Lightly Edited File

Walden ASC Webinar -

Skill Builder Session: Descriptive Statistics

April 14th, 2016

Remote

Walden University

7:00 p.m. - 7:30 p.m. (CDT)

* * * * *

THIS FILE IS BEING PROVIDED IN A ROUGH-DRAFT FORMAT.

COMMUNICATION ACCESS REALTIME TRANSLATION (CART) IS

PROVIDED IN ORDER TO FACILITATE COMMUNICATION

ACCESSIBILITY AND MAY NOT BE A TOTALLY VERBATIM RECORD

OF THE PROCEEDINGS.

* * * * *

CART PROVIDED BY Kristi L. Arntzen, RPR, CBC, CCP

PARADIGM REPORTING & CAPTIONING INC.

612.339.0545

©2015 Walden University Academic Skills Center

For additional help, contact the
Academic Skills Center Tutoring Services:

ASCTutoring@Waldenu.edu

To schedule a tutoring session, go to:
https://waldenu.mywconline.com/
>> Okay, I think the top of the hour, so I think we'll
go ahead and get started.

Step one in this is getting the recording going.

For those of you logged in, we always record these.

We'll explain more later about how to find the
recordings, but I just want to let you all know so that
you do know that it's being recorded.

Um, okay, welcome, everybody, to tonight's Skill
Builder Series.

The topic tonight is descriptive statistics.

My name is Pat Dunn and I'm an instructional support
specialist in the Academic Skills Center at Walden and
I'm joined tonight by my colleague, Kim Palermo-Kielb
who is the instructional support coordinator, is that
correct?

>> That's right, yep.

>> Yes, yes, okay.

Awesome, and Kim really, you know, coordinates the tutoring services that we provide.

So in addition to these webinars, we also provide individual tutors.

So at the end of tonight's presentation, we'll give a little bit more detail on that.

For those of you that have not attended sessions, we have both Skill Builders and a program called stats chat live, the intention is for these to be 30-minute sessions and to be very open-ended.

So I'm gonna give an overview of descriptive statistics, but the idea is to be pretty concise.

I'm gonna actually go to SPSS and Excel, show you a couple of things, but then really open up the chat lines for questions.

You'll notice on the bottom of the screen, actually right below the PowerPoint, there's an option there for actually submitting questions.

©2015 Walden University Academic Skills Center
You can use the chat area for conversation, and Kim will be helping you there.

But when we get to the Q&A, if you could use the Q&A for questions, that would be awesome.

That way I can kind of keep track of the questions.

So with that, so our agenda tonight, again in 30 minutes, this goes very fast.

It goes fast for me.

I'm sure it goes fast for you, Kim, and it'll probably go fast for you in attendance.

Again, we're not gonna be able to cover everything, but because we record these, there will be a way to review.

But the idea is to talk about, really, two different types of descriptives, measures of central tendency, and measures of dispersion.

And of course within that, we have -- we can do descriptive statistics like means and standard deviations and sometimes we have frequency data which are really using counts.

To go again to give you just a quick window into SPSS
and Excel.

And then of course we'll open up for questions.

So why is this important?

Well, descriptive statistics are a part of virtually every research study.

You know, table one, typically in almost any study that you do, whether it's in a receive publication -- peer-reviewed publication or in your own research is gonna a description of your sample or your population.

So it's vital.

Why do we need to do that?

Well, we have assumptions that we make in statistics, assumptions like normality and quality of variance or home no gentlemen Nate of variance among others and descriptives are really important to help you know or figure out whether you're violating any of those assumptions.

So what do you need to know?

Again, you need to know measures of central tendency, the three biggies are the mean, which is the average, also known as the average, sometimes you'll -- you'll
also find the median being used.

The median is the middle value.

And the mode is the most common value.

And of course in a perfectly normal distribution, the mean, the median, and the mode are gonna be, you know, the same number.

In nature, that doesn't always happen.

And I'll show you some examples.

And then of course we also have frequency distribution, if you're counting, say, the number of males and females, you know, it doesn't make any sense to run a mean or a standard deviation of that.

