**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:

Mean = 499.07/5 = 33.27

Variance = 4019.55/15 = 267.97 = 268

Standard Deviation = sqrt(268) = 16.36

Outliers:

Low value outliers: 11.025

High value outliers: 49.385

Possible outliers is: 91.36



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: IRQ = Q3-Q1

Q3 = 12, Q1 = 5

Q3-Q1 = 13-5 = 7

High value Outliers: Q3 + 1.5(8) = 25

Low value Outliers: Q1 – 1.5(8) = -7

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

Ans: positive skewness

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?

Ans: If 2.5 was the value then the box-plot would not have any outliers as the box plot starts from 0 to 20 in representation



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: Mode is most occurred value, then mode will be between 4 to 8.

1. Comment on the skewness of the dataset.

Ans: If the data of the graph is concentrated on left side, then it is call positive skewness.

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.

Ans:Median in boxplot and Mode in histogram.

Histogram provides the frequency distribution so we can see how may times each data point is occurring however boxplot provides the quantile distribution i.e. 50% data lies between 5 and 12.

Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an outlier.

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.)
2. 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: Max P = 0.3 for P(2000). So most likely outcome is 2000

1. Is the venture likely to be successful? Explain
2. What is the long-term average earning of business ventures of this kind? Explain

Ans: Weighted average = x\*P(x) = 800. This means the average expected earning over a long period of time would be 800(including all losses and gains over the period of time).

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

Ans:P(loss) = P(x = -2000) + P(x = -1000) = 0.2. So the risk associated with this venture is 20%.