**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 -

**Plot the data:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

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

plt.show()

A pie chart with numbers and text

Description automatically generated

**Find the outlier:**

sns.boxplot(data)

plt.show()

A graph with a bar and a number of objects

Description automatically generated with medium confidence

**Find mean:**

Data. Mean ()

33.27133333333333.

**Find variation:**

Data. Var ()

287.1466123809524.

**Find std:**

Data.std()

16.945400921222028.



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?

Ans;- (i) IQR = Q3 - Q1 = 12- 5 = 7.5 approximately 7

It means 50% of data points lies in the range of 5 and 12

(ii) The right skewed median is toward the left side it is not normal distribution

(iii) In that case there would be no Outliers on the given dataset because of the outlier the data had positive skewness it will reduce and the data will normal distributed.



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.

Ans - (i) The mode of the data set lie between 5 to 10 and approximately between 4 to 8

(ii) Skewed right. Because mean>median>mode.

(iii)By using them together, you can get a more complete picture of thedata and identify important patterns and trends. They both are right skewed and both have outliners the median can be easily visualized in the box plot where as in histogram mode is more visible

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 call getting misdirected = (1/200)

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

Number of phone calls attempted = 5

Therefore, probability that at least one in 5 attempted call reaches the wrong number is: =1-(199/200) ^5

= 0.025.

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

Ans –

(i) The most likely monetary outcome of the business venture is 2000. This is because the probability of making a profit of 2000 is 0.3, which is higher than the probability of any other outcome.

(ii) Yes, the venture is likely to be 80% successful. This is because the probability of making a profit is 0.8.

(iii) The long-term average earnings of this business is around 800 [sum (X\*P(X))].

Which means on an average the returns will be +800

(iv) Var 86666

Std 294.3