

Quantitative Correlation Research Design Discovering Significant Relationships

Welcome to all the students that have, uh, uh, come to listen to the wondrous Mysteries of Correlation design.

Um, so we're gonna talk a little bit about quantitative correlation research design tonight.

This is not gonna be a terribly esoteric, um, presentation.

This is really designed for students that are developing a dissertation, uh, around quantitative correlation research.

So it's very specific, and I will give examples and I will talk a little bit about how to, um, address this particular type of design as you're going through and doing the various chapters within your manuscript.

So, with that, we'll jump right in.

I always have to start with a little bit of disclaimer about what correlation isn't and what it is.

So first and foremost correlation is the study of Covance, or the, the concurrent variation of two variables in relationship to one another.

So you'll hear the term correlation.

You may hear me use the term Covance.

Um, and those are two terms that are almost synonymous with this particular research design.

You're also gonna hear me use the term inferential statistics on occasion, and that's really what a correlation design is.

It's a specific class of statistics that we refer to as inferential, meaning that we're inferring the value of another variable based on the state of the one we've got.

So you may hear these terms.

They all effectively mean the same thing, and I may even use them interchangeably, uh, as we're going through the presentation.

So don't let that confuse you.

Um, they're essentially, um, terms that you can use in your dissertation you can use as you're writing about your work.

Um, but more importantly, just recognize that they are all essentially the same type of research design.

What I want to do is talk through the various elements of a correlation research effort.

Um, so starting with the problem and purpose statement, how do you structure a problem and a purpose statement to indicate that you are, uh, considering a correlation design.

Um, we'll talk a little bit about how, um, to describe the methodology, not the method, but the methodology and the design alignment within, uh, particularly in chapter one, and how to describe the

variables and the questions and the hypotheses in your chapter to be consistent with a correlation design.

This aspect, uh, called alignment, but across the various aspects of your research proposal is very important.

Um, we're talk a little bit about, um, population and sampling, how to describe those specifically for correlation research.

We're gonna spend a little bit time talking about sampling techniques.

Now this is a little bit off topic for correlation design, but it's one they find many students struggle with as they're putting together their, um, population sample for this type of research.

And so we're gonna talk a little bit about the sampling techniques that can be used for gathering data and information for, um, correlation research.

We're gonna talk a little bit about sampling controls and why they're important, even in correlation design and when they have to be used and when they don't have to be used.

We'll talk a little bit about recruiting and ex uh, accessing your sample.

We're gonna talk about collecting data from the sample.

And there's two different types of statistics that are very important in the correlation design.

One is descriptive statistics, and the other is correlation statistics.

And for a correlation study to be, um, substantive and to have value, you have to have both descriptive statistics and correlation statistics in your, uh, writeup.

And we'll talk a little bit about how to interpret the results.

So a lot to cover. So, um, hang on.

Uh, also know this is being recorded, so you can also always go back and look at it.

And I know there's a lot of information on these slides.

And these slides will be available to you, um, after the presentation, both in recording form and in a PDF form.

So you can go back and refer to them for notes.

And I'm gonna apologize up, apologize up front for the coughing.

I'm still doing a little bit of that after a couple weeks.

So, key objectives of correlation research.

Um, what is included in the design? So first of all, um, the purpose of correlation design is to identify relationships and specifically, uh, patterns within the data.

Um, it is different than regression analysis.

It is different than, um, ex post facto analysis.

It's specifically look ident specifically, um, developed to look for relationships between variables.

Um, correlation design, most correlation designs are not, do not predict outcomes, but they are predict, uh, they do create a prediction.

There's a distinction.

Um, so they, they, while they will predict an outcome, they, you cannot assert that that outcome is because of the other variable.

This, so you'll often hear the term correlation is not causation.

So while you can predict the outcome of the other variable, you cannot assert that because of a B will happen.

B happens in relationship to A, but it doesn't necessarily mean there's a, a causal effect there.

Um, also, um, validity testing.

Uh, we're gonna talk a little bit about how to do validity testing for different types of designs.

And we're gonna talk a little bit about, uh, exploratory analysis and data reductions.

So I saw there were a couple of questions that popped up.

Um, and if you guys are okay, uh, what I'm gonna ask you to do is hold the questions.

You can go ahead and post 'em in the chat, um, and we will get to them, but we may get the end.

I wanna make sure we've got enough time through, uh, all the slides materials.

We, um, but just to reiterate, there will be a recording and you will have access to these slides when we're done.

So don't worry about having to try and screen capture all the slides and, and take lots and lots of notes.

So let's start with a problem statement. Excuse me.

When you're formulating a problem statement, any problem statement for any type of research, um, there's always gonna be five elements within that problem statement.

Now, it's important to remember that a problem statement is different than the issue.

Um, most of us, when we begin developing our chapter one and are formulating the background of our problem, the background of our problem is identification of the issue that issues impact.

A research problem is identifying what information we need to collect and analyze from the population in order to mitigate that issue.

