

## Topics: Descriptive Statistics and Probability

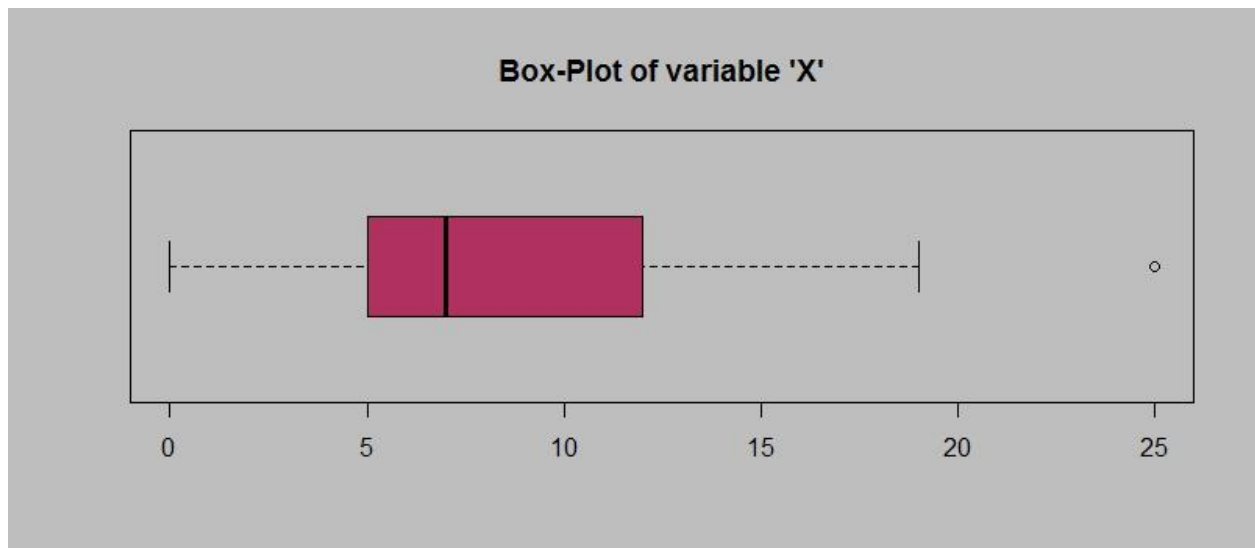
1. Look at the data given below. Plot the data, find the outliers and find out  $\mu, \sigma, \sigma^2$

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%

Answer-:

Please check attached notebook.

2.