You really need to use a frequency table to get the counts.

So here, again, here's central tendency.

So the graph on the left is the textbook graph.

Again, perfectly normal.

And the mean, the median, and the mode are all lined up right in the middle at the very top point.

I have a sample data set that I use, and I just ran a
simple histogram for age, and as you can see, you could argue that this is pretty close to a normal distribution.

But if you actually do the calculations, you'll find in this case that the mean, the median, and the mode are actually different numbers.

The mean in this case is, you know, is kind of in the low 50s, I know I don't have those numbers on there, but you'll just have to trust me.

It's in the low 50s, the median is a little bit higher.

It's in the upper 50s, and the mode actually can see it's the highest bar there is about -- is around 58.

So anyway, it's just a way to kind of see how they're similar and how they're different.

The other one that's also important are the measures of dispersion.

These kind of throw me off.

I heard one time, Kim, that the human brain understands measures of central tendency much more so than we understand measures of dispersion.
You know, a lot of times, we'll say, yeah, the average, you know, my batting average, or my scoring average, so we use averages.

As opposed -- like if you use -- you wouldn't say, you know, my batting average, you know, ranges between .200 and .300 or, you know, I score between 15 and 20 points in a game.

You could represent data that way.

We just don't tend to do it that way.

So then when we see those numbers, it can kind of throw us, because, like, oh, what do these things even mean.

>> Yeah, exactly.

>> So again, I'm just gonna point out some of the real basics.

The ones that you'll see the most often are the standard deviation, the variance, you'll see the standard error or the standard error of the mean, and then also the range.

So again, here -- and here, why is it important?

Why don't we just need to know the mean?
A lot of our tests are just simply testing the means and the difference in the means.

Well, I have two examples here.

With the exact same mean.

But the variation is quite a bit different.

And that would mean quite a bit different in our testing for probability with these two different samples.

So it is really important and critical that you know the variation, however that's measured.

So again, the most common of this is the standard deviation.

We'll get into the details, but it's the square root of the variance.

Okay?

And if you -- if I go back here, again, when we're looking -- or even at this -- at these -- if you look at the -- the blue lines on the left and the right, not on the one in the middle, those are typically representing one standard deviation.

So it's actually defined.
It's the middle -- approximately the middle 68% of
the -- of the scores.

Within a sample.

And then of course you can go out two standard
deviations, which is about 95%.

You can go out three standard deviations, which is
about 99%.

And then if you're getting, you know, way far out, then
it's, you know, much, much smaller, smaller
probabilities.

The standard error is another one.

And it's the standard deviation over the square root of
N, so the number of sample, you know, participants is
important in the calculation of the standard error.

Sometimes it's referred as the standard error, standard
error of the mean, and that's what that is.

Okay, so I'm gonna go now to -- I'm gonna go to SPSS
and Excel.

If I get this right.

Kim, make sure that I have -- that it's on the screen
properly.

>> Okay.

>> Is it showing in the window?

>> Yep.

I see SPSS.

>> Okay.

Awesome.

And for those of you, I know this is gonna be very, very tiny.

You can go full screen.

From your view.

Or, again, you can just, you know, bear with me, and then watch the recording later.

There's no -- any number here that's critical for you to know.

I just kind of want to walk you through how you would calculate both a descriptive and a frequency from the sample data set.

So the way you do this, so this is just a sample.

It's actually a sample of blood pressures, and I have blood pressures, and then I have things like, you know,
gender and ethnicity and actually region of the
country.

You'll notice in SPSS, these are all numbers.

SPSS loves numbers more than words.

So these are actually coded so in my gender, it's -- I
have 1 for males and 2 for females, so it actually
converts it.

But you could actually have the actual names in there,
and it would work the same way.

The way you do it, you go analyze, descriptive
statistics, and it's interesting here, there's actually
two different ways to do this.

It seems like you would go to descriptives, and you
could certainly do that, and let's just take the base
line systolic pressure here.

With SPSS, you can put as many values in here as you
want to.

