

Repeated Measures with Zeroes

Kenneth N. Berk, Illinois State University, kberk@ilstu.edu
Peter A. Lachenbruch, FDA, lachenbruch@cber.fda.gov

Abstract

Consider repeated measures data with many zeroes. For the case with one grouping factor and one repeated measure, we examine several models, assuming that the nonzero data are roughly lognormal. One of the simplest approaches is to model the zeroes as left-censored observations from the lognormal distribution. A random effect is assumed for subjects. The censored model makes a strong assumption about the relationship between the zeroes and the nonzero values. To check on this, you can instead assume that some of the zeroes are “true” zeroes and model them as Bernoulli. Then the other values are modeled with a censored lognormal. A logistic model is used for the Bernoulli p , the probability of a true nonzero. The fit of the pure left-censored lognormal can be assessed by testing the hypothesis that p is 1 (Moulton and Halsey, 1995). The model can also be simplified by omitting the censoring, leaving a logistic model for the zeroes and a lognormal model for the nonzero values. This is approximately equivalent to modeling the zero and nonzero values separately. In contrast to the censored model, this model assumes only a slight relationship (a covariance component) between the occurrence of zeroes and the size of the nonzero values. The models are compared in terms of an example with data from children’s private speech.

1. Introduction

Suppose you have repeated measures with lots of zeroes. Otherwise, the data are all nonnegative. Consider the simple situation with one between-subjects factor and one within-subjects factor. What can be done for a reasonable analysis?

One way of dealing with the zeroes is to regard them as the result of left censoring. This may be sensible in terms of the way that the data are obtained. The data may be the result of limited measurements, and it may be assumed that many if not all of the zeroes would be positive if only the measurements were more extensive. This is the case with private speech (talking to themselves) data for children. They are observed in a limited number of 10-second intervals, and a zero means that none of the intervals had any private speech. Presumably, increasing the number of intervals would result in fewer zeroes.

How does censoring apply here? Let y_{\min} be the lowest positive observation. Then regard the zeroes as having been left censored at this value. A smaller value could be used, but it should be clear that use of y_{\min} will maximize the likelihood.

Can censoring account for all of the zeroes? In the case of the private speech data, one can imagine that some children may not exhibit any private speech no matter how much they are observed. In a similar situation, Moulton and Halsey (1995, 1996) use a Bernoulli model to allow for additional observations at the low end. They treat each observation as responding or not responding on a Bernoulli basis, and then the responders are regarded as left censored observations from a known distribution such as lognormal or gamma. Here the lognormal will be used. This means that a zero observation can be a Bernoulli nonresponder or it can be a responder that is a left censored value from the lognormal distribution. As demonstrated by Moulton and Halsey, the Bernoulli model offers the opportunity to test the adequacy of the censored model. It is just a matter of testing whether the Bernoulli part is needed at all, which is the question of whether there is zero probability of nonresponse.

Lachenbruch (1977) compared independent groups when the data are positive, except for a bunch of zeroes. He did a binomial test to compare zeroes/nonzeroes and then a nonparametric test for the medians of the positive values, which yielded an approximately normal statistic for each. He then added the squares of these two values to get an approximate chi-square with two degrees of freedom, under the null hypothesis of no difference between groups.

Feuerverger (1979) modeled cloud seeding responses by using a Bernoulli distribution for the presence of rain and a gamma distribution for the amount of rain. The logistic distribution is used to model the Bernoulli probability. In this model the presence of rain and the amount of rain are estimated separately. The model of Moulton and Halsey (1996) for antibody response is similar but more complicated because they included left censoring.

Zhou and Tu (1999) compared K independent groups, assuming a Bernoulli model for the zeroes and a lognormal model for the positive values. Their null hypothesis is that the means of the K distributions are the same, and they test this using a likelihood ratio test.

