**STAT 2507 Lab Assignment # 1 Summer 2013**

**Last Name** ______________________ , **First** ______________

**Student #** ______________

**<u>Due</u>: June 10, 2013, prior to the start of class** **Total mark=100**

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 <u>prices</u> of a sample of 200 used cars, in dollars.**

1. Draw a stem-and-leaf plot of these 200 prices and <u>use your plot to answer</u> the following questions:
(a)[1] The maximum price is <u>$6300</u>
(b)[1] The minimum price is <u>$ 900</u>
(c)[1] The median price is <u>$ 5700</u>
(d)[1] 20% of the prices are less than <u>$ 4600</u>
(e)[1] What is the shape of the distribution of the prices of used cars? Answer: <u>skewed to left</u>

2. Use the 'describe' command to answer parts (a), (b), and (c) of Question 1.
Answer: (a)[1] <u>$ 6332.5</u>, (b)[1] <u>$ 950.5</u>, (c)[1] <u>$ 5612.3</u>, (d)[4] The average price of a used car is <u>$ 5304.1.</u> The standard deviation is <u>$ 959.6.</u>

3.
(a)[2] What proportion of the prices are within 2 standard deviations of the mean price (i.e., fall in the interval $\bar{x} \pm 2s$)? <u>Answer: $5304.1 \pm 2(959.6) = (3384.9, 7223.3)$185/200</u>
(b)[2] Answer part (a) using Tchebysheff's Theorem. <u>At least 150/200.</u>
(c)[2] Answer part (a) using the Empirical Rule. <u>Approximately 190/200 .</u>

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) [3] 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 $ ______ and $ 1800 and $3000.

(b)[4] 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)[1] the median is approximately equal to $ 9000.

(ii)[1] The interquartile range is approximately equal to $ 9800-7800=2000.

(iii)[1] The range of the data (new prices) is approximately equal to $ 10300-1800=8500.

(d)[3] Obtain a histogram of the new prices and comment on the shape of the distribution skewed to left.

5. The link click here gives an experiment of tossing two fair dice (one green and one red). Let $C$ be the event that the green die shows a number less than or equal to 2.

Let $D$ be the event that the red die shows a number less than or equal to 3.

(a) [2] (Press Reset) Obtain $P(C)$= 12/36

(b)[2] (Press Reset) Obtain $P(D)$= 18/36

(c)[2] (Press Reset) By clicking on the simple events in which the green die has a number at most 2 and the red one has a number at most 3, obtain $P(C \cap D)$= 6/36

(d) [2] (Use parts (b) and (c)) If you know that the red die showed a number less than or equal to 3, then obtain the probability that the green shows a number less than or equal to 2, by clicking on the relevant simple events in the applet?

I.e., what is $P(C|D) = 6/18$

(e) [2] What is your conclusion about the relation between the events $C$ and $D$? Independent as $P(C|D) = P(C)$

6. The values (1 to 10) in the first column of the Open Office file (see file: Data.xls) are the values that random variable $x$ takes (copy and paste them in C1 of your Minitab worksheet) and the second column contains $p(x)$ (copy and paste this column from Open Office file in C2 of your Minitab worksheet)

(a) [3] Obtain the mean of $x$, i.e., $\mu = E(x)$ by $Calc \rightarrow Calculator \rightarrow sum(C1 * C2)$. What is $\mu$? 1.998047 by clicking $Calc \rightarrow Calculator \rightarrow sum((C1 * *2) * C2) - (sum(C1 * C2)) * *2$. obtain $Var(x) = \sigma^2$. What is $\sigma^2$ ? 1.962886 What is $\sigma$? 1.40103 (b) [3] Calculate the interval $(\mu - 2\sigma, \mu + 2\sigma)$? $(1.998047 - 2(1.40103), 1.998047 - 2(1.40103)) = (-0.804013, 4.800107)$

(c) [3] What values of $x$ are inside the interval in part (c)? 1, 2, 3, and 4

**Solve the following questions.**

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

   (a)[1] The brands of ice cream that you purchase (qualitative)

   (b)[1] The daily high temperature for the last four weeks, (quantitative continuous)

   (c)[1] The amount of sugar consumed by Americans in one yea, (quantitative continuous)

   (d)[1] Number of brothers and sisters you have, (quantitative discrete)

