Fitting Multilevel Hierarchical Mixed Models Using PROC NLMIXED

Raghavendra Rao Kurada
SAS Institute Inc.

ABSTRACT

Hierarchical nonlinear mixed models are complex models that occur naturally in many fields. The NLMIXED procedure's ability to fit linear and nonlinear models with standard or general distributions enables you to fit a wide range of such models. SAS/STAT® 13.2 enhanced PROC NLMIXED to support multiple RANDOM statements, enabling you to fit nested multilevel mixed models. This paper uses an example to illustrate the new functionality.

INTRODUCTION

Mixed modeling techniques are one of the most common tools used for analyzing multilevel data. SAS/STAT software includes several procedures (such as the MIXED, GLIMMIX, and NLMIXED procedures) that can fit these mixed models. You can use the GLIMMIX procedure to fit generalized linear mixed models, in which the conditional distribution of the response given fixed and random effects belongs to an exponential family of distributions. You use a link function to link the expectation of this conditional distribution to the linear combination of fixed effects, random effects, and covariates. Zhu (2014) provides a comprehensive study of using PROC GLIMMIX to fit multilevel models. However, if you want to fit any general linear or nonlinear mixed model that is beyond the class of mixed models that PROC GLIMMIX can fit, then you use the NLMIXED procedure. You can also use PROC NLMIXED to fit any model in the PROC GLIMMIX class of models.

PROC NLMIXED is a flexible and powerful procedure that can fit linear and nonlinear mixed models that have standard or general distributions. Also, the procedure provides a class of optimization routines, from which you can choose one that is suitable for your problem type. Wolfinger (1999) provides a basic introduction to the NLMIXED procedure.

The term nonlinear models broadly refers to models in which the parameters of the conditional distribution of the response given the covariates occur nonlinearly in the model. If these parameters are the function of both fixed and random effects, then the models are known as nonlinear mixed models. Most software packages are restricted to fit a class of these nonlinear mixed models in which the conditional expectation of the response variable is a function of nonlinear fixed and random effects. These models are known as nonlinear mixed effects (NLME) models. PROC NLMIXED fits general nonlinear mixed (GNLM) models, in which the parameters of the conditional distribution of the response are a nonlinear function of fixed and random effects. GNLM models are more general than NLME models, and NLME models are one subclass of GNLM models. An example of a GNLM model that does not belongs to the class of NLME models is a model whose conditional distribution of the response variable follows a Weibull distribution with shape and scale parameters, where the scale parameter is a nonlinear function of fixed and random effects.

PROC NLMIXED handles random effects through the RANDOM statement. Wolfinger (1999) pointed out that the NLMIXED procedure was limited in handling nested or crossed random effects, because you could specify only one RANDOM statement before SAS/STAT 13.1. Littell et al. (2006) provided a way of using one RANDOM statement to fit a particular class of nested nonlinear mixed models, but their method relied on complex programming. In SAS/STAT 13.2 and later versions, you can specify nested random effects by using more than one RANDOM statement in PROC NLMIXED, enabling you to fit a nested nonlinear mixed model. This paper provides an example that shows you how to use multiple RANDOM statements in PROC NLMIXED to fit nested nonlinear mixed models, and it provides details about the computation that is involved in fitting these models.

The rest of this paper is organized as follows: Section 2 introduces the nonlinear mixed models that you can fit with PROC NLMIXED by showing an example that uses a nonlinear mixed model that has one level of random effects. Section 3 provides a brief theory about the nonlinear and nested nonlinear mixed models that you can fit with PROC NLMIXED. Section 4 illustrates the PROC NLMIXED enhancement that enables you to use multiple RANDOM statements to fit a nested nonlinear mixed model that has two levels of nested random effects. Finally, the paper is concluded with a brief discussion.

NONLINEAR MIXED MODEL WITH ONE-LEVEL RANDOM EFFECTS

Theophylline Data

Consider data from Pinheiro and Bates (1995) that were collected to study how the drug theophylline disperses through the body. Twelve patients were chosen randomly for this experiment, and the theophylline drug was administered orally to each patient. After the drug was administered, each patient's serum concentration was measured 11 times over a 25-hour period. The data are shown in the following **Theoph** data set:

```
data Theoph;
   input patient time conc dose;
   datalines;
 1  0.00  0.74 4.02
 1  0.25  2.84 4.02
 1  0.57  6.57 4.02
 1  1.12 10.50 4.02
 1  2.02  9.66 4.02
 1  3.82  8.58 4.02
 1  5.10  8.36 4.02
 1  7.03  7.47 4.02

   ... more lines ...

12 24.15  1.17 5.30
 ;
```

The **Theoph** data set includes the following variables:

- **patient**: patient identifier

- **time**: time at which the serum concentration was measured

- **conc**: serum concentration

- **dose**: initial dose given to the patient

Let y_{ij} denote the serum concentration in patient i at time j, for $j = t_1, t_2, ...t_i$ and $i = 1, 2, ..., 12$. Denote D_{ij} as the dose given to patient i at time j. Because the dose is administered only once, the dose value is the same for all time points for a patient in the data set. This study results in a two-level data structure. Consider the times when the measurements are observed as level-1 units, and consider the patients as level-2 clusters. Note that the level-1 units are nested within level-2 clusters. The serum concentrations collected at several time points on the same patient are correlated, so any analysis for these data should take this correlation into account.

First-Order One-Compartment Model

A simple plot of the serum concentrations against time for each individual illustrates the data. The following SGPLOT procedure program creates series plots for each patient over time:

```
proc sgplot data = Theoph;
   series x = time y = conc/group=patient;
run;
```

