Biostat/Stat 571
Winter 2010
P. Heagerty

Exercise #7
Due: March 3, 2010

Reading:
- ○ Verbeke & Molenberghs, Chapter 15 & 16
- • Diggle, Liang & Zeger, Chapter 7 & 8

1. **LMM and Missing Data**

Introduction: Evaluation of new compounds for the treatment of schizophrenia often involve longitudinal measurements of disease symptoms. The Positive and Negative Syndrome Scale (PANSS) is an instrument devised to measure changes in psychopathology of psychotic (usually schizophrenic) conditions over time. It is now the predominant instrument used to measure change in pharmacological trials of antipsychotic agents.

- A longitudinal study was conducted to compare six treatment groups:

```
Treatment codes:
   1=haloperidol              2=placebo
   3=risperidone10            4=risperidone16
   5=risperidone2             6=risperidone6
```

Data:

- The response is PANSS score measured over a 3 month period. A larger score indicates a poorer response.

- There were approximately 87 subjects assigned to each of the treatment groups.

- Measurements were taken at times: -1 0 1 2 4 6 8 weeks (where 0 is the date of randomization).

Note: For this analysis consider only the following (3) treatment groups: placebo; haloperidol; and risperidone, formed by combining risperidone 6 and risperidone 10.

Questions:

(a) Construct appropriate numerical and graphical summaries to display systematic and random variation in the PANSS score. Is there evidence that certain treatment regimes appear to offer benefit?

(b) Summarize the drop-out pattern by treatment group. Use simple numerical summaries of the dropout time distribution for each of the three treatment groups.

(c) Selection Model $(D \mid Y)$: Use CRM (discrete survival) models to determine if "level" and "trend" in the response are predictors of drop-out. That is, use both Y_{ij} (level) and $Y_{ij} - Y_{ij-1}$ (trend) as

predictors of drop-out at time t_{j+1} (regress R_{ij+1} on Y_{ij} and $Y_{ij} - Y_{ij-1}$ where $R_{ij} = 1$ if a subject is observed at time t_{ij} and 0 otherwise). Consider a subject a "drop-out" at time t_j if this is the first time that they are missing the outcome variable (this is a form of missingness that is called "monotone" missingness where once $R_{ij} = 0$ we have $R_{ij'} = 0$ for all $j' > j$). Also consider the "change" $Y_{ij} - Y_{ij-1}$ and see if both the "level" (Y_{ij}) and the change predict dropout. What does this analysis suggest about the missing data mechanism?

(d) Pattern (Mixture) Model ($Y \mid D$): An alternative way to see if response and dropout are related is to plot the longitudinal response pattern for each given dropout time. That is, separate the subjects according to their drop-out time and then plot the mean response up until their drop-out. Based on these plots is there evidence that subject who drop-out are different than those that remain in study?

(e) We have seen that semi-parametric methods can lead to biased estimates when the data are not missing completely at random (MCAR). However, if we assume that the data are missing at random (MAR), that is, the missingness is predicted by the observed Y's but **not** the unobserved Y's, then maximum likelihood analysis using the correctly specified response model will yield valid estimates. Assume that the missingness is not informative (ie. assume MAR) and use LMM and ML to determine if there are differences at 8 weeks depending on the type and dose of medication received. (Suggestion: if we are interested in group differences at 8 weeks then use of `weeks-8` as the time variable is useful).

(f) Plot residuals versus time for each of the treatment groups and comment. Finally, summarize your analysis stating appropriate conclusions and limitations.

(g) Compare your analysis using LMM to analysis using GEE — specifically fit your regression model using *independence* and *exchangeable* correlation models. Please comment on the whether the inference using GEE appears to differ substantially from that obtained using LMM.