Regression Analysis of Grouped Counts and Frequencies

Using the Generalized Linear Model

by

Stefany Coxe

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2011 by the
Graduate Supervisory Committee:

Leona Aiken, Co-Chair
Stephen West, Co-Chair
Mark Reiser
David Mackinnon

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Coarsely grouped counts or frequencies are commonly used in the behavioral sciences. Grouped count and grouped frequency (GCGF) that are used as outcome variables often violate the assumptions of linear regression as well as models designed for categorical outcomes; there is no analytic model that is designed specifically to accommodate GCGF outcomes. The purpose of this dissertation was to compare the statistical performance of four regression models (linear regression, Poisson regression, ordinal logistic regression, and beta regression) that can be used when the outcome is a GCGF variable.

A simulation study was used to determine the power, type I error, and confidence interval (CI) coverage rates for these models under different conditions. Mean structure, variance structure, effect size, continuous or binary predictor, and sample size were included in the factorial design. Mean structures reflected either a linear relationship or an exponential relationship between the predictor and the outcome. Variance structures reflected homoscedastic (as in linear regression), heteroscedastic (monotonically increasing) or heteroscedastic (increasing then decreasing) variance. Small to medium, large, and very large effect sizes were examined. Sample sizes were 100, 200, 500, and 1000.

Results of the simulation study showed that ordinal logistic regression produced type I error, statistical power, and CI coverage rates that were consistently within acceptable limits. Linear regression produced

type I error and statistical power that were within acceptable limits, but CI coverage was too low for several conditions important to the analysis of counts and frequencies. Poisson regression and beta regression displayed inflated type I error, low statistical power, and low CI coverage rates for nearly all conditions. All models produced unbiased estimates of the regression coefficient.

Based on the statistical performance of the four models, ordinal logistic regression seems to be the preferred method for analyzing GCGF outcomes. Linear regression also performed well, but CI coverage was too low for conditions with an exponential mean structure and/or heteroscedastic variance. Some aspects of model prediction, such as model fit, were not assessed here; more research is necessary to determine which statistical model best captures the unique properties of GCGF outcomes.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure | Page

Chapter 1

Introduction

How many days per week do you exercise for 30 minutes or more? Never? Once or twice per week? About every other day? Most days? Every day? Questions of this type, with their accompanying response scales, are common in many areas of the social sciences. However, problems arise when this type of variable is used as an outcome in a regression model. Using a grouped count or grouped frequency (GCGF) variable such as the one presented above as an outcome leads to violations of the assumptions of linear regression. The assumptions of models that were designed to be used for categorical outcomes, such as Poisson regression and ordinal logistic regression, are also violated. Despite the regularity with which GCGF variables are encountered in the social sciences, there is currently no single analytic model that is designed to accommodate their specific, unique properties.

The purpose of this dissertation is to compare the statistical performance of four regression models that can be used when the outcome variable is characterized as a GCGF; specifically, the statistical power, type I error, and confidence interval (CI) coverage for these models were examined. A simulation study was used to determine the empirical power, empirical type I error rates, and empirical CI coverage rates for these regression models under several sample size and effect size conditions, as well as different outcome mean and variance structures.

Grouped counts and grouped frequencies are widely used in psychology, particularly in social and clinical psychology. The Monitoring the Future scales (Johnston, Bachman, & O'Malley, 2003), the Child Report of Parent Behavior Inventory (CRPBI; Schaefer, 1965), and the Acculturation Rating Scale for Mexican-American (ARSMA; Cuellar, Arnold, & Maldonado, 1995) are examples of scales used in psychology that include GCGF variables. GCGF variables are ordered, categorical, and typically have a specific potential range of numerical values associated with each response option. For example, the item presented is scored on a 0 to 4 scale and has the response options of 0 (Never), 1 (Once or twice per week – 1 to 2 times per week), 2 (About every other day – 3 to 4 times per week), 3 (Most days – 5 to 6 times per week), and 4 (Every day – 7 times per week). There is relatively little methodological literature on items of this type (but see Nagin (1997) for examples).

There are several different statistical models available to analyze a GCGF outcome. Each regression model has strengths and weaknesses when applied to the analysis of GCGF outcomes, so it is unclear which method should be used. The simplest and most commonly used method is linear regression. Linear regression (Cohen, Cohen, West, & Aiken, 2003; Neter, Kutner, Nachtsheim, & Wasserman, 1996) is familiar and easy to interpret when all of its assumptions are met. These assumptions require that the outcome variable be continuous and conditionally normally distributed; however, GCGF variables are non-continuous and likely to be non-normally distributed.

Another method of analysis that may be used for GCGF outcomes is ordinal logistic regression (Agresti, 2002; Allison, 1999; Fahrmeir & Tutz, 2001; Hosmer & Lemeshow, 2000). The outcome options are treated as ordered (but not necessarily equally wide or equally spaced) categories. For ordinal logistic regression, the predicted outcome is the probability of being in a specific category or higher relative to being in a lower category. Prediction can also be thought of as the probability of crossing the threshold from one category to the next higher category. One issue with this method is that, like linear regression, it assumes that predictors have a constant effect on the probability of crossing a threshold, regardless of which pair of categories is being considered. For the example above, that would mean that a predictor has the same effect on the transition from zero (0) days of exercise to 1 – 2 days of exercise per week as it does on the transition from 5 – 6 days of exercise to 7 days. This assumption may not always be appropriate.

Poisson regression (Cameron & Trivedi, 1998; Gardner, Mulvey, & Shaw, 1995; Long, 1997) is typically used for count outcomes, that is, when the outcome takes on only discrete, non-negative values. For count outcomes, Poisson regression is a superior method to linear regression in terms of statistical power and type I error, especially when the mean of the outcome is small. It is unclear whether this advantage persists when the outcome counts are grouped into potentially unequally spaced categories. Poisson regression assumes that the variance of the outcome increases with the mean of the outcome, specifically, that the outcome variance equals the outcome mean. The grouping of GCGF variables leads to increased variance within each category (relative to the ungrouped counts or frequencies) because multiple values of a variable are placed into a single category in GCGF, potentially violating this assumption of the mean-variance relationship.

Beta regression (Kieschnick & McCullough, 2003; Paolino, 2001; Smithson & Verkuilen, 2008) is a less-commonly used method for outcomes that have both upper and lower bounds; it is often used for proportion or percentage outcomes. One advantage of beta regression over the other methods described is that it is extremely flexible regarding the error structure of the outcome. The variance of the outcome can be heteroscedastic and is modeled *separately* from the mean structure, offering an advantage over the homoscedasticity assumption of linear regression and the stringent variance structure of Poisson regression. A weakness of beta regression for GCGF variables is that, like linear regression, the model actually assumes a continuous outcome.

Given that there are several models available for GCGF outcomes and the fact that none of them are perfectly matched to the specific properties of GCGF outcomes, it is desirable to assess the statistical performance of these different models. It is also likely that the properties of the GCGF may vary such that a certain model may be preferable in certain circumstances. Factors that are expected to affect the performance of the models include the mean structure of the relationship between the predictor and the outcome, the conditional variance structure of the outcome, the effect size, and sample size.

Chapter 2 outlines the assumptions of linear regression that are relevant to the outcome variables, with particular attention paid to how categorical outcome variables (such as GCGF outcomes) can violate these assumptions.

Chapter 3 outlines the three other regression models that are proposed for use with GCGF outcomes: ordinal logistic regression, Poisson regression, and beta regression. These models are all members of the generalized linear model family; generalized linear models are often used when the outcome is categorical or otherwise does not meet the assumptions of linear regression. The assumptions of each model and how GCGF outcomes may meet these assumptions are described.

Chapter 4 covers the measurement properties of GCGF outcomes, particularly with respect to the types of statistical analyses that can be performed. This chapter also describes an alternative approach to determining the statistical analysis to be performed, based on the degree of similarity between the assumptions of a statistical model and the properties of the outcome variable.

Chapter 5 describes the concepts of statistical power, type I error, and CI coverage. This chapter also describes the two commonly used tests of regression coefficients for which empirical power will be determined: the Wald test and the likelihood-ratio test.

Chapter 6 describes the details of the statistical simulation study that was used to generate data, analyze the data using the four regression models, and determine power, type I error, and coverage for each of the models. Chapter 7 presents the results of this simulation study. Chapter 8 discusses the results and implications of the simulation study.

Chapter 2

Linear Regression

Assumptions

Multiple regression analysis (Cohen et al., 2003; Neter et al., 1996) is a statistical system for relating a set of independent variables to a single dependent variable. Fixed effects linear regression using ordinary least squares estimation is the most common form of regression analysis. Multiple regression predicts a single continuous dependent variable as a linear function of any combination of continuous and/or categorical independent variables. Assumptions that are directly related to the predictors in multiple regression are minimal; we assume only that predictors are measured without error and that each predictor is fixed, that is, the values of each predictor are specifically chosen by the experimenter rather than sampled from all possible values of the predictor. However, there are additional assumptions of multiple regression that are related to the errors; these assumptions are much more critical.

Estimation of linear regression coefficients typically takes place using ordinary least-squares estimation. The linear regression model with $p + 1$ terms (including p predictors plus the intercept) and n subjects is of the form $\mathbf{Y} = \mathbf{XB} + \mathbf{e}$, where **Y** is the $n \times 1$ vector of observed outcome values, **B** is the $(p+1)\times 1$ vector of estimated regression coefficients, **X** is the $n \times p$ matrix of observed predictors, and **e** is the $n \times 1$ vector of unobserved errors. The Gauss-Markov Theorem (Neter et al., 1996) states that, in order for least-squares estimates to be the best linear unbiased estimates (BLUE) of the population regression coefficients, three assumptions about the errors must be met. First, the conditional expected value of the errors must be equal to zero. That is, for any value of the predictors X, the expected value of the errors is 0.

(1) $E(e_i \mid \mathbf{X}) = 0$

Second, the errors must have constant and finite conditional variance, σ^2. That is, for any value of the predictors *X*, the variance of the errors is σ^2.

(2) $Var(e_i \mid \mathbf{X}) = \sigma^2 < \infty$

This property of constant variance is known as homoscedasticity. Third, errors for individual cases must be uncorrelated:

(3) $Cov(e_i, e_j) = 0$, where $i \neq j$.

These three assumptions are necessary to ensure that the estimates of the regression coefficients are unbiased and have the smallest possible standard errors (i.e., they are BLUE).

In order to make valid statistical inferences about the regression coefficients, one final assumption must be made about the errors. Tests of statistical significance and the construction of confidence intervals for regression coefficients require an assumption to be made about the distribution of the errors. For linear regression, the errors are assumed to be normally distributed. Together with assumptions (1) and (2) above, this means that the errors are assumed to be conditionally normally distributed with a mean of zero and constant variance σ^2:

(4) $e_i \mid \mathbf{X} \sim N(0,\sigma^2)$

A consequence of this additional assumption of normally distributed errors is that assumption (3) above is replaced with the stronger assumption that individual errors (across cases or individuals) are independent (Neter et al., 1996).

Violations of Assumptions

Categorical variables (including GCGF variables) are common in many substantive areas, either variables that are naturally categorical or continuous variables that have been classified into two or more discrete categories. GCGF outcomes are an example of the latter kind of categorical outcome. Common types of categorical variables are binary variables, ordered or unordered categories, and counts. An example of a naturally categorical variable is biological gender; an individual can belong to only the male class or the female class. An example of a continuous variable that is categorized is SAT score. An individual's score on the SAT is a continuous variable, but colleges often determine a minimum SAT score for admission, such that students scoring below that minimum are not accepted. This leads to a categorical variable that indicates qualified or not qualified (based on the continuous SAT score).

Heteroscedasticity. When categorical variables serve as dependent variables, the assumptions of ordinary linear regression are typically violated. First, the errors of the linear regression model will be *heteroscedastic*; that is, the variance of the errors is not constant across all values of the predicted dependent variable. For example, the error variance of binary and count variables is dependent on the predicted score. The error variance of a binary variable, $\sigma^2 = \hat{\pi}(1 - \hat{\pi})$, is largest at a predicted value of $\hat{\pi} = 0.5$ and decreases as the predicted value approaches 0 or 1; the error variance of a count variable often increases with increases in the predicted value. A consequence of heteroscedasticity is biased standard errors. Conditional standard errors may be larger or smaller (depending on the situation) than those in the constant variance case; Gardner, Mulvey, and Shaw (1995) state that applying linear regression to count data typically results in standard errors that are too small. Incorrect standard errors result in biased Wald tests because *z*-tests and *t*-tests of parameter estimates involve dividing the parameter estimate by the standard error of the parameter estimate.

Non-normality. Second, the errors will not be normally distributed, attributable to the limited observed values that a discrete outcome variable may take on. For example, when the observed criterion is binary, only taking on values of 0 or 1, the error value for a predicted value $\hat{\pi}$ is also binary; the error for that predicted score can only take on values of $(1-\hat{\pi})$ or $(0-\hat{\pi})$. In this case, the errors are conditionally discrete. A discrete variable cannot be normally distributed, so the errors cannot be normally distributed. Non-normally distributed errors make the typical statistical tests and confidence intervals on the regression coefficients invalid because these tests are based on normal distribution theory.

Linearity

Ordinary linear regression assumes a model that is both linear in the parameters and linear in the variables (Cohen et al., 2003, p. 193-195). Linear in the parameters means that the predicted score is obtained by multiplying each predictor by its associated regression coefficient and then summing across all predictors. A relationship that is linear in the parameters is exemplified by the linear regression equation:

(5) $$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p .$$

Linear in the variables means that the relation between the predictor and the outcome is linear. In other words, a plot of the relation between the predictor *X* and the outcome is approximately a straight line. Linear regression can also accommodate some types of non-linear relations. Non-linear polynomial relations are allowed by including predictors raised to a power. A quadratic relation between the predictor *X* and the outcome can be incorporated into a linear regression by including X^2 as a predictor. If the relation between *X* and the outcome is quadratic, the relation between X^2 and the outcome will be linear, so the model will still be linear in the variables. When the relation is in fact quadratic, omitting this higher order term in a linear regression model results in model misspecification.

If the relationship between predictors and the outcome is non-linear and is not accommodated by powers of the predictors, estimates of the linear regression coefficients and the standard errors will be biased (Cohen et al., 2003, p. 118). In this case, linear regression is not the appropriate analytic approach. Non-linear relations between predictors and the outcome are common for discrete and categorical outcome variables. For example, consider predicting a binary outcome, the probability of purchasing a new car versus a used car as a function of household income. An increase in income of $20,000 will increase the likelihood of purchasing a new car a great deal for households with an income of $50,000, but probably has little effect on the likelihood of purchasing a new car for a household with an income of $500,000. If the relationship between the predictors and the dependent variable is not linear, the linear regression model will be misspecified for two reasons. First, the relation between the predictor and the outcome is non-linear, so the form of the relation is misspecified. Second, the linear regression model is inappropriate for binary outcomes, so the model itself is misspecified. While a non-linear relationship between the predictor and the outcome such as the one described above can in some cases be resolved by transforming the predictor (e.g., by taking the natural logarithm of the income predictor, see Cohen et al., 2003, Chapter 6), the combination of a non-linear relationship and the binary outcome leads to the conclusion that linear regression is not the appropriate choice for analysis.

For outcome variables with upper and/or lower bounds, another consequence of using a linear model when the relationships between the predictors and the outcome are non-linear is that predicted criterion scores may fall outside the range of the observed scores. This is a problem particular to bounded categorical variables, which are often undefined and not interpretable outside their observed limits. For example, when the outcome variable is binary, predicted scores are probabilities and can only range from 0 to 1. Predicted values that are less than 0 or greater than 1 cannot be interpreted as probabilities. For a model of count data, predicted values less than 0 are not interpretable because an event cannot occur a negative number of times. Count variables may also be bounded at both ends, for example, the number of days in a week in which an event occurs.

Chapter 3

Generalized Linear Models (GLiMs)

The generalized linear model (GLiM), developed by Nelder & Wedderburn (1972) and expanded by McCullagh & Nelder (1983), extends linear regression to a broader range of outcome variables. Models in the GLiM family can be used for a variety of categorical outcomes, including binary outcomes, ordered categories, and counts. For this reason, GLiMs are a reasonable solution to the problem of analysis of GCGF outcomes.

The GLiM introduces two major modifications to the linear regression framework. First, it allows transformations of the *predicted* outcome, accommodating a potentially non-linear relationship between the dependent variable and the predictors via a link function. Second, the GLiM allows error structures (i.e., conditional distributions of the outcome) in addition to the normal distribution error structure assumed by linear regression.

Three Components of a GLiM

There are three components to the generalized linear model – the random portion, the systematic portion, and the link function. The random portion of the model defines the error distribution of the outcome variable. The error distribution of the outcome variable refers to the conditional distribution of the outcome given the predictors. GLiM allows any discrete or continuous distribution in the exponential family; the most common include the normal, exponential, gamma, beta, binomial, multinomial, and Poisson distributions. Other distributions exist in the exponential family, but are more rarely used in GLiMs.

The systematic portion of the model defines the relation between η, which is some function of the expected value of Y, and the predictors in the model. This relationship is defined as linear in the variables, e.g., $\eta = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p$, so the regression coefficients can be interpreted identically to those in linear regression: a 1-unit change in X_1 results in a b_1 unit change in η, holding all other variables constant.

The link function relates the conditional mean of Y, also known as the expected value of Y, E($Y|X$), or μ, to the linear combination of predictors (previously stated as equal to η). The link function allows for non-linear relations between the predictors and the predicted outcome. Several link functions are possible, but each error distribution has a special link function known as its canonical link. The canonical link satisfies special properties of the model, makes estimation simpler, and is the most commonly used link function. For example, the natural log (ln) link function is the canonical link for a conditional Poisson distribution. The logit or log-odds is the canonical link for a conditional binomial distribution, resulting in logistic regression. The canonical link for the normal error distribution is identity (no transformation) resulting in linear regression. In this framework, linear regression becomes a special case of the GLiM. For the case of linear regression, the error distribution is a normal distribution and the link function is identity. A wide variety of generalized linear models are possible, depending on the proposed conditional distribution of the outcome variable.

Ordinal Logistic Regression

Ordinal logistic regression is an extension of binary logistic regression to 3 or more categorical outcomes. Binary logistic regression (Agresti, 2002; Fahrmeir & Tutz, 2001; Hosmer & Lemeshow, 2000) is a commonly used and appropriate analysis when the outcome variable is binary, meaning that the outcome takes on one of two mutually exclusive values, such as alive or dead, diseased or well, pass or fail. Binomial logistic regression is a GLiM with binomial distribution error structure and logit link function. The probability mass function for the binomial distribution,

(6) $$P(Y = y \mid n, \pi) = \frac{n!}{y!(n-y)!}\pi^{y}(1-\pi)^{n-y},$$

gives the probability of observing a given value, *y*, of variable *Y* which is distributed as a binomial distribution with parameters *n* and π. For this distribution, *n* represents the number of observations and π represents the probability of an individual observation being a case (i.e., belonging to a specifically chosen category of the outcome). The mean of this distribution is $n\pi$ and the variance is $n\pi(1-\pi)$.

