

STAT 2507 Solutions (-plots) to Assignment # 1 (Chapters 1&2) Winter 2013

Last Name _____, First _____

Student # _____

Lab group: very important (D_1, \dots , or D_8 , E_1, \dots , or E_8 , F_1, \dots , or F_5 .) _____

Due for sec. D and E on Tuesday. Feb. 5th, and sec. F on Wed. Feb. 6th, IN CLASS.

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

Notes:

1. Please read the note “Minitab and SAS Access” under the link click here.
2. Minitab data used in the lab part below are in the Minitab file Stat2507Lab1data on Culearn.
If you are using one of the computers in the lab room HP3393, please save the data file to “Labdir” using instructions in the note above.
3. For your printing needs from the lab room, please read “Printing” note in the link above.

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

1. (United Fund) Column *Donation* records the yearly charitable contributions (in dollars) to the United Fund for a group of 40 employees at a public university.
 - (a) Construct a stem-and-leaf plot for the data.
 - (b) What is the group’s median donation?—79—
 - (c) What fraction of donations is below 60 dollars?— $8/40=1/5$ —
 - (d) What are the smallest and the largest donations?—24— and —125—
 - (e) Quarter of the donations are above what value?—84—
2. (United Fund, continued.)
 - (a) Construct a boxplot for the data above.
 - (b) What shape does this data have?Skewed to the left

- (c) Do you see any outliers? (what value, if any?)—Yes. \$24——
3. (Recurring illness) The length of time (in months) between the onset of a particular illness and its recurrence was recorded for $n = 50$ patients (see column *Time*).
- (a) Construct a percent histogram for the data.
- (b) Would you describe the shape as roughly symmetric, skewed to the right, or skewed to the left?—Skewed to the right.——
- (c) Give the percentage of recurrence times less than 6 months.—48%——
- (d) Find the range—32-02=31.8—— Use it to approximate the standard deviation $s \approx R/4 = 31.8/4 = 7.95$ — Compare to the exact value of $s = 7.67$. Quite close.—
- (e) Construct a boxplot for this data and check for any outliers—32——. If any, compute their z -score—3.08——
- (f) Would you use Chebychev's theorem or empirical rule for this data?——Chebychev—
—Explain—Since distribution is not fairly symmetric——.
4. Column *Ages* contains ages (in months) at which 50 children were enrolled in a preschool.
- (a) Use *desc* command to get the mean \bar{x} =39.08 and the standard deviation s = 5.986— of this set of data. Using the stem-and-leaf of this data what is the percentage of ages that fall in the intervals $\bar{x} \pm s$?31/50=62% and $\bar{x} \pm 2s$?49/50=98%.
- (b) What shape does the data have?-Rather symmetric———Is all this in line with the empirical rule?-Yes.— Explain-Empirical rule says one should have about 95% between $\bar{x} \pm 2s$. We got 98%, which is even better—— Roughly, how do the mean and median confirm the data shape?-Pretty much close to each other: 39.08 and 39——
- (c) Use *desc* command to find the five-number summary of the *Ages* data Min=30, Q1=34.75, Media=39, Q3=43, and max=55.
- (d) 25% of ages above what value?-43 months—. How old was the oldest student?—55—. 50% of ages between what values? 34.75—— and 43. The interquartile range is 8.25.

Part II Comprehension questions

1. Identify each of the following variables as categorical, discrete, or continuous. Use space left.
- (a) Blood type for a randomly selected person. —Categorical——.

- (b) Amount of snow (in cm) of the next snow storm in Ottawa. —Continuous—.
- (c) Daily exchange rate of Canadian dollar versus US dollar—Continuous—.
- (d) The gender of the next newborn at a local hospital in Ottawa.—Categorical—.

2. Consider the following observations 1,0, 5, 10, -5, 3.

- (a) Compute \bar{x} , the mean of these data, and s , the standard deviation, and find the median value. Use space left below to show all your work.

$$\bar{x} = \frac{1 + 0 + 5 + 10 + -5 + 3}{6} = 2.33$$

$$S^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1} = \frac{1 + 0 + 25 + 100 + 25 + 9 - (1 + 0 + 5 + 10 - 5 + 3)^2/6}{5} = 25.46.$$

$$s = \sqrt{S^2} = \sqrt{25.46} = 5.04, \text{ and Median}=(1+3)/2=2.$$

- (b) Compute the z -score for the observation 10. Would you consider this value as an outlier? Why or why not? Use space below to show your work. z -score for observation 10 is $(10-2.33)/5.04=1.52$. Since z score is not above 3, we do not consider this observation as an outlier.

- (c) A company interested in lumbering rights for a certain tract of slash pine trees is told that the mean diameter of these trees is 35 cm with a standard deviation of 7 cm. Assuming the distribution of diameters is roughly symmetric, what percent of the trees will have diameters between 21 and 55 centimetres? Use space left below to show your work.

Observe that $(21, 49)=(35-2 \times 7, 35+2 \times 7)=(\bar{x}-2 \times s, \bar{x}+2 \times s)$. According to empirical rule, such an interval would contain approximately 95% of data (if data symmetric, which is the assumption). So, the answer is 95%.