

### Homework 4 (covering Statistics Lectures 5 and 6)

To complete this assignment, prepare a MATLAB script called `homework4.m` along with any necessary accompanying function `.m` files. Then, run the MATLAB `publish` command (e.g. `publish('homework4.m')`) to run the script and generate HTML output showing the results. Turn in a print-out of the HTML output (e.g. from your web browser) and also a print-out of any function `.m` files that you write.

*Hint:* In your script file, place `%%` on a line by itself at each point where you want the HTML output to show figures and command-window text. Please note that your code should be commented (where necessary), including documentation of any functions that you write.

---

**Problem 0.** Download the `.mat` file at <http://artsci.wustl.edu/~kkay/psych5007/Homework4.mat> (you will need this file to complete the problems below).

**Problem 1.** The `data1` variable contains 1,000 data points, and the `regressors1` variable contains 100 regressors that might explain these data points. (Note: a constant regressor is already included as the 100th regressor.) Consider a linear model that attempts to predict the data as a weighted sum of the regressors. Apply this model to the data, using bootstrapping (500 iterations) to obtain 95% confidence intervals on the parameters of the model. Visualize these confidence intervals on a figure.

**Problem 2.** We have 20 subjects and 100 measurements of  $x$  and  $y$  from each subject. The `xdata` variable contains the  $x$ -measurements, and the `ydata` variable contains the  $y$ -measurements. We want to see whether there is a nonlinear (quadratic) relationship between the  $x$  variable and the  $y$  variable. To do this, for each subject, use 10-fold cross-validation to estimate (1) the accuracy of a linear model ( $y = ax + b$ ) and (2) the accuracy of a quadratic model ( $y = ax^2 + bx + c$ ). Use  $R^2$  to quantify accuracy. Make a scatter plot that compares the accuracy of the linear model and the accuracy of the quadratic model (there should be 20 dots on this scatter plot).