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%

Ans. There is one outlier in the data which is obtained from the box plot 91.36% is the outlier in the data, which can be seen from the box plot. Mean is 33.2713, standard deviation is 16.9454, and variance is 287.1466.

2.

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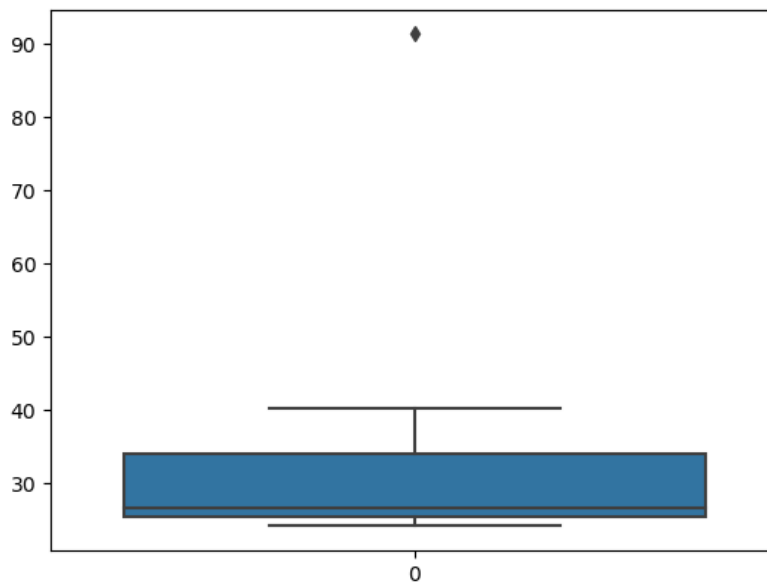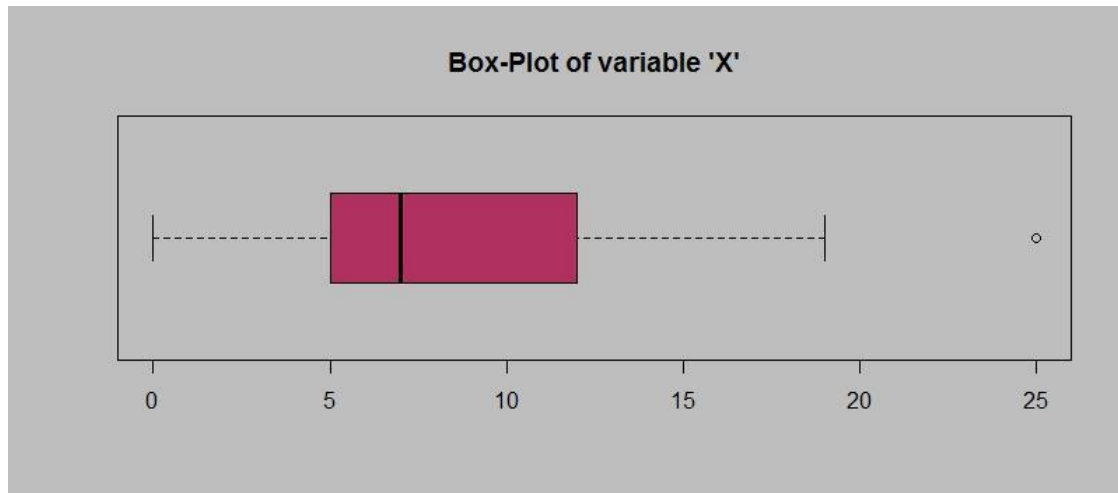
Answer the following three questions based on the box-plot above.

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

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

(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. For the box plot shown in the figure,

(i) IQR = Q_3 - Q_1 = 12 - 5 = 7. IQR shows the range in which the middle 50% of the data lies (25% to 75% in the data set).

(ii) The data set is right skewed or positively skewed.

(iii) If it was found that the data point 25 is actually 2.5, there will not be any outlier, as the calculated value of LB is -5.5 (= 5 - 1.5 x 7) which is lower than the minimum value (0) in this case. However, the Q1, Q2, Q3, LB, and UB will be slightly affected due to inclusion of one more data point. But magnitude of this change in values will depend on the total number of data points in the data set.

3.