Note that unlike the normal distribution, which has independent mean and variance parameters, the variance of the binomial distribution is dependent on the mean. Additionally, the variance of the distribution is dependent on the probability of a success; this will be important for interpretation of this model as well as the ordinal logistic regression model. When n is very large and π is near 0.5, the binomial distribution resembles a normal distribution; it is bell-shaped and symmetric, though it is still a discrete distribution.

The canonical link function for the binomial distribution is the logit. The logit is a mathematically convenient function that allows the logistic regression model to have a linear form. The logit is defined as the natural log of the odds, where the odds is the probability of an event occurring divided by the probability of the event not occurring. The formula for the logit is

$$(7) \qquad \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right),$$

where $\hat{\pi}$ is the predicted probability of an event occurring. As mentioned above, an advantage of GLiM is that it allows a non-linear relation between predicted values and predictors. Figure 1 illustrates the non-linear relation between probability and logit.

For binary logistic regression, observed outcome values are typically coded 1 (case or success) or 0 (non-case or failure), but predicted values are in the probability metric. Predicted probabilities ($\hat{\pi}$)

are continuous but bounded by 0 and 1. Probabilities can also be algebraically converted to odds, that is, $odds = \left(\frac{\hat{\pi}}{1-\hat{\pi}}\right)$, the probability of an event occurring divided by the probability of the event not occurring. For example, if the probability of being a case is 0.75, the odds of being a case is 0.75/0.25 = 3; an individual is 3 times more likely to be a case than a non-case. The logit is the natural log (ln) of the odds, so

(8) $$\text{logit} = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right),$$

where $\hat{\pi}$ is the predicted probability of being a case.

The ordinal logistic regression model (also known as the ordered logit model or the cumulative logit model; Agresti, 2002; Fahrmeir & Tutz, 2001; Hosmer & Lemeshow, 2000; Allison, 1999) generalizes binomial logistic regression to outcome variables that have 3 or more *ordered* categories. One example of an outcome with ordered categories is education, with outcome choices of high school diploma, college diploma, and post-graduate degree. These three options for the outcome variable are distinct and have an inherent ordering, where a college diploma indicates more education than a high school diploma and a post-graduate degree indicates more education than a college degree. Researchers in the social sciences also use Likert-type scales as outcomes; Likert-type scales contain ordered categories, such as strongly disagree, disagree, neutral, agree, strongly agree.

The ordinal logistic regression model is a GLiM with a multinomial error distribution and logit link function that is estimated using $(a-1)$ binary logistic regression equations, where a is the number of ordered categories of the dependent variable. Compared to the multinomial logistic regression model (not discussed here), which is a model for *unordered* categories, the ordinal logistic regression model has several important properties that make it the preferred model choice for many ordered outcomes. Specifically, the ordinal logistic regression model requires that the probability of crossing each threshold from a lower category to the next higher category (e.g., from strongly disagree to disagree; from disagree to neutral) is constant across all category thresholds. Therefore, the ordinal logistic regression model does not become more difficult to interpret with more predictors. The ordinal logistic regression model gains only 1 regression coefficient for each additional predictor because the effect of that predictor is the same regardless of which threshold is being crossed; in contrast, multinomial logistic regression model gains $(a-1)$ regression coefficients for each additional predictor because the effect of the predictor also depends upon which threshold is being crossed. Additionally, if the outcome options are ordered and certain assumptions are met, the ordinal logistic regression model has substantially more statistical power than the multinomial logistic regression model.

The ordinal logistic regression model takes into account the fact that the outcome has a specific ordering. This ordering is reflected in the predicted outcomes for each of the (a – 1) equations. The ordinal logistic regression model characterizes the *cumulative* probability of an individual being in a certain category *or a higher category*. For example, if the outcome has five categories, such as the Likert scale described above, there would be four equations estimated. For each equation, the predicted outcome would be the natural log of the probability of belonging to a specific category or higher divided by the probability of belonging to all lower categories. The predicted outcomes for these four equations would be:

(9) $$ln\left(\frac{\hat{\pi}_{strongly\ agree}}{\hat{\pi}_{strongly\ disagree}+\hat{\pi}_{disagree}+\hat{\pi}_{neutral}+\hat{\pi}_{agree}}\right)$$

(10) $$ln\left(\frac{\hat{\pi}_{agree}+\hat{\pi}_{strongly\ agree}}{\hat{\pi}_{strongly\ disagree}+\hat{\pi}_{disagree}+\hat{\pi}_{neutral}}\right)$$

(11) $$ln\left(\frac{\hat{\pi}_{neutral}+\hat{\pi}_{agree}+\hat{\pi}_{strongly\ agree}}{\hat{\pi}_{strongly\ disagree}+\hat{\pi}_{disagree}}\right)$$

(12) $$ln\left(\frac{\hat{\pi}_{disagree}+\hat{\pi}_{neutral}+\hat{\pi}_{agree}+\hat{\pi}_{strongly\ agree}}{\hat{\pi}_{strongly\ disagree}}\right)$$

Each equation compares the probability of being in a certain category or higher to the probability of being in all lower categories. Another way of thinking about ordinal logistic regression is in terms of thresholds. The regression equation corresponding to the predicted outcome in equation (9) describes the probability of an individual crossing the threshold from the “agree” category up to the “strongly agree” category. Likewise, the regression equation corresponding to the predicted outcome in equation

(11) describes the probability of an individual crossing the threshold from the "disagree" category up to the "neutral" category.

Proportional odds assumption. The ordinal logistic regression model has an additional assumption related to the effect of regression coefficients on the transitions between outcome categories that is known as the *proportional odds* or *parallel regressions assumption*. The proportional odds assumption states the all (a – 1) equations have the same regression coefficient for the same predictor; intercepts are allowed to change as a function of transition between adjacent dependent variable categories. Conceptually, this means that a predictor variable has the same effect on *moving up a category* or *crossing the threshold to the next higher category*, regardless of location in the ordering of categories. Different intercepts for each equation essentially allows for the fact that different proportions of the sample will be in each outcome category. The ordinal logistic regression model for an outcome with 3 outcome options would be estimated by the following 2 equations:

(13) $$ln\left(\frac{\hat{\pi}_3}{\hat{\pi}_1+\hat{\pi}_2}\right) = b_{0,3} + b_1X_1 + b_2X_2 + \cdots + b_pX_p$$

and

(14) $$ln\left(\frac{\hat{\pi}_2+\hat{\pi}_3}{\hat{\pi}_1}\right) = b_{0,23} + b_1X_1 + b_2X_2 + \cdots + b_pX_p.$$

Note that, except for the intercepts, the regression coefficients are the same in both equations. The same regression coefficient, b_1, is used to specify the effect of X_1 in both equations.

Poisson Regression

Poisson regression (Cameron & Trivedi, 1998; Gardner, Mulvey, & Shaw, 1995; Long,1997) is the appropriate analysis when the outcome variable is a count of the number of events in a fixed period of time. The probability mass function for the Poisson distribution,

$$(15) \qquad P(Y = y \mid \mu) = \frac{\mu^y}{y!} e^{-\mu},$$

gives the probability of observing a given value, *y*, of variable *Y* that is distributed according to a Poisson distribution with parameter μ. For the count variable *Y*, μ is the arithmetic mean number of events that occur in a specified time interval; the Poisson distribution would yield the probability of 0, 1, 2, …, *k* events, given the mean μ of the distribution. The Poisson distribution differs from the normal distribution (used in linear regression) in several ways that make the Poisson more attractive for representing the properties of count data. First, the Poisson distribution is a discrete distribution which takes on a probability value only for non-negative integers. In contrast, the normal distribution is continuous and takes on all possible values from negative infinity to positive infinity, not just positive integers. Second, count outcomes typically display increasing variance with increases in the mean. This property is known as heteroscedasticity of variance; it is a violation of the previously mentioned assumption of linear regression and can result in severely biased standard error estimates if linear regression is applied to count data. The Poisson distribution is specified by only one parameter, μ, which defines *both* the mean and the variance of the distribution; that is, the mean and the variance of the Poisson distribution are equal. In contrast, the normal distribution requires two independent parameters to be identified: the mean parameter, μ, and the variance parameter, σ^2. The fact that the mean and variance of the Poisson distribution are completely dependent on one another can be useful in modeling count outcomes.

A Poisson distribution with a high expected value (as a rule of thumb, greater than 10) begins to roughly resemble a normal distribution in shape and symmetry. However, the Poisson distribution is still discrete and has identical values for the mean and variance. Figure 2 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 roughly 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: $\ln(\hat{\mu}) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p$ where $\hat{\mu}$ is the predicted count on the outcome variable, given the specific values on the predictors $X_1, X_2, \ldots, X_p$. The use of GLiM with the Poisson error structure resolves the major problems with applying linear regression to count outcomes, namely non-constant variance of the residuals, non-normal conditional distribution of residuals, and out-of-range prediction.

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 linear regression. The residuals are conditionally Poisson distributed because for any value of the predicted mean ($\hat{\mu}$), the residuals are distributed according to expression (15). 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, since the observed values are counts, the residuals may take on only a limited set of values.

Beta Regression

Beta regression (Kieschnick & McCullough, 2003; Paolino, 2001; Smithson & Verkuilen, 2008) is a type of analysis that expands generalized linear models and is discussed in McCullagh & Nelder (1989), Chapter 10. Beta regression differs from the previously presented GLiMs (ordinal logistic regression and Poisson regression) because it models the mean and variance of an outcome using two different regression equations: one equation models the mean structure of the outcome, whereas the other equation models the variance structure of the outcome. These mean and variance models may have different link functions, different sets of predictors, and different prediction equations for the conditional mean and variance. The mean and variance models are combined into a single error structure based on the beta distribution; this combination of mean and variance in the variance structure allows the modeling of heteroscedasticity (i.e., prediction of non-constant variance).

Beta regression is useful for a wide variety of variables that are not necessarily discrete but also do not meet some of the assumptions for normally distributed outcome variables; these include variables that have upper or lower bounds, excessive skew, or excessive heteroscedasticity. Unlike the other GLiMs discussed here, the outcomes for which beta regression is used are typically not categorical. A common use for beta regression is the modeling of proportions (e.g., Brehm & Gates, 1993; Kieschnick & McCullough, 2003), but beta regression can also be used to model extremely skewed, heteroscedastic, or even U-shaped outcomes.

The error structure for beta regression is the standard beta distribution, with probability density function:

$$(16) \qquad f(Y \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1},$$

where a and b are both shape parameters for the distribution and $\Gamma(x)$ is the gamma function of x, which is equal to $(x-1)!$ or $(x-1)(x-2)\cdots(2)(1)$. For a *standard* beta regression, the predicted values range from 0 to 1, inclusive, but the beta distribution can be adapted to fit any other ranges of predicted values. The beta distribution is an extremely versatile distribution that can take on a wide variety of shapes. The beta distribution is U-shaped if both *a* and *b* are less than 1, unimodal if both *a* and *b* are greater than 1, monotonically increasing if *a* is 1 or greater and *b* is less than or equal to 1, and monotonically decreasing if *a* if 1 or less and *b* is greater than or equal to 1. The versatile shape of the beta distribution means that it can be used to model a variety of error function shapes that cannot be adequately modeled by other regression models, such as logistic regression.

The parameterization of the beta distribution shown in equation (16) does not easily lend itself to modeling, because both the *a* and *b* parameters are shape parameters, not location and spread parameters. For modeling of proportions, Smithson and Verkuilen (2008) suggest re-parameterizing the distribution into mean and precision parameters, μ and ϕ, respectively, given by

$$(17) \qquad \mu = \frac{a}{a+b}$$

and

$$(18) \qquad \phi = a + b,$$

where the variance of the distribution is a function of both the mean and the precision parameter. The precision parameter is somewhat analogous to a variance parameter in that it reflects the spread of the observed values around the mean; however, the precision parameter is the inverse of a variance parameter (i.e., high variance is associated with low precision). Note that, like many other GLiMs, the mean and precision of the beta distribution are not independent; the expressions for the mean (μ) and the precision (ϕ) both contain the shape parameters, *a* and *b*.

The beta regression model actually has two different prediction equations: the mean/location model and the precision/dispersion model. A logit link is typically used to model the mean of the outcome, which lies between 0 and 1 for a proportion, so the location is modeled as

(19) $$ln\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p,$$

where $X_1, \cdots, X_p$ are the *p* predictors of the mean structure. The link function can be inverted (see Cohen et al., 2003, p. 488 for a complete explanation) to show the relationship between the predicted mean (here, a proportion) and the predictors rather than the relationship between the *logit of the predicted mean* and the predictors that is shown in Equation (19). Inverting the link function produces the expression for the predicted mean value, which is a proportion:

(20) $\hat{\mu} = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$. This is the model for the mean or location parameter, μ.

The precision parameter, ϕ, is modeled using a separate equation, with potentially different predictors. The precision parameter must always be positive, so it is typically modeled using a natural log link. The precision parameter is modeled as:

(21) $$ln(\hat{\phi}) = \delta_0 + \delta_1 Z_1 + \cdots + \delta_k Z_k;$$

note that there are different regression coefficients for this portion of the model ($\delta_0, \cdots, \delta_k$), as well as potentially different predictors ($Z_0, \cdots Z_k$). The precision parameter reflects how accurate or precise estimates are; high precision means that values are highly accurate and focused. Variance or dispersion is the inverse of precision; high variance or dispersion means that values are *not* focused or accurate. Since we are accustomed to thinking in terms of dispersion and variance rather than precision, some authors (e.g., Smithson & Verkuilen, 2008) use this fact to ease interpretation and model the dispersion (σ) as the inverse of the precision parameter. Therefore, we can present the dispersion as:

(22) $ln(\hat{\sigma}) = -(\delta_0 + \delta_1 Z_1 + \cdots + \delta_k Z_k)$.

(Algebraically, ln(1/*x*) = -*x*). Inverting this link function produces the expression for the predicted dispersion value,

(23) $\hat{\sigma} = e^{-(\delta_0 + \delta_1 Z_1 + \cdots + \delta_k Z_k)}$.

To be more explicit about the way that both the location and the variance are modeled jointly, one can examine the log-likelihood function for the beta regression model. The log-likelihood function for beta regression for an individual is

(24) $\ln L(a, b \mid y_i) = \ln\Gamma(a+b) - \ln\Gamma(a) - \ln\Gamma(b) + (a-1)\ln(y_i) + (b-1)\ln(1-y_i)$.

It can be shown algebraically from equations (17) and (18) that

(25) $a = \mu\sigma$

and

(26) $b = \sigma - \mu\sigma$.

Inserting the expected values for μ and σ from equations (20) and (23) into equations (25) and (26) for *a* and *b*, and in turn inserting those expressions into the log-likelihood function produces the log-likelihood function to jointly model the mean and dispersion. Of note in expression (23), information about *both* the relationship between the predictors and the mean and the relationship between the predictors and the dispersion are involved in the log-likelihood (expression (24) above) and estimation of parameters. The fact that separate (though related) information about the mean and the dispersion means that the beta regression model should be much more flexible than Poisson regression and ordinal logistic regression models in correctly capturing the unique properties of some outcome variables.

A beta regression model will produce two sets of regression coefficients: one for the model of the mean and one for the model of the dispersion. Each set of regression coefficients can be interpreted according to their corresponding link function. For example, the mean model uses a logit link function, so the regression coefficients for the mean model are interpreted in a manner similar to logistic regression. For logistic regression, results are commonly discussed in terms of the odds ratio, e^b. A 1-unit increase in the predictor *X* multiplies the odds being a case by the odds ratio. Dispersion model regression coefficients are often not interpreted (e.g., Ferrari & Cribari-Neto, 2004; Kieschnick & McCullough, 2003), but it is important to note the meaning of a significant regression coefficient in the dispersion/precision model. A significant regression coefficient implies that that predictor significantly predicts *variation* in the outcome, that is, that the predictor models heteroscedasicity. Because of the versatility of the beta distribution, the dispersion function can take on a wide variety of forms, including constant variance (i.e., homoscedastic like linear regression), increasing variance with increases in the predictor, or variance that increases then decreases as a function of the predictors.

Chapter 4

Grouped Counts and Grouped Frequencies

Outcome variables in the social sciences can take on a variety of forms. Common outcomes include binary variables, counts, ordered categories, and proportions and other bounded variables. Additionally, some variables may not fit clearly into a single group for the purposes of choosing an appropriate analysis method. One type of outcome variable that fits this description is grouped counts or grouped frequencies (GCGF). This type of variable may be used when an exact count or frequency is unknown or difficult for an individual to estimate or remember. An example of a GCGF variable is the number of cigarettes that an individual smokes per day; options may include 0, 1-3, 4-10, 11-20, and more than 20. Another example is a variable reflecting how many minutes per day an individual exercises; in this case, options may be less than 15 minutes, 15 to 30 minutes, 30 minutes to 60 minutes, and more than 60 minutes. In both situations, a true count or frequency exists, but responses are categorized into pre-determined (and sometimes arbitrary) ranges.

Measurement Properties

For GCGF outcome variables, the choice of an appropriate analysis technique is unclear. One reason for this confusion is that, historically, statisticians have recommended choosing an analysis technique based on the “level of measurement” of the outcome. The levels of measurement suggested by Stevens (1946) are based on the mathematical operations that can be meaningfully performed on a set of numbers. These four levels of measurement are known as (a) nominal, (b) ordinal, (c) interval, and (d) ratio, with nominal allowing the fewest and most limited mathematical operations and ratio allowing the most. Nominal variables are simply named categories with no inherent order, such as religions or political parties. Ordinal variables are named categories with some innate ordering, such as rankings. For ordinal variables, the order reflects position but a difference of one rank is not necessarily consistent across the range of rankings. Interval variables are ordered and have consistent difference between values across the range of the variable, but they do not have a meaningful zero-point, so ratios of scale values cannot be compared. A common interval level variable is the Fahrenheit temperature scale: a difference of 15 degrees means the same thing whether that difference is between 10 and 25 degrees or between 70 and 85 degrees, but the zero-point is arbitrary, so ratios of temperatures are not meaningful. Ratio level variables have all of the properties of interval variables with the added property of a meaningful zero point, allowing meaningful ratios of values. The Kelvin temperature scale is a ratio level variable: zero degrees K represents zero molecular activity, so ratios of temperatures can be meaningfully compared. For example, 40 degrees K represents twice the molecular activity of 20 degrees K, just as 20 degrees K represents twice the molecular activity of 10 degrees K.

Since the publication of Stevens (1946), these four levels of measurement have been viewed as strong guidelines for determining the allowable mathematical operations, and therefore the allowable statistical calculations, that can be performed on a variable. For example, calculating the mean of a variable requires that the variable be measured at an interval level of measurement or higher. Linear regression is generally held to be appropriate for continuous, interval-level or ratio-level outcome variables. However, many cases of the application of linear regression to lower-than-interval-level variables exist: for example, the linear probability model is the application of linear regression to a binary outcome. This often occurs because correct analysis methods are unknown (such as for GCGF outcomes), under-studied (in psychology, this includes many GLiMs besides logistic regression), or difficult to implement (such as beta regression for proportions, which requires writing separate programs or "tricking" existing, complex procedures in SAS).

Researchers in psychology and other areas have discussed the true utility of Stevens' four levels of measurement. Many have found them to be limited and inadequate for classifying many types of variables (for example, Chrisman, 1998; Velleman & Wilkinson, 1993). For example, counts are often used as outcomes in psychological studies. Count variables are ordered, categorical, and have a meaningful zero value. Therefore, they share properties with both ordinal variables (ordered and categorical) and ratio variables (meaningful zero), while not having some properties that are typical of ratio variables, such as being continuous. Stevens (1946) considered counts to be ratio level variables.

