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%

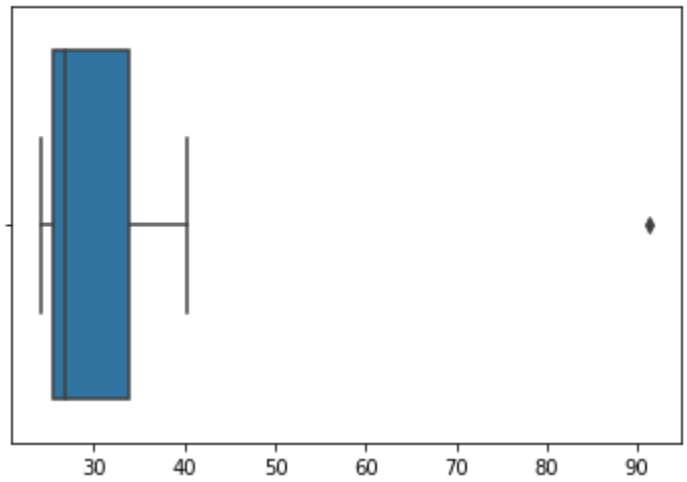
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
x=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00])
```

```
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 outliars
sns.boxplot(x)
```

```
<AxesSubplot:>
```



Questions referred to from *Aczel A., Sounderpandian J., Complete Business Statistics (7ed.)*

```
# Mean
x.mean()
```

```
33.27133333333333
```

```
# Vairance
x.var()
```

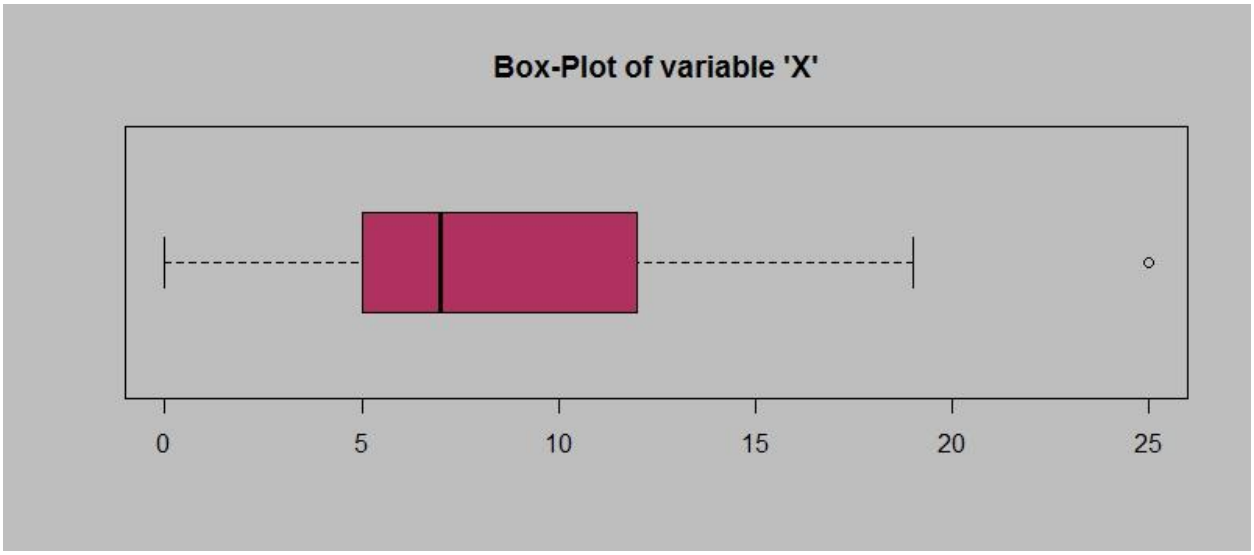
```
287.1466123809524
```

```
# Standard Deviation
x.std()
```

```
16.945400921222028
```

Outlier is 91.36%,Morgan Stanley.

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.

- IQR=Q3-Q1=12-5=7(Approximately)
- This value implies that most of the data lies in this range.

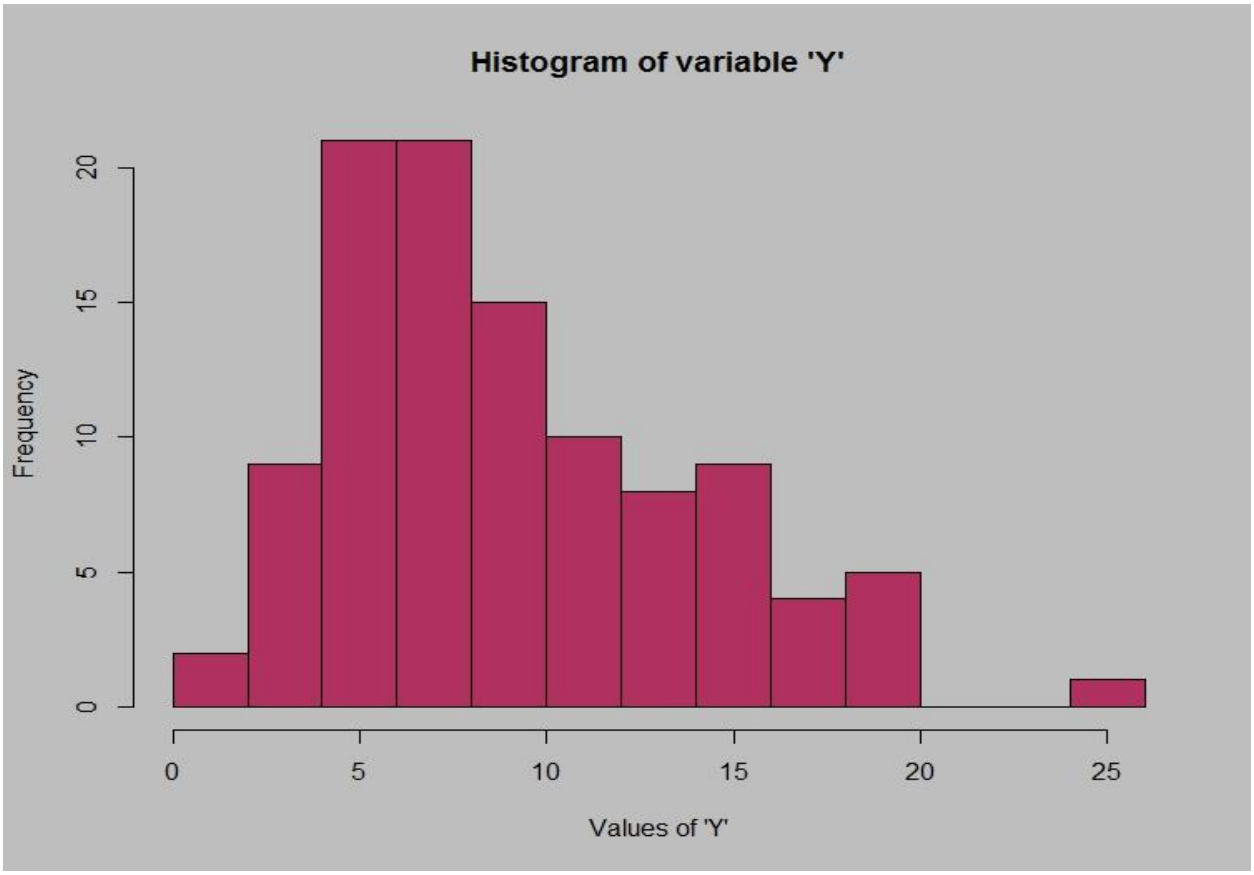
(ii) What can we say about the skewness of this dataset?

- Skewness of the data is ppositive.

(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?

- In thaty case the value 2.5 will liw in lower whisker part and positiveness of skew will decrease slightly.
- Median value will also increase slightly.

3.



Answer the following three questions based on the histogram above.

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

- Between 4 to 8.

(ii) Comment on the skewness of the dataset.

- Skewness is positive.

(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.

- Above histogram & boxplot are positively skewed.
- Both of them have the value 25 as outlier.
- Both of them are likely to have mode & median in similar interval.

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 that at least one in 5 attempted call reaches the wrong number is 0.025

Let us assume

E= The call is misdirected

F= The call is not misdirected

then probability of the event E is

$p(E)=1/200$

Therefore,

$P(F)=1-p(E)=199/200$

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

$=1-(199/200)^5$

$=0.025$

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?

- It can be seen from the above table that for x = 2000, the value of P(X) most. Hence, the most likely monetary outcome of the business venture is x = 2000.

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

- Yes,the venture is likely to be successful as the weighted average is positive.

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

x	P(x)	X*p(x)
-2,000	0.1	-200
-1,000	0.1	-100
0	0.2	0
1000	0.2	200
2000	0.3	600
3000	0.1	300
		Long-term average=800

As per the data given the long-term average is 800.

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

- Standard Deviation is a good measure of the risk involved in a venture,because standard deviation and risk is vice-versa.