Angel Villalobos

Okay, we're going to go ahead and get started. Welcome everyone.

Thank you for joining us today. Before we begin, I would like to acknowledge that a lot of us you'd be feeling a lot of heaviness due to recent violent events, and I want to recognize and make space for that the SAMHSA Office of Behavioral Health Equity, and the NNED's National Facilitation Center would like to take a moment to extend our deepest condolences to the parents, caregivers, families, teachers, and community members who are experiencing the loss of their Robb Elementary School children and their teachers. Our hearts are with all these, all those impacted by the senseless strategy that took place on my May 24th and condemn this and other recent acts of gun violence. Let us take a moment of silence to honor their lives and the grief of their communities.

Thank you. We recognize that this is a particularly difficult time for a nation and would like to encourage everyone to take care of themselves and check in on their mental well-being, SAMHSA's disaster distress helpline is available 24/7 to provide support to anyone experiencing emotional distress. We invite you to utilize this and other resources provided in the chat box.

Welcome everyone and welcome to the Data 101 Workshop.

Due to the size of the webinar, all participant lines are on mute. We encourage you to share your perspectives and comments through the chat box. If you have specific questions, please use the Q&A feature.

This may include tech issues or questions for our featured speaker Marlon Daniel.

As questions may get lost in the regular chat box, we encourage you to use this feature. Closed captioning is available throughout today's discussion at the bottom of the screen using the live transcript icon.

This webinar is being recorded and shared through Facebook Live and on the NNED's Facebook page the recording slides and related resources, including those mentioned on today's webinar will be available on the net website, information, on how to access the recording, and these resources will be sent out to all registers after the webinar.

We'd also like to remind participants that the Q&A is for Marlon to share his experience and give broad and general advice. Our intention is not to comment and give advice on any specific application or scenario. We may not have answers to some questions, and we will take notes and get back to you later.

Great, cool welcome everybody, once again. My name is Angel Villalobos, my pronouns are he him and I will be the facilitator for today's workshop. I am joined by the NNED project manager, Alina Taniuchi.

The NNED is a network of diverse racial, ethnic, gender, and sexual minority community organizations that strive for behavioral health, equity for all individuals, families, and communities. The NNED has more than 5,200 members, which includes over 1,300 partner organizations or community-based organizations across the U.S., territories, and sovereign tribal nations. The NNED highlights and shares new programs or interventions to build a capacity of its members and partner organizations. One way to do this is to virtual workshops like this one.

It's now my pleasure to introduce Marlon Daniel, who is our guest speaker, featured Speaker, for today. He is a statistician and contracting officer representative with the population Surveys Branch Center for behavioral health, statistics, and quality, with the Substance Abuse and Mental Health Services Administration or SAMHSA. He has over 15 years of experience as a statistician health services. researcher and academic instructor. Marlon has self-position supporting the Department of Defense, the Howard University Center for Minority Health Services Research, and the United States Department of Agriculture.

He permanently joined the Federal service as a statistician, with the office of quality, improvement, data, evaluation, and division of the Health Resources and Services administration of the Department of Health and to human services in this position he was responsible for contract management as a level 2 COR, was a lead staff instructor for statistical and research methods, and provided statistical consultation and study design support to division leaders and researchers across the bureau of primary healthcare. In this current position Marlon is a level 3 contracting officer representative. We're responsible for the administration of high dollar value, mission essential agency contracts providing support to staff and statistics and statistical programming in theory, providing guidance to contractors on survey, research, survey, design, data, collection, estimation, epidemiology, and data quality.

Marlon is certified in public health and holds an MPH or master’s in public health and bio statistics and Master of Health Administration from Melinda University and the Bachelor of Science in Behavioral Science from Manchester University. Once again, it is my pleasure to introduce and welcome Marlon Daniel to today's Data 101 Workshop.

Marlon Daniel

Okay, Well, good afternoon. everyone. My name is Marlon Daniel. Thank you very much Angel for that great introduction. And we're going to get started right now.

It's a little weird for me if you could imagine because I am teaching or speaking to a screen, and I can't really get any feedback. So, we're going to do the best we can here. So, this Statistics 101.

There's no expectation at this end of this this presentation that you'll go out and do your own research, or you will go and write a R01 grant, or anything of that level but what we're hoping to accomplish here is to give you a little bit of information, so that you can be a little more conversant, more comfortable so that you can actually start your process of research and statistical analysis. So how do we start with these statistics?

Now I have highlighted here the word science the science of collective describing that interpreting data, and with my call to say it really is a science.

