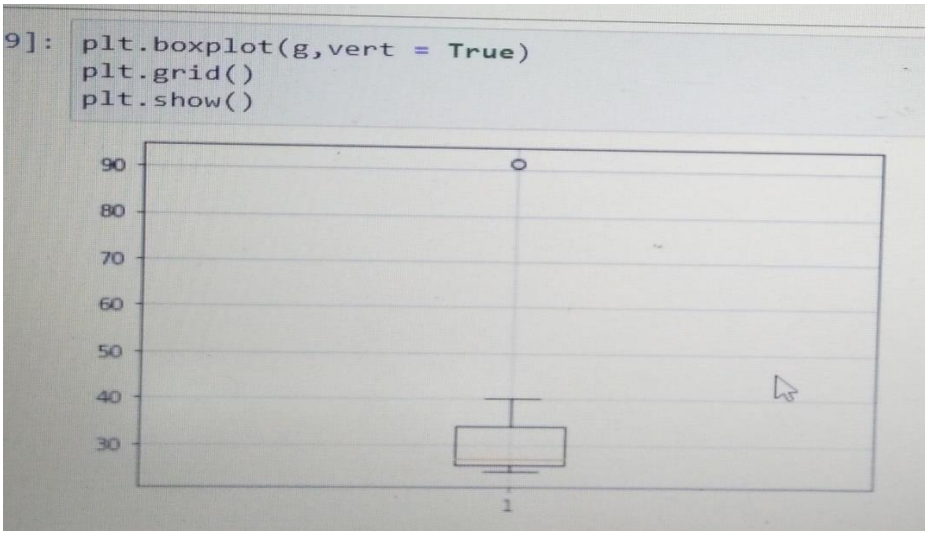
Topics: Descriptive Statistics and Probability

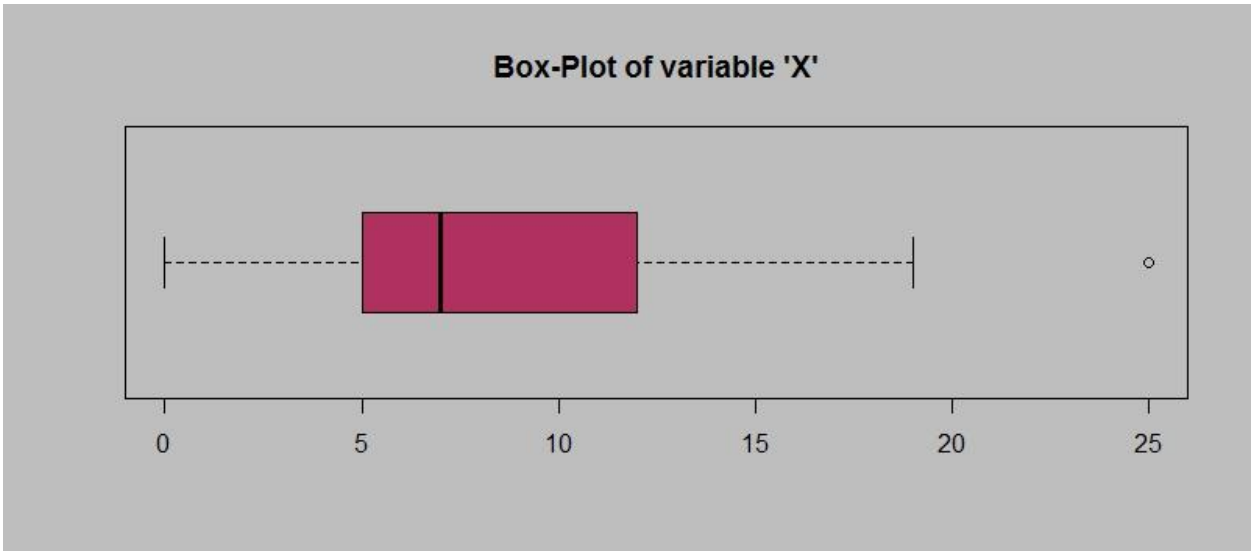
1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

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%



Outlier = Morgan Stanley = 91.36%
Mean = 33.271333
Standard Deviation = 16.945401
Variance = 287.1466123809524

2.

