

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 [5]: 1 import pandas as pd
```

```
In [6]: 1 import seaborn as sns
```

```
In [7]: 1 import matplotlib.pyplot as plt
```

```
In [8]: 1 data=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])
```

```
In [33]: 1 names=["Allied Signal","Bankers Trust","General Mills","ITT Industries","J.P.Morgan & Co.","LehmanBrothers","Marriott","MCI",
```

```
In [34]: 1 %matplotlib inline
```

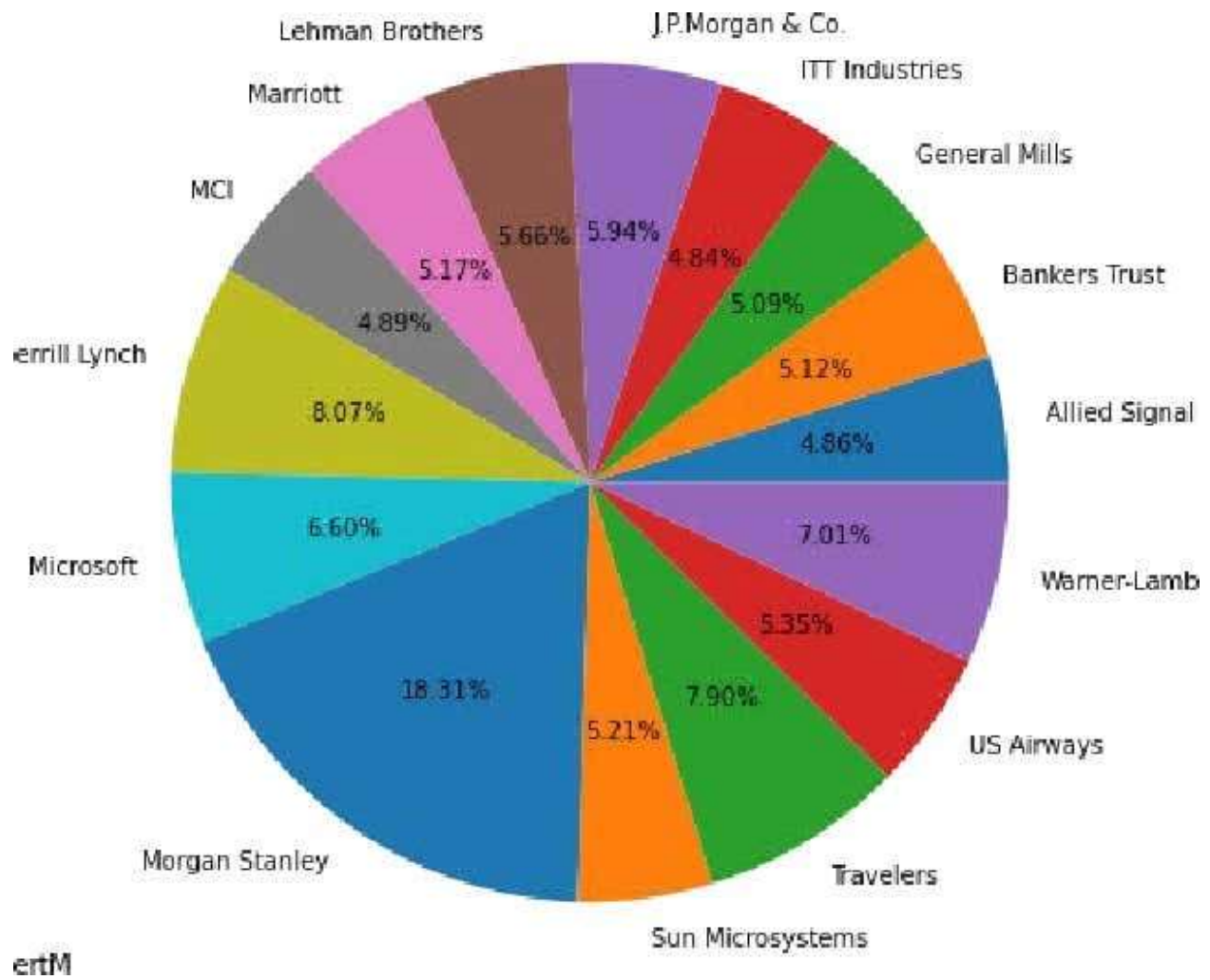
```
In [35]: 1 figure=plt.figure(figsize=(8,8))
```

...

