

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2004 - 16th Annual Conference Proceedings

ANALYZING BINOMIAL DATA IN A SPLIT-PLOT DESIGN: CLASSICAL APPROACHES OR MODERN TECHNIQUES?

Liang Fang

Thomas M. Loughin

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Fang, Liang and Loughin, Thomas M. (2004). "ANALYZING BINOMIAL DATA IN A SPLIT-PLOT DESIGN: CLASSICAL APPROACHES OR MODERN TECHNIQUES?," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1162>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

ANALYZING BINOMIAL DATA IN A SPLIT-PLOT DESIGN: CLASSICAL APPROACHES OR MODERN TECHNIQUES?

Liang Fang and Thomas M. Loughin

Department of Statistics, Kansas State University, Manhattan, KS 66502

ABSTRACT: Binomial data are often generated in split-plot experimental designs in agricultural, biological, and environmental research. Modeling non-normality and random effects are the two major challenges in analyzing binomial data in split-plot designs. In this study, seven statistical methods for testing whole-plot and subplot treatment effects using mixed, generalized linear, or generalized linear mixed models are compared for the size and power of the tests. This study shows that analyzing random effects properly is more important than adjusting the analysis for non-normality. Methods based on mixed and generalized linear mixed models hold Type I error rates better than generalized linear models. Whole-plot tests tend to be conservative in some cases, but these tests can be improved by removing the lower bound of zero from variance parameter estimation or by increasing the number of whole-plot replications. Mixed model methods tend to have higher power than generalized linear mixed models when the sample size is small. However, they perform equally well as the sample size becomes large.

1. INTRODUCTION

Binomial data are often generated in split-plot experimental designs in agricultural, biological, and environmental research (Milliken and Johnson, 1992). For example, in entomology, mortality of insects treated with pesticides is measured on wheat of different moisture contents under each of several different temperatures, where chambers are set at different temperatures and pesticides are assigned to containers within the chambers (Fang et al., 2003). In agriculture, germination rates are observed for different types of spores under several levels of temperature, where temperature is the whole-plot factor and spore type is the subplot factor (Huang et al., 2001).

To analyze the data from these experiments, we are confronted with two challenges: non-normality and random effects. No method of analysis proposed to date provides an exact solution to this problem. Mixed model methods are standard ways to analyze split-plot designs, but they assume sampling from a normally-distributed population. Transformations are often used with the mixed models to stabilize the potentially heterogeneous variances due to the binomial sampling (Zar, 1999), but the transformed data are then assumed to have been sampled from a normal population. Additional techniques have been developed that can model non-normal data directly and are available to general users in popular statistical software packages. Generalized linear models are widely used to analyze binomial data (Agresti, 1996), but they contain no mechanism for modeling random effects, and tests from these models are based on asymptotic approximations that may not be adequate in typically-sized biological experiments. Generalized linear mixed models can analyze binomial data with random effects (Collett 1991, Cox and Snell 1989), but tests from these models are also based on asymptotic approximations, and their small-sample properties are not well understood. Thus, all methods of analysis involve some sort of approximation.

A similar problem occurs when we try to analyze Poisson-distributed count data in a split-plot experiment. Grossardt (MS report, 2003) studies Type I error rates of eleven methods for analyzing count data in split-plot designs and finds that mixed model methods and the generalized linear mixed model method maintain Type I error rates better than generalized linear model methods that have no mechanism to account for random effects. Beckman and Stroup (2003) and Sui and Stroup (2001) compare five methods for analyzing binomial or multinomial data in repeated measures experiments where the issues of non-normality and correlated measurements occur. They find similarly that the classical mixed model approach performs better than methods based on the generalized linear and generalized linear mixed models.

Among all the methods above, the mixed model method with an angular or Freeman-Tukey transformation is widely used by non-statistical researchers in many disciplines, due to its simplicity and long-time existence. The methods with generalized linear or generalized linear mixed models are drawing a lot of attention recently because they are now available in popular statistical software. More and more consulting statisticians recommend these “new” methods to researchers (Garrett et al., 2004); however, to our knowledge, no comprehensive study has been done to assess and compare these methods and ensure that these new methods are, indeed, improvements over their classical counterparts.

The objective of this article is to compare various popular methods of analysis for binomial data from split-plot designs and to provide immediate and practical suggestions to the readers. A simulation study is designed to study the Type I error rates and power of the whole-plot, subplot, and interaction tests. In a preliminary study, seven factors that may affect rejection rates of these tests are studied and screened for a thorough follow-up study. The follow-up study with selected factors and methods provides a more thorough investigation of Type I error rates and power of the tests. General recommendations are made at the end of the article.

2. MODELS AND METHODS

Suppose that N binomial trials are observed on each experimental unit in a split-plot design, and let Y_{ijk} represent the count of successes from the subplot unit receiving subplot treatment j in the k^{th} whole-plot unit receiving whole-plot treatment i . $i = 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K$. Let π_{ijk} represent the corresponding probability of success.

