**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% |

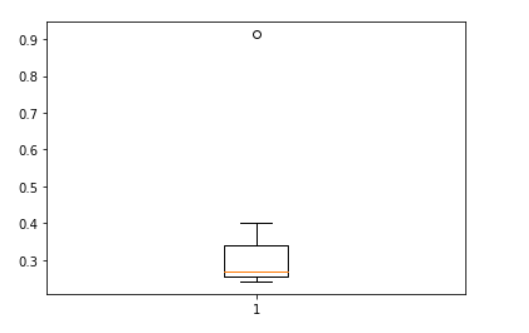
Step 1: Upload the above data into an Excel file

Step 2: Read the Excel with the (pd.read\_excel)

Mean: data['MX'].mean() = 0.3327133333333333

Standard Deviation: data['MX'].std() = 0.16945400921222029

Variance: data['MX'].var() = 0.028714661238095233



We can see from the above Boxplot that the data has one outlier and that is ‘Morgan Stanley’.



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?

Answer:

(i): IQR = Q3 – Q1 = 12 – 5 = 7

IQR or Interquartile Range is the range or area in a Boxplot that represents fifty percentage of the data.

(ii) Skewness = Right Skewed

(iii)If 25 were to be seen as 2.5, then there would be no outliers in the above boxplot.



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.

Answers:

1. Mode would lie in between the bins (4-6) and (6-8)
2. Skewness = Right Skewed as the tail is towards the right
3. Both graphs complement each other in one way, i.e, both provide information about the skewness of the dataset.
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.)

Answer:

Let’s say Probability of a misdirected call = A = 1/200

For one in five attempted calls, probability of misdirect = 1 – P(A BAR)

= [1 – (199/200)]^5

= 0.000025

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

Answers:

1. 2000
2. There is a sixty percent chance of a profit. So, yes, the venture is likely to be successful.
3. ((-2000\*1)+ (-1000\*1)+ (1000\*2)+ (2000\*3)+ (3000\*1) / 6) = 8000/6 = 1333.33
4. Good measure risk can be attributed to a certain venture having a higher probability of profit when compared to probability of loss. In this case, probability of profit equals 60% as opposed to probability of loss, i.e, 20%.