Journal of Personality Assessment, *91*(2), 121–136, 2009

ISSN: 0022-3891 print / 1532-7752 online
DOI: 10.1080/00223890802634175

Routledge
Taylor & Francis Group

STATISTICAL DEVELOPMENTS AND APPLICATIONS

The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives

STEFANY COXE, STEPHEN G. WEST, AND LEONA S. AIKEN

Department of Psychology, Arizona State University

Count data reflect the number of occurrences of a behavior in a fixed period of time (e.g., number of aggressive acts by children during a playground period). In cases in which the outcome variable is a count with a low arithmetic mean (typically <10), standard ordinary least squares regression may produce biased results. We provide an introduction to regression models that provide appropriate analyses for count data. We introduce standard Poisson regression with an example and discuss its interpretation. Two variants of Poisson regression, overdispersed Poisson regression and negative binomial regression, are introduced that may provide more optimal results when a key assumption of standard Poisson regression is violated. We also discuss the problems of excess zeros in which a subgroup of respondents who would never display the behavior are included in the sample and truncated zeros in which respondents who have a zero count are excluded by the sampling plan. We provide computer syntax for our illustrations in SAS and SPSS. The Poisson family of regression models provides improved and now easy to implement analyses of count data.

[Supplementary materials are available for this article. Go to the publisher's online edition of *Journal of Personality Assessment* for the following free supplemental resources: the data set used to illustrate Poisson regression in this article, which is available in three formats—a text file, an SPSS database, or a SAS database.]

Many researchers in psychology and the behavioral sciences have theoretical questions that involve count variables. A *count variable* is a variable that takes on discrete values (0, 1, 2, . . .) reflecting the number of occurrences of an event in a fixed period of time. A count variable can only take on positive integer values or zero because an event cannot occur a negative number of times. There are numerous examples of the use of count variables in psychology and the behavioral sciences. Clinical and health psychologists have modeled the number of depressive symptoms that a child exhibits (Computerized Diagnostic Interview Schedule for Children; National Institute of Mental Health, 1997), the number of alcoholic drinks consumed per day (Armeli et al., 2005), the number of readmissions to alcohol detoxification programs (Shanahan et al., 2005), and the number of cigarettes smoked by adolescents (Siddiqui, Mott, Anderson, & Flay, 1999). Sociologists have modeled the number of complaints made by citizens to their local law enforcement office (Worrall, 2002) and the number of accidents occurring at intersections (Chin & Quddus, 2003). Psychologists interested in personality assessment have modeled the number of disciplinary incidents reported among a group of prison inmates (Walters, 2007).

Using count variables in ordinary least squares (OLS) regression may potentially pose problems. Standard (fixed effects) OLS regression has minimal assumptions regarding predictors in the model, so a count variable can be used as a predictor in OLS regression with only one caution. If the variance of the predictor is very small, as may easily occur in count variables with a small range, the regression coefficient for that predictor will be very unstable and will have a large standard error (Cohen, Cohen, West & Aiken, 2003). However, this is not an issue specific to count predictors; any predictor with a small variance will result in an unstable regression coefficient.

Different problems may occur when a count variable is used as an outcome or criterion variable in OLS regression. When the mean of the outcome variable is relatively high (often defined as greater than 10 as a rule of thumb), OLS regression can typically be applied to a count outcome with minimal difficulty. However, when the mean of the outcome is low, OLS regression produces undesirable results including biased standard errors and significance tests (Gardner, Mulvey, & Shaw, 1995).

The purpose of this article is to provide an introduction to appropriate regression models for count outcomes, models that have maximal statistical power while maintaining the proper Type 1 error rate. We first review the specific problems that may occur when standard OLS regression is applied to count outcomes with a low mean. We then introduce Poisson regression, a family of alternative regression models that is more appropriate for outcome variables with low counts. We initially illustrate the application of standard Poisson regression with an example and discuss its interpretation. We then turn to two forms of complications that may occur. First, actual data may have too much variability to be represented by standard Poisson regression, a problem termed *overdispersion.* We discuss and illustrate two variants of Poisson regression designed to address this problem: overdispersed Poisson regression and negative binomial regression. Second, problems may arise in Poisson regression

Received January 2, 2008; Revised June 4, 2008.
Address correspondence to Stephen G. West, Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ 85287–1104; Email: sgwest@asu.edu

associated with counts of zero. We briefly consider two variants of problems with zeros: (a) the inclusion of a group of people who would never display the behavior (e.g., nonsmokers), and (b) people who could potentially display the behavior but who are excluded by the sampling plan (e.g., a study of only those people who are admitted to the hospital). These variants of basic Poisson regression provide more accurate estimates and tests of hypotheses when the assumptions of the basic Poisson model are violated. To help enable researchers to utilize these techniques in their own research, we also provide the computer syntax for our illustrations in SAS and SPSS, standard packages widely used by psychologists.[1] Finally, we consider the strengths and limitations of the Poisson family of regression models for count data.

Assumptions of OLS Regression

Standard OLS regression makes important assumptions related to the model errors that must be met for accurate statistical inference (Cohen et al., 2003, chap. 4). Model errors for each case are defined as the difference between an individual's observed score and their unobserved expected score. Errors are therefore unobserved. Model residuals are estimates of model errors and can be used to assess violations of OLS regression assumptions. Residuals are defined for each case as the difference between the individual's observed score (Y_i) and the individual's predicted score ($\hat{Y}_i$), or $\hat{e}_i = Y_i - \hat{Y}_i$. The three major assumptions regarding the errors for OLS regression are (a) conditional normality, (b) homoscedasticity (or constant variance), and (c) independence. *Conditional normality* of the errors means that for any value of the predictor X, the conditional distribution of errors given X is a normal distribution. *Homoscedasticity* of the errors means that the variances of the conditional distributions of the errors at all values of X have the same common variance, σ^2. *Independence* means that there is no clustering of the data into groups or contexts that would provide information about the values of other errors. These three properties (conditional normality, homoscedasticity, and independence) describe the *error structure* of OLS regression.

Count outcome variables can violate the first two assumptions of OLS regression in several ways. Count variables often violate the assumption of constant variance by displaying increasing conditional variance with increases in the value of the predictor. For example, consider the number of children in a family. Many industrialized countries currently have a mean of fewer than 2 children per family, so there is little variation in the number of children per family in these countries (e.g., birth rates: Germany, 1.40 children per woman; Japan, 1.23 children per woman; United Kingdom, 1.66 children per woman[2]). In contrast, many countries with economies based on subsistence agriculture tend to have much larger mean family sizes (e.g., Laos, 4.59 children per woman; Nigeria, 5.45 children per woman; Uganda, 6.84 children per woman). As the extent of subsistence agriculture in a country increases, the variation in number of children per family would be expected to increase. This statistical property is known as *heteroscedasticity;* it leads to biased standard errors and biased tests of significance when OLS regression is applied. The conditional distributions of count variables also tend to be positively skewed and kurtotic,[3] with many low-count observations and no observations below zero. These conditions of heteroscedasticity and non-normal conditional distributions mean that statistical significance tests such as t tests of regression coefficients will be biased and inefficient for count outcomes (Gardner et al., 1995; Long, 1997). This result means that the actual Type 1 error rate may not match the stated Type 1 error rate (typically, $\alpha = .05$) and that the statistical power to detect true effects may be affected.

Overview of Poisson Regression

Poisson regression is a member of a family of analyses known as the *generalized linear model* (GLiM; see Dobson, 2002; Fahrmeir & Tutz, 2001; Fox, 2008; McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972). GLiM generalizes OLS regression for use with many different types of error structures and dependent variables. The GLiM family of analyses can provide accurate results for data sets having binary, ordered categorical, count, and time to failure (or success) dependent variables. The GLiM introduces two major modifications to the OLS framework. First, it allows transformations of the predicted outcome, which can linearize a potentially nonlinear relationship between the dependent variable and the predictors. This modification implies that the predicted scores can be in a different metric (unit of measurement) than the observed dependent variable scores. In Poisson regression, the observed scores are counts, and the predicted scores are the natural logarithms of the counts. In GLiM, there is a special transformation function called the *link function* that relates the metric of the predicted scores to the metric of the observed criterion scores. In Poisson regression, the link function is the natural log (i.e., $\log_e$ or ln).

Second, the GLiM is flexible in error structure. The *error structure* describes the conditional distribution of the errors around the predicted value. OLS regression assumes a conditional normal error structure, whereas GLiM allows for a variety of other error structures. The foundation for Poisson regression is the Poisson distribution, which is used to represent the distribution of the errors. The Poisson distribution is a member of a set of probability distributions called the *exponential family.* For all distributions in the exponential family, the height of the probability curve for a specific value of Y, called the *probability density,* contains an exponential function (i.e., the constant e = 2.718+ raised to a power). The normal distribution is a familiar member of the exponential family. The equation for the normal distribution is

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}. \quad (1)$$

The probability density or height of the normal distribution depends on two parameters, the mean μ and the variance σ^2.

[1] STATA is another commercial program that is widely used by economists and sociologists for the analysis of count data. It is typically not available in many psychology departments.

[2] Retrieved December 27, 2007 from http://www.nationmaster.com/graph/peo_tot_fer_rat-people-total-fertility-rate.

[3] *Skewness* is the extent of asymmetry of the distribution and is indexed in the population by $\gamma_1 = \mu^3/\sigma^3$, where μ^3 is estimated by $1/N\,\Sigma(X_i - \bar{X}_i)^3$, and σ is the population standard deviation. Excess positive *kurtosis* refers to distributions having sharper peaks and longer tails than the normal distribution. Excess kurtosis relative to the value 0 for the normal distribution is indexed in the population by $\mu^4/\sigma^4 - 3$. Positive kurtosis (long tails) is associated with bias in estimates of standard errors.

These two independent parameters, μ and σ^2, completely specify the normal distribution.

The Poisson distribution differs from the normal distribution in several ways that make the Poisson more attractive for representing the properties of count data. First, the Poisson distribution is a discrete distribution that takes on a probability value only for nonnegative integers; this characteristic of the Poisson distribution makes it an excellent choice for modeling count outcomes, which only take on integer values of 0 or greater. The probability mass function for the Poisson distribution,

$$P(Y = y|\mu) = \frac{\mu^y}{y!}e^{-\mu}, \tag{2}$$

gives the probability of observing a given value, y, of variable Y that is distributed as a Poisson distribution with parameter μ. y! is y factorial = y (y – 1) (y – 2) . . . (2)(1). For the count variable Y, the "number of disciplinary incidents in a prison group" (Walters, 2007), μ is the arithmetic mean number of incidents that occur in a specific time interval; the Poisson distribution would yield the probability of 0, 1, 2, . . . incidents, given the mean μ of the distribution. In contrast, the normal distribution is continuous and takes on all possible values from negative infinity to positive infinity and not just zero and positive integers.