Three types of models are used in the study:

- Mixed model: $Y_{ijk} = \lambda_0 + \lambda_i^w + w_{k(i)} + \lambda_j^s + \lambda_{ij}^{ws} + s_{jk(i)}$
- Generalized linear model: $Y_{ijk} \sim \text{Binomial}(N, \pi_{ij}), \log(\pi_{ij}/(1 - \pi_{ij})) = \lambda_0 + \lambda_i^w + \lambda_j^s + \lambda_{ij}^{ws}$
- Generalized linear mixed model: $Y_{ijk} \sim \text{Binomial}(N, \pi_{ijk}), \log(\pi_{ijk}/(1 - \pi_{ijk})) = \lambda_0 + \lambda_i^w + w_{k(i)} + \lambda_j^s + \lambda_{ij}^{ws} + s_{jk(i)}$

In each model, λ_0 is the intercept; λ_i^w is the whole-plot effect for whole-plot treatment i ; λ_j^s is the subplot effect for subplot treatment j ; λ_{ij}^{ws} is the interaction of the i^{th} whole-plot and j^{th} subplot treatment; $w_{k(i)}$ is a random error associated with unit k of whole-plot treatment i ; and $s_{jk(i)}$ is a random error associated with the unit receiving subplot treatment j on whole-plot unit $k(i)$. The whole-plot and subplot errors are independent of each other and each has an identical and independent normal distribution with mean zero and variance σ_w^2 and σ_s^2 , respectively. Note that the interpretation of all parameters depends on the context of the model in which they appear. In Model (a) they are direct effects on the counts, while in (b) and (c) they are effects on the log-

odds of success. In all cases, it is assumed that all binomial counts are based on the same number of trials, N , so that analysis of the counts is equivalent to analysis of the proportions.

Derived from the three models, there are seven analysis methods studied in this article:

- 1) Mixed model (a) with original data as response (MIXED)
- 2) Mixed model (a) with angular-transformed binomial data as response (MIXEDT)
- 3) Generalized linear mixed model (c) (GLMM)
- 4) Generalized linear model (b) without random effects using Type III likelihood ratio tests (GLM)
- 5) Generalized linear model (b) without random effects using Wald tests (GLMW)
- 6) Generalized linear model (b) without random effects and with overdispersion adjustment (GLMO)
- 7) Generalized linear model (b) with random effects represented by correlated subplots (GLMGEE)

These methods represent a wild range of popular or recommended analysis approaches for split-plot designs and binomial data.

All computations are done using SAS[®]. PROC MIXED is used to implement the mixed model methods. The generalized linear mixed model (GLMM) is fitted with the GLIMMIX[®] macro that is available on the SAS[®] website (McCulloch 1996). The generalized linear model methods are implemented with PROC GENMOD. Type III likelihood ratio tests and Wald tests are available to test the treatment effects in GENMOD. The GLMO method uses Pearson scaling (Agresti, 1996) in an effort to adjust for the overdispersion likely to result from the failure to model the random effects. As an alternative to modeling the random effects directly, one can model the correlated subplot units as repeated measures with a compound symmetry covariance structure. Model fitting and testing are done using the generalized estimating equations approach of Liang and Zeger (1986) and achieved in SAS[®] using the REPEATED statement in the GENMOD procedure. Only Wald tests are available with this analysis in SAS[®].

3. SIMULATION STUDY

An initial screening experiment is done to identify factors that influence the type I error rates of the various analysis methods and to determine which methods are robust in the sense of holding their type I error rates under varying circumstances. Follow-up studies are done to further quantify the effects of important factors and to explore the extent to which sample sizes and small mean counts affect the methods. Finally, power is assessed for those methods that satisfactorily hold their type I error rates.

3.1 GENERATE BINOMIALLY DISTRIBUTED DATA FOR THE TYPE I ERROR RATE STUDY

Assuming the whole-plot structure of the experiment is a completely randomized design, a plausible model for binomial data in a split-plot design is that the response $Y_{ijk} \sim \text{Bin}(N, \pi_{ijk})$, where π_{ijk} is generated from a logit-linear model of the form

$$\log(\pi_{ijk} / (1 - \pi_{ijk})) = \lambda_0 + \lambda_i^w + w_{k(i)} + \lambda_j^s + \lambda_{ij}^{ws} + s_{jk(i)} \quad (3.1)$$

where all terms are defined as in the models of the previous section. The terms of λ_i^w , λ_j^s , and λ_{ij}^{ws} are set to be zero in the study of Type I error rates, and set to selected values for the study of

power (see Section 3.5). The whole- and sub- plot errors, $w_{k(i)}$ and $s_{jk(i)}$, are simulated from normal distributions $N(0, \sigma_w^2)$ and $N(0, \sigma_s^2)$ respectively, using the SAS[®] RANNOR function. The SAS[®] RANBIN function is used to simulate the binomial data Y_{ijk} 's, $i = 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K$. For each combination of study factors, the data are simulated 1000 times.

