

ST 762, HOMEWORK 6, FALL 2009

These problems are to be turned in on the due date.

1. Recall the data in Homework 5, Problem 3, from a clinical trial studying the effectiveness of a treatment for patients with respiratory illness. See that problem for a description of the study and the contents of the data file. As in that problem, let Y_{ij} be the respiratory status for patient i at week t_{ij} , where $t_{ij} = 0, 1, 2, 3, 4$ weeks for $j = 1, \dots, 5$ for all subjects and $i = 1, \dots, 111$. Let $\delta_i = 0$ if patient i was assigned to placebo and 1 if assigned to active treatment. Let $w_{ij} = 0$ if $t_{ij} = 0$ and $w_{ij} = 1$ if $t_{ij} > 0$.

In that problem, we adopted a population-average perspective, focusing on the model

$$E(Y_{ij}|\delta_i) = P(Y_{ij} = 1|\delta_i) = \frac{\exp(\beta_1 + \beta_2 w_{ij} + \beta_3 w_{ij} \delta_i)}{1 + \exp(\beta_1 + \beta_2 w_{ij} + \beta_3 w_{ij} \delta_i)}, \quad j = 1, \dots, 5. \quad (1)$$

We also assumed a working model for the marginal covariance matrix for the Y_{ij} given δ_i , and we fit the overall population-average mean-covariance model using “GEE-1” methods, with the covariance parameters estimated different ways as implemented in SAS `proc genmod`, `proc glimmix`, and the `nlinmix` macro.

An alternative approach is to adopt a subject-specific perspective, and instead start with a model of the form

$$E(Y_{ij}|\delta_i, b_i) = P(Y_{ij} = 1|\delta_i, b_i) = \frac{\exp(\beta_1 + \beta_2 w_{ij} + \beta_3 w_{ij} \delta_i + b_i)}{1 + \exp(\beta_1 + \beta_2 w_{ij} + \beta_3 w_{ij} \delta_i + b_i)}, \quad (2)$$

where $b_i \sim \mathcal{N}(0, D)$ is a scalar random effect with variance D . Following the discussion on pages 355–356, this is a *different* model from (1) in the sense that it implies a different specification for $E(Y_{ij}|\delta_i)$ upon integration over the distribution of the random effect. Moreover, the interpretations of the parameters $\beta_1, \beta_2, \beta_3$ in (1) are *different* from those for $\beta_1, \beta_2, \beta_3$ in (2).

- (a) Give an interpretation of the parameters $\beta_1, \beta_2, \beta_3$ in (2). In particular, what issue does β_3 address?
 - (b) Fit model (2) to the data using `proc glimmix` with the `method=mspl` option and again with the `method=quad` option. (You will have to consult the documentation for a description of what these options do.)
 - (c) Fit the model again using `proc nlmixed` with the default option for carrying out the integration (adaptive Gaussian quadrature).
 - (d) Fit the model again using the `glimmix` macro with `method=ml` in the `procopt` statement.
 - (e) Compare the results of the fits in (b)–(d). Which ones would you expect to be similar, and are they? Also compare the results to those obtained from fitting the model (1) in Homework 5. Are the results comparable? Should they be?
2. Interferon- α -2b (IFN) is one of many experimental therapies being evaluated for treating patients infected with hepatitis C virus (HCV). A key way of evaluating such treatments is in terms of how they impact the within-patient *dynamics* of the HCV virus. That is, investigators are interested in understanding the behavior of *viral load*, which is roughly a measure of the concentration of virus in the patient’s system, following the start of IFN therapy.

Researchers have developed mathematical models in the form of systems of differential equations to characterize the dynamics, i.e., how the virus behaves in the body and how the body reacts and how IFN therapy may affect these processes over time. In a recent article in the journal *Science* (Volume 282, October 2, 1998, p. 103), Neumann et al. propose such a model to describe the dynamics over the first two days of IFN therapy. If the differential equations are solved, they imply that, if $V(t)$ is the viral load at time x (days) following the start of IFN therapy at day $x = 0$,

$$V(x) = V_0[1 - \epsilon + \epsilon \exp\{-c(x - t_0)_+\}], \quad 0 \leq x \leq 2, \quad V_0, c, \epsilon > 0, \quad (3)$$

where $x_+ = x$ for $x > 0$, and 0 otherwise; V_0 is the viral load at day 0 (prior to start of therapy); c is the *virion clearance rate*; t_0 is the so-called *pharmacological delay* such that decay in viral load is not seen until IFN has made sufficient progress in distributing through the body; and ϵ is an *efficacy* parameter. The efficacy parameter has the interpretation that the effect of IFN therapy is to reduce the production of new virus particles, or *virions*, in the system by the fraction $(1 - \epsilon)$; if $\epsilon = 1$ then the therapy is said to be “perfect.” Here, we will take the pharmacological delay to be known: $t_0 = 0.20$ days.

In the file `hcvmix.dat` on the class web page you will find data from 30 HCV-infected subjects who began IFN therapy on day 0. The first column is subject id number, the second column is day (measured from day 0) at which a measurement of HCV viral load (copies of HCV viral RNA per ml) was taken, and the third column is the viral load itself divided by 10^6 . For each subject, viral load measurements were taken over the first two days of therapy.

For subject i , assume that $f(z_{ij}, \beta_i)$ is given as above, where z_{ij} is the j th day value for this subject, $\beta_i = (\beta_{1i}, \beta_{2i}, \beta_{3i})^T$, $\beta_{1i} = \log(V_{0i})$ is the logarithm of the viral load at day 0, $\beta_{2i} = \log(c_i)$ is the log of virion clearance rate, and $\beta_{3i} = \log(\epsilon_i)$ is the log of efficacy. Viral loads are well-known to be subject to constant coefficient of variation at the individual level. Moreover, the researchers were willing to assume that within-subject correlation is negligible. They proposed the following subject-specific model. Letting Y_{ij} be the j th viral load measurement for the i th subject,

$$E(Y_{ij}|z_{ij}, \beta_i) = f(z_{ij}, \beta_i), \quad \text{var}(Y_{ij}|z_{ij}, \beta_i) = \sigma^2 f^{2\theta}(z_{ij}, \beta_i),$$

where θ is known. The variance parameters σ^2 and θ are common across subjects, reflecting the belief of a similar pattern of intra-subject variation. The second-stage model is given as

$$\beta_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i,$$

where \mathbf{A}_i and \mathbf{B}_i are (3×3) identity matrices, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$, and $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})^T$.

The researchers would like to fit this model and get estimates and standard errors for the typical values of log day-0 viral load, log clearance rate, and log efficacy.

You will help them by doing this three ways:

(a) Use the SAS macro `nlinmix` to fit this model to the data using the “refined” approximation discussed on pages 449–452 of the notes. Specify `method=m1` in the `procopt` statement. Because `nlinmix` cannot estimate intra-individual variance parameters, set θ equal to some likely value of your choice. (You may wish to revise your choice after seeing the answers to (b) and (c).)

(b) Use the R function `nlme()`. Let `nlme()` estimate the variance power parameter θ from the data, as in the example on the web page, so you can see if the data support the investigators’

choice of $\theta = 1$. Full information on the current syntax for the `nlme()` function can be found at <http://cran.r-project.org/>.

(c) Use SAS `proc nlmixed` with the default method of integration (adaptive Gaussian quadrature), and again let the procedure estimate the variance power parameter θ from the data, as in the example on the web page. You will likely want to use estimates from (a) or (b) above as starting values. You will need to visit the `nlmixed` documentation to see how to tell the procedure that you have a model with 3 random effects.