It's really something that helps us all communicate something That's important. Something that we're researching something that we're trying to understand something we're trying to communicate to others. I always tend to say that this entire feel the statistics and all that we work on, all we do. It comes down to one thing telling a story we're trying to tell a story to our end users to our grantees to our organizations or communities we're trying to tell the story of what this data is trying to say, and how to save.

So, I want to bounce around a little bit and there are certain things I want to touch on quickly. We're not going to go through all these slides or here for your reference for you to come back on later on. And you know, read up on them hopefully, it's a pathway for you to you know, with your actually double deeper and have a deeper dive into statistics and research methods.

So, in the research process, or even in the grant writing process you know we are, we're starting with a question. We're trying to answer a question we're trying to ask a question, and we typically call it our research question, and we decide what information is needed to add to the question, what data we need to collect assemble the data, assembles information, have data tables or databases, or whichever way you want to collect your data.

Then you analyze the data using whichever methods. that you. are using, depending on the type of variables and the type of data you've collected 200 and interpretation of this is the biggest part is the last part of telling the story. There are 2 main areas of statistics. Okay, there are descriptive statistics which most of us are very familiar with, and that's describing sample data just telling what it is.

What's information you and then, as inferential statistics, and that's making decisions and drawing conclusions about a population.

So, the entire process of statistical analysis is right looking at a sample that you've collected and you're making an inference on a population that you may or may not know now, here's some basic vocabulary that we're going to start. Okay, we have the data element which is the basic unit that you're trying to measure, or you're collecting data on in the survey work that I do with the National Survey and drove me some help.

You may call that the respondent individual or a sampled person, and then you have a variable which the characteristic of each individual has multiple variables.

For instance, I am, you know I am 45 and I'm short, and I'm a statistician. Those are variables with the characteristics of me.

And then data is the value of that variable. Okay. So, I just mentioned my height, my age. Those are the variables and the data that associated with those variables which would be how tall I am, what my weight is, what my age might be. And then data which is cool because data and data singular and through well, is a set of singular databases collected for the variables and from each element. So that's all your data.

So, you have your data, which is the individual value and your entire data, and that entire data then becomes their data set. And most of you who have collected data in the past, you know, using Excel or Microsoft access, maybe some are more advanced to use SPSS or SAS or data or are all these are programs that help you analyze and build your data set. And I left this one database here, you know. If from a pure computer science perspective might not be the best definition. But I have it here because a lot of folks speak that way. In which you use a data based enter data and managed player. And again, that comes down to how you organization is set up on how you're analyzing data.

So, you can see right here. I have these well, 4 variables, and we have Id, which is, of course, my Id for each respondent to each element of our data set, and we have the variable age in the first individual 23 and then we have the variable bio stats I think, when I use.

When I first created this particular slide, I was teaching about stats and I was asking if a person had bows that's in the past, and then they have another variable here, and then responded or element Number 4 their cholesterol levels 187 so here's an example here that we can talk about our research. We…this is something that we tend to see a lot in our data mailbox. These kinds of questions a researcher wants to use the Nita survey to determine the average number of times in the last 30 days. A respondent drank alcohol sort of variable is in this number. Is the number of times outcomes can, see? We call it non-alcohol as a variable name, and each element is each respondent, so that the data in that singular form will be the number of times for each respondent and then the data in the platform will be the set of values in the sample.

All your data with respect to all the elements of all respondents in your data, set, there are 2 the same kind of thing. I mean, you can break these down on the categories, but we typically break them onto 2 categories, qualitative and quantitative, And I know especially so many of you who do grant-based work are probably familiar with one or the other of both. So, qualitative data is data that describes a lot of minimal population.

It's, you can do you can do everything arithmetic, operations in it. You can add them. You can define them. You can't use them for …Sorry I got this backward. I meant to say qualitative data as the very you can't do mathematic operations. You can't add or average a qualitative variable

you know, and those qualitative variables tend to be attributed the variables and character categorical variables: race, ethnicity, whether or not you receive health insurance. Those are qualitative nature. Do you know they are describing an element of the population?

Now, quantitative data which is something that most people are familiar with. When you hear the word statistics it quantifies element, it's something that you could do a mathematical operation on.

You can take an average of a quantitative variable; you can define a quantitative. You can transform one so loads of the main differences. So, this leads to some folks doing qualitative research, and some for to quantitative research.

And if you really like to have the best of everything you could do, mixed methods, research which allows you to use qualitative data and quantity to do because the thing about these 2 different types of variables. And these 2 types of data types are that quantitative data which is what everybody tends to know about.

