STAT 2507 A/BIT 2000 A **Assignment 1 (Chs. 1 and 2)** **Fall Term 2014**

Last Name ———————————— **First Name** ————————————

Student # ———————————— **Lab Section (Important!)** ——————

Due: Tuesday, October 7, in class.

NOTE: Late assignments will not be accepted. Use spaces left to answer all questions. Total mark = 100. Print and attach all graphs you are asked to generate.

Part I. Lab Questions

1. **Service times.** To decide on the number of service counters needed for stores to be built in future, a supermarket chain wants to obtain information on the length of time (in minutes) required to serve customers. To find the distribution of such time, a sample of 60 customers service times was recorded (see the first column in the Excel file). Copy and paste the data in this column in a Minitab spreadsheet (which you need to open first).

   (a) [2] Construct a stem-and-leaf plot for the data by clicking **Graph→ Stem-and-Leaf**.

   (b) [2] What is the customer's service median time? **1.2.**

   (c) [2] What fraction of the service times are less than or equal to 1 min? **24/60=0.4.**

   (d) [2] What are the smallest and the largest service times? **0.2** and **5.2.**

   (e) [2] 25% of the service times are above what value? $\mathbf{Q_3 = 2.4}$**.**

2. **Service times (continued).**

   (a) [2] Construct a dotplot for the data above by clicking **Graph→ Dotplot→ Simple**.

   (b) [2] What shape does this data have? **Skewed to the right.**

   (c) [2] Do you see any outliers? **Yes. The value 5.2 (larger than 4.9) seems to be an outlier.**

3. **Measurements.** Copy and paste the data in the 2nd column of the Excel file to your Minitab work spreadsheet. This data corresponds to some $\mathbf{n = 120}$ measurements. Now activate the command mode by first clicking on the session window and then choosing **Editor→ Enable Commands** on the top menu bar.

   (a) Use *desc* command to find the mean [2] $\bar{x}$= **8.403** and the standard deviation [2] s= **2.263** of this set of data. Using the stem-and-leaf plot of this data, what is the percentage of measurements that fall in the intervals [2] $\bar{x} \pm s$? **86/120=71.67%** and [2] $\bar{x} \pm 2s$? **113/120=94.17%.**

   (b) What shape does the data have? [2] **Skewed to the left.** Is all this in line with the Empirical Rule? [1] **No.** Explain. [3] **The distribution of the data is not mound-shaped and symmetric, and therefore the Empirical Rule cannot be used.**

4. **Ages of pennies.** In column C of the Excel file, ages of 50 pennies (= current year − the year on the penny) were recorded. Copy and paste this data into your Minitab work spreadsheet.

(a) What is the median age of the pennies? [2] **4.00.** What is the average age of the pennies? [2] $\bar{\mathbf{x}} = \mathbf{8.36}$.

(b) Based on the values obtained in part (a), would you say that the distribution of the ages is symmetric, skewed to the left, or skewed to the right? [2] **Skewed to the right.**

(c) Construct a boxplot for the penny's ages by clicking on **Graph** → **Box-And-Whisker** → **Simple**. Are there any outliers? [2] **No. There are no outliers.**

(d) Does the boxplot confirm your description of the shape of the distribution in part (b)? [1] **Yes.** Explain. [2] **The median is closer to Q_1, and it is smaller than the mean.**

(e) Based on the boxplot, 25% of the pennies are approximately older than what value? [2] $\mathbf{Q_3 = 17.5}$.

5. **Mileage.** The data set named "mileage" in the Excel file corresponds to the number of liters per 100km for selected 20 medium-sized cars. Copy these data into your Minitab work spreadsheet.

   (a) Construct a frequency histogram of these data. How would you describe the shape of the distribution? [2] **Skewed to the left.**

   (b) Use *desc* command in the Minitab command's session to find the mean [1] $\bar{\mathbf{x}} = \mathbf{9.655}$, median [1] $\mathbf{m = 9.750}$, standard deviation [1] $\mathbf{s = 0.804}$,minimum [1] **7.9**, maximum [1] **11.3**, lower quartile [1] $\mathbf{Q_1 = 9.4}$, upper quartile [1] $\mathbf{Q_3 = 10.075}$ of the mileage.

