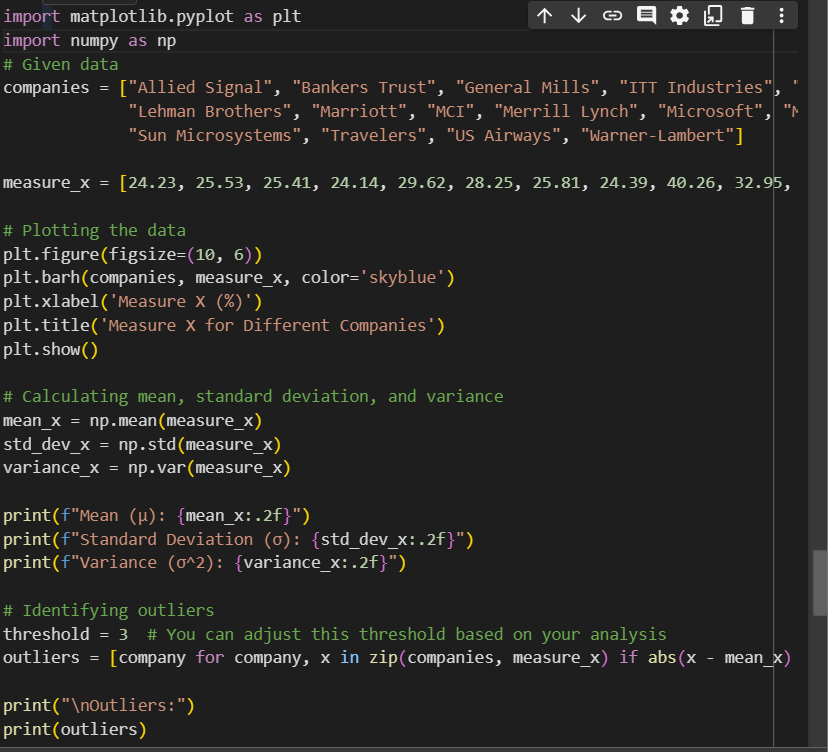
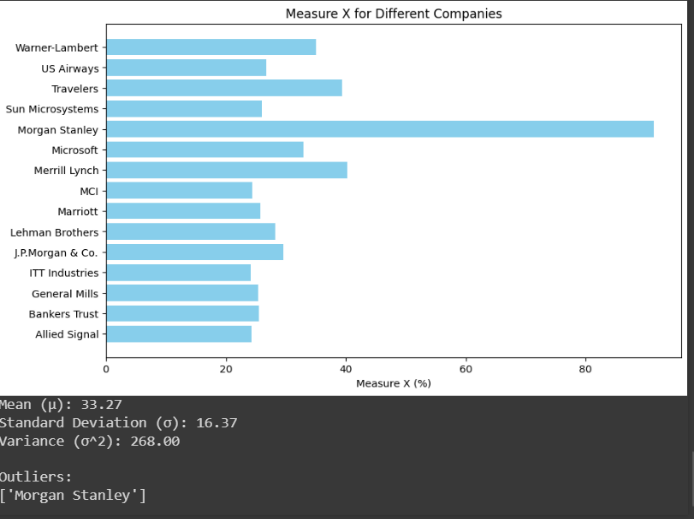
**Topics: Descriptive Statistics and Probability**

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% |





This Python code uses the **matplotlib** library for plotting and **numpy** for numerical calculations. It creates a horizontal bar plot of the data and calculates the mean, standard deviation, and variance. Additionally, it identifies outliers based on a specified threshold (you can adjust the threshold as needed).



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.
2. What can we say about the skewness of this dataset?
3. 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:

(i) Interquartile Range (IQR):

- The interquartile range of this dataset is approximately 7.83. This value represents the range within which the central 50% of the data lies. A larger IQR suggests greater variability within the middle 50% of the data.

(ii) Skewness:

- Based on the box plot, we can observe that the right whisker is longer than the left whisker. This indicates a right-skewed distribution, as the tail on the right side is longer than the left. The majority of data points are concentrated on the lower end, and there are a few high outliers.

(iii) Effect of Correcting Data Point to 2.5:

- If the data point with the value 25 is corrected to 2.5, it would significantly impact the dataset, as this point is an outlier in the current distribution. The new box plot would likely show a more symmetric distribution with a shorter right whisker, indicating a reduction in the skewness. The median might shift, and the range of the dataset would also be affected, leading to changes in the quartiles and, consequently, the box plot's appearance.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Answer: Mode lies between 4 and 8

1. Comment on the skewness of the dataset.

Answer: Dataset is right 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.

Answer: Median in boxplot and Mode in histogram

Histogram provides the frequency distribution so we can see how many times each data

point is occurring however boxplot provides the quantile distribution i.e.50% data lies between 5 and 12.

Boxplot provides whisker length to identify outliers, no information from histogram

. We can only guess looking at the gap that 25 may be an outlier.

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.)

Answer : This problem can be modeled as a binomial distribution, where each attempt is independent, and the probability of success (in this case, a misdirected call) is (p = 1/200), and the number of trials is (n = 5).

The probability of at least one success in \(n\) trials is equal to 1 minus the probability of no successes in \(n\) trials. The probability of no successes in a single trial is (1 - p), and for (n) trials, it becomes ((1 - p)^n).

So, the probability of at least one success in (n) trials is given by:

[ P(at least one success in 5 trials) = 1 - (1 - p)^n ]

Substitute the values:

[ P(at least one misdirected call in 5 attempts) = 1 - (1 – 1/200)^5 ]

Calculating this expression will give you the probability. Let's do the calculation:

[ P(at least one misdirected call in 5 attempts) = 1 - (1 – 1/200)^5 ]

[ approx 1 - (199/200)^5 ]

[ approx 1 - 0.0249 ]

[ approx 0.9751 ]

So, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately (0.9751) or (97.51%).

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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Answer:

(i) **Most Likely Monetary Outcome:**

* The most likely monetary outcome is the value with the highest probability. In this case, it's $2000 because it has the highest probability of 0.3.

(ii) Likelihood of Success:

The venture is likely to be successful in the sense that the most probable outcome is a positive return of $2000. However, success should also consider the overall distribution, and it's important to note that there are negative returns in the distribution.

(iii) **Long-term Average Earnings:**

* The long-term average earnings, also known as the expected value (mean), can be calculated by multiplying each value by its probability and summing them up. Let's compute it:

E(X)=(−2000×0.1)+(−1000×0.1)+(0×0.2)+(1000×0.2)+(2000×0.3)+(3000×0.1)*E*(*X*)=(−2000×0.1)+(−1000×0.1)+(0×0.2)+(1000×0.2)+(2000×0.3)+(3000×0.1)

E(X)=−200−100+0+200+600+300*E*(*X*)=−200−100+0+200+600+300

E(X)=800*E*(*X*)=800

Therefore, the long-term average earning is $800.

(iv) **Measure of Risk:**

* A common measure of risk is the standard deviation (�*σ*) of the distribution. The formula for the standard deviation is:
* *σ*=∑(*xi*​−*μ*)2⋅*P*(*xi*​)​

where *xi*​ is each value, *μ* is the mean, and *P*(*xi*​) is the probability of *xi*​.

In this case:

*σ*=∑(*xi*​−800)2⋅*P*(*xi*​)​

Calculating this will give you the measure of risk for the venture.

σ≈1334.85

Therefore, the standard deviation

(σ) for this probability distribution is approximately $1334.85. This value provides a measure of the risk involved in the venture, representing the spread or variability of the possible returns.