**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% |

***R-code***

library(stringr)

company<-read.csv(file.choose())

View(company)

company$ValueX <- as.numeric(str\_remove(company$Measure.X, "[%]"))

#Remove ValueX column

#company <- subset(company, select = -c(ValueX))

cat("\n Mean of Company = ",mean(company$ValueX), "\n")

cat("\n Standard Deviation of Company = ",sd(company$ValueX), "\n")

cat("\n Variance of Company = ",var(company$ValueX), "\n")

boxplot(company$ValueX)$out

cat("\n Outlier of the data is = ",boxplot(company$ValueX)$out, "\n")

boxplot(company$ValueX, main = "Boxplot MeasureX", col = "LightBlue", pch = 19)

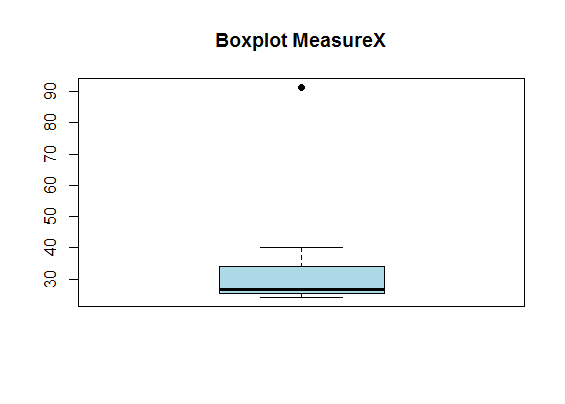
***R-output***

Mean of Company = 33.27133

Standard Deviation of Company = 16.9454

Variance of Company = 287.1466

Outlier of the data is = 91.36





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.

Quartile Range = Max-Min = 19-0 = 19

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

The data has a long tail on the right hence the data is positively/right skewed.

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?

If the new data point is 2.5, then the mean will change, median, IQR would not be affected much



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The highest frequency are from points 4 to 8, Thus the mode would lie between values 4 to 8

1. Comment on the skewness of the dataset.

The data has a long tail on the right hence the data is positively/right skewed.

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.

Boxplot will point out the outliers. We can find the median and the quartiles from a boxplot and we get the mode from a histogram.

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.)

Since 1 in 200 calls are misdirected, the probability of a call getting misdirected = 1/200 = 0.005

Thus probability of a call not getting misdirected = 1 – 0.005 = 0.995

P(atleast one in five attempted calls reaches wrong number) = 1 – P(no call in 5 attempted reaches wrong number)

= 1 – (0.995\*0.995\*0.995\*0.995\*0.995)

= 0.0248

Probability that at least one in five attempted telephone calls reaches the wrong number = 0.0248

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?

Since the highest probability of 0.3 is for $2000, the most likely monetary outcome = $2000

1. Is the venture likely to be successful? Explain

We find the Expected Value

Expected Value = -2000\*0.1 + -1000\*0.1 + 0\*0.2 + 1000\*0.2 + 2000\*0.3 + 3000\*0.1

= 800

Since $800 is the expected value and it is positive, the venture is likely to be successful.

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

We find the Expected Value

Expected Value = -2000\*0.1 + -1000\*0.1 + 0\*0.2 + 1000\*0.2 + 2000\*0.3 + 3000\*0.1

= 800

Long term average earning = $800

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