Week 7
Video Transcripts

Video 1 (12:23): Introduction to Confidence Intervals (CIs)

All right! Hello again, and we are going to talk about confidence intervals for estimating the mean. So, the key ideas, right, that we're going to eventually come ...come up with, again, is that we're going to talk about confidence intervals as a way of quantifying sampling error around our estimate. So, instead of giving you a point estimate– let's say, people watch TV for an average of an hour per day, we will actually quote an interval and we're going to say something like, we are 95% confident that people watch TV anywhere between 75 minutes and 125 minutes a day.
So essentially, we are going to give you an...an assessment of the margin of error around the point estimate. So, this is the idea. And, today we are going to do it– the application will be sort of measuring blood sugar level in the context of, sort of people that are diabetic, which is a big...big problem, both in the United States, and...and globally. So, here is a setting. So, there are two types of diabetes especially, type two is a major...major condition and epidemic in modern living. So, the statistic that I have here is that about one in ten adults suffers from diabetes in the, in...in...in the United States. That is about 26...26 million people, to be precise. Most of them have type two diabetes. And...and, type two diabetes is more easily treatable and preventable.
But, nevertheless has significant side effects, and one has to continuously monitor it, and...and take proper medication. It–the—this condition alone and...and its diagnosis and treatment in the U.S. is...is something that costs about 300 billion dollars. So, there is a genetic component to that but, there are also lifestyle issues, like diet, weight, activity that sort of compound and cause increase in...in blood sugar levels. All right! So, we will actually talk a little bit about that. And in particular, we are going to talk about how do you test for diabetes? How do you test or monitor your blood sugar level over time? So, there are essentially two types of tests. One of them is, you know, you take a blood sample from typically a finger, and you estimate— and you measure what your blood sugar level at the time that you took the sample.
And, typically there are sort of guidelines about what should that be. If you took the blood sample, let's say after you woke up but before you ate your breakfast, so that would be after–in a fasting type of period. Or, whether you took your blood sugar level, let's say a few hours after you ate your meal, which is sort of when you are in the ingestion/digestion phase of your...of your daily routine. So, you know, there is a...a lower level, if that is taken during a fasting period, and a slightly higher threshold, if it was right after a meal. But ideally, you should not be violating that level all the time. So, that's one thing. Another thing that people can do is–you can do, a test that is sort of doing a time average over a long period of time. And, it has a slightly different threshold expressed, slightly differently, that is, the 'A1C' test, but roughly speaking, saying the same thing. And thus, the 'A1C' threshold of six and a half percent, it's roughly corresponds to about 140 milligrams per decilitre in this sort of standard blood sugar measurement test. All right!

So, that's the setting. And, this is an example of, let's say a person whose 'A1C' was seven, and you can see the various measurements that we were making over time. This is a period of four days, and...and...and this is an example of you take a measurement in the morning, midday, evening, night. All right! And, some of these measurements are away from meals, or...or some of these measurements are after you...you had some of your meals, and that's why you're there a bit higher.

Now, one thing that you notice in this graph is that these things fluctuate, right? Quite a bit. First of all, they fluctuate depending on whether it happens in a period–long time after your previous meal but, it also fluctuates based on whether you're exercising in the morning, whether you ate–what you ate in your meal.

So...so these are...these measurements are sort of stochastic over time and that is why, what people do is they take a lot of these measurements, and then they average them out. In particular, we are going to follow this routine. We are going to assume you take four measurements per day, like the graph, and you are going to do it over two weeks. So that would give us, 14 days time...14 days times 446 observations. And, let's assume for now that your sample estimate was 135 milligrams per deciliter, and the standard deviation of our measurements– so, this is these sort of fluctuating blood sugar measurements, the standard deviation here, is of the order of 25 milligrams per deciliter.

