**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: Refer to Set\_1\_Q1.ipynb**



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.

**Ans:**  (First Quantile Range) Q1 = 5 and (Third Quantile Range) Q3 = 12

IQR = 12 -5 = 7

Median = 7

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

**Ans:**  Right-Skewed median is towards the left side so, it is not normal distribution.

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?

Ans: If the data point with the value 25 is actually 2.5 then it would be no Outliers because of the outlier the data had positive skewness it will be normal distributed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:** The mode of this data set lie in approximately between 4 to 8 .

1. Comment on the skewness of the dataset.

**Ans:**  Right-skewed (positively skewed), the tail on the right side (the larger values) is longer. Mean is typically greater than the Median.

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.

**Ans:** Both graphs are right-skewed and both have outliers, We can see median visualized in box plot but in histogram mode is more visible.

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.)

**Ans:** 200 long-distance telephone calls is misdirected, So probability p = 1/200

Probability of call not misdirected: q =1-p

= 1 – 1/200

= 199/200

Probability for at least one in five attempted telephone calls reaches the wrong number.

n =5

P(At least one misdirected in five attempts)=1−(199/200)5

= 0.02475124687

So, the probability for at least one in five attempted telephone calls reaches the wrong number is approximately 0.02475124687

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?

**Ans:** Most likely monetary outcome is the value with the highest probability.

In the table $2000 and P(x) : 0.3 is the monetary outcome with the highest probability compare to other data.

1. Is the venture likely to be successful? Explain

**Ans:** Yes.

* Venture is likely to be successful, we can consider the cumulative probabilities. If the cumulative probability of positive returns (values greater than 0) is high, the venture is more likely to be successful.
* Calculate the cumulative probability for positive returns:

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

= 0.2 + 0.3 + 0.1

= 0.6

* Cumulative probability of positive returns is 0.6 or 60%. Therefore, the venture is likely to be successful.

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

**Ans:** The long-term average earning:

= Sum (x \* P(x))

= (-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

The long-term average earnings for business ventures is $800.

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

**Ans:**  Measure of risk is the standard deviation which indicates the degree of variability in the returns. The formula for the standard deviation is:

σ =√ (−2000−800)2 x 0.1 + (−1000−800)2 x 0.1 + (0−800)2 x 0.2 + (1000−800)2 x 0.2 + (2000−800)2 x 0.3 + (3000−800)2 x 0.1​

= √2152000​

= 1466.287

* A higher standard deviation indicates higher variability or risk in the returns.
* A lower standard deviation indicates lower variability or risk in the returns.