# A marginalized multilevel random effects model for longitudinal semi-continuous data

Gül İnan

*Department of Statistics*
*Middle East Technical University*
*06800-Cankaya, Ankara, Turkey*
*ginan@metu.edu.tr*
0000-0002-3981-9211

**Abstract**

In this study, we proposed a marginalized multilevel random effects model for analysis of longitudinal semi-continuous data. We investigated the performance of the proposed model through a Monte Carlo simulation study under scenarios. The results of the simulation study showed that the proposed model has some favourable statistical properties.
*Keywords:* Augmented zeros; correlated random effects; marginalized models.

**Öz**

***Uzunlamasına kısmi-sürekli veri için marjinalleştirilmiş çok seviyeli rastgele etkili bir model***

*Bu çalışmada, uzunlamasına kısmi-sürekli veri için marjinalleştirilmiş çok seviyeli rastgele etkili bir model önerdik. Önerilen yöntemin başarısını farklı senaryolar altında bir Monte Carlo benzetim çalışmasıyla inceledik. Benzetim çalışmasının sonuçları önerilen modelin yeterli istatistiksel özelliklere sahip olduğunu göstermiştir.*
*Anahtar sözcükler: Artırılmış sıfırlar; ilişkili rastgele etkiler; marjinalleştirilmiş modeller.*

## 1. Introduction

A semi-continuous response variable is a mixture of zeros and positive continuous outcomes that often tend be right-skewed. The semi-continuous response variable can be observed in many disciplines from medical cost studies to insurance claim studies. For example, it is well-known that cost claim data in insurance studies has a semi-continuous nature where the data is divided into two parts: i) whether a claim is occurred or not (resulting in a considerable number of zeros) and ii) the amount of claims given the occurrence of the claim (resulting in a positive valued continuous data).

While traditional regression models working under normal distribution assumption cannot be realistic for semi-continuous data (e.g., may lead to negative predictions), transformation techniques can be used to make the semi-continuous data normally distributed to be able to use these models. However, since transformation changes the scale of the data, this approach may result in lack of interpretation. Furthermore, some commonly used transformations such as a square root transformation leave zeros unchanged and a log-transformation cannot be applied on zeros. Other approaches may be omitting the zeros from the data analysis, which may result in efficiency loss, and adding small positive values to zeros. To avoid from these problems, the recent literature focused on two-part regression models which

accommodate both zero and non-zero values. Recent literature review in this context is available in Neelon et al. [10], Neelon et al. [11], Farewell et al. [2], and Smith et al. [14].

Within the framework of analysis of longitudinal semi-continuous data, where data is collected over time, Olsen and Schafer [12] proposed a two-part random effects model. In their seminal paper, the first part of the model is for predicting the probability of presence of a zero value and the second part of the model is for predicting the mean of non-zero values. Each part of the model is linked to each other via including normally distributed and correlated random effects into the whole model. Following Olsen and Schafer [12], Tom et al. [17] (as a rejoinder of Su et al. [16]) proposed a two-part random effects model, where the first-part of the model (a logistic regression for presence of a zero) now includes a random intercept from a bridge distribution and the second-part of the model (a linear regression for the mean of non-zero values) includes a random intercept from normal distribution. Rodrigues-Motta et al. [13] proposed a two-part random effects model, where the first-part of the model (logistic part for presence of a zero) includes a random intercept from a normal distribution and the second part of the model (linear model for the mean of non-zero) takes form from a broader class of distributions such as Weibull, gamma, log-normal conditionally on a normally distributed random-intercept. Smith et al. [15] also proposed a two-part random effects model, where the first-part of the model (logistic part for presence of a zero) includes a random intercept from a normal distribution and the second part of the model now models the overall mean of zero and non-zero values conditionally on a normally distributed random intercept via a regression model with log link function.

The development of a new regression model may depend on the question of interest as well as the characteristics of the motivating data. In this study, we take the marginalized multilevel models (MMMs) with random effects (Heagerty [5], Heagerty and Zeger [6]) under investigation since they build separate regression models for longitudinal mean and associations of responses to combine the strengths of marginal and conditional models. As a consequence of this model building structure, MMMs take the population-averaged interpretation and robustness of marginal regression parameters from marginal models, whereas they take likelihood-based inference capabilities and flexible specifications for longitudinal associations from conditional models (Griswold et al. [4]). Several extensions of MMMs for longitudinal binary, count, and non-Gaussian continuous data are available in literature. Its extension to longitudinal count data with excess zeros has been discussed by Lee et al. [9] and Kassahun et al. [8] (with a rejoinder by Inan [7]), however, to the best of our knowledge, there is no currently published work on marginalized multilevel models for analysis of longitudinal semi-continuous data except the one by Zhang et al. [18], where they extend the model of Smith et al. [15] to longitudinal semi-continuous data with an overall marginal mean inference.

In this study, we propose a marginalized multilevel random effects regression model for analysis of longitudinal semi-continuous data. The rest of the paper is organized as follows. In Section 2, we present the proposed model and we assess the performance of the proposed method in Section 3. Concluding remarks are given in Section 4.

## 2. Proposed model

Let $Y_{it}$ denote the response of $i$th subject (i = 1,.., N) at $t$th time point (t = 1,…, $n_i$) such that $Y_{it}$ takes any value within the interval [0, + ∞). Then the distribution of $Y_{it}$ is assumed to follow a mixture of the degenerate distribution at zero and of gamma distribution in the following form:

$$\Pr(Y_{it} = y_{it} | p_{0,it}, \mu_{it}^c, \upsilon) = \begin{cases} p_{0,it} & \text{if } y_{it} = 0 \\ (1 - p_{0,it}) f(y_{it} | \mu_{it}^c, \upsilon) & \text{if } y_{it} > 0, \end{cases} \tag{1}$$

where $p_{0,it}$ is the probability of observing a zero value and the probability density function (pdf) of a non-zero response variable is assumed to follow a mean parameterized Gamma distribution of form:

$$f(Y_{it} = y_{it} \mid \mu_{it}^c, \upsilon) = \frac{1}{y_{it} \, \Gamma(\upsilon)} (y_{it} \frac{\upsilon}{\mu_{it}^c})^\upsilon \, e^{-(y_{it} \frac{\upsilon}{\mu_{it}^c})}, \quad \text{for } y_{it} > 0, \qquad (2)$$

where $\mu_{it}^c$ ($\mu_{it}^c > 0$) is the (conditional) mean of the gamma distribution, $\Gamma(.)$ denotes the gamma function, and $\upsilon = \frac{1}{\phi}$ ($\phi > 1$) is the inverse precision parameter, which is held constant across subjects and within the measurements of a subject (e.g., $\upsilon_{it} = \upsilon$ for all i and t) for simplicity.

Under marginalized multilevel random effects regression models framework, a logistic regression model for the marginal mean of zero-part and a gamma regression model with a random-intercept for the (conditional) mean of non-zero part are assumed, respectively, as follows:

$$\begin{aligned} \text{logit}(p_{0,it}) &= \mathbf{X}_{0,it}^T \gamma \\ \log(\mu_{it}^c) &= \Delta_{it} + b_i, \end{aligned} \qquad (3)$$

where logit(.) and log(.) denote the logit and log link functions, respectively, $\mathbf{X}_{0,it} = (X_{0,it1}, ..., X_{0,itp})^T$ is the $p \times 1$ vector of covariates associated with probability of observing a zero value, and $\gamma = (\gamma_1, ..., \gamma_p)^T$ is the corresponding vector of regression coefficients.

The term $\Delta_{it}$ in the non-zero part of (3) is a subject and time specific connector function, which will be clarified later. The logarithm of the mean of the non-zero part is conditioned on the subject-specific random-intercept $b_i$, which is assumed to capture the heterogeneity between the subjects and to follow a standard normal distribution with mean 0 and variance $\sigma^2$. Conditional on the random effects $b_i$, all measurements of each subject are assumed to be independent of each other. Note that although more random effects (e.g., random slopes) can be included into the model, a random intercept form of model is assumed here for simplicity. Furthermore, to avoid from over-parameterization and to obtain computational easiness in the derivation of $\Delta_{it}$, the logistic regression model for $p_{0,it}$ is assumed to be free from random effects as in Bandyopadhyay et al. [1] and Galvis et al. [3].