You know, when you hear about t tests, and regressions, and a-nova, and all these things type analyses using quantitative variables like age, or with 8 or number of times you went into the movies last year. These are quantities. Those are quantitative natures where qualitative data, it’s all the story. Quantitative data tells you something happened. You know, we see… Oh, there was a difference in in blood pressure between Group A and group B. And that's a quantitative approach to dealing things but we don't know what cause this change in blood pressure between these two, or the difference in blood pressure between these 2 and so we're where that's where qualitatively the country. Now it says, okay, we see there's a difference we're going to investigate. Why there are differences between these two groups with respect to blood pressure, so we can even take it a little further.

Now, qualitative deal with data, so we see we have narrative. You have no one no longer, or no we're going to break these different categories down in a moment. Then we've qualitative which is discrete and continuous narrative deal information recorded a narrative form open-ended question. You know, many times, we have in our surveys, we have all these, you know, variables. One will be a bunch of categories that you can select, and then variable tool the other.

So, specify, and so we call those that's narrative information that's information. we could sometimes use sometimes very rich to help us understand more about the question that we're trying to ask nominal variables are variables that have discrete categories. Okay, categories cannot be ranked. Our racking is subjected, which I pretty much just tend to say, you know it should rack them.

For instance, we can't write general, we should never reject yeah blood type. So, these are variables that are nominal in nature. They have discrete categories, but they it should not, it cannot be right.

Okay, a princess of mental health diagnosis where is one until we typically call it a dichotomous response. And that's something that some of you we have encountered in your time. We're looking at dichotomous variables you want to know the difference between group that said “Yes,” and we just said “no,” or you want to admit the pro no you want to predict or estimate the likelihood or the odds of someone saying, yes. So now, with orbital variables. These are variables that can be read there is a theoretical justification as to why these variables should be right.

So, for instance, socioeconomic status some will say you're ranking from port the middle class. the rich. you can wrap them up and down cancer staging. Of course, you have stage 1, stage 2, stage 3, stage 4, and typically one of the one of the ways to look at ordinary variables is that you typically have to start. You have to be in one category before you can get to the other one.

That's why social economic status is a little bit of a strange one. But you said you find that one in many textbooks. Also, on the one baker faces skill that if you ever been to the doctor's office. Sometimes you'll see something under while saying please tell me we we're paying is how what level of pain you're experiencing right now. And that's another oral variable because you assume that 0 is the most level of pain or hurt, and 5 is, you know, exclusion pain hurts the voice.

Okay, so qualitative variables. These are the ones that most of us are in count. Most of us in common. When we hear statistics you know discrete variables, you know variables that are whole numbers. You can't divide for instance, number ICU admissions within a month. Can't have half an ICU admission right or admitted or not.

And then you have continuous variables, which are numeric variables that have an infinite number of buttons possible in each integer. So you have 1, 2, 3, 4, 5, but between 1 and 2, your infinite values between there and then we could break those out even further.

We have racial variables. That's a variable it continues to be able to accurate by absolute 0 where at 0 the measure is absent. That's the way to think about for instance you can have weightlessness. So, weight is an example of racial work whereas an interval variable. There's no absolute 0 point where we're at 0 it means if absent of the measure, for instance, temperature 0 degrees means it. There's not an absence of there's not a that there is no temperature, you know. it's there's not an absence of temperature. It's just that's where that scale is anchored 0.

Open negative values also. Okay, so before we get to the descriptive statistics part, this is the part where we start to get into the stats. What we talked about so far was really just for us to get an understanding.

Okay, here are variables here. What the variables do here is what the variables can and cannot say. But descriptive statistics. Now we started getting statistics. Now we're starting to talk about how we use this data and started from all we down to grade school.

We always remember something called the average calculate the app what's the average, and averages are what we could we say in mathematics? It's arithmetic.

Mean is a measure of central tendency you know what it is it's a way for us to numerically determine where to set off. Our data is with the middle of the data it's where most of the data is clustered. So, we typically see the mean, the median, and the mode as our measures of central tendency.

So, here's a mean. We all know the calculated mean, so I'm not going to go too far into that the need is most commonly used measure of central tendence. The challenge with the mean is that it's highly affected by extreme outliers. Okay, in other words, if you have a, if you have a very, very high, or very, very low value, what it does, it can shift the mean. And so, we tend to prefer using the mean when the data is what we will call a normal distribution, or it has that bell-shaped curve, and even the example on the screen right there. It may not look the most bell shape of the status stations will consider that bill shape for our purposes.

So, on the median. the median takes the middle value of the data, and so the data is essentially divided into equal parts, and that middle part is what we call our immediate. we tend to use it for not for skewed or non-normal data. Remember, I said just now, that the problem with the mean is that it can easily be it's skewed. The word we use here x or are influenced by extreme values.