2. Bernoulli and Lognormal with Left Censoring

The model combines elements of logistic modeling for the Bernoulli success probability, the normal distribution for the logs of the observations, and censoring. For simplicity assume two groups of subjects and three measurements on each subject. Let y_{ijk} be an observation from the k th measurement on the j th subject in the i th group of subjects, where the y_{ijk} are all nonnegative, and let y_{min} be the smallest positive value. Let p_{ijk} be the Bernoulli probability of a nonzero value and let z_{ijk} be 1 for a zero value and 0 for a nonzero value. We use a logistic model for the Bernoulli probability. That is, for the log odds of the p_{ijk} we have the linear model

$$\ln(p_{ijk}/(1-p_{ijk})) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \delta_{ij}$$

where x_1 is an indicator for the groups, x_2 and x_3 are indicators for the three measures, and $x_4 = x_1 * x_2$, $x_5 = x_1 * x_3$ are interactions. The random effect δ_{ij} is assumed to be normal, mean 0,

constant variance τ^2 , with independence between subjects. There is a similar linear model for the mean μ_{ijk} of $\ln(y_{ijk})$,

$$\mu_{ijk} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + \varepsilon_{ij}$$

The standard deviation of $\ln(y_{ijk})$ is assumed to be a constant σ . Here the random effect ε_{ij} is assumed to be normal, mean 0, constant variance v^2 , with independence between subjects. Indeed, together the random effects are assumed bivariate normal with independence between subjects. Call the covariance $\bar{\omega}$.

Given the random effects, the likelihood contribution for a single observation is

$$f(y_{ijk} | \delta_{ij}, \varepsilon_{ij}) = [(1 - p_{ijk}) + p_{ijk} \Phi(\frac{\ln(y_{\min}) - \mu_{ijk}}{\sigma})]^{z_{ijk}} [p_{ijk} \phi(\frac{\ln(y_{ijk}) - \mu_{ijk}}{\sigma})]^{1 - z_{ijk}}$$

Here ϕ is the standard normal density and Φ is the standard normal cumulative distribution function. The likelihood contribution is a product of two terms, where the first corresponds to the zero observations and the second corresponds to the nonzero observations. Omitting the Φ (censoring) term would almost separate the binomial and lognormal modeling, as treated in the two-part models of Lachenbruch (1977). There would still be a slight coupling from the covariance $\bar{\omega}$.

The conditional likelihood for a single individual is obtained as a product

$$g(y_{ij1}, y_{ij2}, y_{ij3} | \delta_{ij}, \varepsilon_{ij}) = \prod_k f(y_{ijk} | \delta_{ij}, \varepsilon_{ij})$$

Letting $q(\delta_{ij}, \varepsilon_{ij})$ be the bivariate normal density of the random effects, the marginal likelihood for a single individual is

$$h(y_{ij1}, y_{ij2}, y_{ij3}) = \int g(y_{ij1}, y_{ij2}, y_{ij3} | \delta_{ij}, \varepsilon_{ij}) q(\delta_{ij}, \varepsilon_{ij}) d\delta_{ij} d\varepsilon_{ij}$$

Finally, the likelihood to be maximized is

$$L = \prod_{i,j} h(y_{ij1}, y_{ij2}, y_{ij3})$$

Fortunately, the new NLMIXED procedure from SAS handles the hard work of maximization.

3. An Example

Consider data from a study of private speech of children (18 males and 15 females) who were followed from first through third grades by Bivens and Berk (1990). Each child was observed in around 100 intervals of duration ten seconds in first grade, and the percentage of

intervals containing offtask private speech was recorded (the children were supposed to be doing their math, so offtask private speech is talking to oneself about something other than math). Similarly, each child is given an offtask private speech score in the second and third grades. Thus, the within-subject measure is grade. The between-subject measure is sex.

3.1. Likelihood Results

Here are results for the full model.

Table 1. Full Model with Bernoulli and Censoring

Model	loglike	$-2*\Delta(\text{loglike})$	df	p	effect
(a) con, sex, grade, int	-102.95				
(b) con, sex, grade	-105.35	4.80 (b-a)	4	.31	int
(c) con, sex	-116.22	21.74 (c-b)	4	.00023	grade sex
(d) con grade	-117.87	25.04 (d-b)	2	3.7E-6	sex grade
(e) con	-127.56	19.38 (e-d)	4	.00066	grade
		22.68 (e-c)	2	1.2E-5	sex

The interaction is clearly not significant, but the main effects clearly are significant. There is a decline in offtask private speech from first to third grades and boys use offtask private speech much more than girls, as shown in Table 7 and Table 8.

