Solutions

STAT 2507 Assignment # 1 (Chapters 1, 2 and 3) Winter 2015

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

Student # ————————

Lab group: very important ————

Due: Sections E, F and H, Monday February 9
Section G Tuesday February 10

The Assignments are to be handed IN CLASS. No late Assignment will be accepted. No electronic submission will be accepted.

Part I. Lab questions. Use only the blanks left to answer lab questions.

1. (Students' Weights) Column *Weights* contains the weights (to the nearest tenth of a kilogram) of 30 students.

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

   (b) [2] What is the median weights of the students? ————————Answer $Q_2 =$ 60.85

   (c) [2] What fraction of weights below 60 kilograms? ———————— Answer 9/30

   (d) [2] What are the smallest and the largest weights? ——— and ———————— Answer Min= 56.30 and Max= 65.70

   (e) [2] Quarter of the weights are above what value? ———————— Answer $Q_3 = 61.75$

2. (Histogram Data)

   (a) [2] Construct a frequency histogram for "Histogram.Data" in the data file by clicking on **Graph → histogram**

   (b) [2] Circle the shape of the frequency histogram of the data. **Skewed to the left** **Symmetrical** **Skewed to the right** Answer Skewed to the left

(c) [2] What is the relationship between the mean and the median of this data set? Answer the mean is smaller than the median (the mean is pulled towards the left while the median is not hence the mean is smaller than the median)

(d) [2] Now activate the command mode by clicking **Editor → Enable Commands**. Use the command *desc* for the column (Histogram) data. The mean is ———— and the median is ———— Answer mean= 7.778 and median= 8.150

(e) [2] Would you use Chebychev's theorem or empirical rule for this data?———— —Explain———————————————————————————— -.

Answer Chebychev's theorem, since the data are skewed

3. (Milage) The data in the column "Milage" are the milage in kilometers of 20 cars on one liter gasoline.

(a) [2] Construct a boxplot for the data in column Milage by clicking on **Graph → box plot**.

(b) [2] Do you see any outliers? How many? ———— Answer Yes, 1 outlier

(c) [2] Look at the boxplot and at the measurements and List the value(s) you think can be outliers.———— Answer: The largest values 18.1

(d) [2] Compute lower fence ———— and upper fence ———— Answer Upper fence= $Upper fence = 9.975 + 1.5 * (9.975 - 9.175) = 11.175$ and Lower fence= $9.175 - 1.5 * (9.975 - 9.175) = 7.975$

(e) [2] Compare your candidates for outliers in part (c) to the lower and the upper fence. —————————————————————————— Answer: $18.1 > Upper fence = 11.175$.

4. (Regression) The data in columns "Brain" and "Body" are averages of weights of brain and body weights of a number of mammal species.

(a) [2] Compute the correlation coefficient between the Brain and the Body weights? ———— Answer $r = 0.934$

(b) [2] What is the equation of regression line when the Body weight is used to predict the Brain weight.———————————————— Answer $Brain = 0.9029 * Body - 56.855$

(c) [2] Use the preceding regression line to predict the Brain weight of an specie with Body weight=110. ——— Answer 0.9029*110 - 56.855=**42.464**

(d) [2] What is the equation of regression line when the Brain weight is used to predict the Buddy weight. ——— Answer $Body = 0.9665 * Brain + 91.0044$.

(e) [2] Predict the Body weight of a mammal specie with Brain weight=5. ——— Answer $0.9665 * 5 + 91.0044 = 95.8369$.