Then the marginal distribution of $Y_{it}$ is assumed to follow a mixture of the degenerate distribution at zero and of gamma distribution in the following form:

$$Pr(Y_{it} = y_{it} \mid p_{0,it}, \mu_{it}^m, \upsilon) = \begin{cases} p_{0,it} & \text{if } y_{it} = 0 \\ (1-p_{0,it}) \, f(y_{it} \mid \mu_{it}^m, \upsilon) & \text{if } y_{it} > 0, \end{cases} \qquad (4)$$

where $p_{0,it}$ is as defined above and $\mu_{it}^m$ is the marginal mean of non-zero part defined as follows:

$$\log(\mu_{it}^m) = \mathbf{X}_{\text{cont},it}^T \beta, \qquad (5)$$

where $\mathbf{X}_{\text{cont,it}} = (X_{\text{cont,it1}},...,X_{\text{cont,itp}})^T$ is the $r \times 1$ vector of covariates associated with the marginal mean of non-zero part and $\beta = (\beta_1,...,\beta_p)^T$ is the corresponding vector of regression coefficients which have population-based inferences.

### 2.1. Calculation of $\Delta_{it}$

Since any conditional expectation can be written in terms of marginal expectation, it is possible to link the conditional mean of the non-zero part of the model to the marginal mean model of that via integrating the conditional expectation over the distribution of random effects such that:

$$E(Y_{it}) = E_{b_i}(E(Y_{it}|b_i))$$
$$(1\text{-}p_{0,it})\mu_{it}^m = \int (1\text{-}p_{0,it})\mu_{it}^c f(b_i)d_{b_i} , \tag{6}$$

where $f(b_i)$ is the standard normal distribution with mean 0 and variance $\sigma^2$. Substituting marginal and conditional means in (3) and (5), respectively, into (6) gives the following expression:

$$\frac{\exp(\mathbf{X}_{\text{cont,it}}^T\beta)}{1+\exp(\mathbf{X}_{0,it}^T\gamma)} = \int \frac{\exp(\Delta_{it}+b_i)}{1+\exp(\mathbf{X}_{0,it}^T\gamma)}f(b_i)db_i . \tag{7}$$

Solving (7) for $\Delta_{it}$ provides the following closed-form expression:

$$\Delta_{it} = \mathbf{X}_{\text{cont,it}}^T\beta - \frac{\sigma^2}{2}. \tag{8}$$

Apparently, $\Delta_{it}$ is a function of both $\beta$ from marginal part and $\sigma^2$ from conditional part and connects both levels two each other. The detailed information on derivation of $\Delta_{it}$ is available in the Appendix.

### 2.2. Likelihood formulation

Let $\theta = (\beta, \gamma, \upsilon)$ denote the unknown parameters in the proposed model. The joint marginal likelihood of the data $Y = (Y_1,...,Y_N)$ with $Y_i = (Y_{i1},...,Y_{in_i})$ can be defined as follows:

$$L(\theta|Y,b) = \prod_{i=1}^{N}\int\prod_{t=1}^{n_i}(p_{0,it})^{I_{\{y_{it}=0\}}}\{(1\text{-}p_{0,it})f(y_{it}|\mu_{it}^c,\upsilon)\}^{I_{\{y_{it}>0\}}}f(b_i)d_{b_i} , \tag{9}$$

where $I(.)$ denotes the indicator function, $p_{0,it}$ and $\mu_{it}^c$ are as defined in (3) along with $\Delta_{it}$ in (8), and $f(b_i)$ is the standard normal distribution with mean 0 and variance $\sigma^2$. The maximum likelihood estimation of the model parameters will be achieved via SAS NLMIXED procedure.

## 3. Simulation study

### *3.1 Simulation design*

We carried out a Monte Carlo simulation study for assessment of proposed model. In this sense, we assumed two different sample sizes (N=100 and 200) with three different probability of observing a zero value ( $p_{zero}$ = 10%, 15%, and 20%).

