ICF Transcription
Power Analysis Webinar Part 2

Lily: Welcome and thank you for your interest in this Power Analysis training session. My name is Lily Zandniapour, I'm a Research and Evaluation Manager at AmeriCorps. And on behalf of the Office of Research and Evaluation at AmeriCorps, I am delighted that we can provide you with this second training in our three level series of training modules on Power Analysis for Program Evaluation.

The first level training webinar provided an introduction to Power Analysis. This modules covers the basic mechanics of power analysis and provides more technical content on the topic. The third level training module's focus will be on applied power analysis.

This training series is delivered by our evaluation, training, and technical assistance provider, NORC of the University of Chicago. For some years now we have partnered with NORC to provide support to our grantees, and strengthen their evaluation studies so they produce credible and quality evidence about the programs and interventions they implement, and that the agency supports.

We hope that you find this training, as well as the other two modules on this topic, helpful in your work. Dr. Carrie Markovitz who is a Principal Research Scientist at NORC, and Project Director of our Evaluation, Training, and Technical Assistance work, lead us in this training and introduce our presenter, Dr. Eric Hedberg. Thank you.

Carrie: Thank you, Lily. NORC, at the University of Chicago, has been collaborating with AmeriCorps' Office of Research and Evaluation for almost a decade, to strengthen the existing evaluation guide and some tools for AmeriCorps' state and national applicants and grantees.

Over the years our TA team has assisted and supported numerous AmeriCorps grantees with their evaluation plans, design and implementation challenges, instrument development, and reporting. And through this work we have identified common areas of need, and then worked with the Office of Research and Evaluation on developing tools, webinars, classes, and presentations to fill these gaps.

So this class empowers one of several classes on evaluation topics that were developed over the past ten years by NORC, and are currently available online on AmeriCorps' website. These classes include information on topics from logic model development to drafting research questions and addressing other design topics, to budgeting and managing and evaluation. We encourage everyone to check out these other classes based on their own evaluation needs.

Now I would like to introduce Dr. Eric Hedberg. For Dr. Hedberg, power is a major methodological topic of interest. Most known for his work on evaluation design, Dr. Hedberg is an interdisciplinary quantitative methodologist. His research interests include several areas of methodology related to evaluation and analysis and he recently authored a sage little green book on statistical power analysis.

Dr. Hedberg is an accredited professional statistician by the American Statistical Association, and he is a sociologist. His current areas of research include investigating the design of evaluations in education and criminology, in addition to measuring social capital through social network contextual effects.

Dr. Hedberg has authored or co-authored over thirty methodology-focused papers and books that have appeared in education, medical, and criminological journals, while also contributing to numerous reports and presentations at major research conferences. Dr. Hedberg earned his PhD in Sociology from the University of Chicago. So now I'd like to pass this on to Dr. Hedberg to do his presentation.

Eric: Thank you so much, Carrie, for that kind introduction, and here we begin our Level II series on the basic mechanics of power analysis. Just again to emphasize that this Level II sits within a larger sequence of courses about power analysis. Level I, which is a broader introduction to power analysis for program evaluation, which you've either attended or you've seen the previous recording for. And after this, more detail is provided in the Level III webinar, which is the more applied power analysis. So I encourage you to check out the other two levels, in order of course.

In this level, our overview is, we're really going to start getting into a little bit more of the meat, a little bit more of the mechanics, sort of how power analysis works and sort of why it's important. In Level I, it was really sort of a sales pitch to sort of really encourage you that power analysis was an important thing to do and now we're going to get into much more of a certain detail as to why.

So we're going to briefly go over some of the introduction, some of which you've seen before, but now we're going to add some additional detail. And some of this additional technical material is going to be important for program staff and evaluation staff to sort of understand and at least sort of become a little bit more familiar and comfortable with a little bit of the jargon.