What you'll do is you'll click options.

And you have some standard options that come up here.

So we have a mean, a standard deviation, minimum,
maximum, let's put in the variance.

Let's put in the range, and let's put in the standard error or the mean.

So you click continue.

And there's our table.

So we have -- we have a mean of 133.461, the standard error of .3126, standard deviation of 20.3441, and a variance of 413.

Now, let's say we wanted to try it the other way.

Go, analyze, and go to frequencies.

You would think that the frequencies would be really designed more for like the count.

You can actually do the same thing.

You put the same value in.

Click on statistics.

And you actually get the same options here, mean, and actually we get a couple more, we get the mean, the median, the mode, put standard deviation, variance, range, we'll put all these same values in.

You could also get what are called skewness and kurtosis which is really more like the shape of the
distribution.
Click continue, and an important thing here, if you're running these descriptive statistics, then uncheck the frequency tables.
Click here.
And you'll notice that although it's -- looks differently in the table, the numbers actually are exactly the same, okay?
We still have the same mean, the same standard deviation, the same standard error and the same variance.
It's just laid out differently.
So you can actually get the same data doing it both ways.
You can also get the frequencies.
So again, go back to analyze.
Descriptive statistics.
And frequencies.
Let's remove the blood pressure value.
And let's put in -- I have a variable for region.
So it's where these people live.

Region of the country.

I am now gonna check display frequency tables.

And you think about it.

It doesn't make any sense to run these descriptives on -- on a categorical variable.

So I'm gonna uncheck these.

Click continue.

And I'm gonna click okay.

And now I have a frequency table.

So it's not giving me a mean, a standard deviation.

What's it's giving me is the counts, and of course you can convert the counts to percentages.

All right?

So now I'm gonna go -- I'm gonna do this in Excel, and then I'm gonna go back to our screen, and then we'll start answering questions.

So now you should an Excel spreadsheet.

That I've put together.

It's actually the exact same data set.

Exact same files.
And what I'm gonna do, because not everybody is familiar with Excel.

Many of you are.

I totally get that, so this is like literally a review.

But for those of you that aren't, so if you wanted to run that same -- you can do it a couple of ways.

Just put in the equals sign, and you can start, like, if you know it's gonna be the average, just start typing it in.

If you don't know, if you're in -- if Excel is new to you, then go up to the formulas tab, and click insert function.

And you could even start typing in like mean or average and it's gonna pull up.

And since I practiced, I have my average right here.

And I had the cell ranges already loaded in.

And it says 133.

And of course, if I expand this, then I get my -- and I have to make this a little bit wider.

I get my same 133.1461.
So exact same as what we got in SPSS.

I wanted to go back and do the same thing with standard deviation.

Go to formulas.

It's -- you know, once you find these, it's helpful because -- and what you can do, like, there's a bunch of standard deviations, and the P means for a population, the S means for a sample.

I just use the regular ST-DEV.

And lo and behold, I get the exact same number. Very cool.

Okay.

The one that isn't available in here is the standard error of the mean.

Okay?

So again, if I -- but if you know your -- if you know your numbers well enough, remember the -- the standard error is the standard deviation, so I'm just gonna write -- this is -- so this is just a demonstration on how to do a formula.

So I'm gonna label this standard deviation.
And I have an N.

And my N was 4236.

And I have -- so I want to -- I want to do the square root.

So, again, I would go -- just click equal.

URT, that's square root.

If I click here, and I can just say that's in cell 4240.

Now I have my square root, and now I can put in my formula.

So equals, so now it's a product.

So now I have -- so we have standard deviation in cell, um, B-4239.

And it says, okay, I got it right there.

Over, and I have my square root of N, which is B-4241.

And sure enough, I have -- if you go back to my SPSS, I get the same number.

It's -- it was .312.

So you can actually reproduce descriptive statistics in Excel.
Some of the inferential statistics are harder to run in Excel.

But you can run all of the descriptives, even frequencies.

You could actually run a -- set up a pivot table, so just click insert.

Create a pivot table.

