Homework 1

IGP 304: Statistics for Biomedical Research (Spring 2007)

due on January 29, 2007

The following questions relate to the *1.4.11.Sepsis.dta* data set, which can be found in "Statistical Modeling for Biomedical Researchers", 2nd Ed., *in press*, by William Dupont. The data set can be downloaded from the web *http://biostat.mc.vanderbilt.edu/dupontwd/wddtext/* or the homework section of this course web.

A short description of the 1.4.11.Sepsis.dta data set is as follows. Bernard et al. (1997) performed a randomized clinical trial to assess the effect of intravenous ibuprofen on mortality in patients with sepsis. People with sepsis have severe systemic bacterial infections and sepsis is a life threatening condition. One measure of a patient's mortality risk is the Acute Physiology and Chronic Health Evaluation (APACHE) score.

1. List the names and labels of all variables in the *1.4.11.Sepsis.dta* data set. (*hint* help labels)

2. What are the numeric values of the *race* variable? Which races do these numeric codes represent? Can you answer this question without opening the data editor?

3. List the APACHE score and baseline temperature of the six patients with the lowest APACHE scores. List the APACHE score, fate and ID number of all black patients whose APACHE score is 35 or greater.

4. Draw dot plots of baseline temperature in black and white patients. Draw these plots on a single graph. Do not include people of other races. Where does Stata obtain the title of the y-axis of your dot plot?

5. Draw box plots of temperature at two hours in treated and untreated patients.

6. Consider treated patients whose race is recorded as "other". Test whether these patients baseline temperature is significantly different from their temperature after two hours. What is the *P*-value associated with this test? How many degrees of

freedom does it have? What is a 95% confidence interval for the true change in temperature among this group of subjects?

7. Test whether baseline APACHE score is different in treated and untreated patients. What is the *P*-value associated with this test? How many degrees of freedom does it have? What is a 95% confidence interval for the true difference in APACHE score between treated and untreated patients. Why is this test important in a clinical trial of the efficacy of ibuprofen in septic patients?

8. (*Bonus point:*) Draw the time trend plot in [EMS] Fig. 3.13. The data set can be downloaded from the course web *http://biostat.mc.vanderbilt.edu/LeenaStataNotes*.