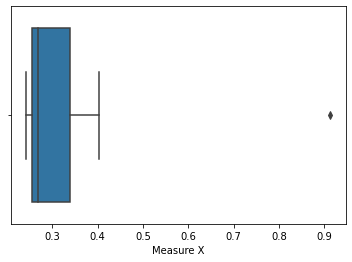
**Topics: Descriptive Statistics and Probability**

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% |

Firstly copy the data in excel file

Then read the excel file in jupyter notebook

= 0.3327133333333333 (Mean)

= 0.16945400921222029 (Variance)

0.028714661238095233 (Standard Deviation)



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?
   1. IQR is the range between upper quartile (Q3) and lower quartile (Q1)

IQR= Q3-Q1= 12-5 = 7

50% of the data lies between IQR

* 1. Skewness is Positive/Right Skewed
  2. There will be no outlier if the value of 25 was actually 2.5. So, mean and median needs to be re-calculated to see if there is any shift in data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The mode will most probably lie between 4-8 as most of the data lie there.

1. Comment on the skewness of the dataset.

Skewness is Positive/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.

There is an outlier of the value 25 and both of them have positive skewness

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.)

X = probability of 1 call misdirected out of 200

Probability of occurring of X = 1/200

P(X)= 1/200

Probability of having at least one successful call will be

1-P(X)= 1-1/200= 199/200= 0.967

As every event is independent of other event the probability will be

1- (0.967)^5

0.02475 = 2.5% chance.

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

Firstly copy paste data in excel file then read it in notebook

* + 1. $2000 as it has the highest probability of occurrence
    2. If Success == positive returns as a measure ,then there is a 60% probability that the venture would be successful (0.3+0.2+0.1=0.6=>0.6\*100=>60%).
    3. (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000 \*0.3)+(3000\*0.1)=800 the long-term average earning for these type of ventures would be around $800
    4. A good measure to evaluate the risk would be variance and standard deviation of the variable x

Var = 3500000

Sd = 1870.83

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