# 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%
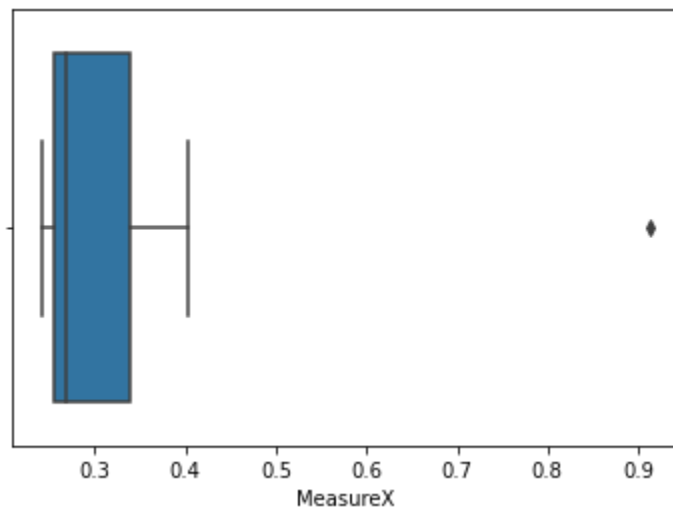
Ans) $mean - \mu$= **0.33** ,standard deviation - σ = **0.169** , variance - σ^2 = **0.028** , Outlier is **Morgan Stanley=91.36%**

Box-Plot of variable 'X'
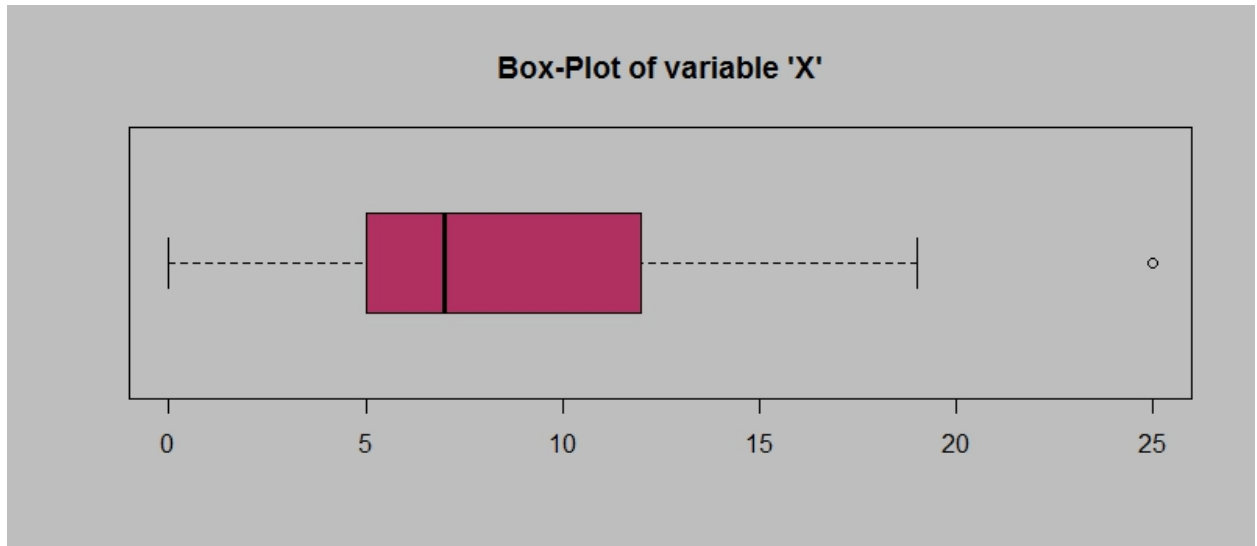
2.

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)   Upper limit Q3= 12.5 , Lower Quartile Q1= 5 , so IQR = Q3- Q1= 7.5 which implies Max Density of Data is 7.5 between 5 to 12.5

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

Ans)   The data is right Skewed which is positive Skewness.

(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 data will be Normally Distributed if we found out that the 25 was actually 2.5 which was previously an outlier affecting the data to be right skewed before.

3.

