# Session 8 Assignment 1

**Question**

1. mins 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 1 ->**

The R code to describe the above question is

```
>noc<-c("Allied Signal","Bankers Trust","General Mills", "ITT Industries","jP Morgan","Lehman
Brothers","Marriott","MCI","Merrill Lynch","Microsoft","Morgan Stanley","SunMicros","Travelers","US
Airways","Warner-Lambert")
>measure<-c(24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00)
>data<-data.frame(noc,measure)
>data
>plot(data, main="Company & Measure X", xlab ="Name of Company")
>sd(measure)
>var(measure)
>y<-sd(measure)
>z<-<-y^2
>z
```

Using this the data frame we get is like below :

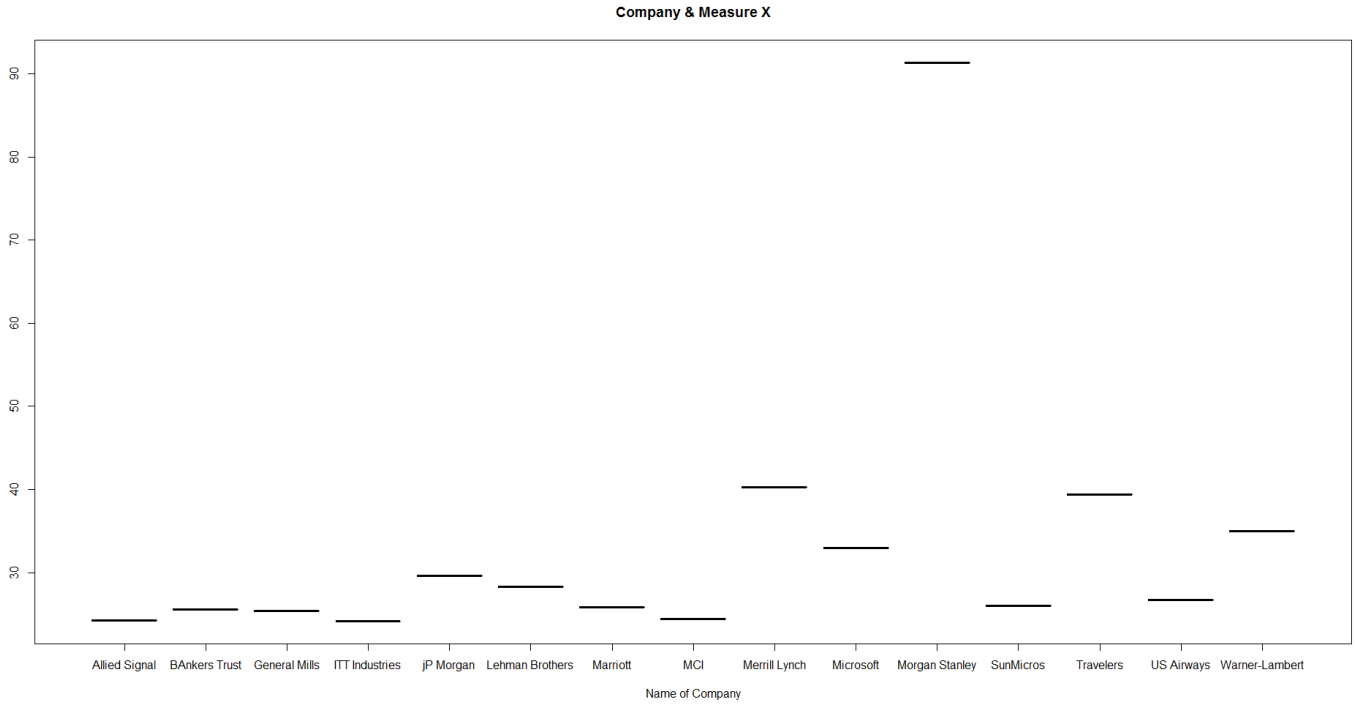
| | noc | measure |
|---|---|---|
| 1 | Allied Signal | 24.23 |
| 2 | BAnkers Trust | 25.53 |
| 3 | General Mills | 25.41 |
| 4 | ITT Industries | 24.14 |
| 5 | j.P. Morgan & Co. | 29.62 |
| 6 | Lehman Brothers | 28.25 |
| 7 | Marriott | 25.81 |
| 8 | MCI | 24.39 |
| 9 | Merrill Lynch | 40.26 |
| 10 | Microsoft | 32.95 |
| 11 | Morgan Stanley | 91.36 |
| 12 | Sun Microsystems | 25.99 |
| 13 | Travelers | 39.42 |
| 14 | US Airways | 26.71 |
| 15 | Warner-Lambert | 35.00 |