3.2 A PRELIMINARY STUDY OF TYPE I ERROR RATES

The seven factors studied in the preliminary study are either directly or indirectly involved in the F or Chi-square tests or in the data simulation process. These factors and their chosen levels are listed in Table 1. The number of replications of each whole-plot treatment (K), number of whole-plot levels (I), and number of subplot levels (J), jointly determine the size of the data set. Levels of I, J , and K are chosen to represent typical “small” and “moderately large” agricultural experiments.

Levels of λ_0 in (3.1) must be chosen. When λ_i, λ_j , and λ_{ij} are all zero, $\log(\pi_{ijk}/(1 - \pi_{ijk})) \sim N(\lambda_0, \sigma_w^2 + \sigma_s^2)$. Thus, λ_0 represents the median value of $\log(\pi_{ijk}/(1 - \pi_{ijk}))$, and hence $\pi = \exp(\lambda_0)/(1 + \exp(\lambda_0))$ represents the median value of the distribution of π_{ijk} . Levels of π are chosen to create both symmetric and skewed binomial distributions.

Levels of N must be chosen as well. Because sample size considerations for chi-square tests are typically based on expected counts, N is set implicitly through the relation $\mu = N\pi$, where μ is the median of the expected counts, $N\pi_{ijk}$, generated by (3.1). Values of μ are chosen to represent cases where normal approximation to the binomial ought to be reasonable. Smaller values of μ are considered in Section 3.4.

The values of σ_w^2 and σ_s^2 in (3.1) also have little intrinsic meaning. Instead, σ_w^2 and σ_s^2 are controlled implicitly by specifying the amount of variation they impart on the generated data.

Because $\text{Var}(\log(\pi_{ijk}/(1 - \pi_{ijk}))) = \sigma_w^2 + \sigma_s^2$, the interval $\lambda_0 \pm 1.96\sqrt{\sigma_w^2 + \sigma_s^2}$ contains the central 95% of the distribution of $\log(\pi_{ijk}/(1 - \pi_{ijk}))$. By transforming this interval, we get an interval that contains 95% of the distribution of π_{ijk} ,

$$\frac{1}{(1 + e^{-\lambda_0 + 1.96\sqrt{\sigma_w^2 + \sigma_s^2}})}, \frac{1}{(1 + e^{-\lambda_0 - 1.96\sqrt{\sigma_w^2 + \sigma_s^2}})} \quad (3.2)$$

Thus, the width of interval (3.2) is representative of the variability of π_{ijk} (and hence $\mu_{ijk} = N\pi_{ijk}$) across experimental units, and therefore also influences the variability of the binomial counts. To specify σ_w^2 and σ_s^2 , the spread rate (s) of π_{ijk} is defined as amount of variation imparted to π_{ijk} by the random effects, expressed as a percentage of the median value π ,

$$s = (\text{half of width of interval (2.2)} / \pi) * 100\% \quad (3.3)$$

To distinguish the relative amounts of whole-plot and subplot variation, the ratio, r , is defined as

$$r = \sigma_w^2 / \sigma_s^2 \quad (3.4)$$

For each specified median probability of success (π), spread rate (s) and ratio (r), we can get the whole-plot and subplot variances, σ_w^2 and σ_s^2 , by jointly solving equations (3.3) and (3.4). Levels of s are chosen to represent relatively high and low overall variability and levels of r are chosen to allow variation in the relative sources of variability. The levels listed in Table 1 result in the values of σ_w^2 and σ_s^2 given in Table 2.

An initial simulation experiment is conducted to screen out factors that do not have an appreciable impact on the error rates, and also to identify which methods work consistently well. To keep simulation times reasonable, a fractional factorial experiment is designed based on all

seven factors. The experiment has 16 runs and achieves resolution IV (Wu and Hamada, 2000), meaning in particular that all main effects are aliased only with three-factor and higher interactions. Observed Type I error rates for this experiment are in Table 3.

