples from the experimental group and the control group: blood samples, urine samples, blood pressure, heart rate, anything that is measurable depending on the disease. And finally, you have to tabulate that and you have to do a statistical test. But at this stage, two issues we have to keep in mind.

First, how a scientist’s subjective bias can influence the result. Notice that, if the scientist has to count the number of parasites or number of bacteria in blood samples or urine samples or whatever, then it is easy to commit errors. Because you are counting

microscopic particles, microscopic things, it is easy to overcount or undercount. And if you have a belief, then depending on the belief, you are likely to over count or undercount.

That means, if you believe that the drug A actually is a cure of disease B, then you would undercount the sample from the experimental group and you overcount the sample from this control group. These things happen often unknowingly, but sometimes also knowingly. In order to avoid that, what is done is, the person who actually does the counting or the person who does the measurement, if it is measurement of the heart rate or the blood pressure, then it is somebody is doing the measurement, that person does not know if that particular patient is from the experimental group or the control group. How is it done? That particular person is not the experimenter, that particular person is not privy to the information as to who is in the experimental group and who is in the control group. He does not know.

He only does the test, he or she only measures. And this kind of a test, where the experimenter does not know whether a particular subject is in the experimental group or in the control group, is called a ‘single blind test’.

So, this eliminates any possibility of experimenter bias. Experimenters can also have bias. I have already said that this is very important for the experimenter to understand that he or she might have that bias unknowingly. So, the way to eliminate that bias is that, if he is doing the measurement, then he or she should ask somebody else to label the individual patients to be either a member of the experimental group or the control group.

And this should be completely mixed. It is not that experimental group people in is one room, control group people in another room. They should be completely mixed; so, that there is no way to distinguish between them. And only the person who has labeled, he or she knows who is in which group. The person who is making the measurement does not know.

There is another possibility. After all, the patient is also a human being. He or she has a mind. And if the person knows that I have the disease B, for example I have typhoid and I have not being given any drug, then he or she might think that I am not likely to be cured. That state of mind will stand in the way of a normal cure. So, it is necessary that

the experimented subject also should not know whether he or she belongs to the experimental group or the control group.

How can we ensure that? That is ensured by giving the drug to the experimental group, but giving ‘something’ to the control group. Something that looks, feels, and tastes like the drug, but is not the drug. It is called ‘placebo’. Placebo effect is the effect where a patient, believing that he is getting a drug, often gets cured by his own psychological effect on the body. That is called the placebo effect.

The people in the control group are administered a placebo. Something that looks, feels tastes like the drug. So, the control group people would also feel that they are being given the drug, but that is actually not the drug. And when you do that, then neither the experimenter nor the experimental subject knows whether a particular subject belongs to the experimental group or the control group. Then it is called a ‘double blind test’.

In critical experiments, for example in drug discovery, the double blind test is a rule. It is a norm followed everywhere. You have to do a double blind test in order to prove the efficacy of a drug. So, these are the things that are necessary precautions in order to avoid what I just stated as confirmation bias. The experiment has to be set up in that way.

Now, I have taken a specific example of drug discovery in illustrating these necessary conditions in setting up the experiment. But you would notice that these are true in all hypothesis testing situations.

Depending on your field you should try to figure out how to do this. How to do a proper sampling, how to do a division into experimental group and control group, whether we apply the completely randomized design or a randomized block design, how do you do a single blind test, a double blind test.

How do you make yourself not know which sample I am testing? Is it from the experimental group or the control group? How do you ensure that you yourself do not know? These things have to be decided by the experimenter, and have to be very clearly specified in the paper. Then only the paper stands any chance of being accepted, because these are the standard procedures to be followed in any scientific experiment.

So, a few very important things we discussed. Things that often people do not notice. Sometimes we will find in newspapers reported that this particular drug has been shown to be effective in this and that. Immediately you should ask: how was the experiment done? Did they collect sufficient number of samples or patients? Did they properly divide into the different groups: experimental and control group? Did they really administer the drug on the experimental group, and did not administer the drug on the control group? Did they do a single blind test or a double blind test?

Unless you are satisfied that things were done in a scientific way, you should not accept the results, and this is something that very often happens. This is applicable not only in physics, but also chemistry. Not only in biology or medical science, but also the other areas of biology, other areas of physics, chemistry, everywhere. This, in general, is how the methodologies work.

For example, in case of a chemistry experiment or a physics experiment, the experimented subject is not a conscious being. And so, the question of double blind test will not appear there, because the experimented subject is some kind of inanimate object, or maybe in case of biology, dogs, cats, bacteria—these are not conscious beings. And therefore, there is no question of a double blind test. Then you have to do a single blind test. But always you have to check, in your particular field which is the scientifically accepted procedure.