The resulting plot is shown in Figure 1.
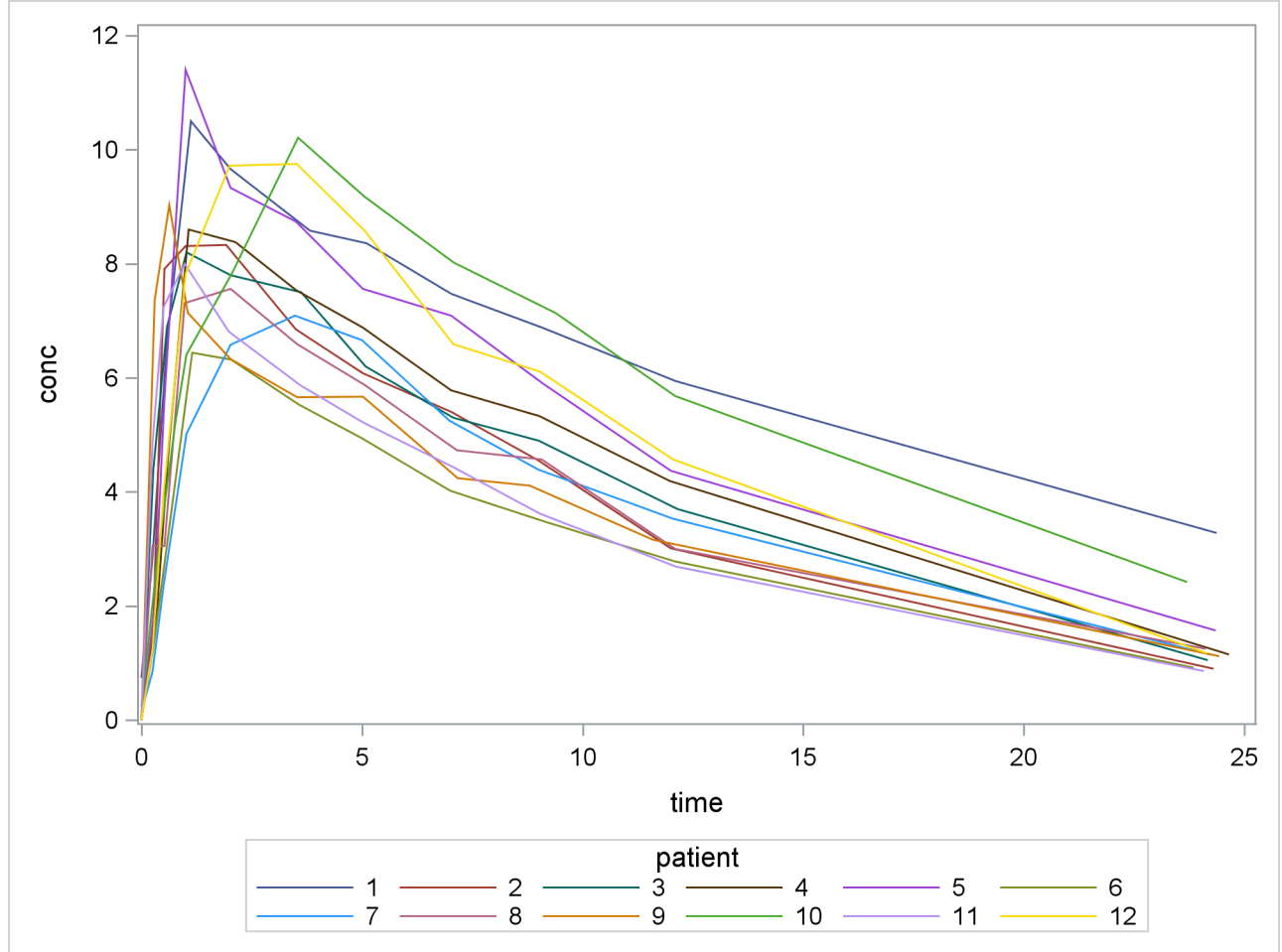
Figure 1 Individual Serum Concentration Plot

Figure 1 shows a nonlinear profile for each patient. The serum concentration for each patient increases rapidly at the beginning of the time period and reaches a maximum before gradually decreasing. This nonlinear behavior is described using a one-compartment model with first-order absorption and elimination from the pharmacokinetics (PK) field. A typical first-order one-compartment model can be written as

$$y(t) = \frac{Dk_e k_a}{Cl(k_a - k_e)}[e^{(-k_e t)} - e^{(-k_a t)}] + e(t)$$

where $y(t)$ is the response measured at time t, $e(t)$ is the Gaussian error term, and D is the dose given to the patient. The constants Cl, k_a, and k_e are the parameters of the one-compartment model, and they represent the clearance, absorption rate, and elimination rate, respectively. The clearance is further described as the volume of the drug eliminated from the body per unit of time. In addition, the first-order one-compartment model is usually fitted with the following reparameterization:

$$Cl = \exp(\beta_1)$$
$$k_a = \exp(\beta_2)$$
$$k_e = \exp(\beta_3)$$

The parameters (Cl, k_a, k_e) are sometimes called the population one-compartment model parameters or population PK parameters. Assume that there is a true first-order one-compartment model for the population of patients with unknown parameters (Cl, k_a, k_e). However, observe from Figure 1 that each patient has his or her own nonlinear behavior that deviates from the true population first-order one-compartment model. To model these random deviations, assume the "patient-specific" one-compartment model, whose parameters can be constructed by introducing random

effects. Pinheiro and Bates (1995) constructed the patient-specific parameters as

$$Cl_i = \exp(\beta_1 + b_{1i})$$
$$k_{a_i} = \exp(\beta_2 + b_{2i})$$
$$k_{e_i} = \exp(\beta_3)$$

where

$$\left(\begin{array}{c} b_{1i} \\ b_{2i} \end{array} \right) \sim N \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right) \right)$$

You can include a b_{3i} random effect for the patient-specific elimination parameter k_{e_i} but Pinheiro and Bates (1995) did not assume that k_{e_i} is significantly different for all the patients. Hence, only Cl_i and k_{a_i} are assumed to be patient-specific parameters, and the following one-compartment mixed model is fit to the theophylline data:

$$y_{ij} = \frac{D_{ij} k_{e_i} k_{a_i}}{Cl_i (k_{a_i} - k_{e_i})} [e^{(-k_{e_i} j)} - e^{(-k_{a_i} j)}] + e_{ij} \quad j = t_1, t_2, ..., t_i \text{ and } i = 1, 2, ..., 12$$

Here the error terms e_{ij} are assumed to be distributed as a normal random variable with mean 0 and variance σ^2. The parameter vector for the first-order one-compartment model is $\theta = (\beta_1, \beta_2, \beta_3, \sigma^2, \sigma_1^2, \sigma_{12}, \sigma_2^2)$. Based on the normality assumption, you can construct the likelihood $L(\theta)$ for the first-order one-compartment model and maximize the likelihood with respect to θ to obtain the maximum likelihood (ML) estimates. You can fit this first-order one-compartment model by using PROC NLMIXED to obtain the ML estimates of θ.

