

1.5: Confounding Variables

Learning Objectives

- A confounding variable is a variable that may affect the dependent variable. This can lead to erroneous conclusions about the relationship between the independent and dependent variables. You deal with confounding variables by controlling them; by matching; by randomizing; or by statistical control.

Due to a variety of genetic, developmental, and environmental factors, no two organisms, no two tissue samples, no two cells are exactly alike. This means that when you design an experiment with samples that differ in independent variable X , your samples will also differ in other variables that you may or may not be aware of. If these confounding variables affect the dependent variable Y that you're interested in, they may trick you into thinking there's a relationship between X and Y when there really isn't. Or, the confounding variables may cause so much variation in Y that it's hard to detect a real relationship between X and Y when there is one.



Fig. 1.5.1: American elm, *Ulmus americana*.

As an example of confounding variables, imagine that you want to know whether the genetic differences between American elms (which are susceptible to Dutch elm disease) and Princeton elms (a strain of American elms that is resistant to Dutch elm disease) cause a difference in the amount of insect damage to their leaves. You look around your area, find 20 American elms and 20 Princeton elms, pick 50 leaves from each, and measure the area of each leaf that was eaten by insects. Imagine that you find significantly more insect damage on the Princeton elms than on the American elms (I have no idea if this is true).

It could be that the genetic difference between the types of elm directly causes the difference in the amount of insect damage, which is what you were looking for. However, there are likely to be some important confounding variables. For example, many American elms are many decades old, while the Princeton strain of elms was made commercially available only recently and so any Princeton elms you find are probably only a few years old. American elms are often treated with fungicide to prevent Dutch elm disease, while this wouldn't be necessary for Princeton elms. American elms in some settings (parks, streetsides, the few remaining in forests) may receive relatively little care, while Princeton elms are expensive and are likely planted by elm fanatics who take good care of them (fertilizing, watering, pruning, etc.). It is easy to imagine that any difference in insect damage between American and Princeton elms could be caused, not by the genetic differences between the strains, but by a confounding variable: age, fungicide treatment, fertilizer, water, pruning, or something else. If you conclude that Princeton elms have more insect damage because of the genetic difference between the strains, when in reality it's because the Princeton elms in your sample were younger, you will look like an idiot to all of your fellow elm scientists as soon as they figure out your mistake.

On the other hand, let's say you're not *that* much of an idiot, and you make sure your sample of Princeton elms has the same average age as your sample of American elms. There's still a lot of variation in ages among the individual trees in each sample, and if that affects insect damage, there will be a lot of variation among individual trees in the amount of insect damage. This will make it harder to find a statistically significant difference in insect damage between the two strains of elms, and you might miss out on finding a small but exciting difference in insect damage between the strains.

Controlling confounding variables

Designing an experiment to eliminate differences due to confounding variables is critically important. One way is to control a possible confounding variable, meaning you keep it identical for all the individuals. For example, you could plant a bunch of American elms and a bunch of Princeton elms all at the same time, so they'd be the same age. You could plant them in the same field, and give them all the same amount of water and fertilizer.

It is easy to control many of the possible confounding variables in laboratory experiments on model organisms. All of your mice, or rats, or *Drosophila* will be the same age, the same sex, and the same inbred genetic strain. They will grow up in the same kind of containers, eating the same food and drinking the same water. But there are always some possible confounding variables that you can't control. Your organisms may all be from the same genetic strain, but new mutations will mean that there are still some genetic differences among them. You may give them all the same food and water, but some may eat or drink a little more than others. After controlling all of the variables that you can, it is important to deal with any other confounding variables by randomizing, matching or statistical control.

Controlling confounding variables is harder with organisms that live outside the laboratory. Those elm trees that you planted in the same field? Different parts of the field may have different soil types, different water percolation rates, different proximity to roads, houses and other woods, and different wind patterns. And if your experimental organisms are humans, there are a lot of confounding variables that are impossible to control.

Randomizing

Once you've designed your experiment to control as many confounding variables as possible, you need to randomize your samples to make sure that they don't differ in the confounding variables that you can't control. For example, let's say you're going to make 20 mice wear sunglasses and leave 20 mice without glasses, to see if sunglasses help prevent cataracts. You shouldn't reach into a bucket of 40 mice, grab the first 20 you catch and put sunglasses on them. The first 20 mice you catch might be easier to catch because they're the slowest, the tamest, or the ones with the longest tails; or you might subconsciously pick out the fattest mice or the cutest mice. I don't know whether having your sunglass-wearing mice be slower, tamer, with longer tails, fatter, or cuter would make them more or less susceptible to cataracts, but you don't know either. You don't want to find a difference in cataracts between the sunglass-wearing and non-sunglass-wearing mice, then have to worry that maybe it's the extra fat or longer tails, not the sunglasses, that caused the difference. So you should randomly assign the mice to the different treatment groups. You could give each mouse an ID number and have a computer randomly assign them to the two groups, or you could just flip a coin each time you pull a mouse out of your bucket of mice.