So, the question is, is this person diabetic, and should he get treated. And, in particular, what that would mean is, what is the likelihood that his true blood sugar level is above the permissible threshold of 140. Right! So, that is...that is the question we're after. So, what do we know? We know that the estimate 'X' bar is 135, but it's noisy. It has error. All right! And, the threshold is 140. So, the question would be whether, what's the probability, let's say or what is the likelihood that my measurement of 135, because of the error, the reality is the true number I was measuring may have been higher than the actual threshold of...of 140. So, that is really what we're after here.

So, let's see. I want to know–I want to figure out, what is the true blood sugar level of that individual, and I'm going to denote it by mu. So, that's the mean of some kind of distribution and I would like to know, in particular, whether the true mean is above my threshold of 140. And, what I did is I measured and I got a sample estimate of, here, 135, sigma was 25, and 'n' was 56. All right! And...and the question is somehow, given the fact that each measurement is noisy and given the fact that I didn't collect too many samples, only 56, what is the likelihood that, you know, by mistake, or by sheer luck, I computed 135, even though the truth was higher. All right! So, that is what I'm after. And...and here is a calculation, right! So what do I know? I know for that from the central limit theorem, which is what we've been talking about in the last two sessions, 'X' bar is centered at the truth, let's say mu, the unknown mu, and...and let's suppose for...for a moment that I wanted to figure out, what is the likelihood that you would be between mu plus 5 and mu minus 5, all right! Or, what's the likelihood, essentially, that the error is small? All right! So, that would be this probability here, right, which is exactly what I just wrote here.

And...and, if I wanted to evaluate this, evaluate as follows. So that, I'll write as 1 minus twice, that probability, which is the probability that 'Z' is bigger than 5, divided by sigma, sigma is 25, over the square root of 56, so that is 1 minus twice the probability that Z is bigger than 1.49, and that is 0.86. So, what do I know? I computed that the probability, right, that the error is less than or equal to 5, is 86%. In other words, there's a 14% probability that the error is bigger than 5, in...in...in...in absolute value. All right! Now, when I put the confidence interval, I actually flip that statement. So, instead of...instead of talking about what is the probability that my error is, sort of exceeds 5 milligrams per deciliter, instead what I'm going to do is I'm going to flip my statement. I'm going to say something of the...of the following sort.

I'm going to say, I'm 95% confident, all right, that the true mean, here, is, let's say, contained in an interval, and the interval will look like my sample estimate 'X' bar, which is 135 plus or minus some amount, and the plus or minus amount will be some kind of multiple. It's going to be a multiplier that comes from the 95% times the standard error. All right, and so, if I want 95% probability or confidence that means that I'm going to have .025 probability on either tail, and that means that this distance needs to be plus or minus 1.96 standard errors. And, this is how I get my confidence interval. I'm going to say 'X' bar plus or minus 1.96 standard errors on each side. So, we...we have already seen the standard error formula, which here would have been 25 over square root of 56.

And, 1.96 is the multiplier that I get from the 95% confidence level, which we saw also in the previous session. So essentially, rather than saying I want to fix the margin of error to be plus or minus 5–and I'm going to tell you what's the probability that I'm going to fall in that...in that range, instead, I am going to now fix the confidence level, e.g., 95%, and I'm going to tell you what's the corresponding margin of error. So I want you to always be 95% confident. Now, tell me what is the width of that interval, essentially, which is what we're doing?

Video 2 (4:20): CIs at Different Confidence Levels

And, this is a picture of what essentially, I'm...I'm...I'm talking about. So, if I wanted, for example, to be 90% confident, you know, if I keep making those measurements, essentially nine out of ten times, the true parameter would fall within the interval, okay. The–if I'm–when I'm...when I'm saying, I'm 90% confident, is analogous to saying that 90% of the time if I repeated the experiment, if I repeated the sampling, right, the true parameter will lie in the 'X' bar plus or minus 1 point...1.645, times sigma, over square root of 'n' confidence interval.

