**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% |

**Solution:** For Code/Plot Refer - Jupyter Notebook PDF Attached

Through line and boxplot draw in jupyter it can be seen there is **one outlier in this dataset.**

μ = **0.332**

σ **= 0.1685**

σ2 = **0.0284**

2.



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.

**Solution:**

Inter Quartile Range (IQR) = Upper Quartile – Lower Quartile

**IOR = 12 - 5 = 7**

IOR value implies **the distance / difference between Q3 and Q1**, also the IOR range contains the 50% of datapoint.

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

**Solution:**

As it can be seen there is one outlier present on the upper extreme of the boxplot, so this outlier comes in the right-hand side tail, hence it **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?

**Solution:**

If it was found that the data point with the value 25 is actually 2.5, then boxplot will not show any outlier **(No Outlier)** as 25 is outlier datapoint. Also, there would be no skewness in the dataset.

3.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Solution:**

The mode of this dataset lies from **4 to 8**

1. Comment on the skewness of the dataset.

**Solution:**

As it can be seen there is one datapoint i.e., 25 presents in the right-hand side tail, hence it **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.

**Solution:**

Both the graphs have **one outlier present in their dataset**, and **both the graphs seems to have Positive Skewness**. It can be said that the **dataset of both the graphs are same.**

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.)

**Solution:**

Number of Calls = 5, n = 5,

p = 1/200

q = 199/200 - at least one in five attempted telephone calls reaches the wrong number

= 1 - (199/200) ⁵

= **0.02475**

**probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475**

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?

**Solution:**

The most likely monetary outcome of the business venture is which has the maximum probability i.e.,

**$2000 because it has the maximum probability of P(x) = 0.3**

1. Is the venture likely to be successful? Explain

**Solution:**

Calculating the Probability of failure and success,

When business earning is negative,

P(f) = 0.1 + 0.1 + 0.2 = **0.4** – taking probability of negative earning

When business earning is Positive,

P(s) = 0.2 + 0.3 + 0.1= **0.6** – taking probability of positive earning

As the **P(s)>P(f), the venture is likely to be successful**

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

**Solution:**

Expected Value E(x) = Σ xi \* P(xi)

Expected Earning E(x) = Σ Earning \* Probability

E(x) = Σ (-2000\*0.1) + (-1000\*0.1) + (0\*0.2) + (1000\*0.2) + (2000\*0.3) + (3000\*0.1)

E(x) = **800**

long-term average earning of business ventures = **$800**

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

**Solution:**

For Code Refer - Jupyter Notebook PDF Attached,

Measuring the variability of the given Probability to determine the risk involved in a venture of this kind,

The Variability of the Probability is: 0.006 So,

**The Good Measure of risk involved is: 0.006**