

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:

```
In [3]: data.mean()
```

```
Out[3]: Measure X    0.332713  
dtype: float64
```

```
In [4]: data.std()
```

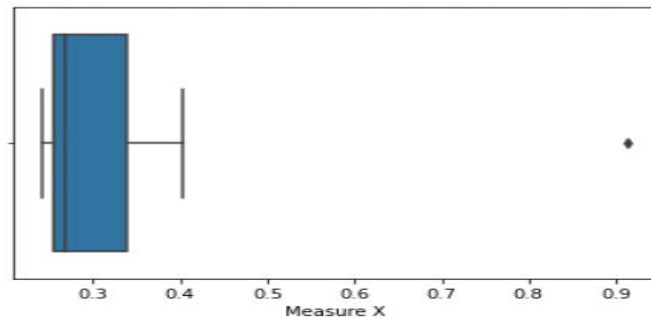
```
Out[4]: Measure X    0.169454  
dtype: float64
```

```
In [5]: data.var()
```

```
Out[5]: Measure X    0.028715  
dtype: float64
```

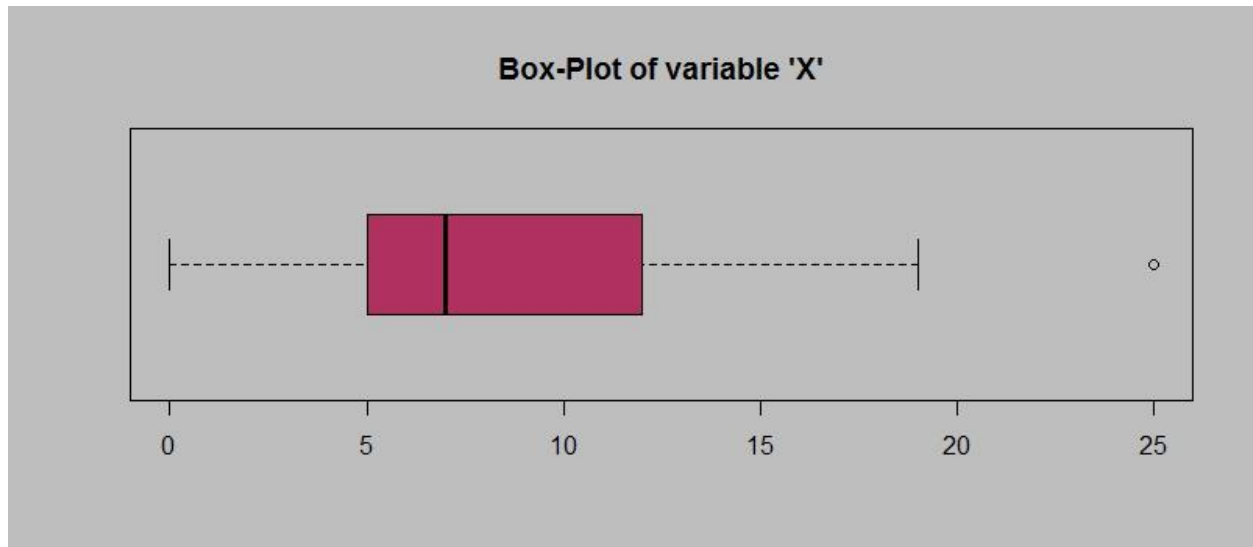
```
In [8]: sns.boxplot(data['Measure X'])  
plt.show()
```

C:\Users\shash\anaconda3\lib\site-packages\seaborn_drg: x. From version 0.12, the only valid positional argument will result in an error or misinterpretation. warnings.warn()



- Mean = 0.332
- Standard deviation = 0.169
- Variance = 0.0287
- The outlier is 91.36%

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:

$$\begin{aligned}\text{IQR} &= Q3 - Q1 \\ &= 12 - 5 \\ &= 7\end{aligned}$$

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

ANS:

