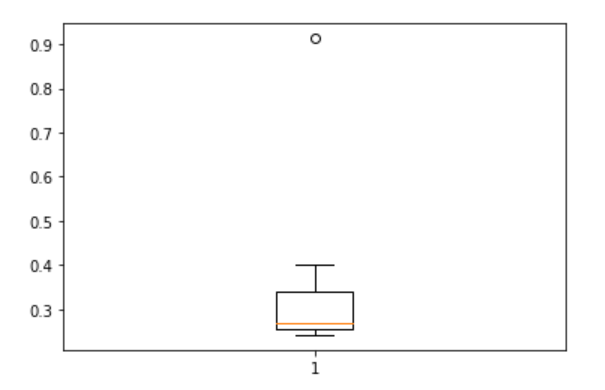
**Topics: Descriptive Statistics and Probability**

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: There is one outlier the given data set.**

**plt.boxplot(df['Measure X'])**



**Mean = 33.271**

**df\_1.mean().round(3)**

**Standard deviation = 16.945**

**df\_1.std().round(3)**

**Variance = 287.147**

**df\_1.var().round(3)**



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:** **IQR = 12-5 = 7**

Interquartile range is the amount of spread in the middle 50%, percent of a dataset.

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

**Ans:** **Skewness is towards right side i.e., positive skewness.**

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 value of 25 is actually 2.5 then there won’t be any outliers in the given dataset.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:** **Mode lies between 5 and 7**

1. Comment on the skewness of the dataset.

**Ans: Skewness is towards right side i.e., -ve Skewness**

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: We can identify the Outliers in the data set.**

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:**

**n = 5**

**P(x) = 1/200**

**P(y) = 1-P(x) = 199/200**

**Probability that at least one in five attempted telephone calls reaches the wrong number = 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?

**Ans: 2000**

1. Is the venture likely to be successful? Explain

**Ans: Yes, the venture is likely to be successful as the probability of gain is 0.2 + 0.3 + 0.1 i.e., 60%.**

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

**Ans: X1\*P(x1) + X2\*P(x2) + X3\*P(x3) + X4\*P(x4) + X5\*P(x5) + X6\*P(x6)**

**(-200) + (-100) + 0 + 200 + 600 + 300 = 800**

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

**Ans: Average Earning = 800**

**Standard Deviation = 294.392**

**Risk involved in the venture is comparatively less because the average earnings are more compared to Standard deviation.**