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: Lori Morrow, RMR, CCP, CRR, CLR
Paradigm reporting & Captioning Inc.
612.339.0545
Caption@paradigmreporting.com

>> Nancy: Welcome, everyone.
I'm hoping that you can hear me okay.
And I want to thank you for --
>> Dr. Zin: Good morning.
>> Nancy: Oh, thank you.
Thank you for joining the Academic Skills Center for this presentation of simple Logistic Regression, part one of a series of presentations on logistic regression hosted by Dr. Zin Htway.
Before we get started, I want to mention that this session is being recorded.
If you'll give me a second, I will start that recording.
There we go.
And this presentation will be available for future viewing on the Academic Skills Center website. We'll provide the URL for this at the end of the session.
Let's go over a few housekeeping rules.
Everyone has a presentation area and a captioning area.
So you'll be able to see -- you'll be able to read
what everyone is saying in a chat area at the bottom of the screen.
And in the center of the screen to the right you'll see a question and answers area where if you have any questions for Dr. Zin, you'll want to type them in that area for us.
We have muted the microphones, and we will not be using the raise hand feature.
If you want to expand the viewing area, you will see a square, four-arrowed box at the top of the screen. This will expand the presentation area to full screen on your computer which can then be minimized with the same control at the top right-hand corner of the full screen.
Again, thank you for joining the Academic Skills Center today for the first part of our series on Logistic Regression hosted by Dr. Zin.
Dr. Zin completed his Ph.D. from Walden University in December 2014 in Public Health Epidemiology. He also holds an MBA in healthcare management from Western Governor's University and an undergraduate degree in clinical science - cytotechnology from California State University - Dominguez Hills. His background is in pathology and cytopathology. He's currently a supervisor and operations manager of an anatomic pathology laboratory and HAZMAT safety officer at a local hospital. He also lectures at California State University Channel Islands in the Department of Biology/Natural Sciences and is an instructional support specialist in Walden's Academic Skills Center assisting students with statistics and SPSS with a focus on doctoral students.
Please welcome Dr. Zin.
>> Dr. Htway:  Good morning, and thank you for the kind introduction, Nancy.
For today's StatsChat Live webinar I want to provide a brief overview application, SPSS utilization, and APA style write-up of simple logistic regression for doctoral research.
As many of you are probably aware, logistic regression
is a mathematically complicated statistical family of tests. As Nancy mentioned, this session is the first of a four-part series. I will cover some of the basic aspects of logistic regression in this session, and then we'll introduce additional aspects of logistic regression in the StatsChat Live sessions that follow. So logistic regression, also called a logit model, is a test stat used to mathematically model dichotomous outcome variables in which there are only two possible outcomes from a set of independent variables which can be discrete and/or continuous limited to a single independent variable and a single dichotomous dependent variable. So logistic regression differs from linear regression in many ways, firstly, there needs not to be a linear relationship between the dependent and independent variables. Secondly, the independent variables need not be multivariate normal. This is not a concern for simple logistic regression. However, for multiple logistic regression, multivariate normality yields a more stable solution. Thirdly, homoscedasticity refers to the assumption that the dependent variable experiences similar amounts of variants across a range of values for an independent variable is not a concern. And lastly, both ordinal and nominal data can be independent variables in logistic regression. So here are some of the assumptions for logistic regression. First, binary logistic regression requires the dependent variable to be binary, and ordinal requires the dependent variable to be ordinal. Since logistic regression assumes that the function of the probability of Y, which is the probability of the event occurring, it is necessary that the dependent variable is coded accordingly. What that means is that when you go into SPSS and you give a value for your dependent variable you would
