**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: Below code is done in R language**

**library(readxl)**

**> company <- read\_excel("D:\\DataScience\\Assignements\\company.xlsx")**

**> View(company)**

**> mean(company$`MeasureX`)**

**[1] 0.332**

**> sd(company$MeasureX)**

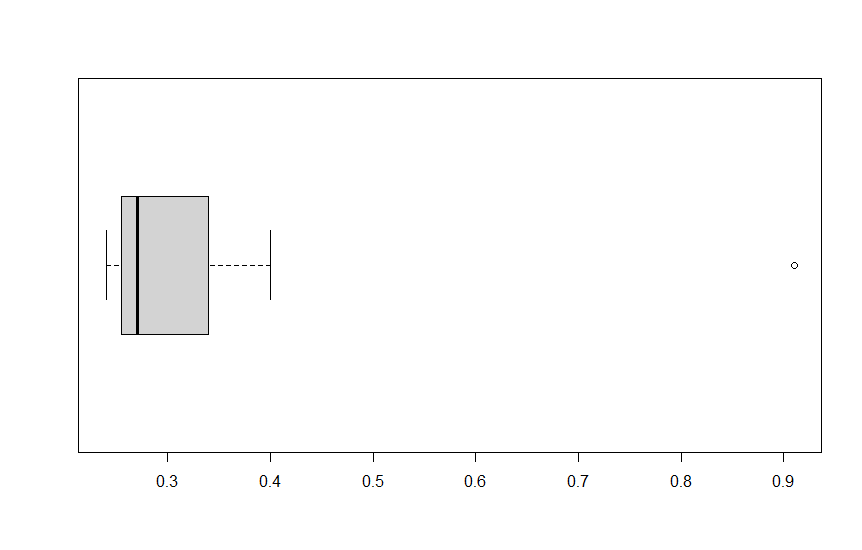
**[1] 0.1685315**

**> var(company$MeasureX)**

**[1] 0.02840286**

**boxplot(x = company$`MeasureX`, horizontal = TRUE)**

* **Basis the observation “Morgan Stanley” is the outlier with 91:36 Measure**





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. -> **12-5 = 7**
2. What can we say about the skewness of this dataset? = **The data is skewed to the right**
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected? **The median will be moved a little to right and there will be no outliers in the box plot**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* **Mode lies between 4 to 8 (highest peak of the histogram)**

1. Comment on the skewness of the dataset.

* **Data is right skewed as the values of ‘Y’ are reducing toward right**

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.

* **Histogram provides frequency information of each data grouped by and fails to note exact outlier of the dataset**
* **Box plot provides exact information of the outlier while it cannot graph data group by frequencies**

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.)

* **Let X be The call is misdirected**

**Probability P(X) = 1/200**

**= 1 – P(X bar)**

**= 1 – (199/200)\* (199/200)\* (199/200)\* (199/200)\* (199/200)**

**= 1 - (199/200)^5**

**= 0.025**

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?

* **x = 2000**

1. Is the venture likely to be successful? Explain

* **Yes, 60% chances of returns are positive**

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

* **Long term average = summation (x\*p(x)) = ((-2000\*1)+ (-1000\*1)+ (1000\*2)+ (2000\*3)+ (3000\*1) / 6) = 8000/6 = 1333**

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

* **Risk = negative returns on the amount invested where the probability is 0.1 + 0.1 = 20%**