The choice of a level of measurement is further confused when “natural” variable types are manipulated in some way, as is the case with GCGF outcomes. The standard method for choosing an appropriate statistical analysis relies heavily on a somewhat arbitrary number of measurement levels that may or may not be appropriate for all types of variables. In contrast, the choice of an appropriate statistical analysis may also be based on the degree of match between the outcome variable and the analysis (Velleman & Wilkinson, 1993). This latter method of choosing an analysis method may prove to be more useful when analyzing outcome variables that do not fit cleanly into the four standard measurement levels; among these scale formats are grouped counts and grouped frequencies.

Analysis Approaches

Choosing an appropriate statistical analysis based on the degree of match between the outcome variable and the analysis requires a careful examination of each method. Specifically, one must determine what the model underlying the analysis assumes concerning the outcome variable. Along the same lines, one must determine how any potential mismatch between outcome properties and analysis requirements will affect the model results. For GCGF outcomes, four analysis methods are considered: (a) linear regression, (b) ordinal logistic regression, (c) Poisson regression, and (d) beta regression. This section presents the properties of each method that make it a desirable choice for use with GCGF outcomes, as well as any potential problems that may be encountered.

Linear regression. As described in detail above, linear regression assumes that the outcome being analyzed is unbounded and conditionally normally distributed, with errors having a conditional mean of zero and constant variance of σ^2. The advantage of linear regression for GCGF outcomes is that it is easy to use and interpret and is the standard method of analysis in many areas of psychology. The disadvantage of using linear regression for GCGF outcomes is that counts and frequencies (and therefore their grouped counterparts) are likely to have non-normal conditional distributions and be heteroscedastic. Additionally, using linear regression for these types of outcomes can easily result in out-of-bounds predicted values, since counts and frequencies have a lower bound of zero. Residuals for a linear regression model for grouped counts and grouped frequencies will also not be normally distributed due to the discrete nature of the outcome.

Ordinal logistic regression. Logistic regression and ordinal logistic regression can be interpreted in a latent variable framework that is conceptually very similar to that of linear regression. For ordinal logistic regression, the ordered categories are described as being based on an underlying continuous latent variable; the observed categories are defined by thresholds or cut points. The latent variable is assumed to be conditionally distributed according to the logistic distribution (which is bell-shaped, symmetric, and similar in shape to the normal distribution) and homoscedastic. GCGF outcomes are typically skewed and heteroscedastic like the counts and frequencies underlying them, so the assumption of homoscedasticity in the ordinal logistic regression model poses the same problems as linear regression.

Poisson regression. Poisson regression assumes that an outcome is non-negative, conditionally Poisson-distributed, and heteroscedastic in a strict manner, such that the conditional mean of the outcome is equal to the conditional variance of the outcome. Poisson regression is the preferred method of analysis for count outcomes because the Poisson distribution can model the skew, heteroscedasticity, and lower bound that are commonly seen in counts. One potential drawback of using Poisson regression for GCGF outcomes is that the grouping of the outcome will cause distortion of the multiplicative effect seen in Poisson regression. To clarify, Poisson regression assumes a multiplicative effect of predictors, that is, that $E(Y|X=x+1) = e^b \times E(Y|X=x)$, where e^b is the exponentiation of the regression coefficient for X. This multiplicative relation seen in raw counts may be distorted when the outcome is coarsely grouped into categories of different sizes (e.g., 0, 1-2, 3-5, 6-10). Specifically, this distortion may manifest as unobserved heterogeneity of the outcome variance because the variability of the outcome represents variability of different values of the underlying count or frequency; for example, the variability of the “3-5” category is a combination of the variance for the values of 3, 4, and 5.

Beta regression. Beta regression assumes an outcome that is continuous with both upper and lower bounds. Heteroscedasticity of error is allowed by this model, but not required. One advantage of beta regression compared to the others is that it is much more flexible about the error structure. Since beta regression models the variance structure separately from the mean structure, the errors may be homoscedastic (as in linear regression) or heteroscedastic (as in Poisson regression); the errors need not follow a strict pattern of heteroscedasticity such as that seen in Poisson regression. However, since the beta distribution is a continuous distribution and GCGF outcomes are discrete, beta regression faces many of the same problems as linear regression; namely, the residuals are not able to closely follow the continuous beta distribution.

Chapter 5

Statistical Power

This study examines the statistical power of the four previously described regression models to detect the effect of a predictor on a GCGF outcome. Two related concepts, type I error and confidence interval coverage, are also examined. Statistical power refers to the probability of detecting an effect in a sample given that the effect does in fact exist in the population (Cohen, 1988; Maxwell, 2000; Maxwell, Kelley & Rausch, 2008). Type 2 error rate is the probability that a true effect in the population is *not* detected in the sample; statistical power is 1 minus the type 2 error rate, $1-\beta$. Adequate power (typically taken as $1-\beta \geq .80$) reflects the ability to detect true effects.

Statistical power is determined by three factors: sample size, effect size, and type I error rate (Cohen et al., 2003). Statistical power can be increased by increasing sample size or by increasing the standardized effect size; for example, the addition of covariates, refined measurement that reduces error variance, and optimal design approaches that sample a wide range of values on the predictor are common methods used to increase the effect size of interest.

Type I error has an obvious relationship to statistical power. While statistical power indicates how likely one is to detect an effect in a sample that actually exists in the population, type I error indicates how likely one is to detect an effect in a sample when that effect *does not* actually exist in the population. A type I error rate that is close to the nominal value (e.g., alpha = 0.05 in most studies) indicates that the likelihood of finding a significant result in error is appropriately low.

In regression models, the regression coefficient is a “point estimate” of the regression coefficient parameter in the population; the regression coefficient is a single number that is supposed to reflect the population value. An alternative or complementary estimate of the population effect is a confidence interval. A confidence interval provides a range of values which should contain the population parameter with some degree of confidence. If a very large number of samples of the same size were taken from the same population, a 95% confidence interval should capture the population parameter in 95% of the time. The 95% is known as the “confidence level;” confidence interval coverage refers to how closely the empirical confidence level (for example, the proportion of replications in a simulation study in which the population value is contained in each confidence interval) matched the nominal confidence level (typically 95% or 90%).

Statistical Power in Linear Regression

In linear regression, there are two types of significance tests for which one might want to determine power. The first is an omnibus test of the prediction by the entire model with the null hypothesis, $H_0 : \rho^2_{multiple} = 0$. The second is a test of an individual regression coefficient with the null hypothesis, $H_0 : \beta_j = 0$. The present research focuses on the single predictor test; for completeness, the omnibus test is also presented here. Both the omnibus test and regression coefficient tests in linear regression are Wald-type tests, that is, the estimate of the parameter is divided by its standard error, with the result being compared to a t-distribution to determine statistical significance. Effect size, f^2, for the omnibus test in linear regression is based on the $R^2_{multiple}$ of the model (Cohen, 1988), using the relation:

(27) $$f^2 = \frac{R^2_{multiple}}{1-R^2_{multiple}}.$$

The effect size, f^2, ranges from 0 to infinity. For a test of a single parameter such as a single regression coefficient, the effect size is based on the $R^2_{multiple}$ and the $R^2_{multiple}$ for a model with the predictor of interest removed, using the relation:

(28) $$f^2 = \frac{R^2_{multiple}-R^2_{multiple(-j)}}{1-R^2_{multiple}},$$

where $R^2_{multiple(-j)}$ is the $R^2_{multiple}$ for a model in which the predictor of interest, predictor *j*, has been excluded. Nominal type I error rate is typically fixed before the study, usually at 0.05 (two-tailed) for studies in the behavioral sciences. Equations and tables (for example, in Cohen, 1988, and Cohen et al., 2003) as well as statistical software (e.g., G*Power, Faul, Erdfelder, Lang & Buchner, 2007) exist to determine the power of a study with a given type I error rate, sample size, and effect size.

Two distinct *F*-distributions are employed to determine statistical power for linear regression. The first represents the null hypothesis of no variance accounted for (i.e., no effect) and the second represents the alternative hypothesis of some non-zero variance accounted for (i.e., some non-zero effect). The null hypothesis is represented by a central or standard *F*-distribution that is familiar from statistical testing. This is the distribution that supplies the critical *F*-value for statistical tests. The alternative hypothesis is represented by the non-central *F*-distribution. The non-central *F*-distribution is shifted to the right of the standard *F*-distribution by an amount determined by the non-centrality parameter. The non-centrality parameter, λ, is determined by effect size and sample size using the relation

(29) $\lambda = n \times f^2$.

The area of the central *F*-distribution that is to the right of the critical *F*-value is the alpha (α) value or the type I error rate. The area of the non-central *F*-distribution that is to the right of the critical *F*-value is the statistical power for the test. Larger effect sizes and larger sample sizes will push the non-central *F*-distribution for the alternative hypothesis farther to the right of the central *F*-distribution, meaning that more of the non-central *F*-distribution is beyond the critical *F*-value and the test has more power.

Statistical Power for GLiMs

Statistical power for generalized linear models cannot be calculated in the same way as linear regression for several important reasons. First, standardized effect size measures for GLiMs are not as well defined as those for linear regression. An examination of the multiple pseudo-R^2 measures for GLiMs (e.g., see West, Aiken, & Kwok, 2003; DeMaris, 2002; Menard, 2001) shows that there is not a single measure of effect size that is appropriate, interpretable, and unbiased across all GLiMs. Second, Wald tests are generally not considered the most appropriate statistical tests for GLiMs. Many software programs (e.g., SAS and SPSS) produce Wald tests for regression coefficients in GLiMs. However, Hauck and Donner (1977), Vaeth (1985), and others have shown that Wald tests behave in a peculiar manner in GLiMs, especially in small samples and in tests of individual parameters (i.e., tests of regression coefficients). Likelihood ratio (LR) and Score tests are often preferred to Wald tests for testing both individual parameters and omnibus hypotheses in GLiMs. Due to the difficulty of easily implementing the appropriate Score test, this study focuses on only the LR test as an alternative to the Wald test.

Much of the research on power for GLiMs occurs in areas outside of psychology. GLiMs, especially logistic regression and count or rate models, are often used in medicine and epidemiology; this is reflected in the large number of articles on power for GLiMs that are found in journals that focus on biological and medical research methods (e.g., *Biometrics*, *Biometrika*, and *Statistics in Medicine*). Much of this research on power for GLiMs has focused on tests of individual regression coefficients, rather than on tests of overall model fit. Many areas of medical research are concerned with the effect of an individual predictor, such as a treatment group, rather than the overall predictive power of a set of predictors.

Likelihood Ratio Test

The likelihood ratio test (Chernoff, 1954; Wilks, 1938) is a nested model test that compares the deviance (or "lack of fit") of a model in which the parameter of interest (for example, a regression coefficient) is estimated to a model in which the parameter of interest is fixed to θ_o. A significant test indicates that the parameter is significantly different from θ_o; for example, to test whether a regression coefficient is significantly different from 0, the value of θ_o is set equal to 0. The likelihood ratio test statistic is given by

(33) $LR = D(M_0) - D(M_\beta)$,

where $D(M_\beta)$ is the deviance of the model with the parameter estimated and $D(M_0)$ is the deviance of the model in which the parameter is fixed to θ_o. For a test of a single regression coefficient, the test statistic has an asymptotic chi-square distribution with 1 degree of freedom. If the LR test statistic exceeds the critical value of the chi-square distribution, the regression coefficient is statistically different from θ_o.

Statistical power for test statistics with a chi-square distribution is conceptually similar to that described above for test statistics with an F-distribution: the power of the test is the area of the non-central chi-square distribution that exceeds the critical chi-square value on the central chi-square distribution. The primary area of research on power and the LR test focuses on proper estimation of the non-centrality parameter. The LR test is an asymptotic method, so proper estimation of the non-centrality parameter in non-infinite samples is extremely important.

Snapinn and Small (1986) examined very small sample estimation of the non-centrality parameter for the LR test for ordinal logistic regression. In small samples ($n < 50$), this method had more appropriate type I error rates than the standard LR test, though type I error was slightly higher than the nominal value. Self, Mauritsen, & Ohara (1992) examined adjustments to the non-centrality parameter for the LR tests for several different GLiMs, focusing on the special situation of case-control models. Because the Self et al. (1992) method focuses on the case-control model, it makes very specific assumptions about the predictors and is essentially limited to categorical predictors with few response options. Shieh (2000b) expanded the Self et al. (1992) method to allow for continuous as well as categorical predictors. Both methods give more accurate sample size estimates than the standard LR test.

Much of the research on LR test adjustment for the purposes of calculating power and sample size focuses on special case uses of the

GLiM, either in type of outcome (such as ordinal logistic regression) or in study design (such as very small samples or case-control studies). This results in limited generalization of results to other outcome types or study designs. General, practical guidelines for required sample size in GLiMs are therefore unavailable. Additionally, the fact that research focuses on special cases means that the conclusions are often incongruent across methods.

Wald Test

Despite admonitions that the Wald test for regression coefficients is biased and underpowered compared to the LR test (e.g., Hauck & Donner,1977; Vaeth, 1985), a great deal of the research on power and sample size determination for generalized linear models focuses on the Wald test. The Wald test is widely used in many areas including psychology and is readily available from statistical software packages. Some researchers approach the problem of power for the Wald test via specific models within the GLiM family, such as logistic regression; others seek a more unified solution based on the shared properties of all GLiMs.

The classic source for power and sample size in logistic regression is Whittemore (1981); Whittemore (1981) assumed a small proportion of “cases” on the outcome in order to simplify calculation of a covariance matrix of the regression coefficients, providing estimates of the variance of the regression coefficient for Wald tests. Hsieh (1989) used the methods developed by Whittemore (1981) to produce extensive tables of sample sizes for logistic regression that are widely used in psychology and other areas. Signorini (1991) expanded on Whittemore's (1981) methods to determine sample sizes required for Poisson regression. Shieh (2000a) showed via simulation that the LR test methods developed by Self et al. (1992) provide better estimates of the sample size required for logistic regression than the Wald test methods of Whittemore (1981). Shieh (2001) later provided refinements to the Wald test methods of Whittemore and Signorini for both logistic regression and Poisson regression.

Strickland and Lu (2003) and Tsonaka, Rizopoulos, and Lesaffre (2006) focus on important special cases of GLiMs; specifically, both studies focus on randomized treatment-control studies with binary or bounded (i.e., proportion) outcomes. These studies use the odds ratio from logistic regression as a measure of effect size; the odds ratio measure of effect size makes these methods somewhat less attractive because they cannot be easily generalized to models besides logistic regression. In simulations, the Strickland and Lu (2003) method overestimated the sample size required for specified power, as compared to empirical power levels, particularly when the effect size was moderate to large (i.e., an odds ratio greater than or equal to 1.5 by their definition).

Newson (2004) describes a "generalized power method" that allows for estimation of power and sample size with any outcome type that can be analyzed with a GLiM. Newson's (2004) method uses the raw mean difference (as opposed to a standardized mean difference) as a measure of effect size; the “influence function” for an outcome type (for example, a binary outcome for logistic regression) depends on this mean difference and the sample size. The generalized power method described in Newson (2004) has been implemented as the POWERCAL command in Stata software.

Statistical Power for GCGF outcomes

It was unclear how information on statistical power for GLiMs might translate to statistical power for grouped count and grouped frequency outcomes. The research presented above on statistical power for GLiMs generally employed simulated data that satisfies the distributional assumptions of the model. For example, Self et al. (1992) simulated a binary outcome to determine the performance of their method for logistic regression and a count outcome to determine the performance of their method for Poisson regression. GCGF outcomes do not satisfy the assumptions of any of the four analysis models tested here; for each of the analysis models (linear regression, ordinal logistic regression, Poisson regression, and beta regression), the model is actually deliberately misspecified when using a GCGF outcome.

Chapter 6

Method

The purpose of this dissertation was to compare several regression models used for outcome variables that are characterized as grouped counts or frequencies. Specifically, the statistical power, type I error rates, and CI coverage for these models were examined. This study used a statistical simulation to determine empirical (observed) power, type I error, and coverage for each regression model in several different conditions.

Factors that may affect the performance of the models include the (1) the mean structure of the relationship between the predictor and the outcome, (2) the conditional variance structure of the outcome, (3) the magnitude of the relationship between the predictor and the outcome, (4) sample size, and (5) the type of predictor (either continuous or binary). Each of these factors was varied in this simulation study; the factors were crossed in a fully factorial design.

Data Generation

All data were generated using SAS 9.2 software (SAS Institute, 2008). For each condition, 1000 replications were conducted. A single predictor X was generated as a normally distributed variable with a mean of 0 and variance of 1. These values of X were used for the continuous predictor conditions; for the binary X conditions, X was dichotomized using a median split following generation of the Y outcome variable. The outcome Y was created as a function of X and the magnitude of the relationship between *X* and *Y* (the effect size), using either a (1a) linear or (1b) exponential mean structure and either a (2a) homoscedastic, (2b) heteroscedastic and increasing, or (2c) heteroscedastic and football-shaped variance, as described in detail below.

Effect size. In order to compare effects in both linear models and non-linear models, a measure of effect size was determined that is measured in the same units for both linear and non-linear models. A common measure of effect size is Cohen's *d* (Cohen, 1988), which is the standardized difference between two group means. Cohen's *d* is calculated as

$$(34) \qquad \frac{\bar{x}_1 - \bar{x}_2}{s}.$$

Cohen never explicitly defined the value of *s* when the groups have different standard deviations; Hedges (1991) further defined the standardized effect size to use the *pooled* standard deviation for the two groups as the denominator. When the two groups have equal standard deviations, Cohen's d and Hedges' g are equal. A similar measure of effect size can be constructed to compare a linear mean structure (as seen in linear regression) and a non-linear, exponential mean structure (as seen in Poisson regression), but several modifications need to be made to make the measure comparable across models. Note that in order to compare linear and nonlinear models, this method differs from the typical effect size used for nonlinear models. For example, the effect size for logistic regression is typically defined in terms of the odds ratio (Hosmer & Lemeshow, 2000); the effect size for Poisson regression is sometimes defined as the multiplicative effect for a 1-unit change in the predictor (e^b; Long, 1997) or as the response rate ratio ($e^b/e^b{}_0$; PASS software; Hintze, 2011).

First, to make the measure have a common metric for models that have different transformations of the predicted score, the predicted score for each model must be in the original metric of the outcome. For example, the predicted value for the standard form of Poisson regression is the natural log of the predicted count, $ln(\hat{Y})$. A predicted score in the original outcome metric can be obtained by raising the transformed value to a power of e, such that

(35) $e^{\ln(\hat{Y})} = \hat{Y}$.

Predicted scores for the non-linear (Poisson regression-like) mean structure can be transformed back into the original outcome metric; predicted scores for the linear (OLS-like) mean structure are already in the original outcome metric.