The probability of a specific count also depends on the variance of the number of counts. In fact, the Poisson distribution is specified by only one parameter μ. The parameter μ defines both the mean and the variance of the distribution; both the mean and variance equal μ. That the mean and variance are equal will often be useful in modeling count outcomes, which typically display increasing variance with increases in the mean. In contrast, the normal distribution requires two independent parameters to be identified—the mean parameter μ and the variance parameter σ^2.

The Poisson distribution increasingly resembles the normal distribution as the expected mean value becomes larger. As a rule of thumb, a Poisson distribution with an expected value greater than 10 approaches a normal distribution in shape and symmetry. However, the Poisson distribution is still discrete and has a single parameter that describes both the mean and variance. Figure 1 shows the probability of each number of events for several different values of μ. Notice how the distributions with very low means are right skewed and asymmetric; the distribution with a mean of 10 appears nearly symmetric. The variances of distributions with higher means are larger.

Poisson regression is a GLiM with Poisson distribution error structure and the natural log (ln) link function. The Poisson regression model can be depicted as

$$\ln(\hat{\mu}) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p, \tag{3}$$

where $\hat{\mu}$ is the predicted count on the outcome variable given the specific values on the predictors $X_1, X_2, \ldots, X_p$. Recall that ln refers to the natural logarithm, b_0 is the intercept, and b_1 is the regression coefficient for the first predictor, X_1. The use of GLiM with the Poisson error structure resolves the major problems with applying OLS regression to count outcomes, namely, nonconstant variance of the errors and non-normal conditional distribution of errors.

Equation 3 looks very much like an OLS regression equation. There is a linear relationship between each predictor and the

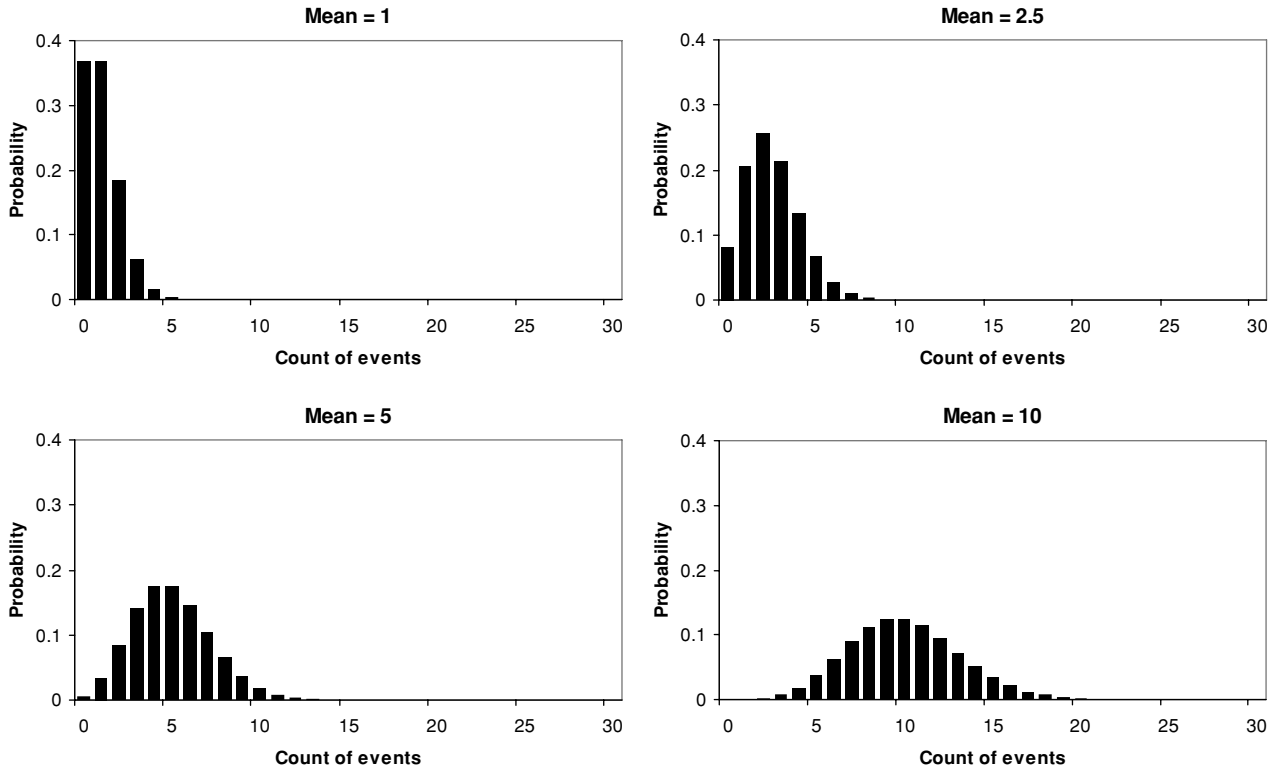

FIGURE 1.—Poisson distributions with different values of the mean parameter.

predicted score just as in OLS regression. What differs from OLS regression is that the predicted score is not itself a count but rather the natural logarithm of the count. Thus it is said that Poisson regression is "linear in the logarithm." It turns out that all the types of regression analysis subsumed under the GLiM model have this same property—there is a form of the regression equation that is linear. For those familiar with logistic regression, there is a form of the logistic regression equation that is linear; specifically, logistic regression is "linear in the logit," that is,

$$logit(\hat{\pi}) = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p, \quad (4)$$

where $\hat{\pi}$ is the predicted probability of being a case. The regression coefficients of these linear equations are simple and easy to interpret given familiarity with OLS regression coefficients. These linear equations, however, have the disadvantage that the predicted scores are not in the same metric as the observed dependent variable scores; rather, the predicted scores are in a less familiar transformed metric. For Poisson regression, the transformed metric is the natural logarithm of counts rather than counts themselves.

Assuming a conditionally Poisson error distribution also means that the residuals of a Poisson regression model are assumed to be *conditionally* Poisson distributed rather than normally distributed as in OLS regression. A discrete distribution such as the Poisson distribution will represent the discrete nature of the residuals that must occur with a discrete outcome. Otherwise stated, because the observed values are counts, the residuals may take on only a limited set of values.

Interpretation of Coefficients

Recall that the Poisson regression model is of the following form: $\ln(\hat{\mu}) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p$. One interpretation of the regression coefficients is in terms of $\ln(\hat{\mu})$ and is similar to that of OLS regression. A 1-unit increase in X_1 results in a b_1 unit increase in $\ln(\hat{\mu})$, holding all other variables constant. This interpretation is straightforward but has the disadvantage of interpreting the change in the units of a *transformation* of the outcome (i.e., the natural logarithm of the predicted count). This interpretation may be of lesser interest when the counts representing the outcome reflect a meaningful scale. Otherwise stated, the researcher would like to talk about how the predictors are expected to affect *the number of times the event occurs*. A second interpretation in terms of the count variable can be obtained following some algebraic manipulation of the regression equation. Raising both sides of the previous equation to the power of e results in

$$e^{\ln(\hat{\mu})} = e^{(b_0+b_1X_1+b_2X_2+\cdots+b_pX_p)}. \quad (5)$$

Note that performing the same operation on both sides of an equation does not change the equality of the two sides of the equation. A property of e and the natural log is that $e^{\ln(x)} = x$, so the left side of the Equation 5 can be simplified, resulting in

$$\hat{\mu} = e^{(b_0+b_1X_1+b_2X_2+\cdots+b_pX_p)}. \quad (6)$$

Now we have an equation that shows the effect of the predictors on the actual predicted count, but it is not yet obvious how each of the predictors contributes to the expected count. A property of exponents is that $x^{a+b+c} = x^a x^b x^c$, so the single term on the right side of the equation can be broken up into several smaller parts, resulting in an equivalent equation:

$$\hat{\mu} = e^{b_0}e^{b_1X_1}e^{b_2X_2}\cdots e^{b_pX_p}. \quad (7)$$

Now we can see that changes in a predictor result in *multiplicative* changes in the predicted count. This contrasts with OLS regression in which changes in the predictor result in *additive* changes in the predicted value. To further clarify the interpretation, we can look at the term for a single predictor such as X_1: $e^{b_1X_1}$. Using the property of exponents shown previously, we can examine the effect of a 1-unit change in X_1 on the outcome:

$$e^{b_1(X_1+1)} = e^{b_1X_1+b_1} = e^{b_1X_1}e^{b_1}. \quad (8)$$

The e^{b_1} term in Equation 8 is the effect of a 1-unit change in X_1 on the outcome. For a 1-unit increase in X_1, the predicted count ($\hat{\mu}$) is *multiplied by* e^{b_1}, holding all other variables constant. Note that the unstandardized (raw) regression coefficient is exponentiated. Given that the variance (standard deviation) of the outcome variable depends on the predicted value in Poisson regression, exponentiation of the standardized coefficient does not yield an easily interpretable value.

Example: Drinks Consumed by College Students on Saturday Night

This example uses simulated data that were created to closely match the characteristics of data originally collected by DeHart, Tennen, Armeli, Todd, and Affleck (2008) in a study of alcohol consumption by students at a university in the northeastern United States. The observed outcome is the report of the number of alcoholic drinks that an individual consumes on one particular Saturday night during the study. Commonly in studies of alcohol consumption, one alcoholic drink is defined as one 12 ounce can or bottle of beer, one 5-ounce glass of wine, one 12-ounce wine cooler, or 1.5 ounces of liquor straight or in a mixed drink. In the DeHart et al. (2008) data set, the students completed an eight-item subscale of sensation seeking (excitement seeking, facet E5 of extroversion) from Costa and McCrae's (1992) Revised NEO Personality Inventory. Each answer could potentially range from 1 (low) to 7 (high) on sensation seeking. The actual range of the mean scores for the eight items in the simulated data was from 3.19 to 6.49. As with many Likert-type scales, no meaningful value of 0 is defined for this scale. Also in the data set is the participant's gender, coded 0 = female and 1 = male. For our simple illustrative example, we initially hypothesized that sensation seeking would have a positive relationship with the number of alcoholic drinks consumed. Later we consider a two predictor model in which both sensation seeking and gender are used to predict the number of alcoholic drinks consumed. The SAS and SPSS computer syntax for the full set of analyses of the two-predictor model are presented in Appendixes A and B, respectively.

