**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% |
| Sun Microsystems | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

**ANS:**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

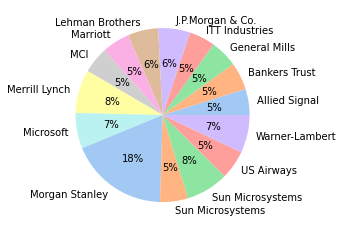
df=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00])

name=('Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers','Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Sun Microsystems','US Airways','Warner-Lambert')

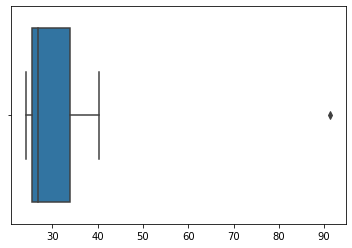
colors = sns.color\_palette('pastel')

plt.pie(df, labels = name, colors = colors, autopct='%.0f%%')

plt.show()



sns.boxplot(df)



np.mean(df)

33.27133333333333

np.std(df)

16.370812590976932

np.var(df)

268.00350488888887



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:**

1. IQR=Q3-Q1 = 12 – 5 = 7. This much data resides in the middle of the scores.
2. The data is Right Skewed.
3. The new box-plot will have no outliers.



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 lies in 4-6 and 6-8 bins.
2. The data is Right Skewed.
3. Both the graph has outlier, both graphs are right skewed.
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.)

**ANS:**

P(x)=1/200

Probability that at least one in five attempted telephone calls reaches the wrong number = 1-(199/200)+(199/200)+(199/200)+(199/200)+(199/200)

= 1-(0.995+0.995+0.995+0.995+0.995)

= 0.0247

Probability that at least one in five attempted telephone calls reaches the wrong number

Is 0.0247.

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:**

1. The most likely monetary outcome of the business venture is 2000 as it has the highest probability of 0.3
2. Yes the chances of venture to be successful is more because returns has more values in positive. i.e. (0.2+0.3+0.1) = 0.6
3. (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1) = 800.
4. The good measure is that the returns are more in profit as compared to loss.