**Name: Saloni Surendra Alshi**

**Assignment 2: Basic Statistics\_Level-2**

**SET 1**

**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% |

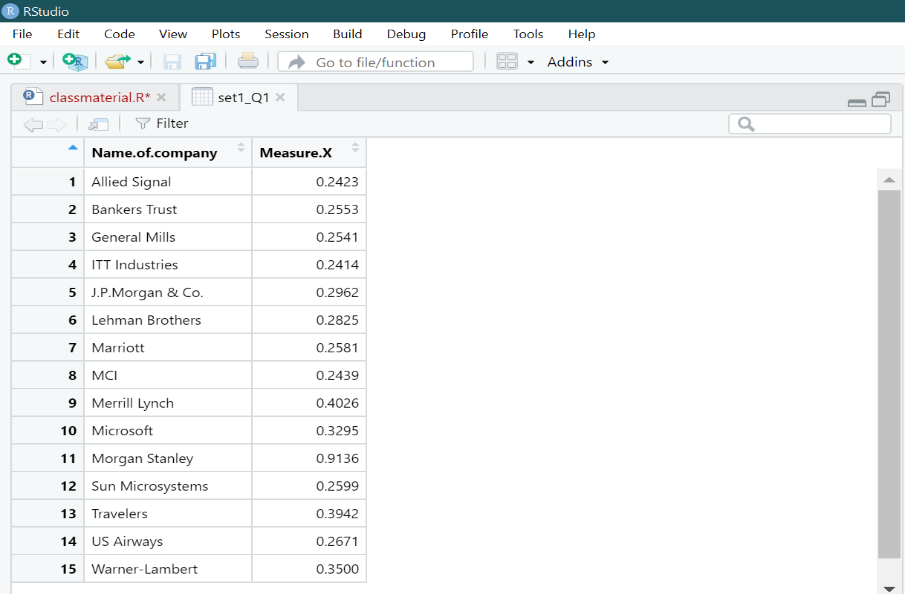
**Answer:**

**Using R-studio,**

**Imported data using excel sheet.**

> set1\_Q1 <- read.csv("~/courses/data science/assignment/assignment2/set1\_Q1.csv")

> View(set1\_Q1)



> # Finding mean, standard deviation, variance

> mean(set1\_Q1$Measure.X)

[1] 0.3327133

> var(set1\_Q1$Measure.X)

[1] 0.02871466

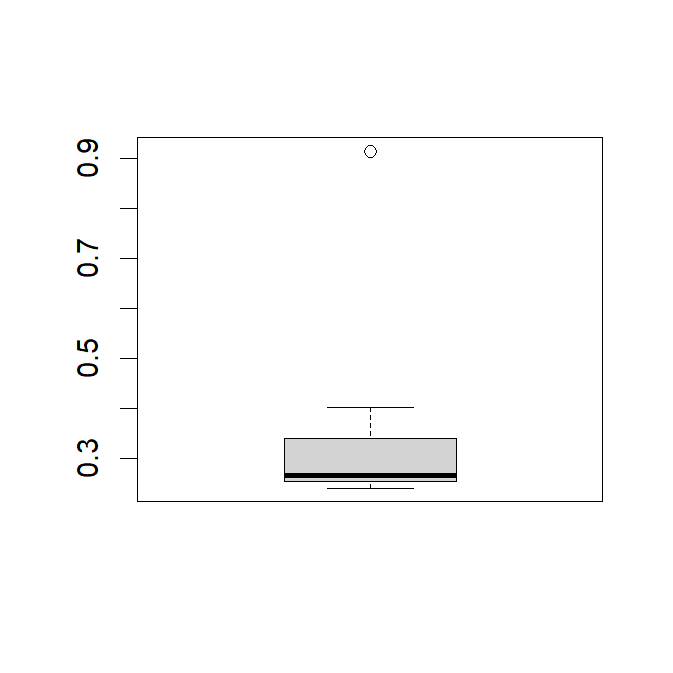
> sd(set1\_Q1$Measure.X)

[1] 0.169454

> #plotting data

> # Boxplot is used to find the outliers

> boxplot(set1\_Q1$Measure.X)



There is one outlier involved in it.

> #To find outliers involving

> boxplot(set1\_Q1$Measure.X)$out

[1] 0.9136

Morgan Stanley is an outlier of 0.9136 i.e 91.36%

**Also, done in python.**



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.

**Answer:**

Lower quartile (Q1) = 5

Upper quartile (Q3) =12

IQR=Q3-Q1=12-5=7 (approximately)

Inter-quartile range contains 50% of observations.

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

**Answer:**

The nature of skewness of the data is right skewed (i.e positively skewed) because median is closer to the bottom quartile (Q1).

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?

**Answer:**

There would be no outliers in the given dataset.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Answer:**

Mode of this dataset lies approximately in between the range of 4 to 8.

1. Comment on the skewness of the dataset.

**Answer:**

Here, most of the observations are concentrated on the left hand side and very few observations are are concentrated on the right hand side. Therefore, we called it as right skewed histogram. It is also known as positively skewed histogram.