Parameter Estimation

The following statements fit the first-order one-compartment mixed model to the theophylline data:

```
proc nlmixed data=Theoph;
   parms beta1=-3.22 beta2=0.47 beta3=-2.45
      s2b1 =0.03 cb12 =0 s2b2 =0.4 s2=0.5;
   cl = exp(beta1 + b1);
   ka = exp(beta2 + b2);
   ke = exp(beta3);
   mu = dose*ke*ka*(exp(-ke*time)-exp(-ka*time))/cl/(ka-ke);
   model conc ~ normal(mu,s2);
   random b1 b2 ~ normal([0,0],[s2b1,cb12,s2b2]) subject=patient;
   estimate 'pop clearance'         exp(beta1);
   estimate 'pop absorption rate'  exp(beta2);
   estimate 'pop elimination rate' exp(beta3);
run;
```

The PARMS statement provides the initial values for all the parameters in the model. The next three SAS® programming statements define the clearance, absorption, and elimination rate parameters. Then the next SAS programming statement defines the mean of the one-compartment model and assigns it to the variable **mu**. The MODEL statement specifies the conditional normal distribution for the response variable **conc** with mean **mu** and variance **s2**. The RANDOM statement specifies the random effects **b1** and **b2** and their joint distribution as the bivariate normal with mean $(0, 0)$ and a variance matrix. It is sufficient to specify the lower triangular part of the variance matrix. The level-2 clusters (patients) are stored in the **patient** variable in the **Theoph** data set. The cluster variable information is specified in the SUBJECT= variable option. The ESTIMATE statements request that PROC NLMIXED estimate the population PK parameters. Selected outputs from the fitted model are shown in Figure 2 and Figure 3.

Figure 2 Parameter Estimates

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gradient
beta1	-3.2268	0.05950	10	-54.23	<.0001	-3.3594	-3.0942	-0.00009
beta2	0.4806	0.1989	10	2.42	0.0363	0.03745	0.9238	3.645E-7
beta3	-2.4592	0.05126	10	-47.97	<.0001	-2.5734	-2.3449	0.000039
s2b1	0.02803	0.01221	10	2.30	0.0446	0.000828	0.05524	-0.00014
cb12	-0.00127	0.03404	10	-0.04	0.9710	-0.07712	0.07458	-0.00007
s2b2	0.4331	0.2005	10	2.16	0.0561	-0.01354	0.8798	-6.98E-6
s2	0.5016	0.06837	10	7.34	<.0001	0.3493	0.6540	6.133E-6

The "Parameter Estimates" table in Figure 2 provides the maximum likelihood estimates of the parameters from the fitted nonlinear mixed model and their standard errors. The population PK parameters (Cl, k_a, k_e) can be estimated by using the estimates of beta1, beta2, and beta3. The "Additional Estimates" table in Figure 3 provides the population PK parameter estimates.

Figure 3 Additional Estimates

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
pop clearance	0.03968	0.002361	10	16.81	<.0001	0.05	0.03442	0.04494
pop absorption rate	1.6171	0.3216	10	5.03	0.0005	0.05	0.9004	2.3337
pop elimination rate	0.08551	0.004383	10	19.51	<.0001	0.05	0.07574	0.09527

These population PK parameter estimates are useful in developing the theophylline drug dosage recommendations for the entire population. For more information about the usage of estimated population PK parameters, see Food and Drug Administration (1999).

The two-level theophylline data are fitted using a nonlinear mixed model with one-level random effects. This is a typical example of nonlinear mixed models that you can fit by using one RANDOM statement in PROC NLMIXED . Similarly, you might need to fit a nested nonlinear mixed model that has random effects of two or more levels. SAS/STAT 13.2 enhanced PROC NLMIXED to support multiple RANDOM statements, enabling you to fit these nested nonlinear mixed models. The next section provides a brief theory of nonlinear and nested nonlinear mixed models that you can fit with PROC NLMIXED.

PROC NLMIXED: THEORY AND METHODS

Nonlinear Mixed Models

This section explains the general framework of the nonlinear mixed models that you can fit with PROC NLMIXED. Suppose that y_{ij} denotes the response variable that is observed at occasion j on subject i, and suppose that $\mathbf{y}_i$ denotes the vector of all measurements for the ith subject, with a total of s independent subjects. The occasions at which the response measurements are observed for each subject are denoted as the level-1 units, and the subjects are denoted as level-2 clusters. The subscript j is used for the level-1 units, and the subscript i is used for the level-2 clusters. The set of explanatory variables for each subject is denoted as $\mathbf{x}_i$. The measurements that are observed at several time points on the same subject are correlated. To account for this dependency of the observations within the same subject, assume that there exist latent random-effect vectors $\mathbf{v}_i$, called level-1 random effects. Also, assume that an appropriate model that links $\mathbf{y}_i$ and $\mathbf{v}_i$ exists. Using this link, the joint probability density function of $(\mathbf{y}_i, \mathbf{v}_i)$ is obtained as $p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{v}_i)q(\mathbf{v}_i|\mathbf{d})$. If $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{d})$ represents the parameter vector, then the marginal likelihood function is the following integral:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{s} \int p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{v}_i)q(\mathbf{v}_i|\mathbf{d})d\mathbf{v}_i$$

PROC NLMIXED approximates this integral by using a numerical integrator and uses a library of optimization routines to maximize the likelihood with respect to $\theta = (\beta, \mathbf{d})$. For more information about the integral approximation methods and the optimization techniques that are available in PROC NLMIXED, see the "Details" section in the chapter "The NLMIXED Procedure" in *SAS/STAT User's Guide*.

For the theophylline data that are considered in the previous section, y_{ij} is the measurement of serum concentration of the ith patient that is observed at the jth time point. These observations are collected on $s=12$ subjects. The time points when the serum concentrations are measured are level-1 units, and the patients are level-2 clusters. The dose values for each patient, D_{ij}, are the independent covariates. The latent random effects $\mathbf{v}_i = (b_{1i}, b_{2i})$ are the level-1 random effects in the first-order one-compartment model. The conditional distribution of y_{ij} given $\mathbf{v}_i$ is normal with mean

$$\frac{D_{ij} k_{e_i} k_{a_i}}{Cl_i (k_{a_i} - k_{e_i})} [e^{(-k_{e_i} j)} - e^{(-k_{a_i} j)}] \quad \text{for} \quad j = t_1, t_2, ..., t_i \text{ and } i = 1, 2, ..., 12$$

where

$$Cl_i = \exp(\beta_1 + b_{1i})$$
$$k_{a_i} = \exp(\beta_2 + b_{2i})$$
$$k_{e_i} = \exp(\beta_3)$$

and variance σ^2. This conditional distribution provides the link between $\mathbf{y}_i$ and $\mathbf{v}_i$. Also, $\mathbf{v}_i = (b_{1i}, b_{2i})$ follows a bivariate normal distribution with mean $(0, 0)$ and the following covariance matrix D:

$$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Hence the parameter vector for this nonlinear mixed model is $\theta = (\beta_1, \beta_2, \beta_3, \sigma^2, \sigma_1^2, \sigma_{12}, \sigma_2^2)$.

Nested Nonlinear Mixed Models

This section extends the framework that is described in the previous section to multilevel data, which have more than two levels of units and need to be fitted by using more than one level of random effects. Models that have more than one level of nested random effects are known as nested nonlinear mixed models. SAS/STAT 13.2 adds the ability to fit nested nonlinear mixed models in PROC NLMIXED.

