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	c
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: Step 1: Converting Percentage Data:

```
In [24]: import numpy as np

         data_percent = [
             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
         ]

         data_numeric = np.array(data_percent) / 100

In [25]: data_numeric

Out[25]: array([0.2423, 0.2553, 0.2541, 0.2414, 0.2962, 0.2825, 0.2581, 0.2439,
                0.4026, 0.3295, 0.9136, 0.2599, 0.3942, 0.2671, 0.35  ])
```

Step 2: Calculate Mean, Standard Deviation, and Variance:

```
In [29]: data_numeric.mean()

Out[29]: 0.3327133333333333

In [31]: data_numeric.std()

Out[31]: 0.16370812590976933

In [32]: data_numeric.var()

Out[32]: 0.026800350488888885
```

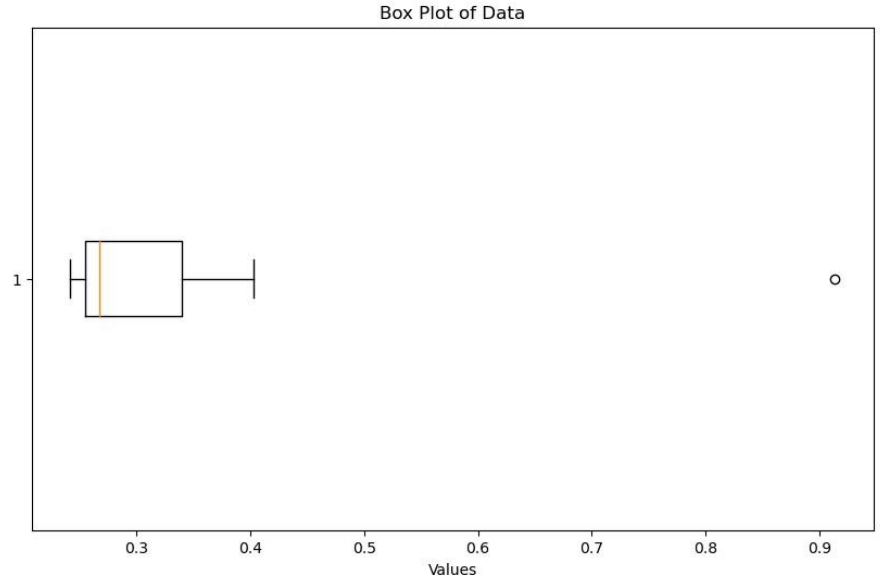
Questions referred to from *Aczel A., Sounderpandian J., Complete Business Statistics (7ed.)*

