

5-2016

# Generalized Linear Model Analyses for Treatment Group Equality when Data are Non-Normal

Harvey J. Kesleman

*University of Manitoba, Harvey.Keselman@umanitoba.ca*

Abdul R. Othman

*Universiti Sains Malaysia*

Rand R. Wilcox

*University of Southern California*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Kesleman, Harvey J.; Othman, Abdul R.; and Wilcox, Rand R. (2016) "Generalized Linear Model Analyses for Treatment Group Equality when Data are Non-Normal," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 1 , Article 4.  
DOI: 10.22237/jmasm/1462075380

## ***Invited Article***

# **Generalized Linear Model Analyses for Treatment Group Equality when Data are Non-Normal**

**H. J. Keselman**

University of Manitoba  
Winnipeg, Manitoba

**Abdul R. Othman**

Universiti Sains Malaysia  
Penang, Malaysia

**Rand R. Wilcox**

University of Southern California  
Los Angeles, CA

---

One of the validity conditions of classical test statistics (e.g., Student's  $t$ -test, the ANOVA and MANOVA  $F$ -tests) is that data be normally distributed in the populations. When this and/or other derivational assumptions do not hold, the classical test statistic can be prone to too many Type I errors (i.e., falsely rejecting too often) and/or have low power (i.e., failing to reject when the null hypothesis is false) to detect treatment effects when they are present. However, alternative procedures are available for assessing equality of treatment group effects when data are non-normal. For example, researchers can use robust estimators instead of the usual least squares estimators to test that treatment effects are equivalent across groups. As well, recent advances in statistical methodology allow researchers to test for equality of treatment group effects by assuming other distributional shapes for the data. One class of such analyses is generalized linear model techniques. On the other hand, researchers can adopt sequential analyses where they first assess the normality assumption and then depending on the result determine the type of analysis that should be adopted. The purpose of the present study was to compare the above approaches for assessing equality of treatment group effects in the presence of non-normal data. Simulation results which were based on various non-normal distributions and the values of group variances and sample sizes revealed that sequential analysis coupled with a generalized linear model solution were just as prone to inflated or depressed rates of Type I error as the classical ANOVA  $F$ -test.

*Keywords:* Tests for equality of treatment effects, non-normal data, multi-group problem, goodness-of-fit statistic, skewed and kurtotic data, 5-point Likert data, familywise Type I error control

---

---

*Dr. Keselman is Professor Emeritus in the Department of Psychology. Email him at: [Harvey.Keselman@umanitoba.ca](mailto:Harvey.Keselman@umanitoba.ca). Dr. Othman is a Professor of Statistics in the School of Distance Education. Dr. Wilcox is a Professor of Psychology.*

## Introduction

Classical procedures that test for the equality of treatment effects across independent groups [e.g., Student's Two Independent Sample t-test, the analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) *F*-tests] provide reliable and valid results when their derivational assumptions hold or nearly hold. All of the aforementioned procedures require that (1) the observations are independent of one another (the independence of observations assumption), (2) the data are distributed normally in each of the treatment populations (the univariate and multivariate normality assumption), and (3) the population variances/covariance matrices are equal across treatment groups (the homogeneity of variances/covariance matrices assumption). When these derivational assumptions hold, the probabilities associated with performing a test of significance are exact; that is, the probabilities of making a Type I error (falsely rejecting the null hypothesis of treatment group equality) and Type II error (falsely accepting the null hypothesis) are known to the researcher. As well, the traditional effect size (ES) statistics and their confidence intervals (CIs) would be valid also.

However, when the assumptions do not hold, either individually or jointly, the actual Type I error probability, as well as the Type II probability, will not be equal to their intended values (e.g., say 0.05 and 0.20, respectively). The consequence is that researchers no longer know the probabilities of inferential error associated with their test of significance for treatment group equality. That is, these errors of inference can be divergently different than those set by the researcher. For example, the probability of committing a Type I error, which is traditionally set at 0.05, could in actuality be 0.60. A consequence of this would be that, when the null hypothesis of treatment group equality has been rejected, researchers would not know whether the hypothesis has been rejected erroneously because the probability of a Type I error is much larger than was set, as a result of violating one or more of the derivational assumptions, or whether the hypothesis was rejected because the population treatment effects are truly not equal across treatment groups. Clearly controlling the probabilities of statistical inference (i.e., Type I and II errors) is paramount when using inferential statistical procedures to assess the credibility of research hypotheses. Moreover the values for ES statistics and their CIs could be inaccurate as well.

Based on the preceding, researchers in the behavioral, biological, and health sciences are encouraged to assess whether derivational assumptions hold prior to using a test of significance (e.g., see Lix & Keselman, 2009; Lix, Keselman, & Hinds, 2005; Kirk, 2013). Indeed, it is well known that psychological, biological,

## GENERALIZED LINEAR MODEL ANALYSES

and medical data are rarely normal in form (e.g., see Lix & Keselman, 2009; Micceri, 1989). For example, many psychologists collect reaction time data; this data typically will be positively skewed (not symmetric as would be the case with a normal distribution) because of large outlying values in the right hand tail of the distribution indicating large reaction times for some subjects.<sup>1</sup> As well, Likert-type data is frequently collected in social science and medical research and such data is clearly non-normal.

With regard to the assumption that data in the treatment populations are normally distributed, researchers can use a test for normality (e.g., Shapiro-Wilk, Kolmogorov-Smirnov – see D’Agostino & Stephens, 1986; SAS Institute, Inc., 2010b, p. 357). If the results of the preliminary test indicate that the empirical data in each treatment group conforms to a theoretical normal distribution, researchers can go on to test for equality of treatment group effects with the *t*- or *F*-test (assuming that the other assumptions are examined and believed to be true as well); additionally, the size of ES statistics and their CIs would be accurate and meaningful. However, if the result of the test for normality indicates the empirical data are not normally distributed within treatment groups, researchers must take remedial action.

Until recently, the choices available to researchers for dealing with non-normal data were to: (1) transform their data to achieve normality (see Box & Cox, 1964), (2) use a procedure that does not require the data to be normally distributed (a rank transformation test – see Akritas, Arnold, & Brunner, 1997), or (3) use a procedure (Welch-James –  $WJ_t$ ) that can perform accurately in the presence of non-normality (e.g., procedures that use robust estimators, such as trimmed means and Winsorized variances, rather than the usual least squares estimators in the test statistic) (see Erceg-Hurn, Wilcox, & Keselman, 2013; Keselman, Algina, Lix, Wilcox, & Deering, 2008a; 2008b). Prior empirical research indicates that these alternative methods can be quite successful in controlling Type I and Type II errors, and accurate in calculating ES statistics and setting CIs around them when data are non-normal in the treatment populations (see Keselman et al., 2008a; 2008b; Lix & Keselman, 1998). Nonetheless, some of the better approaches [e.g., alternative (3)] are not available in the major statistical packages and therefore are not typically used by researchers (Keselman et al., 2008b have provided a software program to implement the  $WJ_t$  procedure).

However, now researchers have methods for dealing with non-normal distributions when testing hypotheses via generalized linear models (i. e., GLM – see Stroup, 2013). In particular, the SAS Institute, Inc. (2010a; 2010b) and PASW (SPSS, Inc., 2009) systems of statistical programs allows users to conduct tests for

equality of treatment group effects where the data need not be presumed to be normal in form (see e.g. the SAS GLMMIX procedure), through generalized linear models (GLMs) (see Breslow & Clayton, 1993; Nelder & Wedderburn, 1972; Stroup, 2013). Indeed, GLMs allow users to perform tests for treatment group equality for data that can be lognormal, exponential, beta, or gamma, etc. distributed. Furthermore, researchers can even adopt a GLM (e.g., see SASs GLIMMIX procedure) that specifies that data in the treatment groups has varied forms; e.g., in some groups the data are normally distributed while in others they are exponentially distributed. Clearly, this approach for dealing with non-normal data could provide researchers with a remarkably reliable and valid way of testing for treatment group equality where data are not normal in form. However, researchers typically do **not** know how their data are distributed in the parent populations and, accordingly, must rely on (statistical) methods to determine the appropriate shape of their data.<sup>2</sup> The two most popular statistical packages employed by behavioral sciences and biological researchers (e.g., the SAS and SPSS systems), in addition to providing appropriate plots of the empirical data (e.g., normality probability plots, box-plots, etc.), provide users with test statistics that can be used to test for normality and tests that examine the fit of the data to various theoretical distributions.

Accordingly, whether GLM analyses will work well when adopting this two stage strategy will depend, in part, on how good these preliminary tests for normality and fit perform. Evidence regarding the accuracy of these tests is varied (e.g., see Dufour, Farhat, Gardiol, & Khalaf, 2010; Kowalchuk, Keselman, Wilcox, & Algina, 2006; Keselman, Othman, & Wilcox, 2013; 2014; Rochon & Kieser, 2011; Schoder, Himmelmann, & Wilhelm, 2006). For example, in the one-group pretest-posttest-design, Schoder et al. (2006) found that the Kolmogorov-Smirnov (K-S) goodness of fit test did not have enough power to detect non-normal distributions unless sample size was greater than 100. On the other hand, in the same design, Keselman et al. (2014) found that the Anderson-Darling (A-D) goodness of fit test did have sufficient power ( $\geq 0.80$ ) if researchers set the level of significance at values greater than 0.05 (e.g.,  $\alpha = 0.15$  or  $\alpha = 0.20$ ). Again comparing goodness of fit statistics for detecting non-normal distributions in the multi-group (3) problem, Keselman et al. (2014) also found that the A-D test provided good power to detect non-normal distributions even while adopting familywise Type I error control over the three tests of normality with overall significance controlled at 0.15 and 0.20.