As a starting point, we used standard OLS regression to predict the number of alcoholic drinks consumed from the continuous measure of sensation seeking. This analysis permits

TABLE 1.—Regression coefficients, standard errors, and p values for example using only sensation seeking to predict number of alcoholic drinks consumed.

	(1) OLS Regression	(2) Poisson Regression	(3) Overdispersed Poisson	(4) Negative Binomial
Intercept	−.4514	−.1403	−0.1403	−0.0826
Standard error	0.9991	0.2128	0.3591	0.3524
p value	0.6514	0.5096	0.6960	0.8147
Sensation seeking	0.6523	0.2315	0.2315	0.2205
Standard error	0.1907	0.0397	0.0669	0.0667
p value	0.0006	<.0001	0.0005	0.0009
Scale $= \sqrt{\phi}$	N/A	1.0000	1.6873	1.0000
Alpha	N/A	0	0	0.7179
Deviance	N/A	1151.6884	404.5057	452.5635
R^2 or Pseudo R^2	0.0284[a]	0.0296[b]	N/A	N/A

Note. OLS = ordinary least squares; N/A = not applicable. In the column labeled "(1)," the predicted scores are in the form of a count of number of drinks; in columns labeled "(2)"–"(4)," the predicted scores are in the form of the natural logarithm of the number of drinks.

[a] R^2. [b] Pseudo R^2.

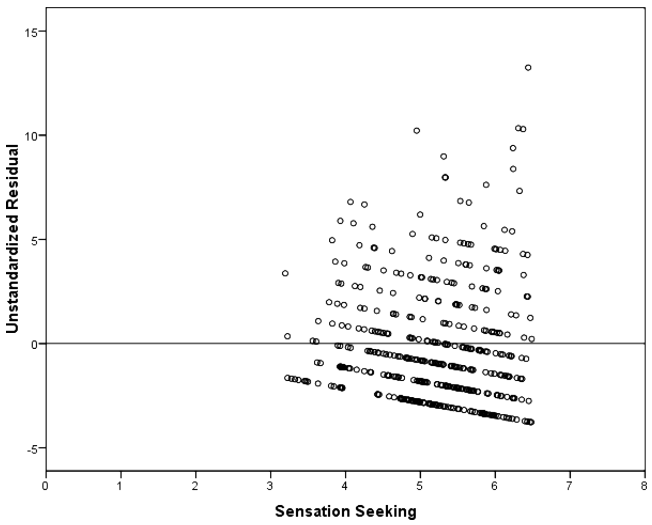

FIGURE 2.—Scatter plot of ordinary least squares residuals as a function of sensation seeking. The variability of the residuals increases as values of sensation seeking increase. The residuals fall in a set of parallel straight lines with a slightly negative slope termed *stripes.* For any specified value of sensation seeking, there is a greater density of residuals with smaller than with larger values, indicating the positive skew of the distribution of the residuals conditional on sensation seeking.

comparison of the later results of the Poisson family of models to those of OLS regression; use of OLS regression represents the standard, nonoptimal practice in much of current literature in psychology. The column labeled "(1)" in Table 1 shows the regression coefficients for this model. Sensation seeking significantly predicts number of alcoholic drinks consumed. A 1-unit increase in sensation seeking resulted in an average of 0.6523 more drinks consumed. Additionally, note that the intercept, that is, the predicted number of drinks consumed for an individual who scored 0 on the sensation-seeking measure, is negative, specifically, –0.45 drinks. Of course, a negative count is impossible. This illustrates yet another potential problem with applying OLS regression to count dependent variables—the predicted scores will sometimes be out of range of actual counts, specifically, impossible negative values. Figure 2 shows the scatter plot of the OLS residuals versus observed scores on the sensation predictor for this model. Heterogeneity of variance is clearly present (i.e., there is more variability in the residuals at the higher sensation seeking scale values than at the lower values), indicating that OLS regression is not appropriate for these data. In addition, note that the residuals are arrayed in a set of parallel continuous lines, termed *stripes,* each of which has a slightly negative slope. Stripes occur in scatter plots with count data because the outcome can only take on integer rather than continuous values, limiting the potential values that may be taken by the residuals.

We then used standard Poisson regression to predict the number of alcoholic drinks consumed from sensation seeking. The column labeled "(2)" of Table 1 shows the regression coefficients for the Poisson regression model. The linear form of the prediction equation is $\ln(\hat{\mu}) = -0.1403 + 0.2315\text{sensation}$, in which the predicted scores are in the form of the logarithm of counts, rather than actual counts, as in OLS regression. In this form, the unstandardized coefficient $b_0 = -0.14$ is the predicted logarithm of the counts when sensation seeking is 0, and $b_1 = 0.23$ is the predicted change in the logarithm of the counts corresponding to a 1-unit change in sensation seeking. These interpretations are straightforward but are in an unfamiliar metric. To interpret the results in the original count metric, we exponentiate the equation for the $\ln(\hat{\mu})$, resulting in the equation for the predicted number of alcoholic drinks consumed: $\hat{\mu} = e^{-0.1403}e^{0.2315\text{sensation}}$. The exponentiation of the intercept, $e^{-0.1403} = 0.87$, is the predicted number of alcoholic drinks consumed by a person who has a score of zero on the sensation-seeking measure; note that this is a positive value as opposed to the negative value from the OLS regression. Given that the measure of sensation seeking does not include zero (indeed, the minimum value of sensation seeking observed in this data set was 3.19), this intercept value has little importance in this example.[4] The exponentiation of the regression coefficient for sensation seeking, $e^{0.2315} = 1.26$, is the predicted multiplicative effect of a 1-unit change in sensation seeking on number of alcoholic drinks consumed. A person with a sensation seeking score of 5 is expected to consume, on average, 1.26 *times* as many drinks as a person with a sensation seeking score of 4.

Figure 3 shows a scatter plot of the observed data with the predicted regression line for both the OLS and Poisson regressions. OLS regression produces a straight line (solid) representing the predicted values. In contrast, there is slight curvature in the Poisson regression prediction line (dotted) in this data set, reflecting the exponential or multiplicative nature of the model. In other data sets in which there is a stronger relationship between the predictor and outcome, a smaller mean on the outcome, or both, the curvature in the Poisson regression prediction line can be more dramatic (see, e.g., Cohen et al., 2003, Figure 13.4.2, p. 529).

[4]The *intercept* is the predicted value of the outcome when each of the predictors is equal to zero. In this illustration, if the measure of sensation seeking were rescaled so that zero had a meaningful value (e.g., if it were rescored as a 0 to 6 scale or it were mean centered, $\text{sensation}_C = \text{sensation} - \text{mean[sensation]}$), then the intercept would be interpretable. The lack of attention of the scaling of the predictors unfortunately makes the intercept a parameter of no interest in many behavioral science applications (see Cohen et al., 2003; Wainer, 2000).

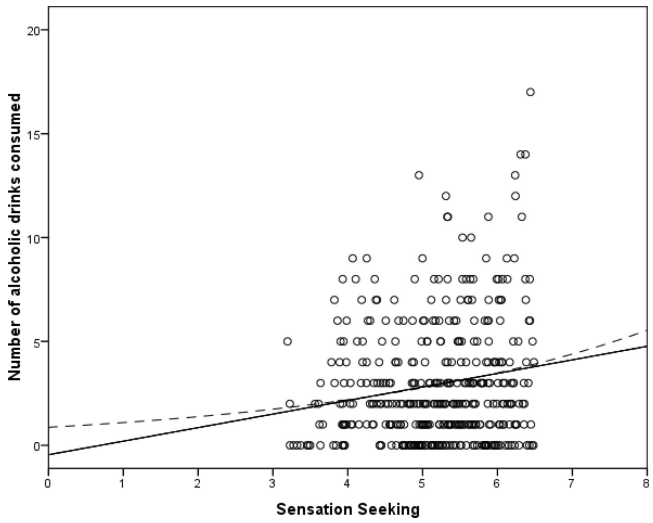

FIGURE 3.—Scatter plot of outcome versus sensation seeking. The straight solid line represents the ordinary least squares fit; the slightly curved dotted line represents the fit of the Poisson regression model.

MODEL FIT AND GAIN IN PREDICTION

Parameter estimation for GLiMs such as Poisson regression employs maximum likelihood methods as opposed to the OLS estimation used for OLS regression. Maximum likelihood estimation finds the population parameters that are most likely to have produced the observed data. (See Enders, 2005, for a more complete explanation of maximum likelihood estimation that is accessible to behavioral science researchers.) In essence, maximum likelihood estimation works by considering many different parameter values, calculating the probability that the observed data came from a population with those parameters, and choosing as estimates the value of each parameter that yield the highest probability of having produced the observed data. After all parameters have been estimated for the model, it is useful to have a measure of how adequately the model accounts for the data.

R^2 and Pseudo R^2 Measures

In OLS regression, the squared multiple correlation or R^2_{multiple} (or equivalently R^2) is a measure of the proportion of variation in the outcome that is accounted for by the predictors. More generally, the R^2 reflects the extent to which the model accounts for the observed data. The calculation of the R^2 measure in OLS regression is based on the variation of the criterion, also called the *total sum of squares* (*SS*). Variation in the criterion is the sum of the squared deviations of the scores on the dependent variable around their mean, $\sum (Y - \bar{Y})^2$. OLS regression completely partitions total *SS* into a portion that is explained by the regression model (explained *SS*) and a portion that remains unexplained (residual *SS*). The R^2 is calculated as the proportion of variation accounted for by the model, or explained *SS* divided by total *SS*.

Poisson regression, like most models in the GLiM family, has no direct analogue to R^2. However, any model estimated with maximum likelihood methods will produce a *deviance* value for the model, which can be used to assess fit of the model. The deviance value differs from an R^2 value in two important ways. First, whereas the R^2 value can indicate absolute fit of the OLS model without reference to other models (i.e., proportion of variation in the outcome variable that is accounted for by the model), the deviance is a relative measure, so it can only be interpreted in relation to another model. Second, the R^2 value reflects the goodness of fit of the model or how well the model accounts for the data. In contrast, deviance represents badness of fit of the model or how much worse the model is than a perfectly fitting model (i.e., a model that has as many predictors as participants and therefore can predict all values perfectly). A large deviance is undesirable, reflecting that the model is much worse than, or deviates a great deal from, the perfect model.

In Poisson regression, the total variation in the outcome can *not* be completely partitioned into explained and unexplained portions[5] (Cameron & Trivedi, 1998). Instead, the deviance of the model can be used to calculate a pseudo-R^2 measure. These pseudo-R^2 measures help to assess how well the model predicts the outcome. The pseudo-R^2 measures actually represent the proportional *reduction in deviance* due to the inclusion of the predictors. Recalling that a large deviance reflects poor fit relative to a perfect model, the pseudo-R^2 measures tell researchers how much closer they have gotten to that perfect model. They do not represent the proportion of variation accounted for by the model (as R^2 does in OLS regression).

