# CPH 931 — Fall 2008 — Dr. Charnigo

## Written Assignment 2

Written Assignment 2 is due on Tuesday 07 October at the end of lecture.

[50] 1. Consider the data in {DSST.xls}, previously discussed in Lectures 3 and 4. Revise the linear mixed model from Lecture 4 to include the following additional interaction terms: DRUG*ACTIVITY, SSSTAT*DRUG*ACTIVITY.

[10] a. Specify $m$ and define $X_1, \ldots, X_m$ for the revised linear mixed model.

[10] b. State (in terms of $\beta_1, \ldots, \beta_m$) and test the null hypothesis that DRUG and ACTIVITY do not interact within either stratum of SSSTAT.

[10] c. State (in terms of $\beta_1, \ldots, \beta_m$) and test the null hypothesis that ACTIVITY is altogether irrelevant to the number of CORRECTTRIALS.

[10] d. For someone with a low sensation seeking personality, what is the estimate of the expected difference in the number of CORRECTTRIALS from TIME = 2 on placebo following a high sensation seeking activity to TIME = 4 on amphetamine following a low sensation seeking activity? Is the estimate significantly different from zero?

[10] e. For someone with a high sensation seeking personality, what is the estimate of the expected difference in the number of CORRECTTRIALS from TIME = 3 on amphetamine following a high sensation seeking activity to TIME = 4 on placebo following a low sensation seeking activity? Is the estimate significantly different from zero?

[50] 2. Consider the data in {Mercury.xls}, previously discussed in Lecture 5. Let $Y$ denote log 3-year standard mercury, $X_1$ denote pH, $X_2$ denote log chlorophyll, $Z_1$ denote log average mercury, $Z_2$ denote log minimum mercury, and $Z_3$ denote log maximum mercury. Note that the variables in {Mercury.xls} are not log-transformed, so you will need to create $Y$, $X_2$, $Z_1$, $Z_2$, and $Z_3$ in SAS.

[10] a. Fit the linear regression model with response variable $Y$ and explanatory variables $X_1, X_2$ (using ordinary least squares) based on the 43 records with no missing values. Report the parameter estimates and their standard errors.

[10] b. Generate five complete data sets via multiple imputation, involving $X_1, X_2, Y$ in the multiple imputation. Report the filled-in values of $Y$ for Lake Annie in each of the five complete data sets.

[10] c. Using the five complete data sets from part b, fit the linear regression model with response variable $Y$ and explanatory variables $X_1, X_2$. Derive overall parameter estimates and standard errors from the five sets of results. Compare the overall parameter estimates and standard errors obtained here to those acquired in part a.

[10] d. Generate five complete data sets via multiple imputation, now involving $Z_1, Z_2, Z_3$ in the multiple imputation along with $X_1, X_2, Y$. Report the filled-in values of $Y$ for Lake Annie in each of the five complete data sets.

[10] e. Using the five complete data sets from part d, fit the linear regression model with response variable $Y$ and explanatory variables $X_1, X_2$. Note that $Z_1, Z_2, Z_3$ are not explanatory variables in the linear regression model even though they were used in the multiple imputation. Derive overall parameter estimates and standard errors from the five sets of results. Compare the overall parameter estimates and standard errors obtained here to those acquired in parts a and c.