**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% |

1. In python file



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?
4. i) inter-quartile range = Q3-Q1

= 12-5 = 7

In descriptive statistics, the interquartile range is a measure of statistical dispersion. It is defined as the difference between the 75th and 25th percentiles of the data.

ii) This is right (probably negatively) skewed distribution as mean lies on left side and tail is longer on right side.

iii) The value 25 upon change to 2.5, there shall no outlier and it may be normally distributed.



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.
4. i) The mode shall lie in range 5-10.

ii) This is right (probably negatively) skewed distribution as mean lies on left side and tail is longer on right side.

iii) The boxplot for this dataset shall display outliers, mean, median and IQR values. However, the Histogram depicts mean, mode, frequency. Mutual feature would be skewness.

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. In Python file
3. 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
5. i) It is likely to be successful as for $2,000 outcome holding 0.3 probability, which is the highest.

ii)The venture shall be profitable as P(x>0) + P(x>1000) + P(x>2000) + P(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 (80% chance of profit)

iii) The long-term average earning be around $800.

iv) The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution.