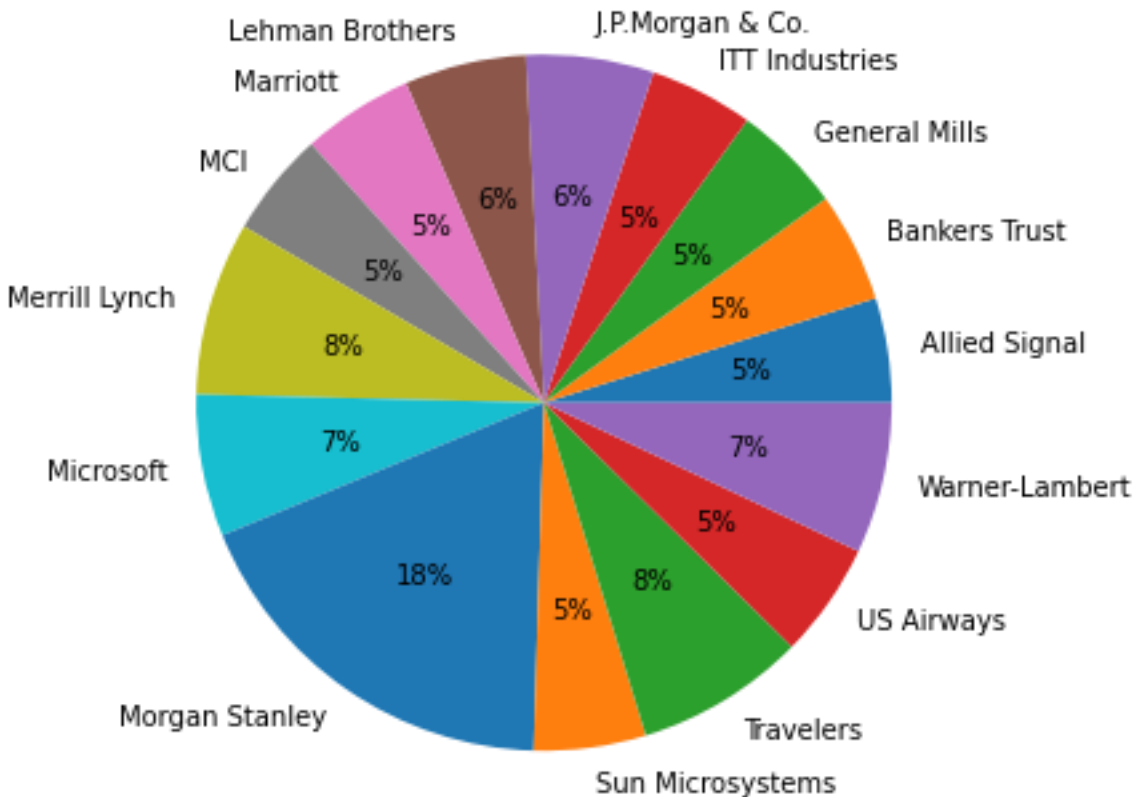


Topics: Descriptive Statistics and Probability

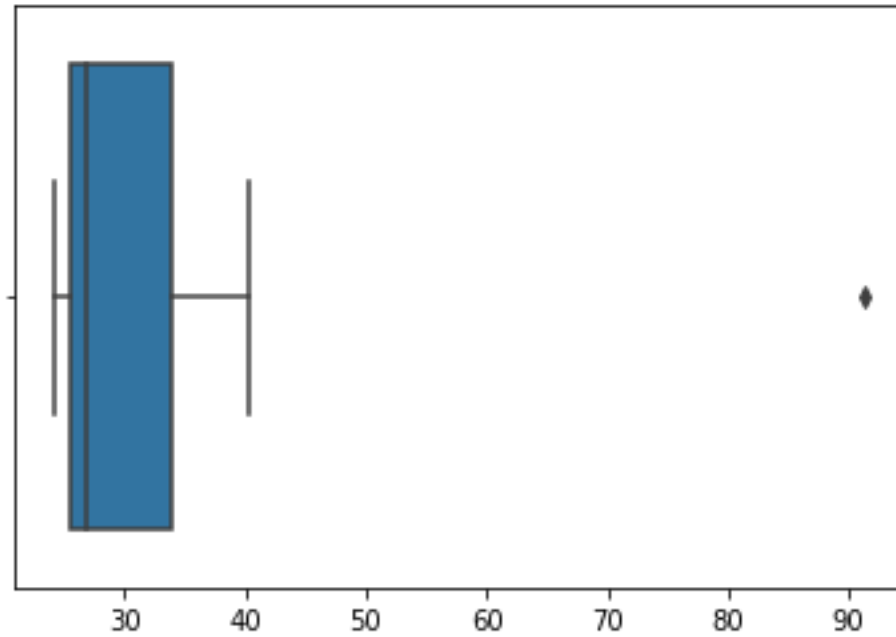
1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

