**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% |

measure\_x.mean()

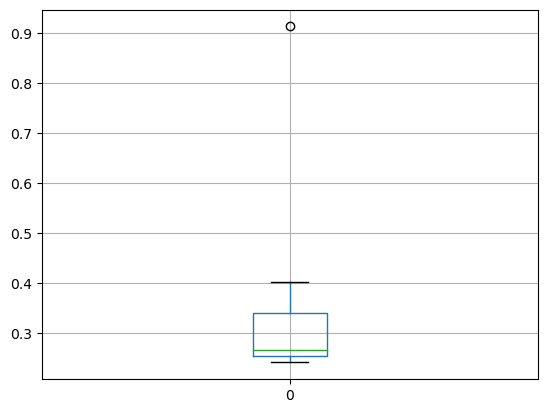
µ = 0.332713

measure\_x.std()

σ = 0.169454

measure\_x.var()

σ2 = 0.028715



q1 = measure\_x.quantile(0.25)

q3 = measure\_x.quantile(0.75)

IQR = q3 -q1

outliers = measure\_x[((measure\_x<(q1-1.5\*IQR)) | (measure\_x>(q3+1.5\*IQR)))]

0 NaN

1 NaN

2 NaN

3 NaN

4 NaN

5 NaN

6 NaN

7 NaN

8 NaN

9 NaN

10 0.9136

11 NaN

12 NaN

13 NaN

14 NaN

Thus, 0.9136 is an outlier



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 = Q3 -Q1

Q1 = 5

Q3 = 12(approx.)

IQR = 12 – 5 =7

The middle half of this data falls under IQR.

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

Ans. The data is right skewed or positively skewed. Which means the data is more concentrated in the left and has a long right tail.

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. In that case, there will be no outliers and we have different value of Q1 ,Q2 , Q3.



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 will lie between 4 to 8.5

1. Comment on the skewness of the dataset.

Ans. The data is positively skewed as the data is more concentrated towards the left.

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.

1. We can see 25 as outlier in both the graphs.
2. The data in both graphs are positively skewed.
3. The bulk of data is between 5 to 12.
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. The probability of a single call being misdirected = 1/200.

The probability of a single call not being misdirected = 1 - 1/200 = 199/200.

Since the telephone call attempts are independent, the probability that all five calls are not misdirected = (199/200)5

The probability that at least one call reaches the wrong number = 1 - (199/200)5 ≈ 0.01249

So, the probability that at least one in five attempted telephone calls reaches the wrong number 1.249%(approx.).

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. The most likely monetary outcome of the business venture is the value with the highest probability, the monetary outcome of 2000 has the highest probability of 0.3.

1. Is the venture likely to be successful? Explain

Ans. For an initial investment of $0: Expected Net Outcome = (-2000 \* 0.1) + (-1000 \* 0.1) + (0 \* 0.2) + (1000 \* 0.2) + (2000 \* 0.3) + (3000 \* 0.1) = $700

Since the expected net outcome is positive, 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 earnings of business ventures of this kind can be calculated by multiplying each monetary outcome by its corresponding probability and then summing up these products:

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

So, the long-term average earnings for business ventures of this kind would be $900.

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

Ans. A good measure of the risk involved in a venture of this kind is the standard deviation.