**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 =** 0.3327133

**Standard Deviation =** 0.169454

**Variance =** 0.02871466

Plot the data.

> boxplot(x = company$`Measure X`, horizontal = TRUE)

From the Observations,

Morgan Stanley is an outlier of 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.

Answer = Approx. IQR = 13-5 = 8

The **interquartile range** (**IQR**) **is** the difference between the upper (Q3) and lower (Q1) quartiles, and describes the middle 50% of values when ordered from lowest to highest.

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

Answer = The distribution of the data 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?

Answer = Median will be changed. There will be no outliers present. The distribution will be approximately normal as the mean, median will change if the point 25 is actually 2.5.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

= mode will be in between 4 – 7. Approx. 6

1. Comment on the skewness of the dataset.

= The data is positively or right skewed. Mean will be greater than median and mode.

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.

= Both the graphs are of same dataset as both plots resembles each other.

: 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?

= The most likely outcome of this business venture is a return of $2000 as it has the highest probability of occurrence.

1. Is the venture likely to be successful? Explain

= If the venture can maintain for long term business then eventually it will be successful since the probability of non-negative return is higher than 0.50 and the expected value for return is a positive number ($800).

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

=From the above question requirement we have to consider similar business ventures of this type whose distribution of the returns is similar to this venture. In that case we say that the expected value of returns to this particular venture is the required average.

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

Therefore the long-term average earning for these type of ventures would be around $800.

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

= Good measure is Positive returns (profits) probability tends to be more than loss

60% probability of profits

20% probability of loss.