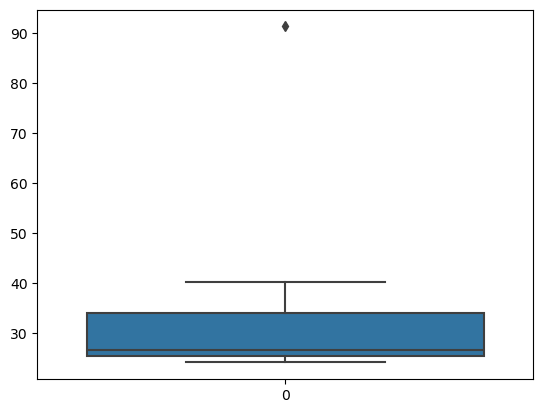
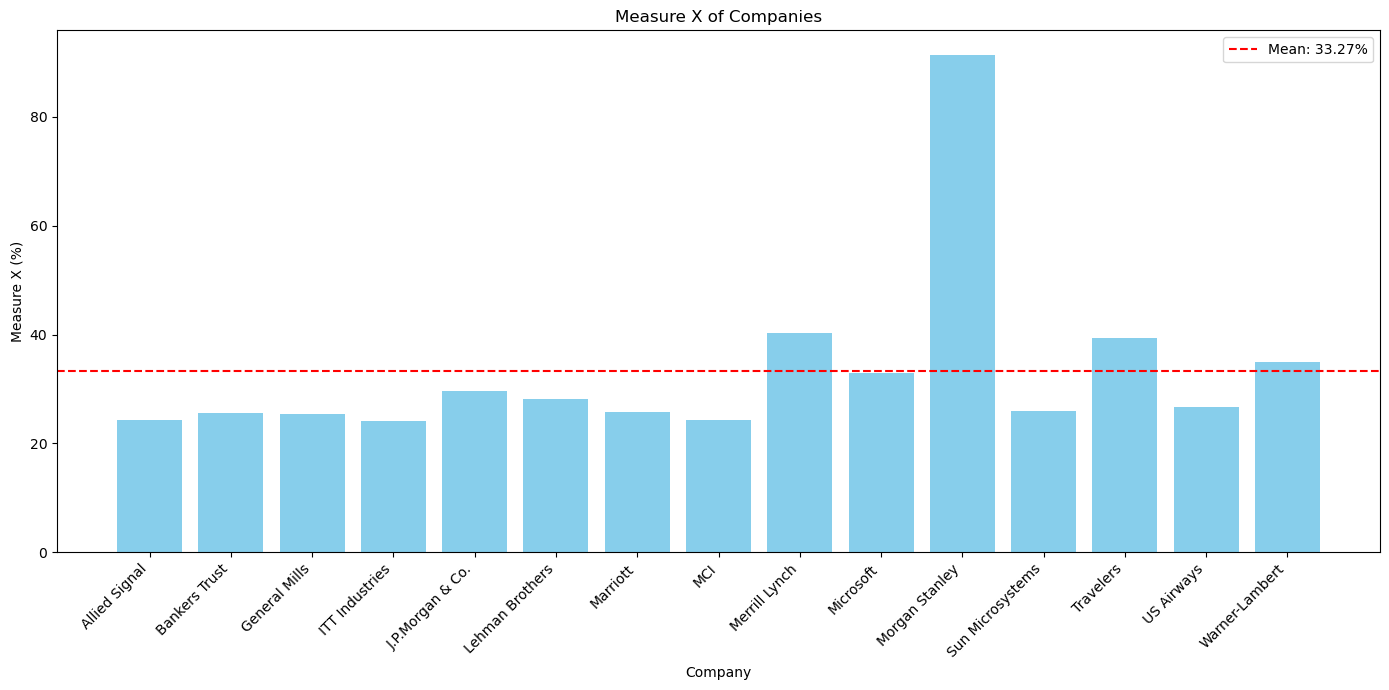
**Topics: Descriptive Statistics and Probability**

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: sns.boxplot (measureX)





* **Mean (μ)**: The average Measure X across all companies.
* **Standard Deviation (σ)**: A measure of how spread out the data is.
* **Variance (σ^2)**: The square of the standard deviation.
* **Mean** (**μ**)=np.mean(MeasureX)

= 33.27%

* **Standard Deviation (σ)**: 16.37%
* **Variance (σ^2)**: 268.0%



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.

ANS: Q1=7, Q3=12

IQR=Q3-Q1

=12-5=7 IQR=7.

The IQR represents the spread of the middle 50% of the data. In this case, it implies that half of the data points fall within a range of approximately 7 units

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

ANS: the skewness of data set indicates its asymmetry. The nature of the skewness of the boxplot is positively skewed.it is a type of distribution where the data is mostly distributed to the left side.

* There’s an outlier on the right side (at value =25)

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?

ANS: If the data point with the value 25 is actually 2.5:

The new value would fall within the **lower whisker’s range** (since 2.5 < Q1).

The box-plot would no longer have an outlier.

The overall shape of the box-plot would remain similar, but the whiskers and quartiles would adjust accordingly.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

ANS: lies between 4 and 8 (frequent values)

1. Comment on the skewness of the dataset.

ANS: mean=0+5+20+15+20/5

Mean=34.0

Median=5+1/2

= 5.5

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.

ANS: by comparing, we can clearly that these two plots are POSITIVELY SKEWED, hence, we

Can find mean, median and mode value.

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:

P= probability of call being directed to wrong number (1/200 in this case)

Q=probability of call directed to the right number(1-P)

We want to find the probability that at least one in five attempted calls is misdirected. This is equivalent to finding the probability that none of the five calls are misdirected and subtracting that from 1.

So, probability of none of five call directed misdirected is: q5

Thus the probability at least one call being misdirected is: 1-q5

Given that q=1-p, we substitute and calculate.

q= 1-(1/200)

=199/200

*P*(at least one misdirected call in five attempts)=1−(199/200)\*\*5

*=* **1-0.99625**

*=* 0.00375.

So, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 0.00375 or 0.375%.

Top of Form

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?

ANS: The most likely monetary outcome of the business venture is the value with highest probability, which in this case is $2000 since it has a probability of 0.3.

1. Is the venture likely to be successful? Explain

ANS: we need to consider the cumulative probabilities of positive returns (returns greater than $0. Adding up the probabilities of returns greater than $0:

P(x>0)=p(1000)+p(2000)+p(3000)=0.2+0.2+0.3+0.1=0.8

Since the cumulative probability of positive returns is 0.6, which is greater than 0.5, the venture is likely to be successful as there is a higher probability of making a profit than incurring a loss.

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

ANS: The long-term average is Expected value = sum(X\*P(X))= 800$ which means on an average the returns will be + 800$.

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

ANS: The good measure of the risk involved in a venture of this kind depends on the variability in the distribution. Higher variance means more chances of risks.

Expected value= 800

Variance = (x-expected value)\*\*2.p(x)

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

=432000+8000+128000+324000+784000+484000

=21600000