

Homework # 3

Biostatistics 210

DUE: November 17, 2009

0. BACKGROUND

This assignment will give you practice in developing a prognostic model.

1. DOWNLOAD

Download the dataset hw3.dta from the course website. The dataset is from a study which attempted to build a prognostic model for survival in primary biliary cirrhosis. The variables are

- dead5yr (1=died in 5yrs, 0 = survived 5 years) [the outcome variable]
- asictes (asictes: 1=present, 0 =absent)
- hepatom (hepatomegaly: 1=present, 0 =absent)
- edema (edema: 1=present, 0 =absent)
- spiders (1=present, 0 =absent)
- age50 (1=age greater than or equal to 50, 0=age less than 50)
- txtype (1=cadaveric donor, 0 = living donor)
- type “codebook” if you are unclear on the variables

2. ASSIGNMENT

Your assignment is to select a set of variables which appear to be predictive of 5-year survival in primary biliary cirrhosis.

Do this by performing a series of logistic regression models. Start with all predictors (the 5 variable model below), calculate the AIC and BIC using the command

```
estat ic
```

after your logistic regression.

# of Variables	AIC	BIC
5		
4		
3		
2		
1		

Discuss which model appears optimal based on the minimizing the BIC and AIC. How does this compare with choosing a model based on whether it has p-value < 0.05 in the model?

Which predictors would you select in your prediction model?