Histogram of variable 'Y'
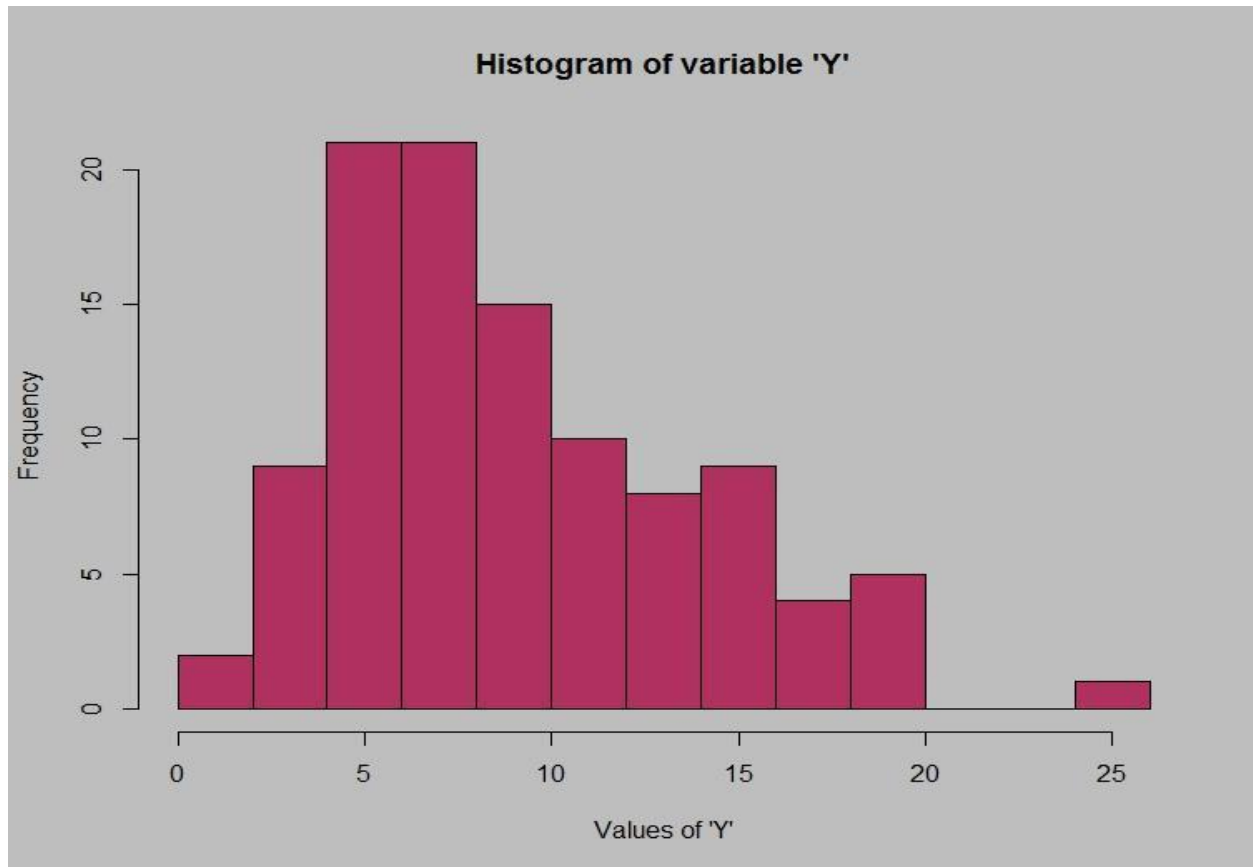
Answer the following three questions based on the histogram above.
(i)   Where would the mode of this dataset lie?
(ii) Comment on the skewness of the dataset.
(iii) Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

Ans. (i) The mode of the data set lies in the location where bins 2 and 3 are located. i.e., for y-values 4-8.
(ii) The data set is positively skewed or right skewed.
(iii) The box plot and histogram combinedly shows us that the data set is positively skewed, with an outlier lying at y-value of 25. The histogram shows us how the distribution looks like and the box plot confirms the distant occurrence of value 25 in the y-axis as 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.)

> Ans. p = 1/200, q = 199/200, n = 5. For probability for at least one in five attempted calls to reach the wrong number, we can say that probability is 1 - probability that no calls reach the wrong number.
> So, required probability, P = $1 - {}^{n}C_{r}\,(p)^{r}\,(q)^{(n-r)} = 1 - {}^{5}C_{0}\,(1/200)^{0}\,(199/200)^{5}$
> $$= 1 - (199/200)^{5} = 1 - 0.975248753121875$$
> $$= 0.02475124687812502$$

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

> Ans. (i) Since the highest probability is associated with $2000 (probability of 0.3), the most likely monetary outcome of the business venture is $2000
>
> (ii)For being successful, the probability is 0.2+0.3+0.1 = 0.6. But for being loss, the probability is 0.1+0.1 = 0.2. Since the probability of being a success is more,tThis venture is likely to be successful.
>
> (iii) E(X) = $\sum\limits_{i=1}^{n} x.P(x)$

= -2000 x 0.1 + -1000 x 0.1 + 0 x 0.2 + 1000 x 0.2 + 2000 x 0.3 + 3000 x 0.1

= -200 -100 + 0 + 200 + 600 + 300

= 800

The long term average earning of business ventures of this kind is $800

(iv) Variance = $E(X^2) - (E(X))^2$

$$E(X^2) = \sum_{i=1}^{n} X^2 \cdot P(X)$$

= 2000^2 x 0.1 + 1000^2 x 0.1 + 0 + 1000^2 x 0.2 + 2000^2 x 0.3 + 3000^2 x 0.1

= (4 + 1 + 0 + 2 + 12 + 9) x 10^5

= 2800000

So, variance = 2800000 - 800^2 = 2800000 - 640000 = 2160000

Since variance is very high, hence risk is high.