Consider a Poisson regression model that contains no predictors. This model will have one regression coefficient, an intercept b_0, which represents the average count. This model is analogous to an OLS regression model with no predictors other than the intercept b_0; the predicted score for every individual is $\bar{Y}$, the arithmetic mean count averaged across all individuals. Cameron and Windmeijer (1997) suggested that the deviance of a Poisson regression model with only an intercept (i.e., no predictors) can be completely separated into two nonoverlapping parts. Recall that deviance is a measure of lack of fit; this measure is reduced by adding predictors to the intercept-only model if the predictors have some accuracy in accounting for the outcome. The first part is the deviance that remains to be accounted for after predictors are included (the *unexplained deviance* or *deviance of the fitted model*). The second part represents the amount by which the deviance is reduced by adding predictors that account for the outcome (or *explained deviance*); the explained deviance of the fitted model is the deviance of the intercept-only model minus the deviance of the fitted model. Then

$$R^2_{\text{deviance}} = 1 - \frac{\text{deviance(fitted\_model)}}{\text{deviance(intercept\_only)}}. \quad (9)$$

[5]Three problems arise with calculating R^2 for Poisson regression in the same manner as R^2 for OLS regression (Cameron & Trivedi, 1998). First, the convenient partitioning of the total sum of squares, $\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2$, breaks down. Because of the nonlinear relationship between $\hat{Y}_i$ and Y_i, the cross-product term $+2 \sum (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$ will not equal zero. Second, because maximum likelihood estimation does not minimize the residual sum of squares, R^2 can be less than 0 or greater than 1. Indeed, adding a predictor variable can potentially decrease R^2. Third, the interpretation of R^2 is problematic when the variance around the regression line changes as a function of the predicted value.

Like the R^2 measure from OLS regression, this pseudo-R^2 measure is bounded by 0 and 1 and typically increases as predictors are added.

For the preceding example of the one-predictor equation with sensation seeking as the only predictor, the deviance for the model can be found in the column labeled "(2)" of Table 1. The deviance is 1151.6884. The deviance of a model with no predictors (i.e., the intercept-only model) is 1186.7614 (not shown in Table 1). Then

$$R^2_{\text{deviance}} = 1 - \frac{\text{deviance(sensation)}}{\text{deviance(intercept\_only)}} = 1 - \frac{1151.6884}{1186.7614} = 0.0296. \quad (10)$$

Including the sensation-seeking predictor reduced the deviance by about 3% compared to using no predictors.

We can also compute an R^2_{deviance} measure for the two-predictor equation containing sensation seeking and gender as predictors. From the column labeled "(2)" of Table 2, the deviance with the two predictors is 959.4582. Then

$$R^2_{\text{deviance}} = 1 - \frac{\text{deviance(sensation + gender)}}{\text{deviance(intercept\_only)}} = 1 - \frac{959.4582}{1186.7614} = 0.1915. \quad (11)$$

Including both predictors reduced the deviance by 19% compared to using no predictors. This is analogous to a squared multiple correlation for a two-predictor OLS regression equation.

Other pseudo-R^2 measures have been developed for GLiMs. Cox and Snell (1989) developed a pseudo-R^2 measure using the likelihoods for the intercept-only model and the tested model to assess improvement from the intercept-only model. A problem with the Cox and Snell measure is that unlike R^2 from OLS and the pseudo-R^2 for GLiM, the Cox and Snell index does not have a maximum value of 1. Nagelkerke (1991) adjusted the Cox and Snell measure to have a maximum value of 1 by dividing the Cox and Snell measure by its maximum value. The Nagelkerke fit measure has a minimum of 0 and a maximum of 1.

Gain in prediction for nested models. In OLS regression, one thinks in terms of gain in prediction by the addition of one or more predictors to an equation already containing at least one predictor. In GLiM, one thinks of additional reduction in deviance by the addition of predictors to an existing equation. We compare two models, one of which is *nested* under the other. Two models are *nested* if the more inclusive model contains *all* predictors present in the simpler model. Continuing with our example of prediction of alcohol consumption, suppose that the researcher wants to examine the proportional reduction in deviance by the addition of gender to an equation containing only sensation seeking as a predictor. In this case, the sensation seeking only model is nested within the gender plus sensation seeking model; the sensation seeking only model serves as the base model. The deviance of the one-predictor model, given in the column labeled "(2)" of Table 1, is 1151.6884; the deviance of the two-predictor model, given in the column labeled "(2)" of Table 2, is 959.4582. The proportional reduction in deviance by the addition of gender to sensation seeking is given by

$$R^2_{\text{deviance}} = 1 - \frac{\text{deviance(sensation + gender)}}{\text{deviance(sensation)}} = 1 - \frac{959.4582}{1151.6884} = 0.1669. \quad (12)$$

Adding gender to sensation seeking has reduced the deviance by almost 17%. This is analogous to an R^2 for gain in prediction in OLS regression.

TABLE 2.—Regression coefficients, standard errors (*SE*), and *p* values for example using sensation seeking and gender to predict number of alcoholic drinks consumed.

	(1) OLS Regression	(2) Poisson Regression	(3) Overdispersed Poisson	(4) Negative Binomial
Intercept	−2.2325	−0.7888	−0.7888	−0.6436
Standard error	0.9360	0.2148	0.3258	0.3329
p value	0.0171	0.0002	0.0155	0.0532
Sensation seeking	0.7719	0.2608	0.2608	0.2351
Standard error	0.1749	0.0388	0.0589	0.0616
p value	<.0001	<.0001	<.0001	0.0001
Gender	2.3699	0.8395	0.8395	0.8224
Standard error	0.2685	0.0629	0.0954	0.0956
p value	<.0001	<.0001	<.0001	<.0001
Scale = $\sqrt{\phi}$	N/A	1.0000	1.5169	1.0000
Alpha	N/A	0	0	0.5105
Deviance	N/A	959.4582	416.9831	459.7743
R^2 or Pseudo R^2	0.1868[a]	0.1915[b]	N/A	N/A

Note. OLS = ordinary least squares. N/A = not applicable. In the column labeled "(1)," the predicted scores are in the form of a count of number of drinks; in columns labeled "(2)"–"(4)," the predicted scores are in the form of the natural logarithm of the number of drinks.

[a] R^2. [b] Pseudo R^2.

Tests of Significance of Model Fit

In OLS regression, we have a test of significance of the R^2 and, in addition, a test of significance of gain in prediction by the addition of predictors. For both, the test statistic is the *F* test. In contrast, in GLiMs, the test statistic is a chi-square test. The chi-square test is the difference between two deviances; the first deviance is that from a base model, and the second deviance is that from a more complete model. The chi-square test examines the reduction in deviance from the addition of one or more predictors to a base model. Consider first the single predictor model with sensation seeking as the only predictor. The base model is the null model containing only the intercept, with deviance 1186.7614, shown previously. The deviance of the model containing sensation seeking is 1151.6884. The chi-square test statistic for the contribution of sensation seeking to prediction is given as $\chi^2(1\ N = 400) = 1186.7614–1151.6884 = 35.073$. The degrees of freedom for the chi-square test equal the number of predictors added to the base model to form the more complete model: here only one, sensation seeking. The tabled critical value of χ^2 with 1 *df* at $\alpha = 0.05$ is 3.84; sensation seeking makes a significant contribution.

We may also test whether the two predictors sensation seeking and gender in combination make a significant contribution to prediction. Again the base model is the null model with deviance 1186.7614. The model deviance from the two-predictor model is 959.4582. The chi-square test statistic for the contribution of sensation seeking plus gender to prediction is given as $\chi^2(2, N = 400) = 1186.7614 - 959.4582 = 227.3032$, $p < .0001$. The 2 *df* for this test equal the number of predictors added to the base model, here sensation seeking and gender. The two predictors together make a significant contribution.

Significance of gain in prediction. We may also test for gain in prediction by the addition of one or more predictors to a model containing at least one predictor. In our example, the deviance of the one-predictor model minus the deviance of the two-predictor model is distributed as chi-square with degrees of freedom equal to the number of parameters difference (i.e., the difference in number of predictors in the base model vs. the enhanced model). There is one additional parameter estimated in the two-predictor model compared to the one-predictor model, so the degrees of freedom for this test equal 1. If the two-predictor model fits much better than the one-predictor model, the deviance reduction will be large, and the chi-square statistic will be large. If the two-predictor model does not fit much better than the one-predictor model, the deviance reduction will be small, and the chi-square statistic will be small. Here, a significant chi-square value means that the two-predictor model fits better than the one-predictor model (i.e., the additional predictor is needed). The deviance of the one-predictor model, given in the column labeled "(2)" of Table 1, is 1151.6884, and the deviance of the two-predictor model, given in the column labeled "(2)" of Table 2, is 959.4582, yielding $\chi^2(1, N = 400) = 1151.6885 - 959.4582 = 192.2302$, $p < .0001$. This significant chi-square statistic indicates that the model that includes gender in addition to sensation seeking fits much better than the model with only sensation seeking as a predictor.

Models Containing Interaction Terms and Polynomial Terms

As in OLS regression, terms can be added to the Poisson regression model to represent interactions and curvilinear effects. For example, if we wish to add an interaction between predictor variables X_1 and X_2 to the regression equation, it would be represented as $\ln(\hat{\mu}) = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$. Note that these effects are most easily interpreted in the metric of the logarithm of the count. In this metric, b_0 is the intercept, the predicted ln(count) when $X_1 = 0$ and $X_2 = 0$, b_1 is the slope of X_1 when $X_2 = 0$, b_2 is the slope of X_2 when $X_1 = 0$, and b_3 is the linear × linear interaction. Or, if we wish to study a possible curvilinear effect of X_1, it would be represented as $\ln(\hat{\mu}) = b_0 + b_1X_1 + b_2X_1^2$. Here, b_0 is the intercept, b_1 is the slope of a tangent line to the quadratic curve when $X_1 = 0$, and b_2 is related to rate of acceleration of the curve. These interpretations in the logarithm of the count metric closely parallel the usual interpretations of interactions and quadratic effects in OLS regression (Aiken & West, 1991; West, Aiken, & Krull, 1996). However, typically these more complex models are extremely difficult to interpret when exponentiated to the original count metric.