The methods with mixed and generalized linear mixed models generally hold their Type I error rates well for all tests. However, for whole-plot tests, these methods tend to be conservative, especially when there are not many replications (K is at the low level). The methods with generalized linear models tend to be liberal for whole-plot tests, with estimated Type I error rates as high as 0.998 for uncorrected methods (GLM and GLMW), 0.66 for the overdispersion-corrected analysis (GLMO), and nearly 0.5 for the GLMGEE method. For subplot and interaction tests, the GLM, GLMW, and GLMGEE continue to be excessively liberal, while GLMO is somewhat better (all estimated Type I error rates ≤ 0.094). The results in this study indicate that methods with mixed or generalized linear mixed models, where random effects are modeled directly, hold Type I error rates better than the methods with generalized linear models, which model non-normality but not random effects. This finding is similar to the one found by Grossardt (2003) for analyzing Poisson data in a split-plot design, by Beckman and Stroup (2003) for analyzing binary repeated-measures data, and by Sui and Stroup (2001) for analyzing multinomial repeated-measures data. T-tests done to test the factor effects show that the main effects of μ , K , and s are possibly significant for the three methods that best maintain their error rates (MIXED, MIXEDT, and GLMM). Follow-up studies focus on the effects of these three factors on the MIXED, MIXEDT, and GLMM methods.

In addition to the angular transformation, the Freeman-Tukey transformation is also used in the study. The results are very similar to those from the angular transformation and not reported here. Similarly, for the GLMGEE method, three different variance-covariance structures (unstructured, compound symmetry, and auto regression) are used in the study. Only the results of the method with unstructured variance-covariance matrix are reported here because others' results are quite similar.

3.3 A FOLLOW-UP STUDY OF TYPE I ERROR RATES

In Table 3, an interesting result is observed that Type I error rates for whole-plot tests tend to be conservative when replicates are low, especially when combined with low expected counts or low spread rates. For binomial data, low counts mean low variability, which could lead to negative whole-plot variance component estimates in SAS[®] PROC MIXED. Even with high counts, negative estimates can result from having too little data to accurately estimate variance components. SAS[®] handles this problem by setting the lower boundary of variance component estimates to be zero. Stroup and Littell (2002) note that this strategy makes the estimates realistic but induces a bias upon them, which leads to conservative whole-plot and liberal subplot tests. This is exactly what is observed in Table 3. They suggest removing the lower bound of zero to reduce bias and improve the tests.

The percentages of zero estimates of the whole plot variance components in our simulation study are reported on Table 3. Runs 1-6, 8, and 10, where the conservative whole-plot and liberal subplot tests occur, have generally high percentages of zero estimates of the whole plot variance components. To assess the effect of this phenomenon on Type I error rates, a NOBOUND option is used to remove the lower bound of zero and the simulations are rerun.

Only the apparent significant factors, μ , K , and s , are studied at both high and low levels. The rest of the factors, π , I , J , and r , are fixed to be 0.25, 5, 5, and 1.

After removing the lower bound on variance component estimates, all whole-plot, subplot, and interaction tests for MIXED, MIXEDT, and GLMM hold Type I error rates within a 95% confidence interval of 0.05 (Table 4). Compared to the tests with the lower bound left intact (Table 3), this adjustment improves the tests' ability to hold Type I error rates dramatically, especially for the whole-plot and interaction tests. All further simulations in this article are performed without the lower bound.

3.4 TYPE I ERROR RATES WITH SMALL EXPECTED COUNTS

Normal linear mixed models assume sampling from normal populations, which is violated with binomial counts. The quality of the normal approximation to the binomial depends on the expected count (Agresti, 1996). For $\pi \leq .05$, the closeness of the approximation deteriorates as $\mu \rightarrow 0$. The initial simulations are conducted under favorable circumstances ($\mu \geq 5$) where the normal approximation might be expected to hold reasonably well. To study the type I error rates of the selected methods under small expected counts, values of μ are reduced to 5, 2.5, 1, and 0.5. The numbers of whole-plot replications, K , are 3, 10, and 20 in expectation that the effects of small expected counts within a plot may be modified by increasing the number of plots. Other factors (π , I , J , r , and s) are set to be 0.25, 5, 5, 1, and 0.1, respectively, and data are simulated as before. The observed rejection rates for this study are in Table 5. Mixed model methods (MIXED and MIXEDT) still hold Type I error rates close to 0.05 in all the simulations except one slightly low value of 0.035. GLMM tends to be conservative when μ is smaller than 5, and the convergence rates drop as well. The trends are offset somewhat when the number of replications increases. However, when μ is extremely small (0.5), the tests are still conservative when the number of replications is increased to 20, which is higher than what is normally used in agricultural and biological experiments.

3.5 A STUDY OF POWER

The power of the three methods (MIXED, MIXEDT, and GLMM) that hold Type I error rate properly is examined in this power study. The factors K , I , J , π , s , and r are fixed to be 3, 3, 3, 0.25, 0.25, and 1 respectively. The median count, μ , is varied at the levels of 1, 5, and 100. The resulting numbers of binomial trials, N , are 4, 20, and 400 respectively.