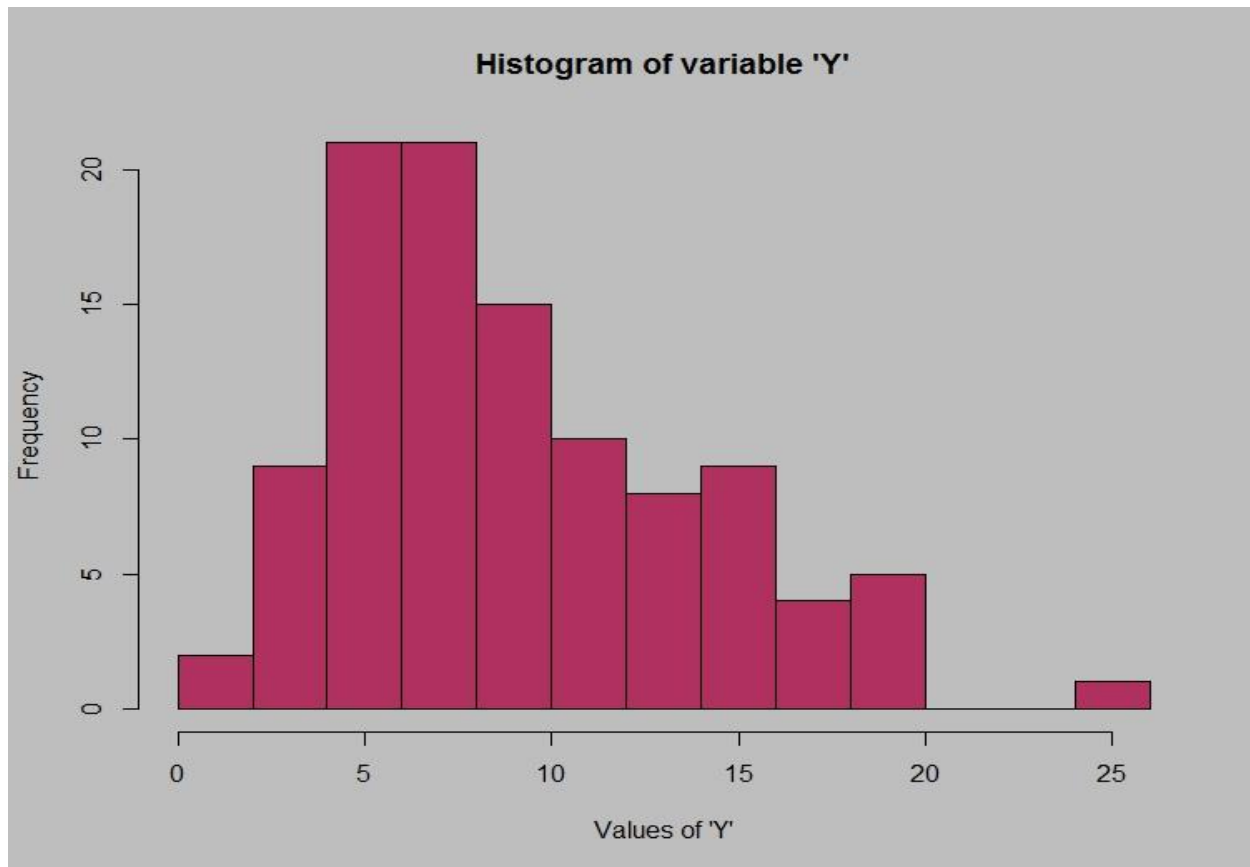
From the above box plot, most of the data towards the left side, hence data is right side.

- (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 value is 2.5, then there will be no outliers in the data.
- The median will shift to next positive value, since it is plotted using median

3.



Answer the following three questions based on the histogram above.

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

ANS:

It lies between 4 to 8.

- (ii) Comment on the skewness of the dataset.

ANS:

Most of the data lies towards left side, it is right 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:

- Boxplot helps to find the outliers, median, interquartile range and maximum and minimum values.
- Histogram helps to find the frequency of occurrence in the data.
- Compared to boxplot, the data distribution is clearly shown in Histogram
- Skewness can be determined easily in boxplot than histogram
-

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:

Probability of getting wrong number (p) = $1/200$

Probability of not getting wrong number (q) = $199/200$

By using binomial distribution equation,

When $x=0$, $p(x) = 0.975$

When $x \geq 1$, $p(x) = 1 - 0.975$
 $= 0.025$

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:

$$\begin{aligned} \text{Outcome} &= (-2000 \cdot 0.1) + (-1000 \cdot 0.1) + (0 \cdot 0.2) + (1000 \cdot 0.2) + (2000 \cdot 0.3) \\ &\quad + (3000 \cdot 0.1) \\ &= 800 \end{aligned}$$

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

ANS:

The most likely outcome is 800, hence the venture is successful

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

ANS:

$$\begin{aligned} \text{Outcome} &= (-2000 \cdot 0.1) + (-1000 \cdot 0.1) + (0 \cdot 0.2) + (1000 \cdot 0.2) + (2000 \cdot 0.3) \\ &\quad + (3000 \cdot 0.1) \\ &= 800 \end{aligned}$$

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

ANS:

$$V(x) = 2160000$$

$$Sd(x) = 1469.69$$