That thing here, is the multiplier for a 90% confidence interval, which I know from what we did in the normal distribution. So, if I want 90% probability, inside here, I have '0.05' on each tail, and that correspond to 1.645 standard errors away from the mean, okay! So, that's the picture that goes with that. So, essentially what we have found is that every confidence interval, and that's sort of very general, would look like my...my estimate. So, I use the data, I got an estimate, 135, in this case, and then I'm going to pair it by saying plus or minus some amount. And, the way I'm going to get this margin of error, it will depend on the multiplier–is going to be some kind of multiple of the standard error. And, the standard error is a function of the estimator, right, so that's, let's say, sigma over square root of 'n' in this example, and the multiplier–it depends on the confidence level. So, what we had is for 95%, we had .025 on each tail, and that corresponds to 1.96. For 90%, we just talk...said in the previous slide, it's 1.645. For 99, it's going to be 2.58, etc. all right! So, that's really the notation. If we want a confidence level or, that is, a level alpha confidence interval, we'll have alpha over two probability on each tail, and, a shorthand notation that we typically, often use for this multiplier is 'Z' alpha over two.

All right! So essentially, we get an estimate, and then we pair it with a confidence interval, and the confidence interval always looks like plus or minus some amount.
And, the plus or minus amount is some multiple of the standard error. Again, we compute an estimator, we choose a confidence level, we find the multiplier. And then, we compute the standard–the confidence interval, as such. If we know sigma, we use it. If we don't know sigma, we typically use a sample to re-estimate sigma, all right.

Video 3 (5:31): CIs at Different Confidence Levels

Now, I want to make a very small correction to what I've said, so far, and that's the following. When—if sigma is unknown, right, so I want– the standard error formula looks like sigma over square root of 'n'. Now, if sigma is unknown, I will either– I will typically go back to the data, and I'm going to estimate it using the data. But, when I estimate sigma using the data, I'm clearly making an error. So, I made an error in 'X' bar. Now, I'm making another error in sigma, in my sigma estimation. And...and, I want to somehow account for that in my confidence interval, because had I said 'X' bar plus or minus 1.9 standard errors and I knew the true standard error that would be one thing, but if I'm giving you a standard error estimate itself that may be wrong, I may have to do something with my confidence interval.
And typically, the thing that...that I would do is, if I knew that my...my sigma estimate may have some error build into it, what I should do is I should take my old...my old confidence interval and just make it a little bit wider, right. And, the way I'm going to do it is, I'm going to get a different multiplier. So instead of, let's say, doing 1.96 standard errors, I'm going to do 2, or 2.1, or 2.2. So essentially, I'm going to say, look because you don't know what you're doing in your estimation of sigma, I'm not going to let you get away with 1.96 standard errors, but maybe I'm going to increase that and make it something a little bit wider.
Typically–and a rule of thumb is that we can ignore this correction, if the sample size is at least 30, okay! And, one thing to note is that this correction is not–never valid, if we're talking about discrete or ordinal data, like in polling, all right! So, let me just show you where this thing comes from. It comes from a distribution, we call it a 't'–'t-distribution' and a 't-table'. You can get it in Excel, in [inaudible]...in Python. I'm going to show you a table in...in a bit, and you're going to see how that looks. The 't-distribution' look as follows. So, it...it almost looks like the normal distribution, right. So, the normal distribution will be the black curve, and what you see is that the 't-distribution' with different degrees of freedom, the number of degrees of—well, let me just write that, degrees of freedom is 'n' minus 1, 'n' is the sample size, used to estimate sigma.
All right! So, if I don't have a lot of data, what we, what...what...what I get is, I get this blue curve. This is 6– 5 degrees of freedom means six sample points. So, I realize that I have a little bit more probability in the tails, like, if 'n' is small, I have more probability in the tails. And, what that would mean, that would mean that the multipliers would go further out. So if I...if I wanted, let's say, to have a 5% or two and a half percent probability in the tail and...and my tail is not decaying fast enough, I need to go further out, so that's what...that's what happens. As the sample size...as the sample size increases, the 't-distribution' converges to the standard normal. Okay! So, all of these differences go away.

And, here is a picture of the 't-multipliers'. So here, I have the degrees of freedom, and that's as I said, 'n'–sorry minus 1. And...and I only give you, here, tail...tail probabilities. So here, for example, if I wanted two and a half percent probability in the tail, what happens is, if I have infinite number of points, I get 1.96. Let's say, if I have 20 degrees of freedom, I would quote– so with 'n' is 21, I would come here, I would quote here instead of 1.96, I would say about, 2.1 standard errors. If I only had five points, I would say 2.57, etc. So, I'll make the interval a little bit wider to account for the extra error in estimating sigma. So, I'm doing exactly the intuitive thing. It's just that the whole point here is, where do I get these multipliers. I get them from the 't-distribution'.

Video 4 (5:48): Interpretation of CIs

So, I want to make a few quick additional comments before we quit that segment. So, the first thing is confidence intervals always look like estimate plus or minus sort of a...a...a margin of error. And, there are three things that determine the margin of error, right! So, the first thing is the multiplier, which depends really on the confidence level, right! And, if you want to be– if the confidence level increases, right, the multiplier–if you want to be more...if you want to be more confident, the multiplier will go up, right, because I need to make the...the interval wider. So, that's the first thing. The next thing is it depends obviously on the sample size, right, because what do we know. We know that the standard error looks like sigma over square root of 'n'.

So, as 'n' grows, right, the standard error goes down, like 1 over square root of 'n', all right! So, that's intuitive, the more–if I have a lot of data, then my confidence intervals will become narrower, because I...I will have smaller error. And, the other thing is the variability of the underlying data. So, that's sigma in the standard error formula. So, if there's a lot of noise in my observations, then that will cause more error in my estimate, that's the only thing that this says. So, these three things affect the accuracy or the margin of error of my...of my–in my confidence intervals. So, the...the...the...the last thing more or less that I want to do, is I want to see how to pick the sample size to achieve a certain margin of error. And, the way I would do it is I would say, if I want the margin of error, let's say, to be equal– it's equal to sigma over square root of 'n', times some kind of multiplier, right, so this is the standard error, and this is the multiplier, so, if you fix this...if you fix the margin of error, I can then, use this expression to figure out what's the sample size.

So, in...in the example that we had before– suppose, for example, we wanted to be 95% confident that my error will be within plus or minus 10 milligrams per deciliter, all right! So, I told you the margin of error and I told you my confidence level, right, so based on that what do I know. I know that the margin of error looks like sigma over square root of 'n', times 1.96. But, I also told you that this should be equal to 10, right! So, I have this expression sigma is 25–this is sigma, so, I can then, solve for 'n'. All right! So, that's...that's it. More...more generally, we have this expression. So, if I give you how big is the margin of error and I give you the confidence level and you know sigma, then, you can compute the sample size that you need. Recap briefly! What we did last time when we're doing proportions, it's the same story, estimate plus or minus multiplier, times standard...standard error.

The standard error formula changes, and it's the estimated proportion 'p' hat, times 1 minus 'p' hat, over 'n' square root. All right! So, I'm going to stop here. So essentially, what we talked about is when we're doing sampling or estimation, instead of just quoting you sort of the estimate of the number that I...that I got, what we typically do is we give you an interval. And we say, for example, that we're 95% confident that the true parameter lies in a particular interval, so that interval-it's centered around your estimate plus or minus some amount, and that amount depends on how confident you want to be, how many data points you used, and what's the variability of the underlying data, so that's sigma.