**Secondly** the values of **Standard Deviation** & **Variance** and **Standard Deviation Square** is given as below

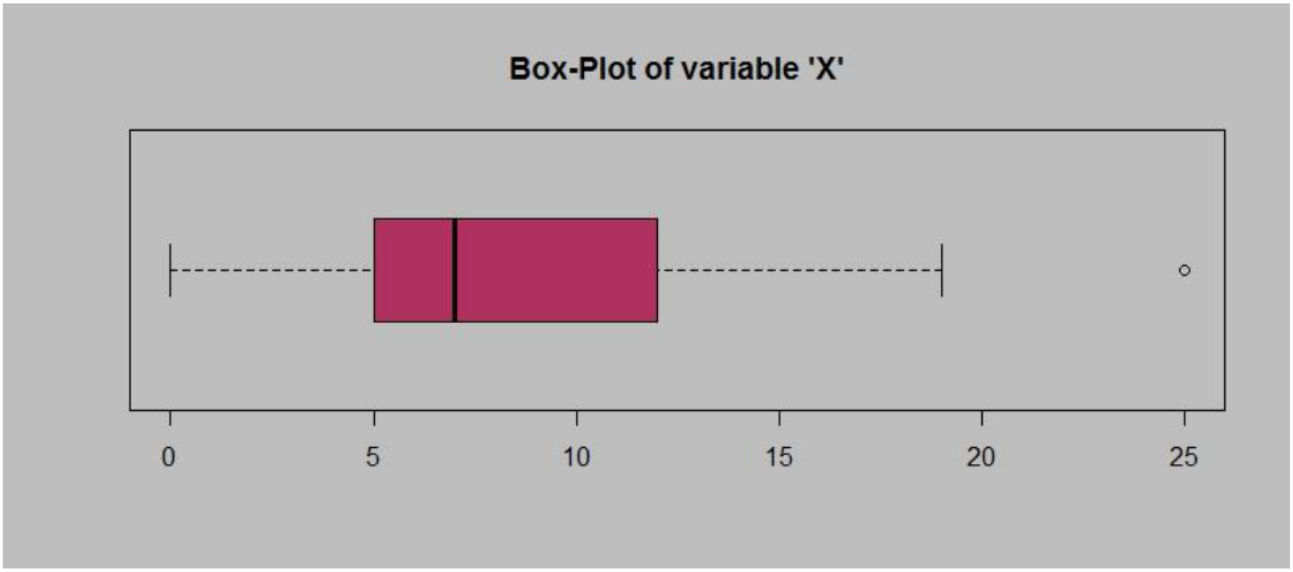
```
> var(measure)
[1] 287.1466
> sd(measure)
[1] 16.9454
```

Also, the value of (Sd)^2 will be same as that of variance which is equal to 287.1466

There is only one outlier in the data which is for **Morgan Stanley,** The data is plotted as shown below



**Question 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?

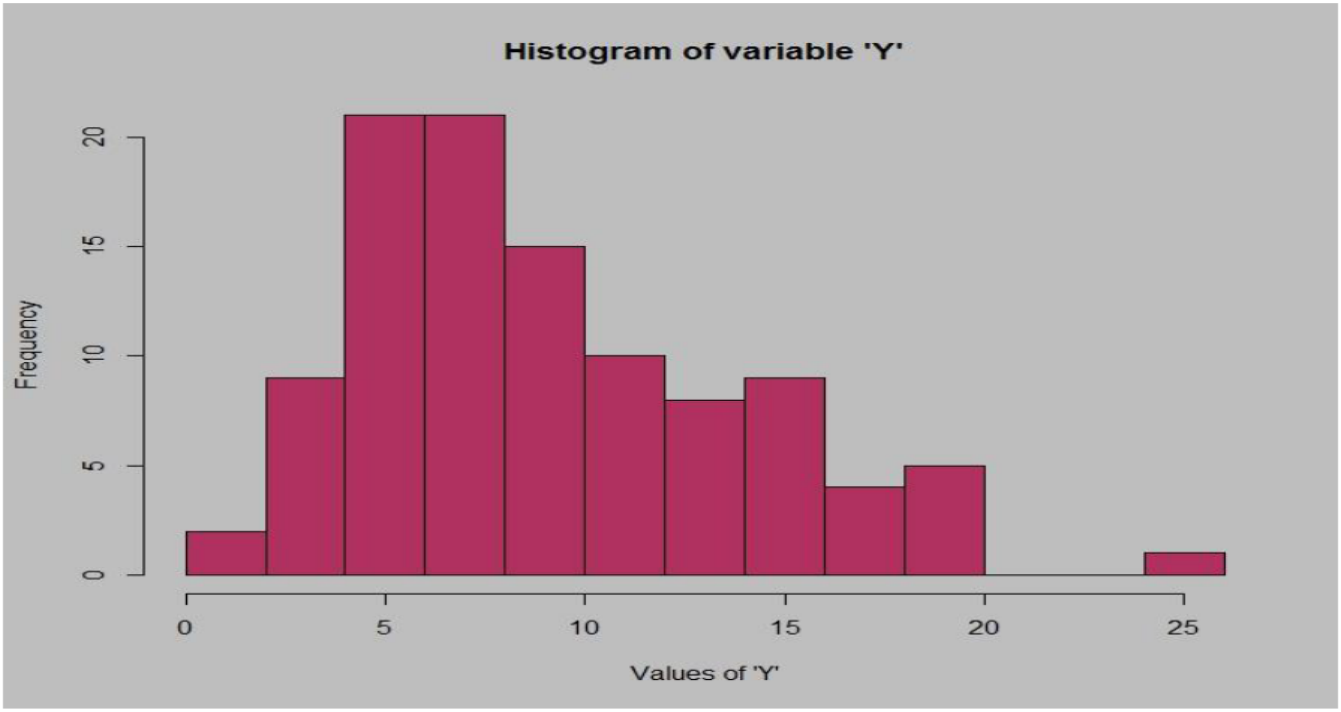
Answer 2:

i) The inter quartile range o the box plot shown here is difference of third quartile - to first Quartile which is given as Q3- Q1 . In this case Q3 = 12 & Q1= 5 Hence inter quartile Range is equal to 12-5 = 7.
The Inter Quartile Range gives us a better measure of spread of the data than the range as it is not affected by outliers present in the data set.

