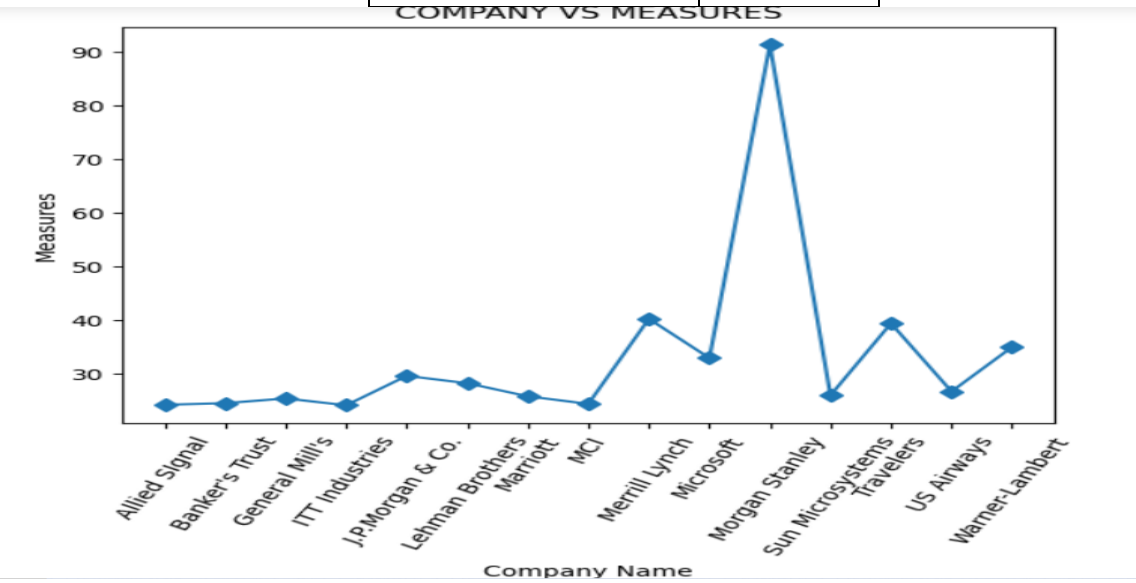
Topics: Descriptive Statistics and Probability

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%



REFER: Basic_Stats2.Set1.Q1.IPYNB

1.Mean, μ= $(\frac{1}{n}) \sum_{i=0}^{n} x_{1+} x_2 + \cdots + x_n$

$=$ (24.23+24.53+25.41+24.14+29.62+28.25+25.81+24.39+40.26+32.95+91.36+25.99+39.42+ 26.71+35.0)/15