So, for example, if my issue is my shoes keep getting untied, what information do I have to collect and analyze about my shoes or shoe strings to help mitigate the problem of my shoes being untied? The purpose of a problem statement is to identify the variables that you need to collect or the center of interest or the central phenomena of your study.

Depending on your type of design and methodology, you have to identify where you're collecting the data from and how that data

population might be delimited so that, uh, it's, it can be, it's attainable, but there's an important thing that occurs in every problem statement, is that there's some indication of the design that you plan to use to collect the data.

That indication is usually some type of key word that is synonymous with the design type.

So in the case of correlation, if you remember, a correlation design is the study of two variables and their relationship with each other.

Um, in the problem statement, um, I always look for the word relationship or Covance or correlation as a key word that indicates that this particular design, uh, is a correlation study.

Uh, and if I'm looking for a correlation study, I'm always gonna also be looking for at least two variables identified within that problem statement.

In this example, I'm talking about shoelace materials and untie rates.

Is there a relationship between the untie rates and the, uh, shoelace materials? Um, in my population? Now, what is my population? In my case, the population is shoes because that's what's having, which is coming untied.

And I've delimited that population to be shoes that are worn by North American, uh, marathon runners during races.

So obviously my issue, which I didn't state here, was that the shoelaces are becoming untied for marathon runners during races in North America.

That's an issue. What I'm looking here is to look at a specific set of variables related to that issue and see if there's a relationship between those variables.

This allows me to, um, uh, analyze data and to be able to draw conclusions that can help mitigate the, the issue that's been identified.

So, identify the issue impact, and then problem statement is, uh, contains that keyword and the variables.

That's what we're looking for. So what's the purpose? Uh, the purpose of correlation research is to examine a relationship between two or more variables.

Correlation research seeks to determine whether or not, or, and to what extent that relationship exists.

So, for example, there might be a positive relationship as one variable increases, the other variable can increase.

There could be a negative relationship, one variable increases as the other variable decreases, and that relationship may be linear, it may be non-linear.

But in any event, the purpose of a correlation study is to examine if a relationship exists, how strong that relationship is, and what direction that relationship is.

So there's an example, uh, purpose statement here.

Purpose of the quantitative correlation study is examination of a relationship between transformational leadership and emotional

intelligence in visual managers in southeastern United States.

That could be a purpose statement, something that you're seeking to understand within your research.

Let's talk a little bit about research questions, um, research questions for correlation research, and for most types of research, typically there are gonna be two or more research questions, um, that you're gonna be looking to examine with your data.

Uh, my example here, I just have a single research question.

What's important, um, with any research question is that I formulate a hypothesis.

And so for this research question, one, I'm formulating two hypothesis.

Um, one is considered the, the, the hypothesis, and one is the null hypothesis or the, um, statement that says this isn't true.

So my research question in this case, is there a correlation between social media usage and self-esteem among adolescents? That's a question. A hypothesis is a statement, though it's not a question.

It's a statement that says, increased time spent on social media platforms is positively correlated with lower levels of self-esteem among adolescents.

So you can see it's a restatement of the research question in the form of a proposed outcome.

You'll also notice on annu hypothesis, I state that opposite, that there is no relationship.

And the intention of a correlation study is to examine whether I can prove the hypothesis or not prove, but I can show, I can demonstrate, um, uh, confidence in the hypothesis or in the null hypothesis.

And so that's the, that's the intention of correlation research is to examine those hypothesis in a, in a disciplined way.

So in quantitative research, um, so we've talked about the problem statement.

We've talked about the purpose statement, we've talked about the research questions and hypotheses.

Um, it says the slides are no longer visible.

Can people see the slides still? I was gonna say that some people can see them still.

Yes, we can see it.

Okay. Okay. I'll, I'll keep going. And Dr.

Hannah, maybe you can pop on, pop back in and maybe you'll be able to see 'em again.

Um, so let's talk a little bit about the, um, types of variables that you may, uh, ex may use within correlation research.

Um, the two most important are the independent and dependent variables.

Um, as I said, most correlation research is done with a at least variable pairs, and one of which is the independent variable or that variable that you change, and you examine the results of the independent variable.

Um, so that, that's two types of variables.

There's some other types of variables that you'll often hear, uh, described within correlation research.

One is, uh, uh, control variables.

You'll hear about moderator variables and mediator variables.

These exist, uh, these types of variables are, um, adjacent variables to the dependent independent variables and may influence the relationship between these two variables.

So if there isn't a, a linear relationship, or the linear relationship is inconsistent across the entire span of the data, then you probably have mediator, mediator, moderator, or control variables present in your analysis.

And it'll be important for you to identify what those are that could possibly be influencing your correlation study.

Let's talk about these a little bit, uh, more.

So this is probably the most important thing to really understand, and this is the most important thing for you to describe in your dissertation.

And I'll tell you why in a moment.

But every variable falls into one of these types of variables.

These are variable types.

Um, variables all have a nature or a type.

They're either ratio variables, interval variables.

So like, um, an interval may mean, uh, a li scale, you know, it's numbers, 1, 2, 3, 4, or IQ scores that don't really have a relationship to a zero value.