Second, for conditions in which the predictor in the regression model is continuous, the numerator cannot be simply defined as the difference between two group means. In keeping with a standardized measure of effect size, we can examine the change in the predicted score for a 1 standard deviation change in the predictor, at the mean of the predictor. That is, the numerator of the measure is defined as

(36) $\hat{Y}_{\bar{X}+0.5SD} - \hat{Y}_{\bar{X}-0.5SD}$,

where the first term is the predicted outcome (in the original outcome metric) for a value of the predictor that is ½ standard deviation above the predictor mean and the second term is the predicted outcome for a value of the predictor that is ½ standard deviation below the predictor mean. For the binary predictor conditions, the binary predictor variable is coded as -0.5 and +0.5.

The denominator of Hedges *g*, given by *s*, is an estimate of the variability of the outcome variable, taking into account that two different groups with potentially different amounts of variability are being observed. In Hedges' *g*, *s* is the pooled estimate of the standard deviation for both groups and is defined as

(37) $s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$,

where n_i and s_i represent the sample size and standard deviation, respectively, of group *i*. Again, for conditions in which the predictor is continuous, the standard deviation for a group is not appropriate. Additionally, a continuous predictor does not allow for a weighted average of multiple variability estimates (for example, the average of the variability of the outcome at 1 SD below the mean and the variability of the outcome at 1 SD above the mean). For all conditions in this study, the denominator of the effect size measure is defined as the standard deviation of the outcome at the mean of the predictor, that is $s_{Y|\bar{X}}$. The effect size measure used for this study, which represents the number of standard deviation units change in the outcome for the 1 SD span about the mean of the predictor, is given by

(38) $\frac{\hat{Y}_{\bar{X}+0.5SD} - \hat{Y}_{\bar{X}-0.5SD}}{s_{Y|\bar{X}}}$.

Multiple effect sizes were generated to evaluate the effect of the magnitude of the effect size on the power of these regression models. Effect sizes examined in this study corresponded to (3a) a 0 standard deviation unit change in the outcome, (3b) a 0.35 standard deviation unit change in the outcome, (3c) a 0.87 standard deviation unit change in the outcome, and (3d) a 1.39 standard deviation unit change in the outcome.[1] The 0 effect size conditions were used to assess type I error rates; the other three effect size conditions were used to assess statistical power. CI coverage was assessed under all effect size conditions.

Mean structure of the outcome. Recall that the predictor was generated to have a mean of 0 and a standard deviation of 1. The outcome variable was generated as a function of the predictor with either a (1a) linear or (1b) exponential mean structure. In order to maintain the previously described effect size equivalence (i.e., the standardized difference between the original-metric mean 0.5 SD below the mean of *X* and the original-metric mean 0.5 SD above the mean of *X*), the linear and exponential mean structures used different regression coefficients.

The linear mean structure was created using the expression

(39) $Y = 3b_{lin}X + 3$,

where Y is the outcome variable, X is the predictor, and b_{lin} is the intended magnitude of the relationship between X and Y. In this study, b_{lin} was not identical to the measure of effect size; b_{lin} took on values of 0, 0.2, 0.5, and 0.8 while the effect sizes were 0, 0.35, 0.87, and 1.39 (see Footnote 1). The exponential mean structure was created using the relation

(40) $Y = e^{(b_{exp}X + b_0)}$,

where b_{exp} is a regression coefficient reflecting the exponential relationship between X and Y, based on the effect size equivalence and b_0 is an intercept value used to equate the means of the linear and exponential mean structure models. In this study, b_{exp} and b_0 took on different values depending on the effect size. The value of b_{exp} was 0, 0.199668, 0.494933, and 0.780071 for effect sizes of 0, 0.35, 0.87, and 1.39, respectively. The value of b_0 was 0 for an effect size of 0 and 1.09861229 for all other effect sizes.

Variance structure of the outcome. The variance structure of the outcome was constructed following the creation of the mean structure. The variance structure of the outcome Y was either (2a) normally distributed and homoscedastic (i.e., constant variance for all values of the predictor), (2b) right skewed and heteroscedastic (i.e., non-constant variance across values of the predictor) following the equidispersion (i.e., conditional variance is equal to the conditional mean) assumption of Poisson regression, or (2c) heteroscedastic “football-shaped” variance structure. Specific variance values were created to maintain equivalent effect sizes for all mean and variance structure combinations. The measure of effect size used here is defined as the mean difference in Y for a 1 SD change about the mean of the predictor divided by the conditional variance of Y at the mean of the predictor; therefore, the mean structures maintain a constant mean difference in Y for both linear and non-linear mean structures, while the variance structures maintain a constant conditional variance at the mean of the predictor, regardless of whether the variance structure is homoscedastic, Poisson-like, or football-shaped.

Intercepts were included in both mean structures for development of the linear and nonlinear mean structures. The Poisson variance structure requires the conditional variance to be equal to the conditional mean. The intercept forced the conditional mean of the outcome to be equal to 3 when the predictor was at its mean value of 0. Therefore, the conditional variance of the Poisson variance structure when the predictor was equal to 0 was also equal to 3. For conditions with Poisson variance structure, the conditional variance was equal to the conditional mean; the conditional mean was set equal to 3 at the mean of the predictor.

For the football-shaped variance conditions, the conditional variance was largest at the mean of the predictor and decreased linearly as the value of the predictor became more extreme. When the predictor was equal to 0, a random residual was drawn from a normal distribution with a mean of 0 and a variance of 3; when the predictor was equal to +3 or -3, a random residual was drawn from a normal distribution with a mean of 0 and a variance of 1.5.

For the homoscedastic variance conditions, the constant variance of the outcome was 3. Regardless of the value of the predictor, a random residual from a normal distribution with a mean of 0 and a variance of 3 was added to the mean value of the outcome. Figures 3 and 4 show the mean structure and variance structure combinations used in this study. Figures 5 and 6 show representative samples (n = 250, d = 0.87) for the 6 mean and variance combinations used in this study

Sample size. Finally, sample size is known to have an impact on statistical power, with increases in sample size corresponding to increases in statistical power. The sample sizes evaluated here were (4a) 100, (4b) 250, (4c) 500, and (4d) 1000.

Predictor type. The predictor X either remained in its original continuous format or was dichotomized into a binary variable following generation of the outcome variable. For the binary predictor conditions, the predictor was dichotomized using a median split; the lowest 50% of the observations were assigned to have a value of -0.5 for the predictor and the highest 50% of the observations were assigned to have a value of 0.5 for the predictor. The binary predictor case parallels a treatment-control experimental study with equal-size groups.

Coarse categorization of the outcome. Following data generation, outcome values were coarsely grouped into 5 categories.

These categories roughly correspond to groupings appropriate to the types of substantive areas in which GCGF outcomes are often used. A common frame for these items is time, such as the past month. For this study, it was assumed that the counts or frequencies had an upper bound of 30, representing a 1-month timeframe. (Other timeframes, such as the past week or the past 2 weeks are also frequently used; for this study, it was decided to use the single timeframe of 30 days.)

For outcomes created using the homoscedastic or football-shaped variance structures, the raw outcome values are continuous; for outcomes created using a Poisson variance structure, the raw outcome values are counts. For the continuous raw outcomes, a categorical integer value was assigned such that values less than or equal to 0.4999999999 were assigned a value of 0, values greater than or equal to 0.5 and less than or equal to 1.4999999999 were assigned a value of 1, etc. Finally, raw outcome values were grouped into the following categories: 0, 1 – 3, 4 – 8, 9 – 15, and 16 – 30. Each individual case was assigned the midpoint for its respective category as its new outcome value; that is, a case with an outcome score of 3 (in the 1 – 3 category) was assigned a new outcome value of 2 and a case with an outcome score of 10 (in the 9 – 15 category) was assigned a new outcome value of 12.

Analysis

The grouped outcome data for the 1000 replications in each condition were analyzed using each of the four modeling techniques described above: linear regression, ordinal logistic regression, Poisson regression and beta regression. Empirical power for each condition was calculated by determining the proportion of non-zero effect size condition replications in which a significant result was obtained, using the Wald test and the likelihood ratio test. Empirical type I error rates for each condition were calculated by determining the proportion of zero effect size condition replications in which a significant result was found.

Linear regression, ordinal logistic regression, and Poisson regression analyses were conducted using SAS PROC GENMOD. Beta regression was conducted using SAS PROC NLMIXED. PROC NLMIXED does not provide LR tests. For beta regression, the type I error and statistical power rates were obtained by re-analyzing all data with a “null” beta regression model (i.e., a model with no predictors); the LR test was conducted by comparing the -2LL from the single-predictor model to the -2LL from the null model. Confidence intervals for the LR tests could not be calculated, so confidence interval coverage is not included for the LR test with beta regression.

Confidence interval coverage rates were calculated as the proportion of the 1000 replications in which the 95% CI around the regression coefficient contained the true population value of the regression coefficient. The population values of the regression coefficients are *not* equal to the regression coefficients used to generate the data because the outcome variable was coarsely categorized following data generation. For each condition, a single replication with a sample size of 1,000,000 was generated. The estimate of the regression coefficient from this single, very large replication ($\hat{\beta}_{1,000,000}$) converges in probability to the population value of the regression coefficient ($\tilde{\beta}$), or $\hat{\beta}_{1,000,000} \overset{p}{\to} \tilde{\beta}$; $\tilde{\beta}$ is the probability limit or plim. The estimate $\hat{\beta}_{1,000,000}$ serves as the population value for the purposes of calculating bias and confidence interval coverage.

Chapter 7

Results

Table 1 shows the type I error, statistical power, confidence interval coverage, and relative bias for OLS linear regression analysis conducted on the ungrouped outcome for the conditions in which the mean structure was linear and the variance was homoscedastic. These conditions match the assumptions of OLS linear regression (i.e., continuous outcome, linear mean structure, and homoscedastic variance), so these values can be used for comparison to results on the grouped outcome. In the ungrouped data, the regression coefficients were unbiased; type I error, statistical power, and CI coverage were within the expected ranges.

Relative Bias

Relative bias of the estimated regression coefficients was assessed for all models in all non-zero effect size conditions. Tables 2 and 3 show the relative bias of the regression coefficient for each model, given by

(41) $\frac{\hat{\beta}-\hat{\beta}_{1,000,000}}{\hat{\beta}_{1,000,000}}$,

for continuous and binary predictors, respectively. Recall that $\hat{\beta}_{1,000,000}$ is the estimate of the regression coefficient obtained using a single replication with 1,000,000 observations and that this estimate converges in probability to a population value of $\tilde{\beta}$.

Shaded cells indicate relative bias of greater than ±5%. Positive numbers indicate that the estimate is larger than the population value;

negative values indicate that the estimate is smaller than the population value. Only the beta regression model showed substantial relative bias in estimating regression coefficients. Specifically, beta regression showed high positive bias (i.e., the estimated regression coefficient was smaller than the population value) for linear mean structure and Poisson-like variance (relative bias ranged from -0.056 to -0.275) and for exponential mean structure and very large effect size (relative bias ranged from -0.072 to -0.217). OLS linear regression, Poisson regression, and ordinal logistic coefficients did not show substantial bias in estimating regression.

Type I Error

Tables 4 and 5 show the type I error rates for the Wald test for continuous and binary predictor conditions, respectively. Bradley's (1978) stringent (type I error rate = [0.045, 0.055]) and liberal (type I error rate = [0.025, 0.075]) criteria for type I error rates were used to assess whether the rates were sufficiently close to nominal levels (i.e., $\alpha = 0.05$). For the Wald test, the linear regression model had appropriate type I error rates for all predictor type, mean structure, variance structure, and sample size conditions. The ordinal logistic regression model also had appropriate type I error rates for all conditions. Type I error rates were appropriate for the Poisson regression model only for conditions in which the mean structure was exponential and the variance structure was Poisson; in other words, the type I error rate was correct for the Poisson regression model when the raw outcome variable closely followed the assumptions of the Poisson regression model. Type I error rates were appropriate for a few conditions of the beta regression model; there was no obvious pattern to the results.

Tables 6 and 7 show the type I error rates for the likelihood ratio test for continuous and binary predictor conditions, respectively. The pattern of results for the LR test was nearly identical to the pattern of results for the Wald test. The linear regression model had appropriate type I error rates for all conditions except when the sample size was 1000, the mean structure was exponential, and the variance structure was football-shaped. The ordinal logistic regression model had appropriate type I error rates for all conditions. The Poisson regression model had appropriate type I error rates only for conditions with an exponential mean structure and Poisson variance structure. The beta regression model had appropriate type I error rates for some conditions with exponential mean structure and a sample size of 100.

Statistical Power

Tables 8 and 9 show the statistical power rates for the Wald test for continuous and binary predictor conditions, respectively. Shaded cells indicate conditions for which type I error rates were unacceptably larger or smaller than the nominal value, per the liberal criteria set forth by Bradley (1978); power rates for these conditions are not readily interpretable.

For the continuous predictor conditions, the linear regression model had adequate statistical power (i.e., ≥0.80) for all conditions except when the effect size was small to medium (0.35), the sample size was 100, and the variance structure was football-shaped; statistical power 0.592 and 0.597 for those conditions. The pattern of statistical power rates for ordinal logistic regression was very similar to that of linear regression; the conditions in which n was equal to 100, the effect size was small to medium, and the variance structure was football-shaped showed statistical power rates of 0.604 and 0.584. Power was also adequate for the Poisson regression model for the conditions which parallel the assumptions of Poisson regression.

For the binary predictor conditions, statistical power was somewhat lower, as expected for a dichotomized predictor. The linear regression model had adequate statistical power for all conditions except when the effect size was small to medium (0.35) and the sample size was 100; and when the effect size was small to medium (0.35), the sample size was 250, and the variance structure was football-shaped. Ordinal logistic regression showed a pattern of statistical power that was very similar to that of linear regression. Power was also adequate for the Poisson regression model for the conditions which parallel the assumptions of Poisson regression, except with a sample size of 100 and a small to medium (0.35) effect size.

Tables 10 and 11 show the statistical power rates for the likelihood ratio test for continuous and binary predictor conditions, respectively. The pattern of results for the LR test was virtually identical to the pattern of results for the Wald test. The linear regression model and the ordinal logistic regression model had adequate power in nearly all conditions with a continuous predictor and in most conditions with a binary predictor. The Poisson regression model exhibited adequate power for the conditions in which the raw outcome was generated following the assumptions of Poisson regression (i.e., exponential mean structure and conditionally Poisson distributed variance structure).

Confidence Interval Coverage

Tables 12 and 13 show the coverage rates for the Wald test for continuous and binary predictor conditions, respectively. Bradley's (1978) stringent ([0.045, 0.055]) and liberal ([0.025, 0.075]) criteria were used to assess adherence to the nominal (α = .05) type I error rate. To determine acceptable CI coverage, the inverse of Bradley's (1978) type I error criteria were used; for example, the stringent criterion has a lower confidence limit of 0.045 for type I error, so the upper confidence limit for coverage was 1 – 0.045 = 0.955. A stringent criterion of coverage = [0.945, 0.955] and a liberal criterion of coverage = [0.925, 0.975] were used to assess whether coverage was near the nominal value of 0.95.

For the linear regression model, CI coverage values were very close to the nominal 0.95 confidence level for most conditions. For a continuous predictor, low CI coverage rates (ranging from 0.765 to 0.915) were observed for linear regression for conditions with large or very large effect sizes and heteroscedastic variance structures, particularly when the mean structure was exponential; for a binary predictor, CI coverage for linear regression was improved and was close to the nominal value for nearly all conditions. For the ordinal logistic regression model, CI coverage values were very close to the nominal 0.95 confidence level for all but one condition with a continuous predictor; the empirical CI coverage rates ranged from 0.915 to 0.975 for a continuous predictor and from 0.939 to 0.971 for a binary predictor. Systematic patterns of adequate coverage were not observed for any other models.

Tables 14 and 15 show the coverage rates for the likelihood ratio test for continuous and binary predictor conditions, respectively. The pattern of results for the LR test was very similar to the pattern of results for the Wald test. The CI coverage rates for the linear regression model followed the same pattern as the CI coverage for the Wald test; several conditions with large or very large effect sizes, exponential mean structures, and heteroscedastic variance structures showed CI coverage rates notably lower than the nominal value. The CI coverage rates for the ordinal logistic regression model were very close to the nominal confidence level, ranging from 0.913 to 0.976 for a continuous predictor and from 0.938 to 0.968 for a binary predictor. For the binary predictor conditions, CI coverage rates for both linear regression and ordinal logistic regression were closer to the nominal values. No systematic pattern of adequate coverage rates were observed for the other models. The SAS NLMIXED procedure does not automatically produce a likelihood ratio test of the regression coefficient, so the LR test was manually computed for beta regression; CI coverage rates for beta regression were not available.

Chapter 8

Discussion

This study examined the statistical performance of four regression models that may be used to analyze grouped count or grouped frequency outcomes: linear regression, ordinal logistic regression, Poisson regression, and beta regression. Of the four models evaluated, linear regression performed well in terms of relative bias, type I error, and statistical power, but did not provide adequate CI coverage for several conditions that are highly relevant to the analysis of count and frequency outcomes (i.e., exponential mean structure and heteroscedastic variance). Ordinal logistic regression performed well in terms of relative bias, type I error, statistical power, and confidence interval coverage, regardless of the type of predictor, sample size, effect size, mean structure, or variance structure. Poisson regression produced type I error rates, statistical power rates and confidence interval coverage rates that were appropriate, but only for conditions in which the ungrouped outcome followed the assumptions of Poisson regression (i.e., exponential mean structure and Poisson variance).

Model fit

One aspect of statistical performance that was not addressed here was model fit. Model fit refers to how closely the predicted outcome values match the observed outcome values, or in terms of a statistical model, how well the model is able to reproduce the observed values. Model fit is assessed by comparing the observed and predicted outcome scores. For example, the chi-square test of model fit uses the squared difference between the observed and expected scores, divided by the expected score (Daniel, 1990). Other measures of model fit that are commonly used in structural equation modeling (such as the CFI, RMSEA, AIC, and BIC) are functions of the chi-square statistic. Ryan (1997) suggests using the correlation between the observed scores and the predicted scores as a measure of model fit; for GLiMs such as Poisson regression which involve transformations of the predicted score, the predicted score should be converted back into the original units of the observed score.

The difficulty in applying these methods of model fit lies in producing predicted scores. Linear regression, Poisson regression, and beta regression can easily produce a single predicted score for each observation. However, ordinal logistic regression does not produce a single predicted score for each observation; ordinal logistic regression produces several predicted probabilities that each indicate the probability of crossing the threshold to the next higher outcome category. For example, if there are 4 outcome categories, an individual observation will have 3 predicted probabilities: the first predicted probability is the probability of crossing the threshold from the first category to the second, the second predicted probability is the probability of crossing the threshold from the second category to the third, and the third predicted probability is the probability of crossing the threshold from the third category to the fourth. It may not always be possible to use these several predicted probabilities to assign a single predicted outcome category. It may also be possible to assess model fit by invoking the latent variable interpretation of the outcome; ordinal logistic regression can also be conceptualized in terms of a single latent variable, the intercepts reflecting latent thresholds in the outcome variable.

Accuracy of prediction is an important aspect of modeling in many areas of the social sciences. Since ordinal logistic regression was clearly the best model choice in terms of the statistical measures assessed here, it would be extremely useful to assess model fit as well. It may be the case that model fit analysis would reveal weaknesses of the ordinal logistic regression model that are not apparent in the type I error rates, power rates, and confidence interval coverage. However, given that the multiple predicted scores produced by ordinal logistic regression are not conducive to assessing model fit, it seemed less valuable to compare only the three poorly performing regression models in terms of model fit.

