**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% |

**Outlier: 91.36**

**μ=33.271**

**16.94**

**σ^2=287.14**



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.

**IQR = 12-5 = 7**

**IQR implies 50% data lies within this range. It is the range of values that resides in the middle of the scores.**

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

**The data is 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?

**The Outlier won’t be the outlier anymore then. The boxplot wouldn’t show any outliers then.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**The mode lies at the left side of the plot. Specifically, the mode value is between 4 and 9.**

1. Comment on the skewness of the dataset.

**The above data is right-skewed. After 10 on x-axis, the data starts skewed to the right side. It is having small frequencies at the end of the graph.**

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.

**The Box plot will show the outliers involved in the given dataset while the histogram shows the frequency distribution of the dataset and also histogram helps to identify whether the data is normally distributed or not. Boxplot will show the median of the dataset with IQR (50% datapoints) while we can easily get the mode value and skewness of the dataset in 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.)

* **probability that at least one in five attempted telephone calls reaches the wrong number = 0.0247**

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?
   * When x=2000
2. Is the venture likely to be successful? Explain

Venture is successful if X is + ve

Hence if X is 1000 , 2000 or 3000

Probability is  0.2 + 0.3 + 0.1 = 0.6

as 0.6 > 0.5 Hence **venture likely to be successful**

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

Expected value =  ∑E(X)P(X)  = 800

long-term average earning of business ventures  = 800 $

venture is  likely to be successful as Expected value is **+ ve   = 800 $**

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

Measure of risk = P(x<0) = P(-1000) + P(-2000) = 0.1 + 0.1 = 0.2

Business will be risky if it goes in loss so measure of risk is P(x<0).