Um, ordinal variables, meaning that you, there's specific order or sequence to the variables.

So, um, uh, bachelor's comes after associate's, master's comes up, bachelor's, uh, doctor comes after master's.

There's an order and sequence those variables, nominal variables.

These are, um, variables that have discreet values, but don't have a necessarily a particular order.

So, male, female, um, single married, divorce widow, they, they're categories of, uh, that are, uh, captured in variables are called nominal variables.

Discrete variables, meaning that, um, there's, there's discrete numbers.

There are no fractional values in those, and you can have continuous variables that have fractional values in them.

Now, why are these important? Understand, and why are they important to describe in your proposal when you're identifying your variables as part of your, um, population sample? Um, it's important to identify the type of variable you have because the types of variables determine the type of statistic that you use when you are looking for a relationship for these variables.

Not all statistics, um, are the same.

Uh, they are, there are different, um, population and sample characteristics that determine which statistic is appropriate for your analysis.

And if you don't describe the types of variables you're using, and you just choose a, a statistic, like a Pearson correlation statistic, it may not be valid because you're using the Pearson.

Uh, if you go into SPSS or Excel and put the numbers in, it'll pop out and give you a value.

But if you're not using Pearson, uh, on the right types of variables or the right types of variable pairs, it may be an erroneous answer.

So one of the first questions I always ask students in oral defenses is, okay, this is the statistic you use.

What were the variables that you put into it? And were those variables appropriate to the statistic and why? Um, that's an important thing to understand and be able to describe when you're doing correlation analysis.

Know your variables. Know what types of variables you're using.

Variables must be defined.

They must be statistic or specific, and their sources must also be clearly defined in your, uh, research proposal.

Um, I've had a lot of research proposals that say, this is my variable.

This is, this is the data value that the type of, uh, data that I'm gonna have in that variable.

And that's as far as they go.

And, and you realize there's absolutely no way to collect that data.

Um, I had, uh, and just fictionally, uh, we had a great conversation about what the right breakfast cereal was for US Presidents across time, fascinating study, uh, could, you could, um, write volumes and books on that study.

There's absolutely no way to collect that data.

So you're very, whenever you're describing, uh, data for a correlation study, you must be able to describe the types of variables.

You must describe, the variable names, the types of variables, and where those variables are coming from and how they're related.

So if you're pulling variables from two different data sources, how can you match them up to do a correlation? So that's gonna be an important part of the, uh, preparation of the data for correlation analysis.

So I always think, what are my variables? What types of variables do I have? And that defines my correlation statistic.

And where am I getting my variables from? I've had several students that have phenomenal proposals and great ideas that they wanna study, but they don't know where they can get the data from, uh, where they, they want to try and get it from a secondary source, and they can't find a secondary source for the two variables that match.

You know, you can't do a correlation analysis between, uh, two ratio variables.

One from 2016 and one from 2019.

They expect to have a useful value.

They have to be paired data that you can, you can do analysis with.

What's the source of your data? Is it connected? So we've talked about problem and purpose.

We've talked about methodology and design, and we've talked about variables, questions, hypotheses.

Um, gonna pop into the next section, and we'll go through this a little faster.

'cause this is gonna talk a little bit about, um, some of the specifics of actually running a design.

Let's talk a little bit about how you describe the population and sample.

Um, this gets back to the, the, um, variable descriptions we had a moment ago.

When you're preparing the population and sample in chapter one, and in chapter three, you have to be able to describe where are your, where's your data coming from? What's the size of that data? What access do you have to the data? What kind of demographics do you have in that data? So for example, um, if you're studying the characteristics of, um, uh, hair, um, hair design, you know, obviously not something good to say for me, but, um, hair designs in different populations know where you're gonna be able to get that data from.

Is this a secondary data source? Or you're gonna have to collect it from a primary data source or do a survey instrument or some other way to collect that data, uh, in a primary manner.

And if you don't have uniform access to the population, and this is, this is important, if you do not have uniform access to the entire population, you cannot do a random sample.

Um, so you may have access to two thirds of the population, uh, and not have access to a, the other third of it, unless you deli your population down, there's no way to do a random sample of that population.

So keep keep in mind, as you describe your population, you're starting to influence the way you're going to be able to sample it.

And it's always important even with any, um, population to be able to understand the demographics of that population.

So if you've got a population that, um, population that you're describing is 50% men and 50% women, and when you sample it, um, you get all women, uh, would that sample be representative of the population? The answer is probably no.

So understanding the demographics of your population, that 50 50 ratio gives you what are called sampling controls that you can use when you're sampling, so that you can assert later on in your analysis that your sample was representative of the population.

So again, I just use a gender demographic, but you can use age demographics, you can use educational demographics, you can use, um, tenure demographics, you can use anything you want as long as you're gonna also collect it and measure it as you sample so that you can make sure that's sample approximates the distribution of the demographics in your population.

That's gonna be an important thing to do at a non-random sample to make sure that your sample is represented with the population sampling.

