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)

> boxplot(Book1$`Measure X`,horizontal = T)
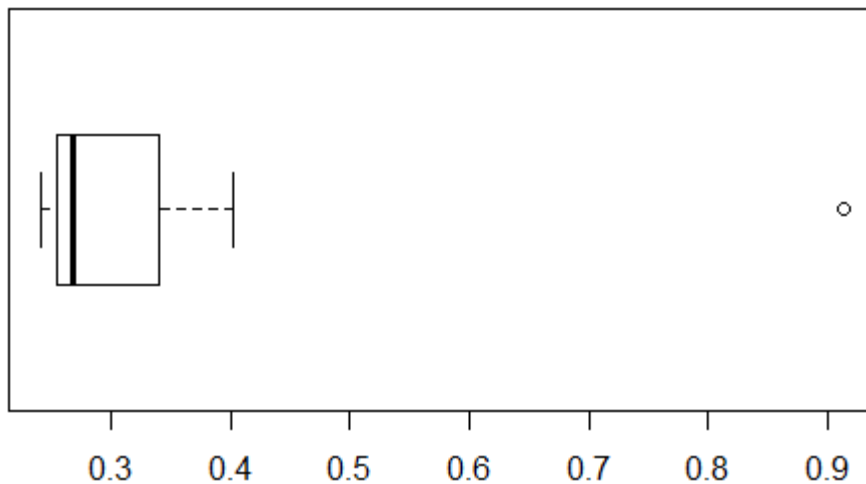
Outliers: Morgan Stanley   91.36%
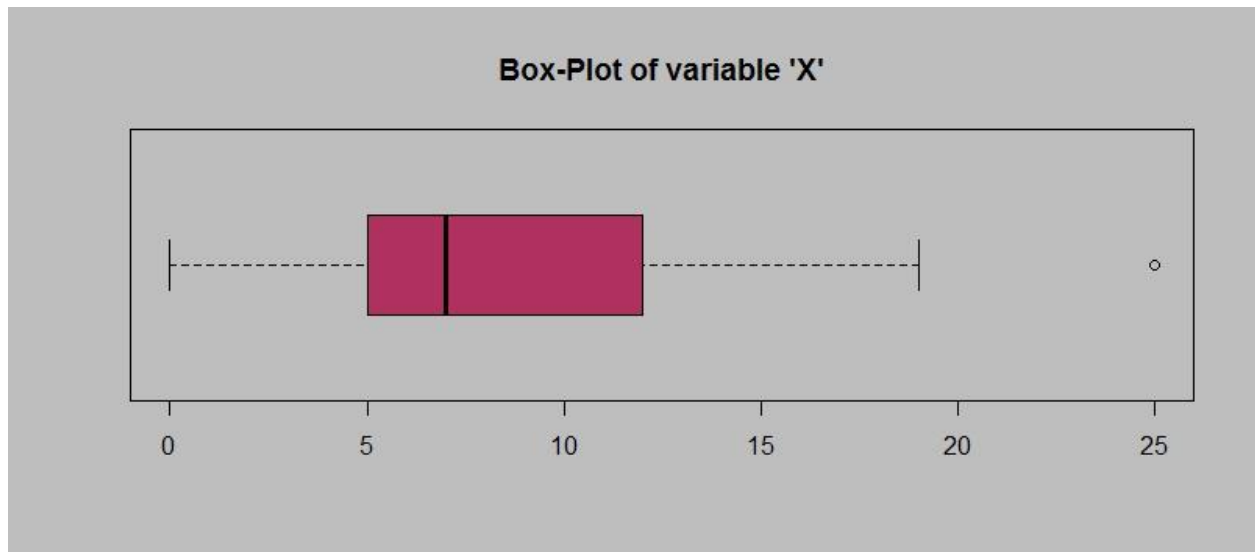
> mean(Book1$`Measure X`)
[1] 0.3327133
> var(Book1$`Measure X`)
[1] 0.02871466
> sd(Book1$`Measure X`)
[1] 0.169454

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.

Ans)   Inter-quartile range is the range between upper quartile (Q3) and lower quartile (Q1).

IQR= Q3-Q1= 12-5 = 7
50% of the data lies between IQR.

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

Ans)   From the above boxplot we can say that the distribution of X is right-skewed or 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?

Ans)   if it was found that the data point is actually 2.5 instead of 25, the outlier in the boxplot will be removed.
Whether the median shifts or not depends on the size of the data.
It will reduce the right skewness of the data.

3.

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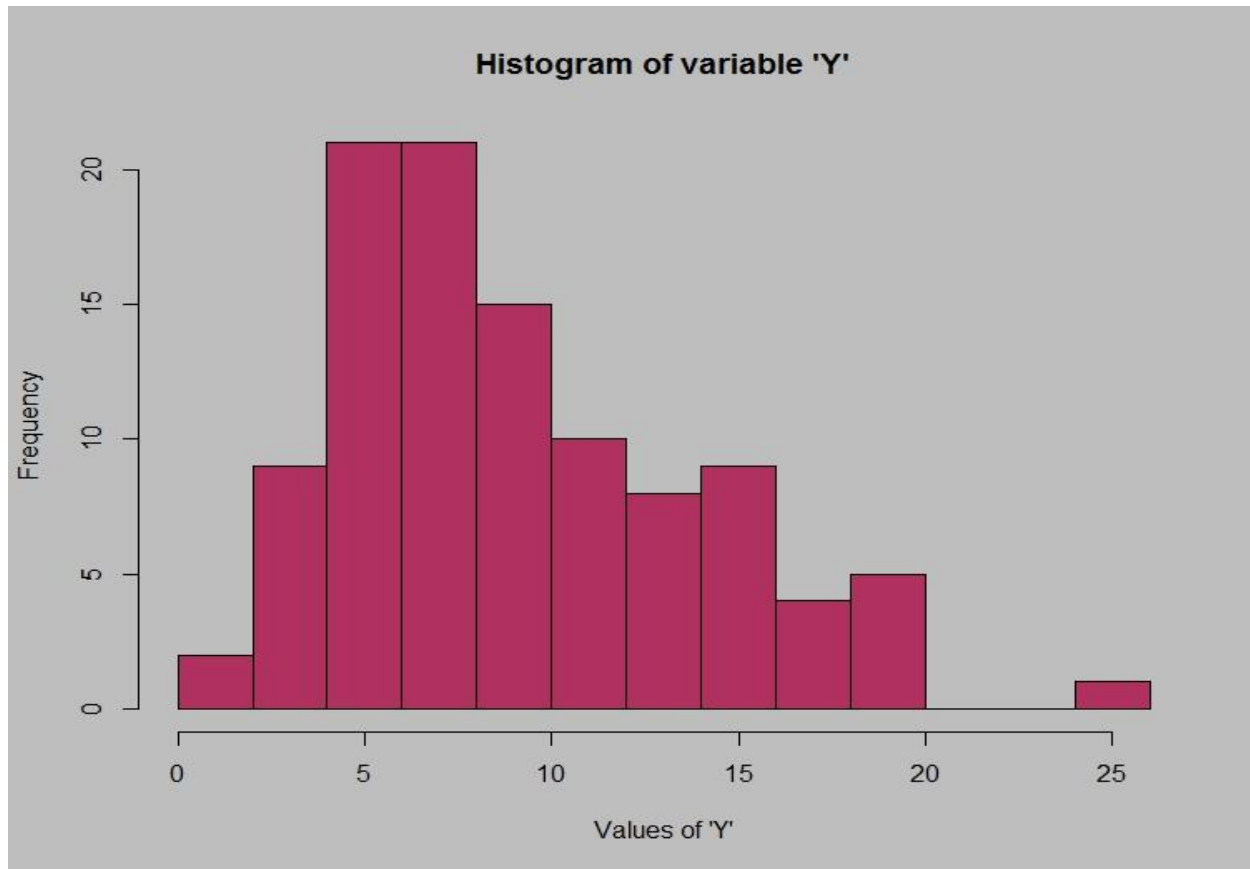
Answer the following three questions based on the histogram above.

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

Ans)   We need to have actual data to get the exact value of the mode. The mode can lie between 4 and 10 because there are many values in this range but this is just an assumption. The 2 bars of the same height doesn't indicate mode every time.

(ii)   Comment on the skewness of the dataset.

Ans)   It is right skewed or +ve skewed.

(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)   from the above histogram and barplot we can confirm an outlier at 25 in Y value. Both the plots indicate the +ve skewness of the dataset.

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)   let us consider the probability of 1 call misdirected out of 200 as event A.
Probability of occurring of event A= 1/200
P(A)= 1/200
Probability of having at least one successful call will be
1-P(A)= 1-1/200= 199/200= 0.967
As every event is independent of other event the probability will be
1- $(0.967)^5$
0.02475 = 2% chance.

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?

Ans)   The most likely outcome of this business venture is a return of $2000 as it has the highest probability of occurrence.

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

Ans)   success of the venture can be defined in multiple ways. But based on the data provided, we can look at positive returns as a measure of success.
The probability distribution gives us an idea about the long-term chances of earning given values of returns (indicated by x). therefore, there is a 60% probability that the venture would be successful. (Note: 0.2+0.3+0.1=0.6=>0.6*100=>60%).

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

Ans)   From the above question requirement we have to consider similar business ventures of this type whose distribution of the returns is similar to this venture. In that case we say that the expected value of returns to this particular venture is the required average.
(-2000*0.1)+(-1000*0.1)+(0*0.2)+(1000*0.2)+(2000 *0.3)+(3000*0.1)=800
Therefore the long-term average earning for these type of ventures would be around $800.

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

Ans)   Risk stems from the possible variability in the expected returns. Therefore a good measure to evaluate the risk for a venture of this kind would be variance or standard deviation of the variable X.

> sd(ex$x)
[1] 1870.829

> var(ex$x)
[1] 3500000
The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky.