RULES:
1. This assignment should be conducted in group.
2. Each group should only consist of TWO members.
3. Assignment should be done in complete (answer all the questions given) and neat.
4. You may use any statistical packages/tools as you prefer to get the result.
5. Submission should follow the date given and should be in hardcopy.

TASK 1: DESCRIPTIVE STATISTICS

Question 1: (allowed for 1 group)

The Woodmill Company makes windows and door trim products. The first step in the process is to rip dimensions (2 x 8, 2 x 10, etc) lumber into narrower pieces. Currently, the company uses the manual process in which an experienced operator quickly looks at a board and determines what rip widths to use. The decision is based on the knots and defects in the wood.

A company in Oregon has developed an optical scanner that can be used to determine the rip widths. The scanner is programmed to recognize defects and to determine the rip widths that will optimize the value of the board. A test run of 100 boards was put through the scanner and the rip widths are identified. However, the boards were not actually ripped. A lumber grader determined the resulting values for each of the 100 boards assuming that the rips determined by the scanner had been made. Next, the same 100 boards were manually ripped using the normal process. The grader then determined the value for each board after the manual rip process was completed. The resulting data, in the file **WOODMILL**, consist of manual rip values and scanner rip values for each of the 100 boards.

a. Develop a frequency distribution for the board values for the scanner and the manual process.
b. Compute appropriate descriptive statistics for both manual and scanner values. Use these data along with the frequency distribution developed in part (a) to prepare a written report that describes the results of the test. Be sure to include in your report a conclusion regarding whether the scanner outperforms the manual process.
c. Which process (scanner or manual) generated the most values that were more than 2 standard deviations from the mean?
d. Which of the two processes has the least relative variability?

Question 2: (allowed for 1 group)

The commercial banking industry is undergoing rapid changes due to advances in technology and competitive pressures in the financial services sector. The data file **BANKS** contains selected information tabulated by Fortune concerning the revenues, profitability, and number of employees for the 51 largest US Commercial Banks in terms of revenues. Use the information in this file to complete the following:

a. Compute the mean, median and standard deviation for the three variables: revenues, profits and number of employees.
b. Convert the data for each variable to z value. Consider Mellon Bank Corporation headquarters in Pittsburgh. How does it compare to the average bank in the study on the three variables? Discuss.

c. As you can see by examining the data and by looking at statistics computed in part (a), not all banks had the same revenue, same profit or the same number of employees. Which variable had the greatest relative variation among the banks in the study?
d. Calculate a new variable: profits per employee. Develop a frequency distribution and histogram for this new variable. Also compute the mean, median and standard deviation for the new variable. Write a short report that describes the profits per employee for the banks.
e. Referring to part (d), how many banks had a profit-per-employee ratio that exceeded 2 standard deviations from the mean?

Question 3: (allowed for 1 group)

Zepolle's Bakery makes a variety of bread types that it sells to supermarket chains in the area. One of Zepolle's problems is that the number of loaves of each type of bread sold each day by the chain stores varies considerably, making it difficult to know how many loaves to bake. A sample of daily demand data is contained in the file called **BAKERY**.
a. Which bread type has the highest average daily demand?
b. Develop a frequency distribution for each bread type.
c. Which bread type has the highest standard deviation in demand?
d. Which bread type has the greatest relative variability? Which type has the lowest relative variability?
e. Assuming that these sample data are representative of demand during the year, determine how many loaves of each type of bread should be made such that demand would be met on at least 75% of the days during the year.
f. Create a new variable called Total Loaves Sold. On which day of the week is the average for total loaves sold the highest?

