**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.

Ans: IQR = 12.4 – 5 = 7.4 (approx.), IQR implies the difference between the 75 percentile and 25 percentile, i.e it implies the spread of the middle most data of a distribution.

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

Ans: Since the Right whisker is longer than the left the data distribution 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?

Ans: The Median will remain same but the IQR would be affected. Also, there will be no outlier then.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: approx. between 4 to 8

1. Comment on the skewness of the dataset.

Ans: The dataset is positively skewed since the right tail is longer than left tail.

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: Box plot provides information about the data range and outliers while the histogram provides information about the frequency of datapoints. They both provides info about the outlier but box plot uses value range but histogram uses frequency.

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: probability of misdirect p = 1/200

Probability of correct q = 1 – 1/200 = 199/200

Num of attempts N = 5

the probability that at least one in five attempted P(1) = 5C1 p^1 q^4 (using P(x)= rCn p^x q^n-x)

= 5 \* (1/200) \* (199/200)^4 = 0.024503737515625

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 is 2000$ as it has maximum probability of 0.3.

1. Is the venture likely to be successful? Explain

Ans: probability of successful venture = 0.2+0.3+0.1 = 0.6

Probability of failed venture = 0.1+0.1+0.2 = 0.5

Since probability of successful venture is greater than Probability of failed venture the venture is likely to be successful.

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

Ans: Expected venture = -2000\*0.1 + (-1000)\*0.1 +0\*0.2+1000\*0.2+2000\*0.3+3000\*0.1 = 800

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

Ans: Expected venture E(x)= 800 , E(x^2) = (-2000)^2\*0.1 + (-1000) ^2\*0.1 +(0) ^2\*0.2+(1000) ^2\*0.2+(2000) ^2\*0.3+(3000) ^2\*0.1 = 2800000

Var = E(x^2) – E(x)^2 = 2800000 – (800)^2 = 2160000 (quite high)

Std = √2160000 = 1469.69 ( Risky since high variance)