Carleton University

School of Mathematics and Statistics

STAT2507 – Assignment 1

Due date – June 04, 2014

Total Marks =100

INSTRUCTIONS

A. Assignments are to be submitted in-class on the due date and prior to beginning of class. No late assignments will be accepted without sufficient advanced notice and a legitimate, documented reason. **No electronic submissions.**
B. You must show all your work. No credit will be given for answers without justification. Only use MINITAB for questions in Part A
C. For questions that require MINITAB, you must include all relevant output with your assignment.

PART A: MINITAB Questions

1) [10]Fifty people are grouped into four categories – A, B, C, and D – and the number of people who fall into each category is shown in the table. First, copy and paste from excel into MINITAB worksheet

Category	Frequency
A	11
B	14
C	20
D	5

   a. [1]What is the experimental unit?
   b. [1]What is the variable being measured? Is it qualitative or quantitative?
   c. [1.5]Construct a pie chart to describe the data (by clicking Graph->pie chart-> check "Chart values from a table -> put category in "Categorical variable" -> put Frequency in "Summary variables" -> click "Ok")
   d. [1.5]Construct a bar chart to describe the data (by clicking Graph-> Bar chart-> select "Values from a table" in drop-down menu named "Bars represent" -> then select "simple" -> click "Ok" -> put Frequency in "Graph Variables" -> Category in "Categorical variable" -> click "Ok")
   e. [1]Does the shape of the bar chart in part (d) change depending on the order of presentation of the four categories? Is the order of presentation important?
   f. [2]What proportion of the people is in category B, C, or D?
   g. [2]What percentage of the people is not in category B?

2) [10]To decide on the number of service counters needed for stores to be built in the future, a supermarket chain wanted to obtain information on the length of time (in minutes) required to service customers. To find the distribution of such a time, a sample of 60 customers service time are given in excel file. Copy and paste the data into MINITAB worksheet.
   a. [2]Construct a stem and leaf plot for the data (Graph->Stem-and-leaf-> put Times in "Graph variables"-> click "Ok")
   b. [1]Describe the shape of the distribution.
   c. [2]What is the median of the customers' service time?

d. [3]What fraction of the service times are less than or equal to 1 minute?
e. [2]What is the smallest and largest of the 60 measurements?

3) [10]Using the service time data,
   a. [2]Construct a relative frequency histogram for super market service times. (Graph-> histogram -> select "simple" -> click "Ok" -> put Times in "Graph variables" -> click "Ok")
   b. [1]Describe the shape of the distribution. Do you see any outliers?
   c. [1]Compare the relative frequency histogram with the stem and leaf plot in Q2). Do the two graphs convey the same information?
   d. [3]Use *desc* command in MINITAB to find the mean, median, standard deviation, minimum, maximum, 1^{st} quartile, 3^{rd} quartile of the service times.
   e. [3]Calculate the z-scores for the smallest and largest service times and decide if any one of them is outliers or not.

4) [10]Ages of 50 pennies are given in Excel. Copy and paste this data from Excel to MINITAB. Answer the following using desc command in minitab.
   a. [1]What is the average age of pennies?
   b. [1]What is median age of pennies?
   c. [1]Based on the results of parts a and b, how would you describe the age distribution of these 50 pennies
   d. [4]Construct a box plot for the data set. [1]Are there any outliers? [1]Does the box plot confirm your description of the distribution's shape, [1]explain? (Graph -> Boxplots -> select "simple" -> put Age(Years) in "Graph variables" -> click "Ok")

PART B: Written Questions

1) [12]Identify each of the following variables as qualitative or quantitative, if quantitative, identify as discrete or continuous.
   a. [2]Ethnic origin of a candidate for public office
   b. [2]Number of people in line at a supermarket checkout counter
   c. [2]Depth of a snowfall
   d. [2]Number of aircraft arriving at the Ottawa International airport in a given hour
   e. [2]Province or territory in which a person lives
   f. [2]Time required to complete a questionnaire

2) [8]The following is the data of 20 healthy volunteers at a cadence of 48 cycles per minute (a cycle consisted of two steps): 87, 101, 109, 91, 79, 78, 80, 112, 96, 94, 95, 98, 90, 94, 92, 107, 96, 81, 98, 96.
   a. [5]Construct a stem and leaf plot to describe the data
   b. [3]Discuss the characteristics of the distribution

3) [16]How much sleep do you get on a typical school night? A group of 10 university students were asked to report the number of hours that they slept on the previous night with the following results: 7, 6, 7.25, 7, 8.5, 5, 8, 7, 6.75, 6.
   a. [4]Find the mean and standard deviation of the number of hours of sleep for these 10 students
   b. [2]Calculate the z-score for the largest value (x=8.5). [1]Is this an unusually sleepy college student?
   c. [2]What is most frequently reported measurement? [1]What is the name for this measure of centre?

d. [5]Construct a box plot for the data. [1]Does the box plot confirm your results in part b? (HINT: Since the z-score and the box plot are two unrelated methods for detecting outliers, and use different types of statistics, they do not necessarily have to [but usually do] produce the same results.)

4) [12]The litres per 100 km (L/100km) for each of 20 medium-sized cars, selected from a production line during the month of March, are as follows: 9.7, 9.9, 9.5, 10.1, 9.8, 10.3, 9.4, 8.9, 10.2, 9.4, 9.5, 9.6, 7.9, 8.1, 9.8, 10.0, 8.9, 9.9, 11.3, 10.9.
   a. [3]Find the mean and standard deviation
   b. Arrange the data from smallest to largest. [2]Find the z-scores for the largest and smallest observations. [1]Would you consider them to be outliers? [1]Why or why not?
   c. [2]What is the median?
   d. [3]Find the lower and upper quartiles

5) [12]A set of bivariate data consists of these measurements on two variables, x and y: (3, 6), (5, 8), (2, 6),
   (1, 4), (4, 7), (4, 6)
   a. [2]Draw a scatter plot to describe the data
   b. [2]Does there appear to be a relationship between x and y? If so, how do you describe it?
   c. [3]Calculate the correlation coefficient, r, using the computing formula given in this section.
   d. [3]Find the best-fitting line using the computing formulas. [1] Graph the line on the scatter plot from part a. [1]Does the line pass through the middle of the points?