The big jargon here is we're going to have a pretty deep discussion about Type I and Type II errors, which are sort of the two ways a study could be wrong and we're going to sort of talk about how each of those errors can threaten a study and how power is very much related to the Type II error. It's sort of the mathematical complement of it.

So Level II is part of, again, this sort of structured three part series. The video before this for Level I sort of really defined and tried to sell you on the idea of statistical power, and it's intended for everybody and after this Level III is going to get much more applied.

We're actually going to, you know, work with some numbers. We're going to work in and really sort of explore how power analyses relate to the design of the study, the analysis, the sample, and all that. And that it's going to sort of end with a little bit more of a qualitative component of really how to sort of write up a power analysis. Or, for those of you watching who are reviewers of proposals, what to look for when you're reading a power analysis.

For the brief review, we want to talk about sort of what is power? And again, power in general is the probability that you'll be able to detect a statistically significant effect from your sample. And this analysis will really help you determine how large your sample size needs to be or, what the smallest effect could be that a given sample could detect.

But ultimately the act of power analysis is really integral to the planning of a study. The power analysis that you do before you collect any data is really a deep discussion that you should have with both the evaluators and the program staff about what you think is going to happen and they're really about estimates of the future.

And because of that, because data are not free, often, power is about using evaluation resources wisely. And I'll say it now and I've said it many times before, as

I’ve said it many times in the past, power analysis should be estimated prior to conducting your study.

So why is power analysis important? Because suppose there’s an actual impact up there? Your intervention is causing a positive effect. However, if you try to validate that intervention with a poorly powered study, you may not protect your impact and have your result as no program evidence on the official record.

However, if you do have a true impact and you have a well-powered study, you are able to detect your impact and you are able to report to the world with reliable evidence that your study produced evidence of program impact.

But again, most of these studies deal heavily with statistics and statistics is the sort of game that we play where we have a population of interest that has a parameter of interest. You know, what is the impact of an intervention on this group of people in the entire world for now and forever?

However, what we're stuck with is a small sample that we used in our study. From that sample we calculate statistics, our best guess about what those parameters are. And the statistical test is that sort of transcendental act of taking our sample and trying to estimate what the parameter is going to be.

Unfortunately, we actually don't know what this parameter is, which is what makes power analysis so important. Because it's one of the things we can do to lend confidence in the study that we eventually conduct, that we have a good estimate of what we think this parameter is.

So, this has all been revealed up until this point things that we talked about in the last. Now we're going to start diving a little bit deeper into really what we mean by reliable, really what we're trying to do with our study and this brings us to the Type I and Type II errors.

Some of you may have heard about these, or, you know, been tested on these in your statistic classes in college and any level really, but these are often forgotten like so many bad memories. So let's sort of dive in and really sort of think about what are the two types of errors.

Ultimately, a study, a report, a pdf you download, or a paper you read in a journal, can either be correct about the world or it could not be correct about the world. So it could be right in a couple of ways. Say the paper says that there is no impact. You know, this study says we really looked at this particular intervention, we don't find evidence that it's effective.

And out there in that sort of parameter that we will never know, maybe that is the correct conclusion. Another way one of these studies could be correct is that if it reports that there is an impact. That, you know, this intervention does have some kind of positive impact on the population. The paper says it's so, and it could be right that that's actually true.

On the flipside, however, there’s also two ways that any study or report or analysis could be wrong. The first way it could be wrong, and we call this Type I Error, is that the study does assert that there is an impact. However in fact, out there in the world, in the parameter that we never know, the actual truth is that there is no impact. We call this Type I Error, and we give it the little Greek letter Alpha.

The second way we could be wrong, Type II Error, is that suppose there actually is an impact, but for one reason or another our study makes the conclusion, writes down on paper, that there is no impact of the intervention when in fact there actually is. That is the second type of way we could be wrong. And we call that Type II Error. We’ll give the little Greek letter Beta - A for I, B for II. And so that’s how we’ll sort of talk about this.

