Math 222 - Project 4 **Due Friday, February 23**

1. A 2001 study looked at whether students took longer to react while talking on a cell phone versus listening to music. In the study, 16 student volunteers played a game that simulated driving a car. Each volunteer drove two times, once with music and the other time while talking on the phone. The students were randomly assigned to whether they used the cell phone or listened to music first. In each situation, the reaction times (in milliseconds) were recorded to see how long it would take students to react to a simulated red light change.

   The file driving.csv contains the data for each student.

   (a) Was this a randomized controlled experiment? What are the explanatory and response variables? Which variables, if any, have been controlled?

   (b) Let X represent the number of students who take longer to respond when talking on a phone. What probability distribution can we use to model X? What are appropriate null and alternative hypotheses for this situation?

   (c) What is the actual value of the statistic X? Is this value statistically significant? Explain in detail how you came to your conclusion and what it means.

   (d) In part (a) and (b), we focused on the statistic X, and the test we carried out was an example of a **sign test**, where you focus on whether a difference is positive or negative, and not on the size of the difference. We could also look at how large the differences are between the reaction times of students when they were using a phone versus when they were just listening to music. Analyze the data and make a t-distribution confidence interval for this difference. Report on your findings. In particular, you should comment on how reliable you think this confidence interval will be.

2. In class, we looked at data from the NHTSA about whether seat belts help prevent fatal injuries in car accidents. Here is data on occupants of vehicles that were involved in fatal car accidents in 2016. The information includes whether the occupant was restrained (wore a seat belt) or not and whether the occupant survived or not (Source: NHTSA CrashStats)

	Restraint Used	Not Restrained
Occupant survived	31,533	5,154
Occupant died	11,282	10,428

   (a) Are people who are wearing seat belts significantly more likely to survive fatal accidents? Carry out an appropriate hypothesis test and explain what the results mean.

   (b) What is the relative risk of dying if you are involved in a fatal accident and you aren't wearing a seatbelt vs. if you are?

   (c) Are there any confounding variables that might be associated with both survival and seatbelt usage? Choose two examples, and describe the effect they might have.

   (d) Make a bootstrap distribution for the relative risk. In order to compute the bootstrap statistic, you will need to choose two bootstrap samples with replacement during each iteration of the for-loop. One bootstrap sample will be for the people who are wearing seatbelts and the other will be for the people who are not. You can use the `rbinom()` command to simulate these re-samples.

   (e) Is the bootstrap distribution for relative risk approximately normal?

   (f) Use the formula

   $$\text{statistic} \pm t^* \text{SE}_{\text{bootstrap}}$$

   to make a 95% confidence interval for the relative risk.

   (g) Make a second bootstrap confidence interval by finding the locations of the 2.5th and 97.5th percentiles in the bootstrap distribution. This is known as the **quantile method**, and you can use the `quantile()` function in R to find the two quantiles. Compare the two methods for making confidence intervals. Are the results similar?