Supplementary Reading 3

Observational Inquiry: Causal Generalizations

1. Samples and Populations: Tulsi-Ginger Tea and Common Colds

Dialogue between **T**eacher and **S**tudent(s)

T: How many of you have designed an experiment before? [Pause] No one? Okay, so this is a good time to start. Now, how many of you have had a drink made up of *tulsi* ('holy basil'), ginger, and honey in warm water?

S1: My grandma used to give it to me when I had a cold.

S2: Me too.

T: [Looking around] From your reaction, it seems that most of you have had that experience. Now listen to a story.

As a young boy, whenever Zin had a cold, his grandmother would boil *tulsi* with ginger, add some honey, and give it to him to drink three times a day. As a grown-up, Zin continued to take the *tulsi*-ginger-honey preparation whenever he had a cold. Once, when Zin's friend Hara had a bad cold, he gave her the same preparation as a remedy.

"How do you know that it cures a cold?" Hara asked.

"That's my grandma's favourite remedy," said Zin. "I've always had it when I've had a cold."

"But does it work? Have you tested it? Is there any reliable *evidence* to show that drinking it cures the common cold?"

S3: Why is she asking him all these questions?

T: (smiling) Are those questions bothering you? You seem irritated.

S3: Teacher, Zin is just trying to help her, and she is being so aggressive. Why does she have to do that?

T: Ah, so that is the problem. You know, the questions Hara asks:

"How do you know it is true?"

"Have you tested it?"

"What is the evidence to show that it is true?"

such questions are a central part of scientific inquiry. And that's what we are learning to do, learning to think like a scientist. My question for YOU, all of you, is, suppose you were in Zin's position. What would you have said to Hara? How would you have answered her questions? How would you have tested the preparation to find out if it works?

S4: Can we do that as a project?

S3: Then we would be doing research.

T: Exactly. How about if you form research teams of 3 or 4, and investigate whether the *tusli*-ginger-honey preparation actually cures the common cold.

Ss: How do we start?

T: Good question. Your task is to come up with a research proposal, implement that proposal, and write a paper on your findings. Can you first come up with a research proposal? Each of you think about this individually, and jot down your ideas. Then discuss your ideas in your group, and consolidate them. Each group must come up with a joint research proposal, and present it in class. We'll then discuss and consolidate the group proposals. How is that?

[The groups encounter problems, and must resolve them before writing the proposal.]

1.1 Controlling for Extraneous Variables

S1: Suppose we give the *tulsi* preparation to a group of 50 people who have a cold. They all recover in five days.

S2: Are you saying that we can conclude from that that the preparation cures the common cold?

S1: Of course.

S3: No, you can't. Suppose a group of 50 people who don't take the preparation also recover in five days: obviously the *tulsi* preparation has no effect.

S4: And if the cold disappears on its own in four days, but takes seven days to disappear among those taking the preparation, we must conclude that the preparation in fact slows down the body's healing mechanism.

S2: What you are saying is, for the investigation of the cause-effect relation to be rigorous, we need two groups. I read somewhere that the group that takes the treatment is the EXPERIMENTAL GROUP, and the other one is called the CONTROL GROUP.

S3: If recovery in the experimental group is significantly better (say, in terms of time or number) than in the control group, then, and only then can the results support the HYPOTHESIS, right?

S2: That's right, if there is no difference between the two groups, then we must conclude that the hypothesis is as yet unsupported by evidence.

S5: But look, sometimes there is improvement among patients just because they have faith in the efficacy of a treatment. So for many types of illnesses, if you give fake remedies to a patient without their knowing that it is fake, the treatment can still be effective in some patients.

S4: I've heard about this. I think it's called the PLACEBO EFFECT. In testing medical treatments, it's essential to control for the placebo effect.

S1: I've heard my Dad call it a 'standard methodological protocol'. You give the treatment to a group, and fake treatment to the other group, the control group, I guess. No one who is part of the experiment knows if they are in the experimental group or the control group.

S4: Under those conditions, of course, if effectiveness in the experimental group is significantly higher, then we conclude that the difference in the effectiveness is because of the treatment.

1.2 Reliability of Observation

S7: Suppose I fall down and break my leg. I can feel the pain. The doctor can see in the X-ray that I have a fracture, and that explains the pain. But what if I have a headache? The doctor has no way of seeing it. And he tells me that I'm imagining a headache. Can he be right?

S8: Did that actually happen?

S7: Yes, more than once. So just as an observer who already believes in the efficacy of a remedy is likely to falsely perceive (greater) improvement in the experimental group, an observer who doesn't believe that there is a problem will not see it. I guess we all have our own biases. Doctors included.

S9: What you are saying is that every observation requires an observer (preferably more than one observer) who observes using the five senses, or a mechanical instrument like a microscope, thermometer, or stopwatch.

S8: In using sensory perception, like we've said, the observer's own biases can influence the observation. To eliminate this bias, it is important that the observer doesn't know who belongs to the experimental group, and who belongs to the control group.

T: Good going. To add to what you've discovered, in many experiments in medical research, neither the patient nor the observer knows who belongs to the experimental group, and who belongs to the control group. These are called *double-blind* experiments. Double-blind experiments are the norm in testing the effectiveness of medical treatments.

S8: What happens when there are different observers, and their observations don't match? So I might say that Zin is tall, and you might say, "No, he's of medium height." I might say, "Noma has a cold," and you might say, "No, she might have been crying." Such variability can be huge when we are talking about the quality of a student essay, or the quality of a candidate for a job. What do we do then?

T: In such cases, the standard practice is to get the perceptions and judgments of a large number of independent observers, and check the range of variability. All else being equal, the less the range of variability, the more reliable the judgment. This won't work if the different observers are influenced by the same source. For example, the media influences the judgments and beliefs of a wide range of people who may not even know one another. It also won't work if members of the group influence one another; for instance, students influence one another's perceptions of who is a good teacher.

S9: The general methodological principle for seems to be that:

The greater the convergence of evidence from independent sources,
the greater the reliability of the conclusion.

T: Excellent!

1.3 Sample Size

S11: My grandmother's *tulsi* remedy works for me, but it doesn't work for my sister. So the effectiveness of the treatment must show be quite a bit of variation across the population. So for any treatment, to make sure that our conclusion is reliable, how many patients do we need in our sample? How do we decide?

S10: What if we have this general methodological principle guiding sample size?

The greater the expected variability in the population,
the larger the size of the sample needs to be.

T: Any research project that involves variability in the data calls for *statistical inquiry*. So that is what you are doing now. You are learning to think like statisticians. Statistics is a way of arriving at conclusions in situations that exhibit *variability in data*.

1.4 Sampling: Biases, Representativeness, and Random Sampling

T: In any medical research project, a treatment is likely to be tested on a sample of patients from the country of the researchers. Now, a treatment that is effective for patients in a tropical region may not be effective for those in the arctic region; one that is effective for vegetarian patients may not work for non-vegetarians; one that works for certain genotypical characteristics may not work for others. So it is important that the sample be not only large, but also *representative* of the population. A sample taken exclusively from a particular country or region may not be representative of the human population.

One solution to this problem is to select the sample randomly from the whole population (*random sampling*). Another strategy is to be aware of the potential *biases* (e.g., geographical region, food habits, genes …) and include an adequate number of subjects from each group.

Most introductory books or websites on the methodological aspects of *statistical inquiry* help with understanding the issues raised above. Some excellent references are:

Richard Rowntree's *Statistics without Tears*;
Darrel Huff's *How to Lie with Statistics*;
Joel West's *Lies, Damned Lies, and Statistics*, and
Steven Levitt and Stephen Dubner's *Freakonomics*

1.5 Observing in order to Explain

As mentioned in the reading, 'Observations and Correlations,' scientific inquiry has two closely related and interacting components: observational inquiry and theoretical inquiry. OBSERVATIONAL INQUIRY makes systematic and careful observations of the world and establishes OBSERVATIONAL GENERALIZATIONS. Here are a few examples of such generalizations:

- When we drop a stone, it comes down in a straight line (not in a wavy or zig-zag path).
- As the length of a simple pendulum increases, its period (the time it takes to move from a location, and return to the original location in the same direction) increases, and vice versa.
- In a given body of gas, temperature increases as its pressure increases, and vice versa.
- Every animal has eukaryotic cells.
- There exist creatures with two, four, and six legs, but not three or seven legs.
- We find winged creatures with two legs and six legs, but not four legs.
- Smoking causes cancer.
- Industrial pollution increases global warming.
- Communalism is one of the causes of violence in human societies.

In scientific inquiry, such statements whose correctness we need to test are called HYPOTHESES. Hypotheses in science are the counterparts of conjectures in mathematics. In mathematical inquiry, a CONJECTURE is a statement that we think is likely to be true, but has not yet been proved. Once a conjecture is proved to be true, it becomes a THEOREM. Here are some mathematical conjectures that have become theorems:

- The circumference of a circle increases as its radius increases, and vice versa.
- The length of any one side of a polygon must be less than the sum of the lengths of the remaining sides.
- For every triangle, there exists exactly one circle that circumscribes it, and exactly one circle that the triangle circumscribes.

Mathematical conjectures like the ones above express correlations; they don't involve causes. Given a circle, neither the corcumference nor the radius is a cause or an effect of the other. Of the nine scientific hypotheses above, each of the first six statements is a *non-causal hypothesis*. Thus, the dropping of the stone and the straightness of its trajectory are neither the cause nor the effect of each other. In contrast, each of the last three statements (smoking causing cancer, industrial pollution causing global warming, and communalism causing violence) is a *causal hypothesis*: it is of the form, "X causes Y", where X is the cause, and Y is the effect.

Observing the world around us carefully and systematically to establish causal and non-causal observational generalizations is one of the most common forms of research in science. In this form of research (observational inquiry),

- we state an observational generalization as a hypothesis to be tested;
- plan a way to gather data that is relevant to the testing;
- gather the data; and
- arrive at a conclusion as to whether we should accept the hypothesis as correct.

Let us return to our initial *question*:

"Does drinking the *tulsi*-ginger-honey preparation cure a cold?"

This can be formulated as: Is the following statement true?

"Drinking the *tulsi*-ginger-honey preparation cures a cold."

As you can see, this is a causal hypothesis. Recall that to investigate this hypothesis, we designed an *experiment*.

Causal hypotheses are typically tested using *experimental inquiry*, one of the staple forms of observational inquiry. The activity of coming up with a research proposal to investigate the *tulsi*-ginger-honey preparation hypothesis was an introduction to *experimental inquiry*.

2 Causal Generalizations: Hotness of Chillies

Moving on to causal generalizations, a few exercises may be the best way to get a grip on them.

Miko and her 12-year old brother Jomo bought hot chilly pepper seeds, and planted some under a tree in their backyard. When the chillies were ready to be eaten, Miko was disappointed because they were not hot. Why were the chillies not hot? Jomo's ***hypothesis*** was that the plants were not watered enough. Miko disagreed. Her hypothesis was that the plants didn't get enough sun. They decided to conduct an experiment to find out who was right.

2.1 *Design an experiment that would tell us whose hypothesis was right.*

1. They were both right.
2. They were both wrong.
3. Jomo was right, and Miko was wrong.
4. Miko was right, and Jomo was wrong.

Describe the experiment, spelling out what kinds of experimental results would support each of the conclusions in (1)-(4) above. NOTE: Don't read further till you have done this exercise.

2.2 Miko and Jomo have two hypotheses between them:

A. If chilly plants do not have adequate water, the chillies won't be hot.
B. If chilly plants do not have adequate sunlight, the chillies won't be hot.

Each of these hypotheses could be either true or false. So, Miko and Jomo need to conduct two experiments, one to test A, and another to text B.

In hypothesis A, there are two VARIABLES, namely, *water* and *hotness* of chillies. According to Jomo's hypothesis, the hotness of the chillies is DEPENDENT on the amount of water that the plant receives. To design an experiment to test this, we need two groups of chilly plants that have the same treatment in all respects except the quantity of water they receive. In Group 1, for instance, the plants receive adequate amounts of water (the EXPERIMENTAL GROUP). In Group 2, the plants receive an inadequate amount of water (the CONTROL GROUP). If the chillies from Group 1 are consistently hotter than those from Group 2, hypothesis A is correct. If not, A is wrong.

