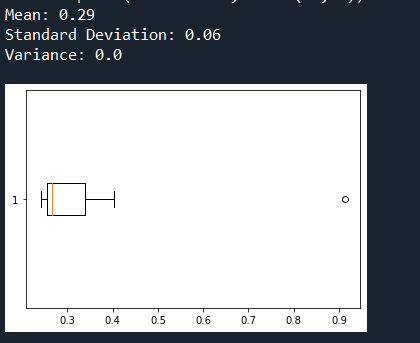
**Topics: Descriptive Statistics and Probability**

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% |

Results:



The outlier is 91.36% which belongs to Morgan Stanley.



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**

1. Inter-quartile range = Q3-Q1=12-5=7. Inter-quartile range represents the spread of a dataset, between the 25th percentile (Q1) and the 75th percentile (Q3), providing a measure of the middle 50% of the data's distribution.
2. It is positively skewed as the median is closer to Q1 than Q3
3. There would be no outliers in the box plots. The points greater than 2.5 move a point right.



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.

**Answer:**

1. The mode of the dataset would probably lie in the 3rd and 4th bins in the historgram.
2. It would be positively skewed as the mean or median is closer to the left tail.
3. We could validate outliers. The median or mean of the dataset could be found from Box plot, and the skewness could be validated from histogram.
4. 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:**

One in 200 long-distance telephone calls is misdirected, which implies that the probability of a call being misdirected = 1/200.

The probability that a call doesn’t get misdirected is = 1 – (1/200) = 199/200.

At least one in five attempted telephone calls implies one or all the 5 calls would be misdirected. Let P(X=n) be the probability of ‘n’ calls being misdirected. Then, the probability that at least one in five attempted telephone calls reaches the wrong number = 1- P(X=0)

Let, each event represent a single call not being misdirected:

Event A: The first telephone call is not misdirected => P(A)=199/200

Event B: The second telephone call is not misdirected=> P(B)=199/200

Event C: The third telephone call is not misdirected=> P(C)=199/200

Event D: The fourth telephone call is not misdirected=> P(D)=199/200

Event E: The fifth telephone call is not misdirected=> P(E)=199/200

Using the multiplicative rule(probability for occurance of all independent events), the probability that none of the calls get misdirected = P(X=0) = P(A and B and C and D and E) = P(A)\* P(B)\* P(C)\* P(D)\* P(E) = (199/200)^5

Therefore, the probability that at least one in five misdirected telephone calls = 1- (199/200)^5

= 0.02475/2.475%

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:**

1. The business venture's most likely monetary outcome is the highest probability value. In this case, the value 2000 has the highest probability of 0.3, making it the most likely outcome.
2. The probability of getting positive returns(1000,2000, and 3000) is 0.6 when compared to that of negative returns(-1000 and-2000) which is 0.2. If the success of the venture can be defined as making a profit, then the venture is more likely to be successful as there is a greater probability of getting positive returns than negative returns.
3. The long-term average earning of business ventures of this kind can be calculated by taking the weighted average of the returns, where each return is multiplied by its respective probability and then added.

EV = = (-2000 \* 0.1) + (-1000 \* 0.1) + (0 \* 0.2) + (1000 \* 0.2) + (2000 \* 0.3) + (3000 \* 0.1) = $800

1. To evaluate the risk involved in the venture, the standard deviation of returns is a good measure. It gives the dispersion of returns from the average. A higher standard deviation indicates a higher risk, as it means a wider range of possible outcomes from the mean. A higher standard deviation implies it is less likely that future outcomes will closely resemble average or expected outcome, the likeliness of occurrence of extreme value

The variance could also be used but standard deviation is better fit as it has the same units as the original data and it gives the average distance of the data points from the mean.

The following formula involving is used when you have a probability distribution with discrete outcomes:

Standard Deviation = √[Σ((X - μ)² \* P(X))]

= √[((-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)]

≈ √[1240000]

≈ 1113.5

So, the standard deviation is approximately 1113.5. This measure quantifies the risk involved in the venture based on the provided probabilities and monetary outcomes.