Okay.
Hello.
This is the recording for logistic regression series 4.
Discriminant analysis and multinomial logistic regression.
(No audio.)
>> Hello, Dr. Zin.
>> Hello.
Morning, Nancy.
How are you doing?
>> I'm doing very well.
Hello, everyone.
We're going to get started in just a couple minutes with our logistic regression series part 4.
So I just want to say thank you all for joining us today from the academic skills center for this presentation of discriminant analysis and multinomial logistic regression.
This is part 4 of a series of presentations on logistic regression hosted by Dr. Zin HTWAY.
So before we get started with Dr. Zin and his presentation, I just want to go through a few housekeeping items.
This session will be recorded and it will be available for future viewing on the academic skills center's website.
The link can be found in the web links for you box.
So let's go over a few things and issues in our room.
Our room has a presentation, which where you see the PowerPoint presentation.
It has a captioning area, which is at the top right-hand corner of your screen.
It has a web link for you and a chat area.
Now, we muted the microphones and we will not be using the raise hand feature for your presentation.
Now, if you want to make your viewing area expanding, you will see a square four-arrow box at the top of your screen.
This will expand the view to full screen.
And then if you want to minimize it again, you hover over the top right-hand corner and you'll see the same control.
And if you click it, it will bring you back to your regular screen.
So again, everyone, thank you for joining the academic skill center today for the fourth part on logistic regression hosted by Dr. Zin HTWAY.
Dr. Zin completed his Ph.D. from Walden University in public health -- oh, my golly.
Sorry.
Epidemiology.
Boy, that was a challenge this early.
I am very sorry.
(Laughing.)
Let me try again.
Dr. Zin completed his Ph.D. from Walden University in public health/epidemiology.
He also holds an MBA from Western Governor's University and a degree in cytotechnology.
Now, he is the operations manager of an atomic pathology laboratory and hazmat safety officer at a local officer.
He also lectures in the department of biology and natural sources and is a support specialist here at the support center with a focus on doctoral students.
Now, please welcome Dr. Zin.
>> For today's life webinar or skill builder series, I want to provide a brief overview, application SPSS utilization and write-up of logistic analysis and doctoral research.
Now, as many of you are aware, these are complicated statistical tests.
Each is appropriate when the dependent outcome is categorical.
The decision to use one test over the other will be dependent on the research question and the independent variables.
So the -- when we consider the discriminant analysis, discriminant analysis is a technique used
to predict membership based upon several variables. There's a linear discriminant analysis, which is the classical form of discriminant analysis. It is used to study the relationship from several predictor variables. And as we move along through the session today, I hope to make that a little bit more clear for all of you.

All right. So when we consider the statistical assumptions of discriminant analysis, the dependent variable should be measured at the categorical level. The continuous variable -- so the dependent variables are the predictors -- all have to be multivariate and normally distributed for each of the populations or each of the groups. There should be no outliers in this. The population variances are the same across all levels of this factor. The participants, of course, are randomly sampled and the scores are independent for all other participants. So there's no relationship. There should be no outliers or highly influential points for any of the factors or any of the variable.