Question 4: (allowed for 1 group)

The Cozine Corporation operates a garbage hauling business. Up to this point, the company has been charged a flat fee for each of the garbage trucks that enter the county landfill. The flat fee is based on the assumed truck weight of 45,000 pounds. In two weeks, the company is required to appear before the county commissioners to discuss a rate adjustment. In preparation for this meeting, Cozine has hired an independent company to weigh a sample of Cozine's garbage trucks just prior to their entering the landfill. The data file **COZINE** contains the data the company has collected.
a. Based on the sample data, what percentile does the 45,000-pound weight fall closest to?
b. Compute appropriate measures of central location for the data.
c. Construct a frequency histogram based on the sample data. Use the $2^k \geq n$ guideline to determine the number of classes. Also, construct a box and whisker plot for these data. Discuss the relative advantages of histograms and box and whisker plots for presenting these data.
d. Use the information determined in parts (a–c) to develop a presentation to the county commissioners. Make sure the presentation attempts to answer the question of whether Cozine deserves a rate reduction.

Question 5: (allowed for 1 group)

The manager of the Cottonwood Grille recently selected a random sample of 18 customers and kept track of how long the customers were required to wait from the time they arrived at the restaurant until they were actually served dinner. This study resulted from several complaints the manager had received from customers saying that their wait time was unduly long and that it appeared that the objective was to keep people waiting in the lounge for as long as possible to increase the lounge business. The following data were recorded, with time measured in minutes:

34	24	43	56	74	20	19	33	55
43	54	34	27	34	36	24	54	39

    a. Compute the mean waiting time for this sample of customers.
    b. Compute the median waiting time for this sample of customers.
    c. Compute the variance and standard deviation of waiting time for this sample of customers.
    d. Develop a frequency distribution using six classes, each with a class width of 10. Make the lower limit of the first class 15.
    e. Develop a frequency histogram for the frequency distribution.
    f. Construct a box and whisker plot of these data.
    g. The manager is considering giving a complementary drink to customers whose waiting time is longer than the third quartile. Determine the minimum number of minutes a customer would have to wait in order to receive a complementary drink.

TASK 2: ESTIMATING SINGLE POPULATION- answer ALL the questions

Question 1:

Most major airlines allow passengers to carry two pieces of luggage (of a certain maximum size) onto the plane. However, their studies show that the more carry-on baggage passengers have, the longer it takes to unload and load passengers. One regional airline is considering changing its policy to allow only one carry-on per passenger. Before doing so, it decided to collect some data. Specifically, a random sample of 1,000 passengers was selected. The passengers were observed, and the number of bags carried on the plane was noted. Out of the 1,000 passengers, 345 had more than one bag.
    a. Based on this sample, develop and interpret a 95% confidence interval estimate for the proportion of the traveling population that would have been impacted had the one-bag limit been in effect. Discuss your result.
    b. The domestic version of Boeing's 747 has a capacity for 568 passengers. Determine an interval estimate of the number of passengers that you would expect to carry more than one piece of luggage on the plane. Assume the plane is at its passenger capacity.
    c. Suppose the airline also noted whether the passenger was male or female. Out of the 1,000 passengers observed, 690 were males. Of this group, 280 had more than one bag. Using this data, obtain and interpret a 95% confidence interval estimate for the proportion of male passengers in the population who would have been affected by the one-bag limit. Discuss.
    d. Suppose the airline decides to conduct a survey of its customers to determine their opinion of the proposed one-bag limit. The plan calls for a random sample of customers on different flights to be given a short written survey to complete during the flight. One key question on the survey

will be: "Do you approve of limiting the number of carry-on bags to a maximum of one bag?" Airline managers expect that only about 15% will say "yes." Based on this assumption, what size sample should the airline take if it wants to develop a 95% confidence interval estimate for the population proportion who will say "yes" with a margin of error of ±0.02?

Question 2:

Paper-R-Us is a national distributor of printer and copier paper for commercial use. The data file called **Sales** contains the annual, year-to-date sales values for each of the company's customers. Suppose the internal audit department has decided to audit a sample of these accounts. Specifically, they have decided to sample 36 accounts. However, before they actually conduct the in-depth audit (a process that involves tracking all transactions for each sampled account), they want to be sure that the sample they have selected is representative of the population.
   a. Compute the population mean.
   b. Use all the data in the population to develop a frequency distribution and histogram.
   c. Calculate the proportion of accounts for customers in each region of the country.
   d. Select a random sample of accounts. Develop a frequency distribution for these sample data. Compare this distribution to that of the population. (*Hint*: You might want to consider using relative frequencies for comparison purposes.)
   e. Construct a 95% confidence interval estimate for the population mean sales per customer. Discuss how you would use this interval estimate to help determine whether the sample is a good representation of the population. (*Hint*: You may want to use the finite population correction factor since the sample is large relative to the size of the population.)
   f. Use the information developed in parts a–e to draw a conclusion about whether the sample is a representative sample of the population. What other information would be desirable? Discuss.