And if we wanted to do that region, click here, and then I'm gonna actually click that into the values.

And now I have my same counts.

In a pivot table.

And if you wanted those instead to be descriptives, I can put my blood pressure in there.

I probably should do that by region.

And then -- and then put the values in.

I can show you that here as an example.

So I've got my average blood pressures by region.

So anyway, that was a quickie.

But we have about ten minutes left.

So I'm going to stop sharing.

And I'm gonna go back to where we started.
And happy to answer questions.

Looks like we had a couple of questions, couple are just hello, getting acquainted.

Question.

Aha.

So the question Dennis has a question about descriptive statistics for multicolinearity.

Very interesting, Dennis.

So I think for those of you that don't know it, it's when you have variables that are highly -- there's a high level of intercorrelation.

And I actually got this in a tutoring session the other night where the person had -- was looking at weight and the body mass index, so for those of you not in health sciences, the body mass index is a ratio of weight to height, it's weight over height squared.

So if you could imagine, there would be a tremendous amount of multicolinearity between weight and the body mass index because it's actually -- one number is actually part of the other number.
There aren't, Dennis, to my knowledge, multicolinearty diagnostics that you could run in Excel, but you could certainly run multicolinearty diagnostics in SPSS. So that might -- it'll require a tutoring session because it's a little bit in the weeds, but like if you're running a linear -- like a multiple linear regression, a multicolinearty diagnostics will be part of the options that you can run there.

Joe has, is this in the spew book too?

>> I don't know what that is.

>> I'm not familiar -- Joe, if you could clarify what spew is.

That might help.

Robin has a question, what is a pivot table?

I know I went incredibly fast.

A pivot table is a function in Excel that can be used to greet groups and -- create groups and categories.

It's a very, very handy thing.

Those of you in the business school I'm sure know about pivot tables.

We actually have on our YouTube channel, we have a
webinar that we did on advanced Excel.

And it shows you how to do pivot tables in there, and

we can get -- I'll have Kim talk in a moment about some

of our resources including our roll accounts.

You can type in, you know, what was, you know, and we

can send you the link or we can help you more so with

that.

>> Okay, Pat, Joe said SPSS book.

>> Aha, got it.

>> Yes.

>> And the question was about --

>> I guess the descriptive statistics?

>> Yes.

There should be -- there should be information in the

book on descriptive statistics.

It's gonna be like the first part of almost any stats

book.

Is gonna talk about descriptives.

>> We have a question, can you please explain the
difference between arithmetic mean and the median?
>> Yeah.

So the mean is the average.

So it's the, you know, if you sum up all of the scores and divide it by the number of scores, that's the arithmetic mean.

The median, you just are looking at the middle value.

So if you had five values, and you ranked them in order from lowest to highest, the third value would be the median.

Next one, can we use -- okay, that was about multicolinearity again.

Yeah, oh, good question.

Using multicolinearity in logistic regression.

So I'm gonna use this as a shoutout to our other colleague, Dr. Zin Htway, who is an expert in logistic regression.

I would actually have to defer to Zin on that and I know he would know the answer.

>> I'm gonna put that in the chat.

And you can reach Zin, first of all, the logistic regression archives are on our skilled builder page, so
you can watch those.

And then hold on, I'm typing this.

And then Zin is at our stats support email.

He answers most of those, I believe.

So that's a great place to get him.

Let me just type that in here.

>> And Michelle has a question about SPSS syntax and APA style.

So when you do your SPSS, whether as a descriptive table, chart, anything that you do, the syntax is always that text right above the output that's like a text description of what you just did.

And the best way to -- you can export it, or you can just simply hover, and copy, so just copy it, and then go to Word and paste it.

A note of caution about APA style.

The SPSS tables are not in APA style.

So I would strongly encourage you, you can go on the Walden -- the Writing Center, which you can get through the myWalden portal, there's a section in there on APA
style, and there's actually a tab for tables and figures, because more and more instructors are taking points off, I've noticed, when you just use the SPSS table because it isn't in APA.

>> Pat, there was a question from Tara.