Power functions and non-centrality parameters for hypothesis tests in split-plot designs are rather complicated. They are functions of the number of replicates, whole-plot levels, subplot levels, and sum of squares of the corresponding treatment effects (Kanji and Liu, 1984). Generally, however, they depend on the treatment effects only through the magnitude of the sum of squared deviations of the effects and not through the pattern of the effects. For convenience, we use the following patterns are used to incorporate the treatment effects in the study:

- (1) Whole-plot effects: $\lambda, \lambda, -2\lambda$
- (2) Subplot effects: $\lambda, \lambda, -2\lambda$
- (3) Whole by sub plot interaction effects:

		Whole plot		
		1	2	3
Subplot	1	$\sqrt{4.5} \lambda$	0	$-\sqrt{4.5} \lambda$
	2	0	0	0
	3	$-\sqrt{4.5} \lambda$	0	$\sqrt{4.5} \lambda$

The sums of squares of the whole and sub plot effects are $3*3*(\lambda^2 + \lambda^2 + (2\lambda)^2) = 54\lambda^2$. The sum of squares of their interaction effect is $3*((\sqrt{4.5} \lambda)^2 + (\sqrt{4.5} \lambda)^2 + (\sqrt{4.5} \lambda)^2 + (\sqrt{4.5} \lambda)^2) = 54\lambda^2$ as well. The values of λ are chosen by “trial-and-error” to achieve a series of rejection rates that fit a smooth curve over the range from 0 to 1. For each of the tests, $\lambda = 0$ represents a study of Type I error rates, and $\lambda > 0$ represents a study of power. One thousand sets of data are simulated from Model (3.1) for each λ and μ combination in a similar way to the Type I error rate study of the previous sections. The rejection rate results are displayed in Figures 1.

Among the three methods tested, none of them stands out when μ is at least 5. The Type I error rates are controlled within the 95% confidence interval of 0.05. However, when $\mu = 1$, the convergence rates of the GLIMMIX[®] macro drop down dramatically (most of them are less than 50%) and all the tests tend to be very conservative. Therefore, no results of GLMM method are shown on the figures when $\mu = 1$. The MIXED and MIXEDT methods have exactly the same Type I error rates when $\mu=1$.

4. CONCLUSIONS

In summary, the methods with mixed (MIXED and MIXEDT) and generalized linear mixed models (GLMM) can hold Type I error rates near a nominal .05 level most of time. The whole-plot tests tend to be conservative when the spread rate and median count are low, in other words, when median counts for all treatments in the experiment are low and close to each other. This difficulty can be overcome by increasing the number of whole-plot replicates in designing the experiment or removing the lower constraint on the mixed models’ variance component estimates in the analysis. Although this latter approach may lead to negative variance component estimates, all tests from MIXED, MIXEDT, and GLMM methods hold Type I error rates close to α . For GLMM method, tests become conservative when μ is smaller than 5. The GLIMMIX macro has serious convergence problems when the median count is smaller than 5 and the number of whole-plot replications is not large. Therefore, mixed model methods are preferred when a lot of small counts occur. The methods based on generalized linear models cannot hold Type I error rates appropriately and should not be used to analyze binomial data in split-plot designs.

While this work demonstrates favorable performance of the classical mixed model approach, we should not rule out the new techniques completely. There are certain limits to this study. The binomial data in the study are simulated with Model 3.1. In real problems, if the mechanism that produces data cannot be represented adequately by this model, different conclusions might be drawn. Also, these simulations are conducted under the assumption that each experimental unit experiences the same number of binomial trials. Possible future work can also be done to compare the methods under unequal N for each treatment combination.

Nonetheless, these results, along with those of Grossardt (2003), Beckman and Stroup (2003), and Sui and Stroup (2001) provide evidence to suggest that normal-based mixed model procedures are reasonably robust against deviations from normality. Also, it appears that proper modeling of random effects is much more important in an analysis than exactly matching the parent distribution of the data.

5. ACKNOWLEDGMENTS

We thank Jeffrey Pontius, Shie-shien Yang, and Dallas Johnson for helpful discussion and/or review of an earlier draft of this article.