TASK 2: ESTIMATING TWO POPULATION PARAMETERS

Question 1: (allowed for 1 group)

Logston Enterprises operates a variety of businesses in and around the St. Paul, Minnesota, area. Recently, the company was notified by the law firm representing several female employees that a lawsuit was going to be filed claiming that males were given preferential treatment when it came to pay raises by the company. The Logston human resources manager has requested that an estimate be made of the difference between mean percentage raises granted to males versus females. Sample data are contained in the file **LOGSTONENTERPRISES**. She wants you to develop, and interpret, a 95% confidence interval estimate. She further states that the distribution of percentage raises can be assumed approximately normal, and she expects the population variances to be about equal.

Question 2: (allowed for 1 group)

A pet food producer manufactures and then fills 25-pound bags of dog food on two different production lines located in separate cities. In an effort to determine whether differences exist between the average fill rates for the two lines, a random sample of 19 bags from line 1 and a random sample of 23 bags from

line 2 were recently selected. Each bag's weight was measured and the following summary measures from the samples are reported:

	Production Line 1	Production Line 2
Sample Size, n	19	23
Sample Mean, $\bar{x}$	24.96	25.01
Sample Standard Deviation, s	0.07	0.08

Management believes that the fill rates of the two lines are normally distributed with equal variances.
 a. Calculate the point estimate for the difference between the population means of the two lines.
 b. Develop a 95% confidence interval estimate of the true mean difference between the two lines.
 c. Based on the 95% confidence interval estimate calculated in part (b), what can the managers of the production lines conclude about the differences between the average fill rates for the two lines?

Question 3: (allowed for 1 group)

A credit card company operates two customer service centers: one in Boise and one in Richmond. Callers to the service centers dial a single number, and a computer program routs callers to the center having the fewest calls waiting. As part of a customer service review program, the credit card center would like to determine whether the average length of a call (not including hold time) is different for the two centers. The managers of the customer service centers are willing to assume that the populations of interest are normally distributed with equal variances. Suppose a random sample of phone calls to the two centers is selected and the following results are reported:

	Boise	Richmond
Sample Size	120	135
Sample Mean (seconds)	195	216
Sample St. Dev. (seconds)	35.10	37.80

 a. Using the sample results, develop a 90% confidence interval estimate for the difference between the two population means.
 b. Based on the confidence interval constructed in part (a), what can be said about the difference between the average call times at the two centers?

Question 4: (allowed for 1 group)

The owner of the A.J. Fitness Center is interested in estimating the difference in mean years that female members have been with the club compared with male members. He wishes to develop a 95% confidence interval estimate. Sample data are in the file called **AJ FITNESS**. Assuming that the sample data are approximately normal and that the two populations have equal variances, develop and interpret the confidence interval estimate. Discuss the result.

Question 5: (allowed for 1 group)

Allstate Insurance is one of the major automobile insurance companies in the country. Recently, the western region claims manager instructed an intern in her department to develop a confidence interval estimate of the difference between the mean years that male customers have been insured by Allstate versus female customers. The intern randomly selected 13 male and 13 female customers from the account records and recorded the number of years that the customer had been insured by Allstate. These data (rounded to the nearest year) are as follows:

Males				Females			
14	9	9	16	3	10	4	7
14	8	12	11	10	5	4	1
6	9	10	7	4	4	6	9
3				2			

    a. Based on these data, construct and interpret a 90% confidence interval estimate for the difference between the mean years for male and female customers.

    b. Based on the confidence interval constructed in part (a), what can be said about the difference between the mean years for male and female customers?

SUBMISSION DATE: 6 APRIL 2017 before 5PM