For researchers who wish to visualize the results of complex Poisson regression analyses in the count metric, we recommend plotting the predicted curves by exponentiating $\ln(\hat{\mu})$ at different values of X_1. For interactions, this involves plotting separate curvilinear relationships of the relationship between Y and X_1 at different values of X_2. If we had found an interaction between sensation seeking and gender in our example, we could plot separate curves for females ($X_2 = 0$) and males ($X_2 = 1$). If instead X_2 were a continuous variable such as age, we could ideally choose meaningful values on the variable (e.g., 20, 30, 40, 50) or choose convenient values within the range of the data (e.g., 1 *SD* below the mean, at the mean, and 1 *SD* above the mean of X_2). Again, we note that the interpretation of the results of complex Poisson regression analyses is far easier in the metric of the logarithm of the count.

Assessing Model Adequacy

In OLS regression, a typical graphical method used to assess model adequacy is to plot the residuals against the observed predictor values as in Figure 2. If there are multiple predictors, the residuals should also be plotted against the predicted value $\hat{Y}$ (Cohen et al., 2003, chap. 4). This plot should show no relationship between the observed value of the predictor and mean of the respective residuals. Additionally, the plot should show constant variance of the residuals across all values of the observed value (homogeneity), a property that is not found in Figure 2; this highlights that OLS regression is not appropriate for these data.

The appropriate graphical method for assessing model adequacy in Poisson regression is to plot the residuals against the predicted outcome values (Cameron & Trivedi, 1998). However, for nonlinear models such as Poisson regression, raw residuals will always be heteroscedastic and asymmetric (Cameron & Trivedi, 1998, p. 141), so alternative types of residuals must be used. Pierce and Schafer (1986) suggested deviance residuals[6] are the best choice for GLiMs, including Poisson regression. Individual deviance residuals are summed to form the model deviance. The plot of the deviance residuals versus predicted values for the two-predictor Poisson regression model is shown in Figure 4 (The SAS and SPSS computer syntax[7] used to generate this plot are presented in Appendixes A and B, respectively). Note that as expected, there is no relationship between the deviance residuals and the predicted value. The lowess fit line on the plot represents the best nonparametric fit of the relationship between the predicted values and the residuals (Cohen et al., 2003, chap. 4). This line relatively closely tracks a residual value = 0, indicating there is no relationship. A nonlinear relationship between sensation seeking and alcohol consumption would be reflected in a curvilinear lowess line.

Another way to assess model adequacy is to compare the predicted values of the outcome to the observed values of the outcome. Long (1997) outlined graphical methods to compare predicted values to observed values including methods to calculate the average predicted proportion for each value of the outcome. That is, of the N cases, how many cases have a

[6]The deviance residual is the square root of the individual contribution of case i to the deviance, $d_i = sign(Y_i - \hat{\mu}_i)\sqrt{2\{l(Y_i) - l(\hat{\mu}_i)\}}$, where the sign($Y_i - \hat{\mu}_i$) preserves the sign of the residual and $l(Y_i)$ is the log of the density of Y_i when $\mu = Y_i$ and $l(\hat{\mu}_i)$ is the log of the density of Y when $\mu = \hat{\mu}_i$.

[7]GENLIN is included in SPSS Version 15 and 16 but only with the Advanced Models module or Graduate Pack.

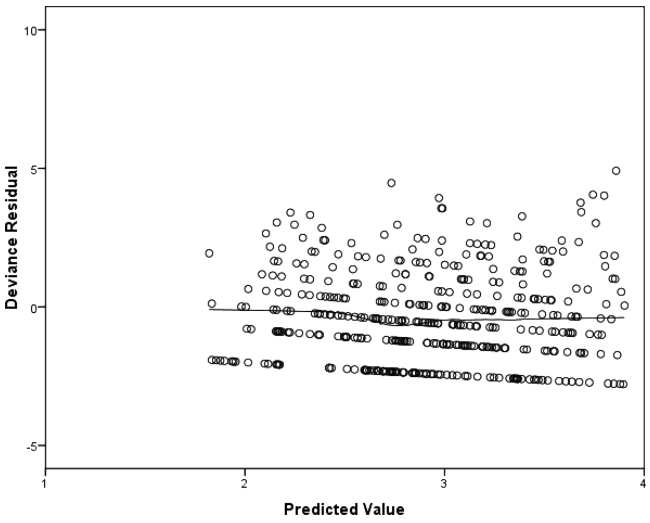

FIGURE 4.—Scatter plot of deviance residuals against predicted value for the two-predictor Poisson regression model. The solid line that closely tracks a deviance residual value of 0 represents the nonparametric lowess fit.

predicted outcome value of 0, of 1, of 2, and so forth? Figure 5 shows the predicted outcome values and the observed outcome values for the single predictor Poisson regression model just discussed. Note that this model seems to underpredict very low and very high values of the outcome and to overpredict middle values.

Diagnostics

Regression diagnostics are a group of statistics that focus on individual cases in the sample to help a researcher detect outliers and poorly fitting cases. For OLS regression, regression diagnostics are well developed and understood, with three main types of diagnostic statistics: *leverage statistics* detect cases that are extreme on the predictors, *distance statistics* detect cases that have large discrepancies between the observed and predicted values on the outcome variable, and *influence statistics* detect cases that have a large influence on the regression coefficients. The diagnostics for distance and influence are based on the key idea of case deletion: The same regression model is estimated with all cases included and then with a specific case (case i) deleted, and the results are compared. Conceptually this would be done N times, once for each case in the data set. For OLS regression, simple relationships exist that allow the full set of diagnostic statistics to be quickly and exactly calculated in a single computer run. The researcher identifies cases with high values on leverage, distance, and influence as potential outliers for further examination. Authors have offered conventions for values on the diagnostic statistics that call for further study of the case and its influence. Cohen et al. (2003, chap. 10) presented a full discussion of these issues.

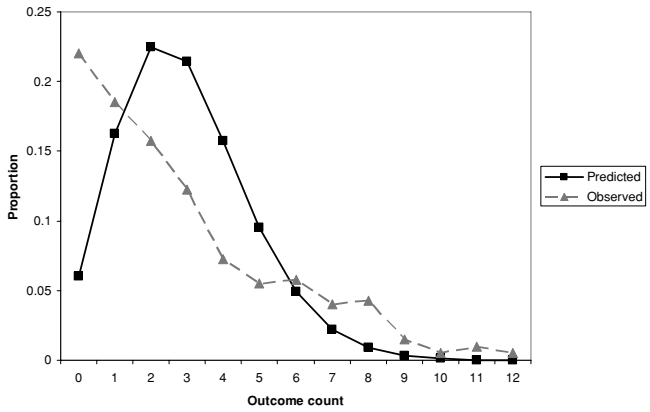

FIGURE 5.—Comparison of observed outcome counts to predicted outcome counts from the single-predictor Poisson regression model.

However, in Poisson regression and other GLiMs, the calculation of diagnostic statistics becomes more complicated, and clear conventions for their interpretation (e.g., cutoff scores) have not been offered. In OLS regression, leverage measures assess how extreme an observation is only on the predictors; the outcome Y is ignored completely. Formally, leverage is a measure of the discrepancy of a case on the set of predictors from the centroid of the predictor space, the point representing the means on all the predictors ($\bar{X}_1$, $\bar{X}_2$, ..., $\bar{X}_p$). Leverage statistics in OLS regression are calculated from a special matrix called the Hat matrix, which is a function of only the predictors in OLS regression. However, in GLiMs, there is a problem with leverage measures. The Hat matrix in GLiMs is not solely a function of the predictors; in fact, in Poisson regression, the Hat matrix is also a function of the mean of the outcome variable for each case, μ_i. Thus the leverage measures for Poisson regression, and for other GLiMs as well, do not have the same meaning as they do in OLS regression and cannot be applied to Poisson regression as they are in OLS regression. Leverage measures are computed in standard software for GLiMs; Appendixes A and B include syntax for SAS and SPSS to save the values of the Hat matrix diagonal using the *LEVERAGE* keyword for both SAS and SPSS.[8]

Distance measures assess the discrepancy between the observed outcome value and the predicted outcome value; all distance measures are based on residuals ($Y_i - \hat{Y}_i$). In OLS regression, one measure used to assess distance for case i is the studentized residual, a special standardized residual. In OLS regression, the computation of the studentized residuals is quite simple; studentized residuals can be calculated for all cases in an analysis simultaneously. In contrast, Fox (2008) noted that in GLiMs exact calculation of studentized residuals requires fitting the regression model to data sets in which each case in turn has been deleted; each of the N deviances would be compared to the deviance of the full model. Such a procedure is computationally intractable in larger data sets. Various procedures that approximate the studentized residual have been suggested. Cameron and Trivedi (1998) suggested estimating the studentized deviance residuals for Poisson regression as

$$d_i^* = \frac{d_i}{\sqrt{1 - h_{ii}}}, \tag{13}$$

[8]Case diagnostics (leverage, Cook's D, and DFBETAS) for PROC GENMOD are available only in SAS Version 9.2 or later. Note that SAS and SPSS will produce slightly different values for deviance residuals and influence because SAS produces leverage values ranging from $1/N$ to 1 and SPSS "centers" the leverage values to range from 0 to $(N - 1)/N$.

where d_i is the deviance residual for case i and h_{ii} is the ith diagonal element of the Hat matrix. Appendixes A and B include syntax for SAS and SPSS to save the values of the raw and studentized deviance residuals using the *STDRESDEV* and *STDDEVIANCERESID* keywords, respectively. Studentized residuals are useful diagnostics; in OLS regression, they signal cases that may be increasing standard errors and reducing power for tests of significance of individual coefficients.

Influence measures assess how much the deletion of a particular case will actually change the values of predicted scores and regression coefficients, that is, affect the results and conclusions from the regression analysis. OLS regression has both global and specific measures of influence. DFFITS and Cook's D assess overall change in the predicted scores from deleting a specific case; DFFITS measures the number of standard deviations by which a case changes its own predicted score when the case is included in versus deleted from the analysis. In addition, OLS regression has measures of influence on the individual regression coefficients; DFBETAS, one for each regression coefficient for each case, assesses the number of standard deviations by which an individual case changes each regression coefficient. DFFITS and DFBETAS conceptually are computed by deleting each case in turn and measuring the change in the predicted score and regression coefficients, respectively. For GLiMs, once again, calculation of exact estimates of these statistics is not computationally possible in larger samples, so procedures that provide approximations to DFFITS, Cook's D, and DFBETAS have been offered (Dobson, 2002; Fox, 2008; Williams, 1987). The Cook's D and DFBETAS measures can be obtained in SAS using the *COOKSD* and *DFBETAS* keywords; Cook's D can be obtained in SPSS using the *COOK* keyword.