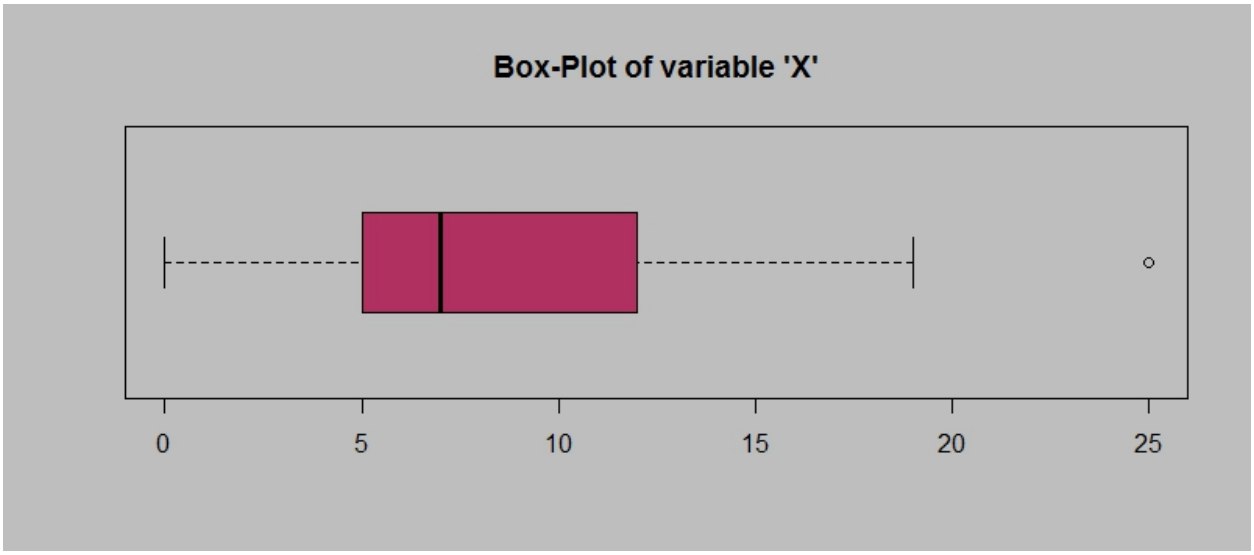
=33.20

2.Variance $\sigma^2 = {}_{i=1}^{n}\sum(X - \bar{X})^2./6$

- ⇨ Here, $\bar{X} = 33.20$
- ⇨ $((24.23 - 33.20)^2* + (24.53 - 33.82)^2* + \ldots\ldots + (35.0 - 33.82)^2)/15$
- ⇨ 269.07

3.Standard Deviation= $\sqrt{\sigma}$=16.40

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:

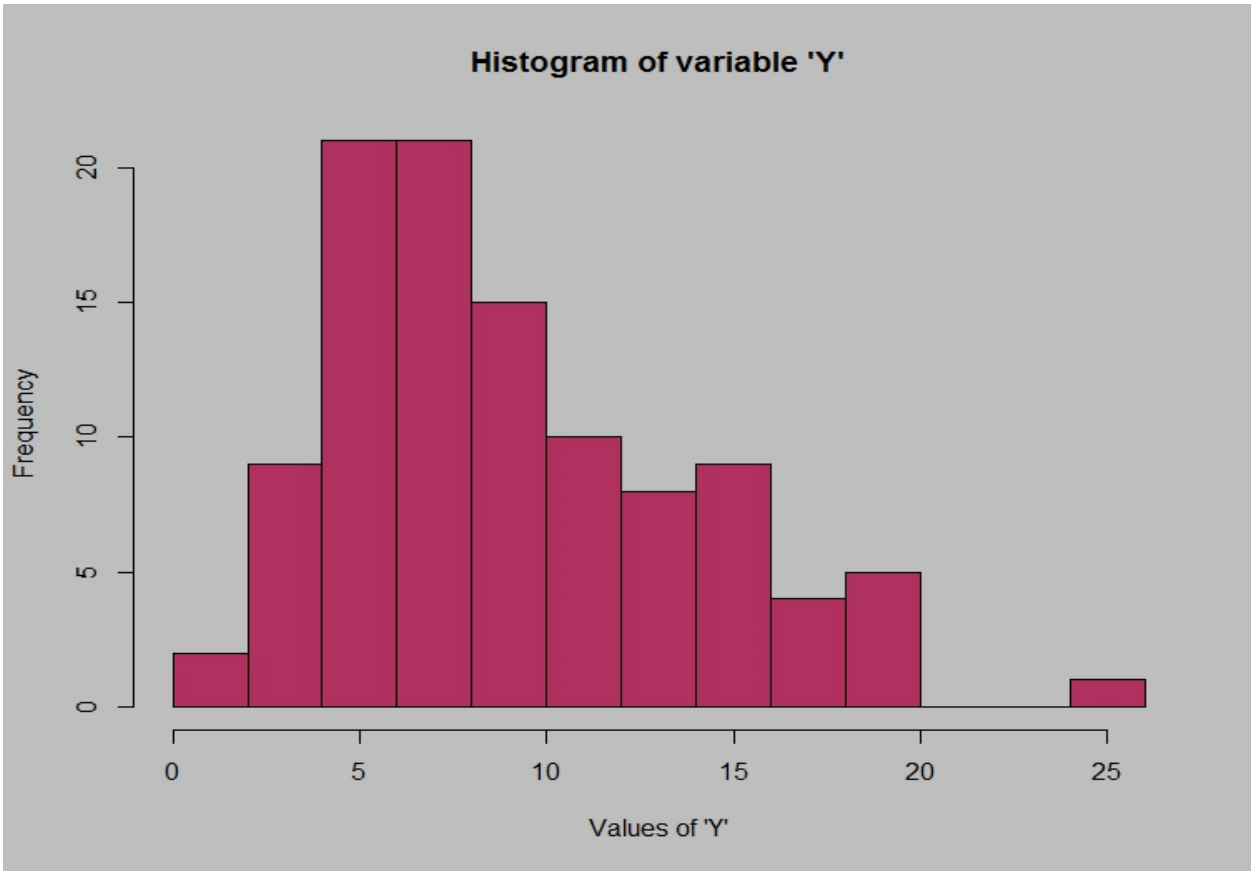
i)As per the data, Q1 is around 5 and Q3 is around 12, so IQR=Q3-Q1 which is around 7. The value 7 means the range or distribution of the data in the given data set.

