STATISTICS 216 Lock
Exam 1 Review – Fall 2014

Name: _____

1. In June 2003 *Consumer Reports* published an article on 30 sport-utility vehicles they had tested recently. They reported some basic information about each of the vehicles and the results of some tests conducted by their staff. Among other things, the article told the brand of each vehicle, its price, and whether it had a standard or automatic transmission. They reported the vehicle's fuel economy, its acceleration (number of seconds to go from zero to 60 mph), and its braking distance to stop from 60 mph. The article also rated each vehicle's reliability as much better than average, better than average, average, worse than average, or much worse than average.

   (a) Identify the cases in the data set.

   (b) List the variables. Indicate whether each variable is categorical or quantitative. If the variable is quantitative, give the units.

   (c) What would the dimensions of the data set be? (number of rows by number of columns)

2. One of the reasons that Monitoring the Future (MTF) project was started was "to study changes in the beliefs, attitudes, and behavior of young people in the United States." Data are collected from 8th, 10th, and 12th graders each year. To get a representative nationwide sample, surveys were given to a randomly selected group of students. In Spring 2004, students were asked about alcohol, illegal drug, and cigarette use.

   (a) Identify the cases in the data set.

   (b) List the variables. Indicate whether each variable is categorical or quantitative. If the variable is quantitative, give the units.

   (c) What would the dimensions of the data set be? (number of rows by number of columns)

3. Determine whether each of the following situations describes a sample or population. Identify the variable collected, whether it is categorical or quantitative, then give the relevant statistic or parameter using correct notation. If it is a sample, is it unbiased? Explain.

   (a) A researcher has identified a beach with a substantial number of driftwood logs. She randomly chooses 30 logs and takes core samples from those logs. Three of the 30 logs have core samples that are less than 10% water.

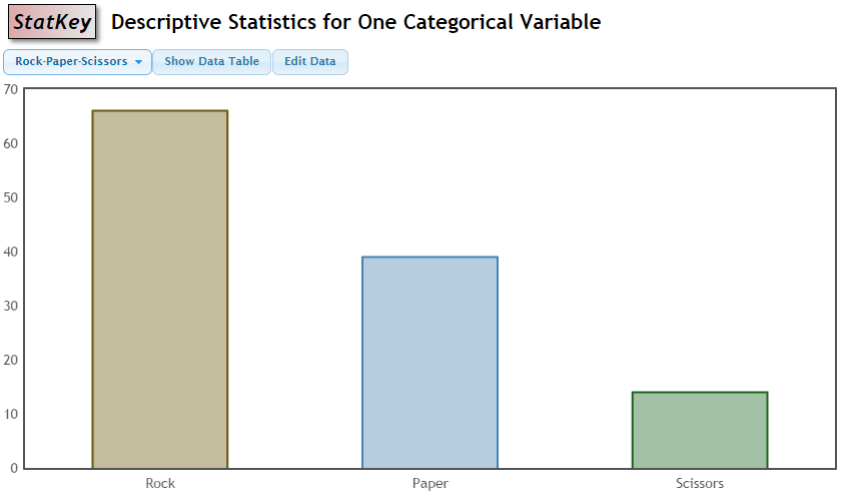
   (b) A professor wants to schedule a review session for an exam. He asks all students enrolled in the course their preferred time, and they all respond, finding that 82% can arrive at 6pm.

4. You are interested in estimating the proportion of MSU students who would support removing all athletic programs from the university.

   (a) Explain how you could select a simple random sample to obtain this estimate.

   (b) Explain how you could select a stratified random sample to obtain this estimate. Be sure to explain how you would stratify the population.

5. A high school student played 119 games of Rock-Paper-Scissors and recorded his opponents choice each time. He displayed the results in a frequency bar chart using StatKey (see below).



(a) What is the approximate relative frequency of people choosing paper?

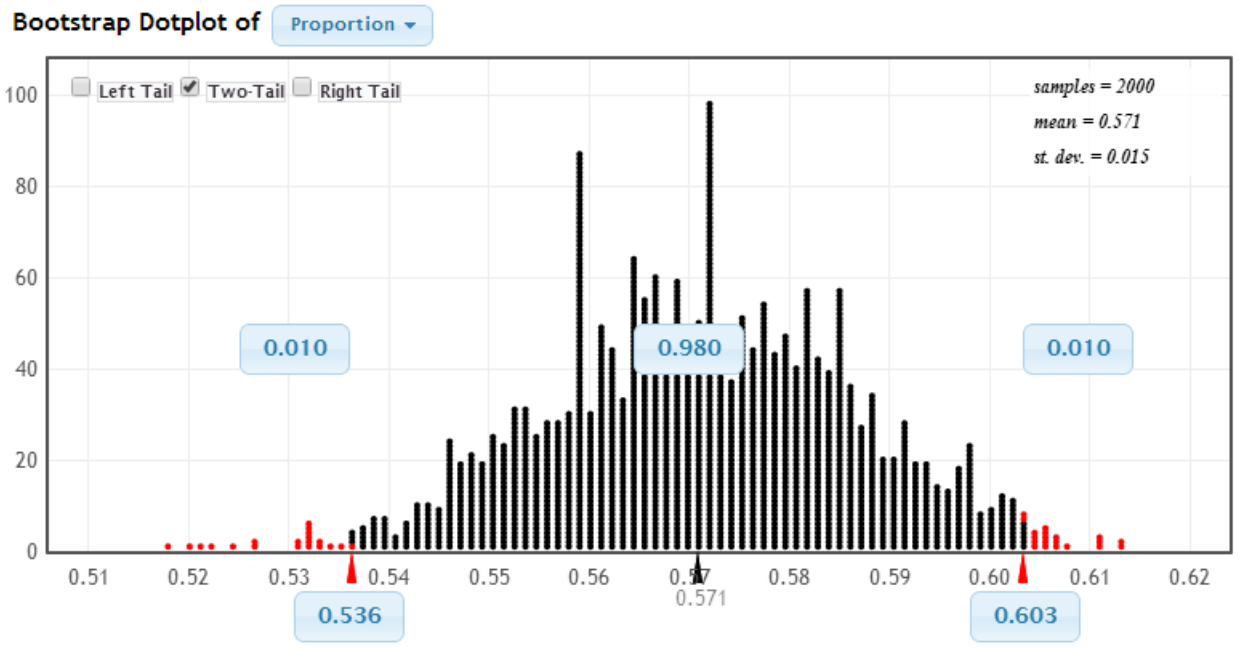
(b) How many people did not choose rock?

(c) Pretend this is representative of what an average person would throw when playing Rock-Paper-Scissors. What should you throw in order to win most often? Explain your choice.

6. A recent national telephone survey published in February 2012 reports that 57% of those surveyed think violent movies lead to more violence in society. The survey included a random sample of 1000 American adults.

(a) Identify the population of interest.

(b) Define the parameter of interest using words and notation.



(c) How is each dot in the plot above generated and what does it represent?

(d) What is the standard error of the estimate?

(e) What is the margin of error for a 95% confidence interval using the information provided?

(f) Calculate and interpret a 95% confidence interval.

(g) Explain why the 98% confidence interval provided on the plot is wider than the 95% confidence interval you created.

(h) Is the proportion of American adults who think that violent movies lead to more violence in society different from 50%?

i. Write the null and alternative hypothesis for this question.

ii. Using the 95% confidence interval you created above, what is your decision? What significance level was used?

iii. Using the 98% confidence interval provided, what is your decision? What significance level was used?

7. The Census Bureau reports that 35.4% of the US population received welfare benefits in 2012. A random sample of 1726 US citizens that same year showed that 41.3% received welfare. A political organization who does not believe the Census Bureau wants to use the sample data to prove the actual percentage of the US population receiving welfare in 2012 is above 38%.

(a) Where would the randomization (null) distribution be centered?

(b) Where would the sampling distribution be centered?

(c) Where would the resampling (bootstrap) distribution be centered?

8. Half-way through President Obama's first term in 2010, Gallop conducted a nationwide poll assessing the proportion of American's who supported the president. In a random sample of 1057 Americans, 583 said they supported President Obama. Does this provide evidence a majority of Americans supported the president in 2010?

(a) Define the parameter of interest in words and notation.

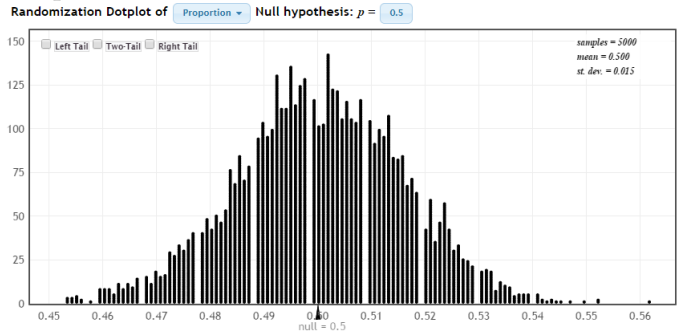
(b) Give the relevant statistic using proper notation.

(c) Write the null and alternative hypotheses in words and notation.
H_0 :

H_A :

(d) Is this a right tail, left tail, or two tail test?

(e) Below is the randomization distribution created in StatKey using 5000 randomization samples.



i. Explain how one dot on this plot is created.

ii. Explain what one dot on this plot represents.

iii. Where is the plot centered? Why does this make sense?

(f) There are 2 dots at least as extreme as the observed result in the right tail and 1 dot at least as extreme as the observed result in the left tail. What is the p-value of the test?

(g) State your conclusion using an $\alpha = 0.05$ significance level in the context of the problem.

(h) Could your conclusion be a Type I error or a Type II error? Explain.

(i) If Gallop polled only 500 Americans instead of over 1000 and reported the same sample proportion of people supporting President Obama, would the p-value of the test increase, decrease, or stay the same. Explain your choice.

9. During the business downturn, Bernie Madoff's investment fund lost value in only 13 of 87 months. During the same time, the S&P500 stock market index dropped in value 28 of the 87 months. A good financial manager can hope to do a little better than a stock market index like the S&P500, but not by much. A simulation of 1000 trials of the number of losing months out of 87 months using the S&P500 numbers gives this table:

months	10	14	16	17	18	19	20	21	22	23
count	1	1	1	5	3	13	24	22	32	45

months	24	25	26	27	28	29	30	31	32	33
count	64	88	83	93	82	61	91	77	54	45

months	34	35	36	37	38	39	40	41	42
count	34	32	20	12	6	5	3	2	1

(a) What is the null hypothesis in notation?
H_0 :

(b) Based on his record (13 of 87) and the plot, quantify the strength of evidence against the claim that the value of Bernie's fund was similar to the stock market's value. Explain how you computed it.

(c) The US House Financial Services Committee wonders if Bernie should be investigated. Based on the above evidence, what would you recommend? Explain your reasoning. (6 pts)