**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

df = pd.DataFrame(pd.read\_excel("C:/Users/prate/OneDrive/Desktop/SET1Q1.xlsx"))

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])

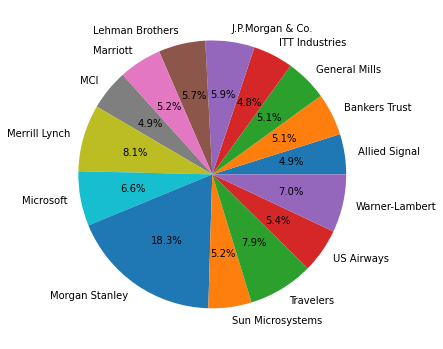
name=['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers',

'Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways', 'Warner-Lambert']

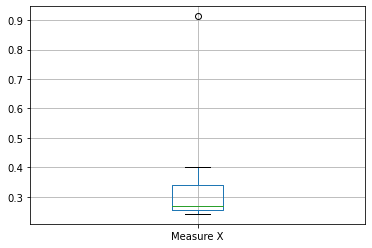
plt.figure(figsize=(6,8))

plt.pie(x,labels=name,autopct='%1.1f%%')

plt.show()



df.boxplot()



#Outlier at .9 (Morgan Stanley)

df.mean()

Measure X 0.332713

dtype: float64

df.std()

Measure X 0.169454

dtype: float64

df.var()

Measure X 0.028715

dtype: float64



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=12-5=7, spread of data

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

Ans. Positive skewed.

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. Upper and lower quartile values will change , median will change and boxplot shift towards positive skewness.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans. Around 5 as highest peak represents the mode in histogram.

1. Comment on the skewness of the dataset.

Ans. 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.

Ans. These are very similar in that they both help to visualize and describe numeric data. Both histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.

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 one misdirected call= 1/200 (Given)(p)

P (Right call) not misdirected calls= Now 1 -1/200=199/200 (q)

5 calls can be made by ⁿCₓpˣqⁿ⁻ˣ ways

Here req. prob. Of at least one misdirected call( n=5 ,x=0)

1 – (199/200)^5=0.024

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 as it has the maximum probability of 0.3

1. Is the venture likely to be successful? Explain

Ans. Yes because it has positive expected value of 800.

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

Ans. Long term average earning is based on expected value = ∑xp(x). here

-200+(-100)+0+200+600+300=+ 800

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

Ans. Positive expected value means low or no risk in this venture.