In the mouse example, you used all 40 of your mice for the experiment. Often, you will sample a small number of observations from a much larger population, and it's important that it be a random sample. In a random sample, each individual has an equal probability of being sampled. To get a random sample of 50 elm trees from a forest with 700 elm trees, you could figure out where each of the 700 elm trees is, give each one an ID number, write the numbers on 700 slips of paper, put the slips of paper in a hat, and randomly draw out 50 (or have a computer randomly choose 50, if you're too lazy to fill out 700 slips of paper or don't own a hat).

You need to be careful to make sure that your sample is truly random. I started to write "Or an easier way to randomly sample 50 elm trees would be to randomly pick 50 locations in the forest by having a computer randomly choose GPS coordinates, then sample the elm tree nearest each random location." However, this would have been a mistake; an elm tree that was far away from other elm trees would almost certainly be the closest to one of your random locations, but you'd be unlikely to sample an elm tree in the middle of a dense bunch of elm trees. It's pretty easy to imagine that proximity to other elm trees would affect insect damage (or just about anything else you'd want to measure on elm trees), so I almost designed a stupid experiment for you.

A random sample is one in which all members of a population have an equal probability of being sampled. If you're measuring fluorescence inside kidney cells, this means that all points inside a cell, and all the cells in a kidney, and all the kidneys in all the individuals of a species, would have an equal chance of being sampled.

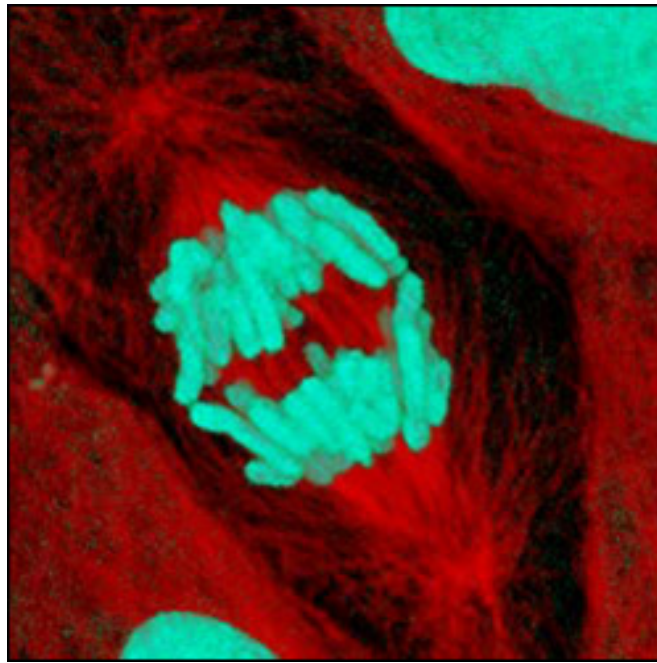


Fig. 1.5.2: Confocal microscope image of a dividing kidney cell.

A perfectly random sample of observations is difficult to collect, and you need to think about how this might affect your results. Let's say you've used a confocal microscope to take a two-dimensional "optical slice" of a kidney cell. It would be easy to use a random-number generator on a computer to pick out some random pixels in the image, and you could then use the fluorescence in those pixels as your sample. However, if your slice was near the cell membrane, your "random" sample would not include any points deep inside the cell. If your slice was right through the middle of the cell, however, points deep inside the cell would be over-represented in your sample. You might get a fancier microscope, so you could look at a random sample of the "voxels" (three-dimensional pixels) throughout the volume of the cell. But what would you do about voxels right at the surface of the cell? Including them in your sample would be a mistake, because they might include some of the cell membrane and extracellular space, but excluding them would mean that points near the cell membrane are under-represented in your sample.

Matching

Sometimes there's a lot of variation in confounding variables that you can't control; even if you randomize, the large variation in confounding variables may cause so much variation in your dependent variable that it would be hard to detect a difference caused by the independent variable that you're interested in. This is particularly true for humans. Let's say you want to test catnip oil as a mosquito repellent. If you were testing it on rats, you would get a bunch of rats of the same age and sex and inbred genetic strain, apply catnip oil to half of them, then put them in a mosquito-filled room for a set period of time and count the number of mosquito bites. This would be a nice, well-controlled experiment, and with a moderate number of rats you could see whether the catnip oil caused even a small change in the number of mosquito bites. But if you wanted to test the catnip oil on humans going about their everyday life, you couldn't get a bunch of humans of the same "inbred genetic strain," it would be hard to get a bunch of people all of the same age and sex, and the people would differ greatly in where they lived, how much time they spent outside, the scented perfumes, soaps, deodorants, and laundry detergents they used, and whatever else it is that makes mosquitoes ignore some people and eat others up. The very large variation in number of mosquito bites among people would mean that if the catnip oil had a small effect, you'd need a huge number of people for the difference to be statistically significant.

