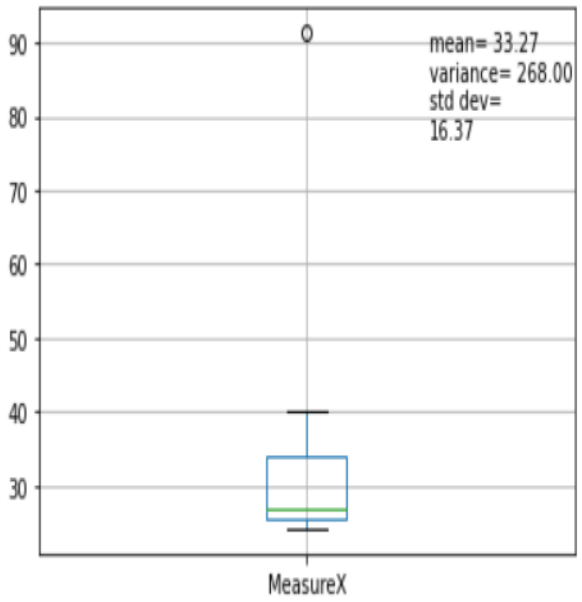
ASSIGNMENT 2

SET 1

MOHAMED IRFAN N

Topics: Descriptive Statistics and Probability

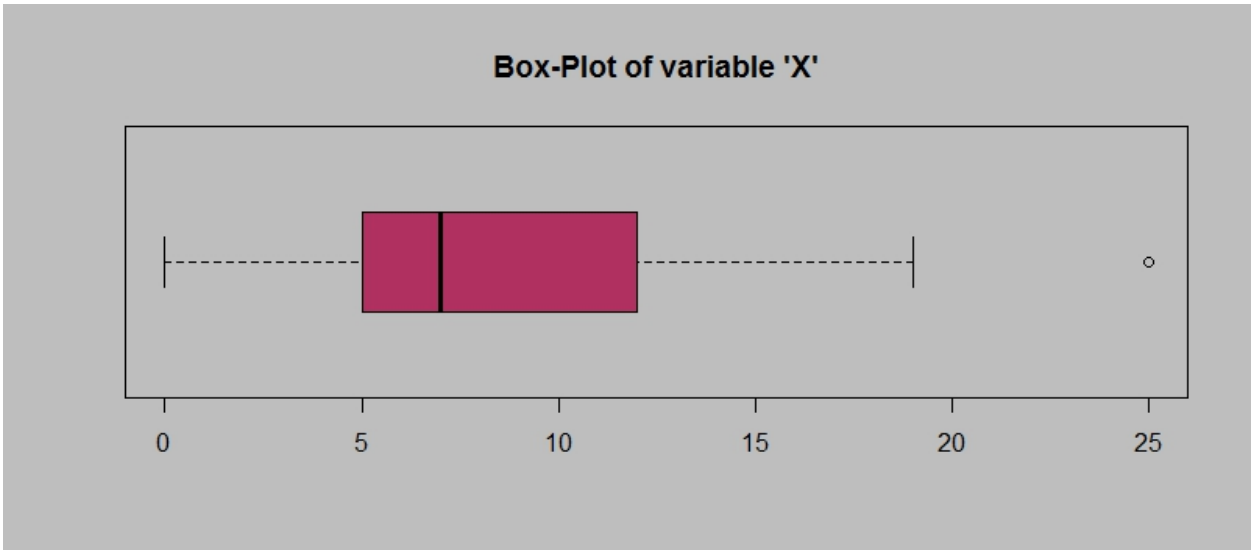
1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2



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%

Mean= 32.27%,
Variance = 268.00,
Std dev = 16.37%
Outlier = Morgan Stanley with 91.36% is the outlier

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: The IGR of the given data set is from 5 to 12 approximately. About 50% of all the data lies in this range and it doesn't not consider outliers or extreme values. Therefore, it could be considered to be more accurate than the range of the data

