Research for Marketing Decisions

Vaibhav Chawla

Department of Management Studies

Indian Institute of Technology Madras

Week - 04

Lecture - 19

Causal Research Design: Experiments and Confounding Variables

Okay, anyways, we'll get started. Good morning, everybody. Okay, so today we are going to learn about experiments. Today we are going to learn about the causal research design so so far we have been discussing we have discussed step number one of the market research process, step number two of the market research process and we are on research design. In research design, we got to know about exploratory, descriptive, and today, we are going to discuss about the causal research design. Exploratory, if you remember, we use in the problem definition step number one. Descriptive and causal we use at step three and onwards.

So today we are going to discuss the causal research design and after that we will be discussing only descriptive and particular method of descriptive which is survey method. Now experimental research design or causal research design is going to be one of the major things that you would learn. It is very, very important because going forward, you might have to do so many experiments day in day out. For example, if you are into an e-commerce company, you would have to, you know, vary the prices to know at what price the sales are getting maximum. So that you can do through experiments.

If you are into FMCG, companies like Coca-Cola, Pepsi, they do taste tests to find out what is more preferred or less preferred. That is done through experimentation. If you are into website design, all marketing companies, branding, brand managers, they are concerned about website design. You would have to do A-B testing, which means you would have to vary the page and see what gets more traction, or crowd, or stickiness, and so on, there comes the experiment. So, experiment is a skill that you will have to learn and carry forward because it is used in so many instances that there is no counting and bad experiments result in

bad inferences and therefore losses. Whereas good experiments results in good inferences and one can actually work on those results to improve the problem. Even something as simple as Let's say a packaging, a brand is there. You are just wearing the font of the packaging and you would want to know which appeals more.

You would have to conduct experiment. With the brand name, brand packaging, shape, anything if you would like to know more about which would appeal more, you can rely on experiments to conduct to derive your inferences. The cause and effect relationships are determined through the causal research design. One can arrive at wrong inferences by badly conducting causal research design or through correlation studies one could assume the cause and effect which might not be there. So if you want to determine the true cause and effect you might have to use the experiments.

I was reading yesterday a study about meditation and yoga. An experiment was conducted for a duration of two months or more. There in contrast to existing understanding that meditation and yoga would quite your ego, which means would decrease your ego, which means would decrease the self-importance, right? General understanding is meditation and yoga might decrease your self-importance, right?

In that particular study done through experimentation, reverse was found. That meditation and yoga actually would enhance your self-importance. Again, one can question those experiments, but if you want to rely on some design for cause and effect, you would have to rely on experimentation. So causal research design is the design which we are going to talk about, and experimentation is the method of determining the cause and effect. Now there are three basic conditions that are required to be met

to say one is the cause and other is the effect. There are three conditions. So, in causal research design, we have to establish one is the cause other is the effect. Cause and effect relationships. For example increase in price result in decrease in sales.

Likewise you can make many hypotheses and then you will have to conduct experiment. Cause and effect relationships. For call and relationships and effect relationships to happen, there are three conditions. What are they? Any idea?

Loud. Very good. So, one of the condition is calls should precede the effect or at least should happen simultaneously with effect. Right? So, one of the condition is calls should happen, either

cause should precede the effect or happen simultaneously with effect. In many of the, you know, chemistry experiments, you would have cause and effect very closely bound with each other that you would not distinguish what is happening first. But what happens is cause should either precede the effect or happen simultaneously with effect. It cannot happen after the effect. That is one of the condition.

What can be other condition? They cannot be manipulated. So when you manipulate the cause, only then you will be able to know the effect, right? Manipulation means you change the cause by keeping all the other possible causes constant. You change one cause to see whether there is an effect.

That is the way of conducting experiments. But there are three conditions that have to be met to say X is the cause of Y. One is X should precede Y. Very good. What can be the other condition? X should be correlated Y with Y. Which means if X changes, Y should change in either of the direction. Positive direction or negative direction.

So time occurrence is one. Second one is it is called concomitant variation or correlation, simply you can say. Why should it be correlated? Like if there is no relation, we can simply just use it again. If we have to say X is a cause of Y, so which means if X would change, something would happen to Y. Then, only then there will be correlation.

If X increases by, let us say, one standard unit, then Y should change by some units, proportionally, proportionally by some units, or fraction of the units, either in positive or negative only then we can say X is a cause of so this is again one of the condition, one is time occurrence of variables, the another one is correlation. Now, the third one is what? Randomization is not the condition, randomization is a way by which we conduct we use to conduct experiment this is one of the you know best practices that we keep in mind but that is not the third condition. X is a cause and Y is effect. So there was actually long back in you know American newspapers there was a news about a particular beach and it was noticed that

Every day there were more people getting drowned and more ice creams were getting consumed. The news reported that it is seen from last several weeks that, there are more deaths on the beach with drowning and there are more ice creams getting consumed. So there is a correlation. But can we say ice cream after eating ice cream... Ice cream is your result.

The correlation doesn't. But there is something else also. External variables. All the other external variables or all the other possible causes should be held constant. Should be controlled.

Controlled when I say should be held constant. And then you know you can see X is a cause of, Only manipulate X. All the other causes should be, possible causes should be held constant. Now in the case of ice cream, it was seen that, the sun was very bright in those days. It was quite sunny and people...

likes to come out on such days, warm sunny days. So people were coming in large numbers. When there is large number of people come, the likelihood of more number of people getting drowned will increase. Large number of people are coming, more ice cream will be getting consumed. So it was another extraneous variable.

The bright sunny day, not the ice cream, which was causing the Right? So, the third condition is absence of any other extraneous variable or simply absence of other possible causes. You have to control other possible causes and look at the one you are concerned with. Change that to see change in the Y. A way to look at this, this is one of the example of control variable is given, the third condition.

When a study was done about education and purchase of fashion clothing. Just look at this table. It was found that this is the percentage. So what can you conclude from the first table or I would say cross tab.

If you see for the high education, the difference between purchasing high fashion clothing and low fashion clothing is quite high. Whereas in the case of low education the difference is lesser than that. So, there is some chance that, we can say that, with with high education, there would be more chances of purchasing more purchasing the fashion clothing. Right?

So, one would be compelled to conclude that high education would lead to high purchase of fashion clothing. Just by looking at this possible correlation, but when income was introduced., so how to control for income income or sorry when income was introduced as an extraneous variable as a one of the possible causes so the researcher wanted to see is it really the education that leads to more purchase of fashion clothing or there is something else might be income, so they want to see whether when they introduce income as a variable, still the effect would hold if they introduce income into the picture will the effect be still there, will the education would still be leading to the purchase of

fashion clothing. So what they did was they introduced income low for low-income and high income they plotted it separately, and what was found that for high education, for the low income one, high education, and low education you see almost similar. For the high income, difference is almost similar.

So which means the effect that you are getting for this one is going away when you are introducing income the effect for high income you see the effect is more for high income there is a huge difference between for high income you see there is a huge difference between for high income for education height is 80% and education low 20 percent and for low education, it is high income sorry high purchase is 76 percent and low purchase is 24 percent so what is getting as a stark difference from the top one is when income is introduced as another extraneous variable the relationship originally which is proposed here which is proposed here would change substantially. So that is what happens a more simple example as I told you was the one when we were wrongly concluding that eating more ice creams is leading to more people getting drowned because there was another variable which was more number of which was bright sunny day which was leading to more number of people so there has to be all other extraneous all other possible causes should be held constant before we conclude about before we conclude about that X is a cause of Y. So there are three conditions that we know now, one is time occurrence X should precede Y second one is

Concomitant variation X and Y should vary together either in positive or negative direction. The third one is absence of any other possible causes which would make us conclude that X is leading to Y. So these are the conditions that we look for to establish the cause and effect relationship. In cross-sectional survey, we do not look at all of these conditions. We look at only correlation. And that is why sometimes there could be a problem that we conclude X is a cause of Y through correlation when it is not.

Because the other two conditions are not met in survey cross-sectional studies. How do you overcome that? One is you conduct the experiment. The other way is in survey, you collect data on all other possible causes as well. And then there is a way to control the effect of other variables on Y. There is a way to do that.

And the logic has to be very strong when we try to say X is a cause of Y. Actually in correlational studies, in the descriptive survey research design, we should never try to say X is a cause of Y. We can say X is positively or negatively related with Y. Although sometimes by mistake, lot of people they try to infer causality through correlation which

should not be done. If you want a causal relationship, it should come through the experiments. Let's try to look at one simple experiment and see what to do and what not to do in an experiment. We'll be reading about many experiments, but just to get started, we will look at this one.

There is a sales organization which has about 500 salespeople, pharma salespeople. And they want to know whether a particular sales training program would impact the sales performance. So they want to know there is an intervention whether a particular sales training program would increase this performance. So X and Y are very clearly given that, we want to know whether X would result in increase in Y. Now they want to conduct experiment.

So what they did was they selected a random sample of 15 salespeople. So they randomly selected 15 salespeople out of 500, and noted down their sales of last three months, took the last they noted down their sales for last three months took the average, and that observation for 15 salespeople is written as O1, At time T1 they took the observation of their sales performance. 15 sales people would have some sales performance average of previous 3 months that is written as O1. At T1 the readings are taken.

Then after taking their reading at after a week the training program is introduced. These 15 sales people were made to undergo a sales training program. Week-long sales training program skill based and knowledge-based, and then after the sales training program so which means the sample is selected random their current performance is noted down, a training program is given after the note the readings were noted. they were made to undergo a sales training program, which is also called a treatment. And then the observations are taken about their sales performance for the next 3 months. And again the average is taken at time T2 which means post 3 months the sales training program which was run for a week.

The average of after the treatment after the sales training program 3 month performance is noted down and at time T2 their average was taken. So this is observation O2. So there you have now O1 for 15 salespeople, the same salespeople you have O2. And if we do O2 minus O1, we will get change in, let us say we have taken their rupee sales. Rather than sales volume, what is written is sales volume.

We can look at sales volume by value and by numbers. By value we are looking at. Let's see. Let's assume that what we have observed is their how much money they brought in, revenue they brought in. That is O1 and O2 in terms of sales revenue they brought in.

And when you do O2 minus O1, you will get change in change in the sales performance. Now can we say O2 minus O1 whether it is positive or negative it is due to sales training program, can we conclude only other variables are constant with respect to that those are not participating in it. Very good so this is one of the ways One of the very simple experiment which is not correct, but I had to include it so that we can identify what are the problems here. But what happens in a simple experiment? You have a random sample.

You take readings about something, about the Y. You introduce an intervention. The X variable you introduce. And then after some time you note down the same Y and then look at Y2 minus Y1, whether it there is a change because of the intervention because of the X variable, that is how very simple understanding of experiments. But in this, we cannot say that that O2 minus O1 is because of whatever increase is there, let us assume that the sales have increased substantially, it is because of the sales training program.

We cannot say that. And why we cannot say that? Because there are, we did not control for extraneous variables. Now, what could be extraneous variable here? Very good.

So, it is possible that It is possible that after this sales training program, when it was going for three months, some disease, some outbreaks, not to the extent of pandemic, but let's say whatever, some outbreak happened and because of that, there were more consumption so such factors such events that happen during the course of experiment that could impact the Y variable, they are called history history they they are clubbed under the history variable or they are clubbed under the history extraneous variable history is an extraneous variable. So whenever you conduct an experiment you have to control for history. Now what is history?

History is any event that happened during the course of experiment that could impact the Y variable. It could be outbreak. It could be let's say government decrees tax on most medicines. And the prices are reduced. So any of such event that happen during the course of experiment which could impact, which could confound we say the effect of X on Y is called is the variable that we put under the history class of extraneous variables.

Now what else? So history is something which we would have taken care of. Here we did not take. What else could be problem? The number of representatives in the territory might have changed also.

They selected 15 sales people. The sample size is 500. I mean the number of workers in the career was 500. Let us say that means it has been increased or decreased. So that would have been the common.

So representativeness would have changed. That could be one of the factors but... So what he is saying is at that point of time 500 sales people were there and a representative sample, let us say we calculated sample size was 15. But then the number would have increased, changed. So the current number would not be a correct representation.

That could be one of the problem but not a very major one. If those who are participating in the study, if they know about the study, they may be intrinsically motivated to perform better. That can create bias. That is also one of the problem. So just one minute.