6. REFERENCES

- Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, New York: John Wiley and Sons.
- Beckman, M., and Stroup, W. W. (2003), "Small Sample Power Characteristics of Generalized Mixed Model Procedures for Binary Repeated Measure Data Using SAS", *Proceedings of the Fifteenth Kansas State University Conference on Applied Statistics in Agriculture*, 2003: 15-35.
- Collett, D. (1991), *Modelling Binary Data*, Chapman and Hall.
- Cox, D. R., and Snell, E. J. (1989), *Analysis of Binary Data*, 2nd edition, Chapman and Hall.
- Fang, L., and Subramanyam, Bh. (2003), "Activity of spinosad against *Rythopertha dominica* (F.) (Coleoptera: Bostrichidae) is not affected by wheat temperature and moisture", *Journal of Kansas Entomological Society*, 76(3): 529-532.
- Garrett, K.A., Madden, L.V., Hughes, G., and Pfender, W.F. (2004). "New Applications of Statistical Tools in Plant Pathology", *Phytopathology*, 94:999-1003.
- Grossardt, B. R. (2003), "A study of Type I Error Rates in the Analysis of Count Data in a Split-plot Design", Masters Report, Kansas State University.
- Huang, Y. J., Toscano-Underwood, C., Fitt, B. D. L., Todd, A. D., West, J. S., Koopmann, B., and Balesdent, M. H. (2001), "Effects of temperature on germination and hyphal growth from ascospores of A-group and B-group *Leptosphaeria maculans* (phoma stem canker of oilseed rape)", *Ann. Applied Biology*, 139: 193-207.
- Kanji, G. K. and Liu, C. K. (1984), "Power Aspects of Split-plot Designs", *The Statistician*, 33(3): 301-311.
- Liang, K.Y, and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22
- McCulloch, C. E. (1996), "Introduction to Generalized Linear Mixed Models", *Proceedings of the Eighth Annual Kansas State University Conference on Applied Statistics in Agriculture*, 1996: 1-17.
- Milliken, G. A. and Johnson, D. E. (1992), *Analysis of Messy Data, Volume I: Designed Experiments*, New York: Chapman and Hall.
- Stroup, W. W., and Littell, R. (2002), "Impact of Variance Component Estimates on Fixed Effect Inference in Unbalanced Linear Mixed Models", *Proceedings of the Fourteenth Kansas State University Conference on Applied Statistics in Agriculture*, 2002: 94-110

Sui, Y., and Stroup, W.W. (2001), "A Comparison of Some Methods to Analyze Repeated Measures Ordinal Categorical Data", *Proceedings of the Thirteenth Kansas State University Conference on Applied Statistics in Agriculture*, 2001: 98-115.

Wu, C.F. J., and Hamada, M. (2000), *Experiments Planning, Analysis, and Parameter Design Optimization*, New York: John Wiley and Sons.

Zar, J.H. (1999), *Biostatistical Analysis*, 4th edition. New Jersey: Prentice Hall.

Table 1. Factor levels in the first set of simulations

Factor	Values	
Mean number of successes (μ)	5	100
Probability of success (π)	0.1	0.5
Number of replications (K)	3	10
Number of whole-plot levels (I)	2	5
Number of sub-plot levels (J)	2	5
Spread rate (variability) of success probs (s)	10%	50%
Ratio of σ_w^2/σ_s^2 (r)	0.25	4

Table 2. Whole- and sub- plot variances used in the first set of simulations

Probability of success π	Spread rate s	Ratio r	Whole-plot variance σ_w^2	Sub-plot variance σ_s^2
0.1	0.1	0.25	0.00064	0.00257
0.5	0.1	0.25	0.00210	0.00839
0.1	0.5	0.25	0.01538	0.06152
0.5	0.5	0.25	0.06284	0.25134
0.1	0.1	4	0.00257	0.00064
0.5	0.1	4	0.00839	0.00210
0.1	0.5	4	0.06152	0.01538
0.5	0.5	4	0.25134	0.06284

Table 3. Factor level combinations and observed Type I error rates for the preliminary simulation study.

<i>K</i>	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
<i>s</i>	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+
μ	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
π	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
<i>r</i>	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-	+
<i>I</i>	-	-	+	+	+	+	-	-	+	+	-	-	-	-	+	+
<i>J</i>	-	+	+	-	-	+	+	-	+	-	-	+	+	-	-	+
run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Method	Whole-Plot Main Effect															
MIXED	0.015	0.009	0.015	0.031	0.029	0.017	0.038	0.025	0.037	0.039	0.040	0.043	0.054	0.051	0.047	0.049
MIXEDT	0.017	0.009	0.015	0.031	0.027	0.016	0.037	0.025	0.042	0.038	0.040	0.042	0.052	0.052	0.046	0.049
GLMM	0.016	0.008	0.016	0.031	0.024	0.015	0.037	0.025	0.038	0.034	0.040	0.043	0.054	0.050	0.048	0.048
GLM	0.048	0.063	0.176	0.309	0.237	0.358	0.751	0.621	0.073	0.063	0.102	0.148	0.151	0.202	0.904	0.998
GLMW	0.046	0.063	0.175	0.308	0.214	0.347	0.751	0.620	0.074	0.062	0.102	0.148	0.151	0.202	0.902	0.998
GLMO	0.067	0.067	0.083	0.155	0.139	0.128	0.351	0.117	0.067	0.062	0.066	0.098	0.097	0.101	0.092	0.663
GLMGEE	0.162	0.173	0.455	0.486	0.498	0.467	0.194	0.192	0.155	0.128	0.068	0.083	0.095	0.081	0.141	0.138
Method	Subplot Main Effect															
MIXED	0.054	0.057	0.049	0.057	0.049	0.057	0.051	0.067	0.050	0.068	0.058	0.052	0.049	0.052	0.047	0.052
MIXEDT	0.055	0.056	0.050	0.057	0.052	0.054	0.052	0.064	0.051	0.065	0.057	0.052	0.049	0.049	0.045	0.052
GLMM	0.051	0.050	0.049	0.057	0.050	0.053	0.055	0.064	0.046	0.065	0.056	0.052	0.050	0.048	0.043	0.053
GLM	0.051	0.051	0.114	0.071	0.066	0.190	0.419	0.558	0.045	0.070	0.061	0.155	0.114	0.052	0.442	0.623
GLMW	0.048	0.041	0.113	0.070	0.064	0.184	0.415	0.557	0.046	0.070	0.061	0.155	0.112	0.052	0.442	0.623
GLMO	0.082	0.074	0.052	0.031	0.042	0.044	0.006	0.064	0.045	0.065	0.042	0.047	0.046	0.019	0.030	0.000
GLMGEE	0.178	0.797	0.406	0.157	0.144	0.401	0.812	0.190	0.105	0.079	0.080	0.163	0.194	0.076	0.064	0.127

