**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:

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

As we can see from the boxplot, an Outlier is confirmed between 0.9-1.0.



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:** The IQR for the above boxplot is 12-5 = 7.The IQR Indicates how spread out the middle 50% of our data is.

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

**Ans:** 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?

**Ans:** As we can see 25 is an outlier. Considering 25 is to be replaced by 2.5, it would be considered as normal data point instead of an outlier. The mean of the dataset will be highly reduced.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:** The mean of dataset will occur where the frequency is highest. In this case it will lie between 4 to 7.5.

1. Comment on the skewness of the dataset.

**Ans:** The data is positively 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:**

* Histogram provides us with frequency distribution of a data point. It shows which data point is occurring how many times in the data set.
* Boxplot is useful in identifying outliers and it indicates how the majority of our data is spread out. In this case 25 can be easily identified.

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:**

One in 200 long-distance telephone calls is misdirected.

Therefore, probability of call misdirecting p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = (nCx)(p^x)(q^(n-x))………………………..………*Formula*

n = 5

p = 1/200

q = 199/200

So, at least one in five attempted telephone calls reaches the wrong number = 1 - none of the call reaches the wrong number

= 1 - P(0)

= 1 - (5C0)((1/200)^0)((199/200)^(5-0))

= 1 - (199/200)^5

= 0.02475

Therefore, probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475 or 2.4%.

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 with the highest probability i.e., 0.3 or 30%.

1. Is the venture likely to be successful? Explain

**Ans:** The probability of venture likely to be successful = P(1000)+ P(2000) + P(3000) = 0.6 or **60%** .

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

**Ans:**

|  |  |  |
| --- | --- | --- |
| x | P(x) | x\*P(x) |
| -2,000 | 0.1 | -200 |
| -1,000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
|  |  | 800 |

The long-term average earning of business = x\*P(x) = **800**.

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

**Ans**: P(-2000) + P(-1000) = 0.2 or 20%.

Therefore, risk associated is 20%.