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%

Plot:



Finding Outliers:



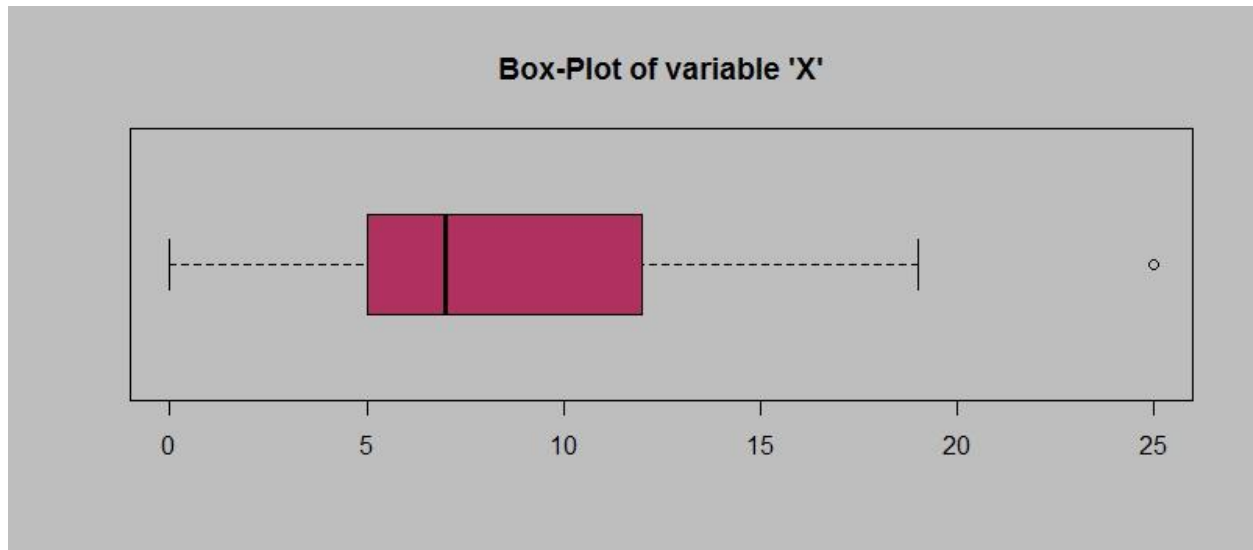
We can conclude that Morgan Stanley is an outlier at 91.36%

Finding Mean, Variance & Std. Deviation:

```
print('μ ',x.mean())  
print('σ = ',x.var())  
print('σ**2 = ',x.std())
```

```
μ  33.27133333333333  
σ  =  287.1466123809524  
σ**2 =  16.945400921222028
```

2.

