What is Regression Analysis?

In short, regression analysis is a statistical method to estimate the relationship between a dependent variable (commonly referred to as “y” or the left side variable)  
 and one or more independent variables (commonly referred to as "x" or the  
regressors). Regression analysis provides a model that allows a user to gain understanding of how the typical dependent variable changes with changes in the independent variables, holding the other independent variables constant. Because of this functionality, regression is typically used to predict or forecast the value of the dependent variable, when values for the independent variables are known.

There are different types of regression models depending on the nature of the data. The simplest type of regression model is a standard linear model. This method may be familiar to a user who have done basic line of best fit work with scatter plots, as seen below.

[insert scatterplot1]

The mathematical process of linear regression creates the equation for the red line on the graph. The process does this by minimizing the square sum of the error terms; one of the error terms is denoted by the blue line. The green dot is an example of a predicted value for x = 86 on the graph. Imagine that the y-axis was the height, the dependent variable, of an individual in pounds, and the x-axis was weight, the independent variable, in inches. Our regression model would then predict that an individual who weighed 86 pounds would be 50 inches tall. Our model seems to predict obese individuals, but the concept of a predicted value is clear.

One of the most important numbers the simple linear regression model spits out is the slope of the estimated line of best fit. The slope value would be in units of pounds/ inches. In this particular example, the slope is close to 0.55 inch. /pound. In other words, the model would predict an individual who was one pound heavier than another, *all other variables constant,* would be 0.55 inches taller.

Why Poisson Regression?

Of course, whenever you include more than one variable and change the characteristics of the predicted variable, the more complicated a model one has to use for accurate results. For example, in the previous illustration the dependent variable was weight—which is a continuous variable (theoretically one could weigh 110 pounds, 110.0001 pounds, 110.0002 pounds…), but what if the dependent variable was a discrete counting number? For our project, our dependent variable—the number of references to God—is a discrete variable for it can have the value 1, 2, 3… but a speech cannot have 1.467 references to God. The Poisson regression model is usually used on data of the counting nature.

One of the primary differences between a simple linear regression model, and a Poisson regression model is what is called the “link function.” A link function is what links the independent variables to the error term. In the standard linear regression, the link function is an identity link, meaning that the coefficients or the “slope terms” can be interpreted as seen above. The link function for the Poisson model is much more complex, with the use of logarithmic functions. The model will be further explained with a hands on approach with the data collected.

How do you interpret the results of the Poisson regression model?

In pure mathematical terms: a one unit change in the independent variable, e.g. the increase of one child, other variables held constant, results in the difference in the logs of the expected counts in the amount of the regression coefficient of that variable. In equation terms:

β= log( μx+1) - log( μx)

Where β is the regression coefficient, μ is the expected count of the dependent variable and the subscripts denote a one unit increase.

Using rules of logarithms, a term called the incident rate ratio (IRR), which has a more direct interpretation as seen in the simple regression model, can be calculated from the Poisson regression coefficients.

Our results

<http://www.ats.ucla.edu/stat/stata/output/stata_poisson_output.htm>

<http://books.google.com/books?id=YKvzt4OZzPsC&pg=PA49&lpg=PA49&dq=regression+on+speeches+of+a+president&source=bl&ots=CI-Jdy4GvV&sig=-Wk1L80-K0ANfE79BkiT0OLPSUM&hl=en&sa=X&ei=clZLU9i-E-TMsQTosoHYAw&ved=0CCYQ6AEwAA#v=onepage&q=poisson&f=false>

“Calculation of the incident rate ratio (IRR) indicates the amount of change in the dependent variable based upon one unit increase in an independent variable, holding all other independent variables constant. An IRR of less than one indicates a negative relationship. “ pg. 48 “The Post-Cold War Presidency” edited by Anthony J. Eksterowicz.

Log(y) = constant + β1x1 +

Y = exp ([the numerical value])