One way to reduce the noise due to confounding variables is by matching. You generally do this when the independent variable is a nominal variable with two values, such as "drug" vs. "placebo." You make observations in pairs, one for each value of the independent variable, that are as similar as possible in the confounding variables. The pairs could be different parts of the same people. For example, you could test your catnip oil by having people put catnip oil on one arm and placebo oil on the other arm. The variation in the size of the *difference* between the two arms on each person will be a lot smaller than the variation among different people, so you won't need nearly as big a sample size to detect a small difference in mosquito bites between catnip oil and placebo oil. Of course, you'd have to randomly choose which arm to put the catnip oil on.

Other ways of pairing include before-and-after experiments. You could count the number of mosquito bites in one week, then have people use catnip oil and see if the number of mosquito bites for each person went down. With this kind of experiment, it's important to make sure that the dependent variable wouldn't have changed by itself (maybe the weather changed and the mosquitoes stopped biting), so it would be better to use placebo oil one week and catnip oil another week, and randomly choose for each person whether the catnip oil or placebo oil was first.

For many human experiments, you'll need to match two different people, because you can't test both the treatment and the control on the same person. For example, let's say you've given up on catnip oil as a mosquito repellent and are going to test it on humans as a cataract preventer. You're going to get a bunch of people, have half of them take a catnip-oil pill and half take a placebo pill for five years, then compare the lens opacity in the two groups. Here the goal is to make each pair of people be as similar as possible in confounding variables that you think might be important. If you're studying cataracts, you'd want to match people based on known risk factors for cataracts: age, amount of time outdoors, use of sunglasses, blood pressure. Of course, once you have a matched pair of individuals, you'd want to randomly choose which one gets the catnip oil and which one gets the placebo. You wouldn't be able to find perfectly matching pairs of individuals, but the better the match, the easier it will be to detect a difference due to the catnip-oil pills.

One kind of matching that is often used in epidemiology is the case-control study. "Cases" are people with some disease or condition, and each is matched with one or more controls. Each control is generally the same sex and as similar in other factors (age, ethnicity, occupation, income) as practical. The cases and controls are then compared to see whether there are consistent differences between them. For example, if you wanted to know whether smoking marijuana caused or prevented cataracts, you could find a bunch of people with cataracts. You'd then find a control for each person who was similar in the known risk factors for cataracts (age, time outdoors, blood pressure, diabetes, steroid use). Then you would ask the cataract cases and the non-cataract controls how much weed they'd smoked.

If it's hard to find cases and easy to find controls, a case-control study may include two or more controls for each case. This gives somewhat more statistical power.

Statistical control

When it isn't practical to keep all the possible confounding variables constant, another solution is to statistically control them. Sometimes you can do this with a simple ratio. If you're interested in the effect of weight on cataracts, height would be a confounding variable, because taller people tend to weigh more. Using the body mass index (BMI), which is the ratio of weight in kilograms over the squared height in meters, would remove much of the confounding effects of height in your study. If you need to remove the effects of multiple confounding variables, there are multivariate statistical techniques you can use. However, the analysis, interpretation, and presentation of complicated multivariate analyses are not easy.

Observer or subject bias as a confounding variable

In many studies, the possible bias of the researchers is one of the most important confounding variables. Finding a statistically significant result is almost always more interesting than not finding a difference, so you need to constantly be on guard to control the effects of this bias. The best way to do this is by blinding yourself, so that you don't know which individuals got the treatment and which got the control. Going back to our catnip oil and mosquito experiment, if you know that Alice got catnip oil and Bob didn't, your subconscious body language and tone of voice when you talk to Alice might imply "You didn't get very many mosquito bites, did you? That would mean that the world will finally know what a genius I am for inventing this," and you might carefully scrutinize each red bump and decide that some of them were spider bites or poison ivy, not mosquito bites. With Bob, who got the placebo, you might subconsciously imply "Poor Bob—I'll bet you got a ton of mosquito bites, didn't you? The more you got, the more of a genius I am" and you might be more likely to count every hint of a bump on Bob's skin as a mosquito bite. Ideally, the subjects shouldn't know whether they got the treatment or placebo, either, so that they can't give you the result you want; this is especially important for subjective variables like pain. Of course, keeping the subjects of this particular imaginary experiment blind to whether they're rubbing catnip oil on their skin is going to be hard, because Alice's cat keeps licking Alice's arm and then acting stoned.

References

1. Picture of elm tree from [Imapix](#).
2. Picture of kidney cell from Indiana Center for Biological Microscopy

This page titled [1.5: Confounding Variables](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [John H. McDonald](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.