Step 3 : Identify Outlier

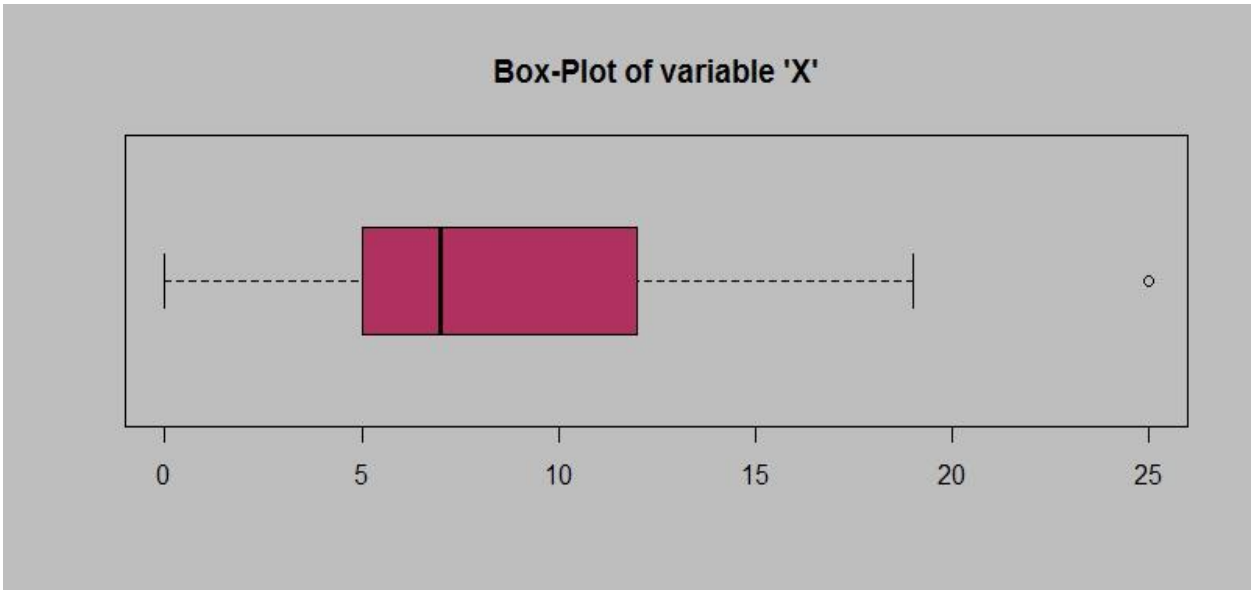
```
In [26]: Q1 = np.percentile(data_numeric, 25)
         Q3 = np.percentile(data_numeric, 75)
         IQR = Q3 - Q1

         lower_bound = Q1 - 1.5 * IQR
         upper_bound = Q3 + 1.5 * IQR

         outliers = [x for x in data_numeric if x < lower_bound or x > upper_bound]
         print("Outliers:", outliers)
```

Step 4: Plot a data

```
In [40]: import matplotlib.pyplot as plt

         plt.figure(figsize=(10, 6))
         plt.boxplot(data_numeric, vert=False)
         plt.title("Box Plot of Data")
         plt.xlabel("Values")
         plt.show()
```



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



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: First Quartile(Q1) = 5,Third Quartile (Q3)= 12 , Second Quartile (Q2)= 7
Interquartile Range QR = Q3-Q1 = 12-5 = 7
Interquartile Range is median value

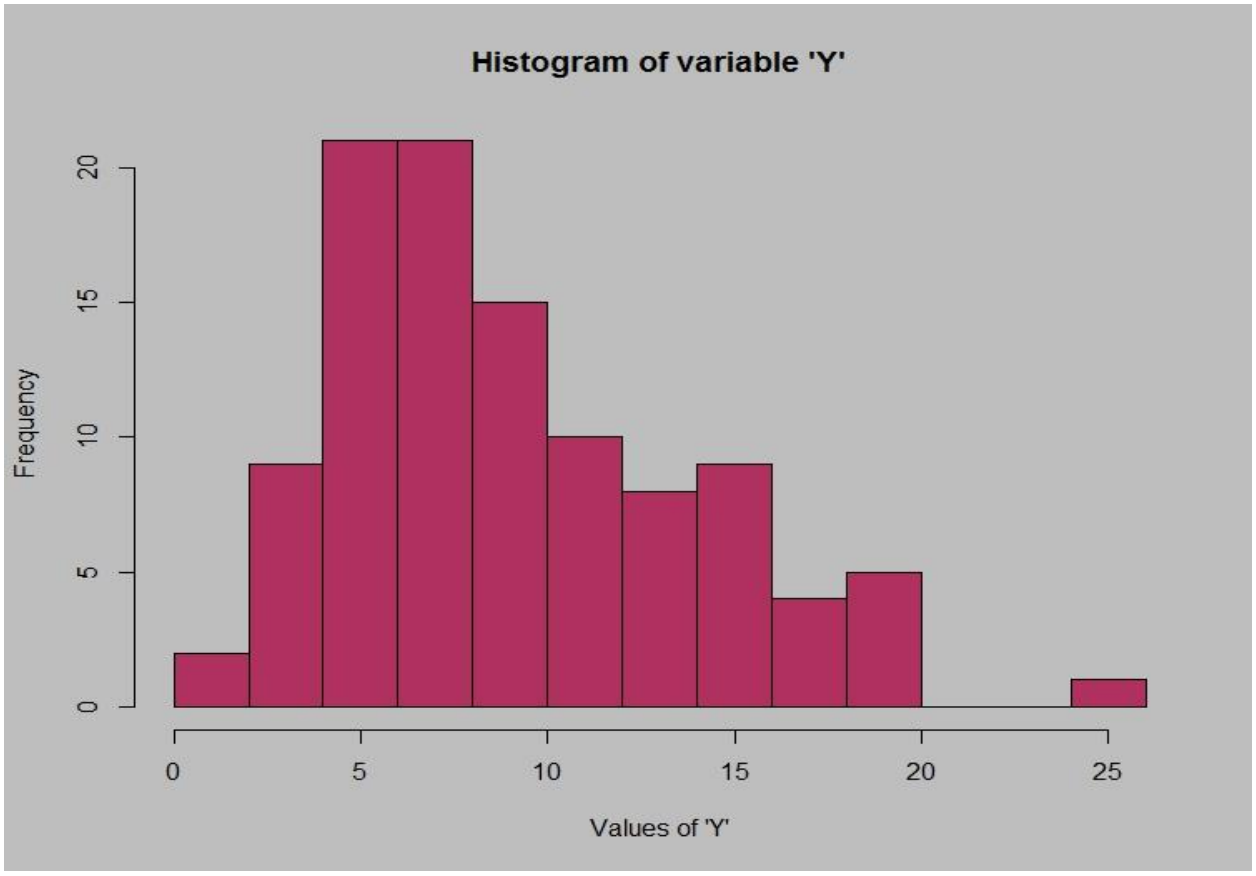
(ii) What can we say about the skewness of this dataset?

