Statistics 354 Midterm – November 2, 2018

This exam has 3 problems and is worth a total of 40 points.

You have 50 minutes to complete this exam. Please read all instructions carefully, and check your answers. Show all work neatly and in order, and clearly indicate your final answers. Answers must be justified whenever possible in order to earn full credit. ***Unless otherwise specified, no credit will be given for unsupported answers, even if your final answer is correct.*** *Points will be deducted for incoherent, incorrect, and/or irrelevant statements. For problems with multiple parts, all parts are equally weighted.*

This exam is closed-book, except that one $8\frac{1}{2} \times 11$ *double-sided page of handwritten notes is permitted. No other aids are allowed.*

You must answer all of the questions in the exam booklet provided.

1. (*8 points*) Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for all i. Let $\hat{\beta}_0$, $\hat{\beta}_1$ denote the least squares estimators of β_0, β_1, respectively. Set $\mu_0 = \beta_0 + \beta_1 x_0$ and $\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$. Determine the distribution of $\hat{\mu}_0$.

2. (*24 points*) The purpose of this problem is to derive a model of multiple regression through the origin; i.e., where it is known a priori that the intercept is zero.

Suppose that we observe data $(x_{i1}, x_{i2}, , y_i)$, $i = 1, \ldots, n$, and postulate that it is appropriate to describe the relationship between x_1, x_2, and y by the regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

where β_1 and β_2 are parameters and $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ for all i.

(a) Carefully determine

(i) a 2×1 vector $\boldsymbol{\beta}$,

(ii) $n \times 1$ vectors $\mathbf{y}$, $\boldsymbol{\epsilon}$, and

(iii) an $n \times 2$ matrix $\mathbf{X}$

so that this model can be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

Let $\hat{\beta}_1$, $\hat{\beta}_2$ denote the least squares estimators of β_1, β_2, respectively. It is a fact that the least squares result derived in class applies to this model: if $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2]'$, then $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$.

(b) Determine the distribution of $\hat{\beta}_1$.

(c) Determine the distribution of $\hat{\beta}_2$.

(d) Determine $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

3. (*8 points*) Consider the multiple linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 I)$. Let $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\hat{\boldsymbol{\mu}} = H\mathbf{y}$, $\mathbf{e} = (1-H)\mathbf{y}$ where $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

(a) Carefully verify that the matrix $\hat{\boldsymbol{\mu}}'\mathbf{e}$ equals the zero matrix.

(b) Carefully explain why $\hat{\boldsymbol{\mu}}$ and $\mathbf{e}$ are independent.