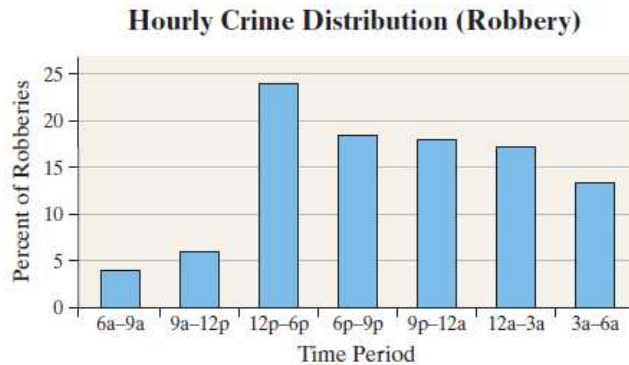


MANB 1123 ASSIGNMENT #1  
**DUE DATE: 23 MARCH 2018 (FRIDAY by 5 PM)**

**RULES:**

1. Each group should only consist of THREE members.
2. Assignment should be done in complete (answer all the questions given) and neat.
3. You may use any statistical packages/tools as you prefer to get the result.
4. Submission should follow the date given and should be in hardcopy. Softcopy should be uploaded into the e-learning.

**Question 1:**



*Source: U.S. Statistical Abstract*

A newspaper article claimed that the afternoon hours were the worst in terms of robberies and provided the following graph in support of this claim. Explain how this graph is misleading.

**Question 2:**

The graphic below is a USA Today type graphic illustrate the women's preferences for shoes.



- a) Which type of shoe is preferred the most? The least?
- b) How is the graphic misleading?

**Question 3:**

For problem (a) until (d) determine the type of sampling use. Give your reasons why you select it as the sampling method.

- a) On Election Day, a pollster for Fox News positions herself outside a polling place near her home and asks the first 50 voters leaving the facility to complete a survey.
- b) An Internet service provider randomly selects 15 residential blocks from a large city and surveys every household in these 15 blocks to determine the number that would use a high-speed Internet service.

- c) Thirty-five sophomores, 22 juniors, and 35 seniors are randomly selected to participate in a study from 574 sophomores, 462 juniors, and 532 seniors at a certain high school.
- d) Officers for the Department of Motor Vehicles pull aside every 40th tractor trailer passing through a weigh station, starting with the 12th, for an emissions test.

**Question 4:**

The following data represent the commuting distances for employees of the Pay-and-Carry Department store.

Commuting Distance (miles)							
3.5	2.0	4.0	2.5	0.3	1.0	12.0	17.5
3.0	3.5	6.5	9.0	3.0	4.0	9.0	16.0
3.5	0.5	2.5	1.0	0.7	1.5	1.4	12.0
9.2	8.3	1.0	3.0	7.5	3.2	2.0	1.0
3.5	3.6	1.9	2.0	3.0	1.5	0.4	6.4
11.0	2.5	2.4	2.7	4.0	2.0	2.0	3.0

- a) The personnel manager for Pay-and-Carry would like you to develop a frequency distribution and histogram for these data.
- b) Break the data into three groups (under 3 miles, 3 and under 6 miles, and 6 and over). Construct a pie chart to illustrate the proportion of employees in each category.
- c) Referring to part (b), construct a bar chart to depict the proportion of employees in each category.

**Question 5:**

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 6:

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:

- Compute the mean, median and standard deviation for the three variables: revenues, profits and number of employees.
- 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.
- 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?
- 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.
- Referring to part (d), how many banks had a profit-per-employee ratio that exceeded 2 standard deviations from the mean?

### Question 7:

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**.

- Which bread type has the highest average daily demand?
- Develop a frequency distribution for each bread type.
- Which bread type has the highest standard deviation in demand?
- Which bread type has the greatest relative variability? Which type has the lowest relative variability?
- 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.
- Create a new variable called Total Loaves Sold. On which day of the week is the average for total loaves sold the highest?

### Question 8:

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.

- Based on the sample data, what percentile does the 45,000-pound weight fall closest to?

- Compute appropriate measures of central location for the data.
- 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.
- 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 9:

An engineer is studying bearing failures for two different materials in aircraft gas turbine engines. The following data are failure times (in millions of cycles) for samples of the two material types.

Material A		Material B	
3.17	5.88	5.78	9.65
4.31	6.91	6.71	13.44
4.52	8.01	6.84	14.71
4.66	8.97	7.23	16.39
5.69	11.92	8.20	24.37

- Determine the sample mean failure time for each material.
- Compute the median failure time for each material.
- Determine the sample standard deviation of the failure times for each material. Which material has its failure times more dispersed?
- Compute the five-number summary (minimum data value,  $Q_1$ ,  $M$  (median),  $Q_3$ , and the maximum data value for each material and draw boxplots for the two materials. Annotate the graph with some general remarks comparing the failure times.
- Describe the shape of the distribution of each material using the boxplot and quartiles.

#### Question 10:

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.

