Problem Set 1

June 7, 2012

Directions: Answer all questions. Clearly label all answers. Show all of your code. Turn in m-file(s) and writeup to me via your dropbox in Sakai (in a folder labeled 'MatlabPS1.1') by 11:59 p.m. on Thursday, July 5, 2012. Writeup may be comments in m-file(s). Put the names of all group members at the top of your writeup or m-file.

1. Initializing variables and practice with basic matrix operations

   (a) Create the following four matrices of random numbers, *setting the seed to* '1234'[1]. Name the matrices and set the dimensions as noted

      i. $A_{10\times 7}$ - random numbers distributed $U[-5,10]$
      ii. $B_{10\times 7}$ - random numbers distributed $N(-2,15)$ [st dev is 15]
      iii. $C_{5\times 7}$ - the first 5 rows and first 5 columns of A and the last two columns and first 5 rows of B
      iv. $D_{10\times 7}$ - where $D_{i,j} = A_{i,j}$ if $A_{i,j} \leq 0$, or 0 otherwise

   (b) Use a built-in Matlab function to list the number of elements of A

   (c) Use a series of built-in Matlab functions to list the number of *unique* elements of D

   (d) Using the reshape command, create a new matrix called E which is the 'vec' operator[2] applied to B. Can you find an easier way to accomplish this?

   (e) Create a new matrix called F which is 3-dimensional and contains A in the first column of the third dimension and B in the second column of the third dimension

   (f) Use the `permute` function to twist F so that it is now $F_{2\times 10\times 7}$ instead of $F_{10\times 7\times 2}$. Save this new matrix as F.

   (g) Create a matrix G which is equal to $B \otimes C$ (the Kronecker product of B and C). What happens when you try $C \otimes F$?

   (h) Save the matrices A, B, C, D, E, F and G as a .mat file named `matrixpractice`.

   (i) Save only the matrices A, B, C, and D as a .mat file called `firstmatrix`.

   (j) Export C as a .csv file called `Cmatrix`.

   (k) Export D as a tab-delimited .dat file called `Dmatrix`.

[1] To set the seed, type the following code: `rand('seed',1234);` and `randn('seed',1234);`

[2] See `http://en.wikipedia.org/wiki/Vectorization_(mathematics)`

2. Practice with loops

   (a) Write a loop that computes the element-by-element product of A and B. Name the new matrix AB. Create a matrix called $AB2$ that accomplishes this task without a loop.

   (b) Write a loop that creates a column vector called `Cprime` which contains only the elements of C that are between -5 and 5 (inclusive). Create a vector called `Cprime2` which does this calculation without a loop.

   (c) Using loops, create a 3-dimensional matrix called X that is of dimension $N \times K \times T$ where $N = 15,169$, $K = 6$, and $T = 5$. For all t, the columns of X should be (in order):

   - an intercept (i.e. vector of ones)
   - a dummy variable that is 1 with probability $.75 * (6-t)/5$
   - a continuous variable distributed normal with mean $15+t-1$ and standard deviation $5(t-1)$
   - a continuous variable distributed normal with mean $\pi(6-t)/3$ and standard deviation $1/e$
   - a discrete variable distribued "discrete normal" with mean 12 and standard deviation 2.19[3]
   - a discrete variable distribued binomial with $n = 20$ and $p = 0.5$

   i.e., let columns 1, 5 and 6 remain stationary over time.

   (d) Use loops to create a matrix β which is $K \times T$ and whose elements evolve across time in the following fashion:

   - $1, 1.25, 1.5, ...$
   - $\ln(t)$
   - $-\sqrt{t}$
   - $e^t - e^{t+1}$
   - t
   - $t/3$

   (e) Use loops to create a matrix Y which is $N \times T$ defined by $Y_t = X_t\beta_t + \varepsilon_t$, where $\varepsilon_t \overset{iid}{\sim} N(0, \sigma = .36)$

3. Reading in Data and calculating summary statistics

   (a) Clear the workspace and import the file `nlsw88.csv` into Matlab. Make sure you appropriately convert missing values and variable names. Save the result as `nlsw88.mat`.

   (b) What percentage of the sample has never been married? What percentage are college graduates?

[3] A discrete normal random variable is properly called a binomial random variable. The distribution described above can be implemented by choosing binomial parameters n and p where $n = 20$ and $p = 0.6$. Use Matlab's `binornd` command to generate this vector of X.

(c) Use the `tabulate` command to report what percentage of the sample is in each race category

(d) Create a matrix called `summarystats` which lists the mean, median, standard deviation, min, max, number of unique elements, and interquartile range (75th percentile minus 25th percentile) of `grade` (highest grade completed). What percentage of grades observations are missing?

(e) Graphically show the joint distribution of industry and occupation

(f) Tabulate the mean wage over industry and occupation categories

4. Practice with functions

(a) Clear the workspace and load `firstmatrix.mat`.

(b) Write a function called `matrixops` that takes as inputs the matrices A and B from question (a) of problem 1 and has three outputs: *(i)* the element-by-element product of the inputs, *(ii)* the product $A'B$, and *(iii)* the sum of all the elements of $A+B$.

(c) Starting on line 2 of the function, write a comment that explains what `matrixops` does.

(d) In the command window, type `help matrixops`. What comes up?

(e) Evaluate `matrixops.m` using A and B from question (a) of problem 1

(f) Just before the first executable line of `matrixops.m` (i.e. right after the first-line comments), write an if statement which gives an error if the two inputs are not the same size. Have the error say "inputs must have the same size."

(g) Evaluate `matrixops.m` using C and D from question (a) of problem 1. What happens?

(h) Now evaluate `matrixops.m` using `ttl_exp` and `wage` from `nlsw88.mat`.