

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

MATLAB In-Class Exercises 1 (covering Statistics Lectures 1 and 2)

Exercise 1. Write a function that calculates the correlation between two input vectors. Use only basic MATLAB functions (e.g. `mean`, `std`, etc.).

Exercise 2. Conduct Monte Carlo simulations in which samples are drawn from a Gaussian distribution with mean 10 and standard deviation 2. For each of various sample sizes ($n = 2, 5, 10, 20, 50, 100, 200, 500, 1000$), draw 1,000 random samples, compute the mean of each sample, and quantify the spread in the results by computing standard deviation. Then make a line plot that shows the relationship between sample size (x -axis) and the uncertainty (spread) of the mean (y -axis). On the same figure, draw a line that shows the theoretical expectation.

Exercise 3. Write a function that accepts an input vector and returns a new vector representing a bootstrap sample drawn from the original vector.

Exercise 4. Modify the function in Exercise 3 to accept another input specifying the number of bootstrap samples to draw. Then have the function return a 2D matrix with different bootstrap samples in the columns of the matrix. Avoid the use of for-loops in this function so that the function is fast.

Exercise 5. Generate some correlated data (two variables). Visualize the data using a scatter plot. Then compute a 2D histogram of these data and visualize the histogram as an image. Compute the marginal distribution of the data (showing the distribution of the data on each variable separately) and then compute the joint distribution that we would expect to observe if the variables were independent. Visualize this distribution as an image.