- 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.
- 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.
- 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  $\pm 0.02$ ?

#### Question 11:

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.

- Compute the population mean.
- Use all the data in the population to develop a frequency distribution and histogram.
- Calculate the proportion of accounts for customers in each region of the country.
- 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.*)
- 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.*)
- 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.

#### Question 12:

Phone Solutions, Inc., specializes in providing call center services for companies that wish to outsource their call center activities. There are two main ways that Phone Solutions has historically billed its clients: by the call or by the minute. Phone Solutions is currently negotiating with a new client who wants to be billed for the number of minutes that Phone Solutions is on the phone with customers. Before a contract is written, Phone Solutions plans to receive a random sample of calls and keep track of the minutes spent on the phone with the customer. From this it plans to estimate the mean call time. It wishes to develop a 95% confidence interval estimate for the population mean call time and wants this estimate to be within 0.15 minutes. The question is how many calls should Phone Solutions use in its sample? Since the population standard deviation is unknown, a pilot sample was taken by having three call centers operated by Phone Solutions each take 50 calls for a total pilot sample of 150 calls. The minutes for each of these calls are listed in the file called **PhoneSolutions**.

- How many additional calls will be needed to compute the desired confidence interval estimate for the population mean?
- In the event that the managers at Phone Solutions want a smaller sample size, what options do they have? Discuss in general terms.

### Question 13:

In a random sample of 678 adult males 20 to 34 years of age, it was determined that 58 of them have hypertension (high blood pressure).

*Source:* The Centers for Disease Control.

- Obtain a point estimate for the proportion of adult males 20 to 34 years of age who have hypertension.
- Construct a 95% confidence interval for the proportion of adult males 20 to 34 years of age who have hypertension. Interpret the confidence interval.
- You wish to conduct your own study to determine the proportion of adult males 20 to 34 years old who have hypertension. What sample size would be needed for the estimate to be within 3 percentage points with 95% confidence if you use the point estimate obtained in part (a)?
- You wish to conduct your own study to determine the proportion of adult males 20 to 34 years old who have hypertension. What sample size would be needed for the estimate to be within 3 percentage points with 95% confidence if you don't have a prior estimate?

### Question 14:

An automotive researcher wanted to estimate the difference in distance required to come to a complete stop while traveling 40 miles per hour on wet versus dry pavement. Because car type plays a role, the researcher used eight different cars with the same driver and tires. The braking distance (in feet) on both wet and dry pavement is shown in the table below.

- Why are these matched-paired data?
- Construct a 95% confidence interval for the mean difference in braking distance on wet versus dry pavement where the differences are computed as "wet minus dry." Interpret the interval.

**Note:** A normal probability plot and boxplot of the data indicate that the differences are approximately normally distributed with no outliers.

Car	1	2	3	4	5	6	7	8
Wet	106.9	100.9	108.8	111.8	105.0	105.6	110.6	107.9
Dry	71.8	68.8	74.1	73.4	75.9	75.2	75.7	81.0

### Question 15:

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 16:**

A television sports commentator wants to estimate the proportion of Malaysian who follows professional football. What sample size should be obtained if he wants to be within 3 percentage points with 95% confidence if;

- (a) He uses an estimate of 52% obtained from a poll?
- (b) He does not use any prior estimates?
- (c) Why are the results from parts (a) and (b) so close?

**Question 17:**

Production engineers at XYZ Manufacturer believe that a modified layout on its assembly lines might increase average worker productivity (measured in the number of units produced per hour). However, before the engineers are ready to install the revised layout officially across the entire firm's production lines, they would like to study the modified line's effects on output. The following data represent the average hourly production output of 12 randomly sampled employees before and after the line was modified:

Employee	1	2	3	4	5	6	7	8	9	10	11	12
Before	49	45	43	44	48	42	46	46	49	42	46	44
After	49	46	48	50	46	50	45	46	47	51	51	49

Identify the point estimate of the mean difference and estimate the 95% confidence interval for the difference for the sampled employees.

**Question 18:**

An article in The American Statistician (M. L. R. Ernst, et al., "Scatterplots for Unordered Pairs," 50 (2010), pp. 260–265) reports on the difference in the measurements by two evaluators of the cardiac output of 23 patients using Doppler echocardiography. Both observers took measurements from the same patients. The measured outcomes were as follows:

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Evaluator 1	4.8	5.6	6.0	6.4	6.5	6.6	6.8	7.0	7.0	7.2	7.4	7.6
Evaluator 2	5.8	6.1	7.7	7.8	7.6	8.1	8.0	8.21	6.6	8.1	9.5	9.6

Patient	13	14	15	16	17	18	19	20	21	22	23
Evaluator 1	7.7	7.7	8.2	8.2	8.3	8.5	9.3	10.2	10.4	10.6	11.4
Evaluator 2	8.5	9.5	9.1	10.0	9.1	10.8	11.5	11.5	11.2	11.5	12.0

Identify the point estimate of the mean difference and estimate the 90% confidence interval for the difference for the sampled patients.