**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:** [ Please Check ‘Ans Q1.ipynb’]



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. Q3 = 12

Q1 = 5

IQR = 12-5 = 7

IQR tells how our data spread in the middle half of our data.

1. From the median value to lower Quartile, data are less spread than the data from median to upper quartile. It may be a right-skewed distribution.
2. Our data sets will be free from outliers. The only change to the box plot is that there will be no points/outliers present after upper limit. This will make the data set a normally distributed.



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. Between 4-8 there might be two modes present inside the data set.
2. It is right-skewed distribution.
3. They both can tell us the skewness of the data. In a histogram, if there is a tail/skewness to the right of the graph, then we can say it’s positive-skewed, while if the upper whisker is larger compared to the lower whisker then we can say it’s positive-skewed and vice versa.
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:**

Calls may get rejected or get connected. There is no other event we can come up with. As this, we have only two outcomes so we can apply Binomial distribution

Probability of misdirected call = 1/200

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

Question is asking about the probability that at least one in five attempted telephone calls reach wrong no.

According to formula

P(At least 1 in 5 missdirected calls) = nCx px qn-x

= 5C1 (1/200)1 (199/200)4

= **0.0245**

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 a good measure of the risk involved in a venture of this kind? Compute this measure

**Ans:**

1. Venture **2000** has the highestprobabilityof 0.3 So it is the most likely monetary outcome
2. If the probability making profits will be more than the loss, we can conclude that venture will be successful.

So probability of making profit will be 0.2+0.3+0.1 = 0.6

While probability of making loss is 0.1+0.1+0.2 = 0.4

As the probability of making profit is more than losses so we can say

‘ **Venture is likely to be successful ‘**

1. The long-term average earnings will be depending on expected values

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

1. For risk we are asked for the variance

Var = E(X2) – (E(x))2

= { (-2000)2 \*0.1 + (-1000)2 \*0.1 + 02 \*0.2 + 10002 \*0.2 + 20002 \*0.3 + 30002 \*0.1} - (800)2

= 2800000 – 640000

= **2160000**

Since variance is very high, there is high risk involved in this kind of venture.