20.

Experiments

Experiments are motivated by our desire to know causation combined with the fact that we typically only have correlations. The cause of a correlation may be the two variables observed to be correlated (X, Y) or to a 3rd variable. If we can control for all relevant third variables in our data, we can ask if the X-Y correlation still holds, and if we have excluded the relevant third variables, then there is a good chance the correlation is due to causation by X or Y.

Approaches to controlling for third variables

There are two approaches when trying to control for third variables. The second is the surest way, but we cannot always carry it out.

A. Sort through the data that nature has provided - these data are correlational. Identify the 3rd variables and use data in which the treatment and control groups are matched for the variables.

B. Gather a new type of data: do an experiment. You gather data in a setting where nature has been altered so that the 3rd variable(s) is controlled a manipulation of nature. You are altering the way that the data are created.

We say that the data used in (A) are correlational; those in (B) are experimental.

Each method has its strengths and weaknesses. The upside of correlational data is that the outcomes of interest have usually happened before you think up the study, and the data may even be available in some fashion before you do the study. They are the only options for studies that you could not do experimentally. The downside of correlational data is that you need to have guessed all the relevant third variables, and then you need to be able to get the data allowing you to control for them. There are many situations in which you can never know if you guessed correctly, so you are never confident. The issues with experiments are different. Experiments offer the benefit of ideal controls that destroy unwanted correlations from hidden

variables. Their downsides are two. First, you cannot apply the treatment ahead of time (except in rare circumstances, as with some accidents). Second, there are many experiments you cannot do, for example because they involve unacceptable manipulations of humans, the scale is too large, or critical events lie in the past.

With experiments, all the ideal data issues still apply: blind, standards, replicates, random. To this list, we now add 'controls.'

Types of experiments

We will distinguish two types of experiments based on whether the third variables are known in advance:

1. **Variables unknown** You do not know the 3rd variables. randomly assign treatment and control groups; 3rd pre-existing variables are not eliminated but they become balanced. This is the approach used in clinical trails: patients are randomly assigned to control vs. treatment group, so all pre-existing variables are distributed equally (on average) between them. Sampling error is of course a possible issue in achieving a true balance of third variables across both groups when sample sizes are small. In our red car example, you would randomly change car color after the car was owned and the driver was in place, and you would assess accident rate immediately, before the owner could sell the car or change any of the other possible variables. By changing car color, it becomes an experiment.

To illustrate experiments of type (1), consider the potential problems with correlational data on accident rate and car color:
The first table illustrates the type of imbalance that may occur. An experiment to destroy these variables would involve randomly changing the color of every car while leaving everything else intact (driver, car type, car age, ...). (We of course could not do such an experiment, at least not

Table 1: Correlational data in which third variables are not controlled

red cars	non-red cars
sporty cars	safe cars
old cars	young cars
risky drivers, showy personalities	safe drivers, dull personalities
drive in accident-prone conditions	drive at safe times

easily.) By randomly assigning color, any and all pre-existing 3rd variables become distributed evenly, as shown in the second table:

Table 2: Correlational data in which third variables are controlled

red cars	non-red cars
sporty + safe cars	sporty + safe cars
old + young cars	old + young cars
risky drivers + safe drivers	risky + safe drivers
showy personalities + dull personalities	showy personalities + dull personalities
drive in accident-prone and safe conditions	drive in accident-prone and safe conditions

Note that this shuffling of third variables works to the extent that the third variables do not change when the color of the car has been randomized; they are pre-existing and stay the same. If changing car color unavoidably altered a variable, then color would be considered causal. For example, if color affected driver aggression, that would change when we changed car color, and color would have to be considered causal.

Other contexts are easy to provide. In a drug trial, you would randomly assign whether a participant received the drug or placebo. Letting people choose whether to take the drug would not control for 3rd variables. In testing whether exam color affects scores, you would randomly assign individuals to exam color. Letting people choose color would not control for 3rd variables.

2. **Variables known** In this second type of experiment, the potential confounding variables are known in advance. The experiment is set up so that one of the relevant variables is controlled, and the data are gathered under those conditions. The control may involve

eliminating the third variable, or it could involve randomization.