We'll talk about that a little bit here, but, uh, identifying a proposed sample, you're always gonna be looking for a minimum sample size.

I think people are familiar if you've done statistics at all, you understand what the concept of a power analysis, when you understand the, the, once you define the size of your, um, uh, delimited population, in order to understand what the minimum sample sizes you're gonna be able to do to get a significant sample, um, you have to do a power analysis to determine that minimum sample.

Why is that a minimum sample and not the def definition of a sample? Um, you could hit the minimum sample size and have a statistically significant sample, but if the sample isn't representative of the population, then it, you're not done sampling.

You may have to continue sampling above the minimum to get to that distribution of demographics that match your population so that you can not only have a statistical statistically significant sample, but also be able to assert that your, um, the analysis is representative of the population because your sample's of the population.

So whatever you describe the sample, particularly in a proposal minimum, um, always describe the minimum sample size based on the power analysis.

We'll talk about that more here in just a minute.

Um, let's talk a little bit more about a population.

This, this big red box that's on here represents the entire population.

I may wanna study. This might be all the people in the United States, but I may not have good access to all the people in the United States, or for some reason or another, I may want to exclude, um, sections of the population that might, um, skew my analysis.

So if I'm looking at studying, um, the US population, I may choose to only look at adults and adults below, um, 62, so that from a city standpoint, they're an accessible, uh, population that has, um, minimal risk.

They're not a high risk population.

So delimited populations, you can delimit your population almost any way you want.

You can delimit it by a state, so you can get a state that you're interested in focusing on to avoid, uh, noise in the data because of

cultural variations or variations in location or climate so that you can have more consistent, reliable results coming out of your data.

The limitation is an important step when you're thinking about how to collect data about your population.

Going back to my shoe example, I was collecting information on shoe laces and shoes, but I was being specific, um, about delimiting it to marathon runners during races.

If I had not done that, I could still do a viable study on the relationship between shoelace material and untie rates.

But the, you can imagine the untie rates of adolescents versus marathon runners is probably different and it would skew and impact my findings.

So limiting that population, um, gives me a, uh, a group of individuals or a set of data that are gonna have a relatively consistent or reliable, uh, finding.

So when I do my power analysis, uh, to determine how big my sample needs to be, I should always do it on my delimited population, um, within the delimited population.

Um, I'm going to sample now.

As I said, the minimum sample size is always gonna be defined by the power analysis that I do.

There are a thousand power analysis tools out there, and we'll talk about a couple of those as we go through this.

Um, they're very simple to run.

You basically put the, the limited population size in, you put a confidence interval, and you put an error rate, and it'll pop out a minimum sample size.

So for correlation studies, there's a lot of different tools out there.

There's very sophisticated tools for doing, um, sampling.

But for correlation analysis, I would use one of the very simple, uh, three parameter tools.

Um, delimited population size, confidence, interval and error rate to get that minimum, uh, size.

Remember, you may continue to sample beyond your minimum sample size to make sure you get a representative population.

Um, but, um, understanding that minimum is the first step that you would go through.

So minimum number, here's a, here's an apple.

Here's an example of a population, um, that I want to make sure I'm getting a representative sample of.

So in this particular case, I'm looking at a, um, sample size of this population that has a distribution of shirt colors.

And so I wanna make sure my population, um, has roughly, it doesn't have to be identical, but approximately the, um, same kind of distribution that I'm looking to have in the overall population so that my sample is representative of the population.

And I also wanna make sure that I have, I understand the statistical significance of the minimum sample.

This particular case, I ran an analysis on 80% confidence interval with a 30% error rate.

Trust me, there isn't a chair out there that's gonna let you do that.

But, uh, 80% typically most of our research studies are gonna be a 90% or 92% confidence interval with an error rate between 5% and 10%.

Uh, that's always something to discuss with your chair.

And the determination of those values depends a lot on the characteristics of your population.

So have that conversation with your chair.

Have the conversation with your URM about what's gonna be a legitimate or acceptable confidence interval and error rate.

Um, but if you're having a really difficult time sampling or getting access to the population, you may, uh, lower the confidence interval from two per, uh, 92% or 95% down to 90%.

You may raise the error rate from 5%, which is your ideal up to 10%.

Um, you want to make sure that you always understand the significance of your sample.

And so whenever you're, and when you're done sampling, when you're done with your analysis, you're gonna recalculate that with your actual sample size so you can understand what you can assert your confidence interval and error ratings of your conclusions.

Again, there's, there's a number of different tools out there for doing, um, sampling.

So here's, or for determining sample size.

So here's, uh, quantitative research.

Um, numbers are collected from surveys of secondary data sources.

In a particular case, the population was described as, uh, design, uh, research, uh, that was color data that was collected primarily, or, and previously, there we go.

There's the word I was looking for previously, delimited, based on the number and variations in the population that could skew the results.

So I might throw out subsets of the data because they represent segments of the dataset that are gonna skew my analysis.

I may delimit my data in that regard.

Um, and so I'm gonna do my populate.

I'm gonna define my size of my population based on the delimited population size, and I'm gonna describe the demographics of that remaining population.

