**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% |
| JPMorgan & 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 – (it is in Jupiter notebook)

**2.a)**



**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 interquartile range (IQR) of the dataset in the image is approximately 15. This means that the middle 50% of the data points fall within a range of 15 units. This implies that the data is relatively evenly distributed, with a few outliers at both ends.

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

**Ans: -**The dataset in the image is positively skewed. This means that the tail of the distribution is longer on the right side than on the left side. This can be seen from the fact that the median (the middle line in the box) is closer to the left whisker than to the right whisker.

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: -** If the data point with the value 25 is actually 2.5, the new box plot would be affected in the following ways:

* The lower whisker would move to the left, to include the new data point.
* The lower quartile would move down, to the median of the lower half of the data.
* The median would move down slightly, to the median of the entire data set.
* The upper quartile would be unaffected.
* The upper whisker would be unaffected.

**b)**



**Answer the following three questions based on the histogram above.**

1. **Where would the mode of this dataset lie?**

**Ans: -**The mode of the dataset in the image would lie between 20 and 25. This can be seen from the fact that the highest bar in the histogram is between 20 and 25. This means that there are more data points between 20 and 25 than any other range of values.

.

1. **Comment on the skewness of the dataset**.

**Ans: -**The dataset in the image is positively skewed. This means that the tail of the distribution is longer on the right side than on the left side. This can be seen from the fact that the median (the middle line in the box) is closer to the left whisker than to the right whisker.

Another way to see the skewness of the dataset is to look at the histogram. The histogram shows that there are more data points at the lower end of the scale than at the higher end. This means that the distribution is skewed to the right.

.

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: -** The histogram details the distribution, revealing shape and patterns, while the box plot offers a concise summary of central tendency, spread, and outliers.

Together, they provide a comprehensive overview of the dataset's characteristics.

In the given image, the positively skewed histogram aligns with the box plot, showing a median closer to the lower quartile and a notable range with outliers.

**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 – (its in Jupiter notebook)

**4)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-:** The most likely monetary outcome of the business venture is the value with the highest probability, which is the mode of the distribution. In this case, the mode is $2,000, as it has the highest probability of 0.3.

1. **Is the venture likely to be successful? Explain**

**Ans:** -Whether the venture is likely to be successful depends on your criteria for success. If success is defined as making a profit, then the venture is likely to be successful because the probability of a positive return (i.e., earning more than $0) is the sum of the probabilities for the outcomes of $0, $1,000, and $2,000, which is 0.2 + 0.2 + 0.3 = 0.7. This means there is a 70% chance of a positive return.

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

**Ans-:** The long-term average earnings, also known as the expected value or mean, can be calculated by multiplying each outcome by its probability and summing them up. In this case:

Expected Value = (-2,000 \* 0.1) + (-1,000 \* 0.1) + (0 \* 0.2) + (1,000 \* 0.2) + (2,000 \* 0.3) + (3,000 \* 0.1)

Expected Value = (-200) + (-100) + (0) + (200) + (600) + (300) = 400

So, the long-term average earnings of business ventures of this kind are $400.

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

**Ans-:** To measure the risk involved in a venture of this kind, you can calculate the variance or standard deviation of the returns. A higher variance or standard deviation indicates higher risk.

Variance = [(x1 - μ) ^2 \* P(x1)] + [(x2 - μ) ^2 \* P(x2)] + ... + [(xn - μ) ^2 \* P(xn)]

Where:

Calculate the variance by plugging in these values and summing them:

Variance = [(-2,000 - 400) ^2 \* 0.1] + [(-1,000 - 400) ^2 \* 0.1] + [(0 - 400) ^2 \* 0.2] + [(1,000 - 400) ^2 \* 0.2] + [(2,000 - 400) ^2 \* 0.3] + [(3,000 - 400) ^2 \* 0.1]

Variance ≈ 1,720,000

To get the standard deviation, take the square root of the variance:

Standard Deviation ≈ √1,720,000 ≈ 1,311.49