**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:

Please refer to the attached Assignment 2 Q1 Jupyter notebook file for the codes. Pie chart and Box plot has been plotted in the same notebook.

Outliers = Morgan Stanley

Mean = 33.2713

Standard deviation = 16.9454

Variance = 287.1466



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. IQR = Q3 – Q1

Q1 = 5

Q3 = 12

IQR = 12 -5 = 7

IQR tells us the range of the middle half of the data.

1. The data is positively skewed
2. If the value 25 is 2.5 then it’ll affect the mean and the median values.

Also, there won’t be any outliers in the data.



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. The mode of this dataset will lie between 4 to 7.5.
2. It is positively skewed.
3. Histograms and box plots complement each other in providing a comprehensive understanding of a dataset by offering insights into different aspects of its distribution, central tendency, and spread. It would help us find the outliers, mode, skewness, and relationships between the median and mean values.
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:

Using the complement probability approach,

Let p be the probability that a call is getting misdirected (p = 1 / 200)

And q be the probability that a call is not getting misdirected (q = 1 - p)

Now, we want to find the probability that at least one in five attempted calls is misdirected.

Which is the completement of the probability that none of the attempted calls are misdirected.

The probability that none of the five attempted calls are misdirected (P(none misdirected)) is given by q^5 (because we assume that the attempts are independent events so q\*q\*q\*q\*q = q^5).

Therefore, the probability that at least one in five attempted calls is misdirected is:

P(at least one misdirected) = 1 - P(none misdirected)

P(at least one misdirected) = 1 - q^5

P(at least one misdirected) = 1 - (1 – 1/200)^5

P(at least one misdirected) = 0.02475

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. The most likely monetary outcome of the business venture is 2000 as it has the highest probability of 0.3.
2. In a financial context, success is measured by the net positive earnings. So, we can find if the venture is likely to be successful by calculating the probability that it’ll yield positive returns.

P(Success) = P (1000) + P (2000) + P (3000)

P(Success) = 0.2 + 0.3 + 0.1 = 0.6

Hence, the probability that it has a 60% chance of giving positive return shows that it is likely to be profitable.

Also, the net earning of the venture is positive (The net earnings is $800)

1. The long-term average earning of business ventures of this kind is calculated by multiplying each return by its probability and summing the products.

= (-2000) \* (0.1) + (-1000) \* (0.1) + (0) \* (0.2) + (1000) \* (0.2) + (2000) \* (0.3) +

(3000) \* (0.1)

= - 200 – 100 + 0 + 200 + 600 + 300

= 800

So, the long-term average earnings are $800.

1. A good measure of risk is the Standard deviation.

Calculation of Standard deviation:

= Square root over (Summation of((x-mean) ^2 \* P(x)))

= Square root over (((- 2000 - 800) ^2 \* 0.1) + ((- 1000 - 800) ^2 \* 0.1) + ((0 - 800) ^2 \* 0.2) + ((1000 - 800) ^2 \* 0.2) + ((2000 - 800) ^2 \*0.3) + ((3000 - 800) ^2 \* 0.1))

= Square root over ((- 2800 ^2 \* 0.1) + (- 1800 ^2 \* 0.1) + (- 800 ^2 \* 0.2) + (200 ^2 \* 0.2) + (1200 ^2 \*0.3) + (2200 ^2 \* 0.1))

= Square root over (784000 + 324000 + 128000 + 8000 + 432000 + 484000)

= Square root over (2160000)

= 1469.69