**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) The value for Morgan Stanley (91.36%) is an outlier.

Mean = 33.271333

Standard Deviation = 16.945

Variance = 287.146



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 is 12 - 5 = 7. This is the range of the middle half of the dataset.

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

Ans) Data is slightly right 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 mean would be smaller, and the data would be slightly more right skewed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans) 4, 5, 6, 7, or a combination of them

1. Comment on the skewness of the dataset.

Ans) 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.

Ans) Both plots provide the same inferences. While the boxplot shows outliers specifically, with the histogram, we have to consider the distribution of the data, and infer the outliers by the gap between it and the remaining data and by it falling outside the estimated distribution assumed by visual inspection. Histogram is a better representation of density at each value of the random variable, providing a better visual representation of the distribution. Boxplot provides quartiles and from them, we can infer the shape of the distribution but not as precisely as with a histogram.

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) 2.475% see notebook

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) $2000 which has the highest probability i.e. 0.3

1. Is the venture likely to be successful? Explain

Ans) The venture is likely to be successful as the probability for profit (0.2 + 0.3 + 0.1 = 0.6) is greater than probability for failure (0.1 + 0.1 = 0.2)

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

Ans) Average earning would be the expected value which is sum of product of each x with its probability

0.1\*-2000 + 0.1\*-1000 + 0.2\*0 + 0.2\*1000 + 0.3\*2000 + 0.1\*3000 = $800

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

Ans) Risk can be described by standard deviation which is the variation around the mean (expected value) which is $800.

np.sqrt((-2000-800)\*\*2\*0.1 + (-1000-800)\*\*2\*0.1 + (0-800)\*\*2\*0.2 + (1000-800)\*\*2\*0.2 + (2000-800)\*\*2\*0.3 + (3000-800)\*\*2\*0.1) = $1469.69