```
In [36]: 1 plt.pie(data,labels=names,autopct='1.2%')
```

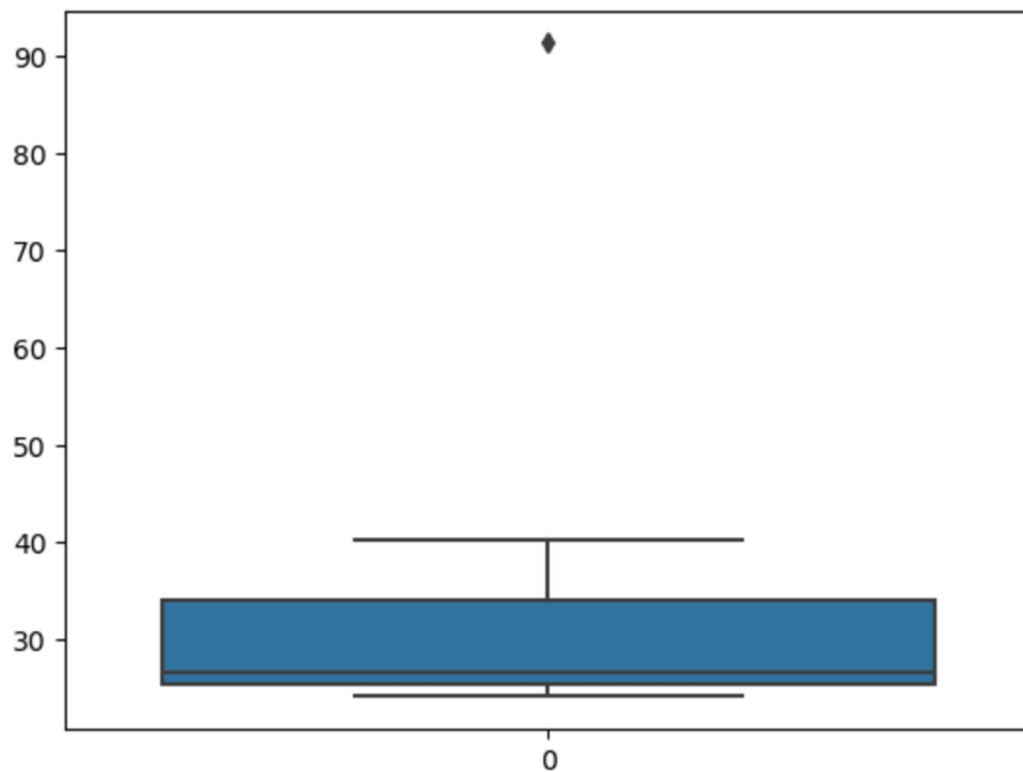
...

```
In [37]: 1 plt.show()
```



```
In [26]: 1 sns.boxplot(data)
```

```
Out[26]: <Axes: >
```



```
In [27]: 1 round(data.mean(),4)
```

```
Out[27]: 33.2713
```

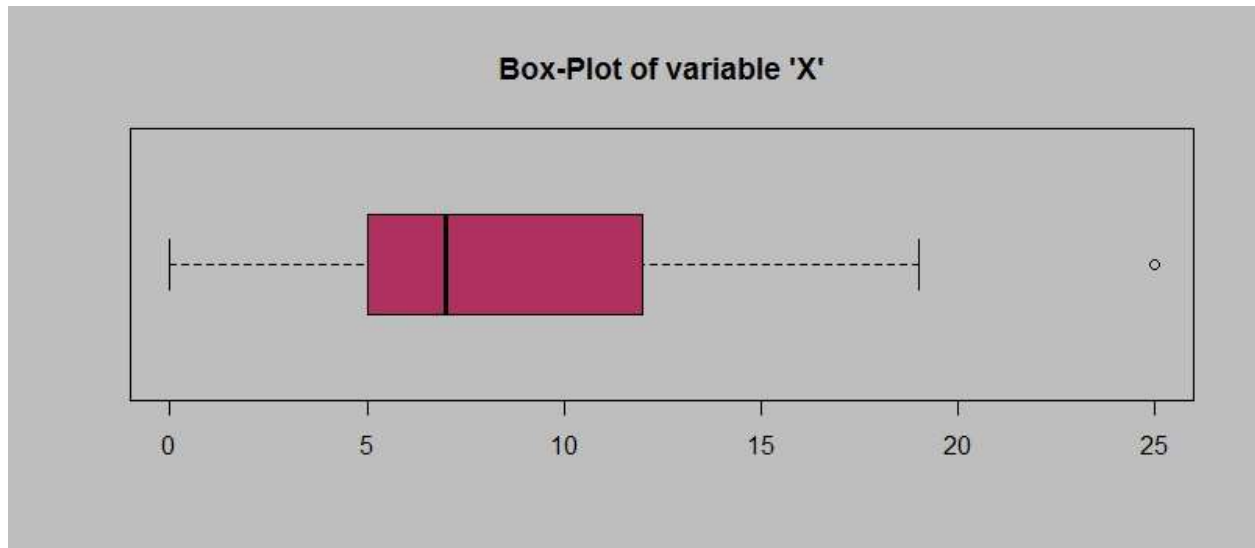
```
In [28]: 1 round(data.std(),4)
```

