EMPIRICAL ESTIMATES OF POWER FOR BINOMIAL DATA WITH MIXED MODELS

R. K. Splan

L. D. Van Vleck

H. D. Hafs

EMPIRICAL ESTIMATES OF POWER FOR BINOMIAL DATA WITH MIXED MODELS

R. K. Splan
University of Nebraska, Lincoln

L. D. Van Vleck
Roman L. Hruska U. S. Meat Animal Research Center, Lincoln

H. D. Hafs
Rutgers University, New Brunswick, New Jersey

ABSTRACT

Observations on return to estrus from anestrus postpartum beef cows were used as the basis for a simulation study to develop a method to determine numbers of locations and animals per treatment per location to achieve a specified power of test. Estimates of among location and total variance were obtained by REML from the data set and then used to generate simulated data for the binomial trait. Each combination of several pre-determined factors was replicated 1000 times. Pre-determined factors were number of locations, number of animals per treatment per location, desired detectable difference due to treatment, alpha-probability level and ratio of among location to total variance. Two methods were used to test for treatment differences. In Method 1, simulated data were analyzed using a mixed model with the variance components used for the simulation based on estimates from the postpartum cow data. For Method 2, variance components were re-estimated from each replicate of the simulated data and used in the mixed model equations. The number of significant differences due to treatment was counted for the 1000 replicates. The fraction of replicates with significant differences is an empirical estimate of the power of the test. The comparison of power of test between the two methods indicates Method 2 may be preferable for empirical estimation of power of test.
Keywords: Simulation, Power, Binomial data

1 INTRODUCTION

Before conducting an experiment or trial, researchers are often concerned with determining the number of experimental units required for detection of treatment differences. If there are large treatment differences, detection is likely even with small samples. Detection of small treatment effects, however, requires large samples (Agresti, 1990). The necessary sample size depends on magnitude of the desired detectable difference, variability of the measured response, type of statistical procedure used, type I error rate and cost involved in sampling (Dowdy and Weardon, 1985). For many experimental designs involving normally distributed data and fixed effects models, determination of required sample size, given a specified level of power of test, is easily available in tabular form (e.g., Schuster, 1990; Gill, 1993b). Even with some non-normal data, tables exist for exact values of power (e.g., Cochran and Cox, 1992). However, analysis with non-normal data and random factors in addition to error may be beyond the experience level of many researchers.

Assume

$$y_{ij} = \mu + a_i + l_j + e_{ij} \qquad (1)$$

is the linear mixed model for a normally distributed, continuous unobservable process with

μ = constant

a_i = fixed effect of ith treatment

l_j = random effect of jth location with $l_j \sim N(0, \sigma^2_l)$

e_{ij} = random residual effect with $e_{ij} \sim N(0, \sigma^2)$

Also suppose the observable response is binary ($b_{ij} = 0$ if unfavorable, $b_{ij} = 1$ if favorable).

For the purpose of this paper, the observable response is detection of estrus in cattle ($b_{ij} = 0$ if the female is not in estrus at time of measurement; $b_{ij} = 1$ if the female is in estrus). The researcher was interested in determining if the proportion of females in estrus in one treatment group were significantly different from the proportion of females in estrus in another treatment group (H_0: $a_1 = a_2$).

Methods for calculating power analytically exist for mixed models of this type. However, assumptions made by these procedures may not always be appropriate. In analytical calculation of power for binomial mixed models, variances are assumed to be known (Dobson, 1990). A researcher may have an approximate idea of the variance, based on previous research, but understandably will not know exactly the variance prior to the experiment. Other methods may be based on large-sample approximations (Agresti, 1990; Gill, 1993a). Therefore, the objective of this study was to develop an empirical method to estimate the number of locations (considered to be random effects) and number of animals per treatment per location to achieve a specified power of test for an experiment with binomial measurements and analysis using linear mixed models.

2 SIMULATION STUDY

A sample data set from a study measuring return to estrus of 362 postpartum beef cows served as the basis for a simulation study to develop a method of sample size estimation given a specified power of test for binomial data in mixed models. In the cow data set, three treatments were used at four locations and return to estrus was recorded as either 0 (did not return to estrus) or 1 (returned to estrus). The estimate of among location variance (σ^2_l) was .012 and of total variance (σ^2_t) was .178 from these data. During the experimental period, 51% of cows returned to estrus.

Estimates of variance components from the original data set served only as a basis for determining variance components for use in the simulation. For simulation, σ^2_t was set to .20. The ratio of among location to total variance was set to .05, .10 or .20 ($\sigma^2_l = .01, .02$ or .04, respectively). To sample from a $N(0,1)$ distribution, variance components were then re-expressed as fractions, with the constraint that total variance equal unity. Thus for the cow data $\sigma^2_l* = \sigma^2_l/\sigma^2_t = .05, .10$ or .20 and $\sigma^2* = \sigma^2/\sigma^2_t = .95, .90$ or .80, respectively, where σ^2 is residual variance. The model used for simulation was

$$y_{ij} = \mu + l_i + e_{ij}, \qquad (2)$$

where y_{ij} is the unobservable continuous response of return to estrus with $y_{ij} \sim N(0, 1)$, l_i is the

random effect of location with $l_i \sim N(0, \sigma^2_l*)$ and e_{ij} is the residual effect with $e_{ij} \sim N(0, \sigma^2*)$. The random effects of l_i and e_{ij} were constructed by multiplying a standard normal random deviate by σ_l* or $\sigma*$, respectively. In the simulation, each combination of several pre-determined factors was replicated. Pre-determined factors included ratio of among location to total variance, number of locations (4, 6, 10 or 20) and number of animals per treatment per location (25, 50, 75 or 100).

For the simulation, two treatments were used. One treatment (a_1) was considered a "control" with a mean of .50 (approximately the mean of the postpartum cow data). The mean of the second treatment (a_2) varied (.50, .55, .60, .65, .70 or .75). The difference between the means of the second and first treatments thus corresponds to the desired detectable treatment difference, δ. Once data were simulated to generate y_{ij}, the underlying normal distribution was truncated to yield a binomial response, b_{ij}. For control observations, the truncation point (q_1) was that value for which $P[y_{ij} \geq q_1 = .5]$. Simulated observations on the underlying normal scale were assigned a value of 1 if they were above q_1 and a value of 0 otherwise. For non-control animals, the truncation point (q_2) was that value for which $P[y_{ij} \geq q_2 = .5 + \delta]$. The simulated continuous observations were assigned a value of 1 if they were above q_2 and a value of 0 otherwise.

The researcher was interested in determining if treatment had an effect on the proportion of cows returning to estrus ($H_0: a_1 = a_2$). In the simulation, the alpha probability level, α, or the probability of detecting a treatment difference given none exists, was set at levels of .010, .025, .050 or .100. Each combination of the pre-determined factors (number of locations, number of animals per treatment per location, σ^2_l*, δ and α) was replicated 1000 times.

After simulated data were reclassified as binomial responses, two methods of analysis were applied. For Method 1, data were analyzed by mixed model methods using the modified versions of the original variance components estimated from the postpartum cow data (e.g. $\sigma^2_l* = .05, .10$ or $.20$ and $\sigma^2* = .95, .90$ or $.80$). For Method 2, variance components were re-estimated from each replicate of simulated data. These variance components were then used in constructing and solving the mixed model equations. Estimates of treatment effects and respective inverse elements from the mixed model equations for each method were then used to determine if there was a significant difference due to treatment. The fraction of significant tests out of 1000 was the empirical power of test.

3 SIMULATION RESULTS

A comparison of Methods 1 and 2 with a treatment difference of .00 is given in Table 1 for an experiment with six locations and 75 animals per treatment per location. If no treatment difference exists, power should equal the Type I error rate. Thus with a treatment difference of zero, the estimated power of test should be equal to the alpha probability level, α. For Method 1, in which variance components used in testing treatment differences were fractions of the total variance used in simulation based on estimates from the original data set, power was over-estimated. For example, for an among location to total variance ratio of .05 and $\alpha = .01$, the estimated power from Method 1 was 2.9 percent. For Method 2, the power of 1.1 percent was in closer agreement with the designated alpha level. The same trend of overestimation of power from Method 1 can be seen in Table 2 for an actual treatment difference of .10 (from a control of .50). As expected, overestimation of power was greatest for smaller alpha levels.

The overestimation of power by Method 1 may be explained by the use of the less appropriate variance component estimates in solving the mixed model equations. The among

location variance used in analysis by Method 1 was calculated as .05, .10 or .20 multiplied by the total variance of .20 used for simulation. Residual variance was then calculated as the difference between total variance and among location variance. However, in Method 2, variance components were re-estimated from the simulated binomial data and used in hypothesis testing.

