Statistics and Data Analysis in MATLAB
Kendrick Kay, kendrick.kay@wustl.edu

MATLAB In-Class Exercises 3 (covering Statistics Lectures 5 and 6)

Exercise 1. Write a function that computes coefficient of determination (R^2). Use the function declaration below.

```
function f = computeR2(model,data)

% function f = computeR2(model,data)
%
% <model> is a 1 x N vector of model values
% <data> is a 1 x N vector of data points
%
% return the coefficient of determination (R^2)
% between <model> and <data>.  this value summarizes
% how well <model> approximates <data>, and has
% an upper bound of 100%.
```

Exercise 2. Generate some data using `x = rand(1,50); y = x + randn(1,50);`. Split the data points in half. Fit a linear model to one half of the data ($y = ax + b$) and compute the R^2 value for the goodness-of-fit. Then quantify how well this fitted model predicts the second half of the data, again summarizing performance using R^2. How do the two values compare? Is the answer the same for other splits of the data?

Exercise 3. Generate some data using `x = rand(1,1000); y = x + randn(1,1000);`. Let's imagine that these data constitute the complete population from which we are sampling. Draw a small sample ($n = 10$) from the population and fit a line that predicts the y-values from the x-values. Visualize this line on a figure. Repeat this for additional samples, plotting all of the lines on the same figure. How stable is the line across different samples? See what happens when you make the sample size larger ($n = 100$).