Name : Samrat Pawar **Branch : Andheri**

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 : Mean = 37.27133
Standard Deviation = 16.9454
Variance = 287.1466

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

jupyter ExcelR Assignment no. 2 (Basic Level Statistics II) Last Checkpoint: 4 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [10]:
```
dx=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']
```

In [15]:
```
#Mean
dx.mean()
```

Out[15]: 33.27133333333333

In [16]:
```
#Standard Deviation
dx.std()
```

Out[16]: 16.945400921222028

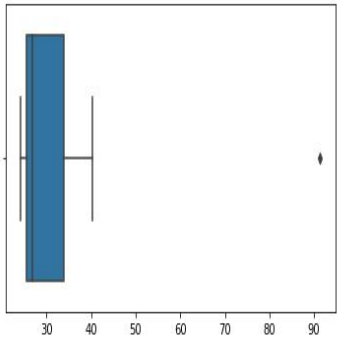
In [17]:
```
#Variance
dx.var()
```

Out[17]: 287.1466123809524

In [12]:
```
#Boxplot
sns.boxplot(dx)
```

```
C:\Users\Computer\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

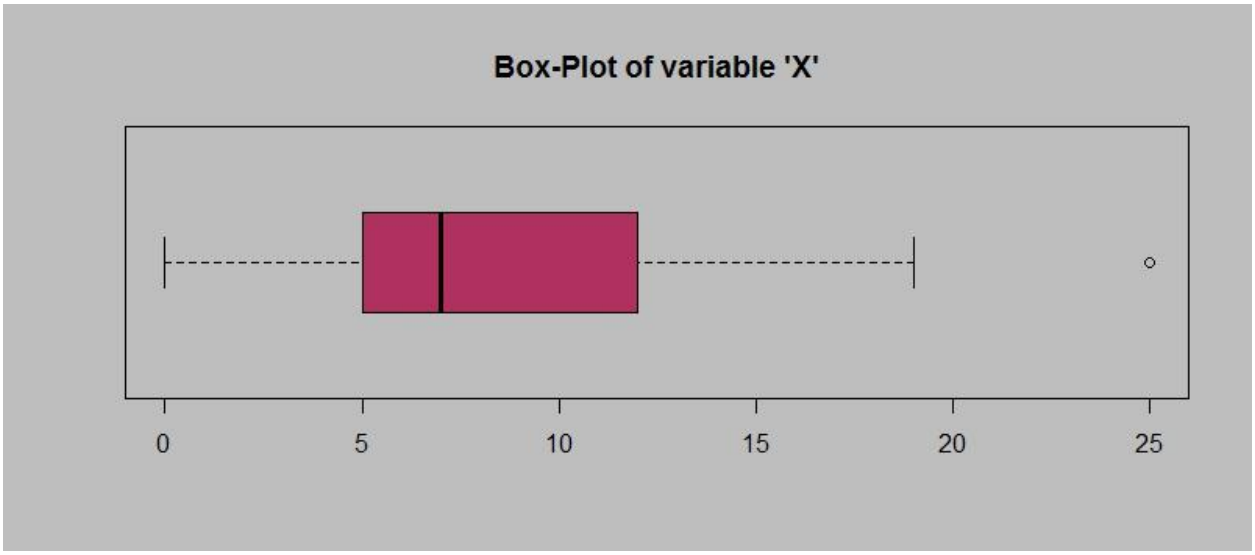
Out[12]: <AxesSubplot:>



In [ ]:
```
#INference : As we there is only one outlier at 92%
```

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

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:

Q3 = Upper Quartile , Q1 = Lower Quartile

IQR = Q3-Q1

= 12-5

= 7

This shows that 50% of the data lies between IQR.

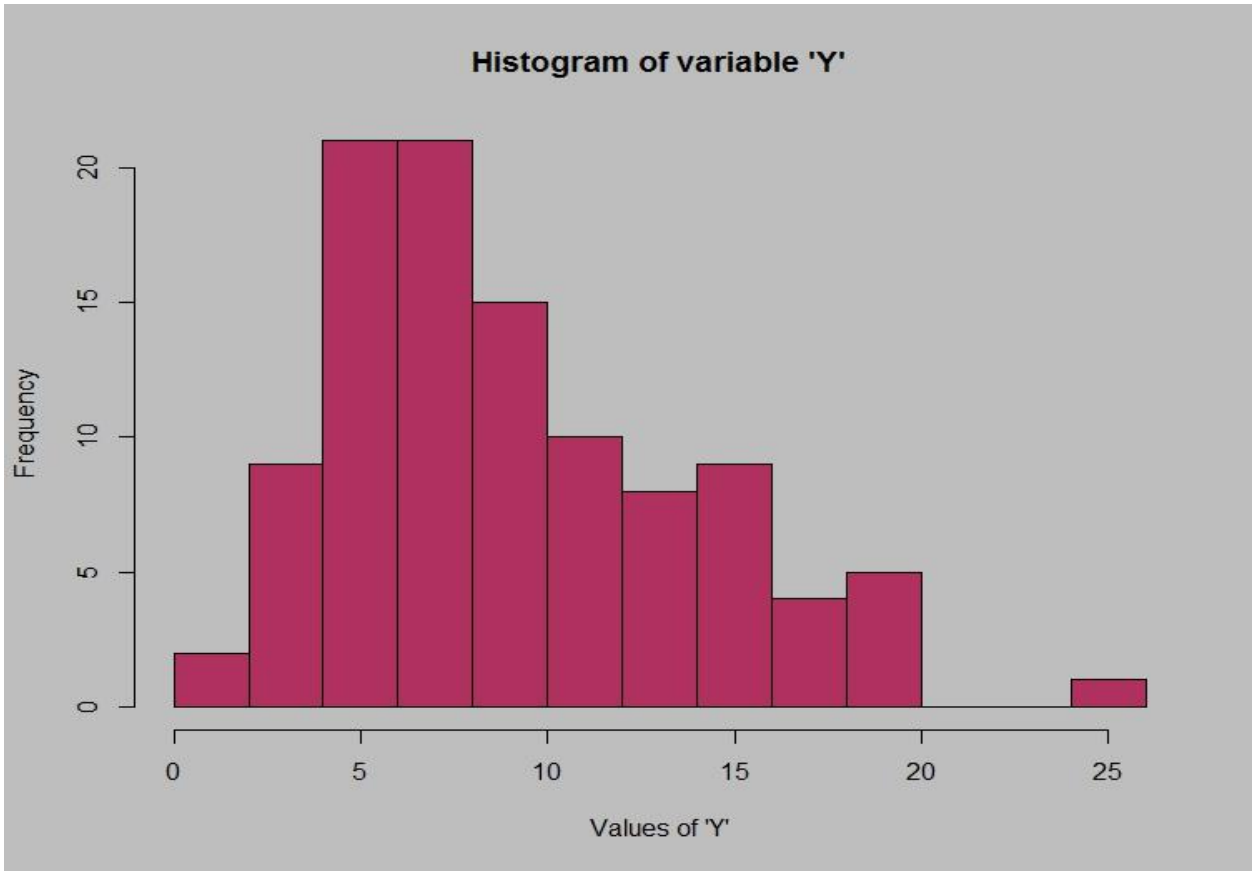
(ii) What can we say about the skewness of this dataset?

Ans : The skewness of the Dataset is Positive.

(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 : As we see there is Outlier on 25 , so if value is 2.5 then there will be no outlier .

3.



Answer the following three questions based on the histogram above.

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

Ans : The mode can lie between 3 and 10 because majority of the entry in this range.
To pin point the actual Mode we will have analyze the data

(ii) Comment on the skewness of the dataset.

Ans : The skewness of the Dataset 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.

Ans : There is an outlier of the value 25 and both the plot has positive skewness.