Consider three-level data that need to be fitted with a nonlinear mixed model that has fixed effects and two levels of nested random effects. Let y_{ijk} denote the response that is observed on the kth level-1 unit from the jth level-2 cluster, which is further nested within the ith level-3 cluster. Let $\mathbf{y}_{j(i)}$ denote the vector of all the measurements for the jth level-2 cluster that is nested within the ith level-3 cluster. Along with the y_{ijk}, a set of explanatory variables, $\mathbf{x}_{ijk}$, are observed. There are s level-3 clusters, and each cluster has s_i level-2 clusters nested within it. An example is $\mathbf{y}_{j(i)}$, which are the heights of students in class j of school i, where $j = 1, \ldots, s_i$ for each i and $i = 1, \ldots, s$. Suppose there exist latent random-effect vectors $\mathbf{v}_{j(i)}$ and $\mathbf{v}_i$ of small dimensions for modeling within cluster covariance. The random effects $\mathbf{v}_i$ are called level-2 random effects, and $\mathbf{v}_{j(i)}$ are level-1 random effects that are nested within level-2 random effects. Assume also that an appropriate model that links $\mathbf{y}_{j(i)}$ and $(\mathbf{v}_{j(i)}, \mathbf{v}_i)$ exists. Then the joint density function in terms of the level-3 clusters can be expressed as

$$p(\mathbf{y}_i | \mathbf{X}_i, \theta, \mathbf{u}_i) q(\mathbf{u}_i | \mathbf{d}) = \left(\prod_{j=1}^{s_i} p(\mathbf{y}_{j(i)} | \mathbf{X}_i, \phi, \mathbf{v}_i, \mathbf{v}_{j(i)}) q_2(\mathbf{v}_{j(i)} | \mathbf{d_2}) \right) q_1(\mathbf{v}_i | \mathbf{d_1})$$

where $\theta = (\beta, \mathbf{d})$, $\mathbf{y}_i = (\mathbf{y}_{1(i)}, \ldots, \mathbf{y}_{s_i(i)})$, $\mathbf{u}_i = (\mathbf{v}_i, \mathbf{v}_{1(i)}, \ldots, \mathbf{v}_{s_i(i)})$, and $\mathbf{d} = (\mathbf{d_1}, \mathbf{d_2})$. As defined earlier in this section, the marginal likelihood function is the following integral:

$$L(\theta) = \prod_{i=1}^{s} \int p(\mathbf{y}_i | \mathbf{X}_i, \theta, \mathbf{u}_i) q(\mathbf{u}_i | \mathbf{d}) d\mathbf{u}_i$$

PROC NLMIXED numerically minimizes the function $f(\theta) = -\log L(\theta)$ over θ in order to estimate θ. Models that have more than two levels of nested random effects follow similar notation. The next section provides an example to illustrate how you use PROC NLMIXED to fit a nested nonlinear mixed model.

NESTED NONLINEAR MIXED MODEL WITH TWO-LEVEL RANDOM EFFECTS

ADEMEX Hospital Admission Counts Data

Consider the hospitalization data from the ADEMEX clinical trial described by Vonesh (2012), in which patients with end-stage renal disease are treated with continuous ambulatory peritoneal dialysis (CAPD). This trial was a randomized controlled trial designed to compare select patient outcomes among patients randomized to receive a high dosage versus a standard dosage of CAPD. Specifically, a total of 965 patients from 24 centers were randomized to either a standard-dosage group (four daily exchanges of 2L of standard peritoneal dialysis solution per day) or a high-dosage group (four or five 2.5L or 3L exchanges per day or five 2L exchanges per day). The primary outcome in the ADEMEX trial was patient survival, but a key secondary outcome was the number of times each patient was hospitalized over the course of follow-up. Along with these counts, the following prognostic factors on each patient were recorded: treatment group, gender, patient age at baseline, diabetes presence at baseline, albumin value at baseline, and the number of months on dialysis prior to randomization. The data set is downloaded from the website `http://support.sas.com/publishing/authors/vonesh.html` and stored in the **Ademex** data set. You can use the following SAS code to obtain the center information for each patient from this data set:

```
data Ademex;
   set Ademex;
   Center = scan(Ptid,1);
run;
```

The **Ademex** data set contains the following variables:

- **Age** : age of the patient at the baseline

- **Albumin** : albumin value at the baseline

- **Diabetic** : diabetes presence indicator (0 = nondiabetic and 1 = diabetic)

- **Hosp** : hospital admission counts

- **PriorMonths** : number of months on dialysis before the baseline

- **Sex** : gender of the patient (0 = male and 1 = female)

- **Trt** : treatment indicator (0 = control or standard dose group and 1 = treatment or high dose group)

- **Ptid** : patient identifier

- **Center** : center identifier

Let y_{ij} denote the hospital admission counts for the jth patient in the ith center for $j = 1, 2, , ..., j_i$, and $i = 1, 2, ..., C$, where C is the number of centers that are chosen randomly from a population of centers and where patients are chosen randomly from each center. Each patient is randomly assigned to a standard- or high-dosage treatment, and the number of times that the patient is admitted to the hospital after receiving the treatment is recorded. The **Hosp** variable in the **Ademex** data set contains these admission counts. By design, this experiment results in multilevel data. Admissions of the patient to the hospital are considered to be level-1 units. The patients are considered to be level-2 clusters that are nested within center, which are considered to be level-3 clusters. Note that all the measurements on the patients within the same center are correlated because they receive the treatment under the same conditions. Hence you should consider this dependency. Also, each patient is randomly chosen from a population of patients within each center. Hence the variation among the patients should be also considered.

This example is a special case according to the theory explained in the previous section. Each patient has only one observation, so the observation vector $\mathbf{y}_{j(i)}$ is a scalar $y_{j(i)}$. Further, the index j refers to the patient and the index i refers to the center. Because the patients are randomly chosen from each center, the notation $y_{j(i)}$ is more appropriate. Denote $\mathrm{Trt}_{ij}, \mathrm{Sex}_{ij}, \mathrm{Age}_{ij}, \mathrm{Diabetic}_{ij}, \mathrm{PriorMonths}_{ij}$, and $\mathrm{Albumin}_{ij}$ as the corresponding covariates that are observed on the jth patient from the ith center. The goal is to compare the hospitalization rate between patients who received higher doses of dialysis solution (treatment group) and patients who received standard doses of dialysis solution (control group) while taking the patient-specific and center-specific dependencies into account.

