STA 6208 – Spring 2003 – Exam 1

Print Name: **SSN:**

All questions are based on the following two regression models, where SIMPLE REGRESSION refers to the case where $p = 1$, and X is of full column rank (no linear dependencies among the predictor variables)

$$\textbf{Model 1:} \quad Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad i = 1, \ldots, n \quad \varepsilon_i \sim NID(0, \sigma^2)$$

$$\textbf{Model 2:} \quad \mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad \mathbf{X} \equiv n \times p' \quad \beta \equiv p' \times 1 \quad \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

1) For **Model 1**, derive the normal equations.

2) A researcher is interested in the relationship between the education level and salaries in rural counties in the U.S. He obtains the percentage of adults over 25 with a college education in each county (X) and the per capita income of the county (Y, in \$1000s). He obtains the following summary statistics, based on a sample of $n = 30$ counties:**(12 points)**

$$\sum(X-\overline{X})^2 = 2207.45 \qquad \sum(X-\overline{X})(Y-\overline{Y}) = 658.37 \qquad \sum(Y-\overline{Y})^2 = 654.86 \qquad \overline{X} = 41.92 \qquad \overline{Y} = 35.83$$

a) Give least squares estimates of the parameters of **Model 1**: β_0, β_1.

b) It can be shown that:

$$\sum(Y - \hat{Y})^2 = \sum(Y - \overline{Y})^2 - \frac{(\sum(X - \overline{X})(Y - \overline{Y}))^2}{\sum(X - \overline{X})^2}$$

Give an unbiased estimate of σ^2.

3) A foam beverage insulator (beer hugger) manufacturer produces their product for firms that want their logo on beer huggers for marketing purposes. The firm's cost analyst wants to estimate their cost function. She interprets β_0 as the fixed cost of a production run, and β_1 as the unit variable cost (or marginal cost). Based on $n = 5$ production runs she observes the following pairs (X_i, Y_i) where X_i is the number of beer huggers produced in the i^{th} production run (in $1000s$), and Y_i was the total cost of the run (in \$1000).

i	X_i	Y_i
1	1.00	3.00
2	2.00	3.50
3	3.00	4.00
4	4.00	4.50
5	5.00	5.00

Obtain the following matrices and vectors: $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{X'X}$, $\mathbf{X'Y}$, $(\mathbf{X'X})^{-1}$.

4) Derive the least squares estimate for β_1 in **Model 1**, when β_0 is constrained to be 0 (regression through the origin).

5) For **Model 2**:

a) Derive the least squares estimate of β.

b) Give the mean vector and variance-covariance matrix for the estimate in a).

6) For **Model 2**:

a) Write $\hat{\mathbf{Y}}$, and $\mathbf{e}$ as linear functions of $\mathbf{Y}$.

b) Show that the matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$ is symmetric and idempotent.

c) Use a) and b) to show that $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$