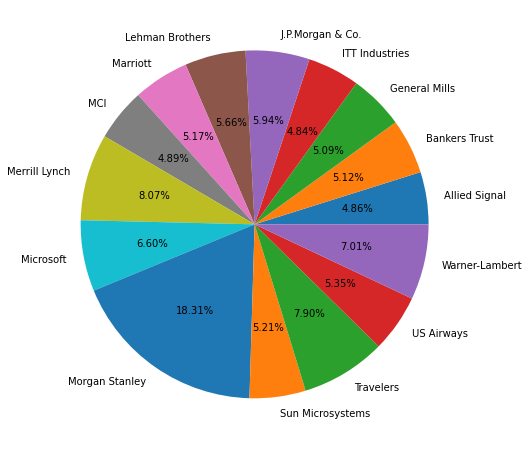
**Topics: Descriptive Statistics and Probability**

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: The results were drawn by using the python program.



By plotting a Box plot we found there is only one outlier.

|  |  |  |
| --- | --- | --- |
| **S.No** | **Parameter** | **Result** |
| 1. | Mean(µ) | 33.271 |
| 2. | Standard Deviation() | 16.945 |
| 3. | Variance () | 287.147 |

Python Code:

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

Data=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00])

Names=["Allied Signal","Bankers Trust","General Mills","ITT Industries","J.P.Morgan & Co.","Lehman Brothers","Marriott","MCI","Merrill Lynch","Microsoft","Morgan Stanley","Sun Microsystems","Travelers","US Airways","Warner-Lambert"]

%matplotlib inline

fig=plt.figure(figsize=(8,8))

plt.pie(Data,labels=Names,autopct='%1.2f%%')

plt.show()

sns.boxplot(Data,orient='h',color='yellow')

Mean = round(data.mean(),3)

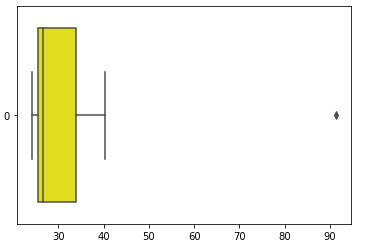
Std = round(data.std(),3)

Var = round(data.var(),3)

print("The Mean is: ",Mean)

print("The Standard Deviation is: ",Std)

print("The Variance is: ",Var)



Box Plot for Finding Outliers.



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?

Ans:

(i) From Above Box plot, We can say that 25 is the outlier.

Median = 7

1st quartile (Q1) = 5

3rd quartile (Q3) = 12

IQR(Q3=Q1) = (12-5) = 7

(ii) The dataset is Positively Skewed

1. If incase the Value is 2.5,
2. There would have been no outliers,
3. Small Effect on Mean and Median
4. Boxplot Moved towards Slightly 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.

Ans:

1. Mode means the most repeated Values. From above Histogram we can say that Mode of dataset lie in between 5 & 8
2. The Skewness of the Dataset is Positively Skewed.
3. 1. If we compare both of the plots, the data will be Positively Skewed.

2. These Plots will give is Mean and Mode.

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:

Probability of Getting Wrong Call =

So, the Probability of Getting Not a wrong call will be = 1- =

Number of phone calls attempted = 5

Therefore, probability that at least one in 5 attempted call reaches the wrong number is:

= 1 -

**= 0.025**

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

Ans:

From the Above Given We can write the probability Table as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **P(X)** | **F(X)= X . P(X)** | **F(X²) = X² . P(X)** |
| -2000 | 0.1 | -200 | 400000 |
| -1000 | 0.1 | -100 | 100000 |
| 0 | 0.2 | 0 | 0 |
| 1000 | 0.2 | 200 | 200000 |
| 2000 | 0.3 | 600 | 1200000 |
| 3000 | 0.1 | 300 | 900000 |

(i) The Maximum outcome for the business venture happens when probability is **0.3** ie., **2000**

(ii) The Business Venture will be successful when the X Value is more than 0 ie., (1000,2000 or 3000)

By summing the probabilities we have 0.2+0.3+0.1 = 0.6 > 0.5

So, Here for X values more Zero in given table we have Successful Venture.

(iii)  The long-term average earning of Business Ventures  =F(X)

F(X) = ∑ X.P(X)  = (**-200)+(-100)+0+(200)+(600)+(300) = 800**

(iv) Risk involved in a venture can be estimated by using Variance.

Var (X) = F(X²)  - { F(X) }²

=   2800000 -   800²

= **2160000**  ( Very High)

SD = =  **1469.6938**

As **Variability is Very high**  hence **Risk is also high.**