P8157: Analysis of Longitudinal Data
Homework #3

Question 1:

In this question you are going to analyze the '`wtloss`' data, available on the course website. Briefly, the data come from a weight loss trial in which K=120 patients were randomized to three treatment arms: dietary counseling at baseline (`diet`=0), dietary counseling at all sudy visits (`diet`=1), and dietary counseling at all visits plus free access to an exercise facility (`diet`=2). Each patient visited the study clinic monthly, for up to 12 months; at each visit their weight was measured.

(a) Use `lme()` to fit two linear mixed effects models, both including a main effect for diet, a main effect for time, and a diet by time interaction:

  (i) random intercepts only

  (ii) random intercepts and random slopes

  Report the results in a table that would be suitable for a clinical journal, and provide precise interpretations of the fixed effects and variance components from model (ii)

(b) Consider conducting a test for whether the random intercepts/slopes model provides a significantly better fit to the data than the random intercept model. Write down the null and alternative hypotheses. In class we learned that this test is non-standard testing scenario, and the likelihood ratio test statistic under the null is a mixture of χ^2_1 and χ^2_2 distributions. In the `lme` help file look up `simulate.lme`. Use this function to simulate the null distribution, setting `n.sim`=1000 and `seed`=1504. Plot this distribution, along with the χ^2_1 distribution and the χ^2_2 distribution, highlighting the 95^{th} percentiles. What do you conclude about the adequacy of the random intercept model as compared to random intercept/slopes model.

(c) Conduct a residual analysis of model (i) from part (a). Report your results in a concise manner and briefly summarize what you conclude, including whether the results from this analysis are consistent with what you concluded from part (b).

(d) Use `geeglm()` to fit the same mean model from part (a) using GEE 1.5, based on: (i) working independence (GEE-I), (ii) working exchangeable (GEE-E), and (iii) working AR-1 (GEE-AR1). Report the results in a table that would be suitable for a clinical journal, and provide precise interpretations of the regression parameter estimates from GEE-E.

(e) State the null and alternative hypothesis for the test of whether the rate of weight loss differs for the treatment groups. Conduct the test for the GEE-E estimator and describe the results using language that would be suitable to a non-biostatistician collaborator

(f) Assuming the mean model is correctly specified, comment on the consistency of the point estimates reported in parts (a) and (d), as well as on the validity of the standard error estimates.

Question 2 (Optional):

In this question we are going to try to understand how much each cluster (or subject) and each observation per cluster (or subject) is weighted by GEE. That is, even though $\boldsymbol{V}_k$ is called a 'working covariance matrix,' it might be more natural to think of it as a *working weighting matrix*, where the weight matrix $\boldsymbol{W}_k$ is equal to $\boldsymbol{V}_k^{-1}$. For simplicity, we'll consider the special case where the response is continuous and the variance is constant (i.e. homogeneous). In that case, the GEE estimating equation is given by:

$$\sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \;=\; \sum_{k=1}^{K} \boldsymbol{X}_k^T \boldsymbol{W}_k(\boldsymbol{Y}_k - \boldsymbol{\mu}_k).$$

Define the total weight given to cluster k as

$$W_{tot,k} \;=\; \sum_{i=1}^{n_k}\sum_{j=1}^{n_k} w_{k,ij},$$

where $w_{k,ij}$ is the $(i,j)^{th}$ element of $\boldsymbol{W}_k$.

(a) Assuming an exchangeable correlation structure with correlation parameter ρ, calculate $W_{tot,k}$ as a function of n_k and ρ using the identity:

$$(a\boldsymbol{I}_m + b\mathbf{1}_m)^{-1} = \frac{1}{a}\left(\boldsymbol{I}_m - \frac{b}{a+mb}\mathbf{1}_m\right).$$

where $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix and $\mathbf{1}_m$ denotes the $m \times m$ matrix of 1's.

(b) From this, derive the form of the relative weight of a person with n_k=10 to one where n_k=5.Calculate this value for ρ=0.9, ρ=0.5, and ρ=0.1 and comment on the trend that you observe.

(c) What do the results in part (b) say about the weight for each subject when working independence is used?

(d) The per-observation weight (per single observation within a person) can be thought of as $W_{tot,k}/n_k$. Using the results from part (b), comment on the trend in the per observation weight received.