**Topics: Descriptive Statistics and Probability**

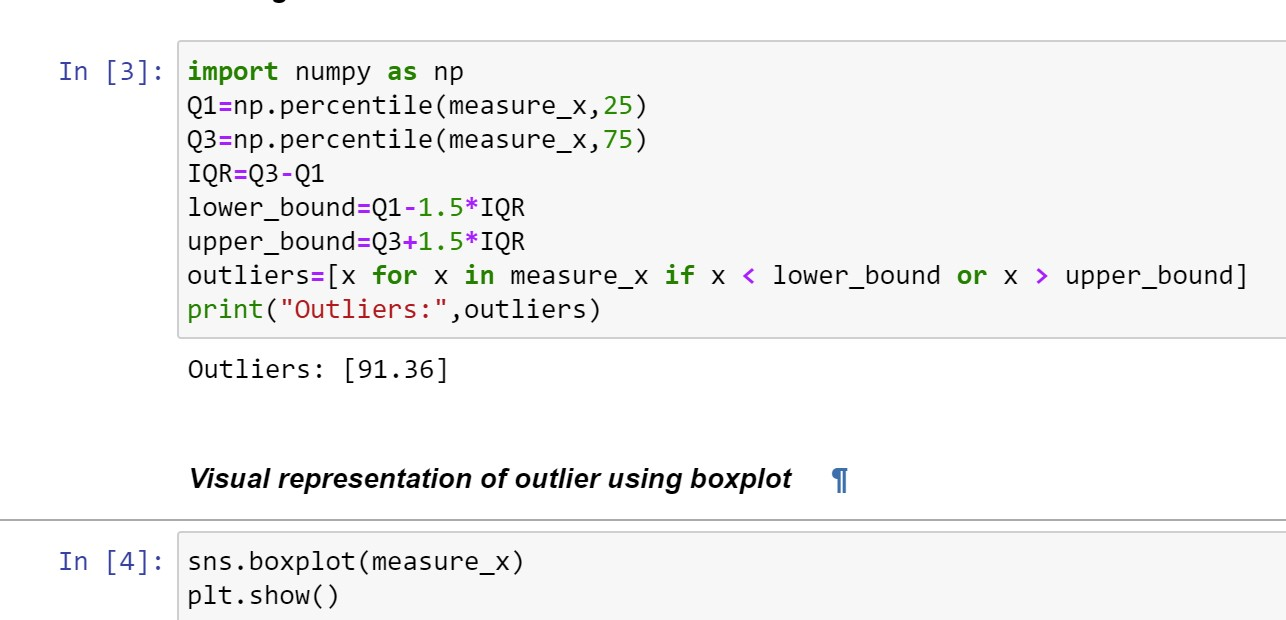
1. Look at the data given below. Plot the data, find the outliers and find out

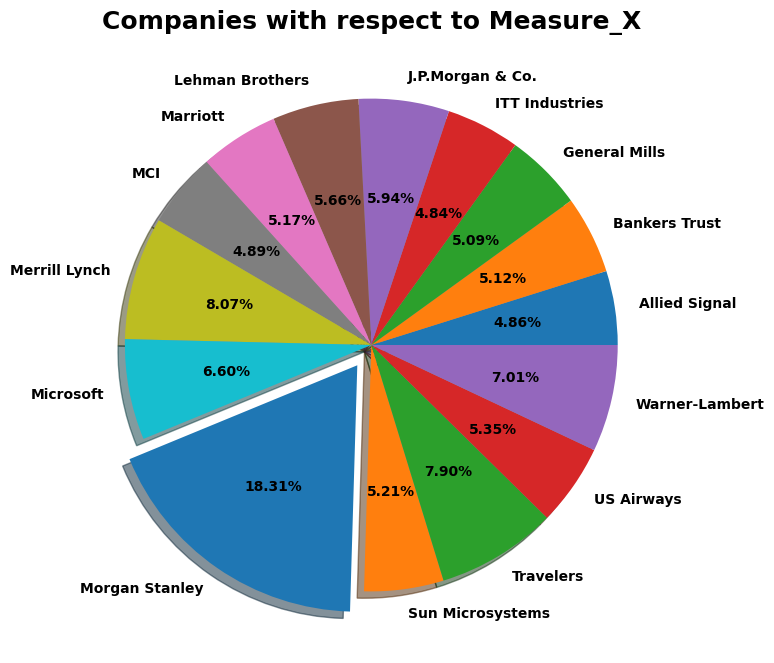
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

