STAT 3502B

Solution-Test # 1

1. Consider the following data,

   80, 87, 74, 63, 179, 65, 84, 72, 64, 82

   such that $\sum_{i=1}^{10} X_i = 850$ $\sum_{i=1}^{10} X_i^2 = 82740$

   a. [2] What are mean and median for this set of data?

   b. [2] What are the values of sample variance and sample standard deviation?

   c. [1] For this set of data, mean better represent the center of data or median? Explain.

   d. [2] Construct a boxplot for these data.

   Sol:

   $\bar{X} = \frac{\sum_{i=1}^{10}}{n} = \frac{850}{10} = 85$

   Ordered data: 63, 64, 65, 72, 74, 80, 82, 84, 87, 179

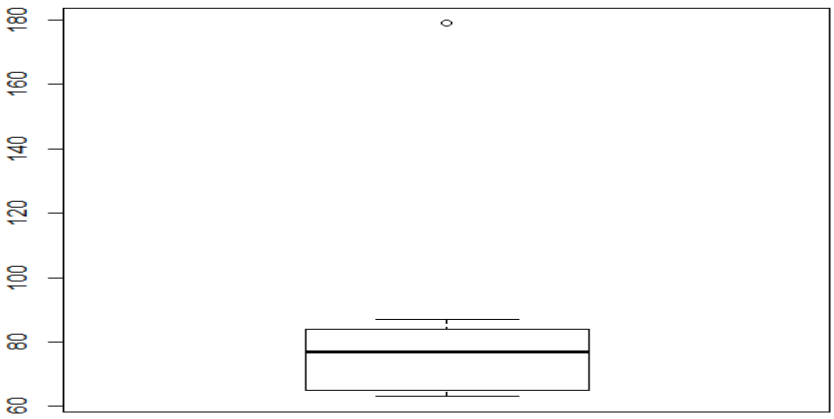
   P.O.median=$\frac{11}{2} = 5.5$, then $\tilde{X} = \frac{74+80}{2} = 77$

   b) $s^2 = \frac{\sum_{i=1}^{10} X_i^2 - \frac{(\sum_{i=1}^{10} X_i)^2}{n-1}}{n} = 1165.556$, and $s = \sqrt{1165.556} = 34.14$

   c) Median, since outlier doesn't have effect on median.

   d) Construct a boxplot for these data.

   Solution: Min=63, Max=179, lower-fourth=65, upper-fourth=84, $f_s = 84 - 65 = 19$



2. 60% of students in an engineering program are from Ottawa, 30% from Kingston and 10% from Toronto. 10% of Ottawa students and 15% of Kingston and 10% of Toronto students drop out of the program,

   a. [3] What is the probability that a randomly selected student will drop out?

   b. [3] If a student drop out, what is the probability that he/she was from Toronto?

   Sol: D: drop out, O: Ottawa, K: Kingston, T: Toronto

   a) $P(D) = P(D|O)P(O) + P(D|K)P(K) + P(D|T)P(T) = 0.6(0.1) + 0.15(0.3) + 0.1(0.1) = 0.115$

   b) $P(T|D) = \frac{P(D|T)P(T)}{P(D)} = \frac{0.1(0.1)}{0.115} = 0.067$

3. A box of candy contains 30 pieces. Twenty-six are made of chocolate and four are made of vanilla.

   a. [3] Five pieces are selected at random, without replacement. What is the probability that four of them are chocolates?

   b. [3] What would be the answer to part (a) if the five pieces are selected at random with replacement (i.e., a selected piece is put back in the box before the next selection is made)?

   Sol:

   a) X: # of chocolate candy and has hypergeometric distribution with $N = 30, M = 26, n = 5$, so

   $P(X = 4) = \frac{\binom{26}{4}\binom{4}{1}}{\binom{30}{5}} = 0.4196$

   b) X: # of chocolate candy and has binomial distribution with $n = 5, p = 26/30$ so $P(X = 4) = 5(\frac{26}{30})^4(\frac{4}{30}) = 0.3761$

4. Suppose that X is a random variable with probability distribution: $P(X = k) = 0.02k$, where k takes the values 8, 10, 12, and 20.

   a. [3] Find the cumulative distribution function (cdf) for X.

   b. [3] Find $E(Y)$ and $V(Y)$ if $Y = 2X - 10$

   Sol: The pmf is

x	8	10	12	20	Total
p(x)	0.16	0.2	0.24	0.4	1

   a)

   The cdf is

$$F(x) = \begin{cases} 0 & x < 8 \\ 0.16 & 8 \leq x < 10 \\ 0.36 & 10 \leq x < 12 \\ 0.6 & 12 \leq x < 20 \\ 1 & x \geq 20 \end{cases}$$

b) Y is a linear function of X, so we need to find $E(X)$ and $V(X)$:

$E(X) = \sum xP(x) = 8(0.16) + 10(0.2) + 12(0.24) + 20(0.4) = 14.16$, so $E(Y) = 2E(X) - 10 = 2(14.16) - 10 = 18.32$

$V(X) = \sum x^2P(x) - (E(X))^2 = 64(0.16) + 100(0.2) + 144(0.24) + 400(0.4) - 14.16^2 = 24.29$, so $V(Y) = 4V(X) = 4(24.29) = 97.16$