Now remove the censoring, so each zero observation is a true zero, and the Bernoulli proportion of zeroes is estimated with a logistic model. Except for the covariance component relating the two random effects, this is equivalent to using for each observation a likelihood that is a product of two likelihoods, where these two likelihoods correspond to separate models for the zero and nonzero values. This is in accord with the Lachenbruch (1977) two-part model in which the zeroes and nonzero values are modeled separately. In the present case, a Bernoulli model is used for the zeroes and a lognormal model is used for the nonzero values.

Table 2. Model with Bernoulli but no Censoring

Model	loglike	$-2*\Delta(\text{loglike})$	df	p	effect
(a) con, sex, grade, int	-106.72				
(b) con, sex, grade	-108.52	3.60 (b-a)	4	.46	int
(c) con, sex	-119.21	21.38 (c-b)	4	.00027	grade sex
(d) con grade	-120.90	24.76 (d-b)	2	4.2E-6	sex grade
(e) con	-130.48	19.14 (e-d)	4	.00074	grade
		22.54 (e-c)	2	1.3E-5	sex

There is essentially no difference in the results from this model, as compared to the model with censoring.

Now remove the Bernoulli term, so the model is just a left-censored lognormal.

Table 3. Model with Censoring but no Bernoulli

Model	loglike	$-2*\Delta(\text{loglike})$	df	p	effect
(a) con, sex, grade, int	-106.91				
(b) con, sex, grade	-109.08	4.34 (b-a)	2	.11	int
(c) con, sex	-118.96	19.76 (c-b)	2	5.1E-5	grade sex
(d) con grade	-122.51	26.86 (d-b)	1	2.2E-7	sex grade
(e) con	-132.16	19.30 (e-d)	2	6.4E-5	grade
		26.40 (e-c)	1	2.8E-7	sex

Here there are half as many parameters because there is no logistic model for the Bernoulli parameter. Because the loglikelihoods are about the same, the p-values are much lower. However, the interaction is still not significant.

Here are results for just the Bernoulli model. This model is concerned only with whether an observation is zero or not.

Table 4. Logistic Model for Zero or not Zero

Model	loglike	$-2*\Delta(\text{loglike})$	df	p	effect
(a) con, sex, grade, int	-47.46				
(b) con, sex, grade	-48.98	3.04 (b-a)	2	.22	int
(c) con, sex	-56.15	14.34 (c-b)	2	.00077	grade sex
(d) con grade	-59.84	21.72 (d-b)	1	3.2E-6	sex grade
(e) con	-66.80	13.92 (e-d)	2	.00095	grade
		21.30 (e-c)	1	3.9E-6	sex

Again, this shows significant main effects but no significant interaction.

Now consider the linear model for the logs of the observations, given that they are positive. This model involves no censoring or Bernoulli terms. One of the six cells is empty, because no third grade girls engaged in offtask private speech.

Table 5. Lognormal Model for the Nonzero Data

Model	loglike	$-2*\Delta(\text{loglike})$	df	p	effect
(a) con, sex, grade, int	-59.25				
(b) con, sex, grade	-59.53	.56 (b-a)	2	.76	int
(c) con, sex	-63.06	7.06 (c-b)	2	.029	grade sex
(d) con grade	-61.31	2.56 (d-b)	1	.11	sex grade
(e) con	-63.71	4.80 (e-d)	2	.091	grade
		5.30 (e-c)	1	.021	sex

Comparing Table 4 and Table 5, we see that the largest effects are mainly in Table 4, and are based on the zero vs. nonzero dichotomy. There are only minor differences in the size of the nonzero values, as shown in Table 5.

Notice that the loglikelihood for the uncensored model (the model with Bernoulli for the zeroes and lognormal for the nonzero values) is approximately the sum of the loglikelihoods for the pure Bernoulli and the pure lognormal models. This is appropriate because for each observation the likelihood for the uncensored model is approximately the product of the likelihoods for the Bernoulli and lognormal models. What makes it approximate is the

covariance between the random effects. It is exact (except for roundoff error) for the cases when the covariance estimate turns out to be zero.

