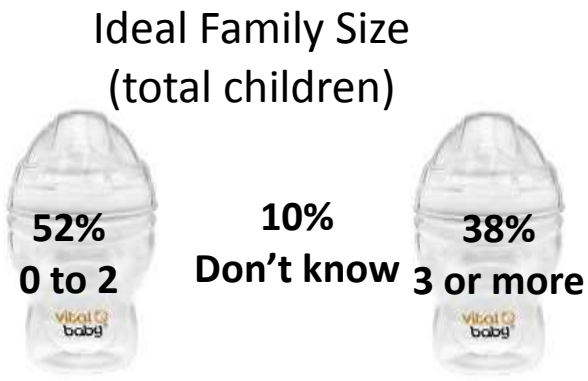
MANB 1123 ASSIGNMENT 1 PART 1

1. **Assignment 1** consists of 2 parts:
   1st part will cover Descriptive Statistics and Estimation and 2nd part will cover Hypothesis Testing
2. This assignment should be conducted **in pairs**.
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 submitted in hardcopy. The softcopy version should be submitted via e-learning

TASK 1: DESCRIPTIVE STATISTICS -

Question 1:

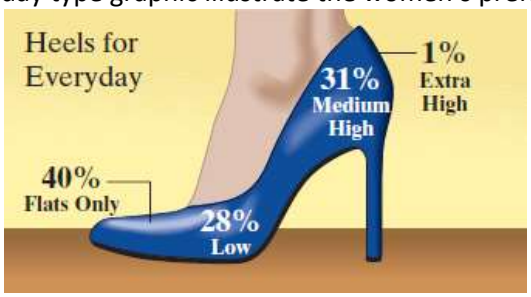
The following USA Today type graphic illustrate the ideal family size (total children) based on a survey of adult American.



a) What type of graphic is being displayed?
b) Describe any problems with the graphic.
c) Construct a graphic that is not misleading and makes the data stand out.

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 graph misleading?

Question 3:

The following data represent the number of cars that arrived at a McDonald's drive-through between 11:50 a.m. and 12:00 noon each Wednesday for the past 50 weeks:

1	7	3	8	2	3	8	2	6	3
6	5	6	4	3	4	3	8	1	2
5	3	6	3	3	4	3	2	1	2
4	4	9	3	5	2	3	5	5	5
2	5	6	1	7	1	5	3	8	4

a) Construct a frequency distribution of the data.
b) Construct a relative frequency distribution of the data.
c) Construct a cumulative frequency distribution of the data.
d) Construct a cumulative relative frequency distribution of the data.
e) Construct a frequency histogram of the data. Describe the shape of the distribution.
f) Construct a relative frequency histogram of the data.
g) What percentage of weeks did exactly three cars arrive between 11:50 a.m. and 12:00 noon?
h) What percentage of weeks did three or more cars arrive between 11:50 a.m. and 12:00 noon?

Question 4:

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?

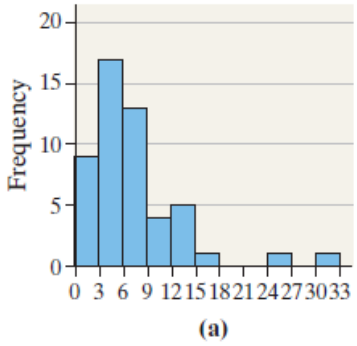
Question 5:

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.

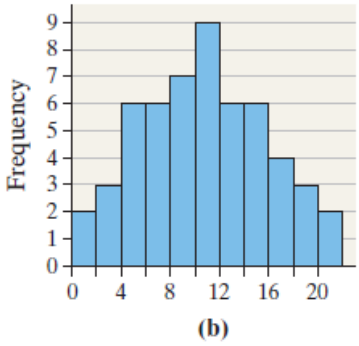
a. Based on the sample data, what percentile does the 45,000-pound weight fall closest to?
b. Compute appropriate measures of central location for the data.
c. Construct a frequency histogram based on the sample data. Use the $2^k \geq n$ guideline to determine the number of classes. Also, construct 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 6:

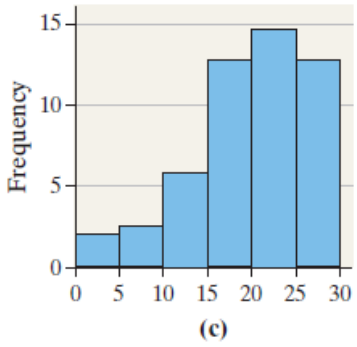
For each of the three histograms shown, determine whether the mean is greater than, less than, or approximately equal to the median. Justify your answer.



(a)



(b)



(c)

Question 7:

The data on the next page represent the travel time (in minutes) to school for nine students enrolled in Sullivan's College Algebra course. Treat the nine students as a population.

Student	Travel Time	Student	Travel Time
Amanda	39	Scot	45
Amber	21	Erica	11
Tim	9	Tiffany	12
Mike	32	Glenn	39
Nicole	30		

a) Determine the population mean for travel time.

b) Find three simple random samples of size 4 and determine the sample mean for travel time of each sample.

c) Which samples result in a sample mean that overestimates the population mean? Which samples result in a sample mean that underestimates the population mean? Do any samples lead to a sample mean that equals the population mean?

TASK 2: ESTIMATION

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.

Question 3:

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 4:

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 5:

The American Society for Microbiology (ASM) and the Soap and Detergent Association (SDA) jointly commissioned two separate studies, both of which were conducted by Harris Interactive. In one of the studies, 1001 adults were interviewed by telephone and asked about their handwashing habits. In the telephone interviews, 921 of the adults said they always wash their hands in public restrooms. In the other study, the hand-washing behavior of 6076 adults was inconspicuously observed within public restrooms in four U.S. cities and 4679 of the 6076 adults were observed washing their hands.

(a) In the telephone survey, what is the variable of interest? Is it qualitative or quantitative?
(b) What is the sample in the telephone survey? What is the population to which this study applies?
(c) Verify that the requirements for constructing a confidence interval or the population proportion of adults who say they always wash their hands in public restrooms are satisfied.
(d) Using the results from the telephone interviews, construct a 95% confidence interval for the proportion of adults who say they always wash their hands in public restrooms.