Stats for Clinical Trials, Math 150
Jo Hardin
Homework # 5, Due: Thurs, 10/7/10

The following questions relate the to the `sepsis.csv` data. [Bernard, G. R., A. P. Wheeler, et al. (1997). "The effects of ibuprofen on the physiology and survival of patients with sepsis. The Ibuprofen in Sepsis Study Group." *N Engl J Med* 336: 912-8.] The goal of the study is to determine whether taking ibuprofen will increase survival and improve health of patients with sepsis (blood poisoning).

The data set contains information on patients from the Ibuprofen in Sepsis trial. The variables are as follows (note: a patient's APACHE score is a measure of the severity of disease, integer between 0 and 71, higher scores imply a more severe disease and a higher risk of death):

$$
\begin{aligned}
treat &= \begin{cases} 0 & \text{if patient received placebo} \\ 1 & \text{if patient received ibuprofen} \end{cases} \\
death30d &= \begin{cases} 0 & \text{if patient was alive 30 days after entry into the study} \\ 1 & \text{if patient was dead 30 days after entry} \end{cases} \\
race &= \begin{cases} 0 & \text{if patient was white} \\ 1 & \text{if patient was black} \end{cases} \\
apache &= \text{baseline APACHE score} \\
id &= \text{patient ID number}
\end{aligned}
$$

Note: to run a logistic regression you'll want to use

```
> sepsis.bu <- sepsis[c(race==1 & treat==0),]
> sepsis.logreg <- glm(death30d~apache, family="binomial", data=sepsis.bu)
> summary(sepsis.logreg)
```

1. Use logistic regression to estimate the probability of death in untreated black patients as a function of baseline APACHE score. [Note: you will have to subset your entire data set to consider only black untreated patients.]

2. What is the odds ratio associated with a unit rise in APACHE score in untreated black patients ? Give a 95% CI for this odds ratio (including interpretation in the context of the problem).

3. Is it possible to predict the probability of death (for black untreated patients) for someone with an APACHE score of 50? If so, give the predicted value. If not, explain why not. Is it a good idea to do such a prediction? Explain.

4. For black untreated patients, what is the APACHE score value that gives you the median survival rate (according to the logistic regression model). Provide a sentence that demonstrates you understand the meaning of the number.

5. Since the term *regression* refers to the mean of a response variable as a function of explanatory variable(s), why is it appropriate to use the term *logistic regression* to describe a proportion or probability as a function of an explanatory variable(s)?

6. Give at least two reasons why the simple linear regression model is inappropriate for describing the regression of a binary response variable on a single explanatory variable.