Is the Bernoulli term needed? In other words, can we get by with just the censored model? One way to answer this is to use the difference in loglikelihoods. Under the null hypothesis of no need for the Bernoulli terms in the model, twice the difference of loglikelihoods should be approximately chi-square. The degrees of freedom should be equal to the difference in the number of parameters. Note that there are two additional random effects parameters that are included when Bernoulli modeling is used. These are the variance component for the logistic model and the covariance between the two random effects. The total additional parameters are the coefficients for the logistic model plus these two random effects. However, when the logistic model is constrained to give the pure censored model, two of these are on the boundary of their domain. One is the variance component, which is set to 0, and the other is the constant coefficient, which must be set to minus infinity to give zero probability of nonresponse. This situation is studied by Self and Liang (1987). In their Case 7 they examine in detail the situation with two parameters constrained to their boundary. The net effect of this is a loss of somewhere between 0 and 2 degrees of freedom. For example, the full Bernoulli censored lognormal model with main effects and interaction, variance components, and covariance component uses 15 parameters. Of these, 8 are not needed in the censored lognormal model, but two are on the boundary, so there are between 6 and 8 degrees of freedom. More explicitly, the difference of loglikelihoods will be approximately a mixture of chi-squares of 6, 7, and 8 degrees of freedom.

Table 6 shows the results for comparison of Table 1 and Table 3. Under "loglike" it shows the likelihood under the full Bernoulli/censored lognormal model and then the likelihood with just the censored lognormal.

Table 6. Differences between the Full Model and the Censored Model

Model	loglike		$2*\Delta(\text{loglike})$	df	p
	Bern	No Bern			
(a) con, sex, grade, int	-102.95	-106.91	7.92	6-8	.24-.44
(b) con, sex, grade	-105.35	-109.08	7.46	4-6	.11-.28
(c) con, sex	-116.22	-118.96	5.48	2-4	.065-.24
(d) con, grade	-117.87	-122.51	9.28	3-5	.026-.098
(e) con	-127.56	-132.16	9.20	1-3	.002-.027

One can argue that only the first two rows are relevant here, because only the first two rows correspond to adequate models. That is, the second row in Table 1 shows no significant difference relative to the complete model in the first row. The first two rows of Table 6 show no significant difference between the full model and the censored model with no Bernoulli logistic modeling. This suggests that one can get by with the much simpler censored model. This cuts out half of the parameters, those for the logistic model with its variance component, plus the component of covariance between the two random effects. It thereby enhances the power of the tests, given that the loglikelihood differences are similar but the degrees of freedom are halved.

Note that Lindsey (1999) takes a different approach. He uses the Akaike Information Criterion (AIC) instead of likelihood ratio tests. Essentially, this minimizes loglikelihood, except for a penalty equal to the difference in the number of parameters. This means that a smaller model is preferred if the difference in loglikelihoods is less than the difference in the number of parameters used. Perhaps this should be corrected for boundary conditions in the way that Self and Liang (1987) describe for the likelihood ratio test. However, proceeding in the usual way, we compare the models in the first rows of Table 1 and Table 3. The difference in loglikelihoods is 3.96, but there are 8 more parameters in the larger model, so AIC prefers the smaller model with no Bernoulli terms. Even if you apply a Self/Liang correction here, the models differ by at least 6 degrees of freedom, so the smaller model is preferred.

Although the results here suggest that one could do fine without the Bernoulli term, this does not say that one should exclude this term a priori when modeling another repeated measures data set. Certainly, some data sets will have “true” zeroes.

It is interesting that the differences are similar in Tables 1, 2, and 3. In other words, the size of the effect, as measured by the difference in loglikelihood, does not depend much on the type of model. Indeed, the loglikelihoods are very similar for Tables 2 and 3. It does not matter much, in terms of loglikelihood, whether you eliminate the Bernoulli modeling or eliminate censoring. Of course, the big difference here is in the degrees of freedom, because the censored model uses fewer parameters.

3.2. Cell Means and the Occurrence of Zeroes

For the full model with both “true” Bernoulli zeroes and censored zeroes (Table 1, a), Table 7 shows the cell means and standard errors for the fitted lognormal distribution. Note that the cell for third grade girls has no nonzero values.

Table 7. Cell Means for Lognormal Model with Bernoulli and Censoring

		Grade		
		1	2	3
sex	male	1.58 (.27)	.72 (.35)	.87 (.39)
	female	.97 (.43)	-1.67 (.44)	-11.01 (2.31)

For the model with both “true” Bernoulli zeroes and censored zeroes, how do the fitted probabilities compare with the actual fraction of zeroes? Table 8 shows the actual zeroes, the Bernoulli probability p and the censored probability Φ . Here Φ is the probability of a left-censored value from the lognormal. The estimated standard deviation is .997 and the smallest nonzero value is .6, so Φ is the probability of a normal value to the left of $\log(.6)$, given the cell mean (from Table 7) and the standard deviation .997.