She wanted to know, is it possible to use descriptive statistics with a weight loss research project? And if so, what would be the best approach?

>> Yeah, great question, Tara.

In almost any research project, you have to do descriptive statistics. You're not doing the descriptives to test for the weight loss itself.

You're doing the descriptives to, you know, really kind of, you know, define your population, and then you would probably do a -- either a T-test or an analysis of variance to actually measure whether there was a significant weight loss or not.

We had a question, how do I get a tutoring session?

So while I'm going through these, Kim, if you could mention just a little bit about tutoring --
>> Sure, because we just have a few minutes.

Yeah, there's a few questions in the A and A -- Q&A,

Pat, and then there's a question from Karen, what's the N?

That's the population 6789 okay, so what I'll start
with putting in the link for how to get into register,
and sign up for our tutoring.

Gonna but that in there.

We use an online platform called WC online and I just
put that link in there.

If you go to that link, you just need to register with
your Walden student I.D. and a password, and you will
have access to the tutor schedules.

For statistics, we have determination tutors,

Dr. Htway, Zin Htway, and Sarah Inkpen.

Sarah's also a dissertation student.

So they're in your shoes, so to speak.

They're wonderful.

We also have a general stats and Excel schedule.

And that's where you can reach Pat.
You can -- we also have Toby and we have a new tutor, Janine and Edward who is also very good with advanced bio stats, and Toby also is very good with bio stats, so if you have any questions with Excel and general stats, more course related work, I would say, I would go to the stats schedule.

It'll say stats in Excel.

And like I said, for determination statistics, I would go to the determination stats schedule.

The tutors, the schedules are open from I believe about 6:00 a.m. to midnight.

They're all day long.

Most tutors do work in the evening because they're students and they also work other jobs, they're graduate assistants, but some work during the day.

Tutoring is now 60 minutes long which is a nice plus because before it was 30 minutes, but we've increased the time as of 4/1.

You can make same-day appointments.

So if you're working on some course work and you got stuck and you need a stats tutor, you can go right in
there and make an appointment.
You can cancel your appointment up to 30 minutes prior.
We also have tutoring in accounting and finance.
And we also have wonderful MS Word and PowerPoint
tutors.
So there's a lot of variation in there for students.
And if you have -- oh, yeah, and if you have any
questions in tutoring, you can reach me at
ASCTutoring@waldenu.edu.
Did I miss anything, Pat?
>> No, and I think if we missed anything, you know,
just use the role accounts that Kim put in there.
Yep.
>> Yeah, I'm sorry, yeah.
That's exactly -- another way to reach a tutor it
through our role accounts.
So we have special email accounts that go directly to
the tutors, they monitor those during the day and
they'll get back to you within 24 hours, 48 hours on
the weekend.
Statsupport@waldenu.edu.

We also have Word support, finance course support.

Everything is actually on our website, and I will put that link in there for everyone so they can go in there on their time to take a look at what we offer.

So hard to say in a few minutes.

>> Yeah.

>> You know?

>> No kidding.

>> Yeah, so that's how -- that link at ASC tutors is where you can reach us.

Pat has a box up here for web links.

Please take our survey and let us know how we're doing, because we're always looking to improve.

We run these throughout the year and Pat and I are always looking for your feedback.

We also have a YouTube channel that is fantastic that has all our stuff, tutorials and webinars and everything we've run, we post on there, the recording.

And there's also a web link for ASC tutoring.

So Pat put some handy links there for everybody.
>> We had two unanswered questions.

Karen, yes, the N, I know I did that very fast.

The N was for the number.

So that was the 4236 was the number of records.

Hassana had a question about, if you have a non-normal distribution with descriptives, so that's where things like the median and the mode can be really helpful.

Like if you had a huge outlier where one value was matter millions but the average is, you know, 20, that's where you would use like a median or even a mode instead of the mean.

So those are examples of that.

Same with the variation.

You know, if you have a huge outlier, You know, sometimes you just eliminate the outlier or you have to factor that in.

