University of California, Davis
Department of Biomedical Engineering

Fall 2019	**Probability and Statistics**	**BIM 105**
David M. Rocke	**for Biomedical Engineers**	**November 14, 2019**

Homework Assignment 6
Due November 21, 2019
Always show your work.
Use MATLAB for problem 3.

1. Cardiologists use the short-range scaling exponent α_1, which measures the randomness of heart rate patterns, as a tool to assess risk of heart attack. The article "Applying Fractal Analysis to Short Sets of Heart Rate Variability Data" (M. Peña et al., Med Biol Eng Comput, 2009:709–717) compared values of α_1 computed from long series of measurements (approximately 40,000 heartbeats) with those estimated from the first 300 beats to determine how well the long-term measurement (y) could be predicted from the short-term one (x). The data were imported into Matlab as a table, alpha1, containing two columns, Short and Long, and 54 observations. Below, on the next page, is the Matlab output from the regression analysis, plus some possible additional needed data. Since the sample size is large, you can use standard normal (z) statistics and p-values.

   (a) Compute a 95% confidence interval for the slope.

   (b) Find a 95% confidence interval for the mean long-term measurement for those with short-term measurements of 1.2.

   (c) Test the hypothesis that the mean long-term measurement for those with short-term measurements of 1.2 is also 1.2. Report the p-value.

   (d) Find a 95% prediction interval for the long-term measurement for a particular individual whose short-term measurement is 1.2.

   (e) The purpose of a short-term measurement is to substitute for a long-term measurement. For this purpose, which do you think is more relevant, the confidence interval or the prediction interval? Explain.

```
>> fitlm(alpha1)

Linear regression model:
    Long ~ 1 + Short

Estimated Coefficients:
                   Estimate       SE       tStat       pValue
                   --------    --------    ------    ----------

    (Intercept)    0.78095     0.098693    7.9129    1.7482e-10
    Short          0.39857     0.077569    5.1383    4.2371e-06


Number of observations: 54, Error degrees of freedom: 52
Root Mean Squared Error: 0.169
R-squared: 0.337,  Adjusted R-Squared: 0.324
F-statistic vs. constant model: 26.4, p-value = 4.24e-06

>> mean(alpha1.Short)

    1.2372

>> var(alpha1.Short)

    0.0898

>> mean(alpha1.Long)

    1.2741

>> var(alpha1.Long)

    0.0423

>> cov(alpha1.Short,alpha1.Long)

ans =

    0.0898    0.0358
    0.0358    0.0423
```

2. Consider the following data:

x	1	2	3	4	5
y	3	2	4	6	5

   By hand, not using Matlab, and showing your work:

   (a) Compute the correlation coefficient.

   (b) Find the least-squares line.

   (c) Find the standard deviation around the least-squares line.

3. The article"Optimization of Medium Composition for Lipase Production by Candida rugosa NCIM 3462 Using Response Surface Methodology" (A. Ragendran and V. Thangavelu, Can J. Microbiol, 2007:643–655) describes a series of experiments in which lipase was produced from a bacterial culture. In each experiment, the rate of lipase production (in μmol per ml enzyme per minute) and the cell mass (in g/L) were measured. The results are in the file `HW6-3.csv` which can be downloaded from my website. Using Matlab for computations when you can, do the following:

   (a) Compute the least-squares line for predicting lipase production from cell mass.

   (b) Compute 95% confidence intervals for β_0 and β_1.

   (c) In two experiments, the cell masses differed by 1.5 g/L. By how much do you estimate that their lipase productions will differ?

   (d) Find a 95% confidence interval for the mean lipase production when the cell mass is 5.0 g/L.

   (e) Can you conclude that the mean lipase production when the cell mass is 5.0 g/L is less than 4.4? Explain. (As always, use a two-sided test.)

   (f) Find a 95% prediction interval for lipase production values when the cell mass is 5.0 g/L.