```
Out[28]: 16.9454
```

```
In [29]: 1 round(data.var(),4)
```

```
Out[29]: 287.1466
```

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 :- Here clearly 25 is the outlier

Median =7

1st quartile =5

2nd quartiles =12

IQR =(12-5)=7

IQR tells us the range of the middle half of the data.

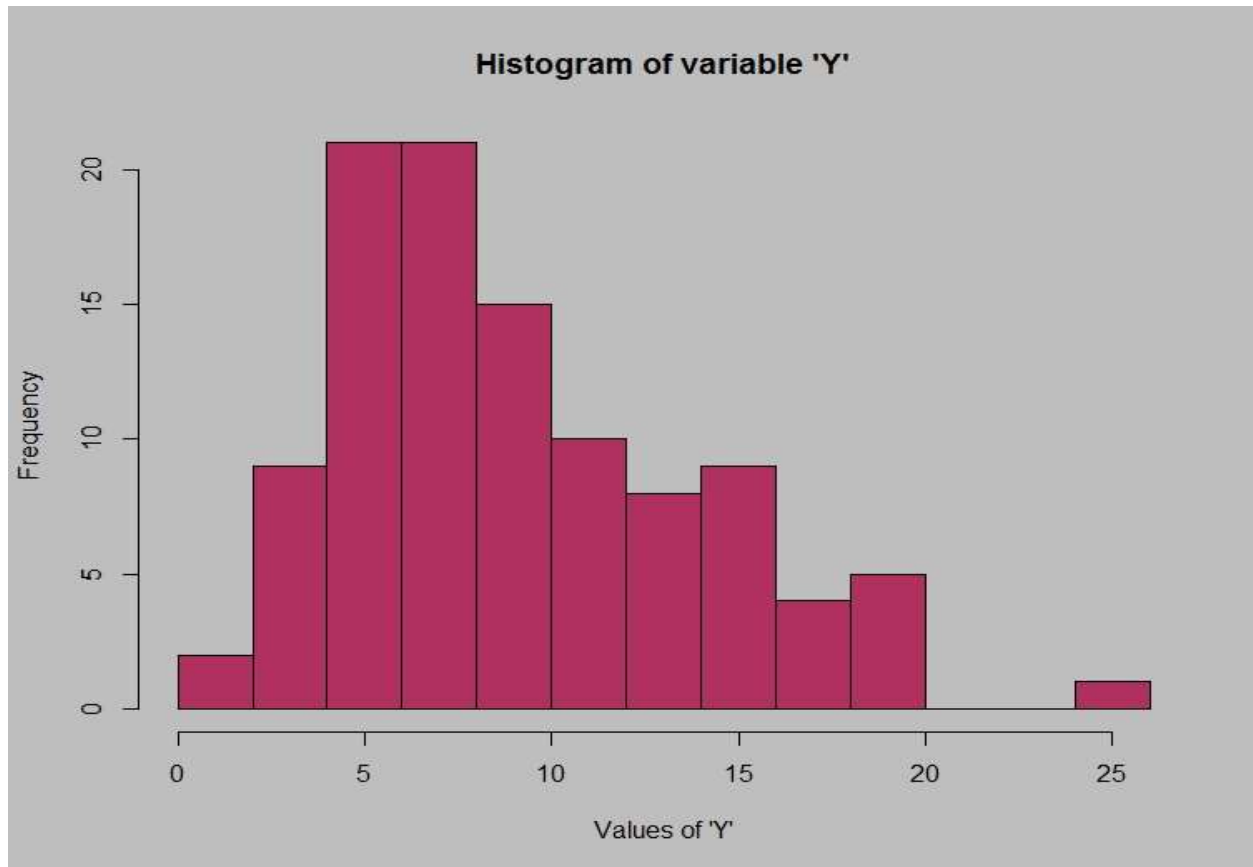
- (ii) What can we say about the skewness of this dataset?

Ans :- 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:- in that case there would have been no outlier , and it might have affected in the values of mean and median slightly . the boxplot might have moved towards rides slightly.

3.



Answer the following three questions based on the histogram above.

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

Ans :- between 5-8 (most frequent data)

(ii) Comment on the skewness of the dataset.

Ans:- it is positively 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 :- By comparing both of them it is very clear that the data would be positively skewed also, would help us finding mean ,mode value.

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 call getting misdirected $= (1/200)$

Hence probability of call not getting misdirected $= (1/200) = 199/200$

Number of phone call attempted $= 5$

Therefore ,probability that at least one in 5 attempted call reaches the wrong number is

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

$$= 0.025$$

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?

Ans:- Here the highest probability is for 2000

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

Ans:- yes because the total earning of the venture is positive in value i.e 800 and highest probability of earning is 2000.

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

Ans:-

X	P(x)	Income (x*p(x))
-2000	0.1	-200
-1000	0.1	-100
0	0.2	0
1000	0.2	200
2000	0.3	600
3000	0.1	300

Total

800

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

X	P(x)	X*P(x)	var 86666
-2000	0.1	-2	std 294.3
-1000	0.1	-1	
0	0.2		
1000	0.2	2	
2000	0.3	6	
3000	0.1	3	

