ST552 Midterm

Winter 2015

Answer the questions in the spaces provided on this exam.

Name: _____

- You have 50 minutes to complete the exam.
- There are 3 questions. Answer all of the questions.
- Please
  - do not look at the exam until I tell you and
  - stop writing when I announce that the exam is over.
- There is one page of statistical tables at the end of the exam. You may remove the page of tables if you desire.

Question	Points	Score
1	15	
2	10	
3	20	
Total:	45	

1. (a) Show that the least squares estimates are unbiased. You should begin by stating the multiple linear regression model in matrix form, along with any assumptions you require. (10)

(b) Imagine the errors are correlated. For example, that $\text{Var}(\boldsymbol{\epsilon}) = \Sigma$, where Σ is a symmetric $n \times n$ matrix. Are the least squares estimates still unbiased? Justify your answer. (5)

2. An experiment was conducted to explore the relationship between the *lifetime* (measured in days) and sexual activity of fruitflys.

   125 fruit flys were divided randomly into 5 treatment groups, each of 25 flys. Each treatment was designed to simulate a different level of sexual activity, with levels: *none*, *one*, *many*, *low* and *high*.

   The *thorax length* of each male was also measured as this was known to affect lifetime.

   One observation in the *many* group was lost.

   The following models were fit:

$$\text{Lifetime}_i = \beta_0 + \beta_1 \text{Thorax Length}_i + \beta_2 \text{one}_i + \beta_3 \text{many}_i + \beta_4 \text{low}_i + \beta_5 \text{high}_i + \epsilon_i$$
$$\text{Lifetime}_i = \beta_0 + \beta_1 \text{Thorax Length}_i + \epsilon_i$$

   where *one*, *many*, *low*, and *high* are indicator variables for the respective treatment groups.

   The two models have **residual sum squares** of 13107 and 22742 respectively.

   (a) Conduct an F-test to compare the two models. (6)

(b) Under what condition would the estimate for β_1 be the same for both models? (4)

3. The following regression model is fit to a subset of Galton's data on the heights of parents and their children:

$$\text{Child's Height}_i = \beta_0 + \beta_1 \text{Father's Height}_i + \beta_2 \text{Mother's Height}_i + \epsilon_i \quad i = 1, \ldots, n$$

where the heights are measured in inches. The subset consists of one male child from each family, for a total of 179 children. Results from the least squares fit are given below.

$$\hat{\beta} = \begin{pmatrix} 20.6 \\ 0.43 \\ 0.29 \end{pmatrix}, \qquad \hat{\sigma} = 2.21, \qquad (X^T X)^{-1} = \begin{pmatrix} 7.4 & -0.1 & -0.1 \\ -0.1 & 0.0009 & -0.0001 \\ -0.1 & -0.0001 & 0.0010 \end{pmatrix}$$

(a) Conduct a t-test of the null hypothesis that $\beta_2 = 0$. (6)

(b) Write a sentence interpreting your result from (a) in context of the study. (4)

(c) What is the predicted value for a future child's height when the father is 68 inches tall and the mother is 64 inches tall? (2)

(d) How would you find a standard error for the estimate in (c)? You need only state the calculation you would do, **do not do the calculation**. (2)

(e) Construct a 95% confidence interval for $\beta_1 - \beta_2$. (6)