Table 8. Zero Probabilities for Model with Bernoulli and Censoring

Grade

	1	2	3
male	3/18=.167	5/18=.278	9/18=.500
	p=.148	p=.185	p=.454
sex	Φ =.018	Φ =.108	Φ =.083
female	8/15=.533	13/15=.867	15/15=1.000
	p=.499	p=.000	p=1.000
	Φ =.068	Φ =.877	Φ =1.000

For the lognormal model with just “true” Bernoulli zeroes and no censoring (Table 2, a), Table 9 shows the cell means and standard errors for the fitted lognormal distribution. Except for the cell for third grade girls with all zero values, the means are all greater than in Table 7, because the left-censored values pull the means to the left.

Table 9. Cell Means for Lognormal Model with Bernoulli and No Censoring

	Grade		
	1	2	3
male	1.62	.93	1.04
sex	(.23)	(.24)	(.29)
female	1.11	-.16	-11.18
	(.33)	(.62)	(0)

For the model with just “true” Bernoulli zeroes and no censoring (Table 3a), how do the fitted probabilities compare with the actual fraction of zeroes? Table 10 shows the actual zeroes and the Bernoulli probability p. Here the Bernoulli p is trying to account for all the zeroes. The differences can be attributed to inaccuracies of estimation.

Table 10. Zero Probabilities for Lognormal Model with Bernoulli and No Censoring

	Grade		
	1	2	3
male	3/18=.167	5/18=.278	9/18=.500
sex	p=.163	p=.275	p=.500
female	8/15=.533	13/15=.867	15/15=1.000
	p=.534	p=.870	p=1.000

For the model with no Bernoulli zeroes and just censored zeroes (Table 3, a), Table 11 shows the cell means and standard errors for the fitted lognormal distribution. Except for the cell for third grade girls with all zeroes, the means here are all less than the corresponding values in Table 9 and Table 7 because the means are influenced by the censored values.

Table 11. Cell Means for Lognormal Model with Left-Censoring

		Grade		
		1	2	3
sex	male	1.15 (.35)	.27 (.36)	-.29 (.39)
	female	-.35 (.43)	-2.27 (.63)	-7.44 (71.37)

For the model with no Bernoulli zeroes but with left-censoring, how do the fitted probabilities compare with the actual fraction of zeroes? Table 12 shows the actual zeroes and the censored probability Φ . Here Φ is the probability of a left censored value from the lognormal. The estimated standard deviation here is 1.471, and the smallest nonzero value is .6, so Φ is the probability of a normal value to the left of $\log(.6)$, given the cell mean (from Table 11) and the standard deviation 1.471.

Table 12. Zero Probabilities for Model with Left-Censoring

		Grade		
		1	2	3
sex	male	3/18=.167 Φ =.130	5/18=.278 Φ =.297	9/18=.500 Φ =.439
	female	8/15=.533 Φ =.456	13/15=.867 Φ =.884	15/15=1.000 Φ =1.000

In Tables 7-12 there are some evident trends. First, there is an association between zero values and low values for the lognormal. That is, if a cell has a high probability of zero, then the cell is likely to have a low mean for the lognormal distribution. Given that the means and the probabilities of zero tend to agree, it makes sense to talk about upward and downward trends in offtask private speech.

The boys engage in more offtask private speech than the girls. With a few exceptions, boys and girls use less offtask private speech as they get older. Given generally downward trends in private speech for both boys and girls, there is no significant interaction. Even though third grade girls display no offtask private speech at all, the models are still able to estimate a mean for the lognormal.

With the association between the occurrence of zeroes and the occurrence of small nonzero values, it makes sense that the left-censored lognormal model would be a pretty good fit. A low mean for the left-censored lognormal implies a high probability of both zero and small nonzero values. On the other hand, if there were a tendency for large nonzero values to occur together with many zeroes, then there would be a need for the "true" zeroes permitted by the more complicated Bernoulli/censored lognormal model.

