Spring 2013, Volume 3, Issue 1

THE STATSWHISPERER

The StatsWhisperer Newsletter is published by Dr. William Bannon and the staff at StatsWhisperer.™ For other resources in learning statistics visit us on the web at: www.statswhisperer.com

Introduction to this Issue

We are often asked "What are the basic steps involved in most statistical analysis projects?" Therefore, we came up with an answer, which is **PUB Magic**! **PUB Magic** is a mnemonic phrase we created to represent four steps involved in most statistical analysis projects. These steps are the <u>P</u>reparation of data, <u>U</u>nivariate analysis, <u>B</u>ivariate analysis, and <u>M</u>ultivariate analysis. This phrase **PUB Magic** refers to how your analysis can be the "<u>M</u>agic" that will enrich your "**<u>PUB</u>**"lication.

The next several issues of the *Research Pathways* newsletter will examine specific techniques within each of these steps. Eventually, when enough methods are discussed, the newsletter series will present a rich view of the techniques, issues, and processes involved in conducting statistical analysis.

This addition of our newsletter will address a very useful statistical method incorporated within the <u>M</u> (Multivariate analysis) quadrant of these four steps. Specifically, this addition will provide an overview of Binary Logistic Regression. A favorite of researchers everywhere!!

However, please remember binary logistic regression is a complex procedure with many details beyond the scope of this newsletter. The newsletter is structured to present the most immediate facts that one needs to know to begin to understand this statistical test. Ideally, the newsletter may facilitate a working knowledge of the technique that will lead to further study.

INSIDE THIS ISSUE ON BINARY LOGISTIC REGRESSION

Binary Logistic Regression: The Rock Star of Regression

Why do we call Binary Logistic Regression the "Rock Star of Regression"? Well, it is because this statistical method is famous! Furthermore, this method truly does rock (metaphorically)!

"The method is famous?" you may ask. Yes, researchers and non-researchers alike know this statistical test worldwide! However, most people do not know that they know this method. For example, most people have heard an odds of something happening mentioned on the TV news. Something like "people that drink fish oil are twice as likely to develop a craving for worms relative to other people." Well, have you ever wondered where they got the value "twice as likely" from? Usually, it is based on a binary logistic regression! See the test is famous, but anonymous!

And the method rocks because it produces an odds ratio statistic that most people around the world understand immediately (at least somewhat)! Even among people that speak different languages! How many facets of research and/or statistics can you say that about?

The ABCs of Binary Logistic Regression

Stated definitively, *Binary logistic regression* is a form of regression which is used when the dependent variable is a dichotomous (e.g., Yes/No) and the independents variables are of any type.

In order to illustrate the properties of this test, I would like to use a relatively simple model we put together for a client several years ago (DeVoe, Bannon, & Klein, 2006). This binary logistic regression model examined which independent variables predicted parental help seeking of mental health counseling for their children after the September 11th terrorist attacks in New York City. So we have a *dichotomous dependent variable, namely did parents seek counseling for their children (Yes/No)*. And independent variables of all types (i.e., continuous and categorical) that will be examined as predictors of the dependent variable.

The binary logistic regression model has many parts with several important details. We will use the analysis presented in Table 1 to help illustrate the most basic elements needed to understand this procedure. However, to reiterate, anyone pursuing an analysis using this procedure is encouraged to consult multiple sources, including text books and teachers, to fully round out their knowledge of this statistical test. This is a great technique, so please feel free to get to know it intimately.

Table 1.
Results Binary Logistic Regression Analysis of Parental Help-seeking of Mental Health Counseling for Their Children After the September 11th Terrorist Attacks in a Sample of 180 Families (87.78% of the Cases Were Classified Correctly)

Variable	B (*SE*)	Wald (X^2)	Odds ratio (95% CI)	*p*
Full child PTSD diagnosis	0.60 (0.60)	1.02	1.83 (0.57-5.86)	.32
Child afraid of new things	1.20 (0.61)	3.84	3.33 (1.00-11.11)	.05
Parental symptoms of depression	0.80 (0.41)	3.83	2.23 (1.00-5.00)	.05
Parental symptoms of anxiety	-0.32 (0.*32*)	1.00	0.73 (0.39-1.36)	.32
Parent sought counseling for self	1.89 (0.75)	6.37	6.63 (1.53-28.82)	.01
Child saw WTC attack in person	2.29 (0.58)	15.29	9.83 (3.13-30.91)	.0001
Child 9/11 media exposure	-0.93 (0.30)	9.51	0.40 (0.22-0.71)	.002

Note. PTSD = posttraumatic stress disorder; WTC = World Trade Center; CI = confidence interval.

Model = X^2 = 54.29, df=7, p<.0001

Pieces of the Binary Logistic Regression Analysis Puzzle

As in most statistical analysis procedures, there are many statistical parameters and values involved with a binary logistic regression model. However, there are a few that are considered *essential* to understand when interpreting a specific model.

Look at it this way; you are describing the regression model like you would describe a restaurant dining experience to a friend. When describing your restaurant dining experience, you likely would mention the main elements of the dining experience, such as the quality and price of the food. You would not likely mention the waiter's shoes (unless you are Imelda Marcos). Thus, regarding our binary regression model, other statistical details exist, but in this newsletter we are just covering the main statistics that are typically reported.

Puzzle piece 1: Odds ratios

If a binary regression model is the rock star of regression, the odds ratio is its greatest hit. As stated earlier, it seems that the odds ratio is the most sought after and understood statistical value in the world. It is the song everyone wants to hear.

What exactly is an odds ratio? An odds ratio is a measure of effect size. The odds ratio is the standard way to report the central results of logistic regression. The value indicates the likelihood of the dependent variable event occurring relative to another factor (the independent variable).

For example, in Table 1 under the variable list, second from the bottom, you will note the variable "Child saw WTC attack in person." This independent variable is dichotomous (Yes/No). Thus, via binary logistic regression, we can examine if a child saw the WTC attack in person (Yes), what are the odds that their parents would seek help (Yes) for them in the form of mental health counseling. Subsequently, under the column marked "Odds Ratio" on that line, you will see the value 9.83, which reflects the fact that children that saw the WTC attack in person were almost ten times (Odds ratio=9.83) more likely to have their parents seek mental health counseling for them, relative to children that did not see the WTC attack in person. We also see under the column marked *p* on this line that probability is below .05 (i.e., .0001), indicating a statistically significant relationship.

Note 1a. Please note the term "relative to children that did not see the WTC attack in person" mentioned in the last paragraph. Whenever you mention who is more or less likely to experience an event, it is important to mention to whom they are being compared. For example, I could have left the last sentence in the above paragraph like this:

…children that saw the WTC attack in person were almost ten times (Odds ratio=9.83) more likely to have their parents seek mental health counseling for them.

But this is incomplete. I should specify more likely than whom? That is why the sentence reads:

…children that saw the WTC attack in person were almost ten times (Odds ratio=9.83) more likely to have their parents seek mental health counseling for them, relative to children that did not see the WTC attack in person.

This may seem like a small issue, but it is a hallmark of attention to detail.

Pieces of the Binary Logistic Regression Analysis Puzzle

Note 1b. The odds ratio between a dichotomous dependent variable and continuous independent variable does not have the same meaning as when an independent variable is categorical.

Specifically, when the independent variables is categorical in a binary logistic regression model, the odds ratio is interpretable as reflecting the degree to which how more or less likely an event might be based on that independent variable.

However, when the independent variable is continuous in a binary logistic regression model, it is not appropriate to interpret the odds ratio as reflecting the degree of likelihood that an event is likely to occur be based on that independent variable.

In other words, if the odds ratio is 2.0, and the independent variable is categorical, it is appropriate to state that the event is twice as likely (2.0) to occur. However, if the odds ratio is 2.0, but the independent variable is continuous, it is not appropriate to state the event is twice as likely (2.0) to occur.

Using Table 1 as an example, please note the second independent variable down from the top "Child afraid of new things," which is answered in a categorical yes/no format. We see the value of *p* is .05, which indicates that a statistically significant relationship with the dependent variable (help-seeking for child). The odds ratio is 3.33. Therefore, it would be appropriate to state that "children that became afraid of new things were over three times more likely (odds ratio=3.33) to have parents engage in help-seeking of mental health counseling on their behalf."

Also, please note the third independent variable from the top "Parental symptoms of depression," which is a continuous measure (from 0 to 4) of symptoms of depression. We see the value of *p* is also .05, which indicates that a statistically significant relationship with the dependent variable (help-seeking for child). The odds ratio is 2.23. However, because this is a continuous independent variable, it would not be appropriate to state the relationship between the independent variable and dependent variable as "being over two times more likely (odds ratio=2.23)." The odds ratio here is still a statistical property, but not one which has the same meaning in terms of stating the likelihood of an event.

Note 1c. An odds ratio greater than 1.0 indicates an increased likelihood, while an odds ratio below 1.0 indicates a reduced likelihood.

For example, in Table 1, the final independent variable is "Child 9/11 media exposure." We first see that the *p*-value is .002, which indicates a statistically significant relationship between the dependent variable and "Child 9/11 media exposure." We also see that the odds ratio is reported as .40. Well, what does that mean? Up to not we have had clearer values reflecting the likelihood of an event. What's up with that?

Actually, there is really no problem. All you need to do is realize two things. First, to reiterate, an odds ration below 1 means the variable is protective or the event. In other words, the independent variable

Pieces of the Binary Logistic Regression Analysis Puzzle

is associated with a reduced likelihood of the event (the dependent variable) occurring. In this instance, the odds ratio indicates that a child having experienced the independent variable (a "Yes" for "Child 9/11 media exposure"), is associated with the dependent variable not occurring (a "No" for parent help-seeking for the child). A sort of curious finding, but that's what the data reflect.

