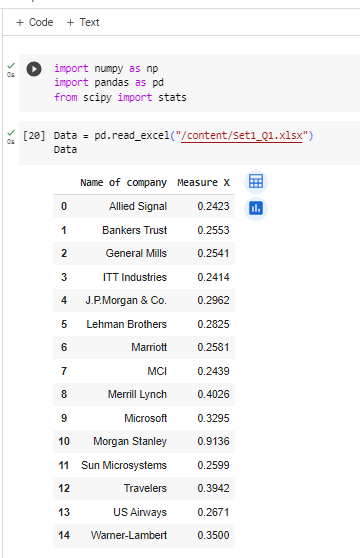
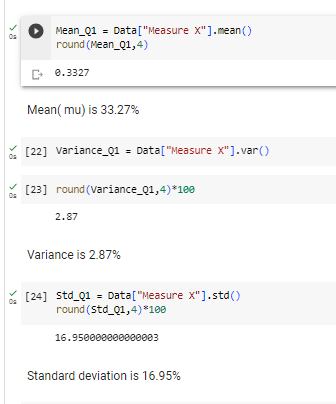
**Topics: Descriptive Statistics and Probability**

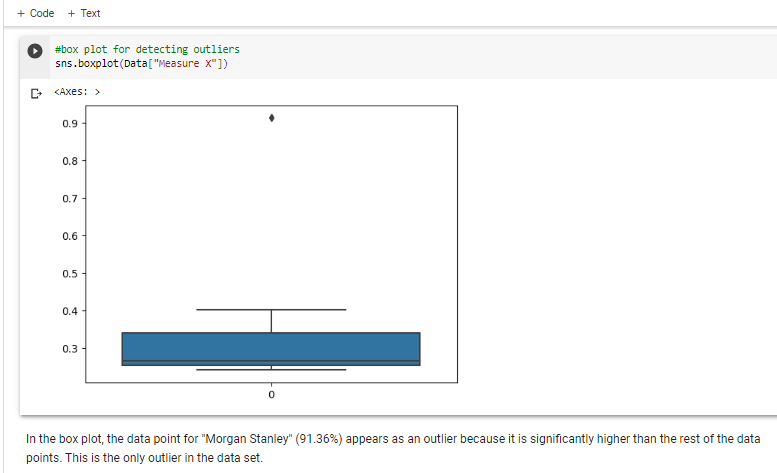
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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** :









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?

**Ans** : Minimum value = 0 , Q1 = 5 , Q2 (med) = 7 , Q3 = 13 , Maximum value = 19 ( approx.)

1. Inter-quartile range of this dataset is Q3-Q1 =13 -5 =8 (approx.).

The inter-quartile range (Q3 - Q1 = 8) implies that the middle 50% of the data points in the dataset have a spread or variability of 8 percentage points around the median.

1. The median to close to Q1 (quartile 1). Therefore, the data is more on the right end. Therefore, positively skewed.
2. The point 2.5 lies between 0 to 5. So, it lies between maximum and minimum value and no outliers are present.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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**:

(i) The mode of a dataset is the value or values that appear most frequently. Here the tallest bar is in the interval of 4-6 and 6-8.

So, mode is 5 and 7.

(ii) The histogram shows more data on the right end. So, it is positively skewed.

(iii) The histogram provides a detailed visual representation of the distribution of the data. It shows how data is distributed across different intervals. It shows the shape of the distribution, whether it's symmetric, skewed (left or right), bimodal, or has other characteristics.

The box plot gives five number summary of data set, specifically the median (Q2), quartiles (Q1 and Q3), and the outliers.

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** :To find the probability that at least one in five attempted telephone calls reaches the wrong number, the concept of complementary probabilitycan be used . We assume that each call attempt is independent.

The probability of a single call being misdirected is 1/200.

The probability of a single call not being misdirected is 1 - (1/200) = 199/200.

Since each call attempt is independent, the probability that none of the five calls are misdirected is (199/200)^5.

P(None of the five calls are misdirected) = (199/200)^5 ≈ 0.9752 (rounded to four decimal places)

Now, to find the probability that at least one in five attempted telephone calls reaches the wrong number, subtract the probability of none being misdirected from 1

P(At least one call is misdirected) = 1 - P(None of the five calls are misdirected)

P(At least one call is misdirected) = 1 - 0.9752 ≈ 0.0248 (rounded to four decimal places)

So, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 0.0248, or 2.48%.

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

**Ans**: (i) The most likely monetary outcome of the business venture can be determined by finding the value of x with the highest probability (P(x)).

In this case, the highest probability is associated with an outcome of $2,000, which has a probability of 0.3. So, the most likely monetary outcome is $2,000.

(ii) To determine if the venture is likely to be successful, we can compute the expected value of the probability distribution

Expected Value

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

E(X) = -200 -100 +0 +200+600+300

E(X) = 800

The expected value is $800, which means that, on average, the business venture is expected to earn $800.Whether this is considered successful or not depends on the context and the initial investment and costs associated with the venture.

(iii) The long-term average earnings of business ventures of this kind are represented by the expected value calculated in part (ii), which is $800. This expected value represents the average outcome over a large number of repetitions of the venture.

(iv) A good measure of the risk involved in a venture of this kind is the standard deviation of the probability distribution. The standard deviation measures the dispersion or spread of the outcomes. A higher standard deviation indicates greater variability and, therefore, greater risk.