Nonetheless, perhaps the most crucial aspect of adopting GLMs is the fact that, in order to use them appropriately (correctly) (e.g., the GLIMMIX procedure),

## GENERALIZED LINEAR MODEL ANALYSES

researchers must correctly specify the ‘link function’ which will determine how the analysis is to be conducted in order to properly assess equality of treatment group effects; a misspecification of the ‘link function’ results in the wrong hypothesis being tested and consequently incorrect results (see Stroup, 2013, Chapter 3).<sup>3</sup>

Thus it must be determined whether testing for normality is a good preliminary strategy prior to assessing treatment group equality. It could be the case that uniformly adopting the ANOVA  $F$ -test, GLM, or  $WJ_t$  might be a better strategy than choosing a method of analysis based on a preliminary test for normality or fit. As well, if users incorrectly specify the link function results will be incorrect (Cerrito, 2005; Stroup, 2013). That is, we do not know whether this approach to handling non-normal data results in better tests for equality of treatment group effects compared to the previously enumerated methods that are available to researchers. Consequently, the purpose of this investigation is to examine this question.

### Methods

A simulation study was conducted to examine GLM analyses when dealing with a single factor one-way univariate design having three groups. Researchers can adopt various GLMs in two ways: One approach first examines the empirical data and compares this data to theoretical distributions, e.g., the normal distribution. Based on this preliminary test, researchers then go on to use GLMs (e.g., GLIMMIX), specifying the shape of the distribution that should be presumed to hold in the population treatment groups (e.g., normal, exponential, lognormal, etc.) through a link function. For example, users could test whether the data are normally distributed in each of the treatment groups and, if the test for normality (SAS provides four tests; users would typically select one – A-D) indicate the data are normally distributed, researchers would then go on to the usual linear model which assumes data are normally distributed (i.e., the ANOVA  $F$ -test) when it tests for equality of treatment group effects (SASs general linear model analysis through the General Linear Model Procedure). If, however, the test for normality does not indicate that the data are normally distributed within each treatment group, researchers then frequently try to determine how the data are distributed. They can accomplish this by examining whether the empirical data conform to other theoretical distributions (e.g., lognormal, exponential, gamma, etc.). If, for example, analyses indicates that the data are exponentially distributed by examining fit statistics (e.g., A-D, Cramer-von Mises (CvM), K-S – see D’Agostino & Stephens, 1986; SAS Institute, Inc., 2010b, p. 279) researchers would then go on to adopt a

GLM model analysis (e.g., GLIMMIX, and through its syntax, tell the program that the data is not normal but exponentially distributed). If one adopts the GLIMMIX procedure it will then perform a test of significance for treatment group equality with estimation procedures that presume the data are exponentially distributed. As indicated in the introduction, GLIMMIX even allows researchers in its estimation methodology to specify that the data are distributed in different forms within each of the treatment groups.

The second manner in which researchers can locate and specify the form of the non-normal data within GLM analyses is to perform a number of analyses specifying different shapes for the data and select the analysis that provides the best fit to the data [SAS allows users to select from a number of fit (Information Criteria) statistics, e.g., Akaike, Schwarz – see Keselman, Algina, Kowalchuk, & Wolfinger, 1998]. That is, within GLM analyses users can request the program to compute fit statistics that then enable users to compare different analyses which specify different shapes for the distribution of the data. Consequently, researchers can compare the fit statistics across the different analyses and pick the results (for the test of equality of treatment group effects) that provided the best fit statistic result.

Thus, GLM model analyses appear to be very versatile tools for examining equality of treatment group effects for data that are not normal in form. However, there is limited published information regarding the effectiveness of the procedure for dealing with non-normal data (see e.g., Schoder et al., 2006; Keselman et al., 2013; 2014); that is, the effectiveness of the procedures for dealing with non-normal data will depend, in part, on how good are the tests for normality, the Information Criteria, and fit in identifying the correct form of the distribution of the data (some published studies report unfavorable results – see e.g., Rochon & Kieser, 2011; Schoder et al., 2006). As a result, if the analyses/procedures cannot identify the correct shape for the distribution(s) of the data to be specified in the link function, GLM analyses will give erroneous conclusions with regard to the test for equality of treatment group effects. However, in many other instances encountered by applied researchers, the nature of the data collected will automatically determine the form of the analysis – the link function. For example, for Likert data, it is obvious that the distribution modelled is the multinomial distribution with cumulative logit as the link function (see SAS Institute, Inc., n.d., p. 64).<sup>4</sup>

Consequently, in our study, we manipulated: (1) the procedure used to assess the shape of a distribution and, in particular, we used the A-D test for normality and also employed the Akaike (1974) fit-statistic (available through the SAS system) (see Keselman, Algina, et al., 1998); (2) the shapes of distributions (15 *g*-and-*h* distributions, 6 contaminated normal mixture models, and 4 multinomial models);



## GENERALIZED LINEAR MODEL ANALYSES

(3) the sample sizes: (i) three groups of equal sizes, 20, 40, and 80 per group; (ii) three groups of unequal sizes with total sample size of 120. The  $g$ - and  $h$ -distributions were simulated for (i) and (ii) (Table 1). However, the contaminated mixed normal and the Likert scale distributions were simulated for total sample size 120 in (i) and (ii) (Tables 2 and 3), and (4) the level of significance for the A-D statistic (i.e.,  $\alpha = 0.15$  and  $0.20$ ).

Choices for non-normal distributions are modifications from Schoder et al. (2006), Zimmerman (2010), and Keselman et al. (2013; 2014). These authors investigated a normal distribution with a single outlier, a normal distribution with 10% of the data containing outliers, skewed distributions with varying skewness, and an ordinal 5-point Likert scale with varying multivariate probabilities (common they state in psychological and medical investigations).

Many non-normal distributions were investigated via  $g$ -and- $h$  distributions (see Headrick, Kowalchuk, & Sheng, 2008; Hoaglin, 1983; 1985; Kowalchuk & Headrick, 2010; Tukey, 1960). These distributions with their values for skewness and kurtosis are enumerated in Table 1. We chose a range of values of  $g$  and  $h$  to cover as broad a spectrum of non-normal distributions that could occur in medical, psychological, and behavioral science experiments (e.g., see Keselman, Huberty, et al., 1998; Micceri, 1989; Wilcox, 2012).

SAS Institute, Inc. (2013) was used to generate  $g$ -and- $h$  data. To generate data from a  $g$  and  $h$  distribution, standard unit normal variables  $Z_{ij}$  were converted to  $g$  and  $h$  distributed random variables via

$$Y_{ij} = \frac{\exp(gZ_{ij}) - 1}{g} \exp\left(\frac{hZ_{ij}^2}{2}\right)$$

when both  $g$  and  $h$  were non-zero. When  $g$  was zero,  $Y_{ij} = Z_{ij} \exp\left(\frac{hZ_{ij}^2}{2}\right)$ . The  $Z_{ij}$

scores were generated by using RANNOR from SAS (2010a). These equations generate symmetric ( $g = 0$ ) and asymmetric distributions ( $g \neq 0$ ), respectively. As Kowalchuk and Headrick (2010) noted, “The parameter  $\pm g$  controls the skew of a distribution in terms of both direction and magnitude. The parameter  $h$  controls the tail weight or elongation of a distribution and is positively related with kurtosis” (p. 63). As well, Type I error rates were investigated when data were obtained from a normal distribution [ $g = h = 0$ , the standard normal distribution (skewness and kurtosis = 0)].



A number of different contaminated mixed-normal distributions, such as those reported in Zimmerman (2010), were examined. Contaminated mixed-normal distributions have one or more outlying values that deviate from the central mean of the distribution by some amount measured in standard deviation units ( $D$ ). For example, Zimmerman (2010) examined a mixed-normal distribution consisting of samples from  $N(0, 1)$  with probability 0.95 and from  $N(0, 400)$  with probability 0.05. Tukey (1960) suggested that outliers are a common occurrence in distributions, and others have indicated that skewed distributions frequently depict psychological data (e.g., reaction time data). Accordingly, eight contaminated mixed-normal distributions were examined that had one, two, or four outlying values which were five or ten standard deviations from the mean value. These distributions are enumerated in Table 2.

Finally, like Schoder et al. (2006), and Keselman et al. (2013; 2014), a 5-point Likert scale was simulated; such data is frequently gathered in medical, psychological (e.g., from clinical, personality, and social psychology) and other behavioral science investigations. As well, there has been much discussion in the medical and social science literatures regarding the analysis of outcomes with 5-point Likert data and classical methods of analysis (see e.g., Roberson, Shema, Mundfrom, & Holmes, 1995; Jameson, 2004; Nanna & Sawilowsky, 1998). Consequently, the same conditions as Schoder et al. (2006) and Keselman et al. (2014) were investigated. Specifically, (1) an even distribution ( $p = .2$  for each category 1 – 5); (2) a symmetric distribution ( $p_1 = 0.1, p_2 = 0.2, p_3 = 0.4, p_4 = 0.2, p_5 = 0.1$ ); (3) a moderately skewed distribution ( $p_1 = 0.5, p_2 = 0.3, p_3 = 0.15, p_4 = 0.04, p_5 = 0.01$ ); and (4) a heavily skewed distribution ( $p_1 = 0.7, p_2 = 0.2, p_3 = 0.06, p_4 = 0.03, p_5 = 0.01$ ). Thus for the 5-point Likert scale data there were 4 multinomial distributions that were simulated (See Table 3).