So, I think that this, if I remember correctly, this is the income distribution of the United States which you can see right here. You have individuals in the first half, you know. First couple of bars are where most of the United States population might be pretty much under $240,000 for household income, and then you have those extreme outliers way out instead of calling the 1%. This is where the 1% state and come from this population. That's way out there, and as highest percent out of earners in the United States. So, because of that the Median is use the meeting because we can't afford, especially when we're making policy level decisions. We can't afford for an individual whose income is so much higher to influence our policy decision. So, we tend to report the median.

Now, unfortunately, more times than not, I see us in our female report with me because that's what people understand. It's easier variable easier measure to work with but there's a reason why, when you hear on the news, for instance, almost every day you hear about Median home values in your area. Well, the reason why Median, in that situation, because in a particular area you may have many houses. Pretty much question around the middle. But you may have that one neighbor who is down the street who has that 2 million dollar home when everyone else has 200,000 homes, and if we're trying to do property taxes, we really don't want your property text to be based on the meeting income or the main home button in your neighborhood, because now your house values now pull all the way up there. But it's not necessarily what it shouldn't so, how do we do this?

How do we calculate the median? Well, you order the data first. And so you go from the lowest vitamin, and if the if your sample or the number of units that you're currently measuring is even that you take the average of the 2 middle observations and if it's odd number of individuals or would your data set on your database is an odd number of respondents, and you may actually use the median just take the middle value So here's an example. So, I have 10 observations right here. Well, we just said that if the number of observations or the element elements are even, then you would take the average of the 2 middle elements. We can, this case should be 3 and 4. Take after 3 and 4 any meetings 3.5.

Now, if you have odd numbers of observations like in second example, 11. Well, now, we can take the middle value which is 4 now interesting thing.

Let me see if that's my next slide no it's not my next slide yet. I'm getting excited. One thing about statisticians that we get excited about those things, and I get excitable teaching and talk about these weird numbers.

Now the mode is the variable, the value that occurs the most frequently in the data set, and we use a lot in qualitative data. We also use it a quantitative also you know we talk about this model multi-modal where the modal class is the interval that has the highest number of observations. So, here's an example of more and more class so we have these so many categories, 1 2, 3, 4, 5, 6, 7 categories, right? And we're looking at the known age distribution of individual sample, the National Survey and drug use. Okay, so what we have here is okay. look at these ends. Look at these percentages, we find that this particular group you could see 45 to 54. They're the largest class, the largest mortal class, and so we are saying that the modal class is here. if we had a bunch of values like in the first example here 1, 2, 2, 2, 3, 4, 5, 6, and 7, and 8, well, we find that 2 is our mode that's our value that's repeated the most in our data.

I am going to it's how much how the degree of smash and symmetry! Oh, a distribution. If a distribution is…if the skewedness is 0, then it's symmetrical where we tend to talk about this when we're talking about the issue negatively, what I wanted this slide. I really wanted you guys if to think about. if the mean mode and median are the same and one could determine, the distribution is normal.

Almost the air quotes normal and has a skewness of zero. So that means we… if our distribution is not skewed, we can use any of those measures of central tendency, and they'll pretty much be around the same cloud. You will take a little bit.

Okay. So now I want to talk about variance. Now many of us heard this term variance, measurements of variation, deviation from the values around the meat. The easiest way to think about is if the numbers are near to me. If variation is small, the numbers are far from the variation is large. So where are we getting? Where are we going with this? So, think about. Imagine you have a circle, everybody. Just imagine for a sec, not You have a circle here, and in the middle of that circle is what we call our meet. Now around the circle we have different points. Imagine, like a star, you have a planet in the millions of stars.

Now, what we might be interested in trying to determine is, on the average, how far is a star from a planet? How far is the individual data plane from the center of the center of its numerical universe. And so, variation now is trying to measure this because if you're all these different little measures that we took all these different stars.

They all come together to creature me, and if all the stars are far from this meeting, or the numbers are far from the average, what happens? Oh, variance is large, and if the numbers are really close to the center of your distribution, you mean your variance is small.

So how is that important? We're going to get to that in a moment. And there they are free measures of variation. Ray, which is largest value.

Man, smaller than some people. also report the range as the smallest value and a large number. and then you have periods which you're going to talk about, because I'm sure you've heard variance multiple times. But if you haven't heard variance, you most likely heard of concept of standard deviation, so range the maximum value minus the minimum value.

Okay, so the cost. So, for instance, I have 10 vials of some drug that I was working on in the in the lab and the concentration of the third, and I have the concentration of the therapeutic agent in these 10 minus 200, all 8 to 12. those are that that's and so what I want to calculate now is my range. So, I have my name, which is 202.4 I don't have a use of measurement, here but my range now is 207. My slide got messed up because that last one should be 207. So sorry about that. It should be 207, minus 195, which will be 12. But sometimes what you might see is in some the literature you might actually see. 195 dash tool set will show our range of values.

