SOLUTIONS TO THE PRACTICE TEST FOR TEST 1 (STAT-461/561)

(1) (a) F, (b) F (the order is wrong), (c) F (it gets rid of selection bias but not the other types of biases such as non-response bias), (d) F, (e) F, (f) F, (g) T, (h) F, (i) T, (j) F, (k) T, (l) T, (m) F, (n) F

(2) (a) Continuous ratio-type quantitative (or numerical), (b) Interval-type quantitative, (c) Ordinal qualitative (or categorical), (d) Ordinal qualitative (or categorical), (e) Nominal qualitative (or categorical), (f) Discrete ratio-type quantitative, (g) Interval-type quantitative

(3) (a) Convenience sampling, (b) Systematic sampling, (c) Cluster sampling, (d) Stratified random sampling

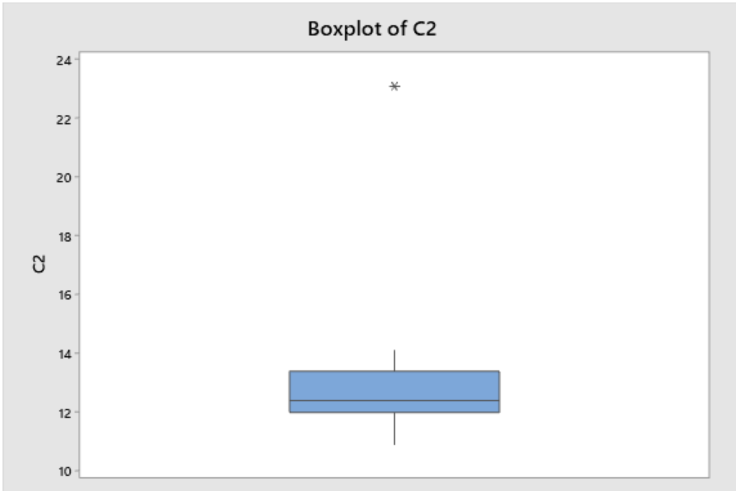
(4) Mean = 2.75, median = (0 + 1)/2 = 0.5, Q1 (the lower quartile) = the median of the data-values below the overall median = (0 + 0)/2 = 0, Q3 (the upper quartile) = the median of the data-values above the overall median = (5+6)/2 = 5.5, IQR = Q3 – Q1 = 5.5 – 0 = 5.5. The histogram will vary from person to person. Draw it yourself.

(5) (a) Stem-and-leaf plot: See below. (b) 5-point summary: Minimum = 10.9, Q1 = 12.0, Median = 12.4, Q3 = 13.4, Maximum = 23.1. (c) Boxplot: see below.

(6) If the data-value 23.1 suddenly becomes 13.1, the mean, the range and the standard deviation will change significantly, whereas the median, the IQR and the MAD will not, because the median, the IQR and the MAD are *robust* to (i.e. not easily changed by) changes in the two extremes of the dataset. Mean = 13.207 for the original data-set and = 12.54 for the changed data-set. Standard deviation = 2.851 for the original data-set and **= 0.813** for the changed data-set. Range = 23.1 – 10.9 = 12.2 for the original data-set and = 14.1 – 10.9 = 3.2 for the changed data-set

(7) Standard deviation of the data-set in #(5) is 2.851. The MAD is the median of the absolute deviations from the overall median 12.4. In other words, the MAD is the median of the numbers |12.4 – 12.4|, |12.2 – 12.4|, |13.4 – 12.4|, |10.9 – 12.4|, …, |12.6 – 12.4|, |11.9 – 12.4|, |23.1 – 12.4|, that is, the median of the numbers 0, 0.2, 1.0, 1.5, …, 0.2, 0.5, 10.7. The answer is MAD = 0.5

Boxplot with outlier detection for the dataset in #(5):



Stem-and-leaf plot for the dataset in #(5)

10| 9

11| 8 9

12| 0 1 2 2 4 6 7

13| 2 4 5

14| 1

15|

16|

17|

18|

19|

20|

21|

22|

23| 1

(8) Correlation 0.941, the equation of the best-fitting (i.e. least-squares) line Y = 0.561 + 0.5919 X, the coefficient of determination (r-squared) = 0.8836 or 88.36%. Since the best-fitting line explains as much as 88.36% of the overall variation in Y, the best-fitting line fits this dataset well.