Effect sizes

The effect sizes examined here are not identical to the small medium, and large Cohen's d effect sizes commonly used in psychology. Table 1 shows the results for linear regressions performed on the ungrouped outcome variable. The non-zero effect sizes used here are approximately a small/medium (0.35), large (0.87), and very large (1.39) Cohen's *d*. Linear regression and ordinal logistic regression had low statistical power for a sample size of 100 and a Cohen's *d* of 0.35; it is likely that the trend toward low statistical power continues and that small and very small effect sizes also have low statistical power. Linear regression had low CI coverage for large and very large Cohen's d; effects of this size are uncommon in some areas of psychology (e.g., clinical, developmental, social, and personality), but are much more common in other areas such as cognitive psychology, behavioral neuroscience, and medicine. For example, in the journal *Statistics in Medicine*, Strickland & Lu (2003) use an odds ratio of 1.5 to represent a moderate to large effect; an odds ratio of 1.5 is approximately equal to a Cohen's d of 0.83 (Chinn, 2001).

Variance Structures and Effect Size

In this study, a measure of effect size was used that is roughly analogous to standard Cohen's *d* and Hedges' *g* measures of effect size. Specifically, effect size was defined as the difference in (original metric) Y for a 1 SD change around the mean of X, divided by the standard deviation of Y *at the mean of X*. For the homoscedastic variance structure, the variance of Y is constant across the range of X, meaning that the effect size measure used here is equivalent to Cohen's *d* and Hedges' *g*. However, two of the three variance structures used in this study (Poisson-like and football-shaped) were heteroscedastic, meaning that the conditional variance of Y varied as a function of X.

The Poisson-like variance structure increases with increasing values of X; however, the variance at the mean of X should roughly approximate the average variance across the range of X. The variance in the football-shaped variance conditions reaches its *maximum value* at the mean of X; therefore, the variance at the mean of X does not approximate the average variance across the range of X for this variance structure. For the football-shaped variance structure, the variance at the mean of X is actually larger than the average variance. This discrepancy between the average variance and the variance at the mean of X results in standard errors that are inappropriately large and decreased statistical power for conditions with football-shaped variance structures. While the measure of effect size was useful given the linear and non-linear mean structures, it shows weaknesses related to unusual non-constant variance conditions.

Proportional Odds Assumption

It was suggested that the good performance of the ordinal logistic regression model may be due to the fact that the data were actually generated to follow the proportional odds assumption of ordinal logistic regression. Recall that the proportional odds assumption states that a predictor variable has the same effect on moving up a category or crossing the threshold to the next category, regardless of the location in the ordering of the categories; that is, a single regression coefficient governs the transition between all pairs of adjacent thresholds across the ordered continuum of categories of the dependent variable. The proportional odds assumption would translate to the presence of a linear effect, where a change in the predictor always results in the same amount of change in the outcome.

If the good performance of ordinal logistic regression model were due *solely* to proportional odds effects, both linear regression and ordinal logistic regression should perform worse for the nonlinear exponential mean structure conditions. In particular, one would expect to observe poorer performance of both linear regression and ordinal logistic regression in the largest effect size conditions with an exponential mean structure; in these conditions, the nonlinear effect would be most pronounced. However, the ordinal logistic regression model showed good statistical performance in both linear and exponential mean structure conditions while linear regression produced low CI coverage for exponential mean conditions.

It is likely that coarse categorization of the outcome leads to the loss of some information about the relationship between the predictor and the outcome. In other words, the nonlinear effects may be absorbed by the coarse categorization. For example, the exponential relationship may by somewhat flattened and linearized by grouping, resulting in a grouped outcome that somewhat approximates the proportional odds assumption, even though the raw outcome does not. As with tests of model fit, an assessment of the proportional odds assumption may reveal weaknesses in the performance of the ordinal logistic regression model related to the proportional odds assumption.

Linear Regression Underperformance

The performance of the linear regression model was generally good, but CI coverage was unacceptable in a number of conditions that are highly relevant for count and frequency outcomes. When the mean structure was exponential, linear regression produced CI coverage rates that were lower than the nominal values; this effect was enhanced when the effect was large or very large (i.e., the exponential effect was more pronounced) and the variance structure was Poisson-like (i.e., monotonically increasing). These points of weakness are important to the analysis of counts and frequencies because counts tend to have an exponential relationship with predictors and also tend to have monotonically increasing heteroscedasticity. Despite the general good performance of linear regression in terms of type I error and statistical power, the fact that the specific weak points of linear regression align so closely with the properties of counts and frequencies makes linear regression a less appealing option for analyzing grouped counts and frequencies.

Probability Limit vs. Theoretical Population Values

In this study, data were generated according to a specific population effect size relationship between X and Y, but following data generation, the outcome was coarsely grouped into several categories.

This coarse categorization means that the effect sizes used to generate the raw data may not reflect the true relationship between X and the coarsely categorized outcome. A probability limit (plim) estimate of the “true” relationship between X and the coarsely categorized outcome was determined for each condition by generating a single replication with a sample size of 1,000,000. The value of the regression coefficient for this single, very large sample was used as the population value when calculating relative bias of regression coefficients and confidence interval coverage. However, the effect size used during data generation, not the plim estimate, was used to determine which conditions were used to assess type I error rates (effect size = 0) and which were used to assess statistical power (effect size > 0).

A comparison of the population effect sizes and their corresponding plim values for linear regression revealed that there were some systematic differences between the population effect sizes used to generate the data and the probability estimates of the population relationship following coarse categorization of the outcome. (Only population and plim values for the linear regression model were compared due to the addition complication of comparing linear and non-linear effects.) Of particular note, the plim estimate was typically larger than the corresponding population value for linear mean structure conditions, but smaller than the population value for exponential mean structure condition. In addition, binary predictor conditions had plim values that far exceeded the population values, for both mean structures.

Relative bias estimates (based on the plim values) revealed very little bias in sample estimates. The lack of bias in comparing the plim values to the individual samples' estimates suggests two possible alternatives. First, the plim estimate is a valid measure of the population regression coefficient for data that are generated with a particular population value and subsequently manipulated in some way (such as coarse categorization). Second, both the plim measure and the individual samples' estimates are biased in a similar way, resulting in agreement between the two numbers. If the first alternative is true, the results of this study and others like it can be accepted in their current state; if the second alternative is true, it would be of interest to compare both plim value and the population regression coefficients to the individual samples' estimates.

Conclusions

Based on the results from this simulation study, the analysis of choice for GCGF outcome variables is ordinal logistic regression. In addition to the statistical performance observed in this study, ordinal logistic regression has several advantages that make it a good analysis choice. It is easily to implement in common statistical packages and is relatively easy to interpret. When the proportional odds assumption is satisfied, there are other properties of the ordinal logistic regression model that make it particularly appealing for GCGF outcomes. According to

Agresti (1996), “When the proportional odds model holds for a given response scale, it also holds with the same effects for any collapsing of the response categories” (p. 215). For GCGF outcomes, this means that slightly different groupings of the outcome count or frequency should not result in substantively different results and conclusions. Given that the specific grouping of counts and frequencies is often arbitrary and a matter of convenience rather than directed planning, this property of ordinal logistic regression is encouraging.

The statistical findings for the four analysis models were very consistent, with ordinal logistic regression consistently performing within the desired ranges for type I error, statistical power, and CI coverage rates. Linear regression performed well in terms of type I error and statistical power, but the low CI coverage in conditions which parallel the properties of counts and frequencies makes linear regression less appealing. More work is needed to assess model fit and accuracy of prediction for all four analysis models; this information would certainly complement the type I error, statistical power, and CI coverage results. Additionally, much of the current research on statistical power in GLiMs focuses on determining the minimum sample size required to obtain 0.80 power. It would be worthwhile to determine the minimum sample size required to have adequate power to detect effects in GCGF outcome variables. These sample size estimates can be compared to established estimates of required sample size for linear regression, Poisson regression, ordinal logistic regression, and beta regression in order to find the penalty involved in coarsely categorizing counts and frequencies.

Table 1. Linear regression on ungrouped counts for linear mean and homoscedastic variance conditions

	n = 100				*n* = 250				*n* = 500				*n* = 1000			
Cohen's *d*	0	0.35	0.87	1.39	0	0.35	0.87	1.39	0	0.35	0.87	1.39	0	0.35	0.87	1.39
Type I error	0.050				0.056				0.048				0.062			
Power		0.919	1.000	1.000		1.000	1.000	1.000		1.000	1.000	1.000		1.000	1.000	1.000
Coverage	0.950	0.942	0.953	0.949	0.944	0.956	0.950	0.954	0.952	0.941	0.951	0.957	0.938	0.948	0.959	0.954
Relative Bias		0.000	-0.002	-0.001		-0.001	-0.001	0.002		0.005	-0.001	0.001		0.001	-0.001	0.001

Table 2. Relative bias in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100			*n* = 250			*n* = 500			*n* = 1000		
		Effect Size	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	-0.016	-0.041	-0.065	-0.010	-0.026	-0.064	0.002	-0.018	-0.058	0.005	-0.009	-0.056
		Poisson-like	-0.056	-0.108	-0.275	-0.025	-0.088	-0.232	-0.010	-0.067	-0.169	-0.007	-0.048	-0.099
		Football-shape	0.014	-0.007	-0.001	0.005	0.000	-0.006	0.012	0.002	-0.003	0.005	0.000	-0.002
	Exponential	Homoscedastic	-0.043	-0.217	-0.045	-0.019	-0.157	-0.007	-0.004	-0.130	-0.002	-0.007	-0.075	0.001
		Poisson-like	-0.036	-0.205	-0.024	-0.021	-0.141	-0.003	-0.005	-0.079	-0.004	-0.002	-0.041	-0.002
		Football-shape	0.000	-0.171	-0.031	-0.008	-0.142	-0.011	-0.002	-0.101	-0.003	-0.001	-0.072	-0.002
Linear	Linear	Homoscedastic	0.007	-0.001	-0.001	0.002	-0.002	0.000	0.011	-0.002	0.000	0.011	-0.004	-0.001
		Poisson-like	0.004	-0.003	0.006	-0.003	-0.005	0.000	0.002	0.000	0.004	0.000	0.002	0.001
		Football-shape	0.020	0.005	0.001	0.007	0.006	0.001	0.008	0.004	-0.001	0.004	0.001	-0.002
	Exponential	Homoscedastic	-0.001	-0.001	-0.002	-0.011	0.002	0.000	-0.003	0.001	0.000	-0.006	0.002	0.003
		Poisson-like	-0.003	0.003	0.003	-0.009	0.007	0.001	0.000	0.006	-0.003	-0.002	0.003	0.000
		Football-shape	0.006	-0.009	-0.004	-0.010	-0.006	-0.007	-0.002	0.001	-0.003	-0.001	-0.002	-0.002
Ordinal	Linear	Homoscedastic	0.036	0.032	0.040	0.014	0.010	0.020	0.017	0.006	0.011	0.013	-0.001	0.008
		Poisson-like	0.037	0.030	0.047	0.006	0.006	0.021	0.005	0.007	0.010	0.003	0.006	0.005
		Football-shape	0.048	0.026	0.030	0.015	0.017	0.010	0.016	0.009	0.007	0.008	0.002	0.002
	Exponential	Homoscedastic	0.030	0.037	0.047	0.003	0.011	0.016	0.005	0.005	0.008	-0.003	-0.001	0.006
		Poisson-like	0.034	0.027	0.038	0.003	0.022	0.016	0.005	0.009	0.008	-0.001	0.005	0.003
		Football-shape	0.034	0.017	0.019	0.000	0.004	0.005	0.003	0.008	0.004	0.001	0.000	0.003
Poisson	Linear	Homoscedastic	0.013	0.008	0.020	0.005	0.004	0.005	0.010	-0.001	0.003	0.009	-0.002	0.002
		Poisson-like	0.004	0.008	0.021	-0.005	0.001	0.013	0.000	0.002	0.008	0.000	0.003	0.005
		Football-shape	0.025	0.011	0.022	0.009	0.010	0.006	0.007	0.007	0.003	0.003	0.002	0.001
	Exponential	Homoscedastic	0.000	0.003	0.006	-0.009	0.000	0.004	-0.002	0.000	0.001	-0.006	0.001	0.002
		Poisson-like	-0.004	0.001	0.012	-0.006	0.008	0.006	0.001	0.002	0.003	-0.001	0.002	0.003
		Football-shape	0.013	-0.002	0.007	-0.007	-0.005	-0.002	-0.002	0.002	0.000	0.000	-0.002	-0.001

Shaded cells are conditions in which the relative bias was greater than ±5%.

Table 3. Relative bias in binary predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100			*n* = 250			*n* = 500			*n* = 1000		
			0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.002	-0.005	-0.004	-0.002	0.006	0.000	0.010	-0.004	0.001	0.006	-0.002	-0.001
		Poisson-like	0.019	0.003	-0.015	0.004	-0.008	-0.014	0.003	-0.003	-0.005	-0.007	-0.001	-0.002
		Football-shape	0.034	0.009	0.007	-0.001	0.011	0.004	0.001	0.001	0.003	0.001	0.006	0.002
	Exponential	Homoscedastic	-0.021	-0.042	-0.059	0.018	-0.021	-0.014	-0.007	-0.010	-0.010	-0.010	-0.001	-0.007
		Poisson-like	-0.010	-0.067	-0.040	-0.004	-0.045	-0.010	-0.017	-0.021	-0.005	0.002	-0.005	-0.001
		Football-shape	0.013	-0.020	-0.042	-0.028	-0.009	-0.013	0.006	-0.008	-0.008	-0.011	-0.004	0.000
Linear	Linear	Homoscedastic	-0.013	-0.018	-0.014	-0.003	0.003	-0.003	0.009	-0.006	0.000	0.009	0.000	-0.003
		Poisson-like	0.009	0.001	-0.003	-0.001	-0.004	-0.003	0.001	-0.003	-0.002	-0.005	0.002	-0.001
		Football-shape	0.005	-0.011	-0.014	-0.017	0.006	-0.001	0.001	0.000	-0.002	-0.008	0.002	0.002
	Exponential	Homoscedastic	-0.030	-0.007	-0.014	0.005	-0.008	0.001	-0.012	0.002	-0.001	-0.009	0.001	-0.002
		Poisson-like	-0.018	0.003	-0.008	-0.007	-0.003	0.000	-0.016	0.005	0.000	0.001	0.004	0.002
		Football-shape	0.001	-0.001	-0.011	-0.026	0.005	-0.002	0.004	0.005	-0.004	-0.010	0.001	0.000
Ordinal	Linear	Homoscedastic	0.012	0.010	0.014	0.006	0.012	0.006	0.011	-0.001	0.006	0.011	0.001	-0.001
		Poisson-like	0.040	0.035	0.050	0.008	0.008	0.010	0.007	0.004	0.007	-0.003	0.006	0.004
		Football-shape	0.027	0.013	0.018	-0.008	0.015	0.007	0.002	0.004	0.002	-0.006	0.007	0.002
	Exponential	Homoscedastic	-0.006	0.025	0.024	0.017	0.003	0.013	-0.007	0.009	0.002	-0.008	0.005	0.002
		Poisson-like	0.007	0.021	0.033	0.002	0.004	0.006	-0.014	0.010	0.004	0.005	0.006	0.004
		Football-shape	0.028	0.024	0.013	-0.018	0.009	0.005	0.008	0.009	0.003	-0.008	0.005	0.003
Poisson	Linear	Homoscedastic	-0.003	-0.011	-0.004	0.000	0.004	0.003	0.010	-0.004	0.002	0.009	0.000	-0.002
		Poisson-like	0.016	0.006	0.009	0.000	-0.003	0.002	0.001	-0.001	0.001	-0.005	0.002	0.000
		Football-shape	0.018	0.001	0.002	-0.011	0.016	0.004	0.006	0.003	0.003	-0.007	0.004	0.003
	Exponential	Homoscedastic	-0.022	0.002	-0.011	0.009	-0.005	0.001	-0.010	0.004	-0.002	-0.009	0.001	-0.003
		Poisson-like	-0.013	-0.001	-0.007	-0.007	-0.002	0.000	-0.016	0.004	0.000	0.002	0.004	0.001
		Football-shape	0.006	0.010	-0.005	-0.025	0.003	0.000	0.005	0.006	0.000	-0.011	0.002	0.003

Shaded cells are conditions in which the relative bias was greater than ±5%.