2. Construct a boxplot for these data and identify any outliers:[10]

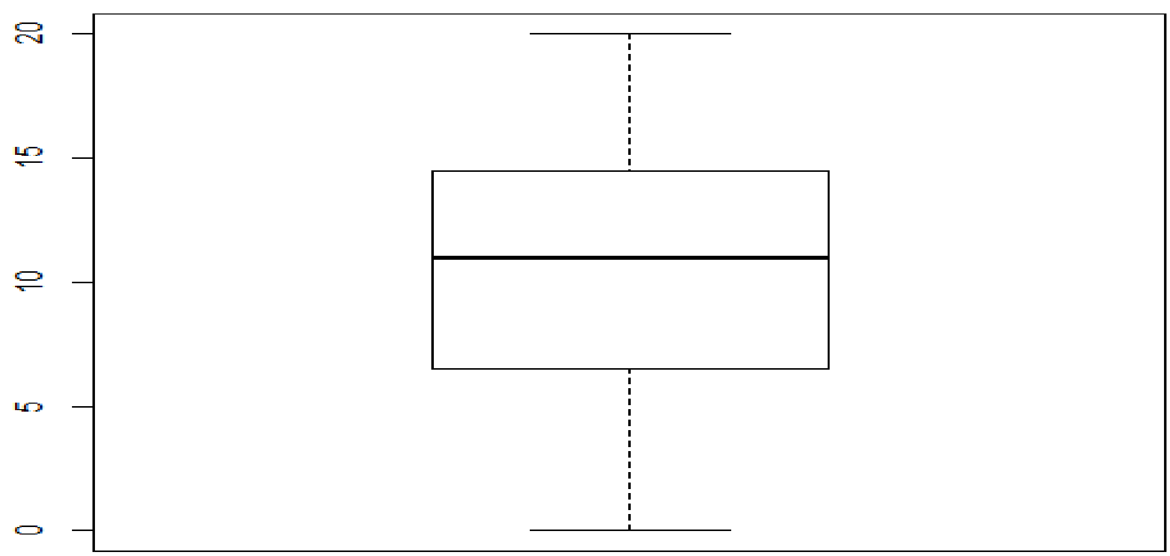
   6, 19, 0, 2, 11, 12, 13, 12, 5, 16, 2, 7, 13, 20, 18, 19, 9, 9, 9

   min = 0, max = 20

   position of the first quartile=$0.25 * (19 + 1) = 5 \rightarrow Q_1 = 6$

   position of the second quartile=$0.5 * (19 + 1) = 10 \rightarrow$ median $= Q_2 = 11$

   position of the third quartile=$0.75 * (19 + 1) = 15 \rightarrow Q_3 = 16$




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

   15, 24, 25, 27, 28, 30, 30, 31, 32, 33, 34, 34, 36, 39, 39, 39, 40, 43, 44, 46, 50, 51, 52, 53, 53, 55, 58, 59, 60, 64, 66,67, 67, 69, 70, 73, 73, 83, 84, 89, 94, 94, 98, 114, 122, 129, 130, 131, 159, 193, 230, 301, 366, 519

   (a)[6] Determine the intervals $\bar{x} \pm s, \bar{x} \pm 2s$, and $\bar{x} \pm 3s$.

$\bar{X} = \frac{\sum X_i}{n} = \frac{4645}{54} = 86.02$ $s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} = \frac{825199 - \frac{(4645)^2}{54}}{53} = 8031$ $s = \sqrt{s^2} = \sqrt{8031} = 89.62$

| $\bar{X} \pm s$ | $\bar{X} \pm 2s$ | $\bar{X} \pm 3s$ |
|---|---|---|
| $86.02 \pm 89.62$ | $86.02 \pm 2(89.62)$ | $86.02 \pm 3(89.62)$ |
| (-3.6, 175.64) | (-93.22, 265.26) | (-182.84, 354.88) |

(b)[4] Find the proportion of the measurements that lie in each of these intervals.

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

(c)[4] How do the percentages obtained in part (a) compare with those given by the Empirical Rule? Should they be approximately the same? Explain.

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. Suppose that $P(A) = 0.6$, $P((A \cup B)^c) = 0.2$, $P(C)$=0.7 and $P(D)$=0.4.
(a) [2] Can $C$ and $D$ be mutually exclusive? Why? Explain. No, since $0.7 + 0.4 = 1.1 > 1$.

(b) [2] If the events $A$ and $B$ are independent then $P(B)$= 0.5.
Since $P(A \cup B) = 1 - 0.2 = 0.8$, if $A$ and $B$ are independent, then

$$P(A \cup B) = P(A) + P(B) - P(A)P(B) \to 0.8 = 0.6 + P(B) - 0.6P(B) \to P(B) = 0.5$$

(c) [2] If the events $A$ and $B$ are mutually exclusive then $P(B)$= 0.2.

Since $P(A \cup B) = P(A) + P(B) - 0 \to 0.8 = 0.6 + P(B) \to P(B) = 0.2$.

.

5. [16] In a particular city, airport A handles 50% of all airline traffic, and airports B and C handle 30% and 20%, respectively. The detection rates of weapons at the three airports are 0.9, 0.5, and 0.4, respectively. If a of passenger at one of the airports is found to be carrying a weapon through the boarding gate, what is the probability that the passenger is using airport A?

Let D='detected'. Use the Bayes' formula to write

$$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)} = \frac{(0.9)(0.5)}{(0.9)(0.5)+(0.5)(0.3)+(0.4)(0.2)} = 0.662$$