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%

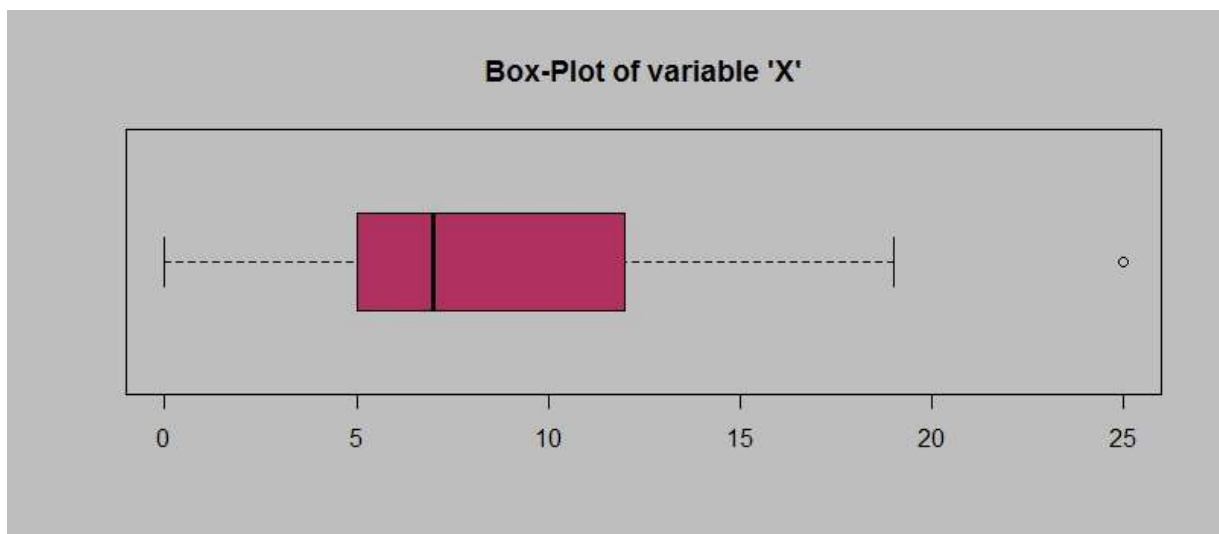Total number of companies = 15
Total measure sum(X)= 544.91
Mean(μ) =544.91/15=33.
Median = 0.2671
Std.Dev(σ)= 16.37
Variance(σ^2)= 268

2.

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.

Ans. Approximately

      (1st Quantile Range) Q1 = 5

      (3rd Quantile Range) Q3 = 12,

      Median (Second Quartile Range) = 7

      (Inter-Quartile Range) IQR = Q3 – Q1 = 12 – 5 = 7

   **Second Quartile Range is the Median Value**

(ii)   **What can we say about the skewness of this dataset?**

      **Ans. Dataset is right / positively skewed, median is towards the left side it is
       not normal distribution**

(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?**

      **Ans: Then the outlier which lies on the value 25 will be in the boxplot towards
       the lower extreme. there would be no Outliers on the given dataset because
       of the outlier the data had positive skewness it will reduce and the data will
       normal distributed**