We assumed the following marginalized multilevel random effects model as the true model:

$$\text{logit}(p_{0,it}) = \gamma_0 + \gamma_1 x_{i1}$$
$$\log(\mu_{it}^c) = \beta_0 + \beta_1 x_{i1} - \frac{\sigma^2}{2} + b_i, \tag{10}$$

where t = 1,2,3,4 indicating that $n_i$ = 4 for each i (i = 1,…,100 or 200), $x_{i1}$ is the subject-specific (time-invariant) covariate generated from standard normal distribution and $b_i$ is the subject-specific random intercept generated from normal distribution with mean 0 and variance $\sigma^2$ = 0.4. The true values of $\beta$ = (1,-0.2) and $\upsilon$ = 8 are assumed to be fixed across all scenarios. The true values of $\gamma$ = (-2.3, 0.5), (-1.7, 0.5) and (-2.4, 0.5) are for $p_{zero}$ = 10%, 15%, and 20%, respectively. For each scenario, 500 data sets are generated based on (10) and the proposed model is fitted using the log of the likelihood defined in (9) via SAS NLMIXED procedure. Under each scenario, bias and mean squared error (MSE) values are computed for each parameter and the results are displayed in Tables 1-3.

### *3.1 Simulation results*

The simulation results are displayed in Tables 1-3. All the results in the tables confirms that the MSE values of all parameters decreases when the sample size increases from 100 to 200. Specifically, the results show that $\gamma$ parameters are estimated with lower MSEs when the percentage of zeros in the data set is 20%. This means the more zeros in the data, the more precise the $\gamma$ parameters. The results also show that the MSE values of $\beta$ parameters are not affected by the percantage of zeros in the data set. On the other hand, since there is only one random intercept term in the model, the variance parameter $\sigma^2$ always results in lower MSEs indicating that the data is rich enough to estimate this parameter.

**Table 1.** Simulation results when $p_{zero}$ = 10%.

| Parameter | True value | N = 100 Bias | N = 100 MSE | N = 200 Bias | N = 200 MSE |
|:---:|:---:|:---:|:---:|:---:|:---:|
| $\gamma_0$ | -2.3 | 0.08 | 0.04 | -0.02 | 0.02 |
| $\gamma_1$ | 0.5 | -0.37 | 0.17 | -0.21 | 0.06 |
| $\beta_0$ | 1.0 | 0.03 | 0.01 | 0.05 | 0.01 |
| $\beta_1$ | -0.2 | -0.03 | 0.01 | 0.04 | 0.00 |
| $\sigma^2$ | 0.4 | -0.03 | 0.00 | 0.04 | 0.00 |
| $\upsilon$ | 8.0 | -0.79 | 1.00 | -0.26 | 0.28 |

**Table 2.** Simulation results when $p_{zero} = 15\%$.

| Parameter | True value | N = 100 | N = 100 | N = 200 | N = 200 |
|---|---|---|---|---|---|
| | | Bias | MSE | Bias | MSE |
| $\gamma_0$ | -1.7 | -0.04 | 0.02 | -0.11 | 0.02 |
| $\gamma_1$ | 0.5 | -0.17 | 0.05 | -0.15 | 0.03 |
| $\beta_0$ | 1.0 | 0.03 | 0.01 | 0.06 | 0.01 |
| $\beta_1$ | -0.2 | -0.02 | 0.01 | 0.05 | 0.00 |
| $\sigma^2$ | 0.4 | -0.02 | 0.00 | 0.04 | 0.00 |
| $\upsilon$ | 8.0 | -0.69 | 0.90 | -0.17 | 0.20 |

**Table 3.** Simulation results when $p_{zero} = 20\%$.

| Parameter | True value | N = 100 | N = 100 | N = 200 | N = 200 |
|---|---|---|---|---|---|
| | | Bias | MSE | Bias | MSE |
| $\gamma_0$ | -1.4 | -0.01 | 0.02 | -0.04 | 0.01 |
| $\gamma_1$ | 0.5 | -0.20 | 0.06 | -0.19 | 0.04 |
| $\beta_0$ | 1.0 | 0.03 | 0.01 | 0.06 | 0.01 |
| $\beta_1$ | -0.2 | -0.03 | 0.01 | 0.04 | 0.00 |
| $\sigma^2$ | 0.4 | -0.01 | 0.00 | 0.05 | 0.00 |
| $\upsilon$ | 8.0 | -0.71 | 0.96 | -0.30 | 0.34 |

## 4. Conclusion

In this study, we proposed a marginalized multilevel random effects model for analysis of longitudinal semi-continuous data. The model provides marginal inference both for zero-part and non-zero part along with subject-specific inferences for the non-zero part. The simulation studies show that the model has desirable properties. We are currently extending this model to longitudinal proportion data augmented with zeros and ones.