We did not investigate the same sample size conditions as Schoder et al. (2006) but did include a reasonable range of values (i.e.,  $n = 20, 40, 80$ ) depending on the condition. Specifically,

- (i) For the 5  $g$ -and- $h$  distributions, sample sizes of 20, 40, and 80 were chosen.
- (ii) For 6 contaminated normal distributions,  $0.95N(0, 1) + 0.05N(0, k)$ ,  $0.9N(0, 1) + 0.1N(0, k)$ , and  $0.975N(0, 1) + 0.025N(0, k)$ ,  $k = 25, 100$ , a sample size of 40 was chosen.

Because in preliminary testing it would be important to guard against a Type II error (falsely accepting the null hypothesis that the data are normal in form), we

selected significance levels of 0.15, and 0.20 when testing assumptions (we believe that the 0.05 level would not be appropriate in preliminary testing). Because some might find that these levels of significance would be too large with multiple tests for normality (one for each group examined), familywise error control with a level of significance of 0.15 or 0.20 was adopted using Hochberg's (1988) sequentially rejective Bonferroni procedure, setting the overall rate of Type I error at either 0.15 or 0.20. Hochberg's procedure, as indicated, has been found to control the overall rate of Type I error over a set of statistical tests and provides greater power to reject non-null hypotheses than the classical Bonferroni method (see e.g., Hochberg & Tamhane, 1987).

### ***Hochberg's Step-Up Sequentially Acceptive Bonferroni***

Hochberg's (1988) step-up Bonferroni procedure is another example of a stepwise Bonferroni method of Type I error control and hence can also be better than the usual Dunn-Bonferroni method (see Kirk, 2013, p. 180). In this procedure, the  $p$ -values corresponding to the  $m$  statistics for testing the hypotheses  $H_{(1)}, \dots, H_{(m)}$  are ordered from smallest to largest  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_m$ . Then, for any  $i = m, m-1, \dots, 1$ , if  $p_i \leq \alpha / (m - i + 1)$ , the Hochberg procedure rejects all  $H_{i'}$  ( $i' \leq i$ ). According to this procedure, therefore, one begins by assessing the largest  $p$ -value,  $p_m$ . If  $p_m \leq \alpha$ , all hypotheses are rejected. If  $p_m > \alpha$ , then  $H_{(m)}$  is accepted and one proceeds to compare  $p_{(m-1)} \leq \alpha / 2$ . If  $p_{(m-1)} \leq \alpha / 2$ , then all  $H_i = (i = m-1, \dots, 1)$  are rejected; if not, then  $H_{(m-1)}$  is accepted and one proceeds to compare  $p_{(m-2)}$  with  $\alpha / 3$ , and so on.

Each condition in the investigation was replicated 5,000 times. We intended to collect Type I error rates for nine methods of testing treatment group equality (preliminary results indicated that we did not need to collect data for the WJ<sub>t</sub> procedure based on trimmed means and Winsorized variances).

### **Nuances to the Study**

Nuances to this study included (1) forcing the estimation and inference method to be maximum likelihood (ML) for all response variables and (2) matching the distributions of the simulated data with those of the response variables, i.e. symmetric data with symmetric response variables and skewed data with the lognormal response variable only (that is, using the proper link function).

## Method

When the response variables were modeled using only the “dist=” option in the model statement (from SASs GLIMMIX Procedure), the default estimation and inference method for lognormal and normal responses is restricted maximum likelihood (REML). For the exponential, gamma, inverse Gaussian, multinomial and tcentral (SASs nomenclature for a central  $t$  distribution) response variables, the default estimation and inference method is ML. In the present study, only the maximum likelihood method was used; hence ML was invoked with SAS syntax when the response variable was lognormal and normal.

## Matching Distribution

Initially, it was observed that the exponential and gamma response variables consistently produced the lowest AIC values, thus dominating the determination of the  $p$ -values of these procedures, usually inflating them, at times by an excessive amount, when most of the  $p$ -values of GLIMMIX modelled on these responses were less than 0.05. To avoid this anomaly, simulated data were matched with the response variable modelled. Thus, instead of considering the minimum AIC of all modeled link function responses, we considered the minimum AIC of symmetric modeled responses on data simulated from symmetric distributions. In the case of skewed data, we stuck to the lognormal response variable only.

## Specific Tcentral Response Variables

The only response variable that can be modeled with parameter values is the tcentral response variable (see SAS Institute, Inc., n. d., p. 66). For  $g = 0$  and  $h = 0.225$  simulated data, we decided to model two tcentral responses:  $t_{(3)}$  and  $t_{(4.039)}$ . The subscripted numbers in parentheses represent the degrees of freedom of the tcentral distribution. Three is the default degrees of freedom (in GLIMMIX), while 4.039 was obtained from solving  $6/(v - 4) = 154.84$ . The left hand side of the equation is the kurtosis of the  $t_{(v)}$  distribution while the right hand side is the kurtosis of the  $g = 0, h = 0.225$  distribution. Similarly, we also modelled the simulated contaminated mixed-normal distribution with specific tcentral response variables by solving equations involving the variance of the  $t$  distribution  $6/(v - 4)$  and the variances of the mixed normal distributions. The degrees of freedom are given in Table 4. Note that, for the first two contaminated mixed-normal distributions, the values obtained for the degrees of freedom are less than 0. Hence only  $t_{(3)}$  response variables were modelled for them. As for the remaining four, both the  $t_{(3)}$  and the  $t$

## GENERALIZED LINEAR MODEL ANALYSES

with the calculated degrees of freedom were modelled. Subsequently, all tcentral response variables modelled with calculated degrees of freedom will be known as tcentral2.

Hence the GLIMMIX procedure was employed in the following ways in this study: Specifically, we adapted the GLIMMIX test under the following conditions:

- (1) When the data were modelled on normal responses with the ML method of estimation and inference (this was our baseline measure = the ANOVA  $F$ -test.),
- (2) Depending on the result from the A-D test for normality (with  $\alpha = 0.15$  and  $\alpha = 0.20$ ), by selecting one of several response variable distributions modelled according to the skew of the simulated data as in Table 1,
- (3) By selecting one of several response variable distributions modelled according to the skew of the simulated data as in Table 1 and depending on the result of the Akaike (1974) model fit statistic,
- (4) Selecting the multinomial option when examining the Likert scale data conditions,
- (5) Selecting the lognormal option with the ML method of estimation and inference when examining data generated from skewed distributions including the  $g$ -and- $h$  distributions,<sup>5</sup>
- (6) Depending on the result from the A-D test for lognormality (with  $\alpha = 0.15$  and  $\alpha = 0.20$ ), by selecting one of several response variable distributions modelled according to the skew of the simulated data as in Table 1.

Thus, the tcentral2 was modelled in Procedures (2), (3), and (6).

## Results

Results were evaluated by adopting Bradley's (1978) criterion for effective Type I error control. According to this perspective, empirical rates of Type I error will be considered well controlled if the empirical value falls within a 0.025 to 0.075 interval. In our tables, values that exceed the upper limit of the interval will be displayed in bold typeface while those values less than the lower limit of the interval will be underscored.

### Symmetric Data

Table 5 presents empirical rates of Type I error when the simulated data were symmetric in shape for the three sample sizes investigated ( $n_1 = n_2 = n_3 = 20, 40,$  or  $80$ ). Of the 72 values reported in this table, only four fell outside Bradley's (1978) criterion and all are liberal ( $> 0.075$ ). The liberal rates ranged in value from 0.0752 to 0.1220. Liberal values occurred when either the A-D normality test was employed or when the Akaike (1974) criterion was used. It should also be noted that rates were well controlled for the baseline test (i.e., the ANOVA  $F$ -test) and when the response variable was set at lognormal.

Table 6 presents rates when the simulated data were skewed in shape for the three sample size cases investigated. Interestingly, regardless of the shape of the data, all empirical rates were contained within Bradley's (1978) interval.

Table 7 presents rates of error for non-normal data where the variances and group sizes were unequal across groups and were positively paired with one another. Of the 52 values reported, when unequal variances were in a 1:1:4 ratio (and sample sizes were 30, 40, and 50), only 14 values were not contained in Bradley's (1978) interval. This number increased to 32 when the unequal variances were in a 1:1:16 ratio (sample sizes were 30, 40, and 50). For the same two unequal variances conditions the non-controlled empirical values were 11 and 31, respectively when sample sizes were larger (i.e., 20, 40, and 60).

Table 8 presents rates for the ANOVA  $F$ -test and GLIMMIX assuming lognormal data in the link function when sample sizes (30, 40, and 50 or 20, 40, and 60) were paired with unequal variances in either a 1:1:4 or 1:1:16 ratio. Again, one can see from the table that many of the empirical values were either conservative ( $< 0.025$ ) or liberal ( $> 0.075$ ). The ANOVA  $F$ -test had 12 non-controlled values while GLIMMIX with the lognormal link function had 11 non-controlled rates.

Table 9 presents rates of error for non-normal data when unequal sample sizes (i.e., 30, 40, and 50 or 20, 40, and 60) were paired with unequal variances (4:1:1 or 16:1:1) in a negative fashion. For negative pairings of unequal group sizes and variances, the empirical rates were almost always outside Bradley's (1978) liberal interval and approached values as high as 0.50.

Finally, Table 10 presents empirical rates of Type I error for the ANOVA  $F$ -test baseline test and the GLIMMIX test assuming a lognormal link function when unequal variances and group sizes were negatively paired. Of the 28 ANOVA  $F$ -test empirical values, only two were contained in Bradley's (1978) interval with liberal values ranging from 0.1046 to 0.9152. Of the 20 GLIMMIX values, all but

one exceeded the upper bound of Bradley's interval with liberal values ranging from 0.1154 to 0.5870.