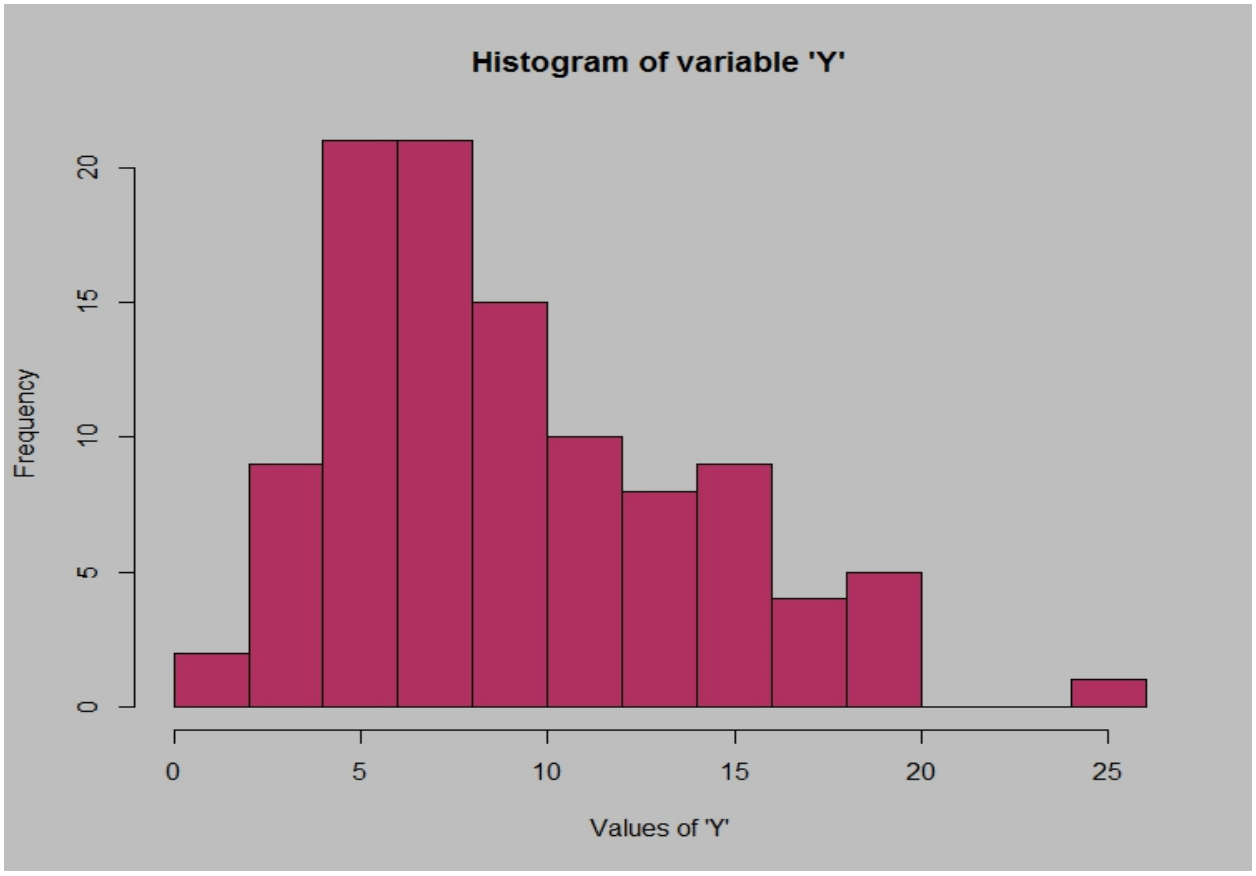
(ii) What can we say about the skewness of this dataset?

Ans: The median in this case would be lesser than the mean. Therefore, this is a right tailed plot with extreme values and outliers to the right side of the graph. The skewness of the dataset must be 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 outlier does not affect the box plot in any way. As per the question, if it is a 2.5, it comes in the lower fence of the boxplot. There could be a shift in the mean depending on the value. However the shift may not be large as 2.5 does not lie in the IQR range.

3.



Answer the following three questions based on the histogram above.

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

Ans: The mode would lie approx. between 4 and 8 where the frequency is greater than 20.

(ii) Comment on the skewness of the dataset.

Ans: It is a positive skewed right tailed distribution. There more positive values towards the right of the distribution.

(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 histogram and box-plot enables us to understand the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points. Histogram provides a better visualization to grasp the probability distribution of a data. Box plots can be more useful when we have to compare several datasets.

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: Probability of call getting misdirected P(E) = 1/200
Prob of call not getting misdirected = 1 – P(E) = 199/200
No of attempts = 5,
Prob that atleast one in 5 attempts misdirected calls P(X) = 1 – (prob that no calls misdirected in 5 attempts)
P(X) = 1 – ((199/200)*(199/200) (199/200)*(199/200) (199/200)) = 1 –((199/200)**5) = 1 - 0.975248753121875
P(X) = 0.02475124687812502 or approx. => 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: P(x) is highest for x = 2000. There it is the most likely monetary outcome of the business venture is x = 2000.

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

Ans: The venture is likely to be successful if P(x >0) is greater than P(x<=0).
P(x>0) = P(x = 1000) + P(x = 2000) + P(x = 3000) = 0.2 + 0.3 + 0.1 = 0.6
P(x<=0) = P(x = 0) + P(x = -1000) + P(x = -2000) = 0.2 + 0.1 + 0.1 = 0.4
P(x>0) > P(x<=0) . Therefore this venture is likely to be successful.

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

Ans: It is required to calculate the expected value of the business venture.
E(x) = (3000*0.1)+(2000*0.3)+(1000*0.2)+ (0*0.2)+ (-1000*0.1)+ (-2000*0.1)
= 300 + 600 + 200 + -100 + -200
E(x) = 800

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

Ans: Standard Deviation is a good measure the calculate the risk involved in this case.
Std Dev = 1469.6938456699068