Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

Name of company	Measure X
Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%
Warner-Lambert	35.00%

https://github.com/venkateshblks/Basic-Statistics-Level-2-Set-1

2.

Box-Plot of variable 'X'
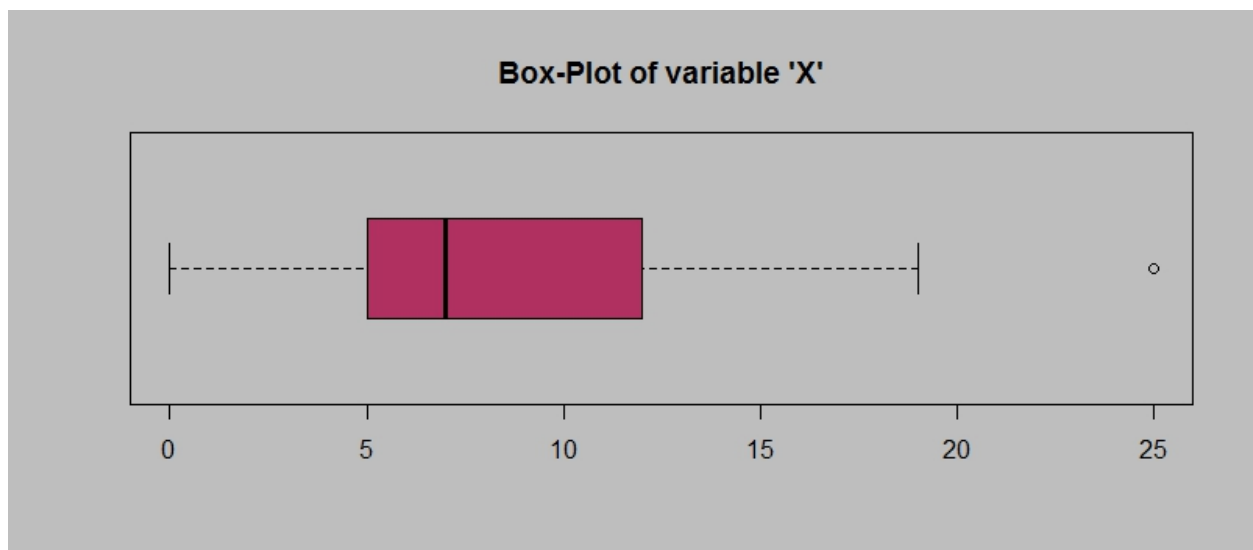
Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