Table 4. Type I error for Wald test in continuous predictor conditions

Analysis Type	Mean Type	Variance Type	*n* = 100	*n* = 250	*n* = 500	*n* = 1000
Beta	Linear	Homoscedastic	0.008	0.004	0.015	0.013
		Poisson-like	0.033**	0.013	0.009	0.015
		Football-shape	0.005	0.003	0.015	0.013
	Exponential	Homoscedastic	0.000	0.000	0.000	0.001
		Poisson-like	0.000	0.000	0.000	0.000
		Football-shape	0.000	0.001	0.004	0.006
Linear	Linear	Homoscedastic	0.062**	0.055*	0.053*	0.058**
		Poisson-like	0.052*	0.061**	0.056**	0.059**
		Football-shape	0.031**	0.043**	0.028**	0.032**
	Exponential	Homoscedastic	0.052*	0.044**	0.056**	0.057**
		Poisson-like	0.052*	0.055*	0.056**	0.065**
		Football-shape	0.040**	0.026**	0.031**	0.025**
Ordinal	Linear	Homoscedastic	0.050*	0.055*	0.052*	0.056**
		Poisson-like	0.042**	0.055*	0.062**	0.057**
		Football-shape	0.033**	0.040**	0.028**	0.029**
	Exponential	Homoscedastic	0.039**	0.045*	0.045*	0.043**
		Poisson-like	0.044**	0.052*	0.054*	0.059**
		Football-shape	0.041**	0.036**	0.031**	0.029**
Poisson	Linear	Homoscedastic	0.102	0.096	0.102	0.106
		Poisson-like	0.084	0.097	0.112	0.098
		Football-shape	0.114	0.115	0.116	0.127
	Exponential	Homoscedastic	0.104	0.112	0.12	0.123
		Poisson-like	0.043**	0.053*	0.057**	0.066**
		Football-shape	0.169	0.170	0.16	0.153

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
Nominal $\alpha = .05$

Table 5. Type I error for Wald test in binary predictor conditions

Analysis Type	Mean Type	Variance Type	n = 100	n = 250	n = 500	n = 1000
Beta	Linear	Homoscedastic	0.002	0.008	0.043	0.080
		Poisson-like	0.008	0.016	0.177	0.388
		Football-shape	0.004	0.028**	0.087	0.091
	Exponential	Homoscedastic	0.001	0.033**	0.092	0.161
		Poisson-like	0.000	0.019	0.051*	0.069**
		Football-shape	0.003	0.030**	0.007	0.001
Linear	Linear	Homoscedastic	0.058**	0.043**	0.046*	0.047*
		Poisson-like	0.051*	0.057**	0.061**	0.050*
		Football-shape	0.056**	0.045*	0.050*	0.045*
	Exponential	Homoscedastic	0.047*	0.060**	0.045*	0.048*
		Poisson-like	0.058**	0.045*	0.047*	0.048*
		Football-shape	0.064**	0.057**	0.035**	0.046*
Ordinal	Linear	Homoscedastic	0.055*	0.042**	0.048*	0.048*
		Poisson-like	0.049*	0.049*	0.061**	0.052*
		Football-shape	0.051*	0.050*	0.048*	0.042**
	Exponential	Homoscedastic	0.052*	0.044**	0.039**	0.051*
		Poisson-like	0.054*	0.050*	0.044**	0.050*
		Football-shape	0.054*	0.060**	0.045*	0.041**
Poisson	Linear	Homoscedastic	0.094	0.093	0.096	0.085
		Poisson-like	0.084	0.093	0.100	0.085
		Football-shape	0.150	0.187	0.156	0.151
	Exponential	Homoscedastic	0.125	0.126	0.130	0.135
		Poisson-like	0.053*	0.042**	0.046*	0.046*
		Football-shape	0.222	0.233	0.227	0.202

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
Nominal α = .05

Table 6. Type I error for LR test in continuous predictor conditions

Analysis Type	Mean Type	Variance Type	*n* = 100	*n* = 250	*n* = 500	*n* = 1000
Beta	Linear	Homoscedastic	0.356	0.457	0.502	0.497
		Poisson-like	0.449	0.487	0.502	0.488
		Football-shape	0.185	0.280	0.466	0.481
	Exponential	Homoscedastic	0.032**	0.376	0.336	0.165
		Poisson-like	0.029**	0.531	0.416	0.414
		Football-shape	0.025**	0.033**	0.345	0.477
Linear	Linear	Homoscedastic	0.062**	0.054*	0.052*	0.057**
		Poisson-like	0.050*	0.061**	0.056**	0.059**
		Football-shape	0.030**	0.042**	0.025**	0.032**
	Exponential	Homoscedastic	0.050*	0.044**	0.056**	0.057**
		Poisson-like	0.050*	0.055*	0.055*	0.063**
		Football-shape	0.040**	0.025**	0.030**	0.024
Ordinal	Linear	Homoscedastic	0.055*	0.055*	0.053*	0.056**
		Poisson-like	0.044**	0.057**	0.063**	0.057**
		Football-shape	0.038**	0.040**	0.029**	0.030**
	Exponential	Homoscedastic	0.045*	0.047*	0.046*	0.043**
		Poisson-like	0.047*	0.054*	0.054*	0.061**
		Football-shape	0.045*	0.036**	0.031**	0.029**
Poisson	Linear	Homoscedastic	0.102	0.097	0.101	0.106
		Poisson-like	0.083	0.097	0.114	0.097
		Football-shape	0.114	0.115	0.115	0.127
	Exponential	Homoscedastic	0.104	0.112	0.120	0.123
		Poisson-like	0.046*	0.052*	0.057**	0.066**
		Football-shape	0.172	0.169	0.160	0.153

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
Nominal α = .05

Table 7. Type I error for LR test in binary predictor conditions

Analysis Type	Mean Type	Variance Type	*n* = 100	*n* = 250	*n* = 500	*n* = 1000
Beta	Linear	Homoscedastic	0.323	0.486	0.517	0.520
		Poisson-like	0.373	0.493	0.502	0.557
		Football-shape	0.158	0.490	0.529	0.515
	Exponential	Homoscedastic	0.405	0.481	0.492	0.548
		Poisson-like	0.032**	0.499	0.468	0.393
		Football-shape	0.347	0.469	0.452	0.304
Linear	Linear	Homoscedastic	0.057**	0.042**	0.046*	0.047*
		Poisson-like	0.051*	0.054*	0.059**	0.050*
		Football-shape	0.054*	0.045*	0.049*	0.045*
	Exponential	Homoscedastic	0.047*	0.060**	0.044**	0.048*
		Poisson-like	0.055*	0.043**	0.047*	0.048*
		Football-shape	0.060**	0.056**	0.035**	0.045*
Ordinal	Linear	Homoscedastic	0.057**	0.043**	0.048*	0.048*
		Poisson-like	0.053*	0.051*	0.061**	0.052*
		Football-shape	0.051*	0.050*	0.048*	0.042*
	Exponential	Homoscedastic	0.052*	0.045*	0.041**	0.051*
		Poisson-like	0.058**	0.050*	0.044**	0.051*
		Football-shape	0.057**	0.060**	0.047*	0.041**
Poisson	Linear	Homoscedastic	0.094	0.093	0.096	0.085
		Poisson-like	0.084	0.093	0.101	0.085
		Football-shape	0.150	0.187	0.156	0.151
	Exponential	Homoscedastic	0.126	0.126	0.131	0.135
		Poisson-like	0.056**	0.042**	0.047*	0.046*
		Football-shape	0.227	0.234	0.227	0.202

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
Nominal $\alpha = .05$

Table 8. Power for Wald test in continuous predictor conditions

Analysis	Mean Type	Variance Type	n = 100			n = 250			n = 500			n = 1000		
			0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.406	0.997	0.999	0.894	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.463	0.989	0.999	0.916	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.107	0.920	1.000	0.432	1.000	1.000	0.911	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.466	0.999	1.000	0.914	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.524	1.000	1.000	0.910	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.119	0.969	1.000	0.469	1.000	1.000	0.926	1.000	1.000	1.000	1.000	1.000
Linear	Linear	Homoscedastic	0.840	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.859	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.592	1.000	1.000	0.938	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.844	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.847	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.597	1.000	1.000	0.943	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ordinal	Linear	Homoscedastic	0.837	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.830	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.604	1.000	1.000	0.946	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.828	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.828	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.584	0.999	1.000	0.948	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
Poisson	Linear	Homoscedastic	0.891	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.901	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.807	1.000	1.000	0.985	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.893	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.888	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.802	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Shaded cells have unacceptably large type I error rates.

Table 9. Power for Wald test in binary predictor conditions

Analysis	Mean Type	Variance Type	n = 100			n = 250			n = 500			n = 1000		
			0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.189	0.957	1.000	0.632	1.000	1.000	0.974	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.284	0.985	1.000	0.722	1.000	1.000	0.972	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.028	0.592	0.985	0.200	0.995	1.000	0.552	1.000	1.000	0.952	1.000	1.000
	Exponential	Homoscedastic	0.200	0.973	1.000	0.685	1.000	1.000	0.970	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.270	0.991	1.000	0.721	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.039	0.713	0.996	0.197	0.998	1.000	0.615	1.000	1.000	0.952	1.000	1.000
Linear	Linear	Homoscedastic	0.630	1.000	1.000	0.947	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.674	1.000	1.000	0.966	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.409	0.967	1.000	0.749	1.000	1.000	0.957	1.000	1.000	0.999	1.000	1.000
	Exponential	Homoscedastic	0.627	1.000	1.000	0.966	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.626	1.000	1.000	0.951	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.425	0.982	1.000	0.756	1.000	1.000	0.972	1.000	1.000	0.999	1.000	1.000
Ordinal	Linear	Homoscedastic	0.625	1.000	1.000	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.650	1.000	1.000	0.963	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.403	0.975	1.000	0.756	1.000	1.000	0.966	1.000	1.000	0.999	1.000	1.000
	Exponential	Homoscedastic	0.617	1.000	1.000	0.969	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.604	1.000	1.000	0.945	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.421	0.971	1.000	0.769	1.000	1.000	0.975	1.000	1.000	1.000	1.000	1.000
Poisson	Linear	Homoscedastic	0.721	1.000	1.000	0.972	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.754	1.000	1.000	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.628	0.995	1.000	0.888	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.729	1.000	1.000	0.987	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.729	1.000	1.000	0.974	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.621	0.999	1.000	0.905	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000

Shaded cells have unacceptably large type I error rates.

Table 10. Power for LR test in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100			*n* = 250			*n* = 500			*n* = 1000		
			0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.841	1.000	1.000	0.734	1.000	0.873	0.692	0.865	0.955	0.788	0.957	0.976
		Poisson-like	0.449	0.795	1.000	0.622	0.834	0.850	0.808	0.915	0.910	0.764	0.972	0.978
		Football-shape	0.581	0.996	1.000	0.633	0.678	0.807	0.522	0.791	0.841	0.539	0.930	1.000
	Exponential	Homoscedastic	0.817	0.999	1.000	0.736	0.803	0.838	0.683	0.939	0.944	0.727	0.944	0.987
		Poisson-like	0.783	1.000	1.000	0.653	0.845	0.903	0.721	0.898	0.952	0.746	0.955	0.992
		Football-shape	0.561	0.985	0.686	0.540	0.631	0.728	0.533	0.666	0.823	0.539	0.711	0.888
Linear	Linear	Homoscedastic	0.838	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.851	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.587	1.000	1.000	0.936	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.841	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.846	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.590	1.000	1.000	0.941	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ordinal	Linear	Homoscedastic	0.843	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.845	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.610	1.000	1.000	0.948	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.838	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.839	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.592	0.999	1.000	0.952	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
Poisson	Linear	Homoscedastic	0.891	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.901	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.807	1.000	1.000	0.985	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.893	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.890	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.801	1.000	1.000	0.984	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Shaded cells have unacceptably large type I error rates.

Table 11. Statistical power for LR test in binary predictor conditions

Analysis	Mean Type	Variance Type	n = 100			n = 250			n = 500			n = 1000		
			0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.709	0.996	0.962	0.772	0.749	0.770	0.610	0.783	0.849	0.656	0.851	0.943
		Poisson-like	0.664	0.999	0.759	0.668	0.719	0.801	0.571	0.749	0.842	0.681	0.872	0.906
		Football-shape	0.420	0.911	0.632	0.516	0.553	0.619	0.512	0.617	0.654	0.556	0.689	0.759
	Exponential	Homoscedastic	0.674	0.989	0.691	0.608	0.743	0.681	0.600	0.748	0.729	0.659	0.804	0.809
		Poisson-like	0.616	0.991	1.000	0.592	0.685	0.691	0.623	0.783	0.742	0.700	0.810	0.840
		Football-shape	0.415	0.901	0.632	0.526	0.592	0.593	0.534	0.642	0.634	0.541	0.698	0.692
Linear	Linear	Homoscedastic	0.625	1.000	1.000	0.947	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.665	1.000	1.000	0.964	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.401	0.967	1.000	0.743	1.000	1.000	0.957	1.000	1.000	0.999	1.000	1.000
	Exponential	Homoscedastic	0.618	1.000	1.000	0.966	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.620	1.000	1.000	0.950	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.415	0.980	1.000	0.753	1.000	1.000	0.969	1.000	1.000	0.999	1.000	1.000
Ordinal	Linear	Homoscedastic	0.634	1.000	1.000	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.664	1.000	1.000	0.965	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.415	0.976	1.000	0.757	1.000	1.000	0.966	1.000	1.000	0.999	1.000	1.000
	Exponential	Homoscedastic	0.625	1.000	1.000	0.970	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.621	1.000	1.000	0.945	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.429	0.971	1.000	0.770	1.000	1.000	0.975	1.000	1.000	1.000	1.000	1.000
Poisson	Linear	Homoscedastic	0.721	1.000	1.000	0.973	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.754	1.000	1.000	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.628	0.995	1.000	0.889	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000
	Exponential	Homoscedastic	0.729	1.000	1.000	0.987	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
		Poisson-like	0.729	1.000	1.000	0.974	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
		Football-shape	0.621	0.999	1.000	0.905	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000

Shaded cells have unacceptably large type I error rates.

Table 12. CI coverage for Wald test in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100				*n* = 250			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.992	0.991	0.898	0.900	0.996	0.997	0.925**	0.870
		Poisson-like	0.967**	0.970**	0.845	0.288	0.987	0.991	0.860	0.041
		Football-shape	0.994	0.999	0.997	0.995	0.997	0.995	0.997	0.999
	Exponential	Homoscedastic	1.000	0.991	0.575	0.747	1.000	0.995	0.158	0.855
		Poisson-like	1.000	0.972**	0.334	0.808	1.000	0.985	0.181	0.844
		Football-shape	1.000	0.999	0.932**	0.811	0.999	0.997	0.769	0.855
Linear	Linear	Homoscedastic	0.937**	0.958**	0.976	0.953*	0.946*	0.961**	0.971**	0.943**
		Poisson-like	0.949*	0.947*	0.953*	0.900	0.939**	0.952*	0.947*	0.899
		Football-shape	0.968**	0.959**	0.971**	0.951*	0.955*	0.965**	0.977	0.966**
	Exponential	Homoscedastic	0.948*	0.945*	0.897	0.749	0.955*	0.959**	0.887	0.778
		Poisson-like	0.948*	0.952*	0.877	0.779	0.945*	0.941**	0.882	0.768
		Football-shape	0.959**	0.962**	0.919	0.837	0.974**	0.970**	0.902	0.821
Ordinal	Linear	Homoscedastic	0.950*	0.952*	0.953*	0.968**	0.945*	0.953*	0.958**	0.962**
		Poisson-like	0.958**	0.955*	0.955*	0.952*	0.944**	0.953*	0.949*	0.959**
		Football-shape	0.966**	0.966**	0.970**	0.958**	0.960**	0.965**	0.969**	0.968**
	Exponential	Homoscedastic	0.960**	0.944**	0.942**	0.931**	0.955*	0.950*	0.937**	0.923
		Poisson-like	0.954*	0.954*	0.953*	0.945*	0.949*	0.944**	0.947*	0.953*
		Football-shape	0.957**	0.968**	0.960**	0.930**	0.964**	0.975**	0.955*	0.945*
Poisson	Linear	Homoscedastic	0.898	0.912	0.929**	0.911	0.904	0.918	0.939**	0.883
		Poisson-like	0.918	0.923	0.917	0.846	0.904	0.930**	0.914	0.851
		Football-shape	0.883	0.868	0.873	0.860	0.886	0.874	0.889	0.867
	Exponential	Homoscedastic	0.896	0.923	0.929**	0.880	0.888	0.924	0.917	0.888
		Poisson-like	0.953*	0.916	0.905	0.891	0.947*	0.897	0.919	0.851
		Football-shape	0.830	0.881	0.874	0.848	0.830	0.879	0.875	0.840

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion

Table 12 continued. CI coverage for Wald test in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 500				*n* = 1000			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.985	0.992	0.940**	0.824	0.987	0.998	0.952*	0.760
		Poisson-like	0.990	0.997	0.826	0.038	0.985	0.983	0.735	0.227
		Football-shape	0.985	0.965	0.998	0.996	0.988	0.937**	0.999	0.993
	Exponential	Homoscedastic	1.000	0.996	0.078	0.838	0.999	0.990	0.265	0.847
		Poisson-like	1.000	0.984	0.424	0.848	1.000	0.961**	0.529	0.836
		Football-shape	0.996	0.968*	0.483	0.854	0.994	0.954*	0.376	0.867
Linear	Linear	Homoscedastic	0.947*	0.951*	0.968**	0.936**	0.942**	0.960**	0.963**	0.940**
		Poisson-like	0.941**	0.955**	0.948*	0.924	0.945*	0.960**	0.944**	0.910
		Football-shape	0.972**	0.969**	0.975**	0.965**	0.968**	0.975**	0.975**	0.965**
	Exponential	Homoscedastic	0.943**	0.960**	0.888	0.754	0.942**	0.967**	0.891	0.776
		Poisson-like	0.944**	0.949*	0.858	0.763	0.938**	0.949*	0.860	0.806
		Football-shape	0.970**	0.958**	0.933**	0.833	0.975**	0.976	0.912	0.833
Ordinal	Linear	Homoscedastic	0.948*	0.939**	0.954*	0.956**	0.944**	0.951*	0.957**	0.952*
		Poisson-like	0.938**	0.958**	0.947*	0.955*	0.941**	0.953*	0.961**	0.956**
		Football-shape	0.971**	0.966**	0.970**	0.972**	0.970**	0.973**	0.974**	0.964**
	Exponential	Homoscedastic	0.955*	0.957**	0.932**	0.915	0.959**	0.968**	0.954*	0.927**
		Poisson-like	0.944**	0.946*	0.946*	0.955*	0.938**	0.953*	0.958**	0.950*
		Football-shape	0.970**	0.964**	0.947*	0.934**	0.971**	0.960**	0.963**	0.927**
Poisson	Linear	Homoscedastic	0.900	0.913	0.923	0.887	0.894	0.921	0.931**	0.892
		Poisson-like	0.886	0.926**	0.919	0.834	0.903	0.933**	0.912	0.834
		Football-shape	0.883	0.885	0.901	0.832	0.863	0.910	0.878	0.861
	Exponential	Homoscedastic	0.880	0.922	0.925**	0.854	0.876	0.945*	0.914	0.853
		Poisson-like	0.944**	0.932**	0.902	0.855	0.933**	0.927**	0.914	0.858
		Football-shape	0.840	0.879	0.880	0.835	0.847	0.865	0.874	0.837

* Meets Bradley's (1978) stringent criterion

** Meets Bradley's (1978) liberal criterion

Table 13. CI coverage for Wald test in binary predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100				*n* = 250			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.998	0.999	0.999	0.997	0.992	0.994	0.985	0.981
		Poisson-like	0.992	0.999	0.993	0.977	0.984	0.988	0.972**	0.960**
		Football-shape	0.996	0.999	1.000	1.000	0.972**	0.981	0.953*	0.939**
	Exponential	Homoscedastic	0.999	0.996	0.967**	0.847	0.967**	0.991	0.941**	0.908
		Poisson-like	1.000	0.993	0.878	0.883	0.981	0.988	0.838	0.909
		Football-shape	0.997	0.999	0.974**	0.892	0.970**	0.972**	0.911	0.919
Linear	Linear	Homoscedastic	0.942**	0.948*	0.940**	0.944**	0.957**	0.944**	0.949*	0.945*
		Poisson-like	0.947*	0.946*	0.928**	0.949*	0.937**	0.956**	0.951*	0.951*
		Football-shape	0.944**	0.958**	0.942**	0.936**	0.954*	0.948*	0.951*	0.940**
	Exponential	Homoscedastic	0.953*	0.943**	0.949*	0.909	0.940**	0.963**	0.942**	0.935**
		Poisson-like	0.944**	0.931**	0.936**	0.922	0.957**	0.941**	0.946*	0.921
		Football-shape	0.936**	0.943*	0.945*	0.920	0.942**	0.946*	0.948*	0.944**
Ordinal	Linear	Homoscedastic	0.945*	0.958**	0.955*	0.966**	0.959**	0.946*	0.960**	0.963**
		Poisson-like	0.951*	0.961**	0.956**	0.971**	0.949*	0.954*	0.955*	0.951*
		Football-shape	0.949*	0.962**	0.949*	0.940**	0.950*	0.959**	0.944**	0.949*
	Exponential	Homoscedastic	0.949*	0.946*	0.955*	0.960**	0.955*	0.968**	0.949*	0.948*
		Poisson-like	0.945*	0.946*	0.960**	0.959**	0.950*	0.946*	0.962**	0.945*
		Football-shape	0.946*	0.944**	0.946*	0.952*	0.940**	0.951*	0.952*	0.958**
Poisson	Linear	Homoscedastic	0.905	0.905	0.865	0.841	0.908	0.899	0.894	0.826
		Poisson-like	0.917	0.921	0.868	0.807	0.911	0.920	0.869	0.810
		Football-shape	0.850	0.856	0.796	0.752	0.815	0.837	0.798	0.764
	Exponential	Homoscedastic	0.877	0.901	0.904	0.828	0.878	0.923	0.882	0.847
		Poisson-like	0.947*	0.899	0.900	0.849	0.959**	0.909	0.890	0.829
		Football-shape	0.780	0.831	0.813	0.756	0.768	0.848	0.813	0.767