Again, the demographics can be anything, uh, any characteristic of the population that will help you determine if your sample represents the population.

You're then gonna take that population size, that delimited population size, and you're gonna do a power analysis to determine what's your minimum sample size you need to collect.

And you're gonna determine the type of sampling technique you're gonna use, convenience, random snowball, et cetera.

We'll talk about those more, uh, in a, in a different webinar.

But, uh, make sure that you use an appropriate sampling technique.

Again, discuss that with your URM to make sure it's appropriate for your population.

And, uh, you're gonna assert that through this, uh, sampling technique.

I'm gonna sample till I at least get the minimum sample size, and then I will continue sampling until I achieve the, the sampling criteria, the distribution, uh, criteria of my, um, in my sample to make sure it's representative of the population.

Always make sure that when you're describing the demographics and the size of your data that you're using, uh, authoritative sources.

Um, so you must have citations and references that support the assertions you have about your population size, the sampling technique you're using, and the demographics of your population.

Most samples, most initial samples, minimum samples you're gonna work with are gonna be a 92% confidence interval and a, and a 10% error rate in most of our business research with primary data.

But if you're working with secondary data that's in high volumes, you may get, um, 95% with 5% error rates because of the, um, uh, characteristics of the overall dataset, uh, to get a higher confidence result.

So, um, many times if you're pulling into your analysis secondary data from a government source, for example, the Department of Education or the, uh, department of Labor Statistics, you're gonna get some fairly large sample sets when you're sampling that, you wanna make sure you're getting at least 95% or 5% error rates, 95% confidence interval, and 5% error rates.

Again, these are bounding boxes, um, generalizations.

Make sure you discuss these, uh, confidence intervals and error rates with your chair and your URM calculating that value.

Um, excuse me, I've got two different, uh, tools here that I use.

I don't promote any one tool, but I do always encourage students to use the simplest tool you can find to get the job done.

Uh, don't overcomplicate it.

'cause every time you complicate it, you're, you're introducing an opportunity for error.

So probably the simplest calculator that's out there, um, that a lot of students use as a SurveyMonkey calculator, you'll notice that it has three variables, population size, confidence, interval margin error, and it'll kick out a sample size.

And again, that'll be the minimum sample size, but there's very sophisticated, uh, general purpose, um, population sample tools.

G Power being one of the most popular.

Um, G Power will allow you to look at all kinds of different variables related to your population, um, whether it's a non-uniform population, whether you're doing temporal data, whether you're doing anex, uh, for an ex post-facto sample, whether you're doing stratify data, so you've got multiple data sets that you're compar comparing.

Um, G power is a very powerful tool if you're doing, um, multiple correlations or, or you're doing other types of quantitative research designs.

It's a great tool. Uh, but there's also some, uh, other tools that are out there as well that you can consider.

Just, I always encourage people, look for the simplest tool you can find.

Make sure that you, you're aware of concurs with that tool and use it.

Um, we don't need to be experts in statistics to do good correlation research significance.

What is significance? You know, we keep talking about minimum, minimum sample size.

Um, minimum sample size means that our results are gonna be significant, meaning that they're more likely to be right than wrong.

Uh, and the error error rate in that data, um, is reasonably small or measurably small so that we have confidence in our conclusions.

We want to be confident, hence the term confident interval.

We want to be confident that when we determine our hypothesis or my, our null hypothesis, uh, is true, that we have the appropriate confidence in that analysis.

And that's part of the reason we have to have a statistically significant, uh, sample size.

Um, again, as I said, you may continue sampling above that minimum, and that's okay.

Um, in fact, it's, it's, it's common that you'll sample above that minimum to get to a representative sample.

Uh, that'll be essential for publication, uh, beyond your dissertation.

But it's also, um, helps you have a level of confidence at your conclusions.

Not only are significant, but represent the population and are gonna be valid when you're determining your recommendations, um, from your findings.

Just to kind of recap, populations can be delimited and sampling techniques, uh, have to be appropriate for the population.

Um, so let's talk a little bit about sampling.

There's different types of sampling, as I mentioned.

Um, you know, the most common that the gold standard for sampling in quantitative research is random sampling, but we almost never have uniform access to the entire population.

So we can sometimes approximate, uh, random sampling with very large portions of the data or very large data sets.

But more commonly, when we're doing primary data research, we're collecting surveys, et cetera.

Um, we're gonna use a sampling technique that will allow us to collect a sample that's significant and a sample that's represented of the population.

These are just a few of the sampling, the common ones, purpose of sampling.

This is where researchers selects participants based on specific criteria relevant to their research.

So you may wade into a population and say, I need you and I need you, and I need you based on some selection criteria that you've predetermined.

You wanna make sure you predetermined that selection criteria so you're not just grabbing the, the people closest to you or the people that are gonna, um, give you the answers you want.

Uh, you want to do everything you can to avoid research or bias and purpose of sampling, uh, is good at that as long as you predetermined your sampling criteria and you stick to it.

Snowball sampling. Excuse me.

Snowball sampling is a technique of seeding the population.

