**Biost 518**
**Applied Biostatistics II**

**Final Examination**

Name: _____ Mailbox: _____

**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.

Problems 1 through 3 pertain to the analyses of data from a hypothetical study investigating the association between sex, age, and race on serum cholesterol. Appendix 1 contains results of these analyses.

1. (10 points) Models A, B, and C each model the effects of age, race, sex, and a race-sex interaction on serum cholesterol. Which of these three models is the better one to use for such purposes? Justify your answer.

2. Using the model you identified in problem 1, answer the following questions
   a. (5 points) What is your best estimate of the expected cholesterol in a black female of age 60?

   b. (5 points) What is your best estimate of the expected cholesterol in a black male of age 50?

   c. (5 points) What is your best estimate of the expected cholesterol in a black male of age 1?

d. (5 points) What is your best estimate of the expected difference in cholesterol when comparing a white female of age 66 to a white female of age 65?

e. (5 points) What is your best estimate of the expected difference in cholesterol when comparing an Asian male of age 60 to an Asian male of age 59?

f. (5 points) What is your best estimate of the expected difference in cholesterol when comparing a white male of age 66 to a white female of age 66?

g. (5 points) What is your best estimate of the expected difference in cholesterol when comparing a black male of age 66 to a white male of age 66?

h. (5 points) What is your best estimate of the expected difference in cholesterol when comparing a black female of age 66 to a white female of age 66?

i. (5 points) What is your best estimate of the expected difference in cholesterol when comparing a black male of age 66 to a black female of age 66?

3. Using the model you identified in problem 1, answer the following questions.

   a. (5 points) Is there statistical evidence that the difference between serum cholesterol in males and

females varies by race? Justify your answer, including the P value you used.

b. (5 points) Is there statistical evidence that the difference between serum cholesterol among whites, blacks, and Asians varies by sex? Justify your answer, including the P value you used.

c. (10 points) Is there a statistically significant difference between the answers to (f) and (i) in problem 2 above? Justify your answer, including the P value you used and how you obtained it.

4. Appendix 2 contains results from analyses of a hypothetical randomized clinical trial comparing three doses of a new drug with respect to systolic blood. Use the information contained in those results to answer the following questions. You may assume that the necessary assumptions for linear regression are valid for any assumption for which there is no direct information contained in the output. Where appropriate, please identify the regression model you used to answer each question.

   a. (5 points) Is there evidence of a statistically significant imbalance in the randomization groups with respect to sex? Justify your answer.

   b. (5 points) Is there evidence of confounding by sex on the effect of dose on systolic blood pressure? Justify your answer.

   c. (5 points) Suppose you decided to model dose as dummy variables. Is there evidence of an effect of the drug on blood pressure? Justify your answer, including the model used to address the question and the P value you used to make a decision. Provide a brief interpretation for the parameters in that model.

d. (5 points) Using the analysis you used in part (c), for what dose groups is there a statistically significant difference in average blood pressures? Justify your answer, including P values.

e. (5 points) Suppose you decided to model dose as a continuous variable. Is there evidence of an effect of the drug on blood pressure? Justify your answer, including the model used to address the question and the P value you used to make a decision. Provide a brief interpretation for the parameters in that model.

f. (5 points) Using the analysis you used in part (e), for what dose groups is there a statistically significant difference in average blood pressures? Justify your answer, including P values.

g. (10 points) Which of the two analyses considered in parts (c) and (e) would you prefer? Justify your answer, briefly stating the issues that you considered in making your decision.

h. (5 points) Is there evidence of confounding by dose on the effect of sex on systolic blood pressure? Justify your answer. Explain why you might get different answers to this question and part b.

5. Consider the problem of evaluating the prognostic value of the nadir PSA on time to relapse. Recall that variable *obstime* measured time until relapse or last follow-up, with variable *inrem* measuring whether the patient was still in remission at last follow-up. Also recall that everyone was followed a minimum of 24 months, thus we could construct a variable *relapse*24 that was an indicator of relapse within 24 months. For each of the following regressions, indicate whether the analysis method would be statistically and scientifically valid. When the analysis is valid, identify the measure of association being compared.

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

   b. A linear regression of nadir PSA (response) on an indicator of relapse within 2 years

   c. A logistic regression of an ian indicator of relapse within 2 years (response) on nadir PSA (predictor)

   d. A proportional hazards regression model of time to relapse as measured by *obstime* on nadir PSA (predictor)