# Degrees of Freedom & Effect Sizes: Crash Course Statistics #28

Crash Course: Statistics
https://youtube.com/watch?v=Cm0vFoGVMB8
https://nerdfighteria.info/v/Cm0vFoGVMB8

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

It's great to have a lot of choices. But sometimes we limit our choices in order to do something productive or meaningful.

Like being on a team project that needs a writer, director, host, camera person, and boom mic holder. If we have 5 different people who can be on that team, after assigning 4 of them positions...the last person doesn't have any freedom to choose theirs. It has effectively been assigned.

If she's willing to give up the freedom to have a choice of positions and take on the great feat of upper body strength that is holding a boom mic, then they have a team that can complete their project. This can happen in statistics, too. Occasionally we have to give up some freedom--degrees of freedom--in order to do something useful with our data.

Degrees of freedom are the number of independent pieces of information we have and Degrees of freedom are an important part of many of the models that we use. In fact, we've also been leaving out another important component of the t-test: effect size. Knowing what degrees-of-freedom and effect-size are and why they matter will help give our t-tests better context.

INTRO In the last few episodes we've covered the general formula for test statistics. And we've gotten pretty good at calculating t-statistics for all sorts of situations: means, proportions, one sample, two sample, paired, unpaired but every time we've needed a p-value, we've let the computer do the work. Which is what we'll continue to do.

But it's important to know that we're not using the same t-distribution every single time. As we've previously discussed, the t-distribution is like the z-distribution, but it has fatter tails, meaning that extreme t-values tend to be slightly more likely. And that's because we don't know the population standard deviation when we calculate a t-statistic, so we estimate it using the sample standard deviation.

This little bit of uncertainty means that we don't have a perfect normal--or z--distribution. Instead we have our fat tailed friend. But with bigger sample sizes, we're better able to estimate population parameters like the mean and standard deviation, so our t-distribution changes its shape to reflect that.

As n--our sample size--gets bigger, we're less and less uncertain about our estimate, and the t-distribution will get closer and closer to z. More information usually means we have a more accurate estimate. Degrees of freedom can help us measure that accuracy.

We choose our t-distribution based on the number of degrees of freedom that we have. Degrees of freedom are the number of pieces of independent information in our data. Let's go to the thought bubble.

After dinner with 2 friends, you all pull out your credit cards to split the bill. Your friend Carmen, who's a bit of math savant, and a bit of a showoff, notices that if you took your credit card numbers as a single 16 digit number, the mean of your three credit card numbers is 4551-9681-7590-9146. She said this really loudly and you're a little nervous that an identity thief might have been lurking nearby and overheard Carmen make her very public declaration.

But there's nothing to worry about! Even though a potential thief has the mean of your credit card numbers, they won't be able to figure out what any of your individual numbers are. In other words, there's a lot of "freedom" around what those numbers could be.

And actually, you'd even be okay if the thief found out Carmen's credit card number. At that point, they could figure out the sum or mean of your and your other friend Eli's cards, but they still couldn't tell what your exact number was. There's still freedom for your credit card number to take on different values.

It could be any of these: BUT as soon as someone knows the mean of all three cards, Carmen's number, and Eli 's number, they'll know exactly what your credit card number is. It's no longer "free" to take on different values. If Carmen's number is this: And Eli's number is this: Then knowing the mean allows anyone to figure out that your number must be this: So you should probably make sure that Eli keeps his number under wraps.

Just to be safe. Thanks Thought bubble. In that example, the three credit card numbers already existed before we started doing any math.

And they are three independent pieces of information. Eli's credit card number has no effect on your credit card number, which has no effect on Carmen's, and so on. But, as soon as Carmen calculated the mean, she used up one of those independent pieces of information.

Once the thief knows the mean, they only need TWO pieces of independent information. (that is n-1 pieces). In this case, once they know any two of the credit card numbers--and the mean--they know all three. So when they learn Carmen's number and Eli's number -- SUDDENLY those numbers can reveal yours.

The thief can figure out your exact credit card number. Since it's no longer independent of the others. To bring it back to our t-tests... when we calculate a mean, we're using up one degree of freedom--or one piece of independent information.

The amount of information that we originally have depends on our sample size--n--which is why you'll often see it in the formulas to calculate degrees of freedom. The more data you have, the more independent information that you have. But every time you make a calculation like a mean, you're using up one piece of independent information.

So, for example, we have data from 100 randomly sampled square miles of avocado orchard, and we've painstakingly counted the number of bees spotted in each sampled square mile over the course of a week. The bee population is declining! We need to be sure avocados are getting pollinated!

The owner of one avocado orchard says that she usually sees 15,000 bees per square mile. So, you set out to analyze your data to see whether you think the bee population has changed. You have 100 pieces of independent data--one measure from each square mile--so, when you calculate the mean number of bees from all 100 square miles, you're using up 1 degree of freedom.

Now that we know the mean number of bees is 16,838, you only need 99 of the bee counts to figure out what the count for the 100th square mile would bee. With a quick one sample t-test, we get our p-value from a t-distribution with 99 degrees of freedom (the black line). If we had less data, say 6 data points, we'd only have 5 degrees of freedom which will give us a slightly different t-distribution with fatter tails (the blue line), and therefore a different p-value.

