**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% |

Ans1:

#mean=33.27133333333333

#std. deviation=16.945400921222028

#variance=287.1466123809524

#from Pie chart we can see that Morgan Stanley is the outlier.

##Code##

import pandas as pd

df=pd.read\_excel('Q1.xlsx')

df

df['Measure'].mean()#33.27133333333333

df['Measure'].std()#16.945400921222028

df['Measure'].var()#287.1466123809524

#Histogram

df['Measure'].hist()

df.dtypes

#scatterplot

df.plot.scatter(x='Measure',y='Name of company')

#Boxplot

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

#piechart

import matplotlib.pyplot as plt

plt.pie(x=df['Measure'],labels=df['Name of company'])

plt.show()

#from Pie chart we can see that Morgan Stanley is the outlier



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.

Ans2(i): IQR=7, it implies that 99.7% of data lies in this range

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

Ans2(ii):data seems to be positively skewed

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

Ans2(iii):2.5 will be in the boxplot and it might become normally distributed



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans3(i): mode=6

1. Comment on the skewness of the dataset.

Ans3(ii): positively skewed

1. 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.

Ans3(iii): we can see that histogram is showing positive(right) skewness, boxplot shows the 25 data point as the outlier and mode<

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.)

Ans4: p=1/200, q=1-1/200=199/200, n=5, r=1

from scipy.stats import binom

bi=binom (n=5, p=1/200) #n, p

#p(r=1)

z=bi.pmf (1)

z

z.round(4)

# probability that at least one in five attempted telephone calls reaches the wrong number is 0.0245

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?

Ans5i: most likely monetary outcome of the business venture would be 2000$ because it has the highest probability.

1. Is the venture likely to be successful? Explain

Ans5ii: From the above table total probability of positive returns is 0.8 i.e., their 80% chance that the venture could be successful.

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

Ans(iii): summation of [x\*p(x)] =800, this 800 shows the net profit of the venture, therefore this venture seems to be profitable in the long term.

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

Ans(iv): From the above table total loss probability is 0.2 which shows that risk is 20%