STAT 446: Midterm Exam
Due: Friday October 11 at 11:59 PM
Name:

Please turn in the exam to D2L and include the R Markdown code *and either* a Word or PDF file with output. While the exam is open book, meaning you are free to use any resources from class, this is strictly an individual endeavor and **you should not discuss the problems with anyone outside the course instructor including class members.** The instructor will answer questions related to the data, expectations, and understanding of the exam, but will not fix or troubleshoot broken code.

1. Subaru Outback Analysis

Suppose you are in the market for a used Subaru Outback, but want to collect some data before making a purchase. You are able to scrape data from craiglist postings for used Outbacks for sale. You are interested in a newer, more reliable car and consider 2011 - 2014 models.

a. (10 points)

Using the data below, construct a confidence interval for the mean price of a 2011 - 2014 Subaru Outback posted in Bozeman. Note, you do not need to consider a finite population correction in this case.

```
cars_MT <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/BZN_Outback.csv')
cars_MT
```

```
## # A tibble: 8 x 5
##   price  year condition odometer paint_color
##   <int> <int> <chr>        <int> <chr>
## 1  8975  2011 excellent   191723 red
## 2 13500  2011 excellent   128000 grey
## 3 10998  2012 excellent   140579 black
## 4 12000  2013 excellent       NA <NA>
## 5 13975  2013 excellent       NA blue
## 6 11995  2014 excellent   149282 blue
## 7 14925  2014 excellent    93117 silver
## 8 15875  2014 excellent    85000 grey
```

b. (10 points)

Suppose you want to know the price of the 2011 - 2014 Subaru Outback to a maximum allowable difference of $200. Consider the `cars_MT` dataset to be a pilot estimate to estimate the sample standard deviation, s. How many vehicles would need to be collected (You don't need to use the fpc here). Include your calculation and a description of the results.

2. (40 points - MT Rainier Climb)

For this analysis we will use data collected on attempted summits of Mt. Rainier. Some additional details are available at https://www.kaggle.com/codersree/mount-rainier-weather-and-climbing-data. Using a subset of this data we will attempt to estimate the proportion of groups that successfully reached the summit of Mt. Rainier using the DC (Disappointment Cleaver) route.

```
rainier <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/climbing_statistics.csv')
rainier <- rainier %>% filter(Route == 'Disappointment Cleaver') %>% mutate(summit = as.numeric(Succeed
```

a (10 points)

There are 2728 climbing groups that attempted to summit Mt. Rainier during 2014 - 2015. You have the ability to randomly sample each group and determine whether (someone in) the group was successful at summiting Mt. Rainier. Suppose you want to estimate the mean population proportion of successful summits, within a maximum allowable difference of .1. How many samples will you need to collect? Describe the process and your rationale, make sure to include any calculations.

b (10 points)

Take a single sample from the Rainier dataset with the size you determined in question 2a. Compute a confidence interval for the point estimator of the population mean. How does the margin of error of your point estimate compare to the maximum allowable difference in part a? Why might they be the same or different?

c (5 points)

Write a short paragraph describing your results in part b. Assume the target audience is a park ranger that has not taken a statistics course.

d (5 points)

The park ranger believes that weather is a major factor in the success or failure of a climbing group. Given the temperature on a particular day, how could this information be used to estimate the population proportion of success on *warm days* or *cold days*? Describe how the sampling procedure and estimators would need to change to calculate these two proportions. You do not need do sample size calculations for this question.

e (10 points)

Now, use the repeated sampling approach to construct a sampling distribution of sample means using the sample size you estimated in part a. That is, draw ~1000 samples of the size from 2a. Create a figure with the sampling distribution, true population mean, and the sample mean from the sample you collected. Describe the figure.