Our p-value of 0.001 tells us that we reject the null that the mean number of bees per square mile is 15,000. And we couldn't find that p-value without knowing our degrees of freedom, because as we mentioned in a previous episode, t-distributions get more and more like a normal distribution as we get more and more

independent information...aka degrees of freedom. In fact, it looks like the number of bees may be higher than it was previously.

Go bees! One thing to note, though: the 1,838 bee increase is statistically significant, but that just means that if the true bee count per square mile was 15,000 then it's unlikely that we'd get a sample mean of 16,838. But it doesn't mean that this difference is practically significant, or all that useful.

An increase of 1,838 bees isn't really that big compared to the standard deviation, 5,420. If on average, we expect bee counts to vary 5,420 bees from the mean, then a change of 1,838 may not be that important to us. For example, say that we treated half the orchard with a bee pheromone...which bees love...and is thought to encourage them to come back.

Our statistical test on the difference between a group of bees exposed to the pheromone and a group not exposed revealed that there was a statistically significant difference of 3,297 bees per square mile between the pheromone and non pheromone groups. But we still need to ask whether a difference of 3,297 bees is useful to the orchard owner? Those pheromones are pricey.

And she wants to make sure that they're worth it. That 3,297 bee per square mile difference is an increase of about 0.6 standard deviations. Remember that almost ALL of the data is within 2 standard deviations of the mean.

So a difference of a little more than half a standard deviation is a big deal.. Maybe those pheromones are worth it. Sometimes statistical significance doesn't give us the whole picture.

You probably already use this kind of reasoning in your real life. Like when you're scrolling through your Instagram feed and see a former Bachelor contestant promoting a hair vitamin. A little Googling tells you that yes, this vitamin does cause a statistically significant increase in hair growth, but only a few nanometers.

Your hair normally grows about 12.7 millimeters a month plus or minus a millimeter. So, this vitamin has what we call a small effect size. Effect size tells us how big the effect we observed was, compared to random variation.

It's really important to pair our p-values with effect sizes, because sometimes, we can get statistically significant effects, but effect sizes that are so small, they don't really matter to us. Let's look at an educational supplement called WOWZERBRAIN!. The creators of WOWZERBRAIN! do an experiment.

They bring 90 kids into their center and randomly assign half of them to get the WOWZERBRAIN! supplemental materials, and the other half as a control group. The control reads age appropriate books for the same amount of time that it takes to go through a WOWZERBRAIN! lesson. Once the data is collected, the WOWZERBRAIN! creators take a look at their data and find out that the kids who took part in the WOWZERBRAIN! intervention had a mean reading score improvement of about 1.329 points and the control group improved an average of 1.265 points.

The first things the WOWZERBRAIN! researchers do is perform a two sample t-test, and find a t-value of -0.21. And a p-value 0.8 -- calculated using a t-distribution with 88 degrees of freedom. So they weren't able to reject the null.

Their effect size - substituted into our equation is only about 0.044, which is pretty small. That means that the kids that got WOWZERBRAIN! materials only had scores that were higher by about 1/23rd of the amount we expect students to vary just by chance. But despite the null result of their t-test, the WOWZERBRAIN! creators look at the raw numbers and see that the kids who got WOWZERBRAIN! did score numerically higher, even though it wasn't statistically significant.

So they, like many researchers and scientists, think to themselves that maybe the reason that the t-test wasn't significant was because they ran an underpowered experiment... with too small of a sample size. Since standard error is scaled by the square root of n then--all things equal--the larger our sample size, the smaller our standard error and the larger our t-statistic will be. So, the researchers wonder whether they could detect an effect if they tested 10,000 children.

And sure enough, with 10,000 kids, they got a t-value of -2.218, with a p-value of 0.02886. Which is small enough to reject the null hypothesis! But notice that their effect size is still the same...about 0.044.

So the intensive WOWZERBRAIN! intervention, still only helped improve average reading scores by 0.064 points. P-values, as you can see, aren't everything. They should always be looked at in the context of other measures, like effect sizes.

P-values tell us whether it's likely something happened by chance alone. Effect sizes help us figure out whether observed effects are practically significant to us. In this case, though the WOWZERBRAIN! creators achieved statistical significance, for many people they may have failed to achieve practical significance.

Parents are unlikely to pay for a year round educational program that only improves test scores by 0.064 points. We talk a lot about p-values, and that's because lots of people use them to do really important things. But they can't stand alone.

P-values are PART of the whole picture and should be paired with other information, like an effect size. It's like trying to buy an apartment based on cost per square foot alone. Sure, maybe you find something for 75 cents per square foot….but it turns out it's right next to the city dump...so maybe you'll pass on that one… And we need degrees of freedom to understand why smaller differences between means can be significant if you have a larger sample size.

The more information you have, the more accurate your estimates are. It's why we might not bat an eye at the fact that two people from two different countries have a height difference of 1 foot, but very surprised if those two countries had an average height difference of 1 foot. And that's about 0.3 meters for you people using the metric system.

Having more accurate information changes the threshold for what's surprising or significant to us. Thanks for watching. I'll see you next time.