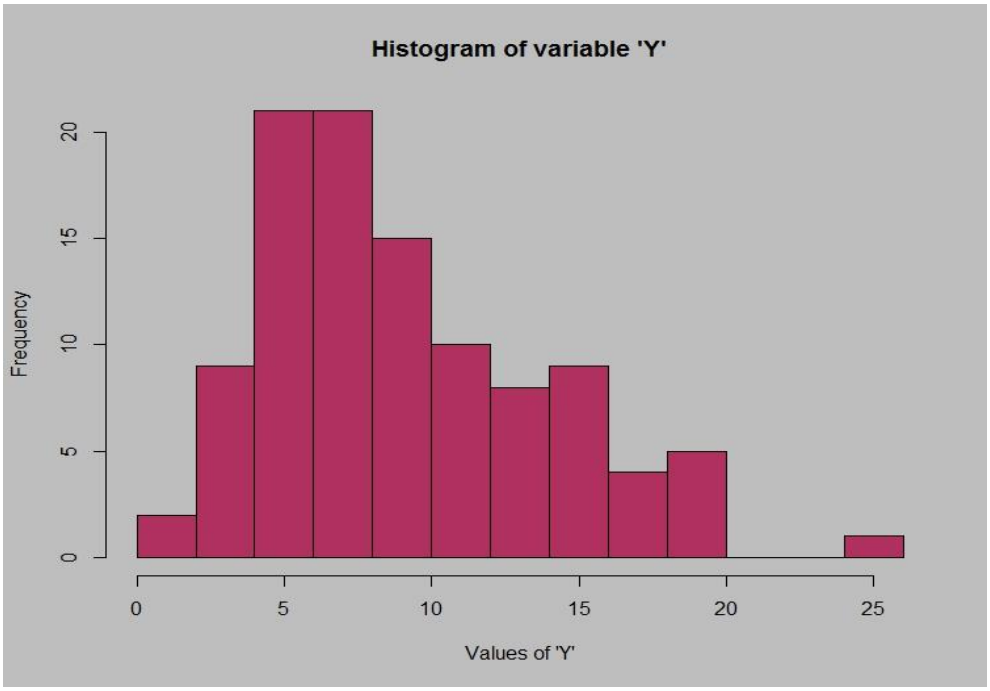


Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.
Solution:- Box plot is applied for continuous dataset. Inter quartile range contains middle 50% data. It has more dense area.
Inter-quartile range = $q_3 - q_1 = 12 - 5 = 07$

(ii) What can we say about the skewness of this dataset?
Solution:- This is negatively skewed data because mode>median>mean.

(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?
Solution:- Data point with the value 25 is an outlier. It has crossed maximum point. Due to this new box plot will be create.



3. Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?
Solution:- Mode of this dataset lies between 5 and 10.

(ii) Comment on the skewness of the dataset.
Solution:- Positive skewness as the tail of the graph lies towards right.

(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.
Solution:- Both the graphs are similar. Histogram gives only approximate values. These graphs provide skewness, mean, median, mode of the dataset.

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.)

Solution:- Probability that the call get misdirected , P=1/200

q=1-p = 199/200

n=5

x=1

P(x) = at least one in five attempted telephone calls reaches the wrong number

Binomial approximation gives,

$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$

P(x) = (nCx) (p^x) (q^n-x) # nCr = n! / r! * (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.0245037

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

Solution:- i) x=2000 is the most likely monetary outcome of the business venture.

ii) Risk probability is less than the profit probability. The probability of getting profit is 60%.

iii) Long term average earning = E(x)= x*p(x)

= (-2000*0.1)+(1000*0.1)+(0*0.2)+(1000*0.2)+(2000*0.3)+(3000*0.1)

= 800

iv) x=[-2000,-1000,0,1000,2000,3000]
np.std(x)
Standard deviation of x=1707.825127659933
np.var(x)
variance= 2916666.6666666665
The large value of variance is $2916666.666 indicates that this venture is highly risky.