Nested Random-Effects Zero-Inflated Poisson Model

Because the response measure consists of counts, it is common to fit a Poisson model to the data. The following SAS code provides the frequency plots (which are shown in Figure 4) for the hospital admission counts within each center:

```
proc sgpanel data = Ademex;
   panelby Center /onepanel layout = panel;
   vbar Hosp;
   colaxis label="Hospital Admission Counts" Fitpolicy= thin;
run;
```

Figure 4 Frequency Plot for Admission Counts
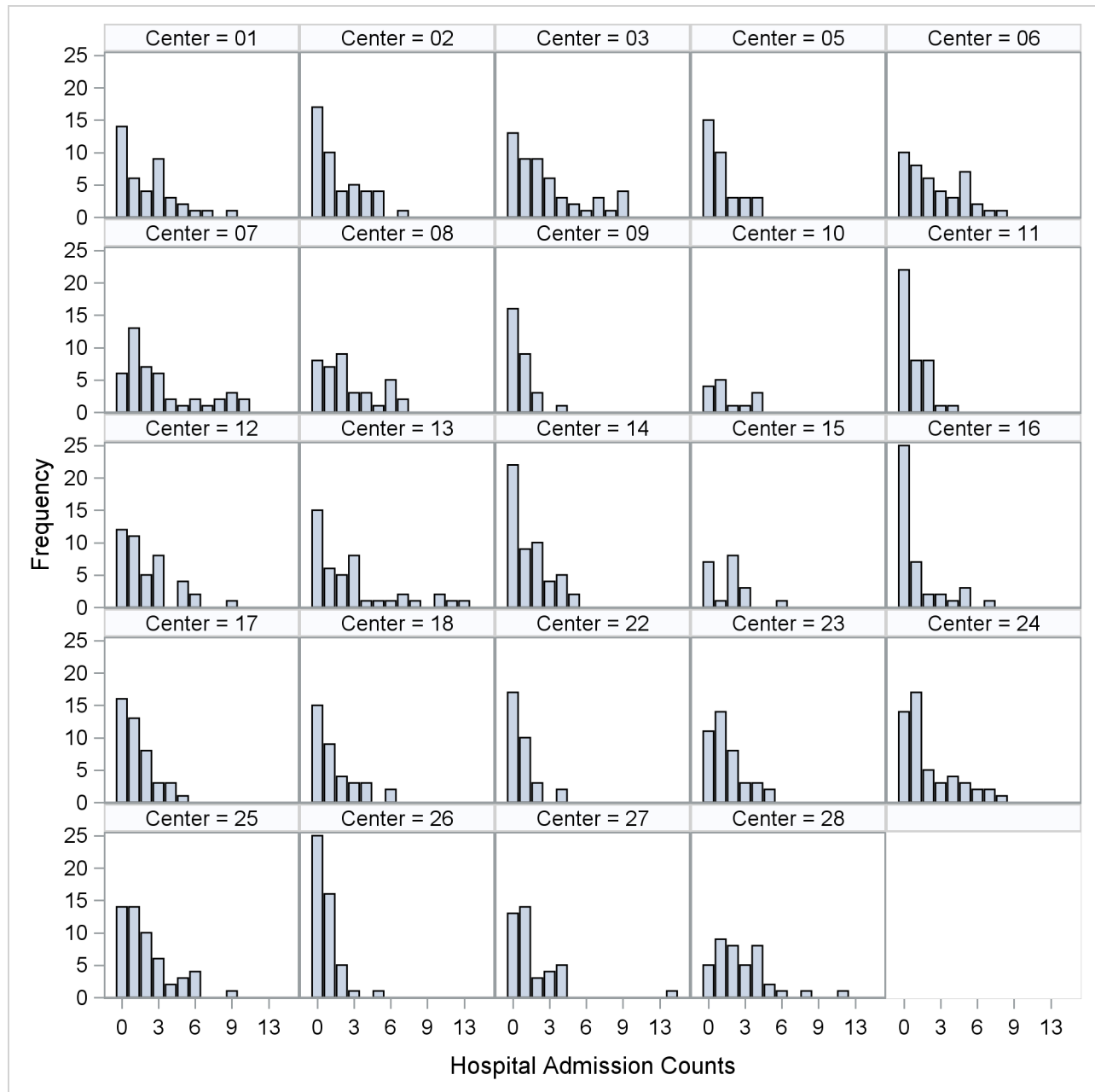
Most centers in Figure 4 have a greater percentage of zero counts than you would expect from a Poisson distribution. To account for excess zeros, you can use the zero-inflated Poisson (ZIP) model. The ZIP random variable is derived based on the idea that the zero counts in a process are distributed between two subpopulations. One subpopulation uses the regular Poisson distribution, and the other uses the distribution that has only zero counts. A typical

8

zero-inflated Poisson random variable is defined as

$$
\Pr(Y = y) = \begin{cases} p + (1-p)e^{-\lambda} & y = 0 \\ (1-p)\frac{e^{-\lambda}\lambda^y}{y!} & y > 0 \end{cases}
$$

where y is the count and p and λ are the parameters. The parameters p and λ are modeled as a function of linear predictors in a ZIP model. In general, a log link function is used for λ and a logit or probit link function is used for p as

$$
\lambda = \exp(\mathbf{x}'\boldsymbol{\beta}) \quad \text{and}
$$

$$
p = \frac{1}{1 + e^{-\mathbf{z}'\boldsymbol{\eta}}} \quad \text{or} \quad p = \Phi^{-1}(\mathbf{z}'\boldsymbol{\eta})
$$

where $\mathbf{x}$ and $\mathbf{z}$ are the covariates that are observed along with the response measures. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are the fixed effects of the model. A detailed discussion of the ZIP models can be found in `https://support.sas.com/rnd/app/examples/stat/GENMODZIP/roots.pdf`. By choosing the logit link function for p, a ZIP model is fitted to the **Ademex** hospital count data. To account for the conditional Poisson variation for the individual patient-level counts, you use the patient-specific random effects. In addition, Figure 4 also suggests that the distributions of counts at each center are different. Choose center-specific random effects along with patient-specific random effects that are nested within center-specific random effects in order to account for the center level variation as

$$
\lambda_{j(i)} = \exp(\mathbf{x}'\boldsymbol{\beta} + v_{j(i)} + v_i)
$$

where v_i are the level-2 random effects and $v_{j(i)}$ are the level-1 random effects that are nested within v_i. The nested ZIP mixed model with patient specific random effects is written as

$$
\Pr(Y_{j(i)} = y_{j(i)} | \mathbf{x}_{ij}, v_i, v_{j(i)}) = \begin{cases} p_{j(i)} + (1 - p_{j(i)})e^{-\lambda_{j(i)}} & y_{j(i)} = 0 \\ (1 - p_{j(i)})\frac{e^{-\lambda_{j(i)}}\lambda_{j(i)}^{y_{j(i)}}}{y_{j(i)}!} & y_{j(i)} > 0 \end{cases}
$$

where

$$
p_{j(i)} = \frac{1}{1 + e^{-(a_0 + a_1\text{Trt}_{ij} + a_2\text{Sex}_{ij} + a_3\text{Diabetic}_{ij} + a_4\text{PriorMonths}_{ij})}}
$$

$$
\lambda_{j(i)} = e^{(b_0 + b_1\text{Trt}_{ij} + b_2\text{Sex}_{ij} + b_3\text{Age}_{ij} + b_4\text{Diabetic}_{ij} + b_5\text{PriorMonths}_{ij} + b_6\text{Albumin}_{ij} + v_{j(i)} + v_i)}
$$

and

$$
v_i \sim normal(0, \sigma_i)
$$

$$
v_{j(i)} \sim normal(0, \sigma_j)
$$

Let $\mathbf{a} = (a_0, ..., a_4)$ and $\mathbf{b} = (b_0, b_1, ..., b_6)$; then the parameter vector is denoted by $\boldsymbol{\theta} = [\mathbf{a}, \mathbf{b}, \sigma_i, \sigma_j]$. The conditional log likelihood for the nested ZIP mixed model can be written as

$$
l(\boldsymbol{\theta} | \mathbf{x}_{ij}, v_i, v_{j(i)}) = \begin{cases} \log(p_{ij} + (1 - p_{ij})e^{-\lambda_{ij}}) & y_{ij} = 0 \\ y_{ij}\log(\lambda_{ij}) + \log(1 - p_{ij}) - \lambda_{ij} - \log(\Gamma(y_{ij} + 1)) & y_{ij} > 0 \end{cases}
$$

Parameter Estimation

You can fit the preceding log likelihood by using multiple RANDOM statements in the NLMIXED procedure. Fitting this model with PROC NLMIXED requires a good set of initial values for the parameters.[1] If you start with bad initial values, then PROC NLMIXED suffers from convergence issues. One simple and good choice for finding the initial values for any mixed model is to ignore the random effects and just fit a fixed-effects model. To find the initial values for the nested ZIP mixed model, you first fit a fixed-effects ZIP model. You can use the following GENMOD procedure statements to fit the fixed-effects ZIP model:

---

[1]Models that are fitted using numerical optimization routines require initial values. Models that have multiple local optima or models that have discontinuity especially require a good set of initial values.

```
proc genmod data=Ademex;
    class Trt Sex Diabetic;
    model Hosp = Trt Sex Age Diabetic PriorMonths Albumin/ dist=zip;
    zeromodel Trt Sex Diabetic PriorMonths;
run;
```

The fixed-effects ZIP model that PROC GENMOD fits has two parts: one is the Poisson model, and the other is the logit model for excess zeros. Figure 5 shows the maximum likelihood (ML) estimates for the excess zero logit model, and Figure 6 shows the ML estimates for the Poisson model.

Figure 5 Parameter Estimates for Excess Zeros Logit Model

The GENMOD Procedure

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates					Wald 95% Confidence Limits		Wald	
Parameter		DF	Estimate	Standard Error			Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.6555	0.2253	-2.0971	-1.2139	54.00	<.0001
Trt	0	1	0.0119	0.1898	-0.3602	0.3839	0.00	0.9502
Trt	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Sex	0	1	0.5352	0.2002	0.1428	0.9276	7.15	0.0075
Sex	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Diabetic	0	1	0.4199	0.2024	0.0233	0.8165	4.31	0.0380
Diabetic	1	0	0.0000	0.0000	0.0000	0.0000	.	.
PriorMonths		1	0.0061	0.0033	-0.0003	0.0126	3.45	0.0633

Figure 6 Parameter Estimates for Poisson Model

Analysis Of Maximum Likelihood Parameter Estimates					Wald 95% Confidence Limits		Wald	
Parameter		DF	Estimate	Standard Error			Wald Chi-Square	Pr > ChiSq
Intercept		1	1.6728	0.1942	1.2921	2.0535	74.18	<.0001
Trt	0	1	-0.0995	0.0579	-0.2129	0.0139	2.96	0.0856
Trt	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Sex	0	1	-0.0341	0.0584	-0.1487	0.0804	0.34	0.5593
Sex	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Age		1	-0.0063	0.0025	-0.0112	-0.0015	6.58	0.0103
Diabetic	0	1	-0.3235	0.0694	-0.4595	-0.1874	21.72	<.0001
Diabetic	1	0	0.0000	0.0000	0.0000	0.0000	.	.
PriorMonths		1	0.0050	0.0011	0.0029	0.0071	21.73	<.0001
Albumin		1	-0.1083	0.0460	-0.1984	-0.0182	5.55	0.0184
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Now you can use PROC NLMIXED to fit the nested ZIP mixed model for the **Ademex** hospital count data as follows:

```
proc nlmixed data = Ademex tech = dbldog;
   parms a0= -1.6555 a1= 0.0119 a2= 0.5352 a3= 0.4199 a4= 0.0061
     b0= 1.6728 b1= -0.0995 b2= -0.0341 b3= -0.0063 b4= -0.3235 b5= 0.0050 b6= -0.1083;
   lp1     = a0 + a1*Trt + a2*Sex + a3*Diabetic + a4*PriorMonths;
   prob    = 1/(1+exp(-lp1));
   lp2     = b0 + b1*Trt + b2*Sex + b3*Age + b4*Diabetic + b5*PriorMonths + b6*Albumin +
             CE + PE;
   lambda = exp(lp2);
   if(Hosp = 0) then
      ll = log(prob + (1-prob)*exp(-lambda));
   else
      ll = Hosp*log(lambda) + log(1-prob) - lambda - lgamma(Hosp+1);
   model Hosp ~ general(ll);
   random CE ~ Normal(0,sc) subject = Center;
   random PE ~ Normal(0,sp) subject = Ptid(Center);
   predict lambda out = p1;
   predict prob   out = p2;
run;
```

A general log-likelihood specification is used in the MODEL statement to specify the log of the mass function from the ZIP model. A linear combination of **a0** to **a4** is assigned to the SAS programming variable **lp1**. Similarly, a linear combination of **b0** through **b6** and the two random effects **CE** and **PE** is assigned to the SAS programming variable **lp2**. The variable **prob** defines the inverse logit link of the variable **lp1**, and the variable **lambda** defines the inverse log link of the variable **lp2**. Then the next four SAS programming statements together define the log of the ZIP probability mass function for each patient. The first RANDOM statement uses the variable **CE** to define the level-2 random effects, v_i. Then the second RANDOM statement uses the variable **PE** to define the level-1 random effects, $v_{j(i)}$, which are nested within level-2 random effects. Both random-effects distributions are specified as normal with mean 0 and variance **sc** and **sp** for the level-2 and level-1, respectively.

