**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: -**

A graph with a blue rectangle

Description automatically generated

Mean: 33.27133333333333

Standard Deviation: 16.945400921222028

Variance: 287.1466123809524

Code for the above boxplot…………

import pandas as pd

import matplotlib.pyplot as plt

# Data

data = {

    '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'],

    'Measure X': [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]

}

# Create DataFrame

df = pd.DataFrame(data)

# Plot boxplot

plt.figure(figsize=(10, 6))

plt.boxplot(df['Measure X'], vert=False)

plt.xlabel('Measure X')

plt.title('Boxplot of Measure X')

plt.show()

# Find outliers

q1 = df['Measure X'].quantile(0.25)

q3 = df['Measure X'].quantile(0.75)

iqr = q3 - q1

lower\_bound = q1 - 1.5 \* iqr

upper\_bound = q3 + 1.5 \* iqr

outliers = df[(df['Measure X'] < lower\_bound) | (df['Measure X'] > upper\_bound)]

print("Outliers:")

print(outliers)

# Calculate mean, standard deviation, and variance

mean = df['Measure X'].mean()

std\_dev = df['Measure X'].std()

variance = df['Measure X'].var()

print("\nMean:", mean)

print("Standard Deviation:", std\_dev)

print("Variance:", variance)



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=Q3-Q1=7

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

**ANS: -** skewness=positive

1. If it was found that the data point with the value 25 is actually 2.5, how would the new boxplot be affected?

**ANS: -** No outliers if the value of the 25 is 2.5 then mean and median should calculate to see is there is any shift in data



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: -**

1)The Mode lie between: 3 and 10

2)Skewness: it is noticed as right skewed data(positive)

3)Comparring both the graphs the values of fro 5-7 gets most occuring values of the

daataset,one outlier of 25,both plots have positive skewness

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: -**

X = probability of 1 call misdirected out of 200

Probability of occurring of X = 1/200

P(X)= 1/200

Probability of having at least one successful call will be

1-P(X)= 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?

**ANS: - $**2000 as it has the highest probability of occurrence

1. Is the venture likely to be successful? Explain

**ANS:-** if Success == positive returns as a measure

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

**ANS:-** (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000 \*0.3)+(3000\*0.1)=800

the long-term average earning for these type of ventures would be around $800

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

**ANS:-** A good measure to evaluate the risk would be variance and standard deviation of the variable x

Var = 3500000 Sd = 1870.83

The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky