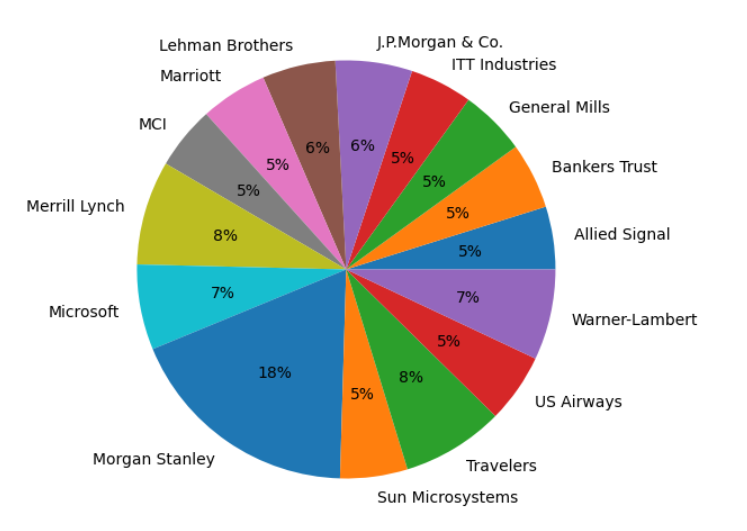
**Topics: Descriptive Statistics and Probability**

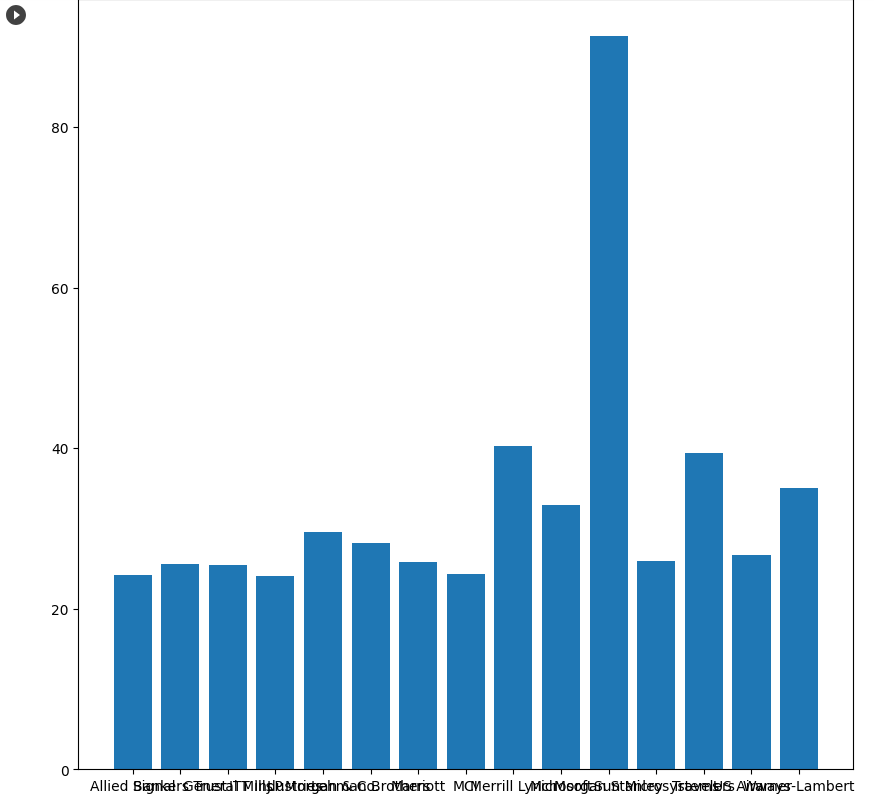
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |
|  |  |

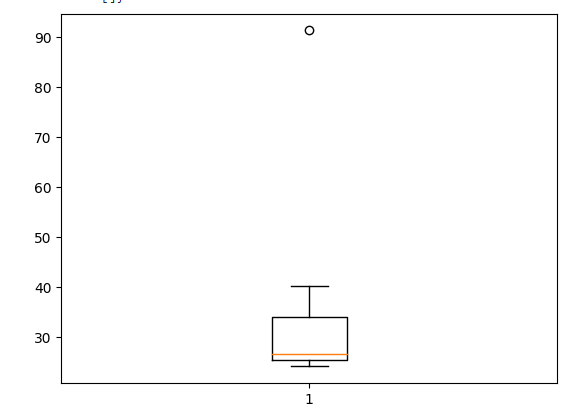
Pie chart



Bar plot



Box plot:



The outlier in the given data can be identified from the boxplot. It can be noticed that the outlier is **91.36%** of **Morgan Stanley**

**Mean: 33.27134**

**Standard Deviation: 16.9454**

**Variance: 287.1466**



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.

Solution: The Q1 value is 5 and Q3 value is approximately 12 from observation. So the interquartile range IQR is Q3-Q1 which is 7. The value implies the range of values that reside in the middle of the scores. It measures the spread or dispersion of middle half of the data

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

Solution: After observing the boxplot we can say that the data is not a normal distribution as the median is closer to the left side of the box. We can say that the data is right skewed as the median is closer towards the left

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?

Solution: If the value 25 is actually 2.5 then there would be no outliers in the boxplot. The data point would be present in the IQR. This would make the median move towards the right. The right skewedness would get reduced.

1. 

Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Solution: By looking at the histogram we can tell that the mode can lie somewhere between 5 and 8 as the frequency of the points is highest between 5 and 8

1. Comment on the skewness of the dataset.

Solution: The data is right skewed

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.

Solution: By comparing both the plots we can observe that both are right skewed and contain outlier at 25. We can also see that the middle range of the data is concentrated between 5 and 12 in histogram which is shown by the boxplot as the IQR

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.)

Solution:

Number of calls=200

Probability of call getting misdirected, p= 1/200

Probability of call not getting misdirected, q= 1- p = 199/200

We can apply binomial distribution:

P(x)=nCX \* px \* q1-x

Probability that 1 in 5 calls get misdirected is P(x)=5C1 \* (1/200)1\*(199/200)4

=5\*0.005\*0.9954

=5\* 0.005 \* 0.9801=0.0245

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?

Solution: The most likely monetary outcome is 2000 as it has highest probability which leads to high expectancy

1. Is the venture likely to be successful? Explain

Solution: The venture is likely to be successful because the probability of the venture obtaining profit is 0.8 or 80%. On the other hand, the probability of incurring a loss is only 0.2 or 20%. So we can say that the business venture is likely to be successful.

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

Solution: To calculate average earning on a long term we can find the sum of x\*P(x)

= -200-100+200+600+300=800

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

Solution: The good measure of the risk involved depends on the variance of the distribution.

Higher variance means more the chance of risk

Var(X)=E(X2)-E(X)2

2800000-8002 = 2160000