The TECH=option in the PROC NLMIXED statement specifies the double-dogleg optimization technique. Because the log likelihood of the nested ZIP mixed model is complex and computationally expensive, PROC NLMIXED takes a considerable amount of time to optimize the log likelihood. In addition, the default PROC NLMIXED optimization technique, quasi-Newton, suffers from convergence issues. So you must choose an optimization technique that provides a proper convergence and completes the optimization in a reasonable time. This paper uses the double-dogleg technique because it converges properly and takes less time when compared to the quasi-Newton technique. The double-dogleg technique is similar to quasi-Newton except that it does not perform a line search at each iteration as the quasi-Newton technique does. For more information, see the "Details" section in the chapter "The NLMIXED Procedure" in *SAS/STAT User's Guide*.

The PARMS statement provides the initial values for the parameters. These values are obtained from the ML estimates of the fixed-effects ZIP model that was fitted by PROC GENMOD and are shown in Figure 5 and Figure 6. In addition, the expected values of the Poisson variable in the ZIP model, $\lambda_{j(i)}$, are requested from PROC NLMIXED by using the PREDICT statement and are stored in the data set **p1**. Also, the excess zeros probabilities $p_{j(i)}$ are requested from PROC NLMIXED by using another PREDICT statement and are stored in the data set **p2**. These predicted values are helpful in model validation, which is discussed in the next subsection. Selected output from the fitted model is shown in Figure 7 and Figure 8.

Figure 7 Model Specifications

The NLMIXED Procedure

Specifications	
Data Set	WORK.ADEMEX
Dependent Variable	Hosp
Distribution for Dependent Variable	General
Random Effects	CE, PE
Distribution for Random Effects	Normal, Normal
Subject Variables	Center, ptid
Optimization Technique	Double Dogleg
Integration Method	Adaptive Gaussian Quadrature

The "Specifications" table in Figure 7 lists the basic information about the data set and the model you have specified. The important information in this table is the Random Effects row, which shows that **CE** and **PE** are the random effects and they are separated by a comma. You read the random-effects information as the variable **CE** denoting the level-2 random effects and the variable **PE** denoting the level-1 random effects that are nested within the level-2 random effects, **CE**. The Subject Variables row provides the information about the cluster levels of the data that you specified in your mixed model. That row shows that the **Center** variable contains the level-3 clusters and the **Ptid** variable contains level-2 clusters that are nested within the level-3 clusters.

Figure 8 Parameter Estimates

						95%		
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Confidence Limits		Gradient
a0	-2.2834	0.9639	875	-2.37	0.0181	-4.1752	-0.3916	-0.09067
a1	0.6524	0.7236	875	0.90	0.3675	-0.7677	2.0726	0.025952
a2	-1.0369	0.6772	875	-1.53	0.1261	-2.3660	0.2921	-0.00889
a3	-0.8955	0.7991	875	-1.12	0.2628	-2.4639	0.6729	0.028670
a4	0.01524	0.008636	875	1.76	0.0779	-0.00171	0.03219	0.95655
b0	0.8956	0.2953	875	3.03	0.0025	0.3159	1.4752	-0.00369
b1	0.1936	0.09046	875	2.14	0.0326	0.01607	0.3711	0.057875
b2	0.03425	0.09214	875	0.37	0.7102	-0.1466	0.2151	0.040402
b3	-0.00563	0.003824	875	-1.47	0.1410	-0.01314	0.001871	-24.1179
b4	0.3280	0.1121	875	2.93	0.0035	0.1079	0.5480	-0.05098
b5	0.004748	0.002017	875	2.35	0.0188	0.000789	0.008707	-19.9987
b6	-0.2097	0.06825	875	-3.07	0.0022	-0.3436	-0.07570	-0.14416
sc	0.1440	0.05975	875	2.41	0.0162	0.02669	0.2612	0.075665
sp	0.4521	0.07883	875	5.73	<.0001	0.2974	0.6068	-0.05957

The "Parameter Estimates" table in Figure 8 lists the maximum likelihood estimates and their approximate standard errors. The p-values for the parameters **sp** and **sc** indicate that the patient-to-patient variation within centers and the center-to-center variation, respectively, are significant. To compare the hospitalization rate between the treatment and control group, use the estimate of **b1**. The p-value for the parameter **b1** from the "Parameter Estimates" table indicates that the treatment effect is significant. It implies that the expected number of hospital admissions for a patient who receives the high dose is $\exp(0.1936) = 1.2136$ times the expected number of hospital admissions for a patient who receives the standard dose while the other covariates are held constant.

Model Validation

One way of assessing the goodness of fit of the nested ZIP model is to compare the observed relative frequencies of the various hospital admission counts to the maximum likelihood estimates of their respective probabilities within each center. Recall that the expected values of the Poisson variable of the nested ZIP model and the excess zero probabilities are obtained using the PREDICT statement and are stored in the **p1** and **p2** data sets, respectively. Using these values and mimicking the SAS code given in https://support.sas.com/rnd/app/examples/stat/GENMODZIP/roots.pdf, you can compute the maximum likelihood estimates of the ZIP probabilities as follows:

```
data p1;
   set p1;
   keep Pred Center Hosp;
   rename Pred = lambda;
run;

data p2;
   set p2;
   keep Pred Center Hosp;
   rename Pred = pZero;
run;

proc sort data = p1; by Center Hosp; run;
```

```
proc sort data = p2; by Center Hosp; run;

data prob;
   merge p1 p2;
   by Center Hosp;
run;

%let max=13;

data prob(drop= i);
   set prob;
   by Center;
   array estProb{0:&max} estProb0-estProb&max;
   array c{0:&max} c0-c&max;
   do i = 0 to &max;
      if i=0 then estProb{i}= pZero + (1-pZero)*pdf('POISSON',i,lambda);
      else estProb{i}= (1-pZero)*pdf('POISSON',i,lambda);
      c{i}=ifn(Hosp=i,1,0);
   end;
run;

proc means data=prob noprint;
   class Center;
   var estProb0 - estProb&max c0-c&max;
   output out=estProb(drop=_TYPE_ _FREQ_) mean(estProb0-estProb&max)=estProb0-estProb&max;
   output out=p(drop=_TYPE_ _FREQ_) mean(c0-c&max)=p0-p&max;
run;

proc transpose data=estProb out=estProb(rename=(col1=zip) drop=_NAME_);
   by Center;
run;

proc transpose data=p out=p(rename=(col1=p) drop=_NAME_);
   by Center;
run;

data Compare;
   merge estProb p;
   by Center;
   if Center = "" then delete;
   Hosp=mod(_N_,&max+1)-1;
   if(Hosp = -1) then Hosp = &max+1;
   label zip = 'Estimated ZIP Probabilities'
           p = 'Observed Relative Frequencies';
run;
```

The **Compare** data set contains the relative frequencies and ZIP probability estimates of the various hospital admission counts within each center. In order to validate the fit of the nested ZIP mixed model, the observed relative frequencies for various hospital admission counts are plotted along with the ZIP probability estimates. These plots are created for each center by using a PANELBY statement in the SGPANEL procedure as follows:

```
proc sgpanel data = Compare;
   panelby Center / onepanel layout = panel;
   scatter x = Hosp y = p /
      markerattrs=(symbol=CircleFilled size=5px color=blue);
   scatter x=Hosp y=zip /
      markerattrs=(symbol=TriangleFilled size=5px color=red);
   colaxis type=discrete label="Hospital Admission Counts" Fitpolicy = thin;
   rowaxis label="Relative Frequencies";
run;
```

The resulting plots are shown in Figure 9. The plots suggest that the observed relative frequencies of various hospital

admission counts from each center are close to the fitted ZIP probabilities and account for the excess zeros very well.