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion

Table 13 continued. CI coverage for Wald test in binary predictor conditions

Analysis	Mean Type	Variance Type	n = 500				n = 1000			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	0.956**	0.975**	0.972**	0.950*	0.920	0.957**	0.966**	0.933**
		Poisson-like	0.823	0.935**	0.943**	0.909	0.612	0.827	0.940**	0.883
		Football-shape	0.913	0.967**	0.924	0.866	0.909	0.979	0.952*	0.810
	Exponential	Homoscedastic	0.908	0.972**	0.893	0.903	0.839	0.954*	0.894	0.920
		Poisson-like	0.949*	0.929**	0.773	0.931**	0.931**	0.850	0.732	0.938**
		Football-shape	0.993	0.923	0.830	0.933**	0.999	0.970**	0.821	0.942**
Linear	Linear	Homoscedastic	0.953*	0.941*	0.950*	0.953*	0.949*	0.952*	0.949*	0.950*
		Poisson-like	0.942**	0.944**	0.955*	0.953*	0.949*	0.955*	0.956**	0.958**
		Football-shape	0.951*	0.937**	0.959**	0.948*	0.956**	0.948*	0.951*	0.930**
	Exponential	Homoscedastic	0.956**	0.953*	0.947*	0.943**	0.950*	0.949*	0.956**	0.930**
		Poisson-like	0.952*	0.952*	0.949*	0.950*	0.952*	0.941**	0.939**	0.930**
		Football-shape	0.966**	0.959**	0.951*	0.938**	0.955*	0.963**	0.954*	0.943**
Ordinal	Linear	Homoscedastic	0.953*	0.946*	0.947*	0.969**	0.954*	0.951*	0.940**	0.966**
		Poisson-like	0.941**	0.946*	0.953*	0.954*	0.947*	0.956**	0.955*	0.955*
		Football-shape	0.951*	0.939**	0.952*	0.953*	0.959**	0.956**	0.953*	0.943**
	Exponential	Homoscedastic	0.960**	0.950*	0.952*	0.963**	0.952*	0.949*	0.960**	0.955*
		Poisson-like	0.956**	0.948*	0.955*	0.952*	0.949*	0.946*	0.953*	0.951*
		Football-shape	0.955*	0.948*	0.955*	0.951*	0.962**	0.959**	0.952*	0.945*
Poisson	Linear	Homoscedastic	0.905	0.882	0.885	0.822	0.915	0.898	0.882	0.828
		Poisson-like	0.897	0.911	0.867	0.812	0.915	0.924	0.881	0.809
		Football-shape	0.844	0.825	0.814	0.780	0.852	0.824	0.828	0.775
	Exponential	Homoscedastic	0.869	0.900	0.879	0.845	0.864	0.908	0.892	0.832
		Poisson-like	0.955*	0.906	0.892	0.842	0.954*	0.908	0.869	0.838
		Football-shape	0.774	0.825	0.823	0.757	0.803	0.851	0.820	0.754

* Meets Bradley's (1978) stringent criterion

** Meets Bradley's (1978) liberal criterion

Table 14. CI coverage for LR test in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100				*n* = 250			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Exponential	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Linear	Linear	Homoscedastic	0.939**	0.959**	0.980	0.957**	0.946*	0.961**	0.972**	0.943**
		Poisson-like	0.950*	0.952*	0.955*	0.903	0.939**	0.955*	0.947*	0.900
		Football-shape	0.969**	0.965**	0.972**	0.955*	0.956**	0.965**	0.978	0.968**
	Exponential	Homoscedastic	0.950*	0.946*	0.899	0.754	0.956**	0.960**	0.889	0.780
		Poisson-like	0.949*	0.952*	0.878	0.781	0.947*	0.941**	0.883	0.769
		Football-shape	0.960**	0.964**	0.923	0.840	0.974**	0.970**	0.903	0.822
Ordinal	Linear	Homoscedastic	0.945*	0.949*	0.939**	0.960**	0.945*	0.949*	0.960**	0.956**
		Poisson-like	0.956**	0.945*	0.949*	0.939**	0.944**	0.949*	0.947*	0.950*
		Football-shape	0.963**	0.967**	0.969**	0.951*	0.960**	0.964**	0.969**	0.964**
	Exponential	Homoscedastic	0.955*	0.938**	0.933**	0.913	0.954*	0.951*	0.939**	0.918
		Poisson-like	0.950*	0.948*	0.952*	0.937**	0.947*	0.941**	0.946*	0.941**
		Football-shape	0.955*	0.964**	0.957**	0.919	0.964**	0.976	0.951*	0.942**
Poisson	Linear	Homoscedastic	0.898	0.912	0.928**	0.909	0.904	0.918	0.938**	0.883
		Poisson-like	0.917	0.923	0.917	0.844	0.904	0.929**	0.914	0.851
		Football-shape	0.884	0.868	0.872	0.858	0.886	0.874	0.889	0.867
	Exponential	Homoscedastic	0.895	0.923	0.929**	0.879	0.888	0.924	0.917	0.888
		Poisson-like	0.953*	0.918	0.906	0.891	0.947*	0.897	0.921	0.851
		Football-shape	0.830	0.881	0.874	0.848	0.829	0.881	0.875	0.840

* Meets Bradley's (1978) stringent criterion

** Meets Bradley's (1978) liberal criterion

N/A indicates that the LR confidence interval was not available for beta regression.

Table 14 continued. CI coverage for LR test in continuous predictor conditions

Analysis	Mean Type	Variance Type	*n* = 500				*n* = 1000			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Exponential	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Linear	Linear	Homoscedastic	0.948*	0.951*	0.970**	0.938**	0.942**	0.960**	0.963**	0.941**
		Poisson-like	0.941**	0.956**	0.950*	0.926**	0.945*	0.960**	0.944**	0.910
		Football-shape	0.972**	0.969**	0.975**	0.965**	0.968**	0.975**	0.975**	0.965**
	Exponential	Homoscedastic	0.944**	0.961**	0.888	0.755	0.943**	0.967**	0.891	0.776
		Poisson-like	0.945*	0.950*	0.859	0.764	0.938**	0.949*	0.860	0.806
		Football-shape	0.970**	0.958**	0.934**	0.834	0.975**	0.976	0.912	0.833
Ordinal	Linear	Homoscedastic	0.947*	0.938**	0.956**	0.952*	0.944**	0.950*	0.956**	0.950*
		Poisson-like	0.937**	0.953*	0.945*	0.956**	0.941**	0.952*	0.955*	0.954*
		Football-shape	0.971**	0.965**	0.971**	0.973**	0.970**	0.974**	0.975**	0.962**
	Exponential	Homoscedastic	0.955*	0.956**	0.934**	0.913	0.959**	0.968**	0.950*	0.926**
		Poisson-like	0.941**	0.947*	0.948*	0.953*	0.938**	0.954*	0.960**	0.949*
		Football-shape	0.969**	0.964**	0.947*	0.932**	0.971**	0.960**	0.960**	0.926**
Poisson	Linear	Homoscedastic	0.900	0.913	0.923	0.886	0.895	0.920	0.931**	0.892
		Poisson-like	0.887	0.926**	0.919	0.833	0.903	0.933**	0.912	0.834
		Football-shape	0.883	0.885	0.901	0.832	0.863	0.911	0.878	0.861
	Exponential	Homoscedastic	0.880	0.922	0.924	0.854	0.877	0.945*	0.914	0.852
		Poisson-like	0.944**	0.933**	0.902	0.855	0.933**	0.927**	0.914	0.858
		Football-shape	0.840	0.879	0.880	0.834	0.847	0.865	0.874	0.837

* Meets Bradley's (1978) stringent criterion

** Meets Bradley's (1978) liberal criterion

N/A indicates that the LR confidence interval was not available for beta regression.

Table 15. CI coverage for LR test in binary predictor conditions

Analysis	Mean Type	Variance Type	*n* = 100				*n* = 250			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Exponential	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Linear	Linear	Homoscedastic	0.943**	0.954*	0.942**	0.948*	0.957**	0.947*	0.950*	0.946*
		Poisson-like	0.950*	0.948*	0.933**	0.953*	0.942**	0.958**	0.951*	0.952*
		Football-shape	0.946*	0.959**	0.944**	0.940**	0.955*	0.948*	0.952*	0.940**
	Exponential	Homoscedastic	0.953*	0.945*	0.952*	0.910	0.942**	0.963**	0.943**	0.936**
		Poisson-like	0.945*	0.936**	0.938**	0.926**	0.957**	0.944**	0.948*	0.922
		Football-shape	0.940**	0.948*	0.947*	0.923	0.944**	0.950*	0.951*	0.945*
Ordinal	Linear	Homoscedastic	0.945*	0.952*	0.954*	0.953*	0.958**	0.945*	0.959**	0.960**
		Poisson-like	0.945*	0.956**	0.941**	0.948*	0.946*	0.952*	0.952*	0.949*
		Football-shape	0.949*	0.961**	0.949*	0.938**	0.950*	0.957**	0.946*	0.947*
	Exponential	Homoscedastic	0.948*	0.943**	0.949*	0.949*	0.955*	0.966**	0.949*	0.949*
		Poisson-like	0.942**	0.943**	0.953*	0.958**	0.950*	0.945*	0.963**	0.941**
		Football-shape	0.943**	0.943**	0.938**	0.948*	0.940**	0.951*	0.951*	0.960**
Poisson	Linear	Homoscedastic	0.905	0.905	0.862	0.838	0.908	0.898	0.894	0.827
		Poisson-like	0.915	0.921	0.866	0.812	0.910	0.920	0.868	0.812
		Football-shape	0.850	0.858	0.792	0.757	0.815	0.838	0.799	0.766
	Exponential	Homoscedastic	0.873	0.902	0.903	0.824	0.876	0.922	0.882	0.846
		Poisson-like	0.945*	0.899	0.900	0.851	0.958**	0.910	0.888	0.827
		Football-shape	0.777	0.831	0.810	0.756	0.766	0.847	0.815	0.769

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
N/A indicates that the LR confidence interval was not available for beta regression.

Table 15 continued. CI coverage for LR test in binary predictor conditions

Analysis	Mean Type	Variance Type	*n* = 500				*n* = 1000			
			0	0.35	0.87	1.39	0	0.35	0.87	1.39
Beta	Linear	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Exponential	Homoscedastic	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Poisson-like	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Football-shape	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Linear	Linear	Homoscedastic	0.953*	0.943**	0.950*	0.953*	0.950*	0.952*	0.950*	0.950*
		Poisson-like	0.942**	0.944**	0.956**	0.953*	0.949*	0.956**	0.956**	0.958**
		Football-shape	0.951*	0.937**	0.959**	0.949*	0.956**	0.948*	0.951*	0.930**
	Exponential	Homoscedastic	0.957**	0.955*	0.947*	0.943**	0.950*	0.949*	0.956**	0.930**
		Poisson-like	0.952*	0.952*	0.950*	0.951*	0.952*	0.941**	0.939**	0.930**
		Football-shape	0.966**	0.959**	0.952*	0.939**	0.956**	0.963**	0.954*	0.943**
Ordinal	Linear	Homoscedastic	0.953*	0.946*	0.946*	0.968**	0.954*	0.950*	0.942**	0.967**
		Poisson-like	0.939**	0.947*	0.952*	0.951*	0.946*	0.956**	0.955*	0.955*
		Football-shape	0.951*	0.939**	0.954*	0.951*	0.959**	0.956**	0.954*	0.944**
	Exponential	Homoscedastic	0.960**	0.949*	0.953*	0.962**	0.952*	0.950*	0.959**	0.956**
		Poisson-like	0.956**	0.950*	0.957**	0.953*	0.948*	0.947*	0.951*	0.950*
		Football-shape	0.954*	0.948*	0.955*	0.949*	0.962**	0.959**	0.950*	0.943**
Poisson	Linear	Homoscedastic	0.905	0.881	0.885	0.823	0.915	0.898	0.881	0.827
		Poisson-like	0.897	0.910	0.867	0.810	0.915	0.924	0.881	0.807
		Football-shape	0.844	0.826	0.814	0.778	0.851	0.824	0.829	0.774
	Exponential	Homoscedastic	0.869	0.900	0.878	0.842	0.864	0.909	0.892	0.831
		Poisson-like	0.954*	0.907	0.892	0.837	0.953*	0.908	0.869	0.838
		Football-shape	0.774	0.825	0.824	0.756	0.803	0.851	0.820	0.753

* Meets Bradley's (1978) stringent criterion
** Meets Bradley's (1978) liberal criterion
N/A indicates that the LR confidence interval was not available for beta regression.

Figure 1. Relationship between probability and logit.

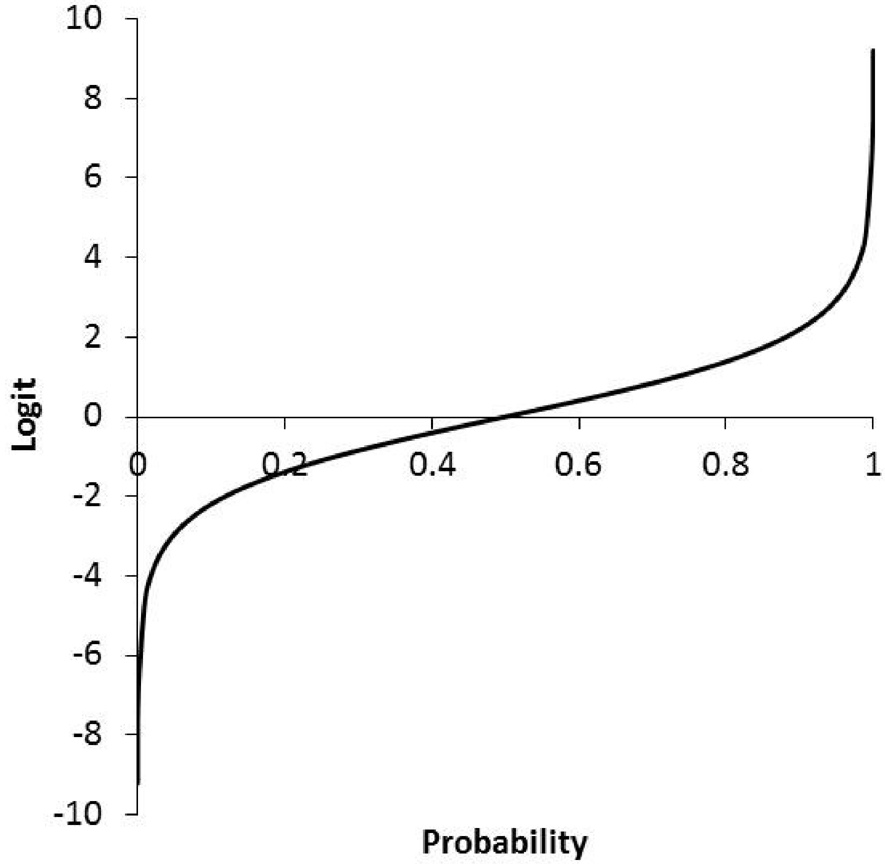
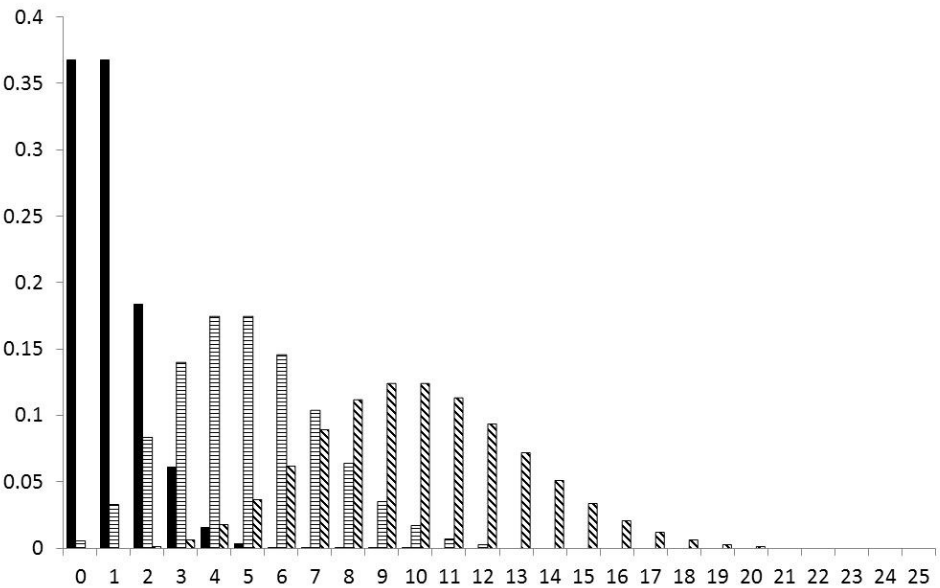


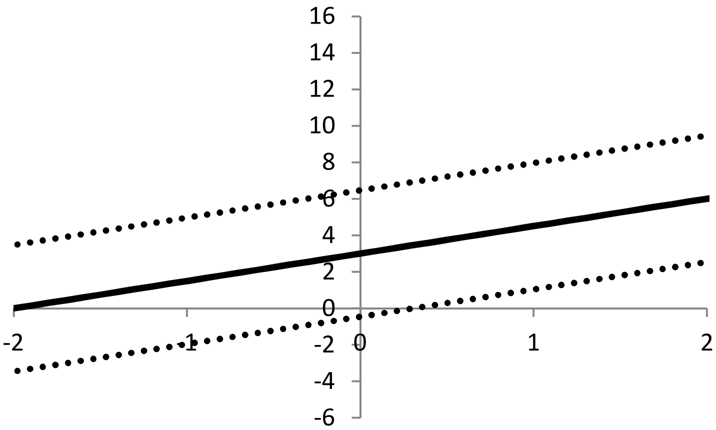
Figure 2. Poisson distributions with means of 1, 5, and 10.



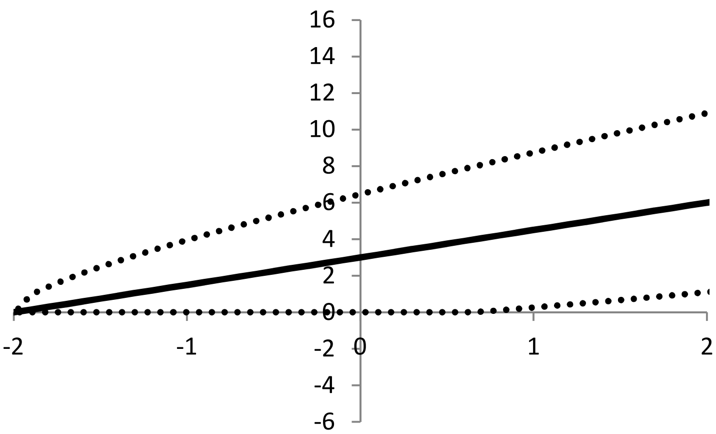
Note: Poisson distributions with mean = 1 (solid bars), mean = 5 (horizontally striped bars), and mean = 10 (diagonally striped bars) are shown.

