Tutorial 7

1. Consider three data sets.

   (1) 19, 24, 12, 19, 18, 24, 8, 5, 9, 20, 13, 11, 1, 12, 11, 10, 22, 21, 7, 16, 15, 15, 26, 16, 1, 13, 21, 21, 20, 19

   (2) 17, 24, 21, 22, 26, 22, 19, 21, 23, 11, 19, 14, 23, 25, 26, 15, 17, 26, 21, 18, 19, 21, 24, 18, 16, 20, 21, 20, 23, 33

   (3) 56, 52, 13, 34, 33, 18, 44, 41, 48, 75, 24, 19, 35, 27, 46, 62, 71, 24, 66, 94, 40, 18, 15, 39, 53, 23, 41, 78, 15, 35

   (a) For each data set, draw a histogram using python and determine whether the distribution is right skewed, left-skewed, or symmetric.

   (b) Compute sample means and sample medians by hand and python. Do they support your findings about skewness and symmetry? How?

2. Solve this question using python. The following data set represents the number of new computer accounts registered during ten consecutive days.

   43, 37, 50, 51, 58, 105, 52, 45, 45, 10.

   (a) Compute the mean, median, quartiles, and standard deviation.

   (b) Check for outliers using the 1.5(IQR) rule.

   (c) Delete the detected outliers and compute the mean, median, quartiles, and standard deviation again.

   (d) Make a conclusion about the effect of outliers on basic descriptive statistics.

3. A program consists of two modules. The number of errors, X, in the first module and the number of errors, Y, in the second module have the joint distribution, P(0, 0) = P(0, 1) = P(1, 0) = 0.2, P(1, 1) = P(1, 2) = P(1, 3) = 0.1, P(0, 2) = P(0, 3) = 0.05. Find
   (a) the marginal distributions of X and Y
   (b) the probability of no errors in the first module
   (c) the distribution of the total number of errors in the program.
   (d) find out if errors in the two modules occur independently.

4. Define X as the height in meters of a randomly selected Canadian, where the selection probability is equal for each Canadian, and denote E[X] by h. Bo is interested in estimating h. Because he is sure that no Canadian is taller than 3 meters, Bo decides to use 1.5 meters as a conservative (large) value for the standard deviation of X. To estimate h, Bo averages the heights of n Canadians that he selects at random; he denotes this quantity by H.
   a) In terms of h and Bo's 1.5 meter bound for the standard deviation of X, determine the expected value and standard deviation for H
   b) Help Bo by calculating a minimum value of n (with $n > 0$) such that the standard deviation of Bo's estimator, H, will be less than 0.01 meters.
   c) Bo would like to be 99% sure that his estimate is within 5 centimetres of the true average height of Canadians. Using the Chebyshev inequality, calculate the minimum value of n that will make Bo happy.
   d) If we agree that no Canadians are taller than three meters, why is it correct to use 1.5 meters as an upper bound on the standard deviation for X, the height of any Canadian selected at random?