Variance components used in Method 1 fail to account for the shift in the data mean (and variance) due to the imposed truncation point associated with the desired treatment difference. Re-estimation of variance components once simulated data were truncated to a binomial response resulted in a more accurate test of treatment differences than when variances were assumed known in Method 1. In general, estimated total and residual variances were larger for Method 2 than for Method 1. Overestimation of power may therefore be explained by the smaller residual variance assumed with Method 1.

4 APPLICATION OF SIMULATION

Simulation and analysis with Method 2 were then applied to another data set. These data were from a similar study investigating onset of first estrus in 317 beef heifers at six locations. The estimate of among location variance was .0002 and of total variance was .1565. The pre-determined factors for simulation were number of locations (4, 6, 10 or 20), numbers of animals per treatment per location (25 or 50), desired detectable difference due to treatment (.00, .05, .10, .15, .20 or .25) plus a control of .50, alpha probability level (.010, .050 or .100) and ratio of among location to total variance (.001, .010 or .050). The estimates of the power of the test by number of locations and number of animals per treatment per location are in Figures 2, 3 and 4 for three combinations of α and variance ratios. As expected, the estimated power increased for larger alpha levels (Figures 2 and 3). Reduction in the variance ratio from .010 to .001 (Figures 3 and 4) did not significantly change estimates of power, due to the small among location variance. With increased variance due to location, a change in estimated power would have been expected with a change in ratio to total variance of a magnitude of 10.

5 SUMMARY

Determination of sample size may be calculated for studies with binomial data and random effects using traditional methods. However, these methods may not be appropriate as they are based on known variances and may involve large sample approximations. Two methods to estimate sample size, given predefined levels of power of test were developed and applied to data on estrus in beef cattle. The more appropriate method tested for treatment differences using estimates of variance components from REML analysis of the simulated binomial data, rather than estimates from the original binomial data set. This method may be useful for determination of sample size for future experiments with binomial data and mixed models.

ACKNOWLEDGEMENTS

REFERENCES

Agresti, A. 1990. Categorical Data Analysis. John Wiley and Sons. New York.

Burstein, H. 1971 Attribute Sampling: Tables and Explanations. McGraw-Hill. New York.

Cochran, W.G. and G. M. Cox. 1992. Experimental Designs. 2nd Ed. John Wiley and Sons. New York.

Dobson, A. J. 1990. An Introduction to Generalized Linear Models. Chapman and Hall. London.

Dowdy, S. and S. Weardon. 1985. Statistics for Research. John Wiley and Sons. New York.

Gill, J. L. 1993a. Design and Analysis of Experiments in the Animal and Medical Sciences. Vol. I. Digicopy Corp. Ann Arbor, MI.

Gill, J. L. 1993b. Design and Analysis of Experiments in the Animal and Medical Sciences. Vol. III. Digicopy Corp. Ann Arbor, MI.

Schuster, J. J. 1990. Handbook of Sample Size Guidelines for Clinical Trials. CRC Press. Boca Raton, FL.

Table 1. Estimated power of test (%) for a treatment difference of .00 with six locations and 75 animals per treatment per location.

Method	$(\sigma^2_l / \sigma^2_t)^b$	alpha level[a]			
		.010	.025	.050	.100
I[c]	.05	2.9	5.0	7.4	15.2
	.10	2.5	4.6	6.9	14.6
	.20	2.2	4.8	7.2	13.6
II[d]	.05	1.1	2.6	4.8	9.2
	.10	1.5	2.5	4.6	9.6
	.20	1.1	2.8	5.0	10.0

[a]Assumed Type I error rate.
[b]Ratio of variance due to locations to total variance used in simulation of binomial data.
[c]Uses variance components used in simulation to test for treatment differences.
[d]Uses variance components estimated from simulated data to test for treatment differences.

Table 2. Estimated power (%) of test for a treatment difference of .10 with six locations and 75 animals per treatment per location.

