**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% |

Outlier = 91.36, Mean = 33.27, Variance = 287.15, Std. Dev = 16.95, Median = 26.71



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.

IQR = Q3 – Q1 = 12-5 = 7. It measures the spread of the middle 50% of values

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

It is right/positively skewed data.

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?

The minimum and the maximum whiskers will remain the same and there would be no Outliers on the given dataset.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Between 4 to 8 (as we can see from above graph that it has higher frequency)

1. Comment on the skewness of the dataset.

It is right/positively skewed data, which means the direction of outliers is in right side & most of data would be on left side of graph.

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.

Histogram generally depicts whether the given dataset follows a normal distribution or not, and also portray a picture if data is right or left skewed. On other hand, box plot helps in determining the outliers, min & max value of data set. And also, how is the distribution of dataset w.r.t quartiles.

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.)

N = 5, p = 1/200 & q = 199/200

Using binomial distribution, we can derive the probability.

Calculated the binomial distribution in python,

Probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475 ~ 2.475%.

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?

Most likely monetary outcome of the business venture is +2000

as it has maximum probability = 0.3

1. Is the venture likely to be successful? Explain

The probability of success totals to 0.6 or 60% for the business venture to earn at least $1000.

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

Expected Value = sum[x\*(P(x))] = 800

Venture is likely to be successful as Expected value is + 800

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

After calculating variability, Var (X) = E(X²) - {E(X)}²

SD = √Var ≈ $1470

As Variability is Quite high, hence Risk is high