You go into the population you find your first participant, you ask them to recommend another participant or recommend two or three other participants that are gonna be willing to participate.

Snowball sampling is a great way to get to a significant sample size, but it's also very prone to bias.

Uh, the challenge with snowball sampling is the people that, um, are gonna be recommended.

Usually you're gonna agree with your first person that you selected, so you're gonna wanna make sure that the seeding process, the initial, um, uh, participants that you recruit have a significant diversity in their, um, opinion and in their positions and in their demographics, so that when they're making recommendations of people that are like them or close to them, that you're still building out a sample that is representative of the population and doesn't have a bias in it, or as as much as possible, you're, uh, gathering as much diversity of opinion in the, uh, in your sample as is possible.

So, always make sure that you've got a very diverse initial seed if you're using snowball sampling.

And that'll be very important to describe clearly in your proposal.

'cause snowball sampling, um, in many journals, and for many URM's is always suspect, um, uh, is very prone to, uh, researcher bias and confirmation bias.

And so we wanna make sure that you've got controls specifically identified in your sampling process to limit that bias.

Convenience sampling, um, researchers, select participants who are readily available and have easy access.

So, for example, if you wade into a university and you start sampling the people in dorm rooms right around you, you are, you are sampling based on convenience.

Now, you may still have a diversity, you may still get representation, but you have a constraint or a limitation in your analysis because you're not able to, um, sample as broadly across the population.

So there is a potential for bias here, uh, even unintentional bias because of proximity.

Uh, and so you wanna make sure you can do whatever you can to control for that or explain at least how that convenience sample is, um, is relatively representative of the population.

This is where, um, those, um, sampling controls are gonna be particularly important, uh, to make sure that you've got a sample that represents the controls of the control values of demographics of the broader population.

A few more, uh, random sampling, as I said, if you've got, particularly if you've got secondary data, a very large population, um, uh, representation of that, uh, data set, and you've got relatively uniform, um, access to that data, random sampling is the gold standard.

Um, you are the, you are least likely to have bias creep into the data.

Like I said, it's very rare in business dissertation research, or most of most dissertation research to have complete uniformed access to the population.

So, um, if you think you do have sufficient ran, uh, access to the population, talking over with your URM before you assert that you're doing random sampling, case study sampling, this is where you are, um, selecting specific cases and individuals within an organization fairly constrained, but you can get some incredible, um, findings on, especially on, uh, specific populations.

Uh, so it's a great way to do it.

Stratified sampling is great if you've got a overall population you're trying to understand, and the demographic variety is fairly narrow.

So, for example, if you're looking at the, the, uh, similarities in relationship in, in perspectives on outcomes based on executives versus individual contributors, you may collect a stratified sample or samples of each of those populations and make sure that they're well represented within your study.

Even if you're collecting data that's identical from those, you want to be able to distinguish between them so you can study the variation among the different strata.

Maximum variation sampling.

This is, again, where you're actively seeking the diversity of the population within your sampling.

Very hard to do. Um, it can yield some, uh, phenomenal results, but it can also result in a lot of noise.

Um, so, um, I love maximum variation sampling.

I don't see it used very often because of how difficult it is to accomplish in large populations in a short amount of time.

That's allowed to us for, um, dissertations.

And then theoretical sampling.

Uh, typically using grounded theory research, theoretical samples selects new participants or cases based on, uh, themes.

So you're actually doing, um, directed sampling as you're doing analysis of the data.

Uh, so it's, uh, it's a fascinating study.

I I haven't seen a use very much at the university here.

I hope to see more students use that, uh, if they're doing grounded theory analysis.

So we just talked through in a couple of minutes all those blue squares.

Um, uh, we talked about the sample.

We talked about how to, uh, the population of the sample.

We talked about sample demographics and aligning them with the, uh, the population demographics, aligning of the sample demographics to make sure your sample's representative, we talked about minimum sample size to make sure it's significant based on confidence interval and error rate.

Uh, we talked about sampling controls to make sure the, um, sample represents the population.

We talked about different sampling techniques that you can use for recruiting and accessing a sample for your population.

Um, and we talked about a collection of data, collecting data from that sample.

Now, uh, we went through that very quickly and there's a lot of variation in there, but, um, there's seldom a lot of options when you, once you've defined your population, the, um, sampling technique kind of arises out of that description of the population as a great conversation to have with your chair and your URM to make sure you're choosing the appropriate sampling technique that's gonna get you to a data set that's gonna have the most value for you during analysis.

So, let's going, let's go through analysis of the data.

First, let's talk about descriptive statistics.

The most over forgotten, or most, uh, most, uh, commonly forgotten aspect of, uh, correlation research is the understanding of the data you've collected.

Uh, so we're gonna talk a little bit about descriptive statistics, which is the summary of the characteristics of your overall data set.

Um, so we're gonna talk about central tendencies, variabilities, distribution patterns.

We're gonna talk about, uh, mean, how, what they mean in your analysis and the selection of your, uh, uh, correlation statistic.

