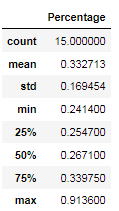
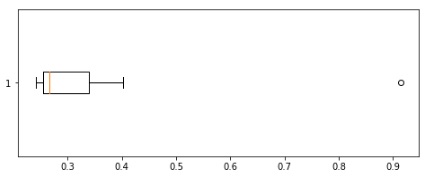
**Topics: Descriptive Statistics and Probability**

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:**





The outlier is Morgan Stanley with 91.36%

Mean: 0.3327

Standard Deviation: 0.1695

Variance: 0.0287

1. 

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?

**Ans:**

1. IQR = 12-5 = 7, this represents the range which contains 50% of the data points.
2. This dataset is Right Skewed i.e. positive in nature.
3. The data point value 2.5 will not be considered as an outlier. The boxplot will range from 0 to 19.

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.

**Ans:**

1. The mode of this dataset will lie between 4 to 8.
2. The dataset is right skewed i.e. value is positive.
3. The Median in boxplot is 7 and Mode in histogram is 4.

Histogram provides the frequency distribution so we can see how many times each data point is occurring. However Boxplot provides the quantile distribution i.e. 50% data lies between 5 and 12.

Boxplot provides whisker length to identify outliers, no such information can be concluded from the histogram. We can only guess by looking at the gap, that 25 may be an outlier.

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:**

Number of Calls(n) = 5, p = 1/200, q = 199/200

P(x) = at least one in five attempted telephone calls reaches the wrong number

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

P(x) = (nCx) (p^x) (q^n-x)

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

P(x) = 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
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

**Ans:**

1. For Max. Value of P = 0.3 for P(2000). So most likely outcome is $2000.
2. P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater than expected returns. So the venture is likely to be successful.
3. Weighted average = x\*P(x) = 800. This means the average expected earnings over a long period of time would be 800(including all losses and gains over the period of time).
4. Variance = [(-2000 - 800)^2 \* 0.1] + [(-1000 - 800)^2 \* 0.1] + [(0 - 800)^2 \* 0.2] + [(1000 - 800)^2 \* 0.2] + [(2000 - 800)^2 \* 0.3] + [(3000 - 800)^2 \* 0.1]

Variance = [(2800)^2 \* 0.1] + [(1800)^2 \* 0.1] + [(800)^2 \* 0.2] + [(200)^2 \* 0.2] + [(1200)^2 \* 0.3] + [(2200)^2 \* 0.1]

Now, we calculate the variance, and then take the square root to find the standard deviation:

Variance ≈ 720,000

Standard Deviation ≈ √(720,000) ≈ 848.5

So, the standard deviation is approximately $848.5, which is a measure of the risk involved in this business venture. A higher standard deviation indicates higher risk.