As one example, imagine that a car runs better after new plugs and higher octane fuel. Is the cause the plugs or fuel? You could do an experiment to understand the cause. The variables are plugs and fuel, because both changed between now and before (Table **??**). To control for plugs while varying fuel, you could use the old plugs and compare mileage when using the two types of fuel (compare cells (1) and (2) in the table) or you could use the new plugs and compare mileage when using the two types of cells (compare cells (3) and (4) in the table). The examples below will illustrate this second type of experiment.

Table 3: Example of correlation in which relevant variables are known

		Fuel	
		old	new
plugs	old	1. poor mileage	2. (expt)
	new	3. (expt)	4. good mileage

Overview. The following table (**??**) summarizes the relationship between the different types of experiments and the extent to which third variables are controlled.

Table 4: Summary of correlations and experiments regarding controls

	Correlational data	Experimental data/manipulation	
How 3rd variables can be controlled	(correctly) sorting the data by third variables	before data are generated, random assignment to treatment vs. control	before the data are generated, fixing the relevant variable
Knowledge of the third variable required?	yes	no	yes
Success in inferring causation assured?	no	yes	only for variables tested

Experiments with Psychics (based on a Nova video shown in class)

The video (excerpts from the Nova 'Secrets of the Psychics") showed or described three different experiments.

Responses to Horoscopes

The first experiment was a horoscope test in which a class of maybe 30 students were led to believe that they had been given personalized horoscopes based on their birth dates and other data. When asked to rate the accuracy of the horoscopes on a scale of 1-5 (5 being the most accurate), the class gave scores of 4 and 5.

The study was an experiment. The students had all be given the same horoscope, but they were unaware of this fact. A similar type of experiment has been done in Bio301D every year as well, cast in the guise of a personality-description. 70% - 75% of the students think that the descriptions ranked from 3-5 on a scale of -5 to 5. Upon discovering the truth of the experiment, people openly admit that they would not have scored it so highly had they known there was no personalization.

To analyze this experiment, two models are being compared and tested:

model 1 peoples responses reflect personalized accuracy of a horoscope

model 2 people's responses reflect expectations of personalized accuracy

In reference to an earlier lecture, we are not using a null model here. We use the equal alternatives approach.

What was the manipulation? Everyone got same horoscope/description descriptions were not personalized. In the normal setting, descriptions are individualized. By 'blinding' the people

who are rating the horoscope (personality description) this experiment controls for expectations – so the expectations are the same as in the normal use of horoscopes. A breakdown of the ideal data features is as follows.

manipulation	everyone got same horoscope
replication:	yes (30 students)
randomization	not relevant all students got same horoscope, so there is no opportunity to randomize
blind	yes, students unaware that all were same necessary to control for expectation
control?	yes by implication of a similar reaction to horoscopes done in the usual fashion
explicit protocol	yes, from the fact that all horoscopes were the same

One assumption in this experiment is that the participants would normally have been given different horoscopes – that they would not all be the same. For example, suppose the horoscope given to everyone was a Pisces horoscope and entire class was Pisces. Then the study would not constitute a manipulation.

The study suggests that the response to horoscopes is not based on personalized accuracy, since we got a favorable response when there was none; model (2) is supported. This does not rule out the possibility that horoscopes do have some accuracy, only that such accuracy is not a requisite for a positive response. Furthermore, (and as suggested in the video), we cannot ever fully rule out that horoscopes have SOME predictive value. There are too many such models to exclude all possibilities, and ruling then all out would be the same as proving that they had no predictive value – which is impossible.

How might we test the the model that horoscope descriptions are no better than random? That model was not tested in the video, but we can imagine how it might be done. One approach would be to generate horoscopes in the usual fashion, and then randomly assign them to individuals (the treatment group) versus give them to the intended recipient (the control group). If the participants were blinded as to what group they belonged, a move favorable response in the

control group would indicate that horoscopes were somehow tailored appropriately to the subject. Alternatively, the different parts of the horoscopes might be randomized in the treatment group to see if that affected responses compared to the control group – that got horoscopes done in the usual fashion. The variable controlled here would again be expectations.

