**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% |

Answer: - Solved in python.



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?

Answer: -

1. Inter quartile range is the difference between Quantile 3 and Quantile 1.

In the given boxplot

Q1 ~ 5

Q3 ~ 12

Therefore, Inter-quantile range = 12-5 = 7

1. The data has an outlier on the right side. Also the median is present on left side from the center which means median<mean. This implies the data is right skewed.
2. The data point having value 25 is currently treated as an outlier. If we change the value to 2.5 it would mean the data would have no outlier. Also skewness will decrease significantly. And data will get closer to normal distribution.



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.

Answer: -

1. As we can see in the histogram, highest frequency is for the two bars between 4 and 8. So, the mode will also lie in that range.
2. We can observe that most of the data is distributed on the left side of the graph. Also mode is on the left side of the graph. But an outlier on the right side would mean, Mean of the data would lie away from median. Hence, the data is right skewed.
3. There are some facts which confirm that dataset used for both the graphs is same. Some of them are: -
4. There is an outlier in both the graphs at 25.
5. Both the graphs show right skewness.
6. Median of both the graphs is on left side from center.
7. 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.)

Answer: - No. of misdirected calls in 200 attempts = 1

Probability that the call is misdirected = 1/200

Parameter n = 5

P (At least 1 misdirected call) = 1 – P (All calls directed correctly)

P (All calls correctly directed) = (199/200) ^5

P (At least 1 misdirected call) = 1 – (199/200) ^5

P (At least 1 misdirected call) ≈ 0.0244

The probability that at least one in five attempted telephone calls reaches the wrong number is approximately is 2.44 %

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

Answer: -

1. Most likely monetary outcome of the business venture is $2000. It has the highest probability of 0.3.
2. Positive returns has a probability of (0.2+0.3+0.1 =) 0.6 or 60% whereas negative return has a probability of (0.1+0.1=) 0.2 or 20%. Hence net returns are positive so the venture is successful.
3. Long term average earning of this venture = 3000\*0.1+2000\*0.3+1000\*0.2-1000\*0.1-2000\*0.1 = $800
4. Large values of variance/standard deviation of monetary outcome represent the high risk involved in the venture. Variance = 3.5, Standard Deviation =1.87