As noted previously, clear conventions for interpreting diagnostic statistics do not currently exist in GLiM, and different approximations used in their calculation may place the values in different metrics (scales). As indicated, there are no recommended cutoff scores for diagnostics in GLiMs as there are for diagnostics in OLS regression. The appropriate strategy for use of diagnostics in Poisson regression is to look for the one or at most several cases that are largest on some diagnostic measure (for DFBETAS, the largest absolute values are considered because a case may substantially raise or lower a regression coefficient). An excellent place to start is with cases that have the highest absolute value of DFBETAS for the theoretically most important variables. These are the cases that have the most effect on the regression coefficients for the tests of theory. If DFBETAS is extreme for a case on a theoretically important variable, the case may be working in one of two opposite ways; the case may be producing the significance of the coefficient of the variable or it may be obscuring a significant effect that is present but not seen when the case is included.

As with all use of diagnostics, if a case is removed from analysis, there must be justification for its removal (e.g., removal from the drinking study of a student with such severe alcohol addiction that the student is in an inpatient detoxification program during the measurement of alcohol consumption and thus yields a consumption score of zero). This advice for the use of diagnostics is no different from that for use of diagnostics in OLS regression—proceed with caution, one case at a time, and ask whether the case is changing the conclusion concerning important theoretical variables in the regression analysis.

Because there will be one diagnostic statistic of each type for every case in the sample, it quickly becomes overwhelming to examine them all. One straightforward, graphical way to interpret these statistics is to construct index plots in which case number is on the x-axis and the value of the diagnostic statistic is on the y-axis. In this way, cases that are very discrepant on one of the diagnostic statistics will "pop out" in the figure, allowing the researcher to easily identify them. Separate index plots are constructed for leverage, distance, global influence, and the measure of specific influence for each predictor. Cases with scores that are high in magnitude *relative* to other cases in the data set deserve careful scrutiny. Cohen et al. (2003, pp. 391–419) described in detail the application of these procedures to OLS regression and Fox (2008, pp. 412–415) described their extension to the GLiM.

Shortcomings of Standard Poisson Regression

There are situations in which observed count data do not meet all of the assumptions of the standard Poisson regression model. There are two situations that are most commonly encountered in practice. First, the individual counts may exhibit more variability than is expected from the Poisson model. Recall that the Poisson distribution has one parameter μ, which characterizes both the mean and the variance of the distribution. Thus, the Poisson model assumes that the conditional mean and variance are equal, a condition known as *equidispersion*. The situation in which the variance is larger than the mean is known as *overdispersion*. Second, there may be fewer or more values of "0" in the conditional distributions of the outcome variable than one would expect in a Poisson distribution with the specified mean and variance. Fewer zeros may occur when the sampling plan excludes members of the population with a score of 0, a problem known as *truncated zeros*. For example, if all participants who did not have an alcoholic drink during the observation night were excluded, this problem would exist. In contrast, the problem of *excess zeroes* occurs when the population being examined includes individuals who would *never* display the behavior. In our example of drinking behavior, excess zeroes would occur if there are nondrinkers in the sample—people who *never* drink for religious, health, or other reasons (e.g., underage drinking is illegal). We initially consider the problem of overdispersion and alternative models that address it. Later, we briefly consider models for truncated zeros and excess zeros.

Overdispersion

Although *equidispersion* is assumed in Poisson regression, actual data can be overdispersed, that is, the conditional variance of the residuals may be larger than the conditional mean (predicted value). If overdispersion is not accounted for, estimates of the standard errors will be too small, test statistics for the parameter estimates will be too large, significance will be overestimated, and confidence limits will be too small. The issue of overdispersion is not present in OLS regression because the normal distribution has two parameters, one defining the mean and one defining the variance or dispersion of the distribution.

Overdispersion occurs for two primary reasons in cross-sectional data.[9] First, there may be individual differences in responses that are not accounted for by the regression model. This problem commonly occurs if an important predictor is omitted from the model. Consider the example described previously in which sensation seeking and gender were used to predict the number of alcoholic drinks consumed per day. When gender is omitted from the model (as it was in our initial analysis) and sensation seeking alone is used to predict number of drinks, the variance in the outcome that would have been explained by gender is considered to be unexplained heterogeneity.

Second, each count that occurs for an individual may not be an independent event as is assumed by the Poisson distribution. This situation is known as *contagion* or *state dependence*. For example, the set of alcoholic drinks consumed by one person are not likely to be independent occurrences. The probability that an individual may consume a first drink in an evening is unlikely to be equal to the probability he or she will consume a second drink in an evening having once consumed a first; the second drink is therefore not independent of the first. There are several options for data that have an overdispersed outcome variable due to unmeasured heterogeneity. Standard Poisson regression may not be appropriate for data in which state dependence is strongly suspected.

Alternatives to Standard Poisson Regression

Overdispersed Poisson Regression Models

The simplest adjustment for overdispersion that can be made to the Poisson regression model is the overdispersed Poisson model (Gardner et al., 1995; Land, McCall & Nagin, 1996; Long, 1997). This model includes a second parameter that is used in the estimation of the conditional variance known as the *overdispersion scaling parameter*, ϕ. The model estimated with this correction now essentially assumes an error distribution that is Poisson with mean μ and variance $\phi\mu$. The scaling parameter ϕ will be greater than 1 if overdispersion is present in the data; ϕ will be equal to 1 if there is equidispersion, and the resulting model is equivalent to the standard Poisson regression model. Finally, ϕ will be less than 1 if the data are underdispersed. Underdispersion is rare in psychological data but may theoretically occur. The amount of dispersion in the model is typically determined by the Pearson chi-square goodness-of-fit statistic[10] (McCullagh & Nelder, 1989), which is a measure of the overall fit of the model provided by the computer output. The calculation of the scaling parameter is given by

$$\phi = \frac{\chi^2_{\text{Pearson}}}{\text{df}}. \tag{14}$$

The overdispersed model allows the conditional variances to be larger than their corresponding conditional means so that the standard errors (which are based on the conditional variances) will be larger than the standard errors in the standard Poisson model by a factor of $\sqrt{\phi}$. Interpretation of coefficients for the overdispersed Poisson model is identical to that of the standard Poisson model. The deviance for this model is also adjusted by the scaling factor; the deviance for the overdispersed Poisson model is equal to the deviance for the standard Poisson model divided by ϕ. The smaller deviance of this model indicates better fit.

Following up the preceding example, an overdispersed Poisson model was used to predict number of alcoholic drinks consumed from the single predictor, sensation seeking. The column labeled "(3)" of Table 1 shows the regression coefficients for the overdispersed Poisson model. The square root of the overdispersion parameter is 1.6873. The value of ϕ is 2.8470, which is substantially larger than 1.00 for the standard Poisson model that assumes equidispersion: The data clearly exhibit overdispersion. Note that the regression coefficients for this model are identical to those of the standard Poisson model, presented in Table 1, column labeled "(2)." However, the standard errors are all larger; each standard error from the Poisson regression in (2) is multiplied by the square root of the overdispersion parameter, 1.6873. For example, the standard error of sensation seeking in the overdispersed Poisson model is equal to 0.0397×1.6873. The regression coefficients and predicted values are identical for these two models, so interpretation of the coefficients is identical as well. A 1-unit change in sensation seeking results in an $e^{0.2315} = 1.26$ times change in the predicted number of alcoholic drinks consumed.

The column labeled "(3)" of Table 2 shows the regression coefficients for the overdispersed Poisson model with both gender and sensation seeking as predictors of number of alcoholic drinks consumed. Again the scale parameter ϕ highlights overdispersion in the data. The square root of the scaling parameter is 1.5169, so ϕ is equal to 2.3010. For all models, gender is coded such that females have a value of 0 on the predictor and males have a value of 1. The intercept is -0.7888, so the exponentiation of it, $e^{-0.7888} = 0.45$, is the predicted number of drinks for a female (gender = 0) who has a score of 0 on the sensation-seeking scale (again, not a possible value). The regression coefficient for sensation seeking is 0.2608, so $e^{0.2608} = 1.30$ is the predicted multiplicative effect of a 1-unit change in sensation seeking on number of drinks consumed, holding gender constant. The regression coefficient for gender is 0.8395. Because gender is dummy coded, a 1-unit change (i.e., from 0 to 1) corresponds to the difference between females (coded 0) and males (coded 1). The exponentiation of the regression coefficient, $e^{0.8395} = 2.32$, is the multiplicative difference in average number of drinks consumed for males versus females, holding sensation seeking constant; males consume, on average, 2.32 *times* more alcoholic drinks on a Saturday night than females.

Assessing whether the second predictor, gender, significantly adds to prediction in the overdispersed Poisson model is not possible here. The model with only sensation seeking as a predictor is *not nested* within the model with both predictors because the scaling parameter ϕ is different between the two models. Likewise, the pseudo-R^2 for the overdispersed Poisson model cannot be calculated because the scaling parameter for the two-predictor model is different from that of the intercept-only model. A nested model test to determine whether the overdispersed Poisson model fits better than the standard Poisson model

[9] Overdispersion also commonly occurs in longitudinal panel studies that result in clustering of the data. We do not discuss alternative analysis models for longitudinal designs such as generalized estimating equation models in this article (see Zeger, Liang, & Albert, 1988).

[10] Some authors (e.g., Allison, 1999) have suggested that another test statistic, the *deviance* chi-square goodness-of-fit statistic, can also be used to determine dispersion of the model. The Pearson chi-square and the deviance chi-square are typically very close in value.

(i.e., a test of overdispersion) is discussed later in the "Tests of Overdispersion" section.

Negative Binomial Regression Models

A second common method for accounting for overdispersion is the negative binomial model (Gardner et al., 1995; Hilbe, 2007; Land et al., 1996; Long, 1997). One shortcoming of the Poisson regression model described previously is that it does not contain an error (disturbance) term that fully parallels the error term found in an OLS regression equation. The standard Poisson model does not allow for heterogeneity among individuals. Often there is additional heterogeneity between individuals that is not accounted for by the predictors in the model and the Poisson error function alone, which results in overdispersion. The negative binomial model accounts for overdispersion by assuming that there will be unexplained variability among individuals who have the same predicted value. This additional unexplained variability between individuals leads to larger variance (than expected by the Poisson distribution) in the overall outcome distribution but has no effect on the mean. This additional variability is conceptually similar to the inclusion of an error term in normal linear regression.