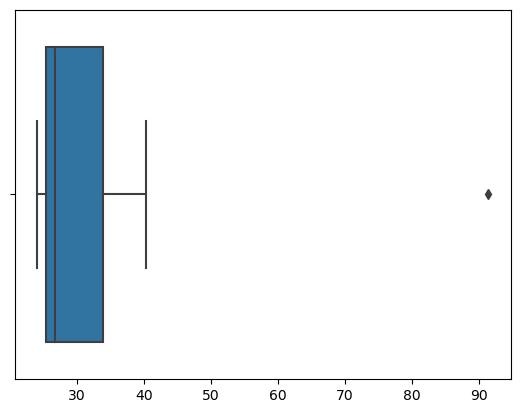
**Plotting pie plot**



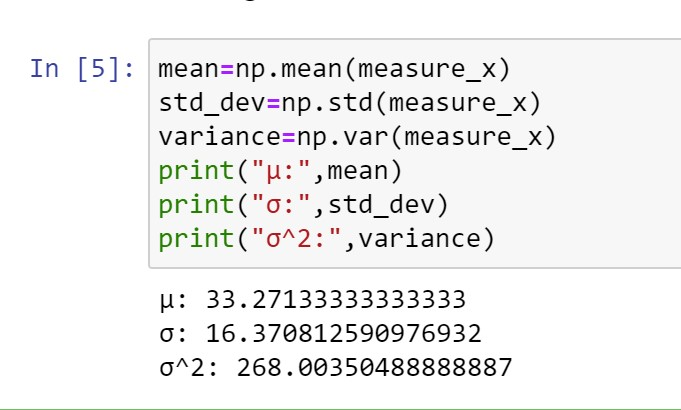
**Finding Outliers**







**Finding mean, standard deviation & variance**



**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.

***Answer)***The interquartile range (IQR) is a measure of the spread of the middle 50% of the data. To approximate the IQR from the box plot, we can subtract the lower quartile (Q1) from the upper quartile (Q3). Looking at the box plot, the approximate values for Q1 and Q3 are:

Q1 ≈ 5

Q3 ≈ 12

Therefore, the interquartile range can be calculated as:

IQR=Q3 - Q1 =12-5= 7

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

***Answer)*** Based on the box plot, we can observe the skewness of the dataset. Skewness refers to the asymmetry of the data distribution. In this case, we can see that the right whisker (upper quartile to the maximum) is much longer than the left whisker (minimum to the lower quartile). This indicates a positive skew, suggesting that the dataset is skewed towards higher values.

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?

***Answer)*** The new box plot would be affected as follows:

* The minimum value would remain the same.
* The lower whisker would be shortened, extending only up to the next smallest value.
* The lower quartile (Q1) would shift downwards.
* The median value would remain the same.
* The upper quartile (Q3) would remain the same.
* The upper whisker would remain the same.
* The maximum value would remain the same.

Overall, the new box plot would show a decrease in the lower whisker length, a lower Q1 value, but the overall shape and spread of the data would not change significantly.



**(b)**Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

***Answer)*** The mode of a dataset represents the most frequently occurring value or values. In the histogram, the mode would correspond to the bin or bins with the highest frequency, indicating the most common range of values in the dataset. By observing the histogram, we can identify the mode based on the tallest bar(s). Approximately between 4 to 8.

1. Comment on the skewness of the dataset.

***Answer)*** The histogram is skewed to the right (tail extending towards higher values), it indicates positive skewness.

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)** The histogram and box plot offer a comprehensive view of the dataset. The histogram provides a detailed distributional view, while the box plot summarizes important statistical measures. By analyzing both plots, we gain the information that they are right skewed and have 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.)

***Answer)*** P (call misdirecting), p = 1/200

P (call not misdirecting), q = 1 - p = 199/200.

P (at least one misdirected call in five attempts) = 1 – P (no misdirected calls in five attempts)

P (no misdirected calls in five attempts) = (q^5)

P (at least one misdirected call in five attempts) = 1 - (q^5) = 1 - (199/200)^5

≈ 0.02479

Therefore, the probability that at least one out of five attempted telephone calls reaches the wrong number is approximately 0.02479, or about 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?

***Answer)*** The most likely monetary outcome of the business venture can be determined by identifying the value with the highest probability. In this case, the value with the highest probability is 2000, with a probability of 0.3. Therefore, the most likely monetary outcome of the business venture is $2,000.

1. Is the venture likely to be successful? Explain

***Answer)*** Expected value = (-2000 \* 0.1) + (-1000 \* 0.1) + (0 \* 0.2) + (1000 \* 0.2) + (2000 \* 0.3) + (3000 \* 0.1)

= -200 - 100 + 0 + 200 + 600 + 300

= $800

If the expected value is positive, it indicates a potential profit and suggests the venture is likely to be successful. Since the expected value is positive ($800), it suggests that, on average, the venture is likely to be successful.

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

***Answer)*** Expected value =$800

Therefore, the long-term average earning of business ventures of this kind is $800.

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

***Answer)*** Variance = [(-2000 - 800)^2 \* 0.1] + [(-1000 - 800)^2 \* 0.1] + [(0 - 800)^2 \* 0.2] + [(1000 - 800)^2 \* 0.2] + [(2000 - 800)^2 \* 0.3] + [(3000 - 800)^2 \* 0.1]

= [1200000 \* 0.1] + [40000 \* 0.1] + [640000 \* 0.2] + [40000 \* 0.2] + [1440000 \* 0.3] + [400000 \* 0.1]

= 120000 + 4000 + 128000 + 8000 + 432000 + 40000

= 720000

Standard Deviation = √Variance ≈ √720000 ≈ 848.53

Therefore, the standard deviation is approximately $848.53, which serves as a good measure of the risk involved in a venture of this kind. It indicates the degree of variability or potential deviation from the expected value. The higher the standard deviation, the higher the risk associated with the venture.