References

- Bivens, J. A., and Berk, L. E. (1990). A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly* 36, 443-463.
- Feuerverger, A. (1979). On some methods for analysis of weather experiments. *Biometrika* 66, 655-658.
- Lachenbruch, P. A. (1976). Analysis of data with clumping at zero. *Biometrische Zeitschrift* 18, 351-356.
- Lindsey, J. K. (1999). *Models for Repeated Measurements*. Oxford, UK: Oxford University Press.
- Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics* 51, 1570-1578.
- Moulton, L. H. and Halsey, N. A. (1996). A mixed gamma model for regression analyses of quantitative assay data. *Vaccine* 14, 1154-1158.
- Self, S. G. and Liang, K-Y (1987). Estimators and likelihood ratio tests under nonstandard conditions. *JASA* 82, 605-610.
- Zhou, X-H. and Tu, W. (1999). Comparison of several Independent population means when their samples contain log-normal and possibly zero observations. *Biometrics* 55, 645-651.

Appendix

This consists of SAS code for fitting the full left-censored lognormal model with a logistic model for Bernoulli zeroes.

```
*TITLE 'PRIVATE SPEECH STUDY - FOLLOW STUDENTS FROM GRADES 1 TO 3';
DATA PRIVATE;
  INPUT (SEX IQ1 IQ3 VERBIQ   OFFTASK1 OFFTASK2 OFFTASK3
        ONTASK1 ONTASK2 ONTASK3  MUMBLE1 MUMBLE2 MUMBLE3 ) (4*3. 9*4.);
LABEL
SEX      = '1 FOR MALE AND 2 FOR FEMALE'
OFFTASK1 = '1ST GRADE OFFTASK PRIVATE SPEECH'
OFFTASK2 = '2ND GRADE OFFTASK PRIVATE SPEECH'
OFFTASK3 = '3RD GRADE OFFTASK PRIVATE SPEECH';
N=_N_ ;
LINES;
  1113113117 4.9 1.315.336.6 9.233.920.838.228.8
  1108113103 5.5 0.0 0.044.515.5 7.524.442.857.0
  1140122121 6.5 0.9 0.056.017.119.219.448.623.9
  1113116112 0.011.2 2.010.1 9.3 1.033.361.146.9
  1115122120 0.0 3.1 0.012.4 1.6 0.726.013.250.0
  1146134132 3.0 1.9 6.321.220.8 0.056.658.564.6
  1136134113 2.8 3.7 0.029.311.1 2.139.565.554.2
  1107105117 6.4 4.6 0.822.028.4 4.124.730.655.3
  1108122132 1.0 0.0 1.134.011.1 1.121.650.121.4
  1119126149 0.9 3.3 1.932.234.717.732.119.038.3
  1132129125 0.0 1.0 0.020.7 8.0 1.348.158.078.5
  112713313128.1 0.7 7.111.015.8 9.919.534.238.1
  1118122136 8.7 0.0 0.0 8.7 0.0 0.019.250.844.3
  1108133107 1.6 0.0 5.4 6.410.319.843.047.711.7
  1103108108 5.1 5.3 0.023.216.5 5.826.324.758.6
  110312011317.0 3.7 0.013.216.0 0.022.745.776.1
  1122144136 4.7 0.0 0.018.9 3.9 7.949.485.976.4
  111110411428.1 3.0 0.925.0 1.610.335.464.248.6
  2102115114 0.0 0.0 0.025.718.622.956.854.737.2
  2 96102102 1.3 1.2 0.036.017.411.645.443.269.8
  2106119113 2.2 0.0 0.027.6 2.6 3.628.732.929.1
  2118114131 0.0 0.0 0.029.7 0.9 4.442.253.560.4
  2108108124 1.3 0.0 0.036.0 1.5 5.620.383.157.8
  2122111117 0.0 0.6 0.035.114.112.020.012.938.7
  2115122120 0.0 0.0 0.042.0 3.3 4.834.064.546.5
  2113113 90 0.0 0.0 0.0 7.6 1.6 0.026.958.850.0
  2109125114 0.0 0.0 0.014.1 0.0 0.048.461.069.6
  2113118109 3.9 0.0 0.025.0 1.5 0.027.656.169.8
  2 82 96102 0.0 0.0 0.020.2 0.0 7.352.669.759.4
  211011211410.1 0.0 0.024.4 2.1 8.3 5.943.722.7
  2121124127 5.2 0.0 0.010.418.4 3.038.569.284.9
