**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% |

**Answers)**

* Outlier : 91.36
* Mean ( : 33.27133
* Standard dev ) : 16.370
* Variance ( : 268.0035

**Code:**

**#storing x values in an array**

import numpy as np

x=np.array([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']

print(x)

print(name)

import matplotlib.pyplot as plt

plt.figure(figsize=(6,8)) #to set size of the pie fig

plt.pie(x,labels=name,autopct='%1.0f%%') #autopct to display the numerical vaalues inside the pie

plt.show()

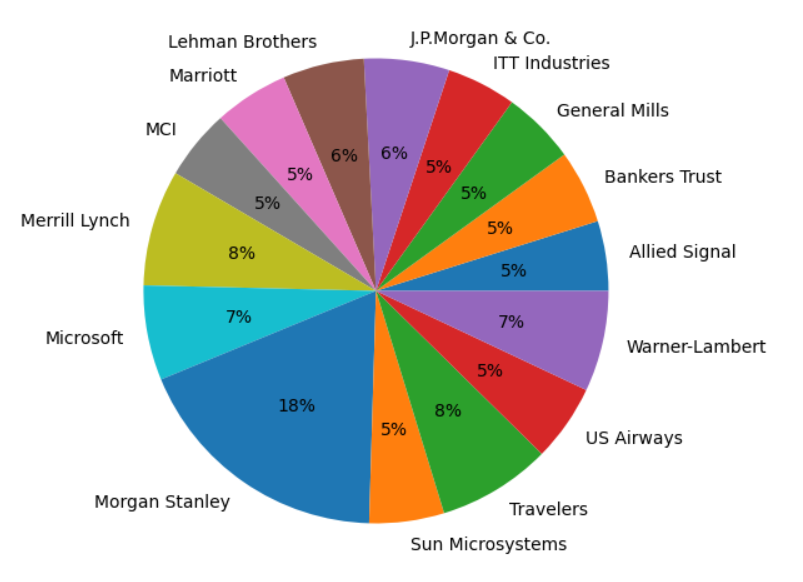
import seaborn as sns

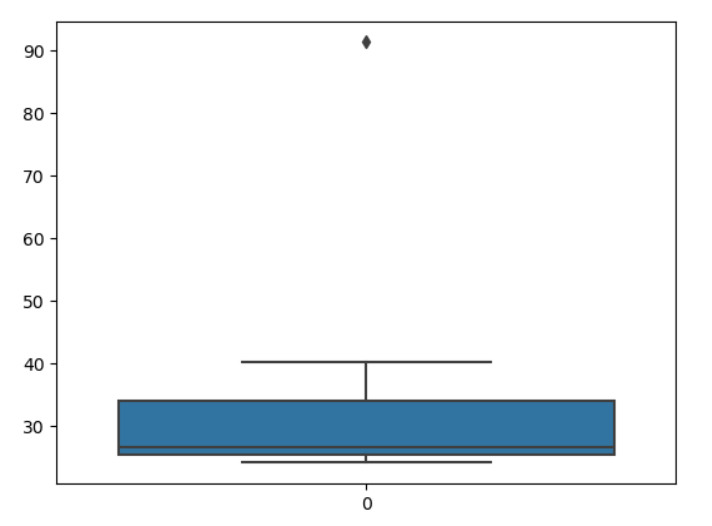
sns.boxplot(x)

print("Mean of x value",x.mean()) #finding mean

print("Variance of x value",x.var()) #finding Variance

print("Standard Dev of x value",x.std()) #standard deviation





Q2)



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)** **Inter-quartile range (IQR) lie between Q3-Q1  = 12-5 = 7. Half of the data points lie in between 5 and 12**

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

**Ans) Positive Skew**

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)** **2.5 is not considered, it ranges from 0 to 20**

Q3)



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans)** **between 4 to 8**

1. Comment on the skewness of the dataset.

**Ans)** **It’s a Right Skew**

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) Median in boxplot and Mode in histogram**

**Histogram provides the frequency distribution so we can see how many times each data point is occurring however boxplot provides the quantile distribution i.e. 50% data lies Between 5 and 12.**

**Boxplot provides whisker length to identify outliers, no information from histogram We can only guess looking at the gap that 25 may be an outlier.**

Q4) 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.)

**Answer)** P = 0.028 (Using MonteCarlo Simulation for 1000 occurances)

**Code:**

d = rep(c("MISS","NO\_MISS"),times=c(1,199)) **#Creating a sample dataset of 200 outcomes, with only 1 event of Mis-routing .**

#event = sample(d,5,replace=TRUE) **#Sample of 5 call scenarios**

event\_1000\_rep = replicate(1000,sample(d,5,replace=TRUE)) **#Repeating experiment 1000**

**times**

i = 1

z=0

while(i<=1000)

{

if("MISS" %in% event\_1000\_rep[,i]) **#Checking if we have even one MISS among the 5 calls**

z=z+1

i=i+1

}

p = z/1000

cat("Probability",p)

Q5) 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)**  **Max. P = 0.3 for P(2000). So most likely outcome is 2000**

1. Is the venture likely to be successful? Explain

**Ans)** **P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater than expected returns. P(Incurring losses) is only 0.2. So the venture is likely to be successful.**

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

**Ans)** **Weighted average = x\*P(x) = 800. This means the average expected earnings over a long period of time would be 800(including all losses and gains over the period of time)**

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

**Ans)** **P(loss) = P(x= -2000)+P(x=-1000)=0.2. So the risk associated with this venture is** **20%**