To illustrate, consider the alcohol study example presented previously in which the number of drinks consumed per day is predicted by both sensation seeking and gender. The standard Poisson model assumes that the outcomes for all individuals with the same values on the predictors are samples from a *single* Poisson distribution with a given mean. That is, the subset of women with a sensation-seeking score of 5 are treated as being alike and modeled by a Poisson distribution with the same mean parameter. The negative binomial model, however, allows the observations of individuals with the same values on the predictors to be modeled by Poisson distributions with different mean parameters. That is, one woman with a sensation-seeking score of 5 may be modeled with a Poisson distribution with a mean of μ_1, whereas another woman with a sensation-seeking score of 5 is modeled with a Poisson distribution with a mean of μ_2.

Note that the data are still modeled using Poisson distributions but that each individual may be represented by a Poisson distribution with a different mean parameter. The variation in individual mean parameters for individuals with the same values on the predictors must be assumed to follow a probability distribution. The negative binomial model uses another standard (although less familiar) probability distribution, the gamma distribution (Freund & Walpole, 1980, pp. 196–197), to represent the distribution of means. In the negative binomial model, the error function is a mixture of two different probability distributions, the Poisson and gamma distributions.

The conditional mean of the outcome, given the values of the predictors, is identical for the standard Poisson model and the negative binomial model. In contrast, the conditional variance of the outcome will be larger in the negative binomial model than in the standard Poisson model. The variance for the negative binomial model is given by $\mu + \alpha\mu_2$ rather than μ as in Poisson regression.[11] The α parameter represents overdispersion in the negative binomial model. If $\alpha = 0$, there is no overdispersion, and the negative binomial model reduces to standard Poisson. An α parameter greater than zero indicates that overdispersion is present; larger values indicate more overdispersion. Interpretation of regression coefficients for the negative binomial model is identical to that for the standard Poisson model. The unexplained heterogeneity (between-individual variation) underlying the overdispersion is partialed out of the effects.

Consider again the alcohol consumption example. We now estimate a negative binomial model with sensation seeking predicting number of alcoholic drinks consumed. Unlike the previous models, the negative binomial model takes into account the fact that there may be unexplained heterogeneity in the outcome (e.g., due to the omission of gender from the model). The column labeled "(4)" in Table 1 shows the regression coefficients for the negative binomial model. The exponentiation of the regression coefficient for sensation seeking, $e^{0.2205} = 1.25$, is the multiplicative effect of a 1-unit change in sensation seeking on number of alcoholic drinks consumed, allowing for heterogeneity between individuals. A person with a sensation-seeking score of 5 is expected to consume, on average, 1.25 *times* as many drinks as a person with a sensation-seeking score of 4. The estimate of α for this model is 0.7179, which is greater than 0, indicating that there is overdispersion in the data.

The column labeled "(4)" in Table 2 shows the regression coefficients for the negative binomial model with both gender and sensation seeking as predictors of the number of alcoholic drinks consumed. Recall that gender is dummy coded and that females are coded 0. The exponentiation of the intercept, $e^{-0.6436} = 0.53$, is the predicted number of drinks consumed by a female with a score of 0 on the sensation-seeking measure, not a meaningful value in this example. The regression coefficient for sensation seeking is 0.2351, so $e^{0.2351} = 1.27$ is the predicted multiplicative effect of a 1-unit change in sensation seeking on number of drinks consumed, holding gender constant. The regression coefficient for gender is 0.8224. The exponentiation of the regression coefficient, $e^{0.8224} = 2.28$, is the multiplicative difference in average number of drinks consumed for males versus females, holding sensation seeking constant; males consume, on average, 2.28 *times* more alcoholic drinks than females. The estimate of α for this model is 0.5105, which is greater than 0, indicating that there is overdispersion in the data.

Similar to the case of overdispersed Poisson regression, one cannot assess the improvement of the model with two predictors over the model with one predictor because the estimated parameter α is different for the two models, so they are not actually nested models. Likewise, a pseudo-R^2 such as was calculated for Poisson regression cannot be calculated for negative binomial models. We discuss a nested model test to determine whether the negative binomial model fits better than the standard Poisson model (i.e., a test of overdispersion) in the next section.

Comparing Overdispersed Models

Tests of overdispersion. The likelihood ratio test (Chernoff, 1954) or the Score test (also known as the Lagrange multiplier [LM] test; Cook & Weisberg, 1983) may be used to assess whether significant overdispersion is present in the data. These two tests are asymptotically equivalent, meaning that they will produce the same result with very large sample sizes. The

[11] This is the negative binomial 2 or NB2 model described by McCullagh and Nelder (1989). The negative binomial 1 or NB1, which has a slightly different variance function, is also discussed by McCullagh and Nelder.

likelihood ratio test is a nested model test that compares the deviance of a model in which the scaling (ϕ) or overdispersion (α) parameter has been fixed to a specific value to the deviance of a model in which the scaling parameter is estimated. The scaling parameter ϕ can be 1 or estimated; the overdispensation parameter α can be 0 or estimated. Comparing the difference in deviances to a chi-square distribution with 1 *df* will determine whether overdispersion is present. If the test is significant, the model with a freely estimated scaling parameter fits better than the model in which the scaling parameter is fixed. Of the three models discussed here, two nested model comparisons of this type are possible—standard Poisson versus overdispersed Poisson and standard Poisson versus negative binomial (Long, 1997, p. 247). The overdispersed Poisson regression model is not nested within the negative binomial model.

For the two-predictor example, the nested model test of standard Poisson versus overdispersed Poisson regression is given by $\chi^2(1, N = 400) = 959.4582 - 416.9831 = 542.4751$, which is significant at $p < .0001$. The overdispersed Poisson model fits better than the standard Poisson model. The nested model test of standard Poisson versus negative binomial is given by $\chi^2(1, N = 400) = 959.4582 - 459.7743 = 499.6839$, $p < .0001$. The negative binomial model fits better than the standard Poisson model. Both of these tests indicate that significant overdispersion is present in the data. A test of whether the negative binomial model fits better than the overdispersed Poisson model is not possible because these two models are not nested.

The LM or Score test uses the likelihood function to determine whether overdispersion is present. This test can be obtained from SAS using the NOSCALE option (see Appendix A). The appropriate use of this test is within the negative binomial model; the test assesses whether the overdispersion parameter in the negative binomial model (α) needs to be freely estimated or can be fixed to 0 (i.e., the standard Poisson model). For the two-predictor example, the LM test statistic is $\chi^2(1, N = 400) = 70.1861$, $p < .0001$. The result of this test indicates that the negative binomial model provides a significantly better fit than the standard Poisson model and that the estimate of α is significantly different from 0.

Comparing Non-Nested Models—Akaike (1973) Information Criterion (AIC) and Bayesian Information Criterion (BIC; Raftery, 1995)

Although no test of the difference in fit of two non-nested models is possible, two closely related measures of the fit of alternative non-nested models have been proposed that can guide model selection. The AIC is composed of two additive parts.[12] The first part is a function of the log likelihood of the proposed model evaluated at the estimated parameter value, here $\hat{\mu}$, that becomes smaller as the model fits better. The second part is a function of the number of parameters (regression coefficients, dispersion parameter) that are being estimated. The model with the smallest AIC value is selected. Because models with a larger number of parameters fit better, this second part penalizes more complex models that use more parameters to achieve the same fit, as indicated by the log likelihood. The BIC follows a similar logic. The first part is identical to the AIC. The second part reflects both the number of estimated parameters and the sample size. In essence, the BIC penalizes complex models based on large sample sizes less than does the AIC. Sample size plays a role in model selection, with more complex models being preferred for large sample sizes, whereas simpler models are typically preferred for smaller sample sizes (MacCallum, 2003). Although the AIC and BIC will typically lead to selection of the same model, the BIC does lead to more accurate model selection than AIC in some cases.

Models That Address Problems With Zeros

Truncated Zeros

One useful expansion of Poisson models is to data in which no zeroes are possible because of the sampling plan. For example, if data are collected from a medical facility, the sample for a study of medical visits will only include individuals who have visited the medical facility at least once. No one in the sample can have a value of 0 for number of visits. Long (1997) described several models that can account for these "missing" zeroes. The modeling approach is to modify the Poisson or negative binomial model to model the probability of a given count on the outcome variable *given that the count is greater than 0* during the observation period.

Structural Zeros: Zero Inflated Models

As mentioned earlier, count outcomes are typically right skewed, so there are many low values of the dependent variable. However, some outcomes may display even more low values than expected by the Poisson distribution. For example, in the alcohol consumption example, the sample may include some never drinkers. These individuals will *always* respond that they consumed zero drinks. In this case, the zeroes that are observed in the sample can be thought of as coming from two groups in the population: never drinkers who produce *structural zeroes* that must always occur and drinkers (including occasional drinkers) who produce zeroes with some probability on the night of the study when the observations are collected.

The best solution to this problem is to attempt to identify the people in the two populations by anticipating the problem in the design of the study. Then those who produce structural zeroes can be identified and eliminated from the data set, whereas those who could *potentially* produce nonzero counts are retained, even if they happen to have an observation of zero in the study. In our illustration, asking questions about lifetime alcohol use and reasons for not drinking would help identify participants who never drink and can be eliminated. This strategy of eliminating the nondrinkers assumes that the population of interest is the people who potentially would consume alcohol.

If information identifying those individuals who are structural zeros is not available or the research interest is in the full population including non-drinkers, zero inflated Poisson models or zero inflated negative binomial models can be used (Greene, 1994; Hall & Zhengang, 2004; Long, 1997). Conceptually, these models have two parts. In the first part, available predictors

[12] The equation for the $AIC = -2 \ln f(y|\hat{\theta}) + 2k$, where $y = (y_1, \ldots, y_n)$ is a random sample of size n, $\hat{\theta}$ is a vector of the maximum likelihood parameter estimates, $\ln f(y|\hat{\theta})$ is the log-likelihood of the current model, and k is the number of estimated parameters in the model. As shown in the equation, the first term is a measure of lack of fit; the second term is a function that penalizes models with a greater number of estimated parameters. The equation for the $BIC = -2 \ln f(y|\hat{\theta}) + k \ln(n)$, where n is sample size. The relative magnitude of the penalty for having more estimated parameters is smaller for larger sample sizes.

are used to estimate the probability that each individual with an observed count of 0 is in each of two latent groups: those that would never perform the behavior (structural zeros) and those who might potentially perform the behavior during another observation session. Logistic regression is often used to estimate these probabilities. Obviously, collection of good background information (e.g., prior drinking history) greatly reduces the uncertainty in the prediction of the latent group to which each individual with a count of 0 should be assigned. The second part of the model is to estimate the Poisson or negative binomial model for the part of the data that does not contain structural zeros. Current approaches estimate both parts simultaneously, adjusting the mean and the variance of the Poisson or negative binomial model in an attempt to eliminate the likelihood that some of the zero observations are in fact structural zeros. Both the zero inflated Poisson and negative binomial models may currently be estimated in the software packages STATA and Mplus (with zero inflated negative binomial becoming available in Version 5.1). The Web site http://www.ats.ucla.edu/stat/MPlus/output/zeroinflatedpoissonreg.htm presents sample computer syntax. At this time, no routine exists in SPSS, and only the experimental PROC COUNTREG exists in SAS for zero inflated count models.[13] COUNTREG does not currently produce many of the relevant fit measures that are necessarily to fully evaluate the model.