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.

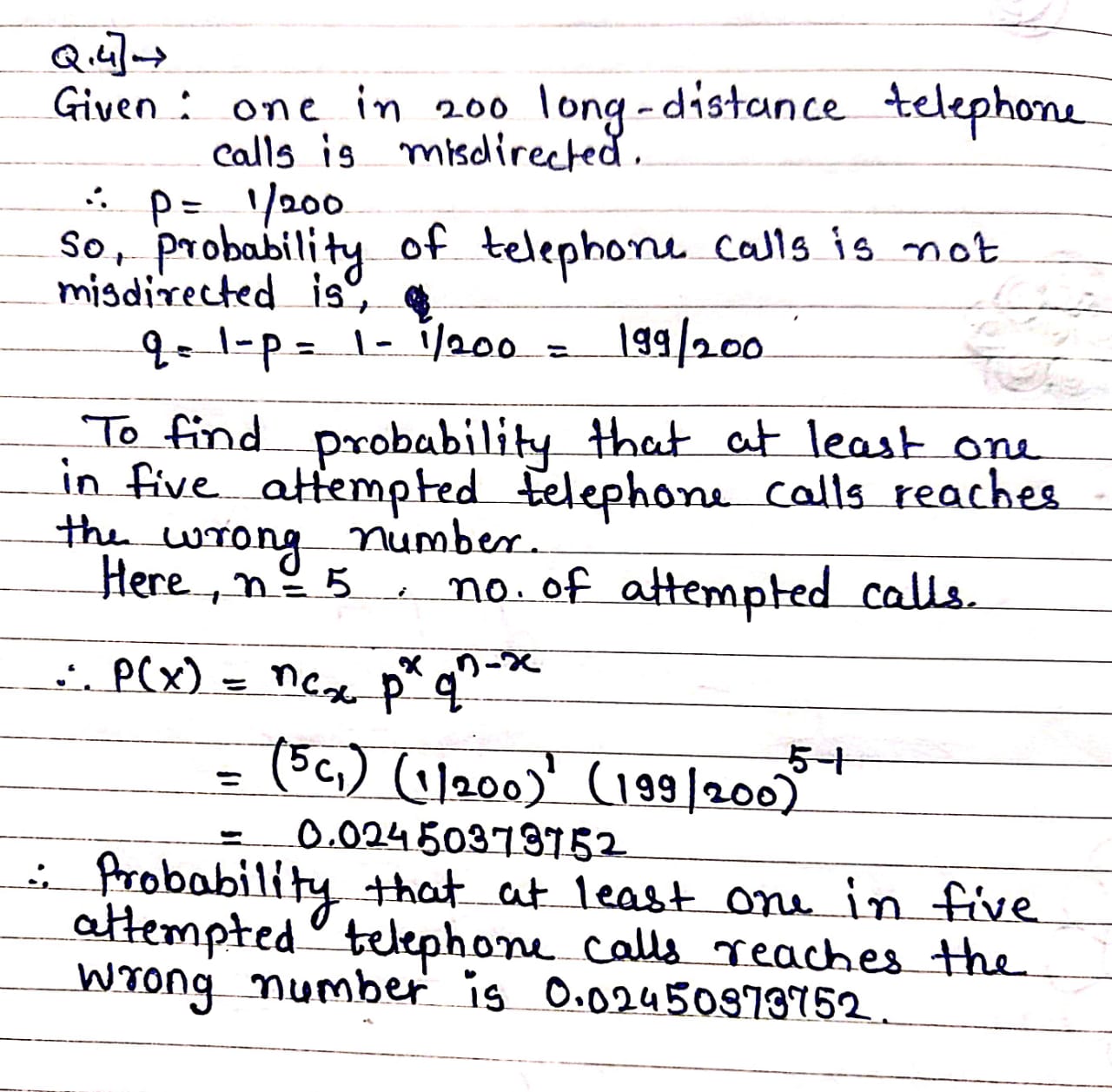
**Answer:**

From the boxplot, we can find median value while mode can easily obtained from the histogram.

Both the graphs show the positive skewness in the data.

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 thwaat 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:**

****

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?

**Answer:**

The most likely monetary outcome of the business venture is $2000 having probability is 0.3 which is maximum compared to the others.

1. Is the venture likely to be successful? Explain

**Answer:**

Yes, venture likely to be successful as it will make profit of zero or more than zero.

i.e, P(X=0)+P(X=1000)+P(X=2000)+P(X=3000)

i.e, 0.2+0.2+0.3+0.1=0.8

means, venture will make profit of 80%

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

**Answer:**

Expected value= sum(x\*P(x))

=800$

Hence, the long-term average earning of business ventures of this kind is 800$

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

**Answer:**

Here, for risk we have to find variance i.e variability in the data.

More the variance more is the risk.

Var(x)= E(x^2)-[E(x)]^2

**=**sum[(x^2)\*P(x)]- [sum[(x)\*P(x)]]^2

=2800000-640000 =2160000