Biost 518
Emerson, Winter 02
February 8, 2002

Biost 518
Applied Biostatistics II

Midterm Examination

Name: _____

Instructions: Please provide concise answers to all questions. Rambling answers touching on topics not directly relevant to the question will tend to count against you. Nearly telegraphic writing style is permissible.

The examination is closed book and closed notes. If you come to a problem that you believe cannot be answered without making additional assumptions, clearly state the reasonable assumptions that you make, and proceed.

Appendix A contains data analysis results from a large cohort study involving 735 subjects aged 65 and older. Data available at study entry included demographic data, as well as selected measures of known risk factors for cardiovascular and cerebrovascular disease. Each subject also had an MRI exam of their brain which was measured for the evidence of cerebral atrophy (shrinking of the brain) on a scale from 0 (none) to 100 (severe atrophy). All subjects were also followed a minimum of 5 years for survival. Variables available for this analysis includes:

- mridate= The date on which the participant underwent MRI scan in MMDDYY format.
- age= Participant age at time of MRI (years)
- male= Indicator of whether participant is male (0= female,1=male)
- race= Indicator of participant's race (1= white, 2= black, 3= Asian, 4= other)
- diabetes= Indicator of whether the participant had been diagnosed with diabetes prior to MRI (0= no, 1= borderline, 2= full diabetes). Diabetes is a disease in which a patient does not regulate his/her blood glucose in a normal fashion. Glucose is the main energy source for our bodies, and in diabetes, the cells lose the ability to take glucose from the blood. Persons with diabetes are at high risk of blindness, kidney disease, heart disease, and other diseases of the circulation system.
- genhlth= An indicator of the participant's view of his/her own health (1= excellent, 2= very good, 3= good, 4= fair, 5= poor).
- atrophy= A measure of global brain atrophy detected on MRI. A 0 means little or no shrinkage of the brain, and 100 is marked atrophy of the cerebrum.
- obstime= The total time (in days) that the participant was observed on study between the date of MRI and death or September 16, 1997, whichever came first.
- death= An indicator that the participant was observed to die while on study. If death=1, the number of days recorded in obstime is the number of days between that participant's MRI and his/her death. If death=0, the number of days recorded in obstime is the number of days between that participant's MRI and September 16, 1997.

1. 20 points In Appendix A, cross out any descriptive statistics which do not provide scientifically meaningful information about the dataset. In the space below briefly explain the main factors that influenced

your decisions.

2. 10 points From the information available in Appendix A, is there any evidence that would suggest that classical linear regression would be inappropriate for detecting associations between mean DSST score and degree of cerebral atrophy? Explain your reasoning?

3. 5 points each part Appendix B contains the results of analyses that might be used to explore the association between mean DSST scores and degree of cerebral atrophy. (For this problem, you may assume that necessary assumptions are satisfied, except as noted.)

   a. What is the best estimate for the mean DSST score among participants with an atrophy grade of 20?

   b. What is the best estimate for the mean DSST score among participants with an atrophy grade of 25?

   c. What is the best estimate for the difference in mean DSST scores between subjects with atrophy grades of 21 and atrophy grades of 20?

   d. What is the best estimate for the difference in mean DSST scores between subjects with atrophy grades of 50 and atrophy grades of 20? Provide a 95% confidence interval for this difference.

   e. Provide an interpretation of the intercept from the regression model, including the statistical and scientific relevance of the estimate.

f. Provide an interpretation of the slope from the regression model, including the statistical and scientific relevance of the estimate.

g. Based on the results of the analysis, what would be your conclusion regarding the existence of an association between mean DSST score and atrophy detected on MRI exam? Explain.

h. Based on the results of the analyses presented in the appendix, what can you say about the presence of a statistically significant correlation between DSST and atrophy grades? Explain.

i. Based on the results of the analyses presented in the appendix, what can you say about the presence of a statistically association between mean atrophy grades and DSST score? Explain.

j. Suppose (for this problem only) that the 735 observations in this data actually represented repeat measurements over a 10 year interval on 300 independent subjects. How might this information be expected to affect your answers to the above questions?

3. Appendix C contains the results of several regression analyses which might be used to explore the association between cerebral atrophy detected on MRI and patient survival. For each of the following analyses, indicate whether the analysis is appropriate to address this question. If it is not, briefly explain why not. If it is, provide a very brief interpretation of the slope (just enough to show me you know what it estimates). In this problem, the variable DeadIn5 was defined as an indicator of death within 5 years of MRI exam.

   a. A linear regression of observation time (response) on atrophy (predictor).

   b. A linear regression of atrophy (response) on an indicator of death within 5 years

   c. A linear regression of atrophy (response) on an indicator of death within 5 years using robust standard error estimates

   d. A logistic regression of an indicator of death within 5 years (response) on atrophy (predictor)

   e. A logistic regression of an indicator of death within 5 years (response) on atrophy (predictor) using robust standard error estimates

f. A proportional hazards regression model of time to death on atrophy (predictor)

4. What would be the conclusions from a t test performed comparing mean atrophy across groups defined by 5 year survival? Explain your answer.