**Topics: Descriptive Statistics and Probability**

**Q.1** Look at the data given below. Plot the data, find the outliers and find out 흁, 흈, 흈**2**

|  |  |
| --- | --- |
| **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% |

**Solution** :-

import pandas as pd

import seaborn as sns

from matplotlib import pyplot as plt

measure\_X=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.9

9,39.42,26.71,35.00])

name\_of\_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.figure(figsize=(8,8))

plt.pie(measure\_X,labels=name\_of\_company,autopct='%1.2f%%')

plt.title('Pie Chart')

plt.show()



**mean, standarad deviation and variance of data**

print("mean of the data",measure\_X.mean())

print("SD of the data",measure\_X.std())

print("variance of the data",measure\_X.var())

**mean of the data: 33.26533333333333**

**SD of the data: 16.948178944175964**

**variance of the data: 287.2407695238095**

**outlier detection**:-

sns.boxplot(measure\_X)

plt.title("Box Plot")



**there is one outlier present in data.**

**Q.2**



Answer the following three questions based on the box-plot above.

* What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

**Solution** :-

First quartile (Q1) = 5

Third quartile (Q3) = 12

IQR = Q3-Q1 = (12-5) = 7

**IQR tells the range of the middle half of the data, that means 50% of data remains in this range.**

* What can we say about the skewness of this dataset?

**Solution** :- positevely skewed

* If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

**Solution** :- In that case there will be no outliers, and it might have affected in the values of mean and median. The boxplot might have moved towards right.

**Q.3**



Answer the following three questions based on the histogram above.

1.Where would the mode of this dataset lie?

**solution**:-

**mode lies between 4 and 8.**

**2.**Comment on the skewness of the dataset.

**solution:- +ve skewed**

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.

**solution**:-

**By comparing both of them it is clear that the data would be positively skewed. Also, it**

**help us to finding mean, mode value.**

**Q.4.**

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.)

**Solution** :-

Probability of call getting misdirected = 1/200

Hence ,

probability of call not getting misdirected = 1 - 1/200=199/200

= 0.995

Number of telephone calls attempted = 5

Therefore,

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

= 1 - (0.995)^5

= 0.0247

**Q.5**

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?

**Solution** :-

**The most likely monetary outcome of the business venture is**

**2000, becoz it has the highest probability**

2. Is the venture likely to be successful? Explain

**Solution** :-

**Yes, becoz the total outcome is positive.**

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

**Solution** :-

**avg earning= summation of x\* p(x)**

**= 800**

**The long-term average earning of business ventures is = 800**

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

**Solution** :-

|  |  |  |
| --- | --- | --- |
| x | P(x) | x\*P(x) |
| -2000 | 0.1 | -200 |
| -1000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 30000 | 0.1 | 300 |
|  | **Total** | **800** |

**Variance = 86666.67**

**SD = 294.3920**