**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% |

Mean : 33.27%

Standard deviation: 0.1637081

Variance: 0.0268004

Outlier: 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) Q3~12, Q1=5.

IQR= Q3-Q1=12-5= 7. This value implies that the middle 50% of the values are spread in this range.

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

(ans) As the median is closer to the bottom of the box, and as the lower whisker is shorter than the upper whisker, we can say that the 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) If the outlier value 25 is actually 2.5, we can say that there is no bad data in this distribution. The new box-plot would be affected in the lower whisker as the corrected point lies below Q1.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

(ans) the mode of the above mentioned plot lies at Y=5 because that value has the highest frequency.

1. Comment on the skewness of the dataset.

(ans) as most of the data lies to the left of the plot, the distribution is said to be 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) If the above histogram and boxplot is plotted for the same data set, they provide information about distribution, skewness of the dataset. The Histogram has information for mode of the data while boxplot has information about the median of the data.

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) P(atleast 1 in 5 attempted telephone calls reach the wrong number)= 1.222566186617772e-8

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, as it has the highest probability

1. Is the venture likely to be successful? Explain

(ans) here overall value of probability = 1

Value of probability of the positive side =0.6

Value of probability of the negative side=0.2

Thus the business venture is likely to be successful

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

(ans) 500

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

(ans) the returns -2000, -1000, 3000 have got the same probabilities. This indicates the good amount of risk of gaining the money and losing the money at the same time. While on 2000 it is stable, and at 1000 and 0, there can only be gain and no loss.