**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% |

1)Mue=mean,sigma=std and sigma^2 =std

A screenshot of a computer

Description automatically generated

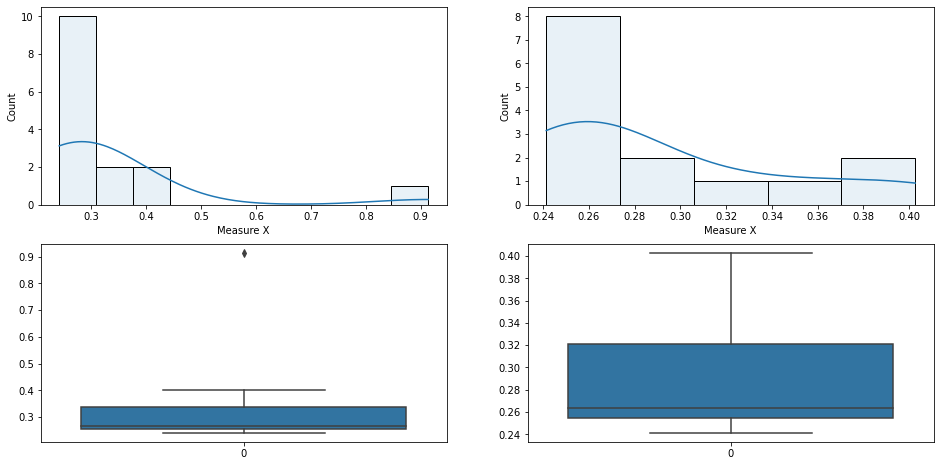
2) there are various methods to detect Outliers->

a) box plot

A screen shot of a graph

Description automatically generated

Left side shows the data with outliers, right side shows the data without outliers



B)IQR range

Anything below 25 percentile-1.5 IQR and 75 percentile +1.5IQR is considered Outlier.



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.

* Interquartile range is 75th percentile-25th percentile of the data, this value represents 50% of the data in the dataset which are close to median.IQR-> 10.3 - 5=5.3

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

* The density of data lying below median is larger and the distribution of the data above or right side of mean is greater, we can say that the data is right/positively 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?

* The outlier (the dot towards the right side) will be removed.The distribution will remain the same though.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* 5 and 6

1. Comment on the skewness of the dataset.

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.

* Box plot will tell the IQR which can be helpful in getting the mid point of data and detecting outliers.
* Histogram will be helpful to visualize the skewness and the variance of the data with its frequency distribution.

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.)

A white board with writing on it

Description automatically generated

Answer=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?

* Monetary or most likely gain in a short period of time will be the one with highest probability i.e 2000

1. Is the venture likely to be successful? Explain

* Assuming successful venture is getting any gains above 0, so the probability of that occurring will be 0.2+0.3+0.1=60%, we can be 60% confident that the venture will be successful and 80% confident that the venture will not have any loss.

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

* Calculate the expected value for this venture i.e

Expected Value = (-2,000 \* 0.1) + (-1,000 \* 0.1) + (0 \* 0.2) + (1,000 \* 0.2) + (2,000 \* 0.3) + (3,000 \* 0.1)

Expected Value = -200 + (-100) + 0 + 200 + 600 + 300

Expected Value = 800

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

* To calculate the risk involved we can calculate the standard deviation of the data

We can do this by using the expected value as the mean and applying that in the formula

Standard Deviation = √[0.1 \* (-2,000 - 800)^2 + 0.1 \* (-1,000 - 800)^2 + 0.2 \* (0 - 800)^2 + 0.2 \* (1,000 - 800)^2 + 0.3 \* (2,000 - 800)^2 + 0.1 \* (3,000 - 800)^2]

Standard Deviation = √[0.1 \* (-2,800)^2 + 0.1 \* (-1,800)^2 + 0.2 \* (-800)^2 + 0.2 \* (200)^2 + 0.3 \* (1,200)^2 + 0.1 \* (2,200)^2]

Standard Deviation = √[784,000 + 324,000 + 128,000 + 8,000 + 432,000 + 484,000]

Standard Deviation = √(2,160,000)

Standard Deviation ≈ $1,470.98

* This shows that with the mean of 800$ and 1470 standard deviation my profit and loss can be calculated, What this implies is that it is a high risk and high reward stratum because there is a equal amount of chance for huge gains or losses.