**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: > library(readxl)**

**> company <- read\_excel("company.csv")**

**> View(company)**

**> mean(company$`Measure X`)**

**[1] 0.3327133**

**> sd(company$`Measure X`)**

**[1] 0.169454**

**> var(company$`Measure X`)**

**[1] 0.02871466**



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= Q3 – Q1 = 12 – 5 = 7**

**IQR represents middle 50% data.**

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

**ANS: Positive skewness**

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 new box plot will have no outliers in it.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**ANS: Mode of this dataset lie on 4-6 and 6-8.**

1. Comment on the skewness of the dataset.

**ANS: Positive skewness.**

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 give idea about skewness of the data, But**

1. **Box plot provides outlier values, which fails to provide by histogram.**
2. **Similarly histogram provides the frequency of datapoints, which fails to provide by box plot**.
3. 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: E: The call is misdirected**

**then probability of the event E is**

**P(E)= 1/200**

**Therefore,**

**Probability that at least one in 5 attempted call reaches the wrong number**

**= 1 - Probability that no attempted call reaches the wrong number**

**= 1 – P(E bar)**

**= 1 – (199/200)\* (199/200)\* (199/200)\* (199/200)\* (199/200)**

**= 1 - (199/200)^5**

**= 0.025**

**Probability that at least one in 5 attempted call reaches the wrong number = 0.025**

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**

1. Is the venture likely to be successful? Explain

**ANS: Yes, there are 60% chances of getting a positive return and 20% chances of**

**negative returns or debts.**

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

**ANS: Long term returns = ((-2000\*1)+ (-1000\*1)+ (1000\*2)+ (2000\*3)+ (3000\*1) /**

**6)= 8000/6 =1333**

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

**ANS: Good measure is, Positive returns (profits) probability tends to be more than negative returns (loss). i.e. 60% probability of profits and 20% probability of loss.**