**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 : Outlier: Morgan Stanley : 91.36%, = MeasureX 33.27133, 16.945400921222028287.1466123809524

Refer Ipynb file attached in mail for plot.



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 = 12-5 = 7. This value implies that te 50% of the dataset values lie between 5 - 12, which are the lower quartile and upper quartile values.
2. What can we say about the skewness of this data-set? Ans: This data-set will be right skewed as there is an outlier on the upper half of data-set.
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: If we found the data point 25 is 2.5, the boxplot will not have an outlier and the plot will slightly shift towards left and the data will no longer be right-skewed as well and will be somewhat closer to a normal distribution.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie? Ans: Mode of dataset will lie between 4 and 8 Value of Y.
2. Comment on the skewness of the dataset. Ans: Dataset will be right skewed.
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: From both the plot we can combinigly conclude that there is an outlier on the datapoint 25, data is right skewed and majority of the points lie between 4 - 12 and except the outlier all datapoints can be mapped between 0-20, also data is somewhat of a normal distribution. Median of data is approx 7 and mode will lie between 4-8.

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: 1 in 200 calls is misdirected.

Let us define event E = call misdirected

P(E) = 1/200

P(Call went through) = P(Q) = 199/200

Number of calls in sample = n = 5

P(x) = (nCx)\*(P(E)^n)\*(P(Q)^(n-x))

P(atleast 1 in 5 reaches wrong number) = 1 - P(none goes to wrong number) = 1 - P(0) = 1 - (5C0)\*((1/200)^0)\*((199/200)^5) = 1 - ((199/200)^5) = 0.02475.

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: Most likely monetary outcome = 2000
2. Is the venture likely to be successful? Explain Ans: P(1000) + P(2000) + P(3000) = 0.2+0.3+0.1 = 0.6.

Business is successful if we consider the consider the monetary return in positive as a measure. So according to data, business will be successful as the probability of positive monetary return is 0.6 which is more than negative returns and zero return.

1. What is the long-term average earning of business ventures of this kind? Explain

Ans: Ʃ(X\*P(X)) = (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.2)+(3000\*0.1) = 800.

Value calculated is expected rate of return per day. In long term the business will have profit.

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

Ans: Measure of risk = P(x<0) = P(-1000) + P(-2000) = 0.1 + 0.1 = 0.2

Business will be risky if it goes in loss so measure of risk is P(x<0).