Statistics 659 - Assignment 7

(due Wednesday, March 23, 2016, 11:59pm)

Instructions:

- Whether you write out the solutions by hand or in a text document, be sure that they are *neat, legible and in order* (even if you choose to solve them in a different order).
- **Type** your name, email address, course number, section number and assignment number at the top of the first page (or cover page).
- Either scan or print your solutions to a **PDF** file under 15MB in size. It must be in a *single* file, not separate files for separate pages. Name the file using your name (for example, I could use twehrly659hw01.pdf) to avoid confusion with other students and/or assignments. *Do not* take a photo of each page and then paste them into a document – this will make your file too big and the results will generally not be very readable anyway.
- Login to your WebAssign account to upload your file. You must do this by **11:59 pm U.S. Central time**, *according to the WebAssign server*, on the due date. We highly recommend that you start the upload at least 15 minutes earlier. You can make multiple submissions but *only the last submission will be graded*.

This assignment covers the material in Chapter 5 discussed in Lectures 21 through 23.

5.1, 5.2 (use the forward selection and backward elimination procedures described in class and compare results), 5.4, 5.5, 5.6, 5.7, 5.10abde, 5.13

Variable Selection Problem

Hosmer and Lemeshow's ICU study considers a sample of 200 subjects who were part of a larger study on the survival of patients following admission to an adult intensive care unit (ICU). The goal is to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients.

Reference: Lemeshow et al. (1988), Journal of the American Statistical Association, 348-356.

1. Use the forward selection, stepwise selection, and backward elimination procedures to select "best" models among those models without interaction terms.

2. Follow the steps that we used to fit the UMARU data in class to develop a logistic regression model (the preliminary main effects model) with variable vital status (`sta`) as the response and the other variables (except `ID`) as predictors. Document your steps in the analysis. Be sure to justify your final model.

3. Discuss the relationship, if any, of these models you found with the automatic variable selection procedure in part 1. with the model you developed in part 2. Also, discuss the similarities and differences of the models.

(Only for students having taken STAT 414, 610 or STAT 630) 5.6