Client Responses to Palm reading

The second experiment described was one of palm reading. In this case, however, the experiment was a case of doing palm readings in two ways: (i) readings done by the book (the control), and (ii) readings done the 'opposite' of the book. The models implicitly tested were similar to those of the horoscope study:

model 1: peoples responses reflect personalized accuracy of a palm reading

model 2: people's responses reflect expectations of personalized accuracy of a palm reading

A breakdown of the ideal data features is as follows:

manipulation	client told opposite of what the book said
replication:	yes (at least 2 clients)
randomization	no (but potentially relevant)
blind	yes, one way. could do a double-blind, if a third party told the palm-reader what to say
control?	yes; people whose palms were read correctly
explicit protocol	yes

Accuracy of picture psychics

The third experiment shown involved measuring the accuracy of supposed psychics who claimed to be able to identify characteristics of an individual from a photograph. This was a manipulation of the natural way of doing things only because the psychics were given no feedback

from their input (until they had completed their description); so no information could be conveyed about how well they were doing. The subject in the photo was known to the experimenter, so the accuracy of the predictions could be checked afterward.

The model being tested here is different than those above. Here a null model is used, that the predictions are no better than random. What constitutes 'random' is not especially clear, but the show gives at least a vague idea, and one could design better experiments to get at randomness.

The main points of this video are

I. it is indeed possible to apply the SM to psychic phenomena.

II When the predictions of psychic methods are vague and/or numerous, we cannot hope to refute all of them. Instead, we use a null model approach to ask if the psychic predictions do better than random or better than some other model that assumes no special powers.

Autism and Facilitated Communication (based on a Frontline video shown in class)

Thousands of children in the U.S. (and in other countries) are born with a severe mental handicap known as autism. No doubt, the classification of autism includes many different specific disabilities, but autistic people are unable to communicate. As a consequence, no one really knows if they are able to learn.

Two decades ago, a revolutionary new method of dealing with autistic children was developed. First used in Australia, it was rapidly adopted by many institutions in the U.S. It is known as Facilitated Communication (FC, for short), and the principle idea behind it is that autistic children can learn and understand but just cannot communicate what has been learned.

FC is a tool to help the person communicate.

FC works like this. The subject (someone with autism) is placed in front of a small keyboard. Next to subject sits an adult (the facilitator), who merely holds the autistic person's arm above the keyboard, so that a finger hangs down and can hit keys one at a time. The keyboard is connected to a printer that prints the letters being typed. In theory, these letters (words) represent the autistic person's thoughts.

When first introduced, FC seemed the most remarkable technology in the history of autism. Autistic people changed from being perceived as severely retarded to being perceived as of normal intelligence, merely needing help to communicate. Social workers were delighted, as were the parents of these children.

Everything with FC seemed fine. Not everyone was convinced that it really worked, but it seemed harmless at best. And it was used for more than just taking tests or learning material – the Facilitators would have conversations with the child. Then, however, things took a turn for the worse. Some of the FC transcripts during conversations graphically indicated sexual abuse. When these transcripts appeared, children were yanked from their parents and put into foster care. Families were being separated by the state because of the typed words. FC was no longer harmless. If the typed words were true, FC was helping to free kids from desperate situations. If the typed words were not true, however, FC was destroying families.

The correlation that everyone considered to be causal

The seductive nature of FC was a striking correlation between what was asked of the child and what words were typed. Many assumed it was causal – that the child was 'causing' the words. Had there not been such a good correlation between the desired answer and the answer obtained, there might have been much less interest in FC. What was apparently ignored in all of this was the possibility that the correlation was due to a third variable – what was asked of the

Facilitator. In the normal FC setting, what was asked of the child was exactly the same of what was asked of the Facilitator – they were given the same information. It was thus ambiguous whether the answers were due to the child or the Facilitator. With the allegations of abuse resulting in legal ramifications, it became necessary to find out who caused the typed answers. This test was not quite the same as deciding if the typed words were true, because even if the words came from the child, they might not be true. But if the words were not coming from the child, then they most certainly were not true.

