

## GX5004: HW 1

To the extent that it helps you to learn, you may work with fellow students on this assignment. R and Python have extensive libraries online that can guide you on this assignment.

Total Number of Points: 20 points

1. Go to <http://www.random.org/integers/> and generate two series of 10 random integers with values between 0 and 9. Call them Y and X. Using R or Python: [7 points]
  - a. Generate a scatter plot of (Y, X).
  - b. Calculate the means of Y and X.
  - c. Calculate the variances of Y and X.
  - d. Calculate the standard deviations of Y and X.
  - e. Calculate the covariance of Y and X.
  - f. Calculate the correlation of Y and X.
  - g. Given the results above, suppose I ask you to predict the value of Y if I give you a value of  $X = 13$ . How would you respond?
  - h. Submit code and results.
2. Recall we discussed the role that random variables play in applied data science. We also discussed the distinction between discrete and continuous random variables. For the random variables below, indicate the more appropriate random variable (discrete or continuous) and why you believe this to be the case. [7 Points]
  - a. The number of taxi rides taken in a month by a NYC resident.
  - b. The speed of a bicyclist on Jay Street.
  - c. The luminosity of light emitted by street lamps in Brooklyn.
  - d. The income of bankers working on Wall Street.
  - e. The number of hotels in Manhattan.
  - f. The ambient sound generated by trash trucks picking up trash at midnight.
  - g. The quality of coffee served in the student lounge.
3. Consider the salaries of bankers on Wall Street. One argument that could be advanced is that Wall Street bankers have high salaries because they have attained high levels of education. Another argument could be advanced that Wall Street banker have high salaries because they have high Intelligence Quotients (IQ). If you had a dataset that provided you with the salaries of a sample of Wall Street bankers, together with their education levels and IQs, discuss how you might explore these arguments. Are there other methods beyond relying on this sample of bankers that could allow you to explore these arguments? (As with most applied data analytics, there is no right or wrong answer to this question. You may find it helpful to consider what we discussed in class to address it.) [6 points]