code zero as a non-event and one as the event. Thirdly, the model should be fitted correctly. All meaningful variables should be included. A good approach to ensure this is to use a step wise method to estimate the logistic regression. Next, the error terms need to be independent. Logistic regression requires each observation to be independent. That's very common to many statistical tests. Next, logistic regression assumes linearity of the independent variability and the log odds. And I'll explain that further as we go along. Next and lastly, logistic regression requires quite large sample sizes. Multiple logistic regression needs at least ten cases per independent variable and some statisticians require at least 30 cases for each parameter to be estimated. In this particular session I will not be discussing sample sizes for logistic regression. I'll save that for our next session. It's a rather complicated mathematical equation. For logistic regression, it's important to have a strong understanding of basic probability. If you need a refresher on probability, my colleague, Dr. Patrick Dunn, recently presented a StatsChat Live! session on the public accountability. It's available for review from the Academic Skills Center home page. For logistic regression, make the effort to become familiar with probability, which is equal to the probability of event written as P event, which is equal to the outcome of interest divided by all possible outcomes. Next is be familiar with odds, which is the probability of an event occurring divided by the probability of the event not occurring. Mathematically, it's written as P divided by 1 minus P. The maximum value of probability of the event occurring is also 1.0, which is symbolized by P.
The range for probability goes from 0 to 1.
So the value of P not occurring is equal to 1 minus P.
The odds ratio, using the term OR, is really a ratio
of two odds where you have odd to the first divided by
odd to the second. And mathematically, it's written
as the group of P1 divided by 1 minus P1 which is
entirely divided by P sub 0 divided by 1 minus P sub
0.
These statistical terms make up a large part of the
language of logistic regression.
Now, getting back to linear regression. In linear
regression we have an independent variable which may
predict a continuous dependent variable.
The statistical analysis calculates a mathematical
model for a straight line.
Y equals B not plus B1 times X where Y is the
dependent variable.
B not is the Y axis intercept.
BY is the slope coefficient.
The statistic regression is similar, but of course
it's different.
Since we cannot graph a dependent variable against an
independent variable to produce a continuous line the
logistic regression begins by plotting the probability
of the dependent variable outcome against the
independent variable.
This gives us a line, an "S" curve line, but still a
continuous line.
The function of "S" curve line is known as pi.
And the formula is shown here.
Note that the exponential is the familiar formula for
a linear regression analysis.
B0 -- or B not plus B1X.
Now, applying the logistic function to the probability
is commonly known as taking the logit of pi.
This function can also be referred to as the logit of
taking the probability of interest.
If we now take the logistical transformation, or the
logit of pi and plot the result against the
independent variable, the result is a familiar Y axis
intercept and slope coefficient.
Note that the middle formula between the equal signs is similar to the formula for odds except probability is replaced by pi.
The exponentiated coefficient, EXP (B1) represents an odds ratio.
To run a simple logistic regression in SPSS, you would click on analyze, get the drop-down menu to go to regression, and then use a side menu and go to binary logistic.
I'm using a fictitious data set for this one.
Now you move the dependent variable to the dependent box and move the independent variable to covariates.
Now, you click on the options tab and select CI for EXP(B) 95%.
The software will now include the 95% confidence interval for the exponential coefficient EXP(B1).
And here we have -- this is the SPSS output for my fictitious data.
They're seen here in the variables for the equation table.
When you run SPSS, you will get a similar table.
The large B towards the left is the unstandardized beta weights, and we can see that for the constant, the unstandardized beta weight is equal to negative 2.503, and we can see that for the predictor variable, the unstandardized beta weight is equal to 0.296.
What that means is when there is a one unit increase in the predictor variable, this is associated with a 0.296 increase in the logit variable which relates back to the probability of the dependent variable or the outcome of interest.
The EXP(B), which is towards the right half, gives us the odds ratio.
What that means is when there is a one-unit increase in the predictor variable, this is associated with a 1.345 times greater likelihood of the dependent variable or the outcome of interest occurring.
Also note that the 95% confidence interval for the odds ratio ranges from 1.254 to 1.443.
Notably, it does not include the value of 1, so we have a good odds ratio here.
Now, in logistic regression, what I have here is a sample template for an APA write-up for simple logistic regression.

So a logistic regression analysis to investigate fill-in-the-blank for your research question was conducted.