So this problem which he said that people who are selected, if they know the intent of the study, they may be intrinsically motivated. which is similar to another variable called main testing effect. Main testing effect we will come to that. Main testing effect which means it happens when the readings are taken pre-treatment and post-treatment and the sample would have intentionally either change their behavior because they know they are part of the experiment, but it doesn't completely fit in this case the main test but it is very similar to main testing effect which is another class of history variable but this we will this is an interesting thing which he said about the intrinsic motivation because the people who are selected

sometime when you are selected to participate in experiment, you might think that you are valued more and therefore, you are intrinsically motivated. It might be because of the intrinsic motivation not because of the sales training program. Right? So, the so what which class of variables it comes in extraneous, we will I'll tell you a little while. Anything more?

Sir, I was saying that what if there is a change in business strategy or management? Event. History. It will come same. History.

Very good example. But again it will come in history. One thing can be like they are visiting only the top hospitals while the other those who are not participating in the experiment are visiting the not so... This is an assumption, but let's hold the assumption that we have randomly selected 15. So, when we randomly select, it is possible that we might have got the people who visit not so good hospitals, not so good territories.

There would be people who would visit good territories. That is what randomization takes care of. But it is again very interesting. So one is we have so far discussed about the history. Another one is maturity.

Maturity means the sample would have gained experience over that period of time. So with experience people improve. Sales people they have gained 3 months of experience during that time, and with more experience people improve. That could be because of the experience. There will be some effect because of the experience not because of the sales training program.

So, we need to take that effect also outside. So, that we did not do. We did not control for maturity. Maturity of the sample during the course of experiment. That is another class of variable.

So, we have history. We have maturity. There is also something called mortality which means if during the if when we are conducting the experiment some someone from the sample, one or two people, they leave the experiment or they leave the company. Then we are, we do not have the same sample that we started with.

So, then the O2 minus O1, we had to take 15 because 15 by calculation is a representative set for 50, 500 sales people. But now it is 13. So, we cannot rightly conclude O2 minus O1 because this sample is The mortality was there. Three, four people, two people, they left the experiment or the company.

Hopefully they did not die because mortality means, but it is mortality in terms of experiment. Test unit mortality. So there are three. One is history, maturity, mortality. Let's say if we would have not selected them randomly.

He would have chosen the ones that the manager would have chosen the ones that he knows very well about. Then where should have been the problem? That is called selection bias. Which is what he was saying. Let's say they selected people who go to good territories.

There is a bias. Good territories, already there is lot of potential. Very good example. So there are four, history, we have to control history, we have to control maturity, we have to control mortality, we have to control selection bias.

There is another called statistical regression. Statistical regression, it is an event, it is a a process which says that people at extremes over time they are more likely to move

towards the mean, which means if for example if you would have selected the best-performing sales people in addition to bringing the selection bias, they would have also brought the statistical regression, which says that if you select so, Virat Kohli cannot score 100 in every match, the more he plays, the more likely he is, his performance, average performance, the more he plays, the more he plays, he cannot, because he cannot score 100 in each match, the more he plays, his score would start little bit coming towards the mean score of all the batsmen that play at his level. turn, third down, one down or two down or whatever.

So it says, statistical regression says that the people at the extremes over time because they cannot be more, generally it is very difficult to become more extreme rather than coming towards the middle mean. They will not come here, but from the extreme generally, it is seen that there would be a change towards the mean. but not further extreme so, if one would have selected top-performing or bottom performing sales people then the change could have been because of statistical regression so, these are the five one is history, maturity, there is one more called instrumentation, instrumentation instrumentation is something like let us say you want to see whether a particular weight training program has an impact on your weight. So you measured your weight on a particular machine, went for a weight training regime. and then came back, but now you are measuring in some other machine.

That is change in instrumentation. In this case let us say, earlier, you were measuring rupee sales, but at the end of the experiment, you are going to look at the volume by numbers, which means not by rupees but how many units you sold for different products. That will be change in instrumentation that should not happen. It could also happen because change in prices.

Change in prices could actually come in history not necessarily in instrumentation. Instrumentation happens when the unit by which you measure Y changes from what you measured pre the intervention. So these are six things that we need to control. We need to take care of when we conduct experiment because these would not allow us to conclude that X is a cause of Y. Are there more?

There are more. There are 2, 3 more problems. So in experiment there are many problems. We cannot just conclude that X is a cause of Y. There are these problems that comes up.