STA/BST 224 Longitudinal Data Analysis

Problem Set 2
DUE: May. 5, 2020 (Tue)

Instruction:

- Please submit it through Canvas. You can scan or take picture of hand-written solution as long as it is clear.
- You can use either R/SAS/Stata or other software. Please include program and important results.
- The grade will be average of all problems. Due to policy of Graduate Studies, all students need to work on the same problems. You can work together on the homework and ask TA for help. However, each of you is responsible for your own statistical programming and for writing-up your solutions in your own words.

1. Prove/derive the following mentioned in lecture notes:

   (a) (Page 46-47 of Note 4) For Exchangeable plus Exponential Variance-Covariance Model:

   $$\epsilon_{ij} = U_i + W_{ij},$$

   where $U_i \sim N(0, \nu^2)$ are independent of each other and independent of W_{ij}'s, W_{ij}s follow an exponential correlation model with autocorrelation parameter α and $\text{var}(W_{ij}) = \delta^2$, i.e.,

   $$W_{ij} = \alpha W_{i,j-1} + E_{ij},$$

   $E_{ij} \sim N\{(0, \delta^2(1-\alpha^2)\}$ are independent of one another and independent to U_i and W_{ij}.
   Compute variance $\text{var}(\epsilon_{ij})$ and correlation $\text{corr}(\epsilon_{ij}, \epsilon_{ik})$.
   (You can use the result of $\text{corr}(W_{ij}, W_{ik})) = \alpha^{|t_{ij} - t_{ij}|}$)

   (b) (Page 13 of Note 4) Provide the proof details after the line of

   $$-\frac{1}{2}\log \sigma^{2(N-p)}|I| - \frac{1}{2\sigma^2} y' A' y$$

   to show that

   $$\hat{\sigma}^2_{ReML} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{N-p}$$

   (You can use the properties that $A'y = y - X\hat{\beta}$ and A is symmetric and idempotent (ie, $AA = A$).)

2. A study was conducted on 41 subjects to compare subjects suffering from panic attacks (group=1) to control subjects (group=2) on a variety of psychophysiological indices in a variety of settings. At each of 11 time points (measured in minutes), anxiety (`anx`) was measured on a scale of 0 to 8, as well as carbon dioxide expiration and pulse rate. The "settings" were either rest, being spoken to about a topic of anxiety to the subject, or asking the subject to hyperventilate. The response of interest here is the anxiety measure (`anx`) (considered as a continuous measure).

   The data `anx.csv` is on the course website. The only goal of this problem is to explore the data and decide how to model the variance-covariance structure of the residuals in a model for anxiety.

   (a) Obtain the residuals in a model for anxiety, ignoring longitudinal structure for now. We will consider a very general mean model, consisting of categorical effects of each time point interacted with each group (22 mean model parameters). Examine correlation matrix and auto-correlation function using the residuals obtained. Explain your exploratory results.

(b) Assume now we want to compare two candidate models: eg, exponential v.s. exchangeable plus exponential (or other two nested candidate VCC models selected based on your findings in (a)). Fit the two models with the flexible mean model in (a) using ReML, and report covariance parameter estimates. Do the LRT to compare the two models. Provide the test statistic, degrees of freedom and P-value. Draw conclusions.