Ans: Right-Skewed median is towards the left side it is not normal distribution

(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 be no Outliers on the given dataset because of the outlier the data had positive skewness it will reduce and the data will normal distributed

2.



Answer the following three questions based on the histogram above.

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

Ans : The mode of this data set lie in between 5 to 10 and approximately between 4 to 8

(ii) Comment on the skewness of the dataset.

Ans: Right-Skewed. Mean>Median>Mode

(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 : They both are right-skewed and both have outliers the median can be easily visualized in box plot where as in histogram mode is more visible.

3. 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 that the probability of a single call being misdirected is 1/200, the probability of a single call being correctly directed is 1 - 1/200 = 199/200.

Assuming independence of attempts, the probability that all five calls are correctly directed is (199/200)^5.

Therefore, the probability that at least one call in five attempted telephone calls reaches the wrong number is:

1 - (199/200)^5 ≈ 0.02469 or about 2.469%.

So, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 2.469%.

4. 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 the one with the highest probability, which corresponds to the value of $2000 with a probability of 0.3.

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

Ans: To determine if the venture is likely to be successful, we need to consider the overall expected value. The expected value is calculated by summing up the product of each outcome and its corresponding probability.

Expected Value (E) = (-2000) * 0.1 + (-1000) * 0.1 + 0 * 0.2 + 1000 * 0.2 + 2000 * 0.3 + 3000 * 0.1

E = -200 - 100 + 0 + 200 + 600 + 300

E = 800

Since the expected value is positive ($800), it indicates that, on average, the business venture is likely to be successful in terms of generating earnings.

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

Ans: The long-term average earnings of business ventures of this kind is given by the expected value, which is $800. This means that, over the long run and considering the given probability distribution, the business venture is expected to earn an average of $800.

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

Ans: A good measure of the risk involved in a venture of this kind is the standard deviation. The standard deviation gives us an idea of how spread out the possible outcomes are around the expected value

Standard Deviation (σ) = √[Σ(xi - μ)^2 * P(xi)]

σ = √[(-2000 - 800)^2 * 0.1 + (-1000 - 800)^2 * 0.1 + (0 - 800)^2 * 0.2 + (1000 - 800)^2 * 0.2 + (2000 - 800)^2 * 0.3 + (3000 - 800)^2 * 0.1]

σ ≈ 1410.67