Controlled Experiments: Crash Course Statistics #9

Crash Course: Statistics
https://youtube.com/watch?v=kkBDa-ICvyY
https://nerdfighteria.info/v/kkBDa-ICvyY

[Complexly theme]

Hi, I'm Adriene Hill and welcome back to Crash Course Statistics.

Famous tech guy and likely future space dweller Elon Musk once told interviewers that there's a high probability that we're all living in a simulation. Now that might sound outlandish, and it's an interesting statement about probability, but today we're gonna focus on the simulation part.

[Intro Music]

Elon Musk, as well as some philosophers, futurists, and technologists argue that it's entirely possible–probable in fact–that in a few thousand years we will be able to create simulations, like World of Warcraft or Mindcraft, that are so real that they'll be indistinguishable from real life. Living in a world like that, there's no way for you to tell if you're living in the "true" world, or a simulation. Or maybe a simulation inside a simulation.

If we were living in Musk's version of a future world, any time we wanted to test something, like whether that bus would have splashed us if you didn't stop to answer a text or whether the new drug we invented can cure lung cancer, all we would need to do is run two identical simulations. In one simulation, we'd give people with lung cancer the new drug and watch to see what happens, taking notes on things like lung capacity and remission rates. While that's happening, we'd do exactly the same in another simulation, except we'd give patients a placebo drug, something that looks and feel like a real drug but isn't, and we'd record our results.

Once both simulations are done, we'd be able to look at our data and have a good idea of whether it works. Now, we can't yet run a parallel universe just to satisfy our curiosity of whether the amount of time we spend sleeping next to our cell phone increases our chance of brain cancer, but we are doing the next best thing. Researchers have started using simulations to study cancer treatments, to predict the impacts of climate change. Researchers are using VR to simulate disaster situations and o help train people how to respond in earthquakes or floods. Thanks Thought Bubble.

Still, we need other ways to answer the burning questions that keep us up at night. And, until scientists can pull through with the necessary technology, we're limited to two main methods of collecting data to answer our questions.

One way to get around our inability to create and destroy universes at will is by using experiments. Experiments try to mimic parallel universes by taking the one universe we do have, and splitting it randomly into groups. Imagine a high school where test scores are low, and we want to see whether buying cappuccino machines for students' families will improve test scores. We don't have two parallel versions of the high school to test our idea, but we can randomly split it into two groups: one group's families will get free cappuccino machines, and the other's won't. And then we let life go one, just like in the simulations, and we record the test scores of both groups.

Because we randomly assigned each person's group, every single family had an equal chance of being in the cappuccino and no-cappuccino groups. Randomness allows us to claim that before the cappuccino makers were given out, there were no systematic differences between the groups. Usually researchers will use a random number generator to assign subjects into random groups.

Random assignment reduces the chance that the bias of the people running the experiment will affect the groups. For example, it prevents them from doing things like putting subjects who they think will respond to caffeine in the cappuccino group, and those who won't in the no-cappuccino groups. It also makes it impossible for people t choose their own groups. That way the coffee-lovers don't all sign up for the free cappuccino machine, while the tea enthusiasts don't. These two problems are call allocation bias and selection bias, respectively.

Because of randomness, it's unlikely, though not impossible, that all the wealthy people ended up in one group and the not-so-wealthy in the other. Or the vegans in one group and the omnivores in the other. Randomness is usually our best method for ensuring that our groups are similar before we give them any treatment. But the slight possibility that our groups might be really uneven is why it's important to replicate experiments. In some situations, researchers can also force the number of wealthy people or the number of vegans to be the same in each group. This is something called randomized block design, but randomness usually does a pretty good job.

In our example, a treatment was either getting a cappuccino machine or not getting a cappuccino machine. In general, treatments are conditions that we want to test, like new medicines or educational interventions, like reading extra books to your kids.

Treatments can also have levels. For example, when we look at whether exercise is related to weight loss, we might have 3 groups: one group does no exercise, one group does 5 hours of exercise a week, and the third group does 10 hours of exercise a week. These levels can help us see whether there's a linear relationship between our treatment and our outcome, or whether 5 hours of exercise is just as good as 10.

One of the treatments in an experiment is usually nothing. These treatments are called controls and they play a huge role in experiments. It's our way of pretending we have a universe where there's no treatment.

Let's say you got a little too eager to get your freshly baked chocolate chip cookies out of the oven, and you didn't put the mit on right and you burned your finger. You put some Neosporin on it, and after a few days, your finger heals. You're so happy that you practically turn into Neosporin's next spokesperson, but it's possible that the burn would have healed just fine on it's own. A control treatment, AKA no treatment, would let us compare what happens between two similar burns: one treated with an antibiotic cream, and one left to heal by itself.

These types of control groups are great at helping us divide up the changes we observe into changes due to treatment and changes that are just due to time or circumstance. If your finger heals faster with ointment than without it, we can more confidently claim that there's a relationship between Neosporin use and burn healing than if we didn't have a control.

Sometimes, we want to control for more than just time and circumstance with our control treatments. With medical trials, you, as the person taking the medicine, would want new medicines to work better than nothing, but you also want to make sure that it's not just the act of doing something that's making you feel better. It's been shown that just the act of taking a fake medication or having a fake surgery, seriously, can make people feel better. These are called placebo effects.