ii)From the above plot it is clear that data distribution is right skewed and most of the data is located near to Q1 and median is located near to Q1.

iii)If the value of 25 is actually 2.5 …the value will be on the left side of the boxplot and will not be considered as outliers…. Also, the value of maximum value of outlier on right side will be Quartile3+(1.5*IQR) which means 22.5 and from the above plot the distribution is approximately till 20 … so there will be no outliers in the above distribution. (Taken the help of Q3 too to compute this question)

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

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:

i)Clearly from the histogram it appears that the distribution is bimodal and it has 2 modes ranging from 5-7 (two values) approximately

ii)The distribution appears to be right skewed and there seems to some outliers on the right end.

iii)From the Q2 Boxplot and this histogram, the distribution is same …. It is reflected by the same type of skewness i.e., right skewness as there are outliers on the right side of distribution. In general, in right skewed distribution, Mean>Meridian > Mode …. Here in this distribution, it is not clear of those values but it approximately shows Mode as 5-7 and Median around 7 …. hence there is more compliments in both plots.

3.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: Here total calls =200, one call is misdirected and others are right

So, in p(success(misdirected))=1/200, q(failure)=199/200, n(samples)=5, x(failure)=atleast 1

For Probability of an event=P(x)= ${}^{n}C_{x} * P^{x} * Q^{(n-x)}$

Here P (at least 1 failure out of 5 calls) + P (no failure calls) = 1 …… (Total Probability=1) -------(1)

=> P (at least 1 failure out of 5 calls) = P (1) + P (2) + P (3) + P (4) + P (5)
=> P (no failure calls) = P (0)

From equation 1,
Hence P (at least 1 failure out of 5 calls) = 1 - P (no failure calls)
P (at least 1 failure out of 5 calls) = 1 - ${}^{5}C_{0} * (\frac{1}{200})^{(0)} * (\frac{199}{200})^{(5-0)}$
P (at least 1 failure out of 5 calls) = 1 – (1*1*0.975248)
P (at least 1 failure out of 5 calls) = 0.024752.

4. 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: Here they have given the probability of returns in a certain business venture …. It involves both probabilities of positive and negative returns.

i) The most out likely outcome of the venture is similar to the Mode of a distribution, if the same distribution is analyzed it shows the out of other values $2000 is having the most probability 0.3 … then is it is most likely outcome of that certain business venture.

ii) Here in order for a venture to be successful…. We must compute the probability …i.e., Difference of Positive return probability and negative return probability…. And the venture will be only successful only when the Positive probability is positive…
Here positive returns are (0,1000,2000,3000) and negatives are (-1000, -2000)
- ⇨ P (To be successful) =P(Positive)
- ⇨ (0.2+0.2+0.3+0.1)
- ⇨ 0.8
- ⇨ 0.8, which shows there is 80% possibility for that venture to be successful.

iii) Long time average can be calculated by summing up all probabilities multiplied by its own return
- ⇨ (1000*0.2) +(2000*0.3) +(3000*0.1) +(-2000*0.1) +(-1000*0.1)
- ⇨ 800, which clearly shows in a given period … for example these returns calculated for one year …. Whatever profit or loss the final outcome when divided by given period shows the venture is having a 600 profit a month after the final computation.

iv) The risk which involved how far the returns deviate from one time to other …. And that deviation can be measured by standard deviation
Variance=$_{i=1}^{n}\sum(\mathrm{X}-\bar{X})^2.P(X)/6$
Here, $\bar{X}$= 600, mean of the distribution

- ⇨ $((-2000-600)^2*0.2+(-1000-600)^2*0.1+(0-600)^2*0.2+(1000-600)^2*0.2+(2000-600)^2*0.3+(3000-600)^2*0.1)/6$
- ⇨ Variance= 2160000
- ⇨ Std.Deviation=$\sqrt{Variance}$
- ⇨ 1469.69, this is the risk from period to other in the venture.