The second thing you need to know is how to convert the odds ratio below 1 into a communicable form. This is simple, but it is often hard to find out how to do this. In essence, the formula is to divide the number 1 by the odds ratio below 1. In our example, we would divide 1 by .40, which equals 2.5.

Converting the odds ratio below 1:
1/the odds ratio=the reduced likelihood
For our example:
1/.40=2.5 times less likely

Thus, we would say that children that experienced 9/11 media exposure (Yes) were two and a half times (odds ratio=2.5) less likely to have parents report help-seeking of mental health counseling after the September 11th 2001 terrorist attacks on the World Trade Center.

Puzzle piece 2: The 95% Confidence Interval (CI)

To put it very briefly, the 95% CI reflects that we are 95% confident that the true value of the odds ratio is between the closed interval presented. For example, please note the independent variable "Child saw WTC attack in person." On that line under the column "Odds ratio (95% CI)," you will see the odds ratio is 9.83 and the 95% CI is 3.13-30.91. This reflects, while the estimated odds ratio given is 9.83, we are 95% confident that the true odds ratio could be between 3.13-30.91. By the way, this is a big interval and is likely related to the small sample size present in the study. Generally, a smaller 95% CI is preferable.

Lastly, within the 95% CI, 1.0 is the null value. Thus, if the 95% CI includes 1.0, the value indicates that the finding may be spurious. You will see in binary logistic regression most statistically significant ($p<.05$) findings evidence a 95% CI that does not include 1.0.

Puzzle piece 3: The Beta and Standard Error

The Beta and Standard Error are reflected in Table 1 as "B(SE)." A discussion of these parameters is beyond the scope of this newsletter. However, I would mention that one basic use of these values is examining if the beta is negative or positive. If the beta is negative there is reduced likelihood of the event (the dependent variable) in relation to the independent variable (this should be accompanied by an odds ration below the value 1). Of course if the beta is positive, this is vice versa.

Puzzle piece 4: The Wald Chi

Briefly the Wald chi, expressed as Wald X^2 in Table 1, is a statistical representation of the magnitude of the relationship between the independent and dependent variables in the binary regression model. It often goes unreported in many studies, but in my opinion it is well worth mentioning.

In Summary (a.k.a, Is Your Model Elvis or Not?)

If you are going to be a rock star, be Elvis (see Figure 1 below)! To many, it seems that being a rock star must be great. However, it also seems apparent that all rock stars are not equal. Some gain widespread notoriety, while some do not. The same is true with research findings. Specifically, some findings are read and soon forgotten, while some are cited by many researchers for decades.

Thus, your job as a researcher is to give your statistical model the ingredients needed to make it a rock star with the notoriety, fame, and staying power of Elvis Presley!

For example, Elvis holds a Las Vegas record. From 1969 until 1976 (the year of his death), he sold out 837 shows (most in arenas the size of a football stadium) for a total of two-and-a-half million people. I am not sure of the equivalent to this in terms of having your research revered and cited by others, but I imagine you'll know when it happens.

What makes the binary logistic regression model Elvis or not? In short the researcher does. The binary logistic regression model is just a series of text and numbers. The researcher gives brilliance, value, and meaning to those letters and values by his or her interpretation of what they mean. This is summer up by one of our main catch phrases at WBA, Inc.:

A great researcher does not get great statistical findings, a great researcher makes statistical findings great.

In other words, once you competently compute your odds ratios, you make them "Elvis" with your brilliant interpretation of the statistical findings.

Figure 1. The King of Rock and Roll, Elvis Presley Performing in Las Vegas

Reference List

DeVoe, E. R., Bannon, W.M., & Klein, T. P. (2006). Post-9/11 Help-seeking by New York City parents on behalf of highly exposed young children. *American Journal of Orthopsychiatry, 76*(2), 167-175.

If you have questions regarding the material presented in this newsletter feel free to contact the staff at WBA, Inc. at the following email address: wb@williambannonassociates.org.

WBA, Inc. is headed by Dr. William Bannon, who is an Assistant Professor at the Mount Sinai School of Medicine, as well as the president of WBA, Inc.

Those of us at WBA, Inc. enjoy providing written materials, as well as private statistical and research consultation to clients. For further information, as well as archived copies of newsletters, visit us on the web at: www.williambannonassociates.org

William M. Bannon, Jr., Ph.D.
1 Gustave L. Levy Lane, #1230
NY, NY 10029

Phone:
(212) 241-0207

Fax:
(646) 596-9610

E-Mail:
wb@williambannonassociates.org

William M. Bannon, Jr., Ph.D.
1 Gustave L. Levy Lane, #1230
NY, NY 10029