Figure 9 Relative Frequencies Comparison Plot
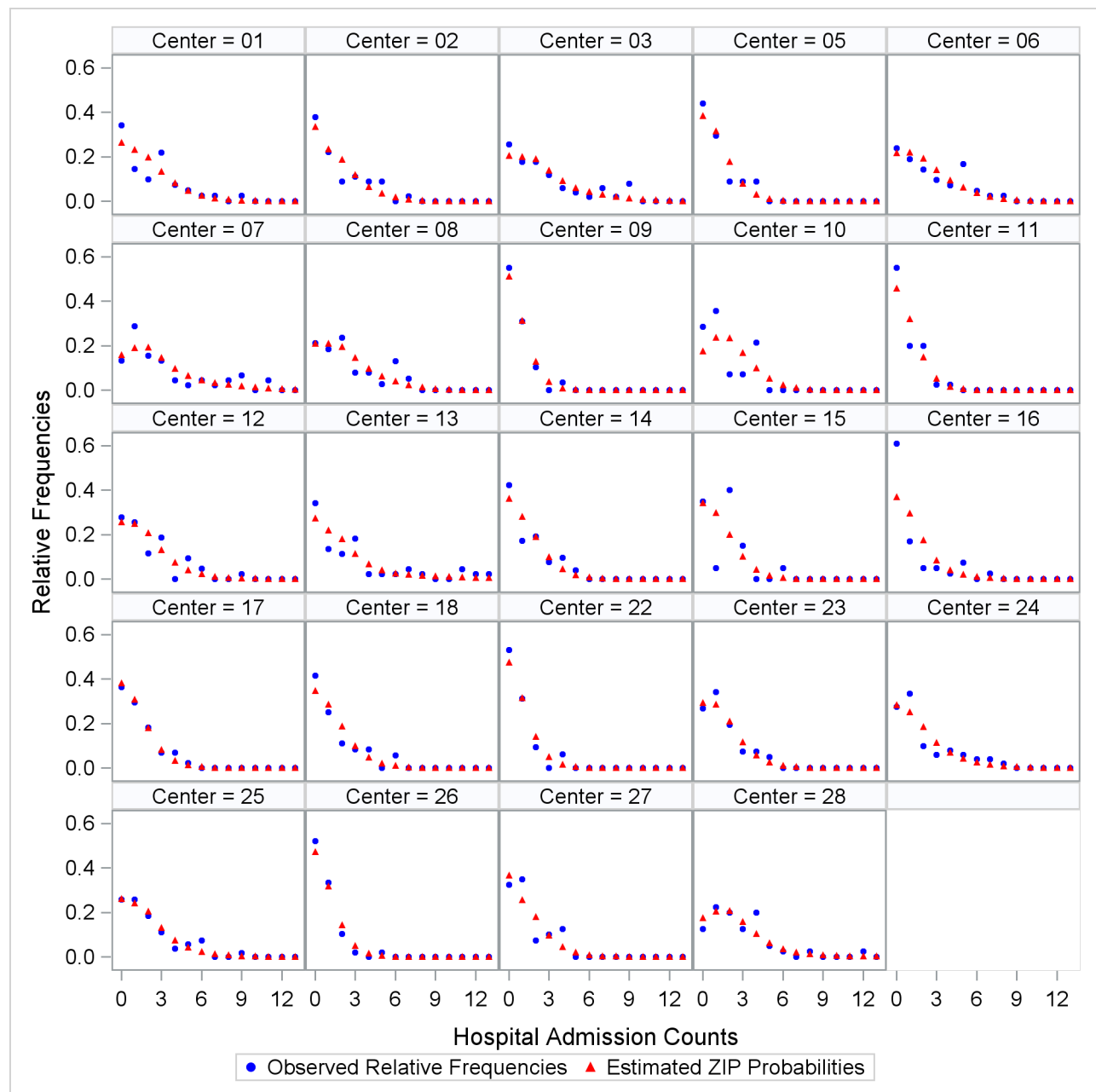
CONCLUSION

When data are collected on nested clusters at multiple levels, observations within the same cluster tend to be correlated. One way of accounting for this correlation is to use "cluster-specific" parameters in the model. Using a set of random effects to construct these cluster-specific parameters is the mixed modeling technique for analyzing multilevel data. This paper illustrates this technique by using the NLMIXED procedure in SAS/STAT. You use the RANDOM statement in PROC NLMIXED to fit the nonlinear mixed model. Because the nested levels increase in the data, nested levels of random effects are required to describe the cluster-specific dependencies in the data set. Models that incorporate nested levels of random effects are called nested nonlinear mixed models. SAS/STAT 13.2 enhanced PROC NLMIXED to support multiple RANDOM statements in order to fit these nested nonlinear mixed models. This paper provides an example to show how to fit a nested nonlinear mixed model that has two levels of nested random effects to the data.

REFERENCES

Food and Drug Administration (1999). "Guidance for Industry: Population Pharmacokinetics." Updated regularly online. http://www.fda.gov/downloads/scienceresearch/specialtopics/womenshealthresearch/ucm133184.pdf.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*. 2nd ed. Cary, NC: SAS Institute Inc.

Pinheiro, J. C., and Bates, D. M. (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics* 4:12–35.

Vonesh, E. F. (2012). *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS*. Cary, NC: SAS Institute Inc.

Wolfinger, R. D. (1999). "Fitting Nonlinear Mixed Models with the New NLMIXED Procedure." In *Proceedings of the 24th Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc. http://www2.sas.com/proceedings/sugi24/Stats/p287-24.pdf.

Zhu, M. (2014). "Analyzing Multilevel Models with the GLIMMIX Procedure." In *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. http://support.sas.com/resources/papers/proceedings14/SAS026-2014.pdf.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Raghavendra Rao Kurada
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
919-531-2083
Raghavendra.Kurada@sas.com