Applied Statistics in Agriculture

<i>K</i>	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
<i>s</i>	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+
μ	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
π	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
<i>r</i>	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-	+
<i>I</i>	-	-	+	+	+	+	-	-	+	+	-	-	-	-	+	+
<i>J</i>	-	+	+	-	-	+	+	-	+	-	-	+	+	-	-	+
run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Method	Whole-Plot by Subplot Interaction															
MIXED	0.069	0.063	0.049	0.064	0.045	0.055	0.055	0.049	0.052	0.056	0.057	0.051	0.051	0.053	0.054	0.045
MIXEDT	0.070	0.062	0.047	0.063	0.037	0.055	0.056	0.047	0.053	0.054	0.056	0.051	0.044	0.056	0.050	0.044
GLMM	0.073	0.054	0.049	0.063	0.037	0.028	0.057	0.047	0.050	0.051	0.057	0.051	0.049	0.051	0.058	0.043
GLM	0.067	0.071	0.177	0.065	0.061	0.377	0.434	0.578	0.054	0.062	0.068	0.150	0.106	0.057	0.858	0.977
GLMW	0.061	0.066	0.174	0.065	0.051	0.340	0.433	0.578	0.047	0.061	0.068	0.150	0.105	0.057	0.857	0.977
GLMO	0.094	0.085	0.074	0.031	0.037	0.050	0.010	0.056	0.061	0.057	0.043	0.047	0.050	0.025	0.027	0.000
GLMGEE	0.200	0.814	0.977	0.443	0.476	0.975	0.833	0.183	0.542	0.141	0.090	0.163	0.174	0.088	0.162	0.545
Method	Percentage of Zero Estimates of the Whole-plot Variance Components															
MIXED	0.513	0.505	0.402	0.200	0.280	0.262	0.011	0.365	0.438	0.481	0.241	0.230	0.333	0.075	0.125	0.000
MIXEDT	0.499	0.506	0.400	0.200	0.268	0.267	0.011	0.365	0.440	0.481	0.243	0.230	0.335	0.072	0.122	0.000
GLMM	0.498	0.490	0.387	0.197	0.268	0.257	0.009	0.348	0.424	0.471	0.237	0.223	0.326	0.072	0.110	0.000

1. Factors with negative signs have low-level values; factors with positive signs have high-level values (see Table 1).
2. Black background means the observed Type I error rates are larger than 0.064 (liberal), gray background means smaller than 0.036 (conservative), and white background means between 0.036 and 0.064.

Table 4. Observed Type I error rates using selected methods without lower bound constraints on covariance parameter estimates.

μ	-	-	-	-	+	+	+	+
K	+	-	+	-	+	-	+	-
s	+	+	-	-	+	+	-	-
run	1	2	3	4	5	6	7	8
Method	Whole-Plot Main Effect							
MIXED	0.043	0.055	0.050	0.050	0.056	0.052	0.046	0.048
MIXEDT	0.040	0.056	0.047	0.052	0.053	0.053	0.045	0.048
GLMM	0.042	0.058	0.049	0.053	0.056	0.052	0.045	0.048
Method	Sub-Plot Main Effect							
MIXED	0.051	0.051	0.061	0.052	0.045	0.057	0.051	0.053
MIXEDT	0.053	0.048	0.055	0.048	0.046	0.055	0.050	0.052
GLMM	0.048	0.048	0.057	0.044	0.048	0.061	0.048	0.056
Method	Whole-Plot by Sub-Plot Interaction							
MIXED	0.044	0.051	0.046	0.062	0.049	0.043	0.045	0.048
MIXEDT	0.045	0.051	0.050	0.063	0.051	0.040	0.044	0.046
GLMM	0.046	0.039	0.040	0.045	0.055	0.049	0.046	0.047

Table 5. Observed Type I error rates with various levels of expected counts and replications