### Conclusion

Generalized linear model analyses (e.g., SASs GLIMMIX procedure) can be used to compare treatment effects across groups when data in the populations are normal or not normal. Having such estimation and testing procedures available to applied researchers should be most beneficial since in applied settings data are not likely to have been drawn from normal populations. Thus, researchers can use generalized linear model analyses to compare treatment effects across groups when data are not normally distributed in the parent populations. This is in contrast to traditional procedures such as the analysis of variance  $F$ -test which compares treatment effects across groups but presumes that data are normally distributed in the parent populations. It is well known that the probability of committing a Type I error (falsely rejecting the null hypothesis of treatment group equality) and the probability to detect treatment effects when they are present (the power of the test) are dramatically affected when the normality assumption does not hold. As indicated in the introduction, it is also well known that nonparametric tests are also negatively affected when data are non-normal and variances are unequal (see e.g., Zimmerman, 2010). Thus, none of these procedures should be adopted to test for equality of central tendency across treatment groups and the availability of alternative generalized analyses is considered an important addition to the researchers arsenal of data analytic techniques.

However, researchers typically do not know how the data in the population distributions are distributed and, accordingly, must rely on informal (e.g., graphs such as normal probability plots) or formal (e.g., tests for normality) methods to determine whether they can adopt the traditional method (i.e., the ANOVA  $F$ -test) of investigating for treatment effects across groups, or adopt more modern methods such as a generalized linear model to make such an assessment (i.e., SASs GLIMMIX procedure). Since informal methods are open to subjectivity of the analyzer, formal methods to analyze whether data are normal or not are recommended (e.g., see SAS Institute, Inc., 2010b; Schoder et al., 2006). Unfortunately, very little is known about the efficacy of adopting formal tests of normality or model fit prior to selecting a method for comparing treatment effects across groups. Prior research regarding the utility of tests for normality are mixed (e.g., see Dufour et al., 2010; Kowalchuk, et al., 2006; Keselman et al., 2013; 2014; Rochon & Kieser, 2011; Schoder et al., 2006) and the evidence regarding the

efficacy of tests for fit (e.g., Akaike, 1974) are not good (e.g., see Dufour et al., 2010; Keselman, Algina, et al., 2006; Keselman et al., 2013; 2014; Rochon & Kieser, 2011; Schoder et al., 2006). Thus, at this time, we do not know whether adopting this two stage strategy will be effective in testing for treatment group equality.

Accordingly, this problem was investigated by sampling data from various non-normal distributions (skewed distributions having different degrees of skewness and kurtosis, multinomial Likert type distributions, and mixed-normal distributions), either having equal variances or having unequal variances and unequal group sizes that were either positively or negatively paired with one another. The ANOVA *F*-test was then compared with GLIMMIX results when the GLIMMIX test was adopted following a significant test for non-normality with the Anderson-Darling test statistic or based on the results from applying the Akaike (1974) goodness of fit statistic. A GLIMMIX solution was also adopted in which we always assumed (through the link function) that the data were lognormal in shape.

It is apparent from the empirical findings that adopting a general linear model approach (e.g., GLIMMIX) procedure to assess treatment group equality across groups does not work when the link function of the data is based on the Anderson-Darling test for distribution shape, the smallest Akiake fit-statistic value, or always assuming a log-normal distribution. It is disappointing to report the generalized linear model solutions were not better at controlling the number of Type I errors as compared to the traditional ANOVA *F*-test. Indeed, for the cases investigated, the ANOVA *F*-test resulted in an inflated (liberal) or deflated (conservative) rate of Type I error 62 percent of the time while the rate for the GLIMMIX solutions varied from 56 to 61 percent. The lowest percentage of Type I errors for the GLIMMIX solution occurred when we presumed in the analysis that the data were lognormal in the population.

Therefore, researchers should adopt a generalized linear model analysis with caution as it will not necessarily provide better Type I error control when data are non-normal. As was indicated by Cerrito (2005), GLIMMIX is a difficult procedure to adopt, and much thought should be given to choosing this method of analysis. As they stated, “While it is possible to use PROC GLIMMIX as the most complex of the models, it is not advisable. Even so, choices as to random versus fixed effects, link function, and covariance matrix still have to be made. Therefore, the investigator should use the simplest procedure that will accommodate the variable choices.” (p. 7) (see also Stroup, 2013, Chapter 3).



## GENERALIZED LINEAR MODEL ANALYSES

All, however, is not lost. As indicated in the introduction, one can successfully test for equality of central tendency across groups when data are non-normal and variances are heterogeneous by adopting the non-pooled Welch-James statistic ( $WJ_t$ ) with robust estimators of central tendency and variability. This is a finding that has been established in many research investigations (e.g., see Keselman et al., 2008a; 2008b; Keselman, Wilcox, & Lix, 2003; Keselman, Wilcox, Lix, Algina, & Fradette, 2007; Keselman, Wilcox, Othman, & Fradette, 2002). As well, the  $WJ_t$  statistic can be applied with bootstrapping methodology (see Erceg-Hurn et al., 2013; Keselman et al., 2002) resulting in very good Type I error control.

### References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19(6), 716-723. doi: 10.1109/TAC.1974.1100705
- Akritas, M. G., Arnold, S. F., & Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92(437), 258-265. doi: 10.1080/01621459.1997.10473623
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211-252. Available from <http://www.jstor.org/stable/2984418>
- Bradley, J. V. (1978) Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x
- Breslow, N. R., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9-25. doi: 10.1080/01621459.1993.10594284
- Cerrito, P. B. (2005, October). *From GLM to GLIMMIX-Which model to choose?* Paper presented at the 13th Annual Conference of the Southeast SAS Users Group, Portsmouth, VA.
- D'Agostino, R. B., & Stephens, M. I. A. (Eds.). (1986). *Goodness-of-fit techniques*. New York, NY: Marcel Dekker.
- Dufour, J., Farhat, A., Gardiol, L., & Khalaf, I. (2010). *Simulation-based finite-sample normality tests in linear regressions* (Unpublished paper). Universite de Montreal, Montreal, Québec.

- Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). Robust statistical estimation. In T. Little (Ed.), *The Oxford handbook of quantitative methods (Vol. 1)* (pp. 388-406). New York, NY: Oxford University Press.
- Headrick, T. C., Kowalchuk, R. K., & Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey *g*-and-*h* transformations and their use for fitting data. *Applied Mathematical Sciences*, 2(9), 449-462. Retrieved from <http://www.m-hikari.com/ams/ams-password-2008/ams-password9-12-2008/headrickAMS9-12-2008.pdf>
- Hoaglin, D. C. (1983). *G*-and-*h* distributions. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences (Vol. 3)* (pp. 298-301). New York, NY: Wiley
- Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distribution. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Exploring Data Tables Trends and Shapes* (pp. 461-515). New York, NY: Wiley.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4), 800-803. doi: 10.1093/biomet/75.4.800
- Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. New York, NY: Wiley.
- Jameson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1212-1218. doi: 10.1111/j.1365-2929.2004.02012.x
- Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). Model selection criteria in the analysis of repeated measurements. *The American Statistician*, 60(3), 210-211. doi: 10.1080/03610919808813497
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. (2008a). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13(2), 110-129. doi: 10.1037/1082-989X.13.2.110
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. (2008b). *Supplemental material for: A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes*. doi: 10.1037/1082-989X.13.2.110.supp
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B.,...Levin, J. R. (1998). Statistical Practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Educational Research*, 68(3), 350-386. doi: 10.3102/00346543068003350

## GENERALIZED LINEAR MODEL ANALYSES

Keselman, H. J., Othman, A. R., & Wilcox, R. R. (2013). Preliminary testing for normality: Is it a good practice? *Journal of Modern Applied Statistical Methods*, 12(2), 2-19. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol12/iss2/2/>

Keselman, H. J., Othman, A. R., & Wilcox, R. R. (2014). Testing for normality in the multi-group problem: Is this a good practice? *Clinical Dermatology*, 2(1), 29-43. doi: 10.11138/cderm/2014.2.1.029

Keselman, H. J., Wilcox, R. R., & Lix, L. M. (2003). A generally robust approach to hypothesis testing in independent and correlated groups designs. *Psychophysiology*, 40(4), 586-596. doi: 10.1111/1469-8986.00060

Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J., & Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology*, 60(2), 267-293. doi: 10.1348/000711005X63755

Keselman, H. J., Wilcox, R., R., Othman, A. R., & Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, 1(2), 288-309. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol1/iss2/38/>

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Kowalchuk, R., K. & Headrick, T., C. (2010). Simulating multivariate *g*-and-*h* distributions. *British Journal of Mathematical and Statistical Psychology*, 63(1), 63-74. doi: 10.1348/000711009X423067

Kowalchuk, R. J., Keselman, H. J., Wilcox, R. R., & Algina, J. (2006). Multiple comparison procedures, trimmed means and transformed statistics. *Journal of Modern Applied Statistical Methods*, 5(1), 44-64. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol5/iss1/6/>

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and non-normality. *Educational and Psychological Measurement*, 58(3), 409-429. doi: 10.1177/0013164498058003004

Lix, L. M., & Keselman, H. J. (2009). Analysis of Variance (ANOVA). In M. W. Kattan (Ed.), *Encyclopedia of medical decision making*. Thousand Oaks, CA: Sage.