Figure 3. Linear mean structure with variance structures.

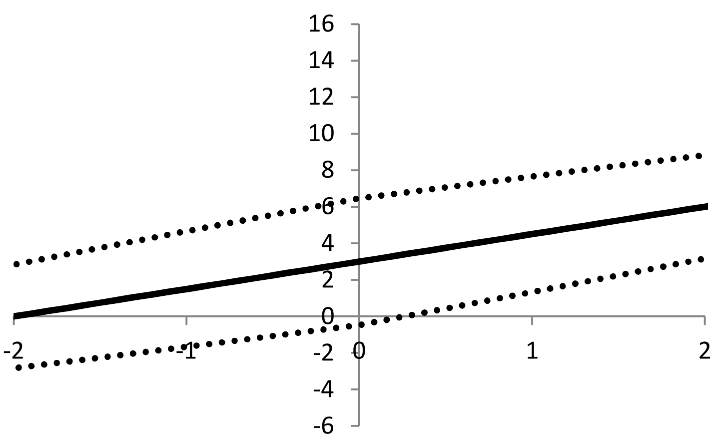
Homoscedastic variance



Heteroscedastic, Poisson-like variance



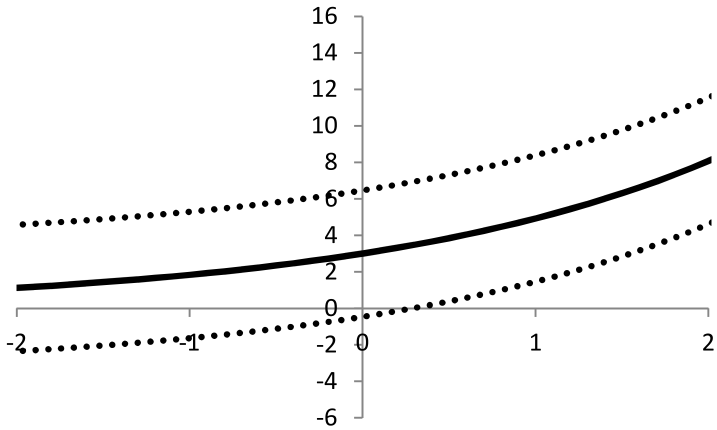
Heteroscedastic, football-shaped variance



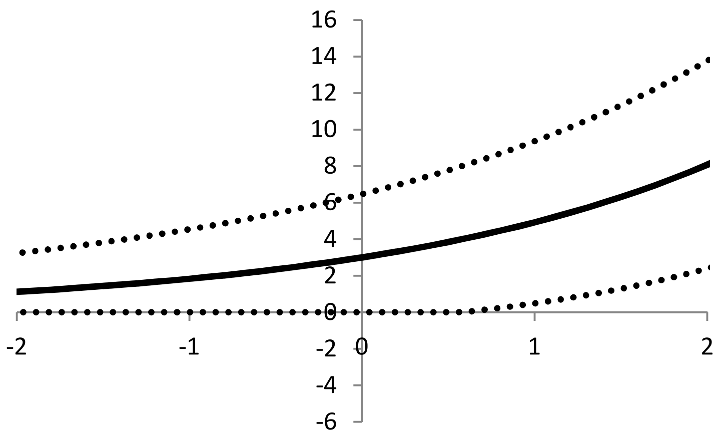
Bold line = mean structure. Dashed lines = 2 standard deviations above and below the mean, based on the variance structure indicated.

Figure 4. Exponential mean structure with variance structures.

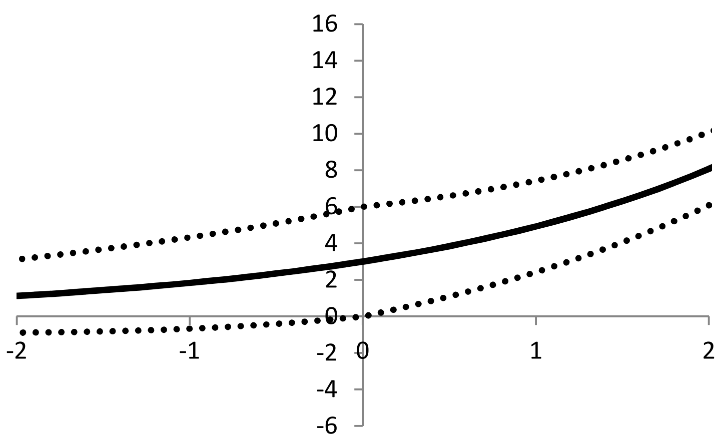
Homoscedastic variance



Heteroscedastic, Poisson-like variance



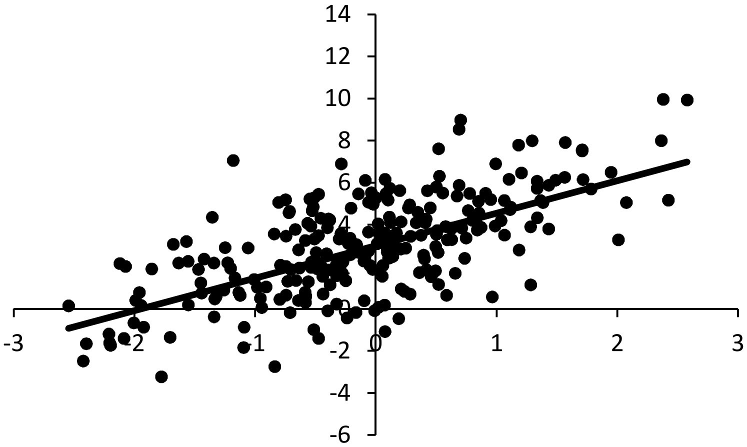
Heteroscedastic, football-shaped variance



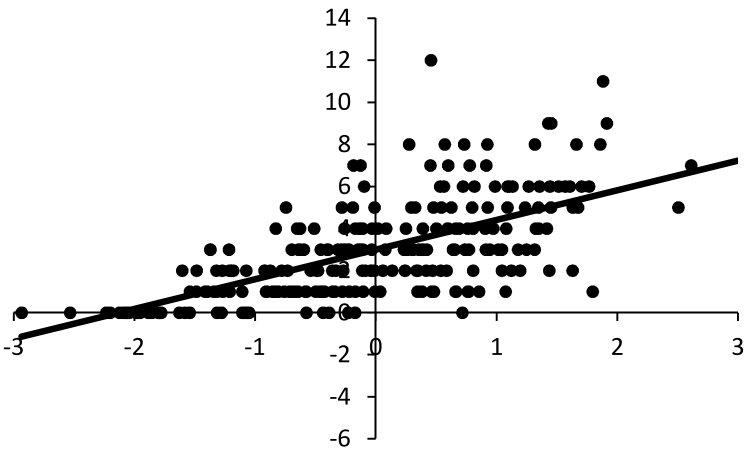
Bold line = mean structure. Dashed lines = 2 standard deviations above and below the mean, based on the variance structure indicated.

Figure 5. Representative samples for linear conditions (n=250).

Homoscedastic variance



Heteroscedastic, Poisson-like variance



Heteroscedastic, football-shaped variance

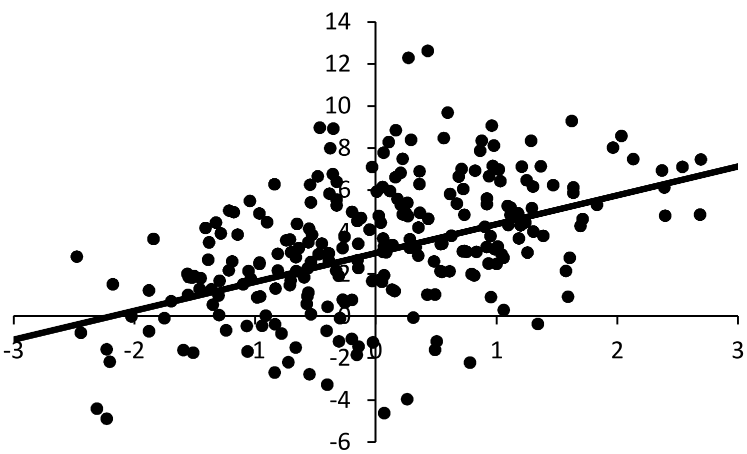
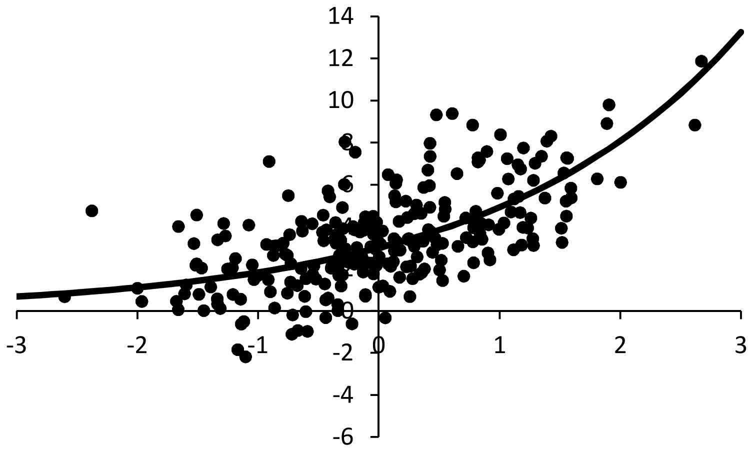
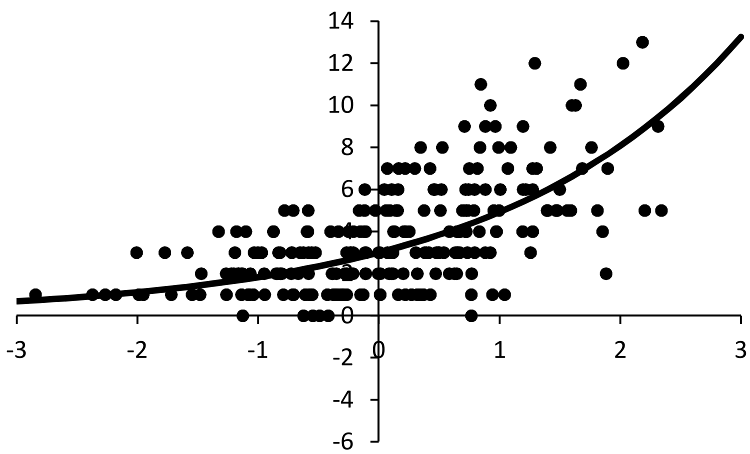


Figure 6. Representative samples for exponential conditions (n=250).

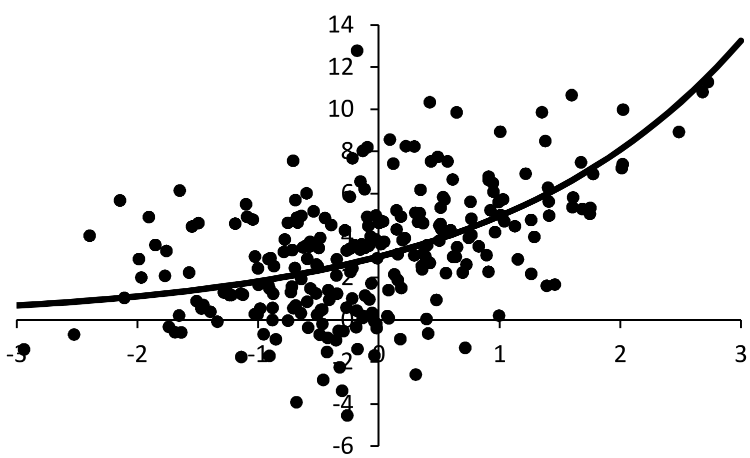
Homoscedastic variance



Heteroscedastic, Poisson-like variance



Heteroscedastic, football-shaped variance



References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley-Interscience.

Allison, P. D., & SAS Institute. (1999). *Logistic regression using the SAS system: Theory and application*. Cary, N.C.: SAS Institute.

Brehm, J., & Gates, S. (1993). Donut shops and speed traps: Evaluating models of supervision on police behavior. *American Journal of Political Science*, 37, 555-581.

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

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

Chinn, S. (2001). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19, 3127-3131.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation rating scale for Mexican Americans-II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Sciences*, 17, 275-304.

Daniel, W. W. (1990). *Applied nonparametric statistics* (2^{nd} ed.). Pacific Grove, CA: Duxbury.

DeMaris, A. (2002). Explained variance in logistic regression: A Monte Carlo study of proposed measures. *Sociological Methods and Research*, 31, 27-74.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, 31, 799-815.

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.

Hauck, W. W. & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, 72, 851-853.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.

Hintze, J. (2011). PASS 11 [computer software]. Kaysville, UT: NCSS, LLC.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Hsieh, F. Y. (1989). Sample size tables for logistic regression. *Statistics in Medicine*, 8, 795-802.

Johnston, L. D., Bachman, J. G., & O'Malley, P. M. (2003). Monitoring the future: Questionnaire responses from the nation's high school seniors. Ann Arbor, MI: The University of Michigan.

Kieschnick, R. & McCullough, B. D. (2003). Regression analysis of variates observed on (0,1): Percentages, proportions, and fractions. *Statistical Modeling*, 3, 193-213.

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

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, 5, 434-458.

Maxwell, S. E., Kelley, K. & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537-563.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London; New York: Chapman and Hall.

Menard, S. (2001). *Applied logistic regression analysis* (2nd ed.) Newbury Park, CA: Sage.

Nagin, D. S. (1997). Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods*, 4, 139-157.

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

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Chicago: Irwin.

Newson, R. (2004). Generalized power calculation for generalized linear models and more. *The Stata Journal*, 4, 379-401.

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, 9, 325-346.

Ryan, T. P. (1997). *Modern regression methods*. New York: Wiley.

SAS Institute (2008). SAS 9.2 [computer software]. Cary, NC.

Self, S. G., Mauritsen, R. H., & Ohara, J. (1992). Power calculations for likelihood raiot tests in generalized linear models. *Biometrics*, 31-39.

Schaefer, E. S. (1965). Children's reports of parental behavior: An inventory. *Child Development*, 36, 417-424.

Shieh, G. (2000a). A comparison of two approaches for power and sample size calculations in logistic regression models. *Communications in Statistics – Simulation and Computation*, 29, 763-791.

Shieh, G. (2000b). On power and sample size calculations for likelihood ratio tests in generalized linear models. *Biometrics*, 56, 1192-1196.

Shieh, G. (2001). Sample size calculations for logistic and Poisson regression models. *Biometrika*, 88, 1193-1199.

Signorini, D. F. (1991). Sample size for Poisson regression. *Biometrika*, 78, 446-450.

Smithson, M., & Verkuilen, J. (2008). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11, 54-71.

Snapinn, S. M. & Small, R. D. (1986). Tests of significance using regression models for ordered categorical data. *Biometrics*, 42, 583-592.

Strickland, P. A. O. & Lin, S.-E. (2003). Estimates, power and sample size calculations for two-sample ordinal outcome under before-after study designs. *Statistics in Medicine*, 22, 1807-1818.

Tsonaka, R., Rizopoulos, D. & Lesaffre, E. (2006). Power and sample size calculations for discrete bounded outcome scores. *Statistics in Medicine*, 25, 4241-4552.

Vaeth, M. (1985). On the use of Wald's test in exponential families. *International Statistical Review*, 53, 199–214.

West, S. G., Aiken, L. S., Kwok, O. (August, 2003). *Hierarchical and generalized linear models: What multiple regression didn't tell us*. Invited address, American Psychological Association.

Whittemore, A. S. (1981). Sample size for logistic regresion with small response probability. *Journal of the American Statistical Association*, 76, 27-32.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9, 60-62.

Footnotes

[1] A programming error led to effect sizes that were different from the intended effect sizes. The intended effect sizes were 0, 0.2, 0.5, and 0.8, corresponding to the standard zero, small, medium, and large Cohen's *d* effect sizes. The actual effect sizes in this study were 0, 0.35, 0.87, and 1.39, corresponding to zero, small/medium, large, and very large Cohen's *d* effect sizes.

APPENDIX A

DATA GENERATION SYNTAX

```
/*********************************************/
/* 2 types of predictor - continuous/binary     */
/* 2 mean structures - linear/exponential       */
/* 4 variance structures - OLS/Poisson/football */
/* 4 effect sizes - 0/0.2/0.5/0.8               */
/* 4 sample sizes - 100/250/500/1000            */
/*********************************************/

/*********************************************/
/* Values for macro variables                   */
/* xtype = 1 for continuous, 2 for binary       */
/* meantype = 1 for linear, 2 for exponential   */
/* vartype = 1 for OLS, 2 for Poisson,          */
/*           3 for football                     */
/* effsize = 0 for 0, 0.2 for 0.2, 0.5 for 0.5, */
/*           0.8 for 0.8                        */
/* sampsize = 100, 250, 500, 1000               */
/*********************************************/

options symbolgen;

%macro gcgf(reps, xtype, meantype, vartype, effsize, sampsize);

%do i = 1 %to &reps;

data one;
keep reps xtype meantype vartype effsize sampsize i absx x mu_x y x_star
y_star;
i = &i;
reps = &reps;
xtype = &xtype;
meantype = &meantype;
vartype = &vartype;
effsize = &effsize;
sampsize = &sampsize;
/* generate x as a normal variate */
do j = 1 to &sampsize;
x_star = rannor(0);
/* generate mean structure of y */
if meantype = 1 then
mu_x = (3 * &effsize * x_star) + 3;
if meantype = 2 and effsize = 0 then
mu_x = exp(0 * x_star);
if meantype = 2 and effsize = 0.2 then
mu_x = exp((0.199668 * x_star) + 1.09861229);
if meantype = 2 and effsize = 0.5 then
mu_x = exp((0.494933 * x_star) + 1.09861229);
else if meantype = 2 and effsize = 0.8 then
mu_x = exp((0.780071 * x_star) + 1.09861229);
if vartype = 2 and mu_x < 0 then mu_x = 0;
/* add variance structure to y */
absx = abs(x_star-0);
if vartype = 1 then
y_star = mu_x + sqrt(3) * rannor(0);
```

```
if vartype = 2 then
y_star = ranpoi(0, mu_x);
else if vartype = 3 then
y_star = mu_x + (((-0.5 * absx) + 3) * rannor(0));
/* chop up y according to these categories:     */
/* 0, 1-3, 4-8, 9-15, 16-30                      */
if y_star le 0.4999999999 then y = 0;
if 0.5 le y_star le 3.4999999999 then y = 2;
if 3.5 le y_star le 8.4999999999 then y = 6;
if 8.5 le y_star le 15.4999999999 then y = 12;
if y_star ge 15.5 then y = 23;
output;
end;
run;

/* for the binary x condition, do a median split of x */
%if &xtype = 1 %then
%do;
data one;
set one;
x= x_star;
run;
%end;
%else %if &xtype = 2 %then
%do;
proc sort data = one;
by x_star;
run;
data one;
set one;
if _n_ le &sampsize/2 then x = -0.5;
else x = 0.5;
run;
%end;

/* Output all data to a file */
data write;
set one;
file "C:\Users\psyripl\Desktop\Dissertation programs\Conditions 1 through
48\estimates.txt" mod;
put reps xtype meantype vartype effsize sampsize i x y absx mu_x x_star
y_star;
run;

%end;
%mend gcgf;
```