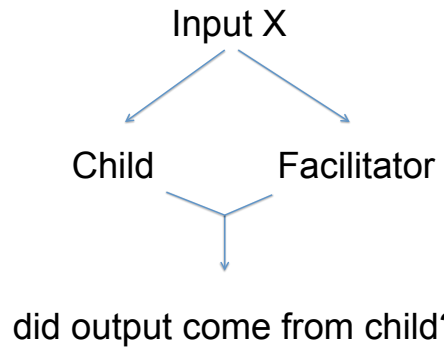

Figure 1:
The normal FC protocol. Both child and facilitator are exposed to the same material, so one cannot tell if

The test to decide who was authoring the words was simple in hindsight: the child and facilitator were shown different objects and the team was then asked to describe what they saw. It would become immediately clear who was typing the words.

FC was a fraud. The test proved decisive: the descriptions were always of what the facilitator saw. Thus the facilitator was responsible for the typing, and any allegations implied about abuse to the child could not have come from the child.

The kind of test used to debunk FC is an experiment. The correlation that had led to widespread acceptance of FC was that the words and answers solicited from the child often matched what was being asked – we got the expected answers. The 3rd variable was that the

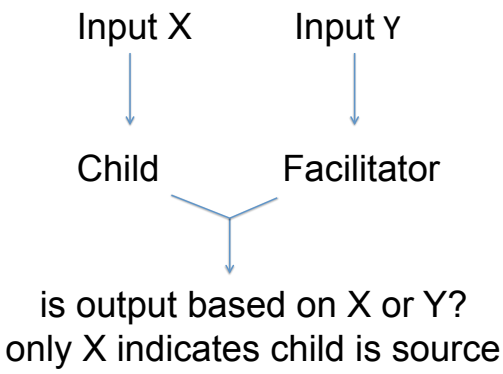

Figure 2:

facilitator also knew what was expected; the way FC was typically done, we had no way of knowing whether the words came from the facilitator or child. The experiment manipulated the normal mode of FC by controlling for this third variable. We wanted trials in which the inquiry of the child was varied but that of the Facilitator was not, requiring that they were shown different information. By deviating from this "natural order" of things in a specific way, it was easy to find out that the child was not typing the words. More specifically, this is an experiment in which we know the relevant third variable in advance and can control for it (Facilitator information) – it is not an experiment in which 3rd variables are controlled by randomization.

The design features illustrated were as follows.

manipulation	Facilitator and child shown different objects
replication:	yes (at 3 levels – multiple locations, multiple types of test, multiple Facilitators and children, multiple attempts to measure communication within a test
randomization	not obviously (could have used random assignment of the order of pictures presented, but probably not needed)
blind	yes, neither the Facilitator nor the child knew what the other was seeing
control?	yes; cases in which the Facilitator and child were shown the same object. This controls shows that the testing environment is working.
explicit protocol	yes

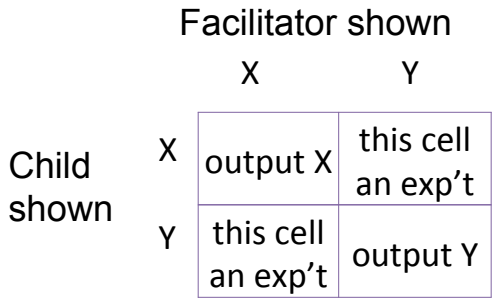

Figure 3:
Interpretation of the FC experiment. The table cells describe the output (words typed) from FC. The variables are the different objects shown to the facilitator and child. The problem is that, in the normal FC environment, both facilitator and child are shown the same thing (upper left and lower right quadrants), so we don't know who is responsible for the output. We need to control for what is shown to the facilitator while varying what is shown the child, or vice versa (upper right and lower left quadrants). These are experiments, because they are deliberate alterations of the normal FC environment, with the purpose of understanding causation.