**Hsci 2117: Homework Assignment #1 – Describing and Summarizing Data**

This problem set utilizes the Season Effect data set used in a study by Turan et al. (2015). It has been cleaned and saved in the csv format. Be sure to download the file accompanying the assignment from blackboard as well as the dataset introduction, study paper, and data dictionary.

Show all work and calculations for all questions on this problem set. Include in your final word document a description of the process and method you used to solve each question. Provide sufficient explanation of how you got the answer to any question. Include relevant screenshots from StatKey as needed.

1. Read the dataset introduction and answer the following questions:
   1. What is the population of interest the researchers wanted to make generalizations about? (5%)
   2. How was a sample from the study population selected? Was random sampling utilized? (5%)

(Hint: Read section 1.2 of the textbook)

* 1. What is the main research question? (5%)
  2. What is the primary outcome variable? Explain how it was measured. (5%)
  3. The introduction notes that patients were “Patients were divided into season surgical procedures depending on their date of surgery”. Is this random allocation? Why, or, why not? (5%)
  4. Explain why randomization is important to the study design. (5%)

(Hint: Read section 1.3 of the textbook)

1. Read the data dictionary and answer the following questions:
   1. List all categorical variables in the dataset (10%)
   2. List all ordinal variables in the dataset (10%)
   3. List all quantitative (discrete or continuous) variables in the dataset (10%)
2. Create a *95% confidence interval estimate for the true proportion* of SSI in all patients. Hint: your answer should include the range of all likely values for the true population proportion. Use the 2\*SE method. Review the StatKey resources from Week 1.
   1. Report the 95% Confidence Interval: \_\_\_\_ to \_\_\_\_. Include a screenshot of StatKey and an explanation of how you calculated the interval. (10%)
   2. How often do you think patients develop an SSI? Explain your answer using everyday English. Support your explanation with statistical evidence. Include some notion of margin of error. (10%)
3. A doctor that works in the hospital thinks that based on her prior experience, colorectal surgery duration is about 4 hours. Conduct a *randomization test for a mean* to assess whether the doctor’s hypothesis is plausible. Hint: Review the StatKey resources from Week 2.
   1. Report the *p*-value. Interpret the *p*-value in terms of rejecting the null hypothesis or failing to reject the null hypothesis. Paste a screenshot of the randomization test from StatKey. (10%)
   2. Is the doctor’s guess about the average surgery duration plausible? Explain why or why not using everyday English. Support and justify your position with statistical evidence. (10%)