Topics: Descriptive Statistics and Probability
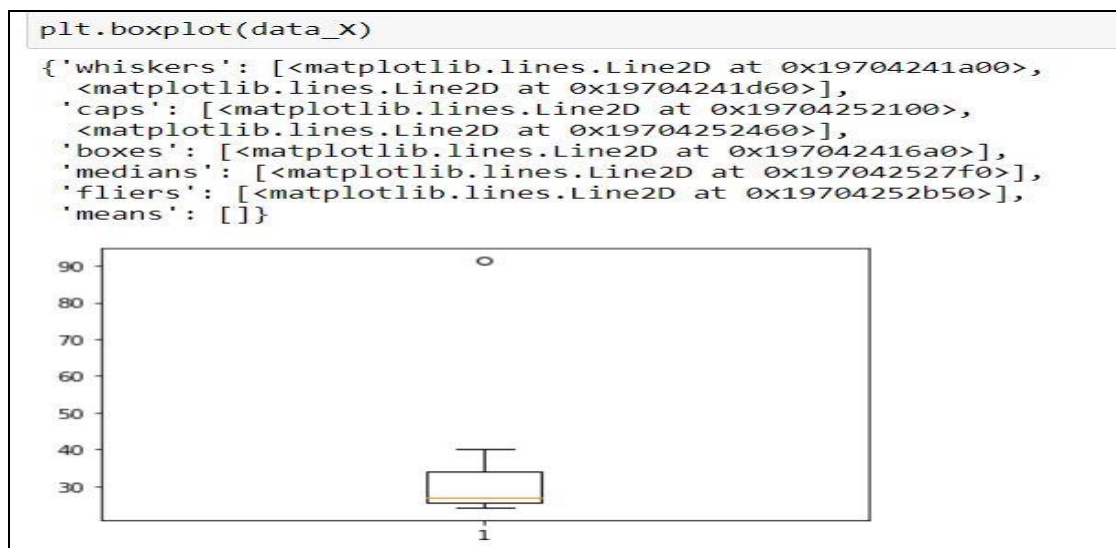
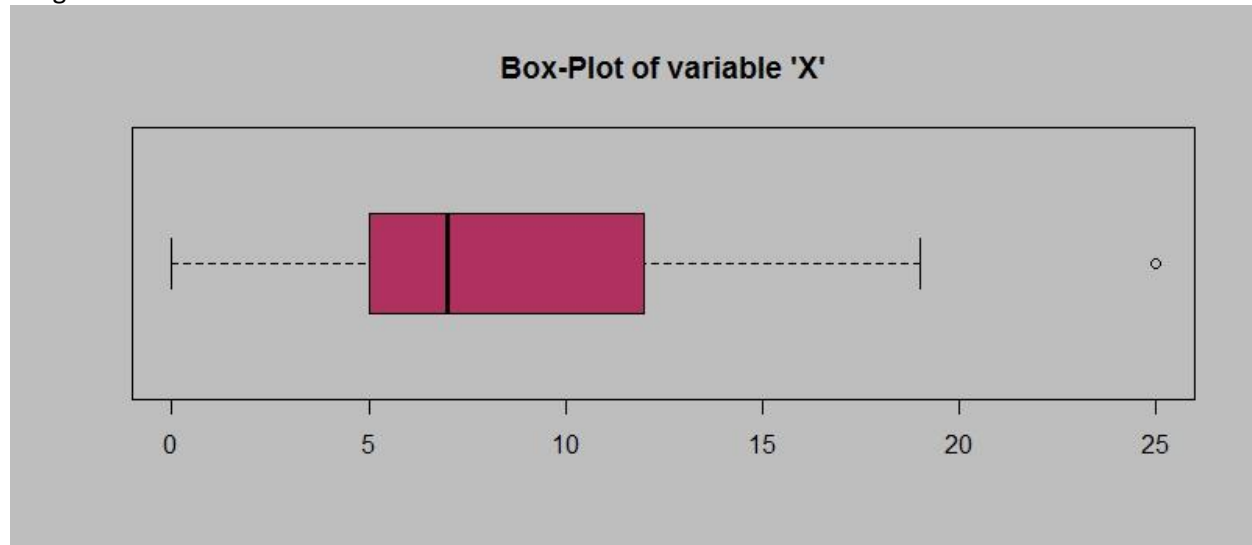1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

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%

Answer = mean = 32.2% variance = 268 std dev = 16.94 outlier = morgan Stanley 91.36%

```
plt.boxplot(data_X)
{'whiskers': [<matplotlib.lines.Line2D at 0x19704241a00>,
  <matplotlib.lines.Line2D at 0x19704241d60>],
 'caps': [<matplotlib.lines.Line2D at 0x19704252100>,
  <matplotlib.lines.Line2D at 0x19704252460>],
 'boxes': [<matplotlib.lines.Line2D at 0x197042416a0>],
 'medians': [<matplotlib.lines.Line2D at 0x197042527f0>],
 'fliers': [<matplotlib.lines.Line2D at 0x19704252b50>],
 'means': []}
```


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.

Answer – the IQR is 8. 50 percent of the data lies within IQR.

(ii) What can we say about the skewness of this dataset?
      Answer – positively skewed

(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?

Answer = the data point 25 is outlier so changing that do not affect the boxplot. Mean can be changed by a smaller margin.

Histogram of variable 'Y'
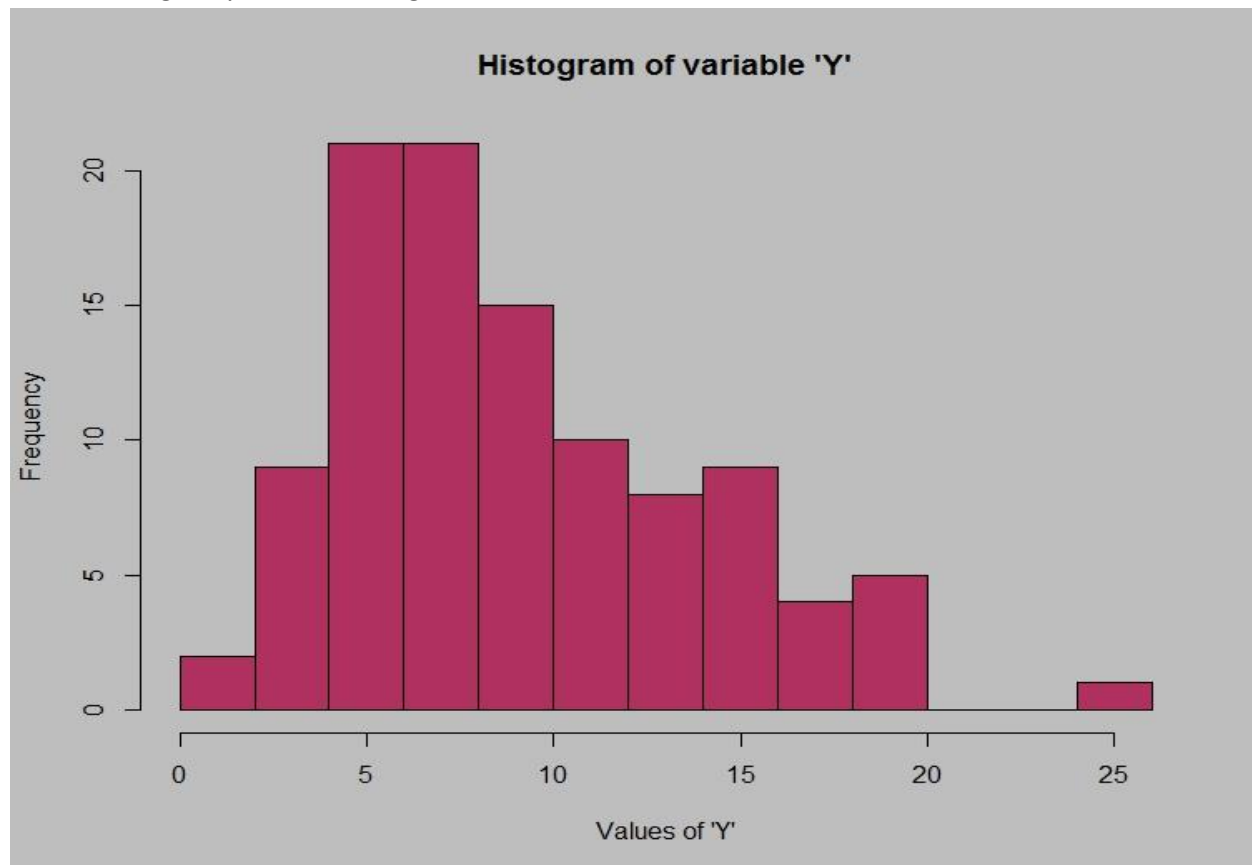
Answer the following three questions based on the histogram above.

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

Answer = the mode lies between 4 to 8 where the frequency is highest.

(ii)   Comment on the skewness of the dataset.

answer – positively skewed. More values to the right.

(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.

Answer - we plot datasets to visualize the mean median mode skewness outliers. Histograms can show the probability distribution of the data set.

3. 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.)

Answer = Probability of call getting mistaken $P(E) = 1/200$

Prob of call not getting mistaken = $1 – P(E) = 199/200$

No of attempts = 5, Probability that atleast one in 5 attempts mistaken calls $P(X) = 1 – $ (prob that no calls mistaken in 5 attempts)

$P(X) = 1 – ((199/200)*(199/200) (199/200)*(199/200) (199/200))$

 $= 1 –((199/200)**5)$

 $= 1 - 0.975248753121875$

$P(X) = 0.02475124687812502$

4. 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?

Answer = 2000 because it has highest prob

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

Answer = yes if if $P(x>0)$ is greater than $P(x<=0)$

$P(x>0) = 0.2+0.3+0.1 = 0.6$

$P(x<=0) = 0.2+0.1+0.1 = 0.4$

Is venture Is likely to be successful.

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

Answer = we need to calculate the expectation

$E(x) = (3000*0.1)+(2000*0.3)+(1000*0.2)+ (0*0.2)+ (-1000*0.1)+ (-2000*0.1)$

  $= 300 + 600 + 200 + -100 + -200$

$E(x) = 800$

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

Answer = standard deviation would be a good measure. Std dev = 1470