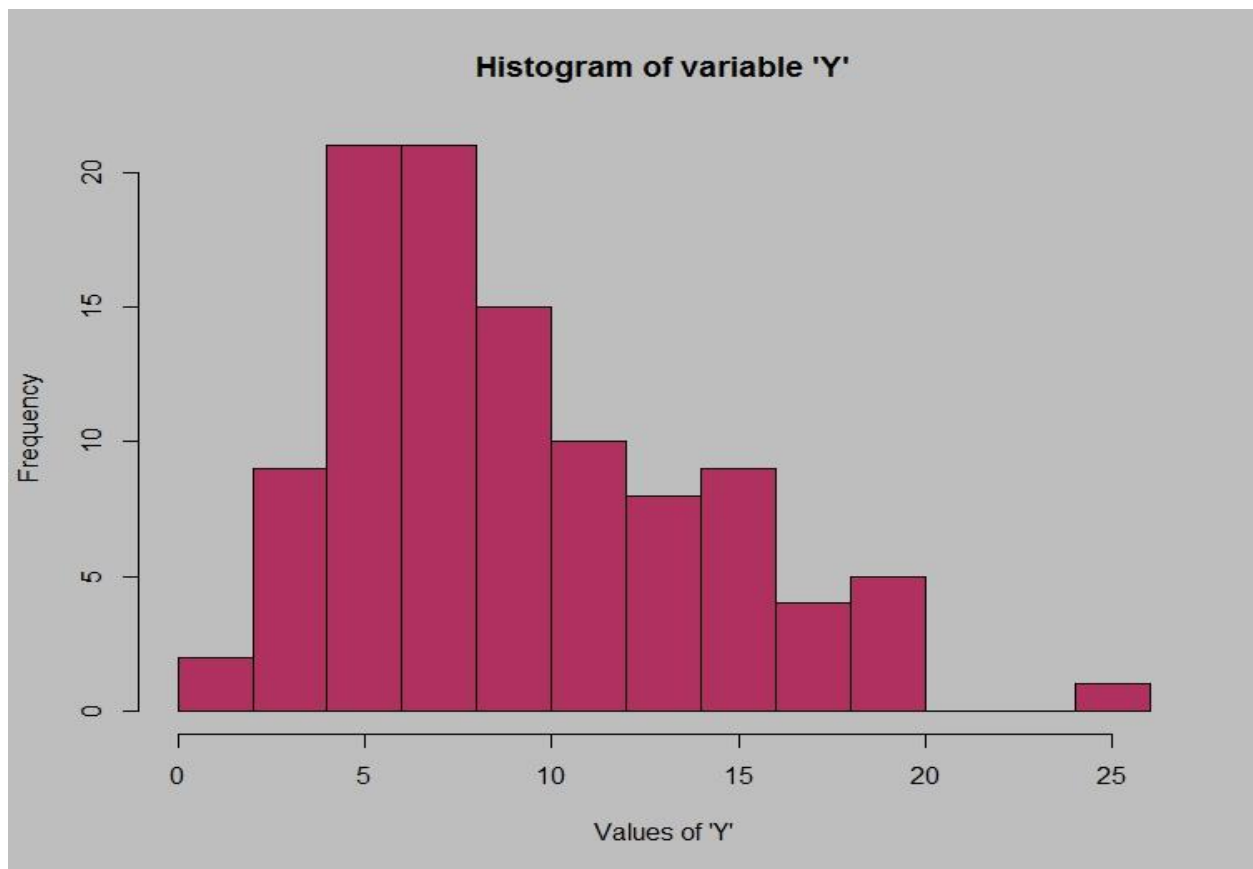
Answer the following three questions based on the box-plot above.

- (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?

**Answer:-**

- 1) Inter-quartile-range =  $Q3 - Q1 = 12 - 5 = 7$  (approximately).  
Almost 50% of data points lie in the range of 5 and 12
- 2) The data set is right side skewed, means positively skewed.
- 3) Box plot will get more skewed to right with this data point addition.  
No change in median value. Outlier get remove & data will more looks like zero skewness & normal distributed type.

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.

Answer-:

- 1) Within the range 4 to 9 the mode of data set lies.
  - 2) The data is positively skewed. Right tail data.
  - 3) Both the plot gives the information about the outliers. More data is lies more near to IQR-1
  - 4) Mode is more than mean value.
- 
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.)

Answer-:

If 1 in 200 long-distance telephone calls are get misdirected ,then

$P = \text{Probability of call misdirecting} = 1/200$

$q = \text{Probability of call not Misdirecting} = 1 - 1/200 = 199/200$

The probability for at least one in five attempted telephone calls reaches the wrong number of calls,

$x=1, n=5$

$P(x) = \text{at least one in five attempted telephone calls reaches the wrong number}$

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

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

$P(1) = ({}^5C_1) (1/200)^1 (199/200)^{5-1}$

$P(1) = 0.0245037$

probability that at least one in five attempted telephone calls reaches the wrong number=0.024503

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

**Answer-:**

- 1) It is the highest probability value 0.3, That is \$2000
- 2) From the given probability value, more than zero profit is more (i.e.  $0.2+0.3+0.1=0.6$ ) .  
Yes, venture likely to be successful.
- 3) The long term average of earning=  
$$= (-2000) * 0.1 + (-1000) * 0.1 + (0) * 0.2 + (1000) * 0.1 + (2000) * 0.3 + (3000) * 0.1$$
$$= \$800$$
- 4) Higher the variance, higher the risk. Depend on the variability of distribution.