  2110122127 3.2 0.0 0.021.1 2.6 4.622.146.742.0
  2 99115103 0.0 0.0 0.0 5.626.0 0.022.220.067.2
data priv99; set private;
  offtask = offtask1; grade=1; output;
  offtask = offtask2; grade=2; output;
  offtask = offtask3; grade=3; output;
  keep n offtask grade sex;
```

```

data priv99; set priv99;
  maxy = max(offtask, .3);
  logmaxy = log(maxy);
  sexmlp1 = 2*sex - 3;
  gradlin= (grade=3)-(grade=1);
  gradq   = (grade=3)+(grade=1) - 2*(grade=2);
  sexlin  = sexmlp1 * gradlin;
  sexq    = sexmlp1 * gradq;

proc nlmixed data=priv99  corr;
  parms gamma0=-1.42 gammal=-3.17 gamma2=-.475 sig=1 sigeps2=.00054
        gamma3=-2.48 gamma4=-2.82 gamma5=-.64 beta0=1.42 betal=-2.80
        beta2=-3.34 beta3=0.28 beta4=-2.02 beta5=-3.17
        sigdel2=.098 r=.99; *these are the initial values;
  bounds sigeps2 sigdel2 >= 0, -1<=r<=1; *bounds for var and covar comps;
  delta0=(logmaxy<log(.5)); *indicator for zeroes;
  mu = gamma0 + gammal*gradlin + gamma2*gradq + gamma3*sexmlp1+
        gamma4*sexlin + gamma5*sexq + eps; *loglinear model;
  mulogit = beta0 + betal*gradlin + beta2*gradq + beta3*sexmlp1 +
        beta4*sexlin + beta5*sexq + del ; *for logistic model;
  p = 1/(1+exp(-mulogit)); *logistic model;
  like = delta0*log((1-p) + p*probnorm((log(.6) - mu)/sig)) +
        (1-delta0)*(log(p)-.5*((logmaxy-mu)/sig)**2 -log(sig*sqrt(8*atan(1)))));
  *log likelihood;
  model logmaxy ~ general(like);
  random eps del ~ normal([0,0],[sigeps2, r*(sigeps2*sigdel2)**.5, sigdel2])
        subject=n; *random effects, variance and covariance components;
  *in the following estimates, m1 refers to males in the first grade, f1 refers to
  females in the first grade, etc.; *mu is an estimated cell mean;
  *prob is an estimated cell Bernoulli probability;
  estimate 'm1 mu' gamma0 + gammal*(-1) + gamma2*1 + gamma3*(-1)+
        gamma4*1 + gamma5*(-1);
  estimate 'm2 mu' gamma0 + gammal*0 + gamma2*(-2) + gamma3*(-1)+
        gamma4*0 + gamma5*(2);
  estimate 'm3 mu' gamma0 + gammal*(1) + gamma2*1 + gamma3*(-1)+
        gamma4*(-1) + gamma5*(-1);
  estimate 'f1 mu' gamma0 + gammal*(-1) + gamma2*1 + gamma3*(1)+
        gamma4*(-1) + gamma5*(1);
  estimate 'f2 mu' gamma0 + gammal*0 + gamma2*(-2) + gamma3*(1)+
        gamma4*0 + gamma5*(-2);
  estimate 'f3 mu' gamma0 + gammal*(1) + gamma2*1 + gamma3*(1)+
        gamma4*(1) + gamma5*(1);
  estimate 'm1 prob' 1/(1+exp(-(beta0 + betal*(-1) + beta2*1 + beta3*(-1)+
        beta4*1 + beta5*(-1))));
  estimate 'm2 prob' 1/(1+exp(-(beta0 + betal*0 + beta2*(-2) + beta3*(-1)+
        beta4*0 + beta5*(2))));
  estimate 'm3 prob' 1/(1+exp(-(beta0 + betal*(1) + beta2*1 + beta3*(-1)+
        beta4*(-1) + beta5*(-1))));
  estimate 'f1 prob' 1/(1+exp(-(beta0 + betal*(-1) + beta2*1 + beta3*(1)+
        beta4*(-1) + beta5*(1))));
  estimate 'f2 prob' 1/(1+exp(-(beta0 + betal*0 + beta2*(-2) + beta3*(1)+
        beta4*0 + beta5*(-2))));
  estimate 'f3 prob' 1/(1+exp(-(beta0 + betal*(1) + beta2*1 + beta3*(1)+
        beta4*(1) + beta5*(1)))); run;

```