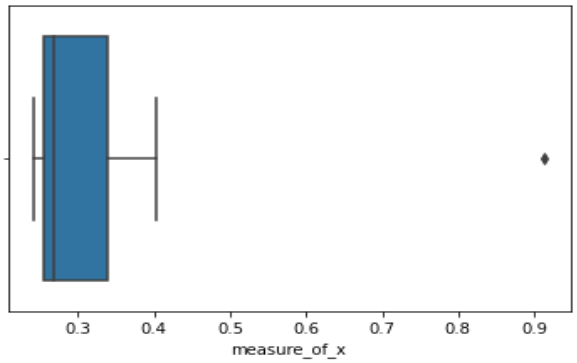
Topics: Descriptive Statistics and Probability

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%

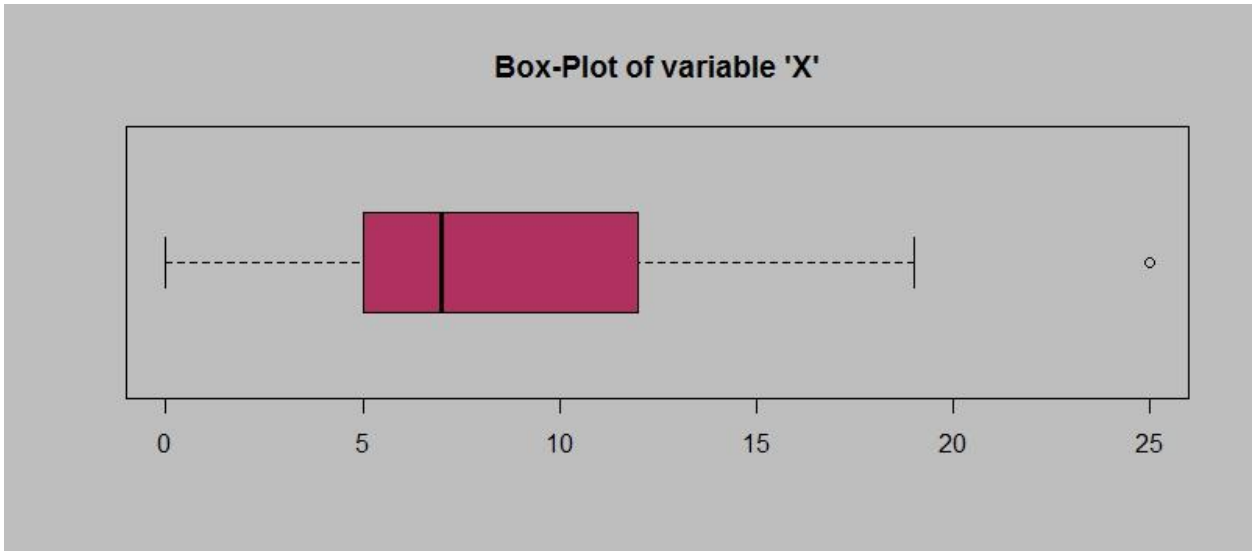
Python code: Copied Data to Excel and saved as df.csv
Importing python liabraries
import pandas as pd
import seaborn as sns
import statistics

- x_data = pd.read_csv('df.csv') # imported the data frame
  x_data
- x_data.measure_of_x.mean() #To find the mean of the given data (μ)
  0.3327 = 32.27
- x_data.measure_of_x.std() # To find the standard deviation value of the data(σ)
  0.1694 =16.94
- statistics.variance(x_data['measure_of_x']) # To find the variance of the data(σ^2)
  0.028714 =287.14
- sns.boxplot(x_data['measure_of_x']) # Data visualization using box_plot to find outlier



From above boxplot diagram we found the outlier value = 0.9136=91.36

2.



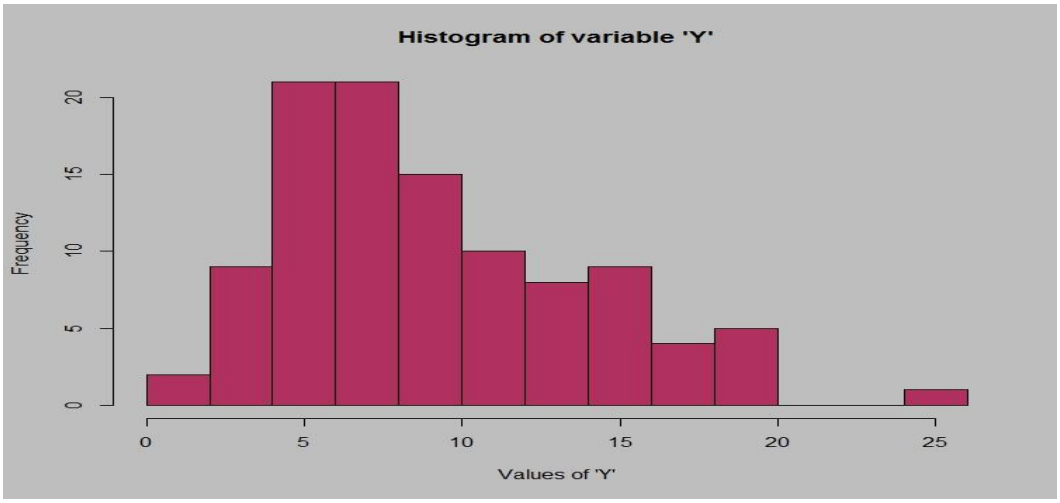
Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.
ANS: IQR = 12-5 = 7(approximately)
This value represents the range which contains 50% of the data points lie in the range of 5 and 12.

(ii) What can we say about the skewness of this dataset?
Ans: The dataset is rightly skewed or 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: The median value remains same,but inter-quartile range will change.

3.



Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?
Ans: The mode of data set lies between 4 and 8

(ii) Comment on the skewness of the dataset.
Ans: The data set is rightly skewed or positively skewed. Most of the data are present in the right side of the plot.

(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:

- We can observed Median in boxplot and Mode in histogram.
- Histogram provides the frequency distribution so we can see how many times each data point is occurring in the plot.
- Boxplot provides the quantile distribution so we can observe 50% data lies between in which range.
- Boxplot provides whisker length to identify the outliers but in histogram we can only guess looking at the grap that 25 may be an outlier.

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:

One misdirected call(wrong call) out of 200

Probability of wrong call = 1/200 =0.005

Probability of not wrong call = 1-0.005=0.995

Probability that at least one in five attempted telephone calls reaches the wrong numbe

= 1-Probability of all five calls are not reaches the wrong number

= $1-0.995^{\wedge 5}$

= 1-0.975

= 0.02475

= 2.47%

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: Expected probability = $\sum$ x * P(X)= $800

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

Ans: Yes the venture likely to be successful. Because when we take P(x>0)

That is 0.2+0.3+0.1 =0.6 and expected value is positive.

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

Ans: 800

Explanation:

[(-2000)+(−1,000) + (0) + (1,000) + (2,000) + (3,000)] / 6

= $500

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

Ans: standard deviation is the good measure of the risk involved in a venture of this kind.

Sd(x) = $1870.829.