**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% |

**Solution:**

Mean= 33.271

Standard Deviation= 16.945

Variance= 287.146

jnn

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.

**Solution:**

Q1=5

Q3=12

IQR=Q3-Q1=7

It means that the 50% of data lies between 5 to 12.

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

**Solution:**

The nature of the skewness is Right 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?

**Solution:**

If the value is 2.5 instead of 25, the value will lie between lower whisker length and Q1. So, there will be no outliers.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Solution:**

Mode of the dataset will be between 5 to 7.

1. Comment on the skewness of the dataset.

**Solution:**

The skewness is Positive or 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.

**Solution:**

Histogram helps to find the frequency of how many data points is occurring whereas boxplot provides the 50% distribution between 5 and 12. From the histogram, we cannot identify outliers whereas the whisker lengths of boxplots help us to find them.

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.)

**Solution:**

from scipy.stats import binom

bi=binom(5,0.005)

1-bi.cdf(0)

0.024

Therefore, the probability of at least one in five attempted telephone calls reaches the wron g number is 0.024.

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?

**Solution:**

2000 with the probability of 0.3.

1. Is the venture likely to be successful? Explain

**Solution:**

At the start of the venture reported losses but later it made profit. So, it can be concluded that the venture is likely to be successful.

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

**Solution:**

The long-term average earning of business ventures of this kind is x\*P(x)=

(-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)=800

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

**Solution:**

The good measure of the risk involved in a venture of this kind =0.2. Because, the loss probability is 0.2