- Lix, L. M., Keselman, H. J., & Hinds, A. M. (2005). Robust tests for the multivariate Behrens-Fisher problem. *Computer Methods and Programs in Biomedicine*, 77(2), 129-139. doi: 10.1016/j.cmpb.2004.09.002
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166. doi: 10.1037/0033-2909.105.1.156
- Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods*, 3(1), 55-67. doi: 10.1037/1082-989X.3.1.55
- Nelder, J. A., & Wedderburn, W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384. doi: 10.2307/2344614
- Roberson, P. K., Shema, S. J., Mundfrom, D. J., & Holmes, T. M. (1995). Analysis of paired Likert data: how to evaluate change and preference questions. *Family Medicine*, 27(10), 671-675. Available from <http://europepmc.org/abstract/med/8582561>
- Rochon, J., & Kieser, M. (2011). A closer look at the effect of preliminary goodness-of-fit testing for normality for the one-sample *t*-test. *British Journal of Mathematical and Statistical Psychology*, 64(3), 410-426. doi: 10.1348/2044-8317.002003
- SAS Institute, Inc. (n.d.) *The GLIMMIX Procedure*. In UCLA Statistical Consulting Group. Retrieved from <http://statistics.ats.ucla.edu/stat/sas/glimmix.pdf>.
- SAS Institute, Inc. (2010a). *SAS/STAT 9.22 user's guide*. Cary, NC: SAS Institute, Inc.
- SAS Institute, Inc. (2010b). *Base SAS 9.2 procedures guide: Statistical procedures* (3rd ed). Cary, NC: SAS Institute, Inc.
- SAS Institute, Inc. (2013). *SAS 9.3 for Unix and Linux*. Cary, N C: SAS Institute, Inc.
- Schoder, V., Himmelmann, A., & Wilhelm, K. P. (2006). Preliminary testing for normality: Some statistical aspects of a common concept. *Clinical Dermatology*, 31(6), 757-761. doi: 10.1111/j.1365-2230.2006.02206.x
- SPSS, Inc. (2009). *PASW statistics for Windows, version 18.0*. Chicago, IL: SPSS, Inc.
- Stroup, W. W. (2013). *Generalized linear mixed models: Modern concepts, methods and applications*. Boca Raton, FL: CRC Press.

## GENERALIZED LINEAR MODEL ANALYSES

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin (Ed.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling*. Stanford University Press: Stanford, CA.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing* (3rd ed.). San Diego, CA: Academic Press.

Zimmerman, D. W. (2010). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *The Journal of Experimental Education*, 67(1), 55-68. doi: 10.1080/00220979809598344

## Footnotes

1. It was suggested by a reviewer of an earlier version of this paper that researchers need not concern themselves about normality/non-normality since the Gauss-Markov theorem guarantees that group means are asymptotically normal under much weaker conditions than assuming Gaussian errors. Such a statement surprised us considerably since it is well known that this is not true. Thus, we strenuously maintain that researchers must attend to whether their data are normal or not.
2. It has been suggested by a reviewer of an earlier version of this paper that researchers would know the appropriate link function for their data and would not need to employ tests of fit, etc. to determine the link function in order to get a correct solution. The first and third authors of this paper each have over forty years of experience teaching in departments of psychology and consulting with applied researchers and it is not our experience that researchers would know a priori the correct link function in order to use a generalized linear model analysis correctly.
3. SASs GLIMMIX procedure uses link functions to let the software know what distribution should be fitted to the data. In other words the GLIMMIX procedure assumes that  $E[\mathbf{Y} | \boldsymbol{\gamma}] = \mathbf{g}^{-1}(X\boldsymbol{\beta} + Z\boldsymbol{\gamma})$ , where  $\mathbf{g}(\bullet)$  is a differentiable monotonic link function,  $\mathbf{Y}$  represents the  $(n \times 1)$  vector of observed data, and  $\boldsymbol{\gamma}$  is an  $(r \times 1)$  vector of random effects (see SAS Institute, Inc., 2010b, p. 2637). For example, if one were to use the SAS syntax DIST=LOGNORMAL, the GLIMMIX procedure would model the logarithm of the response variable as a normal random variable. Thus, the mean and variance are estimated on the logarithmic scale, assuming a normal distribution, that is,  $\log\{\mathbf{Y}\} : N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  (SAS Institute, Inc., 2010b, p. 2725).
4. As we indicated in the introduction the link function in GLIMMIX for some types of non-normal data is known. As an example for Likert data the appropriate link function would be the multinomial distribution. For lognormal data the link function would be the identity function. Thus, in many instances applied researchers would know the correct link function (see SAS Institute, Inc., 2010b).
5. In many psychological investigations reaction time data is collected. Such data is notoriously known to contain large outlying values (i.e., very lengthy delays in reaction times) (see Tukey, 1960). Accordingly,

## GENERALIZED LINEAR MODEL ANALYSES

for such data it would be reasonable to assume that the underlying distribution is lognormal and accordingly we want to investigate the outcome of always assuming that the underlying distribution is of this form when adopting GLIMMIX.



**Appendix: Tables**

**Table 1.** Distributions used in this study with their corresponding measures of skewness and kurtosis values

Distribution	Skewness	Kurtosis
N(0, 1)	0.00	0.00
$g = 0, h = 0.225$	0.00	154.84
$g = 0.5, h = 0$ (lognormal)	1.75	8.90
$g = 1, h = 0$ (lognormal)	6.19	110.94
$g = 0.5, h = 0.5$	120.10 <sup>a</sup>	18393.60 <sup>a</sup>

Note: <sup>a</sup>Undefined skewness and kurtosis calculated from 100,000 simulated values

**Table 2.** Contaminated mixed-normal distributions

$(n_1, n_2, n_3)$	Group Variances	Distribution	Outliers	
			D	#
(40, 40, 40)	(1, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 1)
(40, 40, 40)	(1, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(2, 2, 2)
(40, 40, 40)	(1, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(4, 4, 4)
(40, 40, 40)	(1, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 1)
(40, 40, 40)	(1, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(2, 2, 2)
(40, 40, 40)	(1, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(4, 4, 4)
(30, 40, 50)	(1, 1, 4)	(0.975)N(0, 1) + (0.025)N(0,25)	5	(1, 1, 1)
(30, 40, 50)	(1, 1, 4)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(2, 2, 3)
(30, 40, 50)	(1, 1, 4)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(3, 4, 5)
(30, 40, 50)	(1, 1, 4)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 1)
(30, 40, 50)	(1, 1, 4)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(2, 2, 3)
(30, 40, 50)	(1, 1, 4)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(3, 4, 5)
(30, 40, 50)	(1, 1, 16)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 2)
(30, 40, 50)	(1, 1, 16)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(1, 2, 3)
(30, 40, 50)	(1, 1, 16)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(2, 4, 6)
(30, 40, 50)	(1, 1, 16)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 2)
(30, 40, 50)	(1, 1, 16)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(1, 2, 3)
(30, 40, 50)	(1, 1, 16)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(2, 4, 6)
(20, 40, 60)	(1, 1, 4)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 1)
(20, 40, 60)	(1, 1, 4)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(2, 2, 3)
(20, 40, 60)	(1, 1, 4)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(3, 4, 5)
(20, 40, 60)	(1, 1, 4)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 1)
(20, 40, 60)	(1, 1, 4)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(2, 2, 3)
(20, 40, 60)	(1, 1, 4)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(3, 4, 5)
(20, 40, 60)	(1, 1, 16)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 2)
(20, 40, 60)	(1, 1, 16)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(1, 2, 3)
(20, 40, 60)	(1, 1, 16)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(2, 4, 6)
(20, 40, 60)	(1, 1, 16)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 2)
(20, 40, 60)	(1, 1, 16)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(1, 2, 3)

GENERALIZED LINEAR MODEL ANALYSES

Table 2, continued.

$(n_1, n_2, n_3)$	Group Variances	Distribution	Outliers	
			$D$	#
(20, 40, 60)	(1, 1, 16)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(2, 4, 6)
(30, 40, 50)	(4, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 1)
(30, 40, 50)	(4, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(2, 2, 3)
(30, 40, 50)	(4, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(3, 4, 5)
(30, 40, 50)	(4, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 1)
(30, 40, 50)	(4, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(2, 2, 3)
(30, 40, 50)	(4, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(3, 4, 5)
(30, 40, 50)	(16, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 2)
(30, 40, 50)	(16, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(1, 2, 3)
(30, 40, 50)	(16, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(2, 4, 6)
(30, 40, 50)	(16, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 2)
(30, 40, 50)	(16, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(1, 2, 3)
(30, 40, 50)	(16, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(2, 4, 6)
(20, 40, 60)	(4, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 1)
(20, 40, 60)	(4, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(2, 2, 3)
(20, 40, 60)	(4, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(3, 4, 5)
(20, 40, 60)	(4, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 1)
(20, 40, 60)	(4, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(2, 2, 3)
(20, 40, 60)	(4, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(3, 4, 5)
(20, 40, 60)	(16, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 25)	5	(1, 1, 2)
(20, 40, 60)	(16, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 25)	5	(1, 2, 3)
(20, 40, 60)	(16, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 25)	5	(2, 4, 6)
(20, 40, 60)	(16, 1, 1)	(0.975)N(0, 1) + (0.025)N(0, 100)	10	(1, 1, 2)
(20, 40, 60)	(16, 1, 1)	(0.950)N(0, 1) + (0.050)N(0, 100)	10	(1, 2, 3)
(20, 40, 60)	(16, 1, 1)	(0.900)N(0, 1) + (0.100)N(0, 100)	10	(2, 4, 6)

Note:  $n_i$  stands for the number of observations in group  $i$ ;  $D$  stands for distance in standard deviation units; # stands for the number of outliers in the respective groups

