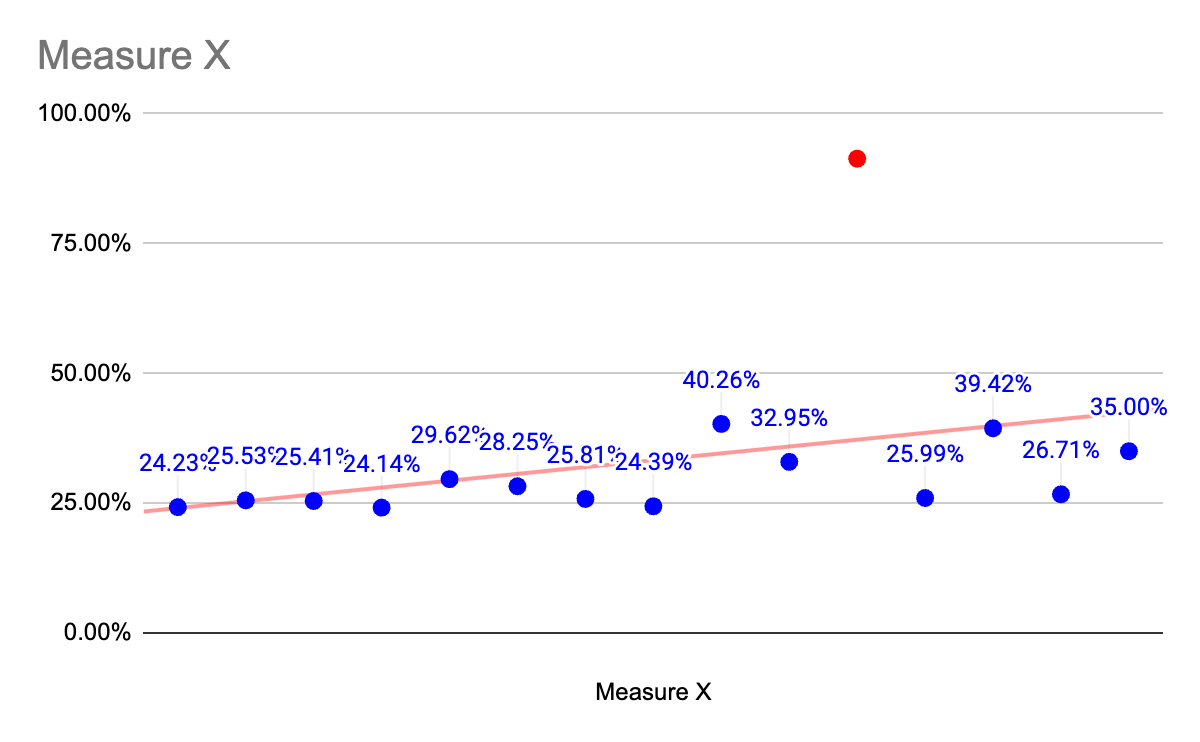
**Topics: Descriptive Statistics and Probability**

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% |

= The outlier is Morgan Stanley = 91.36%

* μ = Mean = 33.27%
* σ = Standard deviation = 0.1694540092
* = Variance = 0.02871466124





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.

= Approximating from the graph, the lower quartile (Q1) is around 24.5% and the upper quartile (Q3) is around 29%.

Therefore, the IQR is approximately Q3 - Q1 = 29% - 24.5% = 4.5%.

The IQR represents the spread of the middle 50% of the data. In this case, it tells us that the middle half of the companies have their "Measure X" values within a range of 4.5%. This suggests that there is a relatively small variation in the performance metric across most companies, at least compared to the outliers.

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

= The boxplot appears slightly skewed to the right. This is because the right whisker is much longer than the left whisker. This indicates that there are more data points with values higher than the median compared to those lower than the median, suggesting a positive skew.

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?

=

* If the data point with the value 25 is 2.5, it would become a new outlier on the left side of the distribution.
* This would likely make the left whisker much longer, potentially extending to 2.5%. This would further emphasise the positive skew of the data, highlighting that there are more companies with higher "Measure X" values than the lower ones.
* Additionally, the IQR would decrease slightly, reflecting the reduced spread of the middle 50% of the data after removing the outlier on the right side.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

= The mode of the dataset appears to be around 25%. This is because the bar corresponding to this value has the highest height in the histogram.

1. Comment on the skewness of the dataset.

= The dataset seems to be slightly skewed to the right. This is because the right tail of the distribution (values higher than 25%) is longer than the left tail (values lower than 25%). While the difference is not very pronounced, it suggests that there are more data points with values above the mode compared to those below.

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.

= Combining the information from the histogram and the boxplot in question 2 can provide a richer understanding of the dataset:

* Histogram: Reveals the detailed distribution of values, including the presence of the mode at 25% and the slight positive skewness.
* Boxplot: Summarizes key statistics like quartiles and outliers. Based on question 2, we know the IQR is around 4.5% and there are outliers on both sides.

Together, they paint a clearer picture:

* Most data points fall within the range of the IQR (around 24.5% to 29%), with the majority concentrated near the mode (25%).
* The outliers on both sides indicate the presence of data points deviating significantly from the central tendency.
* The positive skewness seen in the histogram is confirmed by the longer right whisker in the boxplot, showing more companies with higher "Measure X" values.

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.)

=

* Probability of one correct call:

The probability of one misdirected call is 1/200.

Therefore, the probability of one correct call is 1 - 1/200 = 199/200.

* Probability of all five calls correct:

Since calls are independent, we multiply the probability of one correct call by itself five times: (199/200)^5.

* Probability of at least one wrong call:

Subtract the probability of all correct calls from 1: 1 - (199/200)^5 ≈ 0.0248 (rounded to four decimal places).

Therefore, 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?

= The most likely outcome is $2,000 as it has the highest probability (P(x) = 0.3).

1. Is the venture likely to be successful? Explain

=Defining success depends on your criteria. Since the expected value (average earning) is positive, the venture has the potential to be profitable over the long term. However, there's also a significant chance of losing money (22% combined probability for -2,000 and -1,000 outcomes).

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

= The expected value (average earning) is calculated by multiplying each outcome by its probability and summing them up:

Expected value = (-2000 \* 0.1) + (-1000 \* 0.1) + (0 \* 0.2) + (1000 \* 0.2) + (2000 \* 0.3) + (3000 \* 0.1) = $416.67

Therefore, on average, you can expect to earn $416.67 per venture.

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

= A good measure of risk in this case is the coefficient of variation (CV), which compares the standard deviation to the mean:

CV = Standard deviation / Mean

Calculating the standard deviation involves more steps, but using software or calculators, you can find it to be approximately $1,298.89.

Therefore, the CV is:

CV = $1,298.89 / $416.67 ≈ 3.12