**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:

print('mean:',df['Measure X'].mean())

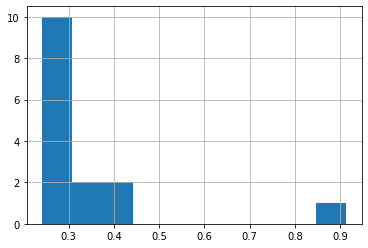
print('std:',df['Measure X'].std())

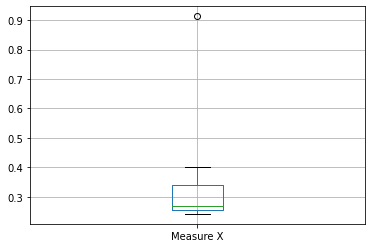
print('variance:',df['Measure X'].var())

mean: 0.3327133333333333

std: 0.16945400921222029

variance: 0.028714661238095233





There is outlier in the dataset



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: Approximately (First Quantile Range) Q1 = 5 (Third Quantile Range) Q3 = 12, (Inter-Quartile Range) IQR = Q3 – Q1 = 12 – 5 = 7

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

Solution: Positive skewness and not normally distributed.

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: Then there is no outliers and data will be distributed normally.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Solution: Mode of data lie between approx( 4 to 8).

1. Comment on the skewness of the dataset.

Solution: Right Skewness; mean>median>mode

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: In both diagram, there is outlier and right skewness and with the help histogram graph visual of skewness are more clear.

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: One in 200 long-distance telephone calls is misdirected

 probability of call misdirecting  p = 1/200

 Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ , Where n=5, p=1/200 and q=199/200

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

(5C1) (1/200)^1 (199/200)^5-1 P(1) = 0.0245037

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

(iv)What is the good measure of the risk involved in a venture of this kind? Compute this measure