Table 3. Multinomial distributions based upon Schoder et al.'s (2006) probabilities simulated as Likert scales

$(n_1, n_2, n_3)$	Group Variances	Description	$(p_1, p_2, p_3, p_4, p_5)$
(40, 40, 40)	(1, 1, 1)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(40, 40, 40)	(1, 1, 1)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(40, 40, 40)	(1, 1, 1)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(40, 40, 40)	(1, 1, 1)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(30, 40, 50)	(1, 1, 4)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(30, 40, 50)	(1, 1, 4)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(30, 40, 50)	(1, 1, 4)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(30, 40, 50)	(1, 1, 4)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(30, 40, 50)	(1, 1, 16)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(30, 40, 50)	(1, 1, 16)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)

**Table 3, continued.**

$(n_1, n_2, n_3)$	Group Variances	Description	$(\rho_1, \rho_2, \rho_3, \rho_4, \rho_5)$
(30, 40, 50)	(1, 1, 16)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(30, 40, 50)	(1, 1, 16)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(20, 40, 60)	(1, 1, 4)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(20, 40, 60)	(1, 1, 4)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(20, 40, 60)	(1, 1, 4)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(20, 40, 60)	(1, 1, 4)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(20, 40, 60)	(1, 1, 16)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(20, 40, 60)	(1, 1, 16)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(20, 40, 60)	(1, 1, 16)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(20, 40, 60)	(1, 1, 16)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(30, 40, 50)	(4, 1, 1)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(30, 40, 50)	(4, 1, 1)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(30, 40, 50)	(4, 1, 1)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(30, 40, 50)	(4, 1, 1)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(30, 40, 50)	(16, 1, 1)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(30, 40, 50)	(16, 1, 1)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(30, 40, 50)	(16, 1, 1)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(30, 40, 50)	(16, 1, 1)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(20, 40, 60)	(4, 1, 1)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(20, 40, 60)	(4, 1, 1)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(20, 40, 60)	(4, 1, 1)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(20, 40, 60)	(4, 1, 1)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)
(20, 40, 60)	(16, 1, 1)	Even	(0.20, 0.20, 0.20, 0.20, 0.20)
(20, 40, 60)	(16, 1, 1)	Symmetric	(0.10, 0.20, 0.40, 0.20, 0.10)
(20, 40, 60)	(16, 1, 1)	Moderate skew	(0.50, 0.30, 0.15, 0.04, 0.01)
(20, 40, 60)	(16, 1, 1)	Heavy skew	(0.70, 0.20, 0.06, 0.03, 0.01)

Note: See the note from Table 2

**Table 4.** Calculated degrees of freedom for the tcentral response variables from the variances of the simulated contaminated mixed normal distributions

Distribution	Variance	Calculated df <sup>a</sup>
(0.975)N(0, 1) + (0.025)N(0, 25)	0.9663	- <sup>b</sup>
(0.950)N(0, 1) + (0.050)N(0, 25)	0.9650	- <sup>b</sup>
(0.900)N(0, 1) + (0.100)N(0, 25)	1.0600	35.333
(0.975)N(0, 1) + (0.025)N(0, 100)	1.0131	154.323
(0.950)N(0, 1) + (0.050)N(0, 100)	1.1525	15.115
(0.900)N(0, 1) + (0.100)N(0, 100)	1.8100	4.469

Note: <sup>a</sup>df = degree of freedom; <sup>b</sup>Negative values, only the default was used

GENERALIZED LINEAR MODEL ANALYSES

**Table 5.** Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method=ML) and various versions of GLIMMIX on symmetric data

Sample Size	Distribution	GLIMMIX				
		Baseline	A-D Normal Test		Akaike	Lognormal
			$\alpha = 0.15$	$\alpha = 0.20$		
20	$g = 0.000, h = 0.000$	0.0598	0.0714	0.0726	0.0682	0.0590
	$g = 0.000, h = 0.225$	0.0544	0.0582	0.0574	0.0616	0.0540
40	$g = 0.000, h = 0.000$	0.0532	0.0636	0.0670	0.0572	0.0530
	$g = 0.000, h = 0.225$	0.0488	0.0554	0.0550	0.0558	0.0500
	(0.975)N(0, 1) + (0.025)N(0, 25)	0.0470	0.0744	<b>0.0752</b>	0.0742	0.0466
	(0.950)N(0, 1) + (0.050)N(0, 25)	0.0472	0.0698	0.0696	0.0692	0.0460
	(0.900)N(0, 1) + (0.100)N(0, 25)	0.0484	0.0600	0.0604	0.0604	0.0484
	(0.975)N(0, 1) + (0.025)N(0, 100)	0.0276	0.0736	0.0736	<b>0.0756</b>	0.0266
	(0.950)N(0, 1) + (0.050)N(0, 100)	0.0274	0.0622	0.0622	0.0622	0.0270
	(0.900)N(0, 1) + (0.100)N(0, 100)	0.0416	0.0442	0.0442	0.0442	0.0370
	Likert-Even	0.0520	<b>0.1220</b>	<b>0.1220</b>	0.0520	0.0514
	(0.2, 0.2, 0.2, 0.2, 0.2)	0.0460 <sup>a</sup>				
	Likert-Symmetric	0.0518	0.0550	0.0550	0.0522	0.0532
	(0.1, 0.2, 0.4, 0.2, 0.1)	0.0448 <sup>a</sup>				
80	$g = 0.000, h = 0.000$	0.0572	0.0676	0.0694	0.0574	0.0560
	$g = 0.000, h = 0.225$	0.0520	0.0530	0.0532	0.0532	0.0522

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable

**Table 6.** Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method = ML) and GLIMMIX with lognormal response on skewed data

Sample Size	Distribution	GLIMMIX	
		Baseline	Lognormal
20	$g = 0.5, h = 0.0$	0.0528	0.0540
	$g = 1.0, h = 0.0$	0.0430	0.0434
	$g = 0.5, h = 0.5$	0.0364	0.0370
40	$g = 0.5, h = 0.0$	0.0544	0.0552
	$g = 1.0, h = 0.0$	0.0442	0.0462
	$g = 0.5, h = 0.5$	0.0330	0.0340
	Likert-Moderate Skew	0.0482	0.0520
	(0.50, 0.30, 0.15, 0.04, 0.01)	0.0442 <sup>a</sup>	
	Likert-Heavy Skew	0.0502	0.0518
(0.70, 0.20, 0.06, 0.03, 0.01)	0.0392 <sup>a</sup>		
80	$g = 0.5, h = 0.0$	0.0540	0.0534
	$g = 1.0, h = 0.0$	0.0466	0.0498
	$g = 0.5, h = 0.5$	0.0320	0.0336

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable

**Table 7.** Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method = ML) and various versions of GLIMMIX on symmetric data when group sample sizes are positively paired with group variances

Sample Size/Variance	Distribution	GLIMMIX				
		Baseline	A-D Normal Test		Akaike	Lognormal
			$\alpha=0.15$	$\alpha=0.20$		
(30, 40, 50) / (1, 1, 4)	$g = 0.000, h = 0.000$	0.0408	0.0568	0.0582	<b>0.0762</b>	0.0442
	$g = 0.000, h = 0.225$	0.0368	0.0536	0.0540	0.0554	0.0384
	(0.975)N(0, 1) + (0.025)N(0, 25)	0.0336	0.0742	0.0748	<b>0.0792</b>	0.0330
	(0.950)N(0, 1) + (0.050)N(0, 25)	0.0316	0.0712	0.0716	0.0736	0.0298
	(0.900)N(0, 1) + (0.100)N(0, 25)	0.0304	0.0570	0.0570	0.0572	0.0286
	(0.975)N(0, 1) + (0.025)N(0, 100)	<u>0.0232</u>	<b>0.0772</b>	<b>0.0772</b>	<b>0.0782</b>	<u>0.0236</u>
	(0.950)N(0, 1) + (0.050)N(0, 100)	<u>0.0198</u>	0.0622	0.0622	0.0622	<u>0.0196</u>
	(0.900)N(0, 1) + (0.100)N(0, 100)	0.0270	0.0420	0.0420	0.0420	<u>0.0226</u>
	Likert-Even	0.0382	<b>0.1444</b>	<b>0.1444</b>	0.0382	<b>0.1358</b>
	(0.2, 0.2, 0.2, 0.2, 0.2)	0.0548 <sup>a</sup>				
	Likert-Symmetric	0.0406	0.0552	0.0552	0.0550	0.1380
	(0.1, 0.2, 0.4, 0.2, 0.1)	0.0498 <sup>a</sup>				
	(30, 40, 50) / (1, 1, 16)	$g = 0.000, h = 0.000$	0.0424	0.0722	<b>0.0774</b>	<b>0.1462</b>
$g = 0.000, h = 0.225$		0.0396	<b>0.0910</b>	<b>0.0912</b>	<b>0.0938</b>	0.0490
(0.975)N(0, 1) + (0.025)N(0, 25)		0.0346	<b>0.1176</b>	<b>0.1184</b>	<b>0.1346</b>	0.0392
(0.950)N(0, 1) + (0.050)N(0, 25)		0.0290	<b>0.1166</b>	<b>0.1168</b>	<b>0.1198</b>	0.0318
(0.900)N(0, 1) + (0.100)N(0, 25)		0.0316	<b>0.1004</b>	<b>0.1004</b>	<b>0.1008</b>	0.0290
(0.975)N(0, 1) + (0.025)N(0, 100)		<u>0.0246</u>	<b>0.1276</b>	<b>0.1276</b>	<b>0.1302</b>	0.0308
(0.950)N(0, 1) + (0.050)N(0, 100)		<u>0.0182</u>	<b>0.1044</b>	<b>0.1044</b>	<b>0.1048</b>	<u>0.0218</u>
(0.900)N(0, 1) + (0.100)N(0, 100)		<u>0.0236</u>	<b>0.0798</b>	<b>0.0798</b>	<b>0.0796</b>	0.0304
Likert-Even		0.0424	<b>0.2188</b>	<b>0.2188</b>	<b>0.2186</b>	<u>0.2402</u>
(0.2, 0.2, 0.2, 0.2, 0.2)		<b>0.0814</b> <sup>a</sup>				
Likert-Symmetric		0.0440	0.0292	0.0292	0.0292	0.0628
(0.1, 0.2, 0.4, 0.2, 0.1)		0.0576 <sup>a</sup>				
(20, 40, 60) / (1, 1, 4)		$g = 0.000, h = 0.000$	0.0258	0.0364	0.0378	0.0564
	$g = 0.000, h = 0.225$	<u>0.0252</u>	0.0400	0.0400	0.0402	0.0254
	(0.975)N(0, 1) + (0.025)N(0, 25)	<u>0.0218</u>	0.0544	0.0550	0.0578	0.0216
	(0.950)N(0, 1) + (0.050)N(0, 25)	<u>0.0158</u>	0.0514	0.0512	0.0520	0.0162
	(0.900)N(0, 1) + (0.100)N(0, 25)	<u>0.0186</u>	0.0390	0.0394	0.0396	0.0180
	(0.975)N(0, 1) + (0.025)N(0, 100)	<u>0.0170</u>	0.0532	0.0532	0.0536	0.0180
	(0.950)N(0, 1) + (0.050)N(0, 100)	<u>0.0102</u>	0.0486	0.0486	0.0486	0.0108
	(0.900)N(0, 1) + (0.100)N(0, 100)	<u>0.0148</u>	0.0290	0.0290	0.0290	0.0154
	Likert-Even	<u>0.0216</u>	0.1084	0.1084	0.0216	0.1046
	(0.2, 0.2, 0.2, 0.2, 0.2)	<u>0.0322</u> <sup>a</sup>				
	Likert-Symmetric	<u>0.0228</u>	0.0418	0.0418	0.0338	0.0960
	(0.1, 0.2, 0.4, 0.2, 0.1)	<u>0.0328</u> <sup>a</sup>				