Power analysis is ultimately trying to control this second type of error. We’re trying to minimize the chance that we missed something, when it actually exists. So, let’s sort of quickly review Type I Error.

This is ultimately what we talk about when we have discussions about statistical significance, stars next to numbers in statistics tables, and so on.

So, Type I Error is the probability of wrongly detecting an impact or concluding that an impact exists when in fact it does not. We call it Alpha. The convention out there is that we don’t want this to be any larger than 5 percent. We don’t want any more than .05 probability that if we make an assertion that there’s an impact, that that assertion is in fact false.

Oftentimes what analysts will do is they’ll incorporate this into the estimation of what’s called a p-value. You know, a lot of statistics programs will sort of give you all sorts of numbers and tests. And the final column, as PSS calls it, sig, other things just call it a p-value, will give you some number. And your well-trained analyst is often scanning that column looking for anything less than .05. Because what that does is that connotes that it’s statistically significant.

However, p-values are often misused or misunderstood, and so I want to stress that the p-value is not a measure of any kind of practical significance. If I were to analyze the data that's as large as the population of the United States and find that someone's, say for color, is statistically significant in its prediction of happiness, it probably would be a small impact practically. But because the sample's so large it would be so precisely measured that we would be statistically significant.

So I don't want anybody to confuse a very small p-value with meaningfully important. And so, we just want to keep our eye on that. And then there's a small note where there's been lots of discussion among the statistical community about actually lowering the convention from .05 to .005, which would make it a lot harder to get that tag of statistical significance.

So, sticking with the world of .05 as the convention, I think we need to be very explicit with what exactly this means. Again, going back to Level I where we had that image of sort of the jar of jelly beans and we wanted to see what was the chance we'd get a grape jelly bean. We need to realize that our data, or your data for the sample, for your one study, represents one out of a very large number of possible samples. And so, what does Alpha .05 really mean?

Well, it's a fraction 1 out of 20 which means that 1 out of every 20 samples will yield a statistically significant result even from a population in which there is no impact. So if we accept .05 as our Type I error rate, many studies – 1 out of 20 – that study a particular intervention, may claim statistical significance when it may or may not be so true.

And so, we have to be very careful when we use our p-values and always sort of judge the power of our study - which is what we're going to talk about soon - as a way to better understand the statistical validity conclusion reliability of our study.

All this discussion of Alpha levels and p-values could be thought of in another way. Often sometimes people are looking at that column in their statistical output, looking for small p-values. Well, another way to go about making an assertion that there’s an impact is to actually look at the hes [phonetic] statistic of a particular result. And often these p-values of .05 are associated with what people who had sort of Stats 101 may remember as being critical values.

These are essentially numbers that a test statistic must exceed for us to be able to reject the null hypothesis that there is no program impact. And these are important ideas to see because these are often also in statistical output, but these numbers that we see here are actually important parts of power analysis.

One thing I want everyone to sort of see is as we look at this table, let’s sort of go row by row here. We see that, you know, alpha is at say .1 a ten percent chance of being wrong, has relatively small critical values. If we have a one-tailed test, right, where we sort of know for a fact that the program will have a positive effect, and we know sort of the negative impact, the test statistic has to be 1.28.

If we do the more typical, and what I recommend, two-tails test, it’s a larger threshold, 1.64. And as the Alpha gets smaller and smaller, you see that these critical values get bigger and bigger. These become harder and harder goals to meet in order to claim statistical significance.

I caution everyone from using say a one-tail critical value at .05, to 1.64, because you see it’s actually the same as a two-tail critical value of .1. And so, I often recommend two-tail tests only, and at least Alpha of .05.

So we’ve just had a deep discussion of what Type I Errors are. Now we’re going to sort of talk about this other error. The chance of being wrong and saying that, you know, there’s no effect, when actually there actually is. So, we call this Type II Error and we give it the symbol B or Beta.

Type II Error is the probability you fail to conclude that an impact exists when in fact out there at the parameter we will never know, it actually is true. So this requires actually a different thinking about probability. Not only is there the set of samples associated with a world in which there's no impact, there's another set of samples, another curve, associated with the world in which there is an impact.

And so, power analysis, as we'll see in the next slide, is ultimately drawing a bunch of curves, cutting them up, and seeing what the areas are. Because power – that probability that you'll be able to detect the effect should it exist, is actually the complement, or 1 minus, your Type II error.

So let's visualize this and sort of, you know, break this down a little bit. In the figure here, you see two curves. The solid black line with the two tails shaded in, this is the curve that's ultimately represented in tables in the back of virtually every Introductory to Statistics book ever printed.

And what this is, is this represents the sampling distribution if we assume no impact. It's centered on zero as you can see and sort of curves and what's shaded are the numbers that are the critical value or greater. So you see that they're shaded at about 2. That's that 1.96 that we talk about. The second curve, the one that's a little purple and blue, this is the curve if we assume that there is an impact.

And you see, what happens is, is we can slice that curve vertically at the critical value of about 2. And we see that the purple, or you know the grape jelly bean that we want, and beyond, that's our power. That's the proportion of that curve that is greater than the critical value, and thus statistically significant. That lighter blue, those are the possible samples we could get, even though there's actually an impact, that are before the critical value and would not be assigned the label of statistically significant.

And so power analysis is really figuring out where to place that sort of blue purple curve, and the further to the right it gets placed, the more powerful your design.

And so, as you see that, you know, our choice of our Type I Error, our choice of Alpha - do we want to have a 5 percent chance of being wrong, or a 10 percent chance of being wrong, or a 1 percent chance of being wrong – has a major impact on power. You could automatically have a more powerful test if you lower your significance threshold. If you say, you know what? .05 seems kind of harsh; I want to have a 10 percent chance of being wrong.

Well, you've automatically increased the power of your study. However, you've also increased the chance that you'll make a false statement. So there's sort of a risk and reward balance that we often have to think about in designing our studies and thinking about power analysis.

Here I sort of show how this sort of works in real time. Here we have four plots and these four plots are represented by, you know, studies that may have small or large impacts. Those are the two rows. The pictures in each of the rows represent the exact same data, the exact same tests.

However the columns are alternative choices as to what we want to set our Type I Error at. The first column, Alpha .05, is sort of our convention; and the second column, Alpha .01, is a more conservative - it's a harder test to beat.

Otherwise, except for that difference, if you look at these different columns within each row, it's pretty much the same curves. Except you'll notice that in the second column, Alpha .01, there's a lot less purple than there is in the first column with the Alpha .05. This is really just to sort of show what I was talking about at risk versus reward.

Oftentimes a similar thing can happen not only with effect sizes but also with the sample sizes. Suppose we have a large sample or a small sample, those are sort of the two rows. Again, the same mechanics of choosing your Type I error can have a similar impact as the difference in effect sizes.

So why would you want to conduct a power analysis? Just as a quick reminder, it’s all about planning, it’s all about resources, it’s all about maximizing the chance that your study will be able to detect the impact that we assume is out there.

And so we want to really think deeply during our power analysis process about what we expect the impacts to be and then we want to think deeply about the design of the sample. How are we going to create our treatment groups? How are we going to create our control groups?

And then, as we do this, as we sort of, you know, meet with the budget personnel on your project and say, I think I need a study that looks like this, and the budget people say this is too expensive. Well, if you make changes based on the budget, you should redo your power analysis to make sure that while you preserved your budget, that you’ve also preserved your statistical power.

And so, the power analysis is not a checkbox. The power analysis is not this quick thing you do just to make sure you cover that base. It's actually an integral part of all the other planning processes you have during your study design and so it's an important check.

Which means, again, this is something that needs to happen prior to collecting any data. This is, you know, you want to plan, you want to budget, you want to do all these things, you want to go through these activities before you ever collect any data.