Now, if the range of values are close to 2.4, we'll say, okay, that means the variance is low, but it's that number, or that interpretation tends to be a little more subjective I would tend to use ranges, mostly with Medians, not needs being that.

There are other ways of measuring the central tendency that work better for the meeting which we start right here, various. And so, it's used to measure how far data points are from the need on an average, don't worry about this Greek in front of you. What I want you to remember from this is that the variance measures how far each data point is from the mean and what that is trying to help us determined here is how precise our estimates. Because if all the estimates are close to the mean that's saying, oh, our estimate is pretty precise.

It's very tight it's very representative of all the values that we base, that that mean on if the variance is large, then we are, we can report our estimate. but we also have that caveat saying yes, this is our estimate or average age was, or app, say, average, conventional levels, 2 or 5.

But the variance is very large and so the true value we're not really sure about and nobody just said, "Okay, he just said true value. What does that mean?"

Okay, well, let's step back for a second and think about what statistics really is about. First. these statistics is really about samples, and which is going to be a slide and look further down. But I want to talk about it right now, really quickly we're take we don't know we don't know the population's information if we knew the population's information all the information with the population, I wouldn't have a job right Now I'd probably do something else that's interesting like, and so what happens here is that we need we?

We try to come up with ways to estimate that's the…what I want to say predict but estimate what phenomena we expect to see what phenomena is in the population.

So, I miss them where we're trying to calculate or precipitate estimate measures of mental health, utilization or mental health prevalence or use of illegal drugs. Well, very well, can't go out in the entire population and ask them a 1% of 1 h document or survey, saying, “Hey, I want you to excellent questions. I mean there are many reasons as to why we'll get that little later. So, because we don't know what's going on in population, we use statistics to try to estimate that now, because we're estimating we have to now hope that we're very precise in this that's what variance comes. So, when you talk to survey statisticians their favorite word is variance.

What's the variance? What's the relative variance? How do we calculate this variance? Because what we want need variance is very important because if the variance is very large. We're not really sure what estimate of estimating the population is really the right estimate of a true estimate.

So that's where…why variance is so important. But because variance, the unit of measurement will be a square term. If you look back here, you see the Greek is a square term there, and there's a reason why it's there, and you. I probably talked about it in the notes, and you could look at it.

And then statistics focused, y squared but because you square. that term you use of measurement ends up being squared. So, in other words, won't be you'll have a mean try to listen right level of mean something that's really bad for testing. But if you, the variance will then be recorded as Milligrams desktop square, which is not really a helpful, you, the measurement, because now your unit measures changed from your mean you want your meeting and your experience, that the same measurement. So, to get rid of, go back to basic algebra to get rid of a square turn. What do you do? You find? A square root, so square root of the variance is now what you call your standard deviation. So, which will then be a unit list. statistic. that explains the how precise your estimate is we're not going to get into this right now.

Here's another one that I think a lot of you would run into, and you your work linguistics have percent 1,000 ports. Well, we've all seen percentiles. I see percentage every time my case we focus come in a percentage of a distribution equal to all the level particular number. So, if you had the 50 percentile that tells you right there that well, your 50% of individual New York for you and individuals score below you and opposite a button in a Grv. For instance, you might

have a score that seems great to Gr. But if 80% of the distribution score better than you did, and you were to 20% out, Percentiles are usually used to compare individual values with a norm, so what we see a lot of times a percentile is like in group charts for children or weight charts. So, Mike, not my son, but a child may have a night. Do you May we? A certain way that their age and yeah, they're percentile rank, or would be they're at 20% weight which means They're actually pretty on the weight because 8% of kids that age will are everywhere, and then 50%.

I'll also come back to when we talk about median 50% out is the meeting the middle of your distribution.

Okay, so I am going to go forward to this slide we don't do rules of thumb because there's a lot of…I don't say that. So, we said best practices. Two went to report different measures of dispersion.

This is standard deviation's usually report to me percentile into portile range I use when you're dealing with ordinary skewed data are you when you're trying to compare to a set of norms into ports, out range, which I skipped this now, but I can just touch right hand use. Describe the central 50% of the data regardless of the shape of the data when it's normal.

The range is used mostly well with meetings. and also, where you're dealing with extreme observations. So, you can have your median home, medium home, Median Home prices in a particular zip code can be 400,000, and with a range of 200,000 to 999, and then the coefficient of variation which we're going to skip I'm not going to talk about that.

That is something that we're going to talk about in a future class. We call it the Cv. in survey sampling and it's a very important term. I want to do it more justice. What I want to talk about really quickly slides its populations, and Samples now given that I'm, I work as a survey statistician I work with a survey population establish a near and dear to you and then we near and dear to. you guys are community-based organizations and you're really concerned about the populations.