GENERALIZED LINEAR MODEL ANALYSES

Table 7, continued.

Sample Size/Variance	Distribution	GLIMMIX				
		Baseline	A-D Normal Test		Akaike	Lognormal
			$\alpha=0.15$	$\alpha=0.20$		
(20, 40, 60) / (1, 1, 16)	$g = 0.000, h = 0.000$	<u>0.0202</u>	0.0384	0.0418	<b>0.0954</b>	0.0390
	$g = 0.000, h = 0.225$	<u>0.0158</u>	0.0428	0.0432	0.0440	<u>0.0186</u>
	(0.975)N(0, 1) + (0.025)N(0, 25)	<u>0.0138</u>	<b>0.0762</b>	<b>0.0774</b>	<b>0.0864</b>	<u>0.0166</u>
	(0.950)N(0, 1) + (0.050)N(0, 25)	<u>0.0122</u>	<b>0.0774</b>	<b>0.0780</b>	<b>0.0804</b>	<u>0.0138</u>
	(0.900)N(0, 1) + (0.100)N(0, 25)	<u>0.0136</u>	0.0614	0.0618	0.0620	<u>0.0112</u>
	(0.975)N(0, 1) + (0.025)N(0, 100)	<u>0.0096</u>	<b>0.0788</b>	<b>0.0792</b>	<b>0.0802</b>	<u>0.0104</u>
	(0.950)N(0, 1) + (0.050)N(0, 100)	<u>0.0052</u>	0.0688	0.0688	0.0688	<u>0.0088</u>
	(0.900)N(0, 1) + (0.100)N(0, 100)	<u>0.0068</u>	0.0444	0.0444	0.0438	<u>0.0118</u>
	Likert-Even	<u>0.0164</u>	<b>0.1614</b>	<b>0.1614</b>	<b>0.1358</b>	<b>0.1900</b>
	(0.2, 0.2, 0.2, 0.2, 0.2)	0.0466 <sup>a</sup>				
	Likert-Symmetric	<u>0.0198</u>	0.0272	0.0272	0.0272	0.0342
	(0.1, 0.2, 0.4, 0.2, 0.1)	0.0314 <sup>a</sup>				

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable

Table 8. Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method=ML) and GLIMMIX with lognormal response on skewed data when group sample sizes are positively paired with group variances

Sample Sizes/Variations	Distribution	GLIMMIX	
		Baseline	Lognormal
(30, 40, 50) / (1, 1, 4)	$g = 0.5, h = 0.0$	0.0416	0.0484
	$g = 1.0, h = 0.0$	0.0552	0.0642
	$g = 0.5, h = 0.5$	0.0254	0.0258
	Likert-Moderate Skew	0.0400	0.0836
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.2230<sup>a</sup></b>	
	Likert-Heavy Skew	0.0456	<b>0.1344</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9570<sup>a</sup></b>	
(30, 40, 50) / (1, 1, 16)	$g = 0.5, h = 0.0$	0.0460	<b>0.0798</b>
	$g = 1.0, h = 0.0$	0.0790	<b>0.1212</b>
	$g = 0.5, h = 0.5$	<u>0.0240</u>	0.0262
	Likert-Moderate Skew	0.0436	0.4360
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.2176<sup>a</sup></b>	
	Likert-Heavy Skew	0.0558	<b>0.5648</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9398<sup>a</sup></b>	
(20, 40, 60) / (1, 1, 4)	$g = 0.5, h = 0.0$	0.0274	0.0334
	$g = 1.0, h = 0.0$	0.0442	<u>0.0538</u>
	$g = 0.5, h = 0.5$	<u>0.0208</u>	<u>0.0216</u>
	Likert-Moderate Skew	0.0206	0.0642
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.1818<sup>a</sup></b>	
	Likert-Heavy Skew	0.0288	<b>0.1124</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9658<sup>a</sup></b>	

Table 8, continued.

Sample Sizes/Variates	Distribution	GLIMMIX	
		Baseline	Lognormal
(20, 40, 60) / (1, 1, 16)	$g = 0.5, h = 0.0$	0.0252	0.0478
	$g = 1.0, h = 0.0$	0.0542	0.0892
	$g = 0.5, h = 0.5$	<u>0.0124</u>	<u>0.0130</u>
	Likert-Moderate Skew	<u>0.0170</u>	<b>0.3738</b>
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.1762<sup>a</sup></b>	
	Likert-Heavy Skew	0.0322	<b>0.5150</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9410<sup>a</sup></b>	

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable

Table 9. Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method = ML) and various versions of GLIMMIX on symmetric data when group sample sizes are negatively paired with group variances

Sample Size/Variance	Distribution	GLIMMIX				
		Baseline	A-D Normal Test			Lognormal
			$\alpha=0.15$	$\alpha=0.20$	Akaike	
(30, 40, 50) / (1, 1, 4)	$g = 0.000, h = 0.000$	<b>0.1024</b>	<b>0.1194</b>	<b>0.1230</b>	<b>0.1454</b>	<b>0.1044</b>
	$g = 0.000, h = 0.225$	<b>0.0922</b>	<b>0.0944</b>	<b>0.0950</b>	<b>0.0928</b>	<b>0.0964</b>
	(0.975)N(0, 1) + (0.025)N(0, 25)	<b>0.0992</b>	<b>0.1268</b>	<b>0.1280</b>	<b>0.1296</b>	<b>0.0976</b>
	(0.950)N(0, 1) + (0.050)N(0, 25)	<b>0.0946</b>	<b>0.1140</b>	<b>0.1138</b>	<b>0.1148</b>	<b>0.0952</b>
	(0.900)N(0, 1) + (0.100)N(0, 25)	<b>0.0908</b>	<b>0.0968</b>	<b>0.0968</b>	<b>0.0964</b>	<b>0.0894</b>
	(0.975)N(0, 1) + (0.025)N(0, 100)	0.0714	<b>0.1238</b>	<b>0.1242</b>	<b>0.1248</b>	0.0650
	(0.950)N(0, 1) + (0.050)N(0, 100)	0.0736	<b>0.1036</b>	<b>0.1038</b>	<b>0.1040</b>	0.0668
	(0.900)N(0, 1) + (0.100)N(0, 100)	<b>0.0834</b>	<b>0.0804</b>	<b>0.0804</b>	<b>0.0804</b>	0.0714
	Likert-Even	<b>0.0970</b>	<b>0.2336</b>	<b>0.2336</b>	<b>0.0984</b>	<b>0.2028</b>
	(0.2, 0.2, 0.2, 0.2, 0.2)	<b>0.1244<sup>a</sup></b>				
	Likert-Symmetric	<b>0.0984</b>	<b>0.0788</b>	<b>0.0788</b>	<b>0.0968</b>	<b>0.2070</b>
	(0.1, 0.2, 0.4, 0.2, 0.1)	<b>0.0866<sup>a</sup></b>				
	(30, 40, 50) / (1, 1, 16)	$g = 0.000, h = 0.000$	<b>0.1450</b>	<b>0.1804</b>	<b>0.1866</b>	<b>0.2698</b>
$g = 0.000, h = 0.225$		<b>0.1402</b>	<b>0.1854</b>	<b>0.1856</b>	<b>0.1866</b>	<b>0.1480</b>
(0.975)N(0, 1) + (0.025)N(0, 25)		<b>0.1430</b>	<b>0.2384</b>	<b>0.2404</b>	<b>0.2538</b>	<b>0.1420</b>
(0.950)N(0, 1) + (0.050)N(0, 25)		<b>0.1420</b>	<b>0.2300</b>	<b>0.2302</b>	<b>0.2348</b>	<b>0.1432</b>
(0.900)N(0, 1) + (0.100)N(0, 25)		<b>0.1398</b>	<b>0.2102</b>	<b>0.2102</b>	<b>0.2110</b>	<b>0.1402</b>
(0.975)N(0, 1) + (0.025)N(0, 100)		<b>0.1172</b>	<b>0.2398</b>	<b>0.2410</b>	<b>0.2426</b>	<b>0.1160</b>
(0.950)N(0, 1) + (0.050)N(0, 100)		<b>0.1252</b>	<b>0.2178</b>	<b>0.2180</b>	<b>0.2180</b>	<b>0.1100</b>
(0.900)N(0, 1) + (0.100)N(0, 100)		<b>0.1378</b>	<b>0.1798</b>	<b>0.1798</b>	<b>0.1796</b>	<b>0.1086</b>
Likert-Even		<b>0.1430</b>	<b>0.3624</b>	<b>0.3624</b>	<b>0.3624</b>	<b>0.3280</b>
(0.2, 0.2, 0.2, 0.2, 0.2)		<b>0.1866<sup>a</sup></b>				
Likert-Symmetric		<b>0.1490</b>	0.0462	0.0462	0.0462	<b>0.1636</b>
(0.1, 0.2, 0.4, 0.2, 0.1)		<b>0.1140<sup>a</sup></b>				



