STAT 2507 Assignment # 1 (Chapters 1&2) Fall 2013

Last Name ——————————————- , **First** ——————————-

Student # ——————————

Total of marks=100. Every question is worth 4 marks.

Part I. Lab Questions. Use only the blanks left to answer lab questions. Print all histograms, boxplots, stem-and-leaf plots, etc, you are asked to generate.

1. (Service times) To decide on the number of service counters needed for stores to be built in 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 time, a sample of 60 customers service times was recorded (see the first column in the Excel file). You are asked to copy and paste data in this column in a minitab spreadsheet (that you have to open first).

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

   (b) What is the customer's service median time? **median=1.2**

   (c) What fraction of the service times are less than or equal to 1 min? **24/60**

   (d) What are the smallest and the largest service times? **0.2** and **5.2**

   (e) Third of the service times are above what value? $Q_3 = 2.4$

2. (Service times, continued.)

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

   (b) What shape does this data have? Skewed to the right

   (c) Do you see any outliers? Yes, the measurement 5.2 (larger than 4.9) seems to be an outlier

3. (Measurements) Copy and paste 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 clicking **Editor→ Enable Commands**

   (a) Use *desc* command to get the mean $\bar{x}$= **8.403** and the standard deviation s= **StDev=2.263** of this set of data. Using the stem-and-leaf of this data what is the percentage of measurements that fall in the intervals $\bar{x} \pm s$? **86/120= 71.66**% and $\bar{x} \pm 2s$? **113/120=0.94%**

   (b) What shape does the data have? **Skewed to the left**. Is all this in line with the empirical rule? **NO** Explain. **The distribution of the data is not mound shaped and symmetric**

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? **4.00** What is the average age of the pennies? $\bar{x} = 8.36$

(b) Based on the values above, would you say that the distribution of the ages is symmetric, skewed to the left, or skewed to the right? **Skewed to the right**

(c) Construct a boxplot for the penny's ages by clicking on **Graph** → **Boxplot** → **Simple**. Are there any outliers? **No Outliers**

(d) Does the boxplot confirm your description of the distribution's shape (in part b)? **Yes** -Explain. **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? $Q_3 = 17.5$

(a) (Mileage) Data 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.

(b) Construct a frequency histogram for these data by clicking **Graph** → **Histogram** → **Simple**. How would you describe the shape of the distribution? **Skewed to the left**?

(c) Use *desc* command in the minitab command's session to find the mean $\bar{x} = 9.655$, median **9.750** , standard deviation $StDev = 0.804$, minimum **7.9**, maximum **11.3**, 1st quartile $Q_1 = 9.4$, 3rd quartile $Q_3 = 10.075$ of the mileage.

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

Part II Comprehension questions

(a) Identify each of the following variables as categorical (qualitative), discrete, or continuous. Use space left.

i. Number of times per year a person catches a cold. **Discrete**

ii. Wind speed (Km/hour) in Chicago. **Continuous**.

iii. The color of a ball drawn from a box containing two red, and 3 white balls. **Qualitative (or Categorical)**.

iv. Monthly unemployment rate in Canada **Continuous** .

(b) Consider the following observations -8, -1.3, -1.8, -1.7, 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.

i. Compute $\bar{x}$, the mean of these data, and s, the standard deviation. Show all your work. Sol: $\bar{x} = \frac{-8-1.3-1.8-1.7+1.8+1.3+1.7-1.2+1.2+1.4-1.4+0.5-0.5+0.2-0.2-1+0.1}{17} \approx -0.523$ and $S^2 = \frac{88.03-\frac{(-8.9)^2}{17}}{16} \approx 5.21$, so $s = \sqrt{S^2} = \sqrt{5.21} \approx 2.28$

ii. Using z-score with mentioning the reason decide wether -8 is an outlier.

Sol: **-8 is an outlier since its z-score is** $\frac{-8+0.523}{2.28} \approx -3.28$. Therefore, $|z-score| > 3$.

(c) Compute the Five-Number summary for the following measurements and identify any outliers using the lower and upper fences.
19, 12, 16, 0, 14, 9, 4, 1, 12, 13, 10, 19, 7, 5, 10.

Sol: Min=0, $Q_1 = 5.00$, median=10.00, $Q_3 = 14.00$ and max=19

Lower fence=$Q_1 - 1.5 * (Q_3 - Q_1) = 5 - 1.5 * (14 - 5) = -8.5$

Upper fence=$Q_3 + 1.5 * (Q_3 - Q_1) = 14 + 1.5 * (14 - 5) = 27.5$

None of the measurements is below the lower fence or above the upper fence. Hence no outliers.