**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% |



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.

Sol: Q1 = 5 , Q3 = 12

IQR = Q3 – Q1 = 12 – 5 = 7

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

Sol:

In this dataset, the distance between lower extreme to first quartile is less than the distance between third quantile to upper extreme. So, the data is Positively skewed.

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?

Sol:

Actually the whole dataset is affected by the outlier 25 value only. If data point with the value 25 is actually 2.5, then there are no outliers in the dataset and also there is no skewness in the data . so, the data is normally distributed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Sol: The mode of this dataset lies in the range of 4 – 8(approximately).

1. Comment on the skewness of the dataset.

Sol: Based on the above Histogram, the data is Positively skewed or right skewed.

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.

Sol: Based on Histogram, we can observe the frequency of the data in the form of bins and

also we can observe how the data is distributed.

Based on box plot, we can find outliers in the dataset.

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.)

Sol: Probability that one in 200 long-distance telephone calls is misdirected p = 1/200 = 0.005

Probability that not misdirected q = 1- p = 1- (1/200) = 1- 0.005 = 0.995

The probability that at least one in five attempted telephone calls reaches the wrong number = > 1 – none of the calls reach wrong number = 1- P(0)

No. of calls = 5

P(X) = (n c X ) \* (PX) \* (q(n – X))

P(0) = (5c0) \* (0.005)0 \* (0.995)(5 – 0 )

= > P(0) = 0.975

1 – P(0) = 1- 0.975 = 0.025

The probability that at least one in five attempted telephone calls reaches the wrong number is

0.025.

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?

Sol: The most likely monetary outcome of the business venture is 2000.

1. Is the venture likely to be successful? Explain

Sol: Probability that venture likely to be successful = 0.2 + 0.3 + 0.1 = 0.6 =60%

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

Sol: Mean = 500

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

Sol: x.var() = 3500000

x.std() = 1870.828

x.mean() = 500