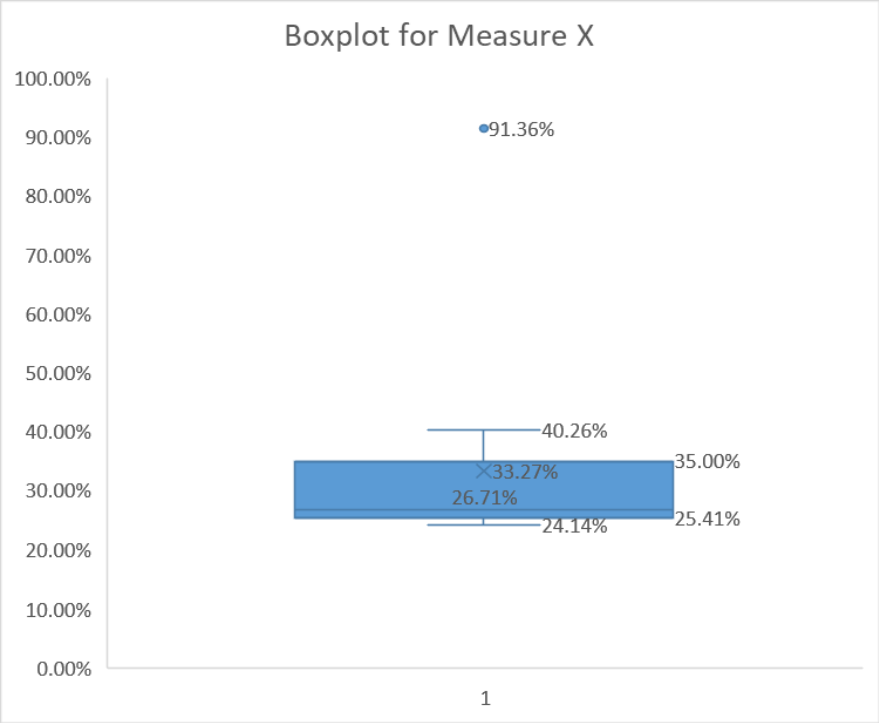
Name – Niranjan Nevase Batch – 5th June

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%

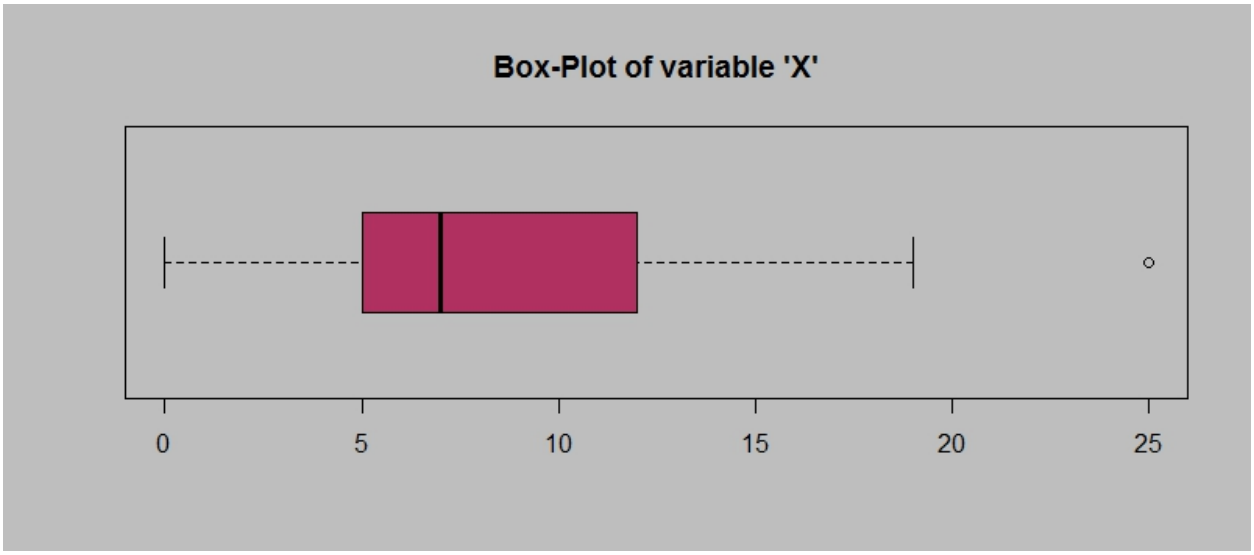


- Outlier = 91.36%
- μ = 33.27%,
- Population, σ = 0.164, σ^2 = 0.027
- Sample, σ = 0.169, σ^2 = 0.029

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

