

**2016 USC Marshall Statistics PhD Screening Exam**  
**Part II: Take Home**

Question 6 (50 points): The ICU data set consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult intensive care unit (ICU). The major goal of this take-home exam is to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients. The data set and a code sheet for the variables are attached. The primary outcome variable is vital status at hospital discharge, STA.

Use this data set to build a logistic regression model that is simple and interpretable. Please provide a report that includes the following components.

- (a) Clearly describe your strategy and criteria for selecting the “best” model.
- (b) Assess the adequacy of the model including individual variables and the overall fit.
- (c) Explain how you would use the final model to obtain the estimated logit and its standard error for each subject. Provide the estimated logit and 95% CI for the first five subjects in your report.
- (d) Explore potential interactions among the selected predictors. Assess the new model and compare it with the one that you obtained previously.