## References

[1]   D. Bandyopadhyay, D.M Galvis, V.H. Lachos, 2017, Augmented mixed models for clustered proportion data, *Statistical Methods in Medical Research*, 26(2), 880-897.
[2]   V.T. Farewell, D.L. Long, B.D.M. Tom, S. Yiu, L. Su, 2017, Two-part and related regression models for longitudinal data, *Annual Review of Statistics and Its Application,* 4, 283-315.
[3]   D.M. Galvis, D. Bandyopadhyay, V.H. Lachos, 2014, Augmented mixed beta regression models for periodontal proportion data, *Statistics in Medicine*, 33(21), 3759-3771.
[4]   M. Griswold, B. Swihart, B. Cao, S. Zeger, 2013, Practical marginalized multilevel models, *Statistics*, 2, 129-142.
[5]   P. Heagerty, 1999, Marginally specified logistic-normal models for longitudinal binary data, Biometrics, 55, 688-698.
[6]   P. Heagerty, S. Zeger, 2000, Marginalized multilevel models and likelihood inference (with discussion), *Statistical Science*, 15, 1-26.
[7]   G. Inan, 2018, Comments on "Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros", *Statistics in Medicine*, 37, 324–326.
[8]   W. Kassahun, T. Neyens, G. Molenberghs, C. Faes, G. Verbeke, 2014, Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros, *Statistics in Medicine*, 33(25), 4402-4419.
[9]   K. Lee, Y. Joo, J.J. Song, D.W. Harper, 2011, Analysis of zero-inflated clustered count data: A marginalized model approach,  *Computational Statistics and Data Analysis,* 55(1), 824-837.

[10]   B,  Neelon, A.J. O'Malley, V.A.  Smith, 2016, Modeling zero-modified count and semi-continuous data in health services research Part 1: Back-ground and overview, *Statistics in Medicine*, 35(27), 5070-5093.

[11]   B. Neelon, A.J. O'Malley, V.A. Smith, 2016, Modeling zero-modified count and semi-continuous data in health services research part 2: Case studies,  *Statistics in Medicine*, 35(27), 5094-5112.

[12]   M.K. Olsen, J.L. Schafer, 2001, A two-part random-effects model for semi-continuous longitudinal data, *Journal of the American Statistical Association,* 96(454), 730-745.

[13]   M. Rodrigues-Motta, et al., 2015, A mixed-effect model for positive responses augmented by zeros, *Statistics in Medicine*, 34(10), 1761-1778.

[14]   V.A. Smith, B.  Neelon, M.L. Maciejewski, J.S. Preisser, 2017, Two parts are better than one: modeling marginal means of semi-continuous data, *Health Services and Outcomes Research Methodology*, 1-21.

[15]   V.A. Smith, B. Neelon, J.S. Preisser, M.L. Maciejewski, 2017, A marginalized two-part model for longitudinal semi-continuous data,  *Statistical Methods in Medical Research*, 26(4), 1949-1968.

[16]   L. Su, B.D. Tom, V.T. Farewell, 2015, A likelihood-based two-part marginal model for longitudinal semi-continuous data, *Statistical Methods in Medical Rese*arch, 24(2), 194-205.

[17]   B.D. Tom, L. Su, V.T. Farewell, 2016, A corrected formulation for marginal inference derived from two-part mixed models for longitudinal semi-continuous data,  *Statistical Methods in Medical Research*, 25(5), 2014-2020.

[18]   B. Zhang, W. Liu, Y. Hu, 2017, Estimating marginal and incremental effects in the analysis of medical expenditure panel data using marginalized two-part random-effects generalized Gamma models: Evidence from China healthcare cost data, *Statistical Methods in Medical Research*, DOI:0962280217690770.