GENERALIZED LINEAR MODEL ANALYSES

Table 9, continued.

Sample Size/Variance	Distribution	GLIMMIX				
		Baseline	A-D Normal Test			Lognormal
			$\alpha=0.15$	$\alpha=0.20$	Akaike	
(20, 40, 60) / (1, 1, 4)	$g = 0.000, h = 0.000$	<b>0.1110</b>	<b>0.1346</b>	<b>0.1372</b>	<b>0.1110</b>	<b>0.1158</b>
	$g = 0.000, h = 0.225$	<b>0.1060</b>	<b>0.0950</b>	<b>0.0948</b>	<b>0.0976</b>	<b>0.1082</b>
	$(0.975)N(0, 1) + (0.025)N(0, 25)$	<b>0.1620</b>	<b>0.1744</b>	<b>0.1738</b>	<b>0.1714</b>	<b>0.1578</b>
	$(0.950)N(0, 1) + (0.050)N(0, 25)$	<b>0.1364</b>	<b>0.1616</b>	<b>0.1620</b>	<b>0.1626</b>	<b>0.1342</b>
	$(0.900)N(0, 1) + (0.100)N(0, 25)$	<b>0.1386</b>	<b>0.1360</b>	<b>0.1362</b>	<b>0.1362</b>	<b>0.1426</b>
	$(0.975)N(0, 1) + (0.025)N(0, 100)$	<b>0.1376</b>	<b>0.1656</b>	<b>0.1656</b>	<b>0.1658</b>	<b>0.1276</b>
	$(0.950)N(0, 1) + (0.050)N(0, 100)$	<b>0.0992</b>	<b>0.1506</b>	<b>0.1506</b>	<b>0.1508</b>	<b>0.0860</b>
	$(0.900)N(0, 1) + (0.100)N(0, 100)$	<b>0.1266</b>	<b>0.1140</b>	<b>0.1140</b>	<b>0.1140</b>	<b>0.1102</b>
	Likert-Even	<b>0.1450</b>	<b>0.2836</b>	<b>0.2836</b>	<b>0.1462</b>	<b>0.2380</b>
	(0.2, 0.2, 0.2, 0.2, 0.2)	<b>0.1664<sup>a</sup></b>				
	Likert-Symmetric	<b>0.1472</b>	<b>0.0952</b>	<b>0.0952</b>	<b>0.1422</b>	<b>0.2546</b>
	(0.1, 0.2, 0.4, 0.2, 0.1)	<b>0.1146<sup>a</sup></b>				
	(20, 40, 60) / (1, 1, 4)	$g = 0.000, h = 0.000$	<b>0.1720</b>	<b>0.2186</b>	<b>0.2258</b>	<b>0.1830</b>
$g = 0.000, h = 0.225$		<b>0.1686</b>	<b>0.1910</b>	<b>0.1910</b>	<b>0.1940</b>	<b>0.1650</b>
$(0.975)N(0, 1) + (0.025)N(0, 25)$		<b>0.2578</b>	<b>0.3156</b>	<b>0.3160</b>	<b>0.3276</b>	<b>0.2570</b>
$(0.950)N(0, 1) + (0.050)N(0, 25)$		<b>0.2440</b>	<b>0.3128</b>	<b>0.3132</b>	<b>0.3174</b>	<b>0.2456</b>
$(0.900)N(0, 1) + (0.100)N(0, 25)$		<b>0.2542</b>	<b>0.2852</b>	<b>0.2858</b>	<b>0.2856</b>	<b>0.2544</b>
$(0.975)N(0, 1) + (0.025)N(0, 100)$		<b>0.2448</b>	<b>0.3198</b>	<b>0.3202</b>	<b>0.3206</b>	<b>0.2092</b>
$(0.950)N(0, 1) + (0.050)N(0, 100)$		<b>0.2146</b>	<b>0.3014</b>	<b>0.3018</b>	<b>0.3020</b>	<b>0.1752</b>
$(0.900)N(0, 1) + (0.100)N(0, 100)$		<b>0.2582</b>	<b>0.2550</b>	<b>0.2550</b>	<b>0.2552</b>	<b>0.1860</b>
Likert-Even		<b>0.2328</b>	<b>0.4850</b>	<b>0.4850</b>	<b>0.4850</b>	<b>0.3728</b>
(0.2, 0.2, 0.2, 0.2, 0.2)		<b>0.2532<sup>a</sup></b>				
Likert-Symmetric		<b>0.2406</b>	<b>0.1056</b>	<b>0.1056</b>	<b>0.1062</b>	<b>0.2426</b>
(0.1, 0.2, 0.4, 0.2, 0.1)		<b>0.1506<sup>a</sup></b>				

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable

Table 10. Type I error rates for treatment group differences using the baseline test (GLIMMIX dist = normal method = ML) and GLIMMIX with lognormal response on skewed data when group sample sizes are negatively paired with group variances

Sample Sizes/Variations	Distribution	GLIMMIX	
		Baseline	Lognormal
(30, 40, 50) / (4, 1, 1)	$g = 0.5, h = 0.0$	<b>0.1046</b>	<b>0.1132</b>
	$g = 1.0, h = 0.0$	<b>0.1102</b>	<b>0.1190</b>
	$g = 0.5, h = 0.5$	0.0676	0.0692
	Likert-Moderate Skew	<b>0.1030</b>	<b>0.1378</b>
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.2962<sup>a</sup></b>	
	Likert-Heavy Skew	<b>0.1054</b>	<b>0.1822</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9240<sup>a</sup></b>	

Table 10, continued.

Sample Sizes/Variances	Distribution	GLIMMIX	
		Baseline	Lognormal
(30, 40, 50) / (16, 1, 1)	$g = 0.5, h = 0.0$	<b>0.1518</b>	<b>0.1894</b>
	$g = 1.0, h = 0.0$	<b>0.1906</b>	<b>0.2402</b>
	$g = 0.5, h = 0.5$	<b>0.1162</b>	<b>0.1154</b>
	Likert-Moderate Skew	<b>0.1456</b>	
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.2944<sup>a</sup></b>	<b>0.5002</b>
	Likert-Heavy Skew	<b>0.1618</b>	
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9152<sup>a</sup></b>	<b>0.5980</b>
(30, 40, 50) / (16, 1, 1)	$g = 0.5, h = 0.0$	0.0460	<b>0.0798</b>
	$g = 1.0, h = 0.0$	0.0790	<b>0.1212</b>
	$g = 0.5, h = 0.5$	<u>0.0240</u>	0.0262
	Likert-Moderate Skew	0.0436	0.4360
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.2176<sup>a</sup></b>	
	Likert-Heavy Skew	0.0558	<b>0.5648</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.9398<sup>a</sup></b>	
(20, 40, 60) / (4, 1, 1)	$g = 0.5, h = 0.0$	<b>0.1126</b>	<b>0.1194</b>
	$g = 1.0, h = 0.0$	<b>0.1130</b>	<b>0.1218</b>
	$g = 0.5, h = 0.5$	0.0714	0.0720
	Likert-Moderate Skew	<b>0.1442</b>	<b>0.1722</b>
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.3230<sup>a</sup></b>	
	Likert-Heavy Skew	<b>0.1512</b>	<b>0.2132</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.8850<sup>a</sup></b>	
(20, 40, 60) / (16, 1, 1)	$g = 0.5, h = 0.0$	<b>0.1796</b>	<b>0.2084</b>
	$g = 1.0, h = 0.0$	<b>0.2164</b>	<b>0.2510</b>
	$g = 0.5, h = 0.5$	<b>0.1286</b>	<b>0.1274</b>
	Likert-Moderate Skew	<b>0.2366</b>	<b>0.5066</b>
	(0.50, 0.30, 0.15, 0.04, 0.01)	<b>0.3242<sup>a</sup></b>	
	Likert-Heavy Skew	<b>0.2536</b>	<b>0.5870</b>
	(0.70, 0.20, 0.06, 0.03, 0.01)	<b>0.8844<sup>a</sup></b>	

Note: <sup>a</sup>GLIMMIX modelled by multinomial response variable