We're gonna talk about, uh, evaluating assumptions in your data and what that means in your analysis and in both in your, um, statistical analysis, but also your, um, determination of findings.

We're gonna talk a little bit about data cleaning.

This is something that, uh, we have to do carefully, but in almost every dataset we're gonna have outliers.

We're gonna have data that is just not believable.

And as you might imagine, you're gonna occasionally get, somebody fills out a survey that, um, says, heck with it, I'm just gonna fill all, all, all the answers as one 'cause I'm gonna hurry and I need to go get some coffee.

Data cleaning is a process of looking at your data set and identifying those outliers and throwing them out so that they don't skew your results, particularly in small data sets.

That's, that's very important.

Uh, you then remove it from your sample, meaning that it's no longer part of your, um, confidence in the data.

Uh, and then comparing variables, descriptive statistics, take a look at the different variables and their relationship between the variables to see if they can be correlated.

So, for example, you may have all of the grass length of, um, the blades of grass in my backyard in the summer of 2016, and you may have all the flower petal sizes in the national forest in Colorado, and you can do a correlation analysis on that.

It'll give you some, it'll give you a, uh, correlation statistic, you'll get a confidence interval, you'll get, but it's meaningless data.

There's no relationship between those data sets.

So if there's a correlation there, it's just coincidental.

So never underestimate the value of assessing your data and making sure that when you're combining data sets that you can legitimately connect the dots.

Um, that means that the variables have to be the same scope.

What is scope? So scope is, um, if you are studying the opinions of, uh, individual workers in a business, and you're comparing it to the job satisfaction statistics of the company, those are two different scopes.

You've got individuals and you've got aggregates of people.

And another one, you can't do a correlation between those because there's no way to connect that data.

So you always have to make sure your data's in the same scope and that there's a relationship between the data.

So comparing the variables and being able to assert that there's a relationship between those variables is part of the descriptive statistics component.

Let's talk about these a little bit.

What is mean, median and mode? So, mean, median and mode are measures of central tendency.

Uh, what does that mean? What it means is we are taking a look at the data to see what characteristics it has.

And when we're looking at mean median and mode, we can start to see whether or not the, um, data is, um, uniformly distributed.

We can start to see if there's outliers in the data, we can start to see if the data can be compared.

And so, understanding the mean media and mode, how similar they are, how far apart they're can start to tell us if there's potential outliers in the data.

So in this particular dataset, you can see the mean median mode.

We're relatively close, but far enough apart that when I plot the data, I can see that there's a couple of outliers in this data.

When I pull those outliers out than these, the mean median mode get much closer together.

And that means that I've got two variables that are related to each other and likely have a relationship between them, linear or not.

The understanding the mean median and mode will help me understand whether or not my data sets can be compared or not without any kind of transform or, uh, cleansing or sanitation of the data.

So always report the mean meaning and mode of your data sets before you do a correlation that'll help the reader, uh, or the evaluator of your data understand whether or not your data set could legitimately be correlated.

Skewness. Um, most regressions are, uh, analyze whether or not there's a linear relationship between the two variables.

And if, um, the two data sets are not normally distributed or not consistently available, you may have one data set that's skewed, uh, to one side or the other.

It doesn't have a nice bell curve shaped to it.

Um, there's two measures that we look at, and these are specific to the variables.

One is called skewness that tells us that the data set is skewed to one side of the distribution or the other.

And we can usually correct skewness through a transformation.

And many of the tools can do transformations as part of the analysis.

But a transformation may mean our, our relationship may still exist but isn't linear, and that's a legitimate finding.

So understanding the skewness helps us understand whether or not we need to transform the data or prepare the data.

For correlation analysis.

Kurtosis is how steep that bell curve is.

Is it peaked up very high or is it very flat? Um, ketosis is typically gonna be something that's gonna tell us the nature of the relationship between the two variables.

If one variable has a much higher ketosis than the other, it's gonna tell us whether it's gonna tell us a lot about the strength of the relationships.

So, um, before you begin to do your correlation analysis, look at the mean median and mode and understand what those mean.

Look at the skewness and kurtosis of your data sets, of all the variables you're including in your analysis, and understand whether those variables match.

Ideally, the skewness and kurtosis should match between the two variables.

They can vary a bit and still have legitimate findings, but, um, you want them to be, um, very similar in, in value.

And ideally if they're skewed, you wanna transform them back so they get to a, um, a more uniform bell curve.

So lots of techniques for transformation that are kind of beyond the scope of this, uh, presentation tonight, but just know that transforms are possible.

So those are the descriptive statistics.

Let's talk a little bit about the correlation itself.

Um, choosing a correlation variable.

Uh, the correlation statistic is dependent on the type of your variables, the nature of the variables.

Are they continuous ordinal, dichotomous, nominal, um, because the re the types of variables determine the type of correlation statistic you can use.

Also check for any, um, uh, correlation statistic, the assumptions that the correlation statistic makes.

So, for example, the Pearson correlation assumes that the data has a linear relationship, and both variables have a normal distribution.

Uh, if either of those conditions don't exist, then Pearson isn't appropriate.

