STAT 2507 Solution-Lab Assignment # 1 Fall 2009

Total mark=100

Note: Since the data are random, the answers may slightly different

Type the following Minitab commands to generate 200 numbers to be saved in a vector called c2:

MTB> *set c1*

DATA> 1:200

DATA> *end*

MTB> *random 200 c11;*

SUBC> *normal 1000 50.*

MTB> *let c2=1000*log(c1)+c11* (c2 contains the data)

Think of these 200 numbers in c2 as the prices of a sample of 200 used cars, in dollars.

1. Draw a stem-and-leaf plot of these 200 prices and use your plot to answer the following questions:

(a)[2] The maximum price is $6300

(b)[2] The minimum price is $ 900

(c)[2] The median price is $ 5700

(d)[2] 20% of the prices are less than $ 4600

(e)[2] What is the shape of the distribution of the prices of used cars? Answer: skewed to the left

2. Use the 'describe' command to answer parts (a), (b), and (c) of Question 1.

Answer: (a)[2] $ 6332.5, (b)[2] $ 950.5, (c)[2] $ 5612.3, (d)[4] The average price of a used car is $ 5304.1. The standard deviation is $ 959.6.

3.

(a)[4] What proportion of the prices are within 2 standard deviations of the mean price (i.e., fall in the interval $\bar{x} \pm 2s$)? Answer: $5304.1 \pm 2(959.6) = (3384.9, 7223.3)$ 185/200

(b)[4] Answer part (a) using Tchebysheff's Theorem. At least 150/200.

(c)[4] Answer part (a) using the Empirical Rule. Approximately 190/200.

4. Now make the following transformation of the used car prices (that are in c2): (new price)=1.6× (price)+200. This can be done in Minitab by using the 'let' command:

MTB> *let c10=1.6*c2+200*

(a)[6] Use a dotplot (click on 'graph' and then a 'Dotplot') of the new prices in the vector c10 to find the two smallest values. They are approximately $ 1800 and $ 3000.

(b)[8] The z-scores corresponding to the two values in part (a) above are -4.49 and -3.70 respectively. Are they outliers? yes, since $|zscore| > 3$.

(c) Use a Boxplot (click on 'graph' and then a 'Boxplot') of the new prices in the vector c10 to answer the following questions:

(i)[3] the median is approximately equal to $ 9000.

(ii)[3] The interquartile range is approximately equal to $ 9800-7800=2000.

(iii)[3] The range of the data (new prices) is approximately equal to $ 10300-1800=8500.

(d)[5] Obtain a histogram of the new prices and comment on the shape of the distribution skewed to left.

Solve the following questions.

1. Identify each of the following variables as either quantitative discrete, quantitative continuous, or qualitative.

   (a)[2] Province or territory where a person lives "qualitative"

   (b)[2] The daily high temperature for the last four weeks "quantitative continuous"

   (c)[2] The amount of sugar consumed by Americans in one year "quantitative continuous"

   (d)[2] Leaves on a maple tree "quantitative discrete"

2. Construct a boxplot for these data and identify any outliers:[14]

   16, 19, 10, 12, 41, 12, 13, 12, 15, 16, 20, 17, 13, 20, 18, 19, 9, 9, 9

   Sol: min = 9, max = 41

   position of the first quartile=$0.25 * (19 + 1) = 5 \rightarrow Q_1 = 12$

   position of the second quartile=$0.5 * (19 + 1) = 10 \rightarrow$ median $= Q_2 = 15$

   position of the third quartile=$0.75 * (19 + 1) = 15 \rightarrow Q_3 = 19$

   $IQR = Q3 - Q1 = 19 - 12 = 7$ lower fence=12-1.5(7)=1.5 and upper fence=19+1.5(7)=29.5

3. Tornadoes cause many deaths each year in the United States. The values for the yearly number of deaths for the 54 years 1950 through 2003 are

   12, 14, 25, 27, 18, 30, 30, 31, 32, 33, 34, 24, 36, 29, 39, 29, 40, 43, 14, 46, 52, 51, 52, 53, 53, 55, 58, 59, 64, 64, 66,17, 67, 69, 70, 75, 73, 83, 84, 99, 94, 94, 98, 114, 122, 129, 130, 131, 159, 193, 230, 301, 366, 519

   (a)[5] Determine the intervals $\bar{x} \pm s, \bar{x} \pm 2s$, and $\bar{x} \pm 3s$.

   Sol:

   $\bar{X} = \frac{\sum X_i}{n} = \frac{4530}{54} = 83.89$ $s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} = \frac{819274 - \frac{(4530)^2}{54}}{53} = 8287.87$ $s = \sqrt{s^2} = \sqrt{8287.87} = 91.04$

$\bar{X} \pm s$	$\bar{X} \pm 2s$	$\bar{X} \pm 3s$
83.89 ± 91.04	$83.89 \pm 2(91.04)$	$83.89 \pm 3(91.04)$
(-7.15, 174.93)	(-98.19, 265.97)	(-189.23, 357.01)

   (b)[5] Find the proportion of the measurements that lie in each of these intervals.

   Sol:

$\bar{X} \pm s$	$\bar{X} \pm 2s$	$\bar{X} \pm 3s$
49/54=0.91	51/54=0.94	52/54=0.96

   (c)[3] How do the percentages obtained in part (a) compare with those given by the Empirical Rule? Should they be approximately the same? Explain.

Sol:

The proportions for Empirical rule are:

$\bar{X} \pm s$	$\bar{X} \pm 2s$	$\bar{X} \pm 3s$
0.68	0.95	0.997

4. The results of three flips for a **biased** coin are observed. The probability of a head occurring is 0.6. Consider the following events:

A: at least two heads are observed B: exactly one tail is observed

C: exactly two tails are observed

(a)[2] List an appropriate sample space S for this experiment.

(b)[3] Are the events of getting a head on flip independent for the 3 flip?

(c)[6] Find $P(A), P(B'), P(B \cap C), P(A \cap B), P(B|A), P(A \cup B)$

(d)[4] Is there a pair of mutually exclusive or independent events among A, B and C? If so which pairs are mutually exclusive or independent?

Sol:

a) [1] S={(HHH),(HHT),(HTH),(THH),(TTH),(THT),(HTT),(TTT)}

b) [1] Yes, event of a head on one flip **is independent** of event of a head on any other flip.

c) [3] $A = \{(HHT), (HTH), (THH), (HHH)\} \rightarrow P(A) = 3(0.6 \times 0.6 \times 0.4) + 0.6 \times 0.6 \times 0.6 = 0.648$

$B = \{(HHT), (HTH), (THH)\} \rightarrow P(B) = 3(0.6 \times 0.6 \times 0.4) = 0.432$

$C = \{(HTT), (THT), (TTH)\} \rightarrow P(C) = 3(0.4 \times 0.4 \times 0.6) = 0.288$

$B \cap C = \varnothing \rightarrow P(B \cap C) = 0,$

$A \cap B = B \rightarrow P(A \cap B) = 0.432$

$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.432}{0.648} = 0.67$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.648 = P(A)$

d) [2] A and C are M.E and also B and C are M.E and they are not independent. A and B are not independent and are not M.E, since $P(B|A) = 0.67 \neq P(B)$.