μ	5			2.5			1			0.5		
K	3	10	20	3	10	20	3	10	20	3	10	20
Methods	Whole-plot Tests											
1	0.051	0.045	0.048	0.045	0.051	0.059	0.044	0.056	0.049	0.048	0.049	0.052
2	0.052	0.048	0.047	0.043	0.053	0.056	0.043	0.054	0.049	0.048	0.049	0.052
7	0.047	0.044	0.050	0.037	0.049	0.057	0.009	0.051	0.046	0.000	0.029	0.048
	Sub-plot Tests											
1	0.045	0.040	0.053	0.060	0.047	0.039	0.038	0.036	0.048	0.047	0.047	0.053
2	0.051	0.038	0.047	0.058	0.041	0.042	0.038	0.035	0.050	0.047	0.047	0.053
7	0.045	0.039	0.050	0.051	0.045	0.037	0.007	0.036	0.051	0.000	0.035	0.052
	Interaction of Whole and Sub plots Tests											
1	0.048	0.044	0.048	0.052	0.042	0.044	0.045	0.055	0.048	0.048	0.052	0.044
2	0.050	0.047	0.048	0.054	0.046	0.041	0.044	0.056	0.048	0.048	0.052	0.044
7	0.033	0.038	0.046	0.028	0.032	0.041	0.004	0.027	0.039	0.000	0.009	0.030
Convergence Rates (%) of Method 7	99.7	100	100	98.5	100	100	76.0	100	100	16.3	97.0	100

The observed Type I error rate is smaller than 0.036 if the cell background is gray, and between 0.036 and 0.064 if the cell background is white.

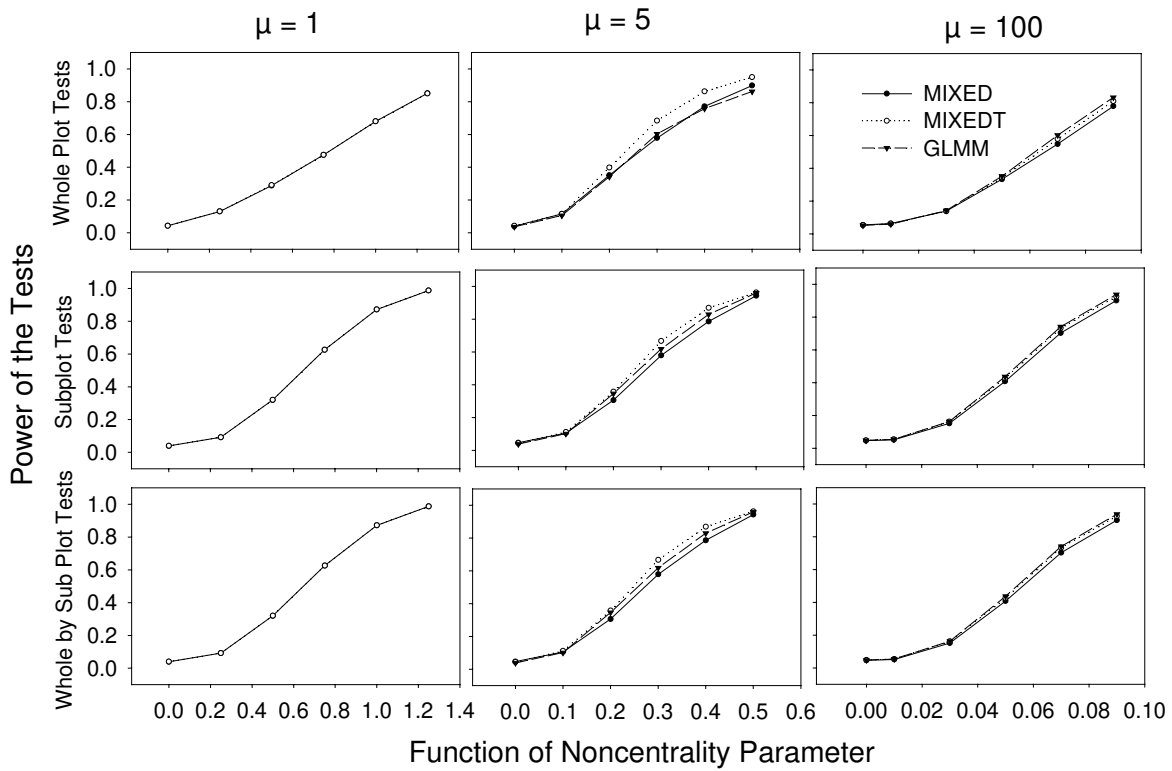


Figure 1. The power is estimated based on 1,000 simulated sets of data. The estimated Type I error rates of all tests shown on the graph are within the 95% confidence interval of 0.05. The GLMM method has very low convergence rate when $\mu = 1$, and the results are not shown here.