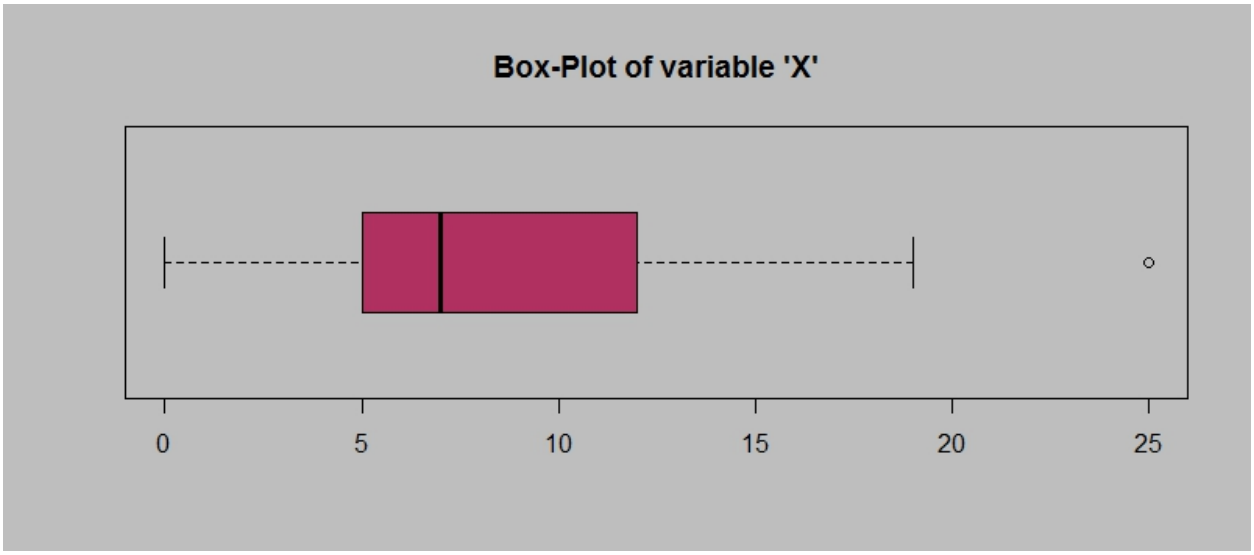
Topics: Descriptive Statistics and Probability

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%

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

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?

