

## Topics: Descriptive Statistics and Probability

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

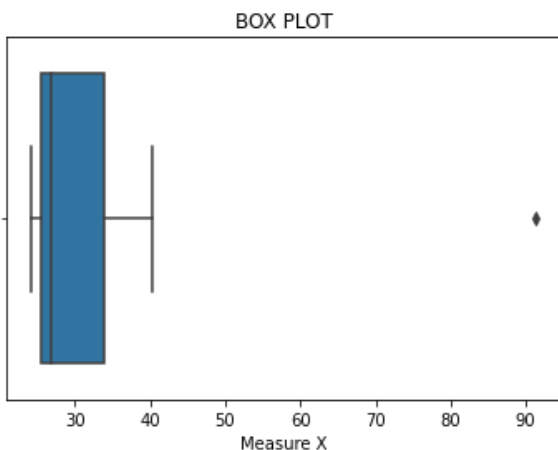
Name of company	Measure X
Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%
Warner-Lambert	35.00%

Mean of the data is 33.2713

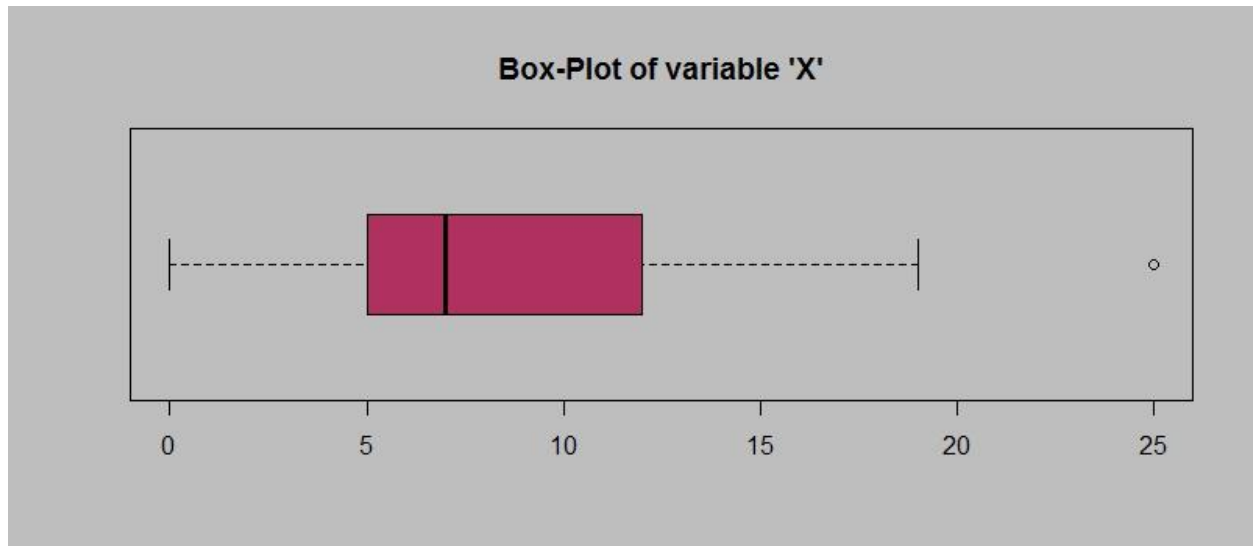
Variance of the data is 287.1466

Standard deviation of the data is 16.9454

Outlier in the dataset is [91.36]



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:  $IQR = Q3 - Q1 = 12 - 5 = 7$  (approximate)

- The middle 50% of the data lies in this range of 7 values. It can be used to find the outliers in the data.

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

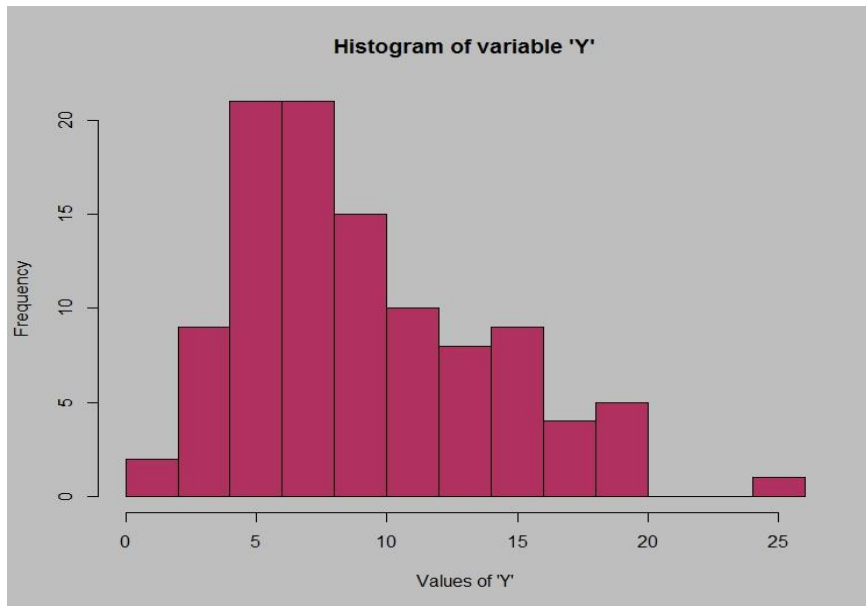
Ans: The data is positive skewed.

- (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?

Ans:

- The 2.5 value would fall in the lower whisker region.
- The Q1, Q3 values would change and move towards positive side.
- The median value increases.
- The data would be slightly decreased from its positive skewness.

3.



Answer the following three questions based on the histogram above.

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

Ans: Between 4 to 7 (approximately)

(ii) Comment on the skewness of the dataset.

Ans: The data is positively skewed(right skewed).

(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:

- Both the graphs have the value 25 as the outlier.
- Both the graphs are positively skewed.
- Both the graphs have median and mode at approximate in the same interval of data.

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:

$$P(x) = {}^nC_x p^x q^{n-x}$$

- $n = 5$
- $p = 1/200$
- $q = 199/200$

at least one in five attempted telephone calls reaches the wrong number

= 1 - none of the call reaches the wrong number

= 1 - P(0)

= 1 -  ${}^5C_0(1/200)^0(199/200)^{5-0}$

= 1 -  $(199/200)^5$

= 0.02475

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: 2000 is most likely monetary outcome of the business venture as it has the max probability.

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

Ans: Yes, the venture is likely to be successful as the weighted average is positive.

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

Ans:

x	P(x)	x*P(x)
-2,000	0.1	-200
-1,000	0.1	-100
0	0.2	0
1000	0.2	200
2000	0.3	600
3000	0.1	300
Long term average =		800

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

Ans:

x	P(x)	$x \cdot P(x)$	$(x - \mu)^2 P(x)$	
-2,000	0.1	-200	784,000	
-1,000	0.1	-100	324,000	
0	0.2	0	128,000	
1000	0.2	200	8,000	
2000	0.3	600	432,000	
3000	0.1	300	484,000	
	Long term average=	800	1,470	= standard deviation