Because after you've collected your data you've moved from the world of possible designs in the future, and much like quantum physics, now you've observed your particle so its properties are set and you can only really deal with the data that you have. And any kind of post-hoc power analysis that some reviewers often ask folks to do, are actually far less informative than most people think.

So how do we design a highly powered study? As I sort of think through these things, I think sort of there's three critical aspects. There's the design of the study. This is sort of the research methods stuff. That there's how you create your groups, what measures do you use, do you want to analyze over time, do you want just one point in time study?

And that study design is going to be integrally related to your statistical analysis, but it doesn't determine the statistical analysis. You still have choices in many cases with how you analyze. So those two really interact.

And finally interactive, this third component, which is your expected impact, what you're actually trying to detect out in the world. These three ingredients work together to produce the power of the study. Which I've tried to represent in this chart, is this sort of study design will then influence your statistical analysis where you still have some choices and then those two things go in and try to detect your impact and those three things produce the power of significance.

And so it's really it's the blue, it's the gray, it's the green, that really work together to produce power.

And another way to sort of think about this, is this sort of balance between power effect size and sample size because most power analyses hold two of these things constant, and try to figure out the third. So if you know what your effect size is going to be, and you know what your sample size is going to be, you can easily ask what's your power going to be?

You could also say, I need my power to be .8, I have a pretty good idea of what my effect size is going to be, then you can ask the question what sample size is going to satisfy those other two criteria. And, I bet you see this coming, you could also say, I know my power, I know my sample, what effect size is required to complete this triangle? And so, you know, power analysis is really sort of balancing all of these things.

As we mentioned before, larger impacts tend to have more power. Because the bigger the impact, the larger your statistical test is going to be, holding everything else constant, which means you have more power. Most cases, larger samples tend to have more power, ore information.

When we get into sort of more complex samples where, say you have students in schools, or you have patients in clinics, things can get a little bit more complicated. But as sort of a broad stroke, more data is more power.

So how we perform a power analysis really involves mostly straightforward computations, but some of them are a little bit more complex and so typically you need specialized software. Power analysis is sometimes hard to do in, say, just Excel. And so, what you will need to do or work with your statisticians to do, is to employ some kind of statistics package like SPSS, Data Stats R, or Python.

And there's even some software that's freely available that you can download for most of the platforms, like PowerUp!, which is ultimately Excel with added macros built in or a freestanding software called G*Power. And then there's lots of other resources available online.

So, there's some specialized software that needs to be required in order to perform these things.

One thing that I want to mention is that it's really important that if you do try any online calculators, that you need to be sure that the calculator is designed to detect a difference between groups, and not a, say, a percentage of people who say yes with a certain margin of error. There's lots of sample size calculator out there often used by, say, marketing firms and they're a totally different analysis than the sort of treatment versus control of your – focused on it in evaluation.

So, again to sort of talk about power analysis as sort of balancing in this triangle, we're sort of talking about Power, Significance, Effect Size, and Sample Size. The reason why Significance isn't really labeled in this graphic is we generally are going to set that at .05 and not touch it. And so it's really sort of a discussion that sort of plays with how much Power we have, what's our level of Power, and just like Alpha at .05 is the convention for the Type I error, a Power level of .8 is the accepted convention for what we want our study to be powered at.

You can always go higher than this. Having a Power of .9 gives you some more assurance; a Power of .7 gives you less assurance. And anything less than .7 is, you know, you're really sort of rolling the dice. So I highly suggest that everyone stick with .8 or higher.

And again, Effect Sizes. We'll talk a little more deeply about this, but, you know, this is basically how much difference does your intervention make, in sort of standard deviation units? And this is often the place where a lot of people will fall in the trap because they'll sort of be constrained by the budget, which means they balance this triangle by assuming a much larger Effect Size than is plausible.

So your expectations of what your Effect Size should be really needs to be tempered by experience and literature. And then finally Sample Size, which is often translated into budgets, how many participants for your study are available.

And so in this example say that there is a mentoring intervention program, and it's for high school age youth, and they want to determine if their program is having an impact on the students', say, school attendance. Well, we can sort of talk to our school district and we can verify - we have about 250 students that our program can impact and then we want to compare them with another 250 students in a control group.

So one key question we could ask is, if we think our Effect Size is .15, how many students are needed to detect an Effect Size of .15? If we want a Power of .8, an 80 percent chance of detecting this effect, and we set Alpha at .05, we would enter them into a calculator – I'll show you an example in a little bit – and the answer would come out that, with just that information, 350 students in the treatment group and another 350 students in the control group.

So our 250 students is not enough. We need more students than we have, to get our Effect Size of .15. If we were to enter this into a calculator, say

PowerUp!, you would see a table like this. And PowerUp! is essentially an Excel file and so you would type in .15 as your Effect Size, and .05 for your Alpha, and type in what you want your Power to be and a lot of these other things you'd sort of leave alone. And it would sort of give you your N Sample Size as 700.

For a second example, let's say we're sort of stuck with the students we have, 250, and so we say, okay, how big does the Effect need to be? And so again we're sort of holding Power and Sample Size constant and now we're saying okay what does the Effect Size need to be?

Well, the answer is about .18 standard deviations. So, you need a little bit more to be able to detect - we need - [unintelligible]. The answer is, so without controlling for stuff, we actually need a larger Effect Size with a smaller Sample. And again this makes sense because with a smaller sample we can't detect such small Effect Sizes, so it needs to be larger than what we were working with before. However it's not that much larger, so I think you would be in pretty good shape.

Here's the examples in the PowerUp!, again, where we sort of see the different levels of Alpha, and we enter the different things, and the final row is what's called the MDES which is the Minimum Detectable Effect Size, which is the .18.

This graph is from another piece of software called G*Power. And again this sort of shows the relationship between Sample Size and Effect Size. We see that for the very, very small Effect Sizes towards the left on the X axis, the .05, boy we need a lot of Sample to be able to detect that. It looks like 12,000.

However, as the Effect Size gets larger and larger and larger we see that actually start to slope down a little bit. And so, if we think the Effect Size is, say, .15, we don't need 12,000, we need more like 2,000 people for our study and it starts to sort of flatten out. So this again gives you sort of a relationship between the total Sample Size and the Effect Size.

Let's go through another example. So now we have a job training intervention for veterans, and we want to determine if the program is increasing the average wage growth for program participants. We again, the Power of .8, we again assume a Type I error of .05 and we need to figure out what's sort of the minimum effect size if the number of trainees who served in a program is 80 veterans.

Well the effect size for a study group of 80 is .32. So, a smaller sample needs a much larger effect size. Here's a picture from G*Power which is another software that's often used and you see that G*Power even shows you, with different colors, but sort of these two curves that we're talking about. It actually tries to visualize this exact power.

But you see in the sort of lower half of this panel, we have input parameters where we talk about Alpha, the tails, the power. We tell how much sample that we think we're going to have. And then in the output it gives all sorts of statistics including, you know, non-centrality parameter which is really what we expect the t-test to be. The critical value that we talked about before and then ultimately the effect size which is, you know, what effect size satisfies the rest of the table.

I’m going to pause for a second, I’m just taking a quick break before I go into the next example. Okay, here we go. A third example. A health education program for low income residents wants to determine if the program is having an impact on basic health knowledge levels. Same parameters as before, .8 Power, Significance we want to assume an Alpha of .05, and we’re going to go into with an Effect Size from previous studies of about .12 standard deviations.

So, what’s left in our triangle? What does our Sample need to be? And here we need to sort of say, hmm, how large does our Sample need to be? And it looks like if we were to plug everything in, that small of an effect size needs a much larger sample. So now we’re talking about 500 residents.