The predictor variable was tested to test that there was no violation of the assumption of the linearity of the logit, the predictor variable.

This is where you fill in the name of your variable. And the p logistic regression analysis was found to contribute to the model.

The unstandardized beta weight equals WWW. Standard error equals XXX.

The wall equals YYY.

The estimated odds ratio favorite, it was either a positive or inverse relationship of nearly N-fold where the exponentiated beta equals ZZZ, and the 95 confidence interval ranges from AAA to BBB for every one unit increase in -- once again, you fill in the blank with your variable.

When I fill in the blanks for my fictitious data, it now reads a logistic regression analysis to investigate if there's a relationship between age retirement planning was conducted.

The predictor variable age was tested apiary to test if there was no violation of the assumption of the linearity of the logit, the predictor age, and the regression analysis was found to contribute to the model.

The unstandardized beta weight to the constant B equals negative 2.503. Standard error equals 0.543. The wall statistic equals 21.948. And P was less than 0.41.

The unstandardized beta weight, beta equals 0.293. The wall statistic equals 67.892, and the P value was less than .001.

The estimated odds ratio, in nearly 35% were the exponentially equals beta 1.345 with the 95% confidence interval ranges from 1.254 to 1.443 with
retirement planning for every one unit increase in age.
So in summary, as stated by Sablok and Larson in 1998, in a simple regression model we relate a covariate X1 to the binary response variable Y in the model log P divided by 1 minus P which is equal to beta naught plus beta 1 times X1 where P is equal to the variability of YX1.
We're interested in testing the null hypothesis where beta 1 is equal to 0 against the alternative hypothesis where beta 1 equals beta prime where beta prime does not equal 0.
That the covariate is related to the binary response variable.
The slope coefficient beta 1 is a change in the log odds for a increase in one unit in the independent variable X1.
So when we speak about the null hypothesis where beta 1 equals 0, that refers to the odds ratio or essentially the slope, so that if there is no change with the independent variable, then we would reject the alternative hypothesis and accept the null hypothesis, but our alternative hypothesis is that there is a change in the odds.
So that's why they write it in as the beta prime not equaling 0 which relates back to beta 1 not equaling 0.
So at this time, we have a few minutes for questions. If you have a question related to the presented program, please type your question into the questions and answer box.
So we have a question here from Lee A. It says please how do I generate data from two to three variables on SPSS software.
So to have two to three variables using the software in the covariate as objection, you would transfer not just the single variable, but you would transfer all of the independent variables that you were interested in.
In the next StatsChat Live! session, I will be covering multiple logistic regression so you'll get a
better explanation there.
And I will also provide more information regarding
some of the details that I skimmed over for simple
logistic regression.
So here we have a question: What is the importance
of SPSS in logistic regression?
The importance of SPSS in the logistic regression is
to actually run the calculations.
Because the mathematics is rather complicated, as you
probably already know, it's better to have the
computer actually run all the analyses for you.
One aspect I also suggest to students is that
essentially beef up their computers with as much RAM
as possible because SPSS to run these calculations can
actually crash your computer, especially if you have a
large data set.
Many students experience this.
I've experienced it. I've learned to actually boost up the RAM memory in
all my desktops as well as my laptop to keep this
going.
We have a question here from Heidi.
My research question is what is the predictive
relationship, if any, between program participation,
yes and no, and inpatient days when controlling for
gender, age, and comorbidity levels.
Logistic or linear regression?
Heidi, since your dependent variable is dichotomous,
the yes or no, which is the program participation, I
recommend that you go with logistic regression.
It would be more appropriate.
And then you would utilize your inpatient days as your
independent variables along with gender, age, and
comorbidity levels.
>> What is the most recent SPSS software that is being
used in 2016?
I believe that the Walden University Research
Department provides students' use of SPSS Version 21.
I'm currently using Version 23.
I don't really believe that there's much difference
for the research that I'm doing with Walden
University, not much change really.
What could be the challenge in using the SPSS, if any? SPSS is actually a very -- is a very powerful software that once you become familiar with its use it becomes easier and easier to use. It's somewhat daunting because it is such a large software with so much capabilities. But if you use it regularly, you'll find that it's incredibly fast, and you can run your analyses over and over again several times making minor adjustments in just a few minutes.