3.
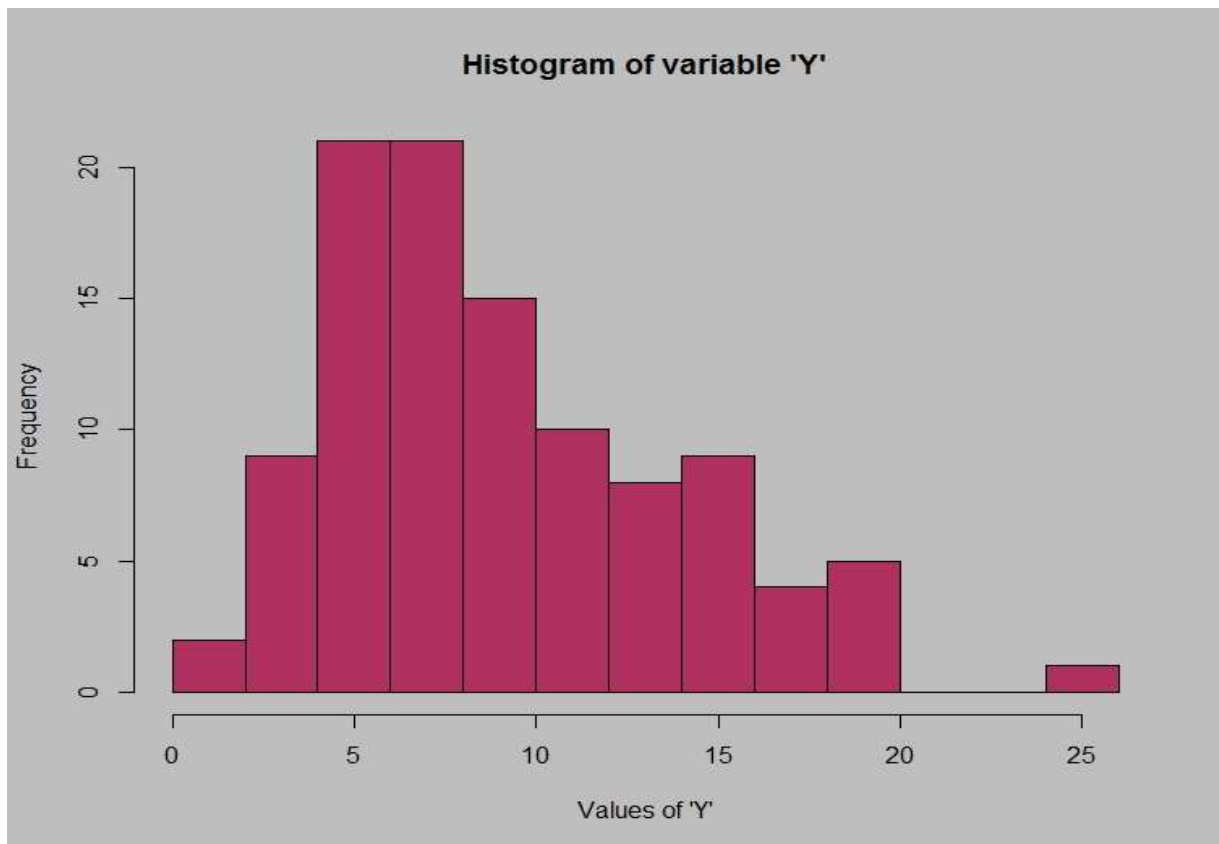
Answer the following three questions based on the histogram above.

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

(ii)   Comment on the skewness of the dataset.

(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.

Ans:(i) The mode of this data set lies in between 5 to 10 and approximately between 4 to 8

(ii) Right-Skewed. Mean>Median>Mode

(iii) The histogram shows whether the data is symmetric or not and the box will show the outlier and IQR of the given data. We can see median in boxplot and mode in histogram. Histogram provides the frequency distribution so we can see how many times each data point is occurring however boxplot provides the quantile distribution i.e., 50% data lies between 5 and 12. Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an outlier.

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.)

prob of call misdirected = 1/200
prob of call not Misdirected = 1-1/200 = 199/200
prob for at least 1/5 attempted calls reaches the wrong number
Number of Calls = 5
n = 5 p = 1/200 q = 199/200
at least one in five attempted calls reach wrong number
= 1 - none of the call reaches the wrong number
= 1 - P(0)
=1 - $^5C_0(1/200)^0(199/200)^{5-0}$

P(1) = 1 - (199/200)^5= 0.02475

Ans: The probability that at least one in five attempted telephone calls reaches the wrong

number =   0.2475

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?

    (ii)   Is the venture likely to be successful? Explain

    (iii)   What is the long-term average earning of business ventures of this kind? Explain

    (iv)   What is the good measure of the risk involved in a venture of this kind? Compute this measure

(I)   Ans: most likely monetary outcome of the business venture is 2000 (prob is 0.3 >others)

(II)   Yes, prob that the venture will make more than 0 or a profit
p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit

(III) long-term average earning of business ventures= 0.8*(sigmax)=0.9*1000-=900
MINIMUM 900 $ WE EARN IN THIS SUCCESSFUL BUSINESS

(IV) good measure of the risk involved= PROBABILTY OF LOSS(-2000,-1000)
p(LOSS)=0.1+0.1=0.2=>20% LOSS MAY HAPPEEN=risk involved