Name – Niranjan Nevase Batch – 5th June

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.

- IQR = 12 – 5 = **7 (approx..)**
- The interquartile range value implies the spread or variability of the middle 50% of the data.

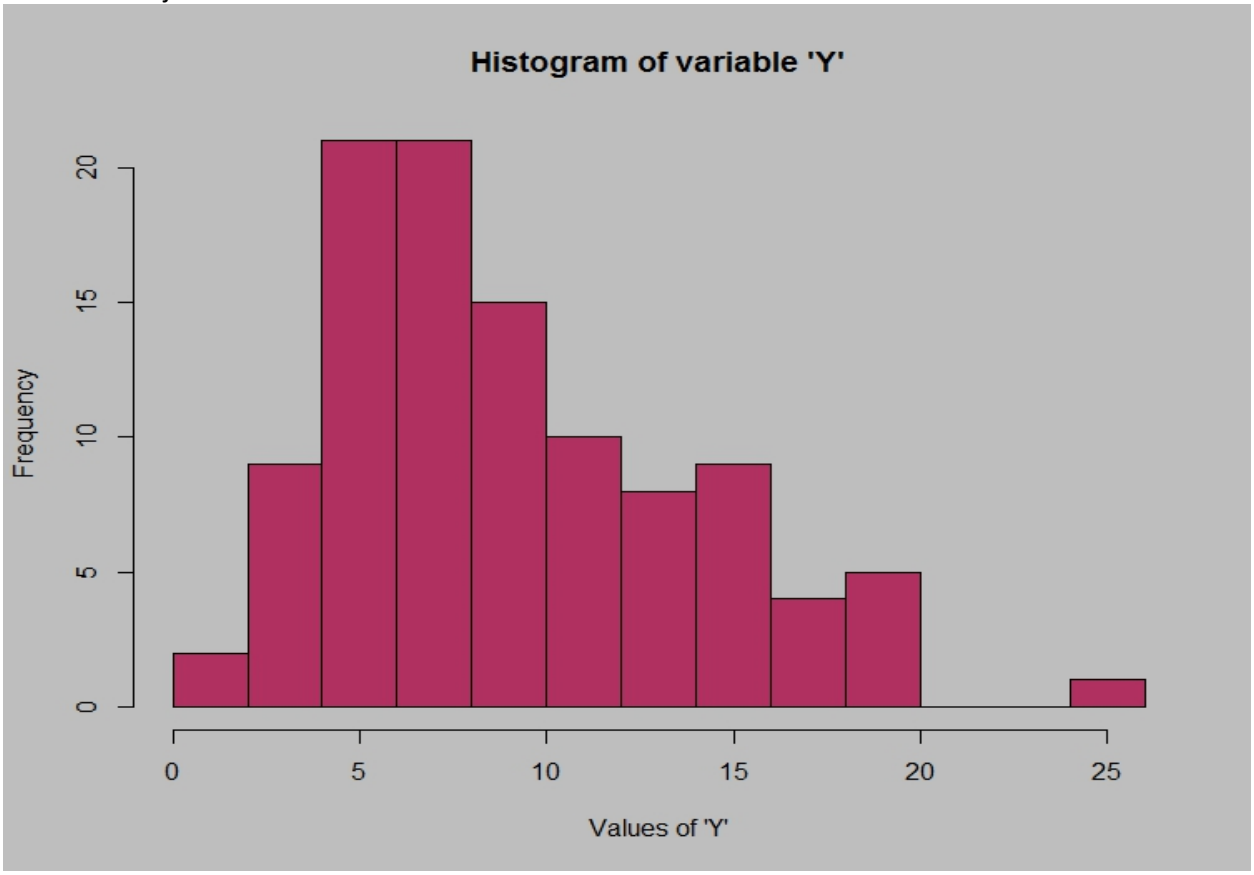
(ii) What can we say about the skewness of this dataset?

- Median is closer to the bottom of the box and the whisker is shorter on the lower end of the box, the distribution is right-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?

- The minimum value of the data would change from 25 to 2.5, which would reduce the range of the data and make the left whisker shorter.
- The first quartile (Q1) and the median of the data would likely decrease, as they are based on the order of the data values. This would shift the box to the left and make it narrower.
- The third quartile (Q3) and the maximum value of the data would not change, as they are not affected by the correction of the data point. The right whisker would remain the same length.
- The interquartile range (IQR) of the data would decrease, as it is calculated by subtracting Q1 from Q3. This would indicate less variability in the middle 50% of the data.
- The new boxplot will have zero outliers.

Name – Niranjan Nevase Batch – 5th June



3. Answer the following three questions based on the histogram above.

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

- The highest peak of the histogram represents the location of the mode of the data set. In above histogram, mode should be aprrox. **20.**

(ii) Comment on the skewness of the dataset.

- Data set has positive skewness, with most of the values concentrated on the left side of the graph and a few large values on the right side.
- The right tail of the histogram is longer than the left tail, indicating that there are some outliers or extreme values in the data.
- A positive skewness also implies that the mean of the data is greater than the median, and both are greater than the 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.

- Both graphs show positive skewness in dataset.
- Both graphs show similar outlier value i.e, 25.

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

Name – Niranjan Nevase Batch – 5th June

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.)

   - We can use the binomial distribution, which models the number of successes in a fixed number of independent trials, each with a constant probability of success. In this case, the number of trials is 5, the probability of success (reaching the wrong number) is 1/200, and we want to find the probability of at least one success.

   The formula for the binomial probability is:
   $P(X=k) = {}^{n}C_{k}\ p^{k}\ q^{n-k}$
   where:
   X is the random variable that counts the number of successes
   k is the number of successes
   n is the number of trials
   p is the probability of success in each trial

   To find the probability of at least one success,
   $P(X=1) = 1 - P(X=0) = 1 - (1 \times 1 \times (199/200)^{5}) = 1 - 0.9752$ **$= 0.0248 = 2.48\%$**

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?

   - The most likely monetary outcome of the business venture, also known as the mode of the probability distribution, is the value with the highest probability mass. In this case, the value 2000 has the highest probability of 0.3, which is the largest among all the probabilities. So, the most likely monetary outcome is **$2,000**.

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

Name – Niranjan Nevase Batch – 5th June

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

Expected value = $\sum x \cdot P(x)$ = **800**

The expected value of the returns on this business venture is $800. This means that on average, we can expect to make a profit of $800 from this venture.

Based on this metric, we can say that the venture is likely to be successful, as it has a positive expected value and can generate profit on average.

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

- The long-term average earning of business ventures of this kind is equal to their expected value.
- As mentioned above, the expected value represents the long-run average of repeated experiments or observations of a random variable. Therefore, if we repeat this business venture many times under similar conditions and probabilities, we can expect to earn $800 on average in each trial.

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

- Possible measure of the risk involved in a venture of this kind is the variance & Standard deviation of its returns.

  $var(X)=\sum(X-\mu)^2 \cdot P(x)$, where $\mu=E(X)$

  var (X) = 2160000

  Std. dev (X) = $1,469.69

- Based on these measures, we can say that the venture has a high risk, as it has a large variance and standard deviation in relation to its mean. This means that the returns are very spread out and unpredictable, and there is a high chance of losing money or making less than expected.

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