Week 4 Homework: Correlation, due Thursday, February 23

Instructions: For all testing hypotheses please make sure that you write down precise statements of your null and alternative hypotheses, decision from the test and conclusion (answer to the question in the problem). You must also write the value of the test statistic, and p-value for every test. Use significance level $\alpha = 0.05$ for all tests. THE ENTIRE HOMEWORK CAN TAKE AT MOST 2 PAGES, 12 pt. Each part of every problem is worth 3 points, for the total of 21 points.

1. **Influence of outliers on the correlations measures.** Consider the data set called cor_lab_data.CSV.

   (a) Compute all three correlation coefficients between x and y.

   (b) One observation in this data is relatively far from the majority of the data. Remove that observation and compute all three correlation coefficients for the new data set.

   (c) How did the values of the correlations change? Which correlation coefficients are more robust, that is not influenced much by outliers?

2. An admissions officer at a small college took a sample of 20 freshmen GPAs at the and of the first year and their ACT scores to evaluate if he can predict the GPA from ACT. The data are in the file cor_hwk_data.CSV on Canvas.

   (a) Use an appropriate graph to make initial assessment of the association between GPA and ACT scores.

   (b) Test if the Pearson corr. coefficient between ACT and GPA is significantly different from zero.

   (c) Test if the Spearman or Kendall (select one and use) corr. coefficient between ACT and GPA is significantly different from zero.

   (d) What is your conclusion? Can the admission officer reasonably predict the freshman GPA from the ACT scores?