Homework 4

Due 11/2/2014

There is less straight programming this week. Take the time to give a good effort to the data-analysis tasks.

1 Linear Algebra Review

Perform the following vector or matrix multiplication operations:

1.
$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}'$$

2.
$$\begin{bmatrix} -1 & 1 & -1 \end{bmatrix}' \begin{bmatrix} 4 & 3 & 12 \end{bmatrix}$$

3.
$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}'$$

4.
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \\ 1 \end{bmatrix}$$

5.
$$\begin{bmatrix} 3 & 6 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

6.
$$\begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

7.

$$\begin{bmatrix} 3 & 1 & -2 \\ 4 & 6 & 3 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 7 & 2 \\ 1 & 1 \end{bmatrix}$$

8. OPTIONAL

$$\begin{bmatrix} 3 & 1 & -2 & 9 \\ 4 & 6 & 3 & 1 \\ 1 & 4 & 5 & 8 \\ 3 & 8 & 9 & 2 \end{bmatrix} \begin{bmatrix} 9 & 1 & 3 & 12 & 2 \\ 1 & 3 & 10 & 8 & 0 \\ 0 & -2 & -8 & 3 & -2 \\ 3 & 0 & 2 & 9 & 11 \end{bmatrix}$$

9. Show that $\mathbf{IX} = \mathbf{XI}$

10. Let $\tilde{\mathbf{b}}$ be the BLUE estimator of β. Demonstrate that $\mathbf{b} = (X'X)^{-1}X'X$ is this estimator $\tilde{\mathbf{b}}$. Start with the following statement:

$$\tilde{\mathbf{b}} = (\mathbf{M} + \mathbf{A})\mathbf{y}$$

where **A** is some (arbitrary) transformation of the standard model matrix. That is, demonstrate why **A** must be zero. Write one sentence on each step of your proof to show that you *know* what is going on.

2 Fox

Complete the following exercises from Fox: 6.1, 6.2, 6.3, 6.4, 6.9, 6.13, 9.4, 9.6 9.7 and 9.10.

3 Matrix Algebra and Regression

1. For the following very simple dataset, calculate the least squares estimates of the slope and intercept BY HAND using matrix algebra, using the identity $\hat{\beta} = (X'X)^{-1}X'Y$. Calculate the sum of squared residuals for this regression. Find the variance-covariance matrix for this dataset, and finally, perform a hypothesis test for the null hypothesis that $\beta = 0$. There should be a lot of ink spilled here.

   Now, enter your data into R and verify your results. Use only matrix and algebra commands here (for example, to calculate the squared residuals you should use a vector of betas and a matrix of Xs to calculate $\hat{Y}$. Subtract $\hat{Y}$ and Y to form a vector of e and use $e'e$ to sum and square.

X	Y
1	0
1	2
3	3
5	4
5	6

2. Once this works for the small dataset, use your R code to work with a dataset of 100 observations and five variables. Randomly generate five X variables, and then specify a linear model that generates Y as a linear function of the X's and an iid random normal error. Use your R code (matrices) to find estimated coefficients, and then verify the results using the "lm" command.

4 Data Work

Get some data from the *incredible* repository that Don Green and Alan Gerber have set up. This is the link: `http://isps.yale.edu/research#.VFGZO1uvc64`. Download a dataset, figure out what is going on inside the variables. Then, pick some new relationships that hasn't been tested in the paper. Use univariate statistics and plots to describe the variables that you think are interesting, describe a *possible* relationship that exists between variables, and then test to see if that relationship exists. Estimate that relationship, go through the steps of a hypothesis test, and conclude something. Some things to consider:

- Most of this data is purpose built for a particular paper, so there might not be a lot of unused variables. Try and find some.
- It would be really easy to just pick a dataset, look at the regressions the Green and Gerber team has run, and change something small.
- It would also be really easy to just data-mine for some trivial correlation, and then make a theory that fits with it.
- Both of these are things happen in the discipline. Don't let it be you.
- Do something interesting! Take a chance!