And so doing those, um, uh, descriptive statistics is important before you do a Pearson correlation statistic.

So look at the variable types, look at which options you have for correlation statistics, and then look at the assumptions that go into forming those statistics, whether it's Pearson or some other correlation statistic to make sure your data fits the assumptions.

And then also take a look at your research question.

Make sure the correlation statistic is appropriate for the type of relationship being answer.

So if your research question or your hypothesis, um, uh, are assuming linearity, but you've got skewed data, then um, you may have to reevaluate how your research question is phrased, um, or how you're doing the, how you're doing that, uh, analysis.

Pearson, uh, Pearson correlation, uh, correlation con, uh, coefficient, again, looks for normal distributions, uh, and it's very sensitive to outliers.

So you wanna make sure you've got clean data.

So again, doing that, um, histogram or doing that mapping and looking at cleaning out outliers is gonna be important to getting a

good result on a piercing correlation.

Spearman's, uh, can be non-parametric suitable for ordinal and non-normal distribu.

It's much less sensitive to, to, uh, uh, outliers compared to Pearson.

So, um, doesn't mean spearin is better or worse, um, but it does mean it's more appropriate for, uh, different types of data.

Uh, Kendall Tau, non-parametric, uh, particularly robust if there's significant outliers, uh, and it doesn't require normality.

Um, and it's often used in small sample sizes.

Um, I'm not gonna go through all of these in great detail for, for the benefit of time, but, um, I did just wanted to let you know there are different correlation statistics that are gonna be, uh, that you have to choose and justify their selection based on the attributes of the variable and the type of data distribution that you have, as in part of your descriptive statistic.

So in a proposal, you would never define the correlation statistic.

You would only define it as part of the analysis.

In chapter four, um, correlation statistics go in a range of minus one to one.

Um, one being a perfect correlation, negative one being a perfect negative correlation or inverse relationship.

Zero meaning there's no relationship at all.

More of those than you might think.

Um, positive correlations, negative correlations, I think we've touched on those briefly, briefly, but the upper row or positive correlations where the, uh, line is up into the right meaning as one variable increases, the other one increases, uh, the negative correlations on the lower side or is one, one number decreases, the other one increases.

Um, and so understanding not only the strength of the relationship, but also the direction of the relationship are two things we always look for in the description of the results.

Um, significance is, again, going back and taking a look at, um, the power analysis of the actual data you're using.

Um, let's see, how am I doing here for time? Um, here's a Q tutorial.

I'm, I'm not gonna do it here, but I'm gonna encourage you people to just go do this little tutorial 'cause it helps you to better understand what a p value is and how to describe the strength of a relationship.

So come back to the notes, run this, go to this URL and run this little tutorial.

It's, it's very enlightening. Whew.

Boy, was that a lot to cover in a short amount of time? Uh, this was a correlation of, or this was a, uh, an overview of correlation research.

And just to recap, you know, your variables, um, know your population, know your sampling, make sure that you understand your variable type before choosing a, um, correlation statistic, and

then be able to articulate and describe what that correlation statistic is telling you about the data.

Um, lot to cover in a very short amount of time, and I, I've got time if people, uh, wanna stick around for some questions.

But, um, wanted to, uh, uh, ask, uh, Dr. Smith if we had any burning questions that, uh, I can answer.

So I don't see any questions in the chat right now, but I would like to, as you have already, open it up for questions.

I know it's a lot to cover and I do want to, uh, remind everybody that in the research methodology group there is a way to request, uh, time during office hours if you've got specific questions about your specific research and your variables.

If you wanna spend some time one-on-one with me, uh, to go over your specific study, I'm happy to do that.

Um, my, uh, link for office hours is, uh, available on the research methodology group and on the calendar out there, and I'm happy to help.

So, any questions? I think they're soaking it in Dr. Rice, they're soaking it in.

It's a lot to cover in a short amount of time.

Yes, so lots of thank yous and very resourceful.

So, um, you might be getting, uh, an uptick on those appointments as a word about quantitative.

Um, I've also posted in the chat the link to the, uh, short feedback survey to let us know how we did.

Um, there was one, uh, person that requested that I send it via email, so I'll send it to everyone that RSVP.

So similar to that, um, email where I, uh, sent a reminder, I'll send it to that same place so that you can share, uh, feedback about this, the, uh, webinar if you're not able to do it now.

So yeah, some great, some great feedback here.

Um, so one last call for questions.

Then you can always post additional feedback in the research, the re the teams, and also as Dr. Rice said, scheduling an appointment with him.

Yeah, and I'm always happy to answer questions in the research methodology group.

In teams, there's a, uh, a, uh, channel there for correlation research.

If anybody's got questions, feel free to answer them there.

And anybody that's a member of that, uh, group is willing to, uh, uh, happy to have jump in and answer questions.

Um, and I'm happy to answer those as well.

But I appreciate everybody hopping on and, and, uh, jumping into a webinar tonight.

This was great and I appreciate everybody's interest in quantitative correlation research.

Fantastic. I'm going to.