Video 5 (7:24): Statistical Significance

We're going to talk about confidence intervals for differences in means and...and proportions, and then statistical significance in the concept of...of...of the 'p-value'. So, the main idea here is when we're looking at data and we're trying to make a statement, we want to assess what the statistical significance, or what's the, sort of, strength over the statistical evidence that we have in front of us in...in...in the data. And...and, the main notion that we're going to have–that we're going to use is going to be something called, 'p-value', which exactly is going to measure the strength or the statistical significance of the observation, all right! And...and, typically small 'p-values' would mean strong evidence and...and in principle, we would look like–we would look for 'p-values' that are less than or equal to .05 or 5%. Now, we're going to return to that later on, but that's really what's coming and that's going to be the main thing that we're going to extract from...from the class. I want to return to the data set that we saw on our first session. And that day we had discussed– we had looked at that data, and we had tried to answer, let's say, whether there is an upward trend in the temperatures and then, we also divided the data set into two periods, and we tried to see whether there is any drift over time. So, I want to return to that second bit. So, what we had seen is that for the first period, the average was 35 and for the second period, it was 38.1. But, we also had some standard deviation, which is sort of a level of accuracy or variation in these observations.
So, what I would like to do is, I would like to compare– I want to compare 38.1 to 35, taking into account the standard error of these measurements, because these are measurements and themselves have error in them, all right! So, paraphrasing that, I want to look at whether the difference between these two is large or relative to the amount of error that we may have in our...in our observations, due to the inherent variability in the data. So the way to do that is, literally, to just build the confidence interval. So, I have two observations, the first one has 'n' equals 100, and the second one has 47 years in the observation period. And, what we have is that the difference between the two sample means is 2.31 degrees, and I want to construct a confidence interval for that.
And, what I know is that my confidence intervals will always look as best guess, that's the 3.1 plus or minus some multiplier like, 1.96, if I'm doing a 95% confidence interval, times the standard error, okay! Now, if these are two independent estimators, then, the variance of the first is sigma squared over 'n', the variance of the second is 'sigma$_Y$' squared over 'm', its own sample size. So, I can just construct the standard error of the difference as... as that. The key...the keyword here is independent. All right! So, the variance of 'X' minus 'Y', if two things are independent, is a variance of 'X' plus the variance of 'Y'. This is really what we're doing here, all right! So, we're going to use that formula, and if we use that formula, we're going to get that the standard error of the difference, all right, is this number .75, which is 3.8 squared, divided by 100, plus 4.4 squared, divided by 47, and that's the standard error of the

difference, which makes the confidence interval be three degrees– 3.1 degrees plus or minus 1.96, times .75, which is 3 plus or minus one and a half.
So, that interval, all right, looks like 1.6 to 4.6. So, that would say that at the 96%–95% level, the change in temperature is significant or is positive, all right...all right! So, going back to the slide. So, what do we have? We have that the confidence interval is 3 degrees plus or minus 1.5. So essentially, what we're saying is that with 95% likelihood, the difference between the two population means is contained in this interval, which is strictly positive–does not contain zero. So, in particular, we're pretty confident that the change in temperature is at least 1.6 degrees or higher. So, it seems like the most recent temperature average is, sort of, higher than the previous year average at the 95% level. All right! Now, we could...we could ask ourselves, so how confident are we on...on that statement. So, we could to pursue that question a little bit further– so, in particular we could, if I want to be more confident, I will start making the confidence interval a little bit wider, and at some point in time, the confidence interval becomes so wide that we'll start to include zero, and at that point in time, we'll need to stop, and that thing will be called, the 'p-value' of that observation.