Models for Rates: Variable Length Observation Periods

Regression models of the type discussed here can also be expanded in other ways. For example, the Poisson regression model assumes observation for all individuals occurs in the same length time period. The worked example in this article used number of alcoholic drinks consumed *in one night* as the outcome. Other examples of outcomes with fixed time periods are number of aggressive acts committed by a child *during a 1-hr play period* and number of cigarettes smoked *per day*. Models exist that can extend Poisson type models to variable time periods. For example, Allison (1999) suggested that including the natural log of the measurement interval as a predictor with regression coefficient equal to 1 allows incorporation of variable time periods and maintains the Poisson error structure of the data. The resulting model is

$$\ln(\hat{\mu}) = \ln(time) + b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p. \quad (15)$$

This model is algebraically the same as a model of rates (i.e., events per unit time):

$$\ln\left(\frac{\hat{\mu}}{time}\right) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p, \quad (16)$$

but maintains the correct error structure. The Poisson distribution only applies to discrete outcomes and so can only be used with Equation 15. In both SPSS and SAS, the natural log of time can be included in this way by including OFFSET = *lntime* in the model statement options (where *lntime* is a variable created by the analyst to be the natural logarithm of the duration of the measurement interval). Time-varying models are related to survival models, which seek to model the time to failure (e.g., time to relapse following treatment for alcoholism) or success (e.g., time to reemployment following a job seeking skills program).

[13] In SAS, zero-inflated models can be estimated using a far more complicated procedure, PROC NLMIXED.

Discussion

The purpose of this article was to introduce the reader to a modern method of analyzing count outcomes: Poisson regression. We began by discussing the problems with using OLS regression to analyze count outcomes and introduced the GLiM, the Poisson distribution, and standard Poisson regression. An example of a Poisson regression analysis that is relevant to personality assessment researchers, predicting the number of alcoholic drinks consumed based on measures of sensation seeking and gender, illustrated the interpretation of the model and evaluation of fit. We discussed the limitations of Poisson regression related to the assumption of equidispersion. Two models that relax the equidispersion constraint, the overdispersed Poisson model and the negative binomial model, were introduced. We expanded the example to include these models. We then introduced models that are useful when there are no zeros or too many zeros. Zero truncated models are useful for data sets in which zeros are systematically excluded by the data collection plan; zero inflated models for studies that include two groups, one of which would always produce a 0 count of the behavior (e.g., nondrinkers). Finally, we briefly discussed models for rates in which the length of observation periods varies across participants.

Other less powerful and less appropriate methods of analyzing count outcomes exist. Prior to the development of Poisson regression, researchers were encouraged to transform their count outcomes to make them more appropriate for use with OLS regression. Common transformations included the square root of the outcome or the natural log of the outcome. There are several reasons why these methods, although simple to implement, are less desirable than optimal Poisson regression models. Count outcomes are typically right skewed, have many low values, and have a narrow range of values. Transforming the outcome does not adequately handle the excess of small values and will typically have little effect on the outcome when the range is very narrow. More importantly, transformation of the outcome cannot completely eliminate heteroscedasticity; Poisson regression allows selection of the correct error structure, which properly accommodates the heteroscedasticity.

In conclusion, we have provided an introduction to regression models that are useful in analyzing count data. These models are attractive alternatives to analyses using standard OLS regression, particularly when the mean of the outcome variable is not large (less than 10 as a rule of thumb). These models provide tests of hypotheses that maximize the power of the statistical test consistent with maintaining the Type 1 error rate at the designated nominal level (e.g., actual $\alpha \approx .05$ when nominal $\alpha = .05$). As illustrated in this article, the interpretation of the results and tests of competing nested models are straightforward. Examination of the plots of the residuals can provide information about the adequacy of the model. Standard computer packages such as SPSS and SAS are increasingly incorporating easy-to-use routines that permit researchers to implement these analyses in their own research. We encourage researchers to consider using

these more optimal approaches whenever count data represent important outcomes in their research.

ACKNOWLEDGMENTS

We thank Howard Tennen and Stephen Armeli for providing us with a portion of the data from DeHart, Tennen, Armeli, Todd, & Affleck (2008). This data set provided the basis on which we developed our simulated data set.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Allison, P. D. (1999). *Logistic regression using SAS: Theory and application*. Cary, NC: SAS Institute Inc.

Armeli, S., Mohr, C., Todd, M., Maltby, N., Tennen, H., Carney, M. A., et al. (2005). Daily evaluation of anticipated outcomes from alcohol use among college students. *Journal of Social and Clinical Psychology, 24,* 767–792.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. New York: Cambridge University Press.

Cameron, A. C., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics, 77,* 329–342.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics, 25*, 573–578.

Chin, H. C., & Quddus, M. A. (2003). Modeling count data with excess zeroes: An empirical application to traffic accidents. *Sociological Methods and Research, 32,* 90–116.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika, 70*, 1–10.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO–PI–R professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). London: Chapman & Hall.

DeHart, T., Tennen, H., Armeli, S., Todd, M., & Affleck, G. (2008). Drinking to regulate romantic relationship interactions: The moderating role of self-esteem. *Journal of Experimental Social Psychology, 44*, 537–538.

Dobson, A. J (2002). *An introduction to generalized linear models* (2nd ed.). New York: Chapman & Hall/CRC.

Enders, C. K. (2005). Maximum likelihood estimation. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1164–1170). Chichester, England: Wiley.

Fahrmier, L., & Tutz, G. (2001). *Multivariate statistical modeling based on generalized linear models* (2nd ed.). New York: Springer-Verlag.

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Freund, J. E., & Walpole, R. E. (1980). *Mathematical statistics* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin, 118,* 392–404.

Greene, W. H. (1994). *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models* [Working paper No. 94–10]. New York: Stern School of Business, New York University, Department of Economics.

Hall, D. B., & Zhengang, Z. (2004). Marginal models for zero inflated clustered data. *Statistical Modeling, 4,* 161–180.

Hilbe, J. M. (2007). *Negative binomial regression*. New York: Cambridge.

Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models: With empirical applications to criminal careers data. *Sociological Methods and Research, 24,* 387–442.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38,* 113–139.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691–692.

National Institute of Mental Health. (1997). *Computerized diagnostic interview schedule for children, IV (NIMH C-DISC IV)*. New York: Author, Columbia University.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A* (General), 135, 370–384.

Pierce, D. A., & Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association, 81,* 977–986.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–196.

Shanahan, C., Lincoln, A., Horton, N., Saitz, R., Winter, M., & Samet, J. (2005). Relationship of depressive symptoms and mental health functioning to repeat detoxification. *Journal of Substance Abuse Treatment, 29,* 117–123.

Siddiqui, O., Mott, J., Anderson, T., & Flay, B. (1999). The application of Poisson random-effects regression models to the analyses of adolescents' current level of smoking. *Preventive Medicine, 29,* 92–101.

Wainer, H. (2000). The centercept: An estimable and meaningful regression parameter. *Psychological Science, 11,* 434–436.

Walters, G. D. (2007). Predicting institutional adjustment with the Lifestyle Criminality Screening Form and the Antisocial Features and Aggression Scales of the PAI. *Journal of Personality Assessment, 88,* 99–105.

West, S. G., Aiken, L. S., & Krull, J. L. (1996) Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64,* 1–48.

Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics, 36,* 181–191.

Worrall, J. L. (2002). If you build it, they will come: Consequences of improved citizen complaint review procedures. *Crime & Delinquency, 48,* 355–379.

Zeger, S. L., Liang, K.-L., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics, 44,* 1049–1066.

APPENDIX A

SAS Syntax

```
/* Syntax for two-predictor Poisson regression */
proc genmod data = one;
model y = gender sensation/dist = poisson link = log;
output out = plots_poi predicted = pred_count resdev =
dev_resid
stdresdev = stud_dev_resid leverage = hat_diagonal
cooksd =_all_ dfbetas =_all_;
run;
/* Syntax for Figure 4 */
symbol1 color = black value = x interpol = sm50s;
proc gplot data = plots_poi;
plot dev_resid * pred_count;
title2 "Deviance residuals vs predicted count for two-predictor
Poisson model";
run;
/* Syntax for index plot of leverage */
symbol1 color = black value = x;
proc gplot data = plots_poi;
plot hat_diagonal * case;
```

```
title2 “Index plot of leverage statistics (Hat diagonal) for all cases”;
run;
/* Syntax for two-predictor overdispersed Poisson regression */
proc genmod data = one;
model y = gender sensation/dist = poisson link = log scale = d;
output out = plots_odpoi predicted = pred_count resdev = dev_resid;
run;
/* Syntax for two-predictor negative binomial regression */
proc genmod data = one;
model y = gender sensation/dist = nb link = log;
output out = plots_nb predicted = pred_count resdev = dev_resid;
run;
/* Syntax for two-predictor negative binomial regression, testing for overdispersion */
proc genmod data = one;
model y = gender sensation/dist = nb link = log noscale;
output out = plots_nb_test predicted = pred_count resdev = dev_resid;
run;
```

APPENDIX B

SPSS Syntax

```
* Syntax for two-predictor Poisson regression.
GENLIN
y
WITH gender sensation
/MODEL gender sensation
INTERCEPT=YES
DISTRIBUTION=POISSON
LINK=LOG
/CRITERIA LIKELIHOOD=KERNEL
/SAVE MEANPRED DEVIANCERESID
STDDEVIANCERESID LEVER AGE COOK.
* Syntax for Figure 4.
GRAPH
/SCATTERPLOT(BIVAR)=MEANPRED WITH
DEVIANCERESID
/MISSING=LISTWISE.
* Syntax for index plot of leverage.
GRAPH
/SCATTERPLOT(BIVAR)=CASE WITH LEVERAGE
/MISSING=LISTWISE.
* Syntax for two-predictor overdispersed Poisson regression.
GENLIN
y
WITH gender sensat
/MODEL gender sensat
INTERCEPT=YES
DISTRIBUTION=POISSON
LINK=LOG
/CRITERIA SCALE=DEVIANCE LIKELIHOOD=KERNEL
/SAVE MEANPRED DEVIANCERESID.
```