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 :

Given : one in 200 long-distance telephone calls is misdirected.

To find : probability that at least one in five attempted telephone calls reaches the wrong number

Solution:

one in 200 long-distance telephone calls is misdirected

=> probability of call misdirecting p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

Probability of having at least one successful call will be

1-P(X)= 1-1/200= 199/200= 0.967

As every event is independent of other event the probability will be

1- (0.967)^5

0.02475 = 2% chance.

probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475

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 : The most likely monetary outcome of the business venture is $2000

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

Ans : Yes . Because there are higher chances of positive returns.
As we see in table 0.3+0.2+0.1 = 0.6*100 = 60%

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

Ans : long-term average earning of business ventures = 800 $
by ,
x*P(x) = (-2000*0.1)+(-1000*0.1)+(0*0.2)+(1000*0.2)+(2000 *0.3)+(3000*0.1)

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

Ans : large value in the standard deviation of the variable x shows that there is high risk involved in this venture.

Var = 3.500000

Sd = 1870.83

In [29]:
```
x =  pd.DataFrame([[-2000,0.1],[-1000,0.1],[0,0.2],[1000,0.2],[2000,0.3],[3000,0.1]] , columns = ['X','P(x)'])
```

In [30]:
```
x
```

Out[30]:

	X	P(x)
0	-2000	0.1
1	-1000	0.1
2	0	0.2
3	1000	0.2
4	2000	0.3
5	3000	0.1

In [37]:
```
x.std()
```

Out[37]:
```
X       1870.828693
P(x)       0.081650
dtype: float64
```

In [34]:
```
x.var()
```

Out[34]:
```
X       3.500000e+06
P(x)    6.666667e-03
dtype: float64
```

In [39]:
```
#Large value in the standard deviation of the variable x shows that there is high risk  involved in this venture.
```

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