Video 6 (4:27): P Value

Let me illustrate that. So, at 95%, we...we constructed a confidence interval, and it did not contain zero. At 98%, the multiplier changes, the standard error stays the same, confidence interval starts to widen. At 99%, it widens further, widens further. At some point in time, in particular at that level of confidence, the standard error becomes big enough, so that the confidence interval will straddle zero. So, in particular at that–at the...at the 99.997% level, we can no longer–cannot–can no longer conclude that 'mu$_x$' minus 'mu$_y$' is bigger than zero, all right! So, we will say that the 'p-value' of this observation is .003%. Okay! So, that's the strength of the observation. Another way of assessing that strength would be to compare this 3.1 to .75, all right! So, that's around 4.15 standard errors. So, we need to go 4.15 standard errors away from the mean, or what we observe, in order for that to cross over zero, and that will end up being related to the 'p-value' as we're going to see.
Defining it–the 'p-value' is, sort of, 100 minus the level of confidence, where the confidence interval straddles zero. So, that in a previous example was the 99.997%, right! Small 'p-values' are good, and in particular, we want them to be smaller than .05, all right! For our data, we conclude that the likelihood, okay so–another interpretation is that the likelihood that a sample...sample mean difference is plus 3.1 degrees when 'mu$_x$' is, in fact, less than or equal to 'mu$_y$', is .003%. So, the probability of observing such a drift in temperature when there is really no change in what's happening in the climate, would have been .003 percentage points. And so, you know, small, all right! Now, we could have computed the 'p-value', as I mentioned earlier, by comparing the 3.1 to the standard error, concluding that we need to be above four standard errors away from zero, and then, the 'p-value'...the 'p-value' is really equal to this number, all right! So, it's going to be twice the probability of being outside, whatever we observed in this calculation.

Video 7 (13:51): A/B Testing Example

I'm going to talk a little bit about doing confidence intervals for different sort of proportions, and show you an applications of something we call, 'A/B testing'. So, to motivate this example, imagine the situation where we're showing a banner ad– this is a banner ad, when people land into the front page of the Financial Times, and let's say, we measure the click-through rates– how many people click on the ad when they...they're shown the ad. And, let's say, this is .01%. Now, we showed a slightly different ad to a different venue.

This is the Wired magazine. So, when you go to that magazine, and I show you a slightly different ad for...for the same product, we observed that .05% of the people that are shown the ad, clicked on it. So clearly, it seems like this is significantly higher than that. The question is, is there something systematic? All right! Now, if you notice, two things happened in this example. First of all, I changed the venue, the...the website, and secondly, I changed the banner ad. All right! So, this...this is a case where, it may be tricky to assess whether the difference is because of the population of users of Wired versus Financial Times, versus the difference in the ad itself.

So, what we would like to do is essentially run a randomized experiment, that's what we typically do. So, we would try to keep everything else equal, and show slightly different ads to statistically similar people and measure the response. And, I would like, let's say, to both be in the Financial Times, both be of the same demographic, both coming in the–around the same time of the day. And, sort of, keep showing these two different samples, different ads, and measure the difference on that. And, in particular, I want to make it as much apples to apples comparison so that, then I can hopefully tease out the difference of the ad design versus everything else that may be changing at the same time.

So, this is an example, let's say, when you...when you search on Airbnb, it could be that the default of the slighter could be different, and...and as you change the default, maybe people eventually, when they navigate, they be more likely or less likely to book a...a stay through that site. So, that–so...so this is sort of the setting, all right! So, I want to control for all systematic differences, demographic, time of day, type of search, type of–you know, and then, I want to tweak only one parameter, let's say, the size of the...of the maximum price or the location of the banner ad or the color or whether I used bold characters versus italics, or what have you. All right! So, I want to do this.

And, in...in internet settings, we call this, kind of, 'A/B tests', and this, sort of, happens–that happened all the time, and...and...and very successfully, all right! So, you know, one factoid is that the Facebook– there are about a thousand of these running simultaneously every day, all right! So, let's see how this would play out, and I have a very simplistic example here. So, I'm going to send out–I'm going to do two email campaigns. And, the first one, I'm going to say, 'Limited time offer! Use promo code ABC123', and then, in the second email, I'm going to say, 'Offer expires on Sunday! Use promo code 123ABC', all right, and then I can actually send it to two samples of, let's say, 50,000 users in that case, and I can measure the response.

So, first of all, how many people clicked on my email and opened it up and visit my website? And then, how many people bought, all right? So, the first–from the first email, I got 1% visit rate. So, I send 50,000 of them–and 1% of them actually visited my website. In the second one, it was half a percent. And then, I got some people to buy, from each of these samples. Now, what I would like to do, is I would like to check whether these differences are statistically significant, all right! Can I tell these things apart? Is it better? Is this thing any better in driving traffic into my website, and ultimately driving sales from this campaign?

So, what I could do, right, is I could compare these two estimates, right! So, I have one estimate that is 1%, and I have another estimate that is half a percent, so, I want to compare this difference with a standard error, and this is my–it's typical to use estimate, plus or minus multiplier, times the standard error. Multiplier is 1.96. The only formula that is changing is this formula for the standard error, and this is sort of the natural formula, which we–what we had a couple of sessions ago. So, this would be, let's say, .01, times 1 minus– sorry–.01 over 50,000, plus point o... .005, 1 minus .005, over 50,000, square root, right! So, that's what we're talking about. And, this is, sort of, pattern matching exactly this. So, if we were to do exactly that calculation, we would, indeed, have, right, .01 minus .005, plus or minus 1.96. And, here I will have .01, times 1, minus .01, over 50,000, plus point... .005, 1 minus .005, over 50,000. And, that's going to be my confidence interval.

So, this is roughly .005 plus or minus 1.96. This is roughly .015 over 50,000, square root. And, that's roughly .5% plus or minus .1%. So, we're 95% confident that the, sort of, click-through rate with email one is anywhere between .4% and .6%, higher. All right! So, the only thing that changed was the formula for standard error here. So, for example, based on that... based on that comparison, we would, again, conclude that the email one leads to significantly higher visit rates to our website, at the 95% confidence level. Now, I could try to compare, whether people actually buy, right! So that would be my, sort of, next...next thing. And, I've done it in the next slide, rather than do it by hand. So, this is .05% minus point o–sorry, 3%. This is this number. And, the standard error is the same thing, is 1.96, times–this big ugly formula. Right! Now, it turns out if you do that, what you end up getting is, you end up getting a confidence interval that itself straddles zero.

And so, at the 95% confidence level, we cannot actually tell whether email one leads to more purchases than email two, and part of the reason is, we... we really just don't have enough purchases, right! Because .05% of 50,000, right–so 1% of 50,000 is 500, so, we're talking about 25 purchases. So, in one case, you had 25 sales. In the other case, you had 15 sales. And, the question is these numbers are still sufficiently close to each other, so that we would not be able to tell them apart, okay! So, at the 95% level, and the 5% significance level, we cannot tell these things apart, in terms of whether you generate more sales with one email versus the other. You may need bigger samples, right! The reality is we typically would want bigger samples, and we typically need a lot more data– and I'm not going to walk you through this calculation but, typically what we may want– you may want to do is you want to break down the data or the sample, let's say, by gender, or by age and by gender, and then, when you realize is that when you start breaking down these numbers by...by gender and by...and by age, the data problem that I was mentioning, actually become very significant.

Right! So, I may want to say is this number significantly different than that number, but the problem is on that demographic, I may not have enough samples, right! I used to have 50,000 samples for the entire market, and now I want to say, "Well, let me focus on male, 18 to 24-year-olds." Well, if I focus on that sample– I certainly don't have 50,000 samples, I may have a few thousand samples, and that will

typically prove not to be sufficient. So, this is the type of data that we had. Or, this is the type of data they had. And...and, the reality is that this is, kind of, not sufficient data size.
So, what you may want to do is just drastically increase the sample size, or the other thing you may want to do is you may want to introduce a model, right! So, in some sense, I mean no one is segmented by age. But, I may want to put a model, where age changes the probability of click-through in some particular way, perhaps, using things from regression that we may see next, all right! All right! So, I'm going to stop with this takeaway slide. So, we talked about differences of means and building confidence intervals, and the main thing here is when the estimators are independent, we change the standard error formula. We did the same thing for proportions. And then, we talked about the strength of the statistical evidence and we called it the 'p-value'. All right! And, small 'p-values' are good.