And if you need help with using SPSS, I strongly suggest that you make an appointment with one of the Academic Skills Center tutors. We have a question from Lee A.

Is it possible to generate a data set on my own for analysis on SPSS software, or must I get it from previous work done?

This is a great question. A lot of students ask how do I make my own data set, especially for doctoral research? My recommendation to students is to actually make your data set, get all of your data, and put it in Microsoft Excel or another spreadsheet program. I use Excel all the time. The reason I recommend you do it in Excel is because Excel has a capability of mathematically adjusting the individual cells across different variables. It makes it much, much easier.

Once you have your data set built in Microsoft Excel, you can actually copy and paste each column into the SPSS data set.

Now, there are ways of using the transform function in SPSS to actually have SPSS open up a Microsoft Excel data set, and then it will convert it to an SPSS data set. However, what I have found is that that only really works with smaller spreadsheets or smaller sets of data.

When you have large sets of data where you have several hundred participants or even thousands of participants, the data winds up becoming corrupted,
and then, once again, there's the possibility that your computer will actually crash because it cannot handle that much programming.

So my recommendation is build it in Microsoft Excel, copy, say, column B because that's your variable of gender, and then just go ahead and paste that directly into the SPSS data set; and then once you have the numbers within SPSS data set, then you go to the variable view, and then you set up all of your variables accordingly to being ordinal or nominal or string values and you put all the values in that way.

If you need additional assistance with creating a data set for SPSS, once again, I recommend making an appointment with an Academic Skills Center tutor. You can make an appointment with me, and I'll walk you through the process.

A question from Barut. Can you comment on the difference between simple and multi-factorial regression with respect to using the variables.

I'm going to save this discussion regarding factorial logistic regression to another session because I would like to stay focused on the simple logistic regression for today.

But we will be doing additional. In about two weeks, I believe we have another session which will cover multiple logistic regression, and then two weeks following will be a program on ordinal logistic regression, and then the last of the four-part series will be on discriminate analysis.

A question.

I tried to use my own data set but with challenges.

I once again recommend you make an appointment with an Academic Skills Center tutor, and the tutors can help you using your -- you know, to find out where the errors are occurring in your data set.

Dr. Zin, can you please add more examples and practice sets to the presentations being posted.

Yeah, if we can do that. And we will definitely do that for the subsequent sessions to follow.

Question: Is logistic regression relevant to study relating to hospital administration?
Rashid, you can use logistic regression for any study, but it comes back to really the research question, what is it that you want to know and how are you going to measure it. SPSS only knows numbers. And so it's how you're going to code those numbers for your independent and dependent variables will determine which statistical test you're going to use. If your dependent variable is binary dichotomous, of course you can use logistic regression. If it is going to be an ordinal, then you can use logistic regression. If it's categorical, you can also use logistic regression.

Okay. I think that we only have a couple of minutes left here. So I'm going to send the presentation back to Nancy.

>> Nancy: Well, thank you very much. And thank you all for joining us today. This has been quite an informative presentation, and I want to thank you, Dr. Zin, for the dynamic session?

>> Dr. Zin: Thank you, Nancy.

>> Nancy: On your screen you'll see additional information about the Walden University Academic Skills Center, including how to reach a tutor if you need free one-on-one support in your studies. Please join our social media sites for updates, tips and events, and if you have any questions please email us at ASCtutoring@waldenu.edu. And I will post that email address in our chat area because if you have any follow-up questions, you can send them there, and we will get them to Dr. Zin, and we will get your answers for you. Again, everyone, thanks for joining us.

Have a good day.

>> Dr. Zin: Thank you, everybody.

Thank you, Nancy.

DISCLAIMER: 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.