I'm taking steps of said population to try to help you understand to tell a story to agencies individual all organizations that you're writing brands to federal oversight agencies populations that sound so you saw me talk I touched on this just now. Briefly, but I’m going to really get dive time.

What are the major purposes of research? One of the major things that you try to do when you tell your stories to make an inference from a sample to a theoretical population. Because again, many times we can't afford to do a survey what we call the census we have for 10 years. Well, that's what the census is doing. No, the census can't ask every individual United States all the questions that we may have interesting in our particular agencies or our particular usage question that we're trying to find in all research.

So, this is why we have to make inferences on this population. Basically, on a sample. There's a lot of maps that goes on side there. Good talk about it if you like. Let's take different definitions inference inferences make it a statement.

Deduction of conclusion or suggestion about all-purpose offers, that's why someone say you know, make it infer something on a population. A population is a set of elements of a finite side that will be studied can be studied, or a large cell collection or items that have some.

So, most of the time what we do here Very nice, you know how in our feeling dhhs, and we are interested in the U.S. population, or we might be interested in the state population or sovereign nations, population, or any of those groups.

I and a sample now is a subset selected to be representative. I should have made representative bold on the line and event representative of the population. That is the biggest assumption that we make about samples, that the sample is representative of the population. And the reason, and that the reason why it has to be representative is the next definition.

Generalizability. What are the results for fines from the study can be assumed to be true for the target population? Well, if the sample is not representative of the target population that we can't to sell infer these sample estimates to the larger population. That's a big problem that a lot of us get into trouble Say, a lot of us is Royal.

We my friendship last week it's something that we get ourselves in trouble, because we tend to always say well in the U.S. But then you ask them, well, tell us what to sample. Talk to talk to a couple of people in my neighborhood. That's not really going to give you a clear understanding of the true population that you're trying to make an inference on. So…yeah, so we have populations. Oh, yes, samples that's not a little bit of Greek for a second. So means this first thing here looks like a new with a long tail. We call it new, and that is what we call the population.

You know the population parameter for the central tendency, and an X bar. That x to the bar over is our mean in a sample it's standard deviation. That's a sigma and then standard deviation of samples. Yes, very remember the relationship between standard deviation and the and various number of you guys remember that we talked about squared and square root. Well, the variance now will be sigma square, and in the sample, s squared, and then the correlation coefficient. Another misuse, word in research and statistics would be role or r in a sample.

So why do we stand? Why is sampling important? First, he said, was going to be studied more quickly in the population, you know, we see a whole lot of it takes to get census estimates back from the census or how long that they implement the census, whereas if we just wanted our, so even you go to the build your local community, and you want to actually study the local community. You want to get all this information sometimes.

It's quicker to actually get a sample of these individuals. Studies samples is much less expensive. I could tell you being the COR of a survey contract, it's very expensive to send people out into the field and collect data extremely expensive. So, it's a lot cheaper to send people try to get 60,000 or 50,000 in the respondents than the 355 million American citizens or individuals who are in the United States study populations is usually impossible. Very true.

Again, cost reasons logistics why there's some well wanting. Some areas are very dangerous to enter and so we can't really enter certain communities. Certain neighborhoods I can't send interviewers into a certain neighborhood, because it's dangerous, and so sometimes we have to do sample based upon we'll talk about that in a moment. Samples can at times be even more accurate than population based on this somebody's here just perfectly. How and was that possible? Remember what I just said about dangerous? Or if difficult, to access areas. So, you're trying to go into say, remote locations, and you know the population. There are a hundred individuals, but in across a 200 square mile area at least 100 individuals scatter. But if you do a sample, and we know, based upon we do things like address listings or so we know all the addresses in that particular area. We can then say, “Okay, instead of trying to go and sample individuals, 200 square mile area, we're going to take a sample of the individuals.” Yeah, of these individuals, and then generalize them to the population. Because what happens is, we, I think, in site 1 one here's something called a healthy worker effect. And all these things so social desirability. What happens is sometimes with population-based studies.

You get individuals who are come, hold to answer questions. and then you have the individual medicine to be involved, and those are the individuals we tend to really want to get, and it's easy to just target your, sorry, me to them as opposed to just doing a full census, and they get a more accurate estimate. For sample is collected correctly. The Arab exhibits W estimate. What do you think this area might be? folks I don't know. Yes, I'm not from the classroom, so what this will be called your various. How different an end of the overall sample mean is from the individual measurement, and samples can reduce heterogeneity.