## Appendix

## A. Calculation of $\Delta_{it}$

Since any conditional expectation can be written in terms of marginal expectation, it is possible to link the conditional mean of the non-zero part of the model to the marginal mean model of that via integrating the conditional expectation over the distribution of random effects such that:

$$E(Y_{it}) = E_{b_i}(E(Y_{it}|b_i)) \tag{A1}$$
$$(1\text{-}p_{0,it})\mu_{it}^m = \int (1\text{-}p_{0,it})\mu_{it}^c f(b_i)d_{b_i},$$

where $f(b_i)$ is the standard normal distribution with mean 0 and variance $\sigma^2$. Substituting marginal and conditional means in (3) and (5), respectively, into equation (A1) gives the following expression:

$$\frac{\exp(\mathbf{X}_{cont,it}^T\beta)}{1+\exp(\mathbf{X}_{0,it}^T\gamma)} = \int \frac{\exp(\Delta_{it}+b_i)}{1+\exp(\mathbf{X}_{0,it}^T\gamma)}f(b_i)db_i$$

$$\exp(\mathbf{X}_{cont,it}^T\beta) = \int \exp(\Delta_{it}+b_i)f(b_i)db_i$$

$$= \exp(\Delta_{it})\int \exp(b_i)f(b_i)db_i$$

$$= \exp(\Delta_{it})\frac{1}{\sqrt{(2\pi\sigma^2)}}\int \exp(b_i)\exp(-\frac{b_i^2}{2\sigma^2})db_i$$

$$= \exp(\Delta_{it})\exp(\frac{\sigma^2}{2})\frac{1}{\sqrt{(2\pi\sigma^2)}}\int \exp(b_i)\exp(-\frac{\sigma^2}{2})\exp(-\frac{b_i^2}{2\sigma^2})db_i$$

$$= \exp(\Delta_{it})\exp(\frac{\sigma^2}{2})\frac{1}{\sqrt{(2\pi\sigma^2)}}\int \exp(-\frac{(b_i-\sigma^2)^2}{2\sigma^2})db_i$$

$$= \exp(\Delta_{it} + \frac{\sigma^2}{2}). \tag{A2}$$

Solving (A2) for $\Delta_{it}$ provides the following closed-form expression:

$$\Delta_{it} = \mathbf{X}_{cont,it}^T\beta - \frac{\sigma^2}{2}$$

(A3)

where $\Delta_{it}$ is a function of both $\beta$ from marginal part and $\sigma^2$ from conditional part and connects both levels to each other.

## B. SAS codes for data generation and model fitting

```
/* Data generation */

data simulated;
call streaminit(123);

/* Preliminaries */
N=200;
nt=4;
g0=-1.4;
g1=0.5;
b0=1;
b1=-0.2;
sigma2=0.4;
sigma=sqrt(sigma2);
nu=8;

do i=1 to N;
/* Generate random effects */
b=rand("Normal",0,sigma);

/* Generate predictor variable */
x1=rand("Normal",0,1);

/* Define the marginal probability for zero-part */
eta_zero_m=g0+g1*x1;
p_zero_m=exp(eta_zero_m)/(1+exp(eta_zero_m));

/* Define the marginal probability, delta, and */
/* conditional probability for continuous-part */

eta_cont_m=b0+x1*b1;
mu_cont_m=exp(eta_cont_m);
delta_cont=eta_cont_m-(sigma2/2);
mu_cont_p=exp(delta_cont+b);

/* Generate longitudinal semi-continuous data */

do j=1 to nt;
idnum=i;
time=j;
y_zero=RAND('BERNOULLI',(1-p_zero_m));
y_cont=(mu_cont_p/nu)*RAND('GAMMA',nu) ;
if y_zero=0 then y_cont=0;
y=y_cont;
output;
```

```
end;
end;
drop i j;
run;

data analysis;
set simulated;
keep idnum  x1 y;
run;

proc sort data=analysis;
by idnum;
run;

/* Model fitting */

proc nlmixed data=analysis;
parms g0=0 g1=0 b0=0 b1=0 sigma2=1 nu=1;
bounds sigma2>0, nu>0;

eta_zero_m=g0+g1*x1;
p_zero_m=exp(eta_zero_m)/(1+exp(eta_zero_m));

eta_cont_m=b0+x1*b1;
mu_cont_m=exp(eta_cont_m);
delta_cont=eta_cont_m-(sigma2/2);
mu_cont_p=exp(delta_cont+b);

if y=0 then ll=log(p_zero_m);
else ll=log(1-p_zero_m)-log(y)-lgamma(nu)+nu*log(y)+nu*log(nu)-nu*log(mu_cont_p)-
(y*(nu/mu_cont_p));
model y~general(ll);
random b ~normal(0, sigma2) subject=idnum;
run;
```