When we look at the assumptions of multinomial logistic regression, the dependent variable should be measured. There is a different set -- there is one of the differences between multinomial and logistic regression. However, the variables must be treated as continuous or categorical. We also have observation. There should be no multiCollinicate. There needs to be a linear relationship between any continuous independent variable and the transformation of the dependent variable because multinomial logistic regression relies on the transformation.
And once again, there should be no outliers or unfluential points. So for today's skill builder session, we're going to be looking at a single research question in the same data set. And we're going to do -- first, we'll do discriminant analysis, then we'll follow up with running multinomials. So our research question is are age, educational attainment, last prenatal visit, and ethnicity factors that predict if a mother will complete a three-shot new vaccination series. We actually have four groups: The first is no vaccination, then first complete, then second shot complete, and then our fourth group is fully vaccinated. We look at our independent variables. The first is age, which is a scale; second being attainment in years, which is also a scale; the third is month of prenatal visit; and the fourth is not used for analysis. And we've got Pacific islander, Asian, native American, and black, and white. So to run discriminant analysis. We go to analyze. We use the dropdown menu and use the side bar to go to discriminant. We move across vaccine shots. And then for our scales, it should be age -- I meant to take out ethnicity because we left that in the variables box. We're just using the three scales. And now, when we -- this is one thing that's important. We have to define the range, actually, for our dependent variable. So we click on the define range button. And here, we're going to put in 0 as the the minute and max is 3. Let's say you're interested in 3 out of the 9, this is where you select those 4 or 3.
Then we just go ahead and click continue. 
And then when we get to -- there we go. 
And then we're just going to hit the okay button to run it. 
At this point, I just want to make a note that discriminant analysis in SPSS has many, many options. 
And what I'm doing here is running the bare minimum. 
As you investigate more into discriminative analysis, you may need more, but since there's so much complexity to it, I'm going to keep it pretty simple this morning. 
So when we run -- we click okay. 
And then we start to get our output. 
And so the first thing that we'll see here is it's a summary of the canonical discriminant functions. 
First of all, this indicates the first or second linear function. 
The number is equal to the number of discriminating variables. 
If there are more groups than variables or one less than the number of variables in the group variable. 
These are the values of the group sums of squares and cross product matrix and the between groups sums of squares and cross product matrix. 
I know it's quite confusing, but this is really what the values mean. 
The values are related to the canonical and describe how much the value weighs. 
So the better the function. 
We have percentage of variance. 
This is the continuous variables found in a given function. 
The proportion of the discrimination of the ability and will sum up to 1. 
And we've got the -- the last column is the canonical correlations of our predictor variables and the grouping variables in vaccination shots. 
The next box that SPSS puts out is the test -- this is the test of functions. 
Now, the test of functions associated with the
functions are all equal to 0.
Right?
In this example, we've got 3 functions, thus, the
first test and the second test are 2 through 3.
And the third is the third correlation alone.
The next is one of the multivariate statistics
calculated.
It is the product of the values of one minus the
canonical correlation squared, actually.
This can test the correlation of a given function as
equal to 0.
And then, of course, we've got the degrees of
freedom and the last column is the SIG.
The smaller canonical correlation is equal to 0, so
we can actually see that our test of functions 1
through 3 was actually significant and 2 through 3
is actually significant, but the third function is
not significant.
And here, we have the -- the standardized canonical
discriminant function coefficients.
And these can be used to calculate the score for any
given case.
It is calculated in the same manner as a predicted
value from a linear regression using the
standardized coefficients and the standardized
variables.
One of the processes is to have SPSS calculate this
out for you or it will -- SPSS will actually create
a table of each of your -- each of your individual
participants and show you where they actually
grouped in terms of your data and where they should
have grouped or the way that the mathematical
calculation in SPSS would have grouped them.
But those tables becomes quite lengthy, but it does
have value in certain research questions.
So next, we have the structural matrix.
This is the canonical structure, also known as
loading or discriminant loading of discriminant
functions.
It shows the dimensions created with the functions.
For those of you not particular with this, the term
canonical has to -- it relates back to cannons to shoot cannon balls.
The mathematics is similar.
You have lines of equations, and those are the functions.
So within SPSS, it looks to see how predictable or how much value and prediction those mathematical lines actually have.
So here, we have the functions of the group centroids.
If we look at the means for the scores by groups, we find that the no vaccine shots group has negative 1.1775.
Then a mean of negative .463.
And those fully vaccinated is .38.
I'm sorry, that should have been three variables.
This is where -- this is a large -- this is a big difference between discriminant analysis.
With multinomial logistic regression, we won't see -- I'm sorry.
We're just looking to see if these predictors can't predict the outcome.
No calculation is required here.
And then, of course, we go on to the second function and indicates that the overall predictor is noticed after the effect of the first function.
Additionally, the original was not significant and did not differentiate among the shots after parsing out the facts of the first and second.
And then we move into logistic regression.
I'm going to go a little quicker here.
We analyze regression, and then we go to multinomial logistic.
And here, we'll put vaccine shots in the dependent box.
Now, factors -- in the factors box, we'll put our categorical, which is ethnicity.
And then we put our scales: Age, education, month of last pre-natal visit.
We have to actually give -- we have to tell which is going to be the reference category.
This is going to be where the calculations are based off on.
In this particular time, I'm going to use the last category as our reference category and then we're going to go ahead and click okay.
Here, we get -- our first model is the model fitting output.
And this is the way that the equation actually fits, and we can see that it is significant.
Our SIG value is less than .001.
We look at our pseudo R square.
Essentially, the higher the value, the better the bit of of fit of the mathematical equation.
When we look at the likelihood ratio test, we can see each of our variables here.
In the first column, we have age, education, prenatal visit, and ethnicity.
We notice that ethnicity is not significant -- not a significant predictor.
And then here, we get our pram better estimates.
And this is the area of multiple -- I'm sorry, multinominal logistic regression, which gives you your output where things get complicated.
But when we look at our first vaccination shot complete, now, our reference is -- our reference value for our dependent variable is actually no shots.
We're trying to see who is getting shots.
I'm sorry.
We ticked the last category.
So when we look at -- when we consider age and we look at the -- I'm sorry.
When we look at age and we see the data in the far right, we see it's 0.0125.
And so every year increase in age, we have an increase in the odds ratio of someone getting their first vaccination shot.
I'm sorry.
Our reference was no shots.
Education for every year, increase in education, they're 60% more likely to get their first
prevaccination shot.
and we see an increase in the first vaccination shot.
With ethnicity, it's all over the board in terms of the odds ratio.
For the second shot, once again, we looked at age. And for every year increase in age, we looked at a 27% increase -- this is in the data column.
Education is positively correlated.
So there's an increase, and then same with the prenatal visit.
We get about a 43% increase.
And once again, ethnicity doesn't really show to be a significant predictor.
And to be fully vaccinated, these also are positively correlated.
So for age, once again, there's an increase.
Education was a little over two-fold, increasing getting fully vaccinated.
And 70% was fully increased.
And once again, ethnicity is not really a predictor here.
And so -- your APA write-up for multinomial logistic regression is longer because you have to address each dependent group and each of your variables.
However, it's definitely doable, it's just going to be substantially longer than what we saw with discriminative analysis.
But once again, between that and regression, you're looking to see whether or not your independent variables dependent variables are predictors of your outcome, of your group.
With multinomial logistic regression, you're looking more for a mathematical model to explain a participant will wind up in a particular group.
So with that being said, we've got a few minutes for any questions.
So if you're willing to, I'd like to take a few questions here.
You can put them into the chat box.
>> I think everybody's been concentrating hard on
what you've been saying, Dr. Zin, because there are have been no questions in the chat box.