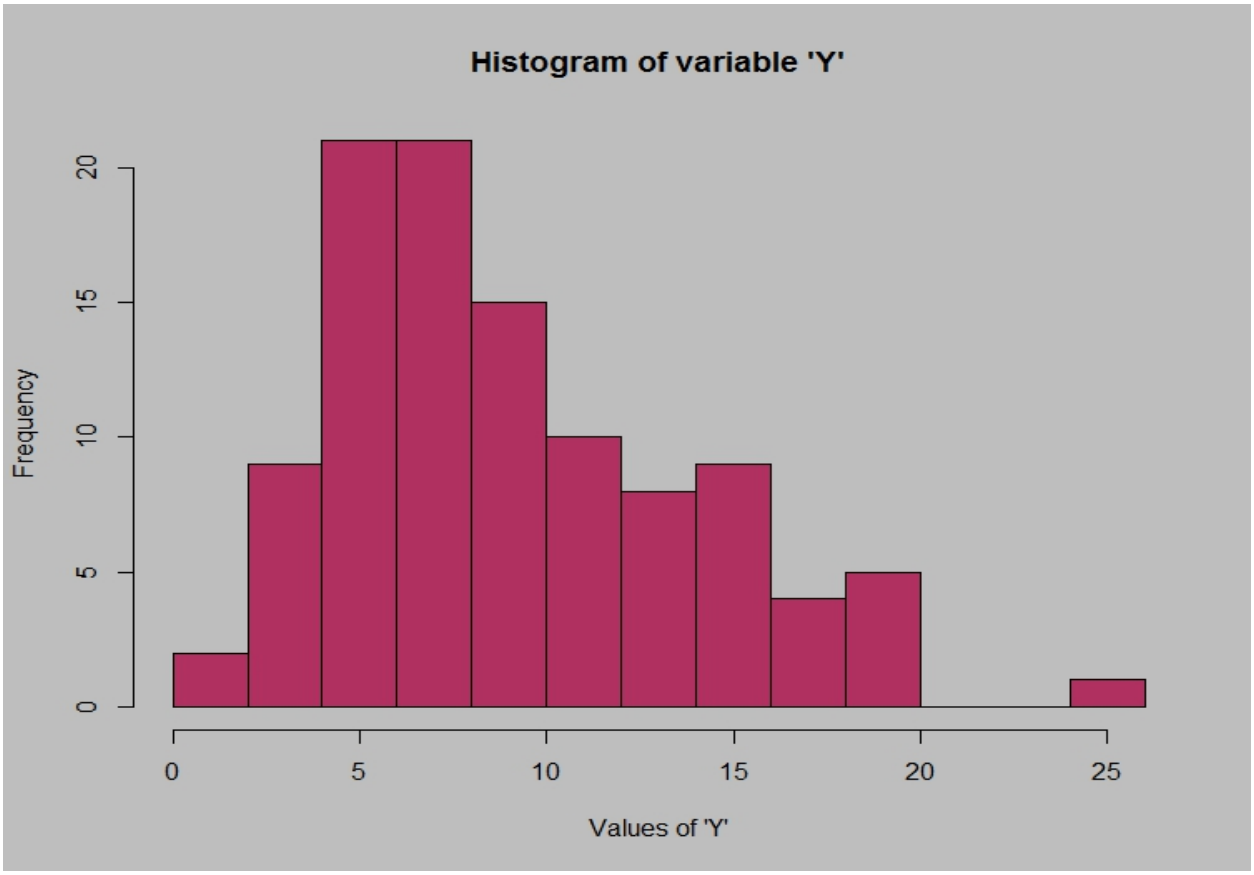
ANS:

1) **The inter-quartile range of the dataset is approximately Q3-Q1=12-5=7. This value is basically where the 50% of the data points of the dataset lie.**

2) **The dataset is clearly right-skewed or positively-skewed, as the median lies closer to the left side quartile and the right whisker is longer than the left one.**

3) **In that case, there would be no outliers at all and the data would not be right-skewed hence there would be normal distribution of the dataset.**

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.

ANS:

1) **The mode of the dataset would lie approximately between 4-8.**

2) **The data is clearly right-skewed as the dataset seems to have extreme points towards the right side or tends to be tailed towards right side. We can also estimate that mean>median.**

3) **The plots seem to be of same dataset as both the plots have 25 as outlier. The box plot helps in visualising the median, whereas histogram visualises the mode.**

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:

The probability of a single call reaching the wrong number is 1/200, and the probability of a call reaching the correct number is 199/200.

Since we are assuming independence of attempts, the probability of none of the five calls reaching the wrong number is (199/200)^5.

Therefore, the probability of at least one in five attempted telephone calls reaching the wrong number is 1 - (199/200)^5.

Calculating this value, we get approximately 0.0244, or about 2.44%.

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:

1) **The most likely monetary outcome of the business venture is x=2000, as it has the maximum probability of 0.3.**

2) **The venture will most probably be successful as the probability that the venture won't make losses is (from 0 to 3000) 0.2+0.2+0.3+0.1 = 0.8; I.e. the venture has 80% probability that it won't make losses.**

3) **The long term average can be calculated as:**
(-2000*01) + (-1000*0.1) + (1000*0.2) + (2000*0.3) + (3000*0.1) = 800. USD

4) **Risk involved:**

Var(x)=E(x^2)-[E(x)^2]

=[x^2*P(x)] - [800]^2

=2800000-640000

=2160000

Std Dev=sqrt(Var(x))

=1470

As Variance is very high (2160000), the risk involved is high.