Part II Comprehension questions

1. Identify each of the following variables as categorical, discrete, or continuous. Use space left.

   (a) [2] Blood type for a randomly selected person. ————————. Answer: Categorical or Qualitative

   (b) [2] Amount of snow (in inch) of the next snow storm in Ottawa ————————. Answer: Quantitative and Continuous

   (c) [2] Daily exchange rate of Canadian dollar versus US dollar ————————. Answer: Quantitative and Continuous

   (d) [2] The number of car accidents in Ottawa area tomorrow ————————. Answer: Quantitative and Discrete

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

   (a) [3] Compute $\bar{x}$, the mean of these data, and S, the standard deviation, and find the median value.
   Solution: $\bar{x} = \frac{1+0+5+10-5+3}{6} = 2.333$ and
   $S = \sqrt{\frac{(1-2.33)^2+(0-2.33)^2+(5-2.33)^2+(10-2.33)^2+(-5-2.33)^2+(3-2.33)^2}{5}} = 5.046.$

   (b) [2] Compute the z-score for the observation 10. Would you consider this value as an outlier? Why or why not?
   Solution: z-score=(10-2.333)/5.046= 1.519. Hence, 10 is not an outlier.

3. The following is a stem and leaf plot for data on the costs (in dollars) of a sample of 31 postal mailings by a company. **Leaf Unit = 0.01**

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 6 6 7 8 9
95	2 3 4 5 5
96	0 5
97	0 4 7
98	
99	
100	0

(a) [4] What are the values of the median, lower quartile (Q1), and upper quartile (Q3) for this data set?
Solution: $Q_1 = 9.23$, $median = 9.46$ and $Q_3 = 9.55$

(b) [8] Construct a box plot for the given data. Using the 1.5*IQR rule, list any potential outliers, if any.
Solution: To construct a box plot we need the five number summary that are $Min = 8.52$, $Q_1 = 9.23$, $median = 9.46$, $Q_3 = 9.55$ and $Max = 10.00$. Also we need $lower\ fence = Q_1 - 1.5 * IQR = 9.23 - 1.5 * (9.55 - 9.23) = 8.75$ and $upper\ fence = Q_3 - 1.5 * IQR = 9.55 + 1.5 * (9.55 - 9.23) = 10.03$.
Since 8.52<lower fence=8.75, the measurement 8.52 is an outlier.

4. [5] 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 symmetrical and bell-shaped, what percent of the trees will have diameters between 21 and 49 centimeters?
Solution: Note that $[21, 49] = [35 - 2 * 7, 35 + 2 * 7]$. Since the distribution is symmetrical

and bell-shaped according to the empirical rule, is approximately 95%

5. We have $n = 1000$ measurements whose mean and variance respectively, are $\bar{x} = 6$ and $S^2 = 9$.

   (a) [5] What is the minimum number of measurements that lie inside $[0, 12]$. Indicate the number and explain how it was obtained.
   Solution: Note that here $[0, 12] = [\bar{x} - 2S, \bar{x} + 2S]$. Using Tchebesheff theorem there are **at least** $1000 * 3/4 = 750$.

   (b) [5] If you know that the shape of the distribution of the measurements is bell-shaped and symmetrical, then approximately how many measurements are in the interval $[3, 9]$. Indicate the number and explain how it was obtained.
   Solution: Note that here $[3, 9] = [\bar{x} - S, \bar{x} + S]$. Using the Empirical rule the number of measurements in this intervals is **approximately** $1000 * 0.68 = 680$.

6. For 10 young patients, catheters were fed from a principal vein into the heart. The necessary catheter length and the patients' height measured.

Patient	1	2	3	4	5	6	7	8	9	10
Height (in inches)	42.8	63.5	37.5	39.5	45.5	38.5	43	22.5	37	23.5
Catheter Length (centimeters)	37	50	34	36	43	28	37	20	34	30

   (a) [5] Obtain the equation of the linear regression of Catheter Length on Height
   Solution: $Catheter\ Length = 9.5431 + 0.6447 * Height$

   (b) [5] Estimate the required Catheter Length for a patient whose Height is 36 inches.
   Solution: $estimated\ Catheter\ Length = 9.5431 + 0.6447 * 36 = 32.752$

   (c) [5] Obtain the equation of the linear regression of Height on Catheter Length
   Solution: $Height = -5.420 + 1.282 * Catheter\ Length$

   (d) [5] Estimate the height of a patient who requires a Catheter with length 22 centimeters.
   Solution: $estimated\ Height = -5.420 + 1.282 * 22 = 22.784$.