**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 : import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

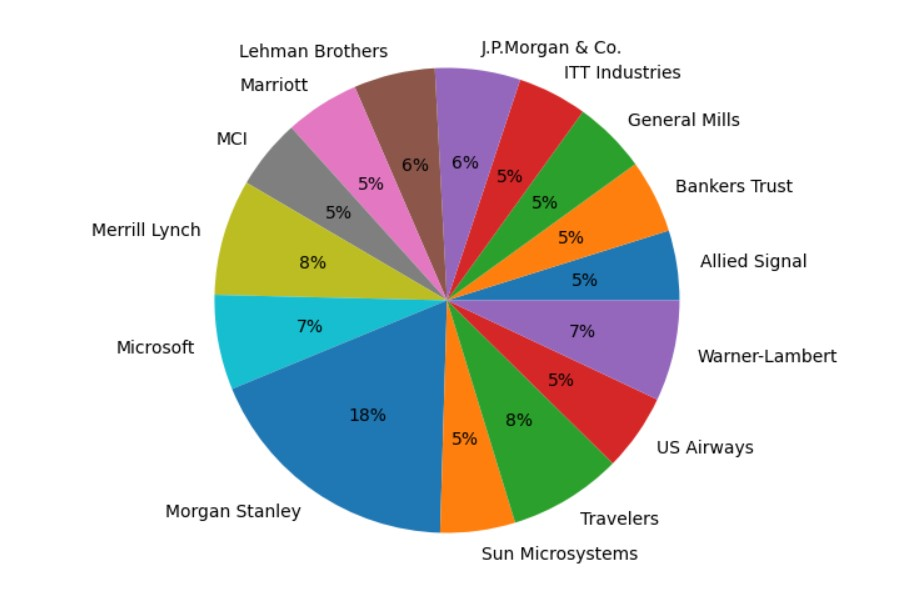
x=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00])

name=['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']

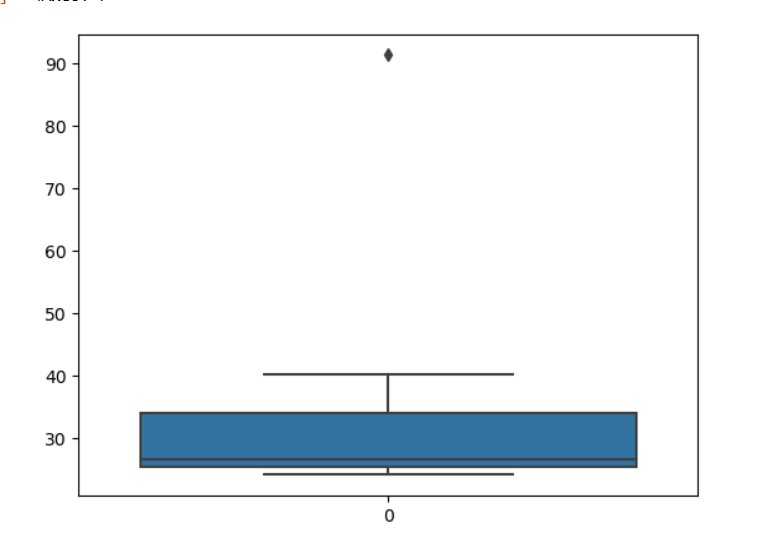
plt.figure(figsize=(6,8))

plt.pie(x,labels=name,autopct='%1.0f%%')

plt.show()



sns.boxplot(x)



x.mean() : 33.27133333333333

x.std() : 16.945400921222028

x.var() : 287.1466123809524



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 : IQR – 7

Lower Quartile (Q1) – 5

Median(Q2) – 7

Upper Quartile(Q3) – 12

Upper Extreme(Q4) – 19

Outlier - 25

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

Ans : Positive skewness

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 : Changing a data point from 25 to 2.5 in a box plot would significantly shift the median , quartiles , and potentially reclassify it as a non-outlier , affecting the box’s shape and whisker length.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans : 4.5 – 8

1. Comment on the skewness of the dataset.

Ans : Positive Skewness

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 : Comparing the two, it’s pretty obvious that the data is skewed to the right (positive

Skewness),and this can be useful for determining the 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 call getting misdirected = 1/200

Probability of call not getting misdirected = 1-(1/200)

Number of phone calls attempted = 5

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

= 1-(199/200)^2

=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?

Ans : 2000

1. Is the venture likely to be successful? Explain

Ans : Yes, because the total earnings of the venture is positive and highest probability of earning is 2000.

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

Ans:

|  |  |  |
| --- | --- | --- |
| x | P(x) | Income[x\*P(x)] |
| -2000 | 0.1 | -200 |
| -1000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
| Total |  | 800 |

The long term average earning for business ventures of this kind is $800. This means that , on average , each venture is expected to yield $800 in earnings when considering the given probability distribution and its associated probabilities.

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

Ans - Expected value(mean) of the returns : E(x) = ∑ [x\*P(x)] = 800

Calculate the variance :

Var(x) = ∑ [x – E(x)^2 \*P(x)

= 113800

Calculate Standard Deviation :

Root of var(x) = 1067.79

So, the standard deviation, which is appro. $1068 , is a good measure of the risk involved

In this business venture. It tells you how much the actual returns are likely to deviate from

The expected return of $800.