17.874
Multivariate Statistics
Spring 2004
Problem Set 1

In this course we will study statistical models for estimating the effects of variables on outcomes of interest – that is, the conditional means of jointly distributed random variables. We will generalize the linear regression model introduced in the first semester statistics course. We will also build on the ideas of estimation and inference presented there.

1. Gelman and King (American Journal of Political Science 1990) define the incumbency effect using a simple regression model. Define and describe this effect. Find at least one other research paper that estimates this effect. Describe the research method used and the estimates. (No more than 2 pages)

2. Define and describe an "effect" that is of particular theoretical interest in your field of research. Briefly state why this effect is of interest or importance. Find at least two estimates provided in existing papers. Describe how the estimates were arrived at. Which estimate do you think is better and why? (No more than 2 pages)

3. Consider the regression

$$Y_i = \alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + \epsilon_i$$

The least squares estimates of the intercept and slope coefficients can be arrived at by choosing values of α, β_1, β_2, ... β_k that minimize:

$$S = \sum_{i=1}^{n} \epsilon_i^2$$

   A. Assume that there is only one independent variable X_1. Derive the least squares estimator of α and β_1.

   B. Assume that there are two independent variables, X_1 and X_2. Derive the least squares estimators of α, β_1, and β_2. [Hint: You can minimize algebra in deriving the slope parameters by deviating each variable from its mean, i.e., $Y_i - \bar{Y}$, $X_{j,i} - \bar{X}_j$. Note that the intercept can be derived by considering the mean of Y.]

   C. Assume that there are three independent variables, X_1, X_2, and X_3. Derive the least squares estimator of α, β_1, β_2, and β_3.

   .... Surely there is a better way.