ST 518: Homework 2

Please create an .Rmd file, then knit it to create a .pdf document containing your answers to the questions. Include your code in the R question.

Please feel free to discuss questions on the discussion board.

R Questions

1. (4 points) It is often argued that victims of violence exhibit more violent behavior toward others. To study this hypothesis, a researcher searched court records to find 908 individuals who had been victims of abuse as children. She then found 667 individuals, with similar demographic characteristics, who had not been abused as children. Based on a search through subsequent years of court records, she was able to determine how many in each of these groups became involved in violent crimes, as shown in the following table.

	Involved in violent crime?	
	Yes	No
Abuse Victim	102	806
Control	53	614

The researcher concluded: "Early childhood victimization has demonstrable long-term consequences for violent criminal behavior."

Conduct your own analysis of the data and comment on this conclusion. Is there evidence of a difference between the two groups? Is the strength of the causal implication of this statement justified by the data from this study?

Conceptual Questions

2. (1 point) During an investigation of the U.S. space shuttle *Challenger* disaster, it was learned that project managers had judged the probability of mission failure to be 0.00001, whereas engineers working on the project had estimated failure probability at 0.005. The difference between these two probabilities, 0.00499, was discounted as being too small to worry about. Is a different picture provided by considering odds? How is that interpreted?

3. (2 points) Suppose that 90% of orange tabby cats are male. Determine if the following statements are true or false, and explain your reasoning.
   (a) The distribution of sample proportions of random samples of size 30 is left skewed.
   (b) Using a sample size that is 4 times as large will reduce the standard error of the sample proportion by one-half.
   (c) The distribution of sample proportions of random samples of size 140 is approximately normal.
   (d) The distribution of sample proportions of random samples of size 280 is approximately normal.

4. (2 points) A 2010 survey asked 827 randomly sampled registered voters in California, "Do you support or do you oppose drilling for oil and natural gas off the coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college.

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

   (a) What percents of college graduates and non-college graduates in this sample do not know enough to have an opinion on drilling for oil and natural gas off the coast of California (i.e., report two percent values)?

   (b) Conduct a hypothesis test to determine whether there is evidence that the proportion of college graduates who do not have an opinion on this issue is different from that of non-college graduates.

5. (1 point) A study of British male physicians noted that the proportion who died from lung cancer was 0.0140 per year for cigarette smokers and 0.00010 per year for nonsmokers. Additionally, the proportion who died from heart disease was 0.00669 for smokers and 0.00413 for nonsmokers. Which response (lung cancer or heart disease) is more strongly related to cigarette smoking, in terms of the reduction in deaths that could occur with the absence of smoking?