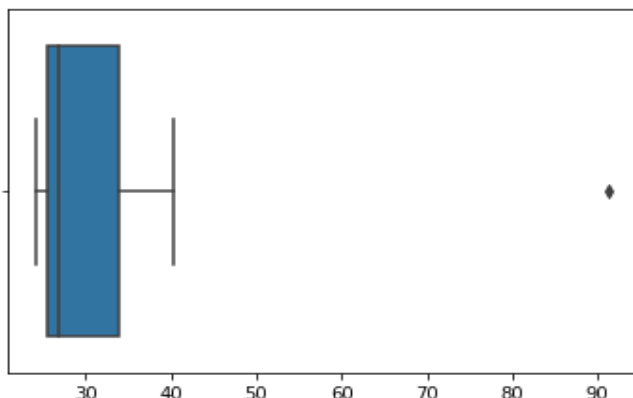


Topics: Descriptive Statistics and Probability

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%

```
import numpy as np
import scipy as sp
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
x=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.0
0])
name=['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers',
'Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US
Airways','Warner-Lambert']
# Box Plot to find outliers
sns.boxplot(x)
```



There is One Outlier which is Morgan Stanley at 91.36%

To find Mean

```
x.mean()
```

```
33.27133333333333
```

To find Variance

```
x.var()
```

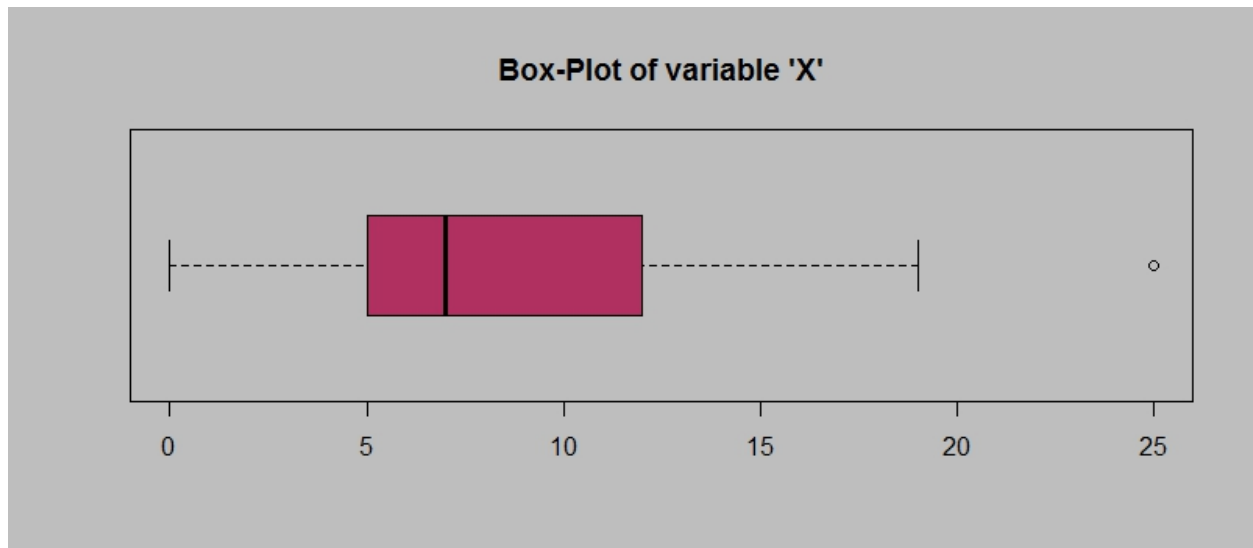
```
287.1466123809524
```

To find Standard Deviation

```
x.std()
```

```
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.

Ans : Inter Quartile Range = $12 - 5 = 7$, IQR represents the range which contains 50% of the data points

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

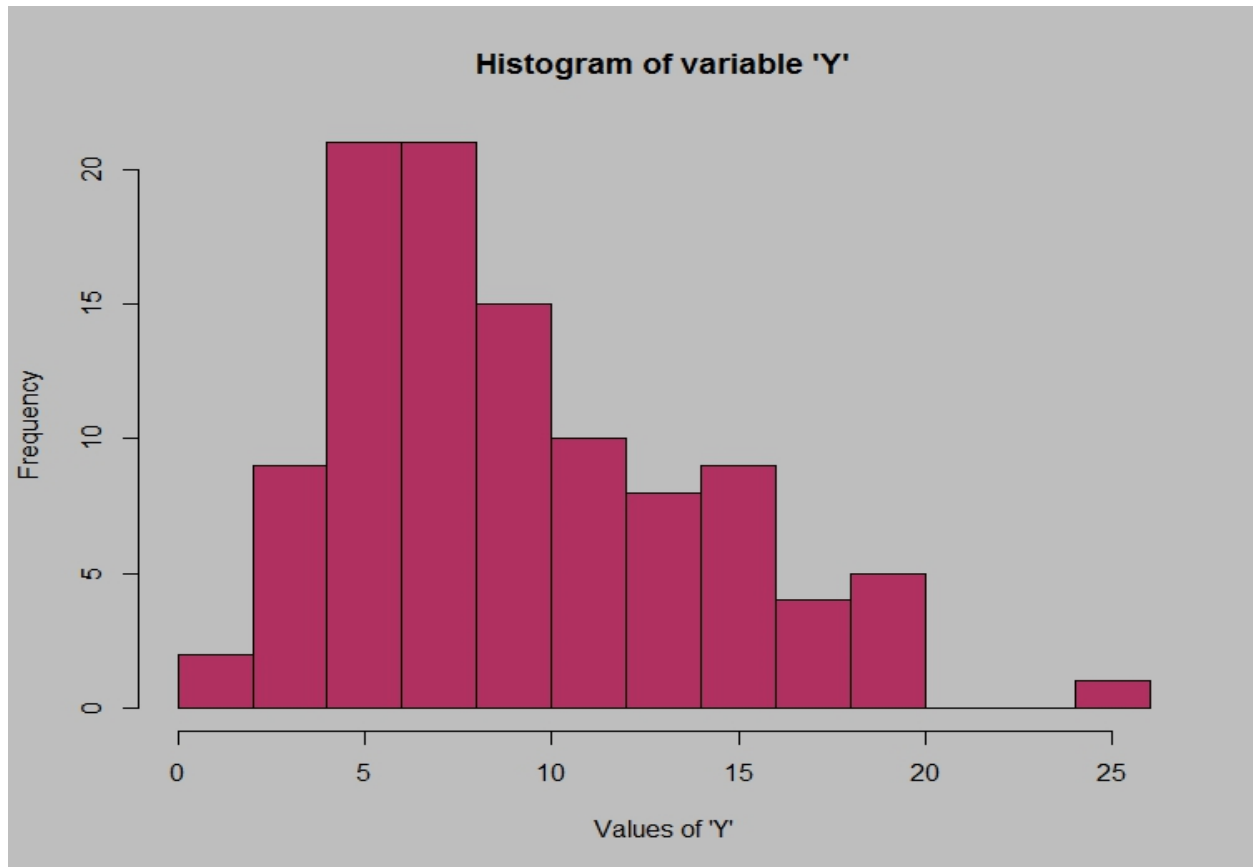
Right skewed

Ans The longer part of the box is to the right the median.

- (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 the data point with the value 25 is actually 2.5, then the outlier will come to the left side of the box, the boxplot will start after the value 2.5 .

3.



Answer the following three questions based on the histogram above.

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

Ans : Mode lies between 4 and 8

(ii) Comment on the skewness of the dataset.

Ans: Right-skewed distribution or positively skewed distribution, most data falls to the right side or positive side

(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: Mode in histogram, Histogram provides the frequency distribution so we can see how many times each datapoint is occurring

Median in boxplot, Boxplot provides whisker length to identify outliers, boxplot provides the quantile distribution 50% data lies between 5 and 12.

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.)

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 : Maximum probability $P = 0.3$ for $P(2000)$. The most likely outcome is 2000
- (ii) Is the venture likely to be successful? Explain
Ans : $P(\text{gain is } x > 0) = 0.6$, implies there is a 60% chance that the venture would yield profits.
 $P(\text{loss is } x < 0) = 0.2$,
venture is likely to be successful
- (iii) What is the long-term average earning of business ventures of this kind? Explain
Ans: Weighted average $= x \cdot P(x) = 800$.
The average expected earnings over a long period of time is 800. Including loss and gain
- (iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure
Ans : $P(\text{loss is } x < 0) = P(x = -2000) + P(x = -1000) = 0.1 + 0.1 = 0.2$
the risk associated with this venture is 20%