Or we could sort of say, you know, if we have 500 residents, how much does – how large does the effect size need to be? And again, we get the same answer as we did before – 500 residents, the minimal detectable effect is about .13. And here we have a final example from PowerUp!, the software which again shows you all the sort of numbers that you enter into this model, and then the final MDES of .125 that you would get out.

So to sort of wrap up this power and the sort of working parts of the power analysis, is I want to sort of bring up something that I've really brought up during the first webinar which is the sort of microphone analogy. Which is basically, your sample, the data that you have, the data that you're working with, basically is your detection instrument. It is your microphone that you are using to try to detect the impact of the program which in this analogy is analogous to the sound.

If your program is going to produce a very, very large impact, then you do not need a very large microphone; a very small sample will probably suffice. Please do the power analysis to make sure. If, on the other hand, you think that your program is going to have a meaningful, but slightly more nuanced impact, you need a much larger microphone in order to detect that.

To summarize what we've talked about over the last thirty minutes or so, let's sort of just start at the beginning. I'm going to pause and try that again. So to summarize everything that we've sort of talked about over the last thirty minutes, let's sort of summarize our Level II webinar here.

First, Power is the chance to find a statistically significant impact. In this set of slides, and of course we sort of learned that Power is really this sort of interplay between what we want our Type II Error to be, and what we want our Type I Error to be.

And well-powered studies, just as a reminder, will increase the confidence in our evaluation findings. Because if we're powered to detect even a small effect, and we do not find an effect, it can help us conclude that maybe this program as it was evaluated may not have an impact versus this more nebulous gut-wrenching conclusion that god, maybe there is an effect but maybe we didn’t have enough sample. That’s not where we want to be. So this is what the power analysis is trying to protect us from.

And the Power analysis and the Power of a study is really the sort of dance between the design of the study, how you analyze the data, and the impacts that exist that you’re trying to detect. Broadly speaking, larger samples can detect smaller impacts or larger impacts, however, the larger studies are really relegated to only having the ability to detect these larger impacts.

Power analysis helps you minimize this Type II error, this Type II error which is missing an effect that’s out there. And then of course, Power analysis much occur before you actually conduct your study. In the many examples we’ve talked about, we’ve used PowerUp! which is a great piece of software. It’s available at causalevaluation.org. They have an Excel version, they have an R version, they have an online version.

And beyond the designs discussed during this series, they also have programs that allow for moderation, which is basically subgroup analysis, or even some mediation analysis, to work on power. If your designs are not going to involve complex samples without clustering, G*Power, which is available on the server website, is another great piece of software. It’s available for Apple computers as well as PC computers.

And in addition to comparing the means between a treatment and control group, it also has routines for various regressions, ANOVA, [phonetic] and even some probability models as well. Beyond software there’s also a lot of books out there, one of which I wrote, which is the bigger picture because these are my slides.

But other really great books out there. The Aberson book is a really good introduction to power analysis that has far fewer equations than my book, but I think really sort of helps people get into it. The Cohen book is, you know, this is the third or fourth edition of Cohen’s classic work of statistical power.

Much of this book is a series of tables for various situations. So if you can't get your hands on software or you find yourself designing an evaluation on a desert island, having this book handy would be a good thing.

There's the book I wrote, which is available from Sage. There's another Sage book, called *How Many Subjects*, which is a great also primer on this. And then there's a more detailed book, *Statistical Power Analysis for the Social and Behavioral Sciences*, which has more equations than my book does, but covers many, many more designs including structural equation model and all sorts of crazy stuff. So these are some really great resources. Again I want to thank you for tuning in to Level II of III of Statistical Power Analysis for Evaluations. Thank you.

Carrie: Thank you so much, Eric, for this informative class on the concept of power and its importance in evaluation, planning, and design. I hope everyone attending now understands the concept of power, and when, why, and how to conduct power analyses. As I mentioned at the beginning of this presentation, this class is the second in a series of three classes on statistical power, so we hope you will take the time to view the other classes which provide additional context and instructions on the process of estimating statistical power.

[End of File]