**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% |

µ=33.2713

σ=16.9454

σ2=287.1654

outliers=91.36

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',

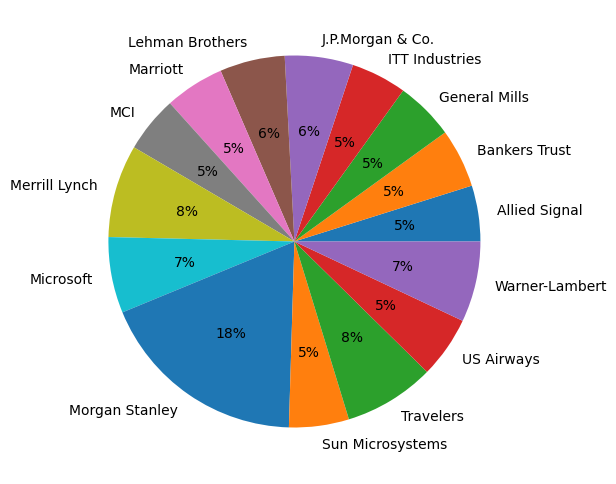
      'Warner-Lambert']

# Pie Plot

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

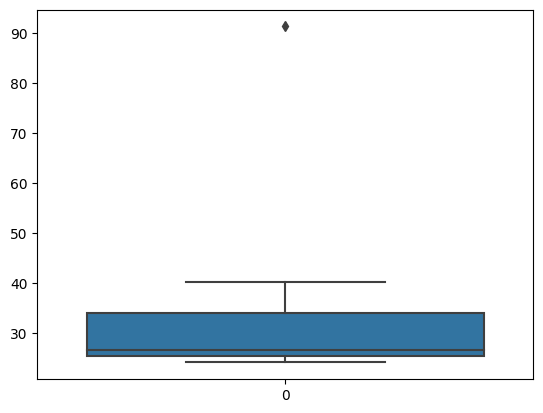
plt.pie(x,labels=name,autopct='%1.0f%%')

plt.show()



# Box Plot to find outliars

sns.boxplot(x)



# Mean

x.mean()

33.271333333333

# Vairance

x.var()

287.1466123809524

# Standard Deviation

x.std()

16.9454009212220

2.



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.

IQR=Q3-Q1=12-7=5 approximately.

It means 50% of data types lie in the range of 5 and 10

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

The dataset is positively skewed. Tail is found extending towards right side of curve.

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?

The median value will remain same but the interquartile range will change moreover there will not have between lie in the same

3.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The mode of dataset lie in left side.

1. Comment on the skewness of the dataset.

The dataset is positively skewed, and towards the left side.

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.

The dataset In question 2 and 3 are similar, both dataset is tail towards right and left side. The median value will remain same but the interquartile range will change moreover there will not have between lie in the same

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.)

In 200 long-distance telephone cells is misdirected

Probability of cells misdirecting P =1/200

Probability of cells not misdirecting =1/200= 199/200

Number of cells =5

P(x)= nC.p^x.q^n-x

N=5

P=1/200

Q=199/200

=1-none of the cell reaches the wrong number

1. P(0)
2. 5C(1/200)0(199/200)^5-0
3. (199/200)

=0.02475

Probability at the least one in five attempted telephone cell reaches the wrong number =0.02475

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 |
|  |  |

E(X) = X\*p(x)

E(X2) = X2\*P(x)

Var (X) = E(X2)-{E(x)}^2

SD = √var

|  |  |  |  |
| --- | --- | --- | --- |
| x | P(x) | E(X) = X\*p(x) | E(X^2) = X^2\*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 |
|  |  | 800 | 2800000 |

1. What is the most likely monetary outcome of the business venture?

$2000 as it has maximum probability 0.3

1. Is the venture likely to be successful? Explain

Hence if x is 1000, 2000, 3000

Probability is 0.2+0.3+0.1=0.6

As 0.6>0.5 hence venture likely to be successful

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

E(X) = X\*p(x) = $800

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

Var (X) = E(X2)-{E(x)}2

2800000- 8002= 2160000

SD = √var

√(2160000)= 1870$