And that's why it's so important to run these descriptives and I believe next time we'll be talking about graphs which will be a way to actually detect where those out liars might be.
Pat, quick question on Likert type scales, are they categorical or continuous data?

Or more for continuous data?

That's a great question, and there's no one answer to that.

Because a lot of people will use a Likert scale even in like a regression equation.

But at its core, I would call a Likert scale an ordinal value.

So an ordinal value is a categorical value that is an ordered set.

Meaning that it -- there is a value from, you know, 1 to 2, you know, 2 is greater than 1, 3 is greater than 2, which is different than in a categorical variable like, let's say, ethnicity where you may have 1 for white, 2 for black, 3 for Hispanic.

Obviously there's no numeric value associated with those.

They're just simply codes to classify.

But in a Likert scale, that's called an ordinal value.

Those numbers actually do have a value and a meaning,
but they’re still kind of technically categorical.

>> So it’s like a ranking?

>> Yeah, so the ordinal, it’s an ordered set, so like

income would be another example.

Or grade in school.

So first grade is lower than second grade and second
grade is lower than third grade.

>> Got you.

Okay, good.

Thanks.

That’s a good explanation.

>> There was a question about SPSS for Walden students.

Our understanding is that all Walden students have
access to SPSS.

>> Yeah, and it is free, so you can get it free through

the Center for Research Quality website has some good
information on how to access SPSS.

And if you already have it and you need to upgrade your
license, they also have information on that as well.

>> Yep.
And if you have questions -- if you have a problem with -- technical questions with SPSS, you can go to the student support team through your Walden student portal.

And chat with them and they can help you on that regard, because we help with SPSS skill-wise, and the tools and whatnot, but we don't really do tech support, so if you run into a download issue or something isn't looking right on your screen, reach out to the student support team.

Yep.

Okay, we're a little bit over time.

I want to thank everybody.

Thank you, Kim, for your support as always.

Yep.

It was -- I enjoy it.

These really do go fast.

They do.

I feel like I'm rushing, but then the 60 minutes ones were a little bit too long.
Yeah.

I'm guessing next year it'll be 45 minutes.

Yeah.

[ Laughter ]

We'll see what the students say.

Yeah.

Exactly.

We'll look at the surveys.

But, yeah, so please join us again for our next webinar.

I want -- I can't remember, I think it's on the 26th?

Is that correct?

If you go to our website, we have a registration page where you can go in there and sign up.

I should know actually.

I was working on that earlier.

And I should know as well.

Yes, April 26th is the stats chat live for graphs.

And our Skill Builder, which Pat has the web link in that box there, but I'll put it in here as well.
If you want to register for that, please do so.

>> Yep, and we're getting ready to roll out the May schedule.

So it'll -- coming soon.

Typically there's two -- two Skill Builders and two stats chats per month.

But -- yeah.

>> I just have a question about the transcript.

Wendy, the transcript will be uploaded onto the -- our website with the recording.

In a few days, give me about 48 hours and I'll get that up for you.

So we'll have the presentation, we'll have the transcript, and we'll have the recording for everybody.

You're welcome.

So what time's your tutoring session, Pat?

>> It's -- it's actually now.

I -- yeah, I had a 7:00, but I told her to wait until 7:30, and then I have three more after that, so I go all the way up to 11:00 my time.

I'm gonna go ahead and stop the recording.
>> Okay.

>> Okay.

We're off recording.

Yeah, no rest for the wicked.

>> No.

No, well, you're in demand.

That's for sure.

>> What happened was when the new schedule came out,

before I even realized, my schedule started filling up

so fast that I didn't even look at, like what was,

like, next week and the following week.

So I already had sessions scheduled from like a week

ago.

So I was like booked solid.

[ THIS TEXT IS BEING PROVIDED IN A ROUGH DRAFT FORMAT.
COMMUNICATION ACCESS REALTIME TRANSLATION (CART) IS
PROVIDED IN ORDER TO FACILITATE COMMUNICATION
ACCESSIBILITY AND MAY NOT BE A TOTALLY VERBATIM RECORD
OF THE PROCEEDINGS. ]