Homework 5

Statistical Inference

1- The General Social Survey asked a random sample of 1,390 Americans the following question: "On the whole, do you think it should or should not be the government's responsibility to promote equality between men and women?" 82% of the respondents said it "should be". At a 95% confidence level, this sample has 2% margin of error. Based on this information, determine if the following statements are true or false, and explain your reasoning.

   a. We are 95% confident that between 80% and 84% of Americans in this sample think it's the government's responsibility to promote equality between men and women.

   b. We are 95% confident that between 80% and 84% of all Americans think it's the government's responsibility to promote equality between men and women.

   c. If we considered many random samples of 1,390 Americans, and we calculated 95% confidence intervals for each, 95% of these intervals would include the true population proportion of Americans who think it's the government's responsibility to promote equality between men and women.

   d. In order to decrease the margin of error to 1%, we would need to quadruple (multiply by 4) the sample size.

   e. Based on this confidence interval, there is sufficient evidence to conclude that a majority of Americans think it's the government's responsibility to promote equality between men and women.

2- A newspaper collects a simple random sample of 450 likely voters in the district and estimates Toohey's support to be 53%. Does this provide convincing evidence for the claim of Toohey's manager at the 5% significance level?

3- In July 2008, the US National Institutes of Health announced that it was stopping a clinical study early because of unexpected results. The study population consisted of HIV-infected women in sub-Saharan Africa who had been given a single dose of Nevirapine (a treatment for HIV) while giving birth to prevent transmission of HIV to the infant. The study was a randomized comparison of continued treatment of a woman (after successful childbirth) with Nevirapine vs. Lopinavir, a second drug used to treat HIV. 240 women participated in the study; 120 were randomized to each of the two treatments. Twenty-four weeks after starting the study treatment, each woman was tested to determine if the HIV infection was becoming worse (an outcome called virologic failure). 26 of the 120 women treated with Nevirapine experienced virologic failure, while 10 of the 120 women treated with the other drug experienced virologic failure.

   a. Create a two-way table presenting the results of this study.

   b. State appropriate hypotheses to test for difference in virologic failure rates between treatment groups.

c. Complete the hypothesis test and state an appropriate conclusion. (Reminder: Verify any necessary conditions for the test.)

4- Two different question phrasings are being used to compare the 2010 healthcare law approval. Depending on how a question is phrased, a person may respond differently. For instance, Pew Research Center conducted a survey with the following question:

As you may know, by 2014, nearly all Americans will be required to have health insurance. [People who do not buy insurance will pay a penalty] while [People who cannot afford it will receive financial help from the government]. Do you approve or disapprove of this policy?

For each randomly sampled respondent, the statements in brackets were randomized: either they were kept in the order given above, or the two statements were reversed. The following table shows the results of this experiment. Create and interpret a 90% confidence interval of the difference in approval.

	Sample size (ni)	**Approve law (%)**	**Disapprove law (%)**	**Other**
"People who cannot afford it will receive financial help from the government" is given second	771	47	49	4
"People who do not buy it will pay a penalty" is given second	732	34	63	3

Table 4.1 Results for a Pew Research Center poll where the ordering of two statements in a question regarding healthcare was randomized.

5- Rock-paper-scissors is a hand game played by two or more people where players choose to sign either rock, paper, or scissors with their hands. For your statistics class project, you want to evaluate whether players choose between these three options randomly, or if certain options are favored above others. You ask two friends to play rock-paper-scissors and count the times each option is played. The following table summarizes the data.

Rock	**Scissors**	**Paper**
43	21	35

Homework 5

Statistical Inference

Use these data to evaluate whether players choose between these three options randomly, or if certain options are favored above others. Make sure to clearly outline each step of your analysis, and interpret your result in the context of the data and the research question.

6- A Restaurant Manager for a Football Park wants to know if there is a relationship between gender (male or female) and the preferred condiment on a sandwich. The following table summarizes the results. Test the hypothesis with a significance level of 5%.

	Condiment			
	Ketchup	**Mustard**	**Relish**	**Total**
Male	15	23	10	48
Female	25	19	8	52
Total	40	42	18	100

7- **(R)** During a five-day workweek, employers are interested in knowing which days employees are absent. It is usually believed by employers that employees are absent equally throughout the week. In this scenario, a random sample of 70 managers is asked what day of the week they had the highest number of employee absences. The results were distributed as in Table 7.1. For the population of employees, do the days for the highest number of absences occur with equal frequencies during a five-day workweek? Test at a 5% significance level. (Using a chi-square test)

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Number of Absences	18	16	10	10	16

Table 7.1 Day of the Week Employees were Most Absent

8- **(R)** Consider the dataset named "survey" in the "MASS" library in R. The Smoke column records the students' smoking habit, while the "Exer" column records their exercise level. Test the hypothesis of whether the students' smoking habit is independent of their exercise level at the 0.05 significance level.