Likewise, to test hypothesis B, we need an experiment in which two groups of chilly plants have the same treatment in all respects except that of sunlight. In the experimental group (Group 3), the plants receive adequate amounts of sunlight, but in the control group (Group 4) they receive an inadequate amount of sunlight. If the chillies from Group 3 are consistently hotter than those from group 4, hypothesis B is correct. If not, B is wrong.

One way to conceptualize a combination of the two experiments to test hypotheses A and B is to use what is called a 'two-by-two matrix':

water \ sunlight	Adequate	Inadequate
Adequate	GROUP 1	GROUP 3
Inadequate	GROUP 2	GROUP 4

We now have four possibilities (P): P-1: A is right, B is wrong. P-3: A and B are both right.
P-2: A is wrong, B is right. P-4: A and B are both wrong.

For each of the potential results below, select the conclusion you would accept.

***Results**: Groups 1 to 4*	***Conclusion***			
1, 3 systematically yield hotter chillies than 2, 4	P-1	P-2	P-3	P-4
1, 2 yield hotter chillies than 3 and 4	P-1	P-2	P-3	P-4
1, 2, 3 yield hotter chillies than 4	P-1	P-2	P-3	P-4
1, 2, 3, 4 yield equally non-hot chillies	P-1	P-2	P-3	P-4

Answers:

***Expected Results**: Groups 1 to 4*	***Conclusion***			
1, 3 systematically yield hotter chillies than 2, 4	P-1			
1, 2 yield hotter chillies than 3 and 4		P-2		
1, 2, 3 yield hotter chillies than 4			P-3	
1, 2, 3, 4 yield equally non-hot chillies				P-4

2.3 *What would you conclude if groups 2, 3, 4 are equally non-hot, and group 1 is hot?*
[NOTE: Read further only after thinking about and working out the answer.]

The result shows that the chillies are hot only when both conditions (adequate sunlight AND adequate water) are met. If the children's claims are:

Miko: Water is the only causal factor (i.e., sunlight is not causal factor)

Jomo: Sunlight is the only causal factor (i.e., water is not a causal factor)

Then the result would show that they are both wrong. However, if their claims are:

Miko: Water is a causal factor (whether or not sunlight is also a causal factor)

Jomo: Sunlight is a causal factor (whether or not water is also a causal factor)

They are both right.

Learning point: To maximize the benefits from our inquiry, it is important to formulate our hypothesis clearly and precisely.

2.4 *Imagine that the experiment used only a single chilli seed for each of the groups 1-4. Can this give rise to a problem? [NOTE: Read further only after thinking seriously about the answer.]*

The seed used for Group 1 (not Groups 2-4) could have come from a plant that has genes for very hot chillies. If so, the variable relevant here is the genes, not the environment. Our conclusions about the CORRELATION between hotness of chillies, the amount of sunlight, and the amount of water would then be wrong.

How would we eliminate the effects of such a variable?

The safest way to eliminate the effects of genes in this investigation is to use a sufficiently LARGE SAMPLE of seeds (say, around 50 seeds) randomly selected from the same batch.

Another would be to use seeds of chillies from the same plant.

2.5 *In your design, did you consider the nature of the soil as a possible factor in the non-hotness of the chillies?* Suppose the soil for A contained nutrients that increased the hotness of the chillies. If so, our conclusion about the relevant correlates of hotness would be unjustified, because the relevant variable is the soil, not water or sunlight. *How would we eliminate the effects of this variable?* One way would be to use the same soil for all the groups.

2.6 Suppose the design of your experiment has taken care to avoid all the potential sources of error and distortion we have looked at so far. After the plants yield chillies, both Jomo and Miko taste the chillies. Jomo finds the chillies from groups 1 and 3 to be hotter than those from groups 2 and 4. Miko finds the chillies from groups 1 and 2 to be hotter than those from groups 3 and 4. They are surprised that their findings are contradictory, and conclude that the subjective nature of their tasting the chillies, combined with each researcher's eagerness to support his/her own hypothesis, could have resulted in the differences in the observational reports.

Have you encountered this problem of 'subjective judgment' before? Can you design a satisfactory strategy to avoid the problem?