

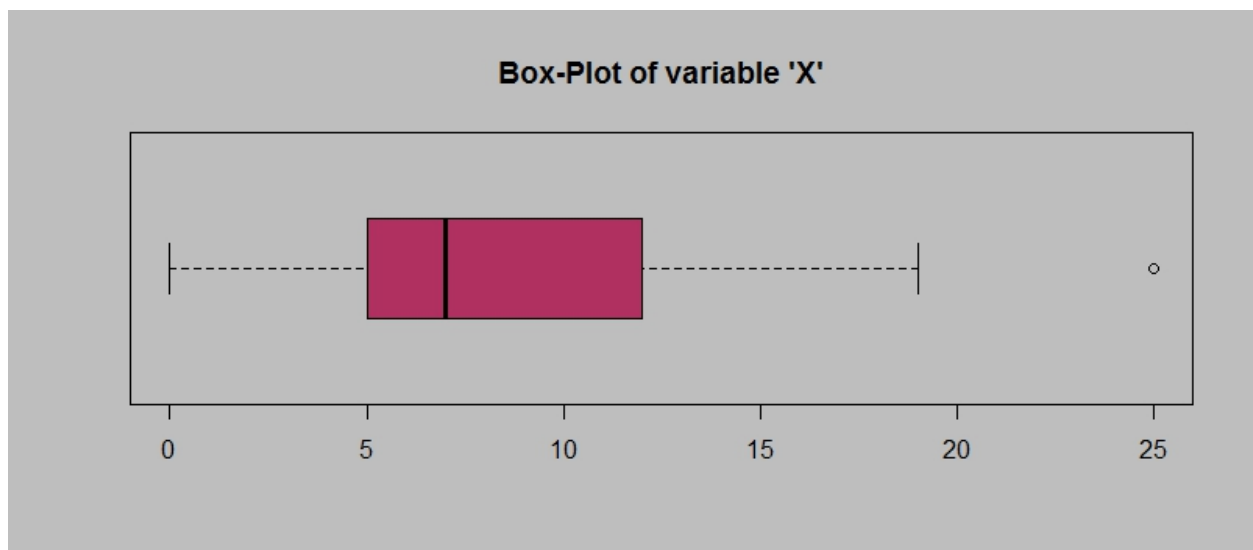
## Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out  $\mu, \sigma, \sigma^2$

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: Given in the attached ipynb file

- 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: From the above box-plot assuming,

$$Q1 = 5 \text{ (approximately)}$$

$$Q3 = 12 \text{ (approximately)}$$

$$IQR = Q3 - Q1 = 12 - 5 \Rightarrow 7$$

It is useful in identifying the spread of data and detecting outliers. IQR provides information about the variability of the middle 50% of the data points of a dataset. It is less sensitive to extreme values.

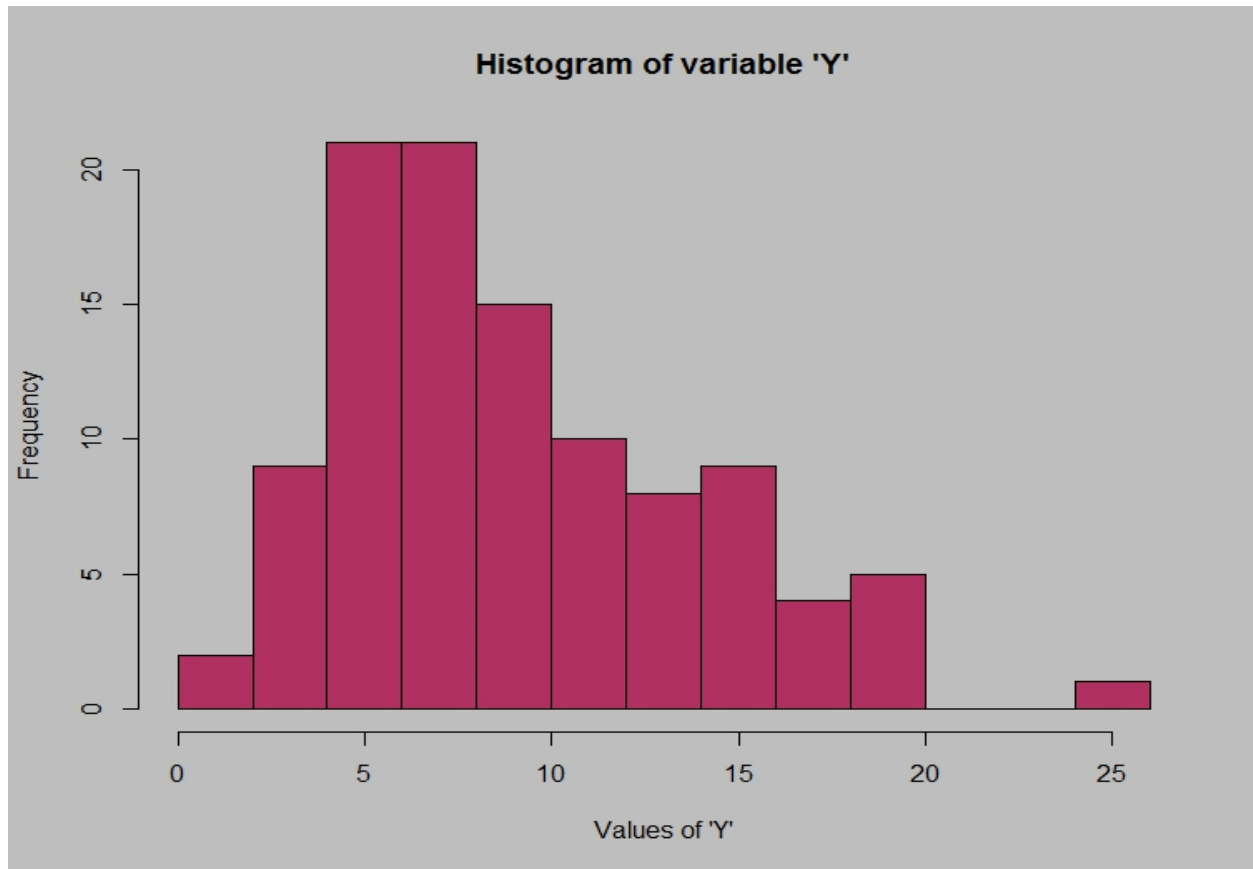
- (ii) What can we say about the skewness of this dataset?

Ans: The above box-plot shows positive/ right skewness where median value is lying towards the bottom of the box and whisker is more towards up side in comparison to bottom side. which means the concentration of data lies towards the lower side of the graph. There is also an outlier in the data.

- (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: If the value is 2.5 then it's mean and median value will be affected. Since it's value is changing from large number so it's mean value will be highly affected in comparison to median value. Since there is more chance mean value will be greater than the median value which results the distribution will right/positive skewed, it means still the distribution will be right skewed.

3.



Answer the following three questions based on the histogram above.

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

Ans: The mode will lie at value 5 (from above histogram) of Y-axis as seen. Bin right above 5 and next to it can be considered as the mode, as it seems same height having higher number of frequency count.

(ii) Comment on the skewness of the dataset.

Ans: The skewness is right/positively skewed as we can see the concentration of the data is more towards the lower side of Y-axis.

(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: From above two plots as we can see the concentration of the data is more between around 5 to 12 which more or less shows the middle 50% of the data lies between 5 to 12 the bin from histogram or dot from boxplot at 25 can be complemented as outlier in the data. Right skewness can be complemented with the whisker size on top of the box.

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 in the attached ipynb file.

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: Given in the attached ipynb file