**Topics: Descriptive Statistics and Probability**

1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |

**Ans**- mean , sigma is sd and sigma square is variance

**Outlier is Morgan Stanley 91.36%.becuase all the company measure are near by 24% to 40% except morgan Stanley.**

> #Finding outliers

> company=read.csv(file.choose())

> mean(company$X)

[1] 0.3327133

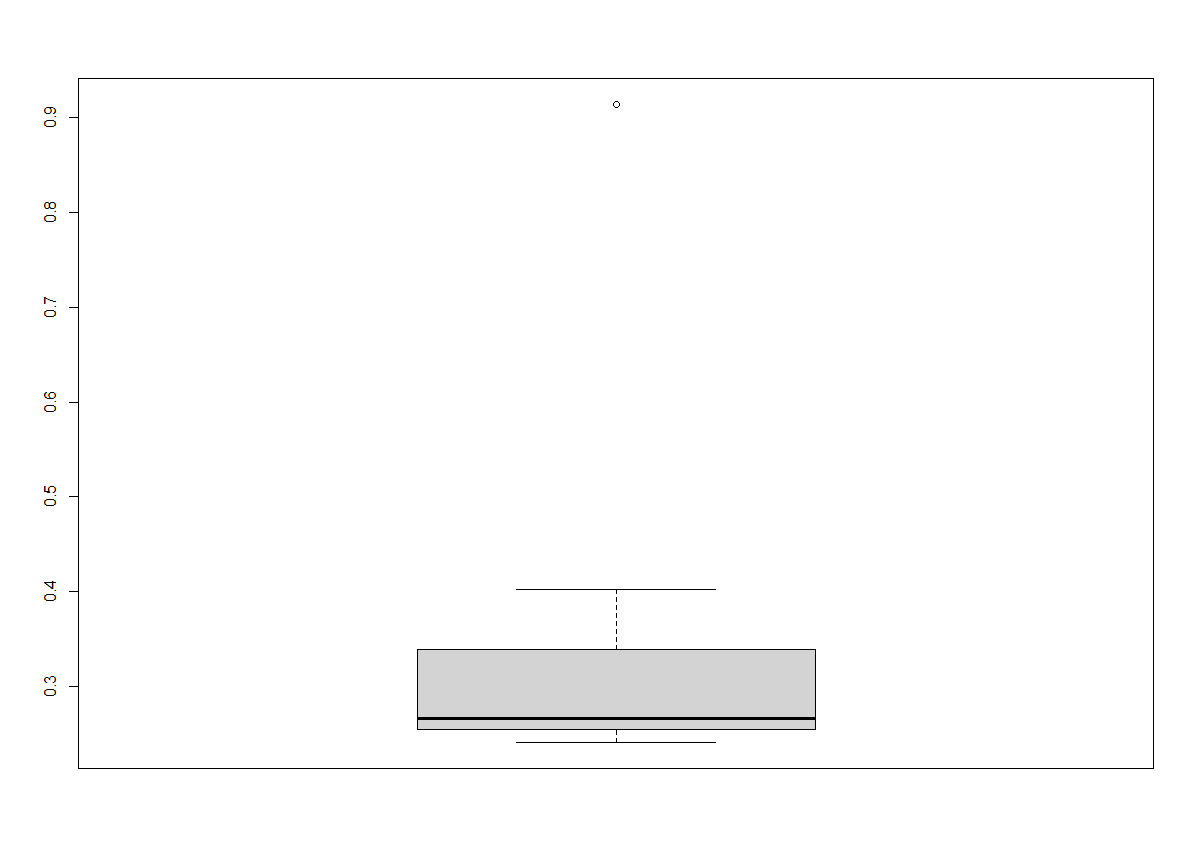
> sd(company$X)

[1] 0.169454

> var(company$X)

[1] 0.02871466

> boxplot(company$X,horizontal = FALSE)





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

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

**Ans**-**IQR = Q3-Q1 here approximately Q3 =12 and Q1 = 5**

**= 12-5**

**=7**

**It means that the data is distributed around from 5 to 12 with median 7. 50% data lies between IQR.**

1. What can we say about the skewness of this dataset?

**Ans** -**So if data is going beyond-6.5 or 23.5 it will be outlier and Skewness of data is positively skewed because most of the data is lying towards right hand side.**

1. 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- If the value 2.5 instead of 25 box plot would be plotted within range and there will be no outlier because 2.5 will come between the 0 to 20 range and Q1 Position will move to left side to 2.5 value .**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**ANS**- **MODE of data set lies between 4 to 16 data values of ‘Y’ . There are two modes which are frequently occurred that are of frequency 9 and frequency 21 approx.**

Comment on the skewness of the dataset.

**ANS**- **Most of the data aligned towards right hand side so plot is positively skewed. There is also present oulier.**

1. 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**- **By plotting boxplot we can find out outlier, Inter quartile range, lower quartile, upper quartile and median and also we can find out the skewness of the data. In this data boxplot showing positively skewed because data is lying towards right hand side of the plot. By plotting histogram we can find the mode easily and also we can determine the skewness and Kurtosis of the data. By plotting both boxplot and histogram it will more useful to make interferences about the dataset.**

1. 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**- **In 200 call 1 call is definitely going to wrong number, so total call left 199**

#expected value for calls

> 199/200\*1+199/200\*2+199/200\*3+199/200\*4+199/200\*5

[1] 14.925

> 14.925/5

[1] 2.985

1. 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 |

1. What is the most likely monetary outcome of the business venture?

**Ans- The most likely outcome of this business venture is a return of $2000 as it has the highest probability that is 0.3.**

1. Is the venture likely to be successful? Explain

**Ans- If the venture can maintain for long term business then eventually it will be successful since the probability of non-negative return is higher than 0.50 and the expected value for return is a positive number ($800). The probability distribution gives us an idea about the long-term chances of earning given values of returns (indicated by x). therefore, there is a 60% probability that the venture would be successful. (Note: 0.2+0.3+0.1=0.6=>0.6\*100=>60%).**

1. What is the long-term average earning of business ventures of this kind? Explain

**Ans- Expected value−$2000 ∗ 0.10 − $1000 ∗ 0.10 + 0 + $1000 ∗ 0.20 + $2000 ∗ 0.30+ $3000 ∗ 0.10 = $800**

**Therefore the long-term average earning for these type of ventures would be around $800.**

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure

**Ans- According probability 60% chance is that the venture will go in profit and 20% change venture will go in loss,so we can say that risk involved in the venture .**

**Therefore a good measure to evaluate the risk for a venture of this kind would be variance or standard deviation of the variable X.**

**> sd(ex$x)**

**[1] 1870.829**

**> var(ex$x)**

**[1] 3500000**

**The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky.**