Project 1

March 1, 2017

You should use the statistical software 'R' to answer the following questions. Failure to use the software will result in a grade of 0 for project 1.
Complete the following problems below. Within each part, include your R program output with code inside of it and any additional information needed to explain your answer. You may need to edit your output and code in order to make it look nice after you copy and paste it into your Word document.

Problem 1

We will use for this problem 1 a dataset collected from the 2014 World Cup Soccer in Brazil. Using your favorite search engine (yahoo, google, bing, ...), get the frequency of various number of goals scored by a team during the first round of play. (Similar to our discussion from class)

1. Find the mean number (or average number) of goals scored during the first round of the 2014 World Cup.
2. Find the Poisson probabilities at each level.
3. Compute the expected frequencies
4. Perform a goodness-of-fit test to check if your Poisson model is appropriate for this data.

Problem 2

In a study of heart disease among males, the 356 subjects were classified according to socioeconomic status and smoking habits. The study recognized

three levels of socioeconomic status (high, middle and low) and three smoking categories (current smoker, never smoked, former smoker). The data are summarized in the following contingency table.

	high	middle	low
current	51	22	43
former	92	21	28
never	68	9	22

(a) Using two different methods, at the significance level of 5%, do these data show that smoking habits and socioeconomic status are dependent or independent?

(b) Use residuals to describe the evidence of association.

(c) Use the partitioning of the chi-square to interpret the association.

(d) Summarize association by constructing a 95% confidence interval for the odds ratio between current or never and whether high or low. Interpret.