   (c) Compute the z-scores for the smallest mileage [2] $\mathbf{z - score} = \frac{\mathbf{7.9-9.655}}{\mathbf{0.804}} = \mathbf{-2.182}$ and for the largest mileage [2] $\mathbf{z - score} = \frac{\mathbf{11.3-9.655}}{\mathbf{0.804}} = \mathbf{2.04}$, and decide if the smallest mileage and/or the largest mileage are/is outlier(s) or not [2] **Neither one is an outlier.**

Part II. Comprehension Questions

1. Identify each of the following variables as categorical (qualitative), discrete, or continuous.

   (a) [2] Number of times per year a person catches a cold. **Discrete.**

   (b) [2] Wind speed (km/hour) in Ottawa in a given day. **Continuous.**

   (c) [2] The colour of a ball drawn from a box containing two red and three white balls. **Qualitative (or categorical).**

   (d) [2] Monthly unemployment rate in Canada over the last 10 years. **Continuous.**

2. Consider the following observations

   8, −1.3, −1.8, 1.8, 1.3, 1.7, −1.2, 1.2, 1.4, −1.4, 0.5, −0.5, 0.2, −0.2, −1, 0.1, 3, −3

   (a) [8] Compute $\bar{x}$, the mean of these data, and the variance s^2. Show all of your work. **Solution:** First, using two decimal places, we get $\bar{x} = \frac{1}{18}\sum_{i=1}^{18} x_i = 0.49$. Next, using the computational formula for the variance s^2, we can write

$$s^2 = \frac{\sum_{i=1}^{18} x_i^2 - \frac{(\sum_{i=1}^{18} x_i)^2}{18}}{17},$$

   where $\sum_{i=1}^{18} x_i^2 = 103.14$ and $\sum_{i=1}^{18} x_i = 8.8$. Now, substituting these values in the above formula for s^2, we get (up to two decimal places) that $s^2 = \frac{103.14 - \frac{(8.8)^2}{18}}{17} = 5.81$.

(b) [4] Using z-score, decide on wether or not 8 is an outlier. Justify your answer.

Solution: The z-score of the measurement 8 is

$$z - score = \frac{8-\bar{x}}{s} = \frac{8-0.49}{\sqrt{5.81}} = 3.12 > 3,$$

and hence the measurement 8 is an outlier.

3. The following is a stem-and-leaf plot for data on the costs (in dollars) of a sample of 30 postal mailings by a company.

85	2
86	
87	
88	
89	
90	
91	3 4 8
92	0 0 1 3
93	1 1 2 5 7
94	0 1 3 3 4 8 9
95	2 3 4 5 5
96	2
97	0 4 7
98	
99	
100	0

Leaf Unit = 0.01

(a) [4] What are the values of the median, lower quartile (Q1), and upper quartile (Q3) for this data set?

(b) [6] Construct a box plot for the given data. Using the 1.5×IQR rule, list all potential outliers (if any).

Solution: (a) We have $Q_1 = 9.225$ and $Q_3 = 9.5425$. (b) To construct a box plot, we need the five-number summary: Min = 8.52, $Q_1 = 9.225$, $m = 9.42$, $Q_3 = 9.5425$ and Max = 10.00. From this, the lower and upper fences are:

$$\begin{aligned} \text{Lower Fence} &= Q_1 - 1.5 \times \text{IQR} = 9.225 - 1.5(9.5425 - 9.225) = 8.74875, \\ \text{Upper Fence} &= Q_3 - 1.5 \times \text{IQR} = 9.5425 + 1.5(9.5425 - 9.225) = 10.01875. \end{aligned}$$

Now, as the value 8.52 is smaller than the lower fence, it may be considered to be an outlier.

4. [12] Compute the Five-Number Summary for the following set of measurements, and identify the outliers (if any) by using the lower and upper fences:

$$19,\ 15,\ 16,\ 0,\ 14,\ 9,\ 4,\ 0,\ 12,\ 13,\ 10,\ 19,\ 7,\ 15,\ 10.$$

Solution: After arranging the observations in an increasing order, it is easy to find that Min = 0, $Q_1 = 7$, $m = 12$, $Q_3 = 15$ and Max = 19.