Author: **Shreyas Shashikant Vaishnav**
Purpose: **Assignment 2 – Set 1**

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:
We have to find firstly outliers so the code for the outliers in the Jupiter notebook is as follows.

```
>> dt = [24.23, 25.53, 25.41, 24.14, 29.62, 28.25, 25.18, 24.39, 40.26, 32.95, 91.36, 25.99, 39.42, 26.71, 35.00]
>>dt
>>ot = []
>>def detect_ot(dt):    # Defining a custom function to find out the outliers
    threshold = 3
    mean = np.mean(dt)
    std = np.std(dt)

  for i in dt:
    z_score = (i-mean)/std
    if np.abs(z_score) > threshold:
      ot.append(i)
  return ot
```

```
ot_pt = detect_ot(dt)
ot_pt
```

[91.36]

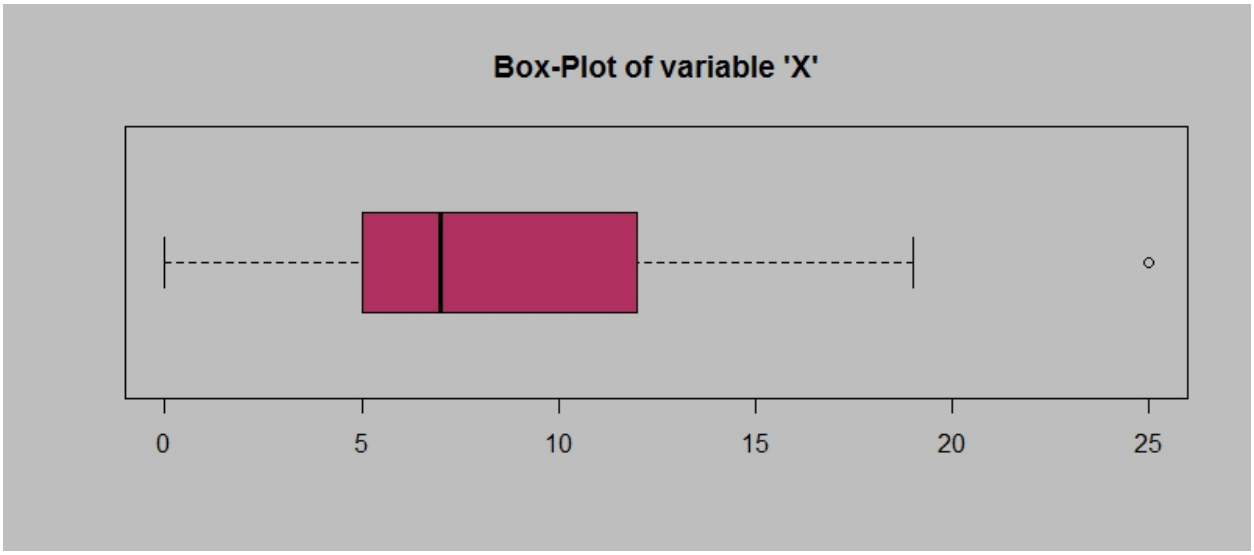
 # plt.plot(x)

stats.mean(x)
33.27133333333333

stats.variance(x)
287.1466123809524

stats.stdev(x)
16.945400921222028

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.

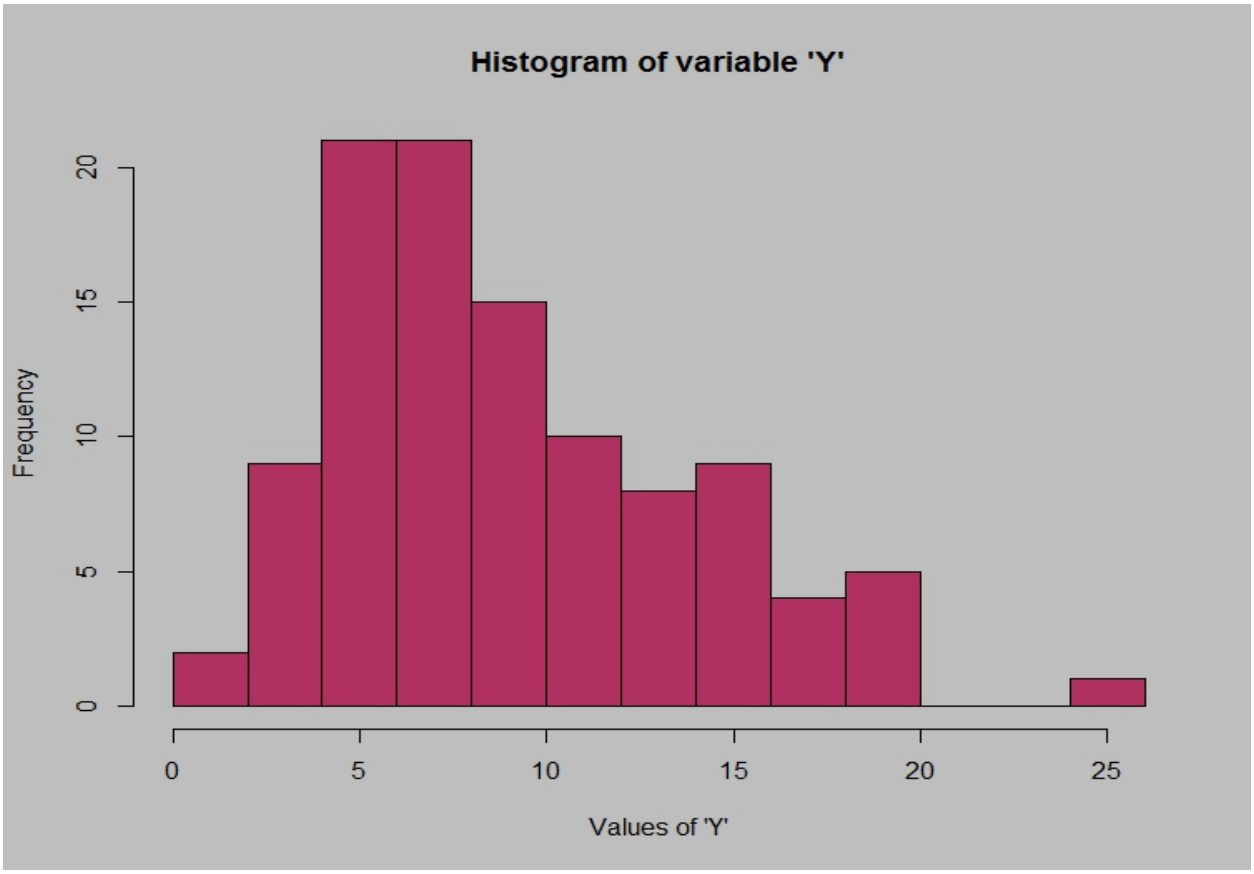
(ii) What can we say about the skewness of this dataset?

(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.

- IQR = Q3-Q2 = 12-5 = 7 and this value 7 is the median of a boxplot.
- This data set is positively skewed.
- If it was found that the data point with the value 25 is actually 2.5 then 2.5 will not be considered an outlier. Because this boxplot starts form zero to twenty.

3.



Answer the following three questions based on the histogram above.

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

(ii) Comment on the skewness of the dataset.

(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.

- Mode of this data set is lie in between 4 to 8.
- This data set is positively skewed and the data represents right tail skewness.
- If boxplot and histogram are plotted for same dataset then in box-plot we easily find out the outliers but in histogram we cannot find we can only guess that 25 is outlier. These plots are shows same skewness for same datasets.

---

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.

Give that, Number of calls = 5 and 200 long-distance telephone calls is misdirected. Then we have to find that the probability that at least one in five attempted telephone calls reaches the wrong number.

In this problem one in 200 long-distance telephone calls is misdirected this implies that,

Probability of call misdirecting p = 1/200 = 0.005 and

Probability of call not misdirecting = 1-1/200 = 199/200 = 0.995

But we know that the binomial distribution is,

$P(X) = P[X = x] = {}^{n}C_{x}\, p^{x}\, q^{n-x}$, x = 0,1,…. And p+q=1

Here, n=5, p = 1/200, q = 1-p = 1-1/200 = 199/200 = 0.995

At least one in five attempted telephone calls reaches the wrong number = 1-none of call reaches wrong number,

= 1 – P [X = 0] = 1 - 5C0 (1/200) ^0 (199/200) ^5⁻0

= 1 - (199/200) ^5

= 1- (0.995) ^5

= 1 – 0.97524875

= 0.02475125 or 2.47%

There-fore the probability that at least one in five attempted telephone calls reaches the wrong number is 0.02475125.

---

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?

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

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

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

Ans.

- For P (200), the maximum p = 0.3 so that the most likely outcome is 2000.
- Here the P(x>0) = 0.6, this implies that there is a 60% chance that the venture would yield greater than expected returns. And the P (Incurring losses) is only 0.2. There-fore the venture is likely to be successful.
- Here we have to find expected value that is weighted average.

And weighted average = x*P(x) = 800. This means the average expected earnings over a long period of time would be 800(including all losses and gains over the period of time).

- Here we have to sum the loss probabilities that is, Probability(loss) = probability (-200) + probability (-100) = 0.1+0.1 = 0.2. This implies that the risk associated with this venture = 20%.
