Statistics and Data Analysis, Fall 2016
Pre-lecture Problems 4

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

Note 1. The deadline of submitting the pre-lecture problem is **14:30, October 12, 2016**. Please submit a hard copy of your work to the instructor in class. Late submissions will not be accepted. Each student must submit her/his individual work. Submit ONLY the problem that counts for grades.

Note 2. Please make your answer as clear (i.e., easy to read) as possible. We reserve the right to take away points when the correctness cannot be easily determined (e.g., when the writing is messy and cannot be easily understood).

1. (0 points) Calculate the mean, median, MAD, variance, standard deviation, and coefficient of variation for the following values of a population:

    $$12, 18, 14, 15, 15, 17, 23, 24, 12, 18.$$

    Note. Though there are MS Excel functions that help people find these numbers, I encourage you to go through all the calculation steps to get a deeper idea about these measurements.

2. (0 points) Let x_1, x_2, ..., and x_{10} be the ten values in Problem 1 and y_1, y_2, ..., and y_{10} be the following ten values of a population

    $$13, 15, 16, 22, 27, 18, 10, 14, 23, 22.$$

    Calculate the correlation coefficient for the two sets of data x and y.

    Note. Though there are MS Excel functions that help people find these numbers, I encourage you to go through all the calculation steps to get a deeper idea about this measurement.

3. (10 points) Consider the two sets of data x and y in the previous two problems.

    (a) (4 points) According to their medians and means, would you conclude that values in one set tends to be larger than those in the other set? Why or why not?

    (b) (4 points) According to their MADs and variances, would you conclude that values in one set is more disperse than those in the other set? Why or why not?

    (c) (2 points) According to their correlation coefficient, please say anything about the correlation between the two sets of data.