STA 303H1S / STA 1002HS: Repeated Measures Practice Problems

1. Using the estimated β's from the "Solution for Fixed Effects", the least squares mean for diet HM at time 2 is $0.9435 + 0.2982 + 0.02423 - 0.04256 = 1.2233$.

2. (a) MODEL 1:

$$\begin{aligned} Y_{ijk} &= \beta_0 + \beta_1 I_{[diet=HG],ijk} + \beta_2 I_{[diet=HM],ijk} + \beta_3 I_{[time=1],ijk} + \beta_4 I_{[time=2],ijk} \\ &+ \beta_5 I_{[diet=HG],ijk} \times I_{[time=1],ijk} + \beta_6 I_{[diet=HG],ijk} \times I_{[time=2],ijk} \\ &+ \beta_7 I_{[diet=HM],ijk} \times I_{[time=1],ijk} + \beta_8 I_{[diet=HM],ijk} \times I_{[time=2],ijk} \\ &+ u_{ij} + e_{ijk} \end{aligned}$$

where Y_{ijk} is LDL for the ith subject on diet j at time k, $k = 1, 2, 3$, $j = 1, 2, 3$, $i = 1, \ldots, n_j$ where n_j is the number of subjects in diet j, $I_{[condition],ijk}$ is 1 if the condition holds for the ijkth observation and 0 otherwise, u_{ij} is the random effect for the ijth subject, and e_{ijk} is the random error for the ijkth observation. In addition to the 9 β's, the following 2 variance-covariance parameters are estimated: $\sigma_u^2 = \text{Var}(u_{ij})$ and $\sigma_e^2 = \text{Var}(e_{ijk})$.

(b) For each model, AIC is "-2 Res Log Likelihood" plus 2 times the number of covariance parameters and BIC is "-2 Res Log Likelihood" plus the number of covariance parameters times the log of 64 (the number of subjects).

(c) H_0 : Model 3 fits the data as well as Model 4; that is, the variances and covariances are the same for each diet
H_a : Model 4 (the model with more parameters) fits the data better; that is, at least one of the variances and covariances are not the same for each diet
The test statistic is $292.6 - 272.7 = 19.9$
From the chi-square distribution with 12 degrees of freedom, the p-value is between 0.05 and 0.10. So there is weak evidence that model 4 fits the data better; that is, there is weak evidence that the variance-covariance parameters differ with diet.

(d) AIC is much smaller for the two models using UN covariance structure. It is smallest for the model using UN and with the same variance-covariance structure for all diets, but in part (c) we saw that there is weak evidence that the model for which the variance-covariance structure differs with diet fits the data better. Either of models 3 or 4 would be a reasonable choice. I'll go with model 3, in the interest of parsimony.

(e) Looking at the "Type 3 Tests of Fixed Effects" for model 3, there is no evidence ($p = 0.8502$) that the effects of diet on the mean value of LDL differ over time. To examine the main effects, it would be best to look at the model with the interaction term removed, but from model 3 it appears that there is no evidence ($p = 0.1572$) of a change in mean of LDL over time and that there is strong evidence ($p = 0.0012$) of differences in the mean of LDL among diets. Thus it appears that the mean LDL differs among the diets at all times in the same way; so it starts different at baseline and this persists.

(f) Again, it would be helpful to see the model with the interaction term removed since it is not significant. But looking at the given output, there is no diet effect for this model. This does not contradict the conclusion in part (e) since the differences among diets are due to differences at baseline and here we have controlled for baseline. In this model we have weak evidence of a difference in mean LDL between times 2 and 3. The model controlling for baseline is typically more powerful than the model in which baseline is used as a response variable and in this case the model controlling for baseline was able to pick up weak evidence of a time effect. (From the plot, we can see that LDL decreases from time 2 to time 3.)

3. (a) From the tests of pairwise differences between the least squares means, we conclude that there is strong evidence ($p = 0.0003$) of a difference in mean LDL between diets HM and LG, moderate evidence ($p = 0.0337$) of a difference in mean LDL between diets HG and HM, and no evidence ($p = 0.1345$) of a difference in mean LDL between diets HG and LG.

   Since LG is used as the reference diet in the model (the diet for which there is no indicator variable), we can use the t-tests for the model coefficients to compare diets HG to LG and diets HM to LG over and above the effects of time. These are the same conclusions as from the differences in least squares means. But from these tests we cannot compare diets HG and HM.

   (b) The spread in the studentized residuals is about the same for each diet-time combination although it may be slightly more for diet HG (which has the predicted means of 3.24 for time 1, 3.31 for time 2, and 3.20 for time 3), particularly when we consider its spread at time 2.

   (You can figure out which predicted mean corresponds to which diet-time combination by using the $\hat{\beta}$'s to find the predicted means.)

   (c) From the plot of the residuals versus the predicted means: There is a potential outlier in diet HG at time 2 with a large, negative residual. It should be looked into further. Without it, it seems reasonable to assume that variances are the same across diets and the model fit is appropriate.

   The normal quantile plot shows only very minor departures from normality. The outlier is evident in the left tail. This outlier is not at all extreme so any concerns about departures from normality are very minor and they don't require remedial measures; the p-values for the tests of fixed effects and the least squares means will still be approximately correct.

4. Model 10.5 from the text has 3 covariance parameters: a variance for each gender plus the variance of the random effect for subject. There are 4 observations (at different ages) for each subject. So the unstructured variance-covariance matrix for each subject has 10 covariance parameters. Since the second model allows for variances and covariances to differ across genders, the second model has a total of 20 covariance parameters. We can use a likelihood ratio test to test that the 17 ($17 = 20 - 3$) restrictions hold, where the restrictions make the variance-covariance matrices for the unstructured model equivalent to those for model 10.5. The null hypothesis is that the restrictions hold (so the 2 models fit the data equally well). The test statistic is the difference in -2 `Res Log Likelihood` $= 14.3$. Under the null hypothesis, this is an observation from a chi-square distribution with 17 df. Since the test statistic is less than the df, you know that the p-value is large. (The actual value is $p = 0.65$; from tables you can estimate $0.1 < p < 0.9$.) So the data are consistent with the simpler covariance structure of model 10.5.