What that means is that, say, for I can with my sample I can minimize using weights or things like that. How many individuals are representative? One particular group say, for instance, we know that we want to get 50%. Hispanic individuals in our sample. but we are in an area that is 80% non-sect way, well samples. We can actually control that and wait our analysis so that we can. Yeah, better distribution of individuals, and then you can adjust that the population again.

That's another longer presentation, and lots of documents and lots of Mediterranean connections. If NNED wants to bring us back, we can talk about that some more.

So how are sample selected? Well, we have 2 different types of main ways to click on the sample, which, as you said, selecting a group of subjects from a study lesson larger, so that individual controls entirely by chance. And we have a not probability sample which is convenient sounds So we're going to move and talk about probability sample.

We have these main samples right here. the simple, random sample, systematic, stratified probability, proportional size, plus a sample. Again. We don't even know about that you just need to know about probability and random stuff.

Another probability sample is what you might tend to find a community when you're trying to just go out in the community. You meet individuals, and you're like hey? can you answer my survey. You have a convenient sample. you intercept sampling, snowball, sampling, spotted, driven sampling and time location sampling.

So, I'm going to stop right here at this slide because I don't want to go beyond this at this point, because you're trying to. Really, I want to come back. So, I want to stop this slide. So, what is random mean? Okay, random. This means that all human influence or is minimized that completely move or unknown.

I'm unknown from selection process in other words. I don't pick a respondent or element based on well, I like him. So, I'm going to pick that personal this person is my age I'm going to pick that.

But let's stop here for now because I want us to get into our Q&A. A little bit. So, let's take that away sir…

Angel Villalobos

Thank you, Marlon. With data I feel like we could go on forever, and then talk about so many different spaces. Thank you so much for or synthesizing that into this quick show and definitely with and it sounds some…someone from the comments would love to have you back into to give us some more information.

So, to everyone here we just joined. my name is Angel Villalobos and I'll be helping facilitate some of these questions with Marlon. These are questions that you all, as attendees and registrants, people that registered submitted questions when you first submitted your registration information.

So, the first question that we identified is: what made the biggest difference for you when you were learning about how to use data elected question resources.

Marlon Daniel

Well, yeah, the cool things you know. nowadays we have the Internet. You have YouTube. There are plenty of resources out there.

A lot of universities - I can't remember the name of that website that offer free classes that you can take at Harvard and Princeton - and all these other myths of these schools. For me the biggest. what meeting? Help me learn how to use data was just to play with it.

You just get a data set and just simple analyses. I want to calculate the need I want to calculate. I want to calculate percentages of proportion. So, of course, something that affect how I learned, and the biggest yet.

Coursera, thank you so much!

What helped me a whole lot in terms of learning statistics and learning methodologies, actually teaching myself. I used to don't laugh at me congrats I would actually stand in front of the mirror and tried to teach myself the lecture that I'm working with that particular date and by 2 teaching myself like talking to myself. And you could talk to my family. I talk to myself all the time You're not crazy. You you're what you're doing is if you can teach yourself, or you could teach someone else. Yeah, you actually, you get a better understanding. I think teaching stats over the years is what really helped me the most in terms of understanding method methods. But for you just clear the data. download data, sets any data sets down available everywhere to download for you to just try to understand it.

Angel Villalobos

Absolutely. Thanks, Marlon. I think that practice, and then presenting in front of others, is always a chapel tool in a variety. And in here it makes sense.

So, the next question that was selected is, how do you share data in an approachable way? I think folks are often times data as something that is really overwhelming and scary sometimes. So what? what? What are some of your tips?

Marlon Daniel

Well, I think going back one of the first things I said in this lecture, which was referring to telling a story, you know, if you go to someone and say, you know 50% of African Americans blah blah blah. No, that's not people don't want to always hear things in that matter. Don't be Number Z as I would say we're telling the story, you know, in our community, we're seeing a particular shame with recruiting individuals being a study based upon what we've seen 50% of African Americans, where, if you couch those numbers within a narrative, people are more acceptable and more open to so hearing about data, don't just put bullets there with numbers that have no context, it's the it's the worst way to share data, and I think whenever you put it in a context, you know. Be on the stats of the group that you're presenting to then that way you tend. They tend to be more open to hearing about an attempt to walk away with it, because now, when I put it in this context, we get it. It sticks, you know. if I just give you a percentage that doesn't really that doesn't really you don't really retain that absolutely thanks.

Angel Villalobos

It's definitely about that storytelling and hearing is that narrative that like what does it actually mean to see what you're using? And how are you sharing that story?

Our next question is, how do you analyze data without solely comparing black, indigenous, and people of color to the white population.

Marlon Daniel

Okay, now, this is a heavy one. Again. But again, this is back to your story. So, you're not compelled to compare black Indians or BIPOC populations to the white population. You're not compelled to do that what this what the reason why you see most analyses using the white population as a reference category.