ii) As we can see that the from the box plot that the data in the box has median more towards the left side of box i.e. towards Quartile 1 ..Hence we can say that the data is skewed positively i.e. towards right from the median

iii) If the outlier that is 25 is in actual 2.5 we will say that it would not make much of a difference just that the median will just shift a little towards left .Hence we an say that the 2nd quartile i.e. Median will a little left towards 1st Quartile.

**Question 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 3

i) The Mode of the data set would lie in the between the range 4-8

ii) Skewness of the data set is towards the right. As we can say that there is tail formation towards the right as shown in the histogram.

iii) If the histogram is of the same data set as the boxplot in the question 2. These two graphs complement each other by stating same information that

a. 25 is the outlier in both the cases

b. **Data is skewed towards the right.** Boxplot and Histogram tells us whether the data is symmetric or not i.e it tells us the measure of skewness of the data.

c. Value of Median can be found out using both type of graphs.

**Question 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 4:**

- P(telephone call is mis-directed) = 1/200
- P(telephone call is not mis directed) = 1 - 1/200
  = 199/200

➢ At least 1 telephone call is mis-directed means that either 1 or 2 or 3 or 4 or 5 calls are mis-directed

we can say that,

**P(at least one telephone is misdirected) = 1 – P(no telephone call is mis-directed)**

**Now,**
**P (No telephone call is mis-directed )** = (1- 1/200)^5
= (199/200)^5
= (0.995)^5
= 0.9752488

**P(at least one telephone is misdirected)= 1 - 0.9752488**
**= 0.2472**

**Question 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 5 :**

**i)**
Most Likely monetary Outcome of the business venture will be :

= Summation x* P(x)
= **(**-2000* P(-2000)**)** + **(** -1000* P(-1000) **)** + **(**0 * P(0)**)** + **(**1000* P( 1000) **)** + **(**2000* P(2000) **)** + **(**3000* P(3000) **)**
= **(**-2000* 0.1**)** + **(** -1000* 0.1 **)** + **(**0 * 0.2**)** + **(**1000* 0.2 **)** + **(**2000* 0.3 **)** + **(** 3000* 0.1**)**
= 800

Hence, we can say that most likely monetary outcome of business venture will be **$800** .

**ii)**
Yes , as per the data given in form of Probabilities the business venture is likely to be successful as probability of positive returns is more than 0.6 i.e chances are 60% for the venture to have postivive return, whereas chances of negative returns are given to be 20% .It is give that 10% chance is that the company will have no Profit No loss Scenario

- P(Positive earnings ) = 0.2 +0.3+0.1
  =0.6
- P(Negative Earnings) = 0.1 + 0.1
  = 0.2

iii) Long term Average Return is basically given by E(x) which is Expected return in long terms .It is given by the formula **E(x)= ∑ x. P(x)** , Which after calculating comes out to be **$800**

iv) Probability distribution provides the basis for measuring the risk of a project. **"the higher the probability distribution of expected future return, the smaller the risk of a given project and the vice versa.".** One of the good measure of risk involved in such type is venture is calculating **Standard Deviation**
The method of finding Standard deviation is as follows

(a) Calculate the mean of expected value of the distribution.

(b) Calculate the deviation from each possible outcome.

(c) Square each deviation.

(d) Multiply the squared deviations by the probability of occurrence for its related outcome.

(e) Sum all the products. This is called variance.

(f) The Standard deviation is square root of Variance.

The smaller the standard deviation, the higher the probability distribution and accordingly the lower the riskiness of the project.

| x | p(x) | 800-x | (800-x)^2 | P(x)* (800-x)^2 | S.d= sqrt(sum/n-1) |
|---|---|---|---|---|---|
| -2000 | 0.1 | 2800 | 7840000 | 784000 | |
| -1000 | 0.1 | 1800 | 3240000 | 324000 | |
| 0 | 0.2 | 800 | 640000 | 128000 | |
| 1000 | 0.2 | -200 | 40000 | 8000 | |
| 2000 | 0.3 | -1200 | 1440000 | 432000 | |
| 3000 | 0.1 | -2200 | 4840000 | 484000 | sd= |
| 3000 | | | sum | 2160000 | 657.267069 |
| | | | sum/n-1 | 432000 | |
| | | | | n=6 | |
| Expected Value= 8oo | | | | n-1= 5 | |

Here we can say that Standarrd Deviation as a measure of risk comes out to be $ 657.26