**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, Variance and Standard Deviation are as

> mean(`Measure X`)

[1] 0.3327133

> #Measures of Dispersion

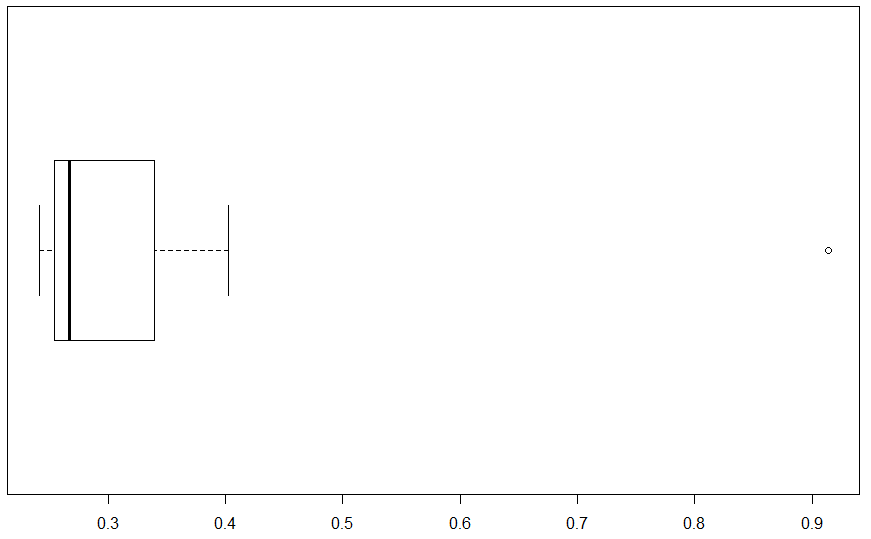
> var(`Measure X`)

[1] 0.02871466

> sd(`Measure X`)

[1] 0.169454

Plot:



Outlier:

bx = boxplot(`Measure X`,horizontal = TRUE)$out

bx #0.9136

So outlier is Morgan Stanley: 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.
2. What can we say about the Skewness of this dataset?
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?

Ans:

1. Inter-quartile range is 7 (12-5)

Inter-quartile means difference between first quartile Q1 and Third Quartile Q3.

1. Skewness is right handed Skewness i.e positive Skewness
2. Shifting data point to 2.5 from 25, will reduce the positive Skewness. Hence outlier will be removed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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:

1. Seems bimodal right skewed distribution. Mod value probably lies between 4-8 as it is having highest frequency.
2. right or positive skewed distribution
3. Both shows right or positive skewed distribution
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.)

Ans: The call is misdirected then probability of the event E is P(E) = 1/200

Therefore, Probability of having success of correct call will be:

=1-p(E) =1-(1/200) =199/200

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

=1-[(199/200)^5]

= 0.02475125

Probability that at least one in 5 attempted call reaches the wrong number is **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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Ans:

1. Most likely outcome will be p(x)=0.3 =$2000 … highest Probability
2. Assuming the production cost is zero, the net profit from this venture will be total of XP(X)

This is $800. (-2000\*0.1+-1000\*0.1+0+1000\*0.2+2000\*0.3+3000\*0.1)

So Venture will be yielding positive output.

1. Long term average is $800, so it is successful. If the probability of positive outcome I increased then it’s a good venture to put forward.
2. The standard deviation of the venture is 1870, whereas profit lies at 800 so it’s a high risky business

sd(X)

[1] 1870.829