Histogram of variable 'Y'
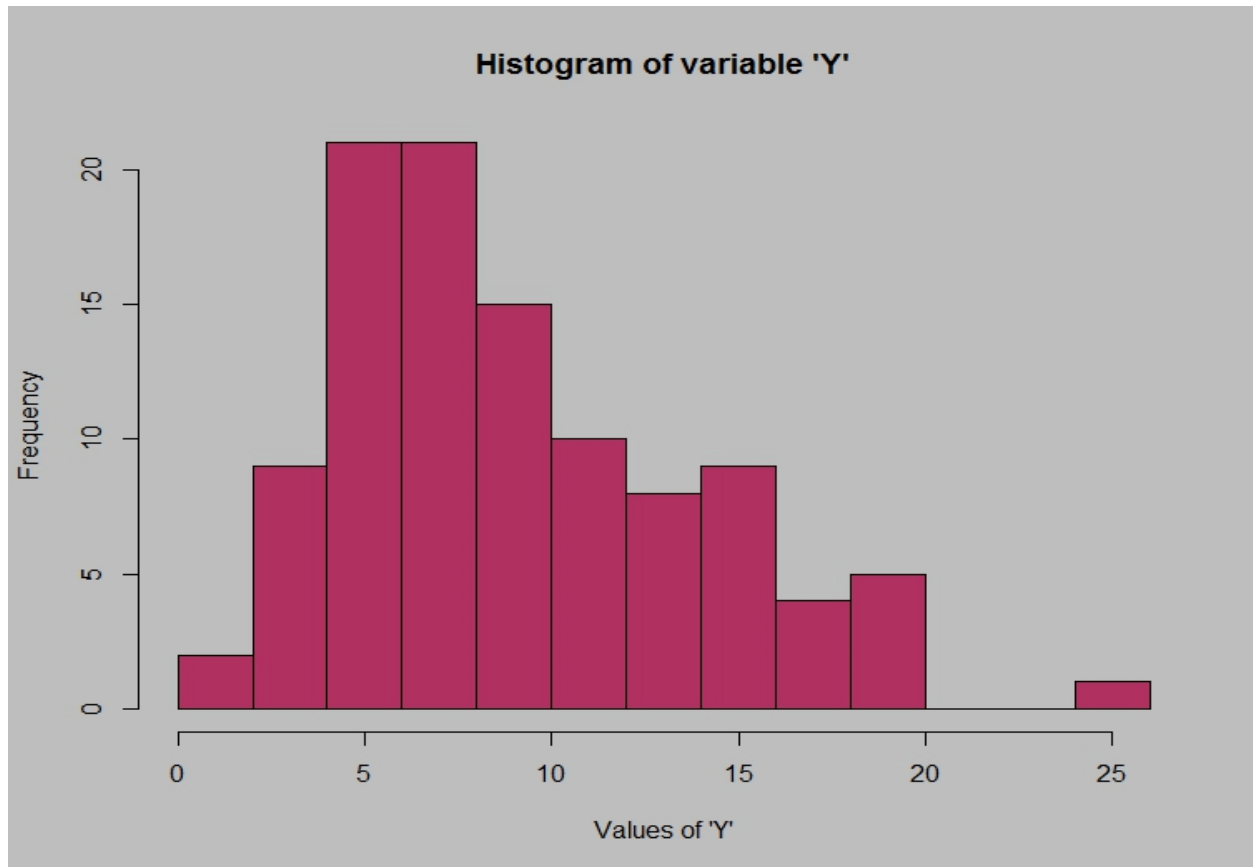
Answer the following three questions based on the histogram above.

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

Ans)   Mode of Dataset would lie between 4.5 -10.

(ii)   Comment on the skewness of the dataset.

Ans)   The Data is right Skewed which is Positive Skewness

(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)   In Boxplot we can't see the mode or frequency of Data distribution which Histogram Explains very well while Boxplot explains the Max Density of Data in Inter – Quartile Range very well compared to Histogram and Boxplot is best to find out the outliers which will help us to understand data more

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 in 200 long-distance telephone calls is misdirected

=> probability of call misdirecting p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/2

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

=1-(199/200)5**

=1-(0.995)5**

=1-0.975

=0.025, therefore the probability of 1 in 5 attempeted telephone calls reaches the wrong number is 0.025

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)   The most likely monetary outcome of the business is 2000$ as its probability is 0.3 which is higher than other probabilities
(ii)   Is the venture likely to be successful? Explain
Ans)   Yes, the venture is likely to be successful as the sum of probabilities of positive(profit) dollars is 0.2+0.3+0.1=0.6 which is 1000$+2000$+3000$ respectively.
(iii)   What is the long-term average earning of business ventures of this kind? Explain
Ans)

X	P(X)	E(X)= X . P(X)	X²	E(X²) = X² . P(X)
-2000	0.1	-200	400000	40000
-1000	0.1	-100	100000	10000
0	0.2	0	0	0
1000	0.2	200	200000	40000
2000	0.3	600	1200000	360000
3000	0.1	300	900000	90000
Sum=	1.0	800	2800000	360000

long-term average earning of business is E(X) which is

E(X) = sum of X.P(X) = $ 800

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

Ans) **Risk involved in a venture is**

Variance of (X) = E(X²) - (E(X))²

= 360000 -   800²

= -280000 (-ve value which indicates low risk)

Standard Deviation = sqrt Variance ≈ $ -529

  Risk is low even if we see the data given probability of getting profit >=0$ is 0.8 which is a good indication that if business venture doesn't work , there is only 0.2 probability to go in loss