Method	$(\sigma^2_l / \sigma^2_t)^b$	alpha level[a]			
		.010	.025	.050	.100
I[c]	.05	86	92	95	97
	.10	86	92	96	99
	.20	87	94	97	98
II[d]	.05	78	86	91	96
	.10	77	86	92	97
	.20	80	89	95	98

[a]Assumed Type I error rate.
[b]Ratio of variance due to locations to total variance used in simulation of binomial data.
[c]Uses variance components used in simulation to test for treatment differences.
[d]Uses variance components estimated from simulated data to test for treatment differences.
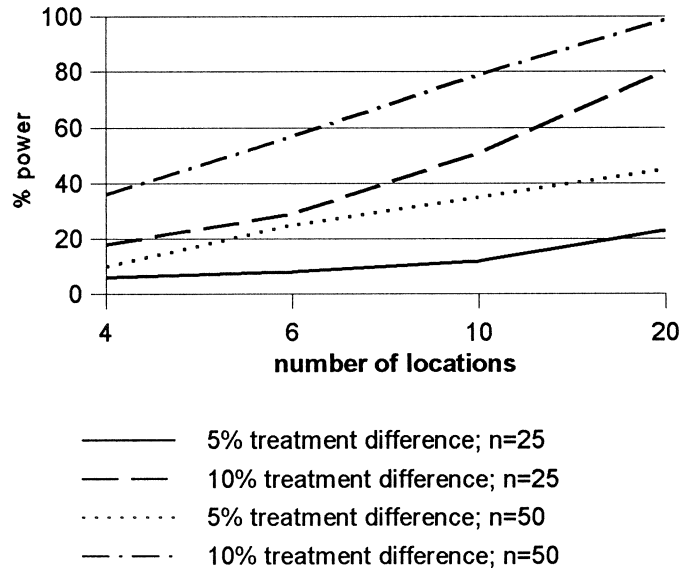
Figure 1. Estimated power of test (%) by number of locations for 25 and 50 animals per treatment per location (n), Type I error rate of .010 and ratio of among location to total variance of .010.
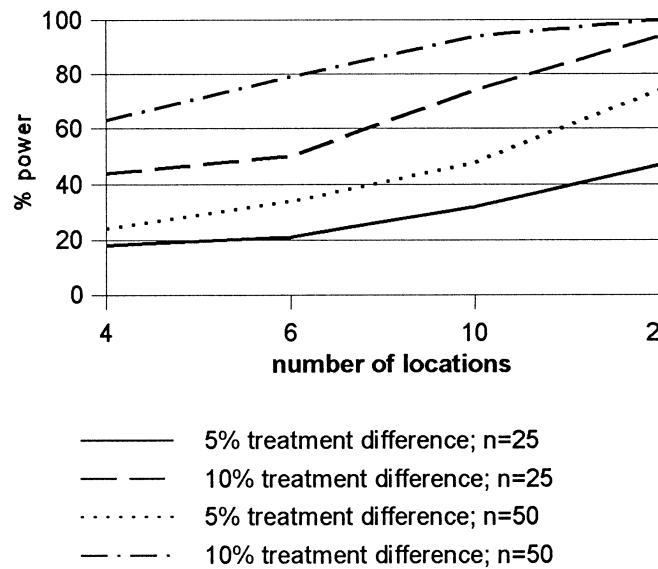

Figure 2. Estimated power (%) of test by number of locations for 25 and 50 animals per treatment per location (n), Type I error rate of .050 and ratio of among location to total variance of .010.
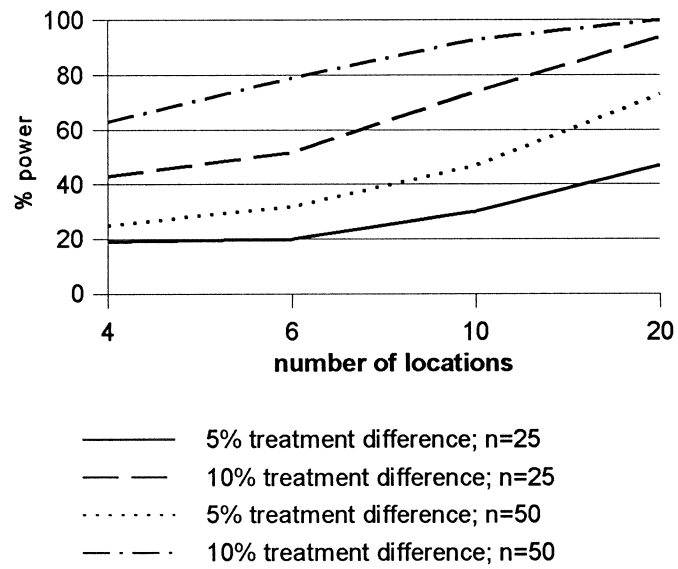
Figure 3. Estimated power of test (%) by number of locations for 25 and 50 animals per treatment per location (n), Type I error rate of .050 and ratio of among location to total variance of .001.