Placebos allow us to control these effects by pretending to treat everyone. Subjects in medical studies are often given sugar pills or saline drips to make it seem like both groups are being treated. Non-medical studies also use placebos. Studies that claim first-person shooter video games, like *Call of Duty*, improve cognition should probably have a control group that plays a non-first-person shooter, like *Mario Kart*. Then, the researchers could ensure that it's not just

the act of participating in research or learning a new video game console that's causing the observed changes. Essentially, good placebos and controls should look and feel as close as possible to the actual treatment, so that the only difference is whatever we're interested in, like first-person shooter video games.

Sometimes, there are undetected factors causing changes in an experiment that you just don't know about. Well chosen placebos and controls allow us to better account for them. Sometimes it's just not possible to shield subjects from knowing the difference between conditions. Take diets, for example. When people agree to be put into a clinical study which compares a group of subjects who do not change their diet to a groups the goes low-carb, it's hard to make both groups think they're doing the same treatment. People can see what they're eating.

When subjects don't know which treatment they're receiving, usually because they're getting a good placebo, they are considered blind to the treatment. An experiment where the subjects don't know what treatment they're getting, but the researchers do, is called a single blind study, which leads to better experiments because all groups experience the act of being treated. But, even in a single blind study, it's possible that researchers might be biased when observing the subjects. The researchers still know which treatment the subjects is getting. A researcher who spent years creating, funding, and planning a study, probably thinks their treatment works. I mean, why would you dedicate your life to developing a drug to cure cancer if you don't think it works? The problem is that belief in the value of the treatment can cause researchers to subconsciously project those beliefs onto their subjects.

In the 1990's, a group of researchers looked at almost two dozen studies about whether sugar cause hyperactivity in children, and they concluded in general, it doesn't. Parents who thought their children were given sugar reported them as being hyperactive, but it turned out they were probably just being kids. When parents are blinded to whether their child received sugar, both groups of parents reported roughly equal levels of hyperactivity. Even when we think our subconscious bias don't get in the way, they still sometimes do.

We try to solve this problem by having double blind studies, in which both the subjects and the researchers have no idea which treatment the subjects is getting. Just like with blinding patients, sometimes it's impossible for the researchers to not be able to tell the difference, but double blind studies are the gold standard in many fields whenever they are physically, and financially, doable.

And, while we still can't bend space and time to make parallel universes, there are a few other ways to pretend we do have them. One is matched-pairs experiments. Just like the name implies, these experiments use pairs of subjects that are very similar and give one of them Treatment A, and the other Treatment B. Identical twins, for example, are about as close as we can get to a parallel universe. Since twins are genetically similar and often grow up in the same situation, we're able to get our treatment groups to be almost exactly the same, or at least way closer than random assignment alone. The more similar the groups are before treatment, the better researchers can spot treatment effects.

We can also have a matched-pairs experiments with non-twins. In this case, each pair is matched on one or more features that the researchers decide are important, like age, race, gender, or weight. Then, these pairs are treated like twins. One is assigned Treatment A, and the other, Treatment B. For simplicity here, we're assuming there are two groups. There could be more, but the best subject researchers can pair you with is yourself. Many experiments will give the same subject multiple treatments, one at a time, to see how they react to each. This ensures that each treatment group is the same, since it's all the same people. This type of matched-pairs design is often called a repeated measures design, and while it comes with its own set of limitations, it can be better than regular random assignment.

And, sometimes, the stars align and we get to see the results of the interesting experiments that we don't have the power to implement ourselves. Recently, the city of Philadelphia passed a sugar tax that would charge companies an extra one and a half cents per ounce of sugary beverage sold. Researchers can't just assign cities to have extra taxes. When cities do vote these things into law, some very interesting things can happen, and they did in the Philadelphia International Airport. Because it straddles the city's border, some of these otherwise identical stores had to comply with the law while others didn't. Through this data, researchers were able to find that contrary to previous assurances, chain stores in taxed terminals raised their prices about .83 cents per ounce more than their non-taxed terminal counterparts. Which means the tax did what it was supposed to: increase the cost of sugary drinks, and encourage people to buy less.

We all consume products that are the result of experimentations, from Amazon's web design to the prescription drugs we take. Which is why it's useful to know why researchers feel like they can make the decisions and claims that they do. Knowing the theory behind experimentation also allows us to be more informed consumers. So, the next time you find yourself staring down the homeopathic medicine section wondering if that effervescent tab, developed by an instructor of students, will really stop your cold, ask how they tested it. Ask who tested it. DFTBA-Q. You know I'm not sure why this isn't taking off. I mean, where's my t-shirt? Anybody?

[Outro Music]

CrashCourse Statistics is filmed in the Chad and Stacy Emigholz Studio in Indianapolis, Indiana, and it's made with the help of all these nice people. Our animation team is Thought Cafe. If you'd like to keep CrashCourse free for everyone, forever, you can support the series at Patreon, a crowdfunding platform that allows you to support the content you love. Thank you to all our patrons for your continued support.

CrashCourse is a production of Complexly. If you like content designed to get you thinking, check out some of our other channels at complexly.com.

Thanks for watching.