It's because it's the largest population, population group in the United States. And then, for historical or sociological reasons, you know, these populations may have better access to healthcare or better education. Or all these other factors. And so, because of that, we tend to want to answer the question.

Biggest question we ask. well, if we're trying to answer questions for equity. Well, we're trying to ask a question. Well, how would these subgroups Are these different into different ratio, ethnic categories? How well do they compare to the white population we want to make sure that the that access and understanding utilization is actually equity Builder? That's equitable between these 2 different you know these groups and the white population.

So, you don't have to always tell a story about that if you're a story is if yeah, yeah, it's dot com.

If your story is about your community, for instance, you want to know in my community, how many individuals have high blood pressure? Oh, that's a different that's a story trying to tell there's nothing to do with the BIPOC population there. If your community is a community of indigenous individuals, then that's a story you try to tell. But yes, that's the reason so you can't analyze without comparing.

But for the most part especially, you want to get a grant in this country, you know you'll probably have to compare to the white population. That's probably Why, you see a lot of that.

Angel Villalobos

Yeah, thanks for lining. And I think this is one of those questions that, could I have many different answers and different theories of, or conversations of, why it is so really, we get your perspective with the question that was given.

I think we have a couple more questions, and maybe we can take Oh, depending on time, one or 2 from the audience: How can small organizations use data and connected to implementing service delivery?

Marlon Daniel

Oh, I mean data drives everything in this world. Data drives all decisions.

So, what you're saying? what you want to find here is say, you have a community member talking about, you want to provide. Let's just now we just went to COVID, and we would like to try to find out how can we best serve our community and getting vaccinations are getting important testing or anything like that?

Well, what you can do in a small organization is understanding your community. Well, what proportion of your community is what community ports in your community? Have poor access to for access to transportation, or whatever.

Well, if you have information on that, that would help you understand? Okay, you know what this particular area in our community this is where we have the most of the elderly population, let's focus on deliver our services to that place. We can set up our vaccination or testing center over there, and we have many tools like that. GIS, so another 2 that we can use, GIS is geographical information system, still not mapping. That helps us, you know. small organizations to help us figure out the smallest organizations to actually get services to where services are required.

Angel Villalobos

Yeah, Thanks, Marlon. I think we have one final question from the pre-asked questions. How can data information go, whether it be to apply for government grants?

Marlon Daniel

And I think this kind of goes with the other one just. I guess, when I read this question to me, it feels like, how do we share our story?

And as you're seeing in the comments in a way that is authentic to who we are while also being able to apply for funds that that are so desperately needed, and that are you know a part of a lot of community-based organizations. Exactly. I mean, you know. any.

Grant that any grant call it goes out or any NOFO they're asking. They're saying the story. hey? we realize this what this is the story that we we're seeing happening. what can you write in your grant to actually answer some of the questions in your story? So. yeah. that's really where a lot of the statistics and research methods coming into the grant application process.

Angel Villalobos

Well, I'd like to just quickly reference one question, and you would absolutely love to follow up with this individual, but sharing over and asked…Can you discuss HIPAA requirements for data collection for program evaluation? And how might differ for research studies?

Marlon Daniel

That's a heavy question that's going to really depend on your organization.

I mean anything, Zach, thanks to past your historical missed these. Everything has to go to an organization that's to go to the IRB Panel - Institution Review Board - any grant is coming out of an organization Here's I'll be involved and an IRB when you're working with your grant writing is that last, that last group that lasts set of individuals to protect your organization from making mistakes With respect to that, when you submit a grant application or a research application, RIB refuses, and they will take a look at them and say, Okay, you know what right?

Here. you are seeing that this particular way of ancient basking this question may be, there are some privacy concerns that we may have here, and we'll set it back to the researcher or send it back to the research group to have that discussion among themselves, try to figure out a way to formulate that question so that it can be more secure.

Angel Villalobos

Thanks, Marlon. Yeah, definitely at a complex question, just want to get a sense of it. So, thank you again, Marlon. So much for joining us today, and for sharing your insights.

If you can look at the comment section there's several comments, and they're appreciating or appreciative of what you did, and wishing that you were their professor back when they were in college.

And thank you all who joined us today we will be including Marlon's information in the slide, and it is currently included, and when we send it out, and just a quick note to stay team for our upcoming series of workshops that are coming up summer 2022, and continuing on, so please stay posted.

And in addition, if you if you'd like to join the Nan Alina has been doing a wonderful job of sharing that information, and as a final thank you to all on behalf of SAMHSA and the NNED, we just really appreciate Marlon and you all for joining us today.

If you have any questions, feel free to reach out to Perry on this slide.