

BIOS663 Final Exam Spring 2011  
Friday, May 6 9am-noon

*Instructions:* Please be as rigorous as possible in all of your answers and show all your work.

Remember that to report a test, you should provide  $H_0$ , the test statistic, the degrees of freedom, the p-value, the decision, and an interpretation of the decision in terms of the subject matter (if possible).

You may not consult with anyone except the instructor for clarification of questions. The work you present should be your work alone. Violation of the honor code will be prosecuted (penalties may include failure of the course and expulsion from the university). Please sign the honor code pledge and submit it with your report.

**Honor Code Pledge: On my honor, I have neither given nor received aid on this examination.**

**Name:**

**Signature:**

**Date:**

1. (40 points total) A group of subjects was recruited to a nutritional study in a medical center at UNC. The data consist of their BMI ( $y = \text{BMI}$ ), daily exercise time ( $x_1 = \text{exercise (in hours)}$ ) and daily vegetable intake ( $x_2 = \text{vegetable (in servings)}$ ). One of the objectives in this study is to estimate how the exercise and vegetable consumption affect BMI. To address the question, we consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

. Let  $\mathbf{X}$  be the associated design matrix of the above model. The data is summarized below:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 200.0000 & 588.7676 & 1033.797 \\ 588.7676 & 2321.7635 & 2951.400 \\ 1033.7973 & 2951.3999 & 7138.232 \end{pmatrix}, \mathbf{X}'\mathbf{y} = \begin{pmatrix} 4647.273 \\ 13561.768 \\ 23709.514 \end{pmatrix},$$

$$\text{and } (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.037523205 & -0.005495959 & -0.003161934 \\ -0.005495959 & 0.001712864 & 0.00008774738 \\ -0.003161934 & 0.00008774738 & 0.0005617387 \end{pmatrix}.$$

- (8 points) A partial ANOVA table is given below. Complete the table.

The GLM Procedure

Dependent Variable: y

Sum of

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	?	85.543209	?	?	-
Error	?	1101.480037	?		
Corrected Total	199	1187.023246			

- (8 points) Compute the least square estimates of the model parameters and their standard errors. Conduct the tests for the significance of each parameter (i.e.,  $H_0 : \beta_1 = 0$ , and  $H_0 : \beta_2 = 0$ ).

- (8 points) Compute the 95% confidence interval for the BMI of individuals who on average exercise 2 hours and eat 6 servings of vegetables daily.

- (8 points) Test  $H_0 : \beta_1 = 3\beta_2$ .

- (8 points) Next we center the exercise and vegetable consumptions at their means, which are 1 hour and 5 servings respectively and refit the data with the new transformed variables. Fill in the cells with ? in the following table.

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	?	?	?	-
newx1	?	?	?	-
newx2	?	?	?	-

2. (40 points total) This study investigates how the four dose levels of Vitamin C (1, 2, 3 and 4 mg) and two delivery methods (orange juice or ascorbic acid) affect the length of odontoblasts (teeth) in 800 guinea pigs. The study is balanced, so for each dose and delivery method combination, 100 pigs are assigned.
- (14 points) first consider the dose variable as categorical and employ an additive model using reference cell coding (where ascorbic acid and dosage 1mg are used as references respectively):

$$y_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \beta x_{4i} + e_i$$

where  $\alpha_i$ s refer the dosage effects and  $\beta$  refers the effect of delivery method. Describe dummy variables  $x_{1i}, x_{2i}, x_{3i}$  and  $x_{4i}$ , based on which write down the cell mean of each group in terms of  $\mu, \alpha_i$ s and  $\beta$  in the following table.

Delivery method	Dose	Mean
Orange juice	1	
Orange juice	2	
Orange juice	3	
Orange juice	4	
Ascorbic acid	1	
Ascorbic acid	2	
Ascorbic acid	3	
Ascorbic acid	4	

- (7 points) If we add the interaction terms between the delivery methods and dosage into the above model and express the new model in matrix notation  $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$ , what are the dimensions of  $\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}$  and  $\mathbf{e}$ ?

- (12 points) Let  $\mu_{orange}$  and  $\mu_{ascorbic}$  be the overall means of the two delivery methods. Write down  $\mu_{orange}$  and  $\mu_{ascorbic}$  for the models with and without interaction terms. Derive the two  $\mathbf{C}$  matrices for testing  $H_0 : \mu_{orange} = 2\mu_{ascorbic}$  under the two models.

- (7 points) Next, treat the Vitamin C dosage as a continuous variable and fit a model with additive effects of the delivery method and vitamin C level, with no interaction. Is this model nested within Model

$$y_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \beta x_{4i} + e_i?$$

If yes, write down  $H_0$  for comparing the two models and derive  $\mathbf{C}$  matrix. What are the degrees of freedom of the corresponding  $F$  test under  $H_0$ ?

3. (20 points total) Table below lists a data set derived from a study on the relationship between incubation temperature for hatching turtle eggs and gender of baby turtles.

Temperature	Male	Female	Total
low	10	40	50
medium	28	22	50
high	34	16	50

To study how the incubation temperature affects the sex of baby turtles, we fit the logistic regression model

$$\text{logit}(p) = \mu + \beta_1 I(\text{temperature} = \text{low}) + \beta_2 I(\text{temperature} = \text{medium})$$

where  $p$  is the probability of hatching a male turtle, and get the following output:

	Estimate	Std. Error	z value	Pr(> z )
intercept	0.7538	0.3032	2.486	0.0129
I(temperature=low)	-2.1401	0.4657	-4.595	4.33e-06
I(temperature=medium)	-0.5126	0.4160	-1.232	0.2179

- (10 points) Estimate the probability that a male turtle hatches from an egg incubated at medium temperature.
- (5 points) What is the estimate of the odds ratio of low vs high temperatures and construct a 95% confidence interval for this odds ratio.

- (5 points) What is the estimate of the odds ratio of low vs medium temperatures. Do you have enough information to construct a 95% confidence interval for this odds ratio? If yes, construct the CI. If not, explain why.