>> It's a complicated test and the math is not something that we learn in school, so -- but anybody --

>> It looks like we have quite a few people typing right now.

>> Okay.

>> So I think we have some questions in there.

>> All right.

Good.

>> Okay.

Here they come.

>> All right.

>> How discriminative analysis is used in survey or test construction?

So when you do a survey, you've got -- the survey will provide you with essentially independent factors or predictors. You're going to use this to determine whether there's any correlation from the answers on the survey and the group that you're actually looking for.

I'm not sure if I answered your question correctly. Let's see.

I think we're waiting for -- a few more people are typing here.

I'm not sure if this will be the test that I need to perform in my research.

I have four independent variables that relate to a dependent variable. You'll have to look to see that you can meet the assumptions of the test and also whether or not these tests will answer your research question. If your dependent variable is categorical and your dependent variables on a scale and normally distributed -- right, have a normal distribution -- then you can use discriminative analysis.

Again, it has to relate to your research question. Distributing the population into different groups. You can use industrialattive analysis for that, but
you'll have to -- you know, on your survey -- you have to be predicting group membership.
So really, it's focused more on the variables.
Can you do certain predictors -- put people in the different pigeon boxes, so to speak?
Is there no other simple way of doing this than SPSS?
Well, it's very powerful, very deep, but it's also very easy to use.
You can run multiple statistical tests within SPSS.
The challenge is becoming familiar with the output and how to read the output.
There's a lot of resources available through the ASC.
We have a number of tutors, as Nancy said earlier.
I'm a tutor for dissertation students.
And you can make appointments with us and send us emails, and we'll try to get back to you.
We're here to help the students complete their schooling.
They're all on scale, the dependent is behavior intention to use technology.
Discriminative analysis is appropriate.
If intention is on the scale or ordinal, you may want to take a look at ordinal logistic regression which will give you the ratio for one reference category or the change in rare you don't to -- or the change in behavioral intention.
Can you elaborate for on the canyons?
It looks at the different independent variables and it creates different mathematical lines on a three-dimensional grid so that -- to see where the lines actually cross.
Because it's determining whether or not these lines can actually predict the -- the outcome or can predict where this person's going to -- this particularly participant is going to land in terms of a group.
And it's -- it is rather complicated and they look to see where the lines actually cross, but this is all done within SPSS.
We don't actually see it.  
It doesn't plot out on a grid, so. 
The template different -- Dr. Zin, is the template different for multiple linear regressions than others you've presented? 
Actually, the template is very similar to the ordinal logistic regression. 
You just have to include essentially different numbers in terms of the -- the output, and you'll also have to structure your write-up -- you have to actually show the -- or explain which is the reference variable and then the -- the odds ratio changes for each of your independent variables. 
And so it becomes much, much longer, but the templates are pretty much the same. 
They're just repetitive. 
Well, I have they're all done here. 
So I'm going to give this back to Nancy.
>> Okay. 
Well thank you very much. 
I want to thank everyone for joining us today and I hope you will be able to join us for additional sessions sponsored by the skill building center. 
Thank you, Dr. Zin, for an informative session. 
>> Thank you, Nancy. 
Thank you, everybody. 
>> On your screen, you will see information about the academic skills center, including how to reach a tutor if you need free one-on-one support for your studies. 
Please join our social media for updates on events and help us improve on our offerings. 
I will post it in the chat box as well. 
If you have any questions about our center or any questions -- further questions for Dr. Zin, please feel free to email us. 
Thank you all again for attending. 
Have a good day.