First quartile(Q1)=5

Third quartile(Q3)=12

Inter-quartile range(IQR)=Q3-Q1=12-5=7

First quartile means 25% of the data points are less than or equal to Q1

Third quartile means 75% of the data points are less than or equal to Q3

Second quartile (Q2)=7 means 50% of the values falling below it and 50% above it and it is median
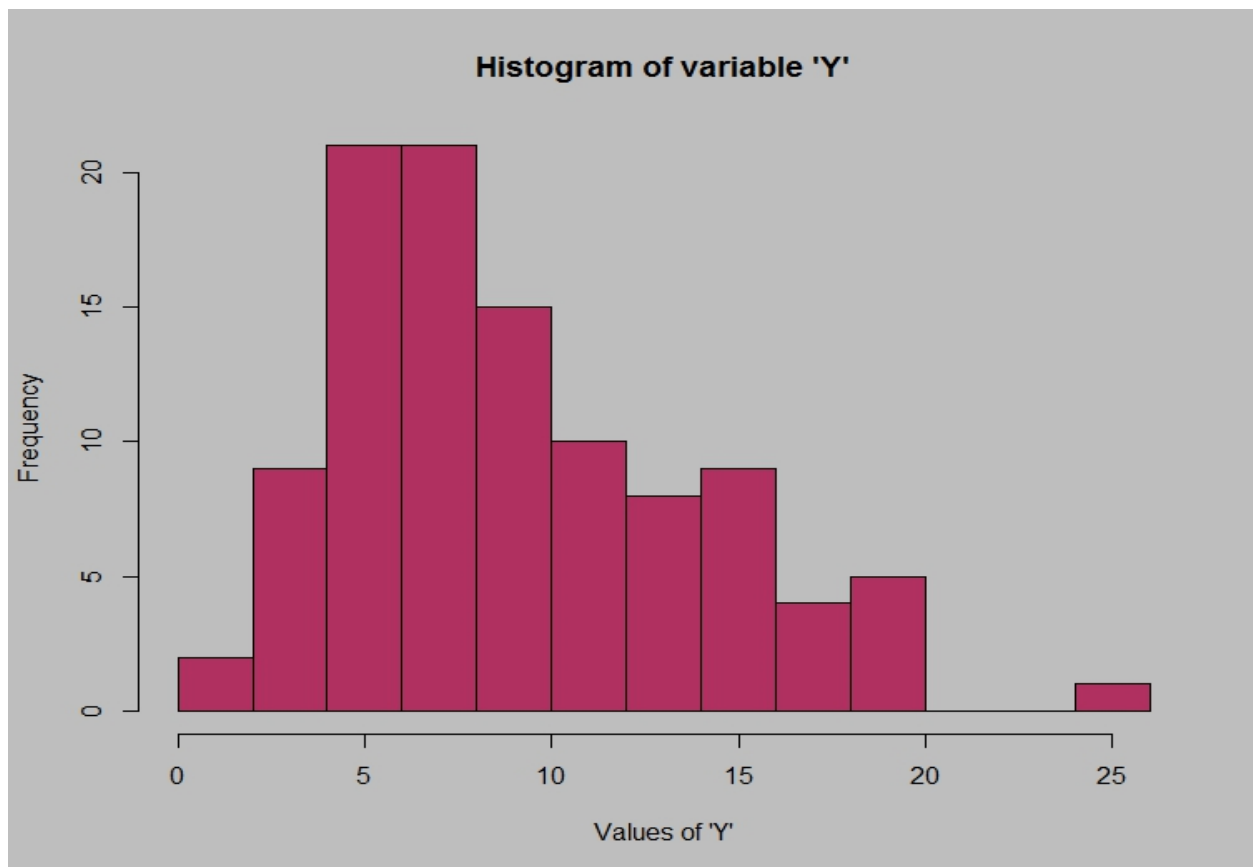
(ii)   What can we say about the skewness of this dataset?

Positive skewness distribution also known as right skewness

(iii)   If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

There would be no outliers

3.

Histogram of variable 'Y'

Answer the following three questions based on the histogram above.

(i)   Where would the mode of this dataset lie?

Between 4 to 8 approximately

(ii)   Comment on the skewness of the dataset.

Positive skewness distribution also known as right skewness so mean>median

(iii)   Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

Both plots are (positive)right skewed and both have outliers and we can find IQR from boxplot and mode from histogram

4. AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that "could happen." Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

   Probability of call misdirected =1/200
   Probability of call not misdirected=1-1/200=199/200
   the probability that at least one in five attempted telephone calls reaches the wrong number=1-none of the call reaches the wrong number
   => 1-nCr(p^r)(q^n-r)
   => 1-5C0(1/200)^0(199/200)^5-0
   = 1-(199/200)^5
   = 0.02475124687

5. Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

x	P(x)
-2,000	0.1
-1,000	0.1
0	0.2
1000	0.2
2000	0.3
3000	0.1

   (i)   What is the most likely monetary outcome of the business venture?
         the highest probability, which is $2000
   (ii) Is the venture likely to be successful? Explain
         venture is likely to be successful because the probabilities associated with positive outcomes is 0.6 (0.2+0.3+0.1) which is greater than the probabilities associated with negative outcomes 0.2(0.1+0.1) So there's a higher chance of making a profit
   (iii) What is the long-term average earning of business ventures of this kind? Explain
         Long-term average earning= (-2000 * 0.1) + (-1000 * 0.1) + (0 * 0.2) + (1000 * 0.2) + (2000 * 0.3) + (3000 * 0.1)
         = -200 - 100 + 0 + 200 + 600 + 300 = 800
         So mean (average earning)= 800
         Long-term average earning = $800
   (iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

higher standard deviation indicates higher variability or risk in the potential returns of the venture. Standard deviation is the square root of the variance so Higher Variance means more chances of risk

var(x)= ((x-mean)^2) * P(x)   where mean is 800

variance= (-2000-800)^2 *0.1 + (-1000-800)^2 *0.1 + (0-800)^2 *0.2

+ (1000- 800)^2 *0.2 +(2000-800)^2 *0.3 +(3000-800)^2 *0.1

=2160000

The standard deviation is 1469.69

Since the deviation is high so the risk involved is also very high