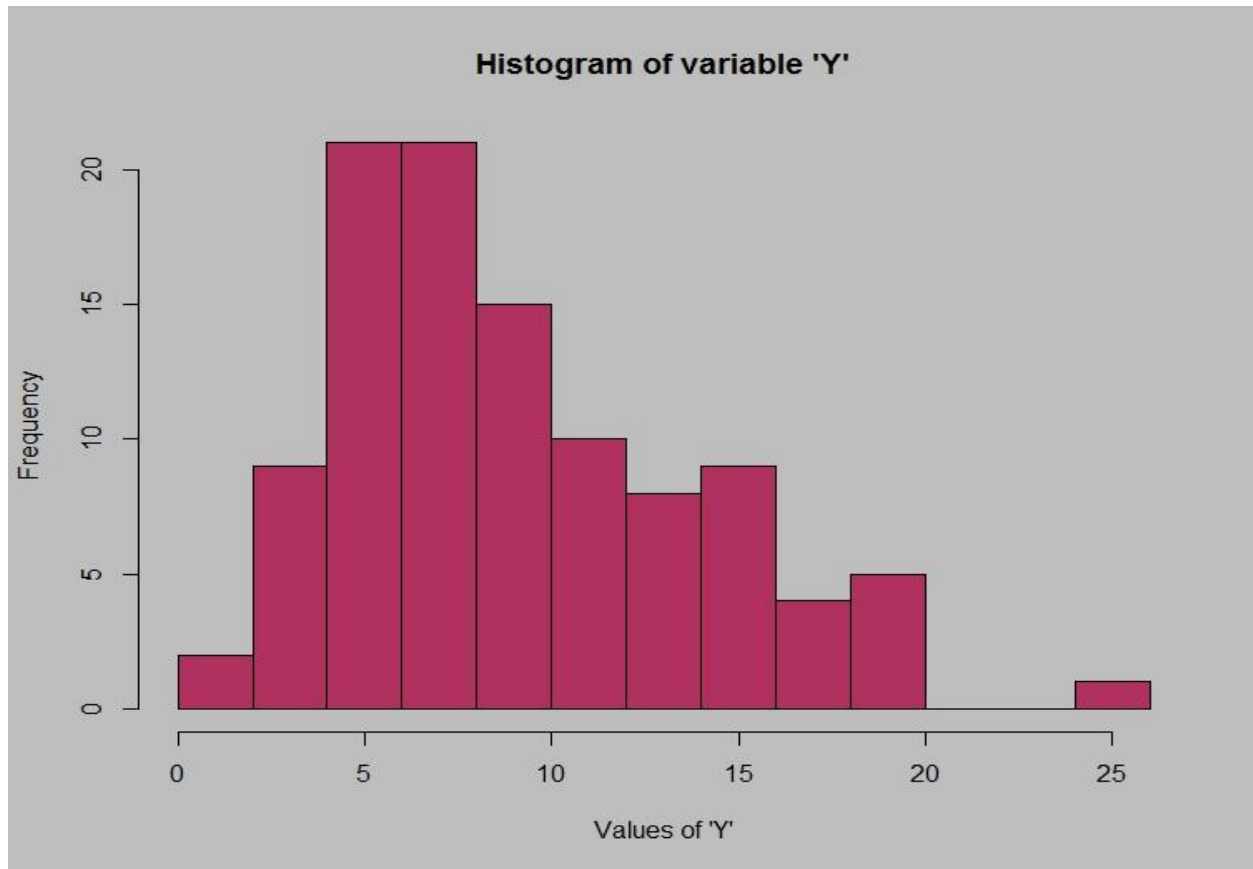


Answer the following three questions based on the box-plot above.

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

- 1)** The interquartile range for the given box-plot is between 5-12. This implies that, 50% of the data lies between 5-12.
- 2)** We can see that, $Q2 - Q1 < Q3 - Q2$, so the dataset is right/positively skewed
- 3)** If data point with value 25 is made 2.5, then the dataset will not have any outliers



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:

- 1)** The mode of the data set will lie between 4-7.5
- 2)** The data set is right/positively skewed
- 3)** We can't find a mode using boxplot, however we can find mode using the Histogram.

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.)

Ans:

Probability that as least five calls reaches the wrong number = 0.025 i.e. 2.5%

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?
- (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:

1) The most likely monetary outcome of the business venture is profit of \$2000 as it have highest probability of 30%.

2) We can see that, the probability of profit is $(0.2+0.3+0.1=0.6)$ i.e. 60%. Thus the venture is likely to be successful.

3) The long term average earning of business venture is:

Earning = $[(0.1)(-2000) + (0.1)(-1000) + 0 + (0.2)(1000) + (0.3)(2000) + (0.1)(3000)] = \800

4) Standard deviation can be a measure of risk involved

Standard deviation = 0.08164

```
In [50]: Profit=pd.Series([0.1,0.1,0.2,0.2,0.3,0.1])
print('Standard Deviation = ',Profit.std())

Standard Deviation = 0.0816496580927726
```