**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 Micro-systems** | **25.99%** |
| **Travelers** | **39.42%** |
| **US Airways** | **26.71%** |
| **Warner-Lambert** | **35.00%** |

**import pandas as pd**

**import matplotlib.pyplot as plt**

**df = 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])**

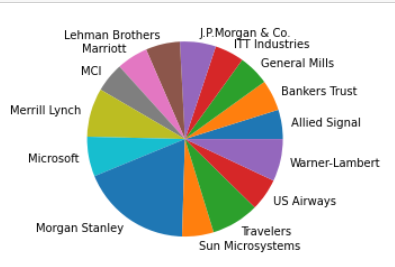
**company = ['Allied Signal','Bankers Trust','General Mills',**

**'ITT Industries','J.P.Morgan & Co.','Lehman Brothers',**

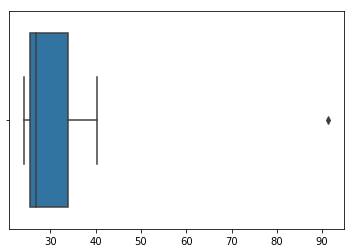
**'Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways','Warner-Lambert']**

**plt.pie(df,labels = company)**

**plt.show()**

****

**plt.boxplot(df)**

****

**There is one outlier at 91% for morgan stanley.**

**df.mean()#mean is 33.27133333333333**

**df.std()#standard deviation is 16.945400921222028**

**df.var()#variance is 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.**
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?**
4. **Inter quartile range(IQR) = 12 - 5 = 7.**

**7 is range of middle half of the data**

1. **The dataset is positively skewed I.e median is closer to the lower quartile.**
2. **As data point at 25 is an outlier,if it was actually 2.5 there would be no outliers in the dataset.**

**Right skewness will also be reduced.**

****

**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.**
4. **The mode of this dataset lies between 4 to 8.**
5. **The dataset is positively skewed I.e most data lies to the right.**
6. **We can clearly see there is an outlier at 25 by looking at both histogram and boxplot(dot indication by boxplot and it is evident looking at the alone data point at 25 which is an outlier in histogram).Also,both plots show that the data positively skewed(for histogram we can see that most of the data is towards the right while for boxplot the median is closer to the lower quartile).**
7. **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.)**

**One in 200 calls is misdirected**

**let us consider the probability of 1 call misdirected out of 200 as event A.**

**Probability of occurring of event A= 1/200**

**P(A)= 1/200**

**Probability of having at least one successful call will be**

**1-P(A)= 1-1/200= 199/200= 0.967**

**As every event is independent of other event the probability will be**

**1- (0.967)^5**

**0.02475 = 2% chance.**

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**
5. **most likely monetary outcome is 2000 as it has the highest probability**
6. **p(x>0) = 0.6 which implies that there is 60% chance that venture will be successful so it is likely that it will be successful.**
7. **Expected value of -2000 = -200**

**Expected value of -1000 = -100**

**Expected value of 0 = 0**

**Expected value of 1000 = 200**

**Expected value of 2000 = 600**

**Expected value of 3000 = 300**

**Long term average earning is +800$ over the future including all profits and losses.**

1. **d = {'x':[-2000,-1000,0,1000,2000,3000],'P(x)':[0.1,0.1,0.2,0.2,0.3,0.1]}**